Investigating adverse effects of psychiatric drugs through data-mining of electronic health records

Iqbal, Ehtesham

Awarding institution:
King's College London

Investigating adverse effects of psychiatric drugs through data-mining of electronic health records

Ehtesham Iqbal

Student Number – 1336802

Department of Biostatistics and Health Informatics,

Institute of Psychiatry, Psychology and Neuroscience

King's College London

This thesis is submitted for the degree of

Doctor of Philosophy

In loving memory

of

my father

'Muhammad Iqbal Javaid'

Abstract

The use of Electronic Health Records (EHRs) in recording the details of patient interactions with healthcare services has generated large amounts of data with great potential for secondary usage in research. However, although the vast information available offers opportunities to improve care by learning from similar patients in parallel situations, there are great challenges in extracting correct and contextually meaningful knowledge due to the free-text, unstandardised and uncertainty-ridden form of clinical text.

The focus of the presented work has been on detecting concepts related to Adverse Drug Events (ADEs) from the EHR using Natural Language Processing (NLP) tools to transform the unstructured text into semantically meaningful annotated knowledge. Specifically, this thesis explored the potential of NLP to identify ADEs from mental health EHRs in order to understand how drugs are working in real-world settings, to complement the current body of knowledge from clinical trials. Four studies were performed on the EHRs of the South London and Maudsley (SLAM) NHS Foundation Trust, with some analyses further performed on two other large psychiatric NHS Trusts: Camden & Islington (C&I) NHS Foundation Trust and the Oxford Health (Oxford) NHS Foundation Trust.

The first study presented means to identify ADEs within an EHR, with a use case in identifying patients who have experienced Extra-Pyramidal Side Effects (EPSEs) at any point and achieved an overall 0.85 precision and 0.86 recall. The second study focused on anchoring ADEs to a point in time and achieved 0.89 precision and 0.86 recall in SLAM and 0.84 precision and 0.87 recall in C&I, contributing to the third study, which built a complete view of the patient medication and Adverse Drugs Reaction (ADR) profile. These methods were applied to study the side effect profile of Clozapine, a potent antipsychotic, in the three large mental health hospitals.

Acknowledgements

I want to express immense gratitude to my supervisors Professor Richard Dobson, Dr Zina Ibrahim and Dr Caroline Johnston, for all the expertise, guidance and support given throughout my PhD. I am grateful to the managers of SGDP centre, Melissa Sollich, Lilly Cleghorn, Chris Parsons, Cathryn Lewis, Francesca Happe and Zarsheesh Divecha for encouraging my early interest in research and the continued support from the SGDP Centre over the years. My thanks also to the SLAM CRIS team and its collaborators for all the technical help. The research presented in this thesis would not have been possible without their help.

I would like to thank my family for their support. To my parents for showing their wisdom and my brothers Usman Iqbal and Ali Iqbal and sister Erum for looking out for me. A special thanks to my wife, Saima, for her love and encouragement.

Declaration

I confirm that this thesis and the research presented therein is my original work. This includes the design and collection of the three health records, planning, pre-processing, statistical analysis and writing. All help received and all published sources consulted are acknowledged appropriately. No part of this thesis has previously been submitted towards any other degree or qualification. Those parts that have been published or submitted for publication are noted as such in the text. I wrote the thesis under the supervision of Professor Richard Dobson and Dr Zina Ibrahim.

Publications

Published articles

Kadra, G., Spiros, A., Shetty, H., **Iqbal, E**., Hayes, R. D., Stewart, R., & Geerts, H. (2018). Predicting parkinsonism side-effects of antipsychotic polypharmacy prescribed in secondary mental healthcare. *Journal of Psychopharmacology, 32(11), 1191-1196.*

Bean, D. M., Wu, H., **Iqbal, E**., Dzahini, O., Ibrahim, Z., M., Broadbent, M., Stewart, R., & Dobson, R. J. (2018). Knowledge graph prediction of unknown adverse drug reactions and validation in electronic health records. *Scientific reports*, *8*(1), 4284.

Iqbal, E., Mallah, R., Rhodes, D., Wu, H., Romero, A., Chang, N., Dzahini, O., Chandra, P., Broadbent, M., Stewart, R., & Dobson, R. J. (2017). ADEPt, a semantically-enriched pipeline for extracting adverse drug events from free-text electronic health records. *PloS one*, *12*(11), e0187121.

Baker, E., **Iqbal, E.,** Johnston, C., Broadbent, M., Shetty, H., Stewart, R., Howard, R., Newhouse, S., Khondoker, M., & Dobson, R. J. (2017). Trajectories of dementia-related cognitive decline in a large mental health records derived patient cohort. *PloS one*, *12*(6), e0178562.

Iqbal, E., Mallah, R., Jackson, R. G., Ball, M., Ibrahim, Z. M., Broadbent, M., Dzahini, O., Stewart, R., Johnston, C., & Dobson, R. J. (2015). Identification of adverse drug events from free-text electronic patient records and information in a large mental health case register. *PloS one*, *10*(8), e0134208.

Wu, H., Hodgson, K., Dyson, S., Morley, K. I., Ibrahim, Z. M., **Iqbal, E**., ... & Sudlow, C. (2019). (Preprint) Contextualised concept embedding for efficiently adapting natural language processing models for phenotype identification

Iqbal, E., Govind, R., Romero, A., Dzahini, O., Broadbent, M., Stewart, R., ... & Ibrahim, Z. M. (2020). The side effect profile of Clozapine in real world data of three large mental health hospitals. PLoS One, 15(12), e0243437.C, Mueller., A, Ahmed., Iqbal, E., S, Pinar & Stewart, R. (submitted) Parkinsonian symptoms in Alzheimer's disease and vascular dementia: Co-morbid features and relationship to adverse outcomes

Al-Harrasi, A. M., **Iqbal, E**., Tsamakis, K., Lasek, J., Gadelrab, R., Soysal, P., ... & Mueller, C. (2021). Motor signs in Alzheimer's disease and vascular dementia: Detection through natural language processing, co-morbid features and relationship to adverse outcomes. Experimental gerontology, 146, 111223.

Accepted Articles

A, Segev., **Iqbal, E**., J, Maccabe., E, Gosney., (submitted) Clozapine induced myocarditis: An electronic health register analysis of incidence, timing, clinical markers and diagnostic accuracy

List of Abbreviations

Abbreviation	Description
A&E	Accident and Emergencies
ADE	Adverse Drugs Event
ADEPt	ADEPt Adverse Drug Event annotation Pipeline
ADR	Adverse Drugs Reaction
AL	Active Learning
BCPNN	Bayesian Confidence Propagation Neural Network
BIDMC	Beth Israel Deaconess Medical Center
Bi-LSTM	Bi-Directional Long Short-Term Memory
BMI	Body Mass Index
BNF	British National Formulary
BRC	Biomedical Research Centre
C&I	Camden & Islington
caTIES	Cancer Text Information Extraction System
CBT	Cognitive Behavioural Therapy
CNS	Central Nervous System
CPK	Creatinine Phosphokinase
CPRD	Clinical Practice Research Datalink
CRF	Conditional Random Fields
CRIS	Clinical Record Interactive Search
CRN	Clinical Research Network
cTAKES	clinical Text Analysis and Knowledge Extraction System
D-CRIS	Distributed Clinical Record Interactive Search
DDI	Drug-Drug Interactions
DL	Deep Learning
ECG	Electrocardiogram
ECHO	Echocardiography
EEA	European Economic Area
EHR	Electronic Health Records
EMA	European Medicines Agency
EMIS	Egton Medical Information Systems
ePJS	Electronic Patient Journey System

EPS	Extrapyramidal Syndromes
EPSE	Extra-Pyramidal Side Effects
EV	EudraVigilance
FAERS	Adverse Event Reporting System
FDA	Food and Drug Administration
FDR	False Discovery Rate
FHIR	Fast Healthcare Interoperability Resources
GATE	General Architecture for Text Engineering
GDPR	European General Data Protection Regulation
HDR	Health Data Research
HES	Hospital Episodes Statistics
HITEx	Health Information Text Extraction
HL7	Health Level 7
ICD	International Classification of Disease
ICSRs	Individual Case Safety Reports
ICU	Intensive Care Unit
ISAC	Independent Scientific Advisory Committee
ISO	International Organisation of Standardisation
KCH	King's College Hospital
KCL	King's College London
KHP	King's Health Partnership
KMCI	KnowledgeMap Concept Identifier
LGPS	Longitudinal Gamma Poisson Shrinker
LONIC	Logical Observation Identifiers Names and Codes
MADE	Medication and Adverse Drug Events
MadLEE	Medical Language Extraction and Encoding System
MAOI	Monoamine Oxidase Inhibitors
MedEx	Medication Information Extraction System
MGPS	Multi-item Gamma Poisson Shrinker
MHRA	Medicine Health Regulatory Agency
MIMIC	Medical Information Mart for Intensive Care
ML	Machine Learning
MMSE	Mini-Mental State Examination
NER	Named Entity Recognition

NHS	National Health Services
NIHR	National Institute for Health Research
NLP	Natural Language Processing
NLTK	Natural Language Toolkit
NSAID	Non-Steroidal Anti-Inflammatory Drugs
OPD	Out-Patient Department
Oxford	Oxford Health
PAMF	Palo Alto Medical Foundation
PCORI	Patient-Centered Outcomes Research Institute
POMR	Problem-Oriented Medical Records
POS	Part of Speech
PPV	Positive Predictive Value
PRR	Proportional Reporting Ratio
RCT	Randomised Control Trial
RNN	Recurrent Neural Network
ROR	Reporting Odds Ratio
SIDER	Side Effects Resource
SJS	Stevens-Johnson syndrome
SLAM	South London and Maudsley
SMI	Severe Mental Illness
SMMSE	Standardise Mini-Mental State Examination
SNRI	Serotonin-noradrenaline reuptake inhibitors
SRS	Spontaneous Reporting System
SSRI	Selective Serotonin Reuptake Inhibitors
STRIDE	Stanford Translational Research Integrated Database Environment
SVM	Support Vector Machines
TCA	Tricyclic antidepressants
UIMA	Unstructured Information Management Architecture
UK-CRIS	United Kingdom - Clinical Record Interactive Search
UMC	Uppsala Monitoring Centre
UMLS	Unified Medical Language System
WBC	White Blood Cells
WHO	World Health Organisation

Table of contents

List of Figures

List of Tables

Chapter 1

1 Introduction

In this thesis, Electronic Health Records (EHR) are used to complement our knowledge about adverse drugs reactions (ADRs) in mental health care settings. It is known that there are ADRs associated with psychotropic drugs even at regular doses (Sengupta et al., 2011). Clinical trials and Spontaneous Reporting Systems (SRS) have limitations, and we hypothesise that many novel ADRs are not captured through these typical methods. This thesis explores how drugs work in real-world data. First, the thesis introduces EHRs, their opportunities, and the challenges of using the EHR data for secondary research. The thesis uses the Clinical Record Interactive Search (CRIS), a de-identified version of the local psychiatric health EHR system at the South London Maudsley NHS Foundation Trust. CRIS is unique as it contains structured and unstructured information where clinicians discuss patient phenotypes in free-text clinical narratives. There are other health data resources available, but they lack the granularity required to study ADRs in real-world settings. There are several tools available for extracting information from clinical text, using different methods from rule-based to machines learning, but limitations include availability, they have not been tested on large scale psychiatric free-text documents and have not been tested in similar settings in other EHR systems. Hence, there is a need to create NLP tools and methods for ongoing pharmacovigilance of psychotropic drug-induced ADRs that are freely available and portable to other psychiatric and general health settings.

1.1 Electronic Health Records

Electronic Health Records (EHRs) are computerised longitudinal medical records containing patient health information collected throughout their contact with healthcare services. The idea of the EHR has been around since the 1960s when Larry Weed introduced the concept of Problem-Oriented Medical Records (POMR) (Jacobs, 2009). However, it was only in 1972, that the first functional EHR system was introduced by the Regenstrief Institute (Tange et al., 1997).

Since the adoption of EHRs, quick access to the patient's health information has become available at the point of care, but it is often not accurate, up-to-date or complete (Staroselsky et al., 2006). Over time, EHRs have become more comprehensive and dynamic, establishing links between departments such as Accident and Emergencies (A&E), laboratories, pharmacies, wards, imaging, Intensive Care Units (ICU) and the Out-Patient Department (OPD) (Bowman, 2013; R. Evans, 2016). These systems shifted the attention from data entry, scheduling and billing to more clinical functions. For example, pharmacy systems were first introduced for prescribing purposes, but more functions were added, such as allergy alerts and drug interactions as typical feature sets. Later these features were made available beyond the pharmacy department.

1.1.1 Secondary Analysis of EHRs

1.1.1.1 Opportunities

Modern EHRs generate a large amount of information in digital format, which represents a complex and diverse collection of patient-centred (Cortada et al., 2012) structured content such as administrative data, demographics, laboratory results, diagnoses, imaging (CT Scan, X-Ray, Ultrasound), and pharmacy data. However, the majority of the information recorded in EHRs comprises unstructured data in the form of free-text clinical notes and admission and discharge summaries. This wealth of data provides an opportunity for secondary analysis of EHRs in areas of pharmacovigilance, such as drug safety and prevention of adverse events, personalised medicine, drug discovery, polypharmacy, disease patterns, comorbidities, drug efficacy, patient stratification, genetic contribution and other socio-economic factors (Data, 2016; Jensen et al., 2012; J. B. Jones et al., 2015; Kadra et al., 2016; G. Wang et al., 2015; Yao et al., 2011).

The secondary use of EHR may supplement the findings of Randomised Clinical Trials (RCT). While RCTs remain the gold standards for establishing causal links between drugs and adverse events, they are limited by their narrow and small populations, short follow-up, and high expenses (Cartwright, 2010; Celi et al., 2014; Humphreys et al., 2013). In contrast, EHRs contain more recording of treatment effects, are less expensive to perform, and provide a large study population and thus help address some of these issues (Angus, 2007; Ioannidis et al., 2001).

1.1.1.2 Challenges

Despite the research potential of the data stored within EHRs, the fact that EHRs are not designed with research usage in mind poses a number of technical and operational challenges in the extraction and usage of EHR data for secondary research purposes. Some of these challenges are summarised below.

1.1.1.2.1 Privacy

EHRs contain identifiable patient information, which is difficult to access by data scientists due to privacy and data security. To deal with this issue, efforts have been made to generate de-identified patient records to facilitate access for researchers. For instance, the United Kingdom (UK) government enable the use of National Health Services (NHS) data for research in a de-identified format through the Caldicott recommendation (Caldicott, 2013). In 2018, the European General Data Protection Regulation (GDPR) was introduced, which includes guidance on the use of health records for research purposes (Grundstrom et al., 2019; Mostert et al., 2016).

1.1.1.2.2 Data completeness

Data completeness describes the extent and nature of missing values within the EHR databases (Liaw et al., 2013). Missing data introduces ambiguity and bias against EHR based studies, which may lead to insufficient data to conclude analysis (Beaulieu-Jones et al., 2018). The data completeness is context-driven data can be missing because the event did not occur, there was a failure to record or incomplete information was recorded, leading to an obstruction of the complete picture of observation. For example, a patient has been prescribed medication without the date and dosage information. In this case, it is hard to establish if the patient is on the medication. There are several methods to manage the missing data. These can include alternative data sources or data triangulation, the substitution of elements that represent missing variables, and statistical methods. If medication, diagnosis and BMI information are not available in the EHR dataset, the pharmacy system within the hospital or primary care research databases can be used to obtain this information. These databases are discussed in section 1.7 of this thesis. Statistical methods are commonly used to handle missing data by removing cases with missing values, replacing missing values with imputation methods and likelihood estimates of missing values (Allison, 2001; Schafer, 1999; Schafer et al., 2002). These

statistical methods have limitations such as effects on sample size, they may not be easily applicable and may produce biased results (Allison, 2001; Talbert et al., 2013).

1.1.1.2.3 Data Consistency

Data consistency relates to the constancy of the data attributes, where two or more attributes are recording the same information for observation at the desired degree of detail within and across databases and datasets (Weiskopf et al., 2013). The data consistency includes data measures (consistency of variables unit of measurement and reference range), the procedure of measurement (documentation of specific variables in the data source) and granularity (degree of detail and the consistency of variable granularity across databases) (Feder, 2018; Kahn et al., 2012; Weiskopf et al., 2013). The data consistency is compromised when data have been recorded in multiple formats, measurements and units due to the different clinical recording behaviour, preferences, procedures, and in the selection and reporting of different variables across single and multiple EHR data sources. For example, the diagnoses for the patients have recorded inconsistency in ICD-9 and ICD-10 codes, the weight of a patient has been recorded on a different scale (kilogram vs pound), lab reports being outside the range of reasonable values, or the medication prescription has been recorded in the brand and generic names.

Clinical recording behaviour in different institutes may differ, and it is dependent on local preferences and procedures. For example, the different EHR settings may follow the different lab reports, symptoms, diagnoses and follow-up procedures. These inconsistencies can significantly affect the reliability of the datasets for secondary use (Botsis et al., 2010). The careful selection of variables and EHR data sources can address the data consistency. Manual review and comparing variables within and across databases by applying data validation rules can identify these inconsistencies (Kahn et al., 2012). Finally, using different data sources (discussed in section 1.1.2) or surrogate variables may supplement variables with less granularity.

1.1.1.2.4 Data Timeliness

The data timeliness or data currency refers to the data quality of being recent. The EHR data should be recorded within the close proximity of an event and should be relevant to the current medical knowledge (Weiskopf et al., 2013). The data entered outside of a specific time frame could impact the research or treatment, leading to incomplete data. The data entry procedures for structured data, lab results, the introduction of new

variables, and old standards becoming obsolete can limit the data's overall utility for research purposes. The timestamp data plays an important role in evaluating data timeliness and measuring the time difference between two related events within or across EHRs. For example, EHR data can be used to study the ADE prevalence rate of a new drug Lurasidone released in the UK market in early 2014. In this case, researchers can preselect the onset of data entry points from early 2014 and remove EHR data beyond those onsets.

1.1.1.2.5 Data Accuracy

The data accuracy can be defined as the degree to which a value in the EHR is a true and precise description of the real-world value (Weiskopf et al., 2013). Accuracy is compromised when healthcare professionals record information in free-text format. In free-text format, clinicians may use a variety of synonyms, proxies or may use the old coding system to describe a procedure or a condition, and follow the recording culture within a hospital. The free-text presents a higher incidence of data inaccuracy (Sukumar et al., 2015).

On the other hand, the accuracy of data is good when healthcare professionals select predefined values such as diagnosis and medication (Bowman, 2013). S. G. Johnson et al. (2016) noted that standardised data entry rules could improve data accuracy. These rules define the expected restrictions placed on a variable such as clinical context, consistency, relationship with other entities and change over time (S. G. Johnson et al., 2016; Kahn et al., 2012).

When the same information is recorded in different formats (predefined or free-text format), it leads to inconsistencies in the EHRs. There is also a possibility of information being entered by an individual incorrectly. For example, Brennan et al. (2012) reported several irregularities and coding errors in the NHS EHR database and proposed careful selection and inspection of EHR data as a result. Other studies have proposed similar means that computer-related errors made by healthcare professionals are common in EHRs settings, and have in the past led to increased mortality rates (Koppel et al., 2005; van Stiphout et al., 2018; Walsh et al., 2006).

The data accuracy assessment can be achieved by using data validation rules and comparing the variables within the database (for internal validity) and comparing the variables with external sources (for external validity). For example, ICD-9 and ICD-10

diagnostics codes are often used to describe medical conditions within free-text documents as a surrogate marker. The move from ICD-9 to ICD-10 has considerably increased the number of available diagnostic codes leading to increased code precision (Sanders et al., 2012). Statistical methods and manual chart review are often used to determine the accuracy of the variable of interest. Simple statistics tests (e.g., mean, median, mode, range, Pearson chi-square, standard deviation) can help determine if variables follow the logical restrictions, distribution, frequencies, and patterns in the dataset (Feder, 2018). Manual chart reviews are often used to determined surrogate marker accuracy by calculating the precision, recall, specificity and negative and positive predictive values (Rosenman et al., 2014; Weiskopf et al., 2013).

1.1.1.2.6 Data Biases

The distribution and completeness of EHR data in longitudinal cohort studies may introduce biases. These biases can be assigned into three categories: selection, information, and confounding bias (Grimes et al., 2002). Health-related information is recorded when a patient gets in touch or admitted to a hospital. Therefore, a patient with long-term and serious illnesses (active patients) will have more recordings of baseline health-related information such as prognoses, diagnoses, ADE and medications, compared to the healthier patients with non-serious illnesses (inactive patients). Including inactive patients in these studies may introduce selection bias (Weber et al., 2017).

When creating a cohort from EHR data, it is important to focus on how to define each individuals' baseline variables (Vassy et al., 2018). For example, it is highly likely that inactive patients will have incomplete ascertainment of full sets of baseline variable information at any given time point. Limiting the analyses only to those patients with complete baseline data at one timepoint reduces the cohort size and excludes the eligible patients, resulting in cohort selection that favours active patients and introduces selection bias (Hripcsak et al., 2011; Weber et al., 2017).

Similarly, different clinicians provide treatment and record their observations, depending on whether the patient is in inpatient or outpatient care. They tend to follow different processes within the EHR. Hripcsak et al. (2013) noted that a different ordering pattern was observed for lab tests by the clinical context in inpatient compared with outpatient surgery events. In another study (Hripcsak et al., 2015), they found the lab tests and medication orders are requested more frequently for sick patients, leading to the possibility of EHR biased towards sick patients. Extending the definition of baseline

timeframe may introduce cohort characterisation leading to information bias, as inactive patients are more likely to have incomplete information at any one instance (Hripcsak et al., 2011).

All the biases discussed above resulting from the way data has been collected within EHR are called healthcare process biases, and examining the distribution of relevant variables helps researchers minimise these bias risks (Levine et al., 2016). Further, inadequate adjustments of covariate that is both predictive of treatment and outcome may result in cofounding bias (Haneuse, 2016). Moreover, inpatients have more recording of the same events such as blood pressure, medication and temperature as compared to outpatients, which can lead to information bias. Similarly, in a psychiatric health setting, outpatient care is usually provided by senior clinicians such as a psychiatrist or a psychologist and contains detailed recordings of the past and current events, however with inpatient care with ward staff tends they tend only to have a more recent recording of events which leads to selection and information bias.

In this thesis, the cohorts are defined to identify ADEs, medication and relevant comorbid factors such as diagnosis, demographics, hospital admission and smoking status. In each subsequent study, the approach is described in examining the baseline timeframe to achieve a higher yield of eligible patients. For example, the diagnosis was determined by widening the timeframe to achieve a higher yield. The decision made to define the baseline timeframe can result in the missing data problem (Newgard et al., 2015), with opposing biases at either end of the spectrum. A narrower timeframe will select the sicker population but a higher yield of available variables, while a longer timeframe will include patients with missing data and may result in misclassification. The data visualisation was performed to determine the impact of different variable rates, yield and minimise the potential biases.

1.1.1.2.7 Format

One of the primary challenges in the secondary use of EHR is to extract information from free-text clinical narratives. Clinicians tend to use free-text clinical narratives for reasoning and observation (Farri et al., 2013), recording details of the patient experience such as co-morbidities, diagnoses, drugs, symptoms and Adverse Drugs Events (ADE). The retrieval and extraction of this information from the free-text is complicated due to the grammatical and spelling errors, use of inconsistent abbreviations and coding systems, as well as the different contexts surrounding the target clinical terms. These issues need

to be carefully addressed when attempting to mine information from free-text by automated natural language processing (NLP).

1.1.1.2.8 Interoperability

Interoperability refers to the exchange of health-related information between two or more systems. Although EHR systems are increasingly adopting interoperability standards such as Health Level 7 (HL7), standard Fast Healthcare Interoperability Resources (FHIR) (Pais et al., 2017), and openEHR Reference Model (Demski et al., 2016), challenges of interoperability between systems remain. EHR providers are continuously encouraged to make their systems adhere to interoperability standards as opposed to the proprietary, closed systems they typically develop.

1.1.1.2.9 Data integration of structured and unstructured sources

In order to have a detailed patient phenotype profile, there is always a need to integrate structured and unstructured information in standardised vocabularies such as Systematized Nomenclature of Human Medicine Clinical Terms (SNOMED CT) for structured clinical vocabulary (Donnelly, 2006), the International Classification of Disease (ICD, 2015) for diagnosis and RxNorm for drug vocabularies in pharmacy management (S. Liu et al., 2005). Efforts have been made to retrieve, annotate and store structured information alongside unstructured information, but it has proven to be difficult due to the heterogeneous nature of EHRs (R. Jackson et al., 2018; H. Wu et al., 2018).

1.1.1.2.10 Scalability

Scalability refers to the technological platform the EHRs are built upon and whether it can support complexities arising from ever-changing needs such as multi-platform support or storing and exchanging a large amount of information (X. Zhang et al., 2013). Over time, the EHRs may use different coding standards and data structures to store information in the databases, leading to the possibilities of an earlier coding system becoming obsolete and data structure differences leading to conversion failure (Gettinger et al., 2012). Therefore, transposing from one EHR system to another imposes the risk of legacy data being lost or partially migrated to the newer system.

1.2 Severe Mental Illness

Severe or Serious Mental Illnesses (SMI) significantly reduce the ability of an individual in one or more life activities (Kessler et al., 2003). SMI diagnoses can be divided into three distinct categories: Schizophrenia (ICD-10: F20-F29, excluding of F25), Schizoaffective (ICD-10: F25) and Bipolar (ICD-10: F31) disorder.

SMI patients have a higher rate of mortality and morbidity; their life expectancy is reduced by ten years compared to the general population due to cardiovascular and infectious diseases and diabetes (Newman et al., 1991; Robson et al., 2007). SMI disorders may impact all areas of daily living for an individual including home, work, social contacts and relationships (Lysaker et al., 2007). Some patients show suicidal behaviour (Radomsky et al., 1999; Spivak et al., 2003). There is no cure for Schizophrenia or Schizoaffective disorder, but the symptoms of these disorders can be controlled with proper treatment, including the use of psychotropic medications.

1.2.1 Schizophrenia disorder

Schizophrenia is a chronic brain disorder that alters the way a person thinks, acts, manifests emotions, perceives reality and relates to others. The symptoms are often categorised into positive and negative symptoms (Poole et al., 2000). Positive symptoms include delusions (believing things that are not real), hallucinations (hearing voices or seeing people or things that are not real), disorganised speech, agitation, and catatonic behaviour. Negative symptoms include emotionlessness (when there is little emotional facial response), avolition (lack of will to do anything), anhedonia (general lack of interest in taking pleasure from previously enjoyed activities which leads to social withdrawal), and lochia (lack of elaborated speech) (Dixon et al., 1999; Leutwyler et al., 2014).

Other symptoms of schizophrenia are cognitive and mood symptoms. These symptoms are typically not as severe as the positive and negative symptoms, which are treated with antipsychotic drugs. Cognitive symptoms include attention deficit and memory lapses, while mood symptoms include euphoria, boredom, grandiose thinking and dysphoria (Bora et al., 2009; Green, 2006).

Schizophrenia patients are categorised into treatment-responsive and treatment-resistant. Treatment-responsive patients have generally prescribed a range of

antipsychotic drugs, whereas Clozapine is the only evidence-based treatment for treatment-resistant patients (Gillespie et al., 2017).

1.2.2 Schizoaffective disorder

Schizoaffective disorder is when an individual has concurrent symptoms of schizophrenia such as hallucinations and delusions and also manifests symptoms of a mood disorder such as depression mania or bipolar disorder. The symptoms of schizoaffective disorders are treated with antipsychotics, mood stabiliser and antidepressants drugs. Studies suggest long-acting monthly injections with medications such as Paliperidone are effective treatments for schizoaffective symptoms (Chue et al., 2016; Fu et al., 2016).

1.2.3 Bipolar disorder

The life of a patient with bipolar disorder is split between two different realities, elation (great happiness) and depression (Phillips et al., 2013). There are different variations of bipolar disorder, but it can be divided into Type 1 and Type 2. Type 1 has extreme highs (manic episodes) alongside the low (depressive episodes). Type 2 involves briefer, less intense periods of manic episodes spread with an extended period of depression. These episodes exceed the typical feeling of joy, causing troubling symptoms such as racing thoughts, sleeplessness, rapid speech, rash action and unnecessary risk-taking behaviour (Weissman et al., 1996).

Without treatment, these manic and depressive episodes become more frequent and intense and take longer to subside. The depressed phase of bipolar disorder manifests in many ways, including low mood, changes in habit and appetite, feelings of worthlessness, sleeping either too much or too little, restlessness and persistent thoughts of suicide. Bipolar patients are usually treated with antidepressant and mood stabiliser drugs.

1.3 Psychotropic drugs

Initially, the aim of this thesis was to investigate ADEs associated with antipsychotics. However, SMI patients are also prescribed psychotropic drugs such as antidepressants (Courtet et al., 2011; Ghaemi et al., 2003), mood stabilisers (McElroy et al., 1987) and hypnotics and anxiolytics (Ilyas et al., 2012). These drugs affect the Central Nervous

System (CNS), changing how the brain processes information such as emotions, thoughts, and behaviour. Therefore, the work was extended to include identification of other psychotropic drugs ADEs in addition to antipsychotics.

The complete list of these drugs along with their primary, secondary and tertiary category and generic and brand names, are available in supplementary Table A.2. The associated ADEs, along with synonyms, phrases and alternative spellings, were generated with the help of pharmacists. Table 1.1 represents a complete list of common to rare ADEs associated with psychotropic drugs and some other generic ADEs added to the list for evaluation purposes. The complete list of ADEs, including synonyms and alternate spellings, are available in supplementary Table A.4.

1.3.1 Antipsychotics

Antipsychotic drugs, also known as neuroleptics, are prescribed to patients with schizophrenia, schizoaffective and bipolar disorder. Antipsychotic drugs are also widely used in dementia patients to treat agitation, delusions and aggression (Gill et al., 2007; L. S. Schneider et al., 2005). Antipsychotic drugs block dopamine channels in the brain, reducing the flow of thoughts in psychotic states. Antipsychotic drugs have two categories, first-generation (also known as typical) and second-generation (also known as atypical). Typical antipsychotic drugs were first developed in the 1950s and are well known for their Extra-Pyramidal Side Effects (EPSEs), such as dystonia, akathisia, tardive dyskinesia and Parkinsonism (Lally et al., 2015; Rummel-Kluge, Komossa, Schwarz, Hunger, Schmid, Lobos, et al., 2010).

Commonly used first-generation antipsychotic drugs are Chlorpromazine, Flupentixol, Haloperidol, Sulpiride and Zuclopenthixol, while the commonly used second-generation antipsychotic drugs are Amisulpride, Aripiprazole, Clozapine, Lurasidone, Olanzapine, Paliperidone, Quetiapine and Risperidone (BNF, 2018; Datapharm Communications Limited, 2017; Taylor et al., 2018).

1.3.2 Antidepressants

Antidepressants are prescribed to patients to treat symptoms of depression, obsessive-compulsive disorder, anxiety and post-traumatic stress disorder. Antidepressant drugs increase the level of neurotransmitters in the brain. However, it is not known how precisely the drug's mechanism of action works (NHS Digital, 2018). Antidepressants

have four major categories. Selective serotonin reuptake inhibitors (SSRIs) are commonly prescribed due to their lower side effects profiles. The widely used SSRIs are Citalopram, Paroxetine, Fluoxetine, and Sertraline. The other categories of antidepressants drugs are Serotonin-noradrenaline reuptake inhibitors (SNRIs), Tricyclic antidepressants (TCAs), Monoamine oxidase inhibitors (MAOIs) and other antidepressants (BNF, 2018; Datapharm Communications Limited, 2017; Taylor et al., 2018).

1.3.3 Mood stabilisers

Mood stabilisers are usually prescribed alongside or without antidepressants to prevent episodes of mania and depression in bipolar disorders (Schloesser et al., 2012). The commonly used mood stabilisers are Carbamazepine, Gabapentin, Lamotrigine, Lithium and Sodium Valproate (BNF, 2018; Taylor et al., 2018)

1.3.4 Hypnotics and anxiolytics

Hypnotics and anxiolytics are sedatives, commonly prescribed to SMI patients and can significantly increase the risk of mortality over a long period of use (Weich et al., 2014). These drugs may cause physical and psychological dependence, and withdrawal of these drugs is difficult (BNF, 2018). Benzodiazepines (Lorzapam and Diazepam), Promethazine and Zopiclone, are the most common hypnotics and anxiolytics (BNF, 2018; Taylor et al., 2018).

1.4 Adverse Drugs Events/Reactions

An ADE is any harmful effect that may occur during the treatment where a causal link with the drug has not been established (but may exist), whereas an ADR occurs in response to drug use and where there is evidence of a causal link. The World Health Organisation (WHO) defines ADRs as: "Any response to a drug which is noxious and unintended and that occurs at doses used in man for prophylaxis, diagnosis or therapy" (World Health Organisation, 1969).

ADRs can be life-threatening, and it is essential to identify them in order to predict the risk for an individual patient. ADRs constitute a significant cause of morbidity mortality and lengthen the stay in hospitals. In the UK, 28% of emergency department visits in hospitals are drug-related (Patel et al., 2002), and during the time from 1999 to

2008, 0.9% of the emergency hospital admission were ADR related, and 1.1% were drug-related (T.-Y. Wu et al., 2010). In 2008, the NHS spent over £2bn for the improper use of medications, and in 2014, the cost went up to £2.5bn (Compass, 2008; Economics, 2014). Many efforts have been made by the NHS to train clinicians to deal with prescription errors. A recent report suggests that annual NHS spending to treat ADR-related hospital admissions is £1bn, £226 million on drug-related poisoning in A&E and £1.1bn on prescription errors (PricewaterhouseCoopers, 2016).

ADRs can be categorised by severity as mild, moderate, severe and lethal. The cost of managing ADRs are directly proportioned to severity (Nivya et al., 2015). In mild ADRs such as headaches and constipation, no treatment is usually required, and the ADR does not lengthen the hospital stay. Moderate ADRs, such as hypertension, prolong the hospital stay and are generally caused by change or addition of drug or dosage. Severe ADRs such as myocarditis require a change in the drug therapy, and lethal ADRs such as organ failure result in the patient's death (Afkat et al., 2016).

Table 1.1: A list of psychotropic drug-induced ADEs

ADEs			
Abdominal pain	Dizziness	Hypertension	QTC Prolongation
Agitation	Diabetic Ketoacidosis (DKA)	Hypokinesia	Rash
Agranulocytosis	Dry mouth	Hypotension	Rhinitis
Akathisia	Dysarthria	Impotence	Sedation
Akinesia	Dyslipidemia	Increased thirst	Shaking
Alopecia	Dyspepsia	Insomnia	Steven Johnson Syndrome (SJS)
Amenorrhoea	Dystonia	Irritable	Skin reactions
Amnesia	Enuresis	Loss of libido	Stomach Pain
Anxiety	Eosinophilia	Muscle pain	Suicidal behaviour
Apnoea	Epilepsy	Muscle twitching	Suicide ideation
Arrhythmia	Erectile dysfunction	Myocarditis	Suicidal tendency
Arthralgia	Extrapyramidal disorder	Nasopharyngitis	Suicide attempt
Ataxia	Extrapyramidal symptoms	Nausea	Sweating
Backache	Fatigue	Neuropathy peripheral	Syncope
Blurred vision	Feeling sick	Neutropenia	Tachycardia
Bradycardia	Fever	Nightmare	Tardive

			dyskinesia
Cardiomyopathy	Flatulence	Neuroleptic Malignant Syndrome (NMS)	Tinnitus
Catatonia	Galactorrhoea	Numbness	Trembling
Confusion	Glycosuria	Nystagmus	Tremor
Constipation	Gynaecomastia	Oedema	Urinary retention
Convulsion	Hallucination	Parasuicide	Vaginal inflammation
Cystitis	Headache	Parkinsonian	Vertigo
Decreased appetite	High CPK	Pericarditis	Vomiting
Dehydration	Hostility	Peripheral oedema	WBC decreased
Delusion	Hyperglycaemia	Pharyngitis	Weight gain
Diarrhoea	Hyper prolactinoma	Pneumonia	Weight loss
Diplopia	Hypersalivation	Polyuria	
Disorientation	Hypersomnia	Pulmonary embolism	

The table represents a collection of common to rare, mild to severe and acute to chronic Adverse Drugs Events (ADE) related to psychotropic drugs. The list has been generated with the help of three pharmacists and two clinicians.
CPK = Creatine Phosphokinase; QTC = Cardio Contraction Time (Q = Q-wave); WBC = White Blood Cells

1.4.1 Pharmacovigilance

Pharmacovigilance is the practice relating to the reporting, assessment, understanding and prevention of ADRs and prescribing errors. The objective of pharmacovigilance is to improve public health and safety in relation to the use of drugs. All drugs go through RCTs before being released on the market, but not all ADRs are detected due to the limited sample size, demographic biases, time and financial constraints of RCTs. Once a drug has been launched, ADRs can be identified through patient self-reports, physical findings by clinicians, laboratory tests and drug history review.

The Spontaneous Reporting System (SRS) is a system for ADR recording in hospital settings. VigiBase is the most extensive pharmacovigilance program run by the WHO from the Uppsala Monitoring Centre (UMC). In this programme, healthcare providers use the Individual Case Safety Reports (ICSRs) in member countries to log medication error and ADRs (WHO, 2017). In the USA, ADRs and medication errors are reported to the Adverse Event Reporting System (FAERS) (FDA, 2018). In the UK, the Medicine Health Regulatory Agency (MHRA) runs a YellowCard Scheme where doctors

and pharmacists can report new incidents of ADRs (MHRA, 2017). The EudraVigilance (EV) is operated by the European Medicines Agency (EMA) and collects ADR related data in the European Economic Area (EEA) (Postigo et al., 2018).

The SRSs have limitations, under-reporting of observed ADRs is common, with reporters not including ADRs that are perceived as non-serious (Hazell et al., 2006; Sarker et al., 2015). There is a likelihood that many novel ADRs are never reported to the system due to challenges around establishing a causal link between the drug and ADE. Also, healthcare staff may feel that ADR reporting is not within the remit of their duties or that a given ADR may already have been reported earlier.

Fortunately, a large volume of the information not reported within the SRS is recorded in free-text clinical notes. Psychiatric clinical notes contain text that details not only clinical problems but also describes patient activities, mood and general observations. Hence, ADE extraction from clinical text through the application of NLP tools can play an essential role in EHR based pharmacovigilance. The next section introduces the field of NLP, a brief description of NLP steps, widely used tools and EHR based pharmacovigilance methods.

1.5 Natural Language Processing

Natural Language Processing is a sub-field of Artificial Intelligence in Computer Science and is centred around the manipulation of unstructured data (e.g., written and speech) in the context of a specific task (Friedman et al., 1999). An NLP system extracts meaningful contextual knowledge from a human language using models that perform tokenisation, sentence boundary identification, dictionary lookup or Named Entity Recognition (NER), morphological segmentation, Part of Speech tagging (POS), chunking and parsing and contextual extraction. The section below gives a brief overview of the NLP steps mentioned.

Tokenisation is the first step of segmentation and text processing. In tokenisation, each segment of the text is divided into a single base form such as date, punctuation, alphabet, space, numerical, separating and symbolic tokens for further annotation purposes. A sentence splitter splits the text into single sentences when a sentence-ending character such as (.,! or?) is found, followed by the space and capital letter. The sentence splitter also considers the word boundaries as a sentence-ending character (.) which are

used within the abbreviations and dates. Clinical text often contains valuable information such as medications, diagnoses and procedures. Tagging this information using dictionary lookups is a common way of making the presence of relevant entities known. Also, in the biomedical field, specialised entities are tagged with the help of ontologies or data dictionaries, such as the Unified Medical Language System Metathesaurus (UMLS) (Bodenreider, 2004). Morphological segmentation is the mapping of various inflections of words back to their root word or initial substring such as (patients to [patient], haven't to [have] [not]). The Porter stemming algorithm (Willett, 2006), NLTK WordNetLemmatizer and the Stanford CoreNLP lemmatiser are the commonly used analysers used to handle word morphology.

In the POS step, grammar tags are applied on each token to identify grammatical category such as (NN for noun, CC for corresponding conjunction, FW for foreign word, JJ for adjective, PP for personal noun, RB for adverb, NN for noun, PP for personal noun, VB for verb SYM for symbol). Chunking also refers to shallow parsing to divide a sentence into high-level segments by grouping the tokens into noun phrases and building a formal structure of a sentence. Finally, parsing, or syntactical analysis, identifies the grammatical relationships by obtaining the parse tree structure of lower syntactic units. Combining these layers of textual information facilitates information extraction tasks. Once the basic structure and named entities are annotated on the free-text, rule-based or Machine Learning (ML) algorithms are applied to extract contextually meaningful target information.

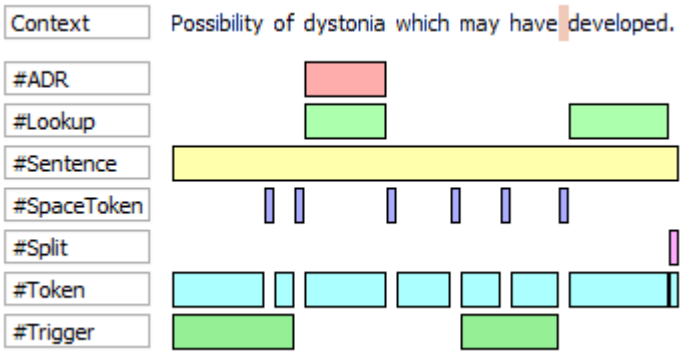

Figure 1.1: Annotation sets created by tools such as the GATE; GATE = General Architecture of Text Engineering

1.5.1 Creating corpora and evaluation metrics

In the biomedical NLP field, annotated documents are often created with the help of one or more human annotators, who are usually clinicians or researchers. Cohen's Kappa statistic is generally used to measure the strength of the agreement between two or more annotators. Cohen's Kappa is defined as

$$K = \frac{P_o - P_e}{1 - P_e}$$

Where is P_o is the observed agreement between the annotators and P_e is the probability agreement by any chance. The Kappa result can be interpreted as follows: values ≤ 0 as representing no agreement and 0.01–0.20 as none to slight, 0.21–0.40 as fair, 0.41– 0.60 as moderate, 0.61–0.80 as substantial, and 0.81–1.00 as almost perfect agreement (McHugh, 2012). The annotated documents act as a gold standard to perform a specific task. They are also used as training data for supervised ML algorithms and as benchmarks to evaluate their performance. For a given corpus, the following metrics are used to evaluate the predictions (annotations) generated by an algorithm:

TP = the true positive counts where the condition is present

TN = the true negative counts where the condition is absent

FP = the false positive counts where the condition is absent but detected as present

FN = the false negative counts where the condition is present but detected as absent

With these values, the classification performance of any given system can be interpreted via statistical measures which use the above metrics to formulate their values. These include precision, recall, specificity, accuracy and F-measure.

The precision or positive predictive value is defined as what proportion of positive cases are correct:

$$Precision = \frac{TP}{\mathrm{TP} + \mathrm{TN}}$$

The recall or sensitivity is defined as what proportion of positive cases are identified correctly:

$$Recall = \frac{TP}{\mathrm{TP} + \mathrm{FN}}$$

The specificity or true negative rate is defined as what proportion of negative cases are identified correctly:

$$Specificity = \frac{TN}{\mathrm{TN} + \mathrm{FP}}$$

The accuracy is defined as what proportion of positive and negative cases are identified correctly:

$$Accuracy = \frac{TP + TN}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}}$$

The F-measure is a harmonic mean of both precision and recall

$$F - measure = \frac{2TP}{2\mathrm{TP} + \mathrm{FP} + \mathrm{FN}}$$

The false positive rate (FPR) is a method used in multiple hypothesis testing to correct multiple comparisons:

$$FPR = \frac{FP}{\mathrm{FP} + \mathrm{TP}}$$

1.5.2 NLP and information extraction Tools

A range of open-source and commercial NLP tools exist to analyse free-text clinical narratives to generate structured findings such as ADEs, medications, diseases, symptoms, and diagnoses, depending on how they are trained. Some of these tools are based on the Unstructured Information Management Architecture (UIMA) (Ferrucci et al., 2009).

In this thesis, the General Architecture for Text Engineering (GATE) was chosen as the core NLP infrastructure for text processing in SLAM CRIS. GATE has been successfully used within SLAM for over a decade, and hundreds of applications have been developed. GATE is an NLP architecture, a framework and a graphical development environment for language engineering. It offers language-processing activities, including

information extraction, building and annotating corpora, and tool evaluation (Cunningham et al., 2011; Cunningham et al., 2013). In this thesis, a modified version of the ConText algorithm was also been used. The ConText algorithm uses regular expressions over pre-indexed clinical conditions and specific sets of words in the text to identify conditions that are negated, hypothetical, historical, or experienced by someone other than the patient (W. W. Chapman et al., 2007). The ConText algorithm is built upon the NegEx algorithm. NegEx is a rule-based algorithm that uses a dictionary of phrases and words indicating negation and implements the dictionary at the sentence level (Chapman et al., 2001).

Another tool, TextHunter, was used to create annotated corpora within this thesis. TextHunter is built on open-source libraries and uses GATE as the core NLP engine. It provides a graphical user interface for a human annotator to search for instances such as a keyword from the clinical text, which can be further annotated into positive, negative and unknown cases. TextHunter also builds and evaluates a range of models against a given task, applying the best performing model to the dataset (M. Ball et al., 2014).

Other tools are available but were not used in this thesis due to the technology they used, and the current infrastructure was not supportive enough to implement them within the SLAM trust. The most commonly used open-source tools for processing EHRs are the clinical Text Analysis and Knowledge Extraction System (cTAKES) built on the Apache openNLP and UIMA framework. Initially developed by the Mayo Clinic, it was later transformed into an open-source Apache project and built upon several analytics engines combining both rule-based and ML methods to extract information from clinical narratives (Savova, Masanz, et al., 2010). Other commonly used tools include Natural Language Toolkit (NLTK) to build python applications for NLP (Bird et al., 2004), spaCy's open-source python library for NER from the free-text (Honnibal, 2017), and MALLET, an open-source Java-based information extraction toolkit which offers a range of ML algorithms for text extraction, such as classification, topic modelling and cluster analysis (A. K. McCallum, 2002).

MedLEE and MetaMap are commercially available tools. MedLEE extracts information from the clinical narratives and stores this information in a controlled vocabulary environment. MedLEE is one of the earliest and commonly used NLP tools for detecting ADRs from clinical text. MedLEE automates the process of encoding the clinical information in narrative patients reports (Friedman et al., 2004). MetaMap

employs a knowledge-intensive approach along with NLP and linguistic techniques to map biomedical text to the Unified Medical Language System (UMLS) Metathesaurus (Aronson, 2001).

1.6 EHR-Based Pharmacovigilance methods

The use of EHRs for post-marketing surveillance, such as detecting ADE-drug signals and ADRs, is an emerging area. EHRs contain more recording of events than spontaneous reporting systems (Cederholm et al., 2015), but challenges and opportunities remain across the field. These challenges include detecting the ADE context, polypharmacy-induced ADEs, integrating heterogeneous data sources, and creating shared corpora (Luo et al., 2017).

Systems such as rule-based are among most the frequently used and have shown good performance (Luo et al., 2017; Shivade et al., 2013). These systems are designed to perform a specific task by following pre-defined patterns and are primarily based on rules and data dictionaries. However, creating large dictionaries and rules is labour intensive. ML methods such as supervised ML can learn to classify documents using labelled data. As supervised ML methods rely on labelled data, domain expert input is still required from clinicians and pharmacists who hold the necessary domain knowledge. More recently, unsupervised ML and have received attention as they do not rely on labelled data (Kreimeyer et al., 2017; Lee et al., 2019; Luo et al., 2017; Yin et al., 2019).

The next section discusses the keyword and triggers phrase search methods, a brief description of NER, rule-based vs ML approaches and their application, statistical methods for signal detection and large scale pharmacovigilance.

1.6.1 Keyword and Trigger Phrase search methods

In early applications, systems were often based on keyword or phrase search and indexed term dictionaries (e.g. keywords corresponding to a disease, ADE or symptoms). These systems were focused on general ADE detection. The keyword and trigger phrase search methods of pharmacovigilance widely used clinical narrative indexed dictionaries which reduces the search time. The trend of this type of pharmacovigilance was later shifted from simple ADE detection to ADE associated with pharmaceutical targets.

Such algorithms were first introduced in the Brigham and Women's Hospital (Boston, USA), where string matching of Micromedex M^2D_2 (Truven Health Analytics, 2013) was applied to identify ADEs in outpatient, structured and free-text clinical notes (Honigman, Lee, et al., 2001; Honigman, Light, et al., 2001). In another study, simple string matching was applied to identify avertable ADEs in ambulatory settings (Gurwitz et al., 2003). The same EHR was used to investigate ADEs resulting from medical management rather than on a patient's underlying condition (Murff et al., 2003).

Field et al. (2004) examined patients aged over 65 and detected possible drug-related incidents. They followed a similar process of string matching in free-text clinical text, as described by (Honigman, Lee, et al., 2001; Honigman, Light, et al., 2001). Cantor et al. (2007) used trigger phrases and customisable grammar structure to identify ADEs in ambulatory care notes for the OPD. In such attempts, the cohort sizes were limited because the level of false positives necessitated a manual review of results.

1.6.1 Named Entity Recognition

Named Entity Recognition (NER) is an active field of NLP in general and biomedical domains since the early 2000s (A. McCallum et al., 2003; Nadeau et al., 2007; Ratinov et al., 2009; Uzuner et al., 2011). NER is an area of NLP that describes the identification and classification of different entities such as medication, symptoms, signs and diseases in clinical text. Community-wide efforts such as BioCreative (Grover et al., 2007) and BioNLP (Alex et al., 2007) have provided many shared tasks such as NER and relation extraction with several systems applying NLP to biomedical text. Some of the efforts in NER are discussed in the section below. The NLP system such as MedLee (Friedman et al., 2004), MetaMap (Aronson, 2001), MedEx (H. Xu et al., 2010), cTAKES (Savova, Masanz, et al., 2010), HITEx (Goryachev et al., 2006), KMCI (Denny et al., 2003) and GATE (Cunningham et al., 2013) are adopted to perform ADE detection tasks with NER components.

1.6.2 Rule-based vs Machine learning Methods

Rule-based and ML are two major approaches for biomedical NLP. Both approaches have advantages and disadvantages. A rule-based system is based on handcrafted rules which mimic the linguistic structure. Rule-based systems are flexible in terms of easy to update with new functions with no significant changes to the core

system. They do not require large, annotated corpus for training. As these rules are primarily based on grammar, these systems understand the linguistic relationship between words to interpret the sentence. Hence, they are popular when handling sentence-level tasks, such as parsing and information extraction and generally better for query analysis. As a result of their manual nature, rule-based systems can typically show high precision and low recall. The disadvantage of a rule-based approach is that it requires domain knowledge expert to craft these rules, which is time-consuming and overtime system can become complicated as the rules can be overriding each other. As domain experts were closely working on the curation of terms dictionary and annotation process, rule-based approach was selected for this thesis.

On the other hand, an ML approach is easy to scale, learn without being explicitly programmed, fast development and have a higher recall. The most significant disadvantage of ML approach is it requires a large amount of training datasets to improve even then the improvements are mainly incremental which reduces the algorithm interpretability and portability (van der Ploeg et al., 2014).

In this thesis, the rule-based approach is selected due to limited training datasets, availability of domain expert, the proximity of ADEs on a single sentence-level and to improve portability.

1.6.2.1 Rule-based systems

Rule-based systems are based on manual rules, usually created by or with the help of domain experts and contain a sequence of events which define a search pattern. In a review by Shivade et al. (2013) identifying phenotypes from clinical text, 24 out of 97 articles were rule-based systems. In a more recent review by Y. Wang et al. (2018), 171 (65%) out of 263 articles/studies used rule-based information extraction systems.

Hazlehurst et al. (2009) conducted a study on outpatients to identify vaccine-related gastrointestinal adverse events. They used MediClass - an automated classification system, and programmed it to identify vaccine-related clinical concepts and linguistic structures used in free-text to extract vaccine-related adverse events (Hazlehurst et al., 2005). Eriksson et al. (2013) described methods to develop an ADE dictionary in Danish clinical narratives and used NER to identify dictionary matches. Further, Eriksson et al. (2014) studied dose-specific ADRs by matching and grouping drugs using the drug dictionary.

X. Wang, Hripcsak, Markatou, et al. (2009) conducted a study on the inpatient department of Presbyterian Hospital, New York. An adjusted version of the MedLEE NLP tool was employed, along with MedDRA symptoms to detect ADE from discharge summaries. They went on to conduct another study (X. Wang et al., 2010) on the same EHR as a data source and ran MedLEE by applying filters (information extraction modules) to capture symptoms and ADRs caused by medication used during hospitalisation. Haerian et al. (2010) used MedLEE with a screening built with domain expert knowledge on discharge summaries for patients with elevated creatine kinase serum.

Haerian et al. (2012) designed an NLP system by focusing on the two life-threatening ADRs, rhabdomyolysis and agranulocytosis. The system identified cases where the event was due to the disease rather than a drug and the system, achieving 0.93 sensitivity and 0.91 specificity. Penz et al. (2007) used the MedLEE and a phrase-matching algorithm to detect central nervous system ADEs. Friedman (2009) used MedLEE to identify drug-related potential ADEs in EHR. E. S. Chen et al. (2008) applied MedLEE and BioMedLEE (L. Chen et al., 2004) and identified disease and drugs from the MEDLINE articles and discharge summaries, applying co-occurrence statistics to evaluate the association between them.

In order to identify negated ADE mentions in clinical text, many studies (Banda et al., 2016; Iyer et al., 2013; LePendu et al., 2013; G. Wang et al., 2015) have used NegEx (W. W. Chapman et al., 2013). NegEx is a rule-based algorithm to identify negation contexts around the target named entity and accommodates a number of languages other than English. Some studies have used their own post-processing rules (Eriksson et al., 2013) and others (Banda et al., 2016; Iyer et al., 2013; LePendu et al., 2013; G. Wang et al., 2015; X. Wang et al., 2010) used extended sets of rule such as ConText (W. W. Chapman et al., 2007) with or without NegEx to determine negation and temporality contexts surrounding ADE mentions. The ConText algorithm is a rule-based algorithm defining the temporal context of an ADE, its negation, and whether the event is related to the patient or someone else.

Bejan et al. (2012) used MetaMap along with NegEx (W. W. Chapman et al., 2001) and ConText (W. W. Chapman et al., 2007) to identify pneumonia-related ADEs from EHR free-text narratives. Teixeira et al. (2016) used MetaMap along with an algorithmic approach to identify hypertensive individuals. Gysbers et al. (2008) used the

Cancer Text Information Extraction System (caTIES), which is built upon MetaMap, as a method for identifying terms suggestive of ADEs. LePendu et al. (2012) used MetaMap to detect patients taking the drug Vioxx and are at risk of getting myocardial infarction.

E. Iqbal et al. (2015) developed a rule-based tool built on GATE to detect extrapyramidal side effects from the free text of psychiatric clinical notes, describe in Chapter 2, and further developed a rule-based pipeline equipped with a modified version of the ConText algorithm (E. Iqbal et al., 2017) to identify psychotropic drug-induced ADEs in the same setting, described in Chapter 3.

Sohn et al. (2011) developed pattern-matching rules manually by examining keywords and expressions surrounding ADE terms to gauge ADEs to drug relationship in the free-text clinical documents of psychiatry and psychology settings. They used regular expression and manual review to map drug names to RxNorm-based concepts. They developed two systems, a rule-based system achieving an F-score of 0.80, and a hybrid system using rules and ML achieving an F-score of 0.75. Q. Li et al. (2014) developed two automated, rule-based systems to detect ADEs and medical errors from the free-text clinical narratives and lab reports in hospital ICU settings.

Other than pharmacovigilance, rule-based information extraction systems have been used in various other studies. For example, Savova, Fan, et al. (2010) applied cTAKES to radiology reports to extract peripheral artery disease cases. They created an expression matching data dictionary for positive, negative, probable and unknown cases and achieved a 0.93 precision, comparing to the NER baseline of 0.46. R. G. Jackson et al. (2017) used GATE along with the rule-based ConText (W. W. Chapman et al., 2007) algorithm to extract SMI symptoms from the free-text clinical narratives.

1.6.2.2 Machine Learning methods

Machine Learning (ML) methods can be divided into three distinct categories, supervised, semi-supervised and unsupervised. In supervised ML methods, the algorithm is trained to learn the pattern from labelled datasets and then applies this model on a test or unseen datasets. In unsupervised ML methods, the algorithm learns directly from the datasets by clustering similar patterns into groups, and in semi-supervised ML methods, the algorithm learns from both labelled and unlabelled datasets.

The most commonly used ML methods in NLP are supervised. In this method, the algorithm learns from manually annotated data and applies the learned model to unseen documents. Support Vector Machines (SVM) is the most commonly used ML method in text annotation, but Conditional Random Forests (CRF) and logistic regression are also in use.

Carroll et al. (2011) used an SVM to identify rheumatoid arthritis. They trained their model on expert-annotated data and achieved 0.94 precision and 0.87 recall. Y. Chen et al. (2013) investigated and concluded that the use of active learning (AL) in the SVM algorithm could reduce the sample size. The study identified rheumatoid arthritis, colorectal cancer, and venous thromboembolism from the free-text documents. Harpaz et al. (2012) used logistic regression after extracting the ADE from EHR narratives and combining their data with AERS. They concluded that combining AERS with EHRs signals could improve the accuracy of signal detection.

Henriksson, Kvist, et al. (2015) used distributional semantics in Swedish health records to identify ADE in free text. First, they identified relevant named entities such as disorders, symptoms and drugs. Next, they labelled the entities with negation, speculation and temporal contexts. In the final stage, they introduced the relationship between labels to identify ADEs by using CRF and showed that distributional semantics improved ADE detection. Henriksson, Zhao, et al. (2015) further experimented with the RFC approach on 27 clinical datasets and demonstrated that the performance significantly improved by multiple distributional semantics obtained from different window sizes. G. Wang et al. (2015) used the text from 9.5 million clinical notes and drug usage and known ADEs as input after labelling them with an annotation application. They used SVM, CRF and other classifiers on training dataset but concluded RFC was superiors in all metrics.

In recent years, advanced ML methods such as Deep Learning (DL) have been used. DL algorithms use a multi-layered artificial neural network to find complicated patterns. DL has been applied in semantic representation and analysis (Yih et al., 2014; J. Zhou et al., 2015), information retrieval (Severyn et al., 2015), entity recognition (Huang et al., 2015) and event detection (Nguyen et al., 2015; Pandey et al., 2017). Pandey et al. (2017) designed a Recurrent Neural Network (RNN) model to enhance to context evaluation of ADE terms. Cocos et al. (2017) developed an RNN model that labels words with ADR in Twitter posts.

Q. Wei et al. (2019) used the MIMIC-III database and used NER and relation classification components such as ADE and medication. They applied and evaluated both DL methods and traditional ML methods and concluded the DL approaches showed superior performance in extracting medications and their attributes such as ADEs. Wunnava et al. (2019) applied RNN and Bi-Directional Long Short-Term Memory (Bi-LSTM) to detect ADEs and medications from the EHR and reported an F-measure of 0.82.

Since 2015, several studies have used an ML approach for pharmacovigilance, mainly due to the availability of EHR data for research purposes, which consequently improves the performance of ML models. Random forests are becoming more popular due to their robust classification performance with other models.

1.6.3 Statistical Methods for signal detection

Statistical methods have been in used for ADE detection since the late ’90s. Statistical methods focus on identifying the common pattern of ADEs, drugs and the association between them.

DuMouchel (1999) applied Bayesian, Multi-item Gamma Poisson Shrinker (MGPS) on the FDA database, and Bate et al. (1998) used the Bayesian Confidence Propagation Neural Network (BCPNN). Both studies concluded that their methods could be used to detect ADR signals in large datasets. S. Evans et al. (2001) developed the Proportional Reporting Ratio (PRR) for use on the YellowCard database and identified 481 signals on newly marketed drugs; 70% of the signals were known ADRs, 13% were related to the underlying disease and 17% required further investigation. X. Wang, Hripcsak, and Friedman (2009) developed a mutual information method to measure the associations generated by NLP processing and reported an overall 81% precision.

Iyer et al. (2013) used data from two different sites: Stanford Translational Research Integrated Database Environment (STRIDE) and Palo Alto Medical Foundation (PAMF). The study estimated the rate of Drug-Drug Interactions (DDI) events among patients on various drugs combinations. Furthermore, Banda et al. (2016) used the same system as described by (Iyer et al., 2013) and developed an algorithm to prioritise drug and ADE associations. M. Liu et al. (2013) performed a comparative analysis of pharmacovigilance methods to detect ADRs from EHRs. They used six signal detection methods: PRP, Reporting Odds Ratio (ROR), Yule’s Q, Chi-Squared, MGPS and BCPN

and concluded all these methods could be used to detect ADRs. However, their ability to detect weak signals is significantly improved by laboratory test results and medication data. Ferrajolo et al. (2014) applied Longitudinal Gamma Poisson Shrinker (LGPS) to identify drug-induced liver injury in children. Banda et al. (2016) used the system by (Iyer et al., 2013) for negation and family history and developed an algorithm for prioritising drug and ADE associations.

Most statistical methods for signal detections are based on hypothesis or prior knowledge and generally have fixed forms and predictors (Luo et al., 2017). Thus, increasing the availability of EHR data does not reflect on the performance of these models.

1.6.4 Large scale pharmacovigilance

The EU-ADR project was initiated in 2008 and aimed to detect signals between ADEs and associated drugs. The project used data mining techniques from 8 EHR systems from four countries (Denmark, Italy, the Netherlands, and the United Kingdom). The system was designed to detect high-frequency signals for known and unknown ADE-drugs associations in the general population (Patadia et al., 2015; Trifiro et al., 2009). In the USA, the Food and Drug Administration (FDA) initiated a pilot program in 2009 (Mini-Sentinel) and an active surveillance system to detect ADRs from EHRs. The system was developed in stages, and in 2014, a fully functional Sentinel system was launched for surveillance (R. Ball et al., 2016).

As can be seen, there are a plethora of emerging approaches for analysing EHR data for ADE detection. Rule bases, statistical models and ML have gained increased attraction. Given the rapid progress, more exciting developments in the NLP-based ADE detections are emerging.

1.7 Health Datasets available for Research

A growing number of publicly released health information and de-identified patient records are available for research purposes in the UK. It is a key aim of major initiatives such as Health Data Research (HDR) UK (HDR UK, 2019). The focus of this thesis is on ADRs in Severe Mental Illness (SMI) patients, mainly in psychiatric health care settings. ADRs are often recorded within the unstructured free-text format, and to the best of our

knowledge, there is no publicly available psychiatric health care resource. Therefore, the work in this thesis was performed through local governance infrastructures of mental health care services. However, some publicly available primary healthcare providers and general practices anonymised health records are discussed here, which may be used along with the psychiatric health EHRs.

In April 2018, HDR UK was launched as a collaboration between 22 universities and research institutes, supported by several public authorities. The HDR provides proficient research services in a secure environment, including NHS administrative and EHRs data (HDR UK, 2019). The database was not in this thesis as the HDR initiative was established after the research was concluded in this thesis.

Databases such as Clinical Practice Research Datalink (CPRD) capture primary care data, and Hospital Episode Statistics (HES) captures secondary care data used as a surrogate data source and improves data quality. The Clinical Practice Research Datalink (CPRD) is an anonymised medical record containing over 19 million patients from approximately 7300 general practices in England. The database is a rich source for health research, providing structured information such as demographics, clinical findings, tests reports, diagnoses and referrals to secondary care. The database is used in over 2000 research studies dating back 30 years (Wolf et al., 2019). Although the CPRD is an excellent resource for health research, some of the information is incomplete, such as diagnosis in primary care records and medication dispensing in secondary care records. Hospital Episodes Statistics (HES), managed by NHS Digital, collects structured data such as demographics, diagnoses, operations and administrative data from all hospital (primary care trusts and psychiatric health trusts) admissions in England. The HES data is published annually and contains longitudinal data from A&E, Admitted patient care, adult critical care, outpatients and maternity (HES - NHS Digital, 2018). The good coverage of diagnosis, demographic, medication and hospital admission episodes are available through local psychiatric health EHR systems. Thus, these data sources were not used in this thesis.

The databases such as Medical Information Mart for Intensive Care (MIMIC) can be used to evaluate the research methods and tools. The MIMIC was established in 2003 and contains de-identified data from Intensive Care Units (ICU) at Beth Israel Deaconess Medical Center (BIDMC) from 2001 to 2012. The current version, MIMIC-III, contains structured and unstructured information for sixty thousand patients (A. E. Johnson et al.,

2016). The MIMIC database provides data such as patient demographics, diagnoses, lab results, clinical notes, cardiology such as Echocardiography (ECHO) and Electrocardiogram (ECG), billing and discharge summaries. Further, researchers can apply NLP tools to extract phenotypic information from the free-text clinical documents. The database is available for access by researchers who are trained to handle sensitive patient information. MIMIC was not used in this thesis, as it only contains general health providers free-text documents.

1.8 EHR systems used in this thesis

1.8.1 South London and Maudsley (SLAM) NHS Foundation Trust

The South London and Maudsley NHS Foundation Trust (SLAM) is one of the largest psychiatric health providers in the UK. The Trust serves a population of over 1.4 million residents living in Croydon, Lambeth, Lewisham and Southwark, providing secondary and tertiary mental health-care. SLAM comprises over 25 community teams, running local clinics and four hospitals, the Maudsley Hospital, Bethlem Royal Hospital, Lewisham Hospital and Lambeth Hospitals (Fernandes et al., 2018; Stewart et al., 2009).

The current EHR used in SLAM, the Electronic Patient Journey System (ePJS), was first introduced in 1999. Before this, multiple unconnected systems were running in the different hospitals of the Trust. The SLAM EHR is typical of many such mental health provider systems in that it stores much of its clinical records and prescribing information in an unstructured free-text format. The EHR has been used widely across all SLAM services since 2006.

1.8.2 Camden & Islington (C&I) NHS Foundation Trust

The Camden & Islington (C&I) NHS Foundation Trust is a psychiatric health provider for two London boroughs of Camden and Islington. The Trust serves a population of around half a million residents living in the Camden & Islington catchment area and some services to the residents of Westminster and Kingston boroughs (Camden and Islington NHS Foundation Trust, 2017; Werbeloff et al., 2018). The Trust operates 20 different sites and two large hospitals at Highgate Mental Health Centre and St Pancras Hospital. An EHR has been in service since 2008. In 2015, the Trust migrated its RiO EHR system to the Carenotes EHR (Digital Health, 2019).

1.8.3 Oxford Health NHS Foundation Trust (Oxford)

The Oxford Health (Oxford) NHS Foundation Trust is a physical and psychiatric health services provider covering a geographic catchment area of 1.9 million residents in Oxfordshire, Buckinghamshire, Swindon and Wiltshire. The current EHR has been in service since late 2009 (Oxford Health NHS Foundation Trust, 2018). In March 2015, the Trust started the migration of its current RiO EHR system to a new EHR system called Carenotes.

1.8.4 Clinical Record Interactive Search (CRIS)

The Clinical Record Interactive Search (CRIS) - a de-identified version of ePJS at SLAM was introduced in 2008 for research purposes (Fernandes et al., 2013; Perera et al., 2016; Stewart et al., 2009). As of June 2018, SLAM CRIS contained over 320,000 patient records comprising over 32 million free-text documents including correspondence, discharge summaries, events, ward progress notes, mental health care plans and mental state formulations (mainly in word or pdf formats) and increasing at a rate of 300,000 new documents per month (Lovestone, 2011). Researchers can apply queries on CRIS to search and extract anonymised data from structured data and unstructured clinical documents.

The CRIS system, including data-processing pipelines and a patient-led governance model, was additionally deployed to C&I and Oxford NHS Trust data to create a research-ready, de-identified version of their EHR systems. As of July 2016, the C&I CRIS dataset contained over 116,000 patient records, comprising over 8 million free-text documents including progress notes, assessment details and correspondence. As of March 2015, the Oxford CRIS contained around 98,000 patient records, comprising over 11 million documents including progress notes, attachments and events. The discrepancy in the dates of the cohorts used from the three Trusts in this thesis is a consequence of the different time frames during which the data was accessed, and when the research was concluded. Figure 1.1 shows the cumulative number of documents from the three NHS Trust by year.

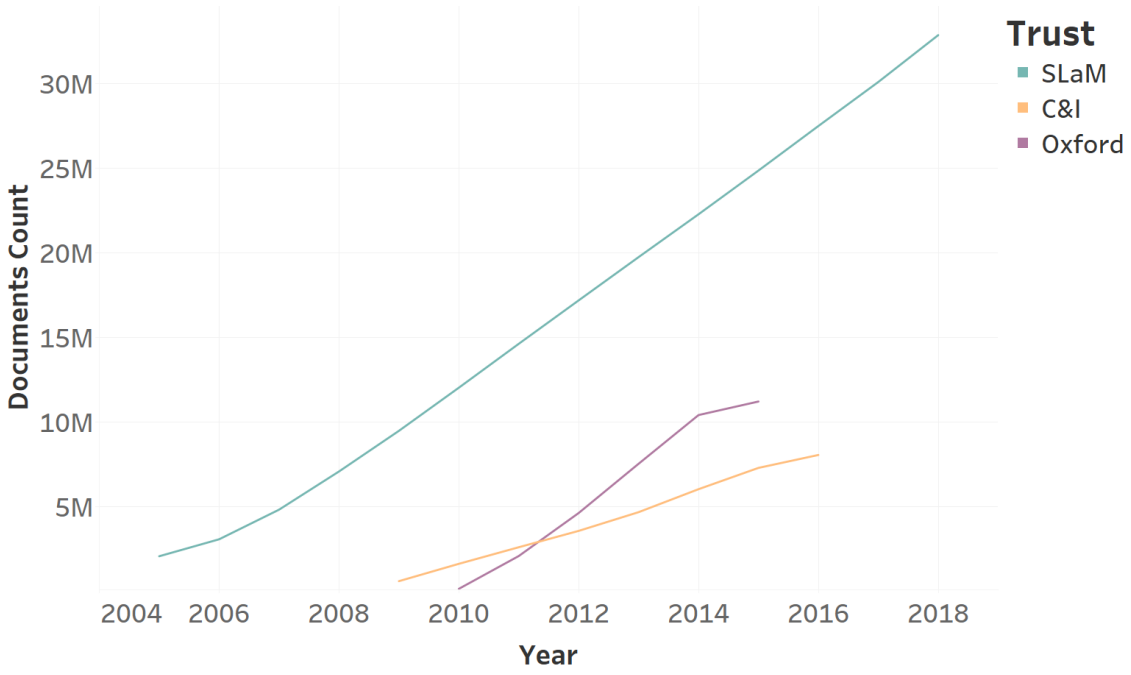

Figure 1.2: The cumulative number of documents (in millions) across SLAM, C&I & and Oxford
SLAM = South London and Maudsley NHS Foundation Trust health record, C&I = Camden and Islington NHS Foundation Trust health record, Oxford = Oxford Health NHS Foundation Trust health record

In SLAM, along with CRIS, the data from other NHS Trusts can be accessible under the collaboration programme Distributed-Clinical Record Interactive Search (D-CRIS) (Mental Health Unit (BRC), 2018). A similar programme UK-CRIS (CRIS Network, 2018), led by Oxford University, build upon D-CRIS and expanded to 14 NHS Mental Health Trusts can also be accessible. The academic groups from King's College London (KCL) are aligned with SLAM and King's College Hospital (KCH) service provisions under the King's Health Partners (KHP) partnership. KHP has a rich biomedical translational research culture among academic, clinical and industrial partners.

1.8.4.1 Demographics: Gender, Age, Ethnicity

In all three Trusts, gender and ethnicity were derived from the last entry recorded. The gender distribution across all three Trusts is shown in Figure 1.2. In SLAM there were 18, in C&I there was 21, and in Oxford, there were 46 distinct ethnic categories. The ethnic categories were divided into four major groups: White, Black, Asian and Others. The ethnic background distributions across three NHS Trusts, are shown in Figure 1.3.

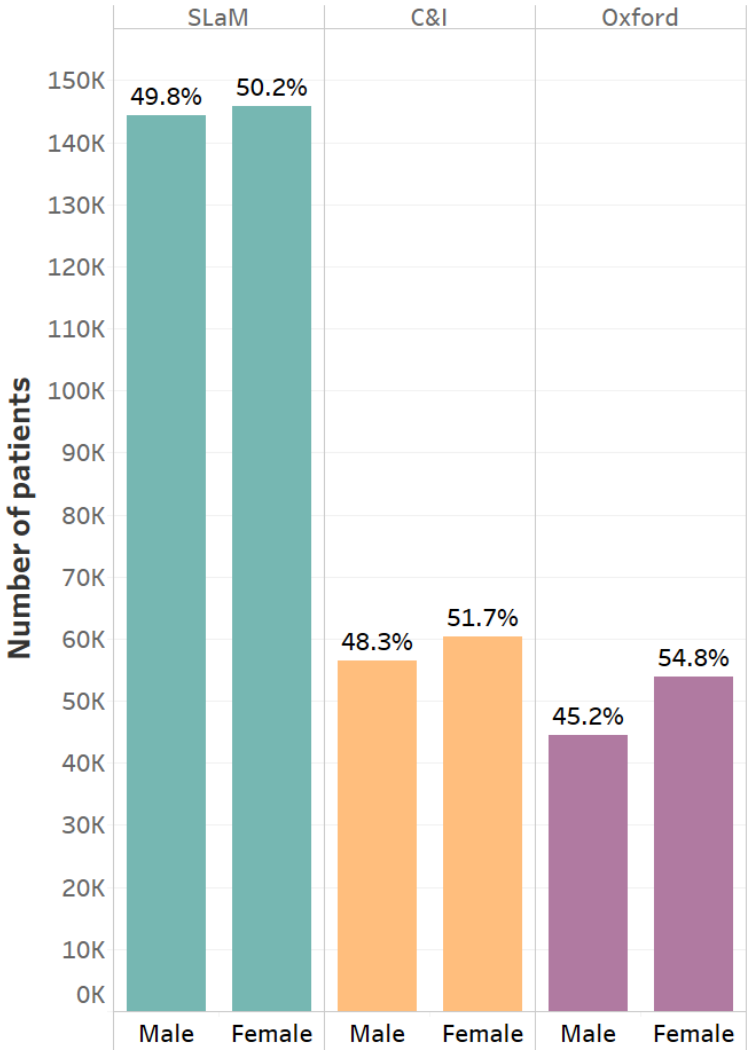

Figure 1.3: Gender distribution of patients in SLAM, C&I & and Oxford
SLAM = South London and Maudsley NHS Foundation Trust health record; C&I = Camden and Islington NHS Foundation Trust health record; Oxford = Oxford Health NHS Foundation Trust health record

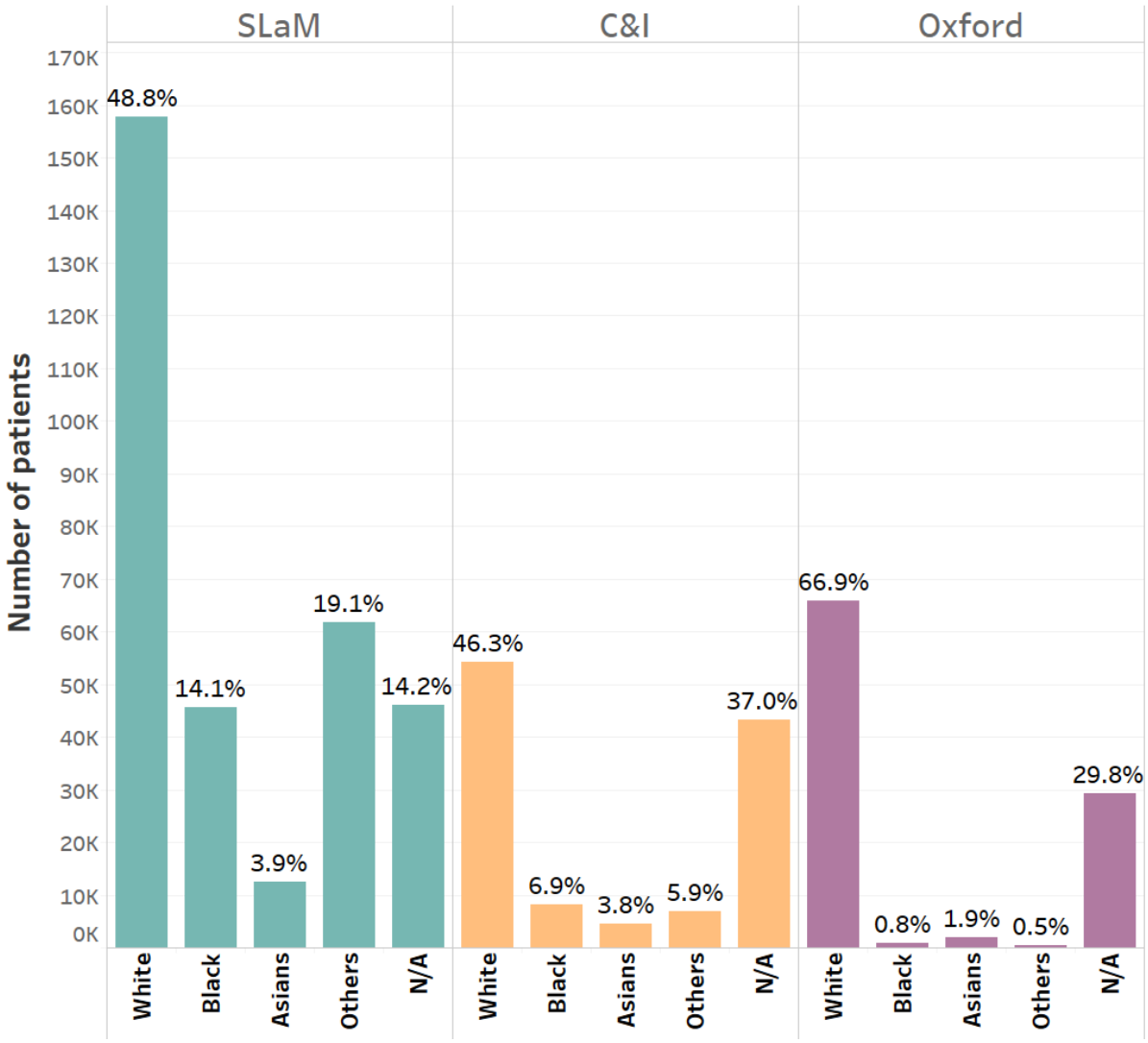

Figure 1.4: Ethnic background of patients in SLAM, C&I and Oxford
SLAM = South London and Maudsley NHS Foundation Trust health record; C&I = Camden and Islington NHS Foundation Trust health record; Oxford = Oxford Health NHS Foundation Trust health record

The cut-off dates used to calculate the age were 30/06/2018 for SLAM, 31/07/2016 for C&I and 31/03/2015 for Oxford, for all the patients except the patients who were deceased. The age for deceased patients was calculated on the available date of death. Overall age distribution among the three Trusts is shown in Figure 1.4. The C&I Trust does not treat children and adolescents, which is why there is a small number of patients in the 'Under 21' age category. A similar overall age distribution can be seen in SLAM and C&I, but Oxford shows a higher proportion of younger and older patients.

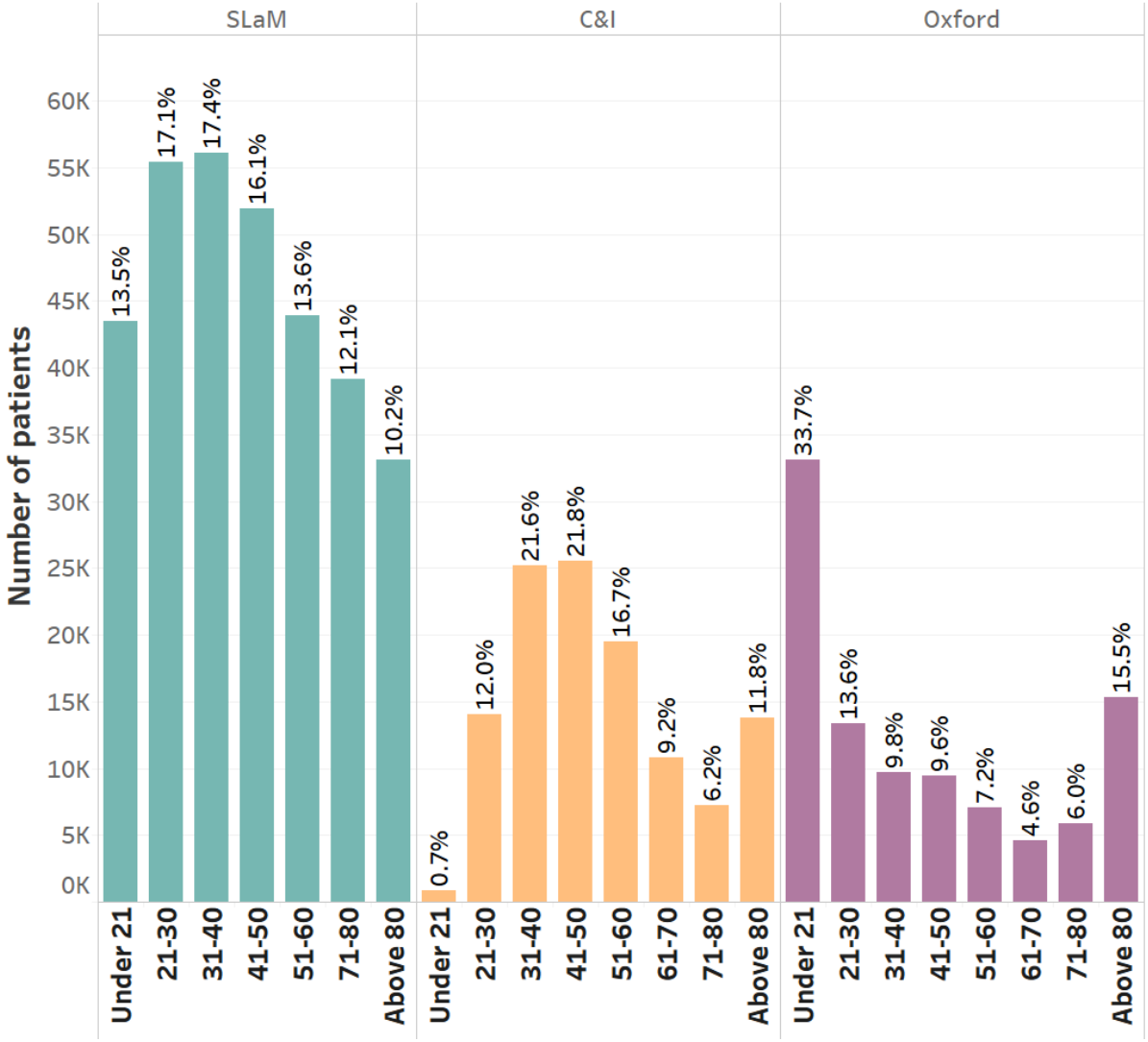

Figure 1.5: Age group distribution of patients in SLAM, C&I and Oxford
SLAM = South London and Maudsley NHS Foundation Trust health record; C&I = Camden and Islington NHS Foundation Trust health record; Oxford = Oxford Health NHS Foundation Trust health record

1.9 Conclusions

Secondary analysis of EHRs has great potential. EHR data is becoming more comprehensive and dynamic with time. It brings opportunities for researchers to explore a vast amount of multidimensional structured and unstructured data along with the challenges. Here CRIS is used as a primary data source as it captures vast phenotypes information in a structured and unstructured format. This work aims to understand how ADEs are recorded in clinical narratives and how they can be extracted by applying

GATE NLP methods. The second focus of this thesis is to identify medication episodes and temporal associations between ADEs and medications to detect possible ADRs. This would enable the use of EHRs for pharmacovigilance and for understanding ADR aetiology and early preventive strategies in real-world psychiatric health care settings.

1.10 Objective

The objectives of this thesis are as follows:

1. Develop NLP algorithms to identify mentions of EPSEs from free-text clinical documents and apply appropriate performance tests such as precision, recall, sensitivity, specificity and F-measure.
2. Develop NLP algorithms by incorporating contextual features and temporal reasoning capabilities to identify ADEs from clinical narratives and application of performance tests such as precision, recall, sensitivity, specificity and F-measure.
3. Develop an algorithm to identify medication episodes and create medication timelines.
4. Determine simple associations between ADEs and medications, infer possible causative relationships (ADRs), and extract potential ADRs signals.
5. Subsequently, conduct descriptive and inferential statistical analyses to stratify the patient population according to medications, diagnoses, ethnicity, age and gender.
6. Apply the developed methods to annotate unseen data in other mental health providers and test the viability of resulting annotations for further studies.

1.11 Thesis overview

The thesis investigates the use of rule-based NLP approaches to detect ADEs from the free-text clinical narratives, on-treatment medication episodes and the association between ADE and medication to deduce possible ADRs. The proposed methods were further applied in other psychiatric health providers to study Clozapine-induced ADRs. Clozapine is the most effective drug for treatment-resistant schizophrenia and is the only antipsychotic drug where routine clinical contact is instructed (Hayes et al., 2014). Therefore, good coverage of patients' phenotypes are recorded within the clinical text.

1.11.1 Chapter 2: Identification of Extrapyramidal side effects from the free-text electronic health records.

The chapter addresses objective one and presents a tool to identify positive from negative ADE mentions, mainly EPSEs and the application of performance metrics such as precision, recall (also known as sensitivity), specificity, accuracy and F-measure. The paper presented in the first part has been published in PLOS One's journal (E. Iqbal et al., 2015). It uses NLP applications applied to EHRs to study the prevalence of EPSEs within subgroups of patients categorised by SMI diagnosis, gender, age and ethnicity in SLAM. The tool identified EPSEs with 0.85 precision and 0.86 recall. In the second part of this chapter, the prevalence of EPSEs are studied in C&I.

1.11.2 Chapter 3: ADEPt, a semantically-enriched pipeline for extracting adverse drug events from Free Text electronic health records

The chapter addresses objective two and investigate ADE characteristics in psychiatric clinical notes. In this chapter, a tool has been developed to detect, annotate and classify ADEs presence as positive or negative, patient-specific or general discussions, and current or past by evaluating the context and temporal reasoning surrounding the ADE annotations. The performance of this tool has been evaluated by precision, recall, accuracy, sensitivity and F-measure. The paper presented in chapter 3 has been published in PLOS One's journal (Iqbal et al., 2017). The tool achieved an overall 0.89 precision and 0.88 recall during internal validation.

1.11.3 Chapter 4: Detecting Adverse Drugs Reaction from the EHR

This chapter addresses objective three and four of this thesis. The chapter discussed the implementation of an Adverse Drug Event annotation Pipeline (ADEPt) NLP tool in two other unseen psychiatric EHR data of C&I and Oxford NHS Foundation Trusts. The tool achieved an overall 0.84 precision and 0.87 recall in external validation at C&I psychiatric EHR. This chapter further discusses the development of on-treatment medication episodes and the association between ADEs and medication to elucidate ADRs possible casual patterns. This approach was implemented in three mental health EHRs of SLAM, C&I and Oxford NHS Foundation Trusts.

1.11.4 Chapter 5: The side effect profile of Clozapine in real-world data of three large psychiatric health providers

This chapter addresses the objective five and six of this thesis. Clozapine is an atypical antipsychotic drug and the only treatment available for treatment-resistant schizophrenia. The chapter investigates the prevalence of Clozapine induced ADRs assessed by demographic, smoking status and hospital admission by applying the methods developed in chapter 4 in SLAM NHS Foundation Trust and further in unseen data of C&I and Oxford NHS Foundation Trusts to test the viability of this method. The method presented in this chapter can be used for other psychotropic drugs in mental health care settings.

Chapter 2

2 Identification of Extrapyramidal Side effects from the free-text electronic health records

2.1 Background

The idea of treating schizophrenia without the debilitating effects of EPSEs is very tempting in clinicians and as well as for patients. EPSEs are movement disorders leading to physical disability, discomfort and distress, and the long-term manifestation of EPSEs contribute to social stigma and isolation for schizophrenia patients, in addition to poor compliance and ultimately poor treatment outcome. The management of schizophrenia requires additional drugs to manage EPSEs that subsequently increase the risk of additional ADEs and drug-drug interactions.

EPSEs are a well-recognised problem that is experienced by patients with first-generation antipsychotics. Casey (1998) noted that 90% of the patients subsequently developed akathisia, dystonia and Parkinsonism and 20% of patients developed tardive dyskinesia with the treatment of first-generation antipsychotics. The introduction of second-generation antipsychotics was met with great expectations of lower incidence of EPSEs due to their mechanism of action, such as lower dopamine receptors affinity. However, different studies have shown the second-generation antipsychotics have caused EPSEs and remains a problem in the treatment of schizophrenia (Casey, 2006; Tarsy et al., 2002; Weiden, 2007). A meta-analysis conducted among first-generation vs second-generation antipsychotics by (J.-P. Zhang et al., 2013) and Leucht et al. (2009) suggested lower incidence of EPSEs for second-generation over first-generation antipsychotics. Other studies also concluded that EPSEs had been linked with both first-generation and second-generation antipsychotics (Leucht et al., 2009; Peuskens et al., 2009). Further studies have proposed there is no difference in first-generation and second-generation of antipsychotics in terms of tolerability and effectiveness with the exception of Clozapine

(Casey, 2006; Haddad et al., 2012; P. B. Jones et al., 2006). Therefore, the search for more effective drugs with a lower incidence of EPSE remains relevant and produces an unmet research requirement to find novel therapeutic strategies for schizophrenia and other psychosis disorders.

This chapter aimed to investigate and characterise the incidence of EPSEs in patients with SMI diagnosis in SLAM dataset. A different EHR system, C&I Trust dataset used by this study for external validation, assessing differences between EHR recording and evaluate the portability of a published algorithm to identify and characterise the incidence of EPSEs. A tool was developed to identify EPSEs and was further generalised and evaluated using other common to rare ADEs in SLAM as Article I discusses. The data access to C&I NHS Foundation Trust was granted after the completion of the work presented in Article I, and by this time, a more enhanced tool equipped with improved rule-based, larger ADE dictionary and modified ConText algorithm, the ADEPt pipeline, was ready for evaluation (presented in chapter 3). Therefore, while Section 2.4 presents the evaluation of the original ADE detection tools to detect EPSEs in SLAM Trust, Section 2.5 uses the ADEPt pipeline to identify EPSEs in the C&I Trust. In both datasets, the ADE detection tools were used to identify positive mentions of EPSEs at any time in a patient record.

2.2 A Brief Overview of EPSEs

Extrapyramidal Side Effects or Extrapyramidal Syndromes (EPS) are a group of four-movement disorders, dystonia, akathisia, tardive dyskinesia and Parkinsonism. EPSEs include symptoms such as involuntary muscle contractions, tics, inner restlessness and tremors affecting parts of the body including the face, mouth, eyes, neck, trunk, pelvis and larynx. EPSEs are common side effects of first-generation antipsychotic medications. Although rarely life-threatening, EPSEs can result in debilitating effects, which can lead to social anxiety and embarrassment. Below is a description of each EPSE:

- Dystonia is an acute disorder that exhibits sustained contractions of muscles, twisting or repetitive movements affecting the neck, jaws, eyes, tongue and trunk. Dystonia is commonly caused by the first generation of antipsychotic medications, which acts on dopamine receptors (Macerollo et al., 2016).

Symptoms of dystonia are treated with anticholinergic medication such as benztropine (Bixler et al., 1987; Stern et al., 1979).

- Akathisia results in intense restlessness and reduced ability to remain motionless, causing the patient, severe discomfort (Serrano et al., 2017). Benzodiazepine is usually prescribed to the patients to treat the akathisia symptoms (Bixler et al., 1987; Lima et al., 2002).
- Parkinsonism is a series of symptoms resembling those developed in Parkinson's disease. Patients suffering from Parkinsonism tend to develop tremors and mask-like faces (Rochon et al., 2005; Savica et al., 2013; Thanvi et al., 2009). Antihistamine and anticholinergic drugs are usually prescribed to treat Parkinsonism symptoms (Katzenschlager et al., 2002).
- Tardive Dyskinesia is a condition that affects the nervous system, which results in symptoms such as repetitive and uncontrollable movements of the jaw and lips, twisting of finger and toes, rocking and jerking of trunks and hips (Dilip et al., 1982; KLAWANS JR, 1973; Schooler et al., 1982). Unfortunately, there is no known medication to treat tardive dyskinesia, forcing clinicians to discontinue first generation of antipsychotic treatments and starting a second-generation antipsychotic treatment such as Clozapine and Risperidone (Caroff et al., 2011; Caroff et al., 2018).

2.3 Identification of EPSEs in SLAM

Article I of this thesis describes the development of ADE detection tool and then the subsequent study of the EPSEs in the SLAM Trust.

Article I

2.4 Identification of Adverse Drug Events from free-text Electronic Patient Records and Information in a Large Mental Health Case Register

PLOS | ONE

RESEARCH ARTICLE

Identification of Adverse Drug Events from Free Text Electronic Patient Records and Information in a Large Mental Health Case Register

Ehtesham Iqbal[1☯], Robbie Mallah[2☯], Richard George Jackson[3,4,5], Michael Ball[3,4,5], Zina M. Ibrahim[1], Matthew Broadbent[4,5], Olubanke Dzahini[2], Robert Stewart[3,4,5], Caroline Johnston[1‡], Richard J. B. Dobson[1,4,5‡*]

1 MRC Social, Genetic & Developmental Psychiatry Centre (SGDP), King's College London, London, United Kingdom, **2** Pharmacy Department, South London and Maudsley NHS Foundation Trust, London, United Kingdom, **3** Department of Health Service & Population Research, Institute of Psychiatry, King's College London, London, United Kingdom, **4** NIHR Biomedical Research Centre for Mental Health, South London and Maudsley NHS Foundation, London, United Kingdom, **5** Biomedical Research Unit for Dementia, South London and Maudsley NHS Foundation, London, United Kingdom

☯ These authors contributed equally to this work.
‡ These authors also contributed equally to this work.
* richard.j.dobson@kcl.ac.uk

OPEN ACCESS

Citation: Iqbal E, Mallah R, Jackson RG, Ball M, Ibrahim ZM, Broadbent M, et al. (2015) Identification of Adverse Drug Events from Free Text Electronic Patient Records and Information in a Large Mental Health Case Register. PLoS ONE 10(8): e0134208. doi:10.1371/journal.pone.0134208

Editor: Christos A. Ouzounis, Hellas, GREECE

Received: June 24, 2014

Accepted: July 8, 2015

Published: August 14, 2015

Data Availability Statement: The terms dictionaries are available to the community at http://git.brc.iop.kcl.ac.uk/rmallah/dystoniaml/. The clinical records themselves are available subject to a collaborative agreement which adheres to strict patient led governance established at South London and The Maudsley NHS Foundation Trust. The data used in this work has been obtained from the Clinical Record Interactive Search (CRIS), a system which has been developed for use within the NIHR Mental Health Biomedical Research Centre and Dementia Unit (BRC/U) at the South London and Maudsley NHS Foundation Trust (SLaM). It provides authorised

Abstract

Objectives

Electronic healthcare records (EHRs) are a rich source of information, with huge potential for secondary research use. The aim of this study was to develop an application to identify instances of Adverse Drug Events (ADEs) from free text psychiatric EHRs.

Methods

We used the GATE Natural Language Processing (NLP) software to mine instances of ADEs from free text content within the Clinical Record Interactive Search (CRIS) system, a de-identified psychiatric case register developed at the South London and Maudsley NHS Foundation Trust, UK. The tool was built around a set of four movement disorders (extrapyramidal side effects [EPSEs]) related to antipsychotic therapy and rules were then generalised such that the tool could be applied to additional ADEs. We report the frequencies of recorded EPSEs in patients diagnosed with a Severe Mental Illness (SMI) and then report performance in identifying eight other unrelated ADEs.

Results

The tool identified EPSEs with >0.85 precision and >0.86 recall during testing. Akathisia was found to be the most prevalent EPSE overall and occurred in the Asian ethnic group with a frequency of 8.13%. The tool performed well when applied to most of the non-EPSEs

researchers with regulated access to anonymised information extracted from SLaM's electronic clinical records system. CRIS is governed by a strict information governance scheme which forbids anyone except for authorised researchers from accessing its records. The CRIS Oversight Committee is the main body which has implemented and currently manages these restrictions. Access to CRIS is restricted to 1) SLaM employees or 2) those having an honorary contract or letter of access from the trust. For further details, and to obtain an honorary research contract or letter of access, contact Adriana Fanigliulo in the R&D office: Adriana.fanigliulo@kcl.ac.uk. Once an honorary contract is established, researchers can only access CRIS once they submit a research project proposal through the CRIS Project Application form. The form is available here: http://www.slam.nhs.uk/about/core-facilities/cris/cris-project-application

Funding: This work was supported by the NIHR Biomedical Research Centre for Mental Health and Biomedical Research Unit for Dementia at the South London and Maudsley NHS Foundation Trust and Kings College London and a joint infrastructure grant from Guy's and St Thomas' Charity and the Maudsley Charity. This study was part funded by the National Institute for Health Research (NIHR) Mental Health Biomedical Research Centre.

Competing Interests: The authors have declared that no competing interests exist.

but least well when applied to rare conditions such as myocarditis, a condition that appears frequently in the text as a side effect warning to patients.

Conclusions

The developed tool allows us to accurately identify instances of a potential ADE from psychiatric EHRs. As such, we were able to study the prevalence of ADEs within subgroups of patients stratified by SMI diagnosis, gender, age and ethnicity. In addition we demonstrated the generalisability of the application to other ADE types by producing a high precision rate on a non-EPSE related set of ADE containing documents.

Availability

The application can be found at http://git.brc.iop.kcl.ac.uk/rmallah/dystoniaml.

Introduction

In the digital era many healthcare providers have transitioned from keeping paper copies of patient and prescription data to electronic records. Although the concept behind electronic health records (EHRs) was primarily to retain documentation of a patient's medical history, it is now apparent that these digital data sets represent a valuable resource for research. However, EHRs are optimised for day-to-day clinical use, not for research, resulting in data sets that are often unstructured, ill-defined and arduous to analyse at scale. Despite these challenges, a number of studies have made use of the rich data in EHRs to mine details relating to adverse drug events (ADEs) for example [1].

Adverse drug reactions (ADRs; ADEs where drug causality is established) are troublesome and potentially fatal outcomes of medication treatment and result in extra expense for health care providers. The ability to mine for, and eventually predict, occurrences of ADRs could have significant patient and cost benefits in the future [2]. A 2004 analysis of 18,820 patients showed that the projected annual costs for ADRs that led to hospital admissions would total 5466m [3]. In addition, a US study reported that there were 2341 ADR related deaths from data collected between 1999 and 2006. Annual mortality rates ranged from 0.08 to 0.12 per 100,000, increasing significantly over time at a rate of 0.0058 per year [4]. After the initial testing phase of a drug, spontaneous reporting systems, such as the UK Yellow Card Scheme, are the primary means for identifying suspected ADRs. These systems are reliant on patient and clinician data entry and many ADRs are under reported [5].

ADE Knowledge Discovery in Electronic Health Records

A number of studies have used text-mining techniques and natural language processing (NLP) tools in EHRs to identify ADEs and establish their causal relationships with drugs. Initially, to detect adverse events from clinical text, simple string matching approaches were applied.

Honigman et al (2001) [6] used notes from the outpatient department of Brigham and Women's Hospital (Boston, USA) to computationally identify ADEs. String matching was used to identify Micromedex M^2D_2 [7] medical data dictionary concepts in: ICD 9 diagnosis codes; patient drug allergy data; computer event monitoring (laboratory tests, prescription data) and free text clinical notes. Possible ADEs were subsequently manually reviewed. The study identified 864 possible ADEs. In a similar approach, Murff et al (2003) [8] investigated

adverse events resulting from medical management records rather than patient's underlying conditions. A computer-based string matching tool was applied to search free-text discharge summaries for trigger words, consisting of a broad range of adverse events. A manual review of the discharge summaries showed 44.8% (327 of 730) of the search term hits were true adverse events representing 131 ADEs. Field et al (2004) [9] conducted a study on patients aged over 65 to detect possible drug-related incidents and identified 1,523 ADEs during a one-year period, of which 421 (28%) were deemed preventable.

In each of these studies, the cohort sizes were limited because the approach required manual review of all results. More recently, studies have taken advantage of NLP tools that have come to replace simple string matching as a major method for detecting adverse events from clinical free text. The cohort sizes of subsequent studies increased, and some studies even applied NLP tools on the whole set of EHR. One such study was performed by Hazelhurst et al (2009) [10], whereby the researchers conducted a study of outpatients to identify vaccine related gastrointestinal adverse events. They used MediClass [11] (an automated classification system) and programmed it to identify vaccine related clinical concepts and linguistic structures used in clinical notes to extract vaccine related adverse events. After encoding the knowledge into MediClass, it detected 319 possible adverse events out of which 181 were true positives (determined upon manual review). However there were some limitations with the study. The manual review was conducted by the author rather than independent coders and the ICD 9 codes do not have good coverage for vaccine related ADEs.

Wang et al (2009) [12] conducted a study using notes from the inpatient department of Presbyterian Hospital, New York. They applied a modified version of the MedLEE NLP tool [13] and used MedDRA symptoms to detect adverse events from discharge summaries. The recall and precision were 75% and 31% and the application detected 132 ADE related to the seven medications. They went on to conduct another study [14] on the same EHR data source and ran MedLEE by applying filters (information extraction modules) to capture symptoms and adverse events caused by using medication during the course of hospitalisation. They applied regular and contextual filters in order to reduce the amount of confusing information. In the regular filter they avoided family history (mother suffered from ADE), past events (patient suffered from ADE last year) and negation (patients shows no signs of ADE). In the contextual filter they kept the clinical information where it was indicated that the drug was administered prior to the adverse event (i.e. establishing the correct time sequencing). Assessment showed that applying the filters improved recall (In Symptoms: from 0.85 to 0.90; ADE: from 0.43 to 0.75) and precision (In Symptoms: from 0.82 to 0.92; ADE: from 0.16 to 0.31).

In another study using the inpatient notes from the department of Presbyterian Hospital, New York, Haerian et al (2010) [15] used the MedLEE natural language processor with a filter that was built with expert knowledge on discharge summaries for patient with elevated creatine kinase serum. They investigated the ADE Rhabdomyolysis resulting from myopathy inducing medication and successfully identified 165 ADE with 96.7% correctly identified rate.

Finally, Eriksson et al. (2013) [16] described methods to develop an adverse event dictionary in Danish clinical narratives. They used Python libraries for NLTK and identified 35,477 unique possible ADEs in a Danish psychiatric hospital's EHR. Manual inspection was performed to validate the ADEs, resulting in precision of 89% and recall of 75%.

The aim of the study described here was to develop a generic natural language processing (NLP) tool for identifying adverse drug events (ADEs) from text fields in English-language mental healthcare records. We define an ADE as any event that could be an ADR; however, at this stage we did not attempt to establish causality from the record (e.g. relating to the agent potentially responsible) but instead simply sought to ascertain the symptom/event itself. The tool was initially built to identify the four key extrapyramidal side effects (EPSEs) associated

with antipsychotic treatment: dystonia, Parkinsonism, akathisia, and tardive dyskinesia. EPSEs are a group of movement disorders ranging from sustained contractions of the muscle, twisting or repetitive movements or abnormal postures in dystonia [17], an inner sensation of restlessness resulting in a patient being unable to remain motionless with akathisia. Parkinsonism, also called Parkinsonresulting in a patient being unable to remain motionless with akathisia[18]. Tardive dyskinesia, associated with long-term antipsychotic use, manifests as slow repetitive movements [19]. Although rarely life threatening, EPSEs can be debilitating leading to social
anxiety and embarrassment, as well as potentially causing non-adherence to medication regimes and risking relapse [20]. In addition, prophylaxis and treatment of EPSEs usually requires further pharmacotherapy and the potential for additional ADRs. Understanding them further and being able to assess the potential for exposure within specific groups is therefore an important challenge. In a second step, the tool was applied to an unrelated mix of rare and common ADRs, described within the Medical Dictionary for Regulatory Activities (MedDRA), an international medical terminology dictionary in wide clinical use. The performance in these 'unseen' ADEs was assessed for generalizability of the approach.

Methods

Data Source

The development of NLP software to detect ADEs was carried out in a large mental healthcare EHR data resource. The South London and Maudsley NHS Foundation Trust (SLaM) is the largest mental health provider in Europe serving a population of over 1.2 million residents from four London boroughs (Croydon, Lambeth, Lewisham and Southwark) [21]. The SLaM EHR, the Electronic Patient Journey System (EPJS), is typical of many such systems in that it stores much of its clinical records and prescribing information in an unstructured free text format. All use of data in our study is covered by a pre-existing ethical approval covering data analysis (Oxford C Research Ethics Committee, reference 08/H0606/71+5; renewed on 4.7.2013 for a further 5 years).

As of October 2012 there were over 200,000 patient records held in EPJS comprising over 20,000,000 free text documents including correspondence, discharge letters and events, increasing at a rate of 300,000 new documents per month. In order to create a resource for research, the Clinical Record Interactive Search System (CRIS) [22], a de-identified version of the EHR, was developed in 2007 and further enhanced with language processing tools to extract information from the vast amount of free text format data stored within this database.

Identification of Extrapyramidal Side Effects

We used the GATE (General Architecture for Text Engineering) NLP framework [23, 24] to develop an application to extract ADE information from free text fields over the whole of CRIS regardless of diagnosis. We trained the ADE tool on detecting EPSEs during its development. First, we defined a dictionary of EPSE ADE terms, including synonyms and alternate spellings. The application initially identifies all mentions of these terms as potential ADEs and then applies a series of rules to remove terms used in a different context. Removal rules can be overridden by 'retain rules' in cases where there is additional clear evidence that the word describes a real ADE. The process is illustrated in the flowchart in Fig 1.

Rules were defined using the Java Annotation Patterns Engine (JAPE) within GATE. Removal rules were written to handle cases where ADE terms were negated; in instances where clinicians were warning about, or monitoring for potential ADEs; names of charities or research organisations for ADEs; mentions of ADEs referring to a subject other than the

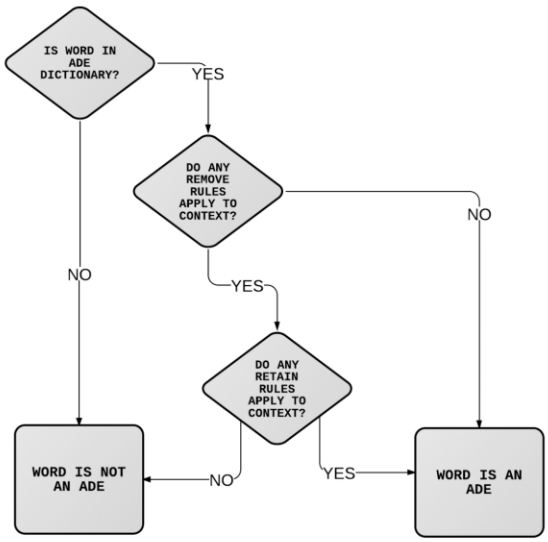

Fig 1. Remove and Retain rules. Flow diagram representing the use of the Remove and Retain rules to identify ADE instances.

doi:10.1371/journal.pone.0134208.g001

patient and cases where mentions indicated uncertainty in diagnosis. It was more important to ensure that identified ADEs were real than to ensure that all ADE mentions were identified, so these rules were developed favouring precision over recall. As ADEs are often mentioned multiple times in a patient record, a missed ADE in one document can be expected to be identified in another document, meaning recall may actually be higher than reported by the tool. To further improve recall we also defined a number of retain rules that could override removal rules when the context made it clear that ADE was present in the patient. Specifically we retained cases where the definite article or a possessive pronoun immediately preceded an ADE: e.g. 'The patient does not think the dystonia was painful'. We also defined a dictionary of commonly used diagnostic phrases that constitute strong evidence of a real ADE: e.g. be expect', We also defin'. Table 1 shows examples of text where Removal and Retain rules were required.

The development phase of the application started with the identification of dystonia ADEs and followed an iterative path whereby rules were developed, the application performance was tested and misclassifications were used to create new and improved rules. (See Fig 2). A different set of manually annotated documents were used for each round of testing. Table 2 shows JAPE rules that were implemented for the dystonia application and corresponding improvement in precision and recall, also shown in Fig 3. Once a plateau had been reached for dystonia, development continued for akathisia, Parkinsonism and tardive dyskinesia EPSEs.

Performance was assessed using a quality assurance (QA) tool built into the GATE software. Batches of 200 documents with the mention of the specific ADE contained within them (e.g. dystonia) were extracted from the CRIS database as XML files. The Text Hunter tool, developed in-house and available online, (http://sourceforge.net/projects/texthunter) was used to enable a clinical pharmacy technician to assign a positive or negative classification to each ADE mention. ADE mentions were classified as positive, even when they were clearly

Table 1. Examples of rule firing annotations. Rules were deployed within the JAPE of GATE.

Text where Removal rules are used	Text where Retain rules are used
ZZZZZ did not have dystonia'	'I asked ZZZZZ if her dystonia had become worsen
'ZZZZZ denied any dystonic reactions'	'I do not think his dystonia had become a probleme
'I discussed the possibility of dystonia with ZZZZZ'	'ZZZZZ denied her dystonia was severe'
'The Dystonia Society have always been a resource for patients'	I discussed the possibility of his dystonia being reduced'
'If Dystonia develops give procyclidine dose'	
'Check for any dystonic reaction'	
'ZZZZZ mother had developed dystonia many years ago'	

doi:10.1371/journal.pone.0134208.t001

indicating a past event. Performance was assessed by considering precision and recall of the application compared to manual annotation. Once the development process was complete, the application was run over all free text fields in CRIS on a high performance computer cluster hosted behind the SLaM NHS firewall, storing the results in a Microsoft SQL server instance for downstream analysis.

Prevalence of Extrapyramidal Side Effects in patients with serious mental illness

To demonstrate the utility of the approach, we investigated the frequency of EPSEs within the 17,995 patients represented on CRIS who had received a diagnosis of a serious mental illness (SMI; schizophrenia, bipolar disorder, schizoaffective disorder) [25] between 2007–2013 which

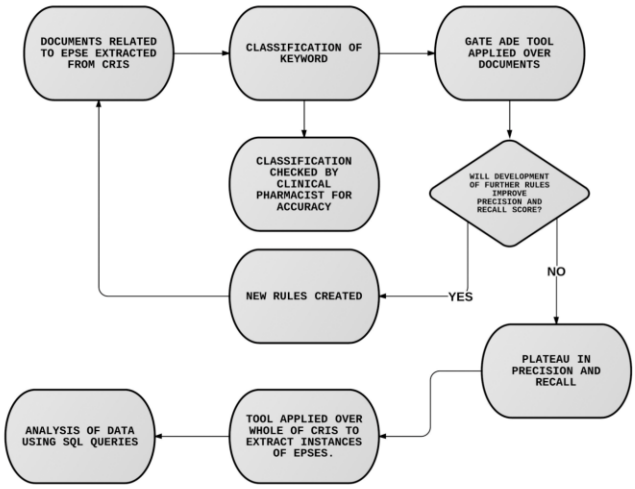

Fig 2. Iterative ADE tool development process. Flow diagram showing the iterative approach taken in development of the tool.

doi:10.1371/journal.pone.0134208.g002

Table 2. Corresponding precision and recall by applying each set of JAPE rules on dystonia corpus.

Stage	JAPE Rule	Purpose of rule	Precision	Recall
	Keyword Search only	Baseline estimates. Not used in iteration process.	62%	98%
1	Negation Before ADE terms	This rule negates the ADE terms which appear before the negation terms. The terms are present in the 'negationbefore' gazetteer.	71%	65%
2	Negation After ADE terms	This rule negates the ADE terms which appear after the negation terms. The terms are present in the 'negationafter' gazetteer.	73%	68%
3	Organisation	This rule negates the ADE terms appear before or after organisation terms (i.e. dystonia workshop). The terms are present in the 'organisation' gazetteer.	74%	68%
4	Symbols and punctuations	These rules negate the ADE terms if symbols and punctuations (like; '?', '/' and '@') appear before or after the ADE terms, with or without any space tokens.	76%	70%
5	Other Person	This rule negates the of ADE terms if other person is discussed in the sentence. The terms are present in the 'people' gazetteer.	77%	70%
6	Monitor	These 2 rules negate the ADE terms if clinicians are monitoring for ADE. The terms are present in the 'monitor' gazetteer.	78%	73%
7	Negative effects	These 7 rules negate the ADE terms where clinicians are explaining, informing, potential and common ADE to the patient. The negative effects terms are present in 'negseffect' gazetteer.	79%	76%
8	Single words around ADE	These 2 rules negates the ADE terms where words appears before the ADE (i.e. 'like', 'rates') and after the ADE (i.e. 'consider')	80%	78%
9	Diagnosis	This rule negate the ADE terms which are actually the diagnosis (i.e. ego dystonia) The terms are present in the 'diagnost' gazetteer.	83%	80%
10	Drugs Effects / Vaccine	These 6 rules negate the ADE terms where clinicians are discussing/explaining the side effects of a drug. The side effects terms are present in the 'druglink' gazetteer.	87%	82%
11	If statement	These 2 rules negate the ADE terms which are hypothetically discussed as a reaction/indication. The ADE indication terms are present in 'ADRin' gazetteer.	89%	85%
12	Retain Rule	These 3 rules un-negate the negation if ADEs terms are present within close proximity of patient name and ADE indication terms present in 'ADRin' gazetteer.	93%	89%

doi:10.1371/journal.pone.0134208.t002

included alive and deceased patients,. We removed deceased patients from this cohort (n = 2087). The diagnosis of each patient was selected as the most recent diagnosis recorded prior to 1st of January 2014. Diagnosis was assigned from a mandatory structured field in the clinical record, which records this information using ICD-10 categories and/or an in-house GATE application that mines text strings associated with diagnosis statements in clinical correspondence [26]. Mortality (an exclusion criterion) was ascertained through routine tracing of past and current cases on EPJS against the national register [27]. The prevalence of EPSEs across groups split by age (on 1st of January 2014), ethnic group, gender and SMI diagnosis subsets were tested using chi-squared tests.

Generic capability of the tool to identify adverse drug events

We tested the generic capability of the unmodified retain and remove rules developed within the tool by applying it to a range of ADEs unrelated to EPSEs but of interest in relation to the treatment of schizophrenia and bipolar disorder: alopecia, convulsions (seizures), hypersalivation, myocarditis, nausea, pneumonia, Stevens-Johnson syndrome, and tachycardia, chosen to represent a range from rare to common and mild to severe.

Annotator agreement

To explore and quantify reliability, a second manual annotator, a clinical pharmacist, independently classified the ADE mentions within a test corpora of documents for two EPSE ADEs, namely akathisia and dystonia, and two non-EPSE ADEs, alopecia and myocarditis. We rated the level of agreement between the two classifiers with a percentage score and Cohendystonia.

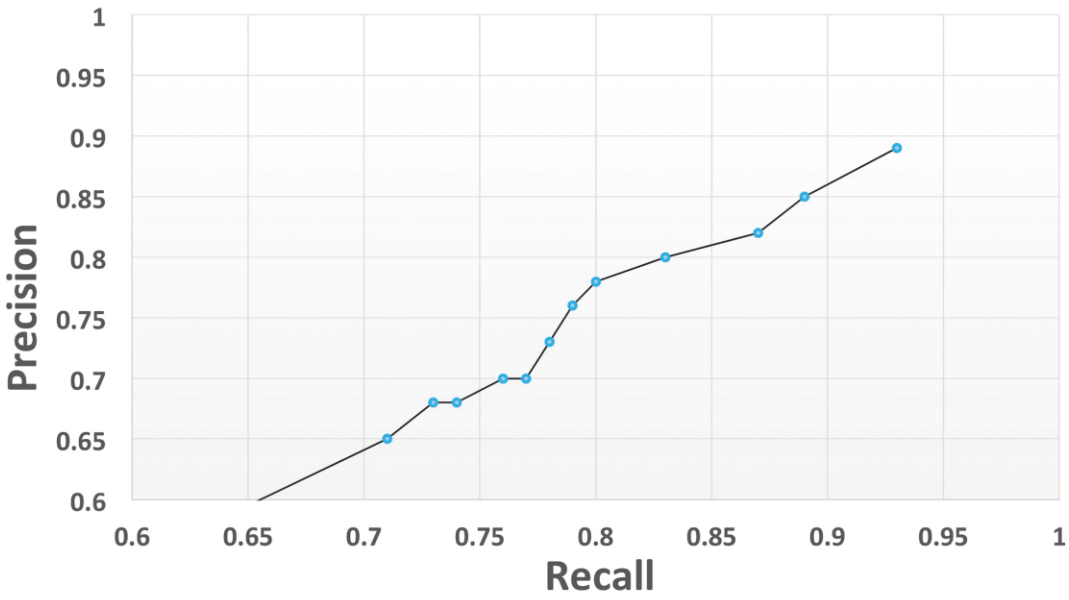

Fig 3. Precision and recall plot for dystonia JAPE rule development. The plot shows the evolution of the performance over the iterative JAPE rule development process for dystonia.

doi:10.1371/journal.pone.0134208.g003

Results

Identification of Extrapyramidal Side Effects

Performance metrics for the NLP applications in test corpora, following the iterative model building step, are displayed in Table 3. The application of the JAPE rules substantially improved precision over a keyword search term alone. Recall statistics were also maintained at satisfactory levels in most instances. The tool performed least well on the Parkinsonism EPSE, reaching a plateau of 0.85 precision and 0.88 recall. The other EPSEs returned precision scores of >0.90 and recall >0.86.

Table 3. Performance metrics for JAPE rules identifying extrapyramidal side-effects (EPSEs).

EPSE	Precision		Recall	
	Using keyword search only	With Remove and Retain rules applied	Using keyword search only	With Remove and Retain rules applied
Dystonia	0.62	0.93	0.98	0.89
Akathisia	0.61	0.92	>0.99	0.86
Parkinsonism	0.58	0.85	0.94	0.88
Tardive dyskinesia	0.89	0.97	>0.99	0.90

doi:10.1371/journal.pone.0134208.t003

Prevalence of Extrapyramidal Side Effects in patients with serious mental illness

Descriptive data on EPSE prevalence in patients with an SMI diagnosis are summarised in Table 4. Akathisia was the most frequently recorded EPSE in all groups. Significant heterogeneity was found for most comparisons although patterns of associations differed between the EPSEs. Akathisia showed no significant differences across the age groups. Dystonia was more commonly identified in younger compared to older patients, whereas the opposite was the case for Parkinsonism and tardive dyskinesia which appeared to be more prevalent in the older age groups with low incidence in the young. Men had higher incidence of recorded dystonia and akathisia than women but there were no gender differences in Parkinsonism or tardive dyskinesia. Considering ethnicity, akathisia and Parkinsonism were most frequently recorded in Asian groups and dystonia and tardive dyskinesia most prevalent in black groups. In terms of diagnosis, all EPSEs were lowest in bipolar patients and highest in schizoaffective disorder patients.

Generic capability of the tool to identify adverse drug events

Table 5 displays performance metrics for the NLP tool applied to non-EPSE ADEs. In summary, the tool performed well over most ADEs, but least well for myocarditis and Stevens-Johnson syndrome. Precision increased with the application of Remove and Retain rules compared to a keyword search.

Annotator agreement

Inter-annotator agreement statistics are summarised in Table 6 and range from 88% (Cohenble ement statistics are summarised in Table-EPSE AD.

Discussion

We describe the development of an NLP tool to identify ADEs within free text fields, such as case note entries and correspondence, from a large mental health EHR-derived database. Text processing rules were initially constructed to identify EPSEs and the distributions of these were described in a sample of patients with serious mental illness. The tool was trained initially on the four principal EPSEs associated with antipsychotic pharmacotherapy, but was developed with the aim of producing a generic set of rules that would be capable of identifying ADEs beyond EPSEs. With this in mind it was important not to over-train the application for EPSE identification specifically. As a result, we found the rules performed well in identifying a range of other ADEs.

A number of challenges were encountered in the development of the application. For example, of the EPSEs targeted, the tool performed least well in identifying Parkinsonism. This was probably because of a higher risk of false positive annotations due to Parkinsonism being mentioned in contexts unrelated to ADE instances (for example, because of Parkinson's disease itself). In general, many instances of potential ADEs were found to be ambiguous, potentially because of diagnostic uncertainty and/or clinical reluctance to record an ADE as definitive. Because of the priority we placed on precision over recall, where there was any doubt around an ADE diagnosis the instance was classified as negative. For this stage of development, the NLP application was designed simply to identify text indicative of a given ADE regardless of timing. Some of the recorded ADEs observed during the manual annotation process related to past instances and this should be considered when interpreting findings. Further development of the application is ongoing to enable future studies dependent on temporal relationships; for example, those investigating timing in relation to medication use.

Table 4. Recorded EPSE frequencies for patients with serious mental illness (SMI) according to demographic status and diagnosis. The numbers reflect a cohort of 12879 patients from 2007 to 2013.

EPSE Total Number of Patients		Dystonia 390	Akathisia 750	Parkinsonism 440	Tardive Dyskinesia 324
Age Group	**Cohort Size**	**EPSE Prevalence**			
Under 21	318	5.97%	8.18%	3.46%	0.63%
21 to 30	2106	4.51%	6.03%	2.71%	1.47%
31 to 40	3018	3.61%	5.40%	2.78%	1.46%
41 to 50	3249	2.65%	6.25%	2.22%	2.28%
51 to 60	2119	2.27%	5.85%	3.21%	2.41%
61 to 70	1129	1.86%	5.93%	6.20%	5.23%
71 to 80	677	1.33%	4.73%	9.31%	7.39%
Above 80	263	1.14%	3.04%	5.70%	4.94%
Chi-Square value (7 df) P-Value		49.568 <0.001	10.648 0.155	123.193 <0.001	10.648 <0.001
Gender	**Cohort Size**	**EPSE Prevalence**			
Male	6969	3.49%	6.50%	3.26%	2.55%
Female	5910	2.49%	5.03%	3.60%	2.47%
Chi-Square value (1 df) P-Value		10.881 <0.001	12.684 <0.001	1.165 0.280	0.092 0.762
Ethnicity	**Cohort Size**	**EPSE Prevalence**			
White	5788	2.32%	6.10%	3.27%	2.16%
Black	4682	4.44%	5.55%	3.55%	3.25%
Asians	861	2.32%	8.13%	6.04%	3.14%
Other	1548	1.81%	4.33%	2.13%	1.29%
Chi-Square value (3 df) P-Value		51.214 <0.001	16.088 <0.001	26.332 <0.001	23.990 <0.001
SMI	**Cohort Size**	**EPSE Prevalence**			
Schizophreniform	8411	3.11%	5.91%	3.40%	2.87%
Bipolar	3208	2.03%	3.99%	2.77%	1.00%
Schizoaffective	1260	5.00%	9.92%	5.16%	4.05%
Chi-Square value (2 df) P-Value		27.867 <0.001	58.342 <0.001	15.607 <0.001	46.399 <0.001

doi:10.1371/journal.pone.0134208.t004

Table 5. Performance metrics for JAPE rules identifying selected other (non-EPSE) adverse drug events (ADEs).

ADE	Precision		Recall	
	Using keyword search only	With Remove and Retain rules	Using keyword search only	With Remove and Retain rules
Alopecia	0.87	0.92	>0.99	0.76
Convulsions (seizures)	0.68	0.88	>0.99	0.81
Hypersalivation (sialorrhea)	0.91	0.95	>0.99	0.82
Myocarditis	0.34	0.64	>0.99	0.69
Nausea	0.68	0.96	0.96	0.74
Pneumonia	0.77	0.84	>0.99	0.82
Stevens-Johnson syndrome	0.29	0.59	>0.99	0.73
Tachycardia	0.85	0.93	>0.99	0.83

doi:10.1371/journal.pone.0134208.t005

Table 6. Inter-annotator agreement test results.

ADE	Agreement (%)	Cohen's Kappa Score
Akathisia	88%	0.74
Alopecia	92%	0.70
Dystonia	96%	0.91
Myocarditis	88%	0.69
Parkinsonian	90%	0.80

doi:10.1371/journal.pone.0134208.t006

Obtaining a good recall score on an ADE was reliant on a broad keyword selection within the gazetteer, incorporating as many descriptions of the ADE as possible. For example, there were a number of alternative spellings of akathisia in the source records (e.g. acasthisia, acathisia, akithisia) which required consideration when developing the gazetteer, and which will need further consideration when applying the tool over the wider MedDRA list of ADEs.

Lower precision and recall statistics were found for the more rare but serious ADEs. This tended to occur because they were more frequently cited in text fields as a warning rather than an occurrence. For example, myocarditis is a rare side effect of clozapine medication [28], but due to its severity, it was often mentioned as a potential consideration or as a recorded warning. These instances were classified as negative in the annotation process but it proved more challenging to produce Remove rules that would identify each one of these false positives.

Despite their importance in mental healthcare and psychopharmacology, EPSEs have been relatively understudied in naturalistic environments [29, 30]. As such this analysis demonstrates the power of secondary use of clinical records for research. However, these data have a number of caveats. Most importantly, the data reflect ADEs that are both recognized and recorded and thus are likely to underestimate the true situation, further reduced by the design of the NLP application to focus on unambiguous instances and ignore tentative terminology. EPSE recognition is also considered to be challenging at a clinical level: for example, the misdiagnosis of akathisia as agitation [31] or dystonia and akathisia as features of the underlying mental disorder [32]. Additionally, a study by Somers et al 2003 reported that spontaneous reporting by physicians and nurses on a geriatric ward revealed considerably fewer ADRs than a patient interview by a pharmacist [33]. However, in the absence of any other means of routine recording of these ADEs, our approach at least allows some scope for surveillance and targeted intervention.

Dystonia was more frequently recorded in the young and in males and reduced linearly with age, supporting previous findings [34]. Akathisia remained relatively consistent in recorded rates through the ages. We were unable to find any previous studies supporting age in being a significant factor in the development of akathisia. Prevalence of recorded Parkinsonism and tardive dyskinesia, on the other hand, display a progressive increase with age, with Parkinsonism displaying a slight dip in the 41–50 group. This increase through the ages is understandable as tardive dyskinesia is more associated with long-term antipsychotic use and Parkinsonism is more common in elderly females [35].

Recorded EPSEs varied noticeably in prevalence between ethnic groups. In particular, akathisia and Parkinsonism were more commonly recorded in patients of an Asian ethnicity whereas dystonia and tardive dyskinesia were more commonly recorded in patients of black ethnicity. There is some evidence that prescribing in psychiatry varies between ethnic groups. While this may reflect differences in hepatic metabolism of these drugs, variations in prescribing may also relate to prejudicial clinical practice [36]. Over 50% of Asian people have intermediate metabolism of cytochrome P450 2D6 subtype (CYP2D6), one of three important enzymes metabolising antidepressants, antipsychotics and benzodiazepines. Poor metabolism

of CYP2D6 leads to higher plasma levels of the drug in question with a consequently raised risk of developing EPSEs [36]. This may, in part, explain the higher recorded frequencies of akathisia and Parkinsonism within the Asian population in our cohort. Black people with a mental illness are more likely to be diagnosed with schizophrenia over non black people and have been found to be both more likely to receive a depot antipsychotic and to receive higher doses of these agents in a study based on SLaM patients in the 1990s [37]. However, more recent studies (based on the same patient population) did not find significant differences in antipsychotic type, dose or any other aspects of antipsychotic prescribing between black and white patients.[38]. Hencehe higher levels of recorded dystonia and tardive dyskinesia observed in black patients in our cohort cannot necessarily be explained by differences in antipsychotic prescribing and this point would require further investigation.

Gender was significantly associated with recorded rates of dystonia and akathisia but not Parkinsonism and tardive dyskinesia. We had expected higher recorded rates of Parkinsonism in females over males, in accordance with our literature findings. There were higher recorded rates of dystonia and akathisia in male patients over female. Male gender is a risk factor for development of dystonia and our results support this [39]. Risk factors for akathisia are not completely understood.

All EPSEs were differentially associated with SMI diagnosis, most noticeably schizophreniform and schizoaffective patients with increased rates of akathisia compared to bipolar patients. This is not surprising as antipsychotics are the most common treatment regimen for schizophreniform and schizoaffective [40], whereas bipolar disorder patients are typically treated with mood stabilisers such as lithium and valproate over antipsychotics [41].

Conclusion

As well as providing important and novel findings on EPSEs, the NLP tool we built demonstrates utility in wider ADE extraction. In the future we will extend and evaluate the tool across ADEs listed within MedDra, to develop and introduce supplementary applications to differentiate current from past events, and to incorporate the ADE application within wider CRIS NLP developments including ascertainment of pharmacotherapy in order to characterise further the profiles associated with higher risk.

The terms dictionaries are available to the community at http://git.brc.iop.kcl.ac.uk/rmallah/dystoniaml/. The records themselves are available subject to a collaborative agreement which adheres to strict patient led governance. We would encourage the community to make contact with the authors to establish a collaboration.

Author Contributions

Conceived and designed the experiments: RD CJ EI RM RJ M. Ball OD RS. Performed the experiments: RM EI. Analyzed the data: RM EI. Contributed reagents/materials/analysis tools: RJ. Wrote the paper: EI RM RD CJ M. Ball M. Broadbent RJ ZI OD RS. Won project funding: RD. Supervised the project: RD CJ.

References

1. Jensen PB, Jensen LJ, Brunak S. Mining electronic health records: towards better research applications and clinical care. Nat Rev Genet. 2012; 13(6):395–405. doi: 10.1038/Nrg3208. WOS:000304200800008. PMID: 22549152
2. Tatonetti NP, Ye PP, Daneshjou R, Altman RB. Data-driven prediction of drug effects and interactions. Science translational medicine. 2012; 4(125):125ra31. doi: 10.1126/scitranslmed.3003377 PMID: 22422992; PubMed Central PMCID: PMC3382018.

3. Pirmohamed M, James S, Meakin S, Green C, Scott AK, Walley TJ, et al. Adverse drug reactions as cause of admission to hospital: prospective analysis of 18 820 patients. Bmj. 2004; 329(7456):15–9. PMID: 15231615
4. Shepherd G, Mohorn P, Yacoub K, May DW. Adverse drug reaction deaths reported in United States vital statistics, 1999–2006. Annals of Pharmacotherapy. 2012; 46(2):169–75. doi: 10.1345/aph.1P592 PMID: 22253191
5. Hazell L, Shakir SA. Under-reporting of adverse drug reactions. Drug Safety. 2006; 29(5):385–96. PMID: 16689555
6. Honigman B, Lee J, Rothschild J, Light P, Pulling RM, Yu T, et al. Using computerized data to identify adverse drug events in outpatients. Journal of the American Medical Informatics Association. 2001; 8 (3):254–66. PMID: 11320070
7. Analytics TH. Micromedex Solutions [cited 2013 14, December]. Available: http://micromedex.com/.
8. Murff HJ, Forster AJ, Peterson JF, Fiskio JM, Heiman HL, Bates DW. Electronically screening discharge summaries for adverse medical events. Journal of the American Medical Informatics Association. 2003; 10(4):339–50. PMID: 12668691
9. Field TS, Gurwitz JH, Harrold LR, Rothschild JM, Debellis K, Seger AC, et al. Strategies for detecting adverse drug events among older persons in the ambulatory setting. Journal of the American Medical Informatics Association. 2004; 11(6):492–8. PMID: 15299000
10. Hazlehurst B, Naleway A, Mullooly J. Detecting possible vaccine adverse events in clinical notes of the electronic medical record. Vaccine. 2009; 27(14):2077–83. doi: 10.1016/j.vaccine.2009.01.105 PMID: 19428833
11. Hazlehurst B, Frost HR, Sittig DF, Stevens VJ. MediClass: A system for detecting and classifying encounter-based clinical events in any electronic medical record. Journal of the American Medical Informatics Association. 2005; 12(5):517–29. PMID: 15905485
12. Wang X, Hripcsak G, Markatou M, Friedman C. Active computerized pharmacovigilance using natural language processing, statistics, and electronic health records: a feasibility study. Journal of the American Medical Informatics Association. 2009; 16(3):328–37. doi: 10.1197/jamia.M3028 PMID: 19261932
13. University C. MedLingMap [cited 2013 08, December,]. Available: http://www.medlingmap.org/taxonomy/term/80.
14. Wang X, Chase H, Markatou M, Hripcsak G, Friedman C. Selecting information in electronic health records for knowledge acquisition. Journal of biomedical informatics. 2010; 43(4):595–601. doi: 10.1016/j.jbi.2010.03.011 PMID: 20362071
15. Haerian K, Varn D, Chase H, Vaidya S, Friedman C, editors. Electronic health record pharmacovigilance signal extraction: a semi-automated method for reduction of confounding applied to detection of rhabdomyolysis. DRUG SAFETY; 2010: S INT LTD 41 CENTORIAN DR, PRIVATE BAG 65901, MAIRANGI BAY, AUCKLAND 1311, NEW ZEALAND.
16. Eriksson R, Jensen PB, Frankild S, Jensen LJ, Brunak S. Dictionary construction and identification of possible adverse drug events in Danish clinical narrative text. Journal of the American Medical Informatics Association. 2013; 20(5):947–53. doi: 10.1136/amiajnl-2013-001708 PMID: 23703825
17. Sanger TD, Delgado MR, Gaebler-Spira D, Hallett M, Mink JW. Classification and definition of disorders causing hypertonia in childhood. Pediatrics. 2003; 111(1):e89–e97. PMID: 12509602
18. McDowell F, Lee JE, Swift T, Sweet RD, Ogsbury JS, Kessler JT. Treatment of Parkinson's syndrome with L dihydroxyphenylalanine (levodopa). Annals of internal medicine. 1970; 72(1):29–35. PMID: 5410397
19. Gerlach J, Casey D. Tardive dyskinesia. Acta Psychiatrica Scandinavica. 1988; 77(4):369–78. PMID: 2898870
20. Misdrahi D, Llorca P, Lancon C, Bayle F. [Compliance in schizophrenia: predictive factors, therapeutical considerations and research implications]. L'Encephale. 2001; 28(3 Pt 1):266–72.
21. Stewart R, Soremekun M, Perera G, Broadbent M, Callard F, Denis M, et al. The South London and Maudsley NHS Foundation Trust Biomedical Research Centre (SLAM BRC) case register: development and descriptive data. BMC psychiatry. 2009; 9:51. doi: 10.1186/1471-244X-9-51 PMID: 19674459; PubMed Central PMCID: PMC2736946.
22. Fernandes AC, Cloete D, Broadbent MT, Hayes RD, Chang C-K, Jackson RG, et al. Development and evaluation of a de-identification procedure for a case register sourced from mental health electronic records. BMC medical informatics and decision making. 2013; 13(1):71.
23. Cunningham H, Maynard D, Bontcheva K, Tablan V, editors. GATE: an architecture for development of robust HLT applications. Proceedings of the 40th annual meeting on association for computational linguistics; 2002: Association for Computational Linguistics.

24. Cunningham H, Tablan V, Roberts A, Bontcheva K. Getting more out of biomedical documents with GATE's full lifecycle open source text analytics. PLoS computational biology. 2013; 9(2):e1002854. doi: 10.1371/journal.pcbi.1002854 PMID: 23408875

25. Johnson DL. Overview of severe mental illness. Clinical Psychology Review. 1997; 17(3):247–57. PMID: 9160175

26. Wu C-Y, Chang C-K, Robson D, Jackson R, Chen S-J, Hayes RD, et al. Evaluation of smoking status identification using electronic health records and open-text information in a large mental health case register. PloS one. 2013; 8(9):e74262. doi: 10.1371/journal.pone.0074262 PMID: 24069288

27. Chang C-K, Hayes RD, Broadbent M, Fernandes AC, Lee W, Hotopf M, et al. All-cause mortality among people with serious mental illness (SMI), substance use disorders, and depressive disorders in southeast London: a cohort study. BMC psychiatry. 2010; 10(1):77.

28. Kakar P, Millar-Craig M, Kamaruddin H, Burn S, Loganathan S. Clozapine induced myocarditis: a rare but fatal complication. International journal of cardiology. 2006; 112(2):E5–E6. PMID: 16808983

29. Kerwin R, Millet B, Herman E, Banki CM, Lublin H, Pans M, et al. A multicentre, randomized, naturalistic, open-label study between aripiprazole and standard of care in the management of community-treated schizophrenic patients Schizophrenia Trial of Aripiprazole:(STAR) study. European Psychiatry. 2007; 22(7):433–43. PMID: 17555947

30. Alvarez E, Bobes J, Gómez J-C, Sacristán JA, Cañas F, Carrasco JL, et al. Safety of olanzapine versus conventional antipsychotics in the treatment of patients with acute schizophrenia. A naturalistic study. European neuropsychopharmacology. 2003; 13(1):39–48. PMID: 12480121

31. Dauner A, Blair D. Akathisia. When treatment creates a problem. Journal of psychosocial nursing and mental health services. 1990; 28(10):13–8. PMID: 1981080

32. Berna F, Timbolschi ID, Diemunsch P, Vidailhet P. Acute dystonia and akathisia following droperidol administration misdiagnosed as psychiatric disorders. J Anesth. 2013; 27(5):803–4. doi: 10.1007/s00540-013-1612-x. WOS:000325620400034. PMID: 23604817

33. Somers A, Petrovic M, Robays H, Bogaert M. Reporting adverse drug reactions on a geriatric ward: a pilot project. European journal of clinical pharmacology. 2003; 58(10):707–14. PMID: 12610749

34. van Harten PN, Hoek HW, Kahn RS. Acute dystonia induced by drug treatment. Bmj. 1999; 319 (7210):623–6. PMID: 10473482

35. Thanvi B, Treadwell S. Drug induced parkinsonism: a common cause of parkinsonism in older people. Postgraduate medical journal. 2009; 85(1004):322–6. doi: 10.1136/pgmj.2008.073312 PMID: 19528308

36. Connolly A. Race and prescribing. The Psychiatrist. 2010; 34(5):169–71.

37. Lloyd K, Moodley P. Psychotropic medication and ethnicity: an inpatient survey. Social psychiatry and psychiatric epidemiology. 1992; 27(2):95–101. PMID: 1594979

38. Connolly A, Rogers P and Taylor D. Antipsychotic prescribing quality and ethnicity—a study of hospitalized patients in south east London. J Psychopharmacol. 2007; 21(2):191–197. PMID: 17329299

39. KEEPERS GA, CASEY DE. Prediction of neuroleptic-induced dystonia. Journal of clinical psychopharmacology. 1987; 7(5):342–5. PMID: 2890672

40. Cascade E, Kalali AH, Buckley P. Treatment of schizoaffective disorder. Psychiatry (Edgmont). 2009; 6 (3):15.

41. Mitchell P, Parker G. Treatment of bipolar disorder. The Medical journal of Australia. 1991; 155(7):488–93. PMID: 1921822

2.5 Identification of Extrapyramidal Side Effects (EPSEs) in Camden & Islington NHS Foundation Trust (C&I)

2.5.1.1 Data source and EPSE cohort

The C&I CRIS is a de-identified version of the C&I EHR containing over 108,000 patient's records, of which approximately 23,000 were receiving ongoing care as of 2017 (Werbeloff et al., 2018). The data was obtained from the C&I CRIS system using the same approach described in Section 2.4 Article I of this Chapter. The C&I SMI cohort (n=4745) was defined where patients received an SMI diagnosis between January 2009 and July 2016. The diagnosis of a patient was retrieved from the most recent diagnosis in the records. Age was calculated as of the first of July 2016 and was further divided into eight distinct categories such as (Under 21, 21-30, 31-40, 41-50, 51-60,61-70, 71-80 and above 80). Gender and ethnicity were derived from the last entry recorded in CRIS. Ethnicity was divided into four major categories, White, Black, Asians and Others. Cases with an unavailable date of birth, ethnicity and gender, or those corresponding to deceased patients, were removed from the cohort.

2.6 Results

Table 2.1, presents the already published findings from SLAM (E. Iqbal et al., 2015) and Table 2.2, presents the findings from the C&I Trust of the prevalence of EPSEs stratified by age (calculated in SLAM as of 1st of January 2014 and C&I as of 1st of July 2016), gender, ethnic group and SMI diagnosis. The stratified groups were tested using chi-square. There was no information available for those under 21 years of age as C&I Trust does not treat children and adolescents. Further, a meta-analysis was performed to increase the statistical power and find the commonality in both datasets, as shown in Table 2.3.

Table 2.1: Recorded EPSE frequencies for patients with SMI diagnosis according to demographic status and diagnosis in SLAM (n=12879)

SLAM NHS Trust - SMI & EPSE Cohort from January 2007 to December 2013						
		SMI Cohort	EPSE's			
			Dystonia	Akathisia	Parkinsonian	Tardive Dyskinesia
Age Groups	Total Patients	12879	390	750	440	324
	Under 21	318	5.97%	8.18%	3.46%	0.63%
	21 to 30	2106	4.51%	6.03%	2.71%	1.47%
	31 to 40	3018	3.61%	5.40%	2.78%	1.46%
	41 to 50	3249	2.65%	6.25%	2.22%	2.28%
	51 to 60	2119	2.27%	5.85%	3.21%	2.41%
	61 to 70	1129	1.86%	5.93%	6.20%	5.23%
	71 to 80	677	1.33%	4.73%	9.31%	7.39%
	Above 80	263	1.14%	3.04%	5.70%	4.94%
	Chi-Square value (7 df) P-Value		49.568 <0.001	10.648 0.155	123.193 <0.001	10.648 <0.001
Gender	Male	6969	3.49%	6.50%	3.26%	2.55%
	Female	5910	2.49%	5.03%	3.60%	2.47%
	Chi-Square value (1 df) P-Value		10.881 <0.001	12.684 <0.001	1.165 0.280	0.092 0.762
Ethnicity	White	5788	2.32%	6.10%	3.27%	2.16%
	Black	4682	4.44%	5.55%	3.55%	3.25%
	Asians	861	2.32%	8.13%	6.04%	3.14%
	Other	1548	1.81%	4.33%	2.13%	1.29%
	Chi-Square value (3 df) P-Value		51.214 <0.001	16.088 <0.001	26.332 <0.001	23.990 <0.001
SMI	Schizophreniform	8411	3.11%	5.91%	3.40%	2.87%
	Bipolar	3208	2.03%	3.99%	2.77%	1.00%
	Schizoaffective	1260	5.00%	9.92%	5.16%	4.05%
	Chi-Square value (2 df) P-Value		27.867 <0.001	58.342 <0.001	15.607 <0.001	46.399 <0.001

EPSE = Extrapyramidal Side Effects; SMI = Severe Mental Illness (SMI); n sample size; SLAM = South London and Maudsley NHS Foundation Trust health record cohort; the mean difference is significant at the 0.001 level.

Table 2.2: Recorded EPSE frequencies for patients with severe mental illness (SMI) according to demographic status and diagnosis in C&I (n=4745)

C & I NHS Trust- SMI & EPSE Cohort from January 2009 to July 2016						
		SMI Cohort	EPSE's			
			Dystonia	Akathisia	Parkinsonian	Tardive Dyskinesia
Age Groups	Total Patients	4745	278	470	292	294
	Under 21	5				
	21 to 30	378	7.9%	7.4%	2.4%	2.4%
	31 to 40	933	8.4%	8.7%	3.6%	3.5%
	41 to 50	1163	6.4%	10.6%	3.8%	4.1%
	51 to 60	1047	5.0%	11.3%	4.3%	6.1%
	61 to 70	598	3.5%	11.5%	10.4%	9.9%
	71 to 80	378	5.0%	10.1%	17.2%	14.0%
	Above 80	243	1.6%	5.3%	13.6%	11.5%
	Chi-Square value (7 df) P-Value		30.167 <0.001	14.978 0.036	158.778 <0.001	95.269 <0.001
Gender	Male	2661	6.5%	10.1%	5.5%	5.8%
	Female	2084	5.0%	9.6%	7.0%	6.7%
	Chi-Square value (1 df) P-Value		5.028 0.024	0.282 0.595	4.159 0.041	1.741 0.187
Ethnicity	White	2939	4.7%	10.2%	6.5%	5.8%
	Black	1186	8.1%	8.7%	5.8%	8.0%
	Asian	360	6.9%	12.8%	5.3%	4.7%
	Other	260	6.9%	7.7%	4.6%	5.0%
	Chi-Square value (3 df) P-Value		18.846 <0.001	7.108 0.069	2.507 0.474	9.076 0.021
SMI	Schizophreniform	2842	6.6%	11.3%	6.2%	7.2%
	Bipolar	1333	4.3%	5.0%	5.2%	3.0%
	Schizoaffective	570	5.8%	14.4%	8.4%	8.4%
	Chi-Square value (2 df) P-Value		9.006 0.011	55.948 <0.001	7.279 0.026	33.687 <0.001

EPSE = Extrapyramidal Side Effects; SMI = Severe Mental Illness (SMI); n sample size; C&I = Camden and Islington NHS Foundation Trust health record cohort; the mean difference is significant at the 0.001 level.

Table 2.3: Meta-analysis from SLAM and C&I comparison group.

	Dystonia	Akathisia	Parkinsonian	Tardive Dyskinesia
Age Groups	33.74 <0.001	10.37 0.0345	30.24 <0.001	32.59 <0.001
Gender	20.84 <0.001	17.96 0.0012	8.92 0.0631	8.13 0.0871
Ethnicity	28.94 <0.001	18.74 <0.001	19.50 <0.001	24.45 <0.001
SMI	25.74 <0.001	33.47 <0.001	24.10 <0.001	23.22 <0.001

SLAM = South London and Maudsley NHS Foundation Trust health record; C&I = Camden and Islington NHS Foundation Trust health record; SMI = Severe Mental Illness (SMI); the mean difference is significant at the 0.001 level.

2.7 Discussion

In lieu of the earlier pipeline developed to detect ADEs presented in Article I of this chapter, the more powerful tool the ADEPt pipeline, which was fully developed by the time access to C&I data was granted, was used to identify EPSEs from the C&I EHR free-text documents. The defined data dictionary of EPSE terms, including synonyms and alternate spellings, were consistent across both Trusts. The rule-based approach was used in both NLP applications as they required a small training set; on the other hand, the ML methods required large datasets to train and improve. These improvements are often incremental, which reduces the algorithm portability and interpretability (van der Ploeg et al., 2014).

The EPSEs are generally understudied (Alvarez et al., 2003; Kerwin et al., 2007) and are the common side effects of the first-generation antipsychotics. Since the introduction of second-generation antipsychotics in 1990, it was widely accepted that the problem of EPSE has diminished. However, research shows that EPSEs are still prevalent with the use of currently available antipsychotics drugs (Rummel-Kluge, Komossa, Schwarz, Hunger, Schmid, Kissling, et al., 2010).

The distribution of data was consistent across both Trusts. Akathisia was the most frequently recorded EPSE. Dystonia was more commonly identified in younger patients as compared to older patients. Parkinsonian and tardive dyskinesia appeared to be more prevalent in the older age groups. The difference in gender group shows that males have

a higher incidence of dystonia and akathisia compared to females. In ethnicity, the Asian population had a higher prevalence of akathisia, and the Black population had a higher recorded prevalence in dystonia.

In both Trusts, some heterogeneity was found for most of the comparisons, although patterns of associations differed between the EPSEs. Age group was the only category where statistical results were consistent across both Trusts. The results show that there is a significant association between the distribution of age and dystonia, parkinsonian and tardive dyskinesia in both Trusts. Both Trusts showed no significant association in gender distribution of Parkinsonian and tardive dyskinesia. The SLAM dataset reports a significant association in gender distribution of dystonia and akathisia, while the C&I dataset does not. Dystonia was the only EPSE where C&I datasets showed a significant association with ethnicity. The SLAM results showed that ethnicity had a significant association with all four EPSEs. Both datasets showed that akathisia and tardive dyskinesia had significant associations with SMI diagnoses. SLAM reports that an SMI diagnosis had a significant association with all four EPSEs.

The results show that dystonia and akathisia are more prevalent in younger age groups. It is also previously reported that the rate of dystonia decreases with age (van Harten et al., 1999). Akathisia is often missed or misdiagnosed as agitation (Dauner et al., 1990), and is therefore under-represented in both datasets. Furthermore, dystonia and akathisia can be misdiagnosed as psychiatric disorders (Berna et al., 2013).

A meta-analysis (see table 2.3) was performed to find the commonality and increase statistical power. Age is significantly associated with dystonia, tardive dyskinesia and Parkinsonian. Gender is significantly associated with dystonia only. Ethnicity shows a significant association with dystonia, akathisia, Parkinsonian and tardive dyskinesia, while SMI diagnosis shows a significant association in dystonia, akathisia, Parkinsonian and tardive dyskinesia.

The study presents several strengths, including the NLP tools achieved an overall 0.89 precision and 0.88 recall in the internal validation in SLAM EHR and an overall 0.84 precision and 0.87 recall in external validation at C&I EHR. Higher precision and recall are essential in large scale phenotype studies in which misclassification of phenotypes can impact on the power of statistical analysis. Another strength of the study is the use of organised, structured data within the EHRs such as demographics and diagnostics codes

which have better data quality and consistency and potentially inform patient experience. Finally, the algorithm was applied in two separate EHRs in two distinct NHS Trusts and hence likely to be portable to other EHR systems.

The study has several limitations, such as recording culture of these EPSEs in different EHR systems. The study also excludes the cases where data is missing, such as demographic and diagnoses which may create selection bias. The small training and testing document set of EPSEs may also limit the algorithm portability but similar performance in the second EHR of the similar settings is reassuring. In addition, as it is an observational study, it cannot identify the causal relationship between the drugs and EPSEs, which is essential to rule out if these EPSEs are caused by a first or second generation of antipsychotic drugs. Finally, the findings of this study may not be replicated in a general health EHR or outpatient settings.

The EPSE analysis was carried out in SLAM and C&I NHS Foundation Trusts. The populations are similar, representing a mixed London background. However, recording and reporting different EPSEs may differ in the different EHR systems. That may lead to discrepancies in findings, as criteria for a subjective ADE may differ among healthcare providers, leading to Berkson bias.

The established method and NLP tool have a broader implication on the psychiatric and NLP community. The method presented in this study can be replicated in other psychiatric EHRs, which build confidence in findings. The consistent results from studies are more likely to represent a reliable claim to the new psychiatric community knowledge. As for the NLP community, researchers are building their work on top of other researchers work. There is a growing need for such information of established method, and experimental setup should be shared openly. Although sharing code and study protocols and parameters are good practices, recent studies show it is not as common as it should be (Mieskes et al., 2019).

The current tool was developed to identify EPSEs from free-text clinical narratives and further evaluated with a range of other ADEs. The temporality aspects of the current tools are limited. Going forward, the next logical step is to develop a tool with a range of essential factors such as context around the ADEs mentions such as if the ADE discussion is hypothetical or patient-specific, temporality features such as if the ADE

mention is current or historical, a range of ADEs related to psychotropic drugs and applicable to different EHR systems.

2.8 Conclusions

In summary, this chapter demonstrated the possibility of using NLP tools to take advantage of EHR data to detect EPSEs and validate the efficacy of an already established method in replicating its performance in other general or psychiatric health EHR datasets.

Chapter 3

3 Detecting Adverse Drug Reactions from Unstructured Psychiatric Clinical Text

3.1 Introduction

Psychotropic drugs are prone to cause ADEs even at the regular doses used to manage psychiatric disorders. These ADEs can lead to poor medication cohesion, social stigma, physical morbidity, noncompliance, and therapy discontinuation (Sengupta et al., 2011; Sridhar et al., 2016). ADE annotation in EHR free-text has been subject to many studies stretching over the past two decades. Several studies have described different pharmacovigilance tools identifying clinical concepts such as negation, temporality (e.g., recent or historical) and certainty, but they used commercial tools such as MedLee and MetaMap and often used in the general hospital settings (Haerian et al., 2012; Harpaz et al., 2010; Iyer et al., 2013; LePendu et al., 2013; Y. Li et al., 2014; X. Wang, Hripcsak, Markatou, et al., 2009). Other studies have also described NLP tools, but these tools are task-specific (mainly commercial) tools, making replication difficult and only limited to single-site settings (Q. Li et al., 2014; Popejoy et al., 2015; Rochefort et al., 2015; S. V. Wang et al., 2017).

The psychiatric health EHRs are investigative nature of treatment progress and contain more narrative text than general health EHRs. The psychiatric clinical text has an abundance of historical, hypothetical, and retrospective text that describe the possibility of an event such as ADE, symptoms and medication. In psychiatric settings, clinicians document the medical history and also the process of eliminating different diagnosis, medication and ADEs. Therefore, temporality and contextual characteristics surrounding these events are essential to role out negation, temporality and certainty of findings.

Hence, there is a need for an NLP tool that is publicly available and embeds negation, contextual information related to temporality (e.g., recent or historical),

hypothetical or who might be the experiencer (e.g., event-related to a patient or a family member). Furthermore, the tool should be scalable and adapted by different health settings (e.g., psychiatric and general), similar settings but different EHR systems, geographical region, reporting styles, and presents consistent results in diverse settings.

The work presented in Chapter 2 of this thesis (E. Iqbal et al., 2015), uses a generic rule-based NLP tool to identify patients who have experienced EPSEs in response to psychiatric treatment during the course of treatment, and therefore, did not identify all mentions of ADEs in a given corpus. This chapter aims to compile features of ADEs that can lead to the design of ML or a rule-based system for annotating ADE mentions in the psychiatric EHRs. The identified features are used to develop a rule-based annotation tool for processing free text psychiatric notes into semantically meaningful annotated knowledge that can be used to answer the onset of ADEs in psychiatric settings. The work presented in this chapter is the continuation of earlier work by employing specific features of psychiatric text to identify all mentions of ADEs into positive (i.e. a patient is suffering from an ADE) and negated (i.e. a patient is not suffering from and ADE). The paper presented in this chapter discussed the characteristics of psychiatric text and the composition of the annotation tool. Furthermore, the tool was applied to four ADRs associated with psychotropic drugs ranging from common to rare and from mild to severe.

Article II

3.2 ADEPt, a semantically-enriched pipeline for extracting adverse drug events from free-text electronic health records

PLOS ONE

RESEARCH ARTICLE

ADEPt, a semantically-enriched pipeline for extracting adverse drug events from free-text electronic health records

Ehtesham Iqbal[1], Robbie Mallah[2], Daniel Rhodes[1], Honghan Wu[1], Alvin Romero[3], Nynn Chang[3], Olubanke Dzahini[2], Chandra Pandey[1,4], Matthew Broadbent[4], Robert Stewart[4,5], Richard J. B. Dobson[1,4,6,7☯], Zina M. Ibrahim[1,4,6,7☯*]

1 The Department of Biostatistics and Health Informatics, Institute of Psychiatry, Psychology and Neuroscience, King's College London, De Crespigny Park, London, United Kingdom, **2** Pharmacy Department, South London and Maudsley NHS Foundation Trust, London, United Kingdom, **3** SLaM BioResource for Mental Health, South London and Maudsley NHS Foundation Trust and King's College London, London, United Kingdom, **4** NIHR Biomedical Research Centre for Mental Health, South London and Maudsley NHS Foundation, London, United Kingdom, Biomedical Research Unit for Dementia, South London and Maudsley NHS Foundation, London, United Kingdom, **5** Department of Health Service & Population Research, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom, **6** The Farr Institute of Health Informatics Research, UCL Institute of Health Informatics, University College London, London, United Kingdom, **7** NIHR Biomedical Research Centre, University College London Hospitals, London, United Kingdom

☯ These authors contributed equally to this work.
* zina.ibrahim@kcl.ac.uk (ZI); richard.j.dobson@kcl.ac.uk (RD)

Citation: Iqbal E, Mallah R, Rhodes D, Wu H, Romero A, Chang N, et al. (2017) ADEPt, a semantically-enriched pipeline for extracting adverse drug events from free-text electronic health records. PLoS ONE 12(11): e0187121. https://doi.org/10.1371/journal.pone.0187121

Editor: Tudor Groza, Garvan Institute of Medical Research, AUSTRALIA

Received: April 30, 2017

Accepted: October 13, 2017

Published: November 9, 2017

Data Availability Statement: The ethical approval to access CRIS data requires the data to be stored behind the hospital's firewall with access monitored by strict guidelines via a patient-led oversight and governance committee. It is due to this restriction that the data cannot be made available within the manuscript, supporting files or via a public repository. However, data access for research purposes is possible subject to approval from the oversight committee and can be initiated by contacting CRIS oversight committee (cris.administrator@slam.nhs.uk).

Abstract

Adverse drug events (ADEs) are unintended responses to medical treatment. They can greatly affect a patient's quality of life and present a substantial burden on healthcare. Although Electronic health records (EHRs) document a wealth of information relating to ADEs, they are frequently stored in the unstructured or semi-structured free-text narrative requiring Natural Language Processing (NLP) techniques to mine the relevant information. Here we present a rule-based ADE detection and classification pipeline built and tested on a large Psychiatric corpus comprising 264k patients using the de-identified EHRs of four UK-based psychiatric hospitals. The pipeline uses characteristics specific to Psychiatric EHRs to guide the annotation process, and distinguishes: a) the temporal value associated with the ADE mention (whether it is historical or present), b) the categorical value of the ADE (whether it is assertive, hypothetical, retrospective or a general discussion) and c) the implicit contextual value where the status of the ADE is deduced from surrounding indicators, rather than explicitly stated. We manually created the rulebase in collaboration with clinicians and pharmacists by studying ADE mentions in various types of clinical notes. We evaluated the open-source **A**dverse **D**rug **E**vent annotation **P**ipeline (ADEPt) using 19 ADEs specific to antipsychotics and antidepressants medication. The ADEs chosen vary in severity, regularity and persistence. The average F-measure and accuracy achieved by our tool across all tested ADEs were 0.83 and 0.83 respectively. In addition to annotation power, the ADEPT pipeline presents an improvement to the state of the art context-discerning algorithm, ConText.

Funding: This paper represents independent research funded by the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. This research was supported by researchers at the National Institute for Health Research University College London Hospitals Biomedical Research Centre, and by awards establishing the Farr Institute of Health Informatics Research at UCL Partners, from the Medical Research Council, Arthritis Research UK, British Heart Foundation, Cancer Research UK, Chief Scientist Office, Economic and Social Research Council, Engineering and Physical Sciences Research Council, National Institute for Health Research, National Institute for Social Care and Health Research, and Wellcome Trust (grant MR/K006584/1).

Competing interests: The authors have declared that no competing interests exist.

Introduction

The data available within EHRs is a potentially valuable resource of information describing patient treatment trajectories at low levels of resolution. However, along with the potential lies significant challenges because a notable portion of the EHR is in free-text form, making it necessary to deploy NLP tools to transform the unstructured text into semantically-meaningful annotated knowledge that can be subsequently used to aide in clinical decision making. As a result, the literature contains many efforts to detect and classify clinical named entities in EHR text [1, 2, 3, 4, 5, 6, 7] and resulting in several tools for information extraction from clinical notes including cTAKES [8], MedEx [9] and MetaMap [10]. Recognition of the importance and value of this task has resulted in the creation of a number of challenges by the *Informatics for Integrating Biology and the Bedside Centre* (i2b2) for clinical entity recognition and classification from free-text clinical records [11], including a psychiatry-specific challenge for extracting symptom severity [12].

The tools developed so far have been used to identify a variety of concepts including medications, symptoms, treatments, tests and dosages. Our interest lies in annotating and classifying a specific concept, namely ADEs, which represent troublesome and potentially fatal outcomes of medication treatment and incur substantial burdens on healthcare providers (with projections of annual ADE-related costs approaching £466m) [13]. We focus on annotating and classifying ADEs associated with antipsychotics, and antidepressants medications for two reasons: 1) In psychiatry, many of the factors leading to variations in individual susceptibility to ADEs remain unknown, making the knowledge mined from any potential tool of great value for research and drug evaluation purposes, and 2) Psychiatric EHRs tend to contain most of the ADE-related knowledge in free-text narratives, and therefore require an NLP annotation pipeline for extraction.

The last decade a number of studies have used NLP to identify adverse drug reactions (ADRs; ADEs where a causative relation with medication is established) of interest in free-text EHR documents [14, 15, 16, 17, 18, 19, 20, 21]. However, the NLP tools developed to obtain the results of these efforts have been study-specific, and at times using commercial tools, and therefore neither replicable nor publically available.

Moreover, ADR detection using Psychiatric EHRs exhibit unique challenges that distinguish them from similar tools operating on general hospital EHRs, Adverse Event spontaneous reporting services [22, 23, 24, 25] literature [26] and social media reports [27, 28]. Apart from the well-studied large amounts of redundancy characterising these notes [29], clinical text contains a plethora of hypothetical and retrospective text, historical discussions as well as text negating possible diagnoses and ADRs. This is a direct consequence of the EHRs being filled with not only direct clinical problems, but also detailed summaries of the patient activities, social and family matters, mood, general observations, discussions or warnings given to the patient about potential side effects. Therefore, modelling and identifying the context of an annotation is essential for correct classification.

In our previous work, we developed a rule-based NLP annotation system using manually-created domain expert rules to identify patients who had experienced Extrapyramidal side effects (Dystonia, Parkinsonism, Akathisia, and Tardive Dyskinesia) and adverse drug events (Alopecia, Convulsion, Hypersalivation, Myocarditis, Nausea, Pneumonia and Tachycardia). The system achieved an overall performance of >0.85 precision on these specific ADEs [30]. However, this work focused on identifying patients who had experienced one of the aforementioned ADEs during the course of treatment, rather than identifying all ADE mentions for a given patient and anchoring them to a specific point of time.

In this paper, we extend our previous work to create a rule-based framework which identifies and annotates temporally anchored mentions of all ADEs present in a given clinical text

corpus. The modular tool builds on the recommendations for concept extraction and classification resulting from the i2b2 Challenge [11] by 1) identifying ADE mentions, 2) classifying ADE mentions (as positive, implying that the ADE is present, or negative, implying that the ADE is absent) and 3) refining the classification using contextual indicators found in the clinical text.

Our tool comprises a multiphase pipeline targeting ADE-specific patterns in psychiatric clinical text. Our easily expandable dictionary currently houses the vocabulary needed to identify 66 common ADEs representing a comprehensive list of antidepressant and antipsychotic ADRs collated by our lead pharmacy partners whose ADRs are a major interest of ours.

Methods

Data source

We acquired data from the Clinical Record Interactive Search (CRIS) [31], a database containing a de-identified replica of the EHRs of four major London, UK-based psychiatric hospitals: 1) The Maudsley Hospital, 2) Bethlem Royal Hospital, 3) Lewisham Hospital and 4) Lambeth Hospital. Conjointly, the four hospitals make up the South London and The Maudsley NHS Foundation Trust, one of the largest mental health provider in Europe serving a population of over 1.2 million patients and storing much of their clinical records and prescribing information in unstructured free text format.

As of January 2017, CRIS contained over 264,000 patient records comprising around 24 million free text documents including correspondence, discharge summaries, events, ward progress notes, mental health care plan and mental state formulations. We extracted 8,321 documents and created 32 corpora, of which around 2,310 documents in four corpora were used for creating the rulebase and remaining around 6,011 documents in 28 corpora for testing, ranging from 130 to 475 documents in each corpus. We created separate corpora for creating the rulebase and testing purposes. The corpus size varies because creating a manual annotation on each mention of ADE is time-consuming and it heavily relied on the availability of clinicians and pharmacist. The size of the corpus left on the expert judgment of clinicians where they agreed the corpus have enough variety of ADEs mention and documents to make a suitable decision. The process we follow as:

1) we extracted all documents within CRIS containing at least one mention of one of 19 ADEs: agitation, akathisia, arrhythmia, galactorrhoea, nausea, myocarditis, cardiomyopathy, constipation, convulsions, diarrhoea, dizziness, dry mouth, hypersalivation, pneumonia, sedation, Steven Johnson syndrome (SJS), tachycardia and weight gain and 2) we further stratified the extracted documents based on the ADE terms mentioned in the document, document length (documents vary in length between a single line and multiple pages) and document types (e.g. discharge summaries, ward progress notes, local GP notes, etc.). The final subset of 8,321 was randomly selected such that the subset contains a variation of document types and lengths for every ADE term.

For creating the rulebase and verification, we manually annotated the 8,321 documents for all mentions of the 19 ADEs in consultation with two clinical and pharmacy researchers who identified and classified mentions of the ADEs. The 19 ADEs were chosen by the clinician and pharmacist represent a range of ADEs ranging from mild to severe, rare to common and short-term to persistent. The selection was based on evidence within the record itself where possible (rare and common), but additionally based on clinical judgement where this was more difficult to derive from the record itself. The level of agreement between the two annotators for all 19 ADEs is given in Table 1 with a percentage representing the agreement and a Cohen's Kappa scores before removing the 1% documents where the length of the free text was

Table 1. Annotation agreement between two clinical annotators. Annotations were retained as the labelled dataset for predictions if the experts annotators agree on the classifications of their mentions.

ADE	Agreement (%)	Cohen's Kappa Score
Agitation	88%	0.65
Akathisia	90%	0.75
Arrhythmia	89%	0.73
Cardiomyopathy	89%	0.76
Constipation	91%	0.78
Convulsions	98%	0.96
Diarrhoea	93%	0.84
Dizziness	93%	0.78
Dry Mouth	89%	0.71
Galactorrhea	92%	0.83
Hypersalivation	94%	0.74
Insomnia	92%	0.75
Myocarditis	89%	0.71
Nausea	85%	0.69
Pneumonia	93%	0.82
Sedation	96%	0.91
SJS	93%	0.82
Tachycardia	94%	0.83
Weight Gain	96%	0.90

https://doi.org/10.1371/journal.pone.0187121.t001

only a single word. In the case of inter-annotator discrepancies, the annotation was reviewed and if ambiguity remained, the document was replaced in the corpus.

Dictionary terms

To accommodate the diversity of writing styles and terminologies used in different hospitals and clinics, and between clinicians and carers, and to account for typographic errors, we compiled a dictionary containing the vocabulary to be used by the different pipeline components. The dictionary also contains variations of ADE-related terms. For example, sedation is a common side effect of antipsychotics. However, depending on context, there are several possibilities to describe sedation such as feeling *sleepy*, *drowsy*, *sleepiness*, *drowsiness*, *sedated*, and *somnolence*. The dictionaries are available for download from our github repository, https://github.com/KHP-Informatics/ADRApp/tree/master/application-resources/ADR.

1. **632 ADE terms**: The ability of the pipeline to recognise and classify additional ADEs is constrained by the ADE terms contained within its dictionary. Our dictionary currently accommodates 66 different ADEs related to antipsychotics and antidepressants, including synonyms and alternate spellings. The vocabulary recogonised by the pipeline is easily extensible with user-provided terms to accommodate additional ADE terms.
2. **2545 drug terms**: derived from the BNF drug dictionary [32], and expanded to include incorrect spellings and updated drug names to reflect coverage within a psychiatric setting (specifically anti-psychotics, anti-depressants, mood stabilisers, hypnotics and anxiolytics).
3. **208 helping terms**: the purpose of which is to indicate ADE occurrences (e.g. 'does have', 'developed' etc.) and include drug administration (e.g. 'taking', 'applying', 'using', 'administering' etc.), monitoring (e.g. 'assess for', 'monitor', 'screen for', 'signs of', 'watch for' etc.),

negative effects (e.g. 'side effect', 'adverse effect', 'SE', 'EPSE' etc.) as well as drug link terms (e.g. 'as it can be', 'as it may', 'if it cause', 'known for', 'may lead' etc.).

4. **660 contextual terms:** these include subject terms (e.g. 'mother', 'patient'), negation terms (e.g. 'does not'), hypothetical terms (e.g. 'if'), temporal terms (e.g. 'previously') and termination terms (e.g. 'however'). These terms are partly derived from the 342 terms used in the ConText algorithm [33], an algorithm used for discerning the context surrounding mentions of medical episodes to aid the classification process.

Populating the dictionaries. With the help of senior pharmacists, we compiled a list of expected ADRs associated with antipsychotics and antidepressants using the British National Formulary (BNF 68 at the time of research), the electronic medicines compendium (eMC) [34], the Maudsley Prescribing Guidelines 11th Edition and the Micromedex Healthcare database. In addition, a list of possible spelling variations and common alternative descriptions used to describe the ADR in clinical practice was also generated.

Development environment

We used the GATE NLP framework [35], which is a development environment for creating language engineering applications. GATE offers language-processing, information extraction and testing tools [36]. We used GATE's Java Annotation Patterns Engine (JAPE) to implement the rule base in all stages of the pipeline [37].

The **A**dverse **D**rug **E**vent annotation **P**ipeline (ADEPt) [38] is composed of four sequentially applied rule-based processing and annotation components. Each component is composed of a set of specialised rules that use co-location and position to correctly annotate ADEs. The overall annotation strategy of the pipeline rests on two observations we made by analysing the structure of the text containing ADE references within the anonymised EHRs:

Observation 1: Clinical text usually takes the form of short delimiter-separated clauses with each clause conveying information about a single ADE-related episode. Delimiters are usually periods, commas and semicolons (with periods being the most commonly used). Therefore, we can use delimiter-separated clauses where the annotations are located as boundaries for classification.

Observation 2: Clinical text contains a plethora of contextual indicators surrounding ADE mentions including hypothetical and retrospective text, historical discussions as well as text negating possible diagnoses and ADEs. Moreover, ADE-housing clauses can contain multiple contextual indicators. Therefore, a specialised context-discerning component which is capable of resolving conflicts among multiple contextual indicators is necessary to correctly classify ADE mentions.

A functional representation of the overall pipeline is shown in (Fig 1). ADEPt begins by using the GATE pre-processing resources to prepare an input document for annotation by tokenizing, splitting sentences and tagging parts of speech. For this component, we designed a rule base that specifically examines clause-level boundaries and splits clauses accordingly. The prepared document is then passed to an ADE-identification component which uses the identified boundaries to locate ADE-related terms, as defined in the dictionary, and produce an initial classification (positive or negative).

The third step comprises a set of rules that refine the annotated ADEs by discerning the context surrounding the identified annotation. These rules are triggered when contextual indicators are found in the clauses containing the ADE annotations. We used contextual references to distinguish the subject, the temporal value and the categorical nature of the annotation (i.e.

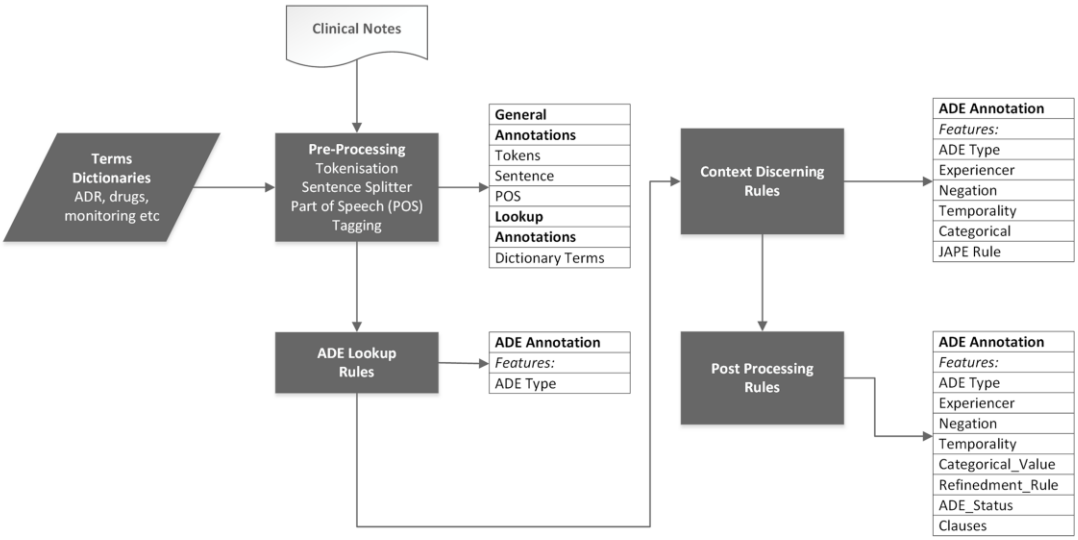

Fig 1. The GATE NLP based ADEPt pipeline comprising four rule-based processing components. The pipeline takes as input EHR clinical notes documents and a dictionary containing all annotation-related terms. The pipeline sequentially applies the four components accumulating new annotations for the target annotation (ADE). The output of the pipeline is a single ADE annotation with six features (ADE type, Experiencer, Negation, Temporality, Categorical_Value, Refinedment_Rule, ADE_status and clause).

https://doi.org/10.1371/journal.pone.0187121.g001

factual, hypothetical or negated) and to identify clauses where the ADE annotations refer to warnings, monitoring, suggestions and ongoing investigations, and implement specialised rules for resolving conflicts among multiple contextual indicators. The pipeline finalises the generated annotations in a post-processing step which attaches metadata summarising the decisions resulting in the final annotation and it's location in the corpus.

The ADEPt pipeline has two distinguishing features: 1) the use of clauses to delimit the annotation scope and 2) the use of ADE-related text properties to define the context surrounding the identified annotations.

Clause-level ADE annotation. Paragraph-level annotation does not offer sufficiently high resolution for our co-occurrence-based annotation system as multiple ADE-containing clauses usually exist in a single paragraph.

Examples: The examples below are adapted from real clinical notes of a single patient and are chosen to be representative of the approach we use to extracting ADEs from free text.

1. *Late shift ZZZZZ has wandered in and out of her room throughout the afternoon. No complaints of dizziness.*
2. *He did not complain of constipation. The patient is still suffering from a light headache.*

The above examples contain multiple-clause paragraphs whereby every clause is delimited by a period (i.e. every clause is a sentence). In the first example, the entire paragraph contains a single ADE mention. Co-location based annotation will examine negation terms co-located with the target ADE within the annotation boundaries and will arrive at the correct decision (a negated mention of dizziness) using both paragraph and sentence-level annotation. This is

because the negation term ('No complaints') is co-located with the ADE in the same paragraph (and sentence).

On the other hand, the second example shows a paragraph containing two ADE mentions (constipation and headache) in two clauses. If ADE annotation proceeds using paragraph boundaries, the negation term '*did not*' will lead to the rejection of both ADEs (constipation and headache). This incorrect annotation is also likely when using models based on 'bag-of-words' that make decisions dependent on the boundaries defined by the tagging algorithm. Bag-of-word models have in the past led to inconsistent tagging and require complex boundary-detection techniques that our clause-level tagging greatly simplifies. Here, sentence-level tagging will appropriately negate the first ADE (constipation) while affirming the second (headache).

The case of multiple ADEs in a single clause. Although clause-level parsing is sufficient in most cases encountered, clinical notes also contain instances of clauses containing multiple ADEs. Manual examination of 8,321 documents showed that approximately 5% of the clauses contained multiple ADE references. Therefore, in order to make correct decisions about ADEs, it is important to accommodate the instances which violate our single-clause, single-ADE assumption.

To approach these situations, we increased the granularity of the annotation process by dividing the clauses containing multiple ADEs into several single-ADE clauses. We did this by creating additional clause-splitting rules to issue splitting actions whenever specific termination terms (as per the dictionary) were present in multi-ADE clauses. For example, the termination term 'but' in the clause below divides it into two independent clauses with headache affirmed in the first and constipation negated in the second.

The patient is still complaining of headache but not constipation.

Discerning context. We designed a specialised component which uses contextual indicators co-located with the ADE terms to refine the value of the identified annotations. The component comprises a rulebase whose constituents are fired when a context-related trigger term which falls within the scope of an identified annotation is found. For example, the trigger term "no" is associated with a negation context. Once a trigger term is detected within the boundary of the annotation, the context value of the annotation is changed to reflect the value associated with the trigger term.

We built our rulebase by extending the open-source ConText algorithm [33], which identifies negations, hypothetical and general discussions, as well as indicators discerning the temporal validity of the identified annotations (whether they are current or historical). We adapted ConText to our ADE detection context in three ways. First, we customised the algorithm such that is uses the same clause-level boundaries as identified in the previous step. This way, all trigger terms which fall within the same clause as the annotation are captured by the algorithm and are directed to the appropriate rule for classification. Second, although ConText comes bundled with a set of triggering terms, many British English and ADE-relevant terms were not present in its dictionary. Therefore, we added 318 additional terms available in the GitHub repository [33, 38] to ConText's working dictionary and classified them according to how they affect the identification of an ADE as Table 2 shows. Categories include negation indicators (e.g. 'excluded' and 'not found'), possibility phrases (e.g. 'most concerned about' and 'rule him out for'), experiencer terms (e.g. 'mother' and 'father'), and temporal and hypothetical indicators (e.g. 'in the past' and 'if'). This is in addition to the termination terms discussed in the previous section.

Table 2. Enrichment of the ConText algorithm trigger terms.

Triggers Terms	ConText Algorithm (n = 392)	Terms Added (n = 318)	Current Terms (n = 710)
Experiencer	29	46	75
Negation	197	216	413
Possibility terms & phrases	28	16	44
Termination	89	17	106
Temporality & Hypothetical	49	23	72

https://doi.org/10.1371/journal.pone.0187121.t002

Finally, although the ConText algorithm is able to correctly classify annotations based on the surrounding context in most instances, there are situations where it is unable to do so. We therefore created an additional set of 26 'removal' and 9 'retention' rules to identify and correctly annotate ADE mentions whose surrounding contextual indicators cannot be properly interpreted by the ConText algorithm. *Retention* rules target annotations that ConText misclassified as negative, and retain them as positive mentions in ADEPT's results, while *removal* rules target annotations that ConText misclassified as positive, marking them as negative mentions in ADEPt's final annotations. Removal rules are overridden by retention rules in cases where there is additional evidence that the discussion of the ADE is positive. 26 removal and 9 retention rules are applied to correct the annotation in the following cases:

1. Retention Rules: The main aim of retention rules is to identify annotations that are surrounded by negation terms but are nevertheless positive mentions. In the ConText algorithm, a negation term is assigned a high priority and will lead to a negative annotation regardless of the existence of additional terms. However, this overgeneralization fails in many instances in clinical text, as negation may not be used to refer to the patient as given in the examples below:

ZZZZZ appeared to be disorientated and not taking his medication.

ZZZZZ restlessness has not worsened on the increased dose of beta-blocker.

The corresponding rules are shown in Fig 2(A) and 2(B). In the figure, the term 'Token' refers to any token not present in our dictionary.

2. Removal Rules: These rules operate on a number of difference clauses, mainly: a) clauses discussing potential ADEs, ongoing investigations, warning, monitoring or explanations, etc. as in Fig 3(A), 3(B) and 3(C) (e.g. I am changing the dose and have warned ZZZZ of dizziness, or Signs of myocarditis, on going investigation or The patient is starting Olanzapine, explained her as it can cause weight gain), b) clauses where there is uncertainty about whether the ADE is present as in Fig 3(D) (e.g. She has had 4 seizures within the last 2 weeks, unstable partial complex seizures), c) questionnaires, which tend to be pervasive in the electronic text as in Fig 3(E) and 3(F) (e.g. Fainting/ dizziness *No * Yes. I become irritable, restless and nervous x 5) and finally d) ADEs that are part of organisational names or addresses Fig 3(G) and 3(H) (e.g. CENTRE FOR ANXIETY DISORDERS AND TRAUMA, sjs@sydenham.lewisham.sch.uk or www.nhs.uk/Conditions/Anxiety).

Fig 2. The retention rules pattern used in the ADEPt pipeline.

https://doi.org/10.1371/journal.pone.0187121.g002

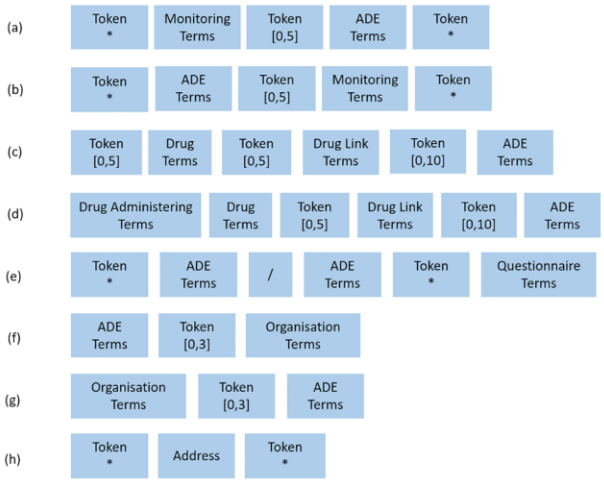

Fig 3. The removal rules pattern used in the GATE NLP based ADEPt pipeline.

https://doi.org/10.1371/journal.pone.0187121.g003

Ruleset building and testing. From the 8,321documents used in this work, we used 2,310 documents uniformly distributed across the four hospital sites to guide the construction of the rulebase. All documents chosen contain mentions of one of the following four ADEs: Akathisia (common and long-term), galactorrhoea (rare, acute and mild), myocarditis (rare and severe) and nausea (acute and common). The four ADEs chosen for rulebase vary in severity, regularity and persistence. The remaining 6,011 documents were used to test the performance of the pipeline on fourteen ADEs. The ADEs tested include the four used for rulebase in addition to dizziness, hypersalivation, pneumonia, sedation, tachycardia, cardiomyopathy, convulsions, diarrhoea, constipation and Steven Johnson syndrome.

Evaluation metrics. We used accuracy, precision, recall and the F-score to evaluate the annotation pipeline (Equations 1, 2, 3 and 4). We also recorded specificity, using it along with recall to examine the shapes of the resulting ROC curves. The metrics rely on true positive (TP), false positive (FP), true negative (TN) and false negative (FN) values, which we defined in terms of the agreement between our pipeline and our annotators consensus for every ADE term found. For example, a true positive annotation is one which, given all associated contextual terms surrounding the annotation, is identified as positive by ADEPt as well as the human annotator. TP is where the subject is the patient, no negation is present and the event is deemed recent. If the match is partial, the ADR is labelled as TN.

Error analysis

We conducted error analysis at each step of the pipeline to enhance performance. The resulting measures are given in Table 3 and include precision, sensitivity, specificity, accuracy and f-measure. These measures are given for each ADR at the following stages: a) 'paragraph', designating the use of paragraphs as delimiters for ADR mentions, b) 'statement', designating the use of statements to delimit ADR mentions, c) 'original ConText Algorithms' designating the use of the off-the-shelf ConText algorithm without modification, d) 'original ConText Algorithms' evaluates the impact of additional vocabulary to the dictionary, and e) 'With Extra

Table 3. Incremental results of akathisia, galactorrhea, nausea and myocarditis.

ADE	Corpus Ref	Total	Precision	Sensitivity	Specificity	Accuracy	F-measure
Akathisia	Paragraph	215	0.73	0.87	0.33	0.69	0.80
	Statement	215	0.76	0.89	0.39	0.73	0.82
	Original ConText Algorithm	215	0.77	0.90	0.43	0.74	0.83
	ConText With extra terms	215	0.94	0.88	0.88	0.87	0.91
	With Extra Terms and Refinement Rules	215	0.96	0.90	0.93	0.91	0.93
Nausea	Paragraph	369	0.84	0.82	0.52	0.74	0.83
	Statement	369	0.86	0.84	0.56	0.77	0.85
	Original ConText Algorithm	369	0.89	0.87	0.67	0.82	0.88
	ConText With extra terms	369	0.93	0.87	0.80	0.85	0.90
	With Extra Terms and Refinement Rules	369	0.95	0.86	0.84	0.85	0.90
Galactorrhea	Paragraph	139	0.59	0.72	0.41	0.57	0.65
	Statement	139	0.62	0.77	0.45	0.62	0.69
	Original ConText	139	0.66	0.81	0.50	0.66	0.73
	ConText With extra terms	139	0.73	0.89	0.61	0.76	0.80
	With Extra Terms and Refinement Rules	139	0.83	0.91	0.78	0.84	0.87
Myocarditis	Paragraph	188	0.28	0.70	0.30	0.41	0.40
	Statement	188	0.29	0.72	0.32	0.43	0.42
	Original ConText Algorithm	188	0.30	0.74	0.34	0.45	0.43
	ConText With extra terms	188	0.40	0.60	0.64	0.63	0.48
	With Extra Terms and Refinement Rules	188	0.51	0.75	0.71	0.72	0.61

https://doi.org/10.1371/journal.pone.0187121.t003

Terms and Refinement Rules' shows the final performance of the pipeline. Guided by the results obtained at each stage, we a) added more vocabulary to accommodate the unidentified ADRs, and b) added and adjusted the refinement rules to create a more generic rulebase.

Results

Here, we present the results of running the pipeline on the manually-annotated 8,321 test set documents for all 19 ADEs. For four ADEs (akathisia, galactorrhoea, nausea and myocarditis), we additionally report the incremental performance increase after each step of the pipeline (Table 3 and Fig 4).

The first and second rows of every ADE in Table 3 show the improvement resulting from using clauses (as opposed to paragraphs) to delimit the annotation scope. Here, we manually examined the TP, TN, FP and FN annotations using clause-level and paragraph-level boundaries and confirmed that: a) none of the correctly-classified ADEs using paragraph boundaries were misclassified using clause boundaries, b) all correctly-classified ADEs using paragraph-level boundaries are proper subsets of those identified by clause-level boundaries.

The last three rows of each ADE show the incremental improvement in classification using contextual indicators found in the text via a) the off-the-shelf ConText algorithm, b) ConText enriched with additional domain-specific terms and finally c) the improved ConText algorithm containing the enriched vocabulary as well as additional refinement rules for conflict resolution and implicit mentions. The incremental improvements are graphically demonstrated using the ROC curves in Fig 4.

Apart from Akathisia, Nausea, Galactorrhea and Myocarditis, we further evaluated ADEPt on fifteen other ADEs (see Table 4). In contrast to our earlier work, the ADEPt pipeline performed well on common ADEs such as constipation, diarrhoea, sedation, hypersalivation,

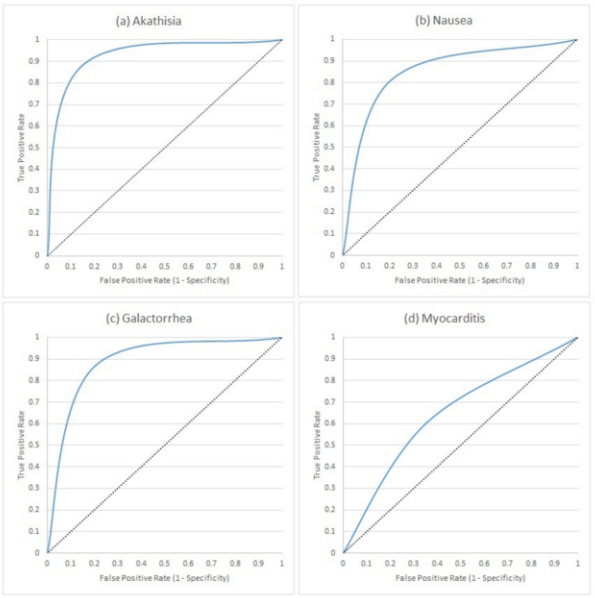

Fig 4. Receiver operating characteristic curves representing the performance of the ADEPt pipeline in identifying akathisia, nausea galactorrhea and myocarditis ADEs from free text. The increments in each graph correspond to 1) our previous work [30], 2) using paragraph boundaries, 3) using clause-boundaries, 4) using unrefined (off-the-shelf) ConText algorithm, 5) adding domain-specific vocabulary to ConText and 6) final refined ConText algorithm.

https://doi.org/10.1371/journal.pone.0187121.g004

tachycardia, pneumonia, sedation and performed least well in identifying convulsions. ADEPt also performed well on rare ADEs such as cardiomyopathy and Steven Johnson syndrome.

The pipeline achieves better performance in common and long-term ADEs than it does with rare and acute ADEs. This directly reflects the number of mentions of the corresponding ADEs in the clinical notes, which affects the number of patterns detected upon manual examination of the 2,310 documents we used to guide the creation of the rules. This variation in performance is also visible upon examining the ROC curves of Fig 4. A final observation from Table 4 is that some rare ADEs (SJS and cardiomyopathy) appear to have more mentions in the clinical notes than common ADEs (when examining the second column of the table), which may appear counterintuitive. However, this increased count is explained by the fact that these two ADEs are severe, driving clinicians to document any warnings or monitoring performed for them, and resulting a large number of negative mentions of these ADEs (as evident by the third column).

Discussion

We created a multi-phase rule-based pipeline for the recognition and classification of named ADEs in free-text psychiatric EHRs. The rulebase was created by manually analysing 2,310 of these documents in collaboration with domain experts to identify patterns of ADE mentions and related contextual text. We constructed the rulebase based on four ADEs (akathisia, nausea, myocarditis and galactorrhea) and evaluated its performance using these four ADEs as

Table 4. Results showing the performance of the ADEPt pipeline in identifying a selection of rare to common ADEs related to antipsychotics and antidepressants drugs.

ADE	Total	TP	TN	Precision	Sensitivity	Specificity	Accuracy	F-measure
Agitation	221	142	32	0.89	0.83	0.65	0.79	0.86
Akathisia	215	132	64	0.96	0.90	0.93	0.91	0.93
Arrhythmia	232	129	61	0.88	0.85	0.77	0.82	0.86
Cardiomyopathy	204	55	109	0.79	0.68	0.88	0.80	0.73
Constipation	475	315	99	0.91	0.91	0.76	0.87	0.91
Convulsions	148	84	37	0.92	0.81	0.84	0.82	0.86
Diarrhoea	221	140	55	0.93	0.90	0.83	0.88	0.92
Dizziness	234	130	96	0.94	0.83	0.90	0.85	0.88
Dry Mouth	211	124	56	0.91	0.87	0.82	0.85	0.89
Galactorrhea	139	68	50	0.83	0.91	0.78	0.85	0.87
Hypersalivation	193	161	18	0.97	0.95	0.78	0.93	0.96
Insomnia	189	119	38	0.84	0.93	0.62	0.83	0.88
Myocarditis	188	40	96	0.51	0.75	0.71	0.72	0.61
Nausea	369	241	75	0.95	0.86	0.88	0.90	0.92
Pneumonia	173	81	75	0.89	0.94	0.93	0.90	0.92
Sedation	189	108	54	0.89	0.89	0.81	0.86	0.89
Stephen Johnson's Syndrome (SJS)	333	68	205	0.60	0.88	0.82	0.82	0.72
Tachycardia	230	192	13	0.96	0.91	0.65	0.89	0.94
Weight Gain	209	108	51	0.92	0.87	0.82	0.86	0.90

https://doi.org/10.1371/journal.pone.0187121.t004

well as 15 ADEs related to antipsychotics and antidepressants drugs in 6,011 unseen clinical text documents by comparing with manual annotations by clinical researchers.

The clinical text contains many surface forms for any ADE mention, necessitating a unifying dictionary to guide the annotation ADE process. In collaboration with pharmacists and clinicians based at the South London and Maudsley NHS Foundation Trust, London, United Kingdom, we compiled a list of 66 ADEs with corresponding surface forms, misspellings and abbreviations. In addition to compiling agreed-upon medical terms from existing resources, the process involved large-scale manual examination as many ADE-related terms are implied rather than explicitly defined, e.g. a clinician may record 'the patient cannot fit in his/her clothes' in lieu of explicitly documenting weight gain.

Overall, the tool performs well compared to general NLP entity recognition systems, specifically the Context algorithm which we used for comparison. However, the performance varied depending on the regularity and persistence of the ADE under investigation. There is still a need to improve the context discerning rules for rare ADEs, which are usually discussed as possibilities in the clinical notes as clinicians usually take a lot of care before ruling out the possibility of a rare ADE and will administer multiple tests for the patients to go through. The current application does not have a high coverage for all the precautionary patterns surrounding multiple ADEs, an issue to address in our current work.

The annotations generated by the tool were compared to the manually-annotated documents prepared by domain experts. In this work, we only trained and tested ADEPt using annotations where the expert annotators agree on the classification of the ADE mention (i.e. whether it is a positive or a negative mention). It would be interesting to see whether the annotations that confused the human annotators (ones where the two experts disagree) will similarly confuse ADEPt. Therefore, part of our ongoing work is to add a third classification category corresponding to ambiguous annotations.

Additional ongoing work includes improving the context-discerning rules to distinguish rare from common ADEs. Moreover, we are investigating the merit of a hybrid approach which learns the rules and examine the effect on performance. Our ongoing work also focuses on linking ADE annotations obtained through ADEPt with medication and prescription information, to create a timeline establishing the associations between ADEs and medication episodes as well as possible drug-drug interactions.

All of the ongoing efforts aim at using the annotated knowledge mined by ADEPt to uncover unknown causal links between administered medications and the undesirable events, therefore making the distinction between ADEs we mine from the clinical text and non-preventable Adverse Drug Reactions (ADRs) which are caused by the medication itself and not due to mismanagement or clinical errors [39, 40].

Limitation

There are a few aspects of the ADEPt pipeline on which we are currently working. One of the issues we faced since the inception of our work is the difficulty of finding experts to annotate the documents and evaluate the annotation results against the annotated documents. As a result, the annotation power of ADEPt has only been evaluated on the 19 ADRs mentioned throughout the manuscript. However, we tried to minimise the effect of the small number of ADRs by selecting those which vary in terms of severity as well as frequency of onset in order to reflect the variations of mentions in the clinical text. Another consequence of the difficulty of finding annotators is that we have used four different annotators throughout the different stages of pipeline development, which may have induced unknown discrepancies in some of the cases.

Another limitation of the ADEPt pipeline is due to the limited number of clinical notes discussing the onset of rare ADEs such as SJS and myocarditis. Due to their rarity (as well as severity), discussions of these ADEs is usually done in the context of potential onset (negative mentions), rather than positive mentions referring to the patient herself. However, for these ADEs, ADEPt achieves better and both precision and sensitivity have improved in SJS simultaneously (0.60 and 0.88) and sensitivity in myocarditis (0.75).

Finally, ADEPt is developed and tested on SLaM's psychiatric clinical notes. Work evaluating its performance on other general or psychiatric notes is currently part of our ongoing work.

Conclusion

The tool described here demonstrates the ability to identify antipsychotics and antidepressants related ADEs from within the free text of psychiatric EHRs. By surfacing ADEs within the routinely collected EHR, we unlock a treasure trove of hitherto inaccessible data describing treatment response that is the first step to tailoring treatment through patient stratification leading to opportunities for novel interventions and studies into the genetic underpinnings of ADEs. The tool is freely available from our online repository at: https://github.com/KHP-Informatics/ADRApp

Acknowledgments

This paper represents independent research funded by the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. This research was supported by researchers at the National Institute for Health Research University College London Hospitals Biomedical Research Centre, and by awards establishing the Farr Institute of Health Informatics Research

at UCL Partners, from the Medical Research Council, Arthritis Research UK, British Heart Foundation, Cancer Research UK, Chief Scientist Office, Economic and Social Research Council, Engineering and Physical Sciences Research Council, National Institute for Health Research, National Institute for Social Care and Health Research, and Wellcome Trust (grant MR/K006584/1).

Author Contributions

Conceptualization: Ehtesham Iqbal, Robert Stewart, Richard J. B. Dobson, Zina M. Ibrahim.

Data curation: Ehtesham Iqbal, Robbie Mallah, Alvin Romero, Nynn Chang, Olubanke Dzahini, Zina M. Ibrahim.

Formal analysis: Ehtesham Iqbal, Chandra Pandey, Richard J. B. Dobson, Zina M. Ibrahim.

Funding acquisition: Richard J. B. Dobson.

Investigation: Ehtesham Iqbal, Robbie Mallah, Honghan Wu, Richard J. B. Dobson, Zina M. Ibrahim.

Methodology: Ehtesham Iqbal, Honghan Wu, Richard J. B. Dobson, Zina M. Ibrahim.

Project administration: Matthew Broadbent, Richard J. B. Dobson, Zina M. Ibrahim.

Resources: Robbie Mallah, Matthew Broadbent, Robert Stewart.

Software: Ehtesham Iqbal, Matthew Broadbent.

Supervision: Olubanke Dzahini, Richard J. B. Dobson, Zina M. Ibrahim.

Validation: Ehtesham Iqbal, Robbie Mallah, Alvin Romero, Nynn Chang, Olubanke Dzahini.

Visualization: Ehtesham Iqbal.

Writing – original draft: Ehtesham Iqbal, Daniel Rhodes, Richard J. B. Dobson, Zina M. Ibrahim.

Writing – review & editing: Ehtesham Iqbal, Richard J. B. Dobson, Zina M. Ibrahim.

References

1. Savova GK, Chute CG. Constructing evaluation corpora for automated clinical named entity recognition.) Informatics; Building . . .. search.informit.com.au; 2007; Available: http://search.informit.com.au/documentSummary;dn=793987578001922;res=IELHEA
2. Chapman WW, Dowling JN, Hripcsak G. Evaluation of training with an annotation schema for manual annotation of clinical conditions from emergency department reports. International journal of medical. Elsevier; 2008; Available: http://www.sciencedirect.com/science/article/pii/S1386505607000147
3. Roberts A, Gaizauskas R, Hepple M, Demetriou G, Guo Y, Roberts I, et al. Building a semantically annotated corpus of clinical texts. J Biomed Inform. 2009; 42: 950–966. https://doi.org/10.1016/j.jbi.2008.12.013 PMID: 19535011
4. Wang Y. Annotating and recognising named entities in clinical notes. Proceedings of the ACL-IJCNLP 2009 Student Research Workshop. Association for Computational Linguistics; 2009. pp. 18–26.
5. Uzuner O, Solti I, Xia F, Cadag E. Community annotation experiment for ground truth generation for the i2b2 medication challenge. J Am Med Inform Assoc. 2010; 17: 519–523. https://doi.org/10.1136/jamia.2010.004200 PMID: 20819855
6. Koeling R, Carroll J, Tate AR, Nicholson A. Annotating a corpus of clinical text records for learning to recognize symptoms automatically. Proceedings of the 3rd Louhi Workshop on Text and Data Mining of Health Documents. 2011. pp. 43–50.
7. Albright D, Lanfranchi A, Fredriksen A, Styler WF 4th, Warner C, Hwang JD, et al. Towards comprehensive syntactic and semantic annotations of the clinical narrative. J Am Med Inform Assoc. 2013; 20: 922–930. https://doi.org/10.1136/amiajnl-2012-001317 PMID: 23355458

8. Savova GK, Masanz JJ, Ogren PV, Zheng J, Sohn S, Kipper-Schuler KC, et al. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. J Am Med Inform Assoc. 2010; 17: 507–513. https://doi.org/10.1136/jamia.2009.001560 PMID: 20819853
9. Xu H, Stenner SP, Doan S, Johnson KB, Waitman LR, Denny JC. MedEx: a medication information extraction system for clinical narratives. J Am Med Inform Assoc. 2010; 17: 19–24. https://doi.org/10.1197/jamia.M3378 PMID: 20064797
10. Aronson AR, Lang F-M. An overview of MetaMap: historical perspective and recent advances. J Am Med Inform Assoc. 2010; 17: 229–236. https://doi.org/10.1136/jamia.2009.002733 PMID: 20442139
11. Uzuner Ö, South BR, Shen S, DuVall SL. 2010 i2b2/VA challenge on concepts, assertions, and relations in clinical text. J Am Med Inform Assoc. 2011; 18: 552–556. https://doi.org/10.1136/amiajnl-2011-000203 PMID: 21685143
12. 2016 Research Domains Criteria (RDoC) for Psychiatry [Internet]. [cited 23 Mar 2017]. Available: https://www.i2b2.org/NLP/RDoCforPsychiatry/
13. Pirmohamed M, James S, Meakin S, Green C, Scott AK, Walley TJ, et al. Adverse drug reactions as cause of admission to hospital: prospective analysis of 18 820 patients. BMJ. 2004; 329: 15–19. https://doi.org/10.1136/bmj.329.7456.15 PMID: 15231615
14. Honigman B, Lee J, Rothschild J. Using computerized data to identify adverse drug events in outpatients. Journal of the American Medical Informatics Association. jamia.oxfordjournals.org; 2001; Available: https://jamia.oxfordjournals.org/content/8/3/254.full
15. Forster AJ, Murff HJ, Peterson JF. The incidence and severity of adverse events affecting patients after discharge from the hospital. Ann Intern Med. Am Coll Physicians; 2003; Available: http://annals.org/article.aspx?articleid=716006
16. Field TS, Gurwitz JH, Harrold LR, Rothschild J, DeBellis KR, Seger AC, et al. Risk Factors for Adverse Drug Events Among Older Adults in the Ambulatory Setting. J Am Geriatr Soc. Blackwell Science Inc; 2004; 52: 1349–1354. https://doi.org/10.1111/j.1532-5415.2004.52367.x PMID: 15271125
17. Hazlehurst B, Naleway A, Mullooly J. Detecting possible vaccine adverse events in clinical notes of the electronic medical record. Vaccine. Elsevier; 2009; Available: http://www.sciencedirect.com/science/article/pii/S0264410X09001686
18. Wang X, Chase H, Markatou M, Hripcsak G, Friedman C. Selecting information in electronic health records for knowledge acquisition. J Biomed Inform. 2010; 43: 595–601. https://doi.org/10.1016/j.jbi.2010.03.011 PMID: 20362071
19. Haerian K, Varn D, Chase H. Electronic health record pharmacovigilance signal extraction: a semi-automated method for reduction of confounding applied to detection of . . .. Drugs. S INT LTD 41 CENTORIAN DR, . . .; 2010;
20. Kuhn M, Letunic I, Jensen LJ, Bork P. The SIDER database of drugs and side effects. Nucleic Acids Res. 2016; 44: D1075–9. https://doi.org/10.1093/nar/gkv1075 PMID: 26481350
21. Warrer P, Jensen PB, Aagaard L. Identification of possible adverse drug reactions in clinical notes: The case of glucose-lowering medicines. Journal of research in. ncbi.nlm.nih.gov; 2015; Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4418138/
22. Han L, Ball R, Pamer CA, Altman RB, Proestel S. Development of an automated assessment tool for MedWatch reports in the FDA adverse event reporting system. J Am Med Inform Assoc. 2017; https://doi.org/10.1093/jamia/ocx022 PMID: 28371826
23. Cai R, Liu M, Hu Y, Melton BL, Matheny ME, Xu H, et al. Identification of adverse drug-drug interactions through causal association rule discovery from spontaneous adverse event reports. Artif Intell Med. 2017; 76: 7–15. https://doi.org/10.1016/j.artmed.2017.01.004 PMID: 28363289
24. Cañada A, Capella-Gutierrez S, Rabal O, Oyarzabal J, Valencia A, Krallinger M. LimTox: a web tool for applied text mining of adverse event and toxicity associations of compounds, drugs and genes. Nucleic Acids Res. 2017; https://doi.org/10.1093/nar/gkx462 PMID: 28531339
25. Sarker A, Gonzalez G. Portable automatic text classification for adverse drug reaction detection via multi-corpus training. J Biomed Inform. 2015; 53: 196–207. https://doi.org/10.1016/j.jbi.2014.11.002 PMID: 25451103
26. Xu R, Wang Q. Large-scale combining signals from both biomedical literature and the FDA Adverse Event Reporting System (FAERS) to improve post-marketing drug safety signal detection. BMC Bioinformatics. 2014; 15: 17. https://doi.org/10.1186/1471-2105-15-17 PMID: 24428898
27. Korkontzelos I, Nikfarjam A, Shardlow M, Sarker A, Ananiadou S, Gonzalez GH. Analysis of the effect of sentiment analysis on extracting adverse drug reactions from tweets and forum posts. J Biomed Inform. 2016; 62: 148–158. https://doi.org/10.1016/j.jbi.2016.06.007 PMID: 27363901

28. Nikfarjam A, Sarker A, O'Connor K, Ginn R, Gonzalez G. Pharmacovigilance from social media: mining adverse drug reaction mentions using sequence labeling with word embedding cluster features. J Am Med Inform Assoc. 2015; 22: 671–681. https://doi.org/10.1093/jamia/ocu041 PMID: 25755127

29. Cohen R, Elhadad M, Elhadad N. Redundancy in electronic health record corpora: analysis, impact on text mining performance and mitigation strategies. BMC Bioinformatics. 2013; 14: 10. https://doi.org/10.1186/1471-2105-14-10 PMID: 23323800

30. Iqbal E, Mallah R, Jackson RG, Ball M, Ibrahim ZM, Broadbent M, et al. Identification of Adverse Drug Events from Free Text Electronic Patient Records and Information in a Large Mental Health Case Register. PLoS One. 2015; 10: e0134208. https://doi.org/10.1371/journal.pone.0134208 PMID: 26273830

31. Stewart R, Soremekun M, Perera G, Broadbent M, Callard F, Denis M, et al. The South London and Maudsley NHS Foundation Trust Biomedical Research Centre (SLAM BRC) case register: development and descriptive data. BMC Psychiatry. 2009; 9: 51. https://doi.org/10.1186/1471-244X-9-51 PMID: 19674459

32. Enabling Cookies: MedicinesComplete [Internet]. [cited 23 Mar 2017]. Available: https://www.medicinescomplete.com/mc/bnf/current/

33. Chapman WW, Chu D, Dowling JN. ConText: An Algorithm for Identifying Contextual Features from Clinical Text. Proceedings of the Workshop on BioNLP 2007: Biological, Translational, and Clinical Language Processing. Stroudsburg, PA, USA: Association for Computational Linguistics; 2007. pp. 81–88.

34. Home—electronic Medicines Compendium (eMC) [Internet]. [cited 23 Mar 2017]. Available: https://www.medicines.org.uk/emc/

35. Cunningham H, Maynard D, Bontcheva K, Tablan V. GATE: An Architecture for Development of Robust HLT Applications. Proceedings of the 40th Annual Meeting on Association for Computational Linguistics. Stroudsburg, PA, USA: Association for Computational Linguistics; 2002. pp. 168–175.

36. Cunningham H, Others. Text Processing with GATE, version 6 ed., University of Sheffield Department of Computer Science, April 2011.

37. Thakker D, Osman T, Lakin P. Gate jape grammar tutorial. Nottingham Trent University, UK, Phil Lakin, UK, Version. 2009;1. Available: https://gate.ac.uk/sale/thakker-jape-tutorial/GATE%20JAPE%20manual.pdf

38. KHP-Informatics. KHP-Informatics/ADRApp. In: GitHub [Internet]. [cited 23 Mar 2017]. Available: https://github.com/KHP-Informatics/ADRApp

39. Gandhi TK, Burstin HR, Cook EF, Puopolo AL, Haas JS, Brennan TA, et al. Drug complications in outpatients. J Gen Intern Med. 2000; 15: 149–154. https://doi.org/10.1046/j.1525-1497.2000.04199.x PMID: 10718894

40. Preventing Medication Errors. Washington, D.C.: National Academies Press; 2007.

3.3 Discussion and Future work

The work presented in Article II of this chapter is the first example of a publicly available tool for detecting and classification of ADEs in psychiatric clinical notes. The rule-based tool was developed by manually annotating and analysing over 2,300 randomly selected documents. The mentions of ADEs, text patterns and contextual properties of the text were identified in the collaboration of domain experts. Four ADEs were selected related to antipsychotic and antidepressant drugs to assess the performance of the tool. The tool was further evaluated in the range of other ADEs from common to rare and acute to chronic in 5,200 unseen clinical documents by comparing the manual annotations.

The tools perform well as compared to general NLP NER systems, mainly the ConText algorithm (W. W. Chapman et al., 2007). However, the performance varied depending on the persistence of ADE, mainly for rare ADEs. Rare ADEs are discussed as a caution, monitoring and possibilities in the clinical notes. The patient must go through many tests before being ruled out as a positive occurrence of an ADE. Currently, there is ongoing work to establish contextual patterns. The terms dictionary (ADEs and contextual terms) can be modified and enhanced with more ADE terms of interest without modifying the rule-based. The tool has been evaluated in a similar psychiatric setting on unseen data and shows good precision and recall and represents tools portability. The research community within the SLAM has shown great interest in using this tool. The ADE dictionary has been extended from ADEs related to antipsychotics and antidepressants drugs to all psychotropic drugs.

In future, the intentions are to improve the rules for discerning context for rare ADEs and a hybrid approach combining the current pipeline with ML methods and assess the effect on performance.

Chapter 4

4 Detecting Adverse Drug Reactions (ADRs) from the EHR

4.1 Introduction

Identifying ADRs from clinical text could help clinicians predict risks and intervene in the future administration of drug treatment, dose alteration and drug withdrawal. Edwards et al. (2000) classified ADRs into six types: “dose-related (Augmented), non-dose-related (Bizarre), dose and time-related (Chronic), time-related (Delayed), withdrawal (End of use), and failure of therapy (Failure)”. ADRs can appear anytime during and after the course a drug therapy. A clinician identifies an ADR by direct observation, results from laboratory reports, self-reported symptoms from patients and observations made by family members. Usually, these findings are recorded in the free-text clinical documents.

The ADR detection from clinical text is a two-step task. First, accurately recognise named entities such as medication, symptoms, adverse events and diseases and secondly, the relationship between identified named entities such as adverse events and medications. In such attempts, many studies have used rule-based clinical tools such as Knowledgemap (Denny et al., 2003), MetaMap (Aronson, 2001) and MedLEE (Friedman et al., 2004). Later, the trend was shifted toward the ML methods, and many studies have reported improved performance. The biomedical NLP community has organised many challenges for NER and relation extraction such as i2b2 (The Center for Informatics for Integrating Biology and the Bedside) (Sun et al., 2013; Uzuner et al., 2011), BioCreative (Grover et al., 2007; C.-H. Wei et al., 2016), SemEval (International Workshop on Semantic Evaluation) (Pradhan et al., 2014) and Medication and Adverse Drug Events from electronic health records (MADE1.0) (D. Xu et al., 2018; Yang et al., 2019). The MADE1.0 consist of three subtasks, (1) NER extraction of ADE and medication and their features, (2) a relation extraction among ADE and medication, (3) combining the subtasks (1) and (2) from a large clinical corpus. The work presented in this chapter is closely based on MADE1.0; however, it was initiated much earlier than open MADE1.0 challenge.

The work presented in Chapter 2 and 3 discussed ADE detection from the clinical text. Identifying the named entity ADE was successfully achieved in Chapter 2 and 3 of this thesis in which the narrative clinical text has been used from free-text clinical notes. This chapter demonstrates the implementation of NLP pipelines group towards the identification of ADRs from free-text clinical notes. These pipelines identify medications and define medication episodes, ADEs and finally an ADR timeline to suggest possible causal relationships between the medication and ADEs in SLAM, C&I and Oxford NHS Foundation Trusts.

4.2 SLAM NLP Capacity

Within SLAM, a range of NLP applications have been developed using two tools primarily, GATE and TextHunter. Over a hundred GATE NLP applications have been developed to date to identify and annotate medical terms from free-text clinical documents. The most widely used GATE applications include medications, diagnosis, smoking status, Standardise Mini-Mental State Examination (SMMSE), Cognitive Behavioural Therapy (CBT), Body Mass Index (BMI), suicide behaviour, negative symptoms, ideation and blood pressure applications. The work presented in this chapter further adds GATE NLP capabilities by developing an Adverse Drug Event (ADE) detection application, building a pipeline to discern medication timelines for patients, and an ADR timeline, which uses temporal precedence to identify possible causal relationships between ADEs and medications.

4.2.1.1 NLP Medication and Diagnosis applications

The NLP medication application is designed to extract the names of medications prescribed to the patient and some dosage information. It distinguishes between current medications, those prescribed at the time of the document was written, and medications that have been prescribed in the past and hypothetical mentions. The application ignores mentions of future medications, as clinicians may include prescription information should the conditions worsen. There is ongoing work to include a daily dosage or the dose given at a single point in time. The medication application has been assessed and validated on Clozapine prescribing, which evaluates the following criteria: (a) which patients have ever taken/been prescribed Clozapine, (b) have patient prescribed Clozapine, (c) and what dose was prescribed. The assessment reported precision on GATE Clozapine annotation

are 0.92 within one month, 0.95 within two months, 0.99 within three months and 0.84 on the correct dosage. The recall reported on GATE Clozapine annotations are 0.81 within one month, 0.82 within two months, 0.91 within three months and 0.78 on the correct dosage (Kadra et al., 2016).

The SLAM CRIS programme has developed an approach to merge data from different sources. For example, medication information is available in ePJS structured fields through links with the pharmacy dispensing system 'Electronic Prescribing and Medicines Administrations' (EPMA) and retrieved by the NLP medication application from the free-text. In the interest of getting higher coverage, this work compiled a combined medication dataset to include medication information from all sources. Figure 4.1 represents a yearly breakdown of the types of documents containing medication information within SLAM, and Figure 4.2 shows the volume of medication information extracted from free-text (unstructured data) and structured fields. In Figure 4.1, the data source 'Medication' represents structured data from the EHR (ePJS), and EPMA represents the pharmacy dispensing structured data.

The ePJS structured fields are not actively used in SLAM. Figure 4.2 shows that over time, an increasing proportion of medication information is extracted from free-text (unstructured data) as compared to structured data. In C&I and Oxford, medication information are mainly obtained from the free-text using the NLP medication application as there is very little structured data available.

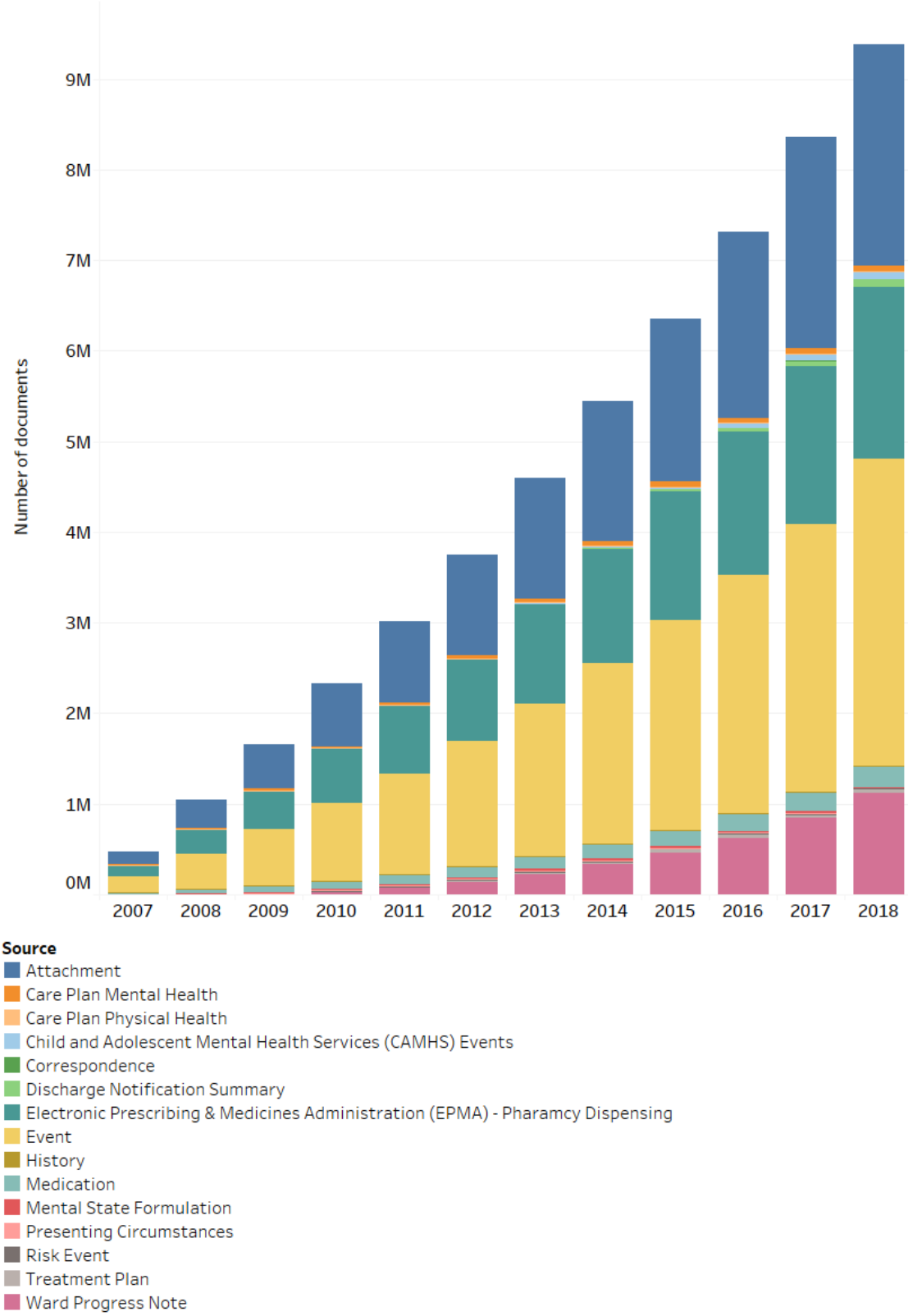

Figure 4.1: Break down of cumulative medication information from the different document types in SLAM; Structured data sources = Electronic Prescribing and Medicines Administrations (EPMA), medication; Free-text data sources = attachment, care plan mental health, care plan physical health, Child and Adolescent Mental Health Services (CAMHS), correspondence, discharge notification summary, event, history, mental state formulation, presenting circumstances, risk event, treatment plan, ward progress note; SLAM = South London and Maudsley NHS Foundation Trust health record

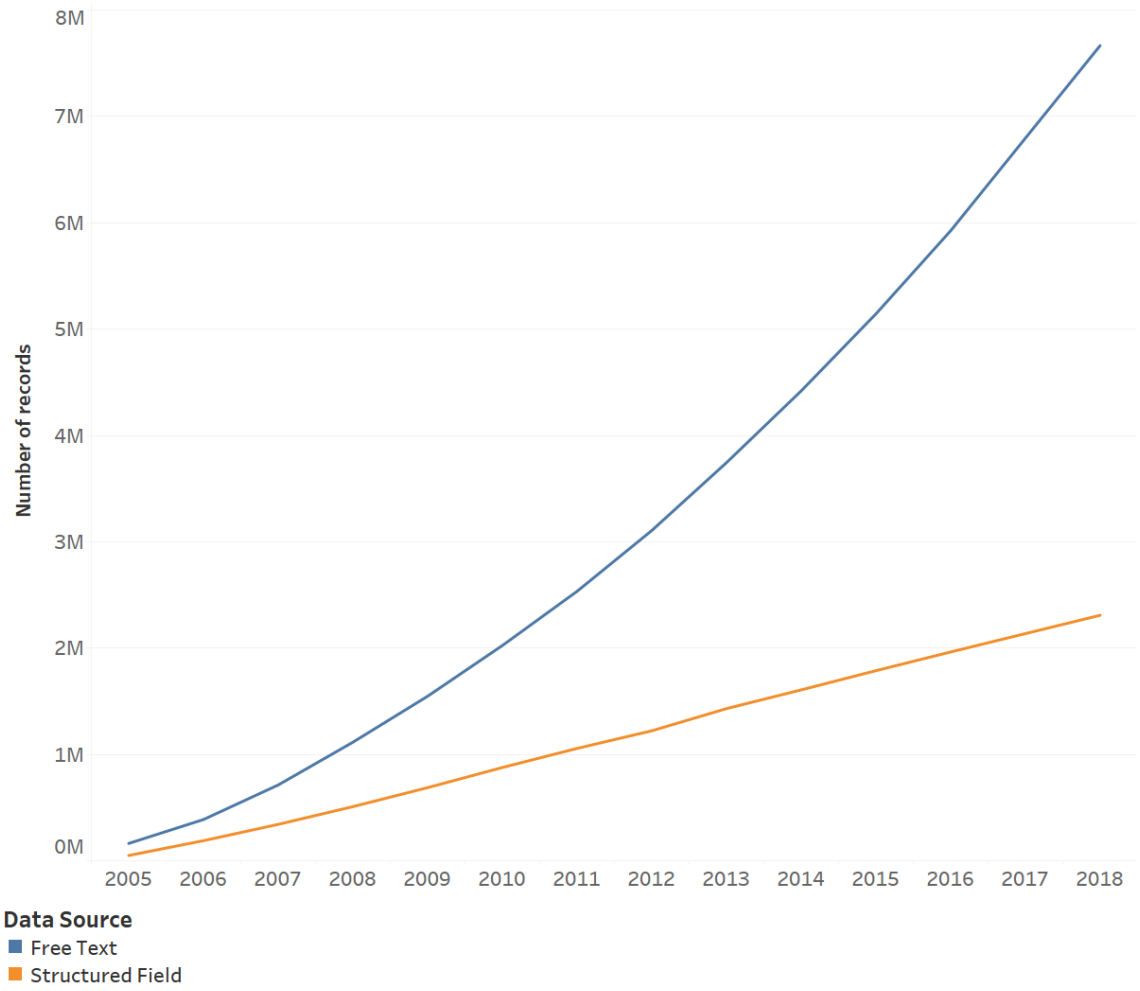

Figure 4.2: Medication information extracted from the free-text (unstructured) and structured fields in SLAM; SLAM = South London and Maudsley NHS Foundation Trust health record

The diagnosis NLP application was intended to supplement diagnosis data obtained from the structured fields in the EHR and was validated using Alzheimer's and SMI diagnoses. Therefore, this work followed a similar approach to that of extracting medication information, combining structured diagnoses with NLP-extracted diagnoses in SLAM. Figure 4.3, represents the yearly breakdown and types of documents containing diagnosis information, and Figure 4.4 shows the volume of information extracted from free-text (unstructured data) by NLP and structured fields for diagnosis.

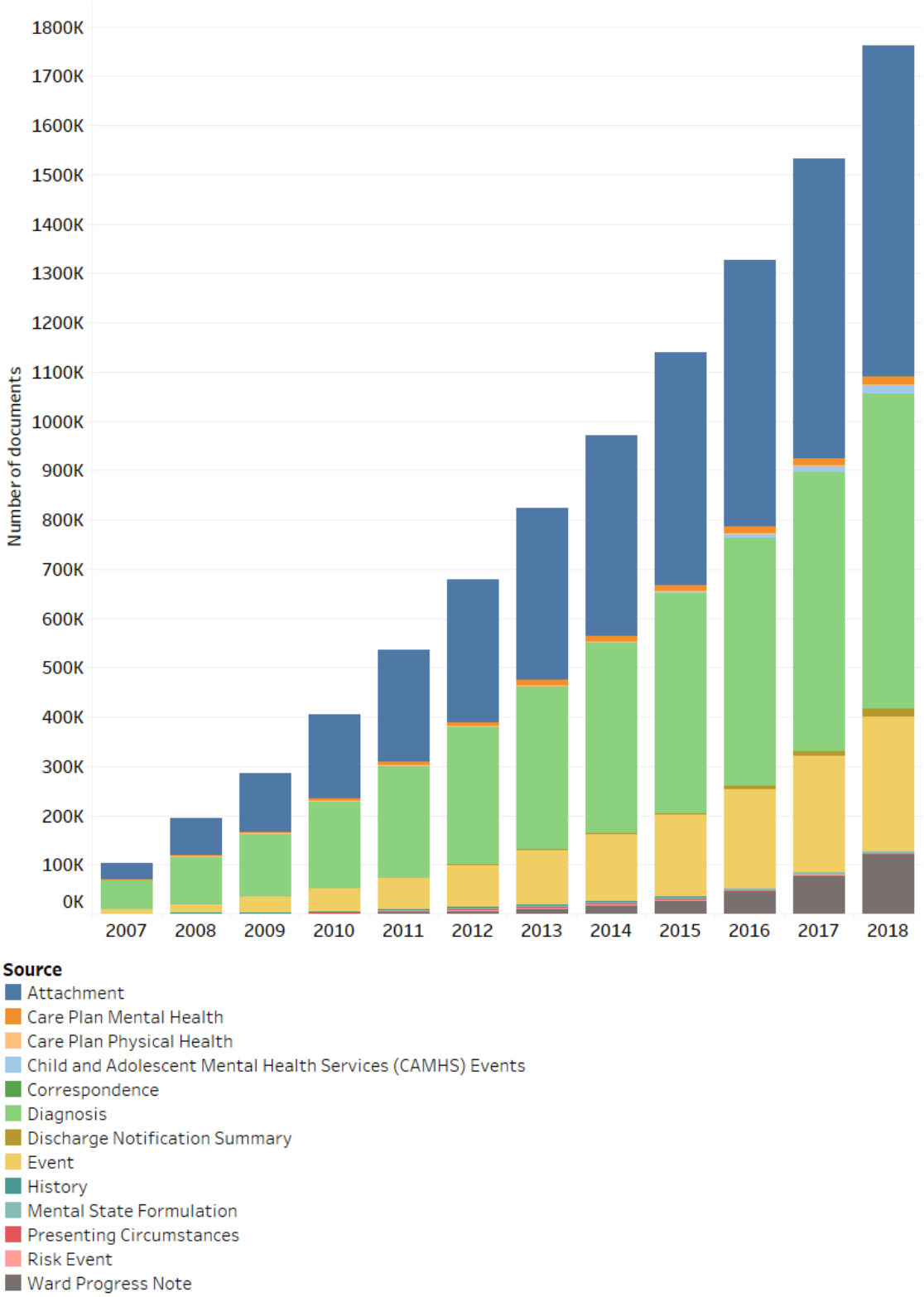

Figure 4.3: Break down of cumulative diagnosis information from the different data sources in SLAM; Structured data sources = diagnosis; Free-text data sources = attachment, care plan mental health, care plan physical health, Child and Adolescent Mental Health Services (CAMHS), correspondence, discharge notification summary, event, history, mental state formulation, presenting circumstances, risk event, ward progress note; SLAM = South London and Maudsley NHS Foundation Trust health record

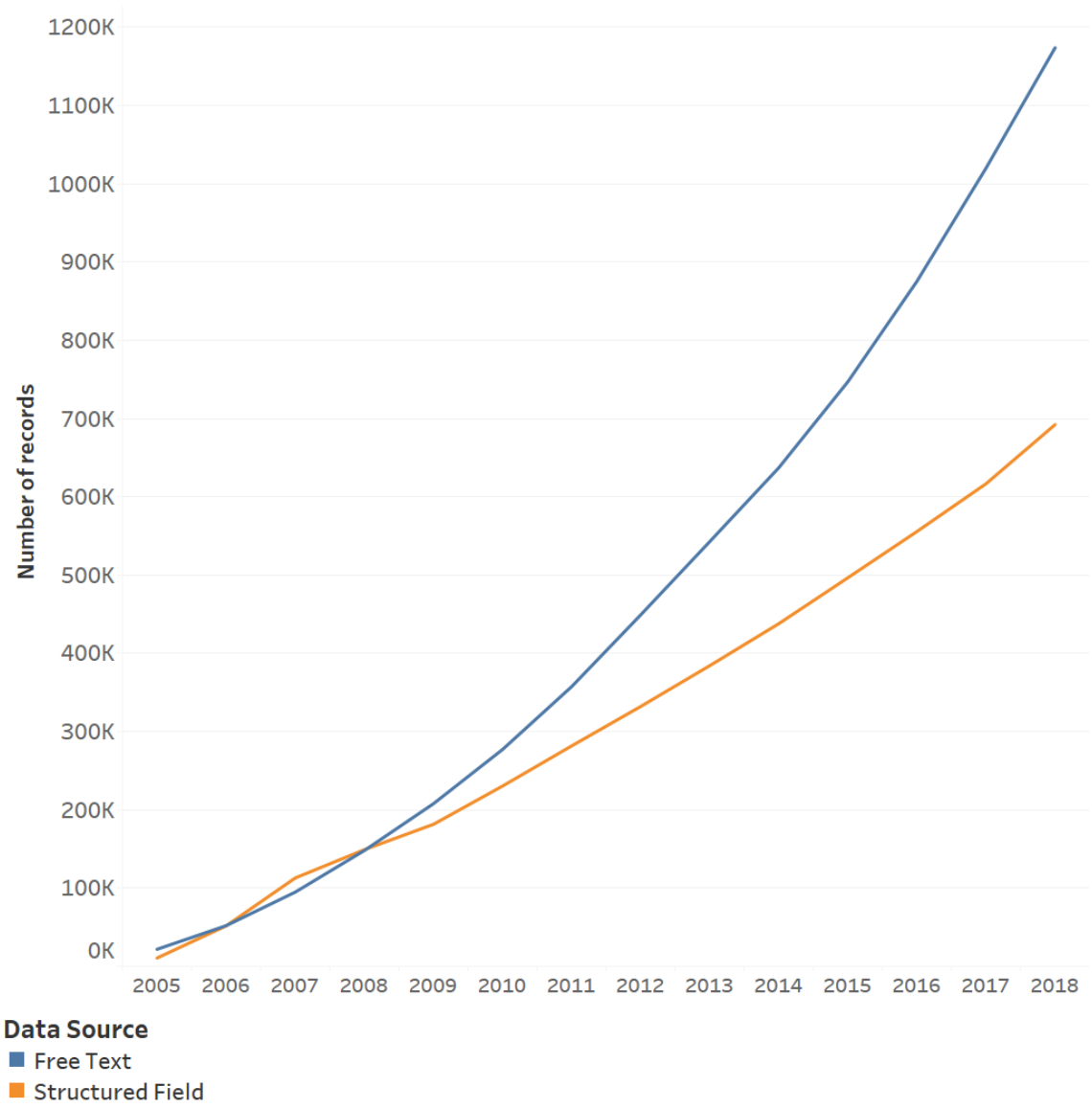

Figure 4.4: Diagnosis information extracted from the free-text (unstructured) and structured fields in SLAM

Table 4.1 shows the volume of ICD10 diagnoses in SLAM, C&I and Oxford Trusts. The diagnoses listed are the most recent for each patient at the time of data extraction. SLAM CRIS contained over 323,713 patient records (as of June 2018), C&I CRIS dataset contained 116,936 patient records (as of July 2016), and Oxford CRIS contained 98,401 patient records (as of March 2015). The different dates across the three Trusts reflect the different timeframes access was granted to the three resources.

In SLAM, the combined diagnosis dataset (structured and unstructured) was used, with 21% of the data being extracted from the free-text (unstructured data) by the NLP

diagnosis application, and 79% of the data was obtained from structured fields. In C&I and Oxford, only structured diagnosis data was used to assert the most recent diagnosis for the patients. As the three NHS Trusts discussed in this thesis are mental health providers, most patients are unsurprisingly annotated with (ICD-10: F00-F99) mental and behavioural disorders. Table 4.1 further breaks down the high-level diagnosis into ten distinct (ICD-10: F00-F99) chapter blocks. In SLAM, the diagnosis breakdown (ICD-10: F00-F99) is available in supplementary Table A.1. In C&I, diagnosis coverage is not as comprehensive as coded diagnoses were not actively used in the early years of EHR adoption.

Table 4.1: Breakdown of diagnoses in SLAM, C&I and Oxford NHS Trusts

ICD 10 Category	Category Name	SLAM (n=323713)	C&I (n=116,936)	Oxford (n=98401)
F00/(G30-G32)	Dementia in Alzheimer's disease	2.67%	1.91%	7.56%
F01-F09	Mental disorders due to known physiological conditions	4.43%	2.80%	5.95%
F10-F19	Mental and behavioural disorders due to psychoactive substance use	8.01%	3.36%	2.49%
F20-F29	Schizophrenia, schizotypal, delusional, and other non-mood psychotic disorders	5.68%	3.61%	4.23%
F30-F39	Mood [affective] disorders	10.33%	4.46%	12.79%
F40-F48	Anxiety, dissociative, stress-related, somatoform and other nonpsychotic mental disorders	8.56%	2.69%	6.98%
F50-F59	Behavioural syndromes associated with physiological disturbances and physical factors	2.55%	0.10%	1.20%
F60-F69	Disorders of adult personality and behaviour	1.29%	1.50%	3.06%
F70-F79	Intellectual disabilities	0.74%	0.05%	0.66%
F80-F89	Pervasive and specific developmental disorders	2.01%	0.11%	2.69%
F90-F98	Behavioural and emotional disorders with onset usually occurring in childhood and adolescence	4.60%	0.27%	4.67%
F99	Mental disorder, not otherwise specified	11.22%	0.05%	1.76%
Z00-Z99	Factors influencing health status and contact with health services	13.51%	0.79%	4.23%

Others	Any Other Diagnosis	4.12%	0.31%	3.10%
Not Available	Not Available	20.28%	77.99%	38.62%

The tables represent the diagnosis break down in three NHS Foundation Trust with the diagnosis break down into F00-F99 - Mental and *behavioural disorder which further breaks down into twelve categories (F00 and G30-G32); F01-F09; F10-F19; F19-F29; F30-F39; F40-F49; F50-F59; F60-F69; F70-F79; F80-F89; F90-F99 and F99), Z00-Z99, any other diagnosis and not available. In SLAM the diagnoses were extracted from structured and unstructured fields (NLP application) and in C&I and Oxford the diagnoses were extracted from the structured fields. SLAM = South London and Maudsley NHS Foundation Trust health record; C&I = Camden and Islington NHS Foundation Trust health record; Oxford = Oxford Health NHS Foundation Trust health record; n = number of patients at the time research was concluded*

4.3 Implementation of the ADE pipeline

The ADEPt pipeline was implemented to extract ADEs from the free-text clinical documents in SLAM, C&I and Oxford NHS Foundation Trusts. In SLAM, the ADEPt pipeline was evaluated against a large number of ADEs, and an overall 0.89 and 0.88 recall was achieved. In order to extract ADEs, a dictionary was created with the help of pharmacists and clinical researchers to contain 110 ADEs. These ADEs are known to be associated with antipsychotics, antidepressants, mood-stabilisers, hypnotics and anxiolytics. The detailed discussion of how the pipeline was developed in SLAM is available in Chapter 3, Article II of this thesis.

The ADEPt pipeline was also implemented and evaluated in C&I using corpora of 1000 manually annotated documents. These corpora contained 10 ADEs (100 documents each) ranging from mild to severe, rare to common and acute to chronic. Comparing the ADEPt pipeline with the manual annotations, the ADEPt pipeline achieved an overall 0.84 precision and 0.87 recall in C&I. The documents were extracted from the ADEPt pipeline results table. The results table include the text where ADE mention is present. Table 4.2 shows the performance of the ADEPt pipeline in C&I CRIS. A similar performance was seen when evaluating ADEPt for common, mild and acute ADEs such as agitation, akathisia, constipation, headache and weight gain in both C&I and SLAM. However, the ADEPt pipeline misclassifies rare ADEs, due to the reasons discussed in Chapter 3.

Table 4.2: Performance of the ADEPt pipeline in C&I NHS Trust

ADE	Total	True Positive	True Negative	Precision	Sensitivity	Specificity	Accuracy	F-measure
Agitation	100	48	37	0.87	0.86	0.84	0.86	0.85
Akathisia	100	61	24	0.90	0.88	0.77	0.89	0.85
Constipation	100	63	21	0.88	0.90	0.70	0.89	0.84
Galactorrhea	100	53	28	0.85	0.84	0.76	0.85	0.81
Headache	100	64	23	0.93	0.89	0.82	0.91	0.87
Insomnia	100	67	27	0.96	0.96	0.90	0.96	0.94
Myocarditis	100	21	49	0.53	0.66	0.72	0.58	0.70
Sedation	100	47	39	0.89	0.85	0.87	0.87	0.86
Stevens-Johnson Syndrome	100	17	51	0.41	0.68	0.68	0.52	0.68
Weight Gain	100	60	29	0.91	0.92	0.83	0.92	0.89
Overall	**1000**	**501**	**328**	**0.84**	**0.87**	**0.78**	**0.85**	**0.83**

Results showing the performance of the ADEPt pipeline in identifying a selection of rare to common ADEs related to antipsychotics and antidepressants drugs; C&I = Camden and Islington NHS Foundation Trust health record

In Oxford, an earlier version of ADEPt with a smaller dictionary was used to extract 66 ADEs. Due to resource constraints, manually annotated corpora was not created for the Oxford dataset, and no evaluation was performed.

Table 4.3 shows a subset of common to rare ADE percentages in SLAM, C&I and Oxford NHS Trusts. The table highlights the total number of ADEs and their status (positive or negative) classified by the ADEPt pipeline at the time of data extraction. Anxiety, agitation, fatigue, sedation and disorientation were the most common and highest recorded ADEs in SLAM, C&I and Oxford NHS trusts. The ADEPt pipeline classified these ADEs as positive mentions in 71-78% of the mentions. Abdominal pain, nausea, and EPSEs such as parkinsonian, akathisia, dystonia and tardive dyskinesia showed that around 50% of these ADEs are positive mentions. In rare and life-threatening ADEs such as Stevens-Johnson syndrome (SJS), myocarditis and pericarditis, there were approximately 30% positive mentions. It was observed that the rare ADEs were generally

discussed within clinical text as warnings or speculative side effects. The overall percentages of common to rare ADEs were similar in SLAM, C&I and Oxford NHS Foundation Trusts.

Table 4.3: Percentages of ADEs in SLAM, C&I and Oxford NHS Foundation Trusts

ADEs	Trust	Total	Positive (%)	Negative (%)
Anxiety	SLAM	3,429,102	71	29
	C&I	820,822	76	24
	Oxford	1,298,363	74	26
Agitation	SLAM	1,472,544	72	28
	C&I	340,592	76	24
	Oxford	386,901	73	27
Fatigue	SLAM	791,064	76	24
	C&I	179,723	78	22
	Oxford	260,939	77	23
Sedation	SLAM	470,153	72	28
	C&I	106,378	72	28
	Oxford	113,505	72	28
Disorientation	SLAM	306,366	78	22
	C&I	91,356	78	22
	Oxford	113,966	78	22
Nausea	SLAM	104,735	50	50
	C&I	20,993	53	47
	Oxford	29,635	57	43
Abdominal pain	SLAM	88,740	49	51
	C&I	20,419	52	48
	Oxford	31,884	57	43
Parkinsonian	SLAM	14,793	54	46
	C&I	3,982	56	44
	Oxford	4,936	57	43
Akathisia	SLAM	16,399	52	48
	C&I	3,679	59	41
	Oxford	3,430	52	48
Dystonia	SLAM	12,923	52	48
	C&I	3,125	55	45
	Oxford	2,508	48	52
Tardive dyskinesia	SLAM	7,890	59	41
	C&I	2,249	63	37
	Oxford	1,457	55	45
Myocarditis	SLAM	3,279	27	73
	C&I	430	24	76
	Oxford	324	33	67
Stevens-Johnson Syndrome	SLAM	1,284	37	63
	C&I	349	35	65
	Oxford	280	38	62

Pericarditis	SLAM	1,067	28	72
	C&I	220	39	61
	Oxford	294	31	69

In three large psychiatric *health providers, the ADEPt pipeline classified anxiety, agitation, fatigue, sedation and disorientation as positive mentions ranging from 71-78%; abdominal pain, nausea, parkinsonian, akathisia, dystonia and tardive dyskinesia classified around 50% positive mentions, rare ADEs such as Stevens-Johnson syndrome, myocarditis and pericarditis were the least recorded and generally discussed within clinical texts as warnings or speculative side effects; ADEs = Adverse Drugs Events; SLAM = South London and Maudsley NHS Foundation Trust health record; C&I = Camden and Islington NHS Foundation Trust health record; Oxford = Oxford Health NHS Foundation Trust health record; Total = number of instances*

4.4 Medication Episode Algorithm: Medication start and stop dates

The GATE medication NLP application (Hayes et al., 2014; Kadra et al., 2016; Perera et al., 2016) only inferred mentions of medications in the clinical notes and did not identify timeframes during which a patient was taking a given medication. The medication episode algorithm developed in this work adds the capability of inferring periods of 'on-treatment' medication episodes to the NLP applications.

A drug dictionary was created for 260 drugs commonly used for Mental, Behavioural and Neurodevelopmental Disorders. The drug dictionary consists of 11 drug categories: Antidepressant, Antidiabetic, Antiepileptic, Antihypertensive, Antipsychotic, Dementia, Hypnotics & Anxiolytics, Lipid Regulatory, Mood Stabilizer, Non-Steroidal Anti-Inflammatory and Parkinson. The dictionary used the British National Formulary (BNF) and electronic Medicines Compendium (eMC) (Datapharm Communications Ltd, 2017; Joint Formulary Committee, 2016) to generate a robust list of all the available brand names. In order to increase coverage, discontinued brand names were included. The algorithm has been previously used in a dementia study to identify trajectories of cognitive decline in the SLAM EHR (Baker et al., 2017).

A functional representation of the medication episodes algorithm is shown in Figure 4.5. The algorithm works in two stages. In the first stage, it maps all the brand names onto generic drug names, by combining information from the GATE medication application and drug dictionaries. In the second stage, the algorithm sorts the records by

date and measures the duration between consecutive dates. These dates include prescription dates and the positive mention indicating that a patient is taking/continuing a particular drug, extracted by the GATE NLP medication app, and act as data points or prescribing dates. The algorithm only considers the positive mentions of drugs, all other mentions classified by medication application, such as hypothetical, future, and negative mentions were ignored as they do not represent a patient is on a particular drug. The hypothetical mentions were excluded as the clinicians may write that a patient should prescribe a particular drug if their conditions worsen, but that may never happen to the patient. The future mentions of medications were ignored as they also relate to the event that did not occur. The negative mentions were ignored as a patient is not currently on a particular drug.

The threshold for determining a medication episode between two consecutive prescribing dates is set to 42 days (6 weeks). Although the common practice for prescribing psychotropic drugs is 28 days (4 weeks), they can be prescribed for up to 42 days based on patient availability. If the gap between two consecutive dates is less than 42 days (6 weeks), the algorithm searches for the next date until it finds the date where the difference between two dates is greater than 42 days, or no data point is available. In each episode, the algorithm counts the number of data points it used to conclude an episode. Once the difference between two dates is over 42 days, the algorithm concludes the episode by the last available date and starts a new episode. The threshold can be changed according to the prescribing practices associated with the medication being studied.

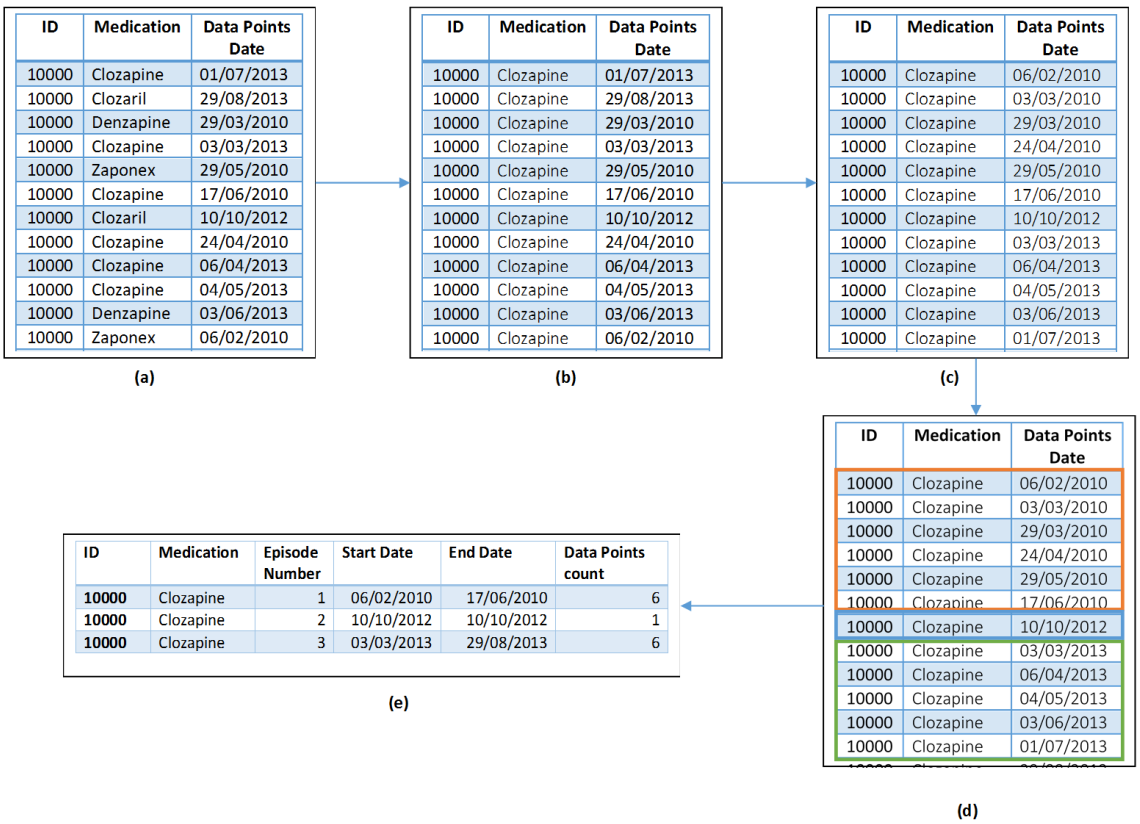

*Figure 4.5: The medication start and stop dates algorithm **(a)** Pre-staged, The algorithm retrieve the drug names and prescription date for a single patient, **(b)** brand names are converted into generic names, **(c)** sorting data into prescription chronological order, **(d)** the algorithm identifies an episode by measuring the distance between consecutive prescription dates, **(e)** medication episodes are created with a clear start and stop dates.*

4.4.1 Validation

The on-treatment medication episodes generated by the algorithm were validated against a manually-curated Clozapine cohort in SLAM. Legge et al. (2016) created a cohort of 300 Clozapine patients with start and end dates. A set of 100 patients who started Clozapine treatment between January 2008 and December 2013 were randomly selected from the manually-curated cohort. The same patients were selected from the algorithm-generated results to compare their episode start and end date. It was found that the selected 100 patients were present in the algorithm-generated cohort. The start dates were compared, and the results showed that 87 out of 100 patients had their algorithm-generated start dates within +/- 10 days of the manually-curated start dates. The end dates were compared, and the results showed that 82 out of 100 patients had their algorithm-generated end dates within +/- 10 days of manually-curated end dates.

Validation results showed that many of the manually-curated medication episodes were split into multiple algorithm-generated episodes. A manual medication episode was generated by only considering the length of the period during which a patient was taking Clozapine, without any consideration of data points corresponding to prescription dates or positive mentions of the drug, from which one would infer that patient is taking/continuing a particular drug. The density of data points plays a crucial role in splitting single manually-generated episodes into multiple algorithm-generated medication episodes.

In addition, the performance of the algorithm was evaluated on a random set of episodes from psychotropic drugs. The validation process included two pharmacists who manually curated each episode by combining consecutive data points. The pharmacists were supplied with brand and generic names of medications as well as a CRIS front-end query. The CRIS front-end is a web interface platform for non-technical users to develop and execute search queries on the CRIS database on structured fields and unstructured clinical documents. The extracted information was used to create data points which were used to curate medication episodes. The initial plan was to validate 100 medication episodes, but the number was deemed too high for manual validation, and the validation process stopped after two episodes. It was observed that a patient on medication for two years to six years could have between 60 to 1500 data points, depending on how active they are within the Trust. Table 4.4 shows the algorithm-generated episodes with medication names, episode start and end dates and the number of data points.

After reviewing the results from the two validated episodes, it was decided to only validate episodes that are no longer than 24 months and have no more than 60 data points. A set of 40 episodes were validated via the manual curation of episodes by combining consecutive data points. The validation results showed that 35 out of the 40 episodes had correct start and end dates with an uncertainty of +/-10 days.

Table 4.4: shows the random selection of medication episodes and the data points considered to draw the episode start and end date.

Drug	Episode Start Date	Episode End Date	Data points
Risperidone	25/02/2009	18/12/2014	1031
Diazepam	07/03/2011	18/12/2014	1010
Diazepam	06/07/2006	22/05/2017	608
Diazepam	07/05/2012	27/08/2014	120
Olanzapine	20/06/2011	21/11/2014	120
Risperidone	29/01/2007	22/02/2010	119
Venlafaxine	28/04/2009	19/08/2010	119
Risperidone	04/02/2012	04/04/2013	85
Diazepam	02/11/2011	09/07/2012	45
Ramipril	19/11/2010	12/07/2012	44

4.4.2 Enhancement

Additional rules were added to the medication start and stop date algorithm to enhance performance further. Episodes consisting of single data points (e.g. A, B and C in Figure 4.2) are ignored by the algorithm, as they do not indicate continuation or starting a medication. It was also observed that such episodes/data points are likely to be false-positive mentions of a drug identified by the GATE medication application, or a patient was not active within the Trust for a long time.

Ad-hoc domain rules were also added with the help of pharmacists to merge two or more consecutive episodes into a single episode. For example, the episodes were merged where the duration between two consecutive episodes is no longer than 60 days, and the episode is longer than 120 days. Figure 4.2 illustrates the enhancements steps.

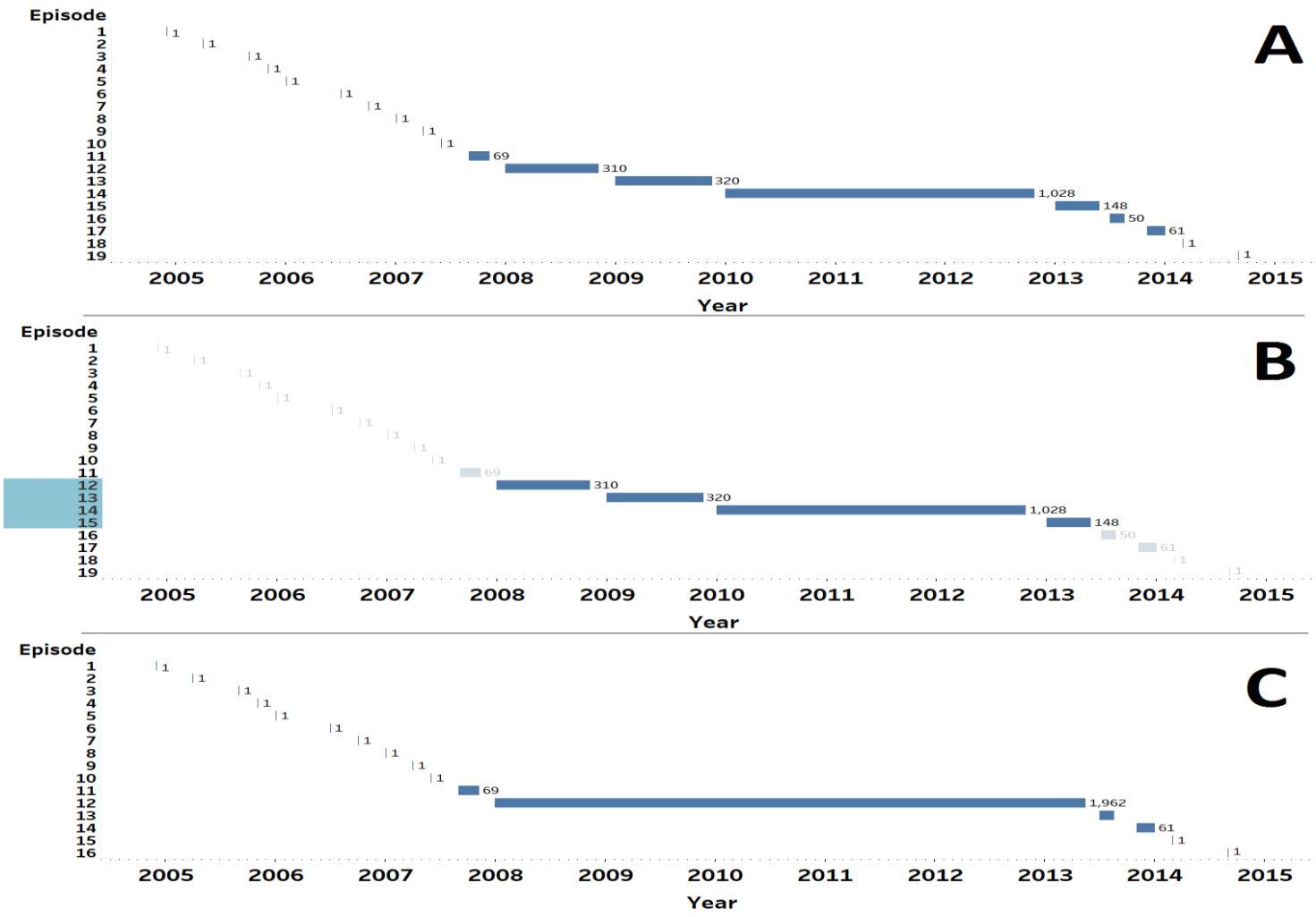

Figure 4.6: The figure shows the medication history of a patient taking Olanzapine from 2005 to 2015. ***Section A*** *shows the medication history generated by the algorithm. Each bar shows a medication episode start date, stop date and duration of the episode.* ***Section B*** *shows that episode 12 to 15 qualifies the enhancement algorithm criteria.* ***Section C*** *shows a new episode created by merging episode 12 to 15.*

4.4.3 Implementation

The on-treatment medication algorithm was implemented in SLAM, C&I and Oxford NHS Foundation Trusts. In SLAM, the medication data included structured data (Pharmacy dispensing data and ePJS) and unstructured data processed by the GATE NLP medication application. It is worth noting that most of the structured medication data were derived from the pharmacy system. In C&I and Oxford NHS Foundation Trusts, the structured data was not available for access. As a result, their medication information was based on data extracted from the GATE NLP medication application. In order to validate a medication treatment, a patient needed to be prescribed a medication for at least six weeks.

Table 4.5 represents the summary of the medication algorithm results for prescribed medications in SLAM (n=134723) from January 2007 to June 2016, C&I (n=56329) from January 2009 to July 2016, and Oxford (n=51260) from January 2010 to March 2015. The results broadly represent the commonly used primary drug categories in SLAM, C&I and Oxford. The drug categories from most prescribed to least prescribed are antidepressants, antipsychotics, hypnotics & anxiolytics, antihypertensive, dementia, lipid regulatory, mood stabiliser, Non-Steroidal Anti-Inflammatory Drugs (NSAIDs), Parkinson and antidiabetic. Antihypertensive, antidiabetic, and lipid regulatory drugs are prescribed to patients to help manage chronic ailments and NSAIDs are used to manage pain. The results show that SLAM has the highest number of patients being prescribed antipsychotic drugs, followed by C&I and finally, Oxford. This is partly due to SLAM having the highest proportion of SMI patients, followed by C&I and Oxford. Overall, Oxford NHS Trust showed the highest proportion of older patients and the highest number of patients being prescribed dementia drugs.

Table 4.5: Summary table for medication algorithm results in the three Trust categorised by primary category.

Primary Category	**SLAM % (n=134723)**	**C&I % (n=56329)**	**Oxford % (n=52160)**
Antidepressant	53.56	45.89	57.80
Antidiabetic	3.91	3.97	2.54
Antiepileptic	9.62	4.83	3.78
Antihypertensive	14.83	15.04	12.49
Antipsychotic	51.74	43.98	35.86
Dementia	5.17	6.01	10.39
Hypnotics and Anxiolytics	41.55	31.59	30.41
Lipid Regulatory Drugs	6.19	6.47	4.95
Mood Stabilizer	9.24	8.36	8.34
Non-Steroidal Anti-Inflammatory Drugs (NSAIDs)	3.01	2.80	2.59
Parkinson	4.70	4.48	2.63

Results show percentages of patients who have been on medication for at least six weeks in SLAM (n=134723) between January 2007 to June 2016, C&I (n=56329) between January 2009 to July 2016 and Oxford (n=51260) between January 2010 to March 2015; n = number of patients at the time research was concluded

The complete results are available in supplementary Table A.3. The three most prescribed primary drug categories are antidepressants, antipsychotics and hypnotics & anxiolytics in SLAM, C&I and Oxford. Broadly, supplementary table A.3 shows that Selective Serotonin Reuptake Inhibitors (SSRIs) is the most common secondary category in antidepressants - these include Citalopram, Sertraline and Mirtazapine. In antipsychotics, the atypical drug category is more commonly prescribed; this includes Aripiprazole, Olanzapine, Quetiapine and Risperidone. In the typical drug category, Olanzapine is more commonly prescribed, especially in SLAM and C&I. Diazepam, Lorazepam, Promethazine and Zopiclone are more commonly prescribed drugs from the hypnotics & anxiolytics category.

4.5 ADR Timeline

The ADR timeline is a set of algorithms that define associations or signals between ADEs and drugs. The ADR timeline is generated by combining positive mentions of ADEs extracted by the ADEPt pipeline with the medication algorithm results. The ADR timeline can help to identify new signals between the possible and unknown causal relationships among the ADEs and drugs by establishing the temporal precedence of medication episodes over ADEs.

The ADR timeline algorithm is written in the SQL query language. The codes are available in CRIS repository. The pipeline consists of multiple steps that filter, arrange and manipulate data to identify ADE-Drug associations. As only document dates are available with no timestamps of data entry, ADEs discussed multiple times on the same date are collapsed into a single ADE event. The algorithm also discerns positive mentions from negative mentions and arranges the ADE data in chronological order for each patient. This is done by querying the medication algorithm to obtain a list of drugs the patient was on per date and creating a single drug-ADR event for each drug. A single-drug ADR event represents one ADE-Drug association (a positive hit for both ADE and medication at any given date). For every single drug-ADR event, the pipeline also adds the following at the time of the ADR event: age, ethnicity (stratified into four major categories), gender, and drug information (such as primary, secondary and tertiary category). Subsequently, the pipeline aggregates single-drug ADR events into a multi-drug ADR event to construct drug-drug interactions. This comprises of all the drugs that a patient was prescribed on the date of the ADE event. The flow of the process is explained in Figure 4.7.

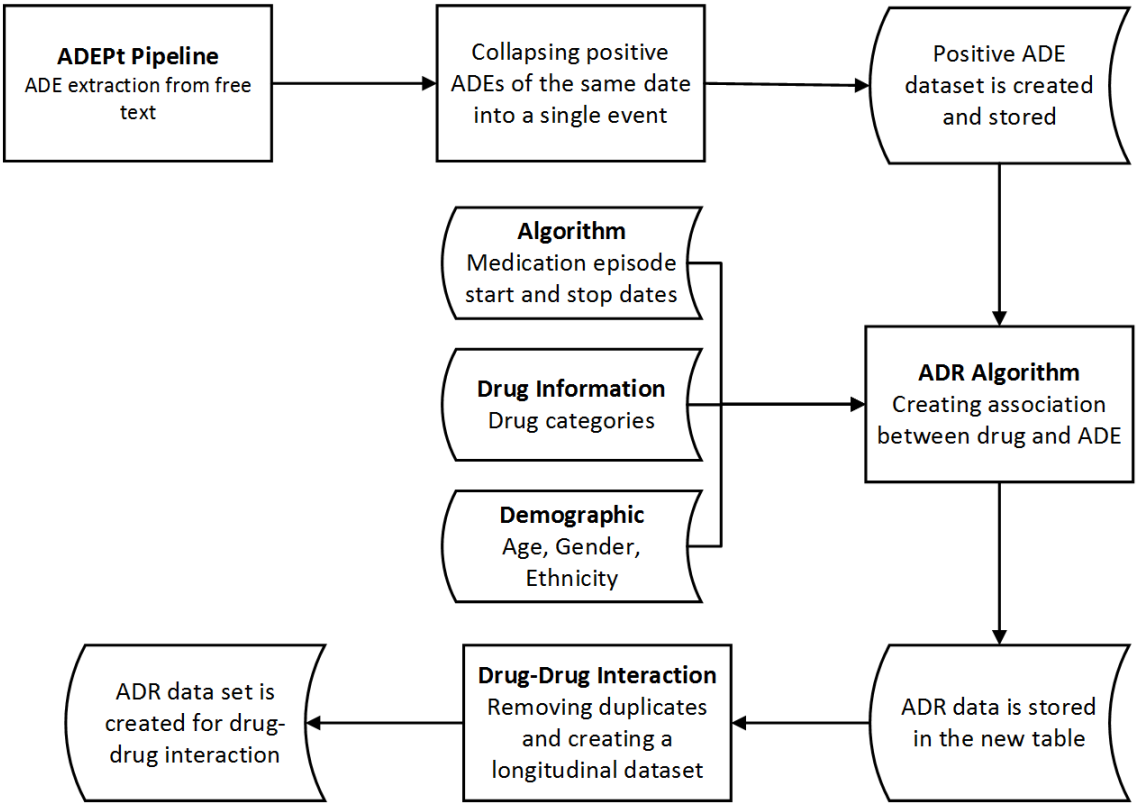

Figure 4.7: ADR timeline algorithm: the process shows the multistep approach to combine the ADE information with the on treatment medication episodes. The process starts by collapsing the positive ADEs of the same date into a single event and storing this information into a table. In the next stage, the ADR algorithm combines on-treatment medication episodes along with drug dictionary and demographics and creates a single ADE and Drug event. In the next stage, the pipeline aggregates the single-drug ADR event into a multi-drug ADR event to construct drug-drug interactions, which represent all the drugs that a patient was prescribed on the data of the ADE events

4.5.1 Implementation and validation

The ADR algorithm was implemented across the three NHS Trusts and was evaluated in SLAM and C&I to assess its performance in detecting associations between drugs and ADEs; access limitations prohibited evaluation in the Oxford Trust. A set of 300 cases were randomly selected from each trust, and manual validation was carried out by two annotators in SLAM and one annotator in C&I reading through the clinical notes according to the following criteria:

a) The presence of ADE is a true positive.

b) The associated medication episode annotation was a true positive

A true positive ADR is defined as the event where the ADE and medication are true positives. For each evaluation set, the PPV is defined as the percentage of ADRs that are true positive. The next focus was on False Discovery Rate (FDR), which defined the confidence level that a given ADR is a true positive. The FDR is related to PPV through the relation of FDR = 1-PPV. By defining the above criteria, the algorithm achieved a 0.89 PPV (95% CI) in SLAM and a 0.87 in C&I. The FDR (95% CI) was 0.1 in SLAM and 0.12 in C&I.

In SLAM, the level of agreement between the two annotators for drugs and ADEs are given in Table 4.6 with a percentage representing the agreement and Cohen's Kappa scores.

Table 4.6: Annotation agreement between two clinical annotators in SLAM.

	Agreement (%)	**Cohen's Kappa Score**
ADEs	93%	0.56
Drugs	94%	0.55

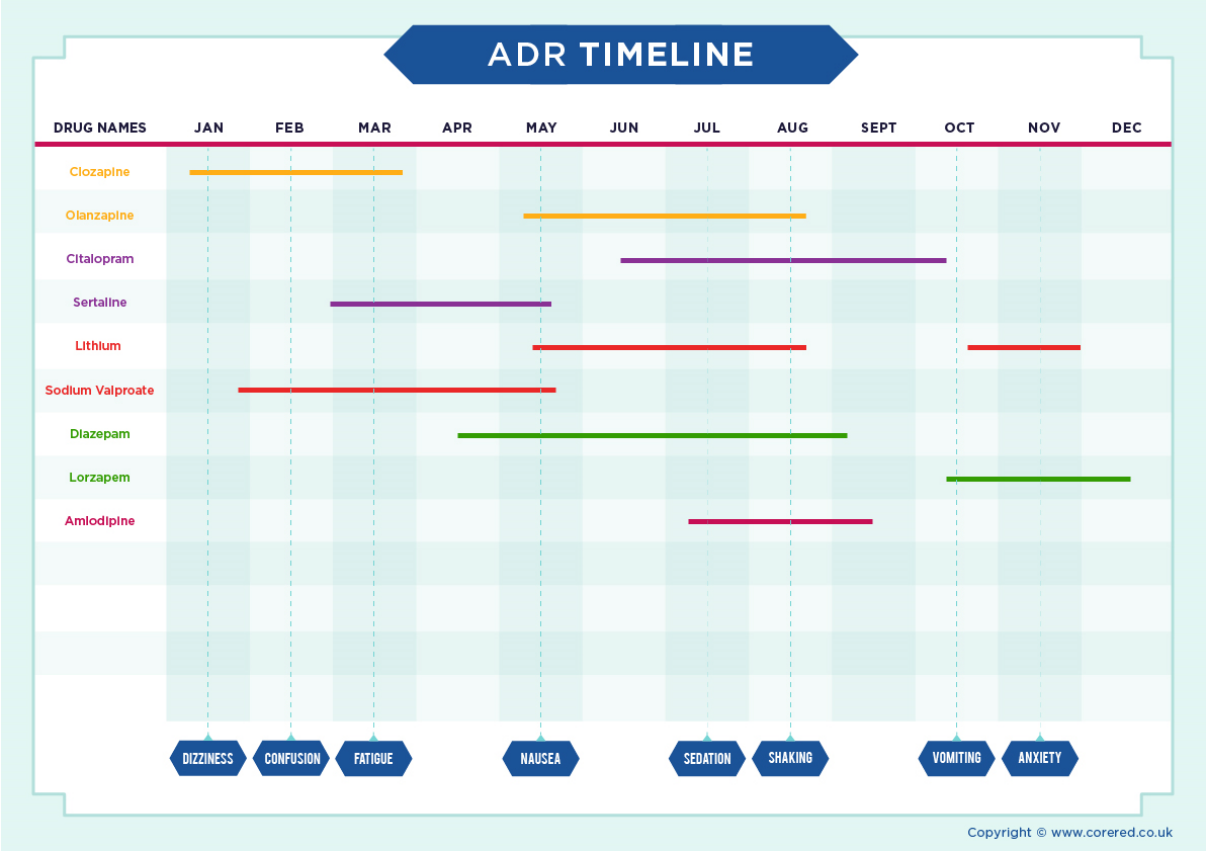

Figure 4.8: shows the ADR timeline for one patient over 12 months. The medication taken over this period in on y-axis and all the ADRs they suffered over this period is on the x-axis. The vertical dotted lined denote the drugs the patient was on during the ADR.

4.6 Discussion

ADRs are a major cause of mortality, morbidity and prolonged hospital stay. Additional interventions in response to ADRs are a significant burden on health services worldwide. ADRs have received significant attention from researchers since the early 1990s (Nebeker et al., 2004) and have become a major safety priority (Shojania et al., 2002). The ability to extract, and eventually predict, occurrences of ADRs could result in significant patient and cost benefits (Tatonetti et al., 2012). Detecting previously unknown ADRs can help health organisations and potentially reduce drug-related morbidities and mortalities. Moreover, despite the high incidence rates of ADRs, as well as their high rates of mortality and morbidity, clinicians are often unable to recognise and appropriately react to ADR incidents (Saposnik et al., 2016).

The assertion of classification (positive vs negative) and temporality (present, past, or retrospective) are active research areas in biomedical NLP. ADR extraction from clinical text relies on a number of components: i) NER of ADEs, drugs, symptoms and disorder, ii) being able to discern whether these named entities have been stated with negation, iii) the temporal constraints around their mentions, and finally, iv) being able to identify possible relations connecting multiple named entities.

In hospital settings, the patient accumulates a large number of documents over time. For example, if a patient is prescribed a medication, such event is recorded in one clinical note, but the ADE caused by this medication may occur at a later date and will be mentioned in a different document. Therefore, the temporal aspects of the annotations play a vital role in asserting a link between medications and ADEs. It is also essential to have a co-reference resolution between medications and ADEs. For instance, if a patient is on a medication such as anti-epileptic, not taking co-reference into account may result in associations such as the anti-epileptic drug causes epilepsy. In this work, the co-reference model was not developed.

There are a number of strengths and potential avenues and for further development and use of the ADR pipeline. Statistical and ML methods can be further applied to differentiate observed from expected ADR cases identified by the pipeline presented here. One can also explore data disproportionality to reveal the associations of interest according to demographics, smoking status, BMI, education, alcohol use, employment, welfare status, blood pressure, blood results, suicide behaviour, symptoms, CBT and

hospital admission status. The output of the ADR pipeline can be used to predict incidence rates, occurrences of ADRs within a specific period, such as the start of the drug therapy, treatment duration and the reasons for treatment withdrawal. The data can also be used to compare the expected frequencies with external data sources such as SIDER (Kuhn et al., 2015). The ADE and drugs signal detected in the ADR algorithm have considerably overlapped with other sources such as FAERS and SIDER, which implies the importance of EHR based pharmacovigilance and can be prioritised for follow-up epidemiolocal studies.

There can be some limitation of this work which should be addressed accordingly. Firstly, the GATE medication application is only assessed for single drug, Clozapine and has not been evaluated with any other drugs, although the work is in progress with other psychotropic drugs. Secondly, assessment of medication episodes and ADR timeline algorithm should have completed using expert reviews from a group of clinicians, nurses and pharmacist. Most importantly, although it is part of ongoing work, the current algorithms are not incorporating medication strength, average daily dose, dosage change information for the medication in question.

The current research shows that ML methods such as SVM and CRFs models have been used for relation extraction, and they can be further evolved into drug-ADE signal and drug-drug Signals. For example, Tang et al. (2013) used SVM models to extract temporal information, J. Xu et al. (2016) used the CRFs models to develop a system for chemical-induced disease challenge. Recent studies from the biomedical NLP domain have shown that deep learning models, such as CNN models, have outperformed traditional ML methods on relation extraction. In such attempts, MADEx (Yang et al., 2019) combined CNN and SVM, and their hybrid approach achieved good performance on identifying ADE, medication and their relation.

The proposed work has been replicated in the three NHS Trusts and has raised a significant interest within the research community in SLAM as well as in other Trusts. A number of studies have used these algorithms in their research; for example, the medication algorithm was used in a dementia study (Baker et al., 2017). The ADR timeline was used in a study to validate knowledge graph prediction and evaluation of ADRs in EHRs (Bean et al., 2017). The knowledge graph uses the semantic metadata and represents data in the context of interlinked entities such as ADE, protein targets, indication and drugs. The ADEPt pipeline was used in a study of antipsychotic

polypharmacy and parkinsonism side-effects (Kadra et al., 2018). Another study (Parkinsonian symptoms in Alzheimer's disease and vascular dementia: Co-morbid features and relationship to adverse outcomes) used the ADEPt pipeline and medication algorithm (submitted). Furthermore, the ADR timeline was used to conduct the systematic surveillance of Clozapine-related ADRs in three NHS Trusts (accepted).

4.7 Conclusions

ADRs are injuries caused by drugs and can result in serious consequences, including mortality. ADRs are closely monitored by regulatory authorities and postmarketing pharmaceutical surveillance such as clinical trials and SRS. However, due to the biases present in clinical trials such (e.g. sample size, demographics) and their high expenses, many ADRs are not recognised in early clinical trials and are only observed once the drug has been made available to the public. Clinical notes are written in a free-text format and are a rich source of ADE and medication information. This chapter presented NLP tools, which have been developed to extract ADE and medication information from the free-text. The temporal association between ADEs and medication plays an important role in discovering potential ADRs.

Chapter 5

5 The side effect profile of Clozapine in real-world data of three large psychiatric health providers

5.1 Introduction

Clozapine is an atypical, also known as a second-generation antipsychotic drug. It is widely recognised as the gold standard in the treatment of schizophrenia and the most effective antipsychotic drug in the management of treatment-resistant schizophrenia (Schulte, 2003; Stahl, 2000). Nevertheless, Clozapine is an underutilised medication (Kar et al., 2016), with only 54% of all eligible patients being prescribed Clozapine in the UK (Mortimer et al., 2010). One of the primary reasons for its limited use is the concern over its side effects, some of which are potentially fatal and require frequent monitoring (Krupp et al., 1992). Two common side effects associated with Clozapine are weight gain and hypoglycaemia, which together can lead to type II diabetes (Gianfrancesco et al., 2002; Nielsen et al., 2010). Others include abdominal pain, agitation, akathisia, amnesia, blurred vision, confusion, constipation, convulsions, delirium, delusion, diarrhoea, dizziness, dry mouth, enuresis, fatigue, fever, headache, heartburn, hallucination, hyperkinesia, hypersalivation, hypertension, hypotension, insomnia, nausea, rash, restlessness, seizures, sleeplessness, sweating, syncope, tachycardia, tremor, vomiting and decrease in White blood cells (WBC) (Gürcan et al., 2017; Miller, 2000; Safferman et al., 1991; Young et al., 1998). These side effects can be dose-related (Flanagan, 2008; Taylor et al., 2009). A more severe side effect is agranulocytosis, a severe and dangerous lowered WBC count. Therefore, as soon as the patient starts taking Clozapine, blood monitoring begins to determine whether or not the patient is at risk of agranulocytosis (Alvir et al., 1993; Amsler et al., 1977; Idänpään-Heikkilä et al., 1977; Krupp et al., 1992). Other severe but rare side effects include myocarditis, neutropenia, cardiomyopathy, Creatinine Phosphokinase (CPK) increase, hepatic narcosis and SJS (De Fazio et al., 2015).

Several studies have reported Clozapine-induced ADRs but have limitations; these include the duration of the study (6 weeks to 2 years), cohort size (31 to 110 patients)

and the number of the ADRs discussed (1 to 18) (Angermeyer et al., 2001; Hodge et al., 2008; Hynes et al., 2015; Lieberman et al., 1992; C. Schneider et al., 2014; Takeuchi et al., 2016). Limited work has been done to characterise a large population of patients experiencing Clozapine-induced ADRs.

Little work has been done to understand the relationship between age and Clozapine-induced ADRs. The literature suggests that there is a positive relationship between weight gain and cardiovascular risks, mainly myocarditis and cardiomyopathy (Haas et al., 2007; Leadbetter et al., 1992). In young patients, several short-term and long-term studies have reported Clozapine-induced ADRs in different hospital settings. However, those only covered a single or a few ADRs at a time, rather than a range of ADRs (Fleischhaker et al., 2006, 2008; Frazier et al., 1994; Gerbino-Rosen et al., 2005; Kumra et al., 1996; Sporn et al., 2007; Turetz et al., 1997; Wehmeier et al., 2004; Wudarsky et al., 1999).

Schizophrenia patients are more likely to be smoker as compared to patients with other psychiatric disorders (Dickerson et al., 2013). They are more likely to be heavy smokers (25 or more cigarettes daily) compared with only 11% of the general population of smokers (C. Kelly et al., 2000; Lohr et al., 1992). The chemicals in cigarette smoke induce enzymes that accelerate the metabolism of antipsychotic drugs (Lucas et al., 2013; S.-F. Zhou et al., 2009), requiring a higher dosage of Clozapine in smokers compared to non-smokers. There is no study showing comprehensive profiling of Clozapine-induced ADRs in the smoker vs non-smoker patient population.

Males and Females exhibit different responses to drug treatment (Anthony et al., 2002; Rademaker, 2001). Studies suggest that women are at a higher risk of ADRs than men, and ADR-related hospital admissions are more common in females and older patients (Hofer-Dueckelmann et al., 2011; Rodenburg et al., 2011). Some studies suggest that weight gain, hypertension, dyslipidemia and other metabolic syndromes are more prevalent in female Clozapine patients, whereas cardiac-related ADRs such as arrhythmia, QT prolongation are more prevalent in males (Choi et al., 2007; Furukawa et al., 2007).

Although often ignored, ethnicity plays a significant role in response to psychotropic medication (Chaudhry et al., 2008; Lin et al., 1986). Studies suggest Clozapine is used much less in the black population compared to any other ethnicity due

to the lower normal range of WBC count (Chaudhry et al., 2008; D. L. Kelly et al., 2006). Clozapine is also known to cause Type II diabetes in the black and Asian population (Citrome, 2004; Cohen, 2004; Nielsen et al., 2010).

This chapter demonstrates the use of the ADEPt pipeline to detect adverse events and medication episodes from the clinical text to enhance the understanding of adverse effects related to Clozapine. The chapter characterises the population experiencing ADRs with Clozapine and compares the results with the SIDER, complementing what is already known about demographics, smoking status and hospital admissions, and to find out how different subgroups are most likely to experience ADRs when administered Clozapine. SIDER contains side effects information from RCTs and systems such as the FAERS ("U.S. Food and Drug Administration,"). As of January 2018, SIDER contained 23 different sources on Clozapine adverse effects such as post-marketing, FDA, labels, Medsafe and Health Canada. Although SIDER collates data from a number of sources, much of its data is from RCTs, which are typically run on small and narrow populations, with little information on how medications work in real-world settings.

5.2 Material and methods

5.2.1 Data Sources

The study uses data of patients receiving care in SLAM (January 2007 to December 2016), C&I (June 2009 to December 2015) and Oxford (January 2010 to December 2014) NHS Foundation Trusts. The combined data from the three Trusts comprises over 500,000 patient records and over 50 million documents. Both structured and unstructured data was used within the EHRs to identify periods of on-off Clozapine medication, and then to identify subsequent Adverse Drug Reactions.

5.2.2 The Algorithm: Medication start and stop dates

As described in Chapter 4, Section 4.4, an algorithm was developed to infer ‘on-treatment’ medication periods. The study presented here uses the algorithm to identify patients actively taking Clozapine for at least three months.

5.2.3 Mining Adverse Events from clinical text

As described in Chapter 3, Article II, the ADEPt pipeline was used to mine ADEs from the free-text (E. Iqbal et al., 2017). An earlier version of ADEPt was applied to extract 66 ADEs on the Oxford trust dataset. Later on, the ADE dictionary was extended to 110 ADEs and was applied to extract 110 ADEs from the SLAM and C&I trust dataset (more recent work used a larger ADE dictionary). Both versions are using the same rulebase except the later version has a larger ADE dictionary.

5.2.4 Associations between Medications & ADRs: Formulating an ADR Timeline

As described in Chapter 4, Section 4.5, the study set out to uncover associations between medication episodes and adverse events by combining the medication timeline with the adverse events mined by the ADEPt pipeline. This is done in a number of steps: First, multiple discussions of an ADE on any one day are collapsed into a single event, taking into account negative mentions (to be filtered out). The ADR algorithm then queries the medication timeline to identify drugs that the patient was taking at the time of an adverse event and then creates an ADR event.

5.2.5 Statistical analysis

The study used the chi-square statistical method with Bonferroni correction to quantify the significance of ADR associations in relation to gender, ethnic background, age, and smoking and hospital admission status. The data taken into account was based on a monthly interval after starting the drug clozapine. R programming language version 3.2.4 was used to conduct statistical analysis.

5.3 Clozapine cohort and associated variables

Once periods of prescribing and associated possible ADRs were identified, the prevalence of ADEs across different subpopulations within the EHR was explored. As such, age, gender, ethnicity, smoking status, hospital admissions (inpatient/outpatient) were extracted for each patient in each Trust where data was available, as shown in Table 5.1. Although efforts were taken to replicate the study in all three NHS Trusts, the

identification and extraction of age groups, smoking status and inpatients/outpatients in Oxford NHS Trust data was not successful due to the resource constraints.

The date of birth, gender, ethnicity, smoking status, hospital admission status and diagnosis were retrieved from CRIS. Age was calculated on the date the patients started Clozapine and was further classified into eight distinct categories. Gender and ethnicity were derived from the latest entry recorded in each NHS Trust. Ethnicity was divided into four major groups, white, black, Asians and others. Smoking status was calculated six months before and after the date of starting Clozapine treatment. Hospital admission status was measured by looking into the patient admission and discharge data. If the patient started Clozapine during their hospital admission, they were classified as an inpatient.

The diagnoses were labelled using six categories according to ICD-10 codes. Three of these categories came from SMI, Schizophrenia (ICD-10: F20-F29) excluding (ICD-10: F25), Schizoaffective (ICD-10: F25) and Bipolar (ICD-10: F31). These diagnoses were extracted with a commonly used algorithm in SLAM EHR dataset to extract relevant diagnoses from the correct ICD chapter. The same algorithm was applied in C&I and Oxford trust datasets. The other three categories were any mental, behavioural and neurodevelopmental disorders (ICD-10: F01-F99) excluding SMI patients, as well as any other (or unavailable) diagnoses. Diagnoses were collected six months before and after the first prescription of Clozapine. In cases where no recorded diagnosis was identified, the search time span was increased until a diagnosis was established.

Table 5.1: Cohort characteristics of SLAM, C&I and Oxford NHS Trust, showing a breakdown of gender, ethnic background, age groups, smoking status, hospital admission status and diagnosis

Cohort	SLAM	C&I	Oxford	Total
Size	1760	561	514	2835
Gender				
Male	1167 66.3%	357 63.6%	342 66.5%	1866
Female	593 33.7%	204 36.3%	172 33.4%	969
Ethnic Background				
White	821 46%	347 62%	426 83%	1594
Black	704 40%	120 21.39%	20 4%	844
Asian	93 5.3%	41 7.31%	41 8%	175

Other	142 8%	53 9.45%	27 5.25%	222
Age Group				
Under 21	57 3.2%		12 2.33%	69
21-30	422 24%	27 4.81%	96 18.68%	545
31-40	488 28%	141 25.13%	155 30.16%	784
41-50	479 27%	168 30%	126 24.51%	773
51-60	233 13%	135 24%	98 19%	466
61-70	62 3.5%	67 12%	22 4.28%	151
71-80	18 1%	21 3.74%	4 0.78%	43
Above 80	1 0.06%	2 0.35%	1 0.19%	4
Smoking Status				
Smoker	1039 59%	360 64%		1399
Non-Smoker	721 41%	201 36%		922
Hospital Admission Status				
Inpatient	737 42%	114 20%		851
Outpatient	1023 58%	447 80%		1470
Diagnosis				
Schizophrenia (ICD-10: F20-F29) excluding (ICD-10: F25)	1355 77%	411 73%	356 69%	2122 75%
Schizoaffective (ICD-10: F25)	260 15%	45 8%	61 12%	366 13%
Bipolar (ICD-10: F31)	44 3%	26 5%	11 2%	81 3%
Any mental, behavioural and neurodevelopmental disorders (ICD-10: F01-F99) excluding SMI patients	32 2%	17 3%	11 2%	60 2%
Any other Diagnosis Excluding (ICD-10: F01-F99)	54 2%	42 7%	39 8%	135 5%
Diagnosis Not Available	15 1%	21 4%	36 7%	72 2%

SLAM = South London and Maudsley NHS Foundation Trust health record; C&I = Camden and Islington NHS Foundation Trust health record; Oxford = Oxford Health NHS Foundation Trust health record; n sample size

5.4 Results

The study identified 2835 patients who have taken Clozapine for at least three months in three large psychiatric health Trusts. Table 5.1 presents characteristics of the three cohorts with respect to gender, ethnic background, age groups, smoking status, hospital admission status and diagnosis.

The study results were compared with SIDER, where possible, to understand the prevalence of Clozapine-induced ADRs in the real-world EHR data. These results are summarised in Table 5.2 for 33 ADRs.

The ADR timeline algorithm was evaluated in SLAM and C&I and achieved a 0.89 PPV (95% CI) and 0.1 FPR (95% CI) in SLAM and a 0.87 PPV (95% CI) and 0.12 FPR (95% CI) in C&I. In SLAM the level of agreement between the two annotators for ADEs and Drugs is given in Chapter 4, Section 4.5, Table 4.6 with a percentage representing the agreement and a Cohen's Kappa scores.

The results are stratified into monthly intervals from the initiation of Clozapine treatment, three months prospective and three months retrospective. The columns (Three Months Earlier, Two Months Earlier, One Month Earlier, One Month Later, Two Months Later, and Three Months Later) show the percentages in each monthly interval. The last two columns (SIDER Low End and SIDER High End) show SIDER reporting from different clinical trials and FDA studies.

Table 5.2: Clozapine-induced ADRs in SLAM, C&I and Oxford NHS

ADR	Trust	Three Months Earlier	Two Months Earlier	One Month Earlier	One Month Later	Two Months Later	Three Months Later	SIDER Low End	SIDER High End
Agitation	SLAM	17.61	22.10	26.53	46.59	32.56	26.99		
	C&I	13.37	17.83	18.36	43.14	28.34	21.03		
	Oxford	14.59	15.76	16.34	34.24	25.10	20.62		
	SIDER							4.00	
Fatigue	SLAM	12.67	14.83	15.85	43.58	35.80	30.51		
	C&I	10.34	12.30	13.37	41.18	29.23	26.56		
	Oxford	9.73	11.87	12.06	35.21	27.43	26.85		
	SIDER								
Sedation	SLAM	12.67	12.16	14.83	43.86	35.51	29.83		

	C&I	5.17	9.09	9.09	38.15	26.56	21.93		
	Oxford	7.20	8.37	9.34	31.52	21.40	18.48		
	SIDER							25.00	46.00
Dizziness	SLAM	2.78	4.20	4.09	16.59	13.13	11.19		
	C&I	3.21	3.39	3.74	18.18	13.73	9.09		
	Oxford	3.89	4.09	4.47	17.70	13.04	10.12		
	SIDER							12.00	27.00
Hyper salivation	SLAM	1.19	1.48	2.10	14.32	13.24	11.31		
	C&I	1.07	1.43	0.53	14.26	6.95	7.66		
	Oxford	0.97	0.78	1.56	12.65	10.70	5.84		
	SIDER							1.00	48.00
Feeling sick	SLAM	4.66	4.94	6.48	14.32	11.19	9.09		
	C&I	3.74	3.92	3.03	10.52	7.13	7.66		
	Oxford	3.89	5.25	5.06	14.20	9.73	7.20		
	SIDER								
Weight gain	SLAM	3.75	4.43	5.06	15.34	10.91	10.34		
	C&I	2.50	3.39	1.96	11.76	6.60	6.24		
	Oxford	3.50	3.31	3.70	11.28	9.92	7.78		
	SIDER							4.00	56.00
Tachycardia	SLAM	2.27	2.05	2.50	15.40	12.95	9.94		
	C&I	1.43	1.43	0.89	11.23	8.38	6.95		
	Oxford	0.78	1.36	1.56	10.89	10.51	7.59		
	SIDER							11.00	25.00
Confusion	SLAM	4.72	5.51	6.08	13.92	8.47	6.76		
	C&I	3.57	6.24	5.53	12.66	6.77	5.88		
	Oxford	2.53	3.89	3.89	9.92	6.42	5.25		
	SIDER							3.00	
Constipation	SLAM	1.76	1.99	2.16	12.27	11.70	9.49		
	C&I	1.07	2.50	1.78	11.41	7.13	5.70		
	Oxford	0.58	0.97	1.36	10.31	7.78	7.78		
	SIDER							10.00	25.00
Headache	SLAM	4.20	4.55	5.45	12.44	8.18	5.91		
	C&I	2.32	3.57	4.28	9.27	6.42	4.63		
	Oxford	3.89	3.89	4.09	10.89	8.37	7.59		
	SIDER								
Insomnia	SLAM	3.92	4.03	5.17	10.40	6.48	4.03		
	C&I	3.57	3.39	3.74	8.91	3.39	4.28		
	Oxford	5.84	4.86	5.84	8.37	6.81	4.09		
	SIDER							20.00	33.00
Hyperprolactinaemia	SLAM	3.18	3.64	4.20	8.52	5.06	4.15		
	C&I	1.60	1.78	2.67	8.20	4.10	3.57		
	Oxford	3.70	4.09	4.28	8.75	4.47	4.86		
	SIDER								
Shaking	SLAM	3.13	2.95	3.92	9.55	5.40	5.06		
	C&I	1.78	1.96	3.74	6.06	3.92	2.85		
	Oxford	2.92	3.31	3.89	7.78	4.47	4.86		

	SIDER								
Vomiting	SLAM	2.56	2.50	3.01	8.86	6.82	5.00		
	C&I	2.14	2.50	2.85	6.77	4.99	4.63		
	Oxford	1.75	2.72	2.92	7.59	5.25	5.06		
	SIDER							3.00	17.00
Hypertension	SLAM	2.05	2.22	3.13	9.15	5.74	4.60		
	C&I	0.71	0.71	1.60	7.13	4.63	2.67		
	Oxford	1.36	1.36	1.56	5.06	4.28	2.14		
	SIDER							4.00	12.00
Abdominal pain	SLAM	1.88	1.99	2.56	8.01	6.02	4.72		
	C&I	0.89	0.89	1.60	3.92	3.57	3.39		
	Oxford	1.75	1.36	1.75	7.39	4.47	5.64		
	SIDER							4.00	
Convulsion	SLAM	1.36	1.70	1.82	7.05	4.94	4.03		
	C&I	0.53	0.53	0.36	2.85	2.14	1.07		
	Oxford	1.36	1.36	1.56	6.42	3.11	2.72		
	SIDER							3.00	
Backache	SLAM	1.14	1.59	2.44	4.94	3.35	2.73		
	C&I	1.43	1.25	1.96	5.35	3.03	3.03		
	Oxford	1.17	1.17	1.36	5.84	3.89	2.92		
	SIDER							5.00	
Nausea	SLAM	1.14	1.08	1.19	6.08	5.23	3.69		
	C&I	0.89	1.43	0.36	4.63	3.57	3.57		
	Oxford	0.97	0.58	1.36	4.86	3.70	2.92		
	SIDER							3.00	17.00
Hypotension	SLAM	0.51	0.97	0.80	5.00	2.95	2.56		
	C&I	0.18	0.53	0.18	3.57	2.32	1.78		
	Oxford	0.58	0.78	0.78	5.64	3.50	2.92		
	SIDER							9.00	38.00
Fever	SLAM	1.02	1.14	1.65	6.36	4.43	3.13		
	C&I	0.89	0.89	0.53	3.74	2.67	0.89		
	Oxford	0.39	0.78	0.58	3.11	2.72	2.33		
	SIDER							4.00	13.00
Enuresis	SLAM	1.02	0.80	1.25	4.20	3.92	3.24		
	C&I	0.71	1.07	1.07	4.10	1.43	1.25		
	Oxford	1.36	0.58	1.36	4.86	4.47	3.50		
	SIDER								
Dry mouth	SLAM	1.08	1.53	1.65	4.66	3.69	2.33		
	C&I	1.25	1.25	1.07	3.92	2.14	0.89		
	Oxford	1.36	1.36	1.56	3.89	1.36	2.33		
	SIDER							5.00	20.00
Diarrhoea	SLAM	1.08	1.31	1.36	4.72	3.58	2.56		
	C&I	0.71	1.25	0.18	3.03	3.39	3.03		
	Oxford	1.17	0.78	1.36	4.09	3.70	2.53		
	SIDER							2.00	
Rash	SLAM	1.25	1.59	2.05	3.64	2.95	2.27		

	C&I	1.25	1.25	0.89	4.28	1.96	2.14		
	Oxford	0.97	1.17	1.17	3.70	2.33	1.36		
	SIDER								
Dyspepsia	SLAM	0.74	1.08	0.91	3.92	3.13	3.69		
	C&I	0.36	0.53	0.53	4.10	2.67	2.50		
	Oxford	0.19	0.58	0.78	3.50	4.09	3.70		
	SIDER							8.00	14.00
Stomach pain	SLAM	1.93	1.76	1.93	4.94	3.52	3.52		
	C&I	0.89	1.25	0.89	3.39	2.85	2.14		
	Oxford	1.56	0.78	0.78	2.14	0.97	0.97		
	SIDER								
Sweating	SLAM	1.08	0.97	1.36	4.43	4.26	2.84		
	C&I	0.53	0.53	0.53	2.85	2.14	1.96		
	Oxford	1.17	0.97	0.97	2.72	1.36	1.95		
	SIDER							6.00	
Tremor	SLAM	1.48	1.99	2.95	5.51	3.52	3.47		
	C&I	1.60	1.78	2.14	3.92	1.96	2.14		
	SIDER							6.00	
Neutropenia	SLAM	0.80	0.80	0.74	5.34	2.73	2.61		
	C&I	0.00	0.18	0.53	1.60	0.89	1.07		
	SIDER								
Akathisia	SLAM	0.80	0.91	0.74	2.67	1.36	0.80		
	C&I	0.00	0.53	0.00	1.25	1.07	0.53		
	Oxford	0.97	0.78	0.97	1.36	1.17	0.97		
	SIDER							3.00	
Blurred vision	SLAM	0.34	0.91	0.63	2.05	1.25	1.02		
	C&I	0.89	0.53	0.71	1.25	0.36	0.89		
	Oxford	0.19	0.39	0.39	1.56	1.56	1.17		
	SIDER							5.00	
	Legend	0	10	20	30	40	50		

The results are shown in percentages (%) and broken down by ADRs (ADRs = Adverse Drugs Reactions), Trusts (SLAM = South London and Maudsley NHS Foundation Trust health record; C&I = Camden and Islington NHS Foundation Trust health record; Oxford = Oxford Health NHS Foundation Trust health record) and SIDER. The columns (Three Months Early, Two Months Early, One Month Early, One Month Later, Two Months Later, and Three Months Later) shows the percentages in each monthly interval. The last two columns (SIDER Low End and SIDER High End) shows the SIDER reporting.

The complete results stratified into monthly intervals are provided in the supplementary material of Chapter 5 for gender (Supplementary Table B.1), ethnicity (Supplementary Table B.2), and age groups (Supplementary Table B.3), smoking status (Supplementary Table B.4) and hospital admissions status (Supplementary Table B.5).

The statistical analysis was first carried out separately for each of the psychiatric health providers. The results are available in Chapter 5 (Supplementary Table B.6). The p-value of the test is adjusted through Bonferroni correction. Gender and ethnicity showed no significant association with any ADRs. Age group showed significant association with agitation, fatigue, feeling sick, sedation and tachycardia. In hospital admission, of the 33, 15 ADRs (abdominal pain, agitation, confusion, dizziness, diarrhoea, fatigue, headache, hypersalivation, hypotension, hypertension, insomnia, sedation, tachycardia, tremor and vomiting) were found to have a significant association. Smoking status showed that out of 33 ADRs, 8 ADRs (abdominal pain, agitation, confusion, dizziness, diarrhoea, fatigue, sedation and tachycardia) were found to have a significant association.

The datasets from the three Trusts were later combined, and chi-square statistical test was performed to estimate the average effect of ADRs in each monthly interval. The combined analysis shows significant frequency distribution after Bonferroni p-value adjustment in the categorical variables (gender, ethnicity, age group, hospital admissions and smoking status) and a number of the ADRs as follows:

a) Gender showed significant associations (See Figure 5.1) with backache, constipation, diarrhoea, fatigue, feeling sick, hyperprolactinaemia and stomach pain. The results demonstrate that dizziness, weight gain, constipation, abdominal pain, backache, diarrhoea, hypotension, hyperprolactinemia, fatigue, and enuresis were more prevalent in females.

b) Ethnic background showed no significant associations with any ADRs.

c) Age group showed significant associations (See Figure 5.2) with agitation, fatigue, feeling sick, sedation, shaking, tachycardia and weight gain. The results show that agitation, sedation, dizziness, insomnia, convulsions, tachycardia and tremor were more prevalent in patients under 30 years of age. ADRs such as dry mouth, enuresis and hyperprolactinemia were prevalent in patients over 40 years of age, and dizziness, hypotension and hypertension were more prevalent in patients who are over 60 years old.

d) Hospital admission status showed significant associations (See Figure 5.3) with abdominal pain, agitation, akathisia, backache, confusion, constipation, convulsion, diarrhoea, dizziness, dry mouth, dyspepsia, enuresis, fatigue, feeling sick, fever, headache, hyperprolactinemia, hypersalivation, hypertension, hypotension, insomnia,

nausea, rash, sedation, shaking, stomach pain, sweating, tachycardia, tremor, vomiting and weight gain.

e) Smoking status showed associations (See Figure 5.4) with abdominal pain, agitation, backache, confusion, convulsion, diarrhoea, dizziness, dyspepsia, enuresis, fatigue, feeling sick, fever, headache, insomnia, sedation, shaking, stomach pain, sweating, tachycardia, vomiting and weight gain. The results are available in Chapter 5, Table C.7 supplementary material.

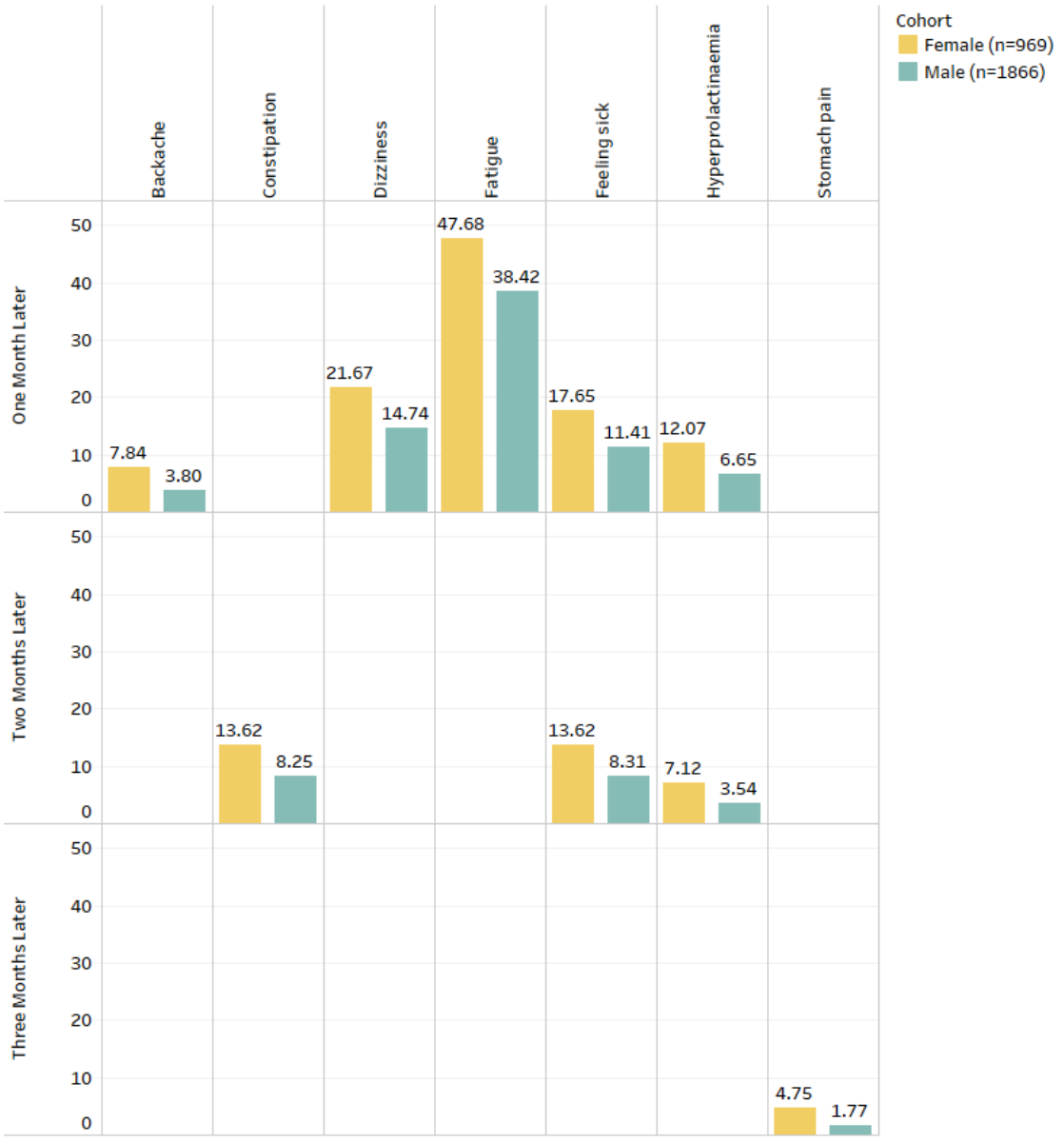

Figure 5.1: Frequency distribution of statistically significant ADRs (after Bonferroni p-value adjustment) from the combined analysis in gender for three months after starting the drug Clozapine.

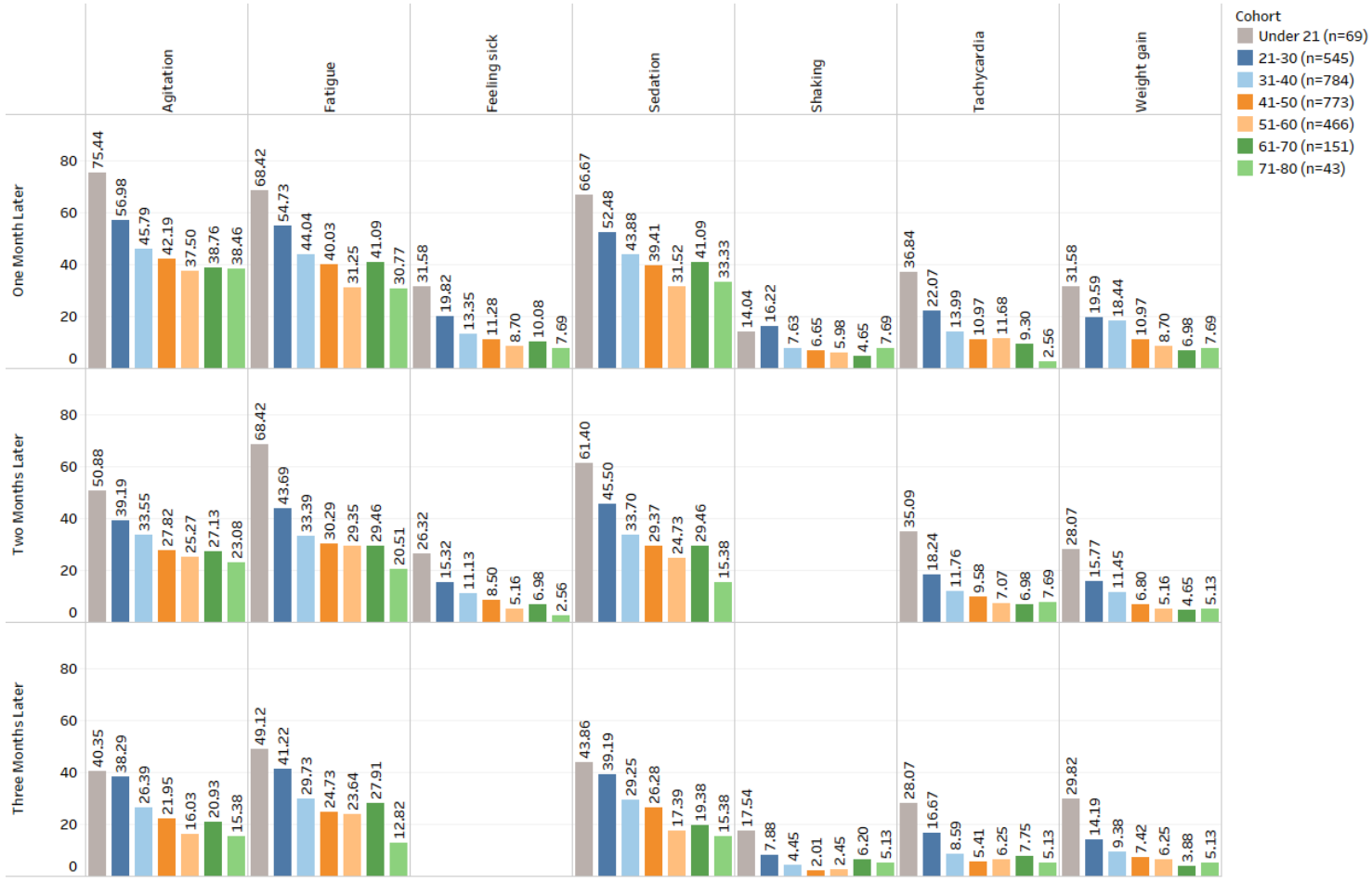

Figure 5.2: Frequency distribution of statistically significant ADRs (after Bonferroni p-value adjustment) from the combined analysis in age groups for three months after starting the drug Clozapine.

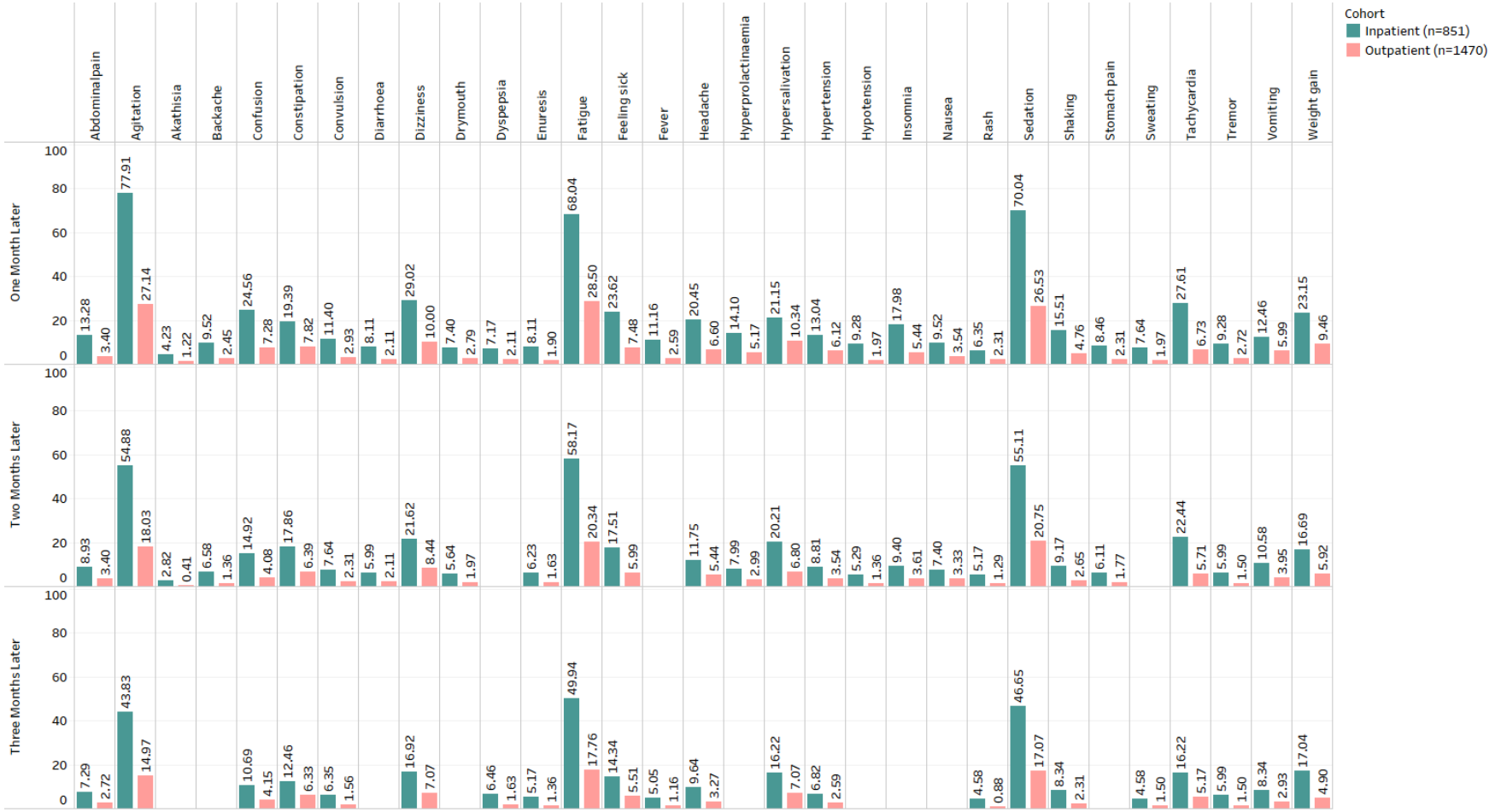

Figure 5.3: Frequency distribution of statistically significant ADRs (after Bonferroni p-value adjustment) from the combined analysis in hospital admission for three months after starting the drug Clozapine.

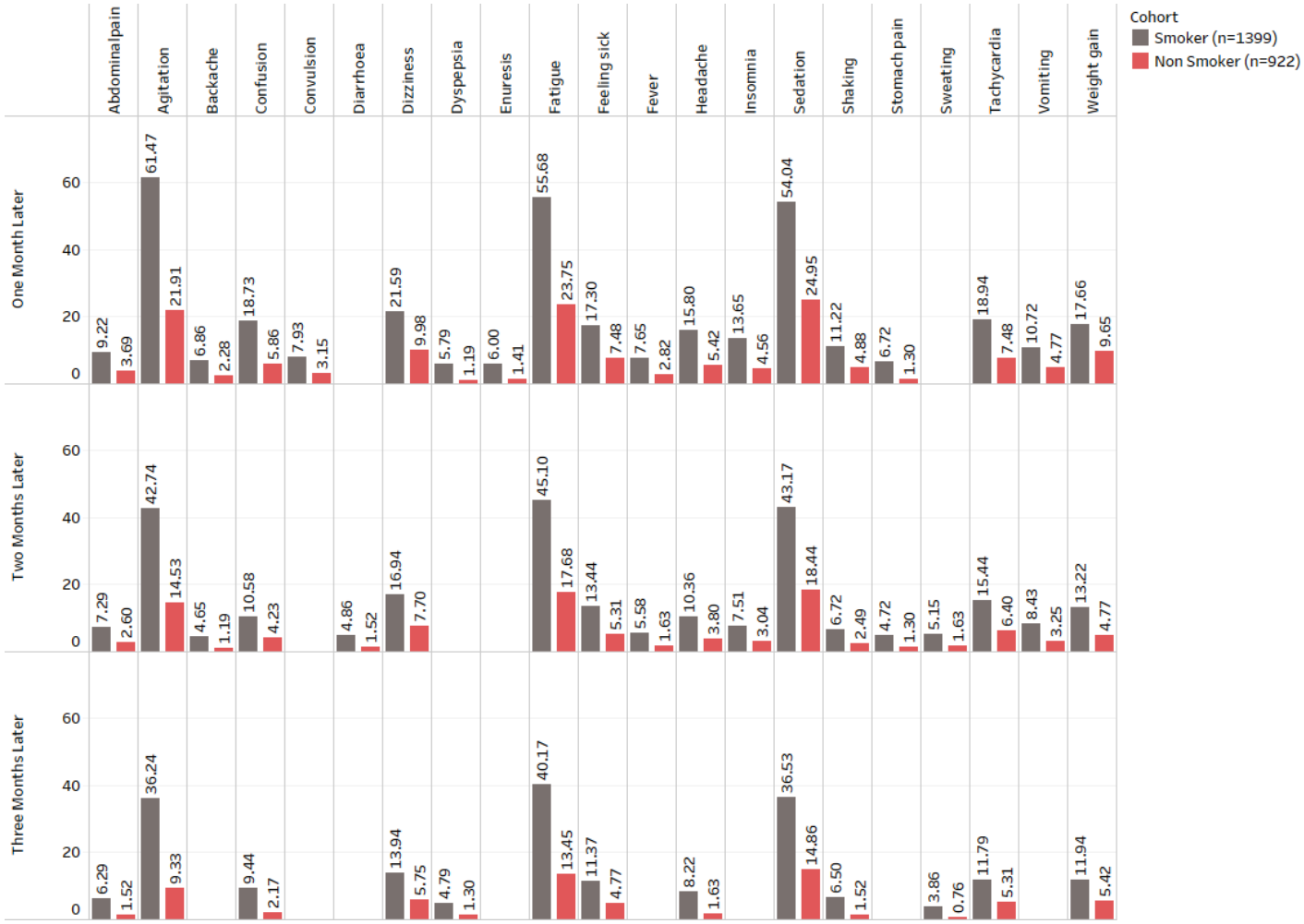

Figure 5.4: Frequency distribution of statistically significant ADRs (after Bonferroni p-value adjustment) from the combined analysis in smoking status for three months after starting the drug Clozapine.

5.5 Discussion

The study presents a medication continuity timeline (start and stop dates) for patients under Clozapine treatment to obtained detailed insight of Clozapine-induced ADRs using data from three large UK-based psychiatric health Trusts comprising over 50 million documents and over half a million patients. The study uses the ADEPt NLP pipeline (Chapter 3) to extract ADEs from free-text psychiatric EHRs, as well as a set of algorithms for creating a medication continuity timeline (Chapter 4). The timeline was used to investigate associations between medications and ADEs, characterising ADR susceptibility with respect to patient demographics, hospital admission and smoking status. Furthermore, the medication algorithm was used to assert the start date and measure the length of drug therapies.

The algorithm found 2835 patients in the three Trusts that started and continued Clozapine treatment for at least three months. The work presented in this chapter has been run inclusively on all available 110 ADEs in the ADEPt pipeline. For this study, 33 ADRs were selected for further analyses, as these were known side effects associated with Clozapine. The results were compared with SIDER, an online side effect resource. SIDER only reports on 25 out of the 33 ADRs. The 8 ADRs not reported in SIDER are fatigue, feeling sick, headache, hyperprolactinemia, weight loss, shaking, rash and stomach pain. These 8 ADRs were added as they showed higher coverage across three Trusts, as shown in Table 5.2. The prevalence of ADRs was assessed over gender, ethnic background, age groups, smoking status and hospital admission.

The data was stratified into demographics information, mainly gender, ethnic background, age groups, hospital admission and smoking status. Sedation, fatigue, agitation, hypersalivation, tachycardia, constipation, dizziness and weight gain are the most common (Frogley et al., 2012; Gareri et al., 2008; M. M. Iqbal et al.; Raja, 2011; Raja et al., 2014) and highest recorded ADRs in all three NHS Foundation Trusts. When comparing hospital admission data (inpatients vs outpatients), the results show that the inpatient group has a higher recording of any ADRs when the patient starts Clozapine therapy compared to the outpatient group. This is due to the inpatient group being more frequently monitored and recorded by clinicians. A similar pattern was observed when comparing smoking status (smoker vs non-smoker).

Rare ADRs such as agranulocytosis, myocarditis, SJS, cardiomyopathy and pericarditis are reported in the analysis. However, rare ADRs require much attention before they can be declared as positive. Currently, two of the on-going studies are using this dataset for myocarditis and agranulocytosis by going through the clinical notes manually to improve ADE assertions. Currently, work is underway on the new antipsychotic Lurasidone, as well as drug-drug interactions with Clozapine.

The future work focuses on extending the analyses to more extended periods, other psychotropic drugs such as Lurasidone, drug-drug interaction of Clozapine, dosage related information and effect. The future studies aim to understand the risk factor of Clozapine-induced ADRs related to the start and during therapy, misuse, dependence, discontinuation and withdrawal. Further research will be carried out to understand the prevalence and cluster of patients and ADRs by adding more patient-related features such as demographics, polypharmacy, Body Mass Index (BMI), education, mobility, employment status, welfare benefits, homelessness, blood results and alcohol use.

5.6 Limitation

This study was first conducted and evaluated using SLAM's psychiatric clinical notes. In SLAM, evaluation of each step was carried out manually by at least two annotators, and an inter-annotator agreement was achieved where possible. Due to limited resources, several challenges were faced when implementing the work in C&I and Oxford Trusts. The ADEPt pipeline and ADR timeline were not manually validated in Oxford Trust and the results presented in the C&I Trust are based on a single annotator. Although the proposed work has shown good results, lower precision and recall were found in rare ADEs as they are frequently recorded as warnings, potential and suspected occurrences.

5.7 Conclusions

Clozapine is becoming a more popular drug of choice due to tremendous evidence of its effectiveness in treatment-resistant schizophrenia. Since 2008 there is an upward trend to understand and study Clozapine-induced ADRs which is associated with major awareness, wider availability and rise in pharmacovigilance (Chiappini et al., 2020; Nielsen et al., 2012; Verdoux et al., 2016). A number of studies have reported Clozapine-induced ADRs but were limited in size and ADRs in question. One of its kind in terms of

cohort size and the variety of ADRs, this study characterises and provides insight into Clozapine-induced ADRs on a large population of patients across three large psychiatric health providers. As well as providing novel findings, the proposed method demonstrates the utility of wider ADR extraction beyond Clozapine, and the study can be replicated using any psychotropic drug. In the future, this work will be expanded to define extended periods of on-treatment episodes to understand risk factors associated with comorbidity, Clozapine-induced ADRs during and initiation of therapy, and withdrawal.

Chapter 6

6 Summary, Discussion and Conclusions

6.1 Summary of Principal findings

The thesis aimed to improve the understanding of how ADRs can be extracted from the free-text clinical documents. Clinical trials, SRS and post-marketing surveillance are the most common methods to identify ADRs, but both are limited by narrow patient populations and being sporadic. These limitations can be addressed by analysing EHRs to uncover the prevalence of ADRs in real clinical settings using large patient populations. In this thesis, a variety of components were studied, and analyses were performed in health records made available for research through CRIS at SLAM, C&I and Oxford NHS Foundation Trusts. The medication, diagnosis, ADE and ADR NLP applications were developed and validated using SLAM CRIS and were further implemented in C&I and Oxford NHS Trusts. CRIS contains a large volume of free-text documents, full of clinical narratives, and automating the extraction of ADE information presented challenges. This thesis reports the results of two iterations of the development of a tool for ADE detection from clinical text. The first iteration (Chapter 2) focused on detecting mainly EPSEs from clinical text and studied the prevalence of EPSEs in patients with SMI diagnoses. The second iteration (Chapter 3) comprised a broader coverage of ADEs related to psychotropic drugs. Temporal constraints and context-discerning rules were imposed on the discovered ADEs in both iterations but were expanded in the second iteration. The thesis (Chapter 4) further discussed the identification of on-treatment medication episodes and associations between medications and ADEs to uncover possible ADR events. Having developed tools linking ADEs to medications to uncover possible ADRs, (Chapter 5) presents a study to examine Clozapine-induced ADRs in SLAM, C&I and Oxford NHS Foundation Trusts. Studying Clozapine-induced ADRs is driven by the fact it is the only treatment available for treatment-resistant schizophrenia. It is hoped that through the analysis of data collected in EHRs, we will develop a better understanding of treatment outcomes and aetiology in a real-world setting. Comparing the tool's results with SIDER confirmed findings of the methods developed within this thesis in the context of Clozapine.

6.2 Discussion

EHR-based pharmacovigilance has been discussed in Chapter 1 of this thesis in great length. However, an extended critique of work where it is compared with other NLP development is discussed here. First, the ADEPt pipeline is freely available and has been implemented in three other psychiatric health providers, representing portability. The performance of this tool has not been evaluated in general health settings. Secondly, the medication episodes algorithm can be used on its own to determine medication-related information (such as start and stop date, treatment duration and polypharmacy) and along with the ADEPt pipeline, and it has a use case to assess causality and determined possible ADR events. The developed pipeline can be used as an active ADR surveillance system.

The work presented in this thesis can be extended to other ADEs and drugs of interest by adding them into the current piece of work as it stands. The ADEs and drugs terms can be found on the free-text clinical notes, and the NLP tools presented can be used to identify these terms. However, pharmacovigilance implementation of NLP explicitly attributed where drugs and ADEs are present on the same document (A. B. Chapman et al., 2019; Jagannatha et al., 2019). The current models do not explicitly consider events of casualty and have poor performance when identifying the association between ADEs and drugs (Dandala et al., 2019; D. Xu et al., 2018). The medication timeline presented in this thesis explicitly determined how long a patient is taking a drug, and the ADE timeline algorithm determined the association between the ADEs and drug and generates and possible ADR event, regardless of both events not being presented on the same document. The recent work which incorporated temporality on ADEs and drugs into signal detection of ADRs are widely based on the structured data (Harpaz et al., 2012; Pacurariu et al., 2015; Star et al., 2015). However, the work presented in this thesis has used both structured and unstructured data. The ADR algorithm has been optimised on patient level, across event assessment presented in different documents such as ADEs and drugs and identifies positive ADE caused by a drug. The algorithms presented in this thesis are portable, have already been tested in two other psychiatric health providers and can be adopted by any English narrative EHR system, in any other country.

6.2.1 Implications

This thesis detailed the development of NLP tools to detect and annotate ADEs (Chapters 2 and 3), medication episodes and possible ADRs (Chapter 4) from clinical text in psychiatric settings and presented a case study of identifying Clozapine-induced ADRs in three large psychiatric health providers (Chapter 5). The section discusses key implications of this thesis.

6.2.1.1 Detecting ADEs from clinical text

The ADEPt pipeline is now a part of a quarterly concept-extraction cycle managed by the SLAM NLP team, adding new ADEs from free-text clinical documents every cycle. It is worth noting that the ADEPt pipeline can be used in general health settings; however, independent validation of this work is essential.

ADEPt has been used in different studies, some of which have already been published. In one study, the ADEPt pipeline was used to study Clozapine-induced rare cardiac ADEs, in which it identified 115 suspected myocarditis complications from 2007 to 2015. These preliminary findings were put forward as an MSc Project aiming to establish the risk rates of cardiac complications of Clozapine compared with other antipsychotics. The student worked with the guidance of a psychiatrist and a cardiologist from KCH. Out of 115 cases, 90 possible myocarditis cases were found. The project is underway in the second year now with a larger cohort of 290 patients from 2007 to mid-2018; their findings will be published once the project is concluded.

In another study, ADEPt was used to identify patients with Parkinsonism side-effects (n=832) who were receiving two or more antipsychotics drugs for at least six months in SLAM (Kadra et al., 2018). Additionally, ADEPt was used to establish a cohort of over 11,000 patients with dementia, given an Alzheimer's disease diagnosis, in SLAM, and was used to identify Parkinsonian motor symptoms (bradykinesia, Parkinsonian gait, rigidity, tremor) at the time of dementia diagnosis (manuscript in preparation - Parkinsonian symptoms in Alzheimer's disease and vascular dementia). Finally, a subset of the validated ADEPt ADEs was used in knowledge graph prediction of unknown ADRs. The study also used the manually curated corpora to validate its results (Bean et al., 2017).

6.2.1.2 The medication timeline can be used to identify medication episodes

The medication timeline developed in this thesis can be used to identify when a patient started a medication, how long they were on a certain medication, change of a medication in a similar category and polypharmacy. However, dosage information was not a part of the current piece of work.

In Baker et al. (2017), the drug timeline was used to extract 45 medications in the category of dementia, lipid regulatory, antihypertensive, antidepressants, antipsychotics, antidiabetic, NSAIDs, mood stabilisers and antiepileptics to study the dementia-related cognitive decline in SLAM and C&I Trusts. The medication timeline was used to establish the medication continuation six months pre and postdate of the first Standardise Mini-Mental State Examination.

Furthermore, the medication timeline was used to identify prescribed drugs at the time of Parkinsonian symptoms, such as antipsychotics (with low, moderate and high potential of EPSE), antidepressants and hypnotics and anxiolytics (manuscript in preparation - Parkinsonian symptoms in Alzheimer's disease and vascular dementia).

6.2.1.3 The ADR timeline can be used to detect possible ADR events

The ADR timeline establishes the temporal precedence underlying medication and ADEs to identify known and unknown ADR signals. The ADR timeline can be used to study patient subpopulations, as it is equipped with demographic information. Furthermore, the ADR timeline can be used to study drug-drug interaction, withdrawal effects of a medication, and chronic and acute ADRs. The ADR timeline can also be aligned with other phenotypes such as smoking and hospital admission status, as demonstrated in (Chapter 5) for the Clozapine use case. In SLAM, the dataset created for Clozapine-induced ADRs is currently being used by another PhD project aiming to expand the current cohort and adding more categorical variables such as BMI, education, mobility, employment status, welfare benefits, homelessness, blood results and alcohol use. The project further make use of this data to develop a prediction model for Clozapine-induced neutropenia leading to agranulocytosis.

Moreover, the ADR timeline was used to evaluate the knowledge graph prediction of unknown ADRs. The proposed model was validated against the ADR timeline for a drug-ADR association. In this study, validation was carried out when a patient prescribed a single drug and an ADR was reported within 30 days (Bean et al., 2017).

More generally, the work presented in this thesis highlights the importance of replication of already defined methods. Through Chapter 2 to 5, the thesis showed that the already established methods could be replicated in similar settings.

6.2.2 Limitations

In addition to limitations specific to each application, which have been outlined in the corresponding chapters of this thesis, a number of general limitations are presented here.

EHRs are dynamic troves of information and a number of issues that should be considered when working with EHR data, as outlined in Chapter 1. These issues include completeness, biases, accuracy, inconsistency, change over time and human error (Hripcsak et al., 2012). Data completeness is influenced by the fragmented collection of EHR data (Kush et al., 2008). Patients only get in contact with the hospitals when they have health problems. Therefore, the sickest group of patients will generally have more data as compared to healthier patients. In psychiatric settings, clinicians tend to record not only detailed description of the medical problem but also information relating to medications, diagnoses, test results, social activities, as well as information given by caregivers and family members. Psychiatric clinical notes are often transcribed from what patients have stated about their clinical problems, which can be speculative as many mental health patients can be delusional. As a result, extra care must be taken to differentiate true positive from false positive mentions of ADEs.

EHR data have limitations that can be inherent throughout the study design and tools discussed within this thesis, such as missingness of diagnosis, ADEs, drugs, and symptoms in active vs inactive patients, leading to bias findings. These limitations have been looked at thoroughly before designing research studies presented in this thesis, but they still exist due to the nature of EHR. These limitations may extend beyond many researchers' capabilities and still be prevalent for any intended use of these tools for current and future pharmacovigilance and epidemiological studies. The EHR data is ever-evolving, new variables are added, and old data variables are obsolete such as diagnosis codes. Feder (2018) describes three general recommendation which should be implemented in all EHR based studies. First, all researchers should know data sources such as data format (structured or unstructured), data currency, data timelessness, and precise knowledge of data source that can help researchers anticipate any potential data quality issues. A data dictionary should be maintained that describes all the data variables

(structure, data type, intended used and any potential issues of which a researcher should be aware). Second, a research plan deals with data quality issues based on the prior knowledge of EHR and other secondary data sources such as pharmacy system. Researchers should include manual and statistical procedures for handling missing values and data quality issues such as the wrong date of birth, lab reports extreme values. Third, these findings should be documented and reported in the studies and intended use of this data for further studies.

Two approaches were considered when creating the ADE NLP application, an ML and a rule-based approach. The ML approach required minimal manual coding but gave imprecise results due to the nature of psychiatric settings as it needed a broad representation of examples. The ML ADE NLP was developed on SLAM CRIS, and 0.77 precision and 0.67 recall were achieved. In order to improve performance, it was decided to take the rule-based approach (Chapter 2, Article I and Chapter 3 Article II). This aspect of the application was developed with the close aide of pharmacists who possessed the domain knowledge and the understanding of psychiatric clinical notes, achieving an overall 0.89 precision and 0.88 recall. The limitation of the rule-based approach is that it is time-consuming to develop, requires a large degree of manual coding and domain expertise, and is not generally transferrable to other domains. Also, rule-based systems used attribute-specific and pattern matching rules, which are often rigid in dealing with variations in language expressions, resulting in high precision, but low recall. To partially alleviate the issue with low recall, this thesis developed generic rules rather than ADE-specific ones.

One of the challenges in identifying ADE concepts from clinical text was creating the ADE dictionary along with synonyms, alternate spellings and different phrases used by clinicians to describe an ADE. The ADE dictionary was created with the help of a pharmacist and clinicians who have a number of years of experience in using clinical notes. In the ADEPt pipeline, the ADE terms present in MedDRA were also added in the dictionary. Although the ADEPt pipeline has shown good results, generalising it to other domains faces the inherent limitation of requiring domain expertise to expand the current ADE dictionary.

Another critical limitation of the ADE detection work is that it only classifies ADE mentions using the context surrounding them in the clinical note and does not consider the measured readings and laboratory test results to conclude the presence and severity of

a certain ADE (e.g. fever, blood pressure, High Creatine Phosphokinase (CPK) level and prolonged Cardio Contraction Time (QT)).

The EPSE analysis (Chapter 2) in this thesis was carried out using the EHR cohorts in SLAM and C&I. Although the size of the cohort is an advantage as the cohort consists of large numbers of patients spread over a number of years, the results showed that many patients experiencing EPSEs have a missing diagnosis, which can lead to biased samples and analyses. The results showed high cases of akathisia as compared to other EPSEs. This may be due to the misdiagnosis of agitation, which can often be mistaken for akathisia as both are symptoms of schizophrenia (Ratey et al., 1984).

Misclassification of ADEs and the association between ADEs and drugs is also a limitation of this study. These may be the result of clinicians not having recognised a patient's symptom as a result of an ADE or failing to include an ADE as a cause of an injury (T.-Y. Wu et al., 2010). In this case, adverse events pipelines were evaluated manually on a larger unseen sample. The incidence of ADRs were assessed on the ADR timeline on a small sample size in SLAM and C&I, where the mention of an ADE is positive, and the patient is currently using a particular drug. The availability of accurate information on drugs, dosage and treatment compliance is essential to rule out an ADR (Vermeer et al., 2016). However, the results presented in this thesis do not consider dosage information.

In the study of Clozapine-induced ADRs in the three large psychiatric health providers (Chapter 5), only a few categorical variables were used for comparison to facilitate replication in all three NHS Trusts. Age groups, smoking and hospital admission status were not included in the Oxford NHS Trust study due to our inability to access these variables. The data from Oxford and C&I was only available in aggregate number format, limiting the applicability of analyses and statical methods. The gold standard manually curated data of Clozapine medication episodes was only available in SLAM. Finally, this study relied on the local researchers to understand CRIS's data and structure in C&I and Oxford, including how the data was collected, stored and the completeness of the different variables used.

6.3 Future direction

The research implications and limitations outlined above highlight the importance of the use of EHRs in research and set the scene for future directions. Embracing the EHR has significantly increased the amount of detailed patient information available today, which was not possible by manually reviewing patient charts. The growing trend of NLP use has improved the quantity and quality of the data available to build large cohorts of interest. The advances in NLP will increase the structured information available within the EHR and open doors for new exploration areas.

Currently, the ADE, medication episodes and ADR pipelines run on an ad hoc basis due to computational constraints. Automating the process will contribute to the continuity of usage and can lead to more routine extractions.

EHR linkage is another avenue to explore. For instance, linking secondary and primary health resources will help to fill the gaps in fragmented data collection, which would allow for a more complete picture of patient health. In addition, non-clinical factors such as diet, sleep patterns and physical activities can be captured using wearable devices and integrated with the data extracted from EHRs. Finally, linking with genotype data can further help understand the aetiology and phenotype-genotype associations.

6.4 What is Next

Regular monitoring of ADRs in psychiatric settings by EHR-based pharmacovigilance methods may help in early detection and reduce the risk associated with ADRs, which improves the quality of care and reduces the treatment time and cost. As a post-marketing surveillance tool, the ADR detection methods presented in this thesis can play an important rule to provide true incidence and prevalence of the ADR profile of drugs. Constant monitoring of ADRs, dosage and drug can improve the effectiveness of drugs compliance. As a long-term pharmacovigilance goal, active collaboration across multiple sites among clinicians that use such tools and share their findings can, in turn, provide early warning signals. Currently, the work is underway to look at the possibilities of combining the current pipeline with machine learning methods.

6.5 Conclusions

ADRs are a leading cause of mortality and morbidity; they lengthen hospital stay and cause a significant burden on healthcare providers worldwide. It is estimated that the prevalence of ADEs is 3.2% in England, 4.8% in Germany and 5.6% in the USA, and most of the English emergency admissions were ADR-related in the fiscal year of 2006 (Stausberg, 2014).

The secondary analysis of EHRs offers many opportunities, as one of the strengths of EHR data is the provision of detailed longitudinal patient clinical history in a structured and unstructured format. In analysing EHRs, studies must account for numerous challenges, some of which have been highlighted in this thesis. The work described in this thesis was mainly developed using SLAM's EHRs with the help of domain experts to extract ADEs, medication episodes and the casual relationship between them to infer possible ADRs. The methods were used to study Clozapine-induced ADRs in three large psychiatric health hospitals. Consequently, the work presented in this thesis can be used to identify ADRs for other medications of interest.

Further improvement in concept extraction from EHR data and data linkage would support a significant understanding of prescribed medications and ADRs prediction, which can improve the findings and design of future studies and much needed therapeutic interventions.

References

Afkat, A., Arshad, H., Samina, F., Shagufta, P., Vineeta, S., & Zubair, A. (2016). Prevalence and severity of adverse drug reactions (ADRs) in patients subjected to different anti-psychotic drugs in an out-patient Department of a Psychiatry Hospital in Kashmir; a prospective observational study. *Int J Pharmacol Clin Sci, 5*(1), 12-16.

Alex, B., Haddow, B., & Grover, C. (2007). *Recognising nested named entities in biomedical text.* Paper presented at the Biological, translational, and clinical language processing.

Allison, P. D. (2001). Sage university papers series on quantitative applications in the social sciences. *Missing data. Thousand Oaks, CA: Sage, 136.*

Alvarez, E., Bobes, J., Gomez, J.-C., Sacristan, J. A., Canas, F., Carrasco, J. L., . . . Group, E. S. (2003). Safety of olanzapine versus conventional antipsychotics in the treatment of patients with acute schizophrenia. A naturalistic study. *European neuropsychopharmacology, 13*(1), 39-48.

Alvir, J. M. J., Lieberman, J. A., Safferman, A. Z., Schwimmer, J. L., & Schaaf, J. A. (1993). Clozapine-induced agranulocytosis--incidence and risk factors in the United States. *New England Journal of Medicine, 329*(3), 162-167.

Amsler, H. A., Teerenhovi, L., Barth, E., Harjula, K., & Vuopio, P. (1977). Agranulocytosis in patients treated with clozapine. *Acta Psychiatrica Scandinavica, 56*(4), 241-248.

Angermeyer, M., Löffler, W., Müller, P., Schulze, B., & Priebe, S. (2001). Patients' and relatives' assessment of clozapine treatment. *Psychological Medicine, 31*(3), 509-517.

Angus, D. C. (2007). Caring for the critically ill patient: challenges and opportunities. *Jama, 298*(4), 456-458.

Anthony, M., & Berg, M. J. (2002). Biologic and molecular mechanisms for sex differences in pharmacokinetics, pharmacodynamics, and pharmacogenetics: Part I. *Journal of women's health & gender-based medicine, 11*(7), 601-615.

Aronson, A. R. (2001). *Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program.* Paper presented at the Proceedings of the AMIA Symposium.

Baker, E., Iqbal, E., Johnston, C., Broadbent, M., Shetty, H., Stewart, R., . . . Dobson, R. J. (2017). Trajectories of dementia-related cognitive decline in a large mental health records derived patient cohort. *PLoS One, 12*(6), e0178562.

Ball, M., Patel, R., Hayes, R. D., Dobson, R. J., & Stewart, R. (2014). *TextHunter–A User Friendly Tool for Extracting Generic Concepts from Free Text in Clinical Research.* Paper presented at the AMIA Annual Symposium Proceedings.

Ball, R., Robb, M., Anderson, S., & Dal Pan, G. (2016). The FDA's sentinel initiative—a comprehensive approach to medical product surveillance. *Clinical Pharmacology & Therapeutics, 99*(3), 265-268.

Banda, J. M., Callahan, A., Winnenburg, R., Strasberg, H. R., Cami, A., Reis, B. Y., . . . Shah, N. H. (2016). Feasibility of prioritizing drug–drug-event associations found in electronic health records. *Drug safety, 39*(1), 45-57.

Bate, A., Lindquist, M., Edwards, I., Olsson, S., Orre, R., Lansner, A., & De Freitas, R. M. (1998). A Bayesian neural network method for adverse drug reaction signal generation. *European journal of clinical pharmacology, 54*(4), 315-321.

Bean, D. M., Wu, H., Iqbal, E., Dzahini, O., Broadbent, M., Stewart, R., & Dobson, R. J. (2017). Knowledge graph prediction of unknown adverse drug reactions and validation in electronic health records. *Scientific reports, 7*(1), 16416.

Beaulieu-Jones, B. K., Lavage, D. R., Snyder, J. W., Moore, J. H., Pendergrass, S. A., & Bauer, C. R. (2018). Characterizing and managing missing structured data in electronic health records: data analysis. *JMIR medical informatics, 6*(1), e11.

Bejan, C. A., Xia, F., Vanderwende, L., Wurfel, M. M., & Yetisgen-Yildiz, M. (2012). Pneumonia identification using statistical feature selection. *Journal of the American Medical Informatics Association, 19*(5), 817-823.

Berna, F., Timbolschi24, I. D., Diemunsch24, P., & Vidailhet123, P. (2013). Acute dystonia and akathisia following droperidol administration misdiagnosed as psychiatric disorders.

Bird, S., & Loper, E. (2004). *NLTK: the natural language toolkit.* Paper presented at the Proceedings of the ACL 2004 on Interactive poster and demonstration sessions.

Bixler, E. O., Kales, A., Brubaker, B. H., & Kales, J. D. (1987). Adverse reactions to benzodiazepine hypnotics: spontaneous reporting system. *Pharmacology, 35*(5), 286-300.

BNF. (2018). Joint Formulary Committee. Retrieved from https://www.medicinescomplete.com/mc/bnf/current/

Bodenreider, O. (2004). The unified medical language system (UMLS): integrating biomedical terminology. *Nucleic acids research, 32*(suppl_1), D267-D270.

Bora, E., Yucel, M., & Pantelis, C. (2009). Cognitive functioning in schizophrenia, schizoaffective disorder and affective psychoses: meta-analytic study. *The British Journal of Psychiatry, 195*(6), 475-482.

Botsis, T., Hartvigsen, G., Chen, F., & Weng, C. (2010). Secondary use of EHR: data quality issues and informatics opportunities. *Summit on Translational Bioinformatics, 2010*, 1.

Bowman, S. (2013). Impact of electronic health record systems on information integrity: quality and safety implications. *Perspectives in health information management, 10*(Fall).

Brennan, L., Watson, M., Klaber, R., & Charles, T. (2012). The importance of knowing context of hospital episode statistics when reconfiguring the NHS. *Bmj, 344*, e2432.

Caldicott, F. (2013). Information: To share or not to share. Information Governance Review. *Information: To share or not to share*.

Camden and Islington NHS Foundation Trust. (2017). Camden and Islington NHS Foundation Trust. Retrieved from http://www.candi.nhs.uk/

Cantor, M. N., Feldman, H. J., & Triola, M. M. (2007). Using trigger phrases to detect adverse drug reactions in ambulatory care notes. *BMJ Quality & Safety, 16*(2), 132-134.

Caroff, S. N., Miller, D. D., Dhopesh, V., & Campbell, E. C. (2011). Is there a rational management strategy for tardive dyskinesia? Decisions should be based on the course of TD and effective control of psychotic symptoms. *Current Psychiatry, 10*(10), 22.

Caroff, S. N., Ungvari, G. S., & Owens, D. G. C. (2018). Historical perspectives on tardive dyskinesia. *Journal of the neurological sciences*.

Carroll, R. J., Eyler, A. E., & Denny, J. C. (2011). *Naïve electronic health record phenotype identification for rheumatoid arthritis.* Paper presented at the AMIA annual symposium proceedings.

Cartwright, N. (2010). What are randomised controlled trials good for? *Philosophical studies, 147*(1), 59.

Casey, D. E. (1998). Effects of clozapine therapy in schizophrenic individuals at risk for tardive dyskinesia. *The Journal of clinical psychiatry*.

Casey, D. E. (2006). Implications of the CATIE trial on treatment: extrapyramidal symptoms. *CNS spectrums, 11*(S7), 25-31.

Celi, L. A., Moseley, E., Moses, C., Ryan, P., Somai, M., Stone, D., & Tang, K.-o. (2014). From pharmacovigilance to clinical care optimization. *Big data, 2*(3), 134-141.

Chapman, A. B., Peterson, K. S., Alba, P. R., DuVall, S. L., & Patterson, O. V. (2019). Detecting adverse drug events with rapidly trained classification models. *Drug safety*, 1-10.

Chapman, W. W., Bridewell, W., Hanbury, P., Cooper, G. F., & Buchanan, B. G. (2001). A simple algorithm for identifying negated findings and diseases in discharge summaries. *Journal of biomedical informatics, 34*(5), 301-310.

Chapman, W. W., Chu, D., & Dowling, J. N. (2007). *ConText: An algorithm for identifying contextual features from clinical text.* Paper presented at the Proceedings of the Workshop on BioNLP 2007: Biological, Translational, and Clinical Language Processing.

Chapman, W. W., Hilert, D., Velupillai, S., Kvist, M., Skeppstedt, M., Chapman, B. E., . . . Deleger, L. (2013). Extending the NegEx lexicon for multiple languages. *Studies in health technology and informatics, 192*, 677.

Chaudhry, I. B., Neelam, K., Duddu, V., & Husain, N. (2008). Ethnicity and psychopharmacology. *Journal of Psychopharmacology, 22*(6), 673-680.

Chen, E. S., Hripcsak, G., Xu, H., Markatou, M., & Friedman, C. (2008). Automated acquisition of disease–drug knowledge from biomedical and clinical documents: an initial study. *Journal of the American Medical Informatics Association, 15*(1), 87-98.

Chen, L., & Friedman, C. (2004). *Extracting phenotypic information from the literature via natural language processing.* Paper presented at the Medinfo.

Chen, Y., Carroll, R. J., Hinz, E. R. M., Shah, A., Eyler, A. E., Denny, J. C., & Xu, H. (2013). Applying active learning to high-throughput phenotyping algorithms for electronic health records data. *Journal of the American Medical Informatics Association, 20*(e2), e253-e259.

Chiappini, S., Schifano, F., Corkery, J. M., & Guirguis, A. (2020). Focus on clozapine withdrawal-and misuse-related cases as reported to the European Medicines Agency (EMA) pharmacovigilance database. *Brain Sciences, 10*(2), 105.

Choi, S., DiSilvio, B., Unangst, J., & Fernstrom, J. D. (2007). Effect of chronic infusion of olanzapine and clozapine on food intake and body weight gain in male and female rats. *Life sciences, 81*(12), 1024-1030.

Chue, P., & Chue, J. (2016). A critical appraisal of paliperidone long-acting injection in the treatment of schizoaffective disorder. *Therapeutics and clinical risk management*, *12*, 109.

Citrome, L. L. (2004). The increase in risk of diabetes mellitus from exposure to second-generation antipsychotic agents. *Drugs of Today*, *40*(5), 445-464.

Cocos, A., Fiks, A. G., & Masino, A. J. (2017). Deep learning for pharmacovigilance: recurrent neural network architectures for labeling adverse drug reactions in Twitter posts. *Journal of the American Medical Informatics Association*, *24*(4), 813-821.

Cohen, D. (2004). Atypical antipsychotics and new onset diabetes mellitus. *Pharmacopsychiatry*, *38*(01), 1-11.

Compass. (2008). Adverse drug reactions waste NHS £2bn. . Retrieved from http://www.compassonline.org.uk/news/item.asp?n=1551

Cortada, J., Gordon, D., & Lenihan, B. (2012). The value of analytics in healthcare: From insights to outcomes. IBM Global Business Services Executive Report. *Annual Institute for Business Value*.

Courtet, P., Samalin, L., & Olié, E. (2011). Antidepressants in bipolar disorder. *L'Encephale*, *37*, S196-202.

CRIS Network. (2018). UK-CRIS Programme. Retrieved from https://crisnetwork.co/uk-cris-programme

Cunningham, H., Maynard, D., & Bontcheva, K. (2011). *Text processing with gate*: Gateway Press CA.

Cunningham, H., Tablan, V., Roberts, A., & Bontcheva, K. (2013). Getting more out of biomedical documents with GATE's full lifecycle open source text analytics. *PLoS computational biology*, *9*(2), e1002854.

Dandala, B., Joopudi, V., & Devarakonda, M. (2019). Adverse drug events detection in clinical notes by jointly modeling entities and relations using neural networks. *Drug safety*, *42*(1), 135-146.

Data, M. C. (2016). *Secondary Analysis of Electronic Health Records*: Springer International Publishing.

Datapharm Communications Limited. (2017). The electronic Medicines Compendium (eMC). Retrieved from https://www.medicines.org.uk/emc/

Datapharm Communications Ltd. (2017). electronic Medicines Compendium (eMC). Retrieved from https://www.medicines.org.uk/emc/

Dauner, A., & Blair, D. T. (1990). Akathisia: When treatment creates a problem. *Journal of psychosocial nursing and mental health services, 28*(10), 13-18.

De Fazio, P., Gaetano, R., Caroleo, M., Cerminara, G., Maida, F., Bruno, A., . . . Segura-García, C. (2015). Rare and very rare adverse effects of clozapine. *Neuropsychiatric disease and treatment, 11*, 1995.

Demski, H., Garde, S., & Hildebrand, C. (2016). Open data models for smart health interconnected applications: the example of openEHR. *BMC medical informatics and decision making, 16*(1), 137.

Denny, J. C., Irani, P. R., Wehbe, F. H., Smithers, J. D., & Spickard III, A. (2003). *The KnowledgeMap project: development of a concept-based medical school curriculum database.* Paper presented at the AMIA Annual Symposium Proceedings.

Dickerson, F., Stallings, C. R., Origoni, A. E., Vaughan, C., Khushalani, S., Schroeder, J., & Yolken, R. H. (2013). Cigarette smoking among persons with schizophrenia or bipolar disorder in routine clinical settings, 1999–2011. *Psychiatric Services, 64*(1), 44-50.

Digital Health. (2019). Two RiO trusts to switch EPR. Retrieved from https://www.digitalhealth.net/2015/04/two-rio-trusts-to-switch-epr/

Dilip, V., Wyatt, R. J., & Need, C. Y. W. E. (1982). Understanding and treating tardive dyskinesia.

Dixon, L., Postrado, L., Delahanty, J., Fischer, P. J., & Lehman, A. (1999). The association of medical comorbidity in schizophrenia with poor physical and mental health. *The Journal of nervous and mental disease, 187*(8), 496-502.

Donnelly, K. (2006). SNOMED-CT: The advanced terminology and coding system for eHealth. *Studies in health technology and informatics, 121*, 279.

DuMouchel, W. (1999). Bayesian data mining in large frequency tables, with an application to the FDA spontaneous reporting system. *The American Statistician, 53*(3), 177-190.

Economics, F. (2014). Exploring the costs of unsafe care in the NHS. *London: Department of Health*.

Edwards, I. R., & Aronson, J. K. (2000). Adverse drug reactions: definitions, diagnosis, and management. *The Lancet, 356*(9237), 1255-1259.

Eriksson, R., Jensen, P. B., Frankild, S., Jensen, L. J., & Brunak, S. (2013). Dictionary construction and identification of possible adverse drug events in Danish clinical

narrative text. *Journal of the American Medical Informatics Association, 20*(5), 947-953.

Eriksson, R., Werge, T., Jensen, L. J., & Brunak, S. (2014). Dose-specific adverse drug reaction identification in electronic patient records: temporal data mining in an inpatient psychiatric population. *Drug safety, 37*(4), 237-247.

Evans, R. (2016). Electronic health records: then, now, and in the future. *Yearbook of medical informatics, 25*(S 01), S48-S61.

Evans, S., Waller, P. C., & Davis, S. (2001). Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports. *Pharmacoepidemiology and drug safety, 10*(6), 483-486.

Farri, O., Monsen, K. A., Pakhomov, S. V., Pieczkiewicz, D. S., Speedie, S. M., & Melton, G. B. (2013). Effects of time constraints on clinician–computer interaction: a study on information synthesis from EHR clinical notes. *Journal of biomedical informatics, 46*(6), 1136-1144.

FDA. (2018). FDA's Adverse Event Reporting System (FAERS). Retrieved from https://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Surveillance/AdverseDrugEffects/default.htm

Feder, S. L. (2018). Data quality in electronic health records research: Quality domains and assessment methods. *Western journal of nursing research, 40*(5), 753-766.

Fernandes, A. C., Cloete, D., Broadbent, M. T., Hayes, R. D., Chang, C.-K., Jackson, R. G., . . . Liebscher, J. (2013). Development and evaluation of a de-identification procedure for a case register sourced from mental health electronic records. *BMC medical informatics and decision making, 13*(1), 71.

Fernandes, A. C., Dutta, R., Velupillai, S., Sanyal, J., Stewart, R., & Chandran, D. (2018). Identifying suicide ideation and suicidal attempts in a psychiatric clinical research database using natural language processing. *Scientific reports, 8*(1), 7426.

Ferrajolo, C., Coloma, P. M., Verhamme, K. M., Schuemie, M. J., de Bie, S., Gini, R., . . . Giaquinto, C. (2014). Signal detection of potentially drug-induced acute liver injury in children using a multi-country healthcare database network. *Drug safety, 37*(2), 99-108.

Ferrucci, D., Lally, A., Verspoor, K., & Nyberg, E. (2009). Unstructured information management architecture (UIMA) version 1.0. *OASIS Standard, OASIS*.

Field, T. S., Gurwitz, J. H., Harrold, L. R., Rothschild, J. M., Debellis, K., Seger, A. C., . . . Bates, D. W. (2004). Strategies for detecting adverse drug events among older

persons in the ambulatory setting. *Journal of the American Medical Informatics Association, 11*(6), 492-498.

Flanagan, R. J. (2008). Side effects of clozapine and some other psychoactive drugs. *Current drug safety, 3*(2), 115-122.

Fleischhaker, C., Heiser, P., Hennighausen, K., Herpertz-Dahlmann, B., Holtkamp, K., Mehler-Wex, C., . . . Warnke, A. (2006). Clinical drug monitoring in child and adolescent psychiatry: side effects of atypical neuroleptics. *Journal of Child & Adolescent Psychopharmacology, 16*(3), 308-316.

Fleischhaker, C., Heiser, P., Hennighausen, K., Herpertz-Dahlmann, B., Holtkamp, K., Mehler-Wex, C., . . . Warnke, A. (2008). Weight gain in children and adolescents during 45 weeks treatment with clozapine, olanzapine and risperidone. *Journal of Neural Transmission, 115*(11), 1599-1608.

Frazier, J. A., Gordon, C. T., McKENNA, K., Lenane, M. C., Jih, D., & Rapoport, J. L. (1994). An open trial of clozapine in 11 adolescents with childhood-onset schizophrenia. *Journal of the American Academy of Child & Adolescent Psychiatry, 33*(5), 658-663.

Friedman, C. (2009). *Discovering novel adverse drug events using natural language processing and mining of the electronic health record.* Paper presented at the Conference on Artificial Intelligence in Medicine in Europe.

Friedman, C., & Hripcsak, G. (1999). Natural language processing and its future in medicine. *Acad Med, 74*(8), 890-895.

Friedman, C., Shagina, L., Lussier, Y., & Hripcsak, G. (2004). Automated encoding of clinical documents based on natural language processing. *Journal of the American Medical Informatics Association, 11*(5), 392-402.

Frogley, C., Taylor, D., Dickens, G., & Picchioni, M. (2012). A systematic review of the evidence of clozapine's anti-aggressive effects. *International Journal of Neuropsychopharmacology, 15*(9), 1351-1371.

Fu, D.-J., Turkoz, I., Simonson, R. B., Walling, D., Schooler, N., Lindenmayer, J.-P., . . . Alphs, L. (2016). Paliperidone palmitate once-monthly injectable treatment for acute exacerbations of schizoaffective disorder. *Journal of clinical psychopharmacology, 36*(4), 372.

Furukawa, T., & Kurokawa, J. (2007). Regulation of cardiac ion channels via non-genomic action of sex steroid hormones: implication for the gender difference in cardiac arrhythmias. *Pharmacology & therapeutics, 115*(1), 106-115.

Gareri, P., De Fazio, P., Russo, E., Marigliano, N., De Fazio, S., & De Sarro, G. (2008). The safety of clozapine in the elderly. *Expert opinion on drug safety, 7*(5), 525-538.

Gerbino-Rosen, G., Roofeh, D., Tompkins, D. A., Feryo, D., Nusser, L., Kranzler, H., . . . Rhinewine, J. (2005). Hematological adverse events in clozapine-treated children and adolescents. *Journal of the American Academy of Child & Adolescent Psychiatry, 44*(10), 1024-1031.

Gettinger, A., & Csatari, A. (2012). Transitioning from a Legacy EHR to a Commercial, Vendor-supplied, EHR. *Applied clinical informatics, 3*(04), 367-376.

Ghaemi, S. N., Hsu, D. J., Soldani, F., & Goodwin, F. K. (2003). Antidepressants in bipolar disorder: the case for caution. *Bipolar disorders, 5*(6), 421-433.

Gianfrancesco, F. D., Grogg, A. L., Mahmoud, R. A., Wang, R.-h., & Nasrallah, H. A. (2002). Differential effects of risperidone, olanzapine, clozapine, and conventional antipsychotics on type 2 diabetes: findings from a large health plan database. *The Journal of clinical psychiatry, 63*(10), 920-930.

Gill, S. S., Bronskill, S. E., Normand, S.-L. T., Anderson, G. M., Sykora, K., Lam, K., . . . Herrmann, N. (2007). Antipsychotic drug use and mortality in older adults with dementia. *Annals of internal medicine, 146*(11), 775-786.

Gillespie, A. L., Samanaite, R., Mill, J., Egerton, A., & MacCabe, J. H. (2017). Is treatment-resistant schizophrenia categorically distinct from treatment-responsive schizophrenia? A systematic review. *BMC psychiatry, 17*(1), 12.

Goryachev, S., Sordo, M., & Zeng, Q. T. (2006). *A suite of natural language processing tools developed for the I2B2 project.* Paper presented at the AMIA Annual Symposium Proceedings.

Green, M. F. (2006). Cognitive impairment and functional outcome in schizophrenia and bipolar disorder. *The Journal of clinical psychiatry, 67*(10), e12-e12.

Grimes, D. A., & Schulz, K. F. (2002). Bias and causal associations in observational research. *The lancet, 359*(9302), 248-252.

Grover, C., Haddow, B., Klein, E., Matthews, M., Nielsen, L. A., Tobin, R., & Wang, X. (2007). *Adapting a relation extraction pipeline for the BioCreAtIvE II task.* Paper presented at the Proceedings of the BioCreAtIvE II Workshop.

Grundstrom, C., Väyrynen, K., Iivari, N., & Isomursu, M. (2019). *Making Sense of the General Data Protection Regulation—Four Categories of Personal Data Access Challenges.* Paper presented at the Proceedings of the 52nd Hawaii International Conference on System Sciences.

Gürcan, G., Şenol, Ş. H., Yağcıoğlu, A. A., & Ertuğrul, A. (2017). Side effects of clozapine and their relationship with clinical variables in patients with schizophrenia. *European Psychiatry*, *41*, S197.

Gurwitz, J. H., Field, T. S., Harrold, L. R., Rothschild, J., Debellis, K., Seger, A. C., . . . Kelleher, M. (2003). Incidence and preventability of adverse drug events among older persons in the ambulatory setting. *Jama*, *289*(9), 1107-1116.

Gysbers, M., Reichley, R., Kilbridge, P. M., Noirot, L., Nagarajan, R., Dunagan, W. C., & Bailey, T. C. (2008). *Natural language processing to identify adverse drug events.* Paper presented at the AMIA... Annual Symposium proceedings. AMIA Symposium.

Haas, S. J., Hill, R., Krum, H., Liew, D., Tonkin, A., Demos, L., . . . McNeil, J. (2007). Clozapine-associated myocarditis. *Drug safety*, *30*(1), 47-57.

Haddad, P. M., Das, A., Keyhani, S., & Chaudhry, I. B. (2012). Antipsychotic drugs and extrapyramidal side effects in first episode psychosis: a systematic review of head–head comparisons. *Journal of Psychopharmacology*, *26*(5_suppl), 15-26.

Haerian, K., Varn, D., Chase, H., Vaidya, S., & Friedman, C. (2010). *Electronic health record pharmacovigilance signal extraction: a semi-automated method for reduction of confounding applied to detection of rhabdomyolysis.* Paper presented at the Drug safety.

Haerian, K., Varn, D., Vaidya, S., Ena, L., Chase, H., & Friedman, C. (2012). Detection of pharmacovigilance-related adverse events using electronic health records and automated methods. *Clinical Pharmacology & Therapeutics*, *92*(2), 228-234.

Haneuse, S. (2016). Distinguishing selection bias and confounding bias in comparative effectiveness research. *Medical care*, *54*(4), e23.

Harpaz, R., Haerian, K., Chase, H. S., & Friedman, C. (2010). *Mining electronic health records for adverse drug effects using regression based methods.* Paper presented at the Proceedings of the 1st ACM International Health Informatics Symposium.

Harpaz, R., Vilar, S., DuMouchel, W., Salmasian, H., Haerian, K., Shah, N. H., . . . Friedman, C. (2012). Combing signals from spontaneous reports and electronic health records for detection of adverse drug reactions. *Journal of the American Medical Informatics Association*, *20*(3), 413-419.

Hayes, R. D., Downs, J., Chang, C.-K., Jackson, R. G., Shetty, H., Broadbent, M., . . . Stewart, R. (2014). The effect of clozapine on premature mortality: an assessment of clinical monitoring and other potential confounders. *Schizophrenia bulletin*, *41*(3), 644-655.

Hazell, L., & Shakir, S. A. (2006). Under-reporting of adverse drug reactions. *Drug safety*, *29*(5), 385-396.

Hazlehurst, B., Frost, H. R., Sittig, D. F., & Stevens, V. J. (2005). MediClass: A system for detecting and classifying encounter-based clinical events in any electronic medical record. *Journal of the American Medical Informatics Association*, *12*(5), 517-529.

Hazlehurst, B., Naleway, A., & Mullooly, J. (2009). Detecting possible vaccine adverse events in clinical notes of the electronic medical record. *Vaccine*, *27*(14), 2077-2083.

HDR UK. (2019). Digital Innovation Hub Prospectus. Retrieved from https://www.hdruk.ac.uk/digital-innovation-hubs/

Henriksson, A., Kvist, M., Dalianis, H., & Duneld, M. (2015). Identifying adverse drug event information in clinical notes with distributional semantic representations of context. *Journal of biomedical informatics*, *57*, 333-349.

Henriksson, A., Zhao, J., Boström, H., & Dalianis, H. (2015). *Modeling electronic health records in ensembles of semantic spaces for adverse drug event detection.* Paper presented at the Bioinformatics and Biomedicine (BIBM), 2015 IEEE International Conference on.

HES - NHS Digital. (2018). Hospital Episode Statistics (HES). Retrieved from https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics

Hodge, K., & Jespersen, S. (2008). Side-effects and treatment with clozapine: a comparison between the views of consumers and their clinicians. *International Journal of Mental Health Nursing*, *17*(1), 2-8.

Hofer-Dueckelmann, C., Prinz, E., Beindl, W., Szymanski, J., Fellhofer, G., Pichler, M., & Schuler, J. (2011). Adverse drug reactions (ADRs) associated with hospital admissions—elderly female patients are at highest risk. *International Journal of Clinical Pharmacologyand Therapeutics*, *49*(10), 577.

Honigman, B., Lee, J., Rothschild, J., Light, P., Pulling, R. M., Yu, T., & Bates, D. W. (2001). Using computerized data to identify adverse drug events in outpatients. *Journal of the American Medical Informatics Association*, *8*(3), 254-266.

Honigman, B., Light, P., Pulling, R. M., & Bates, D. W. (2001). A computerized method for identifying incidents associated with adverse drug events in outpatients. *International journal of medical informatics*, *61*(1), 21-32.

Honnibal, M. (2017). Spacy. In.

Hripcsak, G., & Albers, D. J. (2012). Next-generation phenotyping of electronic health records. *Journal of the American Medical Informatics Association, 20*(1), 117-121.

Hripcsak, G., & Albers, D. J. (2013). Correlating electronic health record concepts with healthcare process events. *Journal of the American Medical Informatics Association, 20*(e2), e311-e318.

Hripcsak, G., Albers, D. J., & Perotte, A. (2015). Parameterizing time in electronic health record studies. *Journal of the American Medical Informatics Association, 22*(4), 794-804.

Hripcsak, G., Knirsch, C., Zhou, L., Wilcox, A., & Melton, G. B. (2011). Bias associated with mining electronic health records. *Journal of biomedical discovery and collaboration, 6*, 48.

Huang, H., Heck, L., & Ji, H. (2015). Leveraging deep neural networks and knowledge graphs for entity disambiguation. *arXiv preprint arXiv:1504.07678*.

Humphreys, K., Maisel, N. C., Blodgett, J. C., Fuh, I. L., & Finney, J. W. (2013). Extent and reporting of patient nonenrollment in influential randomized clinical trials, 2002 to 2010. *JAMA internal medicine, 173*(11), 1029-1031.

Hynes, C., Keating, D., McWilliams, S., Madigan, K., Kinsella, A., Maidment, I., . . . Gaughran, F. (2015). Glasgow antipsychotic side-effects scale for clozapine—development and validation of a clozapine-specific side-effects scale. *Schizophrenia research, 168*(1), 505-513.

ICD. (2015). International Classification of Diseases (ICD). Retrieved from http://www.who.int/classifications/icd/en/

Idänpään-Heikkilä, J., Alhava, E., Olkinuora, M., & Palva, I. (1977). Agranulocytosis during treatment with clozapine. *European journal of clinical pharmacology, 11*(3), 193-198.

Ilyas, S., & Moncrieff, J. (2012). Trends in prescriptions and costs of drugs for mental disorders in England, 1998–2010. *The British Journal of Psychiatry, 200*(5), 393-398.

Ioannidis, J. P., Haidich, A.-B., Pappa, M., Pantazis, N., Kokori, S. I., Tektonidou, M. G., . . . Lau, J. (2001). Comparison of evidence of treatment effects in randomized and nonrandomized studies. *Jama, 286*(7), 821-830.

Iqbal, E., Mallah, R., Jackson, R. G., Ball, M., Ibrahim, Z. M., Broadbent, M., . . . Dobson, R. J. (2015). Identification of adverse drug events from free text electronic patient

records and information in a large mental health case register. *PloS one, 10*(8), e0134208.

Iqbal, E., Mallah, R., Rhodes, D., Wu, H., Romero, A., Chang, N., . . . Stewart, R. (2017). ADEPt, a semantically-enriched pipeline for extracting adverse drug events from free-text electronic health records. *PloS one, 12*(11), e0187121.

Iqbal, M. M., Swati, S., Aneja, A., Iqbal, M. T., Aneja, E., & Haque, S. Most Common Side Effects of Clozapine and Their Management.

Iyer, S. V., Harpaz, R., LePendu, P., Bauer-Mehren, A., & Shah, N. H. (2013). Mining clinical text for signals of adverse drug-drug interactions. *Journal of the American Medical Informatics Association, 21*(2), 353-362.

Jackson, R., Kartoglu, I., Stringer, C., Gorrell, G., Roberts, A., Song, X., . . . Groza, T. (2018). CogStack-experiences of deploying integrated information retrieval and extraction services in a large National Health Service Foundation Trust hospital. *BMC medical informatics and decision making, 18*(1), 47.

Jackson, R. G., Patel, R., Jayatilleke, N., Kolliakou, A., Ball, M., Gorrell, G., . . . Stewart, R. (2017). 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, 7*(1), e012012.

Jacobs, L. (2009). Interview with Lawrence Weed, MD—The Father of the Problem-Oriented Medical Record Looks Ahead. *The Permanente Journal, 13*(3), 84.

Jagannatha, A., Liu, F., Liu, W., & Yu, H. (2019). Overview of the first natural language processing challenge for extracting medication, indication, and adverse drug events from electronic health record notes (MADE 1.0). *Drug safety, 42*(1), 99-111.

Jensen, P. B., Jensen, L. J., & Brunak, S. (2012). Mining electronic health records: towards better research applications and clinical care. *Nat Rev Genet, 13*(6), 395-405. doi:10.1038/nrg3208

Johnson, A. E., Pollard, T. J., Shen, L., Li-wei, H. L., Feng, M., Ghassemi, M., . . . Mark, R. G. (2016). MIMIC-III, a freely accessible critical care database. *Scientific data, 3*, 160035.

Johnson, S. G., Speedie, S., Simon, G., Kumar, V., & Westra, B. L. (2016). Application of an ontology for characterizing data quality for a secondary use of EHR data. *Applied clinical informatics, 7*(01), 69-88.

Joint Formulary Committee. (2016). *British national formulary (BNF) 72* (Vol. 66): Pharmaceutical Press.

Jones, J. B., Weiner, J. P., Shah, N. R., & Stewart, W. F. (2015). The wired patient: patterns of electronic patient portal use among patients with cardiac disease or diabetes. *Journal of medical Internet research, 17*(2), e42.

Jones, P. B., Barnes, T. R., Davies, L., Dunn, G., Lloyd, H., Hayhurst, K. P., . . . Lewis, S. W. (2006). Randomized controlled trial of the effect on Quality of Life of second-vs first-generation antipsychotic drugs in schizophrenia: Cost Utility of the Latest Antipsychotic Drugs in Schizophrenia Study (CUtLASS 1). *Archives of general psychiatry, 63*(10), 1079-1087.

Kadra, G., Spiros, A., Shetty, H., Iqbal, E., Hayes, R. D., Stewart, R., & Geerts, H. (2018). Predicting parkinsonism side-effects of antipsychotic polypharmacy prescribed in secondary mental healthcare. *Journal of Psychopharmacology*, 0269881118796809.

Kadra, G., Stewart, R., Shetty, H., Downs, J., MacCabe, J. H., Taylor, D., & Hayes, R. D. (2016). Predictors of long-term (≥ 6months) antipsychotic polypharmacy prescribing in secondary mental healthcare. *Schizophrenia research, 174*(1), 106-112.

Kahn, M. G., Raebel, M. A., Glanz, J. M., Riedlinger, K., & Steiner, J. F. (2012). A pragmatic framework for single-site and multisite data quality assessment in electronic health record-based clinical research. *Medical care, 50*.

Kar, N., Barreto, S., & Chandavarkar, R. (2016). Clozapine monitoring in clinical practice: beyond the mandatory requirement. *Clinical Psychopharmacology and Neuroscience, 14*(4), 323.

Katzenschlager, R., Sampaio, C., Costa, J., & Lees, A. (2002). Anticholinergics for symptomatic management of Parkinsons disease (Cochrane Review). *The Cochrane database of systematic reviews, 3*.

Kelly, C., & McCreadie, R. (2000). Cigarette smoking and schizophrenia. *Advances in Psychiatric Treatment, 6*(5), 327-331.

Kelly, D. L., Kreyenbuhl, J., Dixon, L., Love, R. C., Medoff, D., & Conley, R. R. (2006). Clozapine underutilization and discontinuation in African Americans due to leucopenia. *Schizophrenia bulletin, 33*(5), 1221-1224.

Kerwin, R., Millet, B., Herman, E., Banki, C. M., Lublin, H., Pans, M., . . . Beuzen, J.-N. (2007). A multicentre, randomized, naturalistic, open-label study between aripiprazole and standard of care in the management of community-treated schizophrenic patients Schizophrenia Trial of Aripiprazole:(STAR) study. *European Psychiatry, 22*(7), 433-443.

Kessler, R. C., Barker, P. R., Colpe, L. J., Epstein, J. F., Gfroerer, J. C., Hiripi, E., . . . Walters, E. E. (2003). Screening for serious mental illness in the general population. *Archives of general psychiatry, 60*(2), 184-189.

KLAWANS JR, H. L. (1973). The pharmacology of tardive dyskinesias. *American Journal of Psychiatry, 130*(1), 82-86.

Koppel, R., Metlay, J. P., Cohen, A., Abaluck, B., Localio, A. R., Kimmel, S. E., & Strom, B. L. (2005). Role of computerized physician order entry systems in facilitating medication errors. *Jama, 293*(10), 1197-1203.

Kreimeyer, K., Foster, M., Pandey, A., Arya, N., Halford, G., Jones, S. F., . . . Botsis, T. (2017). Natural language processing systems for capturing and standardizing unstructured clinical information: a systematic review. *Journal of biomedical informatics, 73*, 14-29.

Krupp, P., & Barnes, P. (1992). Clozapine-associated agranulocytosis: risk and aetiology. *The British Journal of Psychiatry*.

Kuhn, M., Letunic, I., Jensen, L. J., & Bork, P. (2015). The SIDER database of drugs and side effects. *Nucleic acids research, 44*(D1), D1075-D1079.

Kumra, S., Frazier, J. A., Jacobsen, L. K., McKenna, K., Gordon, C. T., Lenane, M. C., . . . Alaghband-Rad, J. (1996). Childhood-onset schizophrenia: a double-blind clozapine-haloperidol comparison. *Archives of general psychiatry, 53*(12), 1090-1097.

Kush, R. D., Helton, E., Rockhold, F. W., & Hardison, C. D. (2008). Electronic health records, medical research, and the Tower of Babel. *N Engl J Med*.

Lally, J., & MacCabe, J. H. (2015). Antipsychotic medication in schizophrenia: a review. *British medical bulletin, 114*(1), 169-179.

Leadbetter, R., & Shutty, M. (1992). Clozapine-induced weight gain: prevalence and clinical relevance. *The American journal of psychiatry, 149*(1), 68.

Lee, C. Y., & Chen, Y.-P. P. (2019). Machine learning on adverse drug reactions for pharmacovigilance. *Drug discovery today*.

Legge, S. E., Hamshere, M., Hayes, R. D., Downs, J., O'Donovan, M. C., Owen, M. J., . . . MacCabe, J. H. (2016). Reasons for discontinuing clozapine: a cohort study of patients commencing treatment. *Schizophrenia research, 174*(1-3), 113-119.

LePendu, P., Iyer, S. V., Bauer-Mehren, A., Harpaz, R., Mortensen, J. M., Podchiyska, T., . . . Shah, N. H. (2013). Pharmacovigilance using clinical notes. *Clinical pharmacology & therapeutics, 93*(6), 547-555.

LePendu, P., Iyer, S. V., Fairon, C., & Shah, N. H. (2012). *Annotation analysis for testing drug safety signals using unstructured clinical notes.* Paper presented at the Journal of biomedical semantics.

Leucht, S., Corves, C., Arbter, D., Engel, R. R., Li, C., & Davis, J. M. (2009). Second-generation versus first-generation antipsychotic drugs for schizophrenia: a meta-analysis. *The Lancet, 373*(9657), 31-41.

Leutwyler, H., Hubbard, E. M., Jeste, D. V., Miller, B., & Vinogradov, S. (2014). Associations of schizophrenia symptoms and neurocognition with physical activity in older adults with schizophrenia. *Biological research for nursing, 16*(1), 23-30.

Levine, M. E., Albers, D. J., & Hripcsak, G. (2016). *Comparing lagged linear correlation, lagged regression, Granger causality, and vector autoregression for uncovering associations in EHR data.* Paper presented at the AMIA Annual Symposium Proceedings.

Li, Q., Melton, K., Lingren, T., Kirkendall, E. S., Hall, E., Zhai, H., . . . Solti, I. (2014). Phenotyping for patient safety: algorithm development for electronic health record based automated adverse event and medical error detection in neonatal intensive care. *Journal of the American Medical Informatics Association, 21*(5), 776-784.

Li, Y., Salmasian, H., Vilar, S., Chase, H., Friedman, C., & Wei, Y. (2014). A method for controlling complex confounding effects in the detection of adverse drug reactions using electronic health records. *Journal of the American Medical Informatics Association, 21*(2), 308-314.

Liaw, S.-T., Rahimi, A., Ray, P., Taggart, J., Dennis, S., de Lusignan, S., . . . Talaei-Khoei, A. (2013). Towards an ontology for data quality in integrated chronic disease management: a realist review of the literature. *International journal of medical informatics, 82*(1), 10-24.

Lieberman, J. A., & Safferman, A. Z. (1992). Clinical profile of clozapine: adverse reactions and agranulocytosis. *Psychiatric Quarterly, 63*(1), 51-70.

Lima, A., Soares-Weiser, K., Bacaltchuk, J., & Barnes, T. (2002). Benzodiazepines for neuroleptic-induced acute akathisia. *The Cochrane database of systematic reviews*(1), CD001950-CD001950.

Lin, K.-M., Poland, R. E., & Lesser, I. M. (1986). Ethnicity and psychopharmacology. *Culture, medicine and psychiatry, 10*(2), 151-165.

Liu, M., Hinz, E. R. M., Matheny, M. E., Denny, J. C., Schildcrout, J. S., Miller, R. A., & Xu, H. (2013). Comparative analysis of pharmacovigilance methods in the detection of adverse drug reactions using electronic medical records. *Journal of the American Medical Informatics Association, 20*(3), 420-426.

Liu, S., Ma, W., Moore, R., Ganesan, V., & Nelson, S. (2005). RxNorm: prescription for electronic drug information exchange. *IT professional, 7*(5), 17-23.

Lohr, J. B., & Flynn, K. (1992). Smoking and schizophrenia. *Schizophrenia research, 8*(2), 93-102.

Lovestone, S. (2011). *Translational research – facilitating experimental medicine in dementia in the UK*. Director Biomedical Research Unit for dementia at Maudsley and King's.

Lucas, C., & Martin, J. (2013). Smoking and drug interactions. *Aust Prescr, 36*(3), 102-104.

Luo, Y., Thompson, W. K., Herr, T. M., Zeng, Z., Berendsen, M. A., Jonnalagadda, S. R., . . . Starren, J. (2017). Natural language processing for EHR-based pharmacovigilance: a structured review. *Drug safety, 40*(11), 1075-1089.

Lysaker, P. H., Davis, L. W., Warman, D. M., Strasburger, A., & Beattie, N. (2007). Stigma, social function and symptoms in schizophrenia and schizoaffective disorder: associations across 6 months. *Psychiatry research, 149*(1), 89-95.

Macerollo, A., & Martino, D. (2016). What is new in tics, dystonia and chorea? *Clinical Medicine, 16*(4), 383-389.

McCallum, A., & Li, W. (2003). Early results for named entity recognition with conditional random fields, feature induction and web-enhanced lexicons.

McCallum, A. K. (2002). Mallet: A machine learning for language toolkit (2002). In.

McElroy, S., Keck, J. P., & Pope, J. H. (1987). Sodium valproate: its use in primary psychiatric disorders. *Journal of clinical psychopharmacology, 7*(1), 16-24.

McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia medica: Biochemia medica, 22*(3), 276-282.

Mental Health Unit (BRC). (2018). D-CRIS. Retrieved from http://www.slam.nhs.uk/research/d-cris

MHRA. (2017). MHRA Yellow Card Scheme. Retrieved from https://yellowcard.mhra.gov.uk/

Miller, D. D. (2000). Review and management of clozapine side effects. *Journal of Clinical Psychiatry*.

Mortimer, A., Singh, P., Shepherd, C., & Puthiryackal, J. (2010). Clozapine for treatment-resistant schizophrenia: National Institute of Clinical Excellence (NICE) guidance in the real world. *Clinical schizophrenia & related psychoses, 4*(1), 49-55.

Mostert, M., Bredenoord, A. L., Biesaart, M. C., & van Delden, J. J. (2016). Big Data in medical research and EU data protection law: challenges to the consent or anonymise approach. *European Journal of Human Genetics, 24*(7), 956.

Murff, H. J., Forster, A. J., Peterson, J. F., Fiskio, J. M., Heiman, H. L., & Bates, D. W. (2003). Electronically screening discharge summaries for adverse medical events. *Journal of the American Medical Informatics Association, 10*(4), 339-350.

Nadeau, D., & Sekine, S. (2007). A survey of named entity recognition and classification. *Lingvisticae Investigationes, 30*(1), 3-26.

Nebeker, J. R., Barach, P., & Samore, M. H. (2004). Clarifying adverse drug events: a clinician's guide to terminology, documentation, and reporting. *Annals of internal medicine, 140*(10), 795-801.

Newgard, C. D., & Lewis, R. J. (2015). Missing data: how to best account for what is not known. *Jama, 314*(9), 940-941.

Newman, S. C., & Bland, R. C. (1991). Mortality in a cohort of patients with schizophrenia: a record linkage study. *The Canadian Journal of Psychiatry, 36*(4), 239-245.

Nguyen, T. H., & Grishman, R. (2015). *Event detection and domain adaptation with convolutional neural networks.* Paper presented at the Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers).

NHS Digital. (2018). Antidepressants. Retrieved from https://www.nhs.uk/conditions/antidepressants/

Nielsen, J., Røge, R., Schjerning, O., Sørensen, H. J., & Taylor, D. (2012). Geographical and temporal variations in clozapine prescription for schizophrenia. *European neuropsychopharmacology, 22*(11), 818-824.

Nielsen, J., Skadhede, S., & Correll, C. U. (2010). Antipsychotics associated with the development of type 2 diabetes in antipsychotic-naïve schizophrenia patients. *Neuropsychopharmacology, 35*(9), 1997-2004.

Nivya, K., Kiran, V. S. S., Ragoo, N., Jayaprakash, B., & Sekhar, M. S. (2015). Systemic review on drug related hospital admissions–A pubmed based search. *Saudi Pharmaceutical Journal, 23*(1), 1-8.

Oxford Health NHS Foundation Trust. (2018). Oxford Health NHS Foundation Trust. Retrieved from https://www.oxfordhealth.nhs.uk/

Pacurariu, A. C., Straus, S. M., Trifiro, G., Schuemie, M. J., Gini, R., Herings, R., . . . Pedersen, L. (2015). Useful interplay between spontaneous ADR reports and electronic healthcare records in signal detection. *Drug safety, 38*(12), 1201-1210.

Pais, S., Parry, D., & Huang, Y. (2017). *Suitability of fast healthcare interoperability resources (FHIR) for wellness data.* Paper presented at the Proceedings of the 50th Hawaii International Conference on System Sciences.

Pandey, C., Ibrahim, Z., Wu, H., Iqbal, E., & Dobson, R. (2017). *Improving RNN with attention and embedding for adverse drug reactions.* Paper presented at the Proceedings of the 2017 International Conference on Digital Health.

Patadia, V. K., Coloma, P., Schuemie, M. J., Herings, R., Gini, R., Mazzaglia, G., . . . van der Lei, J. (2015). Using real-world healthcare data for pharmacovigilance signal detection–the experience of the EU-ADR project. *Expert review of clinical pharmacology, 8*(1), 95-102.

Patel, P., & Zed, P. J. (2002). Drug-Related Visits to the Emergency Department: How Big Is the Problem? *Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy, 22*(7), 915-923.

Penz, J. F., Wilcox, A. B., & Hurdle, J. F. (2007). Automated identification of adverse events related to central venous catheters. *Journal of biomedical informatics, 40*(2), 174-182.

Perera, G., Broadbent, M., Callard, F., Chang, C.-K., Downs, J., Dutta, R., . . . Jackson, R. (2016). 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, 6*(3), e008721.

Peuskens, J., Gillain, B., De Graeve, D., Van Vleymen, B., & Albert, A. (2009). Belgian Schizophrenia Outcome Survey–Results of a 2-year naturalistic study in patients stabilised on monotherapy with olanzapine, risperidone or haloperidol. *European psychiatry, 24*(3), 154-163.

Phillips, M. L., & Kupfer, D. J. (2013). Bipolar disorder diagnosis: challenges and future directions. *The Lancet, 381*(9878), 1663-1671.

Poole, J. H., Tobias, F. C., & Vinogradov, S. (2000). The functional relevance of affect recognition errors in schizophrenia. *Journal of the International Neuropsychological Society, 6*(6), 649-658.

Popejoy, L. L., Khalilia, M. A., Popescu, M., Galambos, C., Lyons, V., Rantz, M., . . . Stetzer, F. (2015). Quantifying care coordination using natural language processing and domain-specific ontology. *Journal of the American Medical Informatics Association, 22*(e1), e93-e103.

Postigo, R., Brosch, S., Slattery, J., van Haren, A., Dogné, J.-M., Kurz, X., . . . Arlett, P. (2018). EudraVigilance Medicines Safety Database: Publicly Accessible Data for Research and Public Health Protection. *Drug safety*, 1-11.

Pradhan, S., Chapman, W., Man, S., & Savova, G. (2014). *Semeval-2014 task 7: Analysis of clinical text.* Paper presented at the Proc. of the 8th International Workshop on Semantic Evaluation (SemEval 2014.

PricewaterhouseCoopers. (2016). *Clinical Pharmacology and Therapeutics -The case for savings in the NHS*. Retrieved from https://www.bps.ac.uk/getmedia/8f24b222-b355-4f4a-8d91-f0a56190381e/The-case-for-savings-in-the-NHS-December-2016.pdf.aspx

Rademaker, M. (2001). Do women have more adverse drug reactions? *American journal of clinical dermatology, 2*(6), 349-351.

Radomsky, E. D., Haas, G. L., Mann, J. J., & Sweeney, J. A. (1999). Suicidal behavior in patients with schizophrenia and other psychotic disorders. *American Journal of Psychiatry, 156*(10), 1590-1595.

Raja, M. (2011). Clozapine safety, 35 years later. *Current drug safety, 6*(3), 164-184.

Raja, M., & Raja, S. (2014). Clozapine safety, 40 years later. *Current drug safety, 9*(3), 163-195.

Ratey, J. J., & Salzman, C. (1984). Recognizing and managing akathisia. *Psychiatric Services, 35*(10), 975-977.

Ratinov, L., & Roth, D. (2009). *Design challenges and misconceptions in named entity recognition.* Paper presented at the Proceedings of the Thirteenth Conference on Computational Natural Language Learning (CoNLL-2009).

Robson, D., & Gray, R. (2007). Serious mental illness and physical health problems: a discussion paper. *International journal of nursing studies, 44*(3), 457-466.

Rochefort, C. M., Verma, A. D., Eguale, T., Lee, T. C., & Buckeridge, D. L. (2015). A novel method of adverse event detection can accurately identify venous thromboembolisms (VTEs) from narrative electronic health record data. *Journal of the American Medical Informatics Association, 22*(1), 155-165.

Rochon, P. A., Stukel, T. A., Sykora, K., Gill, S., Garfinkel, S., Anderson, G. M., . . . Li, P. (2005). Atypical antipsychotics and parkinsonism. *Archives of internal medicine, 165*(16), 1882-1888.

Rodenburg, E. M., Stricker, B. H. C., & Visser, L. E. (2011). Sex-related differences in hospital admissions attributed to adverse drug reactions in the Netherlands. *British journal of clinical pharmacology, 71*(1), 95-104.

Rosenman, M., He, J., Martin, J., Nutakki, K., Eckert, G., Lane, K., . . . Hui, S. L. (2014). Database queries for hospitalizations for acute congestive heart failure: flexible methods and validation based on set theory. *Journal of the American Medical Informatics Association, 21*(2), 345-352.

Rummel-Kluge, C., Komossa, K., Schwarz, S., Hunger, H., Schmid, F., Kissling, W., . . . Leucht, S. (2010). Second-generation antipsychotic drugs and extrapyramidal side effects: a systematic review and meta-analysis of head-to-head comparisons. *Schizophrenia bulletin, 38*(1), 167-177.

Rummel-Kluge, C., Komossa, K., Schwarz, S., Hunger, H., Schmid, F., Lobos, C. A., . . . Leucht, S. (2010). Head-to-head comparisons of metabolic side effects of second generation antipsychotics in the treatment of schizophrenia: a systematic review and meta-analysis. *Schizophrenia research, 123*(2), 225-233.

Safferman, A., Lieberman, J. A., Kane, J. M., Szymanski, S., & Kinon, B. (1991). Update on the clinical efficacy and side effects of clozapine. *Schizophrenia bulletin, 17*(2), 247-261.

Sanders, T. B., Bowens, F. M., Pierce, W., Stasher-Booker, B., Thompson, E. Q., & Jones, W. A. (2012). The road to ICD-10-CM/PCS implementation: forecasting the transition for providers, payers, and other healthcare organizations. *Perspectives in health information management/AHIMA, American Health Information Management Association, 9*(Winter).

Saposnik, G., Redelmeier, D., Ruff, C. C., & Tobler, P. N. (2016). Cognitive biases associated with medical decisions: a systematic review. *BMC medical informatics and decision making, 16*(1), 138.

Sarker, A., Ginn, R., Nikfarjam, A., O'Connor, K., Smith, K., Jayaraman, S., . . . Gonzalez, G. (2015). Utilizing social media data for pharmacovigilance: a review. *Journal of biomedical informatics, 54*, 202-212.

Savica, R., Grossardt, B. R., Bower, J. H., Ahlskog, J. E., & Rocca, W. A. (2013). Incidence and pathology of synucleinopathies and tauopathies related to parkinsonism. *JAMA neurology, 70*(7), 859-866.

Savova, G. K., Fan, J., Ye, Z., Murphy, S. P., Zheng, J., Chute, C. G., & Kullo, I. J. (2010). *Discovering peripheral arterial disease cases from radiology notes using natural language processing.* Paper presented at the AMIA Annual Symposium Proceedings.

Savova, G. K., Masanz, J. J., Ogren, P. V., Zheng, J., Sohn, S., Kipper-Schuler, K. C., & Chute, C. G. (2010). Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. *Journal of the American Medical Informatics Association, 17*(5), 507-513.

Schafer, J. L. (1999). Multiple imputation: a primer. *Statistical methods in medical research, 8*(1), 3-15.

Schafer, J. L., & Graham, J. W. (2002). Missing data: our view of the state of the art. *Psychological methods, 7*(2), 147.

Schloesser, R. J., Martinowich, K., & Manji, H. K. (2012). Mood-stabilizing drugs: mechanisms of action. *Trends in neurosciences, 35*(1), 36-46.

Schneider, C., Corrigall, R., Hayes, D., Kyriakopoulos, M., & Frangou, S. (2014). Systematic review of the efficacy and tolerability of clozapine in the treatment of youth with early onset schizophrenia. *European Psychiatry, 29*(1), 1-10.

Schneider, L. S., Dagerman, K. S., & Insel, P. (2005). Risk of death with atypical antipsychotic drug treatment for dementia: meta-analysis of randomized placebo-controlled trials. *Jama, 294*(15), 1934-1943.

Schooler, N. R., & Kane, J. M. (1982). Research diagnoses for tardive dyskinesia. *Archives of general psychiatry, 39*(4), 486-487.

Schulte, P. F. (2003). What is an adequate trial with clozapine? *Clinical pharmacokinetics, 42*(7), 607-618.

Sengupta, G., Bhowmick, S., Hazra, A., Datta, A., & Rahaman, M. (2011). Adverse drug reaction monitoring in psychiatry out-patient department of an Indian teaching hospital. *Indian Journal of Pharmacology, 43*(1), 36.

Serrano, P., & Musunuri, P. (2017). Akathisia or Anxiety: What is the Difference? *The American Journal of Geriatric Psychiatry, 25*(3), S82-S83.

Severyn, A., & Moschitti, A. (2015). *Learning to rank short text pairs with convolutional deep neural networks.* Paper presented at the Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval.

Shivade, C., Raghavan, P., Fosler-Lussier, E., Embi, P. J., Elhadad, N., Johnson, S. B., & Lai, A. M. (2013). A review of approaches to identifying patient phenotype

cohorts using electronic health records. *Journal of the American Medical Informatics Association, 21*(2), 221-230.

Shojania, K. G., Duncan, B. W., McDonald, K. M., & Wachter, R. M. (2002). Safe but sound: patient safety meets evidence-based medicine. *Jama, 288*(4), 508-513.

Sohn, S., Kocher, J.-P. A., Chute, C. G., & Savova, G. K. (2011). Drug side effect extraction from clinical narratives of psychiatry and psychology patients. *Journal of the American Medical Informatics Association, 18*(Supplement_1), i144-i149.

Spivak, B., Shabash, E., Sheitman, B., Weizman, A., & Mester, R. (2003). The effects of clozapine versus haloperidol on measures of impulsive aggression and suicidality in chronic schizophrenia patients: an open, nonrandomized, 6-month study. *The Journal of clinical psychiatry, 64*(7), 755-760.

Sporn, A. L., Vermani, A., Greenstein, D. K., Bobb, A. J., Spencer, E. P., Clasen, L. S., . . . Lenane, M. C. (2007). Clozapine treatment of childhood-onset schizophrenia: evaluation of effectiveness, adverse effects, and long-term outcome. *Journal of the American Academy of Child & Adolescent Psychiatry, 46*(10), 1349-1356.

Stahl, S. M. (2000). *Essential psychopharmacology: Neuroscientific basis and practical applications*: Cambridge university press.

Star, K., Watson, S., Sandberg, L., Johansson, J., & Edwards, I. R. (2015). Longitudinal medical records as a complement to routine drug safety signal analysis. *Pharmacoepidemiology and drug safety, 24*(5), 486-494.

Staroselsky, M., Volk, L. A., Tsurikova, R., Pizziferri, L., Lippincott, M., Wald, J., & Bates, D. W. (2006). Improving electronic health record (EHR) accuracy and increasing compliance with health maintenance clinical guidelines through patient access and input. *International journal of medical informatics, 75*(10-11), 693-700.

Stausberg, J. (2014). International prevalence of adverse drug events in hospitals: an analysis of routine data from England, Germany, and the USA. *BMC health services research, 14*(1), 125.

Stern, T. A., & Anderson, W. H. (1979). Benztropine prophylaxis of dystonic reactions. *Psychopharmacology, 61*(3), 261-262.

Stewart, R., Soremekun, M., Perera, G., Broadbent, M., Callard, F., Denis, M., . . . Lovestone, S. (2009). The South London and Maudsley NHS Foundation Trust Biomedical Research Centre (SLAM BRC) case register: development and descriptive data. *BMC Psychiatry, 9*, 51. doi:10.1186/1471-244X-9-51

Sukumar, S. R., Natarajan, R., & Ferrell, R. K. (2015). Quality of Big Data in health care. *International journal of health care quality assurance*.

Sun, W., Rumshisky, A., & Uzuner, O. (2013). Evaluating temporal relations in clinical text: 2012 i2b2 Challenge. *Journal of the American Medical Informatics Association, 20*(5), 806-813.

Takeuchi, H., Fervaha, G., & Remington, G. (2016). Consistency between clinician and patient ratings of clozapine-induced side effects. *Schizophrenia research, 174*(1-3), 200.

Talbert, S., & Sole, M. L. (2013). Too much information: research issues associated with large databases. *Clinical Nurse Specialist, 27*(2), 73-80.

Tang, B., Wu, Y., Jiang, M., Chen, Y., Denny, J. C., & Xu, H. (2013). A hybrid system for temporal information extraction from clinical text. *Journal of the American Medical Informatics Association, 20*(5), 828-835.

Tange, H. J., Hasman, A., de Vries Robbé, P. F., & Schouten, H. C. (1997). Medical narratives in electronic medical records. *International journal of medical informatics, 46*(1), 7-29.

Tarsy, D., Baldessarini, R. J., & Tarazi, F. I. (2002). Effects of newer antipsychotics on extrapyramidal function. *CNS drugs, 16*(1), 23-45.

Tatonetti, N. P., Patrick, P. Y., Daneshjou, R., & Altman, R. B. (2012). Data-driven prediction of drug effects and interactions. *Science translational medicine, 4*(125), 125ra131-125ra131.

Taylor, D. M., Barnes, T. R., & Young, A. H. (2018). *The Maudsley prescribing guidelines in psychiatry*: John Wiley & Sons.

Taylor, D. M., Douglas-Hall, P., Olofinjana, B., Whiskey, E., & Thomas, A. (2009). Reasons for discontinuing clozapine: matched, case–control comparison with risperidone long-acting injection. *The British Journal of Psychiatry, 194*(2), 165-167.

Teixeira, P. L., Wei, W.-Q., Cronin, R. M., Mo, H., VanHouten, J. P., Carroll, R. J., . . . Edwards, T. L. (2016). Evaluating electronic health record data sources and algorithmic approaches to identify hypertensive individuals. *Journal of the American Medical Informatics Association, 24*(1), 162-171.

Thanvi, B., & Treadwell, S. (2009). Drug induced parkinsonism: a common cause of parkinsonism in older people. *Postgraduate medical journal, 85*(1004), 322-326.

Trifiro, G., Fourrier-Reglat, A., Sturkenboom, M., Diaz Acedo, C., & Van Der Lei, J. (2009). The EU-ADR project: preliminary results and perspective. *Stud Health Technol Inform, 148*(148), 43.

Truven Health Analytics. (2013). Micromedex Solutions. Retrieved from http://micromedex.com/

Turetz, M., Mozes, T., Toren, P., Chernauzan, T., Yoran-Hegesh, R., Mester, R., . . . Weizman, A. (1997). An open trial of clozapine in neuroleptic-resistant childhood-onset schizophrenia. *The British Journal of Psychiatry, 170*(6), 507-510.

U.S. Food and Drug Administration. Retrieved from https://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Surveillance/AdverseDrugEffects/ucm082193.htm

Uzuner, Ö., South, B. R., Shen, S., & DuVall, S. L. (2011). 2010 i2b2/VA challenge on concepts, assertions, and relations in clinical text. *Journal of the American Medical Informatics Association, 18*(5), 552-556.

van der Ploeg, T., Austin, P. C., & Steyerberg, E. W. (2014). Modern modelling techniques are data hungry: a simulation study for predicting dichotomous endpoints. *BMC medical research methodology, 14*(1), 137.

van Harten, P. N., Hoek, H. W., & Kahn, R. S. (1999). Acute dystonia induced by drug treatment. *Bmj, 319*(7210), 623-626.

van Stiphout, F., Zwart-van Rijkom, J., Versmissen, J., Koffijberg, H., Aarts, J., van der Sijs, I., . . . Egberts, A. (2018). Effects of training physicians in electronic prescribing in the outpatient setting on clinical, learning and behavioural outcomes: a cluster randomized trial. *British journal of clinical pharmacology*.

Vassy, J. L., Ho, Y.-L., Honerlaw, J., Cho, K., Gaziano, J. M., Wilson, P. W., & Gagnon, D. R. (2018). Yield and bias in defining a cohort study baseline from electronic health record data. *Journal of biomedical informatics, 78*, 54-59.

Verdoux, H., Pambrun, E., Cortaredona, S., Coldefy, M., Le Neindre, C., Tournier, M., & Verger, P. (2016). Geographical disparities in prescription practices of lithium and clozapine: a community-based study. *Acta Psychiatrica Scandinavica, 133*(6), 470-480.

Vermeer, N. S., Ebbers, H. C., Straus, S. M., Leufkens, H. G., Egberts, T. C., & De Bruin, M. L. (2016). The effect of exposure misclassification in spontaneous ADR reports on the time to detection of product-specific risks for biologicals: a simulation study. *Pharmacoepidemiology and drug safety, 25*(3), 297-306.

Walsh, K. E., Adams, W. G., Bauchner, H., Vinci, R. J., Chessare, J. B., Cooper, M. R., . . . Landrigan, C. P. (2006). Medication errors related to computerized order entry for children. *Pediatrics, 118*(5), 1872-1879.

Wang, G., Jung, K., Winnenburg, R., & Shah, N. H. (2015). A method for systematic discovery of adverse drug events from clinical notes. *Journal of the American Medical Informatics Association, 22*(6), 1196-1204.

Wang, S. V., Rogers, J. R., Jin, Y., Bates, D. W., & Fischer, M. A. (2017). Use of electronic healthcare records to identify complex patients with atrial fibrillation for targeted intervention. *Journal of the American Medical Informatics Association, 24*(2), 339-344.

Wang, X., Chase, H., Markatou, M., Hripcsak, G., & Friedman, C. (2010). Selecting information in electronic health records for knowledge acquisition. *Journal of biomedical informatics, 43*(4), 595-601.

Wang, X., Hripcsak, G., & Friedman, C. (2009). *Characterizing environmental and phenotypic associations using information theory and electronic health records.* Paper presented at the BMC bioinformatics.

Wang, X., Hripcsak, G., Markatou, M., & Friedman, C. (2009). Active computerized pharmacovigilance using natural language processing, statistics, and electronic health records: a feasibility study. *Journal of the American Medical Informatics Association, 16*(3), 328-337.

Wang, Y., Wang, L., Rastegar-Mojarad, M., Moon, S., Shen, F., Afzal, N., . . . Sohn, S. (2018). Clinical information extraction applications: a literature review. *Journal of biomedical informatics, 77*, 34-49.

Weber, G. M., Adams, W. G., Bernstam, E. V., Bickel, J. P., Fox, K. P., Marsolo, K., . . . Murphy, S. N. (2017). Biases introduced by filtering electronic health records for patients with “complete data”. *Journal of the American Medical Informatics Association, 24*(6), 1134-1141.

Wehmeier, P. M., Schüler-Springorum, M., Heiser, P., & Remschmidt, H. (2004). Chart review for potential features of myocarditis, pericarditis, and cardiomyopathy in children and adolescents treated with clozapine. *Journal of child and adolescent psychopharmacology, 14*(2), 267-271.

Wei, C.-H., Peng, Y., Leaman, R., Davis, A. P., Mattingly, C. J., Li, J., . . . Lu, Z. (2016). Assessing the state of the art in biomedical relation extraction: overview of the BioCreative V chemical-disease relation (CDR) task. *Database, 2016*.

Wei, Q., Ji, Z., Li, Z., Du, J., Wang, J., Xu, J., . . . Zhang, Y. (2019). A study of deep learning approaches for medication and adverse drug event extraction from clinical text. *Journal of the American Medical Informatics Association*.

Weich, S., Pearce, H. L., Croft, P., Singh, S., Crome, I., Bashford, J., & Frisher, M. (2014). Effect of anxiolytic and hypnotic drug prescriptions on mortality hazards: retrospective cohort study. *Bmj, 348*, g1996.

Weiden, P. J. (2007). EPS profiles: the atypical antipsychotics: are not all the same. *Journal of Psychiatric Practice®, 13*(1), 13-24.

Weiskopf, N. G., & Weng, C. (2013). Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. *Journal of the American Medical Informatics Association, 20*(1), 144-151.

Weissman, M. M., Bland, R. C., Canino, G. J., Faravelli, C., Greenwald, S., Hwu, H.-G., . . . Lellouch, J. (1996). Cross-national epidemiology of major depression and bipolar disorder. *Jama, 276*(4), 293-299.

Werbeloff, N., Osborn, D. P., Patel, R., Taylor, M., Stewart, R., Broadbent, M., & Hayes, J. F. (2018). The Camden & Islington Research Database: Using electronic mental health records for research. *PloS one, 13*(1), e0190703.

WHO. (2017). Uppsala Monitoring Centre. The use of the WHO-UMC system for standardised case causality assessment. *WHO, http://www. who-umc. org/graphics/4409.pdf*.

Willett, P. (2006). The Porter stemming algorithm: then and now. *Program, 40*(3), 219-223.

Wolf, A., Dedman, D., Campbell, J., Booth, H., Lunn, D., Chapman, J., & Myles, P. (2019). Data resource profile: Clinical Practice Research Datalink (CPRD) Aurum. *International journal of epidemiology*.

World Health Organisation. (1969). International drug monitoring: the role of the hospital, report of a WHO meeting [held in Geneva from 18 to 23 November 1968].

Wu, H., Toti, G., Morley, K. I., Ibrahim, Z. M., Folarin, A., Jackson, R., . . . Gale, D. (2018). SemEHR: A general-purpose semantic search system to surface semantic data from clinical notes for tailored care, trial recruitment, and clinical research. *Journal of the American Medical Informatics Association, 25*(5), 530-537.

Wu, T.-Y., Jen, M.-H., Bottle, A., Molokhia, M., Aylin, P., Bell, D., & Majeed, A. (2010). Ten-year trends in hospital admissions for adverse drug reactions in England 1999–2009. *Journal of the Royal Society of Medicine, 103*(6), 239-250.

Wudarsky, M., Nicolson, R., Hamburger, S. D., Spechler, L., Gochman, P., Bedwell, J., . . . Rapoport, J. L. (1999). Elevated prolactin in pediatric patients on typical and atypical antipsychotics. *Journal of child and adolescent psychopharmacology, 9*(4), 239-245.

Wunnava, S., Qin, X., Kakar, T., Sen, C., Rundensteiner, E. A., & Kong, X. (2019). Adverse drug event detection from electronic health records using hierarchical recurrent neural networks with dual-level embedding. *Drug safety, 42*(1), 113-122.

Xu, D., Yadav, V., & Bethard, S. (2018). UArizona at the MADE1. 0 NLP challenge. *Proceedings of machine learning research, 90*, 57.

Xu, H., Stenner, S. P., Doan, S., Johnson, K. B., Waitman, L. R., & Denny, J. C. (2010). MedEx: a medication information extraction system for clinical narratives. *Journal of the American Medical Informatics Association, 17*(1), 19-24.

Xu, J., Wu, Y., Zhang, Y., Wang, J., Lee, H.-J., & Xu, H. (2016). CD-REST: a system for extracting chemical-induced disease relation in literature. *Database, 2016*.

Yang, X., Bian, J., Gong, Y., Hogan, W. R., & Wu, Y. (2019). MADEx: a system for detecting medications, adverse drug events, and their relations from clinical notes. *Drug safety, 42*(1), 123-133.

Yao, L., Zhang, Y., Li, Y., Sanseau, P., & Agarwal, P. (2011). Electronic health records: Implications for drug discovery. *Drug discovery today, 16*(13-14), 594-599.

Yih, W.-t., He, X., & Meek, C. (2014). *Semantic parsing for single-relation question answering.* Paper presented at the Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers).

Yin, Z., Sulieman, L. M., & Malin, B. A. (2019). A systematic literature review of machine learning in online personal health data. *Journal of the American Medical Informatics Association, 26*(6), 561-576.

Young, C. R., Bowers Jr, M. B., & Mazure, C. M. (1998). Management of the adverse effects of clozapine. *Schizophrenia bulletin, 24*(3), 381-390.

Zhang, J.-P., Gallego, J. A., Robinson, D. G., Malhotra, A. K., Kane, J. M., & Correll, C. U. (2013). Efficacy and safety of individual second-generation vs. first-generation antipsychotics in first-episode psychosis: a systematic review and meta-analysis. *International Journal of Neuropsychopharmacology, 16*(6), 1205-1218.

Zhang, X., & Zhang, T. (2013). *Achieving scalability in a distributed electronic health record system.* Paper presented at the 2013 Science and Information Conference.

Zhou, J., & Xu, W. (2015). *End-to-end learning of semantic role labeling using recurrent neural networks.* Paper presented at the Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers).

Zhou, S.-F., Yang, L.-P., Zhou, Z.-W., Liu, Y.-H., & Chan, E. (2009). Insights into the substrate specificity, inhibitors, regulation, and polymorphisms and the clinical impact of human cytochrome P450 1A2. *The AAPS journal*, *11*(3), 481-494.

Appendices

A Chapter 4 – Supplementary Material

Supplementary Table A.1 Diagnosis breakdown in SLAM - Mental, Behavioural and Neurodevelopmental Disorders (ICD-10: F00-F99)

ICD 10 Category	ICD 10 Code	Category Name	Patient Total	(%)
Grand Total			200,984	100.00
F00/(G30-G32)	F00	Dementia in Alzheimer's disease	8,647	4.30
F01-F09	F01	Vascular Dementia	3,293	1.64
	F02	Dementia in other diseases classified elsewhere	720	0.36
	F03	Unspecified Dementia	2,694	1.34
	F04	Amnestic disorder due to known physiological condition	54	0.03
	F05	Delirium	2,798	1.39
	F06	Other mental disorders due to known physiological condition	4,297	2.14
	F07	Personality and behavioural disorders due to known physiological condition	238	0.12
	F09	Unspecified mental disorder due to known physiological condition	247	0.12
F10-F19	F10	Alcohol related disorders	12,904	6.42
	F11	Opioid related disorders	5,778	2.87
	F12	Cannabis related disorders	1,065	0.53
	F13	Sedative, hypnotic, or anxiolytic related disorders	192	0.10
	F14	Cocaine related disorders	893	0.44
	F15	Other stimulant related disorders	167	0.08
	F16	Hallucinogen related disorders	65	0.03
	F17	Nicotine dependence	332	0.17
	F18	Inhalant related disorders	20	0.01
	F19	Other psychoactive substance related disorders	4,511	2.24
F20-F29	F20	Schizophrenia	11,908	5.92
	F21	Schizotypal disorder	85	0.04
	F22	Delusional disorders	988	0.49
	F23	Brief psychotic disorder	1,723	0.86
	F24	Shared psychotic disorder	2	0.00
	F25	Schizoaffective disorders	1,684	0.84
	F28	Other psychotic disorder	265	0.13
	F29	Unspecified psychosis	1,739	0.87
F30-F39	F30	Manic episode	490	0.24
	F31	Bipolar disorder	4,449	2.21
	F32	Major depressive disorder, single episode	20,866	10.38
	F33	Major depressive disorder, recurrent	5,235	2.60
	F34	Persistent mood [affective] disorders	940	0.47
	F38	Other mood [affective] disorders	386	0.19

	F39	Unspecified mood [affective] disorder	1,079	0.54
F40-F48	F40	Phobic anxiety disorders	1,285	0.64
	F41	Other anxiety disorders	8,543	4.25
	F42	Obsessive-compulsive disorder	2,701	1.34
	F43	Reaction to severe stress, and adjustment disorders	11,109	5.53
	F44	Dissociative and conversion disorders	1,329	0.66
	F45	Somatoform disorders	1,372	0.68
	F48	Other nonpsychotic mental disorders	1,382	0.69
F50-F59	F50	Eating disorders	5,190	2.58
	F51	Sleep disorders not due to a substance or known physiological condition	131	0.07
	F52	Sexual dysfunction not due to a substance or known physiological condition	2,517	1.25
	F53	Puerperal psychosis	317	0.16
	F54	Psychological and behavioural factors associated with disorders or diseases classified elsewhere	65	0.03
	F55	Abuse of non-psychoactive substances	17	0.01
	F59	Unspecified behavioural syndromes associated with physiological disturbances and physical factors	10	0.00
F60-F69	F60	Specific personality disorders	3,234	1.61
	F61	Mixed and other personality disorders	226	0.11
	F62	Enduring personality changes, not attributable to brain damage and disease	23	0.01
	F63	Impulse disorders	125	0.06
	F64	Gender identity disorder	259	0.13
	F65	Paraphilias	22	0.01
	F66	Other sexual disorders	43	0.02
	F68	Other disorders of adult personality and behaviour	47	0.02
	F69	Unspecified disorder of adult personality and behaviour	183	0.09
F70-F79	F70	Mild intellectual disabilities	860	0.43
	F71	Moderate intellectual disabilities	544	0.27
	F72	Severe intellectual disabilities	322	0.16
	F73	Profound intellectual disabilities	17	0.01
	F78	Other intellectual disabilities	64	0.03
	F79	Unspecified intellectual disabilities	577	0.29
F80-F89	F80	Specific developmental disorders of speech and language	45	0.02
	F81	Specific developmental disorders of scholastic skills	29	0.01
	F82	Specific developmental disorder of motor function	2	0.00
	F83	Mixed specific developmental disorders	25	0.01
	F84	Pervasive developmental disorders	6,347	3.16

	F88	Other disorders of psychological development	8	0.00
	F89	Unspecified disorder of psychological development	52	0.03
F90-F98	F90	Attention-deficit hyperactivity disorders	6,068	3.02
	F91	Conduct disorders	1,354	0.67
	F92	Mixed disorders of conduct and emotions	1,874	0.93
	F93	Emotional disorders with onset specific to childhood	2,760	1.37
	F94	Disorders of social functioning with onset specific to childhood and adolescence	369	0.18
	F95	Tic disorder	230	0.11
	F98	Other behavioural and emotional disorders with onset usually occurring in childhood and adolescence	2,223	1.11
F99	F99	Mental disorder not otherwise specified	36,330	18.08

Supplementary Table A.2 Drug dictionary created for drug timeline

Primary Category	Secondary Category	Third Category	Generic Name	Brand Names
Antipsychotics	Typical		Chlorpromazine	Largactil Chlorpromazine
			Periciazine	Pericyazine
			Perphenazine	Perphenazine Fentazin
			Flupentixol	Flupentixol Fluanxol Depixol
			Haloperidol	Haloperidol Haldol Serenace
			Prochlorperazine	Prochlorperazine Stemetil
			Promazine	Promazine
			Trifluoperazine	Trifluoperazine Stelazine
			Sulpiride	Sulpiride Dolmatil Sulpor
			Zuclopenthixol	Zuclopenthixol Clopixol
	Atypical		Amisulpride	Amisulpride Solian
			Aripiprazole	Aripiprazole Abilify
			Asenapine	Asenapine Sycrest
			Clozapine	Clozapine Clozaril Zaponex Denzapine
			Lurasidone	Lurasidone Latuda
			Olanzapine	Olanzapine Zyprexa Zypadhera
			Paliperidone	Paliperidone Invega Xeplion
			Quetiapine	Quetiapine
			Risperidone	Risperidone Risperdal
			Ziprasidone	Ziprasidone
Antidepressant	Selective serotonin reuptake inhibitors (SSRIs)		Citalopram	Citalopram Cipramil
			Escitalopram	Cipralex Escitalopram
			Paroxetine	Paroxetine

				Seroxat
			Fluoxetine	Fluoxetine Prozac
			Fluvoxamine	Fluvoxamine Faverin
			Sertraline	Sertraline Lustral
	Tricyclic antidepressants (TCAs)		Amitriptyline	Amitriptyline Triptafen
			Clomipramine	Clomipramine Anafranil SR
			Dosulepin	Dosulepin Dothiepin Prothiaden
			Doxepin	Doxepin Sinepin
			Imipramine	Imipramine
			Lofepramine	Lofepramine Feprapax Lomont
			Nortriptyline	Nortriptyline Allegron
			Trimipramine	Trimipramine Surmontil
	Monoamine oxidase inhibitors (MAOIs)		Phenelzine	Phenelzine Nardil
			Tranylcypromine	Tranylcypromine
			Moclobemide	Moclobemide Manerix
			Isocarboxazid	Isocarboxazid
	Others		Reboxetine	Reboxetine Edronax
			Agomelatine	Agomelatine Valdoxan
			Duloxetine	Duloxetine Cymbalta
			Venlafaxine	Venlafaxine Venlafaxine M/R Alventa XL Bonilux XL Depefex XL Foraven XL Politid XL Ranfaxine XL Tifaxin XL Venaxx XL Vensir XL Winfex XL Venlalic XL

				Efexor XL
			Mirtazapine	Mirtazapine Zispin SolTab
Mood Stabilizer			Lithium	Lithium Carbonate Lithium Citrate Priadel Li-Liquid Camcolit Liskonum
			Carbamazepine	Carbamazepine Tegretol
			Sodium Valproate	Sodium Valproate Epilim Depakote
			Lamotrigine	Lamotrigine Lamictal
			Topiramate	Topiramate Topamax
			Gabapentin	Gabapentin Neurontin
Hypnotics and Anxiolytics			Alprazolam	Alprazolam Xanax
			Buspirone	Buspirone
			Chloral hydrate	Chloral hydrate Chloral betain Welldom
			Chlordiazepoxide	Chlordiazepoxide Librium
			Clomethiazole	Clomethiazole
			Flurazepam	Flurazepam Dalmane
			Loprazolam	Loprazolam
			Lorazepam	Lorazepam Ativan
			Lormetazepam	Lormetazepam
			Meltatonin	Meltatonin Circadin
			Meprobamate	Meprobamate
			Nitrazepam	Nitrazepam
			Oxazepam	Oxazepam
			Promethazine	Promethazine Phenergan Sominex
			Sodium Oxybate	Sodium Oxybate Xyrem
			Temazepam	Temazepam
			Zaleplon	Zaleplon Sonata
			Zolpidem	Zolpidem

				Stilnoct
			Zopiclone	Zopiclone Zimovane
Antihypertensi-ve	Diuretics	Thiazides	Bendroflumethiazide	Bendroflumethiazide Aprinox
			Chlortalidone	Chlortalidone Hygroton
			Cyclopenthiazide	Cyclopenthiazide Navidrex
			Indapamide	Indapamide Natrilix Natrilix SR Ethibide XL Tensaid XL Indipam XL
			Metolazone	Metolazone
			Xipamide	Xipamide Diurexan
		Loop Diuretics	Bumetanide	Bumetanide Burinex
			Furosemide	Furosemide Frusol Lasix
			Torasemide	Torasemide Torem
		Potassium-spring Diuretics and aldosterone antagonists	Amiloride	Amiloride Amilamont
			Triamterene	Triamterene Dytac
			Eplerenone	Eplerenone Inspra
			Spironolactone	Spironolactone Aldactone
	Calcium Channel Blocker		Amlodipine	Amlodipine Amlostin Istin
			Diltiazem	Diltiazem Adizem SR Adizem XL Angitil XL Angitil SR Calcicard CR Dilcardia SR Dilzem SR Dilzem XL Kenzem Slozem Retalzem Tildiem

				Tildiem Modified Release Tablets Tildiem Prolonged-Release Capsules Tildiem LA Tildiem Retard Viazem XL Zemtard
			Isradipine	Isradipine Prescal
			Felodipine	Felodipine Cardioplen XL Felogen XL Felotens XL Keloc SR Neofel XL Parmid XL Vascalpha Plendil
			Lacidipine	Lacidipine Motens
			Lercanidipine	Lercanidipine Zanidip
			Nicardipine	Nicardipine Cardene Cardene SR
			Nifedipine	Nifedipine Adalat Adalat LA Adalat Retard Adipine MR Adipine XL Coracten SR Coracten XL Fortipine LA Nifedipress MR Tensipine MR Valni XL
			Verapamil	Verapamil Zolvera Cordilox Securon Half Securon SR Securon SR Univer Verapress Verapress MR 240 Vertab SR 240
	Vasodilator		Ambrisentan	Ambrisentan

				Volibris
			Diazoxide	Diazoxide Eudemine
			Hydralazine	Hydralazine Apresoline
			Iloprost	Iloprost Ventavis
			Minoxidil	Minoxidil Loniten
			Sildenafil	Sildenafil Viagra Revatio Vizarsin Nipatra
			Tadalafil	Tadalafil Adcirca Cialis
	Centrally Acting		Clonidine	Clonidine Catapres Dixarit
			Methyldopa	Methyldopa Aldomet
			Moxonidine	Moxonidine Physiotens
	Alpha Adrenoceptor Blocker		Doxazosin	Doxazosin Cardozin XL Cardura Cardura XL Doxadura Doxadura XL Doxogen XL Raporsin XL Slocinx XL
			Indoramin	Indoramin Doralese Baratol
			Prazosin	Prazosin Hypovase
			Terazosin	Terazosin Hytrin
			Phenoxybenzami-ne	Phenoxybenzami-ne Dibenyline
			Phentolamine	Phentolamine Rogitine
	Renin-Angiotensin	Angiotens-in converting enzyme inhibitors	Captopril	Captopril Ecopace Kaplon Capoten Noyada
			Cilazapril	Cilazapril

				Vascace
			Enalapril	Enalapril Ednyt Innovace
			Fosinopril	Fosinopril
			Imidapril	Imidapril Tanatril
			Lisinopril	Lisinopril Zestril
			Moexipril	Moexipril Perdix
			Perindopril Erbumine	Perindopril
			Perindopril Arginine	Perindopril Arginine Coversyl
			Quinapril	Quinapril Quinil Accupro
			Ramipril	Ramipril Tritace
			Trandolapril	Trandolapril Gopten
		Angiotens-in II receptor antagonist	Azilsartan	Azilsartan Edarbi
			Candesartan	Candesartan Amias
			Eprosartan	Eprosartan Teveten
			Irbesartan	Irbesartan Sabervel Aprovel
			Losartan	Losartan Cozaar
			Olmesartan	Olmesartan Olmetec
			Telmisartan	Telmisartan Micardis
			Valsartan	Valsartan Diovan
		Renin Inhibitors	Aliskiren	Aliskiren Resilez
	Beta Blocker		Propranolol	Propranolol Angilol Bedranol SR Syprol Half Beta Prograne Beta Prograne Slo-Pro
			Acebutolol	Acebutolol

				Sectral
			Atenolol	Atenolol Tenormin
			Bisoprolol	Bisoprolol Cardicor Congescor Emcor
			Carvedilol	Carvedilol Eucardic
			Celiprolol	Celiprolol Celectol
			Esmolol	Esmolol Brevibloc
			Labetalol	Labetalol Trandate
			Metoprolol	Metoprolol Betaloc Lopresor Lopresor SR
			Nadolol	Nadolol Corgard
			Nebivolol	Nebivolol Nebilet
			Oxprenolol	Oxprenolol Trasicor Slow Trasicor Tablets
			Pindolol	Pindolol Visken
			Sotalol	Sotalol Sotacor Beta-Cardone
			Timolol	Timolol Betim
Lipid Regulatory Drugs	Statins		Atorvastatin	Atorvastatin Lipitor
			Fluvastatin	Fluvastatin Lescol Luvinsta XL Sinfatix XL Stefluvin XL Dorisin XL Pinmactil Lescol XL
			Pravastatin	Pravastatin Lipostat
			Rosuvastatin	Rosuvastatin Crestor
			Simvastatin	Simvastatin Simvador Zocor

	Bile Acid Sequestrants		Colesevelam	Colesevelam Cholestagel
			Colestyramine	Colestyramine Questran Questran Light
			Colestipol	Colestipol Colestid
	Ezetimibe		Ezetimibe	Ezetimibe Ezetrol
	Fibrates		Bezafibrate	Bezafibrate Bezalip Bezalip Mono Caberzol XL Fibrazate XL
			Ciprofibrate	Ciprofibrate
			Fenofibrate	Fenofibrate Lipantil Micro Supralip
			Gemfibrozil	Gemfibrozil Lopid
	Nicotinic Acid		Acipimox	Acipimox Olbetam
			Nicotinic	Nicotinic Acid Niaspan
	Omega 3 fatty acid compounds		Omega 3 acid ethyl esters	Omacor Prestylon
			Omega 3 marine triglycerides	Maxepa
Antidiabetic	Sulfonylureas		Glibenclamide	Glibenclamide
			Gliclazide	Gliclazide Diamicron Diamicron MR Laaglyda MR Zicron
			Glimepiride	Glimepiride Amaryl
			Glipizide	Glipizide Minodiab
			Tolbutamide	Tolbutamide
	Biguanides		Metformin	Metformin Glucophage Glucophage SR Diagemet XL Glucient SR Bolamyn SR Glucient SR Metabet SR
	Other Antidiabetics		Acarbose	Acarbose Glucobay
			Exenatide	Exenatide Byetta

				Bydureon
			Linagliptin	Linagliptin Trajenta
			Liraglutide	Liraglutide Victoza
			Nateglinide	Nateglinide Starlix
			Pioglitazone	Pioglitazone Actos Glidipion
			Repaglinide	Repaglinide Prandin NovoNorm
			Saxagliptin	Saxagliptin Onglyza
			Sitagliptin	Sitagliptin Januvia
			Vildagliptin	Vildagliptin Galvus
			Dapagliflozin	Dapagliflozin Forxiga
			Lixisenatide	Lixisenatide Lyxumia
Insulin	Short acting insulin		Insulin	Hypurin Bovine Neutral Hypurin Porcine Neutral Actrapid Humulin S Insuman Rapid
			Insulin Aspart	Insulin Aspart Novorapid
			Insulin Glulisine	Insulin Glulisine Apidra
			Insulin Lispro	Insulin Lispro Humalog
	Intermediat-e and long acting insulin		Insulin Degludec	Insulin Degludec Tresiba
			Insulin Detemir	Insulin Detemir Levemir
			Insulin Glargine	Insulin Glargine Lantus
			Insulin Zinc Suspension	Insulin Zinc Suspension Hypurin Bovine Lente
			Isophane Insulin	Isophane Insulin Hypurin Bovine Isophane Hypurin Porcine Isophane

				Insulatard Humulin I Insuman Basal
			Protamine Zinc Insulin	Hypurin Bovine Protatmine Zinc
	Biphasic Insulin		Biphasic Insulin Aspart	Novomix 30
			Biphasic Insulin Lispro	Humalog Mix25 Humalog Mix50
			Biphasic Isophane Insulin	Hypurin Porcine 30/70 Mix Humulin M3 Insuman Comb 15 Insuman Comb 25 Insuman Comb 50
Parkinson	Dopamine-receptor agonist		Apomorphine hydrochloride	Apomorphine Apo-Go
			Bromocriptine	Bromocriptine Parlodel
			Cabergoline	Cabergoline Cabaser Dostinex
			Pergolide	Pergolide
			Pramipexole	Pramipexole Mirapexin
			Ropinirole	Ropinirole Adartrel Aimpart Eppinix Ralnea Rapinex Raponer Requip Ropinirole
			Rotigotine	Rotigotine Neupro
	Levodopa		Co-Beneldopa	Co-Beneldopa Madopar
			Co-Careldopa	Co-Careldopa Sinemet Duodopa Half Sinemet Stalevo
	Monoamine-oxidase-B inhibitors		Rasagiline	Rasagiline Azilect
			Selegiline Hydrochloride	Selegiline Hydrochloride Eldepryl

				Zelapar
	Cataechol-O-Methyltransferase inhibitors		Entacapone	Entacapone Comtess (Stalevo)?
			Tolcapone	Tolcapone Tasmar
	Amantadine		Amantadine	Amantadine Lysovir Symmetrel
	Antimuscarinic drugs		Orphenadrine	Orphenadrine Disipal Biophen
			Procyclidine	Procyclidine Acrpicolin Kemadrin
			Trihexyphenidyl	Trihexyphenidyl Broflex
	Related Disorders		Piracetam	Piracetam Nootropil
			Riluzole	Riluzole Rilutek
			Tetrabenazine	Tetrabenazine Revocon Tetmodis
	Torsion Dystonia		Botulinum Toxin Type A	Botulinum Toxin Type A Azzalure Bocouture Boox Dysport Vistabel Xeomin
			Botulinum Toxin Type B	Botulinum Toxin Type B NeuroBloc
Non-Steroidal Anti Inflammatory Drugs (NSAIDs)			Aceclofenac	Aceclofenac Preservex
			Acemetacin	Acemetacin Emflex
			Celecoxib	Celecoxib Celebrex
			Dexibuprofen	Dexibuprofen Seractil
			Dexketoprofen	Dexketoprofen Keral
			Diclofenac Potassium	Diclofenac Potassium Boots Joint Pain Relief

				Double Action Pain Relief Voltarol
			Diclofenac Sodium	Diclofenac Sodium Dicloflex Diclomax Retard Diclomax SR Econac Econac SR Econac XL Fenactol Motifene Rheumatac Retard Rhumalgan Solaraze Voltarol Ampoules Voltarol Dispersible Voltarol Gel Voltarol Ophtha Voltarol SR Voltarol Retard Voltarol Suppositories
			Etodolac	Etodolac Eccoxolac Etopan XL Lodine SR
			Etoricoxib	Etoricoxib Arcoxia
			Fenoprofen	Fenoprofen
			Flurbiprofen	Flurbiprofen Froben Ocufen Strefen
			Ibuprofen	Ibuprofen Anadin Ibuprofen Anadin Joint Pain Anadin LiquiFast Anadin Period Pain Anadin Ultra Anadin Ultra Double Burfen Burfen Retard Burfen Syrup Calprofen
			Indometacin	Indometacin

				Indocid Indolar Indometacin
			Ketoprofen	Ketoprofen Oruids Oruvil Axorid Oruvail Powergel
			Mefenamic Acid	Mefenamic Acid Ponstan
			Meloxicam	Meloxicam
			Nabumetone	Nabumetone Relifex
			Naproxen	Naproxen Feminax Ultra Naprosyn
			Piroxicam	Piroxicam Brexidol Feldene
			Sulindac	Sulindac Tenoxicam Mobiflex
			Tiaprofenic	Tiaprofenic Surgam
Dementia			Donepezil	Donepezil Aricept
			Galantamine	Galantamine Elmino Galantamine Galsya Gatalin Lotprosin Luventa Reminyl
			Memantine	Memantine Ebixa Maruxa Nemdatine
			Rivastigmine	Alzest Exelon Nimvastid Voleze
Antiepileptic			Acetazolamide	Acetazolamide DIamox Eytazox
			Clobazam	Clobazam Frisium Perizam Tapclob
			Clonazepam	Clonazepam

			Diazepam	Diazepam
			Eslicarbazepine	Eslicarbazepine Zebinix
			Ethosuximide	Ethosuximide Zarontin
			Lacosamide	Lacosamide Vimpat
			Levetiracetam	Levetiracetam Desitrend
			Nitrazepam	Nitrazepam Mogadon
			Oxcarbazepine	Oxcarbazepine Trileptal
			Perampanel	Perampanel Fycompa
			Piracetam	Piracetam Nootropil
			Phenobarbital	Phenobarbital
			Phenytoin	Phenytoin Epanutin
			Pregabalin	Pregabalin Alzain Lecaent LYRICA
			Primidone	Primidone
			Retigabine	Retigabine Trobalt
			Rufinamide	Rufinamide Inovelon
			Stiripentol	Stiripentol
			Tiagabine	Tiagabine Gabitril
			Vigabatrin	Vigabatrin Sabril
			Zonisamide	Zonisamide Zonegran

Supplementary Table A.3 Medication algorithm results in the three trusts categorised by primary category, secondary and generic names.

Results show percentages of patients who have been on medication for at least six weeks in SLAM (n=134723) between January 2007 to June 2016, C&I (n=56329) between January 2009 to July 2016 and Oxford (n=51260) between January 2010 to March 2015

Primary Category	Secondary Category	Drugs	SLAM (n=134723)	C&I (n=56329)	Oxford (n=52160)
Antidepressant	Tricyclic antidepressants (TCAs)	Amitriptyline	2.0	2.5	2.8
		Clomipramine	0.4	0.4	0.5
		Dosulepin	0.6	0.4	0.3
		Imipramine	0.1		0.3
		Lofepramine	0.5	0.5	0.7
		Nortriptyline	0.9		
		Trimipramine	0.9		
	Monoamine oxidase inhibitors (MAOIs)	Moclobemide	0.9		
	Selective serotonin reuptake inhibitors (SSRIs)	Citalopram	12.2	12.1	14.2
		Escitalopram	1.9	0.8	0.5
		Fluoxetine	7.8	5.4	1.2
		Paroxetine	1.9	1.9	1.2
		Sertraline	9.7	8.6	11.3
	Other antidepressants	Agomelatine	0.1		
		Duloxetine	1.3	0.7	0.8
		Mirtazapine	9.9	8.5	9.3
		Reboxetine	0.2		
		Venlafaxine	4.5	3.6	4.8
Antidiabetic	Biguanides	Metformin	2.5	2.4	1.8
	Other Antidiabetics	Sitagliptin	0.2	0.2	
	Sulfonylureas	Gliclazide	1.0	1.4	0.6
Antiepileptic		Clobazam	0.1		
		Clonazepam	6.9	2.6	2.2
		Levetiracetam	0.3	0.2	
		Nitrazepam		0.3	0.3
		Phenytoin	0.2	0.2	

Primary Category	Secondary Category	Drugs	SLAM (n=134723)	C&I (n=56329)	Oxford (n=52160)
		Pregabalin	1.8	1.4	0.8
Antihypertensive	Alpha-Adrenoceptor Blocker	Doxazosin	0.4	0.4	0.2
	Beta Blocker	Atenolol	1.0	0.9	0.9
		Bisoprolol	1.4	1.4	1.3
		Metoprolol	0.8		
		Propranolol	1.0	1.6	1.1
	Calcium Channel Blocker	Amlodipine	2.8	2.5	1.7
		Diltiazem	0.1		
		Felodipine	0.2		
		Nifedipine	0.2	0.2	
	Centrally Acting	Clonidine	0.2		
	Diuretics	Amiloride	1.0		
		Bendroflumethiazide	0.9	0.9	0.8
		Bumetanide	0.2		
		Furosemide	1.4	1.1	1.3
		Indapamide	0.2	0.2	
		Spironolactone	0.3	0.3	0.2
	Renin-Angiotensin	Candesartan	0.3		0.4
		Enalapril	0.2	0.5	
		Irbesartan	0.1		
		Lisinopril	0.5	0.6	0.9
		Losartan	0.4	0.7	0.3
		Perindopril Erbumine	0.2	0.3	
		Ramipril	2.5	1.9	1.6
	Vasodilator	Sildenafil	0.2	0.3	
		Tadalafil	0.2		
Antipsychotic	Atypical	Amisulpride	2.3	1.6	0.7
		Aripiprazole	6.1	3.9	4.2
		Asenapine	0.9		
		Clozapine	3.0	1.9	1.6
		Lurasidone	0.2		
		Olanzapine	13.7	9.6	9.2

Primary Category	Secondary Category	Drugs	SLAM (n=134723)	C&I (n=56329)	Oxford (n=52160)
		Paliperidone	1.3	0.3	0.4
		Quetiapine	6.2	7.0	8.0
		Risperidone	10.0	9.2	7.2
		Sulpiride	0.6	1.3	0.3
	Typical	Chlorpromazine	0.8	1.6	0.5
		Flupentixol	1.5	1.6	0.9
		Fluphenazine	0.6		
		Haloperidol	3.4	4.6	1.6
		Pipotiazine	0.7		
		Promazine	0.1		
		Trifluoperazine	0.6	0.4	0.3
		Zuclopenthixol	1.4	1.8	0.9
Dementia		Donepezil	3.9	4.7	7.7
		Galantamine	0.3		0.6
		Memantine	1.3	1.7	1.3
		Rivastigmine	0.4	0.8	0.9
Hypnotics and Anxiolytics		Alprazolam	0.8		
		Buspirone	0.3	0.2	
		Chlordiazepoxide	2.0	1.0	0.6
		Diazepam	7.9	9.5	8.3
		Lorazepam	8.5	7.2	7.2
		Melatonin	0.4		0.6
		Nitrazepam	0.2	0.3	0.3
		Oxazepam	0.4		
		Promethazine	9.8	1.4	1.0
		Temazepam	1.1	2.3	3.5
		Zolpidem	0.7	0.4	0.4
		Zopiclone	1.3	9.3	8.1
Lipid Regulatory Drugs	Ezetimibe	Ezetimibe	0.1		
	Omega 3 fatty acid compounds	Omega 3 acid ethyl esters	0.1		
	Statins	Atorvastatin	1.8	2.4	1.2
		Pravastatin	0.2	0.2	
		Rosuvastatin	0.8		
		Simvastatin	3.8	3.9	3.3

Primary Category	Secondary Category	Drugs	SLAM (n=134723)	C&I (n=56329)	Oxford (n=52160)
Mood Stabilizer		Carbamazepine	1.4	1.3	0.8
		Gabapentin	0.8	1.0	0.7
		Lamotrigine	2.3	1.8	1.8
		Lithium	2.3	2.2	3.7
		Sodium Valproate	2.2	1.9	1.9
		Topiramate	0.3		
Non-steroidal anti-inflammatory drugs		Diclofenac Sodium	0.2	0.2	
		Ibuprofen	2.2	1.7	1.9
		Naproxen	0.4	0.8	0.5
Parkinson	Antimuscarinic drugs	Orphenadrine	0.5	0.3	
		Procyclidine	3.6	3.5	1.8
		Trihexyphenidyl	0.1	0.2	
	Levodopa	Co-Penelope	0.9		
		Co-Careldopa	0.2	0.2	0.3

Supplementary Table A.4 ADE Dictionary

The ADEs list, along with synonyms, phrases and alternative spellings were generated with the help of pharmacists and clinicans.

ADEs	Alternate spellings, Synonyms and phrases to describe and ADE
Abdominal pain	Abdo discomfort
	Abdo issues
	Abdo pain
	Abdominal cramps
	Abdominal discomfort
	Abdominal issues
	Abdominal Pain
	Abdominalpain
	Agony in abdo region
	Agony in abdominal region
	Butterflies in stomach
	Chest pain
	Discomfort in abdominal region
	Heart burn
	Heavy feeling in chest
	Pain in abdo region
	Pain in abdominal region
Agitation	Agitated
	Agitation
	Did not remain still
	Easily annoyed
	Feeling agitated
	Fidgety
	Psychomotor agitation
	Restless
	Restlessness
	Unable to relax
	Very tense
Akathisia	acasthesia
	acasthisia
	acathasia
	acathesia
	acathysia
	acisthesia
	acisthisia
	akasthesia
	akasthisia
	akathasia

	akathesia
	akathesic
	akathetic
	akathisia
	akathisic
	akathsia
	akithesia
	akithesic
	akithisia
	akithisic
Akinesia	akinesia
Alopecia	alopecia
	hair loss
	hairloss
Amenorrhoea	Absence of menstrual cycle
	Absence of menstruation
	Amenorhoea
	Amenorrhea
	Amenorrhoea
	Amenorrhoea Absence of period
	ammenorhoea
	Change in period
	Missed period
	No period
	No periods
	Not had period
Amnesia	Amnesia
	acopia
	Anxiety
	anxiety and panic
	Anxious
	constant fear
	constant worry
	feeling dreadful
	feeling fearful
	feelings of panic
	has dreadful feelings
	nervous
	stressed
	tenseness
Apnoea	Apnoea
	apnea
	not breathing
	apneic

Arrhythmia	Arrythmia
	Arrythmogenic
	Cardiac rhythm disturbance
	Cardiacarrhythmias
	Disturbed heart beat
	Dysrhythmia
	Heart rhythm is irregular
	Irregular heart beat
	Irregular heart rate
	Irregular heartbeat
Arthralgia	arthralgia
	joint pain
	arthralgias
	joint pains
	aching joints
	painful joints
	painful joint
	pain Joint
	pain in joint
Ataxia	Ataxia
	ataxias
Backache	Back ache
	Back pain
	Backache
	Pain in the back
Blurred vision	Blur vision
	Blurred vision
	Blurring
	Blurry
	Blurry vision
	Clouded vision
	Eye sight is blurry
	Eye sight is cloudy
	Eye sight is fuzzy
	Fuzzy vision
	Hard to see things
	Less than sharp vision
	Loss of ability to see small details
	Loss of visual acuity
	Sight is impaired
	Vision blurred
Bradycardia	Bradycardia
	Bradycardic
	Heart block

	Low pulse
	Palpitation
	Palpitations
	Skipped heart beat
	Slow heart beat
	Unusual slow heart beat
	Unusually slow heart beat
Cardiomyopathy	cardiomyopathy
catatonia	catatonia
	catatonic
	stupor
Confusion	Confused thoughts
	confusion
	Confusion aggravated
	Gets very confused
	Incoherent
	Lack of clear thinking
	Unclear thoughts
Constipation	bowel movements are difficult
	bowel movements less often than normal
	bowel obstruction
	Can’t open bowels
	cannot open bowel
	constipate
	constipated
	constipation
	difficult bowel movements
	difficult evacuation of feces
	Difficulty in passing stool
	Difficulty in passing stools
	faecal impaction
	fecal impaction
	few bowel movements
	Hard stools
	impacted faeces
	Infrequent bowel movement
	infrequent evacuation of feces
	Not defecated
	Not had a bowel movement
	paralytic ileus
	small stools
	straining to pass stool
	tough to pass stool
	trouble having a bowel movement

	trouble passing stool
	unsatisfactory bowel movement
Convulsion	convulse
	convulsing
	convulsion
	convulsions
	convulsive
	Epileptic
	Grand-mal
	having a fit
	having fits
	seizure
	seizures
	Tonic-clonic
Cystitis	Cystitis
	bladder inflammation
	cystalgia
Decreased appetite	appetite decrease
	appetite decreased
	decrease appetite
	decrease in appetite
	decreased appetite
	poor appetite
	reduced appetite
Dehydration	Dehydration
	excessive loss of water
Delusion	Delusion
	Delusions
Diarrhoea	Diarhea
	Diarrhea
	Diarrhoea
	diarrohea
	Frequent stool
	Frequent stools
	loose stool
	Loose stools
	Watery stool
	Watery stools
Diplopia	Diplopia
	double vision
Disorientation	chaotic
	Disorientated
	disorientation
	disoriented

	no orientation
	not orientated
	poor orientation
	poorly orientated
Dizziness	Disequilibrium
	Dizziness
	Dizzy
	Dizzy spells
	Faint
	Feel like they are going to fall
	Giddiness
	Lightheaded
	Lightheadeness
	Might lose her balance
	Might lose his balance
	Off-legs
	Pre-syncope
	Sense of turning
	Unsteadiness
	Unsteady
	Unsteady on feet
	Whirling sensation
DKA	DKA
	ketoacidosis
Dry Mouth	Burning in the mouth
	Cracked lips
	dry feeling in the mouth
	Dry feeling in throat
	Dry Mouth
	Dry tongue
	Frequent thirst
	Hoarseness
	Mouth Dryness
	No saliva
	Problems speaking
	Sore throat
	Split lips
	Sticky feeling in the mouth
	Tingling on tongue
	Tingling sensation in the mouth
	Xerostomia
Dysarthria	Dysarthria
Dyslipidemia	Abnormal lipids
	Deranged lipids

	Dyslipidemia
	High cholesterol
	High triglycerides
	Hyperlipidaemia
	Hyperlipidemia
	Hypertriglycaeridaemia
	Hypertriglyceridemia
	Raised blood lipids
	Raised cholesterol
	Raised lipids
	Raised triglycerides
Dyspepsia	Belching
	Bloated
	Burning pain behind breastbone
	burping
	Discomfort in abdomen
	Dyspepsia
	Feeling full
	Gastro oesophageal reflux disorder
	GORD
	Heartburn
	Indigestion
	Pain in abdomen
	Regurgitating fluids or food
Dystonia	dystonia
	dystonic
Enuresis	Enuresis
	Overactive bladder syndrome
	Passing urine frequently
	Urinary frequency
	Urinary incontinence
	Urinary urgency
	Wet the bed
	Wetting the bed
Eosinophilia	eosinophilia
	eosinophils increased
	increased eosinophil count
	increased eosinophils
	increased eosinophils count
Epilepsy	epilepsies
	epilepsy
	epilepsy types
	epileptic
	epileptic seizure

	epileptic seizures
	epileptics
	seizure disorder
	seizure disorders
Erectile Dysfunction	Anorgasmia
	Cannot maintain an erection
	Cannot obtain an erection
	cannot satisfy boyfriend sexually
	cannot satisfy girlfirend sexually
	cannot satisfy husband sexually
	cannot satisfy lover sexually
	cannot satisfy partner sexually
	Cannot satisfy wife sexually
	Can't maintain an erection
	Can't obtain an erection
	can't satisfy boyfriend sexually
	can't satisfy girlfirend sexually
	can't satisfy husband sexually
	can't satisfy lover sexually
	can't satisfy partner sexually
	Can't satisfy wife sexually
	Decreased orgasm
	Difficulties maintaining an erection
	Difficulties obtaining an erection
	Does not ejaculate
	Ejaculation problems
	Erectile dysfunction
	Erectile problems
	Failure to reach orgasm
	Inability to achieve an erection
	Inability to achieve an orgasm
	Inability to maintain an erection
	Problems with sexual function
	Sexual difficulties
	Sexual side effects
	Trouble getting a hard on
	Trouble getting an erection
	Trouble sustaining an erection
Extrapyramidal disorder	extra pyramidal disorder
	extra pyramidal disorders
	extra-pyramidal disorder
	extrapyramidal disorder
	extrapyramidal disorders
	extra-pyramidal syndrome

	extrapyramidal syndrome
	extrapyramidal syndromes
Extrapyramidal symptoms	extra pyramidal sign
	extra pyramidal symptom
	extrapyramidal sign
	extra-pyramidal sign
	extrapyramidal symptom
	extra-pyramidal symptom
Fatigue	Asthenia
	Asthenic
	Decreased energy
	Exhausted
	Exhaustion
	Fatigability
	Fatigue
	Knackered
	Lack of energy
	Lacking energy
	Lethargic
	Lethargy
	Loss of energy
	Sleeping excessively
	spaced out
	TATT
	Tired
	Tired all the time
	Tiredness
	Underactive thyroid
	Weariness
	Worn out
	zombie like
Feeling sick	being sick
	Feeling sick
	sick
Fever	Feeling hot
	Fever
	High temperature
	Increased temperature
	Pyrexia
	Pyrexial
	Raised body temperature
	Temperature over 37 degrees
	Termperature over 37°C
Flatulence	farted

	farting
	flatulence
	flatus
	gassy belly
	passing gas
Galactorhoea	Discharge from nipple
	Galactorhoea
	Galactorrhea
	Galactorrhoea
	Milk discharge from breast
	Milk leaking from breast
Glycosuria	Glycosuria
	glucosuria
	glucose urine present
	gluco suria
	glyco suria
Gynaecomastia	Breast enlargement
	breast size increase
	Breast size increased
	Enlarge Breast
	Enlarged breast tissues
	Gynaecomastia
Hallucination	halluci-nation
	hallucination
	halluci-nations
	hallucinations
Headache	cephalagia
	Constant head pain
	head ache
	Head pain
	Headache
	Headaches
	Migraines
	Pain in head
	Throbbing pain in head
High CPK	High CK
	High CPK
	increased CK
	increased CPK
	Increased creatine kinase
	Raised CK
	Raised CPK
	Raised creatine kinase
Hostility	hostility

Hyperglycaemia	Glucose > 7.8 but < 11.1
	Glucose > 7.8 but <11.1
	Glucose >7.8 but < 11.1
	Glucose >7.8 but <11.1
	High blood sugar
	High BM
	Hyperglycaemia
	Hyperglycaemic
	IGT
	Impaired glucose tolerance
	Raised BM
	Raised glucose
Hyperprolactinaemia	High prolactin
	Hyperprolactinaemia
	Increased levels of prolactin
	Prolactin
	Prolactin ++
	Raised prolactin
hypersalivation	drool
	drooling
	hypersalivation
	hyper-salivation
	Increased saliva
	Increased salivation
	ptyalism
	sialorrhea
	Too much saliva
Hypersomnia	excessive sleepiness
	hyper-somnia
	hypersomnia
	increased sleep
Hypertension	Blood pressure is elevated
	Blood pressure is high
	BP is elevated
	BP is high
	Elevated blood pressure
	Elevated BP
	High blood pressure
	High BP
	Hypertension
	Hypertensive
	Raised blood pressure
	Raised BP
Hypokinesia	decreased activity

	decreased motor activity
	hypoactive
	hypoactivity
	hypokinesia
	reduced activity
Hypotension	Blood pressure is low
	BP is low
	Hypotension
	Hypotensive
	Low blood pressure
	Low BP
Impotence	Decreased libido
	Delay ejaculation
	Delayed orgasm
	Difficult in getting erection
	Difficult in maintaining erection
	Difficult in reaching orgasm
	Difficulty in getting arouse
	Ejaculatory incompetence
	Have problems enjoying sex
	Impotence
	Inability to get an erection
	Lack of sexual desire
	Premature ejaculation
	Retarded ejaculation
	Sexual disfunction
	Sexual Dysfunction
Increased thirst	Increased thirst
	thirst increased
	thirst
	thirsty
Insomnia	Acute insomnia
	Broken sleep
	Can’t sleep
	cannot sleep
	Chronic insomnia
	Difficult getting to sleep
	Difficult to sleep
	Difficulty falling asleep
	difficulty sleeping
	Difficulty staying asleep
	Disrupted sleeping patterns
	Disturned sleep
	Early awakening

	Early morning waking (EMW)
	Feeling tired after waking up
	Feeling tired upon waking
	Few hours of sleep
	Generally tired
	Insomania
	Insomia
	Insomnia
	Light sleep
	Loss of sleep
	Lying awake for a long time at night
	Not sleeping
	Secondary insomnia
	Sleep disruption
	Sleeping during day
	Sleepless night
	Trouble going back to sleep
	Trouble sleeping
	Troubled sleep
	Wakes up during the night
	Waking up during the night
	Waking up too early
Irritable	feeling irritable
	hostile
	irritable
	jitteriness
	jittery
loss of libido	decreased sexual desire
	difficulty achieving orgasm during masturbation
	difficulty achieving orgasm during sex
	diminished sex drive
	loss of libido
	loss of sexual desire
	Low sex drive
	not interested in sex
	reduced libido
Muscle Pain	Muscle ache
	Muscle cramp
	Muscle fatigue
	muscle pain
	Pain in muscle
	Stiffness
Muscle twitching	Chorea
	Choreic movements

	Constant muscle contractions
	Corea
	Coreic movements
	Involuntary movements
	Involuntary muscle movements
	Muscle distortion
	Muscle fasciculations
	Muscle jerking
	Muscle jerks
	Muscle spasm
	Muscle spasming
	Muscle twitches
	Muscle twitching
Myocarditis	myocarditis
Nasopharyngitis	Nasopharyngitis
	Naso-pharyngitis
Nausea	nausea
	nauseated
	nauseating
	nauseous
Neuropathy peripheral	Neuropathy peripheral
	peripheral nervous system disorder
	peripheral nervous system disorders
	Peripheral NervousSystem Disorders
	peripheral neuropathies
	peripheral neuropathy
	peripheral neuropathy
	PeripheralNervous System Disorders
Neutropenia	neutropenia
	neutropenias
	neutrophils decreased
Nightmare	Nightmare
	nightmares
	paroniria
NMS	Neuroleptic Malignant Syndrome
	NMS
Numbness	Altered sensation
	chills
	Feels cold
	Feels feverish
	Flu-like
	Numbness
	Pin pricks
	Pins and needles

	Pins prickling
	prickly skin
	Sensation of cold
	Shivering
	Shivers
	Shock like sensation
	Shock-like sensation
	Shocks
	Skin tingles
	Tingle
	Tingling feeling
	Tingling feet
	Tingling sensation
	Tingly
Nystagmus	Nystagmus
Oedema	edemas
	edematous
	oedema
	oedemas
	oedematous
Parasuicide	Parasuicide
	suicide gesture
parkinsonian	parkinsonian
	parkinsonism
	pseudo-parkinsonian
	pseudo-parkinsonism
pericarditis	myopericarditis
	pericarditis
	pleuro-pericarditis
Peripheral edema	Peripheral edema
	Peripheral oedema
	Pitting edema
	Pitting odema
	Pitting oedema
	Swollen arms
	Swollen feet
	Swollen foot
	Swollen legs
Pharyngitis	irritation in the throat
	irritation of the throat
	pharyngeal inflammation
	pharyngitis
	throat inflammation
	throat soreness and irritation

Pneumonia	bronchopneumonia
	bronchopneumonias
	pneumonia
	pneumonias
Polyuria	excessive diuresis
	high urine output
	increase in urinary output
	polyuria
	urine volume increased
Pulmonary embolism	PE
	Pulmonary embolism
QTC	QT prolongation
	QTC
Rash	Hives
	Rash
	Rashes
	Skin Rash
	Urticaria
Rhinitis	Rhinitis
Sedation	drowsiness
	Drowsy
	Feeling sleepy
	Feels sedated
	Over sedated
	Sedated
	Sedation
	Sleepiness
	Sleepy
	Slepiness
	Somnolence
Shaking	Hands shake
	Involuntary shakiness
	Shakiness
	Shaking
SJS	SJS
	stephen johnson
	stephen johnsons
	stephen-johnson
	stephen-johnsons
	stephens johnson
	stephens johnsons
	stephens-johnson
	steven johnson
	steven johnsons

	steven johnston
	steven jonson
	steven-johnson
	steven-johnsons
	stevens johnson
	Stevens Johnsons
	stevens-johnson
	stevens-johnsons
Skinreactions	Acne
	Blotches
	Blothchy skin
	Cutaneous eruptions
	Epidermal eruptions
	Exanthem
	Exanthema
	Maculapupalar
	Skin eruptions
	Skin lesions
	skin reaction
	Skinreactions
	Spotty
Stomach pain	Abdominal ache
	Dull ache in tummy
	Gastritis
	Severe abdominal pain
	Stomach ache
	Stomach cramps
	Stomach pain
	upset belly
	Upset tummy
Suicidal behaviour	Suicidal behaviour
	suicidal behavior
Suicidal ideation	Suicidal ideation
	suicidal thought
Suicidal tendency	Suicidal tendency
Suicide attempt	Suicide attempt
	attempted suicide
	suicidal attempt
Sweating	Cold sweaty hands
	Diaphoresis
	Drenched in sweat
	Hyperhidrosis
	Night sweats
	Perspiration

	Sweating
	Sweating excessively
	Sweats a lot
	Sweats at night
Syncope	Disorders Syncope
	fainting
	faintness
	syncope
Tachycardia	Elevated pulse
	Fast heart beat
	Fast heart rate
	Fast pulse
	Heart beating too fast
	Increased pulse
	Pulse [>100] beats per minute
	Pulse [>100] bpm
	Pulse is elevated
	Racing heart
	Rapid heart beats
	tachycardia
	tachycardic
Tardive Dyskinesia	tardith diskinesia
	Tardith Dyskinesia
	tardive diskinesia
	tardive diskinesias
	Tardive Dyskinesia
	tardive dyskinesias
Tinnitus	Tinnitus
	buzzing in the ears
Trembling	trembling
Tremor	Tremor
Urinary retention	Difficulty urinating
	Problem passing urine
	Urinary retention
Vaginal inflammation	Colpitis
	inflammation of vagina
	vaginal inflammation
	vaginitis
Vertigo	spinning sensation
	vertigo
Vomiting	puking
	Retching
	Throwing up
	Vomit

	Vomiting
	Vomitting
WBC decreased	decreased leukocytes
	decreased WBC
	decreased white blood cell count
	decreased white blood cells
	leukocyte count decreased
	leukocytes decreased
	low WBC
	WBC count decreased
	WBC decreased
	WBC low
	white blood cell count decreased
	white blood cells decreased
Weight gain	eating a lot more
	gain weight
	gained weight
	gaining weight
	increase in body weight
	increased body weight
	increased weight
	put on weight
	putting on weight
	regain weight
	weight gain
	weight gained
	weight gaining
	weight increase
	weight increased
	worried about her weight gain
Weight Loss	Body weight gone down
	Decrease in body weight
	Losing weight
	Lost weight
	Reduced body weight
	Reduction in body weight
	Wasting away
	Weight decrease
	Weight Loss
	Weight Lost
Agranulocytosis	agranulocytosis
	granulocytopenia
	granulopenia
	agranulocytoses

	agran ulocytosis
	gran ulocytopenia

B Chapter 5 – Supplementary Material

Supplementary Table B.1 Gender differences (%)

Clozapine - Gender Differences (%)

ADR	Trust	Cohort	Sub Cohort	Three Months Early	Two Months Early	One Month Early	One Month Later	Two Months Later	Three Months Later	SIDER Low End	SIDER High End
Agitation	SLAM	Trust	Clozapine (n=1760)	17.61	22.10	26.53	46.59	32.56	26.99		
		Gender	Male (n=1167)	17.31	20.31	24.68	44.56	31.71	27.34		
			Female (n=593)	18.21	25.63	30.19	50.59	34.23	26.31		
	Camden & Islington	Trust	Clozapine (n=561)	13.37	17.83	18.36	43.14	28.34	21.03		
		Gender	Male (n=357)	11.20	17.37	18.21	45.66	27.45	22.13		
			Female (n=204)	17.16	18.63	18.63	38.73	29.90	19.12		
	Oxford	Trust	Clozapine (n=514)	14.59	15.76	16.34	34.24	25.10	20.62		
		Gender	Male (n=342)	13.45	13.45	14.62	34.50	25.73	19.30		
			Female (n=172)	16.86	20.35	19.77	33.72	23.84	23.26		
	SIDER	SIDER	SIDER							4.00	
Fatigue	SLAM	Trust	Clozapine (n=1760)	12.67	14.83	15.85	43.58	35.80	30.51		
		Gender	Male (n=1167)	10.80	13.20	13.54	40.19	32.82	28.45		
			Female (n=593)	16.36	18.04	20.40	50.25	41.65	34.57		
	Camden & Islington	Trust	Clozapine (n=561)	10.34	12.30	13.37	41.18	29.23	26.56		
		Gender	Male (n=357)	8.68	10.64	11.76	38.94	29.97	24.93		
			Female (n=204)	13.24	15.20	16.18	45.10	27.94	29.41		
	Oxford	Trust	Clozapine (n=514)	9.73	11.87	12.06	35.21	27.43	26.85		
		Gender	Male (n=342)	9.06	11.40	11.99	31.87	24.85	26.32		
			Female (n=172)	11.05	12.79	12.21	41.86	32.56	27.91		
	SIDER	SIDER	SIDER								
Sedation	SLAM	Trust	Clozapine (n=1760)	12.67	12.16	14.83	43.86	35.51	29.83		
		Gender	Male (n=1167)	12.34	11.48	13.11	41.99	33.68	27.68		
			Female (n=593)	13.32	13.49	18.21	47.55	39.12	34.06		
	Camden & Islington	Trust	Clozapine (n=561)	5.17	9.09	9.09	38.15	26.56	21.93		
		Gender	Male (n=357)	5.60	8.12	8.68	40.06	26.05	23.53		
			Female (n=204)	4.41	10.78	9.80	34.80	27.45	19.12		
	Oxford	Trust	Clozapine (n=514)	7.20	8.37	9.34	31.52	21.40	18.48		
		Gender	Male (n=342)	6.14	7.89	8.77	30.99	21.35	17.54		
			Female (n=172)	9.30	9.30	10.47	32.56	21.51	20.35		
	SIDER	SIDER	SIDER							25.00	46.00
Dizziness	SLAM	Trust	Clozapine (n=1760)	2.78	4.20	4.09	16.59	13.13	11.19		
		Gender	Male (n=1167)	2.06	3.43	3.51	14.31	11.91	10.37		
			Female (n=593)	4.22	5.73	5.23	21.08	15.51	12.82		
	Camden & Islington	Trust	Clozapine (n=561)	3.21	3.39	3.74	18.18	13.73	9.09		
		Gender	Male (n=357)	2.52	3.08	5.04	15.97	11.48	8.12		
			Female (n=204)	4.41	3.92	1.47	22.06	17.65	10.78		
	Oxford	Trust	Clozapine (n=514)	3.89	4.09	4.47	17.70	13.04	10.12		
		Gender	Male (n=342)	2.63	3.22	3.22	14.91	11.11	8.48		
			Female (n=172)	6.40	5.81	6.98	23.26	16.86	13.37		
	SIDER	SIDER	SIDER							12.00	27.00
Hypersalivation	SLAM	Trust	Clozapine (n=1760)	1.19	1.48	2.10	14.32	13.24	11.31		
		Gender	Male (n=1167)	1.46	2.14	2.57	15.25	12.60	10.80		
			Female (n=593)	0.67	0.17	1.18	12.48	14.50	12.31		
	Camden & Islington	Trust	Clozapine (n=561)	1.07	1.43	0.53	14.26	6.95	7.66		
		Gender	Male (n=357)	1.40	1.96	0.56	14.57	7.84	6.72		
			Female (n=204)	0.49	0.49	0.49	13.73	5.39	9.31		
	Oxford	Trust	Clozapine (n=514)	0.97	0.78	1.56	12.65	10.70	5.84		
		Gender	Male (n=342)	0.88	0.29	1.17	10.82	10.23	4.68		
			Female (n=172)	1.16	1.74	2.33	16.28	11.63	8.14		
	SIDER	SIDER	SIDER							1.00	48.00
Feelingsick	SLAM	Trust	Clozapine (n=1760)	4.66	4.94	6.48	14.32	11.19	9.09		
		Gender	Male (n=1167)	4.37	4.54	6.43	11.31	8.83	7.28		
			Female (n=593)	5.23	5.73	6.58	20.24	15.85	12.65		
	Camden & Islington	Trust	Clozapine (n=561)	3.74	3.92	3.03	10.52	7.13	7.66		
		Gender	Male (n=357)	3.64	3.92	3.36	10.08	6.16	7.84		
			Female (n=204)	3.92	3.92	2.45	11.27	8.82	7.35		
	Oxford	Trust	Clozapine (n=514)	3.89	5.25	5.06	14.20	9.73	7.20		
		Gender	Male (n=342)	3.80	3.80	3.80	13.16	8.77	5.56		
			Female (n=172)	4.07	5.81	7.56	16.28	11.63	10.47		
	SIDER	SIDER	SIDER								
Weightgain	SLAM	Trust	Clozapine (n=1760)	3.75	4.43	5.06	15.34	10.91	10.34		
		Gender	Male (n=1167)	3.77	4.46	4.54	13.45	9.08	9.17		
			Female (n=593)	3.71	4.38	6.07	19.06	14.50	12.65		
	Camden & Islington	Trust	Clozapine (n=561)	2.50	3.39	1.96	11.76	6.60	6.24		
		Gender	Male (n=357)	2.52	2.80	1.96	10.08	4.76	6.44		
			Female (n=204)	2.45	4.41	1.96	14.71	9.80	5.88		

Measure Values: 0.00 – 56.00

The results are shown in percentages (%) and broken down by ADRs, Trusts (SLAM, Camden & Islington and Oxford), Cohorts, Sub Cohorts and SIDER reported values.
In Sub Cohort 'Clozapine' represent the total baseline population which further breaks down into Gender 'Male' and 'Female' groups.
The columns (Three Months Early, Two Months Early, One Month Early, One Month Later, Two Months Later, Three Months Later) shows the percentages in each monthly interval. The last two columns (SIDER Low End and SIDER High End) shows the SIDER reporting.

Clozapine - Gender Differences (%)

ADR	Trust	Cohort	Sub Cohort	Three Months Early	Two Months Early	One Month Early	One Month Later	Two Months Later	Three Months Later	SIDER Low End	SIDER High End
Weightgain	Oxford	Trust	Clozapine (n=514)	3.50	3.31	3.70	11.28	9.92	7.78		
		Gender	Male (n=342)	3.51	3.22	3.22	10.53	8.77	7.31		
			Female (n=172)	3.49	3.49	4.65	12.79	12.21	8.72		
	SIDER	SIDER	SIDER							4.00	56.00
Tachycardia	SLAM	Trust	Clozapine (n=1760)	2.27	2.05	2.50	15.40	12.95	9.94		
		Gender	Male (n=1167)	2.40	1.80	2.23	16.28	13.11	10.37		
			Female (n=593)	2.02	2.53	3.04	13.66	12.65	9.11		
	Camden & Islington	Trust	Clozapine (n=561)	1.43	1.43	0.89	11.23	8.38	6.95		
		Gender	Male (n=357)	0.84	1.12	1.12	11.20	8.40	7.28		
			Female (n=204)	2.45	1.96	0.49	11.27	8.33	6.37		
	Oxford	Trust	Clozapine (n=514)	0.78	1.36	1.56	10.89	10.51	7.59		
		Gender	Male (n=342)	0.88	1.17	1.17	10.82	10.53	7.31		
			Female (n=172)	0.58	1.74	2.33	11.05	10.47	8.14		
	SIDER	SIDER	SIDER							11.00	25.00
Confusion	SLAM	Trust	Clozapine (n=1760)	4.72	5.51	6.08	13.92	8.47	6.76		
		Gender	Male (n=1167)	4.46	5.40	6.34	13.37	8.40	7.11		
			Female (n=593)	5.23	5.73	5.56	15.01	8.60	6.07		
	Camden & Islington	Trust	Clozapine (n=561)	3.57	6.24	5.53	12.66	6.77	5.88		
		Gender	Male (n=357)	3.92	6.72	5.32	13.45	7.84	7.28		
			Female (n=204)	2.94	5.39	5.88	11.27	4.90	3.43		
	Oxford	Trust	Clozapine (n=514)	2.53	3.89	3.89	9.92	6.42	5.25		
		Gender	Male (n=342)	2.05	2.63	2.92	8.19	6.14	4.09		
			Female (n=172)	3.49	6.40	5.81	13.37	6.98	7.56		
	SIDER	SIDER	SIDER							3.00	
Constipation	SLAM	Trust	Clozapine (n=1760)	1.76	1.99	2.16	12.27	11.70	9.49		
		Gender	Male (n=1167)	1.11	1.54	1.63	10.45	9.17	8.23		
			Female (n=593)	3.04	2.87	3.20	15.85	16.69	11.97		
	Camden & Islington	Trust	Clozapine (n=561)	1.07	2.50	1.78	11.41	7.13	5.70		
		Gender	Male (n=357)	0.56	2.80	1.68	10.36	7.00	3.92		
			Female (n=204)	1.96	1.96	1.96	13.24	7.35	8.82		
	Oxford	Trust	Clozapine (n=514)	0.58	0.97	1.36	10.31	7.78	7.78		
		Gender	Male (n=342)	0.29	0.88	1.17	8.77	6.43	6.43		
			Female (n=172)	1.16	1.16	1.74	13.37	10.47	10.47		
	SIDER	SIDER	SIDER							10.00	25.00
Headache	SLAM	Trust	Clozapine (n=1760)	4.20	4.55	5.45	12.44	8.18	5.91		
		Gender	Male (n=1167)	3.51	3.26	4.46	10.45	7.46	5.31		
			Female (n=593)	5.56	7.08	7.42	16.36	9.61	7.08		
	Camden & Islington	Trust	Clozapine (n=561)	2.32	3.57	4.28	9.27	6.42	4.63		
		Gender	Male (n=357)	2.24	3.64	3.64	10.08	6.72	4.48		
			Female (n=204)	2.45	3.43	5.39	7.84	5.88	4.90		
	Oxford	Trust	Clozapine (n=514)	3.89	3.89	4.09	10.89	8.37	7.59		
		Gender	Male (n=342)	3.51	3.80	4.09	11.11	8.77	7.31		
			Female (n=172)	4.65	4.07	4.07	10.47	7.56	8.14		
	SIDER	SIDER	SIDER								
Insomnia	SLAM	Trust	Clozapine (n=1760)	3.92	4.03	5.17	10.40	6.48	4.03		
		Gender	Male (n=1167)	3.68	4.03	4.46	9.85	6.00	4.54		
			Female (n=593)	4.38	4.05	6.58	11.47	7.42	3.04		
	Camden & Islington	Trust	Clozapine (n=561)	3.57	3.39	3.74	8.91	3.39	4.28		
		Gender	Male (n=357)	3.64	3.08	4.20	10.08	2.80	3.36		
			Female (n=204)	3.43	3.92	2.94	6.86	4.41	5.88		
	Oxford	Trust	Clozapine (n=514)	5.84	4.86	5.84	8.37	6.81	4.09		
		Gender	Male (n=342)	4.68	3.51	4.39	5.85	5.85	3.51		
			Female (n=172)	8.14	7.56	8.72	13.37	8.72	5.23		
	SIDER	SIDER	SIDER							20.00	33.00
Hyperprolactinaemia	SLAM	Trust	Clozapine (n=1760)	3.18	3.64	4.20	8.52	5.06	4.15		
		Gender	Male (n=1167)	2.66	2.49	2.91	6.43	3.86	3.68		
			Female (n=593)	4.22	5.90	6.75	12.65	7.42	5.06		
	Camden & Islington	Trust	Clozapine (n=561)	1.60	1.78	2.67	8.20	4.10	3.57		
		Gender	Male (n=357)	0.84	0.56	1.96	6.44	2.52	2.52		
			Female (n=204)	2.94	3.92	3.92	11.27	6.86	5.39		
	Oxford	Trust	Clozapine (n=514)	3.70	4.09	4.28	8.75	4.47	4.86		
		Gender	Male (n=342)	2.92	2.34	2.63	7.60	3.51	3.80		
			Female (n=172)	5.23	7.56	7.56	11.05	6.40	6.98		
	SIDER	SIDER	SIDER								
Shaking	SLAM	Trust	Clozapine (n=1760)	3.13	2.95	3.92	9.55	5.40	5.06		
		Gender	Male (n=1167)	2.57	3.00	3.68	9.08	5.23	5.40		

Measure Values: 0.00 – 56.00

The results are shown in percentages (%) and broken down by ADRs, Trusts (SLAM, Camden & Islington and Oxford), Cohorts, Sub Cohorts and SIDER reported values.
In Sub Cohort 'Clozapine' represent the total baseline population which further breaks down into Gender 'Male' and 'Female' groups.
The columns (Three Months Early, Two Months Early, One Month Early, One Month Later, Two Months Later, Three Months Later) shows the percentages in each monthly interval. The last two columns (SIDER Low End and SIDER High End) shows the SIDER reporting.

Clozapine - Gender Differences (%)

ADR	Trust	Cohort	Sub Cohort	Three Months Early	Two Months Early	One Month Early	One Month Later	Two Months Later	Three Months Later	SIDER Low End	SIDER High End
Shaking	SLAM	Gender	Female (n=593)	4.22	2.87	4.38	10.46	5.73	4.38		
	Camden & Islington	Trust	Clozapine (n=561)	1.78	1.96	3.74	6.06	3.92	2.85		
		Gender	Male (n=357)	1.96	1.96	4.48	6.16	3.92	3.36		
			Female (n=204)	1.47	1.96	2.45	5.88	3.92	1.96		
	Oxford	Trust	Clozapine (n=514)	2.92	3.31	3.89	7.78	4.47	4.86		
		Gender	Male (n=342)	2.05	2.92	2.92	7.31	4.39	3.22		
			Female (n=172)	4.65	4.07	5.81	8.72	4.65	8.14		
	SIDER	SIDER	SIDER								
Vomiting	SLAM	Trust	Clozapine (n=1760)	2.56	2.50	3.01	8.86	6.82	5.00		
		Gender	Male (n=1167)	1.97	2.14	3.34	8.57	5.91	4.46		
			Female (n=593)	3.71	3.20	2.36	9.44	8.60	6.07		
	Camden & Islington	Trust	Clozapine (n=561)	2.14	2.50	2.85	6.77	4.99	4.63		
		Gender	Male (n=357)	2.24	2.52	1.40	6.72	4.20	3.64		
			Female (n=204)	1.96	2.45	5.39	6.86	6.37	6.37		
	Oxford	Trust	Clozapine (n=514)	1.75	2.72	2.92	7.59	5.25	5.06		
		Gender	Male (n=342)	2.05	2.34	2.05	7.89	4.09	4.97		
			Female (n=172)	1.16	3.49	4.65	6.98	7.56	5.23		
	SIDER	SIDER	SIDER							3.00	17.00
Hypertension	SLAM	Trust	Clozapine (n=1760)	2.05	2.22	3.13	9.15	5.74	4.60		
		Gender	Male (n=1167)	1.80	1.46	2.49	8.48	5.40	4.54		
			Female (n=593)	2.53	3.71	4.38	10.46	6.41	4.72		
	Camden & Islington	Trust	Clozapine (n=561)	0.71	0.71	1.60	7.13	4.63	2.67		
		Gender	Male (n=357)	0.28	0.28	0.84	6.72	5.04	2.80		
			Female (n=204)	1.47	1.47	2.94	7.84	3.92	2.45		
	Oxford	Trust	Clozapine (n=514)	1.36	1.36	1.56	5.06	4.28	2.14		
		Gender	Male (n=342)	0.88	0.88	0.58	5.56	3.80	2.34		
			Female (n=172)	2.33	2.33	3.49	4.07	5.23	1.74		
	SIDER	SIDER	SIDER							4.00	12.00
Abdominalpain	SLAM	Trust	Clozapine (n=1760)	1.88	1.99	2.56	8.01	6.02	4.72		
		Gender	Male (n=1167)	1.80	1.97	2.31	7.97	6.00	4.20		
			Female (n=593)	2.02	2.02	3.04	8.09	6.07	5.73		
	Camden & Islington	Trust	Clozapine (n=561)	0.89	0.89	1.60	3.92	3.57	3.39		
		Gender	Male (n=357)	0.28	1.40	1.96	3.64	3.36	3.08		
			Female (n=204)	0.98	0.00	0.98	4.41	3.92	3.92		
	Oxford	Trust	Clozapine (n=514)	1.75	1.36	1.75	7.39	4.47	5.64		
		Gender	Male (n=342)	1.46	0.88	1.17	7.60	4.68	4.39		
			Female (n=172)	2.33	2.33	2.91	6.98	4.07	8.14		
	SIDER	SIDER	SIDER							4.00	
Backache	SLAM	Trust	Clozapine (n=1760)	1.14	1.59	2.44	4.94	3.35	2.73		
		Gender	Male (n=1167)	0.94	1.29	1.97	4.03	2.83	2.57		
			Female (n=593)	1.52	2.19	3.37	6.75	4.38	3.04		
	Camden & Islington	Trust	Clozapine (n=561)	1.43	1.25	1.96	5.35	3.03	3.03		
		Gender	Male (n=357)	0.84	1.40	2.24	3.08	2.24	1.96		
			Female (n=204)	2.45	0.98	1.47	9.31	4.41	4.90		
	Oxford	Trust	Clozapine (n=514)	1.17	1.17	1.36	5.84	3.89	2.92		
		Gender	Male (n=342)	0.88	0.88	1.17	3.80	3.22	1.75		
			Female (n=172)	1.74	1.74	1.74	9.88	5.23	5.23		
	SIDER	SIDER	SIDER							5.00	
Nausea	SLAM	Trust	Clozapine (n=1760)	1.14	1.08	1.19	6.08	5.23	3.69		
		Gender	Male (n=1167)	1.03	0.86	1.29	5.06	4.54	2.66		
			Female (n=593)	1.35	1.52	1.01	8.09	6.58	5.73		
	Camden & Islington	Trust	Clozapine (n=561)	0.89	1.43	0.36	4.63	3.57	3.57		
		Gender	Male (n=357)	0.00	1.40	0.28	3.92	2.80	2.80		
			Female (n=204)	2.45	1.47	0.49	5.88	4.90	4.90		
	Oxford	Trust	Clozapine (n=514)	0.97	0.58	1.36	4.86	3.70	2.92		
		Gender	Male (n=342)	0.58	0.58	0.58	3.51	3.22	2.34		
			Female (n=172)	1.74	0.58	2.91	7.56	4.65	4.07		
	SIDER	SIDER	SIDER							3.00	17.00
Convulsion	SLAM	Trust	Clozapine (n=1760)	1.36	1.70	1.82	7.05	4.94	4.03		
		Gender	Male (n=1167)	1.37	1.89	1.63	6.86	4.54	4.03		
			Female (n=593)	1.35	1.35	2.19	7.42	5.73	4.05		
	Camden & Islington	Trust	Clozapine (n=561)	0.53	0.53	0.36	2.85	2.14	1.07		
		Gender	Male (n=357)	0.56	0.56	0.56	3.08	2.24	0.84		
			Female (n=204)	0.49	0.49	0.00	2.45	1.96	1.47		
	Oxford	Trust	Clozapine (n=514)	1.36	1.36	1.56	6.42	3.11	2.72		
		Gender	Male (n=342)	0.88	0.88	1.17	6.43	4.09	2.05		

Measure Values: 0.00 – 56.00

The results are shown in percentages (%) and broken down by ADRs, Trusts (SLAM, Camden & Islington and Oxford), Cohorts, Sub Cohorts and SIDER reported values.
In Sub Cohort 'Clozapine' represent the total baseline population which further breaks down into Gender 'Male' and 'Female' groups.
The columns (Three Months Early, Two Months Early, One Month Early, One Month Later, Two Months Later, Three Months Later) shows the percentages in each monthly interval. The last two columns (SIDER Low End and SIDER High End) shows the SIDER reporting.

Clozapine - Gender Differences (%)

ADR	Trust	Cohort	Sub Cohort	Three Months Early	Two Months Early	One Month Early	One Month Later	Two Months Later	Three Months Later	SIDER Low End	SIDER High End
Convulsion	Oxford	Gender	Female (n=172)	2.33	2.33	2.33	6.40	1.16	4.07		
	SIDER	SIDER	SIDER							3.00	
Hypotension	SLAM	Trust	Clozapine (n=1760)	0.51	0.97	0.80	5.00	2.95	2.56		
		Gender	Male (n=1167)	0.43	0.60	0.34	3.86	2.83	2.49		
			Female (n=593)	0.67	1.69	1.69	7.25	3.20	2.70		
	Camden & Islington	Trust	Clozapine (n=561)	0.18	0.53	0.18	3.57	2.32	1.78		
		Gender	Male (n=357)	0.28	0.56	0.28	2.52	1.96	1.12		
			Female (n=204)	0.00	0.49	0.00	5.39	2.94	2.94		
	Oxford	Trust	Clozapine (n=514)	0.58	0.78	0.78	5.64	3.50	2.92		
		Gender	Male (n=342)	0.00	0.29	0.29	4.09	2.34	1.75		
			Female (n=172)	1.74	1.74	1.74	8.72	5.81	5.23		
	SIDER	SIDER	SIDER							9.00	38.00
Enuresis	SLAM	Trust	Clozapine (n=1760)	1.02	0.80	1.25	4.20	3.92	3.24		
		Gender	Male (n=1167)	0.60	0.51	0.86	3.17	2.74	2.74		
			Female (n=593)	1.85	1.35	2.02	6.24	6.24	4.22		
	Camden & Islington	Trust	Clozapine (n=561)	0.71	1.07	1.07	4.10	1.43	1.25		
		Gender	Male (n=357)	0.28	0.84	0.56	3.36	0.84	1.12		
			Female (n=204)	1.47	1.47	1.96	5.39	2.45	1.47		
	Oxford	Trust	Clozapine (n=514)	1.36	0.58	1.36	4.86	4.47	3.50		
		Gender	Male (n=342)	0.88	0.00	1.17	4.39	3.22	2.05		
			Female (n=172)	2.33	1.74	1.74	5.81	6.98	6.40		
	SIDER	SIDER	SIDER								
Fever	SLAM	Trust	Clozapine (n=1760)	1.02	1.14	1.65	6.36	4.43	3.13		
		Gender	Male (n=1167)	0.77	0.86	1.20	5.14	3.68	3.00		
			Female (n=593)	1.52	1.69	2.53	8.77	5.90	3.37		
	Camden & Islington	Trust	Clozapine (n=561)	0.89	0.89	0.53	3.74	2.67	0.89		
		Gender	Male (n=357)	0.56	0.84	0.56	3.64	1.68	0.84		
			Female (n=204)	1.47	0.98	0.49	3.92	4.41	0.98		
	Oxford	Trust	Clozapine (n=514)	0.39	0.78	0.58	3.11	2.72	2.33		
		Gender	Male (n=342)	0.58	0.88	0.58	3.22	2.05	2.05		
			Female (n=172)	0.00	0.58	0.58	2.91	4.07	2.91		
	SIDER	SIDER	SIDER							4.00	13.00
Diarrhoea	SLAM	Trust	Clozapine (n=1760)	1.08	1.31	1.36	4.72	3.58	2.56		
		Gender	Male (n=1167)	0.86	1.37	0.69	3.86	3.43	1.80		
			Female (n=593)	1.52	1.18	2.70	6.41	3.88	4.05		
	Camden & Islington	Trust	Clozapine (n=561)	0.71	1.25	0.18	3.03	3.39	3.03		
		Gender	Male (n=357)	0.28	0.84	0.28	2.80	3.08	2.24		
			Female (n=204)	1.47	1.96	0.00	3.43	3.92	4.41		
	Oxford	Trust	Clozapine (n=514)	1.17	0.78	1.36	4.09	3.70	2.53		
		Gender	Male (n=342)	1.17	0.29	1.17	2.34	2.63	1.17		
			Female (n=172)	1.16	1.74	1.74	7.56	5.81	5.23		
	SIDER	SIDER	SIDER							2.00	
Drymouth	SLAM	Trust	Clozapine (n=1760)	1.08	1.53	1.65	4.66	3.69	2.33		
		Gender	Male (n=1167)	0.86	1.20	1.46	4.20	3.00	2.23		
			Female (n=593)	1.52	2.19	2.02	5.56	5.06	2.53		
	Camden & Islington	Trust	Clozapine (n=561)	1.25	1.25	1.07	3.92	2.14	0.89		
		Gender	Male (n=357)	0.84	1.40	1.40	4.48	1.68	0.28		
			Female (n=204)	1.96	0.98	0.49	2.94	2.94	1.96		
	Oxford	Trust	Clozapine (n=514)	1.36	1.36	1.56	3.89	1.36	2.33		
		Gender	Male (n=342)	1.17	0.88	1.46	4.39	1.46	1.46		
			Female (n=172)	1.74	2.91	1.74	2.91	1.16	4.07		
	SIDER	SIDER	SIDER							5.00	20.00
Rash	SLAM	Trust	Clozapine (n=1760)	1.25	1.59	2.05	3.64	2.95	2.27		
		Gender	Male (n=1167)	0.86	1.29	1.46	3.17	2.57	1.89		
			Female (n=593)	2.02	2.19	3.20	4.55	3.71	3.04		
	Camden & Islington	Trust	Clozapine (n=561)	1.25	1.25	0.89	4.28	1.96	2.14		
		Gender	Male (n=357)	0.56	1.12	1.12	4.76	1.68	2.24		
			Female (n=204)	2.45	1.47	0.49	3.43	2.45	1.96		
	Oxford	Trust	Clozapine (n=514)	0.97	1.17	1.17	3.70	2.33	1.36		
		Gender	Male (n=342)	1.17	0.88	1.17	2.92	1.75	0.88		
			Female (n=172)	0.58	1.74	1.16	5.23	3.49	2.33		
	SIDER	SIDER	SIDER								
Dyspepsia	SLAM	Trust	Clozapine (n=1760)	0.74	1.08	0.91	3.92	3.13	3.69		
		Gender	Male (n=1167)	0.51	0.77	0.60	3.86	2.74	3.43		
			Female (n=593)	1.18	1.69	1.52	4.05	3.88	4.22		
	Camden &	Trust	Clozapine (n=561)	0.36	0.53	0.53	4.10	2.67	2.50		

Measure Values: 0.00 – 56.00

The results are shown in percentages (%) and broken down by ADRs, Trusts (SLAM, Camden & Islington and Oxford), Cohorts, Sub Cohorts and SIDER reported values.
In Sub Cohort 'Clozapine' represent the total baseline population which further breaks down into Gender 'Male' and 'Female' groups.
The columns (Three Months Early, Two Months Early, One Month Early, One Month Later, Two Months Later, Three Months Later) shows the percentages in each monthly interval. The last two columns (SIDER Low End and SIDER High End) shows the SIDER reporting.

Clozapine - Gender Differences (%)

ADR	Trust	Cohort	Sub Cohort	Three Months Early	Two Months Early	One Month Early	One Month Later	Two Months Later	Three Months Later	SIDER Low End	SIDER High End
Dyspepsia	Camden & Islington	Gender	Male (n=357)	0.00	0.84	0.56	5.04	3.08	3.36		
			Female (n=204)	0.98	0.00	0.49	2.45	1.96	0.98		
	Oxford	Trust	Clozapine (n=514)	0.19	0.58	0.78	3.50	4.09	3.70		
		Gender	Male (n=342)	0.29	0.58	0.58	2.34	3.22	2.92		
			Female (n=172)	0.00	0.58	1.16	5.81	5.81	5.23		
	SIDER	SIDER	SIDER							8.00	14.00
Stomachpain	SLAM	Trust	Clozapine (n=1760)	1.93	1.76	1.93	4.94	3.52	3.52		
		Gender	Male (n=1167)	1.03	1.37	1.20	3.51	2.91	2.14		
			Female (n=593)	3.71	2.53	3.37	7.76	4.72	6.24		
	Camden & Islington	Trust	Clozapine (n=561)	0.89	1.25	0.89	3.39	2.85	2.14		
		Gender	Male (n=357)	0.84	0.84	0.56	2.80	1.96	1.68		
			Female (n=204)	0.98	1.96	1.47	4.41	4.41	2.94		
	Oxford	Trust	Clozapine (n=514)	1.56	0.78	0.78	2.14	0.97	0.97		
		Gender	Male (n=342)	1.46	0.88	0.29	2.34	0.88	0.58		
			Female (n=172)	1.74	0.58	1.74	1.74	1.16	1.74		
	SIDER	SIDER	SIDER								
Sweating	SLAM	Trust	Clozapine (n=1760)	1.08	0.97	1.36	4.43	4.26	2.84		
		Gender	Male (n=1167)	1.20	1.11	1.63	4.37	4.97	3.00		
			Female (n=593)	0.84	0.67	0.84	4.55	2.87	2.53		
	Camden & Islington	Trust	Clozapine (n=561)	0.53	0.53	0.53	2.85	2.14	1.96		
		Gender	Male (n=357)	0.28	0.28	0.56	3.64	2.52	1.40		
			Female (n=204)	0.98	0.98	0.49	1.47	1.47	2.94		
	Oxford	Trust	Clozapine (n=514)	1.17	0.97	0.97	2.72	1.36	1.95		
		Gender	Male (n=342)	1.75	1.17	0.88	2.92	1.17	1.46		
			Female (n=172)	0.00	0.58	1.16	2.33	1.74	2.91		
	SIDER	SIDER	SIDER							6.00	
Tremor	SLAM	Trust	Clozapine (n=1760)	1.48	1.99	2.95	5.51	3.52	3.47		
		Gender	Male (n=1167)	1.03	2.14	3.26	5.74	3.43	3.51		
			Female (n=593)	2.36	1.69	2.36	5.06	3.71	3.37		
	Camden & Islington	Trust	Clozapine (n=561)	1.60	1.78	2.14	3.92	1.96	2.14		
		Gender	Male (n=357)	1.68	2.24	2.52	3.92	1.96	2.52		
			Female (n=204)	1.47	0.98	1.47	3.92	1.96	1.47		
	SIDER	SIDER	SIDER							6.00	
Neutropenia	SLAM	Trust	Clozapine (n=1760)	0.80	0.80	0.74	5.34	2.73	2.61		
		Gender	Male (n=1167)	0.94	0.51	0.69	5.91	3.00	2.49		
			Female (n=593)	0.51	1.35	0.84	4.22	2.19	2.87		
	Camden & Islington	Trust	Clozapine (n=561)	0.00	0.18	0.53	1.60	0.89	1.07		
		Gender	Male (n=357)	0.00	0.00	0.00	1.12	0.84	0.84		
			Female (n=204)	0.00	0.49	1.47	2.45	0.98	1.47		
	SIDER	SIDER	SIDER								
Akathisia	SLAM	Trust	Clozapine (n=1760)	0.80	0.91	0.74	2.67	1.36	0.80		
		Gender	Male (n=1167)	0.77	0.86	0.69	2.74	1.03	0.77		
			Female (n=593)	0.84	1.01	0.84	2.53	2.02	0.84		
	Camden & Islington	Trust	Clozapine (n=561)	0.00	0.53	0.00	1.25	1.07	0.53		
		Gender	Male (n=357)	0.00	0.56	0.00	1.40	1.12	0.56		
			Female (n=204)	0.00	0.49	0.00	0.98	0.98	0.49		
	Oxford	Trust	Clozapine (n=514)	0.97	0.78	0.97	1.36	1.17	0.97		
		Gender	Male (n=342)	1.46	0.58	0.88	1.17	1.17	0.88		
			Female (n=172)	0.00	1.16	1.16	1.74	1.16	1.16		
	SIDER	SIDER	SIDER							3.00	
Blurredvision	SLAM	Trust	Clozapine (n=1760)	0.34	0.91	0.63	2.05	1.25	1.02		
		Gender	Male (n=1167)	0.43	0.94	0.51	1.37	1.20	0.94		
			Female (n=593)	0.17	0.84	0.84	3.37	1.35	1.18		
	Camden & Islington	Trust	Clozapine (n=561)	0.89	0.53	0.71	1.25	0.36	0.89		
		Gender	Male (n=357)	0.84	0.28	0.56	1.12	0.28	0.84		
			Female (n=204)	0.98	0.98	0.98	1.47	0.49	0.98		
	Oxford	Trust	Clozapine (n=514)	0.19	0.39	0.39	1.56	1.56	1.17		
		Gender	Male (n=342)	0.29	0.29	0.29	1.46	1.75	0.88		
			Female (n=172)	0.00	0.58	0.58	1.74	1.16	1.74		
	SIDER	SIDER	SIDER							5.00	

Measure Values: 0.00 – 56.00

The results are shown in percentages (%) and broken down by ADRs, Trusts (SLAM, Camden & Islington and Oxford), Cohorts, Sub Cohorts and SIDER reported values.
In Sub Cohort 'Clozapine' represent the total baseline population which further breaks down into Gender 'Male' and 'Female' groups.
The columns (Three Months Early, Two Months Early, One Month Early, One Month Later, Two Months Later, Three Months Later) shows the percentages in each monthly interval. The last two columns (SIDER Low End and SIDER High End) shows the SIDER reporting.

Supplementary Table B.2 Ethnic background (%)

Clozapine - Ethnic Background (%)

ADR	Trust	Cohort	Sub Cohort	Three Months Early	Two Months Early	One Month Early	One Month Later	Two Months Later	Three Months Later	SIDER Low End	SIDER High End
Agitation	SLAM	Trust	Clozapine (n=1760)	17.61	22.10	26.53	46.59	32.56	26.99		
		Ethnic Background	White (n=821)	16.32	21.80	23.26	38.98	26.31	22.53		
			Black (n=704)	19.03	25.43	27.13	45.45	30.68	26.28		
			Asian (n=93)	18.28	19.35	24.73	45.16	25.81	23.66		
			Other (n=142)	22.54	21.13	30.99	45.77	37.32	30.28		
	Camden & Islington	Trust	Clozapine (n=561)	13.37	17.83	18.36	43.14	28.34	21.03		
		Ethnic Background	White (n=347)	12.10	16.43	18.44	40.63	26.51	19.02		
			Black (n=120)	15.83	25.83	22.50	49.17	32.50	21.67		
			Asian (n=41)	17.07	9.76	14.63	51.22	29.27	31.71		
			Others (n=53)	13.21	11.32	11.32	28.30	22.64	22.64		
	Oxford	Trust	Clozapine (n=514)	14.59	15.76	16.34	34.24	25.10	20.62		
		Ethnic Background	White (n=426)	14.55	15.73	16.20	34.74	25.82	21.36		
			Black (n=20)	10.00	10.00	10.00	35.00	25.00	15.00		
			Asian (n=41)	12.20	17.07	19.51	36.59	21.95	19.51		
			Others (n=27)	22.22	18.52	18.52	22.22	18.52	14.81		
	SIDER	SIDER	SIDER							4.00	
Fatigue	SLAM	Trust	Clozapine (n=1760)	12.67	14.83	15.85	43.58	35.80	30.51		
		Ethnic Background	White (n=821)	13.03	13.89	14.62	35.69	29.72	24.24		
			Black (n=704)	15.20	16.19	17.05	41.62	34.66	28.27		
			Asian (n=93)	15.05	11.83	16.13	45.16	30.11	24.73		
			Other (n=142)	13.38	12.68	12.68	43.66	34.51	31.69		
	Camden & Islington	Trust	Clozapine (n=561)	10.34	12.30	13.37	41.18	29.23	26.56		
		Ethnic Background	White (n=347)	8.65	13.26	11.24	39.77	30.26	24.21		
			Black (n=120)	15.00	11.67	18.33	41.67	30.00	32.50		
			Asian (n=41)	12.20	7.32	14.63	51.22	29.27	36.59		
			Others (n=53)	9.43	5.66	11.32	30.19	24.53	20.75		
	Oxford	Trust	Clozapine (n=514)	9.73	11.87	12.06	35.21	27.43	26.85		
		Ethnic Background	White (n=426)	9.62	12.21	12.44	35.45	26.53	26.53		
			Black (n=20)	5.00	10.00	10.00	30.00	25.00	30.00		
			Asian (n=41)	4.88	4.88	4.88	29.27	24.39	24.39		
			Others (n=27)	22.22	18.52	18.52	44.44	48.15	33.33		
	SIDER	SIDER	SIDER								
Sedation	SLAM	Trust	Clozapine (n=1760)	12.67	12.16	14.83	43.86	35.51	29.83		
		Ethnic Background	White (n=821)	12.91	11.57	14.25	38.49	30.33	23.63		
			Black (n=704)	15.06	13.49	16.62	44.89	35.37	27.56		
			Asian (n=93)	17.20	12.90	11.83	46.24	35.48	23.66		
			Other (n=142)	16.20	14.08	14.79	42.25	38.73	28.87		
	Camden & Islington	Trust	Clozapine (n=561)	5.17	9.09	9.09	38.15	26.56	21.93		
		Ethnic Background	White (n=347)	3.17	9.22	7.20	37.18	25.65	21.04		
			Black (n=120)	10.83	11.67	15.83	39.17	29.17	25.83		
			Asian (n=41)	7.32	4.88	9.76	46.34	31.71	24.39		
			Others (n=53)	5.66	5.66	7.55	24.53	18.87	16.98		
	Oxford	Trust	Clozapine (n=514)	7.20	8.37	9.34	31.52	21.40	18.48		
		Ethnic Background	White (n=426)	7.04	7.75	8.45	30.75	20.66	18.54		
			Black (n=20)	10.00	5.00	15.00	30.00	30.00	30.00		
			Asian (n=41)	2.44	12.20	12.20	31.71	14.63	14.63		
			Others (n=27)	14.81	14.81	14.81	44.44	37.04	14.81		
	SIDER	SIDER	SIDER							25.00	46.00
Dizziness	SLAM	Trust	Clozapine (n=1760)	2.78	4.20	4.09	16.59	13.13	11.19		
		Ethnic Background	White (n=821)	3.29	4.51	3.53	12.67	8.40	6.94		
			Black (n=704)	3.84	5.26	4.12	14.77	9.80	8.10		
			Asian (n=93)	4.30	1.08	3.23	15.05	17.20	7.53		
			Other (n=142)	0.70	3.52	2.11	18.31	14.79	17.61		
	Camden & Islington	Trust	Clozapine (n=561)	3.21	3.39	3.74	18.18	13.73	9.09		
		Ethnic Background	White (n=347)	4.03	3.46	2.88	18.73	14.41	9.22		
			Black (n=120)	2.50	2.50	5.83	19.17	13.33	10.00		
			Asian (n=41)	0.00	7.32	4.88	19.51	17.07	12.20		
			Others (n=53)	3.77	3.77	5.66	11.32	5.66	3.77		
	Oxford	Trust	Clozapine (n=514)	3.89	4.09	4.47	17.70	13.04	10.12		
		Ethnic Background	White (n=426)	3.76	3.99	4.46	18.31	12.91	10.33		
			Black (n=20)	5.00	0.00	5.00	15.00	5.00	10.00		
			Asian (n=41)	2.44	7.32	4.88	9.76	12.20	9.76		
			Others (n=27)	7.41	3.70	3.70	22.22	22.22	7.41		
	SIDER	SIDER	SIDER							12.00	27.00
Tachycardia	SLAM	Trust	Clozapine (n=1760)	2.27	2.05	2.50	15.40	12.95	9.94		
		Ethnic Background	White (n=821)	1.95	2.68	2.44	13.03	11.21	7.80		
			Black (n=704)	2.27	3.13	2.84	15.20	13.07	9.09		

Measure Values: 0.00 – 56.00

The results are shown in percentages (%) and broken down by ADRs, Trusts (SLAM, Camden & Islington and Oxford), Cohorts, Sub Cohorts and SIDER reported values.
In Sub Cohort 'Clozapine' represent the total baseline population which further breaks down into 'White', 'Black', 'Asian' and 'Others' ethnic groups.
The columns (Three Months Early, Two Months Early, One Month Early, One Month Later, Two Months Later, Three Months Later) shows the percentages in each monthly interval. The last two columns (SIDER Low End and SIDER High End) shows the SIDER reporting.

Clozapine - Ethnic Background (%)

ADR	Trust	Cohort	Sub Cohort	Three Months Early	Two Months Early	One Month Early	One Month Later	Two Months Later	Three Months Later	SIDER Low End	SIDER High End
Tachycardia	SLAM	Ethnic Background	Asian (n=93)	0.00	2.15	3.23	18.28	12.90	5.38		
			Other (n=142)	2.11	2.11	1.41	16.20	16.90	12.68		
	Camden & Islington	Trust	Clozapine (n=561)	1.43	1.43	0.89	11.23	8.38	6.95		
		Ethnic Background	White (n=347)	1.44	0.29	1.15	9.22	8.36	5.19		
			Black (n=120)	1.67	4.17	0.83	15.00	8.33	10.00		
			Asian (n=41)	0.00	2.44	0.00	24.39	9.76	7.32		
			Others (n=53)	3.77	0.00	0.00	7.55	5.66	13.21		
	Oxford	Trust	Clozapine (n=514)	0.78	1.36	1.56	10.89	10.51	7.59		
		Ethnic Background	White (n=426)	0.70	1.41	1.64	10.80	9.39	7.28		
			Black (n=20)	0.00	0.00	0.00	20.00	20.00	5.00		
			Asian (n=41)	2.44	2.44	2.44	9.76	9.76	4.88		
			Others (n=27)	0.00	0.00	0.00	7.41	22.22	18.52		
	SIDER	SIDER	SIDER							11.00	25.00
Weightgain	SLAM	Trust	Clozapine (n=1760)	3.75	4.43	5.06	15.34	10.91	10.34		
		Ethnic Background	White (n=821)	4.14	4.26	4.99	13.40	9.87	9.14		
			Black (n=704)	4.83	4.97	5.82	15.63	11.51	10.65		
			Asian (n=93)	1.08	5.38	6.45	19.35	9.68	9.68		
			Other (n=142)	4.93	6.34	4.23	19.01	14.08	12.68		
	Camden & Islington	Trust	Clozapine (n=561)	2.50	3.39	1.96	11.76	6.60	6.24		
		Ethnic Background	White (n=347)	2.88	3.75	2.02	13.26	6.63	5.48		
			Black (n=120)	1.67	2.50	3.33	10.83	7.50	9.17		
			Asian (n=41)	4.88	4.88	0.00	9.76	7.32	9.76		
			Others (n=53)	0.00	3.77	0.00	9.43	7.55	7.55		
	Oxford	Trust	Clozapine (n=514)	3.50	3.31	3.70	11.28	9.92	7.78		
		Ethnic Background	White (n=426)	3.99	3.29	3.52	11.03	10.09	8.45		
			Black (n=20)	0.00	0.00	0.00	20.00	5.00	5.00		
			Asian (n=41)	0.00	0.00	0.00	7.32	4.88	0.00		
			Others (n=27)	3.70	11.11	14.81	14.81	18.52	11.11		
	SIDER	SIDER	SIDER							4.00	56.00
Hypersalivation	SLAM	Trust	Clozapine (n=1760)	1.19	1.48	2.10	14.32	13.24	11.31		
		Ethnic Background	White (n=821)	1.10	1.34	1.34	12.79	11.57	9.99		
			Black (n=704)	1.28	1.56	1.56	14.91	13.49	11.65		
			Asian (n=93)	0.00	2.15	4.30	15.05	12.90	8.60		
			Other (n=142)	2.11	1.41	4.23	16.20	16.20	11.97		
	Camden & Islington	Trust	Clozapine (n=561)	1.07	1.43	0.53	14.26	6.95	7.66		
		Ethnic Background	White (n=347)	0.58	1.44	0.58	15.56	7.49	8.65		
			Black (n=120)	0.83	2.50	0.00	10.83	6.67	7.50		
			Asian (n=41)	0.00	0.00	2.44	17.07	7.32	0.00		
			Others (n=53)	5.66	0.00	0.00	7.55	1.89	7.55		
	Oxford	Trust	Clozapine (n=514)	0.97	0.78	1.56	12.65	10.70	5.84		
		Ethnic Background	White (n=426)	0.23	0.94	1.64	13.38	11.27	6.10		
			Black (n=20)	10.00	0.00	0.00	10.00	5.00	0.00		
			Asian (n=41)	2.44	0.00	0.00	7.32	9.76	7.32		
			Others (n=27)	3.70	0.00	3.70	11.11	7.41	3.70		
	SIDER	SIDER	SIDER							1.00	48.00
Feelingsick	SLAM	Trust	Clozapine (n=1760)	4.66	4.94	6.48	14.32	11.19	9.09		
		Ethnic Background	White (n=821)	4.51	4.38	5.72	12.67	9.38	6.46		
			Black (n=704)	5.26	5.11	6.68	14.77	10.94	7.53		
			Asian (n=93)	8.60	5.38	5.38	10.75	8.60	6.45		
			Other (n=142)	5.63	7.04	8.45	14.79	12.68	11.27		
	Camden & Islington	Trust	Clozapine (n=561)	3.74	3.92	3.03	10.52	7.13	7.66		
		Ethnic Background	White (n=347)	3.17	3.75	3.75	11.53	8.65	7.20		
			Black (n=120)	5.00	6.67	2.50	10.00	3.33	10.00		
			Asian (n=41)	2.44	0.00	2.44	4.88	2.44	2.44		
			Others (n=53)	5.66	3.77	1.89	13.21	5.66	7.55		
	Oxford	Trust	Clozapine (n=514)	3.89	5.25	5.06	14.20	9.73	7.20		
		Ethnic Background	White (n=426)	3.76	4.93	5.16	14.79	9.39	7.75		
			Black (n=20)	0.00	0.00	0.00	10.00	5.00	0.00		
			Asian (n=41)	4.88	4.88	7.32	12.20	7.32	4.88		
			Others (n=27)	7.41	14.81	3.70	11.11	22.22	7.41		
	SIDER	SIDER	SIDER								
Confusion	SLAM	Trust	Clozapine (n=1760)	4.72	5.51	6.08	13.92	8.47	6.76		
		Ethnic Background	White (n=821)	5.36	5.85	5.85	12.55	6.58	5.48		
			Black (n=704)	6.25	6.82	6.82	14.63	7.67	6.39		
			Asian (n=93)	2.15	0.00	4.30	10.75	10.75	3.23		
			Other (n=142)	7.04	3.52	4.93	11.27	9.86	7.75		
	Camden &	Trust	Clozapine (n=561)	3.57	6.24	5.53	12.66	6.77	5.88		

Measure Values: 0.00 – 56.00

The results are shown in percentages (%) and broken down by ADRs, Trusts (SLAM, Camden & Islington and Oxford), Cohorts, Sub Cohorts and SIDER reported values.
In Sub Cohort 'Clozapine' represent the total baseline population which further breaks down into 'White', 'Black', 'Asian' and 'Others' ethnic groups. The columns (Three Months Early, Two Months Early, One Month Early, One Month Later, Two Months Later, Three Months Later) shows the percentages in each monthly interval. The last two columns (SIDER Low End and SIDER High End) shows the SIDER reporting.

Clozapine - Ethnic Background (%)

ADR	Trust	Cohort	Sub Cohort	Three Months Early	Two Months Early	One Month Early	One Month Later	Two Months Later	Three Months Later	SIDER Low End	SIDER High End
Confusion	Camden & Islington	Ethnic Background	White (n=347)	3.75	5.19	6.05	10.95	6.34	4.90		
			Black (n=120)	4.17	8.33	7.50	19.17	11.67	9.17		
			Asian (n=41)	2.44	9.76	0.00	14.63	2.44	4.88		
			Others (n=53)	1.89	5.66	3.77	9.43	1.89	5.66		
	Oxford	Trust	Clozapine (n=514)	2.53	3.89	3.89	9.92	6.42	5.25		
		Ethnic Background	White (n=426)	2.35	3.99	4.23	10.80	5.87	5.16		
			Black (n=20)	0.00	0.00	0.00	10.00	10.00	10.00		
			Asian (n=41)	0.00	2.44	0.00	0.00	9.76	4.88		
			Others (n=27)	11.11	7.41	7.41	11.11	7.41	3.70		
	SIDER	SIDER	SIDER							3.00	
Constipation	SLAM	Trust	Clozapine (n=1760)	1.76	1.99	2.16	12.27	11.70	9.49		
		Ethnic Background	White (n=821)	1.46	1.83	2.31	9.14	9.14	8.04		
			Black (n=704)	1.70	2.13	2.70	10.65	10.65	9.38		
			Asian (n=93)	1.08	1.08	0.00	20.43	17.20	8.60		
			Other (n=142)	4.93	2.82	2.11	14.79	14.79	9.86		
	Camden & Islington	Trust	Clozapine (n=561)	1.07	2.50	1.78	11.41	7.13	5.70		
		Ethnic Background	White (n=347)	1.44	2.59	2.31	12.97	8.07	6.05		
			Black (n=120)	0.00	3.33	0.83	9.17	8.33	6.67		
			Asian (n=41)	0.00	0.00	0.00	14.63	2.44	2.44		
			Others (n=53)	1.89	1.89	1.89	3.77	3.77	3.77		
	Oxford	Trust	Clozapine (n=514)	0.58	0.97	1.36	10.31	7.78	7.78		
		Ethnic Background	White (n=426)	0.70	0.94	1.17	10.09	8.22	7.51		
			Black (n=20)	0.00	0.00	5.00	5.00	5.00	15.00		
			Asian (n=41)	0.00	2.44	2.44	12.20	7.32	9.76		
			Others (n=27)	0.00	0.00	0.00	14.81	3.70	3.70		
	SIDER	SIDER	SIDER							10.00	25.00
Headache	SLAM	Trust	Clozapine (n=1760)	4.20	4.55	5.45	12.44	8.18	5.91		
		Ethnic Background	White (n=821)	4.38	5.24	5.36	10.35	7.80	4.51		
			Black (n=704)	5.11	6.11	6.25	12.07	9.09	5.26		
			Asian (n=93)	2.15	3.23	4.30	13.98	6.45	7.53		
			Other (n=142)	5.63	8.45	5.63	14.08	10.56	7.04		
	Camden & Islington	Trust	Clozapine (n=561)	2.32	3.57	4.28	9.27	6.42	4.63		
		Ethnic Background	White (n=347)	2.31	3.17	3.46	8.65	5.76	4.03		
			Black (n=120)	2.50	5.00	6.67	10.00	6.67	6.67		
			Asian (n=41)	4.88	4.88	2.44	12.20	12.20	4.88		
			Others (n=53)	0.00	3.77	5.66	5.66	1.89	1.89		
	Oxford	Trust	Clozapine (n=514)	3.89	3.89	4.09	10.89	8.37	7.59		
		Ethnic Background	White (n=426)	3.99	3.76	3.99	10.33	8.92	7.75		
			Black (n=20)	5.00	5.00	5.00	5.00	10.00	0.00		
			Asian (n=41)	4.88	7.32	7.32	19.51	4.88	14.63		
			Others (n=27)	0.00	0.00	0.00	11.11	3.70	0.00		
	SIDER	SIDER	SIDER								
Hyperprolactinaemia	SLAM	Trust	Clozapine (n=1760)	3.18	3.64	4.20	8.52	5.06	4.15		
		Ethnic Background	White (n=821)	3.41	4.14	5.12	8.40	5.12	4.02		
			Black (n=704)	3.98	4.83	5.97	9.80	5.97	4.69		
			Asian (n=93)	4.30	3.23	4.30	12.90	3.23	2.15		
			Other (n=142)	3.52	4.93	3.52	8.45	4.23	3.52		
	Camden & Islington	Trust	Clozapine (n=561)	1.60	1.78	2.67	8.20	4.10	3.57		
		Ethnic Background	White (n=347)	0.29	1.44	1.73	6.05	3.46	3.46		
			Black (n=120)	5.00	4.17	5.00	14.17	4.17	2.50		
			Asian (n=41)	2.44	0.00	4.88	12.20	2.44	2.44		
			Others (n=53)	1.89	0.00	1.89	5.66	5.66	5.66		
	Oxford	Trust	Clozapine (n=514)	3.70	4.09	4.28	8.75	4.47	4.86		
		Ethnic Background	White (n=426)	3.52	3.76	3.99	8.45	4.46	3.99		
			Black (n=20)	5.00	0.00	5.00	10.00	5.00	5.00		
			Asian (n=41)	0.00	2.44	2.44	7.32	2.44	7.32		
			Others (n=27)	11.11	14.81	11.11	14.81	7.41	14.81		
	SIDER	SIDER	SIDER								
Insomnia	SLAM	Trust	Clozapine (n=1760)	3.92	4.03	5.17	10.40	6.48	4.03		
		Ethnic Background	White (n=821)	3.29	3.90	5.36	7.80	5.24	3.05		
			Black (n=704)	3.84	4.55	6.25	9.09	6.11	3.55		
			Asian (n=93)	7.53	3.23	9.68	11.83	3.23	4.30		
			Other (n=142)	7.75	4.93	5.63	10.56	7.75	4.93		
	Camden & Islington	Trust	Clozapine (n=561)	3.57	3.39	3.74	8.91	3.39	4.28		
		Ethnic Background	White (n=347)	3.75	3.46	4.03	8.93	3.75	4.03		
			Black (n=120)	4.17	4.17	2.50	8.33	2.50	5.00		
			Asian (n=41)	2.44	2.44	4.88	12.20	4.88	4.88		

Measure Values: 0.00 – 56.00

The results are shown in percentages (%) and broken down by ADRs, Trusts (SLAM, Camden & Islington and Oxford), Cohorts, Sub Cohorts and SIDER reported values.
In Sub Cohort 'Clozapine' represent the total baseline population which further breaks down into 'White', 'Black', 'Asian' and 'Others' ethnic groups. The columns (Three Months Early, Two Months Early, One Month Early, One Month Later, Two Months Later, Three Months Later) shows the percentages in each monthly interval. The last two columns (SIDER Low End and SIDER High End) shows the SIDER reporting.

Clozapine - Ethnic Background (%)

ADR	Trust	Cohort	Sub Cohort	Three Months Early	Two Months Early	One Month Early	One Month Later	Two Months Later	Three Months Later	SIDER Low End	SIDER High End
Insomnia	Islington	Background	Others (n=53)	1.89	0.00	3.77	9.43	3.77	3.77		
	Oxford	Trust	Clozapine (n=514)	5.84	4.86	5.84	8.37	6.81	4.09		
		Ethnic Background	White (n=426)	5.87	4.69	5.63	8.92	6.57	4.69		
			Black (n=20)	10.00	0.00	0.00	5.00	5.00	5.00		
			Asian (n=41)	4.88	7.32	7.32	7.32	12.20	0.00		
			Others (n=27)	3.70	7.41	11.11	3.70	3.70	0.00		
	SIDER	SIDER	SIDER							20.00	33.00
Hypertension	SLAM	Trust	Clozapine (n=1760)	2.05	2.22	3.13	9.15	5.74	4.60		
		Ethnic Background	White (n=821)	2.68	3.17	4.02	10.48	7.19	5.60		
			Black (n=704)	3.13	3.69	4.69	12.22	8.38	6.53		
			Asian (n=93)	1.08	2.15	2.15	8.60	5.38	2.15		
			Other (n=142)	1.41	1.41	2.11	11.27	4.93	4.23		
	Camden & Islington	Trust	Clozapine (n=561)	0.71	0.71	1.60	7.13	4.63	2.67		
		Ethnic Background	White (n=347)	0.58	0.86	2.02	8.07	5.48	2.02		
			Black (n=120)	0.83	0.83	0.83	7.50	3.33	5.00		
			Asian (n=41)	0.00	0.00	0.00	2.44	2.44	2.44		
			Others (n=53)	1.89	0.00	1.89	3.77	1.89	1.89		
	Oxford	Trust	Clozapine (n=514)	1.36	1.36	1.56	5.06	4.28	2.14		
		Ethnic Background	White (n=426)	1.17	1.41	1.41	3.99	3.99	1.88		
			Black (n=20)	0.00	0.00	0.00	0.00	10.00	0.00		
			Asian (n=41)	0.00	0.00	0.00	7.32	2.44	2.44		
			Others (n=27)	7.41	3.70	7.41	22.22	7.41	7.41		
	SIDER	SIDER	SIDER							4.00	12.00
Vomiting	SLAM	Trust	Clozapine (n=1760)	2.56	2.50	3.01	8.86	6.82	5.00		
		Ethnic Background	White (n=821)	2.68	1.95	2.92	7.80	5.60	3.53		
			Black (n=704)	3.13	2.27	3.41	9.09	6.53	4.12		
			Asian (n=93)	4.30	3.23	3.23	5.38	8.60	5.38		
			Other (n=142)	1.41	4.93	4.23	6.34	7.75	4.93		
	Camden & Islington	Trust	Clozapine (n=561)	2.14	2.50	2.85	6.77	4.99	4.63		
		Ethnic Background	White (n=347)	1.73	2.88	2.59	6.63	4.03	3.75		
			Black (n=120)	2.50	2.50	4.17	10.00	8.33	7.50		
			Asian (n=41)	2.44	2.44	2.44	2.44	2.44	0.00		
			Others (n=53)	3.77	0.00	1.89	3.77	5.66	5.66		
	Oxford	Trust	Clozapine (n=514)	1.75	2.72	2.92	7.59	5.25	5.06		
		Ethnic Background	White (n=426)	1.88	2.35	2.58	7.51	5.16	5.87		
			Black (n=20)	0.00	5.00	5.00	5.00	5.00	0.00		
			Asian (n=41)	2.44	4.88	2.44	7.32	7.32	2.44		
			Others (n=27)	0.00	3.70	7.41	11.11	3.70	0.00		
	SIDER	SIDER	SIDER							3.00	17.00
Shaking	SLAM	Trust	Clozapine (n=1760)	3.13	2.95	3.92	9.55	5.40	5.06		
		Ethnic Background	White (n=821)	3.17	2.80	3.29	7.67	4.26	3.41		
			Black (n=704)	3.69	3.27	3.84	8.95	4.97	3.98		
			Asian (n=93)	2.15	2.15	3.23	15.05	9.68	9.68		
			Other (n=142)	4.23	3.52	4.23	7.04	7.75	3.52		
	Camden & Islington	Trust	Clozapine (n=561)	1.78	1.96	3.74	6.06	3.92	2.85		
		Ethnic Background	White (n=347)	1.73	1.15	2.59	4.61	3.17	1.73		
			Black (n=120)	1.67	2.50	5.83	9.17	4.17	4.17		
			Asian (n=41)	0.00	2.44	2.44	2.44	2.44	2.44		
			Others (n=53)	1.89	1.89	5.66	5.66	7.55	7.55		
	Oxford	Trust	Clozapine (n=514)	2.92	3.31	3.89	7.78	4.47	4.86		
		Ethnic Background	White (n=426)	3.29	3.76	4.46	8.45	5.16	4.93		
			Black (n=20)	0.00	0.00	0.00	0.00	0.00	0.00		
			Asian (n=41)	0.00	0.00	0.00	4.88	2.44	2.44		
			Others (n=27)	3.70	3.70	3.70	7.41	0.00	11.11		
	SIDER	SIDER	SIDER								
Abdominalpain	SLAM	Trust	Clozapine (n=1760)	1.88	1.99	2.56	8.01	6.02	4.72		
		Ethnic Background	White (n=821)	1.83	1.95	2.19	6.94	4.38	3.90		
			Black (n=704)	2.13	2.27	2.56	8.10	5.11	4.55		
			Asian (n=93)	1.08	1.08	1.08	8.60	9.68	4.30		
			Other (n=142)	4.23	2.82	2.82	8.45	9.15	5.63		
	Camden & Islington	Trust	Clozapine (n=561)	0.89	0.89	1.60	3.92	3.57	3.39		
		Ethnic Background	White (n=347)	0.58	0.86	0.86	3.46	3.17	2.88		
			Black (n=120)	0.83	0.83	4.17	4.17	4.17	4.17		
			Asian (n=41)	0.00	0.00	0.00	4.88	4.88	4.88		
			Others (n=53)	1.89	1.89	1.89	5.66	5.66	3.77		
	Oxford	Trust	Clozapine (n=514)	1.75	1.36	1.75	7.39	4.47	5.64		
		Ethnic	White (n=426)	1.17	1.41	1.64	7.51	4.93	6.10		

Measure Values: 0.00 – 56.00

The results are shown in percentages (%) and broken down by ADRs, Trusts (SLAM, Camden & Islington and Oxford), Cohorts, Sub Cohorts and SIDER reported values.
In Sub Cohort 'Clozapine' represent the total baseline population which further breaks down into 'White', 'Black', 'Asian' and 'Others' ethnic groups. The columns (Three Months Early, Two Months Early, One Month Early, One Month Later, Two Months Later, Three Months Later) shows the percentages in each monthly interval. The last two columns (SIDER Low End and SIDER High End) shows the SIDER reporting.

Clozapine - Ethnic Background (%)

ADR	Trust	Cohort	Sub Cohort	Three Months Early	Two Months Early	One Month Early	One Month Later	Two Months Later	Three Months Later	SIDER Low End	SIDER High End
Abdominalpain	Oxford	Ethnic Background	Black (n=20)	5.00	0.00	0.00	5.00	5.00	0.00		
			Asian (n=41)	7.32	2.44	0.00	2.44	0.00	4.88		
			Others (n=27)	0.00	0.00	7.41	14.81	3.70	3.70		
	SIDER	SIDER	SIDER							4.00	
Convulsion	SLAM	Trust	Clozapine (n=1760)	1.36	1.70	1.82	7.05	4.94	4.03		
		Ethnic Background	White (n=821)	0.49	1.71	0.85	1.83	2.92	2.31		
			Black (n=704)	0.57	1.99	0.99	2.13	3.41	2.70		
			Asian (n=93)	3.23	1.08	3.23	8.60	3.23	2.15		
			Other (n=142)	2.11	1.41	1.41	9.15	3.52	4.23		
	Camden & Islington	Trust	Clozapine (n=561)	0.53	0.53	0.36	2.85	2.14	1.07		
		Ethnic Background	White (n=347)	0.58	0.29	0.00	2.88	2.31	1.15		
			Black (n=120)	0.83	1.67	1.67	3.33	1.67	0.83		
			Asian (n=41)	0.00	0.00	0.00	0.00	2.44	0.00		
			Others (n=53)	1.89	1.89	1.89	5.66	3.77	3.77		
	Oxford	Trust	Clozapine (n=514)	1.36	1.36	1.56	6.42	3.11	2.72		
		Ethnic Background	White (n=426)	1.64	1.41	1.88	6.34	3.05	3.29		
			Black (n=20)	0.00	0.00	0.00	10.00	10.00	0.00		
			Asian (n=41)	0.00	0.00	0.00	4.88	2.44	0.00		
			Others (n=27)	0.00	3.70	0.00	7.41	0.00	0.00		
	SIDER	SIDER	SIDER							3.00	
Hypotension	SLAM	Trust	Clozapine (n=1760)	0.51	0.97	0.80	5.00	2.95	2.56		
		Ethnic Background	White (n=821)	0.49	0.73	0.37	3.78	1.46	1.46		
			Black (n=704)	0.57	0.85	0.43	4.40	1.70	1.70		
			Asian (n=93)	0.00	1.08	2.15	6.45	2.15	3.23		
			Other (n=142)	0.70	2.11	2.11	9.86	4.93	2.11		
	Camden & Islington	Trust	Clozapine (n=561)	0.18	0.53	0.18	3.57	2.32	1.78		
		Ethnic Background	White (n=347)	0.29	0.58	0.29	4.90	2.59	2.59		
			Black (n=120)	0.00	0.83	0.00	0.83	3.33	0.00		
			Asian (n=41)	0.00	0.00	0.00	4.88	0.00	0.00		
			Others (n=53)	0.00	0.00	0.00	0.00	0.00	1.89		
	Oxford	Trust	Clozapine (n=514)	0.58	0.78	0.78	5.64	3.50	2.92		
		Ethnic Background	White (n=426)	0.47	0.94	0.94	5.63	3.29	3.29		
			Black (n=20)	0.00	0.00	0.00	10.00	0.00	0.00		
			Asian (n=41)	0.00	0.00	0.00	0.00	4.88	0.00		
			Others (n=27)	3.70	0.00	0.00	11.11	7.41	3.70		
	SIDER	SIDER	SIDER							9.00	38.00
Nausea	SLAM	Trust	Clozapine (n=1760)	1.14	1.08	1.19	6.08	5.23	3.69		
		Ethnic Background	White (n=821)	1.58	0.97	0.97	3.78	4.14	2.44		
			Black (n=704)	1.85	1.14	1.14	4.40	4.83	2.84		
			Asian (n=93)	2.15	1.08	2.15	8.60	7.53	2.15		
			Other (n=142)	0.70	1.41	0.00	5.63	8.45	3.52		
	Camden & Islington	Trust	Clozapine (n=561)	0.89	1.43	0.36	4.63	3.57	3.57		
		Ethnic Background	White (n=347)	0.86	1.73	0.00	5.19	4.32	2.88		
			Black (n=120)	0.00	1.67	0.00	5.00	0.83	5.00		
			Asian (n=41)	0.00	0.00	2.44	2.44	4.88	4.88		
			Others (n=53)	1.89	1.89	0.00	3.77	3.77	1.89		
	Oxford	Trust	Clozapine (n=514)	0.97	0.58	1.36	4.86	3.70	2.92		
		Ethnic Background	White (n=426)	0.94	0.70	1.41	4.46	3.29	3.05		
			Black (n=20)	0.00	0.00	0.00	0.00	0.00	5.00		
			Asian (n=41)	2.44	0.00	0.00	9.76	9.76	2.44		
			Others (n=27)	0.00	0.00	3.70	7.41	3.70	0.00		
	SIDER	SIDER	SIDER							3.00	17.00
Backache	SLAM	Trust	Clozapine (n=1760)	1.14	1.59	2.44	4.94	3.35	2.73		
		Ethnic Background	White (n=821)	0.85	1.10	1.58	5.36	2.56	1.34		
			Black (n=704)	0.99	1.28	1.85	6.25	2.98	1.56		
			Asian (n=93)	3.23	4.30	3.23	7.53	5.38	4.30		
			Other (n=142)	3.52	0.70	2.11	6.34	4.23	2.82		
	Camden & Islington	Trust	Clozapine (n=561)	1.43	1.25	1.96	5.35	3.03	3.03		
		Ethnic Background	White (n=347)	1.73	1.15	2.02	6.05	3.46	2.31		
			Black (n=120)	0.83	2.50	3.33	6.67	3.33	5.83		
			Asian (n=41)	2.44	0.00	0.00	0.00	2.44	2.44		
			Others (n=53)	0.00	0.00	0.00	1.89	0.00	1.89		
	Oxford	Trust	Clozapine (n=514)	1.17	1.17	1.36	5.84	3.89	2.92		
		Ethnic Background	White (n=426)	0.94	1.41	1.41	6.10	4.46	3.29		
			Black (n=20)	5.00	0.00	0.00	5.00	0.00	0.00		
			Asian (n=41)	0.00	0.00	0.00	4.88	0.00	0.00		
			Others (n=27)	3.70	0.00	3.70	3.70	3.70	3.70		

Measure Values: 0.00 – 56.00

The results are shown in percentages (%) and broken down by ADRs, Trusts (SLAM, Camden & Islington and Oxford), Cohorts, Sub Cohorts and SIDER reported values.
In Sub Cohort 'Clozapine' represent the total baseline population which further breaks down into 'White', 'Black', 'Asian' and 'Others' ethnic groups. The columns (Three Months Early, Two Months Early, One Month Early, One Month Later, Two Months Later, Three Months Later) shows the percentages in each monthly interval. The last two columns (SIDER Low End and SIDER High End) shows the SIDER reporting.

Clozapine - Ethnic Background (%)

ADR	Trust	Cohort	Sub Cohort	Three Months Early	Two Months Early	One Month Early	One Month Later	Two Months Later	Three Months Later	SIDER Low End	SIDER High End
Backache	SIDER	SIDER	SIDER							5.00	
Fever	SLAM	Trust	Clozapine (n=1760)	1.02	1.14	1.65	6.36	4.43	3.13		
		Ethnic Background	White (n=821)	1.34	1.34	1.58	5.97	4.14	2.92		
			Black (n=704)	1.56	1.56	1.85	6.96	4.83	3.41		
			Asian (n=93)	1.08	0.00	2.15	6.45	3.23	1.08		
			Other (n=142)	0.70	0.70	0.70	9.15	6.34	3.52		
	Camden & Islington	Trust	Clozapine (n=561)	0.89	0.89	0.53	3.74	2.67	0.89		
		Ethnic Background	White (n=347)	1.15	1.15	0.29	3.17	2.88	0.58		
			Black (n=120)	0.83	0.83	0.83	5.83	3.33	2.50		
			Asian (n=41)	0.00	0.00	0.00	0.00	2.44	0.00		
			Others (n=53)	0.00	0.00	1.89	5.66	3.77	1.89		
	Oxford	Trust	Clozapine (n=514)	0.39	0.78	0.58	3.11	2.72	2.33		
		Ethnic Background	White (n=426)	0.23	0.70	0.47	3.05	2.82	2.58		
			Black (n=20)	0.00	5.00	0.00	0.00	0.00	0.00		
			Asian (n=41)	0.00	0.00	2.44	4.88	2.44	0.00		
			Others (n=27)	3.70	0.00	0.00	3.70	3.70	3.70		
	SIDER	SIDER	SIDER							4.00	13.00
Enuresis	SLAM	Trust	Clozapine (n=1760)	1.02	0.80	1.25	4.20	3.92	3.24		
		Ethnic Background	White (n=821)	1.10	0.85	1.46	3.41	2.80	2.07		
			Black (n=704)	1.28	0.99	1.70	3.98	3.27	2.41		
			Asian (n=93)	1.08	1.08	2.15	7.53	3.23	1.08		
			Other (n=142)	1.41	0.00	1.41	4.93	5.63	1.41		
	Camden & Islington	Trust	Clozapine (n=561)	0.71	1.07	1.07	4.10	1.43	1.25		
		Ethnic Background	White (n=347)	0.58	0.86	0.86	3.17	1.73	1.73		
			Black (n=120)	0.00	1.67	1.67	6.67	0.83	0.83		
			Asian (n=41)	2.44	2.44	2.44	7.32	2.44	0.00		
			Others (n=53)	1.89	0.00	0.00	0.00	0.00	0.00		
	Oxford	Trust	Clozapine (n=514)	1.36	0.58	1.36	4.86	4.47	3.50		
		Ethnic Background	White (n=426)	1.41	0.47	1.41	5.16	4.23	2.82		
			Black (n=20)	0.00	0.00	0.00	5.00	5.00	10.00		
			Asian (n=41)	2.44	2.44	2.44	2.44	7.32	4.88		
			Others (n=27)	0.00	0.00	0.00	3.70	3.70	7.41		
	SIDER	SIDER	SIDER								
Drymouth	SLAM	Trust	Clozapine (n=1760)	1.08	1.53	1.65	4.66	3.69	2.33		
		Ethnic Background	White (n=821)	1.22	1.34	1.71	3.53	2.68	1.95		
			Black (n=704)	1.42	1.56	1.99	4.12	3.13	2.27		
			Asian (n=93)	0.00	4.30	0.00	5.38	6.45	3.23		
			Other (n=142)	2.11	3.52	1.41	9.86	3.52	2.11		
	Camden & Islington	Trust	Clozapine (n=561)	1.25	1.25	1.07	3.92	2.14	0.89		
		Ethnic Background	White (n=347)	1.44	1.15	0.86	4.32	2.88	0.86		
			Black (n=120)	0.00	0.00	0.83	2.50	1.67	0.83		
			Asian (n=41)	0.00	4.88	2.44	7.32	0.00	0.00		
			Others (n=53)	3.77	1.89	1.89	3.77	1.89	1.89		
	Oxford	Trust	Clozapine (n=514)	1.36	1.36	1.56	3.89	1.36	2.33		
		Ethnic Background	White (n=426)	1.41	1.41	1.88	4.46	1.41	2.58		
			Black (n=20)	0.00	0.00	0.00	0.00	0.00	0.00		
			Asian (n=41)	0.00	2.44	0.00	2.44	2.44	0.00		
			Others (n=27)	3.70	0.00	0.00	0.00	0.00	3.70		
	SIDER	SIDER	SIDER							5.00	20.00
Diarrhoea	SLAM	Trust	Clozapine (n=1760)	1.08	1.31	1.36	4.72	3.58	2.56		
		Ethnic Background	White (n=821)	0.73	1.10	1.10	3.29	2.19	1.95		
			Black (n=704)	0.85	1.28	1.28	3.84	2.56	2.27		
			Asian (n=93)	1.08	1.08	2.15	2.15	2.15	0.00		
			Other (n=142)	3.52	2.11	0.70	4.93	4.93	2.82		
	Camden & Islington	Trust	Clozapine (n=561)	0.71	1.25	0.18	3.03	3.39	3.03		
		Ethnic Background	White (n=347)	0.29	1.44	0.29	3.75	3.46	3.75		
			Black (n=120)	0.83	1.67	0.00	3.33	5.00	2.50		
			Asian (n=41)	2.44	0.00	0.00	0.00	0.00	0.00		
			Others (n=53)	0.00	0.00	0.00	0.00	1.89	1.89		
	Oxford	Trust	Clozapine (n=514)	1.17	0.78	1.36	4.09	3.70	2.53		
		Ethnic Background	White (n=426)	1.17	0.47	1.41	3.76	3.52	3.05		
			Black (n=20)	0.00	5.00	0.00	10.00	10.00	0.00		
			Asian (n=41)	2.44	0.00	2.44	2.44	2.44	0.00		
			Others (n=27)	0.00	3.70	0.00	7.41	3.70	0.00		
	SIDER	SIDER	SIDER							2.00	
Dyspepsia	SLAM	Trust	Clozapine (n=1760)	0.74	1.08	0.91	3.92	3.13	3.69		
		Ethnic	White (n=821)	0.37	1.34	0.97	3.78	2.19	2.92		

Measure Values: 0.00 – 56.00

The results are shown in percentages (%) and broken down by ADRs, Trusts (SLAM, Camden & Islington and Oxford), Cohorts, Sub Cohorts and SIDER reported values.
In Sub Cohort 'Clozapine' represent the total baseline population which further breaks down into 'White', 'Black', 'Asian' and 'Others' ethnic groups. The columns (Three Months Early, Two Months Early, One Month Early, One Month Later, Two Months Later, Three Months Later) shows the percentages in each monthly interval. The last two columns (SIDER Low End and SIDER High End) shows the SIDER reporting.

Clozapine - Ethnic Background (%)

ADR	Trust	Cohort	Sub Cohort	Three Months Early	Two Months Early	One Month Early	One Month Later	Two Months Later	Three Months Later	SIDER Low End	SIDER High End
Dyspepsia	SLAM	Ethnic Background	Black (n=704)	0.43	1.56	1.14	4.40	2.56	3.41		
			Asian (n=93)	2.15	2.15	0.00	1.08	5.38	1.08		
			Other (n=142)	2.82	0.70	0.70	2.11	2.11	4.23		
	Camden & Islington	Trust	Clozapine (n=561)	0.36	0.53	0.53	4.10	2.67	2.50		
		Ethnic Background	White (n=347)	0.29	0.86	0.29	4.32	2.59	2.88		
			Black (n=120)	0.00	0.00	1.67	2.50	2.50	0.00		
			Asian (n=41)	0.00	0.00	0.00	9.76	7.32	2.44		
			Others (n=53)	1.89	0.00	0.00	1.89	0.00	5.66		
	Oxford	Trust	Clozapine (n=514)	0.19	0.58	0.78	3.50	4.09	3.70		
		Ethnic Background	White (n=426)	0.23	0.70	0.70	3.52	4.69	3.99		
			Black (n=20)	0.00	0.00	0.00	0.00	0.00	0.00		
			Asian (n=41)	0.00	0.00	0.00	2.44	0.00	2.44		
			Others (n=27)	0.00	0.00	3.70	7.41	3.70	3.70		
	SIDER	SIDER	SIDER							8.00	14.00
Rash	SLAM	Trust	Clozapine (n=1760)	1.25	1.59	2.05	3.64	2.95	2.27		
		Ethnic Background	White (n=821)	0.73	1.34	1.22	1.83	1.46	1.34		
			Black (n=704)	0.85	1.56	1.42	2.13	1.70	1.56		
			Asian (n=93)	0.00	3.23	4.30	7.53	0.00	2.15		
			Other (n=142)	4.23	1.41	3.52	3.52	1.41	1.41		
	Camden & Islington	Trust	Clozapine (n=561)	1.25	1.25	0.89	4.28	1.96	2.14		
		Ethnic Background	White (n=347)	1.15	1.15	1.15	4.90	2.31	3.17		
			Black (n=120)	0.00	0.83	0.00	2.50	0.00	0.00		
			Asian (n=41)	4.88	2.44	2.44	4.88	4.88	2.44		
			Others (n=53)	1.89	1.89	0.00	1.89	0.00	0.00		
	Oxford	Trust	Clozapine (n=514)	0.97	1.17	1.17	3.70	2.33	1.36		
		Ethnic Background	White (n=426)	1.17	1.41	1.41	3.99	2.58	1.41		
			Black (n=20)	0.00	0.00	0.00	5.00	5.00	5.00		
			Asian (n=41)	0.00	0.00	0.00	2.44	0.00	0.00		
			Others (n=27)	0.00	0.00	0.00	0.00	0.00	0.00		
	SIDER	SIDER	SIDER								
Stomachpain	SLAM	Trust	Clozapine (n=1760)	1.93	1.76	1.93	4.94	3.52	3.52		
		Ethnic Background	White (n=821)	1.71	1.58	1.58	4.14	3.29	2.68		
			Black (n=704)	1.99	1.85	1.85	4.83	3.84	3.13		
			Asian (n=93)	3.23	3.23	1.08	3.23	3.23	2.15		
			Other (n=142)	3.52	3.52	3.52	6.34	5.63	2.82		
	Camden & Islington	Trust	Clozapine (n=561)	0.89	1.25	0.89	3.39	2.85	2.14		
		Ethnic Background	White (n=347)	0.29	1.73	0.58	2.59	2.31	1.73		
			Black (n=120)	2.50	0.83	2.50	7.50	5.83	4.17		
			Asian (n=41)	0.00	0.00	0.00	2.44	0.00	0.00		
			Others (n=53)	5.66	0.00	0.00	1.89	1.89	1.89		
	Oxford	Trust	Clozapine (n=514)	1.56	0.78	0.78	2.14	0.97	0.97		
		Ethnic Background	White (n=426)	1.41	0.94	0.47	2.58	1.17	1.17		
			Black (n=20)	0.00	0.00	5.00	0.00	0.00	0.00		
			Asian (n=41)	2.44	0.00	0.00	0.00	0.00	0.00		
			Others (n=27)	3.70	0.00	3.70	0.00	0.00	0.00		
	SIDER	SIDER	SIDER								
Tremor	SLAM	Trust	Clozapine (n=1760)	1.48	1.99	2.95	5.51	3.52	3.47		
		Ethnic Background	White (n=821)	1.10	1.71	2.19	3.29	2.92	2.44		
			Black (n=704)	1.28	1.99	2.56	3.84	3.41	2.84		
			Asian (n=93)	1.08	2.15	1.08	8.60	4.30	5.38		
			Other (n=142)	2.82	2.82	3.52	4.93	4.23	4.23		
	Camden & Islington	Trust	Clozapine (n=561)	1.60	1.78	2.14	3.92	1.96	2.14		
		Ethnic Background	White (n=347)	1.44	0.58	0.86	3.75	1.73	1.73		
			Black (n=120)	2.50	3.33	5.00	3.33	2.50	2.50		
			Asian (n=41)	0.00	0.00	0.00	0.00	0.00	2.44		
			Others (n=53)	1.89	3.77	3.77	7.55	3.77	3.77		
	SIDER	SIDER	SIDER							6.00	
Sweating	SLAM	Trust	Clozapine (n=1760)	1.08	0.97	1.36	4.43	4.26	2.84		
		Ethnic Background	White (n=821)	0.61	0.85	1.46	4.14	3.29	1.95		
			Black (n=704)	0.71	0.99	1.70	4.83	3.84	2.27		
			Asian (n=93)	2.15	0.00	1.08	0.00	1.08	3.23		
			Other (n=142)	2.82	0.70	0.70	2.11	2.11	2.11		
	Camden & Islington	Trust	Clozapine (n=561)	0.53	0.53	0.53	2.85	2.14	1.96		
		Ethnic Background	White (n=347)	0.00	0.58	0.29	2.02	2.31	2.02		
			Black (n=120)	1.67	0.83	0.83	5.00	2.50	2.50		
			Asian (n=41)	0.00	0.00	2.44	4.88	0.00	0.00		
			Others (n=53)	1.89	0.00	0.00	1.89	1.89	1.89		

Measure Values: 0.00 – 56.00

The results are shown in percentages (%) and broken down by ADRs, Trusts (SLAM, Camden & Islington and Oxford), Cohorts, Sub Cohorts and SIDER reported values.
In Sub Cohort 'Clozapine' represent the total baseline population which further breaks down into 'White', 'Black', 'Asian' and 'Others' ethnic groups. The columns (Three Months Early, Two Months Early, One Month Early, One Month Later, Two Months Later, Three Months Later) shows the percentages in each monthly interval. The last two columns (SIDER Low End and SIDER High End) shows the SIDER reporting.

Clozapine - Ethnic Background (%)

ADR	Trust	Cohort	Sub Cohort	Three Months Early	Two Months Early	One Month Early	One Month Later	Two Months Later	Three Months Later	SIDER Low End	SIDER High End
Sweating	Oxford	Trust	Clozapine (n=514)	1.17	0.97	0.97	2.72	1.36	1.95		
		Ethnic Background	White (n=426)	0.94	0.94	0.94	3.05	1.17	1.64		
			Black (n=20)	0.00	0.00	0.00	0.00	0.00	5.00		
			Asian (n=41)	0.00	2.44	2.44	2.44	4.88	4.88		
			Others (n=27)	7.41	0.00	0.00	0.00	0.00	0.00		
	SIDER	SIDER	SIDER							6.00	
Neutropenia	SLAM	Trust	Clozapine (n=1760)	0.80	0.80	0.74	5.34	2.73	2.61		
		Ethnic Background	White (n=821)	1.10	0.73	0.85	5.85	3.41	3.17		
			Black (n=704)	1.28	0.85	0.99	6.82	3.98	3.69		
			Asian (n=93)	0.00	1.08	1.08	3.23	2.15	2.15		
			Other (n=142)	0.70	0.00	0.70	6.34	1.41	1.41		
	Camden & Islington	Trust	Clozapine (n=561)	0.00	0.18	0.53	1.60	0.89	1.07		
		Ethnic Background	White (n=347)	0.00	0.00	0.00	0.58	0.29	0.86		
			Black (n=120)	0.00	0.83	2.50	4.17	3.33	2.50		
			Asian (n=41)	0.00	0.00	0.00	2.44	0.00	0.00		
			Others (n=53)	0.00	0.00	0.00	1.89	0.00	0.00		
	SIDER	SIDER	SIDER								
Blurredvision	SLAM	Trust	Clozapine (n=1760)	0.34	0.91	0.63	2.05	1.25	1.02		
		Ethnic Background	White (n=821)	0.49	0.97	0.49	1.34	1.58	1.10		
			Black (n=704)	0.57	1.14	0.57	1.56	1.85	1.28		
			Asian (n=93)	0.00	1.08	0.00	1.08	0.00	0.00		
			Other (n=142)	0.00	0.70	0.70	2.82	2.11	0.70		
	Camden & Islington	Trust	Clozapine (n=561)	0.89	0.53	0.71	1.25	0.36	0.89		
		Ethnic Background	White (n=347)	0.86	0.29	0.58	1.44	0.58	0.86		
			Black (n=120)	1.67	1.67	1.67	1.67	0.00	1.67		
			Asian (n=41)	0.00	0.00	0.00	0.00	0.00	0.00		
			Others (n=53)	0.00	0.00	0.00	0.00	0.00	0.00		
	Oxford	Trust	Clozapine (n=514)	0.19	0.39	0.39	1.56	1.56	1.17		
		Ethnic Background	White (n=426)	0.23	0.47	0.47	1.64	1.64	1.41		
			Black (n=20)	0.00	0.00	0.00	0.00	0.00	0.00		
			Asian (n=41)	0.00	0.00	0.00	0.00	2.44	0.00		
			Others (n=27)	0.00	0.00	0.00	3.70	0.00	0.00		
	SIDER	SIDER	SIDER							5.00	
Akathisia	SLAM	Trust	Clozapine (n=1760)	0.80	0.91	0.74	2.67	1.36	0.80		
		Ethnic Background	White (n=821)	0.49	0.49	0.24	1.58	0.73	0.61		
			Black (n=704)	0.57	0.57	0.28	1.85	0.85	0.71		
			Asian (n=93)	1.08	2.15	3.23	4.30	2.15	1.08		
			Other (n=142)	0.70	0.00	0.00	0.70	0.00	0.70		
	Camden & Islington	Trust	Clozapine (n=561)	0.00	0.53	0.00	1.25	1.07	0.53		
		Ethnic Background	White (n=347)	0.00	0.29	0.00	1.15	0.86	0.29		
			Black (n=120)	0.00	0.00	0.00	0.83	0.00	0.00		
			Asian (n=41)	0.00	0.00	0.00	2.44	2.44	0.00		
			Others (n=53)	0.00	1.89	0.00	0.00	3.77	3.77		
	Oxford	Trust	Clozapine (n=514)	0.97	0.78	0.97	1.36	1.17	0.97		
		Ethnic Background	White (n=426)	0.70	0.70	0.94	1.64	1.41	1.17		
			Black (n=20)	0.00	0.00	0.00	0.00	0.00	0.00		
			Asian (n=41)	4.88	0.00	2.44	0.00	0.00	0.00		
			Others (n=27)	0.00	3.70	0.00	0.00	0.00	0.00		
	SIDER	SIDER	SIDER							3.00	

Measure Values: 0.00 – 56.00

The results are shown in percentages (%) and broken down by ADRs, Trusts (SLAM, Camden & Islington and Oxford), Cohorts, Sub Cohorts and SIDER reported values.
In Sub Cohort 'Clozapine' represent the total baseline population which further breaks down into 'White', 'Black', 'Asian' and 'Others' ethnic groups. The columns (Three Months Early, Two Months Early, One Month Early, One Month Later, Two Months Later, Three Months Later) shows the percentages in each monthly interval. The last two columns (SIDER Low End and SIDER High End) shows the SIDER reporting.

Supplementary Table B.3 Age groups (%)

Clozapine - Age Groups (%)

ADR	Trust	Cohort	Sub Cohort	Three Months Early	Two Months Early	One Month Early	One Month Later	Two Months Later	Three Months Later	SIDER Low End	SIDER High End
Agitation	SLAM	Trust	Clozapine (n=1760)	17.61	22.10	26.53	46.59	32.56	26.99		
		Age Groups	Under 21 (n=57)	31.58	47.37	40.35	75.44	50.88	40.35		
			21-30 (n=422)	20.62	27.96	34.36	55.69	38.86	37.68		
			31-40 (n=488)	17.83	20.08	23.77	45.29	33.20	25.41		
			41-50 (n=479)	16.08	19.21	24.63	42.80	28.60	22.55		
			51-60 (n=233)	12.88	15.88	20.17	37.34	26.18	17.60		
			61-70 (n=62)	12.90	20.97	19.35	35.48	25.81	24.19		
			71-80 (n=18)	16.67	16.67	27.78	33.33	16.67	27.78		
	Camden & Islington	Trust	Clozapine (n=561)	13.37	17.83	18.36	43.14	28.34	21.03		
		Age Groups	21-30 (n=27)	40.74	55.56	59.26	66.67	37.04	40.74		
			31-40 (n=141)	20.57	24.82	18.44	47.52	34.75	29.79		
			41-50 (n=168)	8.33	13.10	16.07	40.48	25.60	20.24		
			51-60 (n=135)	9.63	11.11	12.59	37.78	23.70	13.33		
			61-70 (n=67)	7.46	13.43	19.40	41.79	28.36	17.91		
			71-80 (n=21)	14.29	19.05	19.05	42.86	28.57	4.76		
	SIDER	SIDER	SIDER							4.00	
Fatigue	SLAM	Trust	Clozapine (n=1760)	12.67	14.83	15.85	43.58	35.80	30.51		
		Age Groups	Under 21 (n=57)	26.32	28.07	29.82	68.42	68.42	49.12		
			21-30 (n=422)	18.01	20.85	21.80	53.79	43.84	40.76		
			31-40 (n=488)	11.68	12.50	15.78	43.65	33.40	28.89		
			41-50 (n=479)	10.02	13.36	13.57	39.25	30.69	24.84		
			51-60 (n=233)	8.15	10.30	9.01	31.33	33.05	24.03		
			61-70 (n=62)	6.45	6.45	9.68	35.48	25.81	30.65		
			71-80 (n=18)	22.22	22.22	5.56	22.22	11.11	5.56		
	Camden & Islington	Trust	Clozapine (n=561)	10.34	12.30	13.37	41.18	29.23	26.56		
		Age Groups	21-30 (n=27)	25.93	44.44	55.56	59.26	33.33	40.74		
			31-40 (n=141)	14.18	17.73	15.60	45.39	33.33	32.62		
			41-50 (n=168)	4.17	7.74	7.14	42.26	29.17	24.40		
			51-60 (n=135)	11.11	6.67	12.59	31.11	22.96	22.96		
			61-70 (n=67)	11.94	10.45	10.45	46.27	32.84	25.37		
			71-80 (n=21)	4.76	14.29	9.52	38.10	28.57	19.05		
	SIDER	SIDER	SIDER								
Sedation	SLAM	Trust	Clozapine (n=1760)	12.67	12.16	14.83	43.86	35.51	29.83		
		Age Groups	Under 21 (n=57)	26.32	26.32	28.07	66.67	61.40	43.86		
			21-30 (n=422)	18.48	18.96	21.80	52.13	46.21	38.63		
			31-40 (n=488)	11.68	10.66	13.32	44.47	33.61	29.92		
			41-50 (n=479)	10.23	8.98	10.65	39.46	30.69	26.72		
			51-60 (n=233)	8.15	7.30	11.59	33.05	25.75	19.31		
			61-70 (n=62)	3.23	4.84	11.29	38.71	32.26	24.19		
			71-80 (n=18)	16.67	16.67	16.67	33.33	16.67	11.11		
	Camden & Islington	Trust	Clozapine (n=561)	5.17	9.09	9.09	38.15	26.56	21.93		
		Age Groups	21-30 (n=27)	18.52	25.93	37.04	48.15	25.93	40.74		
			31-40 (n=141)	9.22	14.18	13.48	41.84	34.04	26.95		
			41-50 (n=168)	2.98	6.55	7.14	39.29	25.60	25.00		
			51-60 (n=135)	3.70	4.44	5.19	28.89	22.96	14.07		
			61-70 (n=67)	1.49	8.96	2.99	43.28	26.87	14.93		
			71-80 (n=21)	0.00	4.76	4.76	33.33	14.29	19.05		
	SIDER	SIDER	SIDER							25.00	46.00
Dizziness	SLAM	Trust	Clozapine (n=1760)	2.78	4.20	4.09	16.59	13.13	11.19		
		Age Groups	Under 21 (n=57)	7.02	1.75	5.26	28.07	21.05	15.79		
			21-30 (n=422)	2.84	4.98	4.98	18.25	12.80	12.32		
			31-40 (n=488)	2.05	4.51	2.87	14.14	12.50	8.20		
			41-50 (n=479)	3.13	4.59	3.34	15.03	12.11	10.02		
			51-60 (n=233)	3.00	2.15	5.15	16.74	14.16	15.02		
			61-70 (n=62)	0.00	3.23	8.06	22.58	17.74	16.13		
			71-80 (n=18)	5.56	5.56	5.56	27.78	11.11	11.11		
	Camden & Islington	Trust	Clozapine (n=561)	3.21	3.39	3.74	18.18	13.73	9.09		
		Age Groups	21-30 (n=27)	14.81	14.81	14.81	25.93	14.81	18.52		
			31-40 (n=141)	2.13	4.96	7.80	15.60	20.57	6.38		
			41-50 (n=168)	2.98	1.79	1.19	14.29	7.74	8.93		
			51-60 (n=135)	1.48	1.48	2.22	21.48	13.33	10.37		
			61-70 (n=67)	4.48	4.48	0.00	23.88	17.91	7.46		
			71-80 (n=21)	4.76	0.00	4.76	19.05	9.52	19.05		
	SIDER	SIDER	SIDER							12.00	27.00
Confusion	SLAM	Trust	Clozapine (n=1760)	4.72	5.51	6.08	13.92	8.47	6.76		
		Age Groups	Under 21 (n=57)	8.77	5.26	12.28	21.05	22.81	12.28		

Measure Values: 0.00 – 75.44

The results are shown in percentages (%) and broken down by ADRs, Trusts (SLAM, Camden & Islington and Oxford), Cohorts, Sub Cohorts and SIDER reported values.
In Sub Cohort 'Clozapine' represent the total baseline population which further breaks down into age groups.
The columns (Three Months Early, Two Months Early, One Month Early, One Month Later, Two Months Later, Three Months Later) shows the percentages in each monthly interval. The last two columns (SIDER Low End and SIDER High End) shows the SIDER reporting.

Clozapine - Age Groups (%)

ADR	Trust	Cohort	Sub Cohort	Three Months Early	Two Months Early	One Month Early	One Month Later	Two Months Later	Three Months Later	SIDER Low End	SIDER High End
Confusion	SLAM	Age Groups	21-30 (n=422)	5.69	6.40	6.16	13.27	8.77	6.64		
			31-40 (n=488)	3.69	5.12	6.76	11.89	7.99	6.76		
			41-50 (n=479)	4.38	4.80	4.59	13.78	7.52	6.26		
			51-60 (n=233)	4.72	4.29	5.58	17.17	7.30	8.15		
			61-70 (n=62)	4.84	9.68	8.06	19.35	11.29	3.23		
			71-80 (n=18)	5.56	11.11	5.56	5.56	0.00	0.00		
	Camden & Islington	Trust	Clozapine (n=561)	3.57	6.24	5.53	12.66	6.77	5.88		
		Age Groups	21-30 (n=27)	7.41	25.93	22.22	33.33	3.70	7.41		
			31-40 (n=141)	4.96	9.22	2.84	9.93	7.09	6.38		
			41-50 (n=168)	2.98	4.17	4.17	9.52	5.36	7.14		
			51-60 (n=135)	1.48	1.48	5.19	12.59	8.15	4.44		
			61-70 (n=67)	4.48	7.46	8.96	19.40	8.96	5.97		
			71-80 (n=21)	4.76	4.76	4.76	9.52	0.00	0.00		
	SIDER	SIDER	SIDER							3.00	
Constipation	SLAM	Trust	Clozapine (n=1760)	1.76	1.99	2.16	12.27	11.70	9.49		
		Age Groups	Under 21 (n=57)	1.75	5.26	7.02	19.30	15.79	15.79		
			21-30 (n=422)	1.66	0.24	1.90	13.27	13.27	9.24		
			31-40 (n=488)	1.84	1.84	1.64	9.22	9.22	10.04		
			41-50 (n=479)	1.46	2.51	2.51	11.90	9.39	7.10		
			51-60 (n=233)	1.72	3.00	2.15	12.88	15.88	9.44		
			61-70 (n=62)	1.61	4.84	1.61	19.35	17.74	17.74		
			71-80 (n=18)	11.11	0.00	0.00	27.78	16.67	16.67		
	Camden & Islington	Trust	Clozapine (n=561)	1.07	2.50	1.78	11.41	7.13	5.70		
		Age Groups	21-30 (n=27)	0.00	3.70	3.70	0.00	7.41	0.00		
			31-40 (n=141)	1.42	2.84	1.42	12.06	4.96	6.38		
			41-50 (n=168)	0.60	1.19	1.19	9.52	7.14	5.95		
			51-60 (n=135)	0.74	1.48	0.74	10.37	7.41	5.93		
			61-70 (n=67)	1.49	7.46	4.48	17.91	10.45	5.97		
			71-80 (n=21)	4.76	0.00	4.76	23.81	9.52	4.76		
	SIDER	SIDER	SIDER							10.00	25.00
Hypersalivation	SLAM	Trust	Clozapine (n=1760)	1.19	1.48	2.10	14.32	13.24	11.31		
		Age Groups	Under 21 (n=57)	0.00	3.51	3.51	24.56	15.79	19.30		
			21-30 (n=422)	2.37	1.90	3.55	17.30	13.03	14.45		
			31-40 (n=488)	1.23	1.02	1.23	14.55	14.34	12.70		
			41-50 (n=479)	0.84	1.46	1.25	10.65	11.48	7.31		
			51-60 (n=233)	0.43	1.72	2.58	14.59	14.16	9.01		
			61-70 (n=62)	0.00	0.00	3.23	9.68	14.52	14.52		
			71-80 (n=18)	0.00	0.00	0.00	16.67	11.11	0.00		
	Camden & Islington	Trust	Clozapine (n=561)	1.07	1.43	0.53	14.26	6.95	7.66		
		Age Groups	21-30 (n=27)	3.70	3.70	0.00	7.41	3.70	7.41		
			31-40 (n=141)	2.84	3.55	0.71	17.73	12.06	10.64		
			41-50 (n=168)	0.00	0.00	1.19	12.50	4.17	4.17		
			51-60 (n=135)	0.00	0.00	0.00	14.81	7.41	9.63		
			61-70 (n=67)	1.49	2.99	0.00	17.91	5.97	7.46		
			71-80 (n=21)	0.00	0.00	0.00	0.00	0.00	9.52		
	SIDER	SIDER	SIDER							1.00	48.00
Tachycardia	SLAM	Trust	Clozapine (n=1760)	2.27	2.05	2.50	15.40	12.95	9.94		
		Age Groups	Under 21 (n=57)	10.53	8.77	7.02	36.84	35.09	28.07		
			21-30 (n=422)	3.55	2.84	3.08	21.56	18.48	16.59		
			31-40 (n=488)	1.43	1.43	1.84	14.14	11.89	8.20		
			41-50 (n=479)	1.67	2.09	1.88	12.32	9.81	5.64		
			51-60 (n=233)	0.86	0.43	3.00	11.59	7.73	6.44		
			61-70 (n=62)	0.00	0.00	0.00	4.84	8.06	8.06		
			71-80 (n=18)	11.11	5.56	11.11	5.56	11.11	11.11		
	Camden & Islington	Trust	Clozapine (n=561)	1.43	1.43	0.89	11.23	8.38	6.95		
		Age Groups	21-30 (n=27)	11.11	3.70	7.41	25.93	11.11	14.81		
			31-40 (n=141)	0.71	1.42	0.71	13.48	11.35	9.93		
			41-50 (n=168)	1.19	1.79	0.00	7.14	8.93	4.76		
			51-60 (n=135)	0.00	0.74	0.00	11.85	5.93	5.93		
			61-70 (n=67)	2.99	1.49	2.99	13.43	5.97	7.46		
			71-80 (n=21)	0.00	0.00	0.00	0.00	4.76	0.00		
	SIDER	SIDER	SIDER							11.00	25.00
Weightgain	SLAM	Trust	Clozapine (n=1760)	3.75	4.43	5.06	15.34	10.91	10.34		
		Age Groups	Under 21 (n=57)	10.53	14.04	14.04	31.58	28.07	29.82		
			21-30 (n=422)	5.92	7.58	7.35	19.43	15.17	14.22		
			31-40 (n=488)	4.30	3.48	5.12	20.49	12.50	9.43		

Measure Values: 0.00 – 75.44

The results are shown in percentages (%) and broken down by ADRs, Trusts (SLAM, Camden & Islington and Oxford), Cohorts, Sub Cohorts and SIDER reported values.
In Sub Cohort 'Clozapine' represent the total baseline population which further breaks down into age groups.
The columns (Three Months Early, Two Months Early, One Month Early, One Month Later, Two Months Later, Three Months Later) shows the percentages in each monthly interval. The last two columns (SIDER Low End and SIDER High End) shows the SIDER reporting.

Clozapine - Age Groups (%)

ADR	Trust	Cohort	Sub Cohort	Three Months Early	Two Months Early	One Month Early	One Month Later	Two Months Later	Three Months Later	SIDER Low End	SIDER High End
Weightgain	SLAM	Age Groups	41-50 (n=479)	2.30	2.51	3.34	10.86	7.31	7.93		
			51-60 (n=233)	0.86	3.00	3.00	6.44	4.29	6.87		
			61-70 (n=62)	1.61	1.61	0.00	4.84	8.06	4.84		
			71-80 (n=18)	0.00	5.56	11.11	0.00	5.56	11.11		
	Camden & Islington	Trust	Clozapine (n=561)	2.50	3.39	1.96	11.76	6.60	6.24		
		Age Groups	21-30 (n=27)	0.00	7.41	0.00	18.52	22.22	11.11		
			31-40 (n=141)	2.84	4.26	3.55	11.35	7.80	9.22		
			41-50 (n=168)	2.38	2.98	1.19	11.31	5.36	5.95		
			51-60 (n=135)	3.70	2.96	0.74	12.59	6.67	5.19		
			61-70 (n=67)	1.49	2.99	4.48	8.96	1.49	2.99		
			71-80 (n=21)	0.00	0.00	0.00	14.29	4.76	0.00		
	SIDER	SIDER	SIDER							4.00	56.00
Feelingsick	SLAM	Trust	Clozapine (n=1760)	4.66	4.94	6.48	14.32	11.19	9.09		
		Age Groups	Under 21 (n=57)	7.02	17.54	15.79	31.58	26.32	19.30		
			21-30 (n=422)	7.82	6.40	7.11	20.38	15.64	12.80		
			31-40 (n=488)	3.89	5.53	6.76	12.91	10.86	7.58		
			41-50 (n=479)	3.55	2.92	6.05	11.48	9.81	8.56		
			51-60 (n=233)	2.15	2.58	4.72	9.44	5.58	4.72		
			61-70 (n=62)	3.23	3.23	1.61	11.29	4.84	6.45		
			71-80 (n=18)	11.11	5.56	5.56	5.56	0.00	11.11		
	Camden & Islington	Trust	Clozapine (n=561)	3.74	3.92	3.03	10.52	7.13	7.66		
		Age Groups	21-30 (n=27)	7.41	14.81	11.11	7.41	7.41	22.22		
			31-40 (n=141)	7.09	3.55	4.96	14.89	12.06	8.51		
			41-50 (n=168)	0.60	2.98	2.38	10.71	4.76	7.74		
			51-60 (n=135)	4.44	3.70	0.74	7.41	4.44	4.44		
			61-70 (n=67)	2.99	2.99	1.49	8.96	8.96	7.46		
			71-80 (n=21)	0.00	4.76	4.76	9.52	4.76	4.76		
	SIDER	SIDER	SIDER								
Headache	SLAM	Trust	Clozapine (n=1760)	4.20	4.55	5.45	12.44	8.18	5.91		
		Age Groups	Under 21 (n=57)	15.79	17.54	7.02	12.28	14.04	14.04		
			21-30 (n=422)	5.92	6.40	7.82	17.06	9.72	7.82		
			31-40 (n=488)	3.48	3.48	5.94	12.91	8.20	5.94		
			41-50 (n=479)	3.13	3.55	4.38	10.65	6.47	5.64		
			51-60 (n=233)	3.00	3.00	3.00	8.58	7.30	2.58		
			61-70 (n=62)	1.61	1.61	0.00	6.45	8.06	1.61		
			71-80 (n=18)	0.00	5.56	11.11	11.11	11.11	0.00		
	Camden & Islington	Trust	Clozapine (n=561)	2.32	3.57	4.28	9.27	6.42	4.63		
		Age Groups	21-30 (n=27)	3.70	22.22	3.70	11.11	11.11	0.00		
			31-40 (n=141)	4.26	4.96	5.67	12.77	9.93	9.93		
			41-50 (n=168)	1.79	1.19	4.17	7.74	4.76	4.17		
			51-60 (n=135)	0.74	0.74	3.70	7.41	2.96	0.74		
			61-70 (n=67)	2.99	5.97	4.48	10.45	10.45	5.97		
			71-80 (n=21)	0.00	0.00	0.00	4.76	0.00	0.00		
	SIDER	SIDER	SIDER								
Insomnia	SLAM	Trust	Clozapine (n=1760)	3.92	4.03	5.17	10.40	6.48	4.03		
		Age Groups	Under 21 (n=57)	12.28	8.77	10.53	21.05	7.02	10.53		
			21-30 (n=422)	5.69	4.50	5.92	11.85	7.82	5.69		
			31-40 (n=488)	4.10	4.92	4.51	10.25	6.35	3.07		
			41-50 (n=479)	1.88	3.55	6.05	9.39	7.52	4.38		
			51-60 (n=233)	3.00	2.15	3.86	9.01	3.43	2.15		
			61-70 (n=62)	3.23	0.00	0.00	6.45	1.61	0.00		
			71-80 (n=18)	0.00	5.56	0.00	5.56	5.56	0.00		
	Camden & Islington	Trust	Clozapine (n=561)	3.57	3.39	3.74	8.91	3.39	4.28		
		Age Groups	21-30 (n=27)	7.41	7.41	3.70	14.81	7.41	3.70		
			31-40 (n=141)	5.67	5.67	4.96	9.93	4.26	6.38		
			41-50 (n=168)	2.38	2.98	2.38	5.95	0.00	2.38		
			51-60 (n=135)	2.22	2.22	3.70	8.15	5.93	5.19		
			61-70 (n=67)	1.49	1.49	5.97	13.43	4.48	4.48		
			71-80 (n=21)	9.52	0.00	0.00	9.52	0.00	0.00		
	SIDER	SIDER	SIDER							20.00	33.00
Hypertension	SLAM	Trust	Clozapine (n=1760)	2.05	2.22	3.13	9.15	5.74	4.60		
		Age Groups	Under 21 (n=57)	0.00	1.75	0.00	14.04	10.53	5.26		
			21-30 (n=422)	1.66	1.18	1.66	7.58	4.27	3.79		
			31-40 (n=488)	1.23	1.64	2.25	6.15	4.30	3.48		
			41-50 (n=479)	2.92	2.92	4.59	11.27	5.85	4.80		
			51-60 (n=233)	2.15	2.15	3.86	11.16	6.87	6.44		

Measure Values: 0.00 – 75.44

The results are shown in percentages (%) and broken down by ADRs, Trusts (SLAM, Camden & Islington and Oxford), Cohorts, Sub Cohorts and SIDER reported values.
In Sub Cohort 'Clozapine' represent the total baseline population which further breaks down into age groups.
The columns (Three Months Early, Two Months Early, One Month Early, One Month Later, Two Months Later, Three Months Later) shows the percentages in each monthly interval. The last two columns (SIDER Low End and SIDER High End) shows the SIDER reporting.

Clozapine - Age Groups (%)

ADR	Trust	Cohort	Sub Cohort	Three Months Early	Two Months Early	One Month Early	One Month Later	Two Months Later	Three Months Later	SIDER Low End	SIDER High End
Hypertension	SLAM	Age Groups	61-70 (n=62)	6.45	4.84	6.45	12.90	17.74	9.68		
			71-80 (n=18)	0.00	11.11	11.11	11.11	0.00	5.56		
	Camden & Islington	Trust	Clozapine (n=561)	0.71	0.71	1.60	7.13	4.63	2.67		
		Age Groups	21-30 (n=27)	0.00	0.00	0.00	3.70	3.70	0.00		
			31-40 (n=141)	1.42	0.00	0.00	4.96	7.09	3.55		
			41-50 (n=168)	0.00	0.60	0.00	3.57	2.98	2.98		
			51-60 (n=135)	0.74	0.74	2.22	8.15	2.96	1.48		
			61-70 (n=67)	0.00	2.99	5.97	14.93	5.97	4.48		
			71-80 (n=21)	4.76	0.00	9.52	23.81	9.52	0.00		
	SIDER	SIDER	SIDER							4.00	12.00
Shaking	SLAM	Trust	Clozapine (n=1760)	3.13	2.95	3.92	9.55	5.40	5.06		
		Age Groups	Under 21 (n=57)	5.26	7.02	10.53	14.04	14.04	17.54		
			21-30 (n=422)	6.16	3.55	4.50	15.64	7.58	7.35		
			31-40 (n=488)	1.64	3.07	3.69	8.20	4.30	4.30		
			41-50 (n=479)	2.30	2.51	3.55	7.52	4.18	2.51		
			51-60 (n=233)	2.15	1.29	3.00	6.01	3.86	3.86		
			61-70 (n=62)	1.61	1.61	1.61	6.45	4.84	6.45		
			71-80 (n=18)	5.56	11.11	5.56	0.00	11.11	11.11		
	Camden & Islington	Trust	Clozapine (n=561)	1.78	1.96	3.74	6.06	3.92	2.85		
		Age Groups	21-30 (n=27)	0.00	7.41	14.81	22.22	11.11	14.81		
			31-40 (n=141)	4.26	2.84	4.96	5.67	4.26	4.96		
			41-50 (n=168)	0.60	1.79	2.98	4.17	2.38	0.60		
			51-60 (n=135)	1.48	0.74	1.48	5.93	2.96	0.00		
			61-70 (n=67)	1.49	1.49	1.49	2.99	4.48	5.97		
			71-80 (n=21)	0.00	0.00	9.52	14.29	9.52	0.00		
	SIDER	SIDER	SIDER								
Vomiting	SLAM	Trust	Clozapine (n=1760)	2.56	2.50	3.01	8.86	6.82	5.00		
		Age Groups	Under 21 (n=57)	5.26	7.02	12.28	17.54	15.79	17.54		
			21-30 (n=422)	3.55	3.08	4.27	12.80	9.00	6.16		
			31-40 (n=488)	2.66	3.69	2.46	7.38	7.17	4.30		
			41-50 (n=479)	1.88	1.46	2.09	7.10	5.01	3.76		
			51-60 (n=233)	2.15	0.43	2.15	6.87	4.72	3.86		
			61-70 (n=62)	0.00	0.00	1.61	6.45	3.23	4.84		
			71-80 (n=18)	0.00	5.56	0.00	11.11	5.56	5.56		
	Camden & Islington	Trust	Clozapine (n=561)	2.14	2.50	2.85	6.77	4.99	4.63		
		Age Groups	21-30 (n=27)	11.11	7.41	11.11	11.11	11.11	11.11		
			31-40 (n=141)	3.55	2.84	3.55	9.93	7.80	6.38		
			41-50 (n=168)	0.60	1.19	1.79	6.55	2.98	4.76		
			51-60 (n=135)	1.48	2.22	2.96	2.96	2.96	2.96		
			61-70 (n=67)	1.49	2.99	1.49	7.46	4.48	2.99		
			71-80 (n=21)	0.00	4.76	0.00	4.76	9.52	0.00		
	SIDER	SIDER	SIDER							3.00	17.00
Hyperprolactinaemia	SLAM	Trust	Clozapine (n=1760)	3.18	3.64	4.20	8.52	5.06	4.15		
		Age Groups	Under 21 (n=57)	12.28	14.04	8.77	21.05	17.54	12.28		
			21-30 (n=422)	3.55	4.50	5.45	9.95	6.40	7.11		
			31-40 (n=488)	3.89	4.10	4.30	8.81	4.92	3.89		
			41-50 (n=479)	1.88	2.09	3.34	8.14	3.97	3.13		
			51-60 (n=233)	1.72	2.58	3.00	4.29	3.43	0.86		
			61-70 (n=62)	3.23	1.61	3.23	6.45	1.61	0.00		
			71-80 (n=18)	0.00	0.00	0.00	0.00	0.00	0.00		
	Camden & Islington	Trust	Clozapine (n=561)	1.60	1.78	2.67	8.20	4.10	3.57		
		Age Groups	21-30 (n=27)	7.41	0.00	7.41	14.81	3.70	0.00		
			31-40 (n=141)	2.84	2.84	6.38	12.77	5.67	4.96		
			41-50 (n=168)	0.00	0.60	0.00	7.74	4.17	4.17		
			51-60 (n=135)	1.48	1.48	0.74	5.93	2.96	2.96		
			61-70 (n=67)	1.49	4.48	4.48	4.48	4.48	2.99		
			71-80 (n=21)	0.00	0.00	0.00	0.00	0.00	0.00		
	SIDER	SIDER	SIDER								
Tremor	SLAM	Trust	Clozapine (n=1760)	1.48	1.99	2.95	5.51	3.52	3.47		
		Age Groups	Under 21 (n=57)	1.75	3.51	3.51	12.28	7.02	7.02		
			21-30 (n=422)	0.95	2.37	3.32	5.69	2.13	3.55		
			31-40 (n=488)	1.23	0.61	1.84	4.10	2.46	3.48		
			41-50 (n=479)	1.25	2.09	2.51	4.59	4.18	2.51		
			51-60 (n=233)	2.15	2.15	5.15	6.87	5.15	3.00		
			61-70 (n=62)	3.23	4.84	3.23	9.68	4.84	8.06		
			71-80 (n=18)	11.11	11.11	5.56	11.11	11.11	5.56		

Measure Values: 0.00 – 75.44

The results are shown in percentages (%) and broken down by ADRs, Trusts (SLAM, Camden & Islington and Oxford), Cohorts, Sub Cohorts and SIDER reported values.
In Sub Cohort 'Clozapine' represent the total baseline population which further breaks down into age groups.
The columns (Three Months Early, Two Months Early, One Month Early, One Month Later, Two Months Later, Three Months Later) shows the percentages in each monthly interval. The last two columns (SIDER Low End and SIDER High End) shows the SIDER reporting.

Clozapine - Age Groups (%)

ADR	Trust	Cohort	Sub Cohort	Three Months Early	Two Months Early	One Month Early	One Month Later	Two Months Later	Three Months Later	SIDER Low End	SIDER High End
Tremor	Camden & Islington	Trust	Clozapine (n=561)	1.60	1.78	2.14	3.92	1.96	2.14		
		Age Groups	21-30 (n=27)	3.70	0.00	7.41	7.41	0.00	3.70		
			31-40 (n=141)	2.13	2.84	2.84	2.84	2.13	2.13		
			41-50 (n=168)	0.60	1.19	1.19	2.98	2.38	1.79		
			51-60 (n=135)	0.74	0.74	2.22	2.96	0.74	0.74		
			61-70 (n=67)	1.49	4.48	1.49	7.46	2.99	5.97		
			71-80 (n=21)	9.52	0.00	0.00	9.52	4.76	0.00		
	SIDER	SIDER	SIDER							6.00	
Abdominalpain	SLAM	Trust	Clozapine (n=1760)	1.88	1.99	2.56	8.01	6.02	4.72		
		Age Groups	Under 21 (n=57)	3.51	0.00	1.75	12.28	8.77	5.26		
			21-30 (n=422)	2.84	2.13	3.32	9.24	5.92	4.98		
			31-40 (n=488)	1.02	1.43	1.84	8.61	7.17	5.53		
			41-50 (n=479)	2.51	3.76	2.30	7.52	4.38	4.18		
			51-60 (n=233)	0.43	0.43	3.86	5.58	6.87	3.43		
			61-70 (n=62)	0.00	0.00	0.00	6.45	4.84	3.23		
			71-80 (n=18)	5.56	0.00	5.56	0.00	5.56	5.56		
	Camden & Islington	Trust	Clozapine (n=561)	0.89	0.89	1.60	3.92	3.57	3.39		
		Age Groups	21-30 (n=27)	3.70	0.00	3.70	7.41	3.70	3.70		
			31-40 (n=141)	0.00	1.42	2.84	7.09	4.96	4.96		
			41-50 (n=168)	0.00	0.60	1.19	2.98	3.57	2.98		
			51-60 (n=135)	1.48	0.74	0.74	1.48	1.48	0.74		
			61-70 (n=67)	0.00	1.49	1.49	4.48	4.48	4.48		
			71-80 (n=21)	0.00	0.00	0.00	0.00	4.76	9.52		
	SIDER	SIDER	SIDER							4.00	
Hypotension	SLAM	Trust	Clozapine (n=1760)	0.51	0.97	0.80	5.00	2.95	2.56		
		Age Groups	Under 21 (n=57)	1.75	3.51	3.51	15.79	8.77	12.28		
			21-30 (n=422)	0.47	0.95	0.95	4.98	2.84	3.79		
			31-40 (n=488)	0.20	0.41	0.41	4.10	2.87	1.64		
			41-50 (n=479)	0.21	1.25	0.42	3.76	1.67	0.84		
			51-60 (n=233)	0.86	0.86	1.29	5.58	3.00	3.00		
			61-70 (n=62)	1.61	0.00	0.00	9.68	9.68	4.84		
			71-80 (n=18)	5.56	5.56	5.56	5.56	0.00	0.00		
	Camden & Islington	Trust	Clozapine (n=561)	0.18	0.53	0.18	3.57	2.32	1.78		
		Age Groups	21-30 (n=27)	0.00	0.00	0.00	0.00	0.00	0.00		
			31-40 (n=141)	0.00	0.00	0.71	2.13	4.26	0.71		
			41-50 (n=168)	0.00	0.00	0.00	2.38	0.00	1.19		
			51-60 (n=135)	0.00	0.00	0.00	5.93	2.96	2.96		
			61-70 (n=67)	1.49	4.48	0.00	4.48	2.99	0.00		
			71-80 (n=21)	0.00	0.00	0.00	9.52	4.76	14.29		
	SIDER	SIDER	SIDER							9.00	38.00
Nausea	SLAM	Trust	Clozapine (n=1760)	1.14	1.08	1.19	6.08	5.23	3.69		
		Age Groups	Under 21 (n=57)	3.51	5.26	1.75	7.02	8.77	7.02		
			21-30 (n=422)	1.66	0.95	1.42	8.06	7.11	4.74		
			31-40 (n=488)	0.82	1.64	1.43	5.74	4.30	2.87		
			41-50 (n=479)	0.63	0.63	0.63	5.64	5.22	2.09		
			51-60 (n=233)	1.72	0.43	1.29	4.72	4.29	4.72		
			61-70 (n=62)	0.00	0.00	1.61	3.23	1.61	8.06		
			71-80 (n=18)	0.00	0.00	0.00	5.56	0.00	5.56		
	Camden & Islington	Trust	Clozapine (n=561)	0.89	1.43	0.36	4.63	3.57	3.57		
		Age Groups	21-30 (n=27)	0.00	3.70	0.00	3.70	7.41	3.70		
			31-40 (n=141)	1.42	1.42	0.00	4.26	4.26	2.84		
			41-50 (n=168)	0.00	0.60	1.19	4.76	1.19	4.17		
			51-60 (n=135)	1.48	1.48	0.00	4.44	4.44	3.70		
			61-70 (n=67)	1.49	1.49	0.00	5.97	5.97	2.99		
			71-80 (n=21)	0.00	4.76	0.00	4.76	0.00	9.52		
	SIDER	SIDER	SIDER							3.00	17.00
Fever	SLAM	Trust	Clozapine (n=1760)	1.02	1.14	1.65	6.36	4.43	3.13		
		Age Groups	Under 21 (n=57)	1.75	3.51	5.26	5.26	7.02	7.02		
			21-30 (n=422)	1.42	2.13	2.13	7.11	3.79	4.27		
			31-40 (n=488)	1.02	1.02	1.43	6.76	4.30	1.84		
			41-50 (n=479)	0.84	0.63	0.84	6.68	3.76	3.13		
			51-60 (n=233)	0.43	0.43	2.15	4.72	5.58	2.58		
			61-70 (n=62)	0.00	0.00	1.61	3.23	6.45	3.23		
			71-80 (n=18)	5.56	0.00	0.00	5.56	11.11	5.56		
	Camden & Islington	Trust	Clozapine (n=561)	0.89	0.89	0.53	3.74	2.67	0.89		
		Age Groups	21-30 (n=27)	0.00	3.70	3.70	7.41	7.41	7.41		

Measure Values: 0.00 – 75.44

The results are shown in percentages (%) and broken down by ADRs, Trusts (SLAM, Camden & Islington and Oxford), Cohorts, Sub Cohorts and SIDER reported values.
In Sub Cohort 'Clozapine' represent the total baseline population which further breaks down into age groups.
The columns (Three Months Early, Two Months Early, One Month Early, One Month Later, Two Months Later, Three Months Later) shows the percentages in each monthly interval. The last two columns (SIDER Low End and SIDER High End) shows the SIDER reporting.

Clozapine - Age Groups (%)

ADR	Trust	Cohort	Sub Cohort	Three Months Early	Two Months Early	One Month Early	One Month Later	Two Months Later	Three Months Later	SIDER Low End	SIDER High End
Fever	Camden & Islington	Age Groups	31-40 (n=141)	1.42	2.13	0.71	4.96	4.26	1.42		
			41-50 (n=168)	0.00	0.00	0.00	0.00	1.19	0.60		
			51-60 (n=135)	1.48	0.74	0.00	5.19	1.48	0.00		
			61-70 (n=67)	1.49	0.00	1.49	5.97	2.99	0.00		
			71-80 (n=21)	0.00	0.00	0.00	4.76	4.76	0.00		
	SIDER	SIDER	SIDER							4.00	13.00
Convulsion	SLAM	Trust	Clozapine (n=1760)	1.36	1.70	1.82	7.05	4.94	4.03		
		Age Groups	Under 21 (n=57)	0.00	1.75	3.51	10.53	5.26	8.77		
			21-30 (n=422)	2.37	2.37	1.66	9.95	6.87	4.50		
			31-40 (n=488)	1.02	1.64	1.64	6.56	3.48	4.30		
			41-50 (n=479)	1.46	2.09	2.09	6.26	3.97	3.13		
			51-60 (n=233)	0.86	0.00	2.15	4.72	4.72	2.15		
			61-70 (n=62)	0.00	0.00	0.00	4.84	9.68	8.06		
			71-80 (n=18)	0.00	5.56	0.00	0.00	5.56	5.56		
	Camden & Islington	Trust	Clozapine (n=561)	0.53	0.53	0.36	2.85	2.14	1.07		
		Age Groups	21-30 (n=27)	3.70	3.70	3.70	3.70	0.00	0.00		
			31-40 (n=141)	0.71	1.42	0.71	2.13	3.55	2.13		
			41-50 (n=168)	0.60	0.00	0.00	0.60	2.38	0.60		
			51-60 (n=135)	0.00	0.00	0.00	4.44	2.22	1.48		
			61-70 (n=67)	0.00	0.00	0.00	4.48	0.00	0.00		
			71-80 (n=21)	0.00	0.00	0.00	9.52	0.00	0.00		
	SIDER	SIDER	SIDER							3.00	
Enuresis	SLAM	Trust	Clozapine (n=1760)	1.02	0.80	1.25	4.20	3.92	3.24		
		Age Groups	Under 21 (n=57)	1.75	0.00	0.00	3.51	7.02	8.77		
			21-30 (n=422)	1.42	0.71	0.95	3.79	3.55	2.37		
			31-40 (n=488)	1.23	0.61	0.41	2.66	4.10	4.10		
			41-50 (n=479)	0.84	1.04	2.71	4.38	3.76	2.30		
			51-60 (n=233)	0.43	0.86	0.43	6.87	3.43	3.00		
			61-70 (n=62)	0.00	1.61	0.00	8.06	6.45	4.84		
			71-80 (n=18)	0.00	0.00	11.11	5.56	0.00	0.00		
	Camden & Islington	Trust	Clozapine (n=561)	0.71	1.07	1.07	4.10	1.43	1.25		
		Age Groups	21-30 (n=27)	0.00	0.00	0.00	3.70	0.00	0.00		
			31-40 (n=141)	0.00	2.13	1.42	4.96	2.84	1.42		
			41-50 (n=168)	1.19	0.60	0.60	1.79	0.00	0.60		
			51-60 (n=135)	1.48	0.74	1.48	4.44	0.74	0.74		
			61-70 (n=67)	0.00	1.49	1.49	4.48	1.49	1.49		
			71-80 (n=21)	0.00	0.00	0.00	14.29	9.52	9.52		
	SIDER	SIDER	SIDER								
Backache	SLAM	Trust	Clozapine (n=1760)	1.14	1.59	2.44	4.94	3.35	2.73		
		Age Groups	Under 21 (n=57)	0.00	1.75	3.51	5.26	3.51	1.75		
			21-30 (n=422)	0.95	0.95	1.90	3.79	3.32	3.08		
			31-40 (n=488)	0.61	1.84	2.66	3.69	2.66	2.66		
			41-50 (n=479)	1.67	1.88	2.51	6.26	3.34	2.92		
			51-60 (n=233)	1.29	0.86	2.58	6.01	4.29	1.29		
			61-70 (n=62)	1.61	4.84	1.61	8.06	4.84	4.84		
			71-80 (n=18)	5.56	0.00	5.56	0.00	5.56	5.56		
	Camden & Islington	Trust	Clozapine (n=561)	1.43	1.25	1.96	5.35	3.03	3.03		
		Age Groups	21-30 (n=27)	0.00	3.70	0.00	0.00	0.00	3.70		
			31-40 (n=141)	1.42	1.42	2.13	4.26	2.13	4.26		
			41-50 (n=168)	1.19	0.00	1.79	3.57	2.98	2.98		
			51-60 (n=135)	0.74	0.74	2.22	5.93	1.48	1.48		
			61-70 (n=67)	4.48	2.99	2.99	14.93	8.96	2.99		
			71-80 (n=21)	0.00	4.76	0.00	0.00	4.76	4.76		
	SIDER	SIDER	SIDER							5.00	
Stomachpain	SLAM	Trust	Clozapine (n=1760)	1.93	1.76	1.93	4.94	3.52	3.52		
		Age Groups	Under 21 (n=57)	3.51	5.26	5.26	7.02	3.51	5.26		
			21-30 (n=422)	2.84	2.61	2.61	4.98	3.55	4.27		
			31-40 (n=488)	2.25	1.23	1.64	5.94	3.69	3.69		
			41-50 (n=479)	1.46	1.67	1.46	4.18	4.38	3.34		
			51-60 (n=233)	0.43	0.86	0.86	3.00	2.15	1.29		
			61-70 (n=62)	0.00	0.00	3.23	8.06	0.00	6.45		
			71-80 (n=18)	5.56	5.56	5.56	5.56	5.56	0.00		
	Camden & Islington	Trust	Clozapine (n=561)	0.89	1.25	0.89	3.39	2.85	2.14		
		Age Groups	21-30 (n=27)	7.41	0.00	0.00	3.70	7.41	3.70		
			31-40 (n=141)	0.71	2.13	0.71	3.55	5.67	2.84		
			41-50 (n=168)	0.00	0.00	0.60	4.76	2.38	2.98		

Measure Values: 0.00 – 75.44

The results are shown in percentages (%) and broken down by ADRs, Trusts (SLAM, Camden & Islington and Oxford), Cohorts, Sub Cohorts and SIDER reported values.
In Sub Cohort 'Clozapine' represent the total baseline population which further breaks down into age groups.
The columns (Three Months Early, Two Months Early, One Month Early, One Month Later, Two Months Later, Three Months Later) shows the percentages in each monthly interval. The last two columns (SIDER Low End and SIDER High End) shows the SIDER reporting.

Clozapine - Age Groups (%)

ADR	Trust	Cohort	Sub Cohort	Three Months Early	Two Months Early	One Month Early	One Month Later	Two Months Later	Three Months Later	SIDER Low End	SIDER High End
Stomachpain	Camden & Islington	Age Groups	51-60 (n=135)	1.48	2.22	1.48	2.22	0.74	0.74		
			61-70 (n=67)	0.00	1.49	1.49	1.49	1.49	0.00		
			71-80 (n=21)	0.00	0.00	0.00	4.76	0.00	4.76		
	SIDER	SIDER	SIDER								
Diarrhoea	SLAM	Trust	Clozapine (n=1760)	1.08	1.31	1.36	4.72	3.58	2.56		
		Age Groups	Under 21 (n=57)	3.51	5.26	0.00	7.02	3.51	5.26		
			21-30 (n=422)	0.47	0.71	1.18	4.74	2.84	2.84		
			31-40 (n=488)	1.02	1.23	1.43	3.89	3.89	2.87		
			41-50 (n=479)	1.25	1.04	1.67	4.80	3.76	2.09		
			51-60 (n=233)	0.86	2.15	1.29	5.15	3.43	1.29		
			61-70 (n=62)	1.61	0.00	1.61	6.45	4.84	4.84		
			71-80 (n=18)	5.56	5.56	0.00	5.56	5.56	0.00		
	Camden & Islington	Trust	Clozapine (n=561)	0.71	1.25	0.18	3.03	3.39	3.03		
		Age Groups	21-30 (n=27)	0.00	3.70	3.70	3.70	7.41	11.11		
			31-40 (n=141)	2.13	0.71	0.00	4.26	4.96	2.13		
			41-50 (n=168)	0.00	0.60	0.00	1.79	2.38	2.38		
			51-60 (n=135)	0.74	1.48	0.00	2.96	2.22	2.96		
			61-70 (n=67)	0.00	1.49	0.00	2.99	2.99	2.99		
			71-80 (n=21)	0.00	4.76	0.00	4.76	4.76	4.76		
	SIDER	SIDER	SIDER							2.00	
Drymouth	SLAM	Trust	Clozapine (n=1760)	1.08	1.53	1.65	4.66	3.69	2.33		
		Age Groups	Under 21 (n=57)	1.75	3.51	5.26	10.53	5.26	7.02		
			21-30 (n=422)	1.90	1.18	1.90	5.69	5.69	2.84		
			31-40 (n=488)	0.20	1.23	1.43	4.92	1.84	2.46		
			41-50 (n=479)	0.63	1.67	1.67	3.76	2.92	1.88		
			51-60 (n=233)	1.72	2.15	0.86	3.43	4.72	0.86		
			61-70 (n=62)	3.23	0.00	1.61	3.23	6.45	3.23		
			71-80 (n=18)	0.00	5.56	0.00	0.00	0.00	0.00		
	Camden & Islington	Trust	Clozapine (n=561)	1.25	1.25	1.07	3.92	2.14	0.89		
		Age Groups	21-30 (n=27)	0.00	0.00	0.00	7.41	7.41	0.00		
			31-40 (n=141)	1.42	2.13	2.13	4.96	3.55	0.71		
			41-50 (n=168)	1.19	1.19	1.79	4.17	1.19	1.19		
			51-60 (n=135)	0.74	0.00	0.00	2.22	2.22	1.48		
			61-70 (n=67)	2.99	2.99	0.00	4.48	0.00	0.00		
			71-80 (n=21)	0.00	0.00	0.00	0.00	0.00	0.00		
	SIDER	SIDER	SIDER							5.00	20.00
Rash	SLAM	Trust	Clozapine (n=1760)	1.25	1.59	2.05	3.64	2.95	2.27		
		Age Groups	Under 21 (n=57)	0.00	5.26	7.02	5.26	3.51	1.75		
			21-30 (n=422)	0.95	1.18	2.37	4.27	3.32	2.84		
			31-40 (n=488)	1.23	1.43	1.84	3.89	3.07	2.25		
			41-50 (n=479)	1.67	1.46	1.67	2.71	3.13	2.09		
			51-60 (n=233)	0.86	1.72	1.29	3.43	2.15	2.58		
			61-70 (n=62)	1.61	1.61	3.23	3.23	1.61	0.00		
			71-80 (n=18)	5.56	5.56	0.00	5.56	0.00	0.00		
	Camden & Islington	Trust	Clozapine (n=561)	1.25	1.25	0.89	4.28	1.96	2.14		
		Age Groups	21-30 (n=27)	7.41	7.41	3.70	0.00	3.70	7.41		
			31-40 (n=141)	0.00	0.71	1.42	4.26	1.42	2.84		
			41-50 (n=168)	1.79	1.79	0.60	3.57	2.38	1.19		
			51-60 (n=135)	1.48	0.74	0.00	3.70	1.48	0.74		
			61-70 (n=67)	0.00	0.00	1.49	8.96	2.99	2.99		
			71-80 (n=21)	0.00	0.00	0.00	4.76	0.00	4.76		
	SIDER	SIDER	SIDER								
Neutropenia	SLAM	Trust	Clozapine (n=1760)	0.80	0.80	0.74	5.34	2.73	2.61		
		Age Groups	Under 21 (n=57)	3.51	5.26	1.75	12.28	5.26	3.51		
			21-30 (n=422)	1.66	1.42	1.66	8.77	4.27	4.03		
			31-40 (n=488)	0.61	0.82	0.41	3.28	1.84	1.84		
			41-50 (n=479)	0.21	0.21	0.42	3.97	2.51	2.30		
			51-60 (n=233)	0.00	0.00	0.43	5.58	1.72	2.58		
			61-70 (n=62)	1.61	0.00	0.00	3.23	3.23	1.61		
			71-80 (n=18)	0.00	0.00	0.00	0.00	0.00	0.00		
	Camden & Islington	Trust	Clozapine (n=561)	0.00	0.18	0.53	1.60	0.89	1.07		
		Age Groups	21-30 (n=27)	0.00	0.00	0.00	3.70	0.00	0.00		
			31-40 (n=141)	0.00	0.71	0.71	0.71	1.42	2.13		
			41-50 (n=168)	0.00	0.00	0.60	2.38	0.60	0.60		
			51-60 (n=135)	0.00	0.00	0.74	1.48	1.48	0.74		
			61-70 (n=67)	0.00	0.00	0.00	1.49	0.00	1.49		

Measure Values: 0.00 – 75.44

The results are shown in percentages (%) and broken down by ADRs, Trusts (SLAM, Camden & Islington and Oxford), Cohorts, Sub Cohorts and SIDER reported values.
In Sub Cohort 'Clozapine' represent the total baseline population which further breaks down into age groups.
The columns (Three Months Early, Two Months Early, One Month Early, One Month Later, Two Months Later, Three Months Later) shows the percentages in each monthly interval. The last two columns (SIDER Low End and SIDER High End) shows the SIDER reporting.

Clozapine - Age Groups (%)

ADR	Trust	Cohort	Sub Cohort	Three Months Early	Two Months Early	One Month Early	One Month Later	Two Months Later	Three Months Later	SIDER Low End	SIDER High End
Neutropenia	Islington	Age Groups	71-80 (n=21)	0.00	0.00	0.00	0.00	0.00	0.00		
	SIDER	SIDER	SIDER								
Sweating	SLAM	Trust	Clozapine (n=1760)	1.08	0.97	1.36	4.43	4.26	2.84		
		Age Groups	Under 21 (n=57)	3.51	5.26	3.51	10.53	8.77	7.02		
			21-30 (n=422)	0.71	0.95	2.13	3.55	4.27	2.61		
			31-40 (n=488)	1.02	0.61	1.43	4.51	4.51	3.07		
			41-50 (n=479)	1.46	1.04	1.04	5.22	3.55	2.92		
			51-60 (n=233)	0.43	0.43	0.00	3.86	5.15	2.15		
			61-70 (n=62)	0.00	0.00	0.00	1.61	1.61	1.61		
			71-80 (n=18)	5.56	0.00	0.00	0.00	0.00	0.00		
	Camden & Islington	Trust	Clozapine (n=561)	0.53	0.53	0.53	2.85	2.14	1.96		
		Age Groups	21-30 (n=27)	7.41	0.00	0.00	0.00	7.41	0.00		
			31-40 (n=141)	0.00	0.00	0.71	2.13	1.42	2.13		
			41-50 (n=168)	0.00	1.19	0.00	4.17	1.79	1.19		
			51-60 (n=135)	0.74	0.74	1.48	1.48	2.96	2.22		
			61-70 (n=67)	0.00	0.00	0.00	5.97	1.49	4.48		
			71-80 (n=21)	0.00	0.00	0.00	0.00	0.00	0.00		
	SIDER	SIDER	SIDER							6.00	
Dyspepsia	SLAM	Trust	Clozapine (n=1760)	0.74	1.08	0.91	3.92	3.13	3.69		
		Age Groups	Under 21 (n=57)	0.00	1.75	0.00	5.26	1.75	1.75		
			21-30 (n=422)	0.95	0.95	1.42	5.45	3.55	4.03		
			31-40 (n=488)	1.02	1.43	1.02	4.10	3.48	4.71		
			41-50 (n=479)	0.63	0.84	0.42	3.76	3.13	3.13		
			51-60 (n=233)	0.00	0.86	0.43	2.15	2.15	2.15		
			61-70 (n=62)	1.61	0.00	1.61	0.00	3.23	4.84		
			71-80 (n=18)	0.00	5.56	5.56	0.00	0.00	5.56		
	Camden & Islington	Trust	Clozapine (n=561)	0.36	0.53	0.53	4.10	2.67	2.50		
		Age Groups	21-30 (n=27)	0.00	0.00	0.00	3.70	7.41	3.70		
			31-40 (n=141)	0.71	2.13	0.71	4.26	1.42	3.55		
			41-50 (n=168)	0.00	0.00	1.19	6.55	4.76	3.57		
			51-60 (n=135)	0.00	0.00	0.00	2.96	0.00	0.00		
			61-70 (n=67)	1.49	0.00	0.00	1.49	4.48	2.99		
			71-80 (n=21)	0.00	0.00	0.00	0.00	0.00	0.00		
	SIDER	SIDER	SIDER							8.00	14.00
Blurredvision	SLAM	Trust	Clozapine (n=1760)	0.34	0.91	0.63	2.05	1.25	1.02		
		Age Groups	Under 21 (n=57)	0.00	0.00	0.00	3.51	1.75	1.75		
			21-30 (n=422)	0.71	1.18	0.47	2.37	1.90	1.42		
			31-40 (n=488)	0.41	1.64	0.82	3.07	1.43	1.02		
			41-50 (n=479)	0.21	0.21	0.63	1.46	0.42	0.42		
			51-60 (n=233)	0.00	0.43	0.43	0.43	0.86	0.43		
			61-70 (n=62)	0.00	0.00	1.61	1.61	3.23	3.23		
			71-80 (n=18)	0.00	5.56	0.00	0.00	0.00	5.56		
	Camden & Islington	Trust	Clozapine (n=561)	0.89	0.53	0.71	1.25	0.36	0.89		
		Age Groups	21-30 (n=27)	0.00	0.00	0.00	3.70	0.00	0.00		
			31-40 (n=141)	0.71	0.71	1.42	0.00	1.42	0.00		
			41-50 (n=168)	0.00	0.00	0.00	1.79	0.00	0.60		
			51-60 (n=135)	1.48	0.74	0.00	2.22	0.00	1.48		
			61-70 (n=67)	2.99	1.49	2.99	0.00	0.00	2.99		
			71-80 (n=21)	0.00	0.00	0.00	0.00	0.00	0.00		
	SIDER	SIDER	SIDER							5.00	
Akathisia	SLAM	Trust	Clozapine (n=1760)	0.80	0.91	0.74	2.67	1.36	0.80		
		Age Groups	Under 21 (n=57)	0.00	1.75	0.00	1.75	5.26	0.00		
			21-30 (n=422)	0.71	0.95	1.18	4.74	1.90	0.95		
			31-40 (n=488)	1.02	0.82	0.61	2.05	0.61	0.41		
			41-50 (n=479)	0.84	1.25	1.04	2.30	0.84	0.84		
			51-60 (n=233)	0.86	0.43	0.00	2.15	2.15	1.29		
			61-70 (n=62)	0.00	0.00	0.00	0.00	1.61	1.61		
			71-80 (n=18)	0.00	0.00	0.00	0.00	0.00	0.00		
	Camden & Islington	Trust	Clozapine (n=561)	0.00	0.53	0.00	1.25	1.07	0.53		
		Age Groups	21-30 (n=27)	0.00	0.00	0.00	0.00	0.00	0.00		
			31-40 (n=141)	0.00	0.71	0.00	0.71	1.42	0.71		
			41-50 (n=168)	0.00	1.19	0.00	2.38	1.19	1.19		
			51-60 (n=135)	0.00	0.00	0.00	0.74	0.00	0.00		
			61-70 (n=67)	0.00	0.00	0.00	1.49	2.99	0.00		
			71-80 (n=21)	0.00	0.00	0.00	0.00	0.00	0.00		
	SIDER	SIDER	SIDER							3.00	

Measure Values: 0.00 – 75.44

The results are shown in percentages (%) and broken down by ADRs, Trusts (SLAM, Camden & Islington and Oxford), Cohorts, Sub Cohorts and SIDER reported values.
In Sub Cohort 'Clozapine' represent the total baseline population which further breaks down into age groups.
The columns (Three Months Early, Two Months Early, One Month Early, One Month Later, Two Months Later, Three Months Later) shows the percentages in each monthly interval. The last two columns (SIDER Low End and SIDER High End) shows the SIDER reporting.

Supplementary Table B.4 Hospital admissions (%)

Clozapine - Hospital Admission (%)

ADR	Trust	Cohort	Sub Cohort	Three Months Early	Two Months Early	One Month Early	One Month Later	Two Months Later	Three Months Later	SIDER Low End	SIDER High End
Agitation	SLAM	Trust	Clozapine (n=1760)	17.61	22.10	26.53	46.59	32.56	26.99		
		Hospital Admission	Inpatient (n=737)	30.66	40.43	52.65	77.61	55.36	45.59		
			Outpatient (n=1023)	8.21	8.90	7.72	24.24	16.13	13.59		
	Camden & Islington	Trust	Clozapine (n=561)	13.37	17.83	18.36	43.14	28.34	21.03		
		Hospital Admission	Inpatient (n=114)	11.40	16.67	21.93	79.82	51.75	32.46		
			Outpatient (n=447)	13.87	18.12	17.45	33.78	22.37	18.12		
	SIDER	SIDER	SIDER							4.00	
Fatigue	SLAM	Trust	Clozapine (n=1760)	12.67	14.83	15.85	43.58	35.80	30.51		
		Hospital Admission	Inpatient (n=737)	21.71	26.32	30.26	68.66	59.16	50.88		
			Outpatient (n=1023)	6.16	6.55	5.47	25.51	18.96	15.84		
	Camden & Islington	Trust	Clozapine (n=561)	10.34	12.30	13.37	41.18	29.23	26.56		
		Hospital Admission	Inpatient (n=114)	12.28	6.14	11.40	64.04	51.75	43.86		
			Outpatient (n=447)	9.84	13.87	13.87	35.35	23.49	22.15		
	SIDER	SIDER	SIDER								
Sedation	SLAM	Trust	Clozapine (n=1760)	12.67	12.16	14.83	43.86	35.51	29.83		
		Hospital Admission	Inpatient (n=737)	21.71	22.93	29.58	70.69	56.17	49.12		
			Outpatient (n=1023)	6.16	4.40	4.20	24.54	20.63	15.93		
	Camden & Islington	Trust	Clozapine (n=561)	5.17	9.09	9.09	38.15	26.56	21.93		
		Hospital Admission	Inpatient (n=114)	2.63	6.14	9.65	65.79	48.25	30.70		
			Outpatient (n=447)	5.82	9.84	8.95	31.10	21.03	19.69		
	SIDER	SIDER	SIDER							25.00	46.00
Dizziness	SLAM	Trust	Clozapine (n=1760)	2.78	4.20	4.09	16.59	13.13	11.19		
		Hospital Admission	Inpatient (n=737)	4.61	7.87	8.55	28.36	20.35	16.55		
			Outpatient (n=1023)	1.47	1.56	0.88	8.11	7.92	7.33		
	Camden & Islington	Trust	Clozapine (n=561)	3.21	3.39	3.74	18.18	13.73	9.09		
		Hospital Admission	Inpatient (n=114)	2.63	1.75	4.39	33.33	29.82	19.30		
			Outpatient (n=447)	3.36	3.80	3.58	14.32	9.62	6.49		
	SIDER	SIDER	SIDER							12.00	27.00
Confusion	SLAM	Trust	Clozapine (n=1760)	4.72	5.51	6.08	13.92	8.47	6.76		
		Hospital Admission	Inpatient (n=737)	8.28	10.58	13.16	24.15	14.79	11.26		
			Outpatient (n=1023)	2.15	1.86	0.98	6.55	3.91	3.52		
	Camden & Islington	Trust	Clozapine (n=561)	3.57	6.24	5.53	12.66	6.77	5.88		
		Hospital Admission	Inpatient (n=114)	3.51	5.26	4.39	27.19	15.79	7.02		
			Outpatient (n=447)	3.58	6.49	5.82	8.95	4.47	5.59		
	SIDER	SIDER	SIDER							3.00	
Tachycardia	SLAM	Trust	Clozapine (n=1760)	2.27	2.05	2.50	15.40	12.95	9.94		
		Hospital Admission	Inpatient (n=737)	4.34	4.34	5.29	28.22	23.34	16.96		
			Outpatient (n=1023)	0.78	0.39	0.49	6.16	5.47	4.89		
	Camden & Islington	Trust	Clozapine (n=561)	1.43	1.43	0.89	11.23	8.38	6.95		
		Hospital Admission	Inpatient (n=114)	1.75	1.75	0.88	23.68	16.67	11.40		
			Outpatient (n=447)	1.34	1.34	0.89	8.05	6.26	5.82		
	SIDER	SIDER	SIDER							11.00	25.00
Hypersalivation	SLAM	Trust	Clozapine (n=1760)	1.19	1.48	2.10	14.32	13.24	11.31		
		Hospital Admission	Inpatient (n=737)	1.36	2.17	3.66	21.98	21.57	17.23		
			Outpatient (n=1023)	1.08	0.98	0.98	8.80	7.23	7.04		
	Camden & Islington	Trust	Clozapine (n=561)	1.07	1.43	0.53	14.26	6.95	7.66		
		Hospital Admission	Inpatient (n=114)	0.00	1.75	0.00	15.79	11.40	9.65		
			Outpatient (n=447)	1.34	1.34	0.67	13.87	5.82	7.16		
	SIDER	SIDER	SIDER							1.00	48.00
Weightgain	SLAM	Trust	Clozapine (n=1760)	3.75	4.43	5.06	15.34	10.91	10.34		
		Hospital Admission	Inpatient (n=737)	6.65	7.87	10.04	24.15	18.32	18.05		
			Outpatient (n=1023)	1.66	1.96	1.47	8.99	5.57	4.79		
	Camden & Islington	Trust	Clozapine (n=561)	2.50	3.39	1.96	11.76	6.60	6.24		
		Hospital Admission	Inpatient (n=114)	3.51	4.39	1.75	16.67	6.14	10.53		
			Outpatient (n=447)	2.24	3.13	2.01	10.51	6.71	5.15		
	SIDER	SIDER	SIDER							4.00	56.00
Feelingsick	SLAM	Trust	Clozapine (n=1760)	4.66	4.94	6.48	14.32	11.19	9.09		
		Hospital Admission	Inpatient (n=737)	8.55	9.23	13.43	24.02	18.59	14.79		
			Outpatient (n=1023)	1.86	1.86	1.47	7.33	5.87	4.99		
	Camden & Islington	Trust	Clozapine (n=561)	3.74	3.92	3.03	10.52	7.13	7.66		
		Hospital Admission	Inpatient (n=114)	4.39	0.88	1.75	21.05	10.53	11.40		
			Outpatient (n=447)	3.58	4.70	3.36	7.83	6.26	6.71		
	SIDER	SIDER	SIDER								
Constipation	SLAM	Trust	Clozapine (n=1760)	1.76	1.99	2.16	12.27	11.70	9.49		
		Hospital Admission	Inpatient (n=737)	3.12	3.39	4.07	18.86	18.72	13.16		
			Outpatient (n=1023)	0.78	0.98	0.78	7.53	6.65	6.84		
	Camden &	Trust	Clozapine (n=561)	1.07	2.50	1.78	11.41	7.13	5.70		

Measure Values: 0.00 – 79.82

The results are shown in percentages (%) and broken down by ADRs, Trusts (SLAM, Camden & Islington and Oxford), Cohorts, Sub Cohorts and SIDER reported values.
In Sub Cohort 'Clozapine' represent the total baseline population which further breaks down into 'Inpatients' and 'Outpatients'.
The columns (Three Months Early, Two Months Early, One Month Early, One Month Later, Two Months Later, Three Months Later) shows the percentages in each monthly interval. The last two columns (SIDER Low End and SIDER High End) shows the SIDER reporting.

Clozapine - Hospital Admission (%)

ADR	Trust	Cohort	Sub Cohort	Three Months Early	Two Months Early	One Month Early	One Month Later	Two Months Later	Three Months Later	SIDER Low End	SIDER High End
Constipation	Camden & Islington	Hospital Admission	Inpatient (n=114)	0.00	2.63	2.63	22.81	12.28	7.89		
			Outpatient (n=447)	1.34	2.46	1.57	8.50	5.82	5.15		
	SIDER	SIDER	SIDER							10.00	25.00
Headache	SLAM	Trust	Clozapine (n=1760)	4.20	4.55	5.45	12.44	8.18	5.91		
		Hospital Admission	Inpatient (n=737)	6.51	8.01	10.18	20.90	11.94	9.77		
			Outpatient (n=1023)	2.54	2.05	2.05	6.35	5.47	3.13		
	Camden & Islington	Trust	Clozapine (n=561)	2.32	3.57	4.28	9.27	6.42	4.63		
		Hospital Admission	Inpatient (n=114)	0.00	2.63	5.26	17.54	10.53	8.77		
			Outpatient (n=447)	2.91	3.80	4.03	7.16	5.37	3.58		
	SIDER	SIDER	SIDER								
Insomnia	SLAM	Trust	Clozapine (n=1760)	3.92	4.03	5.17	10.40	6.48	4.03		
		Hospital Admission	Inpatient (n=737)	5.70	7.46	9.50	17.50	9.77	5.43		
			Outpatient (n=1023)	2.64	1.56	2.05	5.28	4.11	3.03		
	Camden & Islington	Trust	Clozapine (n=561)	3.57	3.39	3.74	8.91	3.39	4.28		
		Hospital Admission	Inpatient (n=114)	1.75	4.39	7.89	21.05	7.02	5.26		
			Outpatient (n=447)	4.03	3.13	2.68	5.82	2.46	4.03		
	SIDER	SIDER	SIDER							20.00	33.00
Hyperprolactinaemia	SLAM	Trust	Clozapine (n=1760)	3.18	3.64	4.20	8.52	5.06	4.15		
		Hospital Admission	Inpatient (n=737)	5.02	6.38	7.73	13.84	8.41	6.65		
			Outpatient (n=1023)	1.86	1.66	1.66	4.69	2.64	2.35		
	Camden & Islington	Trust	Clozapine (n=561)	1.60	1.78	2.67	8.20	4.10	3.57		
		Hospital Admission	Inpatient (n=114)	2.63	1.75	1.75	15.79	5.26	3.51		
			Outpatient (n=447)	1.34	1.79	2.91	6.26	3.80	3.58		
	SIDER	SIDER	SIDER								
Hypertension	SLAM	Trust	Clozapine (n=1760)	2.05	2.22	3.13	9.15	5.74	4.60		
		Hospital Admission	Inpatient (n=737)	3.53	4.21	5.97	13.03	8.96	7.19		
			Outpatient (n=1023)	0.98	0.78	1.08	6.35	3.42	2.74		
	Camden & Islington	Trust	Clozapine (n=561)	0.71	0.71	1.60	7.13	4.63	2.67		
		Hospital Admission	Inpatient (n=114)	0.00	0.00	2.63	13.16	7.89	4.39		
			Outpatient (n=447)	0.89	0.89	1.34	5.59	3.80	2.24		
	SIDER	SIDER	SIDER							4.00	12.00
Vomiting	SLAM	Trust	Clozapine (n=1760)	2.56	2.50	3.01	8.86	6.82	5.00		
		Hospital Admission	Inpatient (n=737)	4.21	4.61	5.97	12.62	11.26	8.68		
			Outpatient (n=1023)	1.37	0.98	0.88	6.16	3.62	2.35		
	Camden & Islington	Trust	Clozapine (n=561)	2.14	2.50	2.85	6.77	4.99	4.63		
		Hospital Admission	Inpatient (n=114)	1.75	0.88	2.63	11.40	6.14	6.14		
			Outpatient (n=447)	2.24	2.91	2.91	5.59	4.70	4.25		
	SIDER	SIDER	SIDER							3.00	17.00
Shaking	SLAM	Trust	Clozapine (n=1760)	3.13	2.95	3.92	9.55	5.40	5.06		
		Hospital Admission	Inpatient (n=737)	6.11	5.43	8.14	16.82	9.77	8.68		
			Outpatient (n=1023)	0.98	1.17	0.88	4.30	2.25	2.44		
	Camden & Islington	Trust	Clozapine (n=561)	1.78	1.96	3.74	6.06	3.92	2.85		
		Hospital Admission	Inpatient (n=114)	0.00	0.88	3.51	7.02	5.26	6.14		
			Outpatient (n=447)	2.24	2.24	3.80	5.82	3.58	2.01		
	SIDER	SIDER	SIDER								
Abdominalpain	SLAM	Trust	Clozapine (n=1760)	1.88	1.99	2.56	8.01	6.02	4.72		
		Hospital Admission	Inpatient (n=737)	3.66	3.93	5.29	14.38	8.82	7.73		
			Outpatient (n=1023)	0.59	0.59	0.59	3.42	4.01	2.54		
	Camden & Islington	Trust	Clozapine (n=561)	0.89	0.89	1.60	3.92	3.57	3.39		
		Hospital Admission	Inpatient (n=114)	0.00	0.88	2.63	6.14	9.65	4.39		
			Outpatient (n=447)	0.67	0.89	1.34	3.36	2.01	3.13		
	SIDER	SIDER	SIDER							4.00	
Fever	SLAM	Trust	Clozapine (n=1760)	1.02	1.14	1.65	6.36	4.43	3.13		
		Hospital Admission	Inpatient (n=737)	2.04	2.44	3.26	11.13	6.51	5.70		
			Outpatient (n=1023)	0.29	0.20	0.49	2.93	2.93	1.27		
	Camden & Islington	Trust	Clozapine (n=561)	0.89	0.89	0.53	3.74	2.67	0.89		
		Hospital Admission	Inpatient (n=114)	0.88	0.00	0.00	11.40	4.39	0.88		
			Outpatient (n=447)	0.89	1.12	0.67	1.79	2.24	0.89		
	SIDER	SIDER	SIDER							4.00	13.00
Backache	SLAM	Trust	Clozapine (n=1760)	1.14	1.59	2.44	4.94	3.35	2.73		
		Hospital Admission	Inpatient (n=737)	2.31	3.26	5.16	9.09	6.24	4.48		
			Outpatient (n=1023)	0.29	0.39	0.49	1.96	1.27	1.47		
	Camden & Islington	Trust	Clozapine (n=561)	1.43	1.25	1.96	5.35	3.03	3.03		
		Hospital Admission	Inpatient (n=114)	2.63	1.75	2.63	12.28	8.77	4.39		
			Outpatient (n=447)	1.12	1.12	1.79	3.58	1.57	2.68		
	SIDER	SIDER	SIDER							5.00	
Nausea	SLAM	Trust	Clozapine (n=1760)	1.14	1.08	1.19	6.08	5.23	3.69		

Measure Values: 0.00 – 79.82

The results are shown in percentages (%) and broken down by ADRs, Trusts (SLAM, Camden & Islington and Oxford), Cohorts, Sub Cohorts and SIDER reported values.
In Sub Cohort 'Clozapine' represent the total baseline population which further breaks down into 'Inpatients' and 'Outpatients'.
The columns (Three Months Early, Two Months Early, One Month Early, One Month Later, Two Months Later, Three Months Later) shows the percentages in each monthly interval. The last two columns (SIDER Low End and SIDER High End) shows the SIDER reporting.

Clozapine - Hospital Admission (%)

ADR	Trust	Cohort	Sub Cohort	Three Months Early	Two Months Early	One Month Early	One Month Later	Two Months Later	Three Months Later	SIDER Low End	SIDER High End
Nausea	SLAM	Hospital Admission	Inpatient (n=737)	1.90	1.90	2.17	9.50	7.73	5.02		
			Outpatient (n=1023)	0.59	0.49	0.49	3.62	3.42	2.74		
	Camden & Islington	Trust	Clozapine (n=561)	0.89	1.43	0.36	4.63	3.57	3.57		
		Hospital Admission	Inpatient (n=114)	0.88	0.88	0.00	9.65	5.26	2.63		
			Outpatient (n=447)	0.89	1.57	0.45	3.36	3.13	3.80		
	SIDER	SIDER	SIDER							3.00	17.00
Tremor	SLAM	Trust	Clozapine (n=1760)	1.48	1.99	2.95	5.51	3.52	3.47		
		Hospital Admission	Inpatient (n=737)	2.44	3.80	6.11	9.36	6.38	6.24		
			Outpatient (n=1023)	0.78	0.68	0.68	2.74	1.47	1.47		
	Camden & Islington	Trust	Clozapine (n=561)	1.60	1.78	2.14	3.92	1.96	2.14		
		Hospital Admission	Inpatient (n=114)	0.88	1.75	2.63	8.77	3.51	4.39		
			Outpatient (n=447)	1.79	1.79	2.01	2.68	1.57	1.57		
	SIDER	SIDER	SIDER							6.00	
Convulsion	SLAM	Trust	Clozapine (n=1760)	1.36	1.70	1.82	7.05	4.94	4.03		
		Hospital Admission	Inpatient (n=737)	2.71	3.39	3.80	12.62	8.14	7.19		
			Outpatient (n=1023)	0.39	0.49	0.39	3.03	2.64	1.76		
	Camden & Islington	Trust	Clozapine (n=561)	0.53	0.53	0.36	2.85	2.14	1.07		
		Hospital Admission	Inpatient (n=114)	0.00	0.00	0.00	3.51	4.39	0.88		
			Outpatient (n=447)	0.67	0.67	0.45	2.68	1.57	1.12		
	SIDER	SIDER	SIDER							3.00	
Hypotension	SLAM	Trust	Clozapine (n=1760)	0.51	0.97	0.80	5.00	2.95	2.56		
		Hospital Admission	Inpatient (n=737)	0.81	1.76	1.90	9.77	5.16	3.93		
			Outpatient (n=1023)	0.29	0.39	0.00	1.56	1.37	1.56		
	Camden & Islington	Trust	Clozapine (n=561)	0.18	0.53	0.18	3.57	2.32	1.78		
		Hospital Admission	Inpatient (n=114)	0.00	0.88	0.00	6.14	6.14	1.75		
			Outpatient (n=447)	0.22	0.45	0.22	2.91	1.34	1.79		
	SIDER	SIDER	SIDER							9.00	38.00
Enuresis	SLAM	Trust	Clozapine (n=1760)	1.02	0.80	1.25	4.20	3.92	3.24		
		Hospital Admission	Inpatient (n=737)	1.76	1.22	2.58	8.14	6.78	5.43		
			Outpatient (n=1023)	0.49	0.49	0.29	1.37	1.86	1.66		
	Camden & Islington	Trust	Clozapine (n=561)	0.71	1.07	1.07	4.10	1.43	1.25		
		Hospital Admission	Inpatient (n=114)	0.88	0.88	3.51	7.89	2.63	3.51		
			Outpatient (n=447)	0.67	1.12	0.45	3.13	1.12	0.67		
	SIDER	SIDER	SIDER								
Drymouth	SLAM	Trust	Clozapine (n=1760)	1.08	1.53	1.65	4.66	3.69	2.33		
		Hospital Admission	Inpatient (n=737)	1.76	2.71	3.12	7.46	5.97	3.53		
			Outpatient (n=1023)	0.59	0.68	0.59	2.64	2.05	1.47		
	Camden & Islington	Trust	Clozapine (n=561)	1.25	1.25	1.07	3.92	2.14	0.89		
		Hospital Admission	Inpatient (n=114)	0.88	0.00	0.88	7.02	3.51	1.75		
			Outpatient (n=447)	1.34	1.57	1.12	3.13	1.79	0.67		
	SIDER	SIDER	SIDER							5.00	20.00
Stomachpain	SLAM	Trust	Clozapine (n=1760)	1.93	1.76	1.93	4.94	3.52	3.52		
		Hospital Admission	Inpatient (n=737)	3.53	3.53	4.07	8.68	6.11	5.56		
			Outpatient (n=1023)	0.78	0.49	0.39	2.25	1.66	2.05		
	Camden & Islington	Trust	Clozapine (n=561)	0.89	1.25	0.89	3.39	2.85	2.14		
		Hospital Admission	Inpatient (n=114)	0.00	1.75	0.88	7.02	6.14	2.63		
			Outpatient (n=447)	1.12	1.12	0.89	2.46	2.01	2.01		
	SIDER	SIDER	SIDER								
Dyspepsia	SLAM	Trust	Clozapine (n=1760)	0.74	1.08	0.91	3.92	3.13	3.69		
		Hospital Admission	Inpatient (n=737)	1.22	2.04	2.04	7.06	4.34	6.38		
			Outpatient (n=1023)	0.39	0.39	0.10	1.66	2.25	1.76		
	Camden & Islington	Trust	Clozapine (n=561)	0.36	0.53	0.53	4.10	2.67	2.50		
		Hospital Admission	Inpatient (n=114)	0.00	0.88	0.88	7.89	4.39	7.02		
			Outpatient (n=447)	0.45	0.45	0.45	3.13	2.24	1.34		
	SIDER	SIDER	SIDER							8.00	14.00
Diarrhoea	SLAM	Trust	Clozapine (n=1760)	1.08	1.31	1.36	4.72	3.58	2.56		
		Hospital Admission	Inpatient (n=737)	2.17	2.58	2.85	8.41	5.97	4.75		
			Outpatient (n=1023)	0.29	0.39	0.29	2.05	1.86	0.98		
	Camden & Islington	Trust	Clozapine (n=561)	0.71	1.25	0.18	3.03	3.39	3.03		
		Hospital Admission	Inpatient (n=114)	1.75	1.75	0.00	6.14	6.14	1.75		
			Outpatient (n=447)	0.45	1.12	0.22	2.24	2.68	3.36		
	SIDER	SIDER	SIDER							2.00	
Rash	SLAM	Trust	Clozapine (n=1760)	1.25	1.59	2.05	3.64	2.95	2.27		
		Hospital Admission	Inpatient (n=737)	2.71	3.53	4.48	6.24	5.43	4.88		
			Outpatient (n=1023)	0.20	0.20	0.29	1.76	1.17	0.39		
	Camden & Islington	Trust	Clozapine (n=561)	1.25	1.25	0.89	4.28	1.96	2.14		
		Hospital	Inpatient (n=114)	0.00	1.75	1.75	7.02	3.51	2.63		

Measure Values: 0.00 – 79.82

The results are shown in percentages (%) and broken down by ADRs, Trusts (SLAM, Camden & Islington and Oxford), Cohorts, Sub Cohorts and SIDER reported values.
In Sub Cohort 'Clozapine' represent the total baseline population which further breaks down into 'Inpatients' and 'Outpatients'.
The columns (Three Months Early, Two Months Early, One Month Early, One Month Later, Two Months Later, Three Months Later) shows the percentages in each monthly interval. The last two columns (SIDER Low End and SIDER High End) shows the SIDER reporting.

Clozapine - Hospital Admission (%)

ADR	Trust	Cohort	Sub Cohort	Three Months Early	Two Months Early	One Month Early	One Month Later	Two Months Later	Three Months Later	SIDER Low End	SIDER High End
Rash	Islington	Admission	Outpatient (n=447)	1.57	1.12	0.67	3.58	1.57	2.01		
	SIDER	SIDER	SIDER								
Sweating	SLAM	Trust	Clozapine (n=1760)	1.08	0.97	1.36	4.43	4.26	2.84		
		Hospital Admission	Inpatient (n=737)	1.90	2.04	2.85	7.87	5.97	4.61		
			Outpatient (n=1023)	0.49	0.20	0.29	1.96	3.03	1.56		
	Camden & Islington	Trust	Clozapine (n=561)	0.53	0.53	0.53	2.85	2.14	1.96		
		Hospital Admission	Inpatient (n=114)	0.00	0.00	0.88	6.14	4.39	4.39		
			Outpatient (n=447)	0.67	0.67	0.45	2.01	1.57	1.34		
	SIDER	SIDER	SIDER							6.00	
Neutropenia	SLAM	Trust	Clozapine (n=1760)	0.80	0.80	0.74	5.34	2.73	2.61		
		Hospital Admission	Inpatient (n=737)	1.09	1.63	1.63	8.68	4.07	4.34		
			Outpatient (n=1023)	0.59	0.20	0.10	2.93	1.76	1.37		
	Camden & Islington	Trust	Clozapine (n=561)	0.00	0.18	0.53	1.60	0.89	1.07		
		Hospital Admission	Inpatient (n=114)	0.00	0.00	0.00	2.63	0.00	0.88		
			Outpatient (n=447)	0.00	0.22	0.67	1.34	1.12	1.12		
	SIDER	SIDER	SIDER								
Akathisia	SLAM	Trust	Clozapine (n=1760)	0.80	0.91	0.74	2.67	1.36	0.80		
		Hospital Admission	Inpatient (n=737)	1.09	1.49	1.36	4.75	2.58	1.36		
			Outpatient (n=1023)	0.59	0.49	0.29	1.17	0.49	0.39		
	Camden & Islington	Trust	Clozapine (n=561)	0.00	0.53	0.00	1.25	1.07	0.53		
		Hospital Admission	Inpatient (n=114)	0.00	0.00	0.00	0.88	4.39	1.75		
			Outpatient (n=447)	0.00	0.67	0.00	1.34	0.22	0.22		
	SIDER	SIDER	SIDER							3.00	
Blurredvision	SLAM	Trust	Clozapine (n=1760)	0.34	0.91	0.63	2.05	1.25	1.02		
		Hospital Admission	Inpatient (n=737)	0.41	1.22	1.22	3.39	2.31	1.49		
			Outpatient (n=1023)	0.29	0.68	0.20	1.08	0.49	0.68		
	Camden & Islington	Trust	Clozapine (n=561)	0.89	0.53	0.71	1.25	0.36	0.89		
		Hospital Admission	Inpatient (n=114)	0.00	0.00	1.75	2.63	0.88	0.00		
			Outpatient (n=447)	1.12	0.67	0.45	0.89	0.22	1.12		
	SIDER	SIDER	SIDER							5.00	

Measure Values
0.00 79.82

The results are shown in percentages (%) and broken down by ADRs, Trusts (SLAM, Camden & Islington and Oxford), Cohorts, Sub Cohorts and SIDER reported values.
In Sub Cohort 'Clozapine' represent the total baseline population which further breaks down into 'Inpatients' and 'Outpatients'.
The columns (Three Months Early, Two Months Early, One Month Early, One Month Later, Two Months Later, Three Months Later) shows the percentages in each monthly interval. The last two columns (SIDER Low End and SIDER High End) shows the SIDER reporting.

Supplementary Table B.5 Smoking status (%)

Clozapine - Smoking Status (%)

ADR	Trust	Cohort	Sub Cohort	Three Months Early	Two Months Early	One Month Early	One Month Later	Two Months Later	Three Months Later	SIDER Low End	SIDER High End
Agitation	SLAM	Trust	Clozapine (n=1760)	17.61	22.10	26.53	46.59	32.56	26.99		
		Smoking Status	Smoker (n=1039)	24.35	30.13	36.67	63.04	44.47	38.98		
			Non Smoker (n=721)	7.91	10.54	11.93	22.88	15.40	9.71		
	Camden & Islington	Trust	Clozapine (n=561)	13.37	17.83	18.36	43.14	28.34	21.03		
		Smoking Status	Smoker (n=360)	0.83	24.72	25.56	56.94	37.78	28.33		
			Non Smoker (n=201)	1.49	5.47	5.47	18.41	11.44	7.96		
Fatigue	SLAM	Trust	Clozapine (n=1760)	12.67	14.83	15.85	43.58	35.80	30.51		
		Smoking Status	Smoker (n=1039)	16.65	19.54	21.37	56.98	47.16	41.67		
			Non Smoker (n=721)	6.93	8.04	7.91	24.27	19.42	14.42		
	Camden & Islington	Trust	Clozapine (n=561)	10.34	12.30	13.37	41.18	29.23	26.56		
		Smoking Status	Smoker (n=360)	0.28	16.67	17.78	51.94	39.17	35.83		
			Non Smoker (n=201)	0.50	4.48	5.47	21.89	11.44	9.95		
Sedation	SLAM	Trust	Clozapine (n=1760)	12.67	12.16	14.83	43.86	35.51	29.83		
		Smoking Status	Smoker (n=1039)	16.84	16.46	20.50	56.98	46.29	39.36		
			Non Smoker (n=721)	6.66	5.96	6.66	24.97	19.97	16.09		
	Camden & Islington	Trust	Clozapine (n=561)	5.17	9.09	9.09	38.15	26.56	21.93		
		Smoking Status	Smoker (n=360)	0.00	11.67	12.78	45.56	34.17	28.33		
			Non Smoker (n=201)	0.00	4.48	2.49	24.88	12.94	10.45		
Dizziness	SLAM	Trust	Clozapine (n=1760)	2.78	4.20	4.09	16.59	13.13	11.19		
		Smoking Status	Smoker (n=1039)	3.66	5.87	5.49	20.98	16.55	14.63		
			Non Smoker (n=721)	1.53	1.80	2.08	10.26	8.18	6.24		
	Camden & Islington	Trust	Clozapine (n=561)	3.21	3.39	3.74	18.18	13.73	9.09		
		Smoking Status	Smoker (n=360)	0.28	4.44	5.00	23.33	18.06	11.94		
			Non Smoker (n=201)	0.50	1.49	1.49	8.96	5.97	3.98		
Hypersalivation	SLAM	Trust	Clozapine (n=1760)	1.19	1.48	2.10	14.32	13.24	11.31		
		Smoking Status	Smoker (n=1039)	1.06	1.92	2.50	16.84	15.50	13.47		
			Non Smoker (n=721)	1.39	0.83	1.53	10.68	9.99	8.18		
	Camden & Islington	Trust	Clozapine (n=561)	1.07	1.43	0.53	14.26	6.95	7.66		
		Smoking Status	Smoker (n=360)	0.00	1.67	0.56	15.83	8.33	9.72		
			Non Smoker (n=201)	0.00	1.00	0.50	11.44	4.48	3.98		
Weightgain	SLAM	Trust	Clozapine (n=1760)	3.75	4.43	5.06	15.34	10.91	10.34		
		Smoking Status	Smoker (n=1039)	4.72	5.58	6.26	18.96	14.73	13.09		
			Non Smoker (n=721)	2.36	2.77	3.33	10.12	5.41	6.38		
	Camden & Islington	Trust	Clozapine (n=561)	2.50	3.39	1.96	11.76	6.60	6.24		
		Smoking Status	Smoker (n=360)	0.00	4.72	2.78	13.89	8.89	8.61		
			Non Smoker (n=201)	0.00	1.00	0.50	7.96	2.49	1.99		
Tachycardia	SLAM	Trust	Clozapine (n=1760)	2.27	2.05	2.50	15.40	12.95	9.94		
		Smoking Status	Smoker (n=1039)	2.89	2.60	3.08	20.40	17.32	12.90		
			Non Smoker (n=721)	1.39	1.25	1.66	8.18	6.66	5.69		
	Camden & Islington	Trust	Clozapine (n=561)	1.43	1.43	0.89	11.23	8.38	6.95		
		Smoking Status	Smoker (n=360)	0.00	1.39	1.39	14.72	10.00	8.61		
			Non Smoker (n=201)	0.00	1.49	0.00	4.98	5.47	3.98		
Confusion	SLAM	Trust	Clozapine (n=1760)	4.72	5.51	6.08	13.92	8.47	6.76		
		Smoking Status	Smoker (n=1039)	6.16	7.70	8.85	19.25	11.16	9.72		
			Non Smoker (n=721)	2.64	2.36	2.08	6.24	4.58	2.50		
	Camden & Islington	Trust	Clozapine (n=561)	3.57	6.24	5.53	12.66	6.77	5.88		
		Smoking Status	Smoker (n=360)	0.28	9.44	8.06	17.22	8.89	8.61		
			Non Smoker (n=201)	0.50	0.50	1.00	4.48	2.99	1.00		
Feelingsick	SLAM	Trust	Clozapine (n=1760)	4.66	4.94	6.48	14.32	11.19	9.09		
		Smoking Status	Smoker (n=1039)	6.06	6.64	8.37	18.29	14.73	11.65		
			Non Smoker (n=721)	2.64	2.50	3.74	8.60	6.10	5.41		
	Camden & Islington	Trust	Clozapine (n=561)	3.74	3.92	3.03	10.52	7.13	7.66		
		Smoking Status	Smoker (n=360)	0.00	5.56	4.72	14.44	9.72	10.56		
			Non Smoker (n=201)	0.00	1.00	0.00	3.48	2.49	2.49		
Constipation	SLAM	Trust	Clozapine (n=1760)	1.76	1.99	2.16	12.27	11.70	9.49		
		Smoking Status	Smoker (n=1039)	2.21	2.50	2.79	14.63	15.11	11.26		
			Non Smoker (n=721)	1.11	1.25	1.25	8.88	6.80	6.93		
	Camden & Islington	Trust	Clozapine (n=561)	1.07	2.50	1.78	11.41	7.13	5.70		
		Smoking Status	Smoker (n=360)	0.28	3.61	2.22	13.61	8.61	7.50		
			Non Smoker (n=201)	0.50	0.50	1.00	7.46	4.48	2.49		
Headache	SLAM	Trust	Clozapine (n=1760)	4.20	4.55	5.45	12.44	8.18	5.91		
		Smoking Status	Smoker (n=1039)	5.58	5.87	7.12	17.04	11.16	8.95		
			Non Smoker (n=721)	2.22	2.64	3.05	5.83	3.88	1.53		
	Camden & Islington	Trust	Clozapine (n=561)	2.32	3.57	4.28	9.27	6.42	4.63		
		Smoking Status	Smoker (n=360)	0.00	5.28	6.11	12.22	8.06	6.11		
			Non Smoker (n=201)	0.00	0.50	1.00	3.98	3.48	1.99		
Insomnia	SLAM	Trust	Clozapine (n=1760)	3.92	4.03	5.17	10.40	6.48	4.03		

Measure Values
0.00 63.04

The results are shown in percentages (%) and broken down by ADRs, Trusts (SLAM, Camden & Islington and Oxford), Cohorts, Sub Cohorts and SIDER reported values.
In Sub Cohort 'Clozapine' represent the total baseline population which further breaks down into 'Smokers and 'Non-Smokers'.
The columns (Three Months Early, Two Months Early, One Month Early, One Month Later, Two Months Later, Three Months Later) shows the percentages in each monthly interval. The last two columns (SIDER Low End and SIDER High End) shows the SIDER reporting.

Clozapine - Smoking Status (%)

ADR	Trust	Cohort	Sub Cohort	Three Months Early	Two Months Early	One Month Early	One Month Later	Two Months Later	Three Months Later	SIDER Low End	SIDER High End
Insomnia	SLAM	Smoking Status	Smoker (n=1039)	4.72	5.29	6.35	14.15	8.47	5.29		
			Non Smoker (n=721)	2.77	2.22	3.47	4.99	3.61	2.22		
	Camden & Islington	Trust	Clozapine (n=561)	3.57	3.39	3.74	8.91	3.39	4.28		
		Smoking Status	Smoker (n=360)	0.56	4.44	5.28	12.22	4.72	5.83		
			Non Smoker (n=201)	1.00	1.49	1.00	2.99	1.00	1.49		
Hyperprolactinaemia	SLAM	Trust	Clozapine (n=1760)	3.18	3.64	4.20	8.52	5.06	4.15		
		Smoking Status	Smoker (n=1039)	3.46	4.04	4.81	9.14	5.39	4.72		
			Non Smoker (n=721)	2.77	3.05	3.33	7.63	4.58	3.33		
	Camden & Islington	Trust	Clozapine (n=561)	1.60	1.78	2.67	8.20	4.10	3.57		
		Smoking Status	Smoker (n=360)	0.00	2.22	3.06	10.28	4.72	3.06		
			Non Smoker (n=201)	0.00	1.00	1.99	4.48	2.99	4.48		
Hypertension	SLAM	Trust	Clozapine (n=1760)	2.05	2.22	3.13	9.15	5.74	4.60		
		Smoking Status	Smoker (n=1039)	2.41	2.79	4.23	11.16	7.41	6.35		
			Non Smoker (n=721)	1.53	1.39	1.53	6.24	3.33	2.08		
	Camden & Islington	Trust	Clozapine (n=561)	0.71	0.71	1.60	7.13	4.63	2.67		
		Smoking Status	Smoker (n=360)	0.28	0.56	1.94	7.78	6.11	3.06		
			Non Smoker (n=201)	0.50	1.00	1.00	5.97	1.99	1.99		
Shaking	SLAM	Trust	Clozapine (n=1760)	3.13	2.95	3.92	9.55	5.40	5.06		
		Smoking Status	Smoker (n=1039)	3.27	3.27	5.10	12.51	7.41	7.41		
			Non Smoker (n=721)	2.91	2.50	2.22	5.27	2.50	1.66		
	Camden & Islington	Trust	Clozapine (n=561)	1.78	1.96	3.74	6.06	3.92	2.85		
		Smoking Status	Smoker (n=360)	0.00	2.78	4.72	7.50	4.72	3.89		
			Non Smoker (n=201)	0.00	0.50	1.99	3.48	2.49	1.00		
Vomiting	SLAM	Trust	Clozapine (n=1760)	2.56	2.50	3.01	8.86	6.82	5.00		
		Smoking Status	Smoker (n=1039)	3.18	3.27	3.95	11.16	9.14	6.35		
			Non Smoker (n=721)	1.66	1.39	1.66	5.55	3.47	3.05		
	Camden & Islington	Trust	Clozapine (n=561)	2.14	2.50	2.85	6.77	4.99	4.63		
		Smoking Status	Smoker (n=360)	0.00	3.06	3.33	9.44	6.39	5.83		
			Non Smoker (n=201)	0.00	1.49	1.99	1.99	2.49	2.49		
Abdominalpain	SLAM	Trust	Clozapine (n=1760)	1.88	1.99	2.56	8.01	6.02	4.72		
		Smoking Status	Smoker (n=1039)	2.31	2.60	3.66	10.78	8.18	6.93		
			Non Smoker (n=721)	1.25	1.11	0.97	4.02	2.91	1.53		
	Camden & Islington	Trust	Clozapine (n=561)	0.89	0.89	1.60	3.92	3.57	3.39		
		Smoking Status	Smoker (n=360)	0.00	1.39	2.22	4.72	4.72	4.44		
			Non Smoker (n=201)	0.00	0.00	0.50	2.49	1.49	1.49		
Nausea	SLAM	Trust	Clozapine (n=1760)	1.14	1.08	1.19	6.08	5.23	3.69		
		Smoking Status	Smoker (n=1039)	1.15	0.96	1.64	7.51	5.97	4.62		
			Non Smoker (n=721)	1.11	1.25	0.55	4.02	4.16	2.36		
	Camden & Islington	Trust	Clozapine (n=561)	0.89	1.43	0.36	4.63	3.57	3.57		
		Smoking Status	Smoker (n=360)	0.00	1.94	0.00	5.83	4.44	3.61		
			Non Smoker (n=201)	0.00	0.50	1.00	2.49	1.99	3.48		
Backache	SLAM	Trust	Clozapine (n=1760)	1.14	1.59	2.44	4.94	3.35	2.73		
		Smoking Status	Smoker (n=1039)	1.64	2.21	3.46	6.64	4.91	3.46		
			Non Smoker (n=721)	0.42	0.69	0.97	2.50	1.11	1.66		
	Camden & Islington	Trust	Clozapine (n=561)	1.43	1.25	1.96	5.35	3.03	3.03		
		Smoking Status	Smoker (n=360)	0.00	1.94	3.06	7.50	3.89	4.17		
			Non Smoker (n=201)	0.00	0.00	0.00	1.49	1.49	1.00		
Convulsion	SLAM	Trust	Clozapine (n=1760)	1.36	1.70	1.82	7.05	4.94	4.03		
		Smoking Status	Smoker (n=1039)	1.83	2.12	2.41	9.53	6.35	4.72		
			Non Smoker (n=721)	0.69	1.11	0.97	3.47	2.91	3.05		
	Camden & Islington	Trust	Clozapine (n=561)	0.53	0.53	0.36	2.85	2.14	1.07		
		Smoking Status	Smoker (n=360)	0.00	0.83	0.56	3.33	3.06	1.39		
			Non Smoker (n=201)	0.00	0.00	0.00	1.99	0.50	0.50		
Fever	SLAM	Trust	Clozapine (n=1760)	1.02	1.14	1.65	6.36	4.43	3.13		
		Smoking Status	Smoker (n=1039)	1.25	1.54	2.21	8.47	6.16	4.43		
			Non Smoker (n=721)	0.69	0.55	0.83	3.33	1.94	1.25		
	Camden & Islington	Trust	Clozapine (n=561)	0.89	0.89	0.53	3.74	2.67	0.89		
		Smoking Status	Smoker (n=360)	0.00	1.11	0.83	5.28	3.89	1.11		
			Non Smoker (n=201)	0.00	0.50	0.00	1.00	0.50	0.50		
Tremor	SLAM	Trust	Clozapine (n=1760)	1.48	1.99	2.95	5.51	3.52	3.47		
		Smoking Status	Smoker (n=1039)	1.44	2.21	3.66	6.54	4.43	4.62		
			Non Smoker (n=721)	1.53	1.66	1.94	4.02	2.22	1.80		
	Camden & Islington	Trust	Clozapine (n=561)	1.60	1.78	2.14	3.92	1.96	2.14		
		Smoking Status	Smoker (n=360)	0.56	2.50	3.06	4.72	2.22	3.33		
			Non Smoker (n=201)	1.00	0.50	0.50	2.49	1.49	0.00		
Hypotension	SLAM	Trust	Clozapine (n=1760)	0.51	0.97	0.80	5.00	2.95	2.56		
		Smoking	Smoker (n=1039)	0.67	1.15	1.06	6.16	3.85	3.08		

Measure Values: 0.00 – 63.04

The results are shown in percentages (%) and broken down by ADRs, Trusts (SLAM, Camden & Islington and Oxford), Cohorts, Sub Cohorts and SIDER reported values.
In Sub Cohort 'Clozapine' represent the total baseline population which further breaks down into 'Smokers and 'Non-Smokers'.
The columns (Three Months Early, Two Months Early, One Month Early, One Month Later, Two Months Later, Three Months Later) shows the percentages in each monthly interval. The last two columns (SIDER Low End and SIDER High End) shows the SIDER reporting.

Clozapine - Smoking Status (%)

ADR	Trust	Cohort	Sub Cohort	Three Months Early	Two Months Early	One Month Early	One Month Later	Two Months Later	Three Months Later	SIDER Low End	SIDER High End
Hypotension	SLAM	Status	Non Smoker (n=721)	0.28	0.69	0.42	3.33	1.66	1.80		
	Camden & Islington	Trust	Clozapine (n=561)	0.18	0.53	0.18	3.57	2.32	1.78		
		Smoking Status	Smoker (n=360)	0.00	0.83	0.28	4.17	2.50	2.22		
			Non Smoker (n=201)	0.00	0.00	0.00	2.49	1.99	1.00		
Drymouth	SLAM	Trust	Clozapine (n=1760)	1.08	1.53	1.65	4.66	3.69	2.33		
		Smoking Status	Smoker (n=1039)	1.06	1.92	2.21	5.77	4.72	3.56		
			Non Smoker (n=721)	1.11	0.97	0.83	3.05	2.22	0.55		
	Camden & Islington	Trust	Clozapine (n=561)	1.25	1.25	1.07	3.92	2.14	0.89		
		Smoking Status	Smoker (n=360)	0.00	1.11	1.67	5.00	2.78	0.83		
			Non Smoker (n=201)	0.00	1.49	0.00	1.99	1.00	1.00		
Enuresis	SLAM	Trust	Clozapine (n=1760)	1.02	0.80	1.25	4.20	3.92	3.24		
		Smoking Status	Smoker (n=1039)	1.15	1.15	1.83	6.26	5.10	4.14		
			Non Smoker (n=721)	0.83	0.28	0.42	1.25	2.22	1.94		
	Camden & Islington	Trust	Clozapine (n=561)	0.71	1.07	1.07	4.10	1.43	1.25		
		Smoking Status	Smoker (n=360)	0.00	1.39	1.11	5.28	1.39	1.39		
			Non Smoker (n=201)	0.00	0.50	1.00	1.99	1.49	1.00		
Stomachpain	SLAM	Trust	Clozapine (n=1760)	1.93	1.76	1.93	4.94	3.52	3.52		
		Smoking Status	Smoker (n=1039)	2.60	2.12	2.69	7.31	5.00	4.52		
			Non Smoker (n=721)	0.97	1.25	0.83	1.53	1.39	2.08		
	Camden & Islington	Trust	Clozapine (n=561)	0.89	1.25	0.89	3.39	2.85	2.14		
		Smoking Status	Smoker (n=360)	0.00	1.94	1.39	5.00	3.89	3.06		
			Non Smoker (n=201)	0.00	0.00	0.00	0.50	1.00	0.50		
Rash	SLAM	Trust	Clozapine (n=1760)	1.25	1.59	2.05	3.64	2.95	2.27		
		Smoking Status	Smoker (n=1039)	1.92	2.41	2.60	4.52	4.14	3.27		
			Non Smoker (n=721)	0.28	0.42	1.25	2.36	1.25	0.83		
	Camden & Islington	Trust	Clozapine (n=561)	1.25	1.25	0.89	4.28	1.96	2.14		
		Smoking Status	Smoker (n=360)	0.00	1.94	1.39	5.83	3.06	2.78		
			Non Smoker (n=201)	0.00	0.00	0.00	1.49	0.00	1.00		
Diarrhoea	SLAM	Trust	Clozapine (n=1760)	1.08	1.31	1.36	4.72	3.58	2.56		
		Smoking Status	Smoker (n=1039)	1.15	1.35	1.83	5.49	4.81	3.18		
			Non Smoker (n=721)	0.97	1.25	0.69	3.61	1.80	1.66		
	Camden & Islington	Trust	Clozapine (n=561)	0.71	1.25	0.18	3.03	3.39	3.03		
		Smoking Status	Smoker (n=360)	0.00	1.67	0.00	4.17	5.00	2.78		
			Non Smoker (n=201)	0.00	0.50	0.50	1.00	0.50	3.48		
Dyspepsia	SLAM	Trust	Clozapine (n=1760)	0.74	1.08	0.91	3.92	3.13	3.69		
		Smoking Status	Smoker (n=1039)	0.87	1.15	1.06	5.77	4.43	5.20		
			Non Smoker (n=721)	0.55	0.97	0.69	1.25	1.25	1.53		
	Camden & Islington	Trust	Clozapine (n=561)	0.36	0.53	0.53	4.10	2.67	2.50		
		Smoking Status	Smoker (n=360)	0.00	0.83	0.83	5.83	3.33	3.61		
			Non Smoker (n=201)	0.00	0.00	0.00	1.00	1.49	0.50		
Sweating	SLAM	Trust	Clozapine (n=1760)	1.08	0.97	1.36	4.43	4.26	2.84		
		Smoking Status	Smoker (n=1039)	1.06	1.25	1.83	5.97	5.97	4.33		
			Non Smoker (n=721)	1.11	0.55	0.69	2.22	1.80	0.69		
	Camden & Islington	Trust	Clozapine (n=561)	0.53	0.53	0.53	2.85	2.14	1.96		
		Smoking Status	Smoker (n=360)	0.00	0.83	0.56	3.89	2.78	2.50		
			Non Smoker (n=201)	0.00	0.00	0.50	1.00	1.00	1.00		
Neutropenia	SLAM	Trust	Clozapine (n=1760)	0.80	0.80	0.74	5.34	2.73	2.61		
		Smoking Status	Smoker (n=1039)	0.87	0.87	0.77	6.45	3.66	3.46		
			Non Smoker (n=721)	0.69	0.69	0.69	3.74	1.39	1.39		
	Camden & Islington	Trust	Clozapine (n=561)	0.00	0.18	0.53	1.60	0.89	1.07		
		Smoking Status	Smoker (n=360)	0.00	0.00	0.56	2.22	1.11	1.11		
			Non Smoker (n=201)	0.00	0.50	0.50	0.50	0.50	1.00		
Akathisia	SLAM	Trust	Clozapine (n=1760)	0.80	0.91	0.74	2.67	1.36	0.80		
		Smoking Status	Smoker (n=1039)	1.15	0.96	0.87	3.37	1.64	1.35		
			Non Smoker (n=721)	0.28	0.83	0.55	1.66	0.97	0.00		
	Camden & Islington	Trust	Clozapine (n=561)	0.00	0.53	0.00	1.25	1.07	0.53		
		Smoking Status	Smoker (n=360)	0.00	0.83	0.00	1.94	1.11	0.56		
			Non Smoker (n=201)	0.00	0.00	0.00	0.00	1.00	0.50		
Blurredvision	SLAM	Trust	Clozapine (n=1760)	0.34	0.91	0.63	2.05	1.25	1.02		
		Smoking Status	Smoker (n=1039)	0.19	1.35	0.96	2.21	1.73	1.44		
			Non Smoker (n=721)	0.55	0.28	0.14	1.80	0.55	0.42		
	Camden & Islington	Trust	Clozapine (n=561)	0.89	0.53	0.71	1.25	0.36	0.89		
		Smoking Status	Smoker (n=360)	0.00	0.83	1.11	1.67	0.56	1.11		
			Non Smoker (n=201)	0.00	0.00	0.00	0.50	0.00	0.50		

Measure Values: 0.00 – 63.04

The results are shown in percentages (%) and broken down by ADRs, Trusts (SLAM, Camden & Islington and Oxford), Cohorts, Sub Cohorts and SIDER reported values.
In Sub Cohort 'Clozapine' represent the total baseline population which further breaks down into 'Smokers and 'Non-Smokers'.
The columns (Three Months Early, Two Months Early, One Month Early, One Month Later, Two Months Later, Three Months Later) shows the percentages in each monthly interval. The last two columns (SIDER Low End and SIDER High End) shows the SIDER reporting.

Supplementary Table B.6 Chi-Square tests results

Chi Square Statistics (χ2)

P Value
<0.05
>0.05

ADR	Cohort	Trust	Months: One Month Later	Two Months Later	Three Months Later
Abdominalpain	Gender (1 df)	SLAM	2.30e-28	3.60e-30	1.73e+00
		C&I	5.11e-02	1.16e-02	8.22e-02
		Oxford	5.95e-03	7.89e-03	2.36e+00
	Ethnicity (3 df)	SLAM	1.03e+00	9.10e+00	1.03e+00
		C&I	7.42e-01	1.06e+00	7.96e-01
		Oxford	3.82e+00	2.18e+00	1.60e+00
	Agegroup (7 df)	SLAM	6.31e+00	4.63e+00	2.34e+00
		C&I	8.05e+00	2.75e+00	6.69e+00
	Hospital Admissions(1 df)	SLAM	6.84e+01	1.67e+01	2.46e+01
		C&I	1.20e+00	1.33e+01	1.37e-01
	Smoking Status (1 df)	SLAM	2.55e+01	2.00e+01	2.65e+01
		C&I	1.17e+00	3.03e+00	2.59e+00
Agitation	Gender (1 df)	SLAM	5.51e+00	1.03e+00	1.62e-01
		C&I	2.27e+00	2.73e-01	5.39e-01
		Oxford	6.05e-03	1.29e-01	8.67e-01
	Ethnicity (3 df)	SLAM	7.58e+00	9.10e+00	5.45e+00
		C&I	8.30e+00	2.35e+00	3.78e+00
		Oxford	1.88e+00	9.56e-01	1.12e+00
	Agegroup (7 df)	SLAM	4.86e+01	2.75e+01	4.57e+01
		C&I	9.31e+00	5.90e+00	2.14e+01
	Hospital Admissions(1 df)	SLAM	4.88e+02	2.98e+02	2.21e+02
		C&I	7.66e+01	3.72e+01	1.04e+01
	Smoking Status (1 df)	SLAM	2.74e+02	1.62e+02	1.84e+02
		C&I	7.65e+01	4.28e+01	3.10e+01
Akathisia	Gender (1 df)	SLAM	1.10e-02	2.20e+00	2.44e-27
		C&I	1.29e-03	2.21e-31	9.60e-28
		Oxford	1.62e-02	4.46e-31	1.53e-30
	Ethnicity (3 df)	SLAM	4.56e+00	3.37e+00	2.86e-01
		C&I	1.39e+00	5.82e+00	1.17e+01
		Oxford	1.47e+00	1.25e+00	1.04e+00
	Agegroup (7 df)	SLAM	1.06e+01	1.17e+01	2.89e+00
		C&I	2.99e+00	4.47e+00	2.77e+00
	Hospital Admissions(1 df)	SLAM	1.97e+01	1.24e+01	3.91e+00
		C&I	7.38e-31	1.12e+01	1.64e+00
	Smoking Status (1 df)	SLAM	4.12e+00	9.50e-01	8.16e+00
		C&I	2.54e+00	3.76e-28	1.86e-28
Backache	Gender (1 df)	SLAM	5.62e+00	2.48e+00	1.69e-01
		C&I	8.77e+00	1.41e+00	2.89e+00
		Oxford	6.64e+00	7.63e-01	3.74e+00
	Ethnicity (3 df)	SLAM	1.08e+00	3.08e+00	5.53e+00
		C&I	4.32e+00	1.96e+00	4.11e+00
		Oxford	3.73e-01	2.84e+00	2.10e+00
	Agegroup (7 df)	SLAM	7.47e+00	2.05e+00	3.88e+00
		C&I	1.63e+01	1.05e+01	2.07e+00
	Hospital Admissions(1 df)	SLAM	4.49e+01	3.12e+01	1.35e+01
		C&I	1.19e+01	1.37e+01	4.09e-01
	Smoking Status (1 df)	SLAM	1.47e+01	1.78e+01	4.54e+00
		C&I	8.05e+00	1.77e+00	3.40e+00
Blurredvision	Gender (1 df)	SLAM	6.90e+00	1.58e-03	4.76e-02

Chi Square (χ2) statistics are shown in the results and broken down by ADR, the cohort with a degree of freedom*, three trusts (SLaM, Camden & Islington and Oxford) and further broken down into three months after starting the drug Clozapine.
Adjustment for multiple comparisons: **Bonferroni**.
The mean difference is significant at the **0.05 level** (95% confidence interval for difference). The results in **Red** shows statistically significant p values.

Chi Square Statistics (χ2)

P Value
<0.05
>0.05

ADR	Cohort	Trust	Months: One Month Later	Months: Two Months Later	Months: Three Months Later
Blurredvision	Gender (1 df)	C&I	1.50e-28	9.25e-28	2.48e-31
		Oxford	9.21e-31	1.79e-02	1.83e-01
	Ethnicity (3 df)	SLAM	1.89e+00	1.94e+00	1.47e+00
		C&I	1.46e+00	1.24e+00	1.66e+00
		Oxford	1.80e+00	9.72e-01	1.25e+00
	Agegroup (7 df)	SLAM	7.69e+00	6.84e+00	1.01e+01
		C&I	5.63e+00	5.95e+00	5.70e+00
	Hospital Admissions(1 df)	SLAM	1.03e+01	1.00e+01	2.02e+00
		C&I	1.04e+00	2.71e-02	3.32e-01
	Smoking Status (1 df)	SLAM	1.83e-01	3.88e+00	3.48e+00
		C&I	6.39e-01	1.02e-01	7.46e-02
Confusion	Gender (1 df)	SLAM	7.52e-01	2.90e-03	5.21e-01
		C&I	3.74e-01	1.34e+00	2.82e+00
		Oxford	2.89e+00	3.04e-02	2.11e+00
	Ethnicity (3 df)	SLAM	2.52e+00	3.59e+00	2.63e+00
		C&I	6.07e+00	7.88e+00	3.02e+00
		Oxford	4.92e+00	1.45e+00	1.05e+00
	Agegroup (7 df)	SLAM	8.89e+00	1.86e+01	6.20e+00
		C&I	1.58e+01	3.45e+00	2.47e+00
	Hospital Admissions(1 df)	SLAM	1.09e+02	6.40e+01	3.95e+01
		C&I	2.57e+01	1.67e+01	1.25e-01
	Smoking Status (1 df)	SLAM	5.90e+01	2.30e+01	3.41e+01
		C&I	1.78e+01	6.22e+00	1.22e+01
Constipation	Gender (1 df)	SLAM	1.01e+01	2.08e+01	6.00e+00
		C&I	7.94e-01	2.98e-29	4.92e+00
		Oxford	2.14e+00	2.06e+00	2.06e+00
	Ethnicity (3 df)	SLAM	1.37e+01	8.75e+00	1.09e+00
		C&I	4.91e+00	2.90e+00	1.46e+00
		Oxford	1.38e+00	9.66e-01	2.34e+00
	Agegroup (7 df)	SLAM	1.43e+01	1.38e+01	1.20e+01
		C&I	1.02e+01	2.30e+00	1.82e+00
	Hospital Admissions(1 df)	SLAM	5.01e+01	5.93e+01	1.92e+01
		C&I	1.70e+01	4.80e+00	8.17e-01
	Smoking Status (1 df)	SLAM	1.26e+01	2.77e+01	8.78e+00
		C&I	4.24e+00	2.73e+00	5.13e+00
Convulsion	Gender (1 df)	SLAM	1.15e-01	9.49e-01	6.47e-28
		C&I	2.81e-02	8.94e-31	7.37e-02
		Oxford	4.80e-29	2.36e+00	1.09e+00
	Ethnicity (3 df)	SLAM	3.53e+01	3.50e-01	1.84e+00
		C&I	2.59e+00	7.24e-01	3.45e+00
		Oxford	6.37e-01	4.08e+00	2.97e+00
	Agegroup (7 df)	SLAM	1.09e+01	9.62e+00	9.49e+00
		C&I	8.64e+00	3.88e+00	3.30e+00
	Hospital Admissions(1 df)	SLAM	5.87e+01	2.64e+01	3.13e+01
		C&I	2.46e-02	2.24e+00	2.15e-30
	Smoking Status (1 df)	SLAM	2.30e+01	1.00e+01	2.63e+00
		C&I	4.25e-01	2.90e+00	3.09e-01
Diarrhoea	Gender (1 df)	SLAM	5.15e+00	1.19e-01	7.10e+00
		C&I	2.65e-02	8.22e-02	1.41e+00

Chi Square (χ2) statistics are shown in the results and broken down by ADR, the cohort with a degree of freedom*, three trusts (SLaM, Camden & Islington and Oxford) and further broken down into three months after starting the drug Clozapine.
Adjustment for multiple comparisons: **Bonferroni**.
The mean difference is significant at the **0.05 level** (95% confidence interval for difference). The results in **Red** shows statistically significant p values.

Chi Square Statistics (χ2)

P Value
<0.05
>0.05

ADR	Cohort	Trust	Months: One Month Later	Two Months Later	Three Months Later
Diarrhoea	Gender (1 df)	Oxford	6.68e+00	2.42e+00	6.10e+00
	Ethnicity (3 df)	SLAM	1.63e+00	3.71e+00	2.58e+00
		C&I	3.58e+00	2.76e+00	2.24e+00
		Oxford	2.95e+00	2.45e+00	2.76e+00
	Agegroup (7 df)	SLAM	1.96e+00	1.35e+00	5.70e+00
		C&I	1.86e+00	3.63e+00	6.82e+00
	Hospital Admissions(1 df)	SLAM	3.72e+01	1.98e+01	2.30e+01
		C&I	3.47e+00	2.34e+00	3.41e-01
	Smoking Status (1 df)	SLAM	2.94e+00	1.03e+01	3.32e+00
		C&I	3.40e+00	6.67e+00	4.42e-02
Dizziness	Gender (1 df)	SLAM	1.25e+01	4.17e+00	2.13e+00
		C&I	2.84e+00	3.66e+00	8.14e-01
		Oxford	4.91e+00	2.85e+00	2.50e+00
	Ethnicity (3 df)	SLAM	3.80e+00	1.14e+01	1.83e+01
		C&I	1.87e+00	3.47e+00	2.42e+00
		Oxford	2.36e+00	3.18e+00	2.45e-01
	Agegroup (7 df)	SLAM	1.25e+01	5.22e+00	1.18e+01
		C&I	5.88e+00	1.18e+01	6.99e+00
	Hospital Admissions(1 df)	SLAM	1.25e+02	5.70e+01	3.57e+01
		C&I	2.08e+01	2.96e+01	1.65e+01
	Smoking Status (1 df)	SLAM	3.46e+01	2.54e+01	2.93e+01
		C&I	1.70e+01	1.49e+01	8.96e+00
Drymouth	Gender (1 df)	SLAM	1.36e+00	4.13e+00	5.26e-02
		C&I	4.60e-01	4.75e-01	2.47e+00
		Oxford	3.32e-01	3.32e-30	2.37e+00
	Ethnicity (3 df)	SLAM	1.19e+01	4.01e+00	7.17e-01
		C&I	1.92e+00	1.73e+00	9.71e-01
		Oxford	2.50e+00	1.01e+00	1.79e+00
	Agegroup (7 df)	SLAM	8.32e+00	1.33e+01	9.32e+00
		C&I	3.24e+00	7.54e+00	1.78e+00
	Hospital Admissions(1 df)	SLAM	2.14e+01	1.74e+01	7.12e+00
		C&I	2.68e+00	5.93e-01	2.92e-01
	Smoking Status (1 df)	SLAM	6.51e+00	6.78e+00	1.56e+01
		C&I	2.35e+00	1.20e+00	3.21e-28
Dyspepsia	Gender (1 df)	SLAM	4.28e-03	1.32e+00	4.83e-01
		C&I	1.61e+00	2.70e-01	2.13e+00
		Oxford	3.13e+00	1.36e+00	1.13e+00
	Ethnicity (3 df)	SLAM	3.73e+00	3.57e+00	2.16e+00
		C&I	4.82e+00	4.88e+00	5.47e+00
		Oxford	2.08e+00	3.01e+00	1.05e+00
	Agegroup (7 df)	SLAM	8.18e+00	2.14e+00	4.56e+00
		C&I	5.06e+00	1.11e+01	5.64e+00
	Hospital Admissions(1 df)	SLAM	3.17e+01	5.53e+00	2.44e+01
		C&I	4.10e+00	8.92e-01	9.80e+00
	Smoking Status (1 df)	SLAM	2.20e+01	1.32e+01	1.51e+01
		C&I	6.50e+00	1.05e+00	3.94e+00
Enuresis	Gender (1 df)	SLAM	8.45e+00	1.19e+01	2.28e+00
		C&I	8.94e-01	1.39e+00	1.50e-28
		Oxford	2.43e-01	2.96e+00	5.18e+00

Chi Square (χ2) statistics are shown in the results and broken down by ADR, the cohort with a degree of freedom*, three trusts (SLaM, Camden & Islington and Oxford) and further broken down into three months after starting the drug Clozapine.
Adjustment for multiple comparisons: **Bonferroni**.
The mean difference is significant at the **0.05 level** (95% confidence interval for difference). The results in **Red** shows statistically significant p values.

Chi Square Statistics (χ2)

P Value
<0.05
>0.05

ADR	Cohort	Trust	Months: One Month Later	Two Months Later	Three Months Later
Enuresis	Ethnicity (3 df)	SLAM	4.10e+00	3.10e+00	1.15e+00
		C&I	6.34e+00	1.59e+00	2.01e+00
		Oxford	6.83e-01	8.87e-01	4.54e+00
	Agegroup (7 df)	SLAM	9.63e+00	3.61e+00	1.04e+01
		C&I	8.15e+00	1.50e+01	1.29e+01
	Hospital Admissions(1 df)	SLAM	4.71e+01	2.63e+01	1.82e+01
		C&I	4.10e+00	5.99e-01	3.86e+00
	Smoking Status (1 df)	SLAM	2.53e+01	8.64e+00	5.87e+00
		C&I	2.76e+00	3.47e-29	4.05e-05
Fatigue	Gender (1 df)	SLAM	1.58e+01	1.30e+01	6.66e+00
		C&I	1.79e+00	1.70e-01	1.12e+00
		Oxford	4.58e+00	3.04e+00	7.76e-02
	Ethnicity (3 df)	SLAM	8.55e+00	4.80e+00	5.40e+00
		C&I	4.42e+00	7.38e-01	6.18e+00
		Oxford	1.89e+00	6.25e+00	8.28e-01
	Agegroup (7 df)	SLAM	5.51e+01	5.32e+01	4.80e+01
		C&I	1.12e+01	4.34e+00	7.32e+00
	Hospital Admissions(1 df)	SLAM	3.23e+02	2.99e+02	2.47e+02
		C&I	2.97e+01	3.37e+01	2.09e+01
	Smoking Status (1 df)	SLAM	1.84e+02	1.41e+02	1.48e+02
		C&I	4.69e+01	4.66e+01	4.30e+01
Feelingsick	Gender (1 df)	SLAM	2.48e+01	1.88e+01	1.30e+01
		C&I	8.95e-02	1.02e+00	2.02e-03
		Oxford	6.77e-01	7.63e-01	3.43e+00
	Ethnicity (3 df)	SLAM	2.25e+00	2.23e+00	4.33e+00
		C&I	2.07e+00	5.50e+00	2.64e+00
		Oxford	7.57e-01	5.64e+00	2.08e+00
	Agegroup (7 df)	SLAM	3.65e+01	3.46e+01	2.17e+01
		C&I	4.72e+00	8.55e+00	1.04e+01
	Hospital Admissions(1 df)	SLAM	9.59e+01	6.85e+01	4.86e+01
		C&I	1.55e+01	1.89e+00	2.20e+00
	Smoking Status (1 df)	SLAM	3.18e+01	3.10e+01	1.93e+01
		C&I	1.53e+01	9.13e+00	1.08e+01
Fever	Gender (1 df)	SLAM	8.09e+00	4.06e+00	7.88e-02
		C&I	9.09e-31	2.75e+00	2.48e-31
		Oxford	1.06e-28	1.09e+00	8.99e-02
	Ethnicity (3 df)	SLAM	2.17e+00	1.86e+00	1.67e+00
		C&I	3.91e+00	2.12e-01	3.90e+00
		Oxford	1.10e+00	6.84e-01	1.79e+00
	Agegroup (7 df)	SLAM	2.82e+00	5.06e+00	7.89e+00
		C&I	9.85e+00	6.19e+00	1.55e+01
	Hospital Admissions(1 df)	SLAM	4.69e+01	1.21e+01	2.63e+01
		C&I	2.07e+01	8.92e-01	2.67e-28
	Smoking Status (1 df)	SLAM	1.80e+01	1.69e+01	1.32e+01
		C&I	5.43e+00	4.47e+00	7.46e-02
Headache	Gender (1 df)	SLAM	1.20e+01	2.16e+00	1.91e+00
		C&I	5.32e-01	4.48e-02	3.60e-04
		Oxford	5.15e-03	9.01e-02	2.52e-02
	Ethnicity (3 df)	SLAM	2.77e+00	2.12e+00	2.82e+00

Chi Square (χ2) statistics are shown in the results and broken down by ADR, the cohort with a degree of freedom*, three trusts (SLaM, Camden & Islington and Oxford) and further broken down into three months after starting the drug Clozapine.
Adjustment for multiple comparisons: **Bonferroni**.
The mean difference is significant at the **0.05 level** (95% confidence interval for difference). The results in Red shows statistically significant p values.

Chi Square Statistics (χ2)

P Value
<0.05
>0.05

ADR	Cohort	Trust	Months: One Month Later	Two Months Later	Three Months Later
Headache	Ethnicity (3 df)	C&I	1.44e+00	4.46e+00	2.36e+00
		Oxford	3.99e+00	1.66e+00	6.78e+00
	Agegroup (7 df)	SLAM	1.50e+01	6.23e+00	1.74e+01
		C&I	3.79e+00	1.06e+01	1.62e+01
	Hospital Admissions(1 df)	SLAM	8.18e+01	2.30e+01	3.28e+01
		C&I	1.04e+01	3.21e+00	4.43e+00
	Smoking Status (1 df)	SLAM	4.81e+01	2.91e+01	4.09e+01
		C&I	9.46e+00	3.76e+00	4.07e+00
Hyperprolactinaemia	Gender (1 df)	SLAM	1.87e+01	9.67e+00	1.54e+00
		C&I	3.41e+00	5.17e+00	2.33e+00
		Oxford	1.30e+00	1.61e+00	1.86e+00
	Ethnicity (3 df)	SLAM	2.55e+00	1.81e+00	1.62e+00
		C&I	9.13e+00	8.72e-01	1.24e+00
		Oxford	1.44e+00	9.54e-01	7.01e+00
	Agegroup (7 df)	SLAM	2.01e+01	2.51e+01	2.99e+01
		C&I	9.53e+00	2.26e+00	2.95e+00
	Hospital Admissions(1 df)	SLAM	4.48e+01	2.85e+01	1.89e+01
		C&I	9.72e+00	1.91e-01	9.74e-31
	Smoking Status (1 df)	SLAM	1.07e+00	4.29e-01	1.73e+00
		C&I	5.02e+00	5.97e-01	4.01e-01
Hypersalivation	Gender (1 df)	SLAM	2.25e+00	1.08e+00	7.53e-01
		C&I	2.20e-02	8.56e-01	8.93e-01
		Oxford	2.61e+00	1.10e-01	1.90e+00
	Ethnicity (3 df)	SLAM	2.14e+00	2.89e+00	1.76e+00
		C&I	3.88e+00	2.31e+00	3.88e+00
		Oxford	1.45e+00	1.17e+00	1.68e+00
	Agegroup (7 df)	SLAM	1.44e+01	2.48e+00	2.05e+01
		C&I	7.09e+00	9.81e+00	5.35e+00
	Hospital Admissions(1 df)	SLAM	5.96e+01	7.55e+01	4.34e+01
		C&I	1.39e-01	3.56e+00	4.83e-01
	Smoking Status (1 df)	SLAM	1.27e+01	1.08e+01	1.14e+01
		C&I	1.69e+00	2.40e+00	5.23e+00
Hypertension	Gender (1 df)	SLAM	1.61e+00	5.66e-01	2.52e-03
		C&I	1.06e-01	1.59e-01	7.26e-29
		Oxford	2.62e-01	2.76e-01	1.37e-02
	Ethnicity (3 df)	SLAM	1.80e+00	2.87e+00	3.69e+00
		C&I	2.75e+00	2.42e+00	3.20e+00
		Oxford	1.91e+01	2.67e+00	4.17e+00
	Agegroup (7 df)	SLAM	1.31e+01	2.43e+01	7.58e+00
		C&I	1.98e+01	5.27e+00	3.35e+00
	Hospital Admissions(1 df)	SLAM	2.21e+01	2.32e+01	1.84e+01
		C&I	6.75e+00	2.58e+00	8.92e-01
	Smoking Status (1 df)	SLAM	1.18e+01	1.24e+01	1.67e+01
		C&I	3.93e-01	4.07e+00	2.28e-01
Hypotension	Gender (1 df)	SLAM	8.84e+00	8.51e-02	1.17e-02
		C&I	2.33e+00	2.03e-01	1.53e+00
		Oxford	3.77e+00	3.13e+00	3.74e+00
	Ethnicity (3 df)	SLAM	1.09e+01	8.11e+00	1.71e+00
		C&I	6.57e+00	2.89e+00	4.23e+00

Chi Square (χ2) statistics are shown in the results and broken down by ADR, the cohort with a degree of freedom*, three trusts (SLaM, Camden & Islington and Oxford) and further broken down into three months after starting the drug Clozapine.
Adjustment for multiple comparisons: **Bonferroni**.
The mean difference is significant at the **0.05 level** (95% confidence interval for difference). The results in **Red** shows statistically significant p values.

Chi Square Statistics (χ2)

P Value
<0.05
>0.05

ADR	Cohort	Trust	Months: One Month Later	Two Months Later	Three Months Later
Hypotension	Ethnicity (3 df)	Oxford	4.68e+00	2.23e+00	2.10e+00
	Agegroup (7 df)	SLAM	1.94e+01	1.98e+01	3.35e+01
		C&I	7.03e+00	7.87e+00	2.27e+01
	Hospital Admissions(1 df)	SLAM	5.90e+01	2.01e+01	8.74e+00
		C&I	1.90e+00	7.24e+00	2.33e-30
	Smoking Status (1 df)	SLAM	6.60e+00	6.35e+00	2.30e+00
		C&I	6.26e-01	8.52e-03	5.19e-01
Insomnia	Gender (1 df)	SLAM	9.31e-01	1.09e+00	1.93e+00
		C&I	1.29e+00	5.96e-01	1.45e+00
		Oxford	7.50e+00	1.07e+00	4.84e-01
	Ethnicity (3 df)	SLAM	2.73e+00	2.72e+00	1.55e+00
		C&I	5.79e-01	6.41e-01	2.72e-01
		Oxford	1.29e+00	2.43e+00	3.34e+00
	Agegroup (7 df)	SLAM	1.04e+01	8.15e+00	1.60e+01
		C&I	4.93e+00	1.11e+01	4.23e+00
	Hospital Admissions(1 df)	SLAM	6.74e+01	2.18e+01	5.75e+00
		C&I	2.41e+01	4.46e+00	1.04e-01
	Smoking Status (1 df)	SLAM	3.73e+01	1.58e+01	9.61e+00
		C&I	1.24e+01	4.40e+00	4.92e+00
Nausea	Gender (1 df)	SLAM	5.84e+00	2.89e+00	9.62e+00
		C&I	7.29e-01	1.11e+00	1.11e+00
		Oxford	3.23e+00	3.20e-01	6.76e-01
	Ethnicity (3 df)	SLAM	5.14e+00	6.18e+00	7.44e-01
		C&I	7.45e-01	3.40e+00	1.87e+00
		Oxford	3.67e+00	5.20e+00	1.18e+00
	Agegroup (7 df)	SLAM	4.89e+00	8.34e+00	1.17e+01
		C&I	3.86e-01	6.30e+00	2.45e+00
	Hospital Admissions(1 df)	SLAM	2.49e+01	1.52e+01	5.65e+00
		C&I	6.78e+00	6.60e-01	1.02e-01
	Smoking Status (1 df)	SLAM	8.45e+00	2.45e+00	5.50e+00
		C&I	2.55e+00	1.60e+00	1.95e-29
Neutropenia	Gender (1 df)	SLAM	1.92e+00	6.85e-01	1.00e-01
		C&I	7.35e-01	2.48e-31	7.37e-02
	Ethnicity (3 df)	SLAM	2.07e+00	2.86e+00	2.37e+00
		C&I	7.52e+00	1.04e+01	3.47e+00
	Agegroup (7 df)	SLAM	2.27e+01	8.12e+00	5.55e+00
		C&I	2.46e+00	2.17e+00	2.61e+00
	Hospital Admissions(1 df)	SLAM	2.69e+01	7.78e+00	1.37e+01
		C&I	3.14e-01	3.32e-01	2.15e-30
	Smoking Status (1 df)	SLAM	5.63e+00	7.44e+00	6.43e+00
		C&I	1.46e+00	7.46e-02	3.76e-28
Rash	Gender (1 df)	SLAM	1.77e+00	1.40e+00	1.85e+00
		C&I	2.83e-01	1.00e-01	8.94e-31
		Oxford	1.13e+00	8.44e-01	8.72e-01
	Ethnicity (3 df)	SLAM	1.26e+01	1.65e+00	4.36e-01
		C&I	2.07e+00	5.93e+00	5.56e+00
		Oxford	1.42e+00	2.36e+00	2.92e+00
	Agegroup (7 df)	SLAM	2.41e+00	1.80e+00	2.72e+00
		C&I	5.09e+00	1.74e+00	6.79e+00

Chi Square (χ2) statistics are shown in the results and broken down by ADR, the cohort with a degree of freedom*, three trusts (SLaM, Camden & Islington and Oxford) and further broken down into three months after starting the drug Clozapine.
Adjustment for multiple comparisons: **Bonferroni**.
The mean difference is significant at the **0.05 level** (95% confidence interval for difference). The results in Red shows statistically significant p values.

Chi Square Statistics (χ2)

P Value
<0.05
>0.05

ADR	Cohort	Trust	Months: One Month Later	Two Months Later	Three Months Later
Rash	Hospital Admissions(1 df)	SLAM	2.33e+01	2.56e+01	3.69e+01
		C&I	1.85e+00	9.16e-01	1.99e-03
	Smoking Status (1 df)	SLAM	5.10e+00	1.14e+01	1.03e+01
		C&I	4.92e+00	4.78e+00	1.20e+00
Sedation	Gender (1 df)	SLAM	4.73e+00	4.86e+00	7.36e+00
		C&I	1.30e+00	6.86e-02	1.23e+00
		Oxford	6.74e-02	0.00e+00	4.26e-01
	Ethnicity (3 df)	SLAM	7.22e+00	6.71e+00	4.07e+00
		C&I	5.31e+00	2.72e+00	2.13e+00
		Oxford	2.23e+00	6.06e+00	2.41e+00
	Agegroup (7 df)	SLAM	4.02e+01	5.62e+01	3.95e+01
		C&I	7.92e+00	6.59e+00	1.53e+01
	Hospital Admissions(1 df)	SLAM	3.69e+02	2.35e+02	2.24e+02
		C&I	4.49e+01	3.31e+01	5.81e+00
	Smoking Status (1 df)	SLAM	1.76e+02	1.28e+02	1.09e+02
		C&I	2.25e+01	2.87e+01	2.31e+01
Shaking	Gender (1 df)	SLAM	7.06e-01	1.11e-01	6.44e-01
		C&I	9.31e-31	0.00e+00	4.83e-01
		Oxford	1.51e-01	1.35e-29	4.98e+00
	Ethnicity (3 df)	SLAM	6.41e+00	7.27e+00	8.68e+00
		C&I	4.35e+00	2.70e+00	6.57e+00
		Oxford	2.44e+00	3.07e+00	3.82e+00
	Agegroup (7 df)	SLAM	2.87e+01	1.71e+01	3.25e+01
		C&I	1.70e+01	6.91e+00	2.61e+01
	Hospital Admissions(1 df)	SLAM	7.64e+01	4.60e+01	3.35e+01
		C&I	6.75e-02	3.10e-01	4.19e+00
	Smoking Status (1 df)	SLAM	2.50e+01	1.92e+01	2.81e+01
		C&I	2.98e+00	1.17e+00	2.92e+00
Stomachpain	Gender (1 df)	SLAM	1.42e+01	3.27e+00	1.82e+01
		C&I	5.96e-01	2.00e+00	4.75e-01
		Oxford	1.37e-02	1.53e-30	6.20e-01
	Ethnicity (3 df)	SLAM	1.86e+00	1.98e+00	4.44e-01
		C&I	6.94e+00	5.61e+00	3.55e+00
		Oxford	2.32e+00	1.04e+00	1.04e+00
	Agegroup (7 df)	SLAM	5.32e+00	4.87e+00	6.92e+00
		C&I	2.40e+00	9.41e+00	4.61e+00
	Hospital Admissions(1 df)	SLAM	3.64e+01	2.36e+01	1.45e+01
		C&I	4.46e+00	4.19e+00	1.99e-03
	Smoking Status (1 df)	SLAM	2.91e+01	1.53e+01	6.77e+00
		C&I	6.67e+00	2.92e+00	2.90e+00
Sweating	Gender (1 df)	SLAM	2.89e-03	3.76e+00	1.67e-01
		C&I	1.49e+00	2.74e-01	9.02e-01
		Oxford	1.13e-02	1.62e-02	6.10e-01
	Ethnicity (3 df)	SLAM	6.44e+00	2.71e+00	7.17e-01
		C&I	3.66e+00	1.03e+00	1.01e+00
		Oxford	1.50e+00	4.53e+00	3.57e+00
	Agegroup (7 df)	SLAM	8.64e+00	5.83e+00	5.06e+00
		C&I	5.97e+00	5.04e+00	3.74e+00
	Hospital	SLAM	3.40e+01	8.37e+00	1.33e+01

Chi Square (χ2) statistics are shown in the results and broken down by ADR, the cohort with a degree of freedom*, three trusts (SLaM, Camden & Islington and Oxford) and further broken down into three months after starting the drug Clozapine.
Adjustment for multiple comparisons: **Bonferroni**.
The mean difference is significant at the **0.05 level** (95% confidence interval for difference). The results in Red shows statistically significant p values.

Chi Square Statistics (χ2)

P Value
<0.05
>0.05

ADR	Cohort	Trust	Months: One Month Later	Months: Two Months Later	Months: Three Months Later
Sweating	Admissions(1 df)	C&I	4.19e+00	2.24e+00	2.94e+00
	Smoking Status (1 df)	SLAM	1.32e+01	1.71e+01	1.91e+01
		C&I	2.92e+00	1.20e+00	8.38e-01
Tachycardia	Gender (1 df)	SLAM	1.88e+00	3.93e-02	5.66e-01
		C&I	5.16e-29	6.12e-29	5.54e-02
		Oxford	1.45e-30	4.93e-31	2.52e-02
	Ethnicity (3 df)	SLAM	3.11e+00	3.99e+00	5.13e+00
		C&I	1.08e+01	6.00e-01	6.43e+00
		Oxford	2.11e+00	6.45e+00	5.28e+00
	Agegroup (7 df)	SLAM	4.57e+01	4.79e+01	5.67e+01
		C&I	1.24e+01	3.86e+00	7.55e+00
	Hospital Admissions(1 df)	SLAM	1.58e+02	1.20e+02	6.84e+01
		C&I	2.07e+01	1.15e+01	3.56e+00
	Smoking Status (1 df)	SLAM	4.79e+01	4.20e+01	2.39e+01
		C&I	1.13e+01	2.88e+00	3.59e+00
Tremor	Gender (1 df)	SLAM	2.33e-01	2.79e-02	2.12e-04
		C&I	0.00e+00	0.00e+00	2.74e-01
	Ethnicity (3 df)	SLAM	6.68e+00	1.07e+00	3.55e+00
		C&I	3.78e+00	2.00e+00	1.05e+00
	Agegroup (7 df)	SLAM	1.17e+01	1.18e+01	7.78e+00
		C&I	6.00e+00	2.97e+00	6.81e+00
	Hospital Admissions(1 df)	SLAM	3.48e+01	2.90e+01	2.78e+01
		C&I	7.39e+00	9.16e-01	2.24e+00
	Smoking Status (1 df)	SLAM	4.73e+00	5.47e+00	9.27e+00
		C&I	1.17e+00	7.85e-02	5.35e+00
Vomiting	Gender (1 df)	SLAM	2.72e-01	4.06e+00	1.83e+00
		C&I	3.65e-31	8.73e-01	1.62e+00
		Oxford	3.78e-02	2.11e+00	2.43e-29
	Ethnicity (3 df)	SLAM	2.56e+00	2.08e+00	1.28e+00
		C&I	3.97e+00	4.11e+00	5.11e+00
		Oxford	6.77e-01	4.90e-01	3.67e+00
	Agegroup (7 df)	SLAM	1.83e+01	1.59e+01	2.28e+01
		C&I	6.31e+00	8.01e+00	5.81e+00
	Hospital Admissions(1 df)	SLAM	2.13e+01	3.82e+01	3.49e+01
		C&I	3.98e+00	1.52e-01	3.69e-01
	Smoking Status (1 df)	SLAM	1.59e+01	2.07e+01	9.08e+00
		C&I	1.02e+01	3.36e+00	2.55e+00
Weightgain	Gender (1 df)	SLAM	9.08e+00	1.13e+01	4.76e+00
		C&I	2.24e+00	4.57e+00	6.80e-03
		Oxford	3.82e-01	1.15e+00	1.51e-01
	Ethnicity (3 df)	SLAM	5.03e+00	2.80e+00	2.14e+00
		C&I	1.18e+00	1.49e-01	2.64e+00
		Oxford	2.53e+00	3.96e+00	4.36e+00
	Agegroup (7 df)	SLAM	5.71e+01	4.43e+01	3.87e+01
		C&I	1.96e+00	1.44e+01	6.11e+00
	Hospital Admissions(1 df)	SLAM	7.46e+01	7.03e+01	7.98e+01
		C&I	2.75e+00	6.26e-05	3.62e+00
	Smoking Status (1 df)	SLAM	2.49e+01	3.71e+01	1.99e+01
		C&I	3.82e+00	7.57e+00	8.57e+00

Chi Square (χ2) statistics are shown in the results and broken down by ADR, the cohort with a degree of freedom*, three trusts (SLaM, Camden & Islington and Oxford) and further broken down into three months after starting the drug Clozapine.
Adjustment for multiple comparisons: **Bonferroni**.
The mean difference is significant at the **0.05 level** (95% confidence interval for difference). The results in **Red** shows statistically significant p values.

Supplementary Table B.7 Combine Analysis

Combine Analysis

P value
<0.05
>0.05

ADR	Cohort	Months: One Month Later	Two Months Later	Three Months Later
Abdominalpain	Gender	1.79e-29	2.45e-29	4.09e+00
	Ethnicity	2.26e+00	5.76e+00	2.40e-01
	Agegroup	1.60e+01	5.64e+00	6.85e+00
	Hospital Admission	7.90e+01	3.10e+01	2.57e+01
	Smoking Status	2.52e+01	2.29e+01	2.90e+01
Agitation	Gender	1.14e+00	7.73e-01	9.86e-02
	Ethnicity	1.44e+01	7.62e+00	6.53e+00
	Agegroup	5.96e+01	3.63e+01	7.00e+01
	Hospital Admission	5.58e+02	3.37e+02	2.35e+02
	Smoking Status	3.49e+02	2.04e+02	2.10e+02
Akathisia	Gender	9.11e-03	1.28e+00	6.46e-28
	Ethnicity	3.80e+00	1.63e+00	1.53e+00
	Agegroup	1.33e+01	1.14e+01	1.82e+00
	Hospital Admission	2.01e+01	2.27e+01	7.08e+00
	Smoking Status	6.34e+00	8.24e-01	6.83e+00
Backache	Gender	2.03e+01	5.47e+00	4.79e+00
	Ethnicity	8.09e-01	2.01e-01	7.57e-01
	Agegroup	1.88e+01	7.07e+00	4.71e+00
	Hospital Admission	5.48e+01	4.47e+01	1.27e+01
	Smoking Status	2.35e+01	1.98e+01	8.47e+00
Blurredvision	Gender	5.78e+00	1.55e-31	3.90e-01
	Ethnicity	1.93e+00	1.00e+00	3.22e+00
	Agegroup	5.85e+00	8.72e+00	9.92e+00
	Hospital Admission	1.40e+01	1.37e+01	8.08e-01
	Smoking Status	6.59e-01	4.46e+00	3.94e+00
Confusion	Gender	1.14e+00	6.30e-02	5.81e-01
	Ethnicity	8.71e+00	3.87e+00	4.01e+00
	Agegroup	1.19e+01	2.23e+01	6.62e+00
	Hospital Admission	1.35e+02	8.41e+01	3.66e+01
	Smoking Status	7.72e+01	2.94e+01	4.68e+01
Constipation	Gender	1.33e+01	1.97e+01	1.25e+01
	Ethnicity	8.47e+00	2.99e+00	2.18e+00
	Agegroup	1.81e+01	1.38e+01	8.47e+00
	Hospital Admission	6.69e+01	7.36e+01	2.51e+01
	Smoking Status	1.71e+01	2.92e+01	1.27e+01
Convulsion	Gender	3.72e-03	1.50e-03	2.90e-01
	Ethnicity	1.90e+01	4.94e-01	2.62e+00
	Agegroup	1.62e+01	7.99e+00	1.09e+01
	Hospital Admission	6.68e+01	3.61e+01	3.69e+01
	Smoking Status	2.16e+01	1.25e+01	2.82e+00
Diarrhoea	Gender	1.00e+01	1.63e+00	1.53e+01
	Ethnicity	2.20e+00	2.03e+00	4.86e+00
	Agegroup	1.60e+00	1.54e+00	4.51e+00
	Hospital Admission	4.56e+01	2.27e+01	1.35e+01
	Smoking Status	5.50e+00	1.72e+01	1.82e+00
Dizziness	Gender	2.11e+01	1.10e+01	5.35e+00
	Ethnicity	4.96e-01	6.24e+00	5.63e+00

Chi Square ($\chi 2$) statistics are shown i in the results and broken down into ADR, cohort and further broken down into three months after starting the drug Clozapine.
Adjustment for multiple comparisons: **Bonferroni**.
The mean difference is significant at the **0.05 level** (95% confidence interval for difference). The results in Red shows statistically significant p values.

Combine Analysis

P value
<0.05
>0.05

ADR	Cohort	Months		
		One Month Later	Two Months Later	Three Months Later
Dizziness	Agegroup	1.63e+01	9.25e+00	1.35e+01
	Hospital Admission	1.37e+02	8.03e+01	5.37e+01
	Smoking Status	5.23e+01	4.04e+01	3.82e+01
Drymouth	Gender	4.68e-02	4.21e+00	2.52e+00
	Ethnicity	5.92e+00	1.80e+00	2.13e-01
	Agegroup	1.15e+01	1.49e+01	1.13e+01
	Hospital Admission	2.57e+01	2.15e+01	1.08e+01
	Smoking Status	9.23e+00	8.20e+00	1.28e+01
Dyspepsia	Gender	3.42e-02	1.47e+00	1.89e-01
	Ethnicity	9.22e-01	3.23e+00	2.81e+00
	Agegroup	1.04e+01	6.93e+00	8.02e+00
	Hospital Admission	3.49e+01	7.45e+00	3.68e+01
	Smoking Status	2.97e+01	1.44e+01	1.95e+01
Enuresis	Gender	9.51e+00	1.72e+01	6.12e+00
	Ethnicity	2.72e+00	1.32e+00	4.69e-01
	Agegroup	9.98e+00	4.88e+00	1.28e+01
	Hospital Admission	5.03e+01	3.41e+01	2.78e+01
	Smoking Status	2.82e+01	6.90e+00	5.34e+00
Fatigue	Gender	2.21e+01	1.13e+01	7.24e+00
	Ethnicity	7.31e+00	7.29e+00	5.71e+00
	Agegroup	6.58e+01	6.03e+01	5.82e+01
	Hospital Admission	3.42e+02	3.41e+02	2.67e+02
	Smoking Status	2.30e+02	1.84e+02	1.89e+02
Feelingsick	Gender	2.06e+01	1.92e+01	1.31e+01
	Ethnicity	2.48e+00	3.45e+00	3.90e+00
	Agegroup	4.37e+01	4.54e+01	2.98e+01
	Hospital Admission	1.20e+02	7.68e+01	5.15e+01
	Smoking Status	4.53e+01	3.91e+01	2.94e+01
Fever	Gender	5.80e+00	8.37e+00	2.26e-01
	Ethnicity	7.03e+00	3.29e+00	4.87e+00
	Agegroup	3.78e+00	4.39e+00	1.46e+01
	Hospital Admission	7.18e+01	1.63e+01	3.10e+01
	Smoking Status	2.31e+01	2.15e+01	1.27e+01
Headache	Gender	5.39e+00	6.03e-01	1.67e+00
	Ethnicity	4.82e+00	1.05e+00	3.53e+00
	Agegroup	2.08e+01	1.23e+01	2.49e+01
	Hospital Admission	9.89e+01	2.91e+01	4.02e+01
	Smoking Status	5.70e+01	3.26e+01	4.45e+01
Hyperprolactinaemia	Gender	2.35e+01	1.73e+01	5.82e+00
	Ethnicity	5.77e+00	3.05e+00	1.50e+00
	Agegroup	2.63e+01	2.79e+01	2.85e+01
	Hospital Admission	5.45e+01	2.82e+01	1.63e+01
	Smoking Status	4.15e+00	9.76e-01	5.55e-01
Hypersalivation	Gender	3.51e-01	3.44e-01	2.84e+00
	Ethnicity	2.20e-01	1.72e+00	4.98e+00
	Agegroup	1.45e+01	8.77e+00	2.60e+01
	Hospital Admission	5.05e+01	9.24e+01	4.73e+01

Chi Square ($\chi 2$) statistics are shown i in the results and broken down into ADR, cohort and further broken down into three months after starting the drug Clozapine.
Adjustment for multiple comparisons: **Bonferroni**.
The mean difference is significant at the **0.05 level** (95% confidence interval for difference). The results in Red shows statistically significant p values.

Combine Analysis

P value
<0.05
>0.05

ADR	Cohort	Months		
		One Month Later	Two Months Later	Three Months Later
Hypersalivation	Smoking Status	1.45e+01	1.23e+01	1.58e+01
Hypertension	Gender	1.02e+00	4.02e-01	2.27e-04
	Ethnicity	7.91e+00	5.94e+00	9.29e+00
	Agegroup	1.91e+01	1.41e+01	4.28e+00
	Hospital Admission	3.18e+01	2.80e+01	2.33e+01
	Smoking Status	1.14e+01	1.68e+01	1.58e+01
Hypotension	Gender	1.60e+01	2.07e+00	2.81e+00
	Ethnicity	5.36e+00	3.78e+00	1.90e+00
	Agegroup	2.29e+01	1.91e+01	4.06e+01
	Hospital Admission	6.33e+01	2.91e+01	8.56e+00
	Smoking Status	7.28e+00	5.74e+00	3.13e+00
Insomnia	Gender	1.84e+00	2.86e+00	8.83e-04
	Ethnicity	1.46e+00	4.60e-01	1.21e-01
	Agegroup	1.25e+01	7.69e+00	1.22e+01
	Hospital Admission	9.24e+01	3.24e+01	5.38e+00
	Smoking Status	4.99e+01	1.97e+01	1.52e+01
Nausea	Gender	1.02e+01	4.89e+00	1.23e+01
	Ethnicity	4.02e+00	7.55e+00	5.21e-01
	Agegroup	5.36e+00	1.08e+01	9.62e+00
	Hospital Admission	3.46e+01	1.86e+01	3.65e+00
	Smoking Status	1.12e+01	3.91e+00	4.38e+00
Rash	Gender	1.52e+00	2.87e+00	2.19e+00
	Ethnicity	6.22e+00	1.88e+00	1.13e+00
	Agegroup	4.07e+00	3.07e+00	2.73e+00
	Hospital Admission	2.29e+01	2.92e+01	3.20e+01
	Smoking Status	1.03e+01	1.64e+01	1.21e+01
Sedation	Gender	1.69e+00	3.65e+00	3.76e+00
	Ethnicity	1.45e+01	1.73e+01	1.02e+01
	Agegroup	5.43e+01	7.30e+01	6.45e+01
	Hospital Admission	4.16e+02	2.85e+02	2.33e+02
	Smoking Status	1.91e+02	1.52e+02	1.29e+02
Shaking	Gender	6.72e-01	9.07e-02	6.45e-30
	Ethnicity	3.09e+00	3.73e+00	4.84e+00
	Agegroup	4.40e+01	2.42e+01	4.78e+01
	Hospital Admission	7.70e+01	4.64e+01	4.40e+01
	Smoking Status	2.73e+01	1.98e+01	3.08e+01
Stomachpain	Gender	1.21e+01	5.66e+00	1.98e+01
	Ethnicity	5.74e+00	6.15e+00	4.29e+00
	Agegroup	4.94e+00	8.06e+00	8.00e+00
	Hospital Admission	4.53e+01	3.00e+01	1.61e+01
	Smoking Status	3.62e+01	1.89e+01	9.70e+00
Sweating	Gender	2.49e-01	3.64e+00	9.75e-02
	Ethnicity	7.29e+00	3.87e+00	1.26e+00
	Agegroup	1.07e+01	9.05e+00	5.97e+00
	Hospital Admission	4.31e+01	1.42e+01	1.89e+01
	Smoking Status	1.64e+01	1.81e+01	1.97e+01
Tachycardia	Gender	1.27e+00	5.82e-02	4.77e-01

Chi Square ($\chi 2$) statistics are shown i in the results and broken down into ADR, cohort and further broken down into three months after starting the drug Clozapine.
Adjustment for multiple comparisons: **Bonferroni**.
The mean difference is significant at the **0.05 level** (95% confidence interval for difference). The results in Red shows statistically significant p values.

Combine Analysis

ADR	Cohort	Months: One Month Later	Months: Two Months Later	Months: Three Months Later
Tachycardia	Ethnicity	9.86e+00	6.45e+00	1.34e+01
	Agegroup	6.00e+01	6.15e+01	6.99e+01
	Hospital Admission	1.89e+02	1.43e+02	7.73e+01
	Smoking Status	5.83e+01	4.26e+01	2.71e+01
Tremor	Gender	1.84e-01	1.88e-02	1.32e-01
	Ethnicity	6.38e+00	5.42e+00	6.67e+00
	Agegroup	1.51e+01	9.52e+00	1.19e+01
	Hospital Admission	4.64e+01	3.43e+01	3.43e+01
	Smoking Status	6.03e+00	5.33e+00	1.42e+01
Vomiting	Gender	7.20e-02	7.54e+00	3.11e+00
	Ethnicity	4.73e+00	3.34e+00	4.62e-01
	Agegroup	2.41e+01	2.40e+01	2.53e+01
	Hospital Admission	2.86e+01	3.86e+01	3.27e+01
	Smoking Status	2.49e+01	2.41e+01	1.22e+01
Weightgain	Gender	1.17e+01	1.67e+01	3.55e+00
	Ethnicity	3.75e+00	4.75e+00	5.07e+00
	Agegroup	5.43e+01	6.11e+01	5.27e+01
	Hospital Admission	8.05e+01	6.91e+01	9.23e+01
	Smoking Status	2.81e+01	4.37e+01	2.71e+01

P value
<0.05
>0.05

Chi Square ($\chi 2$) statistics are shown i in the results and broken down into ADR, cohort and further broken down into three months after starting the drug Clozapine.
Adjustment for multiple comparisons: **Bonferroni**.
The mean difference is significant at the **0.05 level** (95% confidence interval for difference). The results in Red shows statistically significant p values.