



BMJ Open Cohort profile: the East London Health and Care Partnership Data Repository: using novel integrated data to support commissioning and research

Amy Ronaldson ^{1,2}, Evangelos Chandakas,³ Qiongwen Kang,³ Katie Brennan,³ Aminat Akande,³ Irene Ebyarimpa,³ Eleanor Wyllie,⁴ George Howard,⁴ Richard Fradgley,⁵ Mark Freestone,² Kamaldeep Bhui ^{2,6}

To cite: Ronaldson A, Chandakas E, Kang Q, *et al.* Cohort profile: the East London Health and Care Partnership Data Repository: using novel integrated data to support commissioning and research. *BMJ Open* 2020;**10**:e037183. doi:10.1136/bmjopen-2020-037183

► Prepublication history for this paper is available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2020-037183>).

Received 22 January 2020
Revised 04 August 2020
Accepted 07 August 2020



© Author(s) (or their employer(s)) 2020. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to

Dr Amy Ronaldson;
amy.ronaldson@kcl.ac.uk

ABSTRACT

Purpose The East London Health and Care Partnership (ELHCP) Data Repository was established to support commissioning decisions in London. This dataset comprises routine clinical data for the general practitioner (GP)-registered populations of two London boroughs, Tower Hamlets and City and Hackney, and provides a rich source of demographic, clinical and health service use data of relevance to clinicians, commissioners, researchers and policy makers. This paper describes the dataset in its current form, its representativeness and data completeness.

Participants There were 351 749 and 344 511 members of the GP-registered population in the two boroughs, respectively, for the financial year 2017/2018. Demographic information and prevalence data were available for 9 mental health and 15 physical health conditions. Prevalence rates from the cohort were compared with local and national data. In order to illustrate the health service use data available in the dataset, emergency department use across mental health conditions was described. Information about data completeness was provided.

Findings to date The ELHCP Data Repository provides a rich source of information about a relatively young, urban, ethnically diverse, population within areas of socioeconomic deprivation. Prevalence data were in line with local and national statistics with some exceptions. Physical health conditions were more common in those with mental health conditions, reflecting that comorbidities are the norm rather than the exception. This has implications for integrated care. Data completeness for risk factors (eg, blood pressure, cholesterol) was high in patients with long-term conditions.

Future plans The data are being further cleaned and evaluated using imputation, Bayesian and economic methods, principally focusing on specific cohorts, including type II diabetes, depression and personality disorder. Data continue to be collected for the foreseeable future to support commissioning decisions, which will also enable more long-term prospective analysis as data become available at the end of each financial year.

Strengths and limitations of this study

- This dataset will allow for the analysis of a relatively young, urban, ethnically diverse, considerably deprived population—groups that are hard to research with traditional survey methods.
- The data are linked across care settings, allowing for the evaluation of integrated interventions.
- Large amounts of data around health service utilisation and associated spend means this dataset can be used to support healthcare commissioning and health service delivery.
- Reliance on primary care data for diagnostic information means that prevalence rates can be underestimated and certain information can be missed.
- As is the case with all routine clinical data, the dataset requires extensive cleaning to remove errors.

INTRODUCTION

The use of observational data is fundamental to healthcare service provision and health research, and increasingly large amounts of routinely collected clinical data are becoming available. The main advantage of these datasets is their comprehensive nature and large patient numbers.¹ The development of these large routine datasets also allow for the application of big data analytic techniques within health research.² These techniques can evaluate population characteristics, identify risk factors and allow for the development of predictive models relating to clinical outcomes.^{1 3} Moreover, these datasets identify areas of high health service utilisation and high-cost patients, which might support decision making around health service delivery for commissioners, policy makers and healthcare providers. To date, several large datasets comprising routine

clinical data have been developed in the UK; these are being used to inform clinical practice and healthcare delivery.^{4–6}

Recently, a large-scale routine clinical dataset has been developed in the Tower Hamlets borough of London, UK, comprising data from the general practitioner (GP)-registered populations of Tower Hamlets and its neighbouring borough, City and Hackney. The East London Health and Care Partnership (ELHCP) Data Repository was developed primarily to support commissioning decisions and health service delivery in East London and comprises rich information relevant to population health and clinical research. These areas of London are unique innercity areas with diverse populations and high levels of deprivation. Tower Hamlets is home to the largest Bangladeshi community in England.⁷ It also has the highest rate of poverty, child poverty and unemployment of any London borough.⁸ What might be of interest to researchers of severe mental illness (SMI) is that the borough of City and Hackney has the highest incidence of psychosis in England.⁹

The aim of this cohort profile is to introduce and describe the dataset in Tower Hamlets and City and Hackney, comparing population demographics, prevalence rates of mental and physical health conditions, health service utilisation and data completeness. We believe that this cohort profile might facilitate the development of other similar datasets. Moreover, these data will prove useful for clinicians, researchers, commissioners and policy makers, particularly those with an interest in understanding health in urban, diverse and deprived areas.

The development of the ELHCP Data Repository

The ELHCP Data Repository was developed by Tower Hamlets Clinical Commissioning Group (CCG) (<https://www.towerhamletscg.nhs.uk/>) and North East London Commissioning Support Unit (NEL CSU) (<http://www.nelcsu.nhs.uk/>), with funding provided by Tower Hamlets CCG and Tower Hamlets Together (THT) Vanguard (<https://www.towerhamletstogether.com/>). Development of the dataset started in 2015 and efforts to expand the dataset (eg, to all areas in North East London) and improve data quality are ongoing.

The dataset comprises up-to-date patient-level data for GP-registered populations in the London boroughs of Tower Hamlets and City and Hackney and is updated every 3 months. All patient data are depersonalised and linked by replacing patient National Health Service (NHS) numbers with unique ID numbers. Data pertaining to geographic area (ie, Lower Super Output Area (LSOA)) and other population-level data are drawn from data aggregated by the Office of National Statistics. Presently, secure data streams are in place for: acute (inpatient, outpatient and emergency) services; community services; mental health services; substance misuse services and primary care services (eg, diagnostic information, prescribing). Efforts are underway to establish secure data streams for social care data, public health data (eg, sexual health) and urgent care data (eg, walk-in-centres). However, challenges surrounding linking social care data mean that sensitive information relating to housing and benefits will not be integrated into the dataset.

As illustrated in [figure 1](#), data for acute, community and mental health services flow to the CCG via NHS digital, The Data Services for Commissioners Regional Office

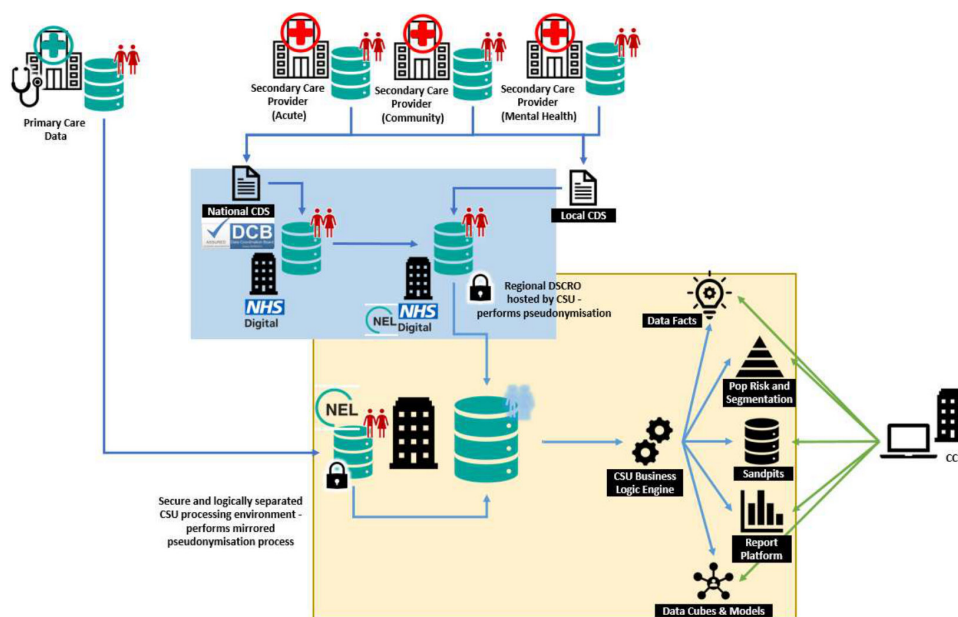


Figure 1 Diagram of data flows from healthcare providers to tower Hamlets and City Hackney CCGs. CCG, Clinical Commissioning Group; CDS, Commissioning Data Sets; CSU, Commissioning Support Unit; DCB, Data Coordination Board; DSCRO, Data Services for Commissioners Regional Office; NEL, North East London; NHS, National Health Service.

(DSCRO) and local CSUs. Pseudonymisation of the data is performed by the DSCRO hosted by the CSU. The data then flows to the local CSU, which formats and organises the data for use at the CCG. Primary care data flows to NEL CSU where a pseudonymisation process is carried out which mirrors that performed by the DSCRO. These data then flow to the local CSU alongside the acute, community and mental health data where they are added to the ELHCP Data Repository.

METHODS

This cohort profile is a descriptive piece detailing the following:

1. Demographic information (eg, age, sex, ethnicity and deprivation).
2. Prevalence rates of mental and physical health conditions.
3. Data relating to health service utilisation in patients with mental health conditions
4. Information about data completeness.

All data presented in this profile were from financial year 2017/2018.

Population demographics

Demographic information for these populations included age, sex, ethnicity and deprivation index. There was also information relating to body mass index (BMI) and smoking status. Information about age, sex and ethnicity came from primary care records. Patients were classified into nine ethnic groups: white or not stated; Indian; Pakistani; Bangladeshi; other Asian; Black Caribbean; Black African; Chinese; other ethnic group. Deprivation was measured using the Index of Deprivation. This measure is provided for LSOAs—small areas that are focused on social homogeneity and a defined population size (1000–1500 residents).¹⁰ The ELHCP dataset allocates each LSOA in Tower Hamlets and City and Hackney to a decile of deprivation with ‘1’ being the most deprived decile. Information relating to BMI and smoking status came from primary care records.

Mental and physical health conditions

Primary care data were used to identify and define 9 mental health conditions and 15 physical health conditions. These conditions are listed in [box 1](#). Variables pertaining to each condition were formulated by NEL CSU using EMIS Read Codes (V.2 and V.3). Lists of the codes used to formulate each variable are available on request. All physical health variables were deemed long-term conditions, and therefore, a patient who had ever had a primary care diagnosis recorded was included in the physical health cohort. This was the same for SMI (ie, schizophrenia, psychosis, bipolar disorder), personality disorder, anorexia, bulimia, dementia and learning disabilities. As depression, low mood and anxiety can potentially be more episodic conditions, further work was done to determine the timelines that should be

Box 1 Mental and physical health conditions recorded in primary care records

Mental health conditions

Depression; low mood*; anxiety; severe mental illness; personality disorder; anorexia; bulimia; dementia; learning disability.

Physical health conditions

Type 1 diabetes, type 2 diabetes, asthma, chronic obstructive pulmonary disorder, cancer (excluding basal cell carcinoma), stroke, coronary heart disease, heart failure, atrial fibrillation, hypertension, peripheral artery disease, chronic kidney disease, liver/pancreas complications (eg, hepatitis, liver disease, chronic pancreatitis), osteoporosis, epilepsy.

*Low mood is often recorded in primary care in place of depression and is therefore important to consider in any analyses relating to mental health.

used to define these cohorts. Based on local clinical and academic advice and available evidence,¹¹ we decided to create two patient cohorts for each conditions—patients who had ever received a diagnosis and patients who had a diagnosis recorded in the last 12 months.

Looking at the impact of the method of definition on activity in secondary care services indicated that the preferable way to define the depression, low mood and anxiety cohorts was to include those who had ever had a diagnosis recorded in primary care. This was partly because only looking at those who had received a diagnosis in the past 12 months meant that we omitted a considerable proportion of patients who likely had common mental health disorders (CMDs, ie, depression and/or anxiety) from our analyses. If required, it is possible to apply defined periods of time to conditions where full prolonged remission or cure is a possibility, for example, cancer.

Prevalence rates for each condition in both Tower Hamlets and City and Hackney were calculated from the data. We assessed population prevalence in the dataset and compared with both the local and the national population from other data sources. The prevalence rates of all 15 physical health conditions across 6 of the mental health conditions were also calculated in order to illustrate the bidirectional links between mental and physical health.

Health service utilisation

Health service utilisation data outlined in this profile covered use of emergency services for exemplary purposes. Data relating to specific mental health conditions were described. Health service utilisation data for elective and non-elective inpatient admissions, and outpatient appointments are also available in the ELHCP Data Repository. These data come from the Secondary Uses Service (SUS) dataset and provide number of attendances/admissions/appointments per patient per year. There is also data pertaining to associated spend (eg, spend per activity, spend per patient per year), which is currently being validated and is therefore not presented in this paper.

Data completeness

As the ELHCP Data Repository comprises routinely collected data from several different data flows, there is considerable variation in completeness across variables. Completeness of certain variables appears to be better in patients with diagnoses of mental/physical health conditions compared with those with no diagnoses. We used type 2 diabetes and chronic obstructive pulmonary disorder (COPD) as exemplar physical conditions, and depression and SMI as exemplar mental conditions to demonstrate how data quality differs across patient groups. We provided completeness counts on several demographic and clinical variables including age, sex, deprivation level, ethnicity, smoking status, alcohol use, BMI, weight (kg), systolic blood pressure, diastolic blood pressure and cholesterol. All variables were from primary care datasets. Alcohol use was measured using a composite variable of combined scores from the Alcohol Use Disorders Identification Test (AUDIT) and AUDIT-consumption (AUDIT-C) completed in primary care.

COHORT DESCRIPTION AND FINDINGS TO DATE

Population demographics

The GP-registered populations for Tower Hamlets and City and Hackney in financial year 2017/2018 were 351 749 and 344 511, respectively. See [table 1](#) for a summary of population demographics for both Tower Hamlets and City and Hackney. Data for age, sex and smoking status were recorded for the complete GP-registered populations in both East London boroughs. Deprivation index was recorded for most of the population (99.8% complete in both boroughs). Ethnicity data was recorded to a good level of completeness at 86.6% and 84.6% in Tower Hamlets and City and Hackney, respectively. BMI data was more problematic and was recorded in the primary care records for less than a quarter of GP-registered patients (Tower Hamlets: 23.4%, City and Hackney: 24.0%).

Demographic data suggested that both Tower Hamlets and City and Hackney are 'young' boroughs relative to English national data¹² (median age 39 years) with mean ages of 32 (median=31 years) and 34 years (median=33 years), respectively. In terms of ethnicity, the majority of the GP-registered population in Tower Hamlets (38.2%) and City and Hackney (46.8%) were of white (or not stated) ethnicity. A considerable proportion of the population of Tower Hamlets were of Bangladeshi ethnicity (27.4%) and a significant amount of the City and Hackney population belonged to 'other ethnic group' (16.8%). Most of the GP-registered population in both East London boroughs were living in areas of high deprivation: in Tower Hamlets and City and Hackney, 58.8% and 54.2% of the population, respectively, were in deciles 1 and 2, indicative of high deprivation. Approximately 20% of the population were recorded in primary care as being smokers. Although BMI is poorly recorded in primary care records, mean BMI scores indicated that

Table 1 Population demographics for the GP-registered populations of Tower Hamlets and City and Hackney

	Tower Hamlets	City and Hackney
GP-registered population, n (%)	351 749 (100)	344 511 (100)
Age (years) (M±SD)	32±17.97	34±19.47
Sex, n (%)		
Male	178 088 (50.63)	168 371 (48.87)
Female	173 656 (49.37)	176 134 (51.13)
Ethnicity, n (%)		
White or not stated	134 411 (38.21)	161 069 (46.75)
Indian	9295 (2.64)	9067 (2.63)
Pakistani	3547 (1.01)	2707 (0.79)
Bangladeshi	96 418 (27.41)	5877 (1.71)
Other Asian	8151 (2.32)	6189 (1.80)
Black Caribbean	4281 (1.22)	16 052 (4.66)
Black African	14 034 (3.99)	30 037 (8.72)
Chinese	10 435 (2.97)	4137 (1.20)
Other ethnic group	23 934 (6.80)	56 140 (16.29)
Deprivation index, n (%)		
High deprivation*	206 723 (58.77)	186 784 (54.22)
Low deprivation†	4955 (1.41)	3134 (0.91)
Smokers, n (%)	69 627 (19.79)	72 371 (21.01)
BMI (kg/m ²) (M±SD)	26.34±8.28	27.06±8.58

*High deprivation refers to the GP-registered population in deciles 1 and 2.

†Low deprivation refers to the GP-registered population in deciles 9 and 10.

BMI, body mass index; GP, general practitioner.

the GP-registered population on average were overweight (Tower Hamlets: 26.3±8.1; City and Hackney: 27.1±8.6).

Prevalence rates of mental and physical health conditions

Prevalence rates for GP-coded diagnoses of mental health conditions in both Tower Hamlets and City and Hackney are provided in [figure 2](#). Prevalence rates were similar across both boroughs for all conditions. Comparative local and national prevalence data for each condition is provided in [table 2](#). We did not provide age-standardised and sex-standardised prevalence rates from the ELHCP Data Repository to allow direct comparison with unstandardised local prevalence data from Public Health England (PHE). The GP-coded prevalence of CMDs was 16.8% and 17.5% in Tower Hamlets and City and Hackney, respectively. In this case, CMD comprised diagnoses of depression, anxiety, as well as low mood. Low mood is often recorded both in place of and alongside depression in primary care and is therefore important to consider in a measure of CMD. For example, in Tower Hamlets of those with a diagnosis of depression, 33% also had a diagnosis of low mood. CMD prevalence was in line with data from the latest Adult Psychiatric Morbidity Survey (APMS), which reported that 17% of the English

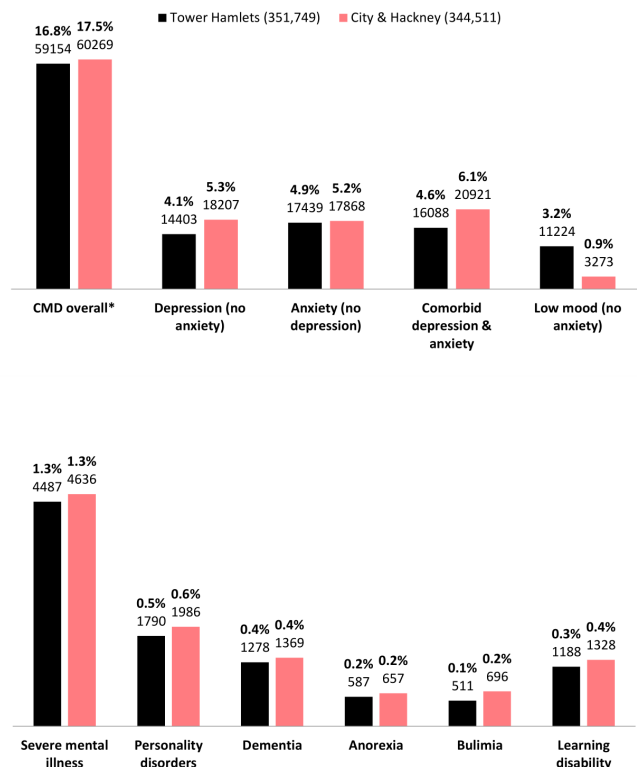


Figure 2 GP-coded prevalence rates of mental health conditions in Tower Hamlets and City and Hackney, financial year 2017/2018. *CMD incorporates depression, anxiety and low mood. CMD, common mental health disorder; GP, general practitioner.

population have a CMD.¹³ However, the GP-coded prevalence of depression in Tower Hamlets (8.7%) and City and Hackney (11.4%) was higher than PHE prevalence data¹⁴ from the same areas from the same financial year (7.3%, 9.6%). This could be due to the GP-coded prevalence in this profile reflecting diagnoses of depression

that were ever recorded, which likely led to somewhat inflated rates.

GP-coded prevalence of SMI (1.3%) from the ELHCP Data Repository was in line with data from the PHE Psychosis Data Report from 2016,⁹ which reported prevalence rates of SMI for both Tower Hamlets and City and Hackney at 1.32% and 1.36%, respectively. Prevalence rates for learning disability were also in line with local and national data provided by PHE.¹⁴ However, there were discrepancies between the GP-coded prevalence of personality disorder, eating disorders and dementia in the ELHCP Data Repository and local and national PHE¹⁴ and APMS¹³ data. What this suggests is that these conditions might be under-recorded in primary care in these boroughs.

Prevalence rates for GP-coded diagnoses of physical health conditions in the ELHCP Data Repository are provided in figure 3. Prevalence rates were similar across boroughs for most conditions. However, primary care records showed that levels of type 2 diabetes were higher in Tower Hamlets and levels of hypertension were higher in City and Hackney. Comparative local and national prevalence data for each condition is provided in table 3. Local and national prevalence data was sourced from PHE public health profiles.¹⁴ Overall, GP-coded prevalence of physical health conditions from the ELHCP Data Repository seems to be in line with PHE local data for the same areas. There is one notable discrepancy, however. Levels of asthma were much higher in the ELHCP Data Repository. This might be due to overdiagnosis of asthma in primary care, which has been reported in both adults¹⁵ and children.¹⁶ Moreover, PHE prevalence data is taken directly from Quality Outcomes Framework (QOF) registers. Prevalence data from QOF registers might differ from data from other sources due to coding or definitional issues, as well as population churn. Within the ELHCP Data Repository, 'liver disease' comprises many liver complications, for example, hepatitis, fatty liver disease, alcoholic liver disease. Therefore, it was not

Table 2 GP-coded prevalence of mental health conditions in Tower Hamlets and City and Hackney compared with local and national prevalence data (sources: APMS 2014, PHE Public Health Profiles, PHE Psychosis Data Report 2016)

Condition	GP-coded prevalence				Local prevalence		National prevalence
	Tower Hamlets	95% CI	City and Hackney	95% CI	Tower Hamlets	City and Hackney	England
CMDs	16.8%	16.69 to 16.94	17.5%	17.37 to 17.62	15.4% ¹⁴	17% ¹⁴	17% ¹³
SMI	1.3%	1.24 to 1.31	1.3%	1.31 to 1.38	1.32% ⁹	1.36% ⁹	0.5%–1.5% ⁹
Personality disorder	0.5%	0.49 to 0.53	0.6%	0.55 to 0.60	–	–	13.7% ^{13*}
Eating disorders	0.3%	0.29 to 0.33	0.4%	0.37 to 0.41	6.9% ¹⁴	7.4% ¹⁴	6.9% ¹⁴
Dementia	0.4%	0.34 to 0.38	0.4%	0.38 to 0.42	5.25% ¹⁴	4.37% ¹⁴	4.33% ¹⁴
Learning disability	0.3%	0.32 to 0.36	0.4%	0.36 to 0.41	0.3% ¹⁴	0.4% ¹⁴	0.5% ¹⁴

*Percentage of sample who scored ≥ 4 on the Standardised Assessment of Personality–Abbreviated Scale (SAPAS).

APMS, Adult Psychiatric Morbidity Survey; CMDs, common mental health disorders; GP, general practitioner; PHE, Public Health England; SMI, severe mental illness

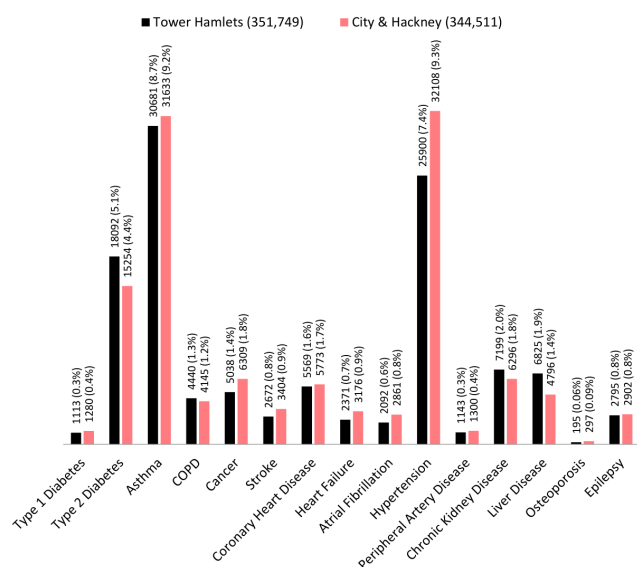


Figure 3 General practitioner-coded prevalence rates of physical health conditions in Tower Hamlets and City and Hackney, financial year 2017/2018. COPD, chronic obstructive pulmonary disorder.

possible to directly compare prevalence data relating to liver disease with local and national statistics as PHE provide individual prevalence rates for each liver disease subtype.

Prevalence of physical health conditions among mental health conditions

Mental and physical health are inextricably linked.¹⁷ Having a mental health condition predisposes an individual to develop problems with their physical health. At the same time, living with a physical health condition can greatly impact mental health. Figure 4A,B illustrates the prevalence of physical health conditions among mental health conditions in the ELHCP Data Repository. Mental conditions included in this analysis were depression, low mood, anxiety, SMI and personality disorder.

Results indicate that levels of all 15 physical health conditions were increased in those with mental health conditions. The distribution of the prevalence of physical health conditions among mental health conditions was similar across boroughs. The most pronounced increases in physical health conditions were seen for those flagged by their GP as having depression, SMI and personality disorder. Some striking findings included the increase in levels of type 2 diabetes and hypertension among patients flagged by their GP as having SMI. Type 2 diabetes is known to be elevated in patients with SMI.^{18 19} This increase might be brought about through use of antipsychotic medications known to be both diabetogenic and linked with weight gain,²⁰ as well as through social deprivation, poor nutrition, chronic stress and lifestyle changes.²¹ Associations between SMI and hypertension reported in the literature have been more mixed.²¹

Table 3 GP-coded prevalence of physical health conditions in Tower Hamlets and City and Hackney compared with local and national prevalence data (source: PHE Public Health Profiles)

Condition	GP-coded prevalence		Local prevalence		National prevalence
	Tower Hamlets	City and Hackney	Tower Hamlets	City and Hackney	England
Type 1 diabetes*	0.3%	0.4%	6.7%	5.7%	6.8%
Type 2 diabetes*	5.1%	4.4%			
Asthma	8.7%	9.2%	4.3%	4.4%	5.9%
COPD	1.3%	1.2%	1.2%	1.1%	1.9%
Cancer	1.4%	1.8%	1.1%	1.4%	2.7%
Stroke	0.8%	0.9%	0.9%	0.7%	1.8%
CHD	1.6%	1.7%	1.5%	1.5%	3.1%
Heart failure	0.7%	0.9%	0.4%	0.7%	0.8%
Atrial fibrillation	0.6%	0.8%	0.6%	0.7%	1.9%
Hypertension	7.4%	9.3%	7.5%	9.3%	13.9%
PAD†	0.3%	0.4%	0.96%	1.01%	1.16%
CKD	2.0%	1.8%	2.5%	2.1%	4.1%
Liver disease‡	1.9%	1.4%	–	–	–
Osteoporosis	0.06%	0.09%	–	–	0.6%
Epilepsy	0.8%	0.8%	0.4%	0.5%	0.8%

*PHE does not provide prevalence rates for diabetes subtypes. Therefore, local and national prevalence rates are for diabetes overall.

†Local and national prevalence data for PAD from PHE were based on modelled estimated prevalence.

‡It was not possible to provide local and national prevalence rates for liver disease as data is provided separately by PHE for disease subtype (eg, hepatitis C, fatty liver disease, alcoholic liver disease).

CHD, coronary heart disease; CKD, chronic kidney disease; COPD, chronic obstructive pulmonary disorder; GP, general practitioner; PAD, peripheral artery disease; PHE, Public Health England.

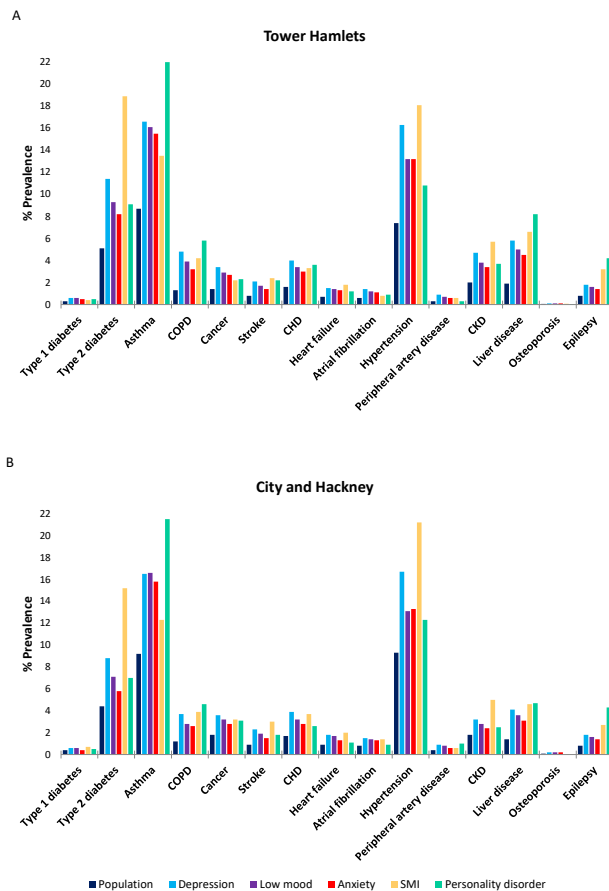


Figure 4 (A) Prevalence of physical health conditions among mental health conditions in Tower Hamlets. (B) Prevalence of physical health conditions among mental health conditions in City and Hackney. CHD,coronary heart disease; CKD, chronic kidney disease; COPD,chronic obstructive pulmonary disorder.

Another striking finding that occurred in both boroughs was the high prevalence of asthma for people with a primary care diagnosis of personality disorder. This may

be linked to smoking, which is known to be increased in patients with personality disorder.²² Personality disorder has been linked to hypochondriasis and perceived breathlessness,²³ which might also be contributing factors to higher levels of asthma diagnosis in primary care for these patients.

Accident and emergency attendances across mental health conditions

We used accident and emergency (A&E) as an example of a health service utilisation outcome measure to illustrate some of the activity data available in the ELHCP Data Repository. Data relating to use of A&E across both boroughs are detailed in table 4. Based on crude attendance rates (per 1000 population), patients with mental health conditions attended A&E 54% and 76% more in Tower Hamlets and City and Hackney, respectively, compared with those without GP-recorded mental health conditions. Increased attendance rates were particularly pronounced in patients with dementia, personality disorder and SMI. Similar health service utilisation data exists relating to inpatient admissions and use of outpatient services.

Data completeness

The ELHCP Data Repository relies on data inputted by healthcare provider organisations (GP practices, clinical coders in acute care settings, etc). It is comprised of routinely collected data that can vary in completeness across variables. Examination of four example physical and mental health conditions illustrated how data completeness varies across patient groups. Table 5 details completeness on key demographic and clinical variables for patients with type 2 diabetes, COPD, depression and SMI in Tower Hamlets. These data were all recorded in primary care databases. Data pertaining to age, sex, deprivation level, ethnicity and smoking status were complete for all patients included in the analysis. Data relating to alcohol use and clinical factors differed between those

Table 4 The impact of mental health conditions on A&E attendances, financial year 2017/2018

Activity (attendances)	Total attendances		% Share		Crude rate (per 1000 population)			
	TH	C&H	TH	C&H	TH	95% CI	C&H	95% CI
	Population	159201	161641	100	100	452	412 to 495	469
All non-mental health	118272	116816	74.3	72.3	412	373 to 453	413	375 to 454
All mental health	40929	44825	25.7	27.7	635	587 to 686	728	677 to 782
CMDs	31625	34215	19.9	21.2	535	491 to 582	568	523 to 616
SMI	3800	4383	2.4	2.7	847	791 to 906	945	886 to 1007
Personality disorder	2529	2914	1.6	1.8	1413	1341 to 1488	1467	1393 to 1543
Eating disorder	584	549	0.4	0.3	554	509 to 602	424	385 to 466
Dementia	1595	1471	1.0	0.9	1248	1180 to 1319	1075	1012 to 1141
Learning disability	796	1293	0.5	0.8	670	621 to 722	974	914 to 1037

A&E, accident and emergency; C&H, City and Hackney; CMDs, common mental health disorders; SMI, severe mental illness; TH, Tower Hamlets.

with and without health conditions, in that the data was considerably more complete for patients registered as having a condition. This improvement in completeness is likely down to the implementation of financial incentive schemes introduced as part of the UK's QOF to improve clinical management of patients with chronic conditions.^{4 24} What this illustrates is that the ELHCP Data Repository is well-suited to perform research on individual physical and mental health conditions using primary care data, where data completeness is to a high standard.

Patient and public involvement

The primary purpose of this project was for the health and care sector to use for system planning and effective commissioning, and therefore public stakeholders in this case comprised primary care managers, commissioners (GPs), public health experts and clinicians in primary and secondary care. A clinical expert group was convened consisting of 15 clinicians, commissioners and public health experts from primary care and secondary care (mental health) services. This group met on a quarterly basis during initial profiling of the cohort and provided feedback into the analysis and emerging clinical and demographic characteristics.

Data security

The dataset contains sensitive patient-level data; therefore, it was and remains important to ensure appropriate protections are in place. All data are deidentified, with no information such as name, home address or NHS number. Also, certain sensitive health data such as records relating to trauma and abuse, or specific sexual health data, are not visible.

To further preserve anonymity and to ensure proportionate and appropriate use of the data, access to records is limited to only being able to view aggregated records where there is a minimum of five individuals. Searches where there are fewer than five records are suppressed. Access to the complete deidentified records is further limited to specific individuals who are employed by, or under contract with, the health and care sector have knowledge of how to use the data effectively and have undergone appropriate training.

This dataset has been compiled for the health and care sector to use for system planning. Currently, plans are being made to deploy the data into an electronic platform for internal use: recognising the possible value this data holds for researchers, there may be scope to support analysis deemed to be in the public interest, with appropriate data protections in place. Currently, permissions must be sought on a case-by-case basis and approved by relevant IG governance groups.

STRENGTHS AND LIMITATIONS

Strengths

One of the main strengths of the dataset is that it provides the opportunity to analyse a relatively young, urban, ethnically diverse and considerably deprived population: groups that are hard to reach with traditional survey methods. Given patient records are linked across care settings, the data can be used to evaluate the impacts of integrated interventions. In this profile, we have illustrated that the ELHCP Data Repository is representative in terms of the prevalence of most mental and physical health conditions compared with other sources of local and national prevalence data supporting its suitability for

Table 5 ELHCP data completeness (%) across physical and mental health conditions in Tower Hamlets, financial year 2017/2018

Variables	T2DM	No T2DM	COPD	No COPD	Depression	No depression	SMI	No SMI
	% complete	% complete	% complete	% complete	% complete	% complete	% complete	% complete
Age	100	100	100	100	100	100	100	100
Sex	100	100	100	100	100	100	100	100
Deprivation	100	100	100	100	100	100	100	100
Ethnicity	100	100	100	100	100	100	100	100
Smoking status	100	100	100	100	100	100	100	100
Alcohol use	94.6	51.1	91.8	52.8	73.1	51.4	89.2	52.8
BMI	99.3	72.6	99.3	73.7	95.4	71.9	97.5	73.7
Weight	99.4	84.6	99.5	85.2	96.7	84.3	98.2	85.2
SBP	99.6	68.8	99.6	69.9	96.4	67.9	98.5	69.9
DBP	99.6	68.7	99.6	69.9	96.3	67.9	98.5	69.9
Cholesterol	99.4	28.8	96.9	31.6	64.9	29.3	84.6	37.8

BMI, body mass index; COPD, chronic obstructive pulmonary disorder; DBP, diastolic blood pressure; ELHCP, East London Healthcare Links Project; SBP, systolic blood pressure; SMI, severe mental illness; T2DM, Type two diabetes mellitus.

real-world research and decision making. This dataset also allows for the analysis of comorbidities and multi-morbidity. Moreover, there is a large amount of data available around health service utilisation and associated spend meaning that the dataset can be used to understand what drives use and spend and support healthcare commissioning and health service delivery.

The dataset is in regular use by commissioners to better understand the health and care needs of the population, highlight where there is scope for improved patient care and evaluate the impact of health and care interventions of patients and the system. This has provided commissioners, and others working on health and care system integration, with more sophisticated evidence and analysis to inform decisions about where to target improvement efforts to maximise benefit to both patients and the system. The findings presented in this cohort profile have highlighted where links between mental and physical health are most prevalent in the population and helped to shape commissioning priorities accordingly. For example, analysis highlighting mental and physical comorbidity supported conversations about expanding Improving Access to Psychological Therapies (IAPT) services within long-term condition pathways, noting that it may be particularly impactful for people with type 2 diabetes, respiratory conditions and cancer. Furthermore, over the next year, Tower Hamlets is investing in a new 24/7 core fidelity crisis service and the transformation of community mental health services to become embedded in primary care networks. These new models will support the assessment and treatment of patients at the right place at the right time. The ELHCP Data Repository described in this cohort profile will be an important tool in evaluating the impact these service changes have on patients and the healthcare system.

In terms of completeness, this paper highlighted that data were considerably more complete for patients with long-term mental or physical illness (ie, type 2 diabetes, COPD, depression, SMI) compared with patients without. This supports the suitability of the dataset for the analysis of patients with long-term conditions.

Limitations

In terms of limitations associated with the dataset, in some cases prevalence data was not in line with local and national data, meaning that it is possible some conditions are under-recorded in primary care, that is, personality disorder, eating disorders and dementia. Also, as prevalence data relies on GP diagnoses, it is not possible to gauge disease duration in patients as the date in which a diagnosis is first recorded might not be the date the patient received the diagnosis. For example, the diagnosis might have been recorded on the day the patient registered with their GP, reflecting that date rather than the date of diagnosis. In terms of data completeness, although this is good for patients with long-term conditions, there is considerable missingness among ‘healthy’ patients in the dataset. This has also been reported in a study of a

large primary care databases in the UK, which partially attributed the discrepancy in missing data to GP incentive schemes.²⁵ It is possible to resolve this with imputation to some degree. As is the case with all routine clinical data, the dataset requires extensive cleaning to remove duplicate entries and spurious values out of clinical range (eg, BMI, blood pressure) prior to analysis. Another limitation involves the amount of data available. Currently, there are two financial years of data that are at an acceptable standard for research and/or decision support. New data continues to be collected and flow into this data set, so the time series available will increase over time. A further limitation surrounds the ability to measure population mobility in a dataset of this kind. One way is to look at the length of GP registration as a proxy for this, but this is problematic as it is not possible to determine whether the mobility is within the borough or whether patients have come from outside.

CONCLUSIONS AND FUTURE COLLABORATIONS

This cohort profile describes a unique patient-linked dataset in East London, which can be used to facilitate the development of similar datasets in other parts of the UK. Moreover, this dataset provides a unique opportunity to analyse a relatively young, urban, diverse and deprived population in East London. The dataset is growing year on year and data quality should improve over time, making this an invaluable resource for researchers, clinicians, commissioners and policy makers. The data are being further cleaned and evaluated using imputation, Bayesian and economic methods, including engagement with machine learning work funded by the Alan Turing Institute to improve and automate the analysis of large-scale electronic medical records (KNIFE project). Queries regarding data access should be directed to Katie Brennan (katie.brennan1@nhs.net) or Nathan Cheetham (nathan.cheetham1@nhs.net) at the NHS Tower Hamlets CCG.

Author affiliations

¹Department of Psychological Medicine, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK

²Centre for Psychiatry, Wolfson Institute of Preventive Medicine, Barts & The London School of Medicine, Queen Mary University of London, London, UK

³NHS Tower Hamlets Clinical Commissioning Group, London, UK

⁴Healthy London Partnership, London, UK

⁵East London NHS Foundation Trust, London, UK

⁶Department of Psychiatry, University of Oxford, Oxford, UK

Correction notice The title of the paper has been corrected.

Twitter Amy Ronaldson @DrRonaldson, Qiongwen Kang @QiongwenKang and Kamaldeep Bhui @ksbhui

Contributors AR drafted the manuscript with input and critical revisions from all authors. QK, KBrennan, EW, GH, RF, MF and KBhui provided supervision throughout. KBhui is the principal investigator and grant holder. Analysis of the data was performed by AA, EC and IE. All authors reviewed the paper before submission.

Funding Funding for this project was granted by East London NHS Foundation Trust (ELFT) (184091163), the Tower Hamlets Together Board, the Health London Partnership Mental Health Transformation Board, and the City and Hackney

Transformation Board. This report is independent research supported by the National Institute for Health Research ARC North Thames.

Competing interests MF received research grants from East London NHS Foundation Trust and The Alan Turing Institute during the conduct of this study.

Patient and public involvement Patients and/or the public were involved in the design, or conduct, or reporting, or dissemination plans of this research. Refer to the Methods section for further details.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data may be obtained from a third party and are not publicly available. The data is housed at the NHS Tower Hamlets Clinical Commissioning Group. Permission to access the data will need to be requested from Katie Brennan (katie.brennan1@nhs.net) or Nathan Cheetham (nathan.cheetham1@nhs.net) at the NHS Tower Hamlets Clinical Commissioning Group.

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

ORCID iDs

Amy Ronaldson <http://orcid.org/0000-0001-8369-9168>

Kamaldeep Bhui <http://orcid.org/0000-0002-9205-2144>

REFERENCES

- Cook JA, Collins GS. The rise of big clinical databases. *Br J Surg* 2015;102:e93–101.
- Bates DW, Saria S, Ohno-Machado L, et al. Big data in health care: using analytics to identify and manage high-risk and high-cost patients. *Health Aff* 2014;33:1123–31.
- Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. *Health Inf Sci Syst* 2014;2:3.
- Correa A, Hinton W, McGovern A, et al. Royal College of general practitioners research and surveillance centre (RCGP RSC) sentinel network: a cohort profile. *BMJ Open* 2016;6:e011092.
- Herrett E, Gallagher AM, Bhaskaran K, et al. Data resource profile: clinical practice research Datalink (CPRD). *Int J Epidemiol* 2015;44:827–36.
- Perera G, Broadbent M, Callard F, et al. Cohort profile of the South London and Maudsley NHS Foundation trust biomedical research centre (SLAM BRC) case register: current status and recent enhancement of an electronic mental health Record-derived data resource. *BMJ Open* 2016;6:e008721.
- Tower Hamlets Council. Ethnicity in tower Hamlets: analysis of 2011 census data, 2013. Available: https://www.towerhamlets.gov.uk/Documents/Borough_statistics/Ward_profiles/Census-2011/RB-Census2011-Ethnicity-2013-01.pdf [Accessed 7 Mar 2019].
- Trust for London. Poverty and inequality data for tower Hamlets. trust Lond, 2017. Available: </data/boroughs/tower-hamlets-poverty-and-inequality-indicators/> [Accessed 3 Jul 2019].
- Public Health England. *Psychosis data report*, 2016.
- Noble M, McLennan D, Wilkinson K, et al. *English indices of deprivation*. London: Communities and Local Government, 2007.
- John A, McGregor J, Fone D, et al. Case-Finding for common mental disorders of anxiety and depression in primary care: an external validation of routinely collected data. *BMC Med Inform Decis Mak* 2016;16:35.
- GOV.UK. Ethnicity facts and figures: age groups. HM government services, 2018. Available: <https://www.ethnicity-facts-figures.service.gov.uk/uk-population-by-ethnicity/demographics/age-groups/latest#download-the-data> [Accessed 6 Jun 2019].
- NHS Digit. Adult psychiatric morbidity survey: survey of mental health and wellbeing, England, 2014. Available: <https://digital.nhs.uk/data-and-information/publications/statistical/adult-psychiatric-morbidity-survey/adult-psychiatric-morbidity-survey-survey-of-mental-health-and-wellbeing-england-2014> [Accessed 11 Jan 2019].
- Public Health England. Public health profiles. Available: <https://fingertips.phe.org.uk/profile-group/mental-health> [Accessed 18 Jun 2019].
- Aaron SD, Vandemheen KL, Boulet L-P, et al. Overdiagnosis of asthma in obese and nonobese adults. *CMAJ* 2008;179:1121–31.
- Looijmans-van den Akker I, van Luijk K, Verheij T. Overdiagnosis of asthma in children in primary care: a retrospective analysis. *Br J Gen Pract* 2016;66:e152–7.
- Doherty AM, Gaughran F. The interface of physical and mental health. *Soc Psychiatry Psychiatr Epidemiol* 2014;49:673–82.
- Das-Munshi J, Ashworth M, Gaughran F, et al. Ethnicity and cardiovascular health inequalities in people with severe mental illnesses: protocol for the E-CHASM study. *Soc Psychiatry Psychiatr Epidemiol* 2016;51:627–38.
- Ward M, Druss B. The epidemiology of diabetes in psychotic disorders. *Lancet Psychiatry* 2015;2:431–51.
- Gianfrancesco FD, Grogg AL, Mahmoud RA, et al. Differential effects of risperidone, olanzapine, clozapine, and conventional antipsychotics on type 2 diabetes: findings from a large health plan database. *J Clin Psychiatry* 2002;63:920–30.
- Osborn DPJ, Wright CA, Levy G, et al. Relative risk of diabetes, dyslipidaemia, hypertension and the metabolic syndrome in people with severe mental illnesses: systematic review and metaanalysis. *BMC Psychiatry* 2008;8:84.
- Zvolensky MJ, Jenkins EF, Johnson KA, et al. Personality disorders and cigarette smoking among adults in the United States. *J Psychiatr Res* 2011;45:835–41.
- Chetta A, Gerra G, Foresi A, et al. Personality profiles and breathlessness perception in outpatients with different gradings of asthma. *Am J Respir Crit Care Med* 1998;157:116–22.
- Kontopantelis E, Reeves D, Valderas JM, et al. Recorded quality of primary care for patients with diabetes in England before and after the introduction of a financial incentive scheme: a longitudinal observational study. *BMJ Qual Saf* 2013;22:53–64.
- Petersen I, Welch CA, Nazareth I, et al. Health indicator recording in UK primary care electronic health records: key implications for handling missing data. *Clin Epidemiol* 2019;11:157–67.