

# Data-Resource Profile: United Kingdom Optimum Patient Care Research Database

Anita Lynam<sup>1</sup>, Charlotte Curtis<sup>1</sup>, Brooklyn Stanley<sup>2,3</sup>, Heath Heatley<sup>3</sup>, Chloe Worthington<sup>2,3</sup>, Emma-Jane Roberts<sup>2,3</sup>, Christopher Price<sup>2,3</sup>, Victoria Carter<sup>2,3</sup>, John Dennis<sup>1</sup>, Andrew McGovern<sup>1</sup>, David Price<sup>2,3</sup>

<sup>1</sup>Momentum Data, Pendragon House, St. Albans, Hertfordshire, UK; <sup>2</sup>Optimum Patient Care, Cambridge, UK; <sup>3</sup>Observational and Pragmatic Research Institute, Singapore

Correspondence: Andrew McGovern, Momentum Data, Pendragon House, St. Albans, Hertfordshire, UK, Email [andy.mcgovern@momentumdata.co.uk](mailto:andy.mcgovern@momentumdata.co.uk)

**Introduction:** Electronic medical records (EMRs) maintained in primary care in the UK and collected and stored in EMR databases offer a world-leading resource for observational clinical research. We aimed to profile one such database: the Optimum Patient Care Research Database (OPCRD).

**Methods and Participants:** The OPCRD, inception in 2010, is a growing primary care EMR database collecting data from 992 general practices within the UK. It covers over 16.6 million patients across all four countries within the UK, and is broadly representative of the UK population in terms of age, sex, ethnicity and socio-economic status. Patients have a mean duration of 11.7 years' follow-up (SD 17.50), with a majority having key summary data from birth to last data entry. Data for the OPCRD are collected incrementally monthly and extracted from all of the major clinical software systems used within the UK and across all four coding systems (Read version 2, Read CTV3, SNOMED DM+D and SNOMED CT codes). Via quality-improvement programmes provided to GP surgeries, the OPCRD also includes patient-reported outcomes from a range of disease-specific validated questionnaires, with over 66,000 patient responses on asthma, COPD, and COVID-19. Further, bespoke data collection is possible by working with GPs to collect new research via patient-reported questionnaires.

**Findings to Date:** The OPCRD has contributed to over 96 peer-reviewed research publications since its inception encompassing a broad range of medical conditions, including COVID-19.

**Conclusion:** The OPCRD represents a unique resource with great potential to support epidemiological research, from retrospective observational studies through to embedded cluster-randomised trials. Advantages of the OPCRD over other EMR databases are its large size, UK-wide geographical coverage, the availability of up-to-date patient data from all major GP software systems, and the unique collection of patient-reported information on respiratory health.

**Keywords:** primary care, electronic health records, medical records, datasets, demography, health outcomes

## Introduction

The National Health Service (NHS) in the UK provides primary care through a network of general practices (GPs). Anyone can register with a GP. Visits are free and also necessary to access specialist care, so services are widely accessed and the vast majority of residents are registered with a GP,<sup>1</sup> and thus primary healthcare records span the full socio-economic spectrum of the UK. These UK electronic primary healthcare records offer a world-leading resource for observational research, a platform to undertake pragmatic clinical trials and other interventional studies, the opportunity to take advantage of variations in the provision of clinical care to maximize the understanding of best practice, and the opportunity to create a "learning healthcare system" that can utilize current information to improve future care.<sup>2-4</sup>

Over the last decade, databases collecting and allowing researchers access to anonymized electronic medical records (EMRs), such as the Optimum Patient Care Research Database (OPCRD; <https://opcrd.co.uk>), Clinical Practice Research Datalink (CPRD; <https://cprd.com>), Oxford Royal College of General Practitioners Research and Surveillance Centre (RCGP-SC; <https://www.rcgp.org.uk/representing-you/research-at-rcgp/research-surveillance-centre>), QResearch (<https://www.qre>

[search.org](https://search.org)), and the Health Improvement Network (THIN; <https://www.the-health-improvement-network.com>), have greatly contributed to research, with 1637 publications citing their use as revealed by PubMed search results. The databases open the door to significant potential for epidemiological intelligence and research using real-world data from routine healthcare delivery, with longitudinal primary care data linked to data from other healthcare settings. OPCRd is a real-world, longitudinal research database of UK electronic primary healthcare records, with coverage of >11.4 million patients since 2010 and growing at an average rate of about 3.2 million patients per year, making it one of the fastest-growing research databases in the UK. It was inceptioned in 2010 by Optimum Patient Care (OPC), a not-for-profit social enterprise founded in 2005. The aims of this study were to provide a cohort profile of OPCRd and an understanding of the underlying resource to both researchers who wish to utilize OPCRd and those researchers reading papers that utilize OPCRd.

## Methods and Participants

### Data-Resource Basics: UK Primary Healthcare System

EMRs have been used in UK general practice since the late 1980s, and by 1996, 96% of general practices' medical records were fully computerized, with most doing complete prescribing from much earlier.<sup>5</sup> The introduction of reimbursement for specific disease clinics in the early 1990s and performance-related funding with the quality and outcome framework (QOF) in 2002 meant all GP surgeries were using computer records for a majority of clinical work and healthcare data recording, and were incentivized to achieve high coding quality in the EMR. In the UK, referrals for secondary care are accessed via a GP. Letters and other outcome documentation from secondary care are routed back to a patient's GP, where diagnoses and treatments are frequently coded into the primary care record, further enhancing capture of health events in the record. Laboratory results are also fully included in primary care records, whether undertaken in primary care or secondary care. Increasingly, biologic prescribing is also becoming well coded in OPCRd, for example, 870 patients with asthma are coded as having a biologic prescribed for asthma.

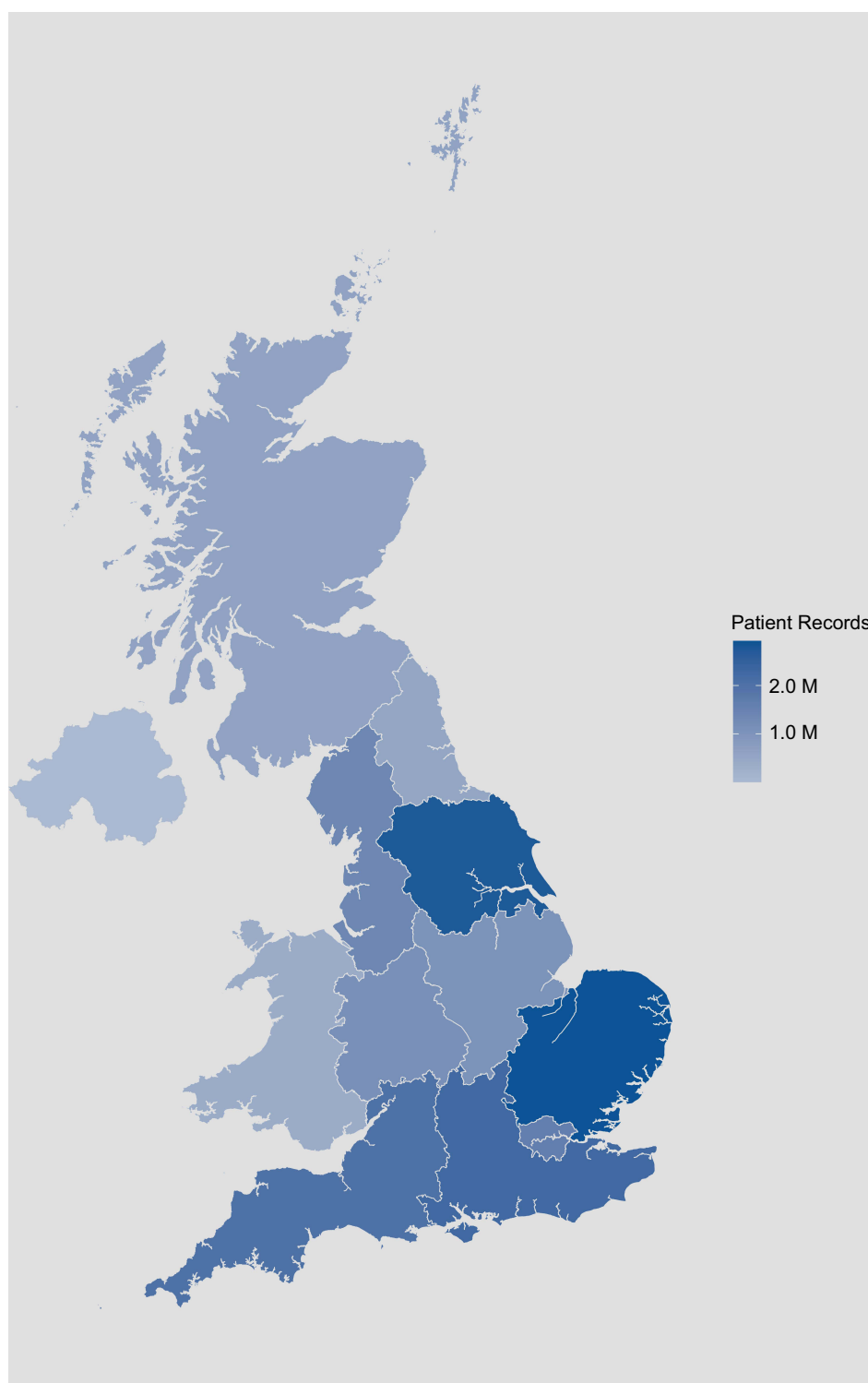
Within the UK, four coding systems have been used to provide EMRs. Read codes were first developed by Dr James Read during the 1980s, with Read Version 1 (4-byte) and Read Version 2 (5-byte) released in the early 1990s. By 1999, the NHS had mandated their use for EMRs. The introduction of Read Version 3 (Clinical Terms Version 3 [CTV3]) expanded the Version 2 code lists to include more areas of clinical practice, but this was never fully implemented due to the introduction of SNOMED CT in 2018.<sup>6</sup> All NHS healthcare providers in England are required to use the SNOMED CT system, with Wales and Scotland also transitioning to SNOMED CT. SNOMED codes can be assigned for most aspects of clinical practice, including findings, symptoms, diagnoses, test results, procedures, treatments, and pharmaceuticals, as well as allowing information to be exchanged across healthcare services, such as details of referrals. SNOMED CT also maps to other health-data systems, including the NHS dictionary of medicines and devices (dm+d), containing unique identifiers and associated textual descriptions for medicines and medical devices used in the diagnosis and treatment of patients in the NHS, and the *International Statistical Classification of Diseases and Related Health Problems* (ICD-11 came into effect in January 2022).<sup>7</sup>

## Results and Findings to Date

### OPCRd: Background, Population Coverage, and Demographic and Clinical Characteristics

The OPCRd was inceptioned in 2010 and is a growing primary care EMR database collecting data from 992 general practices within the UK. It covers >16.6 million patients (approximately 25% of the UK population) across all four countries within the UK (England, Wales, Scotland, and Northern Ireland) (Figure 1), and is broadly representative of the UK population in terms of age, sex, ethnicity, and socio-economic status. Patients have a mean duration of 11.7 years' follow-up (SD 17.50), with a majority having key summary data from birth to last data entry.

Approximately 8.8 million patients were registered at a contributing practice and alive at the time of data extraction (ie, active patients), with over 7.7 million active patients extracted in the last 2 years (Table 1). These are broadly representative of the UK general population in terms of demographic and clinical characteristics, including age (Table 1, Figure 2). Table 2 shows for adults (age ≥18 years) actively registered with OPCRd-contributing practices on



**Figure 1** OPCR patient population by region. The population in the OPCR consists of patients based across the UK, with approximately 25% of the UK population included.

1 July 2022 the prevalence of major QOF-specified chronic conditions (conditions GPs are incentivized to monitor) and the prevalence of major respiratory conditions. In addition to the EMR data, OPCR includes patient-reported (questionnaire) information (PRI) collected as part of the free OPC quality-improvement (QI) and research-support services for GPs in the UK.

**Table 1** Demographic characteristics of patients with valid data and at least 6 months' follow-up in OPCR, 1 July 1 2022

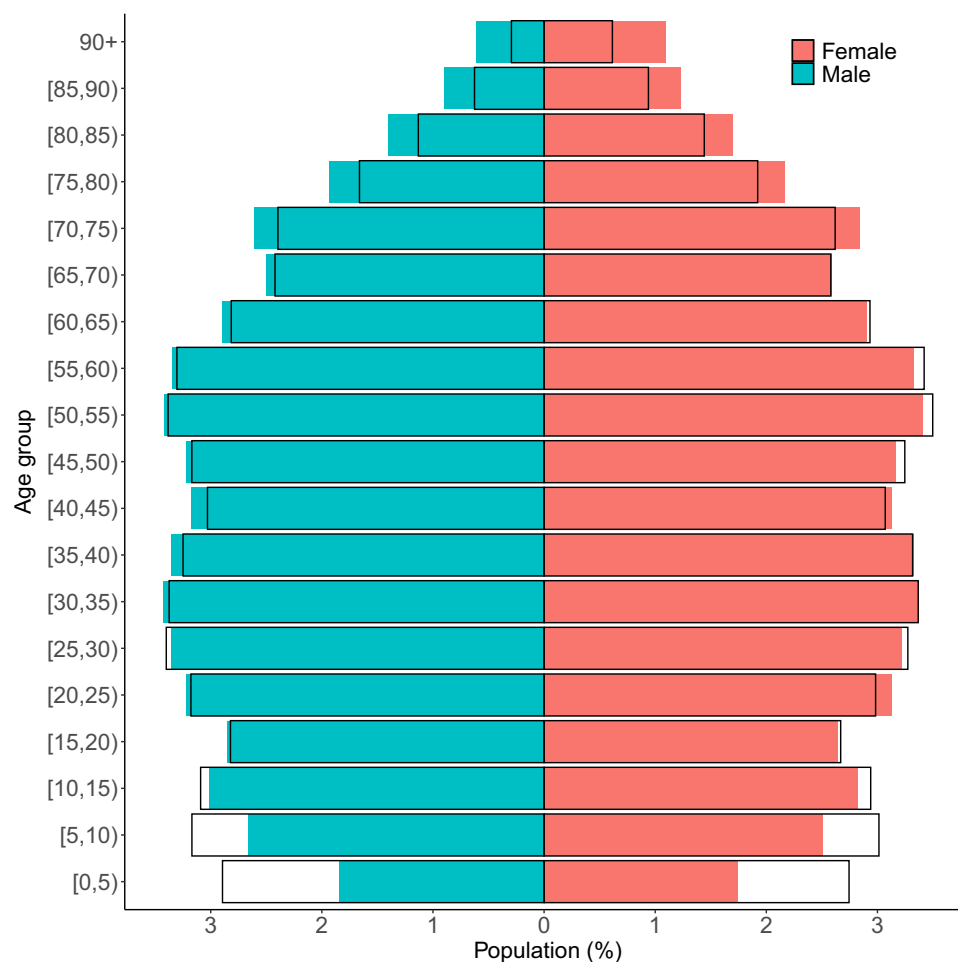
	All patients	Actively registered patients
n	12,008,218	7,766,604
Age (years)	44 (27–64)	43 (25–62)
Follow-up since 1995 (years)	13.1 (9.5–18.1)	7.4 (3.1–14.1)
Sex, n (%)		
Male	5,818,978 (48.5)	3,834,724 (49.4)
Female	6,136,145 (51.1)	3,931,571 (50.6)
Intersex	834 (<0.01)	251 (<0.01)
Ethnicity, n (%)		
White	6,709,297 (55.9)	4,776,423 (61.5)
Black or Black British	241,618 (2.0)	171,119 (2.2)
Asian or Asian British	749,872 (6.2)	518,989 (6.7)
Chinese or other	291,071 (2.4)	197,257 (2.5)
Mixed	159,680 (1.3)	110,899 (1.4)
Not stated <sup>a</sup>	244,391 (2.0)	163,043 (2.1)
Missing	3,612,289 (30.1)	1,828,874 (23.5)
Deprivation quintile, n (%) <sup>b</sup>		
1	1,937,293 (16.1)	1,108,672 (14.3)
2	2,450,954 (20.4)	1,505,419 (19.4)
3	1,579,345 (13.2)	1,031,634 (13.3)
4	1,675,996 (14.0)	1,122,462 (14.5)
5	1,747,048 (14.5)	1,167,357 (15.0)
Missing	2,617,582 (21.8)	1,831,060 (23.6)

**Notes:** <sup>a</sup>Patient chose not to state ethnicity. <sup>b</sup>Deprivation quintile derived using the practice postcode. Data are medians Inter Quartile Range (IQR) unless stated.

## OPCRD Data-Resource Use and Unique Collection of Patient-Reported Information on Respiratory Health

OPCRD has contributed to over 96 peer-reviewed research publications since 2010.<sup>8</sup> A bibliography of publications using OPCR data is available from OPC (<https://optimumpatientcare.org/news>). Many of these studies benefit from additional patient-reported information from >66,000 patients to disease-specific questionnaires including validated questions on respiratory health (Table 3), with specific applications including COVID-19 and heart disease.<sup>9–11</sup> This approach has been used to validate the use of EMR-derived definitions of respiratory conditions in primary care treatment. For example, a study investigating the clinical utility of database-derived asthma-outcome measures in a cohort of patients diagnosed with active asthma found that database-derived severe exacerbations and asthma-control measures were responsive descriptions of a patient's self-reported health status.<sup>12</sup> A study investigating COPD Assessment Test scores also using a combination of EMR and questionnaire data found that these scores were not good predictors of a patient's COPD control status.<sup>13</sup> OPCR EMR data has also been used for epidemiological analysis: a cohort study investigating idiopathic pulmonary fibrosis (IPF) used principal-component analysis to map clinical progression indicators to IPF diagnosis.<sup>14</sup>

More recently, OPCR has also been used to collect key baseline and outcome data for cluster-randomized trials, with interventions randomized at GP level compared to usual care at control practices, using routinely collected primary care EMR data and record linkage for trial outcomes. There are currently three cluster-randomized trials underway within the OPCR. The first, the MAGNIFY COPD study,<sup>15</sup> is a pragmatic trial that aims to evaluate the impact of an interventional package comprising an adherence review, ongoing provision of a dual bronchodilator with an add-on inhaler sensor device, and connected mobile application in patients identified at high risk of exacerbations with historic



**Figure 2** Population pyramid for OPCR and ONS data. Based on July 2022 OPCR data (colored bars) and mid-2020 Office for National Statistics (ONS) data (black bars).

poor treatment adherence (as measured by prescription collection to mono/dual therapy over 1 year,  $n=1312$  patients versus usual care). A total of 176 GPs will be randomized (1:1) to either an adherence-support cluster arm (suitable patients already receiving or initiated Ultibro Breezhaler (indacaterol/glycopyrronium)) will be offered interventional package) or the control cluster arm (suitable patients continue to receive usual clinical care). The intervention will compare time to “treatment failure”, as the first occurrence of one of the following: (1) moderate/severe COPD exacerbation, (2) prescription of triple therapy (inhaled corticosteroid/long-acting  $\beta_2$ -agonist/long-acting muscarinic antagonist [ICS/LABA/LAMA]), (3) prescription of additional chronic therapy for COPD, or (4) respiratory-related death. The trial will also investigate adherence, moderate/severe exacerbations, respiratory-related healthcare resource utilization and costs, and intervention package acceptance rates.<sup>15</sup>

The second is “At-risk registers integrated into primary care to stop asthma crises in the UK” (ARRISA-UK)<sup>16</sup> study is a pragmatic, cluster randomized trial, with nested health economic and process evaluations. This study randomizes 270 GPs from throughout the UK covering over 10,000 registered patients with “at-risk asthma” identified according to a validated algorithm. In the intervention arm practice staff will complete two 45-min eLearning modules and webinar with other practices. On completion of training at-risk patients’ records will be coded so that a flag appears whenever their record is accessed. Practices will receive a phone call at 4 weeks and a reminder video at 6 weeks and 6 months. Control practices will continue to provide usual care. The trial will use primary care records (with linkage to secondary care data) to determine the percentage of at-risk patients with an asthma-related crisis event (accident and emergency attendances, hospitalizations and deaths) after 12 months. It will also capture the time to crisis event, all-cause

**Table 2** Prevalence of major Quality Outcomes Framework (QOF) specified chronic conditions and major respiratory conditions in adults in the active cohort

	Active Adult Population (n=6,189,113)
<b>QOF conditions</b>	
Depression	1,192,997 (19.3)
Hypertension	1,132,542 (18.3)
Hyperlipidaemia	585,616 (9.5)
Type 2 diabetes	430,163 (7.0)
Chronic kidney disease stage 3–5	262,526 (4.2)
Myocardial infarction	110,104 (1.8)
Ischaemic heart disease	325,227 (5.3)
Atrial fibrillation	182,853 (3.0)
Angina	152,387 (2.5)
Revascularisation	70,751 (1.1)
Stroke	146,399 (2.4)
Transient ischaemic attack	70,551 (1.1)
Peripheral vascular disease	50,130 (0.8)
Deep vein thrombosis	57,508 (0.9)
Pulmonary embolism	53,397 (0.9)
<b>Respiratory conditions</b>	
Asthma	1,084,711 (17.5)
Chronic obstructive pulmonary disorder	203,465 (3.3)
Idiopathic pulmonary fibrosis	2878 (<0.1)
Cystic fibrosis	1744 (<0.1)
Interstitial lung disease	87,247 (1.4)
Allergic rhinitis	910,237 (14.7)
$\alpha_1$ Antitrypsin deficiency	2092 (<0.1)

**Notes:** Only individuals with  $\geq 1$  year of follow-up in the primary care record (to ensure major conditions are likely to be accurately recorded). Data are n (%).

**Table 3** Patient-reported information data collected via questionnaire and available for secondary research

Questionnaire title	Condition	Number of patient responses
COVID-19	COVID-19	41,486
Adult asthma	Asthma	19,403
COPD	COPD	9636
MAGNIFY	COPD	3364
Open Clinica	Asthma	756
British Lung Foundation, adult	COPD	730
British Lung Foundation, child	COPD	243
IMP <sup>2</sup> ART	Asthma	202
Admin adult asthma	Asthma	59
Admin child asthma	Asthma	14
Admin COPD	COPD	7

hospitalizations, asthma control and any changes in practice asthma management for at-risk and all patients with asthma, as well as including cost-effectiveness analysis and a mixed-methods process evaluation.

The third is the IMP<sup>2</sup>ART programme — Implementing Improved Asthma Self-Management as Routine (<https://www.ed.ac.uk/usher/imp2art>) — is a National Institute of Health Research (NIHR)-funded intervention to support self-management of asthma in primary care using a combination of patient education, professional training, and organizational support. The trial

components have been developed to provide patient resources, healthcare-provider education via practice facilitators and online learning, and practice support with prioritizing self-management, such as asthma reviews and asthma action plans, as well as a system of patient audit and feedback to help practices monitor their progress. During the trial, 144 GPs will be randomized to routine care or the IMP<sup>2</sup>ART programme for the year-long intervention, with the trial evaluating both the clinical and cost effectiveness of the IMP<sup>2</sup>ART programme.

## OPCRD Structure

### Data Collection

OPCRD is hosted in an SQL Server 2019 environment and receives de-identified data from GPs across the UK, who participate in the OPC QI programmes and use OPC research-support services. OPCRD uniquely integrates with all major UK clinical systems (SystemOne, EMIS, and Vision). It provides a comprehensive picture, including patient demographic information, diagnoses, symptoms, treatments and prescriptions issued, test results and other measurements taken in the practice, as well as referrals. It is a high-quality data source used regularly in clinical, epidemiological, and pharmaceutical research. The dataset has been used for QI projects and research in chronic diseases including, but not limited to, COVID-19, COPD, asthma, and cardiovascular disease.

OPC's focus is on improving the care and management of patients with chronic medical conditions, with an initial interest in respiratory conditions including asthma, chronic obstructive pulmonary disease (COPD) and COVID-19, but recently broadened to cover a number of rare diseases, including Fabry disease and Pompe disease specifically, and moved into other rare diseases per se through an early-indicator algorithm. The OPC QI programme in primary care involves a combined review of de-identified EMRs; patients' responses to disease-specific questionnaires; provision of tailored reports to practices with risk stratification and best-practice support patient management; education workshops, implementation programmes and innovation.<sup>17</sup> The program supports remote patient reviews by completion of key questionnaires that can either replace consultations and QOF reviews or make them easier to deliver by completing pre- (remote) consultation work.

As part of the QI program, OPCRD collects patient-reported responses from validated clinically relevant questions, eg, information about symptom control, triggers, side effects, quality of life, and adherence to medications and self-management guidance. OPCRD currently holds patient-reported data from adult asthma, child asthma, COPD and COVID-19 patient questionnaires from over 66,000 patients, including repeated responses for some patients (Table 3).

### Frequency of EMR Data Collection

Data are extracted incrementally every month from GPs who have agreed to provide data for approved research purposes. Extraction is carried out for all major GP clinical systems in use within the UK, and all three EMR coding systems are supported.

### Ethics

OPCRD has ethics approval from the NHS Health Research Authority Research Ethics Committee (NHS HRA REC ref: 20/EM/0148) to receive and provide anonymized data for research purposes.

## Data Governance and Confidentiality

OPCRD datasets are stored on a secure server. OPCRD operates under strict data security and protection policies in compliance with the General Data Protection Regulation (EU) 2016/679 (GDPR) and the Data Protection Act 2018 (DPA). All data collected are pseudonymized. Data are collected from the contributing GPs using secure extraction software installed on-site at the practices. Only the GP can re-identify their patients for quality-improvement and research-support reports. No direct patient-identifiable or sensitive information is collected or leaves GPs, unless agreed otherwise for the purpose of secondary care data linkage.

Data for patients in England, Wales, and Scotland who have opted out or withdrawn from data sharing via their GP (Type 1 Opt-out) or via the National Data Opt-out scheme in England are not extracted by OPCRD for research or for

data-linkage purposes. Certain sensitive EMR data are also excluded from data extraction, eg, sexually transmitted diseases, gender reassignment, termination of pregnancy, IVF treatment, and physical or psychological abuse.

## Data Linkage

Data in OPCRd can be linked to secondary care datasets and national disease registries, eg, the International Severe Asthma Registry (ISAR) and International Helping Asthma in Real-Life Patients (iHARP). Data linkage with NHS Digital Secondary Care Data and Office for National Statistics death registration data are also possible. OPCRd is approved by the NHS Confidentiality Advisory Group (CAG). During October 2021, the responsibility for the management of the National Disease Registration Service (NDRS), a collection of data on all cancers, rare diseases and congenital anomalies diagnosed each year in England, transferred from Public Health England (PHE) to NHS Digital, and the Data Access Request Service (DARS) can offer researchers access to this data to help improve NHS services. Although mapping to wearable technology data is not currently available in OPCRd, mapping to spirometry diagnostic testing data in near real time has been successfully performed.<sup>18</sup> Mapping against mortality records does not currently exist, but funding to implement this mapping is being sought.

The whole of OPCRd is mapped to the Observational Medical Outcomes Partnership (OMOP) common data model (CDM). This mapping transforms the patient-level data held in OPCRd into a common format that facilitates the generation of real-world evidence. OPCRd is part of the European Health Data Evidence Network (EHDEN) using OMOP CDM to support research (<https://www.ehden.eu/datapartners>).

## Discussion

### Future Opportunities

There is great potential for OPCRd data to be used more widely for epidemiological analysis. OPCRd will support most retrospective observational study designs, including for rare disorders, therapies, and their outcomes. This can be supported by the option of working with GPs to collect new research via patient-reported questionnaire data or by embedding clinical trial models. Other potential uses of OPCRd include feasibility counts for conditions, such as COVID-19, as well as performing post-approval studies and targeted recruitment for interventional trials.

### Strengths and Limitations

OPCRd is a large, broadly representative database with well-characterized patients and detailed up-to-date records (within last 3 months) on primary care information. A key strength of OPCRd over other large UK enhanced healthcare databases is collecting EMR from GPs from England, Wales, and Scotland within the UK, as well as covering patients from all age groups and across the socio-economic spectrum. OPCRd is not limited to data collected from a single specific GP clinical system or coding system: the extraction from all clinical systems across all UK nations means there is less bias of demographics and patient profiles/pathways within OPCRd compared to other healthcare databases. Data are collected monthly and are incremental, so researchers can plan interventions and monitor effectively.

Another strength is the mapping from OPCRd to the OMOP CDM: the advantages of converting patient-level data to an OMOP CDM specification is that only a single definition of a study query is needed to enable execution across multiple and diverse health providers. This makes research very efficient, transparent, reproducible, and scalable. In addition, and uniquely, OPCRd also includes self-reported patient outcomes collected through bespoke patient questionnaires. For example, OPCRd holds unique and enhanced data in respiratory conditions using validated questionnaire responses for over 66,000 patients and has the possibility through the QI service provided to sites to enhance the EMR data with any further info required for a particular study.

Limitations associated with OPCRd are in common with other EMR databases: these include missing data on demographic, key diagnoses, and other clinical information and variation in coding, both between GPs and for different GP clinical systems. At present, free text information is not available to researchers due to limitations on extractions and for data-governance reasons. Information on secondary care may not be comprehensively recorded within the primary care EMR. Primary care data do not usually capture “over the counter” non-prescription medication use, although

information can be gathered on non-prescription medication use through patient-reported information as part of the OPC QI programme.

## Conclusion

OPCRD is a growing UK primary care data resource currently containing EMR records from 992 GPs and some 16.6 million patients, covering approximately 25% of the UK population. These data have been used in >96 peer-reviewed research publications since OPCRD inception in 2010, encompassing a broad range of health conditions. The unique collection of self-reported patient outcomes through bespoke patient questionnaires makes it a valuable EMR data resource for researchers, with great potential to support epidemiological research from retrospective observational studies through to embedded cluster randomized trials.

## Data-Resource Access

OPCRD is governed by the Anonymized Data Ethics and Protocol Transparency (ADEPT) Committee, an independent, non-statutory expert advisory committee, commissioned by the Respiratory Effectiveness Group (REG, <https://www.regresearchnetwork.org/>) to govern the standard of research conducted on real-world research databases. Access to the OPCRD is subject to proposals being reviewed and approved by ADEPT (<https://www.regresearchnetwork.org/adept-committee/>). Protocols are reviewed on the basis of scientific quality, public benefit, ethical considerations, and any risks posed to data subjects. The ADEPT committee may recommend that study-specific research ethics committee (REC) approval be sought if complex ethical issues arise for a particular study. REC approval will be required for any study that includes direct patient involvement. All proposed research must have intent to publish and must have a scientific, medical, or public health basis. Once ADEPT has approved the study, researchers will be required to sign a data-sharing agreement prior to being given access to anonymized datasets from OPCRD. Studies requiring use of OPCRD data will also be required to be registered on recognized study databases, such as the European Network of Centers for Pharmacoepidemiology and Pharmacovigilance (ENCePP; <https://www.encepp.eu>) and ClinicalTrials.gov. The de-identified data required for research are fully anonymized before being provided to the researcher. Additional security and data-protection protocols are put in place to anonymize the data when a client receives unlimited access to their own client database derived from OPCRD. This includes applying additional redaction algorithms on free-text dosing instructions and removing any potential location indicators like practice identifiers and postcodes.

## Data Sharing

See Data Resource Access section of manuscript.

## Ethics Approval

The OPCRD received ethics approval from the NHS Health Research Authority HRA Research Ethics Committee (NHS HRA REC ref: 20/EM/0148) to receive and provide anonymized data for research purposes. This study was approved by ADEPT (ADEPT0121) on 15 February 2021.

## Acknowledgments

This study is based wholly on data from the Optimum Patient Care Research Database ([opcrd.co.uk](http://opcrd.co.uk)) obtained under license from Optimum Patient Care Limited, and its execution was approved by recognized experts affiliated to the Respiratory Effectiveness Group. However, the interpretations and conclusions contained in this report are those of the authors alone.

## Author Contributions

All authors made a significant contribution to the work reported, whether in the conception, study design, execution, acquisition of data, analysis and interpretation, or all these areas, took part in drafting, revising, or critically reviewing the article, gave final approval to 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.

## Funding

Unfunded project management, medical writing, and statistical support were provided by Momentum Data, UK, a specialist real-world evidence company.

## Disclosure

A. Lynam, C. Curtis, J. Dennis and A. McGovern are employees of Momentum Data who were unpaid consultants to OPCRd in connection with the development of this manuscript. A. McGovern reports grants from Pfizer, Eli Lilly, AstraZeneca; personal fees from Boehringer Ingelheim, outside the submitted work. E.J Roberts, V. Carter, C. Price, C. Worthington are employees of Optimum Patient Care, UK. H. Heatley and B. Stanley are employees of Observational and Pragmatic Research Institute, Singapore. D. Price has advisory board membership with AstraZeneca, Boehringer Ingelheim, Chiesi, Mylan, Novartis, Regeneron Pharmaceuticals, Sanofi Genzyme, Thermofisher; consultancy agreements with Airway Vista Secretariat, AstraZeneca, Boehringer Ingelheim, Chiesi, EPG Communication Holdings Ltd, FIECON Ltd, Fieldwork International, GlaxoSmithKline, Mylan, Mundipharma, Novartis, OM Pharma SA, PeerVoice, Phadia AB, Spirosure Inc, Strategic North Limited, Synapse Research Management Partners S.L., Talos Health Solutions, Theravance and WebMD Global LLC; grants and unrestricted funding for investigator-initiated studies (conducted through Observational and Pragmatic Research Institute Pte Ltd) from AstraZeneca, Boehringer Ingelheim, Chiesi, Mylan, Novartis, Regeneron Pharmaceuticals, Respiratory Effectiveness Group, Sanofi Genzyme, Theravance and UK National Health Service; payment for lectures/speaking engagements from AstraZeneca, Boehringer Ingelheim, Chiesi, Cipla, GlaxoSmithKline, Kyorin, Merck, Mylan, Mundipharma, Novartis, Regeneron Pharmaceuticals and Sanofi Genzyme; payment for travel/accommodation/meeting expenses from AstraZeneca, Boehringer Ingelheim, Mundipharma, Mylan, Novartis, Thermofisher; stock/stock options from AKL Research and Development Ltd which produces phytopharmaceuticals; owns 74% of the social enterprise Optimum Patient Care Ltd (Australia and UK) and 92.61% of Observational and Pragmatic Research Institute Pte Ltd (Singapore); 5% shareholding in Timestamp which develops adherence monitoring technology; is peer reviewer for grant committees of the UK Efficacy and Mechanism Evaluation programme, and Health Technology Assessment; and was an expert witness for GlaxoSmithKline. The authors report no other conflicts of interest in this work.

## References

1. Baker C. Population estimates & GP registers: why the difference? House of commons library; 2016. Available from: <https://commonslibrary.parliament.uk/population-estimates-gp-registers-why-the-difference/>. Accessed December 1, 2021.
2. Foley T, Fairmichael F. The potential of learning healthcare systems. the learning healthcare project; 2015. Available from: [https://learninghealthcareproject.org/wp-content/uploads/2015/11/LHS\\_Report\\_2015.pdf](https://learninghealthcareproject.org/wp-content/uploads/2015/11/LHS_Report_2015.pdf). Accessed December 1, 2021.
3. Friedman CP, Wong AK, Blumenthal D. Achieving a nationwide learning health system. *Sci Transl Med*. 2010;2:57cm29. doi:10.1126/scitranslmed.3001456
4. Bradley SH, Lawrence NR, Carder P. Using primary care data for health research in England - an overview. *Future Healthc J*. 2018;5:207–212. doi:10.7861/futurehosp.5-3-207
5. Benson T. Why general practitioners use computers and hospital doctors do not--Part 1: incentives. *BMJ*. 2002;325:1086–1089. doi:10.1136/bmj.325.7372.1086
6. NHS England. SNOMED CT; 2023. Available from: <https://www.england.nhs.uk/digitaltechnology/digital-primary-care/snomed-ct/>. Accessed April 11, 2023.
7. World Health Organization. International classification of diseases (ICD-11); 2023. Available from: <https://www.who.int/standards/classifications/classification-of-diseases>. Accessed April 11, 2023.
8. OPCRd research publications. Available from: <https://opcrd.co.uk/publications/>. Accessed November 12, 2020.
9. Jones R, Davis A, Stanley B, et al. Risk predictors and symptom features of long COVID within a broad primary care patient population including both tested and untested patients. *Pragmat Obs Res*. 2021;12:93–104. doi:10.2147/POR.S316186
10. Kostikas K, Rhee CK, Hurst JR, et al. Adequacy of therapy for people with both COPD and heart failure in the UK: historical cohort study. *Pragmat Obs Res*. 2020;11:55–66. doi:10.2147/POR.S250451
11. Davey P, Kirby MG. Cardiovascular risk profiles of GnRH agonists and antagonists: real-world analysis from UK general practice. *World J Urol*. 2021;39:307–315. doi:10.1007/s00345-020-03433-3
12. Colice G, Chisholm A, Dima AL, et al. Performance of database-derived severe exacerbations and asthma control measures in asthma: responsiveness and predictive utility in a UK primary care database with linked questionnaire data. *Pragmat Obs Res*. 2018;9:29–42. doi:10.2147/POR.S151615
13. Nibber A, Chisholm A, Soler-Cataluña JJ, Alcazar B, Price D, Miravittles M. Validating the concept of COPD control: a real-world cohort study from the United Kingdom. *Copd*. 2017;14:504–512. doi:10.1080/15412555.2017.1350154

14. Thickett D, Voorham J, Ryan R, et al. Historical database cohort study addressing the clinical patterns prior to idiopathic pulmonary fibrosis (IPF) diagnosis in UK primary care. *BMJ Open*. 2020;10:e034428. doi:10.1136/bmjopen-2019-034428
15. Price D, Jones R, Pfister P, et al. Maximizing adherence and gaining new information for your chronic obstructive pulmonary disease (MAGNIFY COPD): study protocol for the pragmatic, cluster randomized trial evaluating the impact of dual bronchodilator with add-on sensor and electronic monitoring on clinical outcomes. *Pragmat Obs Res*. 2021;12:25–35. doi:10.2147/POR.S302809
16. Smith JR, Musgrave S, Payerne E, et al. At-risk registers integrated into primary care to stop asthma crises in the UK (ARRISA-UK): study protocol for a pragmatic, cluster randomised trial with nested health economic and process evaluations. *Trials*. 2018;19:466. doi:10.1186/s13063-018-2816-z
17. Stanley B, Davis A, Jones R, et al. Characteristics of patients in platform C19, a COVID-19 research database combining primary care electronic health record and patient reported information. *PLoS One*. 2021;16:e0258689. doi:10.1371/journal.pone.0258689
18. Alves L, Pullen R, Hurst JR, et al. CONQUEST: a quality improvement program for defining and optimizing standards of care for modifiable high-risk COPD patients. *Patient Relat Outcome Meas*. 2022;13:53–68. doi:10.2147/PROM.S296506

## Pragmatic and Observational Research

Dovepress

### Publish your work in this journal

Pragmatic and Observational Research is an international, peer-reviewed, open access journal that publishes data from studies designed to reflect more closely medical interventions in real-world clinical practice compared with classical randomized controlled trials (RCTs). The manuscript management system is completely online and includes a very quick and fair peer-review system. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <http://www.dovepress.com/pragmatic-and-observational-research-journal>