

## Using general practice clinical information system data for research: the case in Australia

Youens, D<sup>1\*</sup>, Moorin, R<sup>1,2</sup>, Harrison, A<sup>1</sup>, Varhol, R<sup>1</sup>, Robinson, S<sup>1</sup>, Brooks, C<sup>3</sup>, and Boyd, J<sup>4,5</sup>

### Submission History

|            |            |
|------------|------------|
| Submitted: | 04/02/2019 |
| Accepted:  | 26/09/2019 |
| Published: | 30/01/2020 |

<sup>1</sup>School of Public Health, Curtin University, Perth, Australia

<sup>2</sup>School of Population and Global Health, University of Western Australia

<sup>3</sup>Swansea University Medical School, Singleton Park, Swansea, UK

<sup>4</sup>Health Systems & Economics, School of Public Health, Curtin University

<sup>5</sup>Department of Public Health, School of Psychology and Public Health, College of Science, Health and Engineering, La Trobe University

### Abstract

General practice is often a patient's first point of contact with the health system and the gateway to specialist services. In Australia, different aspects of the health system are managed by the Commonwealth Government and individual state / territory governments. Although there is a long history of research using administrative data in Australia, this split in the management and funding of services has hindered whole-system research. Additionally, the administrative data typically available for research are often collected for reimbursement purposes and lack clinical information.

General practices collect a range of patient information including diagnoses, medications prescribed, results of pathology tests ordered and so on. Practices are increasingly using clinical information systems and data extraction tools to make use of this information. This paper describes approaches used on several research projects to access clinical, as opposed to administrative, general practice data which to date has seen little use as a resource for research.

This information was accessed in three ways. The first was by working directly with practices to access clinical and management data to support research. The second involved accessing general practice data through collaboration with Primary Health Networks, recently established in Australia to increase the efficiency and effectiveness of health services for patients. The third was via NPS MedicineWise's MedicineInsight program, which collects data from consenting practices across Australia and makes these data available to researchers.

We describe each approach including data access requirements and the advantages and challenges of each method. All approaches provide the opportunity to better understand data previously unavailable for research in Australia. The challenge of linking general practice data to other sources, currently being explored for general practice data, is discussed.

Finally, we describe some general practice data collections used for research internationally and how these compare to collections available in Australia.

### Keywords

General practice; administrative data; big data, health information systems, medical records systems

## Introduction

Health systems produce large volumes of data and these data are increasingly created and stored in digital formats [1]. Such data can be an invaluable tool both for patient care and research. In Australia the administrative data generated in hospitals has supported research for decades [2], however general practice data has to date seen limited use as a research resource [3].

This paper adds to a limited body of evidence regarding the utility and availability of clinical general practice data for research in Australia. This is based on our experience of working with different sources of general practice data, which we have accessed through different approaches. We describe these approaches in terms of access, usage, challenges and linkage to other collections. We describe the approaches our research

group is familiar with in detail, and provide an overview of other Australian data collections and leading examples internationally.

## Background

### The Australian Health System

General practices are entrenched in communities and are usually the initial interaction people have with the health system. While general practice provides health promotion, prevention, treatment and support; it is also the gateway to specialist services that support a growing population with chronic and complex comorbidities [4].

In Australia primary care and specialist services are pro-

\*Corresponding Author:

Email Address: [david.youens@curtin.edu.au](mailto:david.youens@curtin.edu.au) (D Youens)

vided largely by private providers who are reimbursed by the Commonwealth (Federal) Government on a fee-for-service basis [5]. Australia's universal public insurance system, Medicare, ensures that all citizens and permanent residents have free access to public hospitals and reimburses general practitioners, specialists and other providers for services rendered (though providers may charge additional out-of-pocket costs [6]). It also provides reduced cost access to pharmaceuticals via the Pharmaceutical Benefits Scheme (PBS) [6]. In contrast, public hospitals in Australia are managed and operated by state and territory governments and are funded jointly by these state / territory governments and the Commonwealth Government. A significant recent development is the establishment in 2015 of 31 Primary Health Networks (PHNs) covering the country. These are not-for-profit companies independent of government, established to organise health services in their region. The main roles of the PHNs are to commission services to address gaps and meet prioritised local needs; to work closely with general practitioners and other health professionals to build capacity and the delivery of high quality care; and to integrate services at the local level to improve the patient experience and eliminate duplication [7]. Their key objectives are to improve the efficiency and effectiveness of medical services for patients, particularly those at risk of poor health outcomes, and to improve the coordination of care. The PHNs work with public and private services (including general practice) in their region but do not directly provide services [8].

## Administrative data supporting research in Australia

Administrative data regarding ambulatory care services (including general practice) and medication dispensations are routinely collected in Australia. The Medicare Benefits Schedule (MBS) captures data on all services rendered attracting reimbursement through Medicare [9] and hence includes almost all general practitioner contacts, many diagnostic tests, therapeutic procedures and specialist visits. PBS data include records of all medicines dispensed under the scheme [9]. These data are collected for reimbursement meaning that while they have very good coverage and quality, they often lack detail important for research. For example the PBS records medications dispensed with no information on prescribing, meaning important information needs to be estimated (for example dosage prescribed or the reason for prescription) and there is no information on unfilled prescriptions [10] (relevant if trying to understand, for example, primary non-compliance). The MBS records item numbers indicating reimbursable activities such as general practice consultations, procedures, the completion of pathology tests and so on [9, 11] without any information on context. For example, a record indicating a general practice consultation will not include information on the reason for the visit, the advice offered, or any condition(s) diagnosed. Similarly, a record may indicate the completion of a pathology test but will not hold any information on the results of that test, and in many cases may indicate the completion of any one of a range of tests, which attract the same reimbursement but are otherwise unrelated. Furthermore, where multiple pathology tests are ordered, only the three attracting the highest reimbursements are captured meaning a test which is relevant to a research question may not appear in MBS data. While these

administrative data have certain strengths in their suitability for research (including completeness, whole population coverage, oversight by a single custodian) the limitations described, in particular the lack of clinical and patient information, make them unsuitable for many research questions.

## General practice data

Beyond this administrative data there is a substantial gap in the general practice data available for research in Australia [12]. Previously survey data has provided insight into general practice activity [13] although these surveys were not intended to facilitate whole-sector research linking with data from other sources or longitudinal follow-up of patients.

More suitable clinical data for many research questions is collected through the clinical information system (CIS) software increasingly used by general practitioners since their introduction in the 1990's [14] to manage their patients [3]. This software initially focussed on collecting administrative and business related information, such as billing and scheduling, to assist with daily practice processes. These systems have expanded considerably to also collect clinical information including referrals, prescribed medications, development of patient management plans, pathology tests and other data. As the use of clinical information systems have increased over the past decade, tools to extract the data captured with these systems have been developed. These aid providers through clinical audits of patient cohorts and can highlight opportunities for improvement in business processes and patient care. These tools can extract de-identified patient information from clinical software and aggregate these data into measures related to patient visits, demographics, diagnoses, immunisations, pathology and so on. Many of these provide useful dashboard approaches and have been shown to enable quality improvement within general practice [15]. The development and use of these tools facilitates the extraction of clinical general practice data for research purposes, the focus of this paper.

## Development of data linkage in Australia

Data linkage capabilities within Australia are shaped by the division of health services [16]. Individual states have developed their own linkage centres over several decades and many of these have supported large numbers of research projects [17, 18]. These state-based systems link hospitalisation data with emergency records, disease registries, large surveys, community services and so on. These state-level data can be linked to MBS and PBS (Commonwealth) data on a project by project basis, though as such linkages require the release of linkable fields by one jurisdiction to another the approvals process becomes more complex. The Commonwealth Government established the Population Health Research Network (PHRN) in 2008 to provide Australian researchers with access to linkable de-identified data from a diverse and rich range of health datasets, across jurisdictions and sectors. Meanwhile the Australian Institute of Health and Welfare has been designated an Integrating Authority, meaning that it has the authority to undertake linkage including Australian (Commonwealth) Government data [19]. Through the PHRN Australia now has

a dedicated capability for linkage of administrative and research data across all states and territories [20], though the approvals process for such linkages remains very challenging for researchers [21].

In summary, despite a strong tradition of linked data research in Australia using administrative (predominantly billing) data there are currently major limitations including the lack of information on the clinical management of patients in general practice, and obstacles to research using data from multiple sectors, making research capturing the patients journey through the health system challenging.

## Current approaches to accessing general practice data

Our research group currently accesses general practice data for research through three methods. The first involves working directly with individual practices. The second is through working with general practices in partnership with one of our researcher teams working with the Primary Health Networks (PHNs) in Western Australia. The third is via a centralised collection of general practice clinical information system data collected, maintained by and accessed via the MedicineInsight program run by NPS MedicineWise [22], which is an authority on the quality use of medicines in Australia. We describe the practicalities, benefits and challenges of each of these approaches.

### Working with general practices, directly and via Primary Health Networks

Access to general practice data for research relies in part on the CIS and associated data extraction systems, so a brief overview of these is warranted. As the clinical data collected in general practice has increased in scope, software vendors have developed their clinical information systems largely independently, without standardisation, resulting in a disparate array of methods to store and report information. For example, Best Practice and Medical Director, the two most common general practice management tools in Australia, use the incompatible medical terminology and health coding systems Pyefinch [23] / DOCLE (Doctor Command Language) [24] and the Medical Director termset, a derivative of SNOMED CT (Systematized Nomenclature of Medicine – Clinical Terms) [25]. Other applications such as PractiX, ZedMed and Genie use ICPC-2+ (International Classification of Primary Care) [23].

There is also diversity in the data extraction tools in use in Australia. The most popular is CAT4 (Clinical Audit Tool 4) developed by PenCS [26], which is provided to general practice through subsidised means by most (28 of 31) PHNs. Some PHNs provide POLAR GP (Population Level Analysis & Reporting) [27] as an alternative clinical auditing tool. Individual practices can select alternatives, with some choosing to use the Canning Tool [28] or to participate in the MedicineInsight program, which uses the GRHANITE (GeneRic HeAlth Network Information Technology for the Enterprise) extraction tool [29, 30] and a customised version of cdmNet.

### Access to general practice data

Firstly, we describe experiences in accessing general practice data through working directly with practices.

There are some general steps that researchers can follow if wishing to access data from general practices. Ideally, the project should already have ethical approval from the researcher's institution. Practices should be approached individually and invited to participate in the project. A brief information sheet is important, and practices may be invited to a workshop or information session where the researcher explains their project. For specific low-risk projects that have received ethics approval which require the participation of a cohort of patients (i.e. extraction of de-identified data), the Practitioner and the Practice Principal may consent on behalf of the patient depending on the practice's registration form for new patients. Building trust with general practitioners is crucial for long-term data sharing. Regular sharing of data can be fostered where organisations or researchers are able to provide value to practices, through providing data summaries to practices either as in-person data summary reviews or via data dashboards or other visualisations. These approaches can help practices to increase data literacy and identify areas to target quality improvement activities.

### Involvement of Software Vendors

Some recent developments in patient privacy [31] have prompted clinical software vendors to require oversight of how data captured via their software is used. A data application including the project description and evidence of ethical approval will usually be reviewed by the vendor's committee, and the data request will usually incur a cost. Data extraction can occur in multiple ways. The simplest form of extraction is through predefined templates specific to each clinical software platform, though a unique extraction protocol is required for each platform as there is no agreed standard for extracting data. These templates are limited to what the vendor provides with requests for specific reports often associated with a financial outlay. Where these predefined templates are unsuitable, a customised extraction from the CIS database may be possible, though this requires SQL (Structured Query Language) expertise and governance approval from both the General Practice Principal (owner) and the software vendor. Software vendors are increasingly involved with data extraction requests by restricting data access to only general practice staff. It is becoming commonplace for researchers to work with CIS vendors to obtain data, as vendors want to ensure that data are being used in accordance with ethical approvals and to minimise risks of data breaches [31]. Accessing data this way therefore requires the research team to have sufficient technical skills to extract and store comparable data from across practices, substantial relationship and trust building with practices and administrative work to navigate approvals. These challenges will of course increase the more practices the research involves.

### Involvement of Primary Health Networks

Where researchers need general practice data for their project the PHNs could (though do not necessarily need to) be approached. By involving the PHNs the researchers may get additional support if the project aligns with the PHNs health

initiatives in the region. With the establishment of the PHNs there has been a renewed awareness of the importance and capability of data collected in general practice to improve both patient outcomes and operational efficiencies (i.e. workflows and practices). Prior to this, data extraction across the general practice continuum was sporadic and uncoordinated, making consistent quality improvement a challenge. Routine data extraction is now becoming commonplace in most general practices, with data extracts scheduled at a frequency and time identified by the practice and coordinated by the PHN and the extraction software vendor.

To ensure only approved individuals have access to the data, comprehensive terms of agreement and governance frameworks are defined and applied at both the practice and PHN. These facilitate the reciprocal exchange of data from general practice to PHNs in return for practice specific reports detailing comprehensive insights into practice performance that often include regional comparators. As secondary data custodians, the PHNs may, depending on the agreed governance, further use the aggregated information to help inform population health service planning and policy. Decisions relating to data access for research will depend on the nature of the research project. For example, an ecological analysis of a chronic disease, utilising data aggregated across patients before release from the practice, may be viewed differently from a project requiring longitudinal follow-up of patients and hence individual patient data. For data that are de-identified at the general practice, aggregated and shared with PHNs for continuous quality improvement, the Practice Principal will decide whether information is shared with the PHN.

### Strengths of general practice data

Despite the immediate shortcomings of data collected in general practice, these can provide a diverse range of information to inform policy and practice which are not available elsewhere. The information collected in general practice, ranging from diagnostic, therapeutic, prescription, disease control data and so on, has the potential to complement the data generated in hospitals and elsewhere (e.g. allied health, aged care etc.) to provide a comprehensive view of the patient journey. When combined, these sources offer tremendous opportunities in improving, for example, prevalence and incidence estimates, health services evaluations, outcomes research and economic analyses.

Working with general practices via the means described above involves substantial work in building relationships to facilitate trust and the sharing of data (whether by researchers themselves or by the PHNs). Though this can involve a significant time investment, the building of these relationships can have the added advantage of facilitating the dissemination of research findings back into practices and hence informing the clinical care of patients.

### Challenges in accessing data via general practices

The lack of standardisation in patient information management tools and data extraction tools used across general practice results in substantial variability and inconsistency in the data captured. The variability in the information captured by general practitioners, and between patient information systems

/ extraction tools in the type and coding of information, may not be an issue for individual practices using data for business improvement purposes but can be problematic for researchers aiming to use information aggregated across practices or make comparison between practices.

In addition to the technical aspects mentioned above, there is apprehension from health providers regarding the sharing of data with third parties, including de-identified data. This includes concerns about maintaining patient and provider privacy; a lack of confidence relating to information accuracy and completeness in the data collected; and, in our experience, an uneasiness in providing regulatory bodies with such data. This uneasiness follows examples such as the controversial Quality and Outcomes Framework (QOF) implemented in the United Kingdom (UK) [32] and Care.data which was abandoned over privacy concerns [33].

In the Australian context, the ability of the PHNs to provide data extraction software to practices and to provide snapshots of practice data in comparison to local or regional averages may encourage practices to share their data for quality improvement purposes. However, for individual researchers or research groups working in, for example, the university sector, challenges around data sharing may be more difficult to overcome. One of the advantages of partnering with the PHNs, where this is possible, is that many of the challenges outlined above may have been fully or partially resolved through the prior work of the PHNs. For example, most PHNs that have existing relationships with the practices in their region are providing snapshot reports back to practices, which may encourage data sharing and may have existing governance processes in place.

### External linkages

With the exception of a few exploratory endeavours [34, 35] patient data are not typically linked between practices [35] or to other areas of the health system (e.g. tertiary care) [34] and, as such, the data collected by general practices in Australia cannot currently be used to fully understand patients' interactions with the health system. International examples have demonstrated the power of using data from across the continuum of care in understanding, for example, associations between body-mass index and cancer [36], and risks of myocardial infarction and stroke following acute infection and vaccinations [37].

Although data linkage is not new, it is new to the field of general practice in Australia. The challenges described above in relation to data sharing also apply to data linkage, though given the reliance of data linkage on patient identifiers these issues can be more challenging to overcome. For example, a general practitioner or practice principal who is willing to provide access to de-identified data to support a research project may become much more hesitant where patient names and addresses are required to allow for linkage. Linkage may be facilitated through privacy preserving record linkage, whereby patient identifiers are irreversibly encoded prior to extraction from the practice, and linkage then performed on these encoded data [38]. Even with a technological solution such as this the steps of gaining the confidence and trust of practices remains vital, along with the necessity for a research team to have the technical skills to apply such methods. Of course,



if the data are to be linked to hospitalisation or other data sources, separate applications will also need to be made to the custodians of those data collections which adds further complexity.

Obstructions to data sharing can be the result of legal or legislative barriers but are more often related to understanding the options around data sharing by data custodians. These barriers can often be dismantled by mitigating risks associated with the data sharing process through careful planning, secure protocols and legal agreements. Given the sensitivity of the information involved, and the growing desire to link a broader number of datasets, any risk mitigation models that facilitate broader sharing of data are valued by researchers, data owners and the public.

In cases where linkage to other data sources is desired, the patient is generally required to provide consent for linkage; although the National Health and Medical Research Council provide criteria under which an ethics committee may approve the use of patient data under a waiver of consent, considering the risk to patients and benefits of the research, practicality of obtaining consent for the given project, privacy protections and more [39]. Data linkage can be performed in Australia by linkage facilities under state [40] or Commonwealth Governments [19] or at university-based centres [41]; for a given project the party to perform the linkage will depend on the datasets being linked and the jurisdictions responsible for these data collections.

## NPS MedicineWise MedicinesInsight Data

NPS MedicineWise was established twenty years ago in 1998 with the aim of promoting the quality use of medicines. The MedicinesInsight program was established in 2011 as a quality improvement program to allow consenting general practitioners to assess their patterns of prescribing and patient care and to allow benchmarking at multiple levels [42]. This program involves practices signing up to share clinical data from their clinical information systems to MedicinesInsight monthly, which allows MedicinesInsight to provide insights into aggregated clinical data and provides practices the means to review their own activities. The extracted data includes information on patient demographics, reasons for encounters, conditions, prescriptions, observations, immunisation history and pathology tests, including results where available. Progress notes are not available. NPS MedicineWise has made MedicinesInsight data available to external researchers to support primary health care research following ethical approval. As of July 2017 the program had recruited over 650 practices, which included information from over 3,300 general practitioners and 3.6 million regular patients [43].

### Access

Data are accessed by application to NPS MedicineWise. Liaison staff including biostatisticians and epidemiologists work with researchers to discuss data requests. The release of data requires approval from MedicinesInsight's external independent Data Governance Committee. This committee includes general practitioners, consumer advocates, privacy experts and researchers. Practices enrol in the MedicinesInsight program without the express consent of patients, though participating

practices are provided with information to keep in waiting areas and patients may opt out of having their data reported from the practice to MedicinesInsight.

### Strengths

Practices are only enrolled if they use the patient management systems Medical Director or Best Practice. MedicinesInsight staff perform data cleaning and coding of important information. This includes work to de-duplicate patients with multiple records at a single site, separate 'clinical encounters' from records that are administrative only and identify 'active' and 'regular' patients. Similarly, MedicinesInsight apply geographical information such as remoteness and socioeconomic status using the Australian Statistical Geography Standard Remoteness Areas [44] and the Socio-Economic Indices for Areas [45], respectively. Researchers still need to do additional cleaning of their own as apparent errors remain in the data (for example, impossible values in various free-text fields including clinical observations in some cases). Clinical codes are used for some important fields including Anatomic Therapeutic Classification codes [46] being used for the medicines prescribed and the Logical Observation Identifiers Names and Codes (LOINC) system [47] which is used for pathology data.

A data book and a data dictionary are provided to aid interpretation of data [43, 48]. These provide background information on the collections, data elements (with explanatory notes), governance and ethics, data quality and some worked examples of how the data can be used. Although these are extensive, they lack some detail in comparison to data dictionaries that might be provided by data linkage branches with more substantial experience working with academic researchers. In some cases, the data book and data dictionary lack explanations on data that are automatically coded by the patient information management systems of the practices, such as some clinical observations and pathology requests and results. However, enough information and support is available to help researchers decide if this is a suitable source of data for their project.

Data include a scrambled patient identifier [22], allowing service, pathology, prescription and other data belonging to a single patient to be linked and importantly, allowing for patients to be followed longitudinally. The program has been running since 2011 and, once a practice joins, the full clinical history of patients is available. Information on representativeness of the patient population in the MedicinesInsight data in comparison to the general Australian population is provided for some important fields [48].

MedicinesInsight provides some derived variables on request, notably including patient diagnostic flags for a number of important chronic conditions. These are likely to be useful to many researchers whether aiming to understand prevalence within a population of interest, prescribing within specific cohorts or countless other research questions. While the generated diagnostic flags may not always reflect the cohort a researcher is interested in or the level of detail needed (e.g. a researcher may need the date a diagnosis was made rather than a flag of its presence), detail of the data available gives researchers a great deal of flexibility in constructing their own cohorts and indicators.

## Challenges

As NPS MedicineWise seeks to recover the cost of data provision, researchers are required to pay data access fees, which may present a barrier for some researchers, although this is typical for any researcher wishing to access an existing dataset.

Most of the limitations of MedicineInsight data reflect the fact that it is a collection of the data generated across millions of encounters at hundreds of practices. Inconsistencies in the data captured and / or coded by general practitioners, or at any other step of the data generating process prior to MedicineInsight receiving the data, will ultimately be reflected in the data provided to researchers.

Some important data are provided as free-text fields, for example patient diagnoses and reason for encounter. This can result in a substantial time investment for researchers to identify clinical cohorts of interest, particularly when compared to collections such as hospitalisation data where this information is provided as International Classification of Disease (ICD) codes. The possibility of spelling errors and incorrect use of fields by general practitioners, for example the diagnosis field being used to record symptoms or family histories, can add to the time required to clean the data. This does, however, provide researchers with a high level of control over how they use the data. MedicineInsight do also offer a service of providing cohorts of interest which may save researchers significant time, albeit with added cost.

There are issues in some cases with missing data, for example, pathology test results with no unit of measurement recorded (approximately 8% of tests in the data checked by the authors) and prescription records with no information on medicine strength (approximately 1% of prescriptions) and no Anatomic Therapeutic Class code (4% of records), issues which ultimately reflect the recording of information by general practitioners.

As the MedicineInsight program staff continue to work with both general practices and researchers the data available and supporting documentation will continue to improve.

## Linkage to other data sources

Linkage of MedicineInsight data is currently in its infancy, both in terms of linking general practice data between practices and linking to other data collections. Identifying information such as name, date of birth and address is not collected by MedicineInsight and data linkage can only be implemented using encoded versions of person-identifying information. These cryptographic hashes of the person-identifying information are performed within the practice so no identifying information leaves the practice [22]. Linkage is being explored using the GRHANITE key and Statistical Linkage Key (SLK) matching algorithms. Research is occurring in Victoria to link MedicineInsight data with cancer registry data. The issue of patient identifiers that do not link between practices, meaning that patients who visit multiple providers will appear multiple times in the data, has been explored using algorithms by the University of Melbourne. Longitudinal research is possible for patients who receive their care through a single practice. MedicineInsight are planning work to identify records across practices belonging to the same patient, which will facilitate improved longitudinal research.

## Comparisons to other systems

We present here brief comparisons to other relevant general practice data collections. We describe other Australian collections that we are aware of; and provide comparisons to the UK's Clinical Practice Research Datalink (CPRD) and Wales' Secure Anonymised Information Linkage (SAIL) databank, which represent 'gold standards' in general practice data collections. We also describe systems in Canada at a similar level of development to the approaches described above. Canadian examples provide useful points of comparison on account of the two countries similar federations, and similarities in access to both public hospitals and medical services [49, 50].

## Australian examples

The Melbourne East Monash General Practice Database (MAGNET) presents another collection of general practice data to support research in Australia [51]. This database includes practices within a single region in Melbourne. Practices which agree to participate have their data encrypted and extracted to a data warehouse (without identifying information) where it can then be used for research. Data are fed back to practices for quality improvement, which is the primary purpose of the underlying infrastructure. The data include patient demographics, episodes of care, diagnoses, medications prescribed, observations, investigations ordered and received, immunisations and more, while some practice data is also included. This collection is much smaller than MedicineInsight in terms of practice and patient numbers included. The MAGNET database includes a SLK which allows the same patient to be identified at different practices and hence duplication avoided and may also support linkage to external data collections. MAGNET has been used as the basis for research investigating service use [52], prescribing [53], cohort characterisation [54] and more.

More recently, the Data for Decisions project has been established through the University of Melbourne [55]. This program has been recruiting practices since only 2017 and is described as being in its start-up phase. De-identified data are transferred to a data repository using GRHANITE, and the program may facilitate linkage to different datasets. The data extracted appears comparable to other data collections; with initial projects assessing antimicrobial prescribing, chronic disease programs and disease detection [56]. Data is accessed following approval by an independent Data Governance Committee, and a practice opt-in, patient opt-out model has been adopted.

Australia's earliest collection of CIS data is the General Practice Research Network, administered by the Health Communication Network, publishers of Medical Director. This system extracts information on medications, conditions, history, imaging and pathology tests ordered, observations taken, basic patient demographics and risk factors [57]. This repository is restricted to practices which use a single CIS (Medical Director) and operates on a GP opt-in, patient opt-out model. The dataset includes information for approximately 3 million unique patients of 1,100 GPs [58]. Since 1999 this network has supported publications examining prescribing and vaccination [59], though publications using this resource appear to be

rare, particularly in the last decade.

## International examples

Internationally, one of the most comprehensive sources of primary care data used for research is the Clinical Practice Research Datalink (CPRD) in the UK [60]. The CPRD has provided data and services to support research investigating pharmacovigilance and the use of medicines, informing of health policy and healthcare delivery and exploring disease risk factors for more than thirty years. The CPRD is a primary care database of de-identified medical records from general practices in the UK including over 11 million registered patients split across two datasets [61]. The CPRD has been designed to provide a representative population dataset and is linked to a number of other sources to provide a rich data resource. The CPRD includes information on encounters, immunisations, tests, therapies and patient socio-demographics. Data including diagnoses are recorded using version 2 Read codes. One of the major strengths of the CPRD is linkage to external data sources including hospitalisation, mortality and disease registries for a subset (over half) of practices [60]. Furthermore, the CPRD has a broad ethical approval for observational research using the primary care data and established linkages simplifying data access. Referrals to secondary care, and information fed back from secondary care are also included. Data quality is promoted through the Quality and Outcomes Framework in place in the UK which provides financial incentives for recording of important data items [62]. Similar financial incentives are currently being planned in Australia, which will reimburse practices for participating in quality improvement practices and sharing a minimum dataset demonstrating this with their local PHN [63], however the implementation of these incentives has been subject to delays and it is as yet unknown how data recording and quality may change as a result of the incentives [64].

The SAIL (Secure Anonymised Information Linkage) databank is based in Swansea University Medical School. SAIL works in partnership with researchers and health professionals, aiming to maximise the value of routinely collected individual level data through record linkage and to enable and support health related research [65]. Linked data from different sources is created by the NHS (National Health Service) Wales Informatics Service using the NHS number. The Welsh Demographic Service [66] provides personal information of all persons who have registered with a general practice or received care from health services in Wales [67, 68]. SAIL links a wide range of data including general practice, hospitalisation, national screening programmes, the national cancer registry and more [65]. The primary care dataset contains information on patient encounters with primary care; capturing the signs, symptoms, test results, diagnoses, prescribed treatment, specialist referrals and social aspects relating to the patient's home environment [69]. Similar to NPS MedicinesInsight, SAIL recruits general practices to voluntarily share data; practices sign up to share data without the express consent of patients, though practices are provided with information to keep in waiting areas and patients may opt out of having their data reported to SAIL. As of November 2018, SAIL had recruited 334 general practices (76% of all Welsh practices) which relates to information for over 2.5 million patients. SAIL

has now been used to support a range of research including follow-up of clinical trials [70], evaluation of medication use [71], epidemiological studies [72] and policy evaluations [73].

Currently the general practice data available for research in Australia is comparable to the CPRD and SAIL systems in some ways, including the breadth of data captured, the centralised data application and practice opt-in / patient opt-out consent model. The key point on which these UK systems are further developed than any system in Australia is the availability of linkage between general practice and other data sources, though this capability is developing with regards to the MedicinesInsight data collection and other collections within Australia.

There are also examples of developing primary care linkages in Canada. These include the Electronic Medical Record Administrative Data Linked Database (EMRALD) in Ontario, and the Canadian Primary Care Sentinel Surveillance Network (CPCSSN), which includes practices across most provinces. Like MedicinesInsight, these programs recruit general practices to voluntarily report data, based on their use of CISs. Each of these captures some information on patient encounters with physicians, medications prescribed, laboratory investigations and so on. The EMRALD database includes linkage to other health related administrative, survey, registry, demographic data and more through the comprehensive data holdings of the Institute for Clinical and Evaluative Sciences [74]. This provides a more complete picture of patient's care than is possible through the primary care data collections in Australia, although the system is confined to a single Canadian province [75]. The CPCSSN is larger than EMRALD in the number of practices and patients captured, with over 200 practice sites and 1.5 million patients as at May 2016 [76] though is smaller than the MedicinesInsight collection on these measures. Furthermore, the CPCSSN has recently begun to be used for research projects linking primary care data to other sources such as hospitalisation and census data [76, 77]. Case definitions are applied for a number of chronic conditions, similar to the MedicinesInsight collection, though in the case of the CPCSSN these are supported by the use of ICD-9 codes in Canadian CIS data. The scope of the data collected is otherwise similar. One problem common to both the CPCSSN and general practice data in Australia, is that when patients visit multiple providers they will exist in the database more than once and duplicated patients cannot be differentiated.

## The road ahead

Despite the increased use of CISs in general practice, and extensive collections of administrative data, research covering the whole health system has been hampered by jurisdictional issues obstructing linkages [3, 21], and until relatively recently, the lack of any centralised collections of general practice data. Such collections are now becoming available, and technological advances are developing to help resolve challenges around sharing of patient data for linkage to other sources. Even with technological solutions available, it remains important to build trust among general practitioners that any patient data shared remain confidential and that evidence generated is fed back to practices to support patient care and practice processes. Though evidence on the Australian public's views on data use

is scarce, opinion polls suggest that a vast majority are supportive of health records being used for research [78]. Meanwhile most GPs are supportive of general practice research though there are barriers to involvement [79]. Data sharing models that maintain patient and provider privacy are therefore likely to be valued by all. These data sharing challenges were the focus of an Australian Productivity Commission inquiry (under the Productivity Commission Act 1998) into the benefits and costs of options for increasing the availability and use of public and private sector data by individuals and organisations. The resulting report prepared by the Productivity Commission [80] proposes a new legal and policy framework to allow public and private sector data to flow. The recommendations in the report provide good foundations to build future data sharing and data linkage models.

## Conclusion

Within Australia there is extensive population-level administrative data captured on the use of health and other services. However, the delivery of services by different levels of government, and hence the holding of data by different levels of government, presents challenges to researchers aiming to perform whole of system research. Furthermore, administrative data covering primary care services are generally limited to basic information gathered for reimbursement purposes. Where research questions require the use of more detailed primary care data including patient diagnoses, the ordering of and results of pathology testing and measurement of observations, the prescribing of medications and so on, there are limited options available to researchers. We have described three avenues to access this data from general practices, though there are limitations to each of these methods. One limitation of the approaches described here is the limited ability to link the detailed clinical data from general practices to other sources of information such as hospital admissions and even other general practices although this capacity is emerging.

International examples demonstrate that comprehensive individual-level data capturing primary, secondary and tertiary care can be linked and made available to researchers, though these countries have universal unique identifiers for patients. We hope that the data available to Australian researchers can continue to improve, following these examples.

## Acknowledgements

We thank NPS MedicineWise staff for their time in reviewing the accuracy of content regarding MedicineInsight data.

## Statement on Conflicts of Interest

Two of the authors, JB and SR, are currently working on a research project with NPS MedicineWise. RM sits on NPS MedicineWise's General Practice Insights advisory group.

## Ethics statement

As this is a commentary paper, ethical approval is not required.

## References

1. Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. *Health Information Science and Systems*. 2014;2(3). <https://doi.org/10.1186/2047-2501-2-3>
2. Tew M, Dalziel KM, Petrie DJ, Clarke PM. Growth of linked hospital data use in Australia: a systematic review. *Australian Health Review*. 2017;41:394-400. <https://doi.org/10.1071/AH16034>
3. Canaway R, Boyle DIR, Manski-Nankervis JAE, Bell J, Hocking JS, Clarke K, et al. Gathering data for decisions: best practice use of primary care electronic records for research. *Medical Journal of Australia*. 2019;210(S6):S12-S6. <https://doi.org/10.5694/mja2.50026>
4. Foster M, Fleming J. The Health Care System in Australia. In: Taylor S, Foster M, Fleming J, editors. *Health Care Practice in Australia: Policy, Context and Innovations*. Melbourne: Oxford University Press; 2008.
5. Duckett S, Willcox S. *The Australian Health Care System*. 4th ed. Melbourne: Oxford University Press; 2011.
6. Willis E, Reynolds L, Keleher H. *Understanding the Australian Health Care System*. 3rd ed. NSW: Elsevier; 2016.
7. Department of Health. Fact sheet: Primary Health Networks Canberra: Australian Government; 2018 [Available from: <https://www1.health.gov.au/internet/main/publishing.nsf/Content/Fact-Sheet-Primary-Health-Networks+>].
8. Henderson J, Javanparast S, MacKean T, Freeman T, Baum F, Ziersch A. Commissioning and equity in primary care in Australia: Views from Primary Health Networks. *Health and Social Care in the Community*. 2018;26:80-9. <https://doi.org/10.1111/hsc.12464>
9. van Ommeren M, Merton C, Short G. *Inventory of Australian health data collections*. Canberra: Australian Government Publishing Service; 1991.
10. Paige E, Kemp-Casey A, Korda R, Banks E. Using Australian Pharmaceutical Benefits Scheme data for pharmacoepidemiological research: challenges and approaches. *Public Health Research & Practice*. 2015;25(4). <https://doi.org/10.17061/phrp2541546>
11. Department of Health. *Medicare Benefits Schedule Data Canberra: Australian Government; 2017* [Available from: [http://www.health.gov.au/internet/main/publishing.nsf/Content/PHN-MBS\\_Data](http://www.health.gov.au/internet/main/publishing.nsf/Content/PHN-MBS_Data)].
12. Tran B, Straka P, Falster MO, Douglas KA, Jorm LR. Overcoming the data drought: exploring general practice in Australia by network analysis of big data. *Medical Journal of Australia*. 2018;209(2):68-73. <https://doi.org/10.5694/mja17.01236>



13. Britt H, Miller G. BEACH program update. *Australian Family Physician*. 2015;44(6):411-4.
14. Pearce C. The Adoption of Computers by Australian General Practice - A Complex Adaptive Systems Analysis. *Journal of General Practice*. 2013;1(3). <https://doi.org/10.4172/2329-9126.1000121>
15. Primary Health Care Advisory Group. Better Outcomes for People with Chronic and Complex Health Conditions: Report of the Primary Health Care Advisory Group Canberra: Department of Health; 2015 [Available from: [https://www.health.gov.au/internet/main/publishing.nsf/Content/76B2BDC12AE54540CA257F72001102B9/\\$File/Primary-Health-Care-Advisory-Group-Final-Report.pdf](https://www.health.gov.au/internet/main/publishing.nsf/Content/76B2BDC12AE54540CA257F72001102B9/$File/Primary-Health-Care-Advisory-Group-Final-Report.pdf)].
16. Mitchell RJ, Cameron CM, McClure RJ, Williamson AM. Data linkage capabilities in Australia: practical issues identified by a Population Health Research Network 'Proof of Concept project'. *Aust NZ J Public Health*. 2015;39:319-25. <https://doi.org/10.1111/1753-6405.12310>
17. Holman CDJ, Bass AJ, Rosman DL, Smith MB, Semmens JB, Glasson EJ, et al. A Decade of Data Linkage in Western Australia: Strategic Design, Applications and Benefits of the WA Data Linkage System. *Australian Health Review*. 2008;32(4):766-77. <https://doi.org/10.1071/ah080766>
18. Centre for Health Record Linkage. Publications Sydney: NSW Ministry of Health; 2019 [Available from: <http://www.cherel.org.au/publications>].
19. Australian Institute of Health and Welfare. Our secure linkage environment Canberra: Australian Government; 2019 [Available from: <https://www.aihw.gov.au/our-services/data-linkage/integrating-authority>].
20. Frommer PM, Madriono C, Kemp S, Jenkin R, Reitano R. NCRIS Capability 5.7: Population Health and Data Linkage. Sydney: University of Sydney; 2007. p. 8.
21. Palamuthusingam D, Johnson D, Hawley C, Pascoe E, Fahim M. Health data linkage research in Australia remains challenging. *Internal Medicine Journal*. 2019;49:539-44. <https://doi.org/10.1111/imj.14244>
22. Busingye D, Gianacas C, Pollack A, Chidwick K, Merrifield A, Norman S, et al. Data Resource Profile: MedicineInsight, an Australian national primary health care database. *International Journal of Epidemiology*. 2019. <https://doi.org/10.1093/ije/dyz147>
23. Gordon J, Miller G, Britt H. Reality check - reliable national data from general practice electronic health records Canberra: Deeble Institute for Health Policy Research; 2016 [Available from: [https://ahha.asn.au/system/files/docs/publications/deeble\\_institutue\\_issues\\_brief\\_no\\_18.pdf](https://ahha.asn.au/system/files/docs/publications/deeble_institutue_issues_brief_no_18.pdf)].
24. Oon YK. DOCLE. A new system of medical notation. *Australian Family Physician*. 1988;17(9):772-3.
25. Young J, Eley D, Fahey P, Patterson E, Hegney D. Enabling research in General Practice: Increasing Functionality of Electronic Medical Records. *Australian Family Physician*. 2010;39(7):506-9.
26. PenCS. CAT4 Sydney: PenCS; 2015 [Available from: <http://www.pencs.com.au/products/cat4/>].
27. Outcome Health. POLAR GP / POLAR Explorer Melbourne: Outcome Health; 2016 [Available from: <https://outcomehealth.org.au/polar.aspx>].
28. Improvement Foundation. A change for good - introducing new data extraction software Adelaide: Improvement Foundation; 2018 [Available from: <http://improve.org.au/news-and-media/media-centre/canningtool/#.W805gWgza71>].
29. Boyle DIR, Rafael N. BioGrid Australia and GRHANITE: Privacy-Protecting Subject Matching. In: Hansen DP, Maeder AJ, Schaper LK, editors. *Health Informatics: The Transformative Power of Innovation*: IOS Press; 2011. p. 24-34.
30. Health and Biomedical Informatics Centre. GRHANITE Health Informatics Unit Melbourne: The University of Melbourne; [Available from: <https://www.grhanite.com/projects/>].
31. Office of the Australian Information Commissioner. Notifiable Data Breaches scheme Canberra: Australian Government; [Available from: <https://www.oaic.gov.au/privacy-law/privacy-act/notifiable-data-breaches-scheme>].
32. Roland M, Guthrie B. Quality and Outcomes Framework: what have we learnt? *British Medical Journal*. 2016;354(i4060). <https://doi.org/10.1136/bmj.i4060>
33. Boiten E. Care.data has been scrapped, but your health data could still be shared: The Conversation; 2016 [Available from: <https://theconversation.com/care-data-has-been-scrapped-but-your-health-data-could-still-be-shared-62181>].
34. Ferdousi S. Transformed Health Through Data & Insights Western Sydney: Western Sydney PHN; 2017 [Available from: [https://ahha.asn.au/sites/default/files/docs/policy-issue/shahana\\_ferdousi.pdf](https://ahha.asn.au/sites/default/files/docs/policy-issue/shahana_ferdousi.pdf)].
35. Abhijeet G, Charlton KE, Girdo L, Batterham M. Using data from patient interactions in primary care for population level chronic disease surveillance: The Sentinel Practices Data Sourcing (SPDS) project. *BMC Public Health*. 2014;14(557). <https://doi.org/10.1186/1471-2458-14-557>

36. Bhaskaran K, Douglas I, Forbes H, dos-Santos-Silva I, Leon DA, Smeeth L. Body-mass index and risk of 22 specific cancers: a population-based cohort study of 5.24 million UK adults. *Lancet*. 2014;384:755-65. [https://doi.org/10.1016/S0140-6736\(14\)60892-8](https://doi.org/10.1016/S0140-6736(14)60892-8)
37. Smeeth L, Thomas SL, Hall AJ, Hubbard R, Farrington P, Vallance P. Risk of Myocardial Infarction and Stroke after Acute Infection or Vaccination. *The New England Journal of Medicine*. 2014;351(25):2611-8. <https://doi.org/10.1056/NEJMoa041747>
38. Randall SM, Ferrante AM, Boyd JH, Bauer JK, Semmens JB. Privacy-preserving record linkage on large real world datasets. *Journal of Biomedical Informatics*. 2014;50:205-12. <https://doi.org/10.1016/j.jbi.2013.12.003>
39. National Health and Medical Research Council. National Statement on Ethical Conduct in Human Research, 2007 (Updated 2018). In: Council AR, editor. Canberra: Australian Government; 2018.
40. Western Australia Department of Health. Data Linkage Western Australia Perth: Government of Western Australia; 2019 [Available from: <https://www.datalinkage-wa.org.au/>].
41. Curtin University. Centre for Data Linkage Perth: Curtin University; 2019 [Available from: <https://healthsciences.curtin.edu.au/healthsciences-research/research-institutes-centres/centre-for-data-linkage/>].
42. National Prescribing Service. MedicinesInsight NSW: National Prescribing Service; 2017 [Available from: <https://www.nps.org.au/medicine-insight>].
43. MedicinesInsight. MedicinesInsight Data Book version 1.1. Sydney: NPS MedicineWise; 2018.
44. Australian Bureau of Statistics. Australian Statistical Geography Standard (ASGS) Canberra: Australian Bureau of Statistics; 2018 [Available from: [https://www.abs.gov.au/websitedbs/D3310114.nsf/home/Australian+Statistical+Geography+Standard+\(ASGS\)](https://www.abs.gov.au/websitedbs/D3310114.nsf/home/Australian+Statistical+Geography+Standard+(ASGS))].
45. Australian Bureau of Statistics. Socio-Economic Indexes for Areas Canberra: Australian Bureau of Statistics; 2018 [Available from: <https://www.abs.gov.au/websitedbs/censushome.nsf/home/seifa>].
46. WHO Collaborating Centre for Drug Statistics Methodology. ATC/DDD Index 2019: World Health Organisation; 2018 [Available from: [https://www.whocc.no/atc\\_ddd\\_index/](https://www.whocc.no/atc_ddd_index/)].
47. Forrey A, McDonald C, DeMoor G, Huff S, Leavelle D, Leland D, et al. Logical observation identifier names and codes (LOINC) database: a public use set of codes and names for electronic reporting of clinical laboratory test results. *Clinical Chemistry*. 1996;42(1):81-90. <https://doi.org/10.1093/clinchem/42.1.81>
48. MedicinesInsight. MedicinesInsight Data Book version 1.2. Sydney: National Prescribing Service; 2016.
49. Philippon DJ. Health System Organization and Governance in Canada and Australia: A Comparison of Historical Developments, Recent Policy Changes and Future Implications. *Health Policy*. 2008;4(1):e168-e86. <https://doi.org/10.12927/hcpol.2008.19991>
50. Duckett S. Expanding the breadth of Medicare: learning from Australia. *Health Economics, Policy and Law*. 2018;13:344-68. <https://doi.org/10.1017/S1744133117000421>
51. Mazza D, Pearce C, Turner LR, Leon-Santiago M, McLeod A, Ferriggi J, et al. The Melbourne East Monash General Practice Database: Using data from computerised medical records for research. *Journal of Innovation in Health Informatics*. 2016;23(2):523-8. <http://doi.org/10.14236/jhi.v23i2.181>
52. Turner LR, Pearce C, Brijnath B, Browning C, Lowthian J, Shearer M, et al. General practice utilisation of Medicare Benefits Schedule items to support the care of older patients: findings from the REDIRECT study. *Australian Journal of Primary Health*. 2018;24:54-8. <https://doi.org/10.1071/PY17047>
53. Brijnath B, Xia T, Turner L, Mazza D. Trends in GP prescribing of psychotropic medications among young patients aged 16-24 years: a case study analysis. *BMC Psychiatry*. 2017;17(214). <https://doi.org/10.1186/s12888-017-1375-2>
54. Turner LR, Pearce C, Borg M, McLeod A, Shearer M, Mazza D. Characteristics of patients presenting to an after-hours clinic: results of a MAGNET analysis. *Australian Journal of Primary Health*. 2017;23:294-9. <https://doi.org/10.1071/PY16084>
55. The University of Melbourne. Data for Decisions and the Patron program of research Melbourne: The University of Melbourne; 2019 [Available from: <https://medicine.unimelb.edu.au/school-structure/general-practice/engagement/data-for-decisions>].
56. The University of Melbourne. Projects Melbourne: The University of Melbourne; 2019 [Available from: <https://medicine.unimelb.edu.au/school-structure/general-practice/engagement/data-for-decisions/projects>].
57. Sayer G, McGeechen K, Kemp A, Bhasale A, Horn F, Hendrie L, et al. The General Practice Research Network: the capabilities of an electronic patient management system for longitudinal patient data. *Pharmacoepidemiology and Drug Safety*. 2003;12(6):483-9. <https://doi.org/10.1002/pds.834>
58. Carrington M. Bittersweet findings of blood glucose levels in 467,955 patients in primary care. Melbourne: Baker IDI Heart & Diabetes Institute; 2015.

59. Trinh L, Macartney K, McIntyre P, Chiu C, Dey A, Menzies R. Investigating adverse events following immunisation with pneumococcal polysaccharide vaccine using electronic General Practice data. *Vaccine*. 2017;35(11):1524-9. <https://doi.org/10.1016/j.vaccine.2017.01.063>
60. Herret E, Gallagher AM, Bhaskaran K, Forbes H, Mathur R, van Staa T, et al. Data Resource Profile: Clinical Practice Research Datalink (CPRD). *International Journal of Epidemiology*. 2015;44(3):827-36. <https://doi.org/10.1093/ije/dyv098>
61. CPRD. Clinical Practice Research Datalink: NHS; 2018 [Available from: <https://www.cprd.com/>].
62. Roland M, Guthrie B. Quality and Outcomes Framework: what have we learnt? *BMJ*. 2016(354):i4060. <https://doi.org/10.1136/bmj.i4060>
63. Department of Health. POP QI Incentive guidance Canberra: Australian Government; 2019 [Available from: [https://www.health.gov.au/internet/main/publishing.nsf/Content/PIP-QI\\_Incentive\\_guidance](https://www.health.gov.au/internet/main/publishing.nsf/Content/PIP-QI_Incentive_guidance)].
64. Woodley M. Details of government incentive program revealed. NewsGP. 2019.
65. Ford DV, Jones KH, Verplancke J-P, Lyons RA, John G, Brown Gea. The SAIL Databank: building a national architecture for e-health research and evaluation. *BMC Health Services Research*. 2009;9(157). <https://doi.org/10.1186/1472-6947-9-3>
66. NHS Wales Informatics Service. Handling Patient Details Wales: NHS Wales; 2017 [Available from: <http://www.wales.nhs.uk/nwis/page/52552>].
67. Jones KV, Ford DV, Jones C, Dsilva R, Thompson S, Brooks CJ. A case study of the Secure Anonymous Information Linkage (SAIL) Gateway: A privacy-protecting remote access system for health-related research and evaluation. *Journal of Biomedical Informatics*. 2014;50:196-204. <https://doi.org/10.1016/j.jbi.2014.01.003>
68. Lyons RA, Jones KH, John G, Brooks CJ, Verplancke J-P, Ford DV, et al. The SAIL databank: linking multiple health and social care datasets. *BMC Medical Informatics and Decision Making*. 2009;9(3). <https://doi.org/10.1186/1472-6947-9-3>
69. SAIL Databank. Primary Care GP dataset Swansea: Swansea University; 2018 [Available from: <https://saildatabank.com/saildata/sail-datasets/primary-care-gp-dataset/>].
70. Davies G, Jordan S, Brooks CJ, Thayer D, Storey M, Morgan G, et al. Long term extension of a randomised controlled trial of probiotics using electronic health records. *Scientific Reports*. 2018;8(7668). <https://doi.org/10.1038/s41598-018-25954-z>
71. Charlton RA, Neville AJ, Jordan S, Pierini A, Damase-Michel C, Klungsoyr K, et al. Healthcare databases in Europe for studying medicine use and safety during pregnancy. *Pharmacoepidemiology & Drug Safety*. 2014;23:586-94. <https://doi.org/10.1002/pds.3613>
72. Sayers A, Thayer D, Harvey JN, Luzio S, Atkinson MD, French R, et al. Evidence for a persistent, major excess in all cause admissions to hospital in children with type-1 diabetes: results from a large Welsh national matched community cohort study. *BMJ Open*. 2014;5(4):e005644. <https://doi.org/10.1136/bmjopen-2014-005644>
73. Lyons RA, Rodgers SE, Thomas S, Bailey R, Brunt H, Thayer D, et al. Effects of an air pollution personal alert system on health service usage in a high-risk general population: a quasi-experimental study using linked data. *Journal of Epidemiology & Community Health*. 2016;70(12):1184-90. <https://doi.org/10.1136/jech2016-207222>
74. Institute for Clinical and Evaluative Sciences. Electronic Medical Record Administrative data Linked Database (EMRALD) Ontario: Institute for Clinical and Evaluative Sciences; 2018 [Available from: <https://www.ices.on.ca/Research/Research-programs/Primary-Care-and-Population-Health/EMRALD>].
75. Schultz SE, Rothwell DM, Chen Z, Tu K. Identifying cases of congestive heart failure from administrative data: a validation study using primary care patient records. *Chronic Diseases and Injuries in Canada*. 2013;33(3):160-6.
76. Garies S, Birtwhistle R, Drummond N, Queenan J, Williamson T. Data Resource Profile: National electronic medical record data from the Canadian Primary Care Sentinel Surveillance Network (CPSSN). *International Journal of Epidemiology*. 2017;46(4):1091-2f. <https://doi.org/10.1093/ije/dyw248>
77. Garies S, Soos B, Williamson T, Frost B, Manca D, Duerksen K, et al. Linking primary care EMR data and administrative data in Alberta, Canada: experiences, challenges, and potential solutions. *International Journal of Population Data Science*. 2018;3(4). <https://doi.org/10.23889/ijpds.v3i4.991>
78. Research Australia. Australia Speaks! Research Australia Opinion Polling 2017 2017 [Available from: <https://researchaustralia.org/wp-content/uploads/2017/08/2017-Opinion-Poll-Digital.pdf>].
79. Askew DA, Clavarino AM, Glasziou PP, Del Mar CB. General practice research: attitudes and involvement of Queensland general practitioners. *MJA*. 2002;177:74-7. <https://doi.org/10.5694/j.1326-5377.2002.tb04670.x>
80. Productivity Commission. Data Availability and Use. Canberra: Australian Government; 2017.

## Abbreviations

|          |  |
|----------|--|
| AIHW     | Australian Institute of Health and Welfare                       |
| BEACH    | Bettering the Evaluation and Care of Health                      |
| CAT4     | Clinical Audit Tool 4  |
| cdmNet   | Chronic Disease Management-Net                                   |
| CHeReL   | Centre for Health Record Linkage                                 |
| CIS      | Clinical Information System                                      |
| CPCSSN   | Canadian Primary Care Sentinel Surveillance Network              |
| CPRD     | Clinical Practice Research Datalink                              |
| DOCLE    | Doctor Command Language  |
| EMRALD   | Electronic Medical Record Administrative Data Linked Database    |
| GP       | General Practitioner   |
| GRHANITE | GeneRiC HeAlth Network Information Technology for the Enterprise |
| ICD      | International Classification of Diseases                         |
| ICPC     | International Classification of Primary Care                     |
| LOINC    | Logical Observation Identifiers Names and Codes                  |
| MAGNET   | Melbourne East Monash General Practice Database                  |
| MBS      | Medicare Benefits Schedule                                       |
| NWIS     | NHS Wales Informatics Service                                    |
| NHS      | National Health Services   |
| NPS      | National Prescribing Service                                     |
| PBS      | Pharmaceutical Benefits Scheme                                   |
| POLAR    | Population Level Analysis & Reporting                            |
| PHN      | Primary Health Network   |
| PHRN     | Population Health Research Network                               |
| QOF      | Quality and Outcomes Framework                                   |
| SAIL     | Secure Anonymised Information Linkage                            |
| SLK      | Statistical Linkage Key  |
| SNOMED   | Systematized Nomenclature of Medicine                            |
| SQL      | Structured Query Language  |

