**Department of General Practice**
Faculty of Medicine, Dentistry & Health Sciences

# Primary care data and linkage

## Australian dataset identification & capacity building

22 December 2020

**Report prepared by Department of General Practice for Melbourne Academic Centre for Health (MACH) reporting of the Australian Health Research Alliance (AHRA) study: *Building capacity in data driven health care improvement***

**December 2020**

**Citation:**

Canaway R, Boyle D, Manski-Nankervis J, Gray K and MACH (2020). Primary Care Data and Linkage: Australian dataset mapping and capacity building. A report from the Melbourne Academic Centre for Health for the Australian Health Research Alliance, Melbourne, Australia, pp.54

Department of General Practice, Level 3 / 780 Elizabeth Street, The University of Melbourne, Victoria 3010

# Table of Contents

# Tables

# Acknowledgements

The findings from this study are largely based on survey findings. Some survey respondents opted to be acknowledged in this report; the table below lists them (excluding those who did not complete the survey). Their names and affiliations have been copied verbatim from their completed questionnaire.[1] We would like to thank everyone who responded to the survey, and those who participated in an interview or the workshop.

| | Name | Affiliation |
|---|---|---|
| **1** | | Sydney School of Public health, University of Sydney |
| **2** | | System Information and Analytics Branch, NSW Ministry of Health |
| **3** | | WA Primary Health Alliance |
| **4** | Gloria Antonio | NPS MedicineWise |
| **5** | Dr Nasser Bagheri | Australian National University |
| **6** | Dr Diane Blanckensee | Tweed Health for Everyone Superclinic |
| **7** | Douglas Boyle | University of Melbourne |
| **8** | Rosemary Cooper | Sonic Healthcare |
| **9** | Kirsty Douglas | Australian National University |
| **10** | Oliver Frank | Oakden Medical Centre and the University of Adelaide |
| **11** | Abhijeet Ghosh | COORDINARE: South Eastern NSW PHN |
| **12** | Harriet Hiscock | Royal Children's Hospital |
| **13** | Louisa Jorm | UNSW Sydney |
| **14** | Fabian Kong | Victorian Government Department of Health and Human Services |
| **15** | Prof Amanda Leach | Menzies School of Health Research |
| **16** | Professor Dimity Pond | University of Newcastle |
| **17** | Jan Radford | University of Tasmania |
| **18** | Anthony Scott | The University of Melbourne |
| **19** | Prof Nigel Stocks | Head Discipline of General Practice, University of Adelaide |
| **20** | Elisa Manley Wentworth | Healthcare Ltd, providers of the Nepean Blue Mountain Primary Health Network |
| **21** | Dr Kam Cheong Wong | University of Sydney |

## Funder

This work is funded by a Medical Research Future Fund grant to the Australian Health Research Alliance (AHRA) via the Melbourne Academic Centre for Health (MACH), Advanced Health Research Translation Centre. This work has been undertaken under the auspices of the AHRA Data Driven Healthcare Improvement initiative under the priority area 'Integration of large data sets across the care continuum'.

AHRA is an alliance of Advanced Health Research Translation Centres (NHMRC) and Centres for Innovation in Regional Health, established with funding from the Medical Research Futures Fund. AHRA have prioritised greater use of data assets and their integration into the national health data framework to support healthcare delivery, service improvement and best practice.

[1] The survey instruction read: "Please check the box if you would like to be named in acknowledges of any published results as having contributed to this research" – "Please enter your name and affiliation".

# 1. Executive Summary

In Australia most people receive most of their health care in primary and community health sectors yet access to primary care clinical data for secondary purposes is limited. 'Secondary purposes' refers to data used for reasons other than the purpose for which they were collected, e.g. research, audit, surveillance, quality assurance activities and data linkage. Australia's secondary use of health data is said to be behind many developed countries, and this limits knowledge about the patient journey through the healthcare system [1, 2].

This report details the findings of a study commissioned by the Australian Health Research Alliance (AHRA) in 2018 as part of its objective to increase data use by healthcare professionals and other stakeholders to improve health outcomes. AHRA have prioritised greater use of data assets and their integration into the national health data framework to support healthcare delivery, service improvement and best practice.

**Drawing primarily from survey and interview data from 62 custodians and users of primary care data used for secondary purposes** in Australia, the study explored the primary care-related datasets considered available in Australia and their data linkage status; the quality frameworks being used to assess primary care datasets; and the stakeholder perceptions of enablers and barriers to strategically building benefit and capacity of primary care data use.

Some **106 datasets from eclectic sources were identified** by participants, including some that did not strictly contain primary care clinical data (e.g. health workforce data). The large number of datasets suggests duplication of effort around data collection, cleaning and use. Appendix 7.7 summarises the information provided by participants about these datasets and their purpose. The cost of gaining access to data varied widely but for clinical data appears to be in the tens of thousands of dollars; this could be in addition to having the data recipient provide a service or feedback to general practices that provided the data, or having the Data Custodian be a named co-author of resultant research papers. **A third of data custodians did not use a data quality framework** or were not sure whether they did (section 4.9.24.9). There was lack of clarity around what a data quality framework was and some scepticism about the quality of data in primary care repositories. The breadth of datasets included in this study may mean that they are too variable to allow application of a standard type of data quality framework.

Thematic analysis of interview and survey data yielded a strong narrative that primary care datasets could be better used to **support improved health outcomes by employing data linkage. C**oupled With appropriate research designs, the health sector could make much greater use of existing data to generate evidence to demonstrate needs and opportunities, then act on the new knowledge to enhance services. Data linkage (section 4.5) was repeatedly referred to, as the way to increase the utility or benefit of secondary use of primary care data. No respondent suggested data linkage be avoided, however the inability to link datasets was frequently raised.

Limitations of and barriers to secondary use of primary care data were many. Poor data quality, technical limitations, lack of standards and guidelines and a common data model, lack of leadership, expertise, data availability and understanding of primary care data complexity and context were commonly raised. However, along with data quality issues, **fear and lack of trust** were the most cited barriers to better use of primary care data. Fear underpinned many barriers – fear of poor data security, government attempts to control GPs, privacy and reputational damage – and garnering trust to allay fear underpinned many of the suggested **enablers of better secondary use of primary care data. To build trust requires** transparency of governance and end-to-end data-related processes, strong and transparent leadership, meaningful stakeholder engagement, shared vision, robust data security and privacy protection, and publicised outcomes to demonstrate benefit and provide reassurance that data custodians and end-users are 'doing the right thing'.

**Resources and investment** are required for training and education, incentivising, data quality and tool improvements, workforce upskilling and making data access more affordable to end-users (more research means more potential for good). Time is also an enabler, time to work through a maturation process toward greater acceptance of data sharing and toward greater proficiency in transforming clinically generated data so it can work for a collective greater good (conversion to a common data model is a specific example).

By taking a broad approach to what a primary care-related data set is, we were seeking to hear from anyone who was using data related to primary care – particularly for research purposes. If a dataset, extraction tool or data linkage broker was not named that does not necessarily mean it is not relevant. Some data custodians may have commercial or research interests which led them to protect their dataset. Some gave vague descriptions of their dataset – which could equally be to protect their anonymity. The many gaps in the table of named datasets (Appendix 7.7) reflects the limited information that was shared. Nevertheless, participants were very forthright in their descriptions of benefits, limitations, barriers and enablers of better use of primary care data. Taken as a whole, this study contributes strong support for measures to make better use of primary care data.

# 2. Background

Most Australians receive most of their healthcare in primary care settings from general practitioners (GPs), yet there is limited secondary use of primary healthcare data (including data linkage with other routinely collected data – hospitals, specialist and allied healthcare providers, social services, education, etc). Secondary data use is when data are used for purposes other than the purpose for which it was collected – including for research, audit, surveillance and quality improvement activities. Australia's use of health data is behind many developed countries and this, among other things, limits knowledge about the patient journey through the healthcare system [3]. Primary care data is primarily drawn from electronic medical records (EMRs) which are designed to support clinical, not research, purposes [1, 2].

Despite secondary use being limited, it does occur. Tools for capture and extraction of primary care electronic data have been used for over 10 years, for example, The Canning tool, GRHANITE, POLAR GP, Pen CAT, and cdmNET. My Health Record has increased in prominence since the 2018 change from opt-in to opt-out and the release of the *Framework to guide the secondary use of My Health Record system data* in May 2018 [4]. In 2018, the Quality Improvement Practice Incentive Payment (QI PIP) was introduced, but it was not implemented until August 2019 [5]. The QI PIP now incentivises general practice sharing of data and participation in a data-driven quality improvement program. The regulatory environment for data sharing is also changing. The Australian Government response to the Productivity Commission's [3] report was released on 1st May 2018 with proposed reforms underpinned by a proposed Data Sharing and Release Act (DS&R Act). A discussion paper on the proposed legislative reforms received 68 submissions in September-October 2019 (but at the time of writing the Bill was yet to be introduced).

When linked to other datasets (such as hospital and administrative datasets) primary care data has the capacity to help researchers and policymakers better understand the patient journey through the health system. Yet due to the project-by-project nature of most secondary use of primary care data, across multiple independent institutions, there is little understanding of what data are being used, where, by whom, how it is collected, to what level of granularity, and the barriers to and benefits of its collection and use [2]. Without this 'big picture' it is possible that resources are being duplicated and/or poorly utilised, leading to greater than needed economic burden associated with using primary care data. Whether data quality frameworks are applied to ensure that the clinical data are appropriate for research is also not well known. The range of methods used for extracting and using primary care data for secondary purposes have not been mapped, and therefore little is known about the environments in which secondary use are occurring.

The Melbourne Academic Centre for Health (MACH) is one of a number of National Health and Medical Research Council (NHMRC) recognised Advanced Health Research and Translation Centres [AHRTCs] in Australia. The MACH is a member of the Australian Health Research Alliance (AHRA). AHRA has several national systems level initiatives, one being to *build capacity in data driven healthcare improvement*. They have prioritised greater use of data assets and their integration into the national health data framework to support healthcare delivery, service improvement and best practice. This is part of a growing momentum in Australia to better use existing data both within and outside of the healthcare sector [2, 6].

The study reported here was led by MACH /AHRA member A/Prof Douglas Boyle, University of Melbourne. Its purpose was to scope existing general practice/primary care dataset and data harmonisation activities nationally, and engage stakeholders around gaps, needs and optimal delivery methods for building capacity in data driven health care improvement and data harmonization.

The University of Melbourne is a MACH partner. This report, compiled by University of Melbourne staff, is for MACH reporting to AHRA and, as appropriate, to study informants and other stakeholders.

### 2.1.1. Related analysis of a Data Quality Framework

This study has another component, not detailed here, related to building deeper understanding and trust in routinely collected general practice data. In that component, undertaken by Sandra Henley-Smith (GRHANITE® IT Business Analyst, HaBIC $R^2$, Department of General Practice, The University of Melbourne) with oversight from by A/Prof Douglas Boyle and A/Prof Kathleen Gray, the Data Quality framework established by Kahn and colleagues [7] was tested for practical and systematic data warehouse implementation. The initial pilot work from the Data Quality testing component were presented at the same workshop as were the preliminary findings from this study reported on below. The AHRA-facilitated workshop: *Primary care data for health care improvement and research: approaches to data quality assurance*, held 26th November 2018, at the Mercure Hotel George Street, Sydney, is discussed at section 3.5. Appendix

7.2 provides a summary of the activity related to this work that tested the implementation of the data quality framework. Henley-Smith, Boyle and Gray [8] have published this work.

# 3. Study design and methods

## 3.1. Aims

The research described in this report aimed to:

- Gain a deeper understanding of the availability and secondary use of routinely collected primary care data (primarily general practice data) across Australia.
- Build capacity[2] for data-driven healthcare improvement.

## 3.2. Research questions

- What primary care datasets are available in Australia and are they being linked to other datasets?
- What quality frameworks are being used to assess primary care datasets?
- What enablers and barriers should be considered to maximise the effectiveness of strategies to build benefit and capacity of primary care data?

## 3.3. Investigators

The study was carried out by:

- Associate Prof Douglas Boyle BSc(Hons), PGDip IT, PhD. Director, HaBIC Research Information Technology Unit (HaBIC R[2]), Department of General Practice, The University of Melbourne
- Associate Prof Kathleen Gray BA, MLS, MEnvSc, PhD. Joint appointment, Melbourne Medical School and the School of Computing and Information Systems, The University of Melbourne
- Associate Prof Jo-Anne Manski-Nankervis BSc(Hons), MBBS(Hons), CHIA, PhD, FRACGP.. Department of General Practice, The University of Melbourne
- Dr Rachel Canaway BHSc, MSocHlth, PhD. VicREN Primary Care Research Network Manager, Department of General Practice, The University of Melbourne

### Disclosure statement

A/Prof Douglas Boyle is the Data Steward for the Patron primary care data repository, part of the University of Melbourne's *Data for Decisions* research initiative. He is also the developer of the GRHANITE® data extraction tool and is a member of the AHRA Data-driven Healthcare Improvement Committee. A/Prof Jo-Anne Manski-Nankervis is a researcher involved in *Data for Decisions.* She is currently a member of the RACGP Data Governance Advisory Committee, she has utilised NPS MedicineInsight data for research and is on their MedicineInsights Advisory Group. Dr Rachel Canaway project manages *Data for Decisions* (www.gp.unimelb.edu.au/datafordecisions) which incorporates the Patron primary care data repository. The researchers' experience establishing Patron provided broad understanding of collection and use of general practice data in Australia.

## 3.4. Data collection: survey and interviews

To address the research questions, an online data collection questionnaire was developed to elicit information about datasets, applied use of quality frameworks, data linkage activities and informants' perspectives on use of general practice datasets, including facilitators, barriers, benefits and limitations to use of general practice data. It was piloted

[2] Capacity building is an umbrella term that aggregates many different strategies which may not be mutually exclusive (e.g. partnership development, advocacy, resource allocation, communication and knowledge transfer, leadership development). Capacity building is used "to ultimately improve or strengthen an existing capacity". It is a process "that enhances the ability of the individual, entity or a broader social system to perform effectively in the functions for which they exist". https://www.vichealth.vic.gov.au/~/media/resourcecentre/publicationsandresources/health%20promotion/capacity%20building%20for%20whom%20in%20what%20circumstances_final.ashx Accessed 1 November 2020

before purposive and snowball sampling were used for its dissemination. Informants were invited to complete the questionnaire or be interviewed (to provide greater depth of response and insights) or both. No reimbursement or incentive payments were offered, but informants were given the option to be acknowledged in any reports arising from the study (see Acknowledgements p.iv). Preliminary findings were presented at a workshop attended by members of Advanced Health Research and Translation Centres (AHRTCs) and other stakeholders; workshop discussions aimed to make steps towards a national, unified approach to secondary use of primary care data. With ethics approval and informed consent, workshop participant input was recorded as part of data collection for this study so that their input could form part of the analytical thinking around next steps (see section 3.5).

### 3.4.1. Participant eligibility

Eligible participants were aged 18 years and over and were Data Custodians or Data Users who self-identified as having used an Australian primary care dataset for secondary purposes or having responsibility for a primary care dataset used for secondary purposes. A Data Custodian was someone who considered themselves to be an 'owner' of one or more general practice or other primary care datasets.

### 3.4.2. Questionnaire

The data collection questionnaire was hosted within the University of Melbourne's instance of REDCap (Research Electronic data Capture). REDCap is a secure, web-based application designed to support data capture for research studies [9]. The online questionnaire was progressively piloted by eight people and modifications made based on their responses to increase survey functionality and responder comprehension. The questionnaire employed skip logics to direct the participant based on the sort of respondent, i.e. Data Custodian or Data User. After the Eligibility Check and sorting question, Data Custodians were offered a minimum of 24 questions and Data Users a minimum of 18 questions. The majority of the questions required free text or long answer responses. A full version of all questions is available at Appendix 7.3.

### 3.4.3. Dissemination of questionnaire

Campaign monitor (https://www.campaignmonitor.com) was used to distribute an invitation email and link to the questionnaire. Email recipients were encouraged to on-send (snowball) the invitation to anyone they thought relevant or interested (see Appendix 7.4 invitation). The invitation email was distributed on 2nd October 2018 to 208 recipients.[3] Recipients were identified through the investigators' and AHRA networks and through Google searches to identify people working with primary care data in Australia. On 2nd November 2018 an invitation (and survey link) was sent to Chief Executive Officers of 29 Primary Health Networks (PHNs) by a representative of the Australian Government Department of Health. Lack of prior mapping of primary care datasets led to a broad approach being taken in the identification of potential participants. As potentially eligible participants were identified, additional email invitations were sent. Table 1 outlines the types of organisations that were included in the broad dissemination of the survey link. A reminder email was sent to 300 recipients on 29th October 2018[4] and the survey closed on 9th of November 2018.

### 3.4.4. Participant interviews

Thirteen survey respondents indicated that they would like to contribute more information via phone. All were emailed plain language statements and consent forms; phone / Skype interviews were completed with seven of these people (five did not respond to the request to interview and a suitable time could not be found to interview a 6th person). All interviews were undertaken by RC between the 8th and 22nd November 2018. The interview questions were based on the survey questions, enabling expansion of responses previously given and greater depth and insight in the perspectives and experiences of data users (see Appendix 7.5). All interviews were audio-recorded and transcribed, and the informants were given the opportunity to review their transcript. One audio recording was of poor quality, so a verbatim transcript was not possible.

[3] Campaign Monitor reported that 49% of the emails were opened and 29% clicked on a link. Campaign Monitor also indicated that some recipients forwarded their invitation. The top three forwards resulted in 316, 141 and 56 opens of one of the instances of the invitation.

[4] Campaign Monitor statistics show that the reminder email was opened by 37% of recipients and forwarded multiple time by the recipients.

**Table 1. Type of organisations that received the questionnaire link**

| | Recipients of invitations to complete the survey were associated with the following types of organisations |
|---|---|
| **1** | Aboriginal Medical/Health Organisations |
| **2** | Consumer representative organisations |
| **3** | Known primary care data collection agencies, organisations or related committee |
| **4** | Health data-related agencies and services |
| **5** | Health informatics networks, societies and associations |
| **6** | Health insurers |
| **7** | Primary care data collection software makers |
| **8** | Primary care-related medical councils, colleges, networks and peak bodies |
| **9** | Primary Health Networks (PHNs) throughout Australia |
| **10** | Relevant research organisations or institutes |
| **11** | State, Territory and Commonwealth Departments of Health |
| **12** | Tertiary Departments of General Practice / Family medicine (Medical Schools) |

### 3.4.5. Data analysis

The survey responses were exported from REDCap into an Excel spreadsheet and the data checked for completeness. The cleaned Excel survey dataset was imported into QSR NVivo 12 Plus [10] as a 'classification sheet' with codable (long answer questions) and classifying (demographic or set value response questions) fields. 'Cases' were made for each respondent and respondents' qualitative responses were auto-coded so that each were sorted under the relevant question. The qualitative and demographic data was linked to the survey respondents (cases). The long answer responses were thematically coded. All responses relating to identification of primary care data sets were coded twice to ensure list accuracy. The questionnaire contained 17 categorical response fields which were imported into SPSS Statistics version 26 [11] to generate frequency and contingency tables.

The Interview transcripts were imported into the same NVivo project as the survey responses, facilitating their thematic analysis by the interviewer-researcher. Analysis of the interviews was undertaken subsequent to analysis of the questionnaire data. The interviews provided descriptive excerpts to illustrate issues raised by survey responders and they also introduced additional themes which were incorporated into the analytical summaries of the results and are described in the results below.

## 3.5. Stakeholder workshop

The AHRA-facilitated four-hour *Primary care data for health care improvement and research: approaches to data quality assurance* workshop was held on 26th November 2018 at the Mercure Hotel George Street, Sydney. The workshop participants were expert stakeholders including government/agency employees or academics who may or may not have been data users or custodians but had reasons to be interested in primary care data collection and quality (e.g. policymakers). The workshop was facilitated by the investigators and a MACH Project Officer provided operational and resource support (see workshop agenda at Appendix 7.6).

The workshop was attended by 16 participants including one online via live-stream webinar from the Australian states/territories of Victoria, New South Wales, Australian Capital Territory. Participants were representatives from AHRA partner organisations, Australian Institute of Health and Welfare (AIHW), Primary Health Networks (PHNs) and other primary care data custodians. The workshop discussed the preliminary findings of the primary care survey and the trial implementation of the quality assurance framework. Small group sessions were utilised to gather feedback and consensus from participants on the following questions:

- What does primary care data quality mean to you?
- What are your key concerns about data quality for our future use of primary care data?

- How can Australia reach a consensus on an approach to primary care data quality?
- What would a minimum data quality standard look like or have in it?

Participants were informed that the opinions they expressed during the workshop could be used by the researchers and if they did not want their views to be considered that they were to state it was 'off the record' or not record it on paper. The opinions recorded on paper by each of the four workshop groups, for each of the four questions, were collated and returned to each of the workshop participants. Participants were asked to review the summaries and use tracked changes or comments (in Microsoft Word) to further contribute to or refine the summaries of discussion points. No changes were received. The workshop participant summaries are collated and summarised at section 4.9.1; along with the survey and interview findings, they were used to inform the Recommendations (section 5).

## 3.6. Ethics approval

Approved by the General Practice Human Ethics Advisory Group, The University of Melbourne: ID 1852055. The downloadable plain language statement was included with the survey explaining the study and consent to participate, and interview participants were provided with a plain language information statement and consent form. The consent information for both survey and interview participants included acknowledgement that the study gathered information from a small number of experts in the area of general practice data so it may not be possible to guarantee participant anonymity despite researchers not revealing the identity or personal details about participants.

Workshop participants received a 'Workshop Participant Research Information Statement' that explained the conduct of the workshop and the data collection component of the workshop. Participants who choose not to consent could do so by not contributing to workshop notes (butcher's paper group discussion). No attendee chose not to contribute.

# 4. Findings

## 4.1. Participant characteristics

### 4.1.1. Survey respondents

Some 137 people attempted or 'looked' at the survey between 2nd October and 9th November 2018. The snowball sampling technique made reporting a response rate not possible. Of the attempts, 62 eligible surveys were received, 17 responders marked themselves as ineligible and so exited the survey at the initial eligibility check question,[5] and 58 responders provided insufficient or no data in their survey response – 21 of whom, from a variety of organisation types, requested a summary of results.

Of the 62 eligible questionnaires received, 32 (51.6%) responders branched to the 'Data User' questions and 30 (48.4%) to the 'Data Custodian' questions. All respondents, regardless of type, were asked about their awareness of other general practice or primary care datasets, building capacity of primary care datasets and demographic questions. Table 2 summarises the jurisdiction of respondents. One respondent indicated they were 'national', which may refer to their affiliation with a global health group.

**Table 2. State or territory of survey respondents, N=62**

| | ACT | NSW | NT | QLD | SA | TAS | VIC | WA | Elsewhere (National) | Not stated | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Custodian** | 3 | 10 | 0 | 5 | 1 | 0 | 9 | 1 | 1 | 0 | **30** |
| **User** | 4 | 8 | 3 | 3 | 2 | 2 | 8 | 1 | 0 | 1 | **32** |
| **Totals n (%)** | **7 (11)** | **18 (29)** | **3 (5)** | **8 (13)** | **3 (5)** | **2 (3)** | **17 (27)** | **2 (3)** | **1 (2)** | **1 (2)** | **62 (100)** |

Table 3 shows that the number of respondents affiliated with educational/research institutions was the same as the number affiliated with Primary Health Networks (PHNs) – both had 24 responders (combined 77.4% of the sample). About half of all Data Custodians were affiliated with PHNs (50%), followed by universities/research institutes (27%), whereas half of all Data Users were affiliated with research institutions (50%). Six researchers are also GPs (not accounted for in Table 3).

**Table 3. Types of organisational affiliations of respondents, N=62**

| Organisation | Total n (%) | Data Custodians n (%) | Data Users n (%) |
|---|---|---|---|
| **Educational or Research Institute *** (including universities) | **24** (38.7) | 8 (26.7) | 16 (50.0) |
| **Primary Health Network** | **24** (38.7) | 16 (53.3) | 8 (25.0) |
| **Government** | **6** (9.7) | 1 (3.3) | 5 (15.6) |
| **General Practice*** | **3** (4.8) | 2 (6.7) | 1 (3.1) |
| **Pharmaceutical** | **1** (1.6) | Nil | 1 (3.1) |
| **Health insurer** | Nil | Nil | Nil |
| **Other** (incl. software developer, non-Government/non-University data holder) | **3** (4.8) | 3 (10.0) | Nil |
| **Not stated** | 1 (1.6) | Nil | 1 (3.1) |
| **Total** | **62 (100)** | **30 (100)** | **32 (100)** |

* **Note** that six respondents from research/education institutions also nominated that they worked in general practice. The 3 GPs noted here worked only in general practice.

[5] 'Eligibility check' question: Have you used, or do you have responsibility for, a primary care dataset that is used for secondary purposes. 'Secondary purposes' was defined to included research, audit, surveillance and quality improvement.

The majority of respondents had accessed between one and five Australian primary care datasets for the purpose of secondary data use, and six had accessed 11 or more (mode = accessed 2 such datasets). Three Data Custodians had never accessed a primary care dataset for secondary purposes. Data Custodians can have oversight roles, being ultimately responsible for a dataset managed by their staff or associates (Table 4).

All but four respondents first accessed a primary care dataset for secondary use from the year 2000 onwards (data missing from five Data Users and from 10 Data Custodians). One Data User first accessed a primary care dataset prior to the year 2000 (in 1986) and three Data Custodians first accessed between 1983 and 1995.

**Table 4. Number of different Australian primary care datasets accessed, N=62**

| | None | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11-20 | 21-30 | Missing |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Custodian** | 3 | 2 | 3 | 4 | 4 | 4 | Nil | Nil | Nil | Nil | Nil | 5 | Nil | 5 |
| **User** | Nil | 7 | 7 | 5 | 3 | 5 | Nil | Nil | Nil | Nil | Nil | Nil | 1 | 4 |
| **Totals n (%)** | **3 (4.8)** | **9 (14.5)** | **10 (16.1)** | **9 (14.5)** | **7 (11.3)** | **9 (14.5)** | - | - | - | - | - | **5 (8.1)** | **1 (1.6)** | **9 (14.5)** |

### 4.1.2. Interview participants

The seven interviews were completed between the 8th and 22nd of November 2018, averaging 42 minutes in length (range 24 to 55 minutes). The interview participants were a mix of Data Custodians (n=4) and Users (n=3) from the Australian states of Victoria, New South Wales, Queensland and Tasmania. They represented four different universities, a government health department and two Primary Health Networks (PHNs). Six of the seven interviewees had accessed between one and three primary care datasets for secondary use and one had accessed between 11 and 20. Approximate year of first access of a primary care dataset for secondary use was between 2005 and 2018 for six participants and in the 1990s for one.

## 4.2. Identifying primary care datasets

Data Custodians were asked to name the datasets they had responsibility for and/or where they were located, and Data Users were asked what primary care or general practice datasets they had used for secondary purposes (see the questions in the box below). To protect anonymity, Custodians had the option to describe where their dataset was located and characteristics of it rather than name it – but all named their datasets. All questionnaire respondents were asked to list 'other' general practice or primary care datasets that they were aware of.

**Survey questions asked to identify primary care datasets**

1. *Have you used or do you have responsibility for a primary care dataset that is used for secondary purposes? (including research, audit, surveillance and quality improvement).*
2. For Data Custodians: *What is the name of the dataset you are Custodian of or have responsibility for and/or where is it located? If responsible for more than one, please list all.*
3. For Data Users: *What primary care or general practice datasets have you used for secondary purposes? Please list them and how you gained access to them.*
4. *Please list other general practice or primary care datasets that you are aware of (you may or may not have used them).*

The table at Appendix 7.7 summarises the **106** datasets identified by Data Custodians and Data Users and information that was provided about them such as the localities of the majority of the data, purpose/nature of use and whether the dataset had been linked to other datasets. Information was missing on the nature of secondary data use associated with specifically named datasets. Lack of a check mark in the 'Audit', 'Surveillance', 'Research/Trials' boxes does not mean that the dataset is not used for any of those purposes, it means that the information was not provided by respondents. The datasets most commonly mentioned (by three or more respondents) are outlined in Table 5.

The 106 named 'primary care datasets' included a varied range of datasets indicating wide understanding among Data Users and Custodians of what a primary care-related dataset used for secondary purposes is. Respondents nominated datasets held by, but not limited to:

- government and government agencies *(e.g. Medicare Benefits Scheme [MBS], Pharmaceutical Benefits Scheme [PBS], registry e.g. Australian Immunisation Registry, PHN collected, Australian Bureau of Statistics data, My Health Record, patient reported outcome and experience measures [PROMs and PREMs])*
- Universities and research institutes *(e.g. BEACH, ACCEPt, ASPREE, MABEL, Patron, STAREE, 45 and up study)*
- primary care providers *(e.g. GPs - used for practice audit, reporting to PHNs, sharing with researchers)*
- Aboriginal Community Controlled Organisation held and Aboriginal Medical Services (AMSs)
- community services and community mental health agencies
- Network of Alcohol and other Drugs Agencies data (NADA)
- general practice clinical software vendors *(e.g. MD Heart, Pyefinch dataset [Best Practice])*
- health workforce agencies *(e.g. Australian Health Practitioners Regulation Agency [AHPRA], GP workforce statistics)*

**Table 5. Commonly named primary care datasets used for secondary purposes**

| Number of mentions | Name of primary care dataset used for secondary purposes | Jurisdiction |
|---|---|---|
| **>20** | Primary Health Network collected data (individual datasets held by PHNs) | National |
| **20** | NPS MedicineInsight | National |
| **11** | BEACH (Bettering the Evaluation and Care of Health) data (1998-2016) | National |
| **11** | Outcome Health and POLAR data | NSW, Victoria |
| **11** | Medical Benefits Scheme (MBS) data | National |
| **11** | Pharmaceutical Benefits Scheme (PBS) data | National |
| **9** | PHN related Primary Mental Health Care Minimum Data Set | National |
| **7** | Patron primary care data repository / Data for Decisions (University of Melbourne) | Victoria |
| **4** | Aboriginal Community Controlled Organisations / Aboriginal Medical Services | National |
| **4** | Australian Institute of Health and Welfare (AIHW) held data (**in addition** AIHW was mentioned in relation to access to other AIHW held datasets such as PBS, MBS) | National |
| **3** | Australian Immunisation Register (AIR) | National |
| **3** | University of NSW ePractice-Based Research Network data | NSW |
| **3** | Medical Director (clinical software vendor held data) | National |
| **3** | 10% MBS and PBS sample data (noted by respondents as no longer available) | National |
| **3** | Patient Reported Experience Measures (Australian Bureau of Statistics) | National |
| **3** | My Health Record | National |

NSW = New South Wales

PHN datasets were most commonly mentioned – which is not surprising given that 39% (n=24) of respondents were affiliated with PHNs. The majority of PHNs collected data directly from EMRs using the Pen CS tools (www.pencs.com.au). The single dataset (as opposed to the many individual datasets held by PHNs) most named was NPS MedicineInsight. It is a dataset used for audit, but also available for approved research. BEACH data was also commonly nominated, although like the other datasets collected specifically and primarily for research purposes, it is not strictly a general

practice/primary care-related dataset being used for secondary purposes because BEACH was collected between 1998 and 2016 for the primary purpose of research.[6]

## 4.3. Characteristics and purpose of the datasets

Data Custodians were asked to describe the characteristics and purpose of their dataset(s). Characteristics were mostly conveyed as the data being de-identifiable, identifiable or aggregated, delivered by a third party, and as 'mapped' data (e.g. mapped to SNOMED-CT-AU, WHO ATC Drug Classification System). In two instances the ability to perform data linkage were described, but as indicated in Appendix 7.7 there were others that also were able to link to other datasets. Table 6 summarises the given purpose of the dataset(s) as described by 29 Data Custodian responders, by responders' organisation type.

**Table 6. Purpose of dataset as described by Data Custodians, n=29**

| Organisation type | Purpose of the secondary dataset is / is to… |
|---|---|
| **Primary Health Network (PHN)** | • build capacity of general practices to use their data to inform practice improvement, e.g. for accreditation, better clinical and business outcomes<br>• inform the activities of PHN:<br>  ○ Region-level quality improvement / planning: aggregated use of epidemiological, demographic and service utilisation indicators<br>  ○ Practice-level planning for population health and prevention (needs assessment)<br>  ○ Individual patient level surveillance for reporting back of high-risk patients/patients at risk to practices (incl. mental health care) → intervention for improved patient health care and prevention services<br>• identify opportunities for data quality improvement<br>• report to the Department of Health<br>• identification of practices to participate in clinical service projects<br>• enable some approved research |
| **Education or Research Institution** | • provide reliable general practice data that meets the needs of information users<br>• link GP and cancer hospital data to examine patterns of care across the continuum of care<br>• collect longitudinal data – including of patient GP presentations and medical workforce<br>• provide specific disease surveillance<br>• study and advocate for specific population cohorts<br>• enable use of de-identified GP EMR data for research, policy, teaching |
| **Government** | • monitoring and surveillance including of at-risk population groups |
| **General Practice** | • enabling secondary use by sharing data with My Health Record and PHNs<br>• audit<br>• research<br>• patient care (listed, but this is not secondary use) |
| **Other** | • support quality improvement in primary care<br>• post-market surveillance of medicines<br>• approved research<br>• record and classify general practice encounters (i.e. a vendor clinical practice software)<br>• reporting to Department of Health, Medicare |

Interpretation of the purpose of the dataset, determined from long answer responses to a survey question has limitations such as analyst inability to question the responder on the breath and inclusiveness of their response. Some responders gave detailed descriptions of the primary and secondary purposes for which they use the data, others gave high-level responses – such as 'quality improvement' which can be a label used to capture various activities. The PHN

[6] Between 1998 and 2016 the BEACH study compiled a database of annually collected general practice data from approximately 1,000 GPs who completed details of about 100 consecutive patient encounters on structured paper encounter forms [12, 13].

responders mostly described their organisations as holding datasets to enable quality improvement, and some also described the purpose as including population health and activity planning. Audit, monitoring, surveillance and evaluation are all activities undertaken to enable quality improvement.[7] The structure of PHNs means that each operates independently so that they can "generally end up with things that are relevant to the local context rather than having a product imposed from the top down" (Data User interviewee 6). Just one PHN responder indicated that "data quality improvement" was a goal of use of the data. Education / research institutions mostly described dataset purpose in terms of their research purpose.

Some of the PHN Data Users and two of the PHN Data Custodians indicated that they used their organisational dataset for research projects: "it can be used for approved research" and the "PHN uses this data for… some research type projects". It was not clear whether PHN data related research projects could include external researchers accessing the data. A PHN-employed participant described that their EULA (End-User License Agreement) had a research clause, but also that review of their data sharing arrangement indicated lack of governance structure around data use within the PHN and data sharing with external researchers was unlikely before internal governance structures were strengthened.

## 4.4. Cost of accessing data

As displayed at Table 7, 20% of Data Custodians (n=6) reported a fee for access to their dataset, and 50% (n=16) of Data Users indicated that they had to pay or provide something in return for access to data. Of these 16, 11 reported paying fees for data access and the remaining five provided another kind of payment or service such as providing the Data Custodian with data analytics or other feedback, or assisting to recruit general practice data providers on behalf of the Custodian (in return for access to the resulting curated dataset). Several who paid fees also outlined their requirement to provide the Data Custodian with progress or annual reports and copies of publications related to use of the data.

Twelve of the 16 PHN-associated Data Custodians indicated 'No' to the question: "Do you get anything from people/researchers who get access to the dataset (for example a fee, service, report or other information)", but few allowed data access to anyone outside of their PHN. Across the *non*-PHN Data Custodians, fee for access varied depending on the complexity of the data request, the type of data recipient (student, academic, commercial, government) and number of years of data provided. Some required co-authorship of output in exchange for their in-house analyst assistance, or individualised reports for general practices providing data.

### Table 7. Access costs for primary care data

| N (%) | Custodian charges for access* | User pays for access † |
|---|---|---|
| **Yes** | 6 (20) | 16 (50) ‡ |
| **No** | 21 (70) | 12 (38) § |
| **Not sure** | 1 (3.3) | 4 (12) ¶ |
| **No response** | 2 (6.7) | nil |
| **Total** | **30 (100)** | **32 (100)** |

*** Question asked to Data Custodians:** Are people who access the data required to pay or provide something in return for access?

† **Questions to Data Users:** Did you have to pay and/or provide something back to the data custodian in order to access the data?

‡ **Note:** 11 of 16 respondents reported paying a monetary fee in return for data, 5 'paid' by providing something else in return.

§ **Note**: 5 from Education/Research institutions, 4 from PHNs, 1 government, 1 general practice, 1 from pharmacy.

¶ **Note**: 2 from Education/Research institutions, 2 from PHNs

Most respondents did not list the fee paid or charged, but those given ranged from no fee, $80 flat fee, $100 per randomised participant to $16,000, $20,000 or $30,000, with the fee sometimes being in addition to Data Custodian co-authorship of manuscripts. The Data Users who paid a fee were associated with research/education institutions (n=9),

[7] PHNs were established by the Australian Government with the key objectives 'of increasing the efficiency and effectiveness of medical services for patients, particularly those at risk of poor health outcomes, and improving coordination of care to ensure patients receive the right care in the right place at the right time.' (www1.health.gov.au/internet/main/publishing.nsf/Content/PHN-Home, 14 July 2020). 'Quality improvement', an organisational objective of PHNs, was given by many PHN responders as the purpose of their dataset.

followed by government (n=4), PHNs (n=2) and one not stated. One Data User commented that: "by the time you analyse the data, the sum will have inflated by a few $1000. Products offered by (data provider blinded) usually have an inflation rate similar to that in Venezuela."

Among Data Users who did not (n=12) or were not sure (n=4) if they paid a fee for data access, seven were affiliated with research/education institutions, followed by PHNs (n=6) and one each from three other categories. Those from the research/education institutions all indicated in their written responses that they used the data as part of their job or study. It is likely that for many some form of payment would have been made by the institution for the acquisition or use of the named datasets – for example, several were accessing extracts from NPS MedicineInsight or the BEACH datasets which normally are associated with fee for extract. One of the PHN responders highlighted the problem of "the increasing vendor costs of extraction data tools … (and that) enhancements and extended tool applications are costly".

## 4.5. Data linkage

Data linkage was widely referred to among participants, especially related to increasing the utility or benefit of secondary use of primary care data (see 4.6 and 4.7.2); no one suggested it a limitation or to be avoided, but many barriers were outlined. The problem of "not being able to link datasets to review patient journeys" was frequently raised. Barriers and enablers (see 4.8) included technical issues (lack of/need for reliable IDs for data linkage), governance, capacity, trust and knowledge building – for example, lack of stakeholder agreement and education around the population benefits of linked data. Western Australia was referred to as "the nirvana of data linkage". Some participants were seeking data linkage as a means to provide improved care, monitoring and services for an individual; linkage of identifiable data was referred to as the "holy grail".[8] While academics and policy-maker respondents appeared to be clear about the purpose of data linkage, a PHN interviewee suggested that in their jurisdiction the purpose of data linkage was not always clear, partly because linkage of identifiable data is not an option; so purpose should be clear before it is undertaken:

> *You need a purpose. There is so much data in general practice, and if you want to intervene data linkage is where your best chances are. Linking it to the hospital, you might get a few more diagnoses coming in, but I don't know the value … It still comes down to: What is the intervention in general practice? And that is where we are trying to use the data to inform that. (Interviewee 7)*

Table 8 provides insights from interviewees about data linkage. Some participants discussed data linkage in terms of linking disparate sets of health data whereas others were seeking cross-sectoral data linkage. Barriers to data linkage included organisations/government departments or agencies not wanting to share tranches of data even between one department and another, and the long time it can take to gain access to datasets and determining cost of access.

### Table 8. Perspectives on data linkage

| Themes arising from interviewees |
|---|
| **Linkage of health care datasets:**<br>*We certainly need better data linkage, don't we? Trying to have a look at what happens to an older person through the system, it's improving, but it's very difficult. How many services they access, hospital admissions, transition to nursing home, looking at predictors of those things. A whole lot could be done for people if we could build up profiles of risk factors, and that would be better for the (health) system too. You can't do that until you have a more complete dataset… So data linkage is certainly a big area. (Interviewee 7)*<br>**Cross sectoral data linkage:**<br>*I like data linkage … I've been talking a lot with the Centre for Victorian Data Linkage (CVDL) … Where there are complex systems within multiple organisations, like health, justice, education – all these different systems that run separate things – I think it's really important to understand a person's journey through those different systems. I think the only way you're going to do that is your data linkages … (It's needed) to make decisions about policies around certain subjects, and how you deal with those populations … without that, they're (policy-makers) just going blind. (Interviewee 2)* |

[8] The use of dbMotion™ was suggested by an interviewee as a way to get around data linkage privacy issues for individual patients, using identifiable data related to care plans. This, however, is out of scope of this study which is exploring secondary rather than primary use of patient data.

**Lack of uniform approach and unwillingness to share data:**

*(Health data linkage) at a real high level across the country would ideal, because everyone is covered. But as it is now it's case by case and organisation by organisation, and it's all: 'Do you want to share?' And they say: 'No' or 'Under these conditions' so it's an ongoing battle to get the information you need. (Interviewee 6)*

**Long lead-time to access data and uncertainty of cost:**

*One of the big challenges is, you're typically given a timeline for a (research or evaluation) project - you've got a budget and timeline that you need to stick to. But the people that are providing the data - it's often down the road that you find out it's going to take a bit longer than expected … and it's often hard to get a fix on what it's going to cost. (Interviewee 4)*

*Each tranche of data to link is about $10,000… I think it took about eight months to get, which is not too bad considering the stories I've heard … (my PhD student) waited three years for data on immunisation of post-code level from the Health Department. (Interviewee 1)*

**Lack of resourcing and expertise limiting access and use of government held data:**

*I think a big limitation (of data linkage) is data sharing within the state. Trying to link through Department of Health to Department of Education and Department of Justice is a really tough thing to do. If Data Linkage (Centre for Victorian Data Linkage) can get that overall approval to use that data, it might expedite data sharing a bit better. But the issue is that Data Linkage is such a small department for such a massive need. A data linkage project would take, probably, six months, to get data out. The other issue is … they expect that the requester has the ability to analyse that data. I would say less than 1% of DHHS people have that skill. So, while I think data linkage is good and it's a really valuable tool, it's not really designed, currently, to allow a policy person or a manager, at the DHHS level, to be able to use that data meaningfully. Half the time, they don't even know what they're looking for. Data Linkage will need you to state the variables from each of the data sets that you need to extract. And, I would say that most DHHS people don't even know what those variables are, or even know where to look for it. (Interviewee 2)*

**Lack of provider-patient data linkage:**

*So you have provider level data, hospitals, and doctors, but I want to get the patient-level data to provider level – is actually quite hard in existing data sets … We collect lots about patients but if you can't link it back to providers, then it's actually really hard to look at anything like medical practice … The data in healthcare it's being pushed forward in data linkage, it's all the patient level, it's not at the provider level … I think if you want to change anything you've got to understand doctors' behaviours and if you haven't got that level of data, aggregated to that level, then it's not going to change in terms of health policy and stuff like that. (Interviewee 1)*

## 4.6. Better use of primary care data to improve health outcomes

The strongest narrative arising from the combined comments of the 53 Data Custodian and Data User responders to the question: "How do you think that primary care datasets could be better used to support improved health outcomes?" was that they could be better used by employing data linkage coupled with appropriate research, use the research evidence to demonstrate needs and opportunities, then act on the new knowledge to deliver appropriate services. Appendix 7.9 summarises the responses of the 53 responders, grouped by the type of solution suggested – the solution 'types' (not mutually exclusive) are listed below:

- Research solutions
- Data linkage
- Technical data solutions
- Surveillance and monitoring
- Inform/support general practice
- Health promotion solutions
- Policy, governance and system change.

These 'solutions' should be considered alongside the 'enablers' of better use of primary care data' identified and outlined at section 4.8.2. Responses to this question tended to point to on the ground level 'solutions' such as specific types of research, health promotion, monitoring and surveillance (such as using data to monitor antimicrobial use to reduce antimicrobial resistance, or to identify the top five dispensed drugs for chronic conditions, thereby identifying

the top five chronic conditions and promote health projects to improve those conditions); whereas the enablers at section 4.8.2 tended to be conceived as higher-level interventions.

Technical solutions to support data linkage were suggested by many, including the need for greater standardisation of data and use of unique person identifiers to make data linkage more straightforward. Data governance and access issues raised the need for strong governance structures and approval processes and improved access to primary care data (especially data linked to hospital data and other services). Some responders suggested the need to have all general practices contribute data to a single dataset, even if just core fields, to create a 'national minimum dataset', others focused on support at the general practice level; e.g. using data to provide more clinical decision-making guidelines and educating primary care clinicians and staff in how datasets can be used with the aim to raise awareness of the importance of quality data. One interviewee suggested that more provider/clinician-level data is needed to improve health outcomes, because it is though understanding doctors' behaviours that changes to health policy and then health outcomes can be achieved.

Compulsory sharing or working to increase population acceptance of sharing of health data was suggested as a way to make better use of general practice data for population health gains. The data sharing systems in the Republic of Korea and Denmark were given as examples that Australia might follow to trigger 'dramatic' and positive change:

> *All people in Korea opt into having their data collected for health, when they use the health system. The Korean government extract all that data regularly and they do massive studies around the whole population of Korea… that's very useful for them because they can plan at a national level. That's also what they do in Denmark, they collect all your data, locally, to provide that same level of information. I think that's something that would be really powerful, in terms of Australia's needs, if they were able to do that and protect privacy at the same time. I think it would be amazing. I think the whole system would change dramatically. (Interviewee 2)*[9]

Primary care was described as "the core of where the (health intervention) activities are" so its use can only improve outcomes. With money spent "in niche areas" it was described as "crazy" when monitoring and evaluation are not undertaken, often because data are incomplete due to lack of compulsory sharing of data by GPs. "So much money gets poured in and, you ask them, 'How did you evaluate it?' They don't evaluate it. It's crazy!" One interviewee described their vision for using primary care data to improve health outcomes, from building infrastructure to bringing the community along on the journey and adopting the example of the Scotland:

> *My vision is about building the infrastructure and then the source of questions that could be answered, I think, is multitudinous. So, health data that is collected and coded correctly, improving the software so that it makes it easy, linking it, being able to follow the patient journey, having absolute best practice in terms of security, and then having a really good process for ensuring that sensible questions, that are answerable, are asked of the data with an adequately skilled workforce to do that. That could produce very powerful outcomes in terms of the community getting to know itself better, understanding its issues better, being part of the conversation, which we need to have about bringing everybody along with the hope to improve health outcomes. Health outcomes writ large. It's not about just health. It's about linking with education, linking with correctional services, linking with housing, linking with all of those databases needs to happen. And that isn't a conversation that seems to be happening, it's all just about the health databases at this stage. So, take Scotland as our example and move on for God's sake. (Interviewee 3)*

## 4.7. Benefits and limitations of secondary data use

### 4.7.1. Limitations

Participants' responses to the question: What do you think are the limitations of secondary use of primary care datasets? varied from "too many" to "there are no limitations". Australia was described as "so far behind" other countries (such as Scotland, England, Netherlands) with regard to secondary used of primary care data; lack of an individual identifier

[9] Note that the Danish General Practice Database (DAMD) was terminated in 2014 after reinvestigation of the legal basis for collection of the data was deemed not to meet the country's legal definition for clinical databases which limits such to disease-specific registers. [14] https://www.oecd.org/health/health-systems/Primary-Care-Review-of-Denmark-OECD-report-December-2016.pdf

for person tracking was said to contribute to this. Using data for purposes other than those which they were collected was suggested by some participants as a fundamental limitation:

> *The biggest limitation is calling and purposing the primary care data as 'the secondary use of data'. The data should be collected to use in one way or another way, and there is no such thing as 'primary use' or 'secondary use' of data*. (Data Custodian)

It was suggested that the term 'secondary' data use should not be used at all because: "the integrated data strategy[10] allows mute uses of data, cognisant of each context. So, the limitation is continuously calling it 'secondary use'." The example of BEACH GP data[11] was given as not having this limitation because "it is not really 'secondary analysis' since the purpose was to set up a reliable and valid database for such research."

The suggestion that general practice electronic medical record (EMR) data should not be used for any purpose other than to provide clinical care represented a minority voice, while acknowledging limitations most participants outlined limitations arising from the data being used for a purpose other than that which it was collected, but that in itself was not reason to not acknowledge and work to overcome those limitations. The limitations suggested by 52 respondents, are represented by the following themes; Appendix 7.9 expands on these:

- Poor data quality, reliability and technical limitations
- Poor data utility
- Lack of understanding of the data complexity and its context
- Unequal representativeness of data
- Difficult to access
- Privacy concerns, trust and ownership (an access issue)
- Lack of guidelines, policies, standards and 'common data model'

Poor data quality was emphatically emphasised by one survey respondent who complained about the 'missingness' of data and pervasive use of free text:

> *What's important is that it [accessing primary care data for secondary purposes] is a monumental waster of time. Data cleaning is high cost and low value. The lack of data quality puts any research conclusions on VERY shaky ground; it is often an exercise in futility given the 'missingness' of data and the high volume (pervasive) use of free text narrative entries.* (Data User survey responder)

The complex social and relational environment impacting on GP recording of diagnoses, which impacts data quality, was acknowledged:

> *(There is) a total lack of understanding of the primary care context. I get people bagging me out all the time: "Why don't these GPs just write it down? What's the matter with them!" … (For some sensitive conditions) GPs, don't actually write down that diagnosis. … When I said to the GPs: "Why is that?" they said: "I don't write it down because people don't like us to write it down" or "the relatives don't like it" … In some cases, they've sent them to the geriatrician and the geriatrician put them on anti-dementia drugs… The geriatrician doesn't like to write it down because they send a copy of the letter to the patient. So the (GP) said: "I haven't written it down because the geriatrician didn't write it down and I'm not sure" … Dementia is a highly stigmatised and tricky to diagnose condition… there'd be others in that category … Also, everything goes up into the cloud now, so that's another problem. The patients might not want that up in the cloud, up in My Health Record … I've been deleting things all over the place, that patients don't want anyone to know about it; mental health mainly. (Interviewee 5)*

The lag time to access data, particularly after an intervention and using dated data and aggregated data for planning, were described as strong limitations – particularly problematic if funding for a service/intervention is tied to its evaluation:

> *If you put a lot of energy into putting an intervention into place and there's a long lag time (to get data for conducting evaluation) it could be potentially a year after the official intervention has ended before you*

[10] The respondent did not elaborate on what is the integrated data strategy.

[11] The BEACH study purposefully collected general practice data for research between 1998 and 2016 https://www.sydney.edu.au/medicine-health/our-research/research-centres/bettering-the-evaluation-and-care-of-health.html

*can come out with some solid and conclusive results. And sometimes, the health services may struggle to sustain that service in the absence of the evidence. So the closer to real-time that that data can be provided the better… Where we're delivering interventions and where there's stakeholders and vested interests, you've got to maintain that momentum. (Interviewee 4)*

*Using out of date data to do planning, using aggregated data to try to help individuals. Out of date data unable to identify hot spots. Hotspots (aggregated) at postcode level, whereas it's better to get data at practice level. (Interviewee 7)*

### 4.7.2. Benefits

There were more commonalities between survey respondents in their ideas on benefits of secondary use of primary care data than there were about its limitations. The predominant theme arising was related to benefits gained from the generation of new knowledge and evidence; one Data User said that secondary use of primary care data "fills a big hole in available data about health and health care". Many of the benefits were noted in terms of the *possible* opportunities and the *potential* for benefit, with some stating caveats that the benefits could be realised only "as (data) quality improves" or "if the primary care datasets are of decent quality".

*The potential is considerable. As quality improves, GP data has the potential to provide greater population health information than is available in any other source because virtually everyone over a 1 to 5-year period visits a GP. Furthermore, linkage to other datasets provides the opportunity to study outcomes and evaluate interventions across the whole population and target many groups. If conditions are well recorded, then GP data will also potentially deliver accurate population-based prevalence of a whole range of health conditions linked to interventions and outcomes.* (Data User survey respondent)

The opportunities and potential of "rich", "real life" primary care patient-related data was described as "endless", offering "massive benefits for everything we do". It was suggested that "the creativity of researchers in the use of these data usually produce benefits that no one has thought of"; that secondary use of primary care data is "absolutely vital to addressing rising risk health profiles in Australia", that it can "shed light on aspects of primary care that we would otherwise know nothing about" and that it "will become more and more relevant as healthcare costs lead to a greater emphasis on wellness and prevention." The **benefit of data linkage** was referred to repeatedly and was considered a pre-requisite for the fruition of the far-reaching population health benefits. No one suggested that there were no benefits (data missing from n=9).

Appendix 7.11 summarises the benefits suggested by respondents, grouped by benefit type (also see below). Many of the benefits were ultimately related to enabling improved healthcare and health outcomes. The themes discussed are interrelated and can be broadly described as follows:

- **Potential improvements to provision of care and health outcomes:** Secondary use of primary care data (especially when linked to datasets from other services and sectors) leads to enhanced understanding of population health and service needs which informs evidence-based policy, planning and service delivery to improve efficiency of the overall system, better meet the needs of local populations and so improve health outcomes.
- **Direct and pragmatic research benefits:** Enabling statistically powerful and cost-effective research (no need to allocate resources to recruiting individual participants) among large and representative population cohorts, the data itself being able to generate new research questions, and also contributing to cost effective evaluation of intervention outcomes.
- **Assist with policy and planning related to service provision**: Enable monitoring and evaluation, risk stratification and management, predictive modelling to inform preventive care, needs analysis, tracking of disease outbreaks, workforce planning, competitive benchmarking, leading to evidence-based investments and decision-making around service provision and informing national policies related to health needs of communities.
- **Practice-level benefits:** Enable general practices to review their activities and make business improvements.
- **Intrinsic benefits of the data:** Unique, rich, granular, 'real world', large dataset with extensive population representation that minimises measurement bias in research. Linked data provides a systems view that makes known the unknown.

- **Benefits of possible technical applications:** Data extraction tools have potential for add-on patient-generated data collection through linked apps.

## 4.8. Barriers and enablers of secondary data use

### 4.8.1. Barriers

Questionnaire respondents were asked to outline what they considered to be barriers to better use of primary care data. In some instances, barriers described were similar to described in limitations (e.g. data quality issues), but overall, many more systems and motivation-related barriers were identified. For example, one survey respondent stated the main barrier to be: "a confused 'secondary use' sector who persist in thinking that the data should be provided for secondary purposes, when the clinician ONLY sees value in (minimal) data for primary care delivery purposes; its cart before the horse)" – this comment was an outlier in its type. Barriers variously described by many were succinctly summarised by one: "The main barrier is privacy concerns followed by technology and then data quality". The responses of the 53 responders are summarised at Appendix 7.12; thematically the barriers are categorised as:

- Fear and lack of trust
- Lack of leadership, governance & ethics constraints
- Lack of data availability
- Lack of expertise, experience & motivation
- Barriers to data linkage
- Technical systems barriers
- Health system and research barriers

**Fear, reticence and lack of trust** themes included fear of poor data security, "illegal use of data", government attempts to "control" GPs, and "privacy and reputational and financial damage to the organisation". Many who described privacy and confidentiality concerns linked them to fear. A PHN responder wrote: "Data breach legislation has practices hyper vigilant and wary of data sharing" and another highlighted that "raised concerns about litigation" have become a barrier to practices sharing data. "Unwillingness" and "reticence" of clinicians (and patients) to share data was another theme, which could be related to fear, but was also described as due to general practice not seeing value in secondary use of data. In discussing the possibility that GP vendor clinical software systems could be collectors of GP data, an academic GP interviewee expressed lack of trust, not only of the software vendors as Data Custodians, but also lack of trust of the PHNs and drug companies:

> *The data governance around all of this is scary. I think MedicineInsight do their best to get that right. I don't trust the PHNs to do it right; not because I think they're evil people, but they fumble in my state with everything, so how might they get that right? And I wouldn't trust the drug companies or the medical software producers as far as I could throw them. (Interviewee 3)*

Another interviewee outlined how fear of data being used out of context and potentially used against them was getting in the way of GPs sharing data:

> *I think there is a fear that the data could be used against the GPs. I mean, if you found that one particular practice had a disproportionate amount of mental health prescriptions or stuff like that … that information could be used against the practice and interpreted that way. Our (name withheld) LGA has a higher proportion of registered mental health clients compared to other LGAs across the state. Now, is that because there are more (people with mental health issues), or is it because the services are better and accessibility better? It's that kind of interpretation that you don't really know just from the data. I think there's a fear - a concern maybe from the GPs that the data could be used in a way that is out of context and incorrect. (Interviewee 6)*

Risk adverse Data Custodians (example given of government departments) were also said to lack trust and to fear reputational damage through sharing data:

> *I think it's (the main barrier to secondary primary care data use) trust. I think that's a big thing, of whomever is the Custodian of the data… For governments it's more about trust and they throw the privacy card at you, but it's not about that, it's not about risk of a breach. It's about risk of embarrassing a politician,*

*which is how bureaucrats' brains work. They just close it down. They want zero risks. ... But the benefits are greater than the risks. (Interviewee 1)*

**Leadership, governance and ethics constraints** were frequently cited barriers. For example, lack of national leadership, governance and regulation issues including confused determination or establishment of "who owns the data". Leadership issues were mostly described at a federal level, but lack of GP leadership was also suggested and lack of leadership around engagement with other stakeholders – such as GP clinical software vendors. The voluntary status of data sharing and lack of regulation that would make sharing of general practice data mandatory was described by some as a barrier to its better use. One respondent described "explicit client consent and strict access controls via stringent data governance protocols" to be the only 'real' barrier to use of primary care data, and another that "people aren't clear when you need consent and when you don't (referring to discrepancies between Office of the Information Commissioner versus RACGP guidelines).[12] A Data Custodian explained how lack of clarity around de-identification was a barrier: "We have got such a rich pool of data… I just think this is such a great opportunity for researchers, but then you run into that whole issue around concern about deidentification, and it gets hard again." (Table 9 provides a study of the barrier of appropriate deidentification). Others described lack of transparency around consent models and governance processes, including the need to negotiate cumbersome, slow, expensive processes for gaining ethics and Data Custodian approvals to access data in the first instance. One participant highlighted that as ethics committees become increasingly aware of the complexities associated with secondary use of data, that some uses are being "inappropriately constrained".

## Table 9. A study of issues around de-identification and consent

| Perspective of a PHN-associated interviewee on data deidentification and consent |
|---|
| ***To be honest, I think some of that [data de-identification], from what I have seen, is maybe a bit of a myth.*** *The risk of reidentification is the big nightmare. There is so much confusion. Once information is deidentified, it is no longer under the Health Act, the Privacy Act, but people then say you need consent to get deidentified information, and it just goes around in circles.*<br><br>*There is a lack of understanding around the privacy principles, they are not very clearly explained. When I contacted the Information Commissioner by phone … the advice was: 'If you put posters and brochures up' which we did 'and you have got the opt-out' which we do have 'then that would be enough.' I was pleased I verbally got that information, and then I said: 'Best get that in writing'. So I wrote and posed the same questions, and I got back this list of: 'If APP3 applies…' and it just didn't give me an answer, and more and more lawyers are getting involved in this stuff to get clear answers around what is acceptable. Also, the RACGP Standards talks about taking deidentified data and getting consent from patients to do that (which is at odds with the Information Commissioner).*<br><br>*Even in the Pen CAT, if you put enough filters on, you can find the individuals. So, is it deidentified anymore? I know it is a new space and nobody has got the answers, but when people don't have the answers, they tend to get a bit nervous and pull back from being a bit brave with saying: 'We will interpret it this way'. So the whole deidentification is just so grey. The whole consent from patients is very grey. The way practices collect consent is very variable and very grey. Some practices, if you are with that practice for 20 years you may not sign anything else that says you are happy to share your data. I can't get straight answers to this. So, we had to go back and back and back. The easiest thing is just to … get data from the GP, run it through the analytics and provide it back (and not do any secondary research or analytics for purposes other than feedback to GPs).*<br><br>*It would be nice to have some really clear steps about how you deidentify data, if you do deidentify it, and the levels and when you stop saying it is deidentified. What is the point when you say: 'Actually, it is not deidentified because it can be reidentified?' That is where I am struggling with developing a data set for secondary use … If we provide deidentified data to a researcher, we don't need consent, do we? Maybe, maybe not. People can't seem to come up with an answer. If you were going to do research with patients in practices, yes, you would need ethics and all the rest of it.*<br><br>***It would be nice if there was a document that articulated what the different levels (of deidentification) are, what it is recommended you need, and when is it okay to do one thing and not another.*** *(Interviewee 7)* |

[12] The RACGP were described by a study participant RACGP member to have "not got its house in order in terms of having unified, organised coherent policies" around secondary use of data across its research, eHealth and quality care committees.

**Lack of data availability** for secondary use was much cited. This barrier was related to prohibitive cost of accessing data, researchers not knowing what data was available nor where, sufficient quantity of general practice data not being shared, or when shared the "people who can establish the best outcomes" not having access to it (due to cost or other reasons). Lack of availability also related to lack of certain types of data and to missing data. Some respondents cited the barrier of lack of data to support priority populations (e.g. CALD, ATSI); this was associated with general practice "reluctance to complete key data points such as indigenous status" and because "Aboriginal Medical Services see sharing data with PHNs as both duplicative and none of our business". Data Custodian's refusing access was another barrier to primary care-related research:

> *I tried to get access to (general practice) data (that linked general practice to hospital data) but they (the Data Custodian) kindly declined. I was in a research program… my interest was the correlation between SA2 primary care and hospitalisation rates. I was looking to formulate a question for research but it was contingent on access to the data. (Interviewee 6)*

**Lack of expertise, experience & motivation** related both to analysts and researchers working with data and to clinical staff capturing data. It was highlighted that data analysts need to have some medical knowledge, familiarity with processes involved in collecting the data, and an understanding of the health care system so they can understand the limitations of the data and so avoid reaching inappropriate conclusions. Lack of general practice clinicians and staff knowledge and therefore motivation to collect clean, accurate and complete data was another barrier to its better use; it was said that general practices tended to lack understanding of the benefits of accurate and complete data for driving quality improvements and improved health outcomes. Lack of GP training on how to code and properly use their clinical data capture software (and the software packages not being adequately designed to enable capture of clean data) suggests knowledge, expertise and technical barriers. Lack of shared vision to motivate the building of a healthcare system that uses and learns from many sources of data (not just primary care data) was also raised. With secondary use of data not a GPs primary concern, it was said that: "many GPs are haphazard in their reporting and clinical notes" and that there was lack of drivers or incentives to encourage GPs to improve data quality (note that the data Quality Improvement Practice Incentive Payment [QI PIP - [5]] had not been introduced at the time of this survey in October 2018).

**Barriers to data linkage** were frequently raised as a barrier to better secondary use of primary care data, for example: "Inability to link data to get a fuller picture of the patient journey across sectors - particularly hospital presentations / admissions / discharge". The barriers to linkage were essentially related to governance and what was allowed: i.e. lack of mechanism to link patient-level data, and lack of agreement between stakeholders to enable data linkage (see 4.5).

**Technical systems barriers** related mostly to lack of standardisation of and/or interoperability between data extraction and/or GP vendor clinical software tools (and the limitations of datasets created from these tools) and to general data quality issues due to lack of data 'cleanliness', granularity. "Raw data would be preferable" was a comment from one PHN respondent, suggested to overcome the barriers related to inadequate data granularity and infrequent receipt of data. It's "MESSY data!" was the lament of one Data User: "It is not consistently collected e.g. different GP systems used, systems have not been set up to collect data for secondary use… (therefore) processing and preparing data for analysis can be a lengthy process". Lack of a minimum data set and variable definitions of what is a minimum data set, and the limitations of My Health Record, were also mentioned as barriers to better use. Limitations of GP vendor clinical data collection software systems also present barriers to better use of GP data. The following example is overcome by extraction of GP data for secondary use which can archive and flag changes to patient status, for example:

> *Currently, the two (main GP vendor clinical software) systems … tend to, one more than the other, override existing data so that history is lost sometimes… So, for instance, just a small one is smoking. If you ran a smoking cessation program and a lot of people stopped smoking you would lose the fact that they were originally smokers. They would just get overridden as a non-smoker (whereas our system would store the fact that they smoked, and this was the day they stopped). (Interviewee 7)*

Another found a commonly used commercial data extraction tool and their associated legal and data governance arrangements to be a barrier to better use of primary care data, a barrier not faced when using data extracted using a university-developed tool:

> *They (the PHN) now have a QA (quality assurance) process … and, that's how they've discovered that (corporate data extraction tool company) is a terrible company. They don't care about the data. They are just concerned about having contracts and getting paid money. That's a very different thing to (university*

*developed data extraction tool) which has a really strong public health output. The (corporate) system is not a system that I would ever recommend to anyone. It's also really difficult to change things within the system, (if) the PHNs have already agreed on data (the variables to extract and how they will be collated, then) that's already set in stone, you can't change that now within that contract. (Interviewee 2)*

**Health system and research barriers** preventing better use of primary care data included the small business structure of general practice, patient freedom to attend any practice (and therefore, in absence of data linkage, limiting longitudinal patient records), the system of voluntary data sharing meaning that no complete primary care dataset exists, a workforce within general practice that tends to have high staff turnover which can negatively impact motivation or ability to ensure consistently high quality data capture, lack of timeliness of data release, PHNs' focus on reporting to government agencies (Australian Institute of Health and Welfare and Australian Bureau of Statistics), and **systematic lack of funding to collect data from general practice, analyse it and to support implementation of findings** arising from it. "Clinicians are NOT our data entry clerks" was the message from one PHN data user. Lack of funding for resources to motivate and educate (build capacity of) the general practice to prioritise data input for better data quality, accuracy and completeness, and funding for the academic GP sector to build research capacity.

The poor research funding success rate for primary care-related research was also highlighted, and insufficient funding to take time to clean "very raw and very dirty" general practice data before it can be used – resulting in researchers doing "a lot of work that was unfunded". A Data User interview participant highlighted a PHN's lack of expertise, compared with a university, and the PHNs lack of resources, capacity and awareness "of how big that project is" to adequately undertake public health surveillance using primary care data:

*(There is) lack of GP leadership, a weak academic GP sector with almost no funding of capacity building in the sector; many skill sets are needed including in medical informatics and statistics. Research funding opportunities and success rates are poor for large projects in primary care. To do GP research we need motivation, funds and an easy way of doing it. Primary care data provides a relatively cheap and easier way to obtain significant research outcomes but funding to build the datasets and the trained people to undertake the work need boosting. (Interviewee 3)*

*I think those (PHN) projects where they're not formally expert in doing public health surveillance systems, they probably don't know what they're doing. I think this is the case with a lot of the PHNs. They don't really have the capacity to do it. So, for example, with the (blinded university) studies that (lead researcher blinded) ran – if you look at the enormity of that project and how many people worked on it, you have that same enormity in the PHN level, because they've got so many GPs, but they hire one data analyst who also has to work on other stuff. They maybe apply one or two other people to do the (practice) engagement side. They don't even have an idea of how big that project is. I think it's something that they don't do very well because they don't have capacity about how complex it is, and also, capacity to actually do the recruitment. (Interviewee 2)*

Barriers to timely and cost-effective access to government held data were perceived to be perpetuated by government departments seeking increased funding/resources to raise their internal capacity:

*(Government departments) do have a lot of capacity issues and you (researcher seeking data) get bumped down if there's a government project coming. (I've offered): "We can send you a researcher to be in there with you. He can do it… We've got funded researchers here waiting, why can't you use them?" Why can't we have some secondment scheme for the data analytics where researchers can go in? … Building capacity that way is something that's probably never crossed their minds. They want to actually use the increased demand and data to internally ask for more money … they want to use this increased waiting list that they have for data to say: "You know we haven't got enough people."… These people need incentives to give data out as a default not keep it as a default. (Interviewee 1)*

### 4.8.2. Enablers

Questionnaire respondents outlined many enablers of better use of primary care data. One academic respondent considered there to be "very few (enablers) at the moment" and another 'didn't know', but enablers suggested by 49 other respondents can be grouped around the following non-mutually exclusive categories – see Appendix 7.13:

- Qualities (e.g. trust and altruism)
- Leadership
- Governance
- Partnerships and capacity building
- Technologies and method development, and
- Resources.

**Qualities:** Whereas fear was repeatedly raised as a barrier to better use of primary care data, the idea of generating trust underpinned many of the suggested enablers of better use; i.e. increasing GP and public trust of Data Custodians, users and processes. Enablers suggested that could lead to increased trust included "ensuring transparency in data access and use", clearer line of sight for consumers and GPs on how data are used, ensuring data are protected by robust safeguards, clear leadership, and strengthening engagement and partnerships between key stakeholders. Building trust included having GPs recruit general practices to share data rather than PHNs or people not known to the practice.[13] Altruism was a quality suggested by multiple responders to enabler better use of primary care data – altruism as a driver to prompt GPs to share data 'for the greater good'.

**Leadership** from Data Custodians and/or relevant institutions such as PHNs, universities, general practitioner colleges and even NPS MedicineWise. With regard to showing leadership and overcoming fear, one interviewee considered it important that policy-makers across jurisdictions "believe in", engage with and get behind their data sharing policies as a way to speed up data sharing:

> *I think the first step (to improve current systems of health data sharing) would be that people genuinely shared their data, that would be a big step. I think the fear behind sharing data is that people don't believe that the quality of data is good enough to share. They're a bit afraid that what they're sharing might be wrong. But, I think that's the whole point, that you look at the data – if it's terrible, you improve it. You're never going to know if you don't look at it… I think if the government allowed people to share data, and all the agencies underneath it were engaged in that process – that (would enable) when you ask for the data they say: "Yes, I believe in it" rather than saying: "This is a policy we've never believed in." I think that's really important. If that happened we could probably get into data pretty fast. (Interviewee 2)*

Continued leadership, openness, provision of support with strong governance framework from Data Custodians were suggested as enablers. A multi-institutional Centre for Excellence in use of primary care data, to pool together resources (and data) and expertise to facilitate access to primary care data and build capacity in working with it, was also suggested: bearing in mind that "if you build databases, you've got to maintain them and keep building them".

> **Governance:** Many of the governance-related solutions were focused on achieving data centralisation to enable clear, unambiguous frameworks and direction for governance, adoption of safeguards, and on how to deidentify data (and different 'levels' of deidentification - see Table 9

) so it meets clear legal and ethical standards. Also, to centralise primary care data, its coordination and adoption of a single data extraction tool.[14] Access to larger datasets and reducing duplication of effort due to centralised coordination were secondary benefits of centralisation. The 'long overdue' government Quality Improvement Practice Incentive Payment (QI PIP) was mentioned as an enabler of better use of data (despite it being another 9 months after the close of the survey before QI PIP was introduced). A Data User suggested an accreditation system for data users to enable more efficient access. The system would accredit trusted data users who have "done certain training or have proven themselves over several projects that have protocols in place" so they can bypass or fast-track long approval processes

[13] Interviewee 3 compared the success of GP recruiting of general practices to share data for secondary use to the recruitment model used in the UK by the Clinical Practice Research Datalink (CPRD) where National Health Service / government funded research networks are engaged and GPs are targeted to 'spread the word'. This builds trust and relationships. The message given is that GPs should share data with the CPRD so that their population is represented in studies, but also the altruism of data sharing related to the public good of better disease surveillance, post drug marketing surveillance, epidemiological studies and linking to other datasets to capture the patient journey.

[14] An interviewee noted that while some participants were striving for adoption of a single clinical software system, in Europe this was not a goal – there was acceptance that it "would be nice, but that's not what happens".

to save time and resources and "embed the partnerships" with the Data Custodian. A suggestion for streamlining data governance approvals was as follows:

> *Universities (Data Custodians) need to invest a little bit more, generally speaking, in the type of support they can provide (to people analysing primary care data). Simple things like if there's going to be a legal contract with every dataset and we're increasingly accessing datasets, it may be a more harmonised understanding across all universities and government organisations about what the requirements are so that we could just have some sort of Information Officer at our university sign off and say that they would oversee that they are going to meet these standard expectations of a contract and it's all ready to go. So if we were working across 10 projects and we've got 10 datasets and 10 contracts, not having to start from scratch with everyone … would be an enabler. (Interviewee 4)*

**Partnerships and capacity building** were repeatedly seen as key enablers, particularly related to education and knowledge building among the public and GPs so that the value and application of primary care data can be better understood and therefore lead to greater support and trust. Demonstrating or showcasing the benefits of data use to primary care providers was considered, not only a way to increase trust, but also a means to promote 'best practice' data collection and thereby enable better use of data because the data will be of greater quality. A Custodian of a medical-personnel-related dataset described the intention to "build capacity in the use of the data" by ensuring low cost ($80 flat fee) access and "we don't demand co-authorship". Examples of capacity building between Data Custodians and Data recipients were given, such as:

> *With the Patron dataset, that's my first experience where we've actually had people come in and show us what the data means. If (getting data directly from) a health service… the best you can get is the dataset handed to you and their clinicians may or may not understand the data and the may or may not understand the backend of their data. The complexity of general practice data as such is very valuable to have. The involvement of people (who can interpret the data) or a very strong manual or a set of rules explaining what can and can't be done with the data needs to be provided. (Interviewee 4)*

**Technologies and method development** for enabling better use of primary care data focused on data standardisation, consistency, centralisation, and use of one clinical software or data extraction package or interoperability between them.

**Resource** related enablers included investments in / funding for training, incentive payments, tool improvements, workforce upskilling, making data access more affordable; and time–time to demonstrate good outcomes (and therefore increase awareness of value and promote trust). Time was considered an enabler allowing GPs and PHNs to work through a maturation process from variable acceptance of general practice data sharing and proficiency in its secondary use to "more cohesive" use:

> *I kind of feel like we've had to go through these few years of what I call birthing pains, and we're just coming into young childhood now (with GPs accepting data extraction tools and QI PIP being implemented to incentivise data quality improvement and sharing), and in a couple more years' time we will be entering adolescence and hopefully adulthood, and things will become more cohesive. I just think it's the nature of the beast. If you implement these structures, they have to find their feet and apply themselves locally. (Interviewee 6)*

Ensuring the right team to work with data was described as essential:

> *You've got a database administrator, they download the data and clean it up… the ontology experts work out how to make sense of the data. You've clinician academics who actually know where the data came from and can interpret it … and you've got your statisticians who interpret it … and if you're looking at a larger research effort you might need a methodology expert who can get it across the line. It's not just always about the data, it's a real team effort to use the data effectively. (Interviewee 3)*

GP advisors to non-GP data analysts were considered "essential" to enable correct interpretation of general practice data:

> *I think they (GP advisors) are absolutely essential. Otherwise, well, you can't take that data as some sort of gospel about the real state of affairs out there. It's only what's written and there's a big gap between what's known and what's written, and then there's a gap between 'what is' and what's known, what's identified. (Interviewee 5)*

## 4.9. Data quality and quality frameworks

All questionnaire respondents were given opportunity to share 'more thoughts on data / dataset quality' after they had discussed benefits, limitations, barriers and enablers of better use of primary care data. Solutions thinking was shown (but no new issues or themes were raised that had not previously been raised). The themes raised at this final question included:

- The importance of data analysts understanding why, how and when the data were collected, how it was processed – including coding, cleaning and formatting – and understanding potential biases in the data. To do this there should be:
    - Well developed, transparent standard operating procedures (SOPs) developed for all data collection steps, with input from a range of individuals involved in collecting, preparing and analysing the primary care data, from clinicians to the data analysts/statisticians/end point users.
    - Individuals across the data collection continuum should be trained on use of these SOPs to ensure data quality and consistency in coding and data cleaning processing.
    - There should be continual communication between all the individuals across the continuum of when the data are first collected to the endpoint when analysed and outcomes reported.
    - The analyst can also give feedback on how data quality can be improved to enable better secondary use of primary care data.
- More emphasis on implementation of data quality improvement at the practice level:
    - Involve researchers and secondary users in educating front end primary health care providers so they can appreciate their role and responsibility in data quality and be involved in its secondary use.
    - Resources required for introduction of more meaningful data quality processes integrated into general practice workflow.
- Engage vendors of clinical software to enforce data quality and interoperability, and the same for third party data extraction/analysis software vendors.
- Quality Improvement – Practice Incentive Payments (QI-PIP) and strict, transparent accreditation processes will enforce better data quality (and accountability in primary care)

While primary care data quality was generally seen as poor, a workshop participant cautioned about using the term 'poor quality' without defining what that mean, arguing for the use of 'warts and all' longitudinal data in preference to highly cleaned and curated data:

> *Some longitudinal data includes everything 'warts' and all and is more comprehensive and arguably higher quality than a nicely curated or self-reported dataset where you can't see what is missing.*

A PHN-associated participant described the unreliability of data tables used for reporting which 'changed' seemingly fixed data dates when new inputs were introduced:

> *There was one tangle on one of the systems. Smoking cease date, so we were using that for one of the reports, and then we started doing some data entry and looking at what was happening, if you changed anything that goes in that table, it just updated the smoking cease date. So anybody that was using that to say: "This person stopped smoking" was getting completely the wrong information.*

These final words on processes needed to improve data use and data quality were left by a Data User:

> *When analysing data for research purposes, whether the data were collected specifically for a particular study or came from secondary sources (or combination of both), the analyst needs to understand why, how and when the data was collected, as well as how it was processed, including the coding, cleaning and formatting of the data as all these can potentially introduce biases to the data. There should be well developed, and transparent standard operating procedures (SOPs) for all these data collection steps, with input from a range of individuals involved in collecting, preparing and analysing the primary care data, including the clinicians right through to the data analysts/statisticians/end point users. Individuals across the data collection continuum should also be trained on these SOPs to ensure data quality and consistency in coding and data cleaning processing. There should be continual communication between all the individuals across the continuum of when the data are first collected to the endpoint when analysed and outcome reported. The analyst can also give feedback on how data and data quality can be improved to enable better use of primary care data. (Survey response)*

### 4.9.1. Stakeholder workshop on data quality

The AHRA Workshop: *Approaches to Data Quality Assurance in Australia* that was run as part of this program of work (26th november2018) gathered additional stakeholder perspectives on data quality. Workshop participant suggestions have been summarised as follows:

**What does primary care data quality mean to you**?

- Complete (noting that consideration needed on how 'complete' is defined – because it will not mean all fields of GP data captured)
- Contextual (including understanding the inconsistencies of data from different sources)
- Accurate
- Standardised
- Validated
- Consistent (where the same data means the same thing to every user)
- Transparency of end to end requirements (no allowance of 'illicit' data transformations)
- Timely / current
- Accessible
- Interoperable
- Relevant
- Coherent
- Understanding of what is missing from the data

**Key concerns about data quality for the future use of primary care data were**:

- Unknown quality (lack of documentation)
- Essential to have Data Dictionary including history of data transformations / changes
- Lack of resourcing / funding / time
- GPs collected data is for clinical care, not for research – the minimum data for care is different to the data requirements for research
- Need greater incentive for GPs to share data
- Tell consumers how it really is – i.e. quality data from general practice is not already being used
- Address GPs' perception that good quality data capture is at the expense of quality care

**What are the priority areas to improve data quality?**

- Raise awareness about data quality (more publications on data quality)
    - Clinician engagement
    - Increased consumer engagement around data use
- Foundational governance model
- A comprehensive data quality framework
- Transparency
    - across all bodies working with data
    - of data transformations (Data Dictionaries – including for the workings of the data extraction tools)
- Academic support of and collaboration with PHNs: Greater linkage between Australian Health and Translation Research Centres (AHTRCs) and Primary Health Networks (PHNs)
- Data linkage / integration for insight into patient journey through the health care system
    - Mapping tool to link data from different PHNs / different sources
    - Privacy protecting data linkage keys
- Map current activities to decrease duplication
- Seek patient-centred data rather than general practice-centred data (outcomes will drive quality improvement)

**How can Australia reach consensus on an approach to primary care data quality?**

- National approach
- Structured strategy with common goals
- Incremental and staged change
- Professional leadership and clinician buy-in (consider GP culture and 'what's in it for me' from GP perspective)

- Build clinicians' capabilities and get more GPs and practice staff involved in research (create a toolkit)
    - Have GPs receive and use their own data to drive clinical change (to understand benefit of quality data)
    - Practice-based interventions demonstrating the advantages and positive impact on patient outcomes
- Minimum dataset with consistent terminology and standardised metadata
- Ensure nationally available prescribing and diagnostic coding
- Ensure transparency around data quality framework and underlying data definitions
- Easy to use, validated data capture system (one system for all Australia)
- Take a patient/consumer focus to drive change

**What would a minimum data quality standard look like or have in it?**

- Minimum data specification
- Open source coding
- Data dictionary (accurate description of data manipulation and of contents/context of tables and fields)
- Standardised across clinical software vendors

**Enablers of a minimum data quality standard**

- Work with software vendors to create intelligent systems that are easy to use, ease of population of data fields, data validation (e.g. avoid incorrect birth years)
- Engagement of patients / consumers - encourage them to regularly see their patient summary to pick up quality issues.

### 4.9.2. Data quality frameworks

Data Custodians were asked whether there were any data quality frameworks or tools in place for (any of) the dataset(s) they were responsible for. Two thirds (n=19, 63.3%) responded 'yes', 23.3% (n=7) responded 'no', and 10% (n=3) were 'not sure' (one did not respond). Those 'not sure' included a GP (who explained their secondary use of data as having it incorporated into the NPS MedicineInsight data repository and sharing it with the PHN) and another Custodian who wrote:

> *It depends what you mean by data quality frameworks. We don't have a formal documented DQF (Data Quality Framework) apart from our governance process, and data is managed in-house by our own IT teams*.

The tools, processes, management systems of the 'data quality frameworks' listed by the 19 Data Custodian who reported having such tools/frameworks in place are at Appendix 7.8. Some respondents pointed to published tools or guidelines including the ABS Data Quality Framework, Data 61/CSIRO De-identification Decisions Making Framework, Department of Health Data Governance Framework and adherence to the National Health Act. In the most part, however, custom/bespoke frameworks and local unpublished protocols were described. The types of frameworks / tools described can be categorised as:

- data governance principles and procedures
- privacy policies
- embedded tools in software
- benchmarking and data quality reports being given to general practices that contribute data to the dataset
- workflow for data collectors to improve data input quality
- manuals explaining to users how data are collected and cleaned, and
- published tools (e.g. ABS Data Quality Framework, Data 61/CSIRO De-identification Decisions Making Framework, Department of Health Data Governance Framework)

Appendix 7.7, the above list of data quality framework categories, the variable list of their 'notable' limitations (4.9.3) and Table 10 illustrate lack of consistency, across the sample of responders, in the conceptualisation and application of Data Quality Frameworks and associated data quality tools. It suggests that this concept is not yet sufficiently matured for uniform conceptualisation of what a Data Quality Framework 'looks like' or what one is. Or perhaps the breadth of

datasets included in this study were themselves too variable in nature to allow application of a standard type of data quality framework.

### 4.9.3. Limitations of data quality frameworks

The Data Custodians were asked if they found any *notable* limitations to their data quality frameworks and tools. Eight (42.1%) reported 'yes' there were notable limitations, six (31.6%) that there were not, and one was not sure (four did not respond). Notable limitations included:

- No applicable standard data quality framework
- Lack of defined code sets
- Limitations of SNOMED in practical applications
- Activities limited by internal manpower/resources to support primary care to implement data quality improvements (and workflow issues within practices)
- Implementation of a very comprehensive data quality framework and hence documenting the data quality of collected data is a major task, the tools to implement the framework are still being developed
- Inconsistencies in the extraction tools used by practices and the PHNs leading to data quality issues such as inconsistent results, for example, in age breakdown and diagnosis fields
- The high vendor costs of data extraction tools and their extended tool applications and enhancements
- No clear delineation between levels of de-identification, leading to the issue that most health data needed for health improvement for individuals carries the risk of re-identification
- Technical limitations preventing the data recipient from easily changing extraction fields or customising reports
- Inaccessibility of published data quality frameworks (e.g. too theoretical).

A Data Custodian using CSIRO Data61 (https://data61.csiro.au) as a data quality framework put it into the "too hard basket" and would have preferred some "really clear steps":

> *I found it (CSIRO Data61) quite a lot theory but not very practical. I don't know. Maybe, it is above me. It is all about statistics, and I just find as an end user I had a look at it, and I keep putting it down because it just seems in the too hard basket for me. (Interviewee 7)*

**Table 10. Participants' explanations of Data Quality Frameworks**

| Perspective of participants about Data Quality Frameworks |
|---|
| *We provide a user manual outlining how the data are collected and cleaned to ensure a high-quality dataset. This is our own data quality framework (Survey respondent).* |
| *My interpretation of data quality framework is that it speaks to the business processes of the collectors and the interpretation and use by bodies such as mine [PHN], but that is a nice broad, fluffy statement … The (data quality) frameworks should be allowing for the errant business processes, or the variety of business processes and workflows that exist and seek to standardise them… There is several layers to it. There is the coding and the counting, and there's also the consistency between practices on how they collect and store… There's various degrees, but it is quite confronting for some of them (the general practices). (Interviewee 6)* |
| *For me, it's (data quality frameworks) about understanding your data limitations … the quality is about making sure that you've got enough selection of diverse clinics, so the sample size is representative of a random sample. You don't want to have a biased sample where you're going to get biased results. So, you've got to start with a good base of clinics. And then, once you start extracting your data, you need to understand the staff and the frameworks that are in place, to make sure that data is of good quality … There needs to be some process of quality assurance (QA), to assure that the data you're getting is robust and believable …. So, that's how I see that framework. It's about how confident I can be in that data. (Interviewee 2)* |
| *I don't know if this is exactly what you're thinking about having within a data quality framework, but people who are coming to the Data Custodian obviously don't have the level of expertise that the Custodian will have, generally, so I see that there is a level of responsibility on the Custodian to be able to guide the users significantly … In terms of quality assurance... It's probably a good idea to have a system where the Custodian is providing oversight to make sure that the data is going to be used appropriately by the end-user. (Interviewee 4)* |

### 4.9.4. Strengths and limitations of the study

A clear strength of this study is that the purposeful and snowball sampling recruitment and by taking a broad approach to what a primary care-related data set is, the views of a broad range of known and possible primary care data users and data custodians were included. However, participant self-identification as users or custodians of primary care data used for secondary purposes (with secondary purposes including research, audit, surveillance and quality assurance activities), coupled with varying stakeholder conceptualisation of what is 'secondary use', may have limited participant uptake. Seventeen participants marked themselves 'ineligible' and so exited the survey at the first question and 58 commenced the survey but provided insufficient information to be included. It might be that confusion around what is 'secondary use' of primary care data affected the response rate.

Some respondents enabled insight into varied ways that primary and secondary purpose of data **use** is conceived – i.e. "there is no such thing as 'primary use' or 'secondary use' of data" – highlighting that once general practice data is collected / extracted and curated for another use (research, audit, surveillance, etc) then the new use becomes that data's primary use. It follows then that some of the 75 people who initially entered the survey might have exited or not continued based on not identifying as a secondary user of data. Also, in taking a broad approach to what a primary care-related data set is, we were seeking to hear from anyone who was using data related to primary care – particularly for research purposes. Researchers who had collected primary care-related data for their own studies, again, may not have identified as eligible and exited early. Nonetheless, the 62 respondents identified at least 106 different datasets as relating to primary care, including many bespoke research data collections. If a dataset, extraction tool or data linkage broker was not named (is missing from this report), it does not necessarily mean it is not relevant, it means it was not mentioned by this cohort of study participants. Some data custodians had commercial or research interests which may also have led them to protect their dataset by offing vague descriptions of their dataset (vague descriptions could also be due to a desire to protect anonymity and lack of engagement and time). The gaps in the table of named datasets (Appendix 7.7) reflects the limited information that was shared and therefore limited the appropriateness of 'mapping' available primary care datasets as initially planned. Participants were, however, very forthright in their descriptions of benefits, limitations, barriers and enablers of better use of primary care data, so this study contributes strongly in this area.

Data collection for this study ceased in November 2018 and since then there have been changes to Australia's primary care data environment which are not reflected in the responses made by participants. Notably, the introduction of QI PIP in August 2019 which incentivised GP sharing of the PIP Eligible Data Set to the local PHN, and general practice participation in a program of data-driven continuous quality improvement [15]; and development of the National Primary Health Care Data Asset by the Australian Institute for Health and Welfare [16] and the advent of 'Lumos', NSW Health (https://www.health.nsw.gov.au/lumos/Pages/default.aspx). While these initiatives were known about at the time of data collection, the QI PIP was not implemented and consultation on the National Primary Health Care Data Asset had not yet commenced.

# 5. Recommendations

The purpose of this study was to scope existing primary care datasets (with a focus on general practice) and engage stakeholders around gaps, needs and optimal delivery methods for building capacity in data harmonisation in order to support data driven health care improvement. Participants in the study expressed a shared vision to improve population health outcomes through better use of primary care data. To achieve this, the following recommendations could be considered by AHRA and other key stakeholders involved in the collection, curation, sharing and use of primary care data.

1. **Establish partnerships between University departments, PHNs, government and other data custodians and stakeholder groups**

   Accountable partnerships may address many of the barriers identified in this study, including the need for clinician/academic general practitioner involvement and leadership, implementation of data quality frameworks, training of data analysts, and increased standardisation of data coding systems in general practice EMRs. These partnerships may also open up new possibilities to use data for multiple purposes and reduce duplication of effort.

2. **Establish professional development and career paths for clinicians involved at all stages from data entry to analysis**

   The benefits of optimising data entry into EMRs can be more immediate than secondary use for research and population health and policy purposes. Digital health literacy, including optimising the use of EMRs, should be a key component in both medical student and general practice vocational training. Clinicians need to learn how to realise clinical safety and quality of care benefits, and to achieve this without increasing workload. This will require liaison with bodies including University Medical Schools, the Australian College of Rural and Remote Medicine and the Royal Australian College of General Practitioners.

   Hospitals implementing EMRs are increasingly offering clinical informatics positions to medical staff. There are not such opportunities currently in general practice, though this study has demonstrated the need for related roles and responsibilities. Support for building academic general practice leadership and informatics training will be critical in building general practice capacity in this area. This could incorporate trainee secondments with data custodians in government, Universities and PHNs, with mutual benefits for data analysts in those settings to gain insight from primary care clinicians.

3. **Establish trust through data governance transparency and security**

   The words ‘privacy’, ‘trust’ and ‘governance’ figure prominently in this study, reflecting on how difficult it is to build mechanisms to protect the rights of individuals and data suppliers, and to manage the risks for data custodians. The survey responses indicate inconsistencies in governance, not a surprise given that 106 separate data collections are reported. However this study also supports the view that good practice does exist. Whilst effective and transparent governance mechanisms are challenging, the sector needs to encourage, promote and support the wider adoption of consistent practice in data governance.

   Governance of primary care datasets should be publicly available and include references to data security and data sharing availability and preferences. Data dictionaries that detail data transformations, context and limitations should be available so that determinations about “fit for purpose” can be assessed. Procedures should aim to provide data in a timely manner to approved projects in order to ensure that they can guide population health planning and policy initiatives. Privacy protecting mechanisms for consumers and clinicians are important, but this should be achieved while retaining the ability to support data linkage to facilitate the exploration of the patient journey across the health system and other sectors. In some cases, researchers need access to data that is deemed sensitive. This should not be perceived as a barrier so long as governance, management and security measures are in place commensurate to the level of risk. Measures to help mitigate risk include the provision of secure research

environments, contractual obligations on researchers, mandatory research training and transparent, proactive standard operating procedures.

Above all, we need to strive to maximise transparency on the use of primary care data and to increase communication with those who supply such data and the individuals and communities we are researching.

**4. Establish clarity on data definitions and the use of and purpose of data quality frameworks**

Clarity about terms, definition and taxonomy is required. In particular, definitions of secondary use of data and de-identification are important to clarify so that there is shared understanding of the concept and application of a data quality framework. There is need for a clear, practical, comprehensive data quality framework for primary care data that enables standardisation and assessment of data quality. This is key to understanding if the data are "fit for purpose". Kahn's Harmonized Data Quality Framework [7] is gaining popularity internationally and is already being applied through the work of the AHRA Transformational Data Collaboration (University of New South Wales and The University of Melbourne) and the Australian Institute for Health and Welfare [8]. The implementation of a comprehensive data quality framework for rigorous assessment and documentation of data context and quality requires resources, to achieve meaningful application and information.

**5. Establish a program of further applied research into optimising data for data driven health care improvement**

Since data collection was completed for this study, the Quality Improvement Practice Incentive Payment (QI PIP) has been implemented and the development of a Primary Health Care National Minimum Data Asset by the Australian Institute of Health and Welfare has commenced. It is likely that these initiatives will impact our national conversation with regards to data governance, ownership and use. It is crucial that we leverage these initiatives to further optimise the way the sector is able to work with data for health care improvement.

Implementation of the recommendations from this study would require additional funding to achieve, particularly in the areas of governance uplift, workforce development and the development and implementation of data quality frameworks. Additionally, the utilisation of mechanisms such as common data models to present data from general practice EMRs in a consistent and validated format would enhance data quality assurance, simplify data governance (so that only aggregated data are released), and reduce the costs and complexity of medical research.

Work in these areas is compatible with the aspirations and goals of The Australian Health Research Alliance. The national, open, collaborative nature of the Australian Health Research Alliance places it in a good position to help support uplift at a national level in collaboration with primary care stakeholders, government and consumers.

Potential funders and collaborators include the Medical Research Future Fund, the Australian Research Data Commons (ARDC), State and Federal Government, the NHMRC, the Australian Research Council and Primary Health Networks.

# 6. References


1. Youens D, Moorin R, Harrison A, Varhol R, Robinson S, Brooks C, et al. Using general practice clinical information system data for research: the case in Australia. International Journal of Population Data Science. 2020;5(1).

2. Canaway R, Boyle DI, Manski-Nankervis J-A, Bell J, Hocking J, Clarke K, et al. Gathering data for decisions: Best practice use of primary care electronic records for research. Med J Aust. 2019;210:S12-S6. doi:10.5694/mja2.50026

3. Productivity Commission. Data Availability and Use: Productivity Commission Inquiry Report2017 July 2018; (July):[64 p.]. Available from: https://www.pc.gov.au/inquiries/completed/data-access#report.

4. Australian Government Department of Health. Framework to guide the secondary use of My Health Record system data. Canberra: Australian Government Department of Health; 2018.

5. Department of Health. PIP QI Incentive guidance Canberra: Australian Government; 2020 [cited 2020 1 November]. Available from: https://www1.health.gov.au/internet/main/publishing.nsf/Content/PIP-QI_Incentive_guidance.

6. Australian Institute of Health and Welfare. Primary health care data development: Australian Government; 2020 [cited 2020 1 November]. Available from: https://www.aihw.gov.au/reports-data/health-welfare-services/primary-health-care/primary-health-care-data-development.

7. Kahn MG, Callahan TJ, Barnard J, Bauck AE, Brown J, Davidson BN, et al. A harmonized data quality assessment terminology and framework for the secondary use of electronic health record data. eGEMs. 2016;4(1):18. doi:10.13063/2327-9214.1244

8. Henley-Smith S, Boyle D, Gray K. Improving a Secondary Use Health Data Warehouse: Proposing a Multi-Level Data Quality Framework. EGEMS (Wash DC). 2019;7(1):38-. doi:10.5334/egems.298

9. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap): a metadata-driven methodology and workflow process for providing translational research informatics support. Journal of Biomedical Informatics. 2009;42(2):377-81. doi:10.1016/j.jbi.2008.08.010

10. QSR International. NVivo Qualitative Data Analysis Software [NVivo 12 Plus]. Melbourne2020.

11. IBM. IBM SPSS statistics. 26.0.0.0 ed: IBM Corporation; 2019.

12. The University of Sydney. Bettering the Evaluation and Care of Health (BEACH): National study of general practioner clinical activity 2016 [cited 2020 16 May]. Available from: https://www.sydney.edu.au/medicine-health/our-research/research-centres/bettering-the-evaluation-and-care-of-health.html.

13. Australian Government Department of Health. BEACH - Bettering the Evaluation and Care of Health: a continuous national study of general practice activity. Communicable Disease Intelligence [Internet]. 2003 16 May 2020 [cited 2020 16 May]; (27,3). Available from: https://www1.health.gov.au/internet/main/publishing.nsf/Content/cda-pubs-cdi-2003-cdi2703-htm-cdi2703m.htm.

14. Forde I, Nader C, Socha-Dietrich K, Oderkirk J, Colombo F. Primary care review of Denmark. OECD Health Division; 2016.

15. Australian Government Department of Health. PIP QI Incentive guidance Canberra: Australian Government Department of Health; 2019 [cited 2020 30 July]. Available from: https://www1.health.gov.au/internet/main/publishing.nsf/Content/PIP-QI_Incentive_guidance.

16. AIHW. Developing a National Primary Health Care Data Asset: consultation report. Canberra: Australian Institute of Health and Welfare; 2019. Contract No.: Cat. no. PHC 1.

# 7. Appendices

## 7.1. Glossary and acronyms

| Term | Definition |
|---|---|
| **AHRA** | Australian Health Research Alliance |
| **AHRTC** | Advanced Health Research and Translation Centres – an initiative of the NHMRC |
| **AIHW** | Australian Institute of Health and Welfare |
| **AMS** | Aboriginal Medical Service |
| **ATSI** | Aboriginal and Torres Strait Islander |
| **CALD** | Culturally and Linguistically Diverse |
| **CVDL** | Centre for Victorian Data Linkage |
| **Data Custodian** | The Data Custodian is the Head, Department of General Practice and is ultimately responsible for the data contained in the Patron data repository. |
| **Data Steward** | The Data Steward – the Director of HaBIC $R^2$ – has the delegated responsibility to manage the physical security and other data curation activities of the Patron data collection. |
| **DHHS** | Victorian government Department of Health and Human Services |
| **EMR** | Electronic Medical Record |
| **EULA** | End User Licence Agreement |
| **GP** | General practitioner |
| **GRHANITE®** | A data software tool that enables privacy-protecting extraction, curation and delivery of sensitive data to data storage facilities. It was first developed in 2007 at The University of Melbourne by Douglas Boyle and Siaw Teng Liaw. |
| **HaBIC $R^2$** | Health and Biomedical Informatics Centre, Research Information Technology Unit |
| **MACH** | Melbourne Academic Centre for Health, an NHMRC- accredited venture between Victorian healthcare providers, medical research institutes and The University of Melbourne. Its purpose is to facilitate collaboration between academia and healthcare to accelerate the translation of innovative research into clinical care. |
| **MBS** | Medical Benefits Scheme |
| **MDS** | Minimum Data Set |
| **NADA** | Network of Alcohol and other Drugs Agencies |
| **NHMRC** | National Health and Medical Research Council |
| **PBS** | Pharmaceutical Benefits Scheme |
| **PHN** | Primary Health Network |
| **QA** | Quality assurance |
| **QI PIP** | Quality Improvement Practice Incentive Payment (QI PIP was known about but had not been introduced at the time of the survey) |
| **RACGP** | Royal Australian College of General Practitioners |
| **REDCap** | Research Electronic Data Capture: a secure web-based platform developed by Vanderbilt University |

## 7.2. Summary of the Data Quality Framework activity

### Routinely-collected GP data – data quality frameworks

The primary care data survey highlighted a lack of consistency in approaches to managing data quality in primary care data. This finding was not unexpected, and MACH was researching data quality frameworks in advance of AHRA 2018 activities. This activity was designed to explore options for the development of a practical data quality framework that may be implemented in this sector.

A practical data quality assessment framework derived in-part from Kahn et al's paper: *A Harmonised Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data [7]* was published by the investigators [8].

Implementation of the experimental data quality framework was piloted utilising the University of Melbourne Data for Decisions data repository known as 'Patron', or the Patron primary care data repository (see www.gp.unimelb.edu.au/datafordecisions. This data repository contains data collected from three common GP clinical software systems. It should be noted that the framework was designed to be data warehouse agnostic – i.e. regardless of how a data warehouse is structured, the data quality may be assessed. The framework is extensive with full characterisation of the data warehouse beyond the remit of the project. The project limited itself to examining data commonly utilised in research projects, for example, patient identification, medical encounter recording, diagnosis recording and prescribing.

### Data Analysis and Results

An SQL database with a user interface was created to allow the compilation of the data quality characteristics of the warehouse. This database allowed for the characterisation of the data at the system, table and field level. The characterisation is highly detailed with several dozen separate tests employed in the areas of: context within the original database system, relational conformance, uniqueness and temporal plausibility, value conformance, computational conformance, data completeness, and atemporal plausibility.

The data analysis resulted in the documentation of commonly identified issues such as missing data or limitations to the data due to inconsistent coding and free-text entry. Some more unusual and difficult to spot issues were identified for example in one GP computer system, a table containing medical diagnoses has a field called 'DIAGNOSIS_DATE'. This field contains the date the diagnosis record was saved, it is not the actual diagnosis date. The actual diagnosis date is recorded in a separate table.

This work highlighted several key factors:

- Issues in data that are hard to spot; misconceptions are not just possible but indeed are highly likely.
- Given the subtlety of the issues identified, it can be expected that errors will be getting made in how to interpret the data across different tool providers nationally.
- Data is frequently incomplete; having a physical record of the nature of data errors is fundamental to epidemiology. In most cases access to information about the underlying completeness is not available to the researcher at the time of requesting access to data.
- Data is commonly coded, and these codes vary from provider to provider. Standardisation and the correct interpretation of data collected using different coding mechanisms is highly problematic.
- Applying a data quality framework is very time consuming.

## 7.3. Data collection questionnaire (REDCap online survey)

*Confidential*



# Mapping Australian primary care datasets and building data use capacity

Please complete the survey below.

If you have any questions please email rachel.canaway@unimelb.edu.au or phone Dr Rachel Canaway on (03) 8344 3392 at the Department of General Practice, The University of Melbourne.

Thank you!

---

**Welcome and consent**

There is increasing interest in 'secondary' use of general practice data in Australia, but no one has a clear picture on what data is collected for secondary use, by whom, and what is being done with it.

The results of this survey will provide a clearer picture of primary care datasets in Australia (being used for secondary purposes).

This project is part of national capacity building work to prompt data-driven healthcare improvement. It is funded by AHRA - the Australian Health Research Alliance.

Your completion of this survey will take around 5-15 minutes. This time variation is because many of the questions allow open ended responses (less or more detailed).

On completing and submitting this questionnaire, your consent to participate is implied. Through your consent, you acknowledge that:

1. You understand that this project is for research purposes.

2. You have read and understand the Participation Information and you freely agree to participate (see attached Information Statement).

3. You understand that participation is voluntary, and you can withdraw your data up to one week after you submit the questionnaire (by contacting Dr Rachel Canaway).

4. You understand that the researchers will not reveal your identity or personal details if information about this project is published or presented in any public form; excepting if you wish to be named in acknowledgements (tick box below).

5. You understand that as the study gathers information from a small number of experts in the area of primary care data, it may not be possible to guarantee your anonymity.

If you have any questions please contact Dr Rachel Canaway at The University of Melbourne: (03) 8344 3392 or rachel.canaway@unimelb.edu.au

[Attachment: "Information Statement_Data mapping and linkage_Survey.pdf"]

Please check the box if you would like to be named in acknowledgements of any published results as having contributed to this research: ◯ Yes, I would to be named in acknowledgements

Please enter your name and affiliation: _____



Please check the box if you would like a summary of the results emailed to you at the conclusion of the project: ◯ Yes, please email me a summary of results

Please enter your email address to receive the summary of results: _____

*Page 3 of 11*

## Eligibility check

Some electronic medical records from Australian general practice or other primary care services are used for secondary purposes including research, audit, surveillance, and quality improvement.

Have you used, or do you have responsibility for, a primary care dataset that is used for secondary purposes?

◯ Yes ◯ No

*Page 4 of 11*

## Data Custodian or Data User?

Do you consider yourself a Custodian or 'owner' of one or more general practice or other primary care datasets?

◯ Yes ◯ No

*Page 5 of 11*

## Questions for Data Custodians

What is the name of the dataset you are Custodian of or have responsibility for, and/or where is it located?

_____

If you are responsible for more than one, please list all.

What state(s) and/or localities does the majority of the data come from?

_____

Please describe characteristics and purpose of the dataset(s).

_____
(* Be as brief or detailed as you like.)

Please describe any data governance and consent mechanisms that are in place around the dataset(s).

_____

Are there any data quality frameworks or tools in place for (any of) the dataset(s)?

◯ Yes ◯ No ◯ I'm not sure

Please describe the data quality framework(s) or tools:

_____

Please briefly describe why you are 'not sure' whether there are data quality frameworks or tools in place:

_____

Have you found any notable limitations to the data quality frameworks or tools you have used?

◯ Yes ◯ No ◯ I'm not sure

Please briefly describe the limitations of the data quality framework(s) or tools:

_____

Who can access the dataset and how?

Do you get anything from people/researchers who get access to the dataset?

○ Yes ○ No ○ I'm not sure
(For example a: fee; service; report or other information; etc.)

Please describe what you get in exchange for sharing the data:

Has your dataset been linked with other datasets?

○ Yes ○ No ○ I'm not sure

What other datasets have you linked your dataset to? Please list and also explain what methods / tools were used for data linkage:

Have you found limitations to the tool or methods used for data linkage method(s)? If yes, please explain:

*Page 6 of 11*

Do you intend for your dataset to be linked with other datasets?

○ Yes ○ No ○ I'm not sure

*Page 7 of 11*

## Secondary users of general practice and other primary care datasets

You have indicated that you are not a Data Custodian but you have used general practice or other primary care datasets for secondary purposes.

Please briefly describe the nature of your interaction with these datsets:

What primary care or general practice datasets have you used for secondary purposes? Please list them and describe how you gained access to them:

Did you have to pay or provide a service to access the data? Please explain:

Please describe what elements are important to you when accessing or using primary care / general practice datasets for secondary purposes:

Have you linked any of the general practice / primary care datasets you've used with other datasets?

○ Yes ○ No ○ I'm not sure

What methods or tools for data linkage did you use?

If you encountered limitations related to data linkage tools or method(s) you have used, please describe them here:

*Page 8 of 11*

## Mapping primary care datasets

Please list other general practice or primary care datasets that you are aware of (you may or may not have used them):

(If you are not aware of any other primary care datasets being used for secondary purposes, please say so.)

## Building capacity of primary care datasets - to support improved health outcomes

How do you think that primary care datasets could be better used to support improved health outcomes?

_____

What do you think are the LIMITATIONS of secondary use of primary care datasets?

_____

What do you think are the BENEFITS of secondary use of primary care datasets?

_____

What are the BARRIERS to better use of primary care data?

_____

What are the ENABLERS to better use of primary care data?

_____

Do you think there is an ideal way to link primary care datasets? If yes, please describe it:

_____

If you would like to share more thoughts on data / dataset quality, please do so here:

_____



## About you

What type of organisation do you work for (related to your work with primary care data)?

- ☐ Educational institution
- ☐ GP clinic
- ☐ Government
- ☐ PHN (Primary Health Network)
- ☐ Pharmaceutical
- ☐ Insurer
- ☐ Other

Please describe the type of 'other' organisation(s).

_____

In what state or territory are you based?

- ○ ACT ○ New South Wales
- ○ Northern Territory ○ Queensland
- ○ South Australia ○ Tasmania
- ○ Victoria ○ Western Australia
- ○ Elsewhere

Please describe where 'Elsewhere' is:

_____

About how many different Australian primary care datasets have you accessed for secondary purposes?

- ○ None ○ 1 ○ 2 ○ 3
- ○ 4 ○ 5 ○ 6 ○ 7
- ○ 8 ○ 9 ○ 10 ○ More than 10

In what year (approximately) did you first access a primary care dataset for secondary use?

_____

How did you hear about this survey?

- ○ Direct email from the researchers
- ○ Link was forwarded by a colleague
- ○ RACGP
- ○ Social Media
- ○ Other

(Please choose the closest single option.)

If other, please describe:

_____



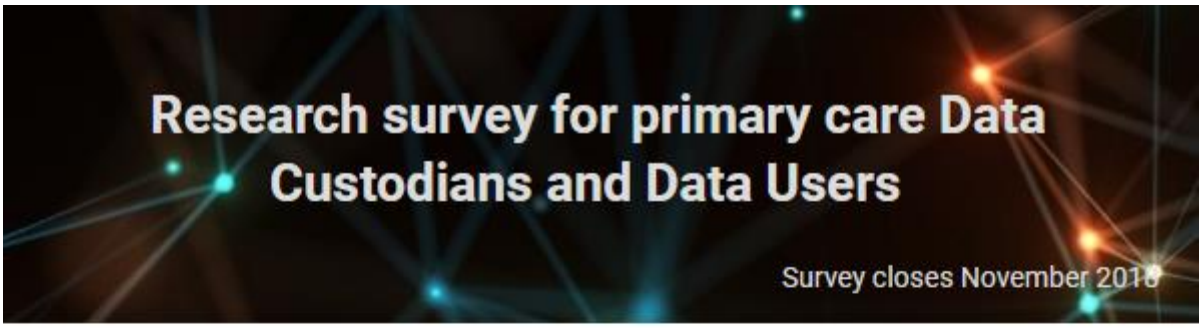
## Do you have more to say?

If you would like to make any other comments please do so here. Or leave your contact details if you would like to contribute further via phone:

_____

Alternatively, phone Dr Rachel Canaway on (03) 8344 3392, 0407 658 012 or email rachel.canaway@unimelb.edu.au

## 7.4. Emailed advertisement for survey participation




AHRA
AUSTRALIAN HEALTH RESEARCH ALLIANCE

**Ethics** ID 1852055
**Funder:** Medical Research Future Fund (MRFF)

**More information:**
Dr Rachel Canaway
The University of Melbourne
(03) 8344 3392
Email

Melbourne Academic Centre for Health

### Help us map Australia's **primary care** data 'black hole'

Data Custodians and Data Users please **click** on the link for more information and to complete the survey. Please also forward this email to your relevant colleagues.

**Survey**

The purpose of this research is to map the existence of Australian primary care datsets (with a focus on general practice), identify data linkage tools and data quality frameworks. Findings will help identify strategies **to maximise the effectiveness of secondary use of primary care data**.

The survey should take 5-10 minutes, depending on the detail of your answers. If you prefer you can contribute by phone.

**Do you know other primary care data users or custodians?**
Please circulate this email among your team. If the link button doesn't work, paste this URL into your browser: https://redcap.healthinformatics.unimelb.edu.au/surveys/?s=Y7R9D338E3

Copyright © 2018 All rights reserved

## 7.5. Interview question theme guide

*This project aims to gain deeper understanding of availability and secondary use of routinely collected general practice data. If you agree to be interviewed, the questions asked will explore the following areas:*

1. Briefly: your professional role and the type of organisation you work for.
2. Whether you are a Data Custodian / Data Steward or a user of general practice data.
3. Identifying general practice datasets in Australia, including their:
    a. Purpose
    b. Governance and consent mechanisms
    c. Data quality frameworks
    d. Data linkage capability
4. What you consider to be the ideal way to link data – and barriers and facilitators of this?
5. What you consider to be the main benefits and limitations of secondary use of general practice datasets?
6. Ways that general practice datasets could be better used to support improved health outcomes – including barriers and enablers to better user.
7. How current systems of healthcare related data sharing could be improved to better prompt research and knowledge discovery?
8. Suggestions of other people in Australia that might be appropriate to be interviewed as part of this study.

## 7.6. Stakeholder workshop agenda


**Primary Care data for healthcare improvement and research:**

**Approaches to data quality assurance in Australia**

Monday 26 November 1:00 - 5:00pm
Mercure Sydney,
818-820 George St, Sydney

**AGENDA**

| Time | Item | Presenter |
|---|---|---|
| 1:00-1:15pm | Welcome and Introductions | A/Prof Douglas Boyle |
| 1:15-1:35pm | Results from the AHRA Research survey for primary care Data Custodians and Data Users (20 minutes) | Dr Rachel Canaway<br>Dr Jo-Anne Manski-Nankervis |
| 1:35-2:00pm | Small group discussions (25 minutes) | |
| 2:00-2:15pm | Feedback from the groups (15 minutes) | |
| 2:15-2:30pm | Break | |
| 2:30-2:50pm | Data Quality Assurance Frameworks research – and results (20 minutes) | A/Prof Douglas Boyle<br>Ms Sandra Henley-Smith |
| 2:50-3:15pm | Small group discussions (25 minutes) | |
| 3:15-3:30pm | Feedback from the groups (15 minutes) | |
| 3:30-3:45pm | Break | |
| 3:45-4:45pm | AHRA Primary Care Data Research Forward Planning<br>• General Priority Areas<br>• MRFF Australian Medical Research and Innovation Priorities 2018-2020: Primary Care Research | TBC<br>A/Prof Douglas Boyle |
| 4:45-5:00pm | Closing remarks | A/Prof Douglas Boyle |

## 7.7. Summary of primary care dataset identified by respondents

| | Primary care dataset named by respondents | Respondent used or owned | Custodian organisation | Notes. Extraction tool | Can it be Linked? | ACT | NSW | NT | QLD | SA | TAS | VIC | WA | Audit | Surveil-lance | Research / trials |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Primary care related data available for secondary use** | | | | | | | | | | | | | | | |
| 1 | 10% MBS and PBS sample data (no longer available) | ✓ | AIHW, Australian Government | | Yes | | | | | | | | | | | ✓ |
| 2 | 45 and Up Study | ✓ | Sax Institute | | Yes via CHeReL | | | | | | | | | | | ✓ |
| 3 | Aboriginal Community Controlled Organisation and Aboriginal Medical Services data, e.g. from VACCHO, NACCHO | | Various | | | | | | | | | | | | | ✓ |
| 4 | AIHW – Website and general data access | | AIHW | | | | | | | | | | | ✓ | | |
| 5 | Apollo - patient register for Sonic Healthcare (available on request) | ✓ | Sonic Healthcare | | Yes | | | | | | | | | ✓ | | ✓ |
| 6 | Australian Sentinel Practice REsearch Network (ASPREN) | ✓ | University of Adelaide | Manual collection | ? | | | | | | | | | | ✓ | ✓ |
| 7 | ATSI Health Organisations: Online Services Report (AIHW) https://www.aihw.gov.au/reports/indigenous-health-welfare-services/health-organisations-osr-key-results-2016-17/contents/table-of-contents | | AIHW | | | | | | | | | | | ✓ | | |
| 8 | Australian Immunisation Register (AIR) – Australian Government | ✓ | Curtin University | N/A | | | | | | | | | | | ✓ | ✓ |
| 9 | BEACH – Bettering the Evaluation and Care of Health (data collection ceased in 2016) | ✓ | University of Sydney | Manual Collection | Yes | | | | | | | | | | | ✓ |
| 10 | ePractice-Based Research Network general practice dataset | | University of NSW | GRHANITE | | | | | | | | | | | | ✓ |
| 11 | MBS – Medicare Benefits Scheme (AIHW) and Medicare Online http://www.mbsonline.gov.au/internet/mbsonline/publishing.nsf/Content/Home | ✓ | AIHW | | Yes | | | | | | | | | ✓ | | ✓ |
| 12 | Medical Director held data | ✓ | Medical Director | | | | | | | | | | | ✓ | | In-house by MD |
| 13 | My Health Record | | AIHW | MyHR | | | | | | | | | | | | Not yet |
| 14 | My Healthy Communities | | AIHW | | | | | | | | | | | | | |
| 15 | NPS MedicineInsight | ✓ | NPS MedicineWise | GRHANITE | Yes to MBS, PBS, | | | | | | | | | ✓ | | ✓ |
| 16 | NSW MoH Western Sydney Data linkage pilot project, 2015-2018 | ✓ | | | Yes via CHeReL | | | | | | | | | | ✓ | ✓ |
| 17 | NT Primary Care Information System (PCIS) – rural and remote health clinics | ✓ | | | Yes | | | | | | | | | | | ✓ |
| 18 | The Patron primary care data repository | ✓ | University of Melbourne | GRHANITE | Yes | | | | | | | | | | | ✓ |
| 19 | PBS -Pharmaceutical Benefits Scheme (AIHW) | ✓ | AIHW | | Yes | | | | | | | | | ✓ | | ✓ |

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 20 | PEN CS | ✓ | Pen CS | | ? | | | | | | | | | | | ✓ |
| 21 | Outcome Health / POLAR Data Space (NSW, Vic, others) –POLAR formerly MAGNET | ✓ | Outcome Health | POLAR | Yes | | | | | | | | | | | ✓ |
| 22 | Monash data MAGNET | | | | | | | | | | | | | | | |
| 23 | ReCEnT – Registrar Clinical Encounters in Training dataset | ✓ | University of Newcastle | Manual collection | | | | | | | | | | | | ✓ |
| 24 | SA Health Omnibus (some questions relate to GP care) | | | | | | | | | | | | | | | |
| 25 | Voluntary Indigenous identifier (VII) (in Medicare Database). Used to generate statistics | ✓ | | | Yes | | | | | | | | | | | |
| 26 | WA Data Link (Data Linkage WA – WA Department of Health data, registries) – primary care not readily available | | | | Yes | | | | | | | | | ✓ | | ✓ |
| 27 | WA Primary Health Alliance (PHN data, AIR, Primary Mental Health Commission) | ✓ | Curtin University | | | | | | | | | | | ✓ | ✓ | ✓ |
| | **Health workforce** | | | | | | | | | | | | | | | |
| 28 | AHPRA – for health workforce | | | | | | | | | | | | | ✓ | | |
| 29 | AMPCo Database of medical practitioners - to draw a random sample of registered GPs across Australia - paid | ✓ | | | | | | | | | | | | | | ✓ |
| 30 | DoH GP Workforce Statistics. http://www.health.gov.au/internet/main/publishing.nsf/content/general+practice+statistics-1 | | | | | | | | | | | | | | | |
| 31 | MABEL – Medicine in Australia Balancing Employment and Life | ✓ | | | Yes | | | | | | | | | | | ✓ |
| 32 | National Health Workforce Data Set (NHWDS) - AIHW | | | | | | | | | | | | | | | |
| | **Software vendors** | | | | | | | | | | | | | | | |
| 33 | Best Practice – Pyefinch dataset | ✓ | | | Yes | | | | | | | | | | | |
| 34 | MD Heart (Medical Director) | ✓ | | | | | | | | | | | | | | |
| 35 | **DOCLE** (**Do**ctor **C**ommand **L**anguag**e)** Used by Medical Director | | | | | | | | | | | | | | | |
| 36 | ICPC International Classification for Primary Care | | | | | | | | | | | | | | | |
| | **PHNs** General practice data; PIP program data; PMHC-MDS Primary Mental Health Care Minimum Data Set | | | | | | | | | | | | | | | |
| 37 | PHN – ACT (1) | ✓ | | | | | | | | | | | | ✓ | ✓ | ✓ |
| 47 | PHNs – NSW (10) – including SPDS project (Sentinel Practices) in SE PHN | ✓ | | PEN CS PATCAT POLAR | | | | | | | | | | ✓ | | Yes |
| 48 | PHN – NT (1) | ✓ | | PEN CS PATCAT | | | | | | | | | | Yes | | |
| 55 | PHNs – QLD (7) Includes Primary Sense – Gold Coast | ✓ | | PEN CS PATCAT Primary Sense | | | | | | | | | | ✓ | | ✓ |
| 57 | PHNs – SA (2) | ✓ | | PENCAT | | | | | | | | | | ✓ | Yes | ✓ |
| 58 | PHN – Tas (1) | | | | | | | | | | | | | ✓ | | |
| 64 | PHNs – Vic (6) | ✓ | | PENCAT, POLAR | | | | | | | | | | ✓ | | ✓ |

| | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 67 | PHNs – WA Primary Health Alliance (Alliance of 3 PHNs) | ✓ | | | | | | | | | | | | ✓ | | Yes |
| | **Mental Health** | | | | | | | | | | | | | | | |
| 68 | Headspace – Datasets HAPI reports | | | | | | | | | | | | | | | |
| 69 | PHNs - PMHC-MDS Primary Mental Health Care Minimum Data Set https://pmhc-mds.com/ | ✓ | | | | | | | | | | | | ✓ | | ? |
| | **Research projects creating data not normally available to others** | | | | | | | | | | | | | | | |
| 70 | Community pharmacy project – extracted direct from health service for trial participants (identified) | ✓ | Deakin University | | Yes | | | | | | | | | | | ✓ |
| 71 | ACCEPt trial data - Custom GP consultation data for research. Australian Chlamydia Control Effectiveness Pilot http://accept.org.au/ 2010-2015. | ✓ | University of Melbourne | GRHANITE | | | | | | | | | | | | ✓ |
| 72 | Custom GP data for research -ASPREE: ASPrin in Reducing Events in the Elderly | | | | | | | | | | | | | | | ✓ |
| 73 | Custom GP data for research -STAREE: STAtins in Reducing Events in the Elderly | | | | | | | | | | | | | | | ✓ |
| 74 | GRAPHC Custom GP data for research - West Adelaide data | ✓ | Australian National University | Canning Tool | | | | | | | | | | | | ✓ |
| 75 | Custom GP data for research -mycoplasma in sexual health clinical antibiotic prescribing. general practice and sexual health clinic data sets for research purposes | ✓ | James Cook University | | | | | | | | | | | | | ✓ |
| 76 | Custom GP data for research – CD-IMPACT (funded by Better Care Victoria) | ✓ | Western Health | PEN CAT | | | | | | | | | | | | ✓ |
| 77 | Custom GP data for co-located paediatrics study | ✓ | Royal Children's Hospital | GRHANITE | | | | | | | | | | | | ✓ |
| 78 | Custom GP data for research – datasets from recruited practices, paid $100 per randomsied patient participant. | ✓ | University of Tasmania | | | | | | | | | | | | | ✓ |
| 79 | Custom GP data for research – Corporate Health Service data – not just ACT | ✓ | Australian National University | | | | | | | | | | | | | ✓ |
| 80 | Custom GP data for research – Dementia related general practice patient data | ✓ | University of Newcastle | | | | | | | | | | | | | ✓ |
| 81 | Custom GP patient data - WEAVE | ✓ | University of Melbourne | | | | | | | | | | | | | ✓ |
| 82 | VCCC Primary Care Linked Dataset | ✓ | Victorian Comprehensive Cancer Centre | | Via BioGrid Australia | | | | | | | | | | | ✓ |
| 83 | National and state datasets, data cost $16K plus manuscripts | ✓ | | | | | | | | | | | | | | ✓ |
| 84 | Wollongong Uni EHR Data | | | | | | | | | | | | | | | ✓ |
| 85 | WA Homeless person's health survey | | | | | | | | | | | | | | | ✓ |
| 86 | Other custom surveys or datasets (e.g. will have to check with Data Custodian before listing – was not listed 104) | | | | | | | | | | | | | | | ✓ |
| | **Other miscellaneous datasets noted** | | | | | | | | | | | | | | | |

| | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 87 | Custom GP data for research ACCESS trial - The Australian Collaboration for Chlamydia Enhanced Sentinel Surveillance 2007-2020. GP consult data | ✓ | Burnet Institute | GRHANITE | | | | | | | | | | | ✓ | ✓ |
| 88 | GPs using their clinical datasets – or sharing with PHNs / universities for secondary use. | ✓ | | | | | | | | | | | | | | |
| 89 | Medicare Online http://www.mbsonline.gov.au/internet/mbsonline/publishing.nsf/Content/Home | ✓ | | | | | | | | | | | | | | ✓ |
| 90 | State, Territory governments – e.g. sentinel data collection, | | | | | | | | | | | | | | ✓ | |
| 91 | DoH – National datasets. | | | | | | | | | | | | | | | |
| 92 | DHHS identifiable datasets (to be accessed by Vic PHNs) | | | | | | | | | | | | | ✓ | | |
| 93 | NSW Government - Aboriginal Affairs: Opportunity, Choice, Healing, Responsibility, Empowerment (OCHRE) | | | | | | | | | | | | | | | |
| 94 | NADA – Network of Alcohol and other Drugs Agencies data | | | | | | | | | | | | | | | |
| 95 | PROMs – Patient Reported Outcome Measures | | | | | | | | | | | | | ✓ | | ✓ |
| 96 | PREMs – Patient Reported Experience Measures (e.g. ABS patient experience survey) | | | | | | | | | | | | | | | |
| 97 | South West Sydney Local Health District Community and primary care datasets (SWSLHD) | | | | | | | | | | | | | | | |
| 98 | Australian Bureau of Statistics | ✓ | | | | | | | | | | | | ✓ | | ✓ |
| 99 | Community service providers | | | | | | | | | | | | | | | |
| 100 | Pharmacy data | ✓ | | | | | | | | | | | | ✓ | ✓ | |
| 101 | Hospital ED and Admissions data | | | | | | | | | | | | | ✓ | | ✓ |
| 102 | Practice accreditation data from accrediting bodies | | | | | | | | | | | | | ✓ | | |
| 103 | Australian Collaboratives data set | | | | | | | | | | | | | | | |
| 104 | Public Health Information Development Unit (PHIDU) – Social Health Atlases | | Torrens | | | | | | | | | | | | | |
| 105 | NT Community Care Information System (CCIS) | | | | | | | | | | | | | ✓ | | |
| 106 | NT Shared Electronic Health Record (eHealth) | | | | | | | | | | | | | ✓ | | |

**Note:** When 'Respondent owned or used' is not ticked that refers to a dataset named but not used by a respondent.

## 7.8. Data Quality Frameworks identified by respondents

*Partially blinded*

| Dataset | Data Quality Framework or tool |
|---|---|
| A PHN | • Benchmarking reports for practices on data quality (used in the past and are being developed). |
| A PHN | • The PHN has data governance and privacy policies and procedures in place. |
| APHN | • Alignment to Data 61/CSIRO The De-identification Decision Making Framework,<br>• Local protocols |
| A PHN | • Tool: PEN Computer Systems suite of data extraction and analysis.<br>• Framework: A specific Data Quality management framework |
| A PHN | • Tools embedded in the PEN software and some of our own validation. |
| A PHN | • We have quality improvement frameworks to help improve the quality of general practice data in the practice |
| A PHN | • Work closely with General Practice to improve data quality in accordance to accreditation using Plan Do Study Act (PDSA) activities, in line with the continual quality improvement framework |
| ASPREN | • Training for GPs and practices.<br>• Audits of data quality with cross checking of results.<br>• External review of results by the Department of Health and other users<br>• External review of CPD activities |
| BEACH | • For detail see the Introduction and Methods of General Practice Activity in Australia 2015-16, and the Decade report https://www.sydney.edu.au/medicine-health/our-research/research-centres/bettering-the-evaluation-and-care-of-health.html<br>• The GP report references many methodological papers resulting from the step wise development of reliability and validity of the BEACH methods over the 20 years prior to its start in 1998, and during the 18 years of the program. |
| A GP practice dataset | • PEN CAT tool is used to audit the data for quality improvement<br>• Secondary use of data = upload to My Health Record and PHN Quality Improvement Program |
| MABEL – health workforce | • Provide a user manual on how the data are collected and cleaned to ensure a high-quality dataset. This framework we developed on the basis of what is done with the HILDA survey. |

| | |
|---|---|
| NPS MedicineInsight | • ABS Data Quality Framework seven dimensions of data quality: institutional environment, relevance, timeliness, accuracy, coherence, interpretability, accessibility.<br>• The quality of data are initially checked with sentinel practices, to ensure correct identification of patients, relevant medicines, tests, conditions and other relevant data elements<br>• Using the National Clinical Terminology Service (NCTS), operated by the Australian Digital Health Agency to ensure the dataset is using the national clinical terminologies such as Australian Medicines Terminology (AMT), LOINC and SNOMED CT-AU<br>• Health Professional Learning teams work with practices to improve the quality and completeness of patient records. Practices receive routine feedback on data quality, including completeness of records, as part of practice reports. |
| Patron - Data for Decisions | • Custom framework based on the conceptual model in Kahn et al's 2016 paper: 'A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data' [7, 8]<br>• Published Data Governance Framework and unpublished Standard Operating Procedures |
| POLAR Data Space | • Governed by the principles of an organisational data governance program (published)<br>• The 10 commandments of Health Data (unpublished) |
| Voluntary Indigenous Identifier (VII) | • A table in the Medicare Database<br>• The Department of Health Data Governance Framework<br>• Part of the Medicare dataset managed under the National Health Act and the Health Insurance Act. |

## 7.9. How primary care datasets could be better used

Summary of responses from 53 responders to the survey question: How do you think that primary care datasets could be better used to support improved health outcomes?

| Theme | How primary care data could be better used to support improved health outcomes |
|---|---|
| **Research solutions** | • Use the data for research that aims to improve health outcomes and will provide evidence / demonstrate outcomes<br>• Improve researcher access to / availability of the data<br>• The big dataset(s) can provide powerful research outcomes that are value for money – pending appropriate and rigorous analysis and interpretation<br>• Use the data to identify service use, need, to target resources and focus on disease forecasting, early intervention and disease prevention<br>• Enhance research through the development of a practice-based research network<br>• Potentially powerful research outcomes that provide value for money<br>• Learn more about prevalence and incidence of chronic disease and comorbidity → inform policy decisions<br>• Use for randomised controlled trials<br>• Share publications/ citations that have used primary datasets to encourage other researchers to use them |
| **Data linkage** | • Ensure linkage of primary care data with other services<br>• Use data linkage to examine how primary care data relates to other parts of the health system, reveal gaps and opportunities for improvements |
| **Technical data solutions** | • Have all GPs 'dump' their data into a centralised database (i.e. larger datasets are required for the data to be useful)<br>• Ensure that a unique person identifier (e.g. IHI) or consistent statistical linkage key is used that follows patient through healthcare, Centrelink, housing etc to enable effective data linkages while remaining de-identified<br>• Use Healthcare Identifiers (e.g. HPI-O) to increase integration / linkability of data.<br>• Improve data quality and reliability so that it has 'at least some features of computable/analysable data'<br>• Data first needs to be standardised because comparability of data from bespoke collections is not known / Standardisation of clinical software systems<br>• Consistency of Data Dictionaries<br>• More timely use of data – e.g. using real time or current data for disease surveillance and identification of people 'at risk' of problems |
| **Surveillance and monitoring** | • Population health surveillance and monitoring including identification of under-utilisation<br>• Monitor use of antimicrobials as an effort to reduce antimicrobial resistance<br>• Monitor patient outcomes as a result of health service changes and changes in clinical management of conditions |
| **Inform / support general practice** | • Inform clinical decision-making guidelines<br>• Focus on tools that support clinicians to understand their own data and automate clinical recommendations<br>• Audit and feedback to practices including about their 'performance' in relation to other practices<br>• Educate general practice in how datasets can be used<br>• Have more quality improvement tools embedded within practice software<br>• Have all PHNs provide structured and locally tailored quality improvement initiatives based on local primary care data |
| **Health promotion** | • Identify the top 5 dispensed drugs for chronic conditions and promote health projects to improve those conditions<br>• Consolidate / aggregate GP datasets to provide 'big picture' view of the nation's health |

| | |
|---|---|
| **Policy, governance and system change** | • Ensure a minimum dataset (core fields) is available from general practice → use the data to improve health outcomes<br>• Clearly define 'ownership', strong governance structures and approval processes, to support agreement from Data Custodians to engage in data linkage and data sharing<br>• Require research projects to clearly define aims and objectives that fit within PHN Performance Framework health outcomes<br>• Provide data to PHNs that is more appropriately able to help them meet their Program Performance and Quality Framework<br>• Use data to feed into development of a Learning Healthcare Systems[15]<br>• Work collaboratively to deliver<br>• Mandate capture of diagnosis for each consultation and referral documentation |

[15] Learning Healthcare System: https://www.sciencedirect.com/science/article/pii/S1532046416301319

## 7.10. Limitations of secondary use of primary care data

Summary of responses to the survey question: What do you think are the limitations of secondary use of primary care datasets?

| Themes | Summary of limitations |
|---|---|
| **Poor data quality, reliability and technical limitations** | • Data inaccuracies, incompleteness, variability, poor coding by clinicians – at practitioner level and between practices<br>• SNOMED-CT not designed for use in primary care: gap between the terms used by GPs and those included in SNOMED, and terms changing over time<br>• Clinical software systems not designed for research purposes<br>• Reliability of data extraction tools not assessed or transparent → relates to lack of transparency around data quality of some primary care datasets<br>• Duplication of patient records (patients 'doctor shopping')<br>• Limited scope of data when only certain fields shared / obtained, data lacking granularity and frequency of download (not meet users' needs).<br>• Differences in available clinical data capture systems → different data quality<br>• Use of probability linkage in analysis – e.g. prescription of Metformin → assumption of diabetes when it may have been prescribed for PCOS<br>• Presence of free text making some data unusable (or too resource intensive to use) |
| **Poor data utility** | • De-identification of data, which can include its aggregation, limiting its utility<br>• Lack and difficulty of linkage to data from other parts of the healthcare system (linkage may not be permitted)<br>• Lack of a minimum primary care / general practice dataset<br>• Lack of patient unique identifier |
| **Lack of understanding of data complexity and its context** | • It is complex to use appropriately<br>• Data end-users not understanding social and clinical context required to appropriately interpret the data → poor interpretation of data<br>• Complex general practice workflows for data processing negatively impacting on data quality<br>• Secondary use of data collected for another purpose is the limitation (data should be collected to use in one way or another) |
| **Unequal data representativeness** | • Lack of data on priority populations CALD, aboriginal communities, under-representation of vulnerable groups, over-representation of the worried well |
| **Difficult to access** | • Some general practices use paper-based records<br>• Insufficient primary care data available for secondary use (lack of GP consent to data sharing)<br>• 'Arduous' nature of data linkage permissions and processes (only getting part of the picture) |
| **Privacy concerns, trust and ownership (an access issue)** | • Lack of community consultation on its use<br>• Poor privacy and consent mechanisms for clinicians sharing data for secondary use<br>• Varying perceptions on who believes they own the data<br>• When shared data are stored off-shore (not always made clear at the outset) |

| | |
|---|---|
| **Lack of guidelines, policies, standards and 'common data model'** | • Lack of national standards for general practice data quality and evaluation – lack of common data model<br>• Lack of interoperability between clinical data capture systems – too many clinical data management systems<br>• Lack of standardised data extraction tools<br>• Lack of standard coding and common terminology<br>• Lack of management and leadership to improve data standards<br>• Legal blocks and governance issues |

The above summary of themes is representative of comments provided by 53 survey responses.

## 7.11. Benefits of secondary use of primary care data

Summary of responses from 53 responders to the survey question: What do you think are the benefits of secondary use of primary care datasets?

| Benefit Type | Summary of respondents' suggestions |
|---|---|
| **Improvements to provision of care and health outcomes** | • Enable greater knowledge to assess and improve services, understand treatment outcomes and improve population health<br>• Understanding service needs at **local levels** (right services in the right place) → targeted services and equity in primary health care<br>• Driver of service quality improvement (through competitive benchmarking)<br>• Lead to evidence-based investments and interventions<br>• Inform national policy (related to health needs of communities) and workforce planning<br>• Contribute to development of 'Learning Healthcare Systems'[16] that enhance conversations about quality of healthcare |
| **Direct, pragmatic research benefits and efficiencies** | • Rapid and cost effective → cost reductions and increased research scale (including not having to recruit individual patients for research)<br>• Shedding light on aspects of primary care that we would otherwise know nothing about<br>• Access to data representing more people than other sources → increasing statistical power of the<br>• Use the data to generate important hypotheses / research questions |
| **Assist with policy and planning related to service provision** | • Understanding, identifying, measuring trends, needs and how patient cohorts access and use health services<br>• Monitoring and evaluating interventions and changes in service provision<br>• Evaluation and management of chronic illness / facilitating disease incidence and prevalence studies<br>• Data for risk stratification, risk management and preventive care (predictive modelling for hospital avoidance)<br>• Provides indicators of quality of care, drug safety, use of medicines, effectiveness of health policy, health care delivery and disease risk factors<br>• Facilitating tracking disease outbreaks – surveillance, including drug reactions, device recalls.<br>• Generating efficiencies in health spending<br>• Provide an evidence-base for investment in tech infrastructure across the health system for effective and efficient care |
| **Technical application** | • Potential for add-on patient generated data collection through apps etc |
| **Practice level** | • Enable general practices to review their activities and make business improvements<br>• Ability to track and improve patient outcomes |
| **Intrinsic benefits of the data** | • Unique, rich, granular and very big dataset that has great population representation (minimises measurement bias in research)<br>• 'Real world' data / population-based evidence<br>• Makes accessible regional-level information<br>• Most people visit general practice in a 5-year period → there is more population health information than any other health data source<br>• When linked to other data:<br>  ○ Systems view / triangulation, creating a 'patient centred view' of data – greater understanding of care pathways, patient needs and service gaps |

[16] Learning Healthcare System: https://www.sciencedirect.com/science/article/pii/S1532046416301319

## 7.12. Barriers of secondary use of primary care data

Summary of responses from 53 responders to the survey question: What are the barriers to better use of primary care data?

| Barrier Type | Summary of barriers identified |
|---|---|
| **Fear and reticence** | • GP concerns about patient privacy impacting ability to collect data<br>• Fear of privacy breaches, 'illegal use of data', poor data security, reputational and financial damage to practices that share data<br>• GP and peak body perceptions that sharing data will lead to government control of GPs<br>• GP unwillingness to share data: Clinicians not seeing value in secondary use |
| **Leadership, governance & ethics constraints** | **Leadership and regulatory issues**<br>• Lack of identification of who is 'in charge' of primary care data; lack of national leadership; federal-state divide<br>• Lack of 'national approach' to data collection and a national data repository<br>• Lack of, or confused, determination of 'who owns the data'<br>• Lack of GP leadership<br>• Lack of regulatory requirement for general practice to share data (the barrier of voluntary sharing of data)<br>• Lack of relationships/engagement between key stakeholders: e.g. Difficulty engaging GP clinical software vendors, research institutions not coordinating effort to optimise data use<br>**Ethics and governance**<br>• Stringent ethical constraints, data governance protocols, data access controls and confidentiality restrictions as barriers to access: e.g. the need to de-identify data, issues of gaining consent/permission to use data<br>• Increasing recognition of ethics and governance complexities coupled with lack of technical knowledge → research being inappropriately enabled or constrained (e.g. by ethics committees)<br>• Lack of transparency of consent models, governance processes and methodologies leading to lack of trust in data sharing<br>• Cumbersome, slow, expensive processes for ethics and Data Custodian approvals to access data<br>• Lack of clarity on what is 'deidentified' data and legal, ethical & governance issues related to sharing such data without explicit patient consent |
| **Lack of data availability** | • Inaccessibility of data for secondary use due to lack of availability and cost barriers<br>• Researchers not knowing what data are available and how to access it<br>• Limited data to support priority populations – e.g. CALD, aboriginal<br>• Poor quality data entry by health professionals – missing data<br>• Lack of longitudinal data at patient level<br>• People who can "establish the best outcomes" not having access to the data<br>• Privatisation of secondary use of data by providers causing a cost barrier |
| **Lack of expertise, experience & motivation** | • Too few clinicians involved in planning data analyses and in reaching research conclusions<br>• General practice staff lacking perception of need / motivation to collect clean, accurate and complete data<br>• Lack of shared vision and capacity to build a Learning Healthcare System that uses many sources of data (not just primary care data) |

| | |
|---|---|
| **Barriers to data linkage** | • Inability of PHNs to link patient data a barrier to better use: "Not being able to link datasets to review patient journeys"<br>• Lack of a reliable IDs for data linkage<br>• Lack of availability of and access to some datasets needed for linkage (stakeholder agreement) |
| **Technical systems barriers:**<br><br>**(Lack of systems to improve data quality and quantity)** | • Lack of structured, standardised EMRs (standardised coding, classification, data definitions, basic data structures across different clinical systems)<br>  ○ Lack of accreditation of GP clinical software vendors → lack of standardisation → poor data quality → decreases data utility<br>  ○ SNOMED is incompletely mapped to lists of terms used by GPs and not all GP vendor clinical software systems are mapped to SNOMED<br>  ○ Lack of interoperability/variability of data extraction tools (PEN CAT, POLAR, GRHANITE) → lack of standardisation/interoperability of data extracted using different tools<br>• Data extractions tools unable to extract from all GP vendor clinical software systems and the lack of standardisation of resulting datasets<br>• Relatively few providers of data warehousing<br>• Limitations and poor data quality: lack of data completeness, cleanliness, granularity, difficulty accessing individual level data.<br>• Poor My Health Record system: "Start again with My Health Record"<br>• Lack of the minimum data set (MDS) and variability of definition of MDS |
| **Health system and research barriers** | • The small enterprise / private business structure of primary care<br>• Patient freedom to attend multiple practitioners → lack of longitudinal patient records<br>• Voluntary data sharing: Not all GPs submit monthly data (lack of complete and current data)<br>• Workforce: High staff turnover in general practice → impacts negatively on data input and data quality<br>• PHN focus on reporting to government rather than focus on serving the population<br>• Timelines; slow release of data – e.g. waiting for MBS dataset release<br>• Funding / Cost: Systematic lack of funding to **collect** and **analyse** data from general practice and to support the **implementation** of findings<br>  ○ Lack of resources/motivation/capacity/education to prioritise data input / improved data quality (accuracy & completeness) at general practice<br>  ○ Lack of resources for research – including for PHNs – to interpret data from health planning perspective and to drive quality improvement<br>  ○ Lack of funding for the academic GP sector for research and skills/capacity building. |

## 7.13. Enablers of secondary use of primary care data

Summary of responses from 51 responders to the survey question: What are the enablers to better use of primary care data?

| Enabler Type | Summary of respondents' suggestions:<br>Enablers of ways to better use primary care data are… |
|---|---|
| **Qualities** | • Trust (GPs and the public need to trust data custodians and users) – reassurance from Data Custodians that data is not being used for unintended purposes<br>• Transparency in data access and use<br>• Growing understanding, awareness and knowledge of the value and application of primary care data<br>• Shared vision<br>• Innovative, forward thinking solutions<br>• Altruism (the desire to share data for the greater good)<br>• Affordable researcher access to data |
| **Leadership from relevant institutions** | • PHNs (because they have relationships with and access to general practice)<br>• Universities (to provide leadership, expertise, professional engagement and general information to promote secondary use)<br>• GP Colleges<br>• NPS MedicineWise (MedicineInsight program) |
| **Governance** | • Unambiguous / agreed strategic framework(s) and clearer processes for data access and use<br>• Robust safeguards (ensure benefits outweigh risks) – including infrastructure / resources for audit of data security and governance<br>• Clear government direction and tighter governance (so data providers have a clearer line of sight of the data)<br>• Centralised coordination and management of GP data / an Australian CPRD (UK's Clinical Practice Research Datalink)<br>• National adoption of a single GP data extraction tool (to decrease duplication of effort)<br>• QI-PIP* (government Quality Improvement Practice-Incentive Payments) to enable improvement of data quality<br>• Clear steps on how to deidentify data, classification of levels of deidentification, including when it stops being deidentified, and what can be done with the various levels of 'deidentified' data |
| **Partnerships and capacity building** | • Engagement between key stakeholders: clinicians, consumers, government, researchers, PHNs<br>• Turn data into knowledge and deliver it back to clinicians for business and care improvement → open pathways for greater secondary use<br>• Cross-sector capacity (e.g. increasing the will to enable data linkage)<br>• Enabling health professionals to benefit from review / audit of their own data → enable better use<br>• Having patients / the public understand the research value of primary care data and support its use (especially when linked to other datasets)<br>• Build on existing public expectation that policy-makers are already using linked data systems to improve services<br>• Develop a research-oriented network for negotiating access to primary care data<br>• Workforce capacity building to use data in research e.g. Data Custodians providing researchers and GPs with support to access and interpret data |

| | |
|---|---|
| **Technologies and method development** | • Advancement in computing hardware and software technology<br>  ○ Interoperability of data collection tools or adopting a single extraction capable of working across multiple vendor software packages<br>  ○ Re-design of clinical data collection software with interoperability and data extraction in mind<br>  ○ Improved portals for practice display of data to encourage continuous quality improvement in data capture and service knowledge<br>• Robust data storage / IT security<br>• Cross-sector technical capability (e.g. the technical capability to enable data linkage)<br>• Data standardisation<br>• Coding consistency and standardisation<br>• Systematic data quality assurance<br>• Aggregation of datasets at state or federal levels<br>• Mechanisms for appropriate data interpretation – e.g. clinician data coders and analysts (to prevent misinterpretation and scandalisation |
| **Resources** | • Funding for:<br>  ○ practices to invest in training, quality improvement, and better use of their own data<br>  ○ PHNs to drive data quality in practices and for them to better interpret the data<br>  ○ curation of accessible and affordable data for secondary use<br>  ○ sustained effort to build a single, properly governed dataset<br>• Incentives (includes incentives for quality data entry by GPs/ QI-PIP, practices investing in the use of their own data for monitoring)<br>• Good extraction tools<br>• Skilled workforce / training: clinicians with health informatics expertise, experienced computer scientists and health informaticians (to support practices to capture quality data and enable use of primary care data)<br>• Training for primary care staff: to demonstrate benefits, educate about data value, teach best practice data collection<br>• Time (time to demonstrate good outcomes)<br>• Increasing access to primary care data<br>• My Health Record (a resource to enable better use of primary care data) |

* The government QI PIP scheme was not introduced at the time of the survey – it was 9 months after survey close that it was introduced in August 2019


Department of General Practice, The University of Melbourne, Victoria 3010, Australia

www.gp.unimelb.edu.au

Copyright 2020