International Journal of Population Data Science

Journal Website: www.ijpds.org

Sussex Integrated Dataset (SID) Data Resource Profile

Elizabeth Ford[1,*], Jon Elsom[2], Paul Molyneux[3], David Crees[2], Yemisi Bakare[2], and Dan Hughes[4]

Submission History

Submitted:	27/10/2025
Accepted:	16/04/2026
Published:	17/06/2026

[1] Department of Primary Care and Public Health, Brighton and Sussex Medical School, Watson Building, Village Way, Falmer, Brighton, BN1 9PH

[2] East Sussex Healthcare NHS Trust, Conquest Hospital, The Ridge, St Leonards-on-Sea, East Sussex TN37 7RD

[3] GRIP Analytics 3rd Floor 86 - 90 Paul Street, London, United Kingdom, EC2A 4NE

[4] NHS Sussex, Sackville House, Brooks Close, Lewes, East Sussex, BN7 2FZ

Abstract

This paper presents a data resource profile of the Sussex Integrated Dataset (SID). SID is a linked dataset of National Health Service (NHS) and adult social care data from across the geographical footprint of Sussex, in coastal Southeast England, and represents around 1.7 million patients. Living in a predominantly rural and coastal area, the population has more elderly people than the national average and is characterised by large differences in deprivation levels particularly in coastal towns. The city of Brighton and Hove has a large LGBTQ+ population and the largest cohort of people living with HIV outside of London.

First linked in 2019, the dataset contains contributions from all NHS providers across East and West Sussex and the city of Brighton and Hove, including routinely collected data from all Sussex hospital trusts, general practitioners (up to 100% coverage), community and mental health trusts, ambulance data and local authority social care services. Most data sources provide information from 2017, with general practice data going back up to 20 years. Data quality is carefully monitored, demographics are harmonised within a "demographic milestones" table, and patient cohorts living with common long-term conditions are curated using a "semantic layer".

Already used widely within NHS Sussex and by Local Authorities for public health, service evaluation, monitoring and planning, the SID is available more widely to use for research and innovation through the NHS Secure Data Environment Network.

Keywords

health data, linked data, routinely collected data, public health, service evaluation, population health management.

Key Features

- The Sussex Integrated Dataset (SID) covers all NHS and adult social care activity for the population of Sussex, including up to 100% of general practice data.
- It was developed as part of a programme to deliver a single health and care record for each person living in Sussex and to help plan and improve services for the local population.
- SID hosts the de-identified and linked data of the 1.7 million GP registered population of Sussex including data flows from primary care, hospital trusts, mental health trusts, community care, ambulance and adult social care.
- Individual patients are pseudonymised at the data source using an encryption key which creates a unique ID; this means they can be traced across all NHS providers, but their identifiers never leave the healthcare environment.
- Data is now available to approved researchers and projects through application via the NHS Secure Data Environment Network and the National Health Data Research Service.

*Corresponding Author:
Email Address: e.m.ford@bsms.ac.uk (E Ford)

https://doi.org/10.23889/ijpds.v8i6.3382

Background to the Geographical Area

Sussex is a large region on the Southeast coast of England, split into three local government jurisdictions, including two administrative counties, East and West Sussex, and the city of Brighton and Hove as a unitary authority. The combined population of Sussex as of the 2021 UK national census was about 1.7 million [1].

There is considerable economic variation within Sussex. In most areas, economic output is close to or significantly lower than the UK average and is far below the average for Southeast England. There is also serious deprivation in Sussex, comparable to the most deprived UK inner city areas. Some areas of Sussex are in the top five per cent of the most deprived in the UK, and in some areas two-thirds of children are living in poverty. Socioeconomic deprivation is especially concentrated in the seaside towns [2].

Current Population Demographics

As the county is divided into two administrative regions (East and West Sussex) and one unitary authority (Brighton and Hove City), population statistics are compiled separately for these areas.

West Sussex

In the 2021 Census, the population of West Sussex was estimated to be 882,700 people, an increase of 9.4% since 2011 [3]. The largest increase in population since the last census (2011) has been older people (aged 65 years and over) – though only in Chichester, Mid Sussex and Horsham districts. The county has a higher-than-average proportion of people over 50 years. While life expectancy has increased overall, health life expectancy (the years an individual can expect to live in good health) has decreased over the last ten years. This is highly influenced by material deprivation: people in the least deprived areas of West Sussex live, on average, 6 years longer than those living in the most deprived areas [4].

East Sussex

In the 2021 census, the population of East Sussex was estimated to be 545,847 [1]. Compared to England and the Southeast, East Sussex has a much older age profile [5]. 26% of the population comprises people aged 65 or over, compared to 18% for England. 93.9% of the county's population identified as White British, compared to the national average of 81.0%. The region has a mix of high and low deprivation, with 7% of lower super output areas (LSOAs; a small geographical area representing roughly 1700 residents) in the most deprived 10% nationally (Figure 1) [5]. East Sussex has consistently had a higher life expectancy for both men and women than the national average, however, there is a gap in life expectancy between the most and least deprived areas of East Sussex, with the biggest causes of death associated with these gaps being circulatory diseases, cancers, and respiratory diseases. For females, this gap amounts to 4.5 years, while for males it is 6.8 years.

The City of Brighton and Hove

Brighton and Hove has an estimated 276,300 residents (2021 census), with a higher proportion of younger age adults than the two surrounding counties (72% are aged 16 to 66 years (64% South East, 65% England)) and fewer older people (10% are aged 67 to 84 years (16% South East, 15% England))[6]. This is because around 50,000 of the city's residents are made up of students at the two universities. While students may only stay in the city for a few years, they are recommended to register with GP services when they arrive (with a dedicated on-campus, university GP surgery) and use local National Health Services (NHS). More than a quarter of residents (26%) are Black and Racially Minoritised (non-White UK/British). While the city has some very affluent areas, the highest concentration of deprivation is in the Whitehawk, Moulsecoomb and Hollingdean areas which are in the 20% most deprived areas in England [7] (Figure 2). The 2021 UK census found that 10.7% of Brighton and Hove residents identified as lesbian, gay, bisexual, pansexual or another non-straight orientation, and around 1% identified as a gender different from their birth sex [8]; both these proportions are much higher than the England average. In 2022, Brighton and Hove had one of the highest diagnosed HIV (human immunodeficiency virus) prevalence rates in England, with 7.3 per 1,000 people aged 15-59 living with HIV (compared to 2.3 in England) [9].

Healthcare Provision in Sussex

Sussex is served by a large network of NHS providers, including 158 General Practices joined within 38 primary care networks, and seven major hospitals [10] (Table 1). In 2022, following the Health and Social Care Act (2022) all Sussex NHS providers came together into an Integrated Care System run by the integrated care board (ICB), known as NHS Sussex [11]. The NHS Sussex ICB planned and delivered joined-up health and care services across the region, serving 1.7 million people. From April 1, 2026, NHS Sussex and NHS Surrey Heartlands Integrated Care Boards (ICBs) merged to form a single, new ICB (NHS Surrey and Sussex). Digital transformation work in NHS Sussex was also supported by the South Central and West NHS Commissioning Support Unit. NHS Commissioning Support Units (CSU) were organisations providing expert, external support services to NHS commissioners, primarily Integrated Care Boards (ICBs). CSUs helped with administrative, technical, and analytical functions such as data analysis, procurement, HR, and IT.

The Sussex Integrated Dataset:

Foundational Decisions

The Sussex Integrated Dataset (SID) was created in 2019, as part of the Our Care Connected (OCC) programme for NHS Sussex [12]. The vision of the programme is to deliver a single health and care record for each person living in Sussex. The SID is one of three OCC products, the others being Plexus, a platform for allowing practitioner to practitioner communications, and My Health Care Record,

Figure 1: Index of Deprivation in East Sussex (taken from [5])

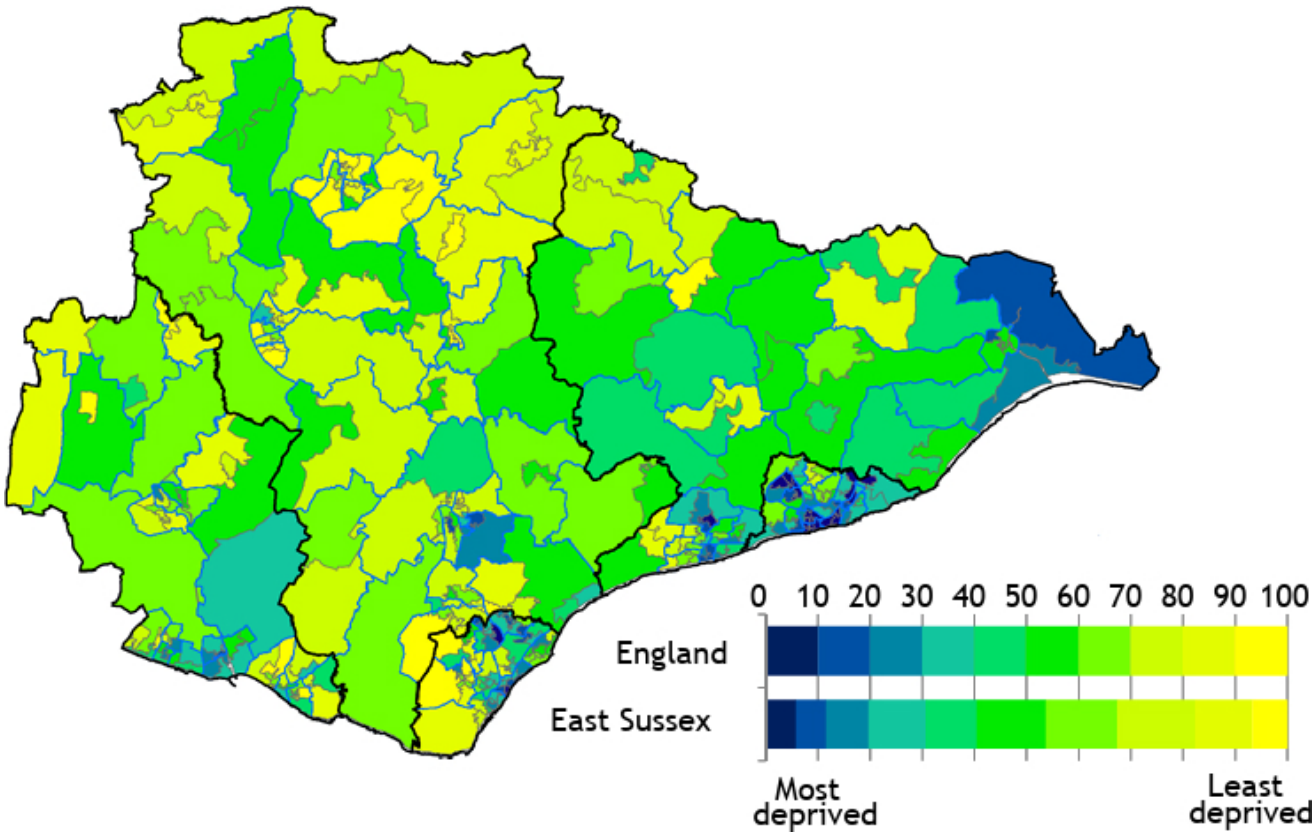
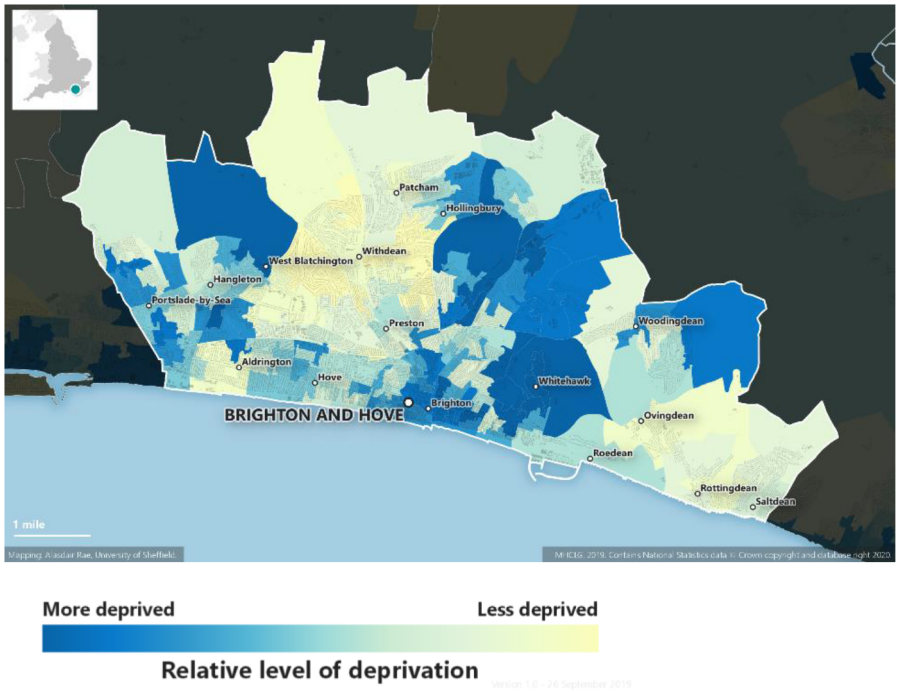


Figure 2: Index of Deprivation in Areas of Brighton and Hove (taken from [7])



allowing patients to access and add to their own health care records.

The SID is a primary and secondary health and social care, integrated, pseudonymised dataset that links routinely collected service user and activity data from health and care providers in Sussex together, to help plan and improve services for the local population.

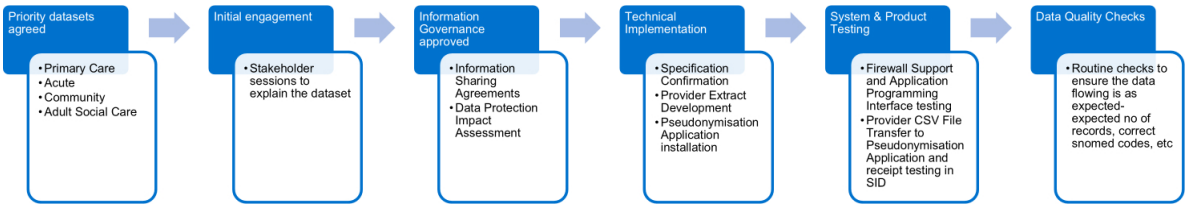
SID Development Process

Following the requirement of local Shared Care Record Programmes to have a secondary use repository, the SID was set up using the process shown in Figure 3. A key decision at initial set up was to enable NHS South Central and West CSU to act as a data provider for all primary care establishments

Table 1: Healthcare Service Provision Across Sussex

NHS services	No of providers
GP Practices	158
Community Pharmacies	315
Dental Contracts	344
Ophthalmic Services	141
Community Clinics	95
Community Hospitals	14
Minor Injuries Units	7
Urgent Treatment Centres	6
111 (out of hours telephone service), Ambulance and Transport	2
Major Hospitals (including Emergency Department)	7
Specialist Hospitals	14

Figure 3: Early Development Process for SID



in Sussex. This meant that the SID only had to deal with the CSU, not 158 separate GP practices, and each practice had a clear automated route through which to provide data to SID.

Legal Basis for Data Processing

The legal basis for processing data for the SID comes from Article 6(1)(e) (Lawfulness of Processing, Performance of a public task) of the General Data Protection Regulation (GDPR) 2018. The data controller derives its authority from the Health and Social Care Act of 2012, Article 9: Processing of special categories of personal data. Specifically this relates to Article 9(2)(h): *Processing is necessary for the purposes of preventive or occupational medicine, for the assessment of the working capacity of the employee, medical diagnosis, the provision of health or social care or treatment or the management of health or social care systems and services on the basis of Union or Member State law or pursuant to contract with a health professional.* For data processing and to prepare anonymised data for researchers, we rely on the GDPR (Articles 6 and 9) provisions for a task carried out in the public interest [13, 14].

Operating Model and Data Security

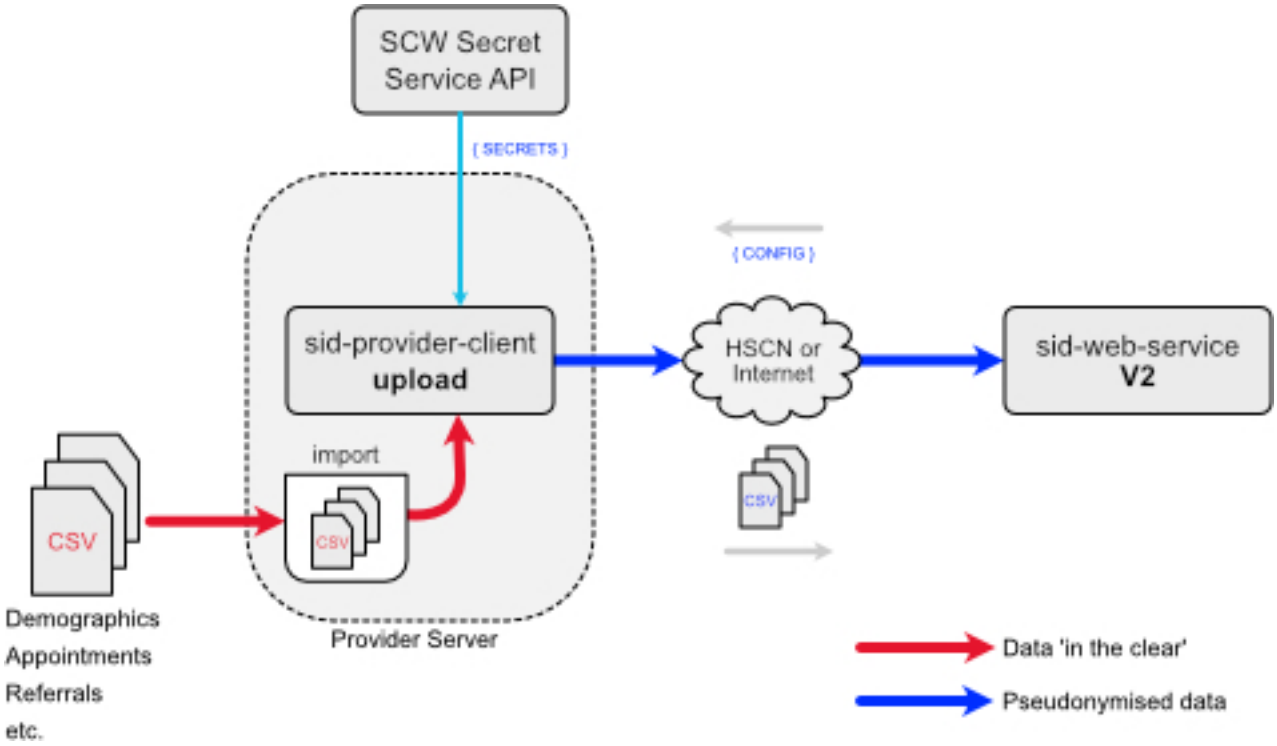
Infrastructure for the OCC programme is delivered by East Sussex Healthcare Trust (ESHT). The initial version of the SID drew in data from providers via a windows-based application called the Pseudonymiser Application (PseudoApp), written in C# version 7, which removed identifiers and replaced them with a unique pseudo-ID. Figure 4 shows how data flows into SID, particularly how the pseudonymisation secrets are fetched from the CSU while the data is still at the NHS provider, ensuring that patient data is pseudonymised before it leaves the provider network.

Having pseudonymised the data at source, the software then calls an Application Programming Interface (API), running on a Microsoft.Net Core version 2.2, that writes the data in JavaScript Object Notation (JSON) format into a Microsoft Structured Query Language (SQL) server data warehouse. The warehouse was initially hosted on premises but has subsequently moved to being part of a Microsoft Azure cloud tenant managed by ESHT.

SID pseudonymises data at source so that no identifiable data leaves a data provider's network. The patient's NHS number and date of birth combination is used to create a pseudo-ID which is given as a unique identifier to link data from multiple datasets. Data is sent using SSL (Secure Sockets Layer), while handshakes between PseudoApp and API, and API and database, are password-protected and timestamp controlled. The consistent generation of a unique identifier from an NHS number and date of birth combination ensures that data for the same patient can be linked from different data sources in a pseudonymised way. In summary, identifiable data is handled in the following ways:

1. Name, address and other identifying fields are not extracted.
2. NHS Number and date of birth are converted into a pseudo-identifier.
3. Date of birth is converted into age.
4. Consultant/clinician seen is converted into unique identifier.
5. Long free text data items are excluded.
6. Small number suppression is applied for all outputs: queries resulting in fewer than five records are obfuscated. For individual analysts this is part of the user

Figure 4: Showing the Pseudonymisation Process: Within NHS Provider Servers, the Pseudonymisation "secrets" are Fetched from SCW CSU before a Pseudonymised ID is Used to Replace the Identifiers



HSCN: Health and Social Care Network (HSCN), a secure network used by the NHS in England to connect healthcare providers.

agreement to access SID, in PowerBI this is controlled within the visualisation settings (PowerBI is a Microsoft business intelligence tool that allows users to visualise and analyse data, turning it into interactive dashboards and reports).

Apart from the pseudonymisation processes that replace identifiable data in the payload, the following measures are taken to ensure data security:

1. Data provider access is controlled through usernames and passwords
2. Data is transferred from provider to SID API using encrypted messaging via SSL
3. A time-dependent authorisation header is included in all subsequent communications following logon. This header is checked when data is sent into the SID; if it cannot be matched the data is rejected.

There have been no data breaches reported so far. In the event of a data breach, the circumstances would be reported to the analytics working group (AWG), Steering, and Capability Boards (see Governance Section).

Data Updates and Lifecycle

Data is uploaded into SID usually on a monthly basis from each data provider. The data is 'pushed' into the platform, the process being initiated by the data provider. Within a monthly cycle the following order is observed:

- Data lands from provider in JSON format
- Data is processed into relational tables upload schema
- Data is standardised and transferred to reporting schema
- Semantic Layer processing creates derived tables
- Results tables for Microsoft PowerBI dashboards are repopulated

Governance and Oversight

Access to and use of the SID is overseen by a number of different boards or committees, the different purposes of which are described in Table 2, along with their frequency of meetings. Figure 5 shows how the groups are formed into a governance structure.

Data Flowing into SID

Data is currently flowing into SID from almost all NHS providers within Sussex, as well as some Adult Social Care data from the Local Authorities. At the time of writing, 100% of general practices in Sussex are now supplying their data to SID; only structured data (clinical codes, data stamps, and administrative data such as admission and discharge dates) are captured. Figure 6 shows the breadth of data within the SID database. Data flow in and are stored in separate files. However, a patient's journey, represented via date-stamped interactions with different NHS providers, can be traced across files using the SID pseudo-ID. Data is not received if a patient has exercised their preference to opt-out via the National Data Opt-Out system [15]. Coverage within SID of the Sussex population is estimated to be near 100%. Up to 100% of

Table 2: Purpose of Each Oversight Group for SID

Governance group	Meeting purpose	Frequency
SID Capability Programme Board	Provide leadership, oversight, and direction. It will ensure that the SID complies with guidance and standards in the delivery of patient and public involvement, has adequate processes in place to comply with the relevant national legislation and local policies that may apply. Ensure that proper Information Governance practice is embedded throughout the SID and delivery of value within boundaries agreed by different stakeholders.	6 weekly
SID Data Ethics Group	To review and enhance the Data Access Request (DAR) pathway for non-Sussex Health and Care (non-SHC) organisations, ensuring ethical and effective use of Sussex Integrated Dataset (SID) data. The group will make recommendations to the SID Capability Board on pathway improvements, potential new datasets, and analytical requirements to maximise SID's impact for the Sussex population. It will also assess and process DAR submissions, oversee data outputs, and escalate governance issues as needed. The Data Ethics Group, comprising partners across the Sussex Health & Care Partnership and allied organisations, will play a key role in supporting decision-making, promoting best practice, and enabling innovative, transformational data use aligned with strategic priorities.	Ad-hoc
Analytics Working Group (AWG)	Oversee and develop the capability of the SID. A partnership of providers with the Sussex Health & Care Partnership and allied partners, they play a supporting and steering role to maximise the analytical capability and reach of the SID, identify best practice and quality, improve access and understanding data analysis and identify opportunities to develop the infrastructure in line with the strategic priorities across Sussex. Evaluate, discuss and approve/reject applications to use the SID for quality evaluation and improvement projects and trusted analysts.	Monthly
SID Steering Group	Ensure that the project is aligned with organisational strategy and makes good use of assets and resources available.	Fortnightly
Data Provider Working Group (DPWG)	This group is a partnership of data providers that ensure the relationship between providers of data and SID is enabling and supportive and can identify opportunities for the SID to streamline data supply processes and provide practical benefit to teams managing data within providers.	2-monthly

GPs flow data and hospital data should cover every user of hospital care in Sussex. In early work, 3 million unique user IDs were found in SID (compared to only 1.7m patients registered with GPs) and considerable work was put into identifying and marking living, current Sussex residents, this work is reported elsewhere [16]

Quality Standards

Data quality is judged and addressed within SID based on the five pillars of data quality: accuracy, completeness, timeliness, consistency and uniqueness. For basic quality control, each item is checked against the information schema data dictionary to ensure that the item is known and represented as a column within the database. Basic validation is applied to distinguish numbers, dates and short text descriptions (e.g. test results). Enhanced validation is used to add derived values e.g. converting ethnicity codes to the NHSE national standard. For these derived variables, row counts are taken at the start of processing and compared at the end, and if there is a difference, exception processing is invoked.

Data quality enhancement is supported by regular reviews with governance groups to look at SID data quality issues, and use of a dedicated data quality dashboard, with example outputs shown in Figure 7, in which data are reviewed by ethnicity, age, sex and Lower Super Output Area (LSOA) and can be compared against other open source datasets about the local population to ensure they match.

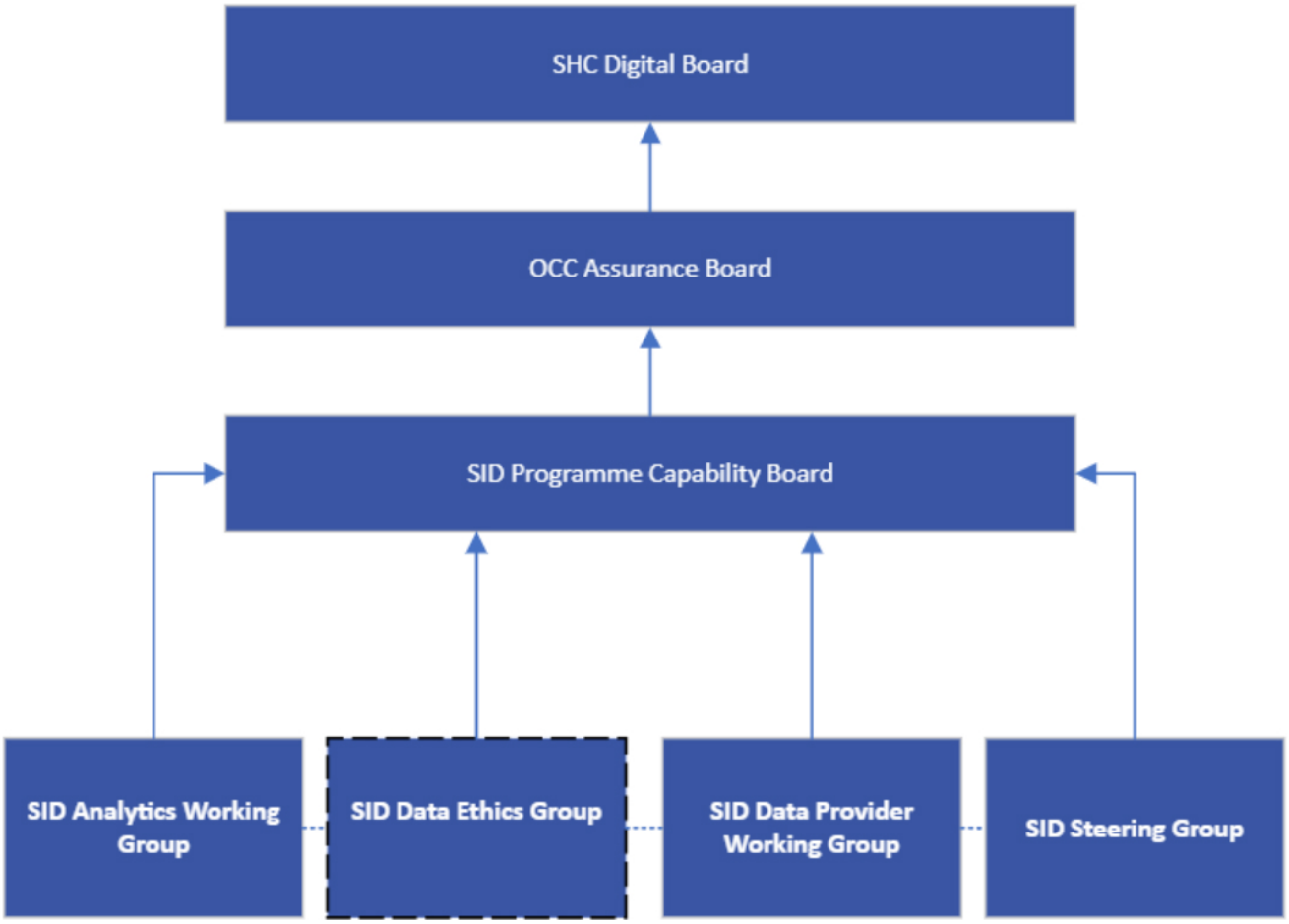
In addition to these quality checks, SID provides a semantic layer that is designed to:

- Make the views of data more consistent
- Apply transformations on data where appropriate to aid reporting
- Remove duplicate records
- Derive calculated data items using published and transparent algorithms

The main elements of the semantic layer are:

- Demographic Milestones
- Condition Milestones
- Enablement of clinical risk stratification (in the first case for falls prevention).

Figure 5: Hierarchy of Oversight and Governance Boards for SID



Demographic milestones are created through this semantic layer processing. They are designed to give a reflection of patient demographics at a particular point in time, i.e. the milestone. Information such as registered GP practice, LSOA of patient residence, and age, is abstracted from different sources of data, and where there are conflicting values, the most plausible value is chosen based on a set of rules. A fuller description of the conflicting values found at the inception of SID is given elsewhere [16]. Resolving scripts have been written which decide the most plausible value by taking the answer from the latest record supplied, and prioritising in the order of primary care, social care, community, mental health and acute (hospital) trusts, taking the first the first non-empty answer. This is calculated for year of birth, year of death, sex, ethnicity, age (at a given date), and current GP practice (at a given date). As an example process, for ethnicity the following rules are applied:

- Ethnicity is calculated for each supplier of data separately.
- The ethnicity field can contain NHS England codes or free text values, e.g. 'White British'.
- The calculator removes codes 'Z' and '99' (not supplied) applies a set of regular expressions to convert the ethnicity to an NHS England code if necessary and returns the first valid value from the latest record.
- The resolver sorts the answers in order of primary care -> acute trusts (explained above) and takes the first value.

For the user, the milestones table and semantic layer allow selection of patients based on attributes such as patients registered at a particular GP practice, patients diagnosed with diabetes, or patients within an age category. These summary totals can be compared to other data sources, e.g. Quality and Outcomes Framework standardised reporting, to validate the data in SID.

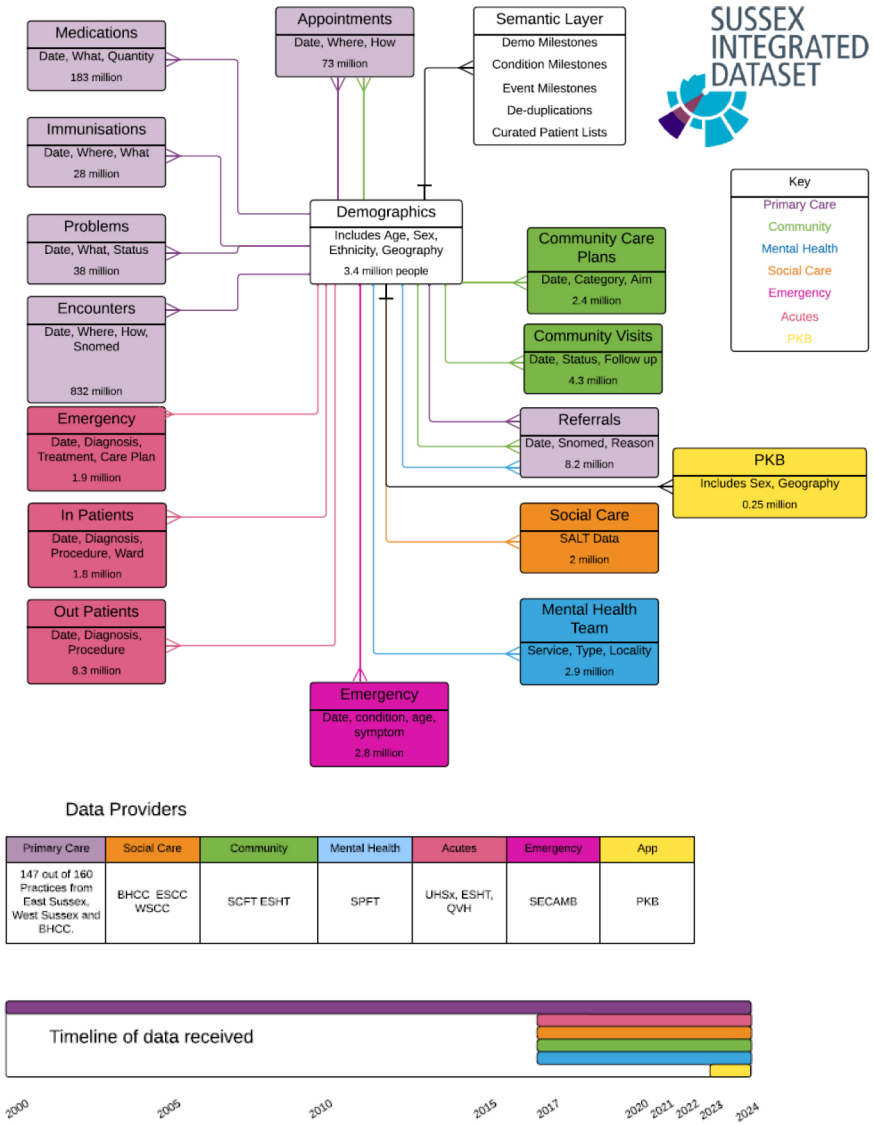
The semantic layer can used in conjunction with other data sources to validate the completeness and consistency of the data in SID. The semantic layer is the first area that users of SID will interrogate to get a basic understanding of the data. The semantic layer has a number of ways to connect back to the underlying provider-supplied data which means that a user of SID can always check the working and apply the derivations of calculated data on the raw data supplied (Figure 8).

Access Arrangements

Access to SID data is provided in four ways:

1. Via the Kent, Medway and Sussex NHS Sub-National Secure Data Environment (KMS SDE) [17] for any research uses. The SDE publishes its metadata on MetaDataWorks [18] and will evaluate applications

Figure 6: Data Sources Flowing into SID



Abbreviations: Snomed: Systematized Nomenclature of Medical Clinical Terms – a coding system used in UK general practice. PKB: Patient Knows Best – patient facing electronic health record. SALT: Short and Long Term Care – a data extraction from social care. BHCC: Brighton and Hove City Council. ESCC: East Sussex County Council. WSCC: West Sussex County Council. SCFT: Sussex Community Foundation NHS Trust. SPFT: Sussex Partnership Foundation NHS Trust. QVH: Queen Victoria Hospital. UHSx: University Hospital Sussex NHS Trust. SECAMB: Southeast Coast Ambulance Service.

from any legitimate research organisation. (For more information about access via this route, see [17]).

2. Through single-project Data Access Requests (DARs) for population health management or service evaluations from within Sussex NHS data-contributing organisations and the Sussex Local Authorities (see Supplementary Figure 1).

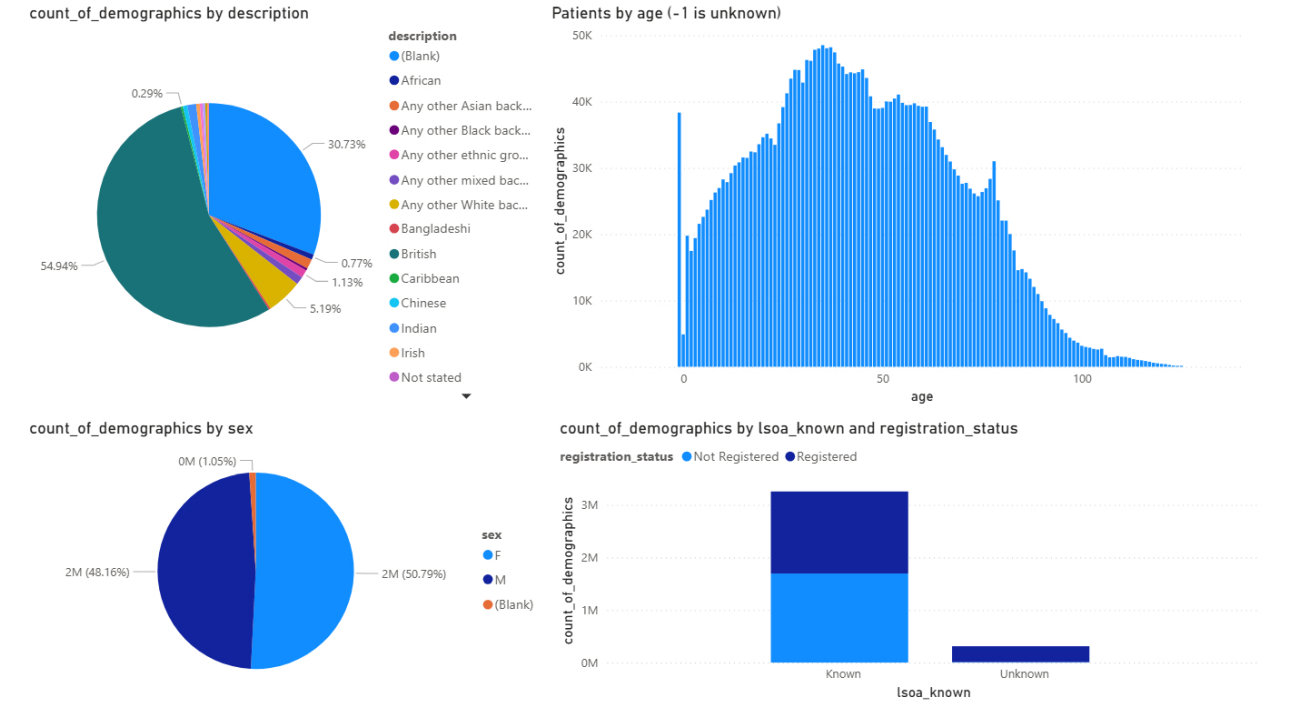
3. Through becoming a Trusted Analyst (TA) within Sussex NHS data-contributing organisations and the Sussex Local Authorities. This trusted status allows continuous access to SID for a range of public benefit related projects (See Supplementary Figure 2).

4. Via various PowerBI dashboards, mainly constructed by the SID development team, examples for which are:

    a. Core20+5 subject areas [19]:
        i. SMI-PHC
        ii. Hypertension
        iii. COPD
        iv. Breast Cancer Screening
        v. Bowel Cancer Screening
        vi. Cervical Cancer Screening
    b. Diabetes
    c. Frailty
    d. Data Quality

Project applications for SID must meet the SID guardrails. These are a set of rules ensuring that data is used safely and securely for the intended purpose. The guardrails from the outset of SID in 2020 have been as follows:

Figure 7: Example Output Dashboard of Data Quality Analysis



- An appropriate and ethical use of the data, that it will deliver clear public benefits.
- The analysis supports health and care planning or service planning or service development or evaluation or commissioning.
- Data will not be used for performance management.
- The 'Sponsor' of the piece of work (Senior Director/ Manager within NHS Sussex/LA) will be clearly identified.
- Data use will be restricted to that specified in the original request.
- Anyone with access to the data will not copy it (or download it) elsewhere.
- There will be no attempt to re-identify patients from data.
- There will be clear plans for addressing potential identification with low numbers when processing the data.
- Data created for dashboards is only visible to relevant health and social care staff.
- Data for non-dashboard use is aggregated, anonymous.
- Applicants must have skills, or access to skills, to analyse the data.
- All staff with access to the data will have up to date information governance training.

Updated guardrails for the use of SID for research and development outside of the Sussex NHS and LA systems, via the Secure Data Environment, are currently being developed by, and agreed with, the SDE patient and public advisory group.

When research is published using SID data, we recommend reporting according to the "Reporting of Studies Conducted using Observational Routinely Collected Health Data (RECORD)" guidelines and checklist [20].

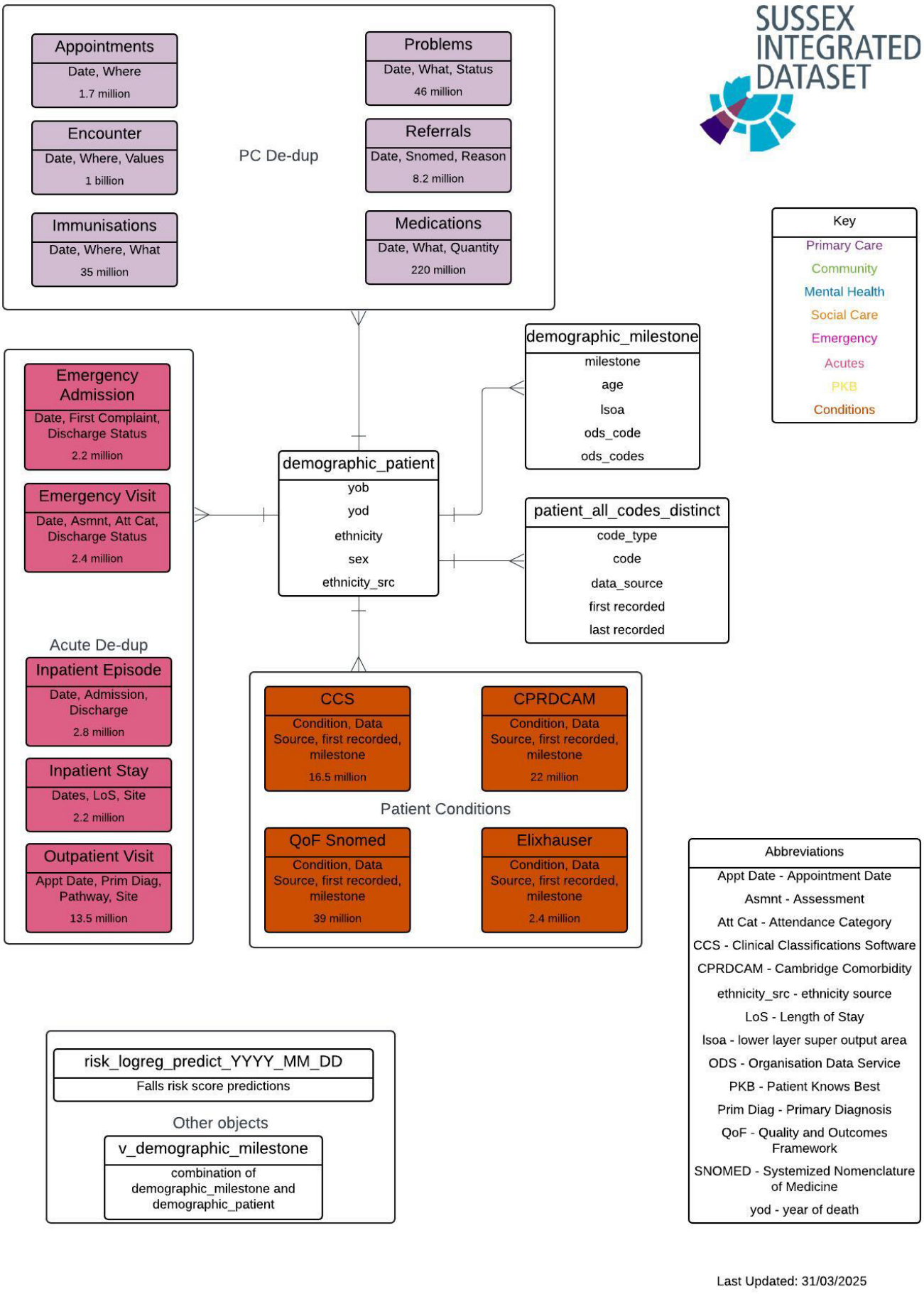
Public and Patient Engagement and Involvement

There have been a number of initiatives to engage with public citizens in Sussex including a set of deliberative focus groups in 2021 [21] and ICB commissioned deliberative engagement which was carried out in 2022. Contributors to these events largely supported use of patient data to improve health service efficiency and resource management, preventative services and out of hospital care, joined-up services and information flows. They suggested that transparency of data usage, ongoing, inclusive and valued involvement of the public in dataset decision-making, and a commitment to building trust, would meet their expectations for responsible data use [21]. These public views underpinned the values and principles of the KMS SDE at set up phase, and engagement is ongoing with the NHS Sussex Digital and Data People's Panel who meet regularly to steer digital transformation in the local NHS.

Case Study Use of SID data – Falls Risk Prediction

In conjunction with Grip Analytics and PA Consulting, the SID has been used to produce a falls risk model (FRM) for injurious falls for patients registered in Sussex who are 65 and over

Figure 8: Schematic of the SID Semantic Layer



years of age. According to national data, Sussex is an outlier for injurious falls [22] and, with an ageing population, there is an imperative to reduce these occurrences. While the UK National Institute for Health and Care Excellence recommends against using falls risk prediction tools for individuals, Sussex aims to use the tool at the Primary Care Network (PCN) or population level to guide and direct resources towards GP registered patients who might benefit from an intervention.

Model Development

Sample: All primary care registered adults aged 65 and over, who had data for complete years in 2022 and 2023, were included; patients who died, arrived in or transitioned out of Sussex during this time were removed from the dataset. The final sample was 280,777 people.

Variables and definitions: Inspired by research developed by Smith et al., [23], and drawing on Smith's study's definitions for each of these markers, 37 clinical concepts were extracted from SID as predictors of falls, as shown in Table 3.

Each of these was modelled as a binary variable on whether there was evidence or no evidence of the clinical concept in the patient's linked NHS records (both primary and hospital care) using lists of SNOMED and ICD codes. Acute conditions and events were ascertained only during the calendar year Jan-Dec 2022. Chronic conditions (osteoporosis, COPD, dementia, asthma, depression, Parkinson's etc) were ascertained throughout the entire patient record history (using QOF business rules code lists [24]).

Model Training

A stepwise logistic regression process was used, with the variables above as predictors, and the binary outcome of a documented fall and/or fracture (drawn from both primary and secondary care) or no evidence of a fall/fracture during the calendar year Jan-Dec 2023. Stepwise variable optimisation was conducted with Akaike Information Criterion (AIC). AIC is a statistical measure that balances the goodness of fit of a model with its complexity, favouring models that fit the data well while avoiding overfitting by penalising the inclusion of unnecessary variables.

To test model accuracy, the data was split between training and test data (80:20) stratified by age group, sex and outcome prevalence. This approach aimed to balance model fit and avoid overfitting. No collinearity was detected. Performance was assessed using Area Under the ROC Curve, a Hosmer-Lemeshow plot for calibration, and a Confusion Matrix.

Results

Of the 280,777 eligible patients entered into the model, 17,949 (6.4%) patients had a fall or fracture captured in primary or secondary care records during the follow up period (2023). Patient characteristics are show in Table 4. Within the group who had a fall in the follow up period, they were more likely to be older, female and live in a higher deprivation area than the cohort as a whole (significance was not tested). Within the falls group, there were high levels of hypertension (54%), diabetes (51%), anxiety or depression (45%), hearing loss (29%), atrial fibrillation (21%), coronary heart disease (18%) and history of stroke or TIA (15%). Fall rates appeared higher in patients with high polypharmacy, and more GP, ED and outpatient visits in the last year.

Model Fit Parameters

The model performed well with an AUC of 0.794, and good overall model fit, evidenced by the Hosmer-Lemeshow plot (Figure 9). Table 5 shows the parameters retained by the regression model and their relative weights in the model. Older

Table 3: Variables Tested in the Falls Risk Prediction Model

Demographic Variables	**Long Term Conditions**	**Previous Falls and Fractures**
Age	Anxiety Depression	Osteoporotic Fracture
Sex	Asthma	Hospital recorded fall
Ethnicity	Atrial fibrillation	Arm/wrist fracture
Employment Decile	Cancer	Spine fracture
	Chronic sinusitis	Fragility fracture
	Connective Tissue Disorder	GP recorded Fall
Healthcare usage	COPD	Pelvis fracture
Polypharmacy	Coronary Heart Disease	Hip fracture
N GP appointments	Dementia	Femur fracture
N Outpatient appointments	Diabetes	
Outpatient appt last month	Epilepsy	
N Emergency Dept visit	Hearing Loss	
	Heart Failure	
	Hypertension	
	Irritable bowel syndrome	
	Parkinsons Disease	
	Psychosis/Bipolar Disorder	
	Stroke and TIA	
	Alcohol problem	

Table 4: Patient Characteristics for Falls Risk

Characteristic	Overall N = 280,777	0 – No fall N = 262,828	1 - Fall N = 17,949
Age		74 (69–80)	82 (76–88)
Sex			
Female	153,138 (55%)	141,753 (54%)	11,385 (63%)
Male	127,639 (45%)	121,075 (46%)	6,564 (37%)
Ethnicity			
A. White British	244,669 (87%)	228,228 (87%)	16,441 (92%)
B. White Irish	2,317 (0.8%)	2,174 (0.8%)	143 (0.8%)
C. White Any Other	10,507 (3.7%)	9,818 (3.7%)	689 (3.8%)
D. Mixed White and Black Caribbean	163 (<0.1%)	153 (<0.1%)	10 (<0.1%)
E. Mixed White and Black African	220 (<0.1%)	211 (<0.1%)	9 (<0.1%)
F. Mixed White and Asian	311 (0.1%)	295 (0.1%)	16 (<0.1%)
G. Mixed Any Other	1,057 (0.4%)	980 (0.4%)	77 (0.4%)
H. Asian or Asian British Indian	1,550 (0.6%)	1,487 (0.6%)	63 (0.4%)
J. Asian or Asian British Pakistani	426 (0.2%)	410 (0.2%)	16 (<0.1%)
K. Asian or Asian British Bangladeshi	166 (<0.1%)	157 (<0.1%)	9 (<0.1%)
L. Asian or Asian British Any Other	1,351 (0.5%)	1,310 (0.5%)	41 (0.2%)
M. Black or Black British Caribbean	526 (0.2%)	487 (0.2%)	39 (0.2%)
N. Black or Black British African	457 (0.2%)	440 (0.2%)	17 (<0.1%)
P. Black or Black British Any Other	261 (<0.1%)	249 (<0.1%)	12 (<0.1%)
R. Chinese	618 (0.2%)	606 (0.2%)	12 (<0.1%)
S. Any Other	1,205 (0.4%)	1,154 (0.4%)	51 (0.3%)
Z. Not stated	14,973 (5.3%)	14,669 (5.6%)	304 (1.7%)
IMD Decile			
1 high deprivation	6,604 (2.4%)	5,942 (2.3%)	662 (3.7%)
2	10,362 (3.7%)	9,481 (3.7%)	881 (4.9%)
3	16,764 (6.0%)	15,410 (5.9%)	1,354 (7.6%)
4	20,345 (7.3%)	18,838 (7.3%)	1,507 (8.4%)
5	30,067 (11%)	28,020 (11%)	2,047 (11%)
6	44,978 (16%)	42,279 (16%)	2,699 (15%)
7	23,824 (8.6%)	22,325 (8.6%)	1,499 (8.4%)
8	40,069 (14%)	37,517 (14%)	2,552 (14%)
9	35,775 (13%)	33,642 (13%)	2,133 (12%)
10 low deprivation	48,484 (17%)	45,967 (18%)	2,517 (14%)
Alcohol problems	7,193 (2.6%)	6,561 (2.5%)	632 (3.5%)
Anxiety or depression	95,032 (34%)	86,904 (33%)	8,128 (45%)
Asthma	37,365 (13%)	34,192 (13%)	3,173 (18%)
Atrial fibrillation	28,660 (10%)	24,969 (9.5%)	3,691 (21%)
Cancer	32,179 (11%)	29,614 (11%)	2,565 (14%)
Chronic sinusitis	8,949 (3.2%)	8,368 (3.2%)	581 (3.2%)
Connective tissue disorder	27,648 (9.8%)	24,968 (9.5%)	2,680 (15%)
COPD	17,083 (6.1%)	15,307 (5.8%)	1,776 (9.9%)
Coronary heart disease	31,840 (11%)	28,688 (11%)	3,152 (18%)
Dementia	10,447 (3.7%)	7,341 (2.8%)	3,106 (17%)
Diabetes	122,685 (44%)	113,600 (43%)	9,085 (51%)
Epilepsy	7,931 (2.8%)	7,213 (2.7%)	718 (4.0%)
Hearing loss	54,241 (19%)	49,005 (19%)	5,236 (29%)
Heart failure	9,990 (3.6%)	8,467 (3.2%)	1,523 (8.5%)
Hypertension	119,962 (43%)	110,295 (42%)	9,667 (54%)
Irritable bowel syndrome	12,268 (4.4%)	11,394 (4.3%)	874 (4.9%)
Parkinsons disease	2,298 (0.8%)	1,819 (0.7%)	479 (2.7%)
Psychosis or bipolar disorder	29,404 (10%)	26,074 (9.9%)	3,330 (19%)
Stroke and TIA	19,500 (6.9%)	16,734 (6.4%)	2,766 (15%)

Continued

Table 4: Continued

Characteristic	Overall N = 280,777	0 – No fall N = 262,828	1 - Fall N = 17,949
Polypharmacy			
0 prescriptions	67,353 (24%)	61,385 (23%)	5,968 (33%)
1-4	61,417 (22%)	59,992 (23%)	1,425 (7.9%)
5-9	75,228 (27%)	72,018 (27%)	3,210 (18%)
10+	76,779 (27%)	69,433 (26%)	7,346 (41%)
Osteoporotic fracture	176 (<0.1%)	119 (<0.1%)	57 (0.3%)
Falls (hospital coded)	5,076 (1.8%)	3,400 (1.3%)	1,676 (9.3%)
Fracture arm or wrist	1,230 (0.4%)	976 (0.4%)	254 (1.4%)
Fracture spine	1,126 (0.4%)	791 (0.3%)	335 (1.9%)
Fracture fragility	1,538 (0.5%)	1,234 (0.5%)	304 (1.7%)
Falls (GP coded)	8,208 (2.9%)	5,114 (1.9%)	3,094 (17%)
Fracture pelvis	95 (<0.1%)	65 (<0.1%)	30 (0.2%)
Fracture hip	1,154 (0.4%)	843 (0.3%)	311 (1.7%)
Fracture femur	185 (<0.1%)	136 (<0.1%)	49 (0.3%)
Major ED visits			
0	238,351 (85%)	226,310 (86%)	12,041 (67%)
1	26,884 (9.6%)	23,743 (9.0%)	3,141 (17%)
2	9,699 (3.5%)	8,276 (3.1%)	1,423 (7.9%)
3	5,843 (2.1%)	4,499 (1.7%)	1,344 (7.5%)
GP appointments			
0	28,035 (10.0%)	27,374 (10%)	661 (3.7%)
1-5	33,215 (12%)	32,115 (12%)	1,100 (6.1%)
6-10	35,678 (13%)	34,214 (13%)	1,464 (8.2%)
11+	183,849 (65%)	169,125 (64%)	14,724 (82%)
Outpatient appointments			
0	159,644 (57%)	152,145 (58%)	7,499 (42%)
1-5	80,168 (29%)	73,919 (28%)	6,249 (35%)
6-10	24,130 (8.6%)	21,785 (8.3%)	2,345 (13%)
11+	16,835 (6.0%)	14,979 (5.7%)	1,856 (10%)
Outpatient in last month	29,654 (11%)	26,818 (10%)	2,836 (16%)

age and female sex were both retained following stepwise elimination, as were many of the previous falls/fracture variables and number of GP, outpatient appointments and emergency department visits. Of the long-term conditions, dementia, stroke and Parkinson's disease were strongly weighted, as were alcohol problems.

Table 6 shows the Confusion Matrix summarising the performance of the model showing relationship between predicted and actual outcomes. With the threshold set at the top 10% of predicted risk (corresponding to a $\geq$13.1% probability of a fall), 5,615 individuals were classified as high risk. Of these, 1,468 experienced a fall within the subsequent 12 months, yielding a positive predictive value (PPV) of 26.1%. At this threshold, the model identified 40.9% of all individuals who experienced a fall during the observation period (sensitivity of 40.9%).

Implementation in Sussex and Reflections

The model has been applied to an identifiable patient dataset from the Bexhill Primary Care Network (PCN), which covers a coastal urban/suburban region of East Sussex, with 3 GP practices, and has a population significantly older than the English average (34.5% aged 65 and over compared to 18.4% in England). The PCN used the tool to identify patients for their "Multi-Disciplinary Team (MDT) for Highest and Ongoing Needs (HONs)" programme. The team consists of two senior clinicians – a Consultant Geriatrician and a Senior GP, supported by a Pharmacist, Dietician and an Allied Health Professional (AHP) from the Joint Community Rehabilitation (JCR) Team. Of 151 patients identified by the falls risk tool as being very high risk, an initial case note review was undertaken by senior clinicians to identify those who would benefit from any intervention. Patients were contacted and consent sought to attend a clinic or a home visit for assessment. The visit would include: medication review (optimisation including de-prescribing); falls assessment; dietary assessment (nutrition and hydration); advice regarding strength training, possible gait retraining and rehab; provision of appropriate equipment to help with mobility; referrals to local gyms; social circumstances review and appropriate advice; and then aim to capture any falls/near miss falls by remote data collection. At time of writing the trial is continuing, with plans to roll out to the rest of East Sussex. The development of Integrated Community Teams (ICTs) and their involvement in this programme has been crucial.

Figure 9: (a) Hosmer-Lemeshow Plot and (b) ROC Curve Showing Model Discrimination

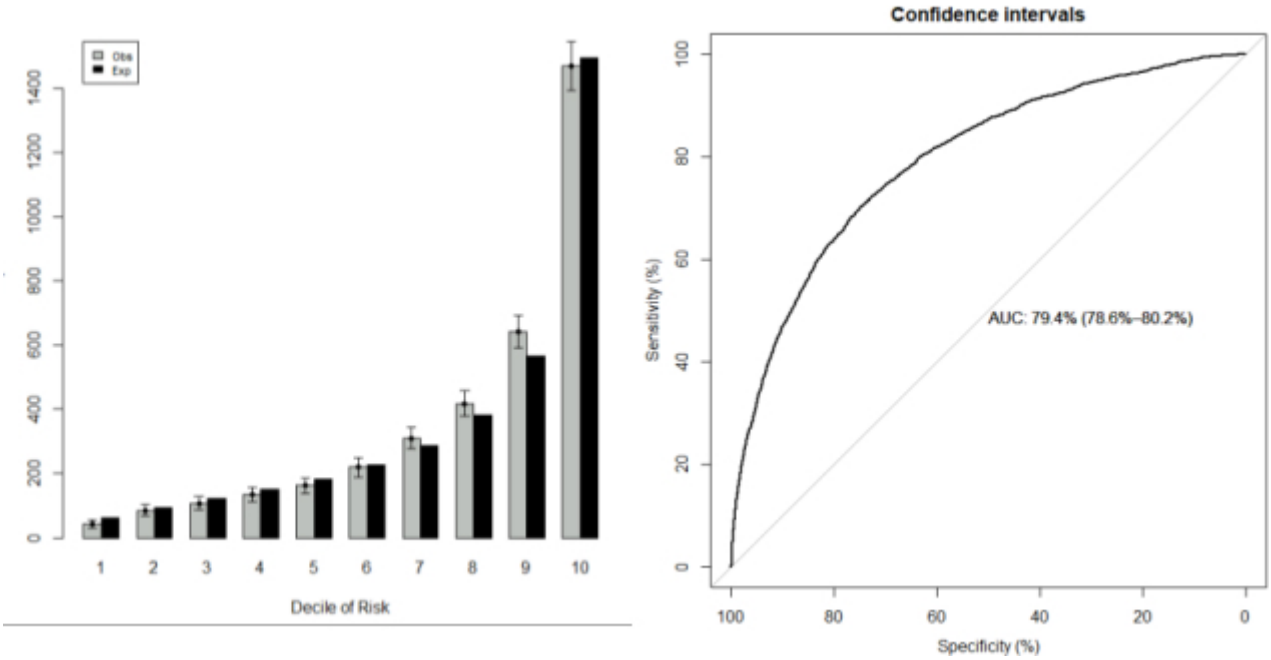


Table 5: Retained Variables and Logistic Regression Weights

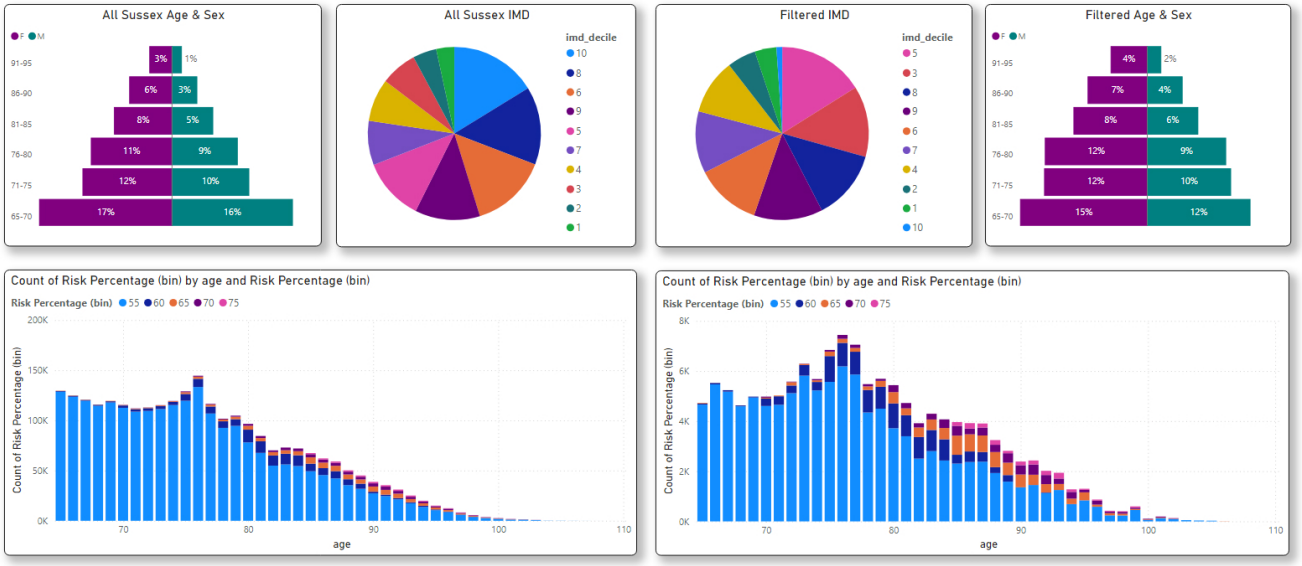
Parameter	Logistic Regression Estimate	Standard Error	P value
Sex (Male)	-0.248	0.0197	<0.00001
Alcohol problems	0.456	0.0515	<0.00001
Anxiety or depression	0.217	0.0208	<0.00001
Asthma	0.0166	0.0256	0.51600
Atrial fibrillation	0.197	0.0258	<0.00001
Connective tissue disorder	0.112	0.0269	0.00003
COPD	0.252	0.0328	<0.00001
Dementia	0.933	0.0306	<0.00001
Diabetes	0.0991	0.0187	<0.00001
Epilepsy	0.136	0.0492	0.00567
Heart failure	0.188	0.0377	<0.00001
Hypertension	0.0613	0.019	0.00128
Parkinsons disease	0.915	0.0652	<0.00001
Psychosis or bipolar disorder	0.212	0.0283	<0.00001
Stroke and TIA	0.282	0.0278	<0.00001
Falls (hospital coded)	0.394	0.0439	<0.00001
Fracture (spine)	0.466	0.0886	<0.00001
fragility fracture GP	0.811	0.0822	<0.00001
Falls (GP coded)	1.29	0.0314	<0.00001
Fracture (pelvis)	0.372	0.302	0.21800
Major ED visits	0.184	0.0127	<0.00001
GP appts 1-5	0.223	0.0574	0.00010
GP appts 6-10	0.294	0.0551	<0.00001
GP appts 11	0.557	0.0476	<0.00001
Outpatient appts 1-5	0.193	0.022	<0.00001
Outpatient appts 6-10	0.292	0.0309	<0.00001
Outpatient appts 11	0.335	0.0348	<0.00001
Age	0.0808	0.0201	0.00006

The model has also been applied to pseudonymised data from the SID for all patients aged over 65, with a dashboard available to planners which shows the demographic split by age, ethnicity and falls risk for all Sussex (left hand side graphics) compared to Bexhill PCN (right hand side graphics) (Figure 10). The aim is to make it easy for health service planners to identify where specific groups of patients are disproportionately more at risk.

Table 6: Confusion Matrix for 10% Patients Identified as Highest Risk by Model

	Fell	No	Yes	Total	PPV
High Risk	No	48,413	2,117	50,530	
	Yes	4,147	1,468	5,615	**26.1%**
	Total	52,560	3,585	56,145	
Sensitivity			**40.9%**		

Figure 10: Example Dashboard Showing Patients at Risk of Falls



Discussion

Lessons Learned So Far

Since the inception of the SID in 2019, progress has been made in improving data quality. In the early uses of the SID [16], we found many conflicting demographic values within each patient, which led to the development of the demographic milestones layer, to automate decision-making around which was the likely truest value for each demographic variable. Creating an automated demographics table, which was populated with weighted values for each calendar year, helped both with identifying reliable and stable demographics and also helped with identifying patients who died or who did not live in the county. Strong teamwork between early analysts of the SID and the core SID team enabled these changes to be made early on [16]. Further learning showed that we could not rely on open-source code lists, developed on other NHS data sources, to accurately identify patients with particular conditions in the SID. Instead, significant checking and adaptation of code lists has gone into providing the semantic layer, with full and transparent documentation available for SID users. In general SNOMED codes are favoured by primary care providers in Sussex over Read codes, and while care has been taken to align code lists with published lists, we have prioritised capturing our patient cohorts to the best standard, compared to other well-established open-source patient lists such as QOF.

The falls risk tool is the first SID project which has led to implementable outcomes, and the implementation of the falls prevention intervention is underway. This model was intentionally developed as a baseline approach, using a core set of routinely available features to support initial implementation and evaluation. As such, there is clear scope for further refinement and improvement in predictive performance. Future iterations could incorporate additional clinically relevant variables, such as prior emergency admissions, drug types within polypharmacy, and indicators of vestibular or balance-related conditions, which are likely to enhance risk stratification and targeting. We would therefore expect subsequent versions of the model to demonstrate improved discrimination and calibration, supporting more precise identification of individuals at highest risk.

In parallel, the SID is also being used to support the identification and targeting of individuals for complex case management. The Johns Hopkins ACG® System, a widely used population health management framework that groups individuals based on diagnoses, morbidity burden and expected healthcare utilisation, has been applied across the SID, with all individuals classified and risk-scored using its suite of algorithms. Within this framework, Patient Need Groups (PNGs) are used to identify those with the greatest complexity. For the complex case management programme, the target cohort comprises individuals classified within PNG 10 (high complexity multimorbidity) and PNG 11 (frailty), alongside those in the top decile of predicted risk of emergency hospital admission. This combined approach enables the identification of patients with both high clinical complexity and elevated risk of acute care utilisation, facilitating more proactive and coordinated care delivery at scale.

The SID has also been used to examine several other pressing issues for the Sussex healthcare system, such as

examining the proportion of mental health beds where the occupant came in through an Emergency Department (ED) admission, the ED usage of people diagnosed with personality disorders, and the number of physical and cardiovascular health checks being offered routinely to people with severe mental illness or learning difficulties; none of these have so far resulted in an intervention, but momentum in the use of SID for driving system change is growing.

Strengths and Limitations

The SID has been operational now for 6 years and we have been able to identify some strengths and weaknesses with the overall approach. Particular strengths are that the SID decision to pseudonymise the data at source has meant that several information governance hurdles, which have affected other similar initiatives, have not affected SID because identifiable data has never been hosted. The lack of patient identifiers leaving clinical environments has also meant a greater uptake from data providers, with a particular strength of SID being its high primary care provider contribution. The pseudonymisation application has been easy for data providers to install and operate, and also means that linkage across sources is simple because every data source contains the same unique identifier for each patient.

Since its development, the semantic layer has enabled reporters to gather sometimes complex data consistently, and to make this quickly available to service planners in the form of dashboards and data visualisations, rather than needing to wait for the publication of complex analysis reports.

However, we have experienced some weaknesses in the chosen approach. The generation of the Pseudo-ID is based on combination of NHS Number and date of birth. The inclusion of date of birth, whilst giving a guarantee of uniqueness, has prevented an automated process for re-identification, which would have been possible if only one unique identifier had been used for the encryption. We are now in the process of remodelling this so that re-identification under very controlled and limited circumstances, can be supported, following appropriate governance and ethical approvals.

Future Work

The ambition of the SID is to provide a holistic view of health-related data for all people in Sussex. As social and environmental factors play key roles in a person's health and well-being, the team is striving to link, include or make accessible any data sets that have a bearing on health such as housing, lived environment, air quality and socioeconomic data. However, because of the lack of identifiable data and the pseudo-ID being based, in part, on an NHS number, we cannot currently link in data at patient level from wider determinants such as non-healthcare or non-social care sources, because without an NHS number in the source data, the pseudo-ID cannot be generated. We are currently working on linking in wider determinants data at the small area level instead [25]. This means looking at open sources of wider determinants of health with data on small geographic areas (such as LSOA), and mapping these onto patient level data in the SID by creating new variables mapped against the patients' LSOA marker. Within the SDE environment, the SID has been part of an application to host patient identifiers for means of new linkages and therefore the capability to match in other wider determinants or administrative data may develop in the future.

Data from the population of Sussex offers a genuinely interesting set of demographics for those interested in wider UK population epidemiology. Sussex has an older age profile compared to many areas, and with a UK-wide ageing population, this may prefigure the UK population in 5-10 years, allowing testing of hypotheses about evolving complex care needs. This could also be considered a limitation: a Sussex cohort will over-represent older adults, with implications for multimorbidity, frailty, dementia prevalence, and patterns of healthcare use. An additional difference compared to the UK population as a whole is the under-representation of minority ethnic groups, although this may match other predominantly rural areas of England. Sussex does represent wide sociodemographic disparities with serious deprivation in some areas comparable to inner city UK areas. Notably Brighton and Hove has a much younger and more diverse population which can balance the white and older demographic of the more rural areas.

Conclusions

The Sussex Integrated Dataset has developed quickly in the span of 6 years to become a nationally leading linked dataset with very high primary care coverage, that can contribute to health research and epidemiology across the UK. Data quality is carefully monitored and curated, and we have strong proof-of-concept projects which show the value of SID for public health, risk prediction and service evaluation and planning. The SID is now available for research via the NHS secure data environment network, and can be used for a range of projects, with the caveat that these must show potential benefits for the population of Sussex. We continue to develop the SID to address regional health inequalities and to address population health issues, to improve outcomes for the whole Sussex population.

Acknowledgements and Funding Statement

EF is funded by the National Institute of Health and Care Research (NIHR) Applied Research Collaboration in Kent, Surrey and Sussex (NIHR200179). The funders had no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Ethical Statement

Ethical approval was not needed for this work as it 1) reports on infrastructure development rather than research and 2) the exemplar analysis reported was carried out for service improvement and de-identified patient records were used.

Conflict of Interest Statement

None declared.

Data Availability Statement

Please see above in the section "Access Arrangements" for how to access the dataset described here. SID is available for approved researchers and projects through the NHS Secure Data Environment Network and the National Health Data Research Service.

AI Disclosure Statement

The authors declare that no generative AI tools were used in the preparation of this manuscript.

References

1. Office for National Statistics. Census 2021 [Available from: https://www.ons.gov.uk/census]
2. Sussex Community Foundation. Sussex Uncovered 3 2019 [Available from: https://sussexcommunityfoundation.org/wp-content/uploads/2024/01/Sussex-Uncovered-3_2019.pdf]
3. West Sussex County Council. Census 2021: Population estimates 2022 [Available from: https://jsna.westsussex.gov.uk/assets/pdf/census-briefing/WSX-census-21-population-estimates-briefing.pdf]
4. West Sussex County Council. West Sussex Joint Strategic Needs Assessment Summary Document 2024/5 2024 [Available from: https://jsna.westsussex.gov.uk/assets/pdf/WS_JSNA_Summary_2024_2025.pdf]
5. East Sussex County Council. State of the County 2023 Focus on East Sussex 2023 [Available from: https://democracy.eastsussex.gov.uk/documents/s51141/Appendix%201%20Focus%20on%20East%20Sussex%202023.pdf]
6. Brighton and Hove City Council. Brighton and Hove demographics 2023 [Available from: https://www.brighton-hove.gov.uk/brighton-hove-city-council-plan-2023-2027/brighton-hove-demographics]
7. mySociety Research. Index of Multiple Deprivation 2019 Brighton and Hove 2019 [Available from: https://research.mysociety.org/sites/imd2019/area/la-brighton-and-hove-city-council/lsoa/]
8. Brighton and Hove City Council. Joint Strategic Needs Assessment Brighton, UK: Brighton and Hove City Council; 2022 [Available from: https://www.brighton-hove.gov.uk/joint-strategic-needs-assessment]
9. UK Health Security Agency. Summary profile of local authority sexual health Brighton and Hove 2024 [Available from: https://fingertips.phe.org.uk/static-reports/sexualhealth-reports/2024/E06000043.html?area-name=Brighton%20and%20Hove]
10. NHS. Sussex Integrated Care System 2023 [Available from: https://hhmicrosites.com/hh-microsites-home/sussex-joint/introduction/nhs-sussex-welcome/nhs-sussex-sussex-ics/]
11. Sussex Health and Care. What is an ICS? 2025 [Available from: https://www.sussex.ics.nhs.uk/about-sussex-health-and-care/frequently-asked-questions/]
12. Sussex Health and Care. Our Care Connected: Sussex Health and Care 2022 [Available from: https://www.sussex.ics.nhs.uk/our-vision/priorities-and-programmes/digital/our-care-connected/]
13. European Union. Art. 9 GDPR Processing of special categories of personal data 2018 [Available from: https://gdpr-info.eu/art-9-gdpr/]
14. European Union. Art 6 GDPR Lawfulness of Processing 2018 [Available from: https://gdpr-info.eu/art-6-gdpr/]
15. NHS Digital. National data opt-out 2018 [Available from: https://digital.nhs.uk/services/national-data-opt-out-programme]
16. Ford E, Tyler R, Johnston N, Spencer-Hughes V, Evans G, Elsom J, et al. Challenges Encountered and Lessons Learned When Using a Novel Anonymised Linked Dataset of Health and Social Care Records for Public Health Intelligence: The Sussex Integrated Dataset. Information. 2023;14(2):106. https://doi.org/10.3390/info14020106
17. NHS England. Kent, Medway and Sussex Secure Data Environment 2025 [Available from: https://digital.nhs.uk/data-and-information/research-powered-by-data/sde-network/kent-medway-and-sussex-secure-data-environment]
18. Kent Medway and Sussex Secure Data Environment. Metadata Browser 2025 [Available from: https://kms.metadata.works/browser/search?include=dataset::datastandard::terminology::dataclass::dataelement]
19. NHS England. Core20PLUS5 (adults) – an approach to reducing healthcare inequalities 2025 [Available from: https://www.england.nhs.uk/about/equality/equality-hub/national-healthcare-inequalities-improvement-programme/core20plus5/]
20. Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, et al. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. PLoS medicine. 2015;12(10). https://doi.org/10.1371/journal.pmed.1001885
21. Ford E, Rees-Roberts M, Stanley K, Goddard K, Giles S, Armes J, et al. Understanding how to build a social licence for using novel linked datasets for planning and research in Kent, Surrey and Sussex: results of deliberative focus groups. International Journal of Population Data Science. 2023;5(3). https://doi.org/10.23889/ijpds.v5i3.2114

22. Department of Health and Social Care. Fingertips Public Health Profiles 2025 [Available from: https://fingertips.phe.org.uk/]

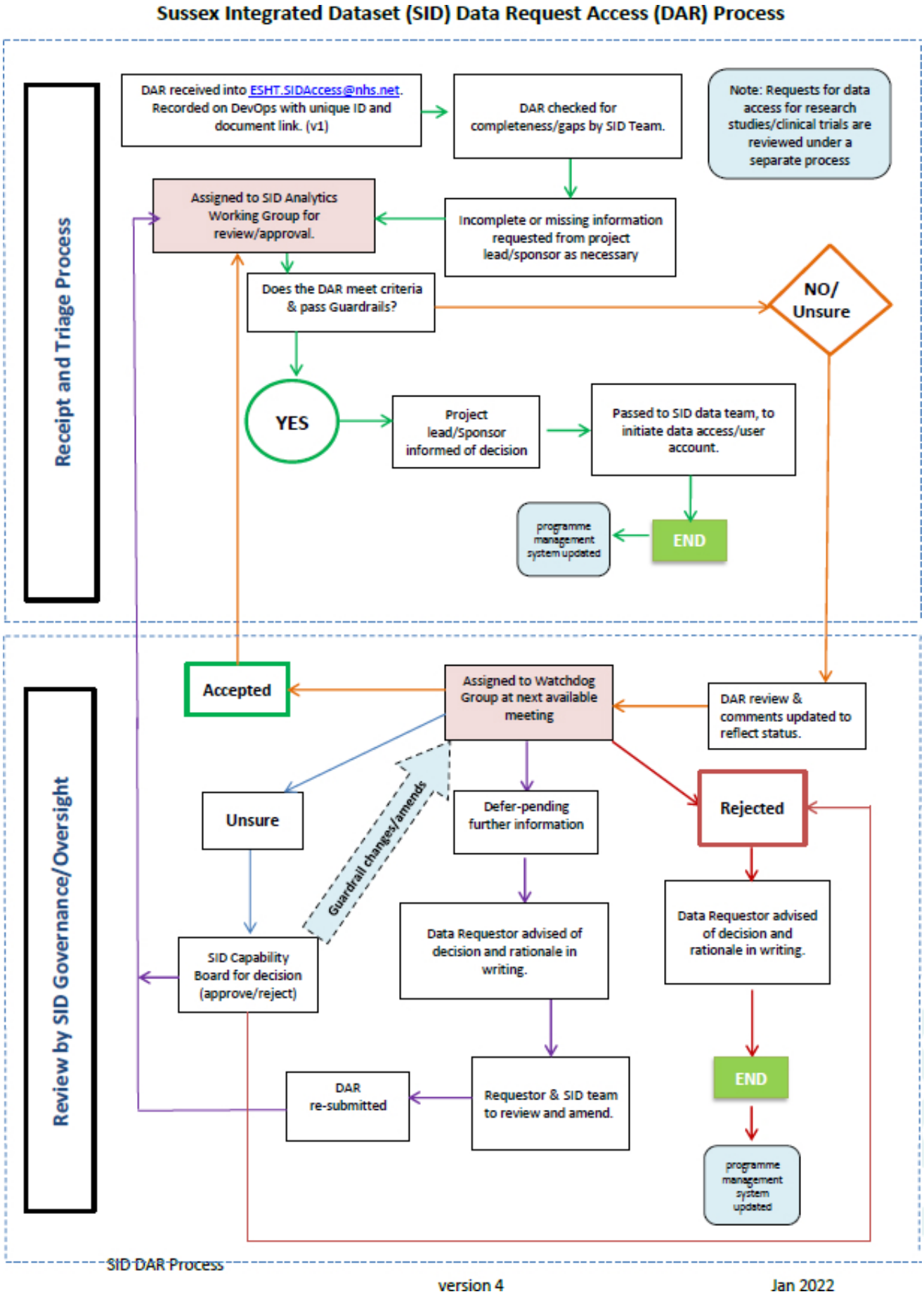
23. Smith MI, de Lusignan S, Mullett D, Correa A, Tickner J, Jones S. Predicting Falls and When to Intervene in Older People: A Multilevel Logistical Regression Model and Cost Analysis. PLOS ONE. 2016;11(7):e0159365. https://doi.org/10.1371/journal.pone.0159365

24. NHS Employers, British Medical Association. 2014/15 General Medical Services (GMS) Contract Quality and Outcomes Framework (QOF): Guidance for GMS Contract 2014/15. 2014.

25. Childs S, Farmer C, George A, Ford E, Rees-Roberts M. Data resource profile: Exploring freely accessible data describing wider determinants of health in England. International Journal of Population Data Science. 2024;8(6):2384. https://doi.org/10.23889/ijpds.v8i6.2384

Supplementary Materials

Supplementary Figure 1: Project Application Process for "Internal" SID users



Supplementary Figure 2: Trusted Analyst Application Process for "Internal" SID users