




Seamless EMR data access: Integrated governance, digital health and the OMOP-CDM

Christine Mary Hallinan ¹, Roger Ward,¹ Graeme K Hart,² Clair Sullivan,³ Nicole Pratt,⁴ Ashley P Ng ^{5,6}, Daniel Capurro,^{2,7} Anton Van Der Vegt,⁸ Siaw-Teng Liaw ⁹, Oliver Daly,² Blanca Gallego Luxan,¹⁰ David Bunker,⁸ Douglas Boyle¹

To cite: Hallinan CM, Ward R, Hart GK, *et al.* Seamless EMR data access: Integrated governance, digital health and the OMOP-CDM.

BMJ Health Care Inform 2024;**31**:e100953. doi:10.1136/bmjhci-2023-100953

Received 29 October 2023
Accepted 14 January 2024

ABSTRACT

Objectives In this overview, we describe the Observational Medical Outcomes Partnership Common Data Model (OMOP-CDM), the established governance processes employed in EMR data repositories, and demonstrate how OMOP transformed data provides a lever for more efficient and secure access to electronic medical record (EMR) data by health service providers and researchers.

Methods Through pseudonymisation and common data quality assessments, the OMOP-CDM provides a robust framework for converting complex EMR data into a standardised format. This allows for the creation of shared end-to-end analysis packages without the need for direct data exchange, thereby enhancing data security and privacy. By securely sharing de-identified and aggregated data and conducting analyses across multiple OMOP-converted databases, patient-level data is securely firewalled within its respective local site.

Results By simplifying data management processes and governance, and through the promotion of interoperability, the OMOP-CDM supports a wide range of clinical, epidemiological, and translational research projects, as well as health service operational reporting.

Discussion Adoption of the OMOP-CDM internationally and locally enables conversion of vast amounts of complex, and heterogeneous EMR data into a standardised structured data model, simplifies governance processes, and facilitates rapid repeatable cross-institution analysis through shared end-to-end analysis packages, without the sharing of data.

Conclusion The adoption of the OMOP-CDM has the potential to transform health data analytics by providing a common platform for analysing EMR data across diverse healthcare settings.

INTRODUCTION

The Observational Medical Outcomes Partnership Common Data Model

Adoption of the Observational Medical Outcomes Partnership Common Data Model (OMOP-CDM) internationally and in Australia has enabled the conversion of vast amounts of complex, and heterogeneous

electronic medical record (EMR) data into a standardised structured data model. The conversion of data has the potential to provide hospitals, health departments, auditors, regulators and universities valuable insights tailored to each institution's needs, both for operational and research purposes. This is achievable as long as the secure utilisation of an institution's EMR clinical and administrative data for purposes beyond its initial collection, known as 'secondary use', is effectively managed and employed.

Such data can be transformative, especially if used to monitor, evaluate and audit healthcare to improve clinical practice, reduce inefficiencies, contribute to the evidence base and develop a 'learning healthcare system' for improved patient care.¹⁻⁴ However, this potential is often not realised due to the inherent complexity of EMR databases—that comprise thousands of data elements across thousands of proprietary tables—where vast amount of data needs to be transformed, cleaned and restructured to make it 'fit' for 'secondary use'.⁵ For highly powered collaborative research, where large volumes of EMR data are combined, use is further constrained by the heterogeneity of each institution's EMR schema⁶; concern over data sharing and privacy breaches and lack of clarity over governance and consent.⁷

The Observational Health Data Sciences and Informatics (OHDSI) consortium⁸ is addressing these challenges through the transformation of each EMR database into the open-source OMOP-CDM, where EMR data elements are translated into the OMOP-CDM using standardised terminologies such as SNOMED-CT,⁹ LOINC¹⁰ or RxNORM.¹¹ Importantly, these transformed data are also able to be securely stored within their dedicated environment, complete with the



© Author(s) (or their employer(s)) 2024. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to

Dr Christine Mary Hallinan; hallinan@unimelb.edu.au

Douglas Boyle; dboyle@unimelb.edu.au

necessary validation, analysis and reporting tools.¹² Given the OMOP-CDM is 'open source', the original source code is freely available to the public. This allows anyone to view, use, modify and distribute the software's source code which fosters collaboration and community-driven development. This 'open-source' approach promotes transparency, innovation and widespread accessibility.

The utility and adoption of the OMOP-CDM

An increasing number of Australian and international organisations are transforming their EMR data into the OMOP-CDM as these converted databases provide health services and researchers a valuable data source to monitoring health service utilisation, contribute to the evidence base through research and develop clinical decision support systems to improve quality of care. Furthermore, it enables researchers to 'scale-up' and 'de-risk' collaborative research, by securely sharing deidentified and aggregated data and executing analyses across multiple OMOP-converted databases, ensuring that patient-level data remains securely firewalled within its respective local site.¹²

The adoption of OMOP-CDM has been on the rise globally, with the conversion of approximately 12% of EMRs worldwide by 2022, which encompasses data from 453 databases, that accounts for more than 928 million unique patient records across 41 countries.¹² This substantial adoption demonstrates the recognition of OMOP-CDM's utility in leveraging EMR data for various purposes.

An Australian OHDSI Chapter has been established to support the use of OMOP and develop collaborations between database stakeholders. OMOP members include clinicians and researchers from the University of Melbourne, the University of South Australia, the University of Queensland and the University of New South Wales and Western Australia.¹³ The Australian databases that have undergone OMOP-CDM conversion include those that contain data from large tertiary hospitals in major cities, specialised hospitals that hold data for children's and cancer care services, joint replacement registries, Australian Electronic Practice-Based Research Network (AU-ePBRN),¹⁴ local health district databases, the Primary Care Audit, Teaching and Research Open Network (PATRON) database¹⁵, pharmaceutical registries, and the Australian Department of Veterans Affairs.¹² However, it is important to acknowledge that this progress is not without its limitations. Currently, there exists a gap in data integration, notably the absence of a seamless linkage between hospital and primary care data OMOP data sources. Despite the comprehensive approach to data integration across various healthcare contexts, the lack of connectivity between these crucial components of the healthcare system represents a constraint. This limitation highlights an area for potential improvement in Australia's data infrastructure. Addressing this gap and establishing effective linkage between hospital and primary care data could lead to even more comprehensive and impactful research outcomes.

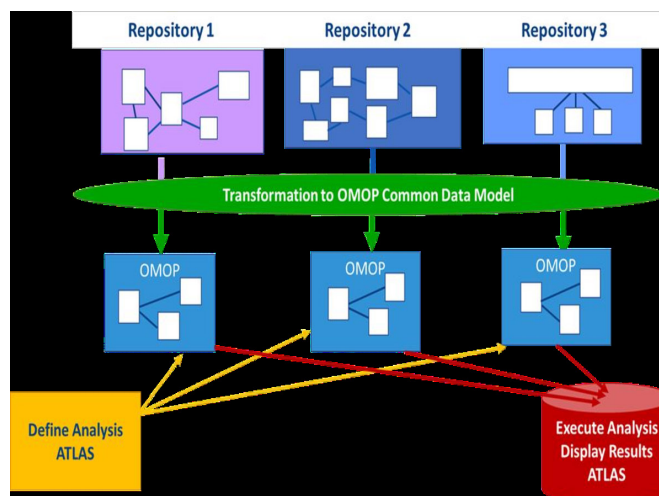


Figure 1 Observational Medical Outcomes Partnership Common Data Model (OMOP-CDM). Adapted from Standardised Data: The OMOP Common Data Model.¹²

Aim

In this overview, we describe the OMOP-CDM, the established governance processes employed in EMR data repositories, and demonstrate how OMOP transformed data provides a lever for more efficient and secure access to EMR data, by health service providers, evaluators, auditors and researchers. Governance, privacy, consent and ethics vary by country or jurisdiction. For this review, we have applied an Australian context, however, the general nature of the guidance here is applicable internationally.

THE OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP COMMON DATA MODEL

The OMOP-CDM: structure and process

The OMOP-CDM can be implemented using many of the existing database management systems. The OMOP-CDM extraction, transformation and loading process converts complex clinical and administrative EMR data into a simplified standard format consisting of 16 data tables and other derived tables.⁸ Through this process, it is important to note the source EMR data are not changed or lost, OMOP conversion just provides a new representation of existing EMR data. For the deployment and installation of the OMOP-CDM into existing information system infrastructure (figure 1), we recommend the OMOP-CDM instance, that model's institution-specific data, is maintained under the existing repository data access and governance mechanisms established by each data custodian.

OMOP, data quality and the principles of Findable, Accessible, Interoperable and Reusable, Collective benefit, Authority, Researcher and Ethics and Five Safes

The use of OMOP-CDM aligns well with the need for systematic data evaluation and adherence to data quality standards and the principles of FAIR (Findable, Accessible, Interoperable and Reusable), CARE (Collective

benefit, Authority, Researcher and Ethics) and the Five Safes.

Before use, OMOP-CDM data undergoes a rigorous data quality assessment process, which includes checks for completeness, concordance, plausibility and currency when compared with the source EMR data.¹⁶ These quality checks are predefined and configured to run on datasets conforming to OMOP standards, and they can be executed using tools such as Achilles, which is accessible via the OHDSI Data Quality Dashboard.¹⁷ In addition, the OMOP-CDM enables researchers to work within a secure and firewalled environment while conducting advanced analytics and prediction techniques. This aligns with the principles of making data 'FAIR, ensuring that data are available for a wide range of research applications.^{18 19} Data accessed through an OMOP-CDM also adheres to the CARE Principles for Indigenous Data. CARE operates within the governance framework established by the custodians of each local data repository. The CARE principles complement FAIR principles by aligning data sharing with the rights and interests of Indigenous Peoples. By adhering to CARE, Indigenous Peoples worldwide gain greater control over their data and the knowledge derived from it, ensuring alignment with their worldviews and the knowledge economy. This framework emphasises the Indigenous Peoples' right to derive value from data while promoting responsible and ethical data usage, for collective and equitable benefit of researchers, evaluators and the broader community.^{18 19}

OMOP data adheres to the 'Five Safes' guiding principles by providing a structured and secure framework for managing and sharing healthcare data while ensuring privacy and security are maintained.²⁰ These frameworks were selected for their compatibility with the principles of ethical research, data quality and data governance. Their widespread adoption and acceptance within the research community make them robust and suitable choices for guiding data management practices in the context of the OMOP-CDM. The responsibility for applying the SAFES framework typically falls on various stakeholders involved in data access and usage, including government agencies, research institutions and data custodians (table 1).

OMOP, DATA GOVERNANCE, ETHICAL REVIEW AND CONSENT OMOP-CDM and governance

By virtue of its design and objectives, the OMOP-CDM enhances the governance of secondary health data, by ensuring data utilisation in both research and healthcare decision-making is ethical, transparent and effective.

With the transformation of EMR data into a standardised structure, the OMOP-CDM ensures there is a uniform representation of these data regardless of the data's original source. This uniformity streamlines data governance and, importantly, eases the complexities associated with conducting single

site studies that contain native EMR data (raw and/or curated), and multisite studies that involve integrating data from various disparate sources.^{21 22} In addition, the common data model emphasises data quality, allowing for consistent checks and ensuring that research data meets the highest standards.^{23 24} The standardised model also ensures that security and privacy protocols are uniformly applied, safeguarding secondary health data from data breaches to maintain patient privacy. Given the structured approach of the OMOP-CDM, an institution can easily implement access controls, thereby ensuring that only authorised parties can access or interact with the data. As a result, the OMOP-CDM acts as a cornerstone for the conduction of rigorous and ethically sound research as it builds trust among stakeholders, mitigates information disparities and encourages the production of high-quality medical evidence for rigorous and ethical research²⁵

Operational use and quality assurance activities in a hospital or healthcare setting

For operational use quality assurance activities where the 'primary purpose is to monitor or improve the quality of service delivered by an individual or an organisation'²⁶ data governance and principles for ethical use apply. However, within healthcare institutions, ethics approval is not mandated for the establishment of the OMOP database or data use, provided:

- ▶ The data being collected and analysed, is coincidental to standard operating procedures with standard equipment and/or protocols.
- ▶ The data are being collected and analysed expressly for the purpose of, maintaining standards or identifying areas for improvement in the environment from which the data were obtained.
- ▶ The data being collected and analysed, is not linked to individuals.
- ▶ None of the triggers for consideration of ethical review are present.²⁶

Research in a university setting

For research use, the data custodian is usually the agency or organisation that commissioned the research and paid for the data collected by the owner (ie, hospital/general practice). Existing local governance principles already developed by custodians can be applied to OMOP standardised data including: (1) data only being made accessible to named researchers on relevant ethics applications approved by the relevant institution, (2) appropriate secure data management strategies for transfer and management of data using password-protected computers or servers with multifactor authentication, (3) data restrictions that align with project scope and objectives and (4) storage of data outputs extracted from the OMOP-CDM as approved by the HREC. For

**Table 1** Guiding principles of FAIR, CARE and the Five Safes

The FAIR guiding principles		
F	Findability	Metadata and data should be easily found by both humans and computers through the assignment of globally unique and permanent identifier to enable the automatic discovery of datasets and services via machine learning. ¹⁸
A	Accessibility	Metadata and data should easily retrieved by authorised and authenticated users via a standard communication protocol. ¹⁸
I	Interoperability	Data from one data source can be integrated with data from other sources so that it can be aggregated into a single, unified view and refers to the intergration and exchange of applications, analysis, storage and workflow processing across different data sources. ¹⁸
R	Reusability	Metadata and data characteristics are specified in detail to enable replication and/or linkage in different settings. Reusability includes the release of data usage licenses, provenance details and disclosure around community standards relevant to the domain. ¹⁸
The CARE principles		
C	Collective benefit	Collective benefit including where the well-being of Indigenous Peoples' rights is of primary concern. ¹⁹
A	Authority	Indigenous Peoples' rights and interests about their peoples, communities, cultures and territories with regard to data are recognised and clearly articulated. ¹⁹
R	Researcher	Researchers have a responsibility to develop and nurture respectful relationship with Indigenous Peoples' from whom the data originate. ¹⁹
E	Ethics	Minimise harm and maximise benefit for Indigenous Peoples', for justice and future use. ¹⁹
The Five Safes framework		
People	Safe People	Is the researcher appropriately trained and authorised to access and use the data? ²⁰
Projects	Safe Projects	Is data used for an appropriate purpose that is valid and of public benefit? ²⁰
Settings	Safe Settings	Does IT access and physical environment prevent unauthorised use? ²⁰
Data	Safe Data	Has appropriate and sufficient protection been applied to the data to avoid risk of disclosure? ²⁰
Outputs	Safe Outputs	Are the statistical results non-disclosive? ²⁰

OMOP converted data that contains linked data, for example, AU-ePBRN where primary care data are linked with hospital admissions data,²⁷ governance and liability procedures would need to be explicitly developed to ensure the governance interests of all institutions are considered.

Consent

In any research, regardless of whether it is conducted by an individual researcher, clinician or collaborative research team, it is imperative to determine the nature of the consent obtained from a patient for the secondary use of their data. This assessment should consider the risks and the potential for psychological, social, economic and legal harm that may arise from data collection, utilisation or any potential breaches.

In Australia, a 'waiver of consent' as per National Health and Medical Research Council (NHMRC) guidelines can be applied to secondary use of health data²⁶ (box 1). Some ethics committees may request an 'opt-out' model, necessitating the consideration

of options for patients who wish to decline to participate.^{26 27} Through the deidentification methods employed by OMOP-CDM, the risks related to data breaches, such as the reidentification of individuals, are significantly reduced. This is achieved by exclusively using aggregated results from OMOP-CDM and by refraining from reporting small cell sizes. Reidentification is further minimised by ensuring only aggregated outputs from OMOP-CDM are used and that small cell sizes are not reported.

Risk mitigation

OMOP mitigates many of the risks of using EMR data for secondary purposes including: (1) replacement of all personal identifiers with a generic number that does not allow reidentification back to the original personal identifier¹²; (2) an option for data custodians to perform analyses on behalf of an individual researcher and auditor (ie, no data release); (3) the use of a user interface tool such as ATLAS, where researcher or auditor access to data in all tables can

Box 1 Consent

If an ethics committee deems a research project or a healthcare evaluation to be of minimal risk to the individual, an exception to obtaining the legislated requirement for patient consent can be managed using a 'waiver of consent.' A 'waiver of consent' can be applied based on a duty of 'easy rescue,' where the potential benefits of data access are considered significant, and the harm associated with the risk of a loss of privacy are considered minimal.³⁰ It is also hypothesised a 'waiver of consent' avoids the consequence of consent bias where individuals who provide informed consent to participate in a study differ in important ways from those who do not consent or choose not to participate.³⁰ Numerous research and evaluation initiatives have employed a 'waiver of consent' approach, allowing for the secondary utilisation of electronic medical record (EMR) data.^{31–34}

Arguments against a 'waiver of consent' considers the societal costs and potential patient harm, against the benefits of patient data utilisation. Costs include privacy breaches per se and the use of data for nefarious purposes, both of which contribute to a heightened risk of eroding trust.³⁵ This includes potential for a loss of informational privacy where an individual's personal or sensitive information is exposed, shared or accessed by others without their consent, or in a manner that violates their expectations of privacy.³⁵ Additional rationale against the application of a 'waiver of consent' stems from the primary rationale for an individual's involvement in research lies in the process of duty of care to obtain, 'informed consent'. This justification is grounded in the idea that depending solely on research and evaluation might not adequately protect the values and interests of those participating. Further to this, informed consent is regarded as a means of building trust, not only in the research and evaluation process itself but in the researcher/clinician understanding of health data use.³⁵

Notwithstanding ethics committee considerations for patient consent, there should also be considerable social engagement across a breadth of stakeholders on research that uses health data, even if it is deidentified. This engagement provides options for the provision of 'social permission' and 'social licence' for consent, where the determination of consent is cocreated by patients and therefore morally legitimised—beyond the limits of law and outside of what is acceptable by an ethics committee—to preserve societal trust.³⁶

Patient and community acceptability of the use of data within their EMR for research and healthcare evaluation indicates, for social licence to be assumed, a breadth of patient and public values, needs and interests should be incorporated into governance frameworks.³⁷

be configured to protect privacy⁸; (4) collaborative analyses are always conducted within each institution's firewalled network⁸; (5) use of standardised terminology only removes potential identifiers in the source terminology and (6) there is an option to obscure dates from view, such that temporal association can be calculated from a relative date (box 2).²⁸

BENEFITS, LIMITATIONS AND CONSIDERATIONS OF OMOP-CDMS

OMOP-converted databases offer a secure and standardised approach to EMR data analysis within an open-source framework, which produces aggregated results which are free of patient identifiers. This eliminates the need for direct access to native EMR data or external

Box 2 Risk mitigation

An access control policy is crucial for ensuring the privacy, management and security of data, especially when it is related to research. This ensures use of data is managed appropriately and underpinned by respect of the rights and expectations of the individuals it represents.

Access control measures include the application of strong passwords that are complex and contain alphanumeric characters as well as symbols; multifactor authentication where data users apply two or more evidence pieces (or factors) to verify their identity; safe connectivity where the standard practice for data access is via devices connected to secure and private networks rather than devices that are connected to public networks; the prompt reporting of data breaches to mitigate the impact of any cybersecurity attack and prevent further vulnerabilities; the verification of ethics approvals before granting access to ensure that research, healthcare evaluation and audit is conducted in an ethically sound manner; and the permissions for data access limited to those researchers and health service evaluators who are authorised and working within the confines of an institution's environment.

A review of data that is due to be transmitted to researchers and health service evaluators provides another important safety check, as does maintenance of version control, where the most recent database is always held as back up. Additional risk management controls include the delivery of explicit instructions to researchers and evaluators on the appropriate use of the dataset; the incorporation of additional hardware authentication such as the YubiKey, Titan, Thetis and Kensington Verimark hardware keys; restricted access to identifiers in the underlying Structured Query Language database and continuous evaluation of anonymisation adequacy instructions on appropriate use for dataset.

sources that have data sharing restrictions, it also sets it apart from the less structured EMR 'data lakes' that contain vast amounts of native and disparate data. These 'data lakes' that lack standardised schemas, make data management and analysis more challenging. In contrast, OMOP-CDM benefits from OHDSI's open-source tools and standardised analytics, by enhancing transparency, reducing coding errors and supporting validation processes.

As an extension of OMOP data conversion, the OHDSI consortium has developed the OHDSI Quality Dashboard, a tool to ensure the quality of data converted into the OMOP-CDM to improve transparency, reduce coding errors and enable validation.²⁹ The OHDSI Quality Dashboard is designed to assess and monitor the quality of data that has been converted into the OMOP-CDM. It provides a set of data quality checks and validation tools that help identify issues or anomalies in the converted data.¹⁶ In doing so, it identifies and addresses data quality issues that may arise during the conversion process by checking data for completeness, consistence, accuracy and adherence to standard terminologies. This ensures that the data in the OMOP-CDM are reliable and suitable for research, analysis, evaluation and audit.

Specialist fields such as oncology and pregnancy have unique data requirements.⁸ For instance, cancer-related data elements can vary among healthcare sites due to clinical practice variations. The OMOP-CDM may not always fully standardise these elements during the mapping

process. To address specialised data needs, the OHDSI community actively develops and shares OMOP-CDM extensions, particularly for specific cancer types.

Specialist fields such as oncology and pregnancy have specific data needs⁸ and data elements may vary among healthcare sites due to differences in clinical practices. The OMOP-CDM may not always fully standardise these elements during the mapping process. To cater to these specialised data needs, the OHDSI community actively creates, refines and disseminates OMOP-CDM extensions that are highly specific to cancer types and treatments, as well as pregnancy episodes and outcomes.

Given the EMR captures similar data across various healthcare sites, such as specific pathology indicators, they may contain different data elements due to differences in pathology classifications (ie, pathology definitions and units). To address these variations and maintain the OMOP-CDM's relevance and flexibility, the OHDSI community also actively develops extension to address variation in pathology phenotypes to preserve the OMOP-CDM's overall compatibility and interoperability across diverse healthcare sites and research projects.

Institutional governance and privacy frameworks have evolved independently alongside the adoption of secondary EMR use,⁷ therefore, achieving a consensus on governance practices across institutions is an ongoing endeavour. This underscores the importance of ongoing collaboration and standardisation in the healthcare data field to ensure that valuable health data can be leveraged effectively and ethically for research and healthcare improvement. Given all the opportunities the OMOP-CDM offers for integrated data governance, these opportunities are limited by lack of standalone funding required for the comprehensive mapping of data from local EMRs to the common format. Despite these challenges, the commitment of the global community to the OMOP-CDM signifies a promising future for standardised health data, which will pave the way to transform healthcare research, evaluate operational processes and facilitate quality improvement within healthcare organisations.

CONCLUSION

Adoption of the OMOP-CDM internationally and locally is well worth the investment, as it enables conversion of large amounts of complex, and heterogeneous EMR data into a standardised structured data model, simplifies governance processes and facilitates rapid repeatable cross-institution analysis through shared end-to-end analysis packages, without the sharing of native data. Combined with pseudonymisation and common data quality assessments, the OMOP-CDM provides a powerful model to support ethical real-world 'big' data research. The continued adoption of OMOP-CDM, ongoing development efforts, and the emphasis on sound governance practices all contribute to the realisation of OMOP's

utility in unlocking valuable EMR data. These factors collectively support a wide range of applications, from health service operational reporting to diverse clinical, epidemiological and translational research projects.

While the adoption of OMOP and the collaborative efforts in data integration in Australia is commendable, there is room for further development in bridging the gap between hospital and primary care data. This ongoing endeavour has the potential to significantly enhance Australia's capacity for data-driven research and improve healthcare outcomes for its population.

Author affiliations

¹Health and Biomedical Informatics Centre, Research Information Technology Unit (HaBIC R2), Department of General Practice and Primary Care, The University of Melbourne Faculty of Medicine Dentistry and Health Sciences, Melbourne, Victoria, Australia

²School of Computing and Information Systems, Faculty of Engineering and Information Technology, Centre for the Digital Transformation of Health, The University of Melbourne Faculty of Medicine Dentistry and Health Sciences, Melbourne, Victoria, Australia

³Queensland Digital Health Centre (QDHeC), Centre for Health Services Research, The University of Queensland Faculty of Medicine, Woolloongabba, Queensland, Australia

⁴Quality Use of Medicines and Pharmacy Research Centre, Clinical and Health Sciences, University of South Australia, Adelaide, South Australia, Australia

⁵Clinical Haematology Department, The Royal Melbourne Hospital, Peter MacCallum Cancer Centre, Melbourne, Victoria, Australia

⁶Sir Peter MacCallum Department of Oncology, The University of Melbourne Faculty of Medicine Dentistry and Health Sciences, Melbourne, Victoria, Australia

⁷Department of General Medicine, The Royal Melbourne Hospital, Parkville, Victoria, Australia

⁸Queensland Digital Health Centre (QDHeC), Centre for Health Services Research, The University of Queensland Faculty of Medicine, Herston, Queensland, Australia

⁹School of Population Health, UNSW, Sydney, New South Wales, Australia

¹⁰Centre for Big Data Research in Health (CBDRH), UNSW, Sydney, New South Wales, Australia

Twitter Christine Mary Hallinan @Cmhallinan, Ashley P Ng @drng and Siaw-Teng Liaw @EIAstee

Collaborators Not applicable.

Contributors The research presented in this paper was a collaborative effort involving multiple authors who contributed to various aspects of the review. The review was conceived by RW, GKH, NP and DBu. CMH structured the review, synthesised the evidence, drafted the manuscript, reviewed the manuscript and completed the final version of the manuscript. RW, GKH, CS, NP, AN, DC, AVDV, S-TL, OD, BGL, DBu and DBo provided evidence and edits for the review, and reviewed the full manuscript when completed.

Funding This study was funded by Australian Government, Australian Research Data Commons (ARDC) Public Sector Bridges Program 'Electronic Medical Records as a National Data Asset' (Doi: 10.47486/PS014).

Competing interests None declared.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data sharing not applicable as no datasets generated.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

 Christine Mary Hallinan <http://orcid.org/0000-0002-0471-4444>

 Ashley P Ng <http://orcid.org/0000-0001-9690-0879>

 Siaw-Teng Liaw <http://orcid.org/0000-0001-5989-3614>

REFERENCES

- 1 Safran C, Bloomrosen M, Hammond WE, *et al*. Toward a national framework for the secondary use of health data: an American medical informatics association white paper. *J Am Med Inform Assoc* 2007;14:1–9.
- 2 Danciu I, Cowan JD, Basford M, *et al*. Secondary use of clinical data: the vanderbilt approach. *J Biomed Inform* 2014;52:28–35.
- 3 Budronis A, Bellika JG. The learning healthcare system: where are we now? A systematic review. *J Biomed Inform* 2016;64:87–92.
- 4 Park K, Cho M, Song M, *et al*. Exploring the potential of OMOP common data model for process mining in healthcare. *PLoS One* 2023;18:e0279641.
- 5 Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. *J Am Med Inform Assoc* 2013;20:144–51.
- 6 Jean-Baptiste L, Mouazer A, Sedki K, *et al*. Translating the observational medical outcomes partnership - common data model (OMOP-CDM) electronic health records to an OWL ontology. *Stud Health Technol Inform* 2022;290:76–80.
- 7 Evans RS. Electronic health records: then, now, and in the future. *Yearb Med Inform* 2016;Suppl 1(Suppl 1):S48–61.
- 8 OHDSI. Observational health data sciences and Informatics data standardization 2021. 2021 Available: <https://www.ohdsi.org/data-standardization/the-common-data-model/>
- 9 SNOMED CT. SNOMED International leading Healthcare terminology, worldwide 2022. 2022. Available: <https://www.snomed.org/>
- 10 LOINC. LOINC the International standard for identifying health measurements, observations, and documents 2022. 2022. Available: <https://loinc.org/>
- 11 National Library of Medicine. Unified medical language system Rxnorm. 2022. Available: <https://www.nlm.nih.gov/research/umls/rxnorm/index.html>
- 12 OHDSI. Our journey: where the OHDSI community has been and where we are going observational health data sciences and Informatics; 2022.
- 13 Electronic medical records National data asset (Internet). 2022. Available: <https://doi.org/10.26188/6295c4a5d7c5c>
- 14 Electronic Practice Based Research Network. Centre for Primary Health Care and Equity, UNSW Sydney, Available: <https://cphce.unsw.edu.au/research/electronic-practice-based-research-network>
- 15 Boyle D, Sanci L, Emery J, *et al*. PATRON Primary Care Research Data Repository. . 2019 Available: <https://medicine.unimelb.edu.au/school-structure/general-practice-and-primary-care/research/data-for-decisions>
- 16 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;4:1244.
- 17 OHDSI. Observational health data sciences and Informatics ACHILLES for data characterization 2022. n.d. Available: <https://www.ohdsi.org/analytic-tools/achilles-for-data-characterization/>
- 18 Wilkinson MD, Dumontier M, Jan Aalbersberg I, *et al*. Addendum: the fair guiding principles for scientific data management and stewardship. *Sci Data* 2019;6:6.
- 19 The Global Indigenous Data Alliance. CARE principles for indigenous data governance. 2022. Available: <https://www.gida-global.org/care>
- 20 ABS. Five safes framework Australian Bureau of Statistics Canberra: Australia. n.d. Available: <https://www.abs.gov.au/about/data-services/data-confidentiality-guide/five-safes-framework>
- 21 Biedermann P, Ong R, Davydov A, *et al*. Standardizing registry data to the OMOP common data model: experience from three pulmonary hypertension databases. *BMC Med Res Methodol* 2021;21:238.
- 22 OHDSI. The book of OHDSI: observational health data sciences and Informatics. 2023. Available: <https://ohdsi.github.io/TheBookOfOhdsi/>
- 23 Blacketer C, Defalco FJ, Ryan PB, *et al*. Increasing trust in real-world evidence through evaluation of observational data quality. *J Am Med Inform Assoc* 2021;28:2251–7.
- 24 Blacketer C, Voss EA, DeFalco F, *et al*. Using the data quality dashboard to improve the EHDS network. *Applied Sciences* 2023;13:11920.
- 25 Kim J-W, Kim C, Kim K-H, *et al*. Scalable infrastructure supporting reproducible nationwide healthcare data analysis toward FAIR stewardship. *Sci Data* 2023;10:674.
- 26 NHMRC. National Statement on ethical conduct in human research (2007) - updated 2018; 2018. National health and medical research Council
- 27 UNSW. The electronic practice based research network Sydney: centre for primary health care and equity medicine. 2021. Available: <https://cphce.unsw.edu.au/research/electronic-practice-based-research-network>
- 28 Hripcsak G, Mirhaji P, Low AF, *et al*. Preserving temporal relations in clinical data while maintaining privacy. *J Am Med Inform Assoc* 2016;23:1040–5.
- 29 Schuemie MJ, Ryan PB, Pratt N, *et al*. Principles of Large-scale Evidence Generation and Evaluation across a Network of Databases (LEGEND). *J Am Med Inform Assoc* 2020;27:1331–7.
- 30 Porsdam Mann S, Savulescu J, Sahakian BJ. Facilitating the ethical use of health data for the benefit of society: electronic health records, consent and the duty of easy rescue. *Phil Trans R Soc A* 2016;374:20160130.
- 31 Tu K, Sarkadi Kristiansson R, Gronsbell J, *et al*. Changes in primary care visits arising from the COVID-19 pandemic: an international comparative study by the International consortium of primary care big data researchers (INTRePID). *BMJ Open* 2022;12:e059130.
- 32 Lu Y, Van Zandt M, Liu Y, *et al*. Analysis of dual combination therapies used in treatment of hypertension in a multinational cohort. *JAMA Netw Open* 2022;5:e223877.
- 33 Ahmadi N, Peng Y, Wolfien M, *et al*. OMOP CDM can facilitate data-driven studies for cancer prediction: a systematic review. *Int J Mol Sci* 2022;23:19.
- 34 Lane JCE, Weaver J, Kostka K, *et al*. Risk of hydroxychloroquine alone and in combination with azithromycin in the treatment of rheumatoid arthritis: a multinational, retrospective study. *Lancet Rheumatol* 2020;2:e698–711.
- 35 Ploug T. In Defence of informed consent for health record research - why arguments from “easy rescue”, “no harm” and “consent bias” fail. *BMC Med Ethics* 2020;21:75.
- 36 Muller SHA, Kalkman S, van Thiel G, *et al*. The social licence for data-intensive health research: towards co-creation, public value and trust. *BMC Med Ethics* 2021;22:110.
- 37 Kalkman S, van Delden J, Banerjee A, *et al*. Patients’ and public views and attitudes towards the sharing of health data for research: a narrative review of the empirical evidence. *J Med Ethics* 2022;48:3–13.