

International Journal of Population Data Science

Journal Website: www.ijpds.org



Swansea University
Prifysgol Abertawe

Data Resource Profile: Public Health Scotland (PHS) Homecare Medicines Dataset: A National Resource for Linked Prescribing Data for Specialist Medicines Prescribed in Hospital Outpatient setting and Supplied via Homecare Services

Amanj Kurdi^{1,2,3,4,5*}, Laura Stobo¹, Morven Millar¹, Will Clayton¹, Andrew Merrick¹, Stuart McTaggart¹, Tanja Mueller^{1,2}, and Marion Bennie^{1,2}

Submission History

Submitted:	17/07/2025
Accepted:	27/10/2025
Published:	xx/xx/20xx

¹ Public Health Scotland, Scotland, UK

² Strathclyde Institute of Pharmacy and Biomedical Sciences, University of Strathclyde, Glasgow, Scotland, UK

³ Department of Clinical Pharmacy, College of Pharmacy, Hawler Medical University, Erbil, Iraq

⁴ College of Pharmacy, Al-Kitab University, Kirkuk 36015, Iraq

⁵ Department of Public Health Pharmacy and Management, School of Pharmacy, Sefako Makgatho Health Sciences University, Pretoria 0204, South Africa

Abstract

Introduction

The Homecare Medicines (HCM) dataset is a national, patient-level dataset developed by Public Health Scotland (PHS) to capture the supply of specialist medicines delivered through homecare services in Scotland. These services are a critical component of outpatient treatment pathways, particularly for long-term conditions requiring specialist care, such as inflammatory arthritis, cancer, and immune-mediated diseases. Prior to 2019, data on homecare prescribing were fragmented and locally held, limiting national analyses.

Methods

The dataset was initially established during the COVID-19 pandemic to identify immunocompromised patients for vaccine prioritisation. Monthly supply-level data are submitted by homecare providers. Each record includes a pseudonymised unique patient identifier, derived through national health person-level data linkage processes and standardised medicine information mapped to the NHS dictionary of medicines and devices (dm+d), including medicine name (brand and/or generic), formulation and supply date, and, where provided, treatment indication. The presence of a unique patient identifier enables deterministic linkage with a range of national datasets, including community and hospital prescribing, hospital admissions, mortality, cancer registry, and demographic indicators.

Results

The HCM dataset is held securely within the PHS national data infrastructure and accessed via the National Safe Haven. As of April 2025, it includes data from five national providers and covers approximately 98% of the Scottish homecare market. The dataset comprises over 1.3 million supply records for more than 41,000 patients since 2019. Data quality is high for core fields, with missingness levels very low—almost all key variables have <1% missing values—and more than 99.9% of records are successfully indexed with the unique patient identifier. Indication data is partially complete and improving. Medicines are coded using standardised drug dictionaries.

Conclusion

Access to the HCM dataset is available through eDRIS subject to Public Benefit and Privacy Panel (HSC-PBPP) approval. The dataset is well suited for studies on medicine utilisation, equity in access, treatment outcomes, and service planning. Ongoing improvements include enhanced indication capture and integration with Scotland's wider digital prescribing infrastructure.

Keywords

home care services; specialty drugs; medical record linkage; Pharmacoepidemiology; observational studies

Key features

- **Unique coverage of specialist outpatient medicines:** The Homecare Medicines (HCM) dataset is the national resource in Scotland capturing patient-level data on dispensed/supplied specialist medicines prescribed in secondary care and delivered directly to patients' homes, a model commonly referred to as "homecare" in the UK NHS context, which account for a substantial proportion of secondary care prescribing costs.
- **Created to support pandemic response and beyond:** The dataset was established during the COVID-19 pandemic to identify clinically vulnerable patients for vaccine prioritisation and has since evolved into a resource for research, surveillance, and service planning and monitoring.
- **Comprehensive and growing dataset:** Held by Public Health Scotland, the HCM dataset includes over 1.3 million supply records for more than 41,000 patients from 2019 onwards, covering approximately 98% of the Scottish homecare market.
- **Rich linkage potential:** The presence of a unique patient identifier enables deterministic linkage to national datasets including community prescribing (Prescribing Information System, PIS), hospital inpatient medicines administration (Hospital Electronic Prescribing and Medicines Administration, HEPMA), hospital admissions, cancer registry, mortality, and demographic indicators.
- **Detailed prescribing data:** Each record contains structured fields on patient demographics, medicine name, dosage (a free text as stated/entered by the prescriber), formulation, supply date, and treatment indication (where available), mapped to the NHS dictionary of medicines and devices (dm+d), British National Formulary (BNF), and anatomical therapeutic chemical (ATC) coding standards.
- **Secure access for approved researchers:** Data access is available via the Public Health Scotland Electronic Data Research and Innovation Service (eDRIS) within the National Safe Haven, subject to approval by the Public Benefit and Privacy Panel for Health and Social Care (HSC-PBPP). Contact: phs.edris@phs.scot.

*Corresponding Author:

Email Address: amanj.baker@strath.ac.uk (Amanj Kurdi)

Background

Medicines Homecare Services (homecare) are a growing component of the healthcare delivery model across the UK and internationally, providing patients with direct-to-home access to specialist medicines initiated in secondary care. In Scotland, this model has experienced rapid expansion over the past decade, with over 41,000 patients receiving medicines via homecare in 2023, accounting for approximately 30% of the national secondary care medicines budget [1]. Homecare medicines are prescribed by specialist clinicians, typically in outpatient secondary care settings, and supplied to patients by third-party providers contracted by NHS Scotland. These providers include private companies commissioned either by the NHS or pharmaceutical manufacturers [1]. Homecare services offer substantial benefits, including increased convenience for patients, reduced burden on hospital outpatient pharmacy services, streamlined medicine supply processes, and perceived cost savings to the health system [1]. These services are particularly important for long-term conditions requiring regular administration of high-cost biologic or specialist medicines, such as those used in oncology, rheumatology, and immunology [2, 3].

Despite the clinical and operational benefits of homecare, until recently, patient-level data on homecare medicines prescribing in Scotland had been fragmented and inconsistently available. Prior to 2019, data could only be accessed locally within NHS Health Boards (regional organisations responsible for planning and delivering healthcare services to their local populations across Scotland) through hospital pharmacy stock control systems or directly from individual homecare providers. This created significant limitations for national service evaluation, equity monitoring, and pharmacoepidemiological research including evaluation of real-world uptake, effectiveness and safety of medicines. This lack of centralised data infrastructure stood in contrast to Scotland's other medicines national datasets for community prescribing (Prescribing Information System, PIS) [4] and hospital inpatient medicines administration (Hospital Electronic Prescribing and Medicines Administration, HEPMA) [5].

The COVID-19 pandemic accelerated efforts to address these data gaps. Many medicines supplied via homecare are immunomodulatory, and patients receiving them were at increased clinical risk during the pandemic. Public Health Scotland (PHS) responded by initiating the development of a national Homecare Medicines dataset (HCM), initially developed to support the identification of clinically vulnerable patients and to understand the impact of COVID-19 on these groups, and later used to support COVID-19 vaccine and treatment prioritisation. Recognising the broader strategic value of the dataset, the scope expanded to align with the Scottish Government's Digital Health and Care Strategy [6]. This strategy promotes a whole-system, data-enabled approach to delivering person-centred care and supports integration across primary, secondary, and community care settings. The aim of this manuscript is to provide a comprehensive data resource profile of the PHS Homecare Medicines dataset, detailing its structure, coverage, linkage capabilities, and utility for population-level

pharmacoepidemiological research and service planning in Scotland and beyond.

Methods

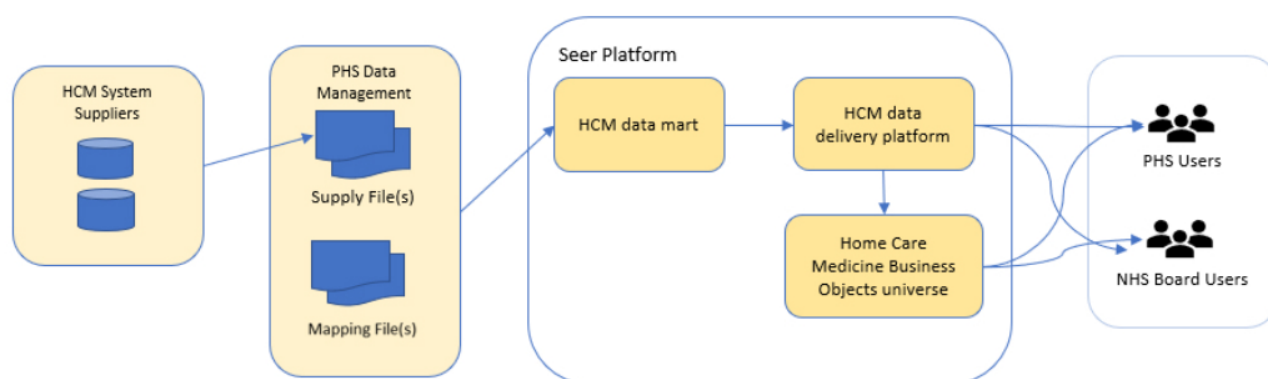
Dataset origin

The HCM dataset was established to capture patient-level information on specialised medicines prescribed in secondary care and delivered through homecare services, a route of medicines provision not represented in existing national prescribing datasets such as PIS or HEPMA. Recognising its broader strategic utility, the dataset is now a near-complete national resource for homecare prescribing data in Scotland. The dataset is compiled from data provided by commercial homecare companies contracted by NHS Scotland to deliver medicines directly to patients' homes.

Data collection and processing

Prescriptions are currently issued on paper and processed by local NHS Health Board homecare medicines teams before transmission to the homecare providers. Each provider submits monthly data extracts to PHS via secure file transfer. Upon receipt, PHS conducts manual validation and data standardisation before integration into its national data infrastructure. Each record includes a pseudonymised unique patient identifier, derived through national health person-level data linkage processes. Deterministic linkage using the Community Health Index (CHI) number has been applied to enhance the dataset with sociodemographic and clinical information from other national sources. The CHI is a population-wide identifier, assigned at birth or upon first registration with NHS Scotland, and is used across all health and care records to support safe and accurate linkage of individual-level data. Through this process, the Homecare dataset can be enriched with information from a range of national datasets including hospital admissions (Scottish Morbidity Records, SMR), the Scottish Cancer Registry, and mortality data from National Records of Scotland (NRS), as well as demographic variables and the ability to identify prescribing events for the same individual over time, and integrate with other national medicines datasets. Data are curated and maintained within a secure PHS corporate data warehouse. All records are retained, enabling robust data linkage and cohort creation for research and audit purposes. Figure 1 provides a schematic overview of the data flows and permissions underpinning the HCM dataset. Supply and mapping files are transferred securely from homecare system suppliers to PHS data management, then ingested into the Seer Platform. Within Seer, the data are processed into the HCM data mart and delivered through a dedicated platform and the Homecare Medicines Business Objects universe, making the data accessible to both PHS and NHS Health Board users under approved governance arrangements. Access controls ensure that NHS Health Board users can only view records relating to patients who either reside within their Board area or have received homecare medicines supplied within their Board, thereby maintaining strict information governance and patient confidentiality.

Figure 1: Data flows and permissions for the Homecare Medicines (HCM) dataset



Results

Dataset size and temporal coverage

The HCM dataset captures national homecare prescribing activity from January 2019 onward. While early submissions varied across providers, consistent national coverage is available from 2020, with increasing provider participation over time. As of April 2025, the dataset comprises monthly submissions from five major homecare providers, with two additional providers in the process of being onboarded. Although the exact number of providers contracted with NHS Scotland is not available as contractual arrangements are subject to change over time, these five providers account for approximately 98% of all homecare services nationally in Scotland, ensuring near-complete coverage of the market [7]. Across all submissions, the dataset encompasses over 1.3 million individual supply records and covers more than 41,000 unique patients across all therapeutic areas served by homecare services.

Structure and composition of records

Each record in the HCM dataset represents a discrete medicine supply event and contains more than fifty structured variables spanning several key domains (Table 1). Patient-level characteristics captured include sex, age, and NHS Health Board of residence, derived through linkage to the national CHI register. Information on NHS Health Board of treatment, and prescribed therapy originates from the homecare providers' data submissions. Comprehensive details on each medicinal product are recorded, with medicine names mapped to the NHS dictionary of medicines and devices (dm+d). Depending on the level of submission, this may include the Virtual Therapeutic Moiety (VTM, the active substance), Virtual Medicinal Product (VMP, the generic product description), or Actual Medicinal Product (AMP, the specific branded product), thereby capturing either generic, branded, and/or therapeutic group information as appropriate. This mapping enables subsequent classification using systems such as the British National Formulary (BNF) and the Anatomical Therapeutic Chemical (ATC) classification. Additional product-level information includes the quantity supplied, date of supply, unique order identifiers, and dosage instructions, which is recorded and submitted as

a free text, exactly as entered by the prescriber and can include details such as strength, frequency, and instructions for use (e.g. "1000 mg paracetamol qds prn"). Where provided, clinical metadata describe the indication for which the medicine was prescribed. Where indication data is provided, entries are grouped into broader condition categories to support exploratory analysis. These groupings are informed by alignment with the International Classification of Diseases, Tenth Revision (ICD-10); for example, rheumatoid arthritis is classified under inflammatory polyarthropathies. In addition, the dataset includes operational metadata such as the identity of the homecare provider, sending location codes and names, timestamps for data loading and updates, and contract references. To support quality assurance and audit, further fields address data integrity and validation, including processing timestamps and indicators of validation status. It is important to note that any data fields deemed confidential under national information governance frameworks are excluded from release to researchers. These include identifiable personal information or commercially sensitive fields such as provider-level contractual details or operational arrangements, which are not necessary for analytical use. The dataset does not currently include information on costs, ancillary services, or delivery charges. Only non-confidential, pseudonymised data are accessible for approved research through secure environments.

Update frequency and data curation

HCM data are submitted monthly by each provider to PHS via secure electronic transfer using predefined data specifications. Submissions are processed centrally by the PHS national medicines team. Standard operating procedures (SOPs) include automated and manual validation routines to identify missing values, resolve duplication, and map drug codes to national dictionaries. Validated records are stored in PHS's secure corporate data warehouse; selected data items are then extracted by the Electronic Data Research and Innovation Service (eDRIS) [8] into a curated version of the HCM dataset, which is further processed to support approved research projects and analytical programmes within the national Trusted Research Environment (TRE). Data curation is ongoing, with improvements made to field standardisation, indication mapping, and metadata documentation as part

Table 1: Summary of key variables in the public health scotland homecare medicines (HCM) dataset

Variable Name	Description	Data Type
Patient Demographics		
Patient UPI Number	Unique Patient Identifier	String
Patient Sex	The sex of the patient at birth expressed as a code e.g. 1 = Male, 2 = Female	String
Patient Sex Desc	The sex of the patient at birth expressed as text e.g. Male	String
Patient Age at Supply Date	The age in years between the Patient Date of Birth and the Supply Date using the Seer standard age derivation.	Numeric
Patient NHS Board Code	9-digit code of the NHS Board associated with the patient's postcode at the Request Received Date using the NHS Board configuration applicable at that time	String
Patient NHS Board Name	NHS Board associated with the patient's postcode at the Request Received Date using the NHS Board configuration applicable at that time	String
Patient Scottish NHS Board Flag	9-digit code of the NHS board associated with the patient's postcode at the Request Received Date using the NHS Board current configuration	String
Patient Postcode Area	Postcode area in which the patient is a resident at the Supply Date.	String
Patient Postcode District	Postcode district in which the patient is a resident at the Supply Date.	String
Patient Postcode Sector	Postcode sector in which the patient is a resident at the Supply Date.	String
Patient Data Zone	The Data zone from the Seer geography reference information for the patient postcode on the record.	String
Patient Intermediate Zone	Associated with the Patient Postcode at the Supply Date.	String
Patient Urban Rural	From the Seer geography reference information for the patient postcode on the record. Used to differentiate between different types of rural and urban areas	String
Patient Urban Rural Group 8	From the Seer geography reference information for the patient postcode on the record. Used to differentiate between different types of rural and urban areas	String
Patient Urban Rural Group 6	From the Seer geography reference information for the patient postcode on the record. Used to differentiate between different types of rural and urban areas	String
Patient Urban Rural Group 2	From the Seer geography reference information for the patient postcode on the record. Used to differentiate between different types of rural and urban areas	String
Geographic and Socioeconomic Indicators		
Prompt Dataset - Deprivation Type	The deprivation type for the respective deprivation record. (SIMD or CARST - Carstairs)	String
Prompt Dataset - Deprivation Year	The deprivation year for the respective deprivation record	Numeric
Prompt Dataset - Deprivation Area Type	The deprivation area basis for the respective deprivation record. (Data zone or Output Area)	String
Prompt Dataset - Deprivation Area Year	The published year for the deprivation area type for the respective deprivation record (Year of respective data zone or output area)	Numeric
Prompt Dataset - Deprivation Area Value	The deprivation area type value (data zone or output area)	String
Prompt Dataset - Deprivation Score	The SIMD Score is an area based measure, calculated at Data Zone level and has seven domains (income, employment, education, housing, health, crime and geographical access). These have been combined into an overall index or score.	Numeric
Prompt Dataset - Deprivation Top 15	A marker (1=yes, 0=no) to determine whether the Data Zone is amongst the top 15% most deprived Data Zones in Scotland based on the SIMD score.	Numeric
Prompt Dataset - Deprivation Bot 15	A marker (1=yes, 0=no) to determine whether the Data Zone is amongst the bottom 15% most deprived Data Zones in Scotland based on the SIMD score.	Numeric
Prompt Dataset - Deprivation Scot Quintile	A categorisation which divides the Scottish population into five equal categories based on the range of SIMD scores so that approximately 20% of the population falls into each quintile (population weighted). Quintile 1 is the MOST deprived, quintile 5 the LEAST deprived.	Numeric

Continued

Table 1: Continued

Variable Name	Description	Data Type
Prompt Dataset - Deprivation Scot Decile	A categorisation which divides the Scottish population into ten equal categories based on the range of SIMD scores so that approximately 10% of the population falls into each decile (population weighted). Decile 1 is the MOST deprived, decile 10 the LEAST deprived.	Numeric
Prompt Dataset - Deprivation Scot Category	The category of deprivation of the Scottish population	Numeric
Prompt Dataset - Deprivation HB Quintile	A categorisation which divides the population of each Health Board into five equal categories based on the range of SIMD scores so that approximately 20% of the population falls into each quintile (population weighted). Quintile 1 is the MOST deprived, quintile 5 the LEAST deprived.	Numeric
Prompt Dataset - Deprivation HB Decile	A categorisation which divides the population of each Health Board into ten equal categories based on the range of SIMD scores so that approximately 10% of the population falls into each decile (population weighted). Decile 1 is the MOST deprived, decile 10 the LEAST deprived.	Numeric
Prompt Dataset - Deprivation HSCP Quintile	A categorisation which divides the population of each HSCP into five equal categories based on the range of SIMD scores so that approximately 20% of the population falls into each quintile (population weighted). Quintile 1 is the MOST deprived, quintile 5 the LEAST deprived.	Numeric
Prompt Dataset - Deprivation HSCP Decile	A categorisation which divides the population of each HSCP into ten equal categories based on the range of SIMD scores so that approximately 10% of the population falls into each decile (population weighted). Decile 1 is the MOST deprived, decile 10 the LEAST deprived.	Numeric
Hist Dataset - Output Area	The Output Area value for which the deprivation values were derived for.	String
Hist Dataset - Output Area Year	The published year of Output Area for which the deprivation values were derived for.	Numeric
Hist Dataset - Deprivation Type	The deprivation type for the respective deprivation record. (SIMD or CARST - Carstairs)	String
Hist Dataset - Deprivation Year	The deprivation year for the respective deprivation record	Numeric
Hist Dataset - Deprivation Area Type	The deprivation area basis for the respective deprivation record. (Data zone or Output Area)	String
Hist Dataset - Deprivation Area Year	The published year for the deprivation area type for the respective deprivation record (Year of respective data zone or output area)	Numeric
Hist Dataset - Deprivation Area Value	The deprivation area type value. (Data zone or Output Area value)	String
Hist Dataset - Deprivation Score	The SIMD Score is an area-based measure, calculated at Data Zone level and has seven domains (income, employment, education, housing, health, crime and geographical access). These have been combined into an overall index or score.	Numeric
Hist Dataset - Deprivation Top 15	A marker (1=yes, 0=no) to determine whether the Data Zone is amongst the top 15% most deprived Data Zones in Scotland based on the SIMD score.	Numeric
Hist Dataset - Deprivation Scot Quintile	A categorisation which divides the Scottish population into five equal categories based on the range of SIMD scores so that approximately 20% of the population falls into each quintile (population weighted). Quintile 1 is the MOST deprived, quintile 5 the LEAST deprived.	Numeric
Hist Dataset - Deprivation Bot 15	A marker (1=yes, 0=no) to determine whether the Data Zone is amongst the bottom 15% most deprived Data Zones in Scotland based on the SIMD score.	Numeric
Hist Dataset - Deprivation Scot Decile	A categorisation which divides the Scottish population into ten equal categories based on the range of SIMD scores so that approximately 10% of the population falls into each decile (population weighted). Decile 1 is the MOST deprived, decile 10 the LEAST deprived.	Numeric

Continued

Table 1: Continued

Variable Name	Description	Data Type
Hist Dataset - Deprivation Scot Category	The category of deprivation of the Scottish population	Numeric
Hist Dataset - Deprivation HB Quintile	A categorisation which divides the population of each Health Board into five equal categories based on the range of SIMD scores so that approximately 20% of the population falls into each quintile (population weighted). Quintile 1 is the MOST deprived, quintile 5 the LEAST deprived.	Numeric
Hist Dataset - Deprivation HB Decile	A categorisation which divides the population of each Health Board into ten equal categories based on the range of SIMD scores so that approximately 10% of the population falls into each decile (population weighted). Decile 1 is the MOST deprived, decile 10 the LEAST deprived.	Numeric
Hist Dataset - Deprivation HSCP Quintile	A categorisation which divides the population of each HSCP into five equal categories based on the range of SIMD scores so that approximately 20% of the population falls into each quintile (population weighted). Quintile 1 is the MOST deprived, quintile 5 the LEAST deprived	Numeric
Hist Dataset - Deprivation HSCP Decile	A categorisation which divides the population of each HSCP into ten equal categories based on the range of SIMD scores so that approximately 10% of the population falls into each decile (population weighted). Decile 1 is the MOST deprived, decile 10 the LEAST deprived.	Numeric
Treatment NHS Board Code	The Health Board code associated with the Location Code at the Supply Date for the current NHS Board configuration (Health Board Code 9 Curr).	String
Treatment NHS Board Name	The Health Board name associated with the Treatment Health Board Code.	String
Medication Details and Classification		
Medication Name	The mapped name of the medicine/regimen prescribed, which can include the formulation and strength	String
Medication Name Submitted	Unaltered name of the medicine/regimen prescribed, which can include the formulation and strength	String
DMD Code	The DMD code of the prescribed medication. Will be VMP or AMP (Actual Medicinal Product, specific branded product)	String
DMD Current Unique Ref	From the Seer dmd reference information for the dmd code on the record.	String
VTM ID	The VTM (Virtual Therapeutic Moiety) ID of the prescribed medication. From the Seer dmd reference information for the dmd code on the record.	String
VTM Name	The VTM (Virtual Therapeutic Moiety, the active substance) name of the prescribed medication. From the Seer dmd reference information for the dmd code on the record.	String
DDD Conversion Factor	The defined daily dose conversion factor. From the Seer dmd reference information for the dmd code on the record.	Numeric
VMP Name	The VMP (Virtual Medical Product, the generic product description) name of the prescribed medication. From the Seer dmd reference information for the dmd code on the record.	String
BNF Code	The BNF code from the Seer BNF reference information associated with the dmd code on the record.	String
BNF Chapter Code	BNF chapter code as defined by the British National Formulary	String
BNF Chapter Description	BNF chapter name as defined by the British National Formulary	String
BNF Chapter	BNF chapter code and descriptions as defined by the British National Formulary	String
BNF Section Code	BNF section code as defined by the British National Formulary	String
BNF Section Description	BNF section name as defined by the British National Formulary	String
BNF Section	BNF section code and description as defined by the British National Formulary	String
BNF Sub Section Code	BNF sub-section code as defined by the British National Formulary	String

Continued

Table 1: Continued

Variable Name	Description	Data Type
BNF Sub Section Description	BNF sub-section name as defined by the British National Formulary	String
BNF Sub Section	BNF sub-section code and description as defined by the British National Formulary	String
BNF Paragraph Code	The BNF Paragraph code from the Seer BNF reference information for the BNF code on the record.	String
BNF Paragraph Description	The BNF Paragraph description from the Seer BNF reference information for the BNF code on the record.	String
BNF Paragraph	BNF paragraph code and description as defined by the British National Formulary	String
ATC Code	The ATC (Anatomical Therapeutic Chemical) code associated with the dmd code on the record.	String
ATC Anatomical Main Group Code	The ATC First Level Code of the 14 groups, e.g. R	String
ATC Anatomical Main Group Description	The ATC First Level Description of the 14 groups, e.g. Respiratory	String
ATC Therapeutic Subgroup Code	The ATC Second Level Code e.g. G04	String
ATC Therapeutic Subgroup Description	The ATC Second Level Description e.g. Urological	String
ATC Pharmacological Subgroup Code	The ATC Third Level Code e.g. L04A	String
ATC Pharmacological Subgroup Description	The ATC Third Level Description e.g. Immunosuppressants	String
ATC Chemical Subgroup Code	The ATC Fourth Level Code e.g. L03AB	String
ATC Chemical Subgroup Description	The ATC Fourth Level Description e.g. Interferons	String
ATC Chemical Substance Code	The ATC Fifth Level Code e.g. B01AB01	String
ATC Chemical Substance Description	The ATC Fifth Level Description e.g. Heparin	String
Supply Information		
Supply Quantity	The quantity of the medicine supplied to the patient. Only whole numbers will be submitted. Can include negative numbers.	Numeric
Supply Date	The date the medicine was supplied	Date Time
Supply Day Short Name	The day component of the supplied date expressed as an abbreviated textual description e.g. Mon	String
Supply Day Name	The day component of the supplied date expressed as a textual description e.g. Monday	String
Supply Day in Week	The day component of the supply date expressed as number in the week i.e. 1-7, where 1 = Sunday	Numeric
Supply Day in Month	The day component of the supply date expressed as number in the month i.e. 1-31	Numeric
Supply Day in Calendar Year	The day component of the supply date expressed as number in the year i.e. 1-366	Numeric
Supply Calendar Year	The year component of the supply date in the format YYYY e.g. 1999	Numeric

Continued

Table 1: Continued

Variable Name	Description	Data Type
Supply Calendar Month Number	The month component of the supply date in the format 1-12 e.g. 1 = January	Numeric
Supply Calendar Month Name	The month component of the supply date as a full textual description e.g. January	String
Supply Calendar Month Short Name	The month component of the supply date as an abbreviated textual description e.g. Jan	String
Supply Calendar Month and Year	The month and year components of the supply date as an abbreviated textual description e.g. Jan-2023	String
Supply Calendar Quarter	The supply date expressed in terms of which calendar quarter it falls. Format is single value 1-4 where 1 = Jan-Mar	Numeric
Supply Calendar Quarter Name	The supply date expressed in terms of which calendar quarter it falls e.g. Jan-Mar	String
Supply Calendar Week	The supply date expressed in terms of which week of the year it falls. Format is numeric values 1-53 from 1 January	Numeric
Supply Day Financial Year	The day component of the supply date expressed as number in the financial year i.e. 1-366, where 1 is 01 April	Numeric
Supply Day Financial Month Name	The month component of the supply date as a full textual description e.g. April	String
Supply Day Financial Month Short Name	The month component of the supply date as an abbreviated textual description e.g. Apr	String
Supply Day Financial Month Number	The month component of the supply date in the format 1-12 e.g. 1 = April	Numeric
Supply Day Financial Month and Year	The month and year components of the supply date as an abbreviated textual description e.g. Apr-2023	String
Supply Day Financial Quarter	The supply date expressed in terms of which financial quarter it falls. Format is single value 1-4 where 1 = Apr-Jun	Numeric
Supply Day Financial Quarter Name	The supply date expressed in terms of which financial quarter it falls e.g. Apr-Jun	String
Supply Day Financial Week	The supply date expressed in terms of which financial week of the year it falls. Format is numeric values 1-53 from 1 April	Numeric
Supply Unique ID	Unique identifier for the medication supplied	Numeric
Record ID	A system generated code to uniquely identify a specific supply event of a prescription for an individual patient	String
Supply ID	A system generated ID to uniquely identify a prescription supply	String
Submitted_dose_instruction	A free text field submitted exactly as entered by the prescriber. This can include details such as strength, frequency, and instructions for use (e.g. "1000 mg paracetamol qds prn")	String
Clinical Metadata		
Indication	The Indication mapped from the submitted Indication.	String
Submitted Indication	The unaltered submitted indication from the homecare supplier's data	String

of dataset maturity. Version control is maintained, and reprocessing of legacy records occurs periodically to align historical data with current definitions.

Data quality and completeness

The dataset demonstrates consistently high data quality, reflecting its primary operational use in medicines supply

and reimbursement, which requires accuracy for contractual and financial purposes. Quality assurance processes include validation of supply and mapping files against defined formats, conformance checks to the NHS dictionary of medicines and devices (dm+d), and cross-validation with reference datasets. Levels of missingness are generally low, with almost all key/core variables having <1% missing values, with the exception of some fields such as dosage

instructions (~13% missing), and ICD-10 codes (~50% missing). Linkage completeness to the CHI is almost complete, with over 99.9% of records have been successfully indexed, enabling deterministic linkage to other national datasets. Together, these indicators support confidence in the accuracy, completeness, and reliability of the dataset for research and analytical use. The volume of records within the HCM dataset has grown steadily since inception, reflecting phased onboarding of providers. Annual totals increased from 116,227 records in 2019 to 237,704 records in 2022, and further to 268,902 records in 2024 (Figure 2). Some fields, particularly indication and prescription source details, are variably populated, reflecting heterogeneity in provider systems and documentation practices. In the case of indication data, a substantial proportion of entries are recorded as 'missing' or 'not provided', likely due to absent or inconsistent information on the original prescriptions. This variability limits completeness and may affect downstream analyses depending on the intended use. Indication data is incomplete in approximately 25% of supply records, with around 11% containing no submitted indication and a further 13% comprising uninterpretable or non-specific entries. Efforts to improve completeness, including provider engagement and upstream data capture enhancements, are ongoing.

Data linkage utility

The inclusion of a unique patient identifier (CHI), derived through national person-level data linkage processes, in nearly all HCM records enables deterministic linkage to a wide range of national health datasets in Scotland. These include community prescribing (PIS) [4], hospital inpatient medicines administration (HEPMA) [5], hospital inpatient and outpatient records (SMR01 and SMR00) [9], cancer registration data (SMR06) [10], and mortality data from National Records of Scotland (NRS) [11]. The dataset also enables connections to births, maternity, and neonatal datasets such as SMR02 [12] and the Scottish Birth Record (SBR) [13], as well as to immunisation [14] and COVID-19 testing [15]. Furthermore, it can be linked to demographic and socioeconomic indicators, including the Scottish Index of Multiple Deprivation (SIMD) and classifications of urban and rural residence. These linkages support detailed longitudinal analyses of treatment initiation, medicine switching, persistence, healthcare utilisation, and the real-world effectiveness of prescribed therapies. This enables a whole-system view of medicines use across the continuum of NHS Scotland healthcare delivery. Harmonisation of variables across the component datasets such as standardising supply dates and drug coding schemes ensures comparability and longitudinal tracking. In addition, the HCM dataset is fully integrated into the Scottish Combined Medicines Dataset (SCoMeD) [16], which facilitates harmonised, cross-sector analysis of prescribing across the homecare, community (PIS), and hospital (HEPMA) settings. This enables a whole-system view of medicines use across the continuum of NHS Scotland healthcare delivery.

Research applications

The HCM dataset has already been utilised in national surveillance and research efforts, most notably during the COVID-19 pandemic, where it played a key role in identifying clinically vulnerable individuals eligible for vaccine prioritisation across NHS Scotland. Since then, it has supported a range of analytical applications to generate health intelligence directed to the health system in Scotland. These have included monitoring the uptake of biosimilar medicines, characterising treatment patterns in therapeutic areas such as rheumatology [17], and assessing variations in prescribing practices across NHS Health Boards, through an NHS access portal to support clinical audit. Data can also be applied in investigating patterns of polypharmacy in patients receiving complex regimens. Building on this foundation, the dataset continues to be explored for its potential to support studies on treatment effectiveness, medicine adherence and persistence, and health economic evaluations, particularly in alignment with Scotland's Value-Based Health and Care policy framework [18]. We acknowledge that the dataset captures only supply data and does not provide structured dosage instructions or information on how often patients take their medicines and therefore cannot in isolation provide a full picture of adherence behaviours. This limitation can be addressed by combining the dataset with complementary study designs, such as patient surveys or qualitative research. In situations where such approaches are not feasible or where rapid insights are needed, assumptions can be made using established methods such as the WHO Defined Daily Dose (DDD) or prescribed daily dose estimates. While these approaches have recognised limitations, they can nonetheless provide useful proxy measures of adherence and help generate timely evidence. Similarly, health-economic analyses would be strengthened by the inclusion of cost data, which are not currently part of the dataset.

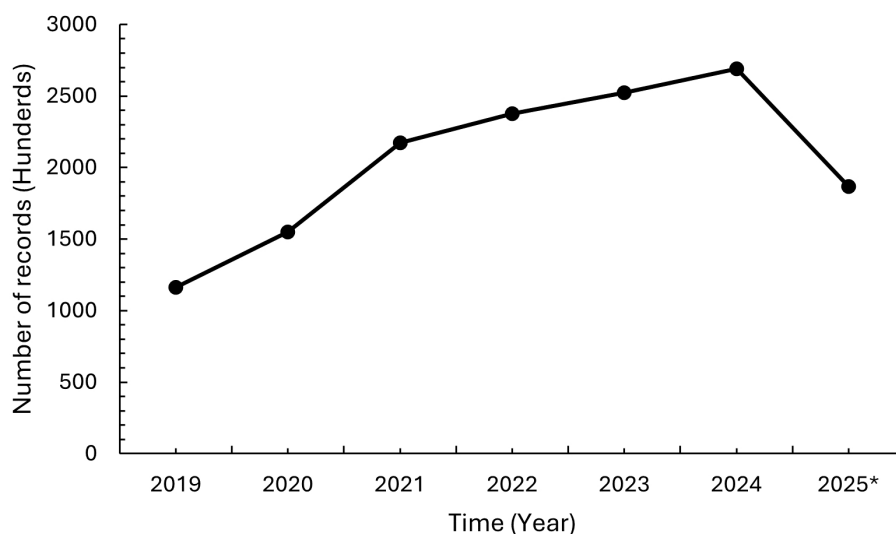
The most recent case example demonstrating the utility of the HCM dataset is a national pharmacoepidemiological study focused on the use of biologic disease-modifying antirheumatic drugs (bDMARDs) in patients with inflammatory rheumatological conditions in Scotland [17]. The study aimed to describe real-world prescribing trends, treatment sequencing, and geographic variation in access to bDMARDs across the country between 2019 and 2023. Leveraging the high coverage and linkage capacity of the HCM dataset, this work provided detailed insights into patient demographics, indication-specific prescribing patterns, and inequalities in treatment access by Health Board and deprivation status. The study also explored longitudinal treatment pathways, including persistence and switching behaviour, offering valuable evidence to inform clinical policy, service delivery, and equitable access to advanced therapies.

Discussion

Summary and importance of the dataset

The HCM dataset is a major step forward in Scotland's capability to monitor, evaluate, and understand the use of specialist medicines delivered through homecare

Figure 2: Trends of annual volume of records in the Homecare Medicines (HCM) dataset, 2019–2025



services. Initially established in response to the COVID-19 pandemic to identify immunosuppressed patients for vaccine prioritisation [1], the dataset has evolved into a comprehensive national resource, capturing over 98% of homecare medicines activity across Scotland [7]. Its integration into the Scottish Combined Medicines Dataset (SCoMeD) has expanded its value, enabling a whole-system view of prescribing across primary, secondary, and ambulatory care [16]. The HCM dataset therefore addresses a previously unmet need for consistent, linkable data on medicines supplied via homecare, a sector accounting for approximately 30% of the Scottish secondary care medicines budget [1].

Key strengths

A key strength of the HCM dataset is its population coverage and completeness. Monthly data submissions from five major homecare providers, governed by national service-level agreements, are processed, validated, and curated by PHS, ensuring high levels of data quality. The presence of a unique patient identifier in over 99.9% of records facilitates deterministic linkage with other nationally maintained datasets, including hospital admissions (SMR01), outpatient records (SMR00), cancer registration (SMR06), mortality data from National Records of Scotland (NRS), maternity and neonatal datasets (SMR02, SBR), vaccination records, and socioeconomic indicators such as SIMD [16, 19]. In addition to its breadth, the dataset offers detailed granularity at the prescription level. Each supply record includes the product name, quantity, dosage instructions (a free text field), supply date, and is mapped to standard drug dictionaries. Where recorded, clinical indications are also mapped to ICD-10-aligned condition groupings, allowing for disease-specific analyses in areas such as rheumatology, and gastroenterology, and neurology.

Emerging research applications

Beyond its initial role in supporting Scotland's pandemic response, the HCM dataset is increasingly contributing to

a broadening spectrum of real-world evidence generation. Its unique structure, linking outpatient specialised medicine supply data with individual-level clinical and demographic information, makes it well-suited for longitudinal pharmacoepidemiological studies and service evaluations. Applications to date have included analyses of biosimilar adoption trends, bDMARDs prescribing variation across NHS Health Boards [17], and the investigation of prescribing complexity in patients with multimorbidity or polypharmacy.

As use cases expand, the HCM dataset is also increasingly aligned with Scotland's Value-Based Health and Care ambitions [18], supporting evaluation of treatment value, resource allocation, and delivery of person-centred care.

Limitations and areas for development

To support continued usability, structured metadata and publicly available variable definitions are regularly maintained and updated. Despite these strengths, the HCM dataset has limitations. While the dataset captures high-quality operational and supply data, some clinically relevant elements are either incomplete or absent. For example, prescriber specialty, patient weight, and treatment response metrics are not recorded. Indication data, while available in many records, is inconsistently captured and populated due to variation in provider systems and input mechanisms. These issues mirror challenges seen in other national prescribing datasets and are the focus of ongoing improvement initiatives [5, 16]. Future work could consider imputing missing indications by leveraging additional data sources. For example, linkage to clinician speciality or clinic codes, or to hospital admission records within SMR inpatient dataset, where the prescription was associated with an inpatient stay, may provide contextual information about the likely indication. Similarly, medicines recorded in PIS could serve as proxies for certain conditions. In addition, imputation approaches based on standard dosing regimens, treatment patterns, or established pharmacoepidemiological methods (e.g. the WHO ATC/DDD framework) may help approximate indication where direct information is not available. While each of these approaches has limitations

and the potential for misclassification, they represent feasible methodological strategies for enhancing the utility of the dataset. Provider-specific variation in data quality, particularly in optional fields, can also impact analysis, although the introduction of a nationally standardised dataset specification has helped mitigate these issues. Continued engagement with suppliers and the progressive digitisation of homecare prescribing, as recommended in Scotland's national strategic reviews, are expected to further enhance the dataset's completeness and consistency [1, 3].

Future directions

Looking ahead, the HCM dataset is excellently positioned to support an expanding range of use cases. Planned developments include onboarding of additional providers, expansion of metadata documentation, and improved capture of treatment indications. Integration with other digital systems such as the HEPMA dataset and electronic prescribing initiatives in outpatient care will enhance interoperability and analytic potential [5, 16]. The dataset is increasingly being used to inform national medicines intelligence, support horizon scanning and cost-effectiveness modelling, and evaluate prescribing practices in line with the Scottish Government's Value-Based Health and Care Action Plan [18].

Data access and governance

Access to the HCM dataset is available for approved research, evaluation, and public health purposes, subject to appropriate governance approvals. The dataset is held within the secure national infrastructure maintained by PHS and accessed via the National Safe Haven, a Trusted Research Environment (TRE) managed in partnership with eDRIS [8]. All applicants seeking access to the HCM dataset must submit a formal request to the NHS Public Benefit and Privacy Panel for Health and Social Care (HSC-PBPP) [19], which reviews applications for compliance with data protection legislation, ethical standards, and alignment with public interest. Applications must include a clear research question, methodology, and justification for data variables requested. Approval is conditional on adherence to strict data security protocols and information governance requirements. Only pseudonymised data are provided for analysis; no directly identifiable or confidential information is made available to researchers. Fields classified as confidential, commercially sensitive, or unnecessary for the stated purposes are excluded from extract datasets. Researchers must access data within the National Safe Haven and are not permitted to download or transfer raw data outside of this secure environment. Further information on data access procedures, including guidance on the application process and associated timelines, is available through the eDRIS website [8] and the HSC-PBPP governance portal [19].

Conclusion

The HCM dataset fills a critical gap in Scotland's national prescribing intelligence infrastructure. Its strengths in population coverage, linkage potential, and clinical relevance

make it a uniquely valuable tool for researchers, policymakers, and clinicians. As the dataset matures and integration with wider systems continues, it will play an increasingly central role in supporting real-world evaluation of specialist medicines, ensuring equitable access, and driving improvements in patient outcomes across the health system.

Acknowledgments

None

Ethics statement

Ethical approval was not required since this study did not involve human participants, nor animals. No data was collected or generated specifically for this project.

Conflict of interest statement

The authors have no conflict of interest to declare

Publication consent

Consent has been gained from the data provider to publish and openly share the data included in this study.

Funding statement

There was no specific funding for this piece of work. The development and implementation of PHS HCM in Scotland is supported by PHS.

Data availability statement

The HCM dataset is held securely by PHS and is not publicly available due to governance and confidentiality constraints. Access to the dataset is restricted to approved researchers via the National Safe Haven, a Trusted Research Environment managed by PHS in collaboration with the Electronic Data Research and Innovation Service (eDRIS). Researchers wishing to access the HCM dataset must submit a data access application to the NHS Public Benefit and Privacy Panel for Health and Social Care (HSC-PBPP). Approval is granted on the basis of public benefit, ethical compliance, and adherence to data protection standards. Only pseudonymised, non-confidential data are provided for approved uses, and data cannot be exported from the secure environment.

Authors contribution

All authors substantially contributed to the conception and design of the work. AK drafted the initial manuscript, with input from all authors. All authors critically reviewed and revised the draft; and approved the final manuscript.

References

- Hannan BK. Independent Review of Medicines Homecare in Scotland. Scottish Government 2025.
- Department of Health. Homecare Medicines Towards a Vision for the Future. 2011.
- Royal Pharmaceutical Society. Professional Standards for Homecare Services 2024. 2024.
- Alvarez-Madrado S, McTaggart S, Nangle C, Nicholson E, Bennie M. Data resource profile: the Scottish national prescribing information system (PIS). *International journal of epidemiology*. 2016;45(3):714-5f. <https://doi.org/10.1093/ije/dyw060>
- Mueller T, Proud E, Kurdi A, Jarvis L, Reid K, McTaggart S, Bennie M. Data Resource Profile: The Hospital Electronic Prescribing and Medicines Administration (HEPMA) National Data Collection in Scotland. *International Journal of Population Data Science*. 2023;8(6):2182. <https://doi.org/10.23889/ijpds.v8i6.2182>
- Scottish Government. Digital Health and Care Strategy. 2021.
- Public Health Scotland. Home Care Medicines (HCM) data 2025 [Available from: <https://publichealthscotland.scot/resources-and-tools/health-intelligence-and-data-management/national-data-catalogue/national-datasets/search-the-datasets/home-care-medicines-hcm-data/>].
- Public Health Scotland. Electronic Data Research and Innovation Service (eDRIS) 2025 [Available from: <https://publichealthscotland.scot/resources-and-tools/health-intelligence-and-data-management/electronic-data-research-and-innovation-service-edris/overview/what-is-edris/>].
- Public Health Scotland. Definitions by SMR record section 2025 [Available from: <https://www.publichealthscotland.scot/resources-and-tools/health-intelligence-and-data-management/national-data-catalogue/smr-data-manual/definitions-by-smr-record-section/smr00-outpatient-attendance/>].
- public Health Scotland. National Datasets: Scottish Cancer Registry (SMR06) 2025 [Available from: <https://publichealthscotland.scot/resources-and-tools/health-intelligence-and-data-management/national-data-catalogue/national-datasets/search-the-datasets/scottish-cancer-registry-smr06/#:~:text=Public%20Health%20Scotland%20is%20responsible,site>].
- Scottish Government. National Records of Scotland 2025 [Available from: <https://www.nrscotland.gov.uk/>].
- public Health Scotland. SMR02 maternity inpatient and day case 2025 [Available from: <https://publichealthscotland.scot/resources-and-tools/health-intelligence-and-data-management/national-data-catalogue/smr-data-manual/definitions-by-smr-record-section/smr02-maternity-inpatient-and-day-case/general-definitions/>].
- public Health Scotland. National Datasets: Scottish Birth Record (SBR) 2025 [Available from: <https://www.publichealthscotland.scot/resources-and-tools/health-intelligence-and-data-management/national-data-catalogue/national-datasets/search-the-datasets/scottish-birth-record-sbr/>].
- public Health Scotland. National Datasets: Vaccinations dataset 2025 [Available from: <https://publichealthscotland.scot/resources-and-tools/health-intelligence-and-data-management/national-data-catalogue/national-datasets/search-the-datasets/vaccinations-dataset/>].
- public Health Scotland. National Datasets: Test and Protect 2025 [Available from: <https://www.publichealthscotland.scot/resources-and-tools/health-intelligence-and-data-management/national-data-catalogue/national-datasets/search-the-datasets/test-and-protect/>].
- Mueller T JL, Stark V, et al. Data Resource Profile: The Scottish Combined Medicines Dataset (SCoMeD). *International Journal of Population Data Science* (In Press). 2025.
- Kurdi A, Proud E, Stobo L, et al. Prescribing pattern of disease-modifying antirheumatic drugs in Scotland (2019–2023): a population-based retrospective cohort study. *Expert Review of Clinical Pharmacology*. 2025;18(7), 519–530. <https://doi.org/10.1080/17512433.2025.2544038>
- Scottish Government. Value based health and care: action plan 2023 [Available from: https://www.gov.scot/publications/value-based-health-care-action-plan/pages/3/?utm_source=chatgpt.com].
- Public Health Scotland. NHS Scotland Public Benefit and Privacy Panel for Health and Social Care (HSC-PBPP) 2025 [Available from: <https://www.informationgovernance.scot.nhs.uk/pbpphsc/>].

Abbreviations

ATC:	Anatomical Therapeutic Chemical Classification
BNF:	British National Formulary
CHI:	Community Health Index
COVID-19:	Coronavirus Disease 2019
dm+d:	dictionary of medicines and devices
eDRIS:	Electronic Data Research and Innovation Service
GP:	General Practitioner
HCM:	Homecare Medicines
HEPMA:	Hospital Electronic Prescribing and Medicines Administration
HSC-PBPP:	Public Benefit and Privacy Panel for Health and Social Care

ICD-10:	International Classification of Diseases, Tenth Revision	SBR:	Scottish Birth Record
NHS:	National Health Service	SIMD:	Scottish Index of Multiple Deprivation
NRS:	National Records of Scotland	SMR:	Scottish Morbidity Record
PHS:	Public Health Scotland	SCoMeD:	Scottish Combined Medicines Dataset
PIS:	Prescribing Information System	TRE:	Trusted Research Environment

