

How to develop a good research question: use cases for hospital electronic medical record data in the Australian Health Data Evidence Network

Jack Janetzki¹, Georgina Kennedy², Jodie Austin³, Blanca Gallego Luxan⁴, Ezra Fass⁵, Nicole Pratt¹ on behalf of the AHDEN network and node leads

¹Adelaide University, Adelaide, Australia

²Ingham Institute, Sydney, Australia

³The University of Queensland, Brisbane, Australia

⁴University of New South Wales, Sydney, Australia

⁵Royal Children's Hospital, Melbourne, Australia

Background

The increasing availability of hospital electronic medical record (EMR) data presents substantial opportunities for generating real-world evidence.¹ However, translating clinical questions into feasible EMR-based research projects requires careful consideration of data availability, quality, completeness, and linkage capabilities.

The Australian Health Data Evidence Network (AHDEN)^{1,2} initiative supports the development and subsequent use of harmonized hospital EMR data for research across multiple health services. Use of large-scale harmonized hospital EMR data has not previously been possible in Australia which has limited use of such data to more localized contexts within hospitals that use the same EMR.^{1,2}

This study describes the early development and refinement of research questions through four use cases within the AHDEN initiative, highlighting key lessons in assessing feasibility and maximizing the value of EMR data.

Methods

Four proposed use cases have been developed within AHDEN using hospital EMR data mapped to the Observational Medical Outcomes Partnership (OMOP) Common Data Model. Initial research questions for each use case were generated in collaboration with clinical and research stakeholders and underwent iterative feasibility assessments through regular monthly meetings (Figure 1). Feasibility considerations so far have included the availability of structured data elements, completeness of clinical information, data linkage requirements, and the accessibility of relevant outcomes within participating hospital systems. Research questions are currently being refined, where necessary, to align with available data resources.



Figure 1 Iterative process for research question design in AHDEN (created with ChatGPT)

Results

Development of each of the four use cases is currently at Step 2 of Figure 1 with a view to progress to Step 3 in the coming months.

The first use case is examining treatment pathways for patients with head and neck cancers. Structured diagnostic, procedural, and treatment data are available in New South Wales and South Australia, supporting the development of clinically meaningful research questions using routinely collected hospital data. While cancer data has not yet been converted to the OMOP CDM at all nodes within AHDEN, the use case is likely feasible in two out of eight states or territories in Australia.

The second use case, from Queensland, initially sought to investigate gestational diabetes and associated maternal and neonatal outcomes. Feasibility assessment identified substantial challenges related to the need for linked maternity, neonatal, and external healthcare datasets, as well as variability in data capture across sites. Consequently, the research question has been refined to examine rates of diabetic ketoacidosis presentation to emergency departments and association with hospital stays, an outcome more readily identifiable within available hospital EMR data across Australia.

The third use case, also from New South Wales, is focusing on characterising heart failure and its treatment. While heart failure diagnoses can easily be identified within structured datasets, important clinical variables required for phenotyping, characterisation of heart failure, disease severity assessment, and outcome evaluation are primarily contained within unstructured clinical notes in the available EMR in NSW. The absence of routinely accessible natural language processing capabilities across hospital sites and nodes limited the feasibility of the original research question and highlighted the importance of understanding the boundaries of structured EMR data.

A fourth use case from Victoria plans to examine the incidence of renal toxicity in paediatric patients receiving intravenous acyclovir for herpes simplex virus-related meningitis or encephalitis. This use case is likely highly feasible within available hospital EMR datasets (although it is still undergoing feasibility assessment) as exposure to intravenous acyclovir, laboratory measures of renal function, and relevant diagnostic information are routinely captured in structured hospital data throughout Australia.

Across all use cases, early feasibility assessment substantially influenced study design, outcome selection, and the ultimate research questions pursued.

Conclusion

Developing high-quality research questions using hospital EMR data requires an understanding of both clinical priorities and the practical limitations of available datasets particularly outside of the main investigators' own state and hospital. The AHDEN experience demonstrates that successful EMR-based research often involves iterative refinement of research questions to align with data availability and quality. Early feasibility assessment can identify potential barriers related to data linkage, completeness, and reliance on unstructured information, ultimately improving study efficiency and the likelihood of generating meaningful real-world evidence.

References

1. Ward, R., Pratt, N., Hart, G., Meyers, I., Sullivan, C., Gallego Luxan, B. and Kennedy, G. (2025) *Australian Health Data Evidence Network (AHDEN): Building a National Data Infrastructure for Standardised, Federated Health Data Research*. Brief report. Observational Health Data Sciences

and Informatics (OHDSI) 2025 Symposium. Available at: [OHDSI Brief Report PDF](#) [Accessed: 4 June 2026]

2. Australian Research Data Commons (n.d.) *Australian Health Data Evidence Network (AHDEN)*. Available at: [Australian Health Data Evidence Network \(AHDEN\)](#) [Accessed: 4 June 2026]