

Using primary care data for research: What are the issues and potential solutions?

**Ron Cheah, Rachel Canaway,
Christine Mary Hallinan,
Lucas de Mendonça,
Jo-Anne Manski-Nankervis**

This article is part of a longitudinal series on research.

Background

Interest in using primary care data for research is growing with increasing recognition of its potential for improving healthcare. Many issues exist, some inherent in the data and others external.

Objective

This paper explores the main issues associated with the use of primary care data for research and proposed solutions to address them.

Discussion

Issues related to the use of primary care data for research are complex. Government reimbursement system administrative data have limitations as they lack clinical detail. General practice electronic medical record data are more suitable; however, challenges include variable data quality and interoperability. There are concerns from general practices and the public about data access and use. Strategies to address these issues include incorporating best-practice principles, implementing standards and data quality frameworks, creating partnerships between data custodians and ensuring robust governance systems exist. Leadership and the will of key stakeholders to reform, with governmental support in implementing required actions, must be prioritised.

INTEREST in secondary use of primary care data for research is growing, as evidenced by the formal establishment of government-led data collection initiatives for research (NPS MedicineWise, www.nps.org.au; The Health Improvement Network, www.the-health-improvement-network.com/data-users; Clinical Practice Research Datalink, <https://cprd.com>) and the increasing number of research publications utilising these data over time. Primary care data, used for non-clinical purposes (secondary use), is typically gathered from administrative and clinical sources through data-sharing agreements with the original data holder. Administrative data includes ambulatory care service and medication dispensing reimbursement data captured from the Australian Medicare Benefits Schedule (MBS, www.mbsonline.gov.au/internet/mbsonline/publishing.nsf/Content/Home) and the Pharmaceutical Benefits Scheme (PBS, www.pbs.gov.au/pbs/home). Clinical data can be obtained by government agencies, universities or other organisations directly from electronic medical record (EMR) systems embedded into clinical information systems in general practices.¹ EMRs have been part of general practice in Australia for decades, with large volumes of data continuously generated and stored.^{2,3} In addition to clinical care, primary care data can be used for a multitude of research activities, such as longitudinal cohort, interventional and comparative studies, big data analytics for randomised controlled trials and predictive modelling.⁴ However, these data are currently underutilised for research

in Australia compared with similar countries, despite the known positive effects,^{4,5} as acknowledged by Australia's Productivity Commission.^{6,7}

Aim

This article aims to explore issues associated with the use of primary care data for research in Australia – in particular data quality, interoperability, linkage and access – and propose solutions to address them.

Issues with primary care data

Administrative data from both the MBS and the PBS have very good coverage and quality; however, they do not contain the necessary detail required for a broad range of primary care research.¹ Details such as patient diagnoses, test results, observation measurements and prescribing instructions, which provide the clinical context needed to answer research questions, are absent from administrative data.^{1,8} This was evident in early studies that linked MBS and PBS data to state healthcare datasets to examine the effects of primary care on hospitalisations and mortality.^{4,9,10} Due to the limited clinical information available, many assumptions needed to be applied to derive meaning from the findings.⁴ Clinical data from general practice EMR systems are more suitable for this purpose; however, these data also carry inherent issues.

EMRs used in Australian general practice were primarily designed to improve administrative and clinical workflows, including Medicare claims management.

The use of captured data for research was not a design consideration.⁴ These EMR systems were developed independently with unique schema for medical terminology and clinical coding, thus preventing direct interoperability,^{4,11} and a reliance on free-text as opposed to coded data entry.¹ The absence of standardised data practices has resulted in inconsistent approaches to storing and reporting information for secondary use. This has led to suboptimal data quality because systems often allow unstructured free-text entries rather than coded ones.^{1,4,11}

Lack of interoperability between EMR systems and data extraction tools and the absence of accreditation to ensure data are standardised to a common data model contribute to varying formats of, and repositories for, data storage.^{1,4,12} As a result, there are challenges when research requires information aggregation of data across practices using different EMR systems and data extraction tools.^{1,4,13} Furthermore, a widely used commercially developed data extraction tool has been described as ‘a barrier to better use of primary care data’ due to its associated inflexible legal and data governance arrangements.¹³

Access to primary healthcare data and linked datasets are major issues for research. General practice EMR data are regularly collected by Primary Health Networks (PHNs), Australian government-funded independent organisations whose role is to assess primary and community healthcare, report to government and commission services for quality improvement purposes.¹⁴ Data gathered are used for quality improvement activities (ie performance feedback) and to inform health service planning and policy development.

Given established pathways for PHN use of EMR data, access to these data for research purposes within the university sector is not as streamlined.^{1,13} Research involving primary care data is often carried out in ‘research silos’, thereby limiting opportunities for ‘big picture’ research collaborations. Data access barriers also limit the use of EMR data for research. These barriers include the protracted time to gain approval from data custodians and for data access once approval is gained.¹³ The reticence of general practitioners and other holders of primary care data to share it can also be attributed to

a general lack of trust linked to fears around potentially poor data security and privacy, questions regarding ownership of data once shared and reputational and financial damage should there be any data breach.⁵ Financial constraints might also prevent secondary use: access fees imposed by custodians might range from a modest flat fee to many tens of thousands of dollars.⁵

Potential solutions

Addressing the issues associated with secondary use of primary care data requires a comprehensive approach, as these issues are multilayered at the data, technology and system levels. The development and application of clinician-, researcher- and consumer-agreed best practice principles for appropriate use of health data are needed to ensure healthcare provider and public trust. Best practices include de-identification of data before it is extracted into a repository, governance committees independent of the data custodian/managers to make decisions about use of the data on behalf of the public, transparent governance processes, robust security systems and provision of the minimum required information to answer research questions.^{4,15-19}

Issues with data quality might be addressed using a workforce approach. Primary care workforce training for best practice data collection has been in place since the rollout of the Australian Government-funded Practice Incentives Program (PIP) Quality Improvement (QI) Incentive;²⁰ notwithstanding its successes, PIP QI is limited by EMR design. Increased provision of clinician health informatician support to ensure appropriate data capture and interpretation⁵ and improved data collection tools that focus on data quality and continuity^{21,22} might also be helpful.

To improve data quality output from EMR systems, a suite of standards must be adopted, owned and implemented at scale. These include defined data models that establish linkages between related data elements, consistency between data element labels and definitions, use of standardised clinical terminologies and classifications, and the introduction of an accreditation process for quality assurance.^{4,11} Data models that support high-quality care already exist; these include HL7 FHIR (www.hl7.org/fhir) and openEHR

(<https://openehr.org>). Additionally, widely used standardised clinical terminologies (SNOMED CT and Australian Medicines Terminology; www.healthterminologies.gov.au/access-clinical-terminology/access-snomed-ct-au) have been mapped to the *International classification of diseases, 10th revision*.²³ The incorporation of data quality frameworks, such as Kahn’s harmonised data quality framework,²⁴ also enables rigorous assessments of data quality, including fitness for purpose assessments, to be performed.¹³ For research use, interoperability challenges between EMR systems and extraction tools can be addressed by mapping data to common data models, such as the Observational Medical Outcomes Partnership Common Data Model,²⁵ and ensuring data extraction packages are capable of working across multiple software packages.⁵

Improved data linkage for research can be attained by establishing accountable partnerships between the various stakeholders, such as universities, PHNs and government.¹³ Provision of incentives and additional funding should also be considered to encourage the sharing of data between these entities. Such partnerships can enable the possibility of a centralised coordination model for primary care data; this will improve research capacity through improved data quality, timely access, reduced duplication of effort and the ability to link to gold standard datasets. Concerns of privacy loss associated with linkage can be mitigated by ‘privacy-preserving record linkage’,²⁶ which involves irreversibly coding patient identifiers prior to extraction and linkage.^{1,4} EMR de-identification, where all patient and provider identifiers in the data are removed, enables data within the EMR to be used or shared in ways that might not otherwise be permitted under the *Privacy Act 1988*.²⁷ Privacy concerns pertaining to public and healthcare provider trust in the secondary use of data in research need further consideration. Consistency and transparency in governance systems in research, including the provision of secure research environments, researcher contractual obligations, sharing of data breach risk mitigation and management strategies with consumers, mandatory research training and proactive standard operating procedures, are necessary to gain this trust.¹³ Effective communication of this information is equally

important to allay fears, especially around data security and sharing availability and preferences.¹³ The Royal Australian College of General Practitioners' checklist for the secondary use of de-identified data²⁸ and guiding principles for managing requests for the secondary use of de-identified data²⁹ are valuable resources that will help general practices manage requests for access to their data. These documents can empower healthcare providers to make informed decisions regarding their EMRs and to overcome any initial doubts or concerns they might have with research-related data requests.

Steps have been taken by the Australian Health Research Alliance's Transformational Data Collaboration³⁰ to address some of these issues: improving health data useability through the development of tools and methods to improve data integration and harmonisation; and increasing user capacity by providing cost-free common data model training for researchers.^{5,30} The success of this initiative relies on widespread professional, consumer and vendor support, along with the establishment of clear and enforced timescales and, potentially, the provision of regulatory incentives to break the status quo. Leadership from key stakeholders (professional bodies, universities, primary health networks and data custodians) with governmental support and funding is required to enable a national, cohesive approach to the development and implementation of standards for general practice EMRs and improve data quality.¹¹ Policy and governance reforms to improve access and linkage between practice and research will enable the aforementioned 'big picture' collaborations through integration of data currently housed in different repositories and more fluid data sharing.⁵ This will improve the current poor data utilisation and reduce the inefficiencies and unnecessarily high economic burden of duplication of effort.¹³

Conclusion

Primary care data are a rich source of information that can contribute to healthcare improvement through research. Unfortunately, many challenges hinder the optimal use of these data. Issues include challenges with data quality and access and

data custodian fears of compromised privacy. Strategies to address these matters include incorporating evidence-based principles of best practice, implementing EMR system standards and data quality frameworks, creating accountable partnerships between data custodians, ensuring the transparency of professional and consumer input and having robust governance systems in place. Leadership from key stakeholders with governmental support in implementing standards across EMR systems and national legislation to ensure harmonisation of health data use must be prioritised.

Key points

- The use of general practice EMR data provides opportunities to undertake large-scale observational research. Poor data quality, limitations in the necessary structures to facilitate interoperability, lack of implementation of best practice for the capture of an 'optimal' dataset, linkage barriers, privacy concerns and limits to access all need to be overcome to facilitate appropriate use.
- Administrative data derived from national healthcare reimbursement schemes contain robust data; however, these data lack the clinical detail required for clinical-related primary care research.
- The application of best practice principles for the appropriate use of health data for research is crucial to establish and maintain trust among data custodians and the public to ensure continued access to data for research.
- Effective leadership from professional bodies, universities, primary health networks and data custodians, along with governmental support, are required to drive the necessary changes to address primary care data issues.
- Australia is said to 'lag behind' comparable countries in the secondary use of health data.

Authors

Ron Cheah BPharm, GCertPharmPrac, GCertHSM, MPharmPrac, FSHP, FANZCAP, Senior Pharmacist, The National Centre for Antimicrobial Stewardship, Melbourne, Vic

Rachel Canaway BHSc, MSocHlth, PhD, Research Fellow, Department of General Practice and Primary Care, The University of Melbourne, Melbourne, Vic

Christine Mary Hallinan BASc, MPH, PhD, GD-Biostat Research Fellow, HaBIC Research Information Technology Unit, Department of General Practice and Primary Care, Faculty of Medicine, Dentistry and Health Sciences, The University of Melbourne, Melbourne, Vic

Lucas de Mendonça MD, General Practitioner; The University of Melbourne, Melbourne, Vic

Jo-Anne Manski-Nankervis BSc (Hons), MBBS (Hons), CHIA, PhD, FRACGP, Associate Professor, Department of General Practice and Primary Care, The University of Melbourne, Melbourne, Vic; Academic General Practitioner, Australian Centre for Accelerating Diabetes Innovations, The University of Melbourne, Melbourne, Vic

Competing interests: RCanaway and CMH work in The University of Melbourne's HaBIC Research Information Technology Unit, which manages technical components of the Patron program of research and undertakes research using the dataset. RCanaway reports consultancy fees paid by the Population Health Research Network to the University of Melbourne for a project to better understand data linkage in Australia; and involvement, between 2017 and 2021, in the development, implementation and administration of The University of Melbourne's Data for Decisions and Patron program of research, which extracts and curates de-identified general practice data for secondary use. J-AM-N declares receipt of personal fees as Chair of the RACGP Expert Committee – Research, which includes advocacy for funding for general practice research; and funding from the Medical Research Future Fund; and funding from the National Health and Medical Research Council. J-AM-N is Director of Torch Recruit Pty Ltd, which has developed software to identify people who are potentially eligible to participate in clinical trials; Chair of the Data Management Committee for the Patron program of research, a general practice data repository managed by The University of Melbourne's Department of General Practice and Primary Care; and has been a member of the advisory committee for NPS MedicinesInsight annual reporting. RCheah and LdeM have no conflicts of interest to disclose. Funding: None.

Provenance and peer review: Commissioned, externally peer reviewed.

Correspondence to: ron.cheah@unimelb.edu.au

References

1. Youens D, Moorin R, Harrison A, et al. Using general practice clinical information system data for research: The case in Australia. *Int J Popul Data Sci* 2020;5(1):1099. doi: 10.23889/ijpds.v5i1.1099.
2. McInnes DK, Saltman DC, Kidd MR. General practitioners' use of computers for prescribing and electronic health records: Results from a national survey. *Med J Aust* 2006;185(2):88-91. doi: 10.5694/j.1326-5377.2006.tb00479.x.
3. Henderson J, Britt H, Miller G. Extent and utilisation of computerisation in Australian general practice. *Med J Aust* 2006;185(2):84-7. doi: 10.5694/j.1326-5377.2006.tb00478.x.
4. Canaway R, Boyle DI, Manski-Nankervis JE, et al. Gathering data for decisions: Best practice use of primary care electronic records for research. *Med J Aust* 2019;210 Suppl 6:S12-16.
5. Canaway R, Boyle D, Manski-Nankervis JA, Gray K. Identifying primary care datasets and perspectives on their secondary use: A survey of Australian data users and custodians. *BMC Med Inform Decis Mak* 2022;22(1):94. doi: 10.5694/mja2.50026.
6. Australian Government Productivity Commission. Data availability and use: Productivity Commission

- inquiry report no. 82. Productivity Commission, 2017. Available at www.pc.gov.au/inquiries/completed/data-access/report/data-access.pdf [Accessed 27 October 2023].
7. Australian Government Productivity Commission. 5-year productivity inquiry: Advancing prosperity, vol. 1, inquiry report no. 100. Productivity Commission, 2023. Available at www.pc.gov.au/inquiries/completed/productivity/report [Accessed 9 October 2023].
 8. Monaghan T, Biezen R, Buising K, Hallinan C, Cheah R, Manski-Nankervis JA. Clinical insights into appropriate choice of antimicrobials for acute respiratory tract infections. *Aust J Gen Pract* 2022;51(1-2):33-7. doi: 10.31128/AJGP-07-21-6073.
 9. Einarsdóttir K, Preen DB, Sanfilippo FM, Reeve R, Emery JD, Holman CDAJ. Mortality in Western Australian seniors with chronic respiratory diseases: A cohort study. *BMC Public Health* 2010;10(1):385.
 10. Einarsdóttir K, Preen DB, Emery JD, Holman CDAJ. Regular primary care decreases the likelihood of mortality in older people with epilepsy. *Med Care* 2010;48(5):472-76. doi: 10.1097/MLR.0b013e3181d68994.
 11. Gordon J, Miller G, Britt H. Deeb Institute Issues Brief No. 18. Reality check - Reliable national data from general practice electronic health records. Australian Healthcare and Hospitals Association, 2016. Available at <https://apo.org.au/sites/default/files/resource-files/2016-07/apo-nid65550.pdf> [Accessed 13 June 2023].
 12. The Royal Australian College of General Practitioners (RACGP). Secondary use of general practice data. RACGP, 2017. Available at www.racgp.org.au/download/Documents/e-health/Secondary-use-of-general-practice-data.pdf [Accessed 13 June 2023].
 13. Canaway R, Boyle D, Manski-Nankervis J, Gray K. 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. Department of General Practice, The University of Melbourne, 2020. Available at https://machaustralia.org/wp-content/uploads/2021/05/Primary-care-data-and-linkage_Dec-2020.pdf [Accessed 13 June 2023].
 14. Australian Government Department of Health and Aged Care (DHAC). What primary health networks do. DHAC, 2022. Available at www.health.gov.au/our-work/phn/what-PHNs-do [Accessed 13 May 2023].
 15. Stockdale J, Cassell J, Ford E. 'Giving something back': A systematic review and ethical enquiry into public views on the use of patient data for research in the United Kingdom and the Republic of Ireland. *Wellcome Open Res* 2019;3:6. doi: 10.12688/wellcomeopenres.13531.1.
 16. Kim KK, Joseph JG, Ohno-Machado L. Comparison of consumers' views on electronic data sharing for healthcare and research. *J Am Med Inform Assoc* 2015;22(4):821-30. doi: 10.1093/jamia/ocv014.
 17. Spencer K, Sanders C, Whitley EA, Lund D, Kaye J, Dixon WG. Patient perspectives on sharing anonymized personal health data using a digital system for dynamic consent and research feedback: A qualitative study. *J Med Internet Res* 2016;18(4):e66. doi: 10.2196/jmir.5011.
 18. Williams H, Spencer K, Sanders C, et al. Dynamic consent: A possible solution to improve patient confidence and trust in how electronic patient records are used in medical research. *JMIR Med Inform* 2015;3(1):e3. doi: 10.2196/medinform.3525.
 19. Consumers Health Forum of Australia and NPS MedicineWise. Engaging consumers in their health data journey. CHF and NPS MedicineWise, 2018. Available at https://chf.org.au/sites/default/files/engaging_consumers_health_data_report_updated.pdf [Accessed 23 June 2023].
 20. Australian Government Department of Health and Aged Care (DHAC). PIP QI Incentive guidance. DHAC, 2021. Available at www1.health.gov.au/internet/main/publishing.nsf/Content/PIP-QI_Incentive_guidance [Accessed 23 June 2023].
 21. Institute of Medicine (US) Subcommittee on Standardized Collection of Race/Ethnicity Data for Healthcare Quality Improvement. Ulmer C, McFadden B, Nerenz DR, editors. Race, ethnicity, and language data: Standardization for health care quality improvement. National Academies Press, 2009.
 22. Chen H, Yu P, Hailey D, Cui T. Identification of the essential components of quality in the data collection process for public health information systems. *Health Informatics J* 2020;26(1):664-82. doi: 10.1177/1460458219848622.
 23. World Health Organization (WHO). International classification of diseases, 10th revision. WHO, 2019. Available at <https://icd.who.int/browse10/2019/en> [Accessed 9 October 2023].
 24. Kahn MG, Callahan TJ, Barnard J, et al. A harmonized data quality assessment terminology and framework for the secondary use of electronic health record data. *EGEMS (Wash DC)* 2016;4(1):1244. doi: 10.13063/2327-9214.1244.
 25. Observational Health Data Sciences and Informatics (OHDSI). Standardized data: The OMOP common data model. OHDSI, 2023. Available at www.ohdsi.org/data-standardization/ [Accessed 9 October 2023].
 26. Randall SM, Ferrante AM, Boyd JH, Bauer JK, Semmens JB. Privacy-preserving record linkage on large real world datasets. *J Biomed Inform* 2014;50:205-12. doi: 10.1016/j.jbi.2013.12.003.
 27. Office of the Australian Information Commissioner. De-identification and the Privacy Act. Canberra, 2018. Available at www.oaic.gov.au/privacy/privacy-guidance-for-organisations-and-government-agencies/handling-personal-information/de-identification-and-the-privacy-act [Accessed 23 June 2023].
 28. The Royal Australian College of General Practitioners (RACGP). Secondary use of de-identified data: A checklist for general practice. RACGP, 2019. Available at www.racgp.org.au/FSEDEV/media/documents/Running%20a%20practice/Security/Secondary-use-of-de-identified-data-A-checklist-for-general-practice.pdf [Accessed 13 June 2023].
 29. The Royal Australian College of General Practitioners (RACGP). Guiding principles for managing requests for the secondary use of de-identified general practice data. RACGP, 2019. Available at www.racgp.org.au/FSEDEV/media/documents/Running%20a%20practice/Security/Guiding-principles-for-managing-requests-for-the-secondary-use-of-general-practice.pdf [Accessed 13 June 2023].
 30. Melbourne Academic Centre for Health (MACH). Transformational data collaboration. MACH, 2021. Available at www.machaustralia.org/ahra-tdc [Accessed 22 June 2023].

correspondence ajgp@racgp.org.au