




Data Resource Profile: IQVIA Medical Research Data (IMRD)

Holly Christina Smith ¹, Juan Carlos Bazo-Alvarez ¹, Louise Pinder², Itohan Evbuomwan², Irene Petersen ¹

¹Department of Primary Care and Population Health, University College London, London, UK; ²Real World Evidence and Commercial & Marketing Solutions, IQVIA, London, UK

Correspondence: Holly Christina Smith, Email Holly.dorning.18@ucl.ac.uk

Abstract: IQVIA Medical Research Data (IMRD) was established to be used for population-based research to provide insights into primary care managed diseases and treatments to help inform healthcare decisions and improve patient outcomes. Uses include determining the burden of illness, treatment patterns and adherence, drug safety, morbidity and mortality patterns, outcomes research, health economics research, risk management, quantifying unmet needs and resource utilisation. IMRD contains longitudinal, non-identified data on more than 6 million people ever registered with a participating primary care practice in England – of these, 2.2 million remain actively registered and contributing data as of January 2024 with a median of 8 years of follow-up data. Both historic and current records are available for research. IMRD contains information on primary care consultations including patient characteristics (eg year of birth and socioeconomic information), clinical events (eg symptoms, diagnoses and referrals), prescribed medications, tests and other health information (eg height and weight) and immunisations. IMRD has NHS Research Authority approval for medical research and treatment analysis. Data can be supplied to external researchers for scientifically approved studies under Data Sharing Agreements. Access is granted by the Scientific Review Committee based on study protocols. All studies are required to show scientific merit, fulfil the research purpose outlined, and demonstrate potential benefit to health and social care.

Keywords: data resource profile, electronic health records, primary care, epidemiology

Data Resource Basics

IQVIA Medical Research Data (IMRD) contains non-identified electronic patient health records supplied from English primary care practices which use EMIS Health or SystemOne clinical management systems and who have agreed to participate in the Medical Research Extraction Scheme (MRES).¹ IMRD can be used for a wide range of research studies, possible applications include determining: the burden of illness, treatment patterns and adherence, drug safety, morbidity and mortality patterns, outcomes research, health economics research, risk management, quantifying unmet needs and resource utilisation. IMRD was newly established in 2021 by IQVIA Ltd which is part of the IQVIA group of companies serving the combined industries of health information technologies and clinical research worldwide. Although IQVIA previously licensed access to data from The Health Improvement Network (THIN), sourced from practices using the Vision clinical system, IMRD is a distinct database and does not reuse any data from THIN. In the UK, IQVIA has collected and supported the research use of non-identified patient data for over 20 years.

IMRD is a large UK-based primary care electronic health record databases and contains information on 6 million registered patients from 189 practices across England, this includes historic and current records. This is roughly 3% of English practices.² This number is set to increase as further primary care practices join the MRES as part of ongoing recruitment from IQVIA. As of January 2024, there were 2.2 million patients actively contributing data to IMRD (they have not died or transferred practice) with a median follow-up time of 8.3 years. IMRD is broadly representative of the UK population in terms of demographics (Table 1). However, there is a slighter greater proportion of people living in the most deprived areas compared to the least. IMRD contains routinely collected patient-level information on



Table 1 Characteristics of Patients Included Within IMRD and Comparisons with National Datasets Where Available/Appropriate

Characteristic	All Patients, n (%)	Active Patients, January 2024, n (%)	National Data, n (%)
Overall	6,057,167	2,220,339	57,690,323 ^a
Sex			
Female	3,104,111 (51.3)	1,096,850 (49.4)	29,407,249 (51.0) ^a
Male	2,952,235 (48.7)	1,123,258 (50.6)	28,283,074 (49.0) ^a
Indeterminate/Missing	821 (0.02)	231 (0.01)	
Age in 2024			
<18	–	425,458 (19.2)	12,669,854 (22.0) ^a
19–29	–	358,649 (16.2)	7,911,774 (13.7) ^a
30–64	–	1,078,721 (48.6)	26,325,608 (45.6) ^a
65+	–	357,511 (16.1)	10,783,087 (18.7) ^a
Indices of multiple deprivation			
1- Most deprived	560,274 (10.1)	206,069 (10.1)	–
2	801,224 (14.5)	292,976 (14.4)	–
3	762,350 (13.8)	273,167 (13.4)	–
4	655,791 (11.8)	233,937 (11.5)	–
5	537,841 (9.7)	198,859 (9.8)	–
6	459,011 (8.3)	167,605 (8.2)	–
7	432,354 (7.8)	164,544 (8.1)	–
8	452,203 (8.2)	171,601 (8.4)	–
9	370,496 (6.7)	143,610 (7.1)	–
10- Least deprived	507,099 (9.2)	179,646 (8.8)	–
Unknown/Missing	518,524	188,325	–
Region			
London	1,983,735 (32.8)	675,480 (30.4)	–
Midlands East England	1,458,377 (24.1)	596,308 (26.9)	–
North of England	907,677 (15.0)	350,121 (15.8)	–
South of England	1,707,378 (28.2)	598,430 (27.0)	–
Follow-up time since practice registration			
Median years (IQR)	–	8.3 (3.1–17.3)	–
Mean years (s.d).	–	12.7 (13.5)	–

Notes: ^aFigures from ONS 2023 mid-year population estimates in England.³

demographics, prescribing, symptoms, procedures, prevention, lifestyle factors and diagnostics. When patients register with a general practice which contributes to IMRD, they are assigned a unique ID within that practice. This allows information on each person to be linked across different data stored within the practice and over calendar time. A family ID is also available which can link patients living within the same address. Patient records are de-identified before data are provided to IMRD and have the potential to be linked via the pseudonymised ID with other datasets, such as secondary care, registries or social care data, using an established pseudonymisation and secure linkage methodology supported by the data suppliers and subject to appropriate project approvals. Although no external datasets are currently linked to IMRD by default, linkage can be undertaken within secure environments by converting personal identifiers (such as NHS number) into the same pseudonymised form in each dataset, allowing records to be matched without sharing or revealing personal identifiers.

Data Collected

IMRD is organised into separate files based on the type of information captured in primary care consultations (Table 2). The different file types can be linked by key identifier variables: practice id, patient id, observation id and staff id to

Table 2 Overview of Data Files and Structure of IMRD

File	Description	Linkage Information
Practice records	Contains practice level information including the area the practice is located and key practice-level dates.	Linked to staff and patient files via <u>practice id</u> .
Staff records	Contains information of the staff member recording the information including staff role (for example GP, nurse etc).	Linked to consultation, observation and drug issue files via <u>staff id</u> .
Patient records	Contains personal demographic information including year of birth (month and year of birth for children), sex, date of practice registration, date of practice transfer and date of death.	Linked to consultation, observation and drug issue files via <u>patient id</u> .
Consultations	Contains information on primary care consultations, including date of consultation and type of consultation (for example telephone consultation, at-home consultation or in-practice consultation etc).	Linked to staff, patient and observation files.
Observations	Contains information on a patient's medical history including symptoms, diagnoses, immunisations tests and lifestyle factors. Clinical information is coded using SNOMED CT.	Linked to consultation files and drug issues by <u>patient id</u> and problem, referrals files by <u>observation id</u> .
Problems	Contains information grouping other files which GP's have determined to be part of the same "problem". For example, diagnoses, tests and drugs relating to a chronic condition.	Linked to observations, referrals, drug issues and consultations.
Referrals	Contains information on referrals to/from the practice.	Linked to observations by <u>observation id</u> .
Drug issues	Contains prescribing information including the generic name of the medication and date of prescription. The dosage, duration and quantity prescribed is also be included where information is available.	Linked to patient, staff, observation and consultation files.

provide a comprehensive picture of patient care. Practice records contain information about the location of each practice and key dates. Staff records detail the role of the healthcare professional (such as GP or nurse) who provided the care. Patient records include demographic information such as year of birth, sex, practice registration, deregistration and death dates. Consultation records include the date and type of each healthcare interaction (eg in-practice or telephone). Observation records provide coded clinical information about medical history, symptoms, and lifestyle factors, and are connected to other files such as consultations, referrals, drug issues, and problem groupings. This information is categorised using SNOMED CT which is a comprehensive clinical terminology used to classify health terms.⁴ Problem records group clinical entries judged by GPs to relate to the same underlying issue, such as a chronic condition. Referral records document transfers to and from the practice. Finally, drug issue records include details of prescriptions, including medication name, dosage, duration, and quantity. The connections across these different record types and over time enable longitudinal tracking and analysis of patient care, diagnoses, and treatments across the healthcare system. Free text comment may also be recorded by clinicians, but these are not currently available for research.

To identify specific symptoms, diagnoses and/or treatments, a systematic approach using code lists⁵ is typically used in collaboration with colleagues who have the relevant expertise in primary care and data management. A code list is a list of specific codes used to find and group relevant information from large databases. For example, if a study is looking at people with dementia, the researchers would create a code list of all the codes that mean "dementia" in the database. This helps ensure that all relevant patients are included in their analysis and that the study can be repeated by others using the same definitions. For some conditions/exposures it is necessary to use different approaches/algorithms alongside code lists to identify the individuals of interest. For example, to identify an individual with depression, a SNOMED CT code identified in an observation record and/or a prescription for an antidepressant drug identified in a drug issues record could be used in the definition.

Data Resource Use

As IMRD is a relatively new database, a handful of studies have been published so far. Publications to-date include studies investigating urine sampling rates of patients with suspected lower urinary tract infections,⁶ antibiotic prescriptions for upper respiratory tract infections,⁷ youth mental health outcomes after SARS-CoV-2 infection long-COVID or long-pandemic

syndrome⁸ and antipsychotic prescribing in people with dementia.⁹ Further studies will become available once they are published here: rwsbibliography.iqvia.com. This database is similar to a previous primary care data source eg THIN and Clinical Practice Research Datalink (CPRD). More than 5000 studies have been published using these data sources and include studies investigating treatment patterns, drug safety, adherence to treatment, morbidity and mortality patterns, outcomes research, health economics research, risk management, quantifying unmet needs and resource utilisation. Selected examples include a study to identify ethnic differences in the prevalence of Type 2 diabetes diagnoses,¹⁰ a case-control study investigating prediagnostic presentations of Parkinson's disease in primary care¹¹ and a study on the relative risk of cardiovascular and cancer mortality in people with severe mental illness.¹²

It is advised that studies using IMRD are written according to the principles set out in "Reporting of Studies using Observational Routinely-collected Data" (RECORD) statement.¹³ This checklist outlines best practice for reporting observational studies.

Strengths and Weaknesses

There are several advantages of using IMRD to conduct research. Firstly, the size of the IMRD dataset (6 million patients) allows researchers to include a large cohort of people for their study without the higher costs associated with primary research studies and without the burden of participation which can mitigate potential recruitment and retention issues experienced in primary studies. This large size also allows the opportunity to analyse rare cases and conduct more detailed subgroup analyses. Secondly, owing to the detailed and longitudinal nature of IMRD, researchers can capture information across a patient pathway (often spanning many years), for example, before and after a diagnosis and across different types of data, for example, comparing prescribing medications alongside changes in weight. Thirdly, using data from primary care practices minimises the impact of recall and selection bias compared to other approaches. For example, other research which requires active recruitment, such as using a survey, may be subject to recall and selection bias. In summary, the "real-world", "in-the-moment" experience captured in IMRD captures the care delivered to a broad cohort of people. Lastly, IMRD provides a simple and efficient data-access review system that is faster than many similar datasets, supporting feasibility studies, student projects, and large-scale epidemiological research; decisions are usually made within four weeks of protocol submission.

As with all Electronic Health Record datasets, there are also some weaknesses with using IMRD. Firstly, only people who are registered with a participating practice and are engaging with primary care services will be included a study. Secondly, only information which is documented in someone's record and appropriately codified is known. Different health needs may be discussed during consultations but may not be recorded and/or coded in their health records. This could be particularly apparent for some healthcare pathways which are not financially incentivised, eg not included in the Quality and Outcomes Framework which is a voluntary "pay-for-performance" programme in England that rewards general practices for the quality of care they provide to their patients, based on specific indicators.¹⁴ As such, this may limit comprehensive documentation in these cases. In addition, information is only included in IMRD if it is raised/addressed during a consultation. As some symptoms can be sensitive (such as incontinence) or the patient may not see them as important/relevant to raise or they may not have time to raise all their concerns during a consultation. This means the community prevalence/burden of health needs may not always be fully captured in primary care records. Thirdly, as researchers can only investigate what is recorded, the context to these records cannot be fully explored. For example, a researcher could identify that someone stopped receiving prescriptions for a medication, but they do not necessarily know if this is because their symptoms resolved or if they stopped because of another reason, such as having negative side effects. Lastly, IMRD is subject to data quality issues such as missing or incorrect information. The amount of missing data for sex (0.02%) and deprivation (8.6%) is low (Table 1). Consideration of data quality criteria/missing data and how it applies to each specific research study needs to be discussed and strategies to mitigate the impact should be implemented by researchers. This could be done by utilising common approaches developed with previous datasets, such as multiple imputation for missing data,¹⁵ identifying Acceptable Computer Usage (ACU) among practices¹⁶ and when practices have Acceptable Mortality Reporting (AMR).¹⁷

Data Resource Access

IMRD has NHS Health Research Authority approval for medical research and treatment analysis (NHS Research Ethics Committee ref 23/EM/0151). Data can be supplied to external researchers for scientifically approved studies under Data Sharing Agreements. This access is granted by the IQVIA Scientific Review Committee based on a review of a study protocol outlining the intended use of IMRD. In these protocols, studies are required to show scientific merit, fulfil the research purpose outlined, and demonstrate potential benefit to health and social care. To cover the expenses of providing its data services, IQVIA charges user licence fees to academic, industry, and government researchers. IMRD data access is available under different licensing arrangements for academic and commercial users. Academic institutions typically benefit from discounted rates to encourage research and publication. Pricing is tailored case-by-case based on factors such as cohort size, study scope, data format and duration of access. Further information on obtaining data access is available here: <https://www.iqvia.com/locations/united-kingdom/solutions/life-sciences-industry-solutions/real-world-solutions/iqvia-medical-research-data>.

Acknowledgement

HCS was funded by the National Institute for Health and Care Research (NIHR) School for Primary Care Research (STOP-THEM project, reference 674). JCB-A was funded by the National Institute for Health Research (NIHR) Three Research Schools Mental Health Programme (grant reference number: MHF012) and School for Primary Care Research (grant reference number: 721).

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Disclosure

Ms Louise Pinder and Ms Itohan Evbuomwan are employed by IQVIA. The authors report no other conflicts of interest in this work.

References

1. IQVIA. IQVIA Medical Research Data (IMRD). Available from: <https://www.iqvia.com/locations/united-kingdom/solutions/life-sciences-industry-solutions/real-world-solutions/iqvia-medical-research-data>. Accessed Apr 25, 2025.
2. Statista. *Number of GP Practices in England From December 2016 to December 2024*. 2025
3. Office for National Statistics. *Dataset: Estimates of the Population for England and Wales*. 2024.
4. NHS England. SNOMED CT. Available from: <https://digital.nhs.uk/services/terminology-and-classifications/snomed-ct>. Accessed Apr 25, 2025.
5. Matthewman J, Andresen K, Suffel A, et al. *Checklist and Guidance on Creating Codelists for Routinely Collected Health Data Research [Version 2; Peer Review: 3 Approved]*. 2024:1–26.
6. Ciaccio L, Fountain H, Beech E, et al. Trends in urine sampling rates of general practice patients with suspected lower urinary tract infections in England, 2015–2022: a population-based study. *BMJ Open*. 2024;14(8):e084485. doi:10.1136/bmjopen-2024-084485
7. Yang Z, Bou-Antoun S, Gerver S, Cowling TE, Freeman R. Sustained increases in antibiotic prescriptions per primary care consultation for upper respiratory tract infections in England during the COVID-19 pandemic. *JAC Antimicrob Resist*. 2023;5(1). doi:10.1093/jacamr/dlad012
8. Bilu Y, Flaks-Manov N, Goldshtein I, et al. Youth Mental Health Outcomes up to Two Years After SARS-CoV-2 Infection Long-COVID or Long-Pandemic Syndrome: a Retrospective Cohort Study. *J Adolesc Health*. 2023;73(4):701–706. doi:10.1016/j.jadohealth.2023.05.022
9. Smith HC, Petersen I, Hayes JF, et al. Antipsychotic prescriptions in people with dementia in primary care: a cohort study investigating adherence of dose and duration to UK clinical guidelines. *Lancet Psychiatry*. 2025;12(10):758–767. doi:10.1016/S2215-0366(25)00261-5
10. Pham TM, Carpenter JR, Sharma M, et al. Ethnic Differences in the Prevalence of Type 2 Diabetes Diagnoses in the UK: cross-Sectional Analysis of the Health Improvement Network Primary Care. *CLEP*. 2019;11:1081–1088. doi:10.2147/CLEP.S227621
11. Schrag A, Horsfall L, Walters K, Noyce A, Petersen I. Prediagnostic presentations of Parkinson's disease in primary care: a case-control study. *Lancet Neurol*. 2015;14(1):57–64. doi:10.1016/S1474-4422(14)70287-X
12. Osborn DPJ, Levy G, Nazareth I, Petersen I, Islam A, King MB. Relative Risk of Cardiovascular and Cancer Mortality in People With Severe Mental Illness From the United Kingdom's. *General Practice Research Database*. 2007;64:2.
13. Benchimol EI, Smeeth L, Guttman A, et al. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. *PLoS Med*. 2015;12(10):1–22. doi:10.1371/journal.pmed.1001885
14. NHS England. Quality and Outcomes Framework (QOF). 2024. Available from: <https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/general-practice-data-hub/quality-outcomes-framework-qof>. Accessed Jun 13, 2025.

15. Pedersen AB, Mikkelsen EM, Cronin-Fenton D, et al. Missing data and multiple imputation in clinical epidemiological research. *Clin Epidemiol*. 2017;9:157–166. doi:10.2147/CLEP.S129785
16. Horsfall L, Walters K, Petersen I. Identifying periods of acceptable computer usage in primary care research databases. *Pharmacoepidemiol Drug Saf*. 2013;22(November 2012):64–69. doi:10.1002/pds.3368
17. Blak BT, Thompson M, Dattani H, Bourke A. Generalisability of the Health Improvement Network (THIN) database: demographics, chronic disease prevalence and mortality rates. *Inform Prim Care*. 2011;19(4):251–255. doi:10.14236/jhi.v19i4.820

Clinical Epidemiology

Publish your work in this journal

Clinical Epidemiology is an international, peer-reviewed, open access, online journal focusing on disease and drug epidemiology, identification of risk factors and screening procedures to develop optimal preventative initiatives and programs. Specific topics include: diagnosis, prognosis, treatment, screening, prevention, risk factor modification, systematic reviews, risk & safety of medical interventions, epidemiology & biostatistical methods, and evaluation of guidelines, translational medicine, health policies & economic evaluations. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use.

Submit your manuscript here: <https://www.dovepress.com/clinical-epidemiology-journal>

Dovepress

Taylor & Francis Group