

OHDSI AI: Generative AI-powered Knowledge Translation of OHDSI Research Literature and Singapore's Cardiovascular Research

Maisie Ng^{1,2}, Cindy Ho^{1,2}, Li Ting Ang^{1,2}, Hang Png^{1,2}, Shuen Lin Tan^{1,2}, Estella Ye^{1,2}, Ismail Mohd^{1,2}, Mengling Feng^{3,4}, Sebastian Maurer-Stroh^{1,5,6}, Johan G Eriksson^{2,7,8,9}, Mukkesh Kumar^{1,2}

Institution(s) of origin:

¹ Bioinformatics Institute, Agency for Science Technology and Research, Singapore

² Singapore Institute for Clinical Sciences, Agency for Science Technology and Research, Singapore

³ Saw Swee Hock School of Public Health, National University of Singapore, National University Health System, Singapore

⁴ Institute of Data Science, National University of Singapore, Singapore

⁵ Yong Loo Lin School of Medicine, National University of Singapore, Singapore

⁶ Department of Biological Sciences, National University of Singapore, Singapore

⁷ Department of Obstetrics and Gynaecology and Human Potential Translational Research Programme, Yong Loo Lin School of Medicine, National University of Singapore, Singapore

⁸ Department of General Practice and Primary Health Care, University of Helsinki, Helsinki, Finland

⁹ Folkhälsan Research Center, Helsinki, Finland

Background

The Observational Health Data Sciences and Informatics (OHDSI) federated network is a multi-stakeholder, interdisciplinary collaborative effort aimed at improving health outcomes through the generation and dissemination of real-world evidence¹. Health evidence needs to be communicated and disseminated in a manner that is clearly understood by health professionals, patients and public or private health managers. At the 2023 OHDSI APAC Symposium, *Knowledge Translation* was a key theme emerging from the round table discussion with the consumers of evidence (Patients & Consumers, Clinicians, Researchers, Regulators & Policy Makers)². Generative Artificial Intelligence (AI) can be an effective strategy for communicating scientific health evidence to the population and/or health managers.

Methods

The Agency for Science, Technology and Research (A*STAR) has pioneered health research with an open interactive Data Vault platform (Figure 1) for hypothesis construction and data-driven discoveries³. The OMOP Data Catalogue^{4,5} in Data Vault enables database level characterizations for knowledge discovery and management in an open science system. The OHDSI AI⁶ in Cardiovascular Data Vault was configured to answer questions from the research literature of OHDSI and Singapore's cardiovascular studies (Figure 2).

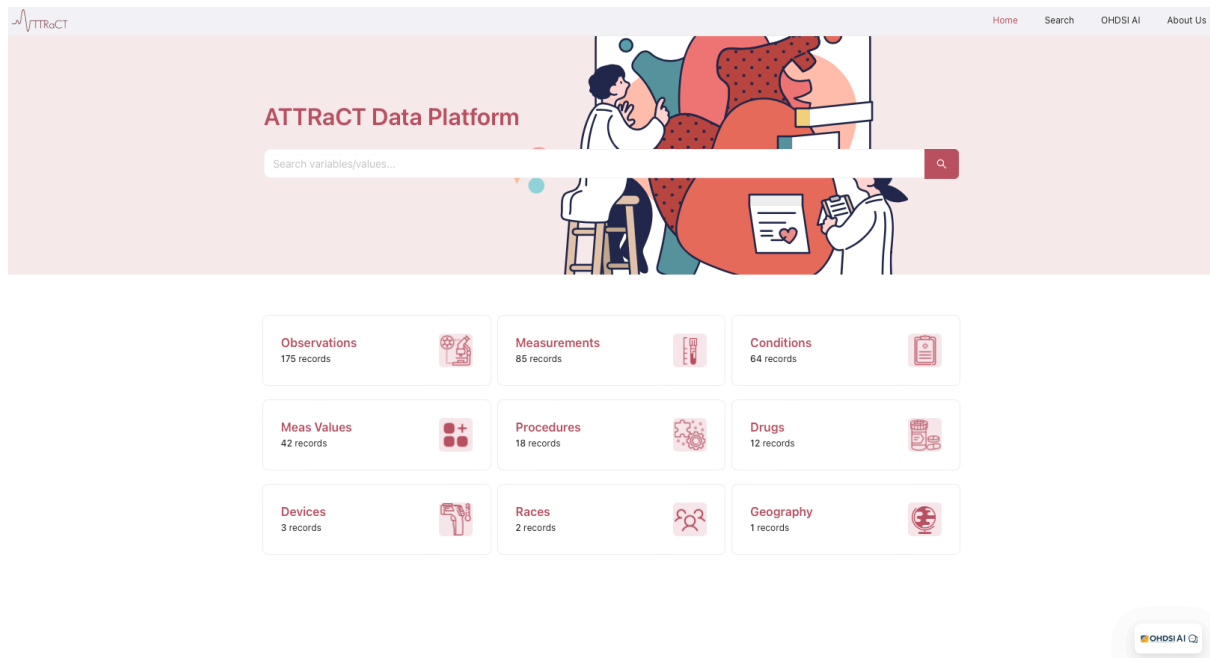


Figure 1. Homepage of Cardiovascular Data Vault

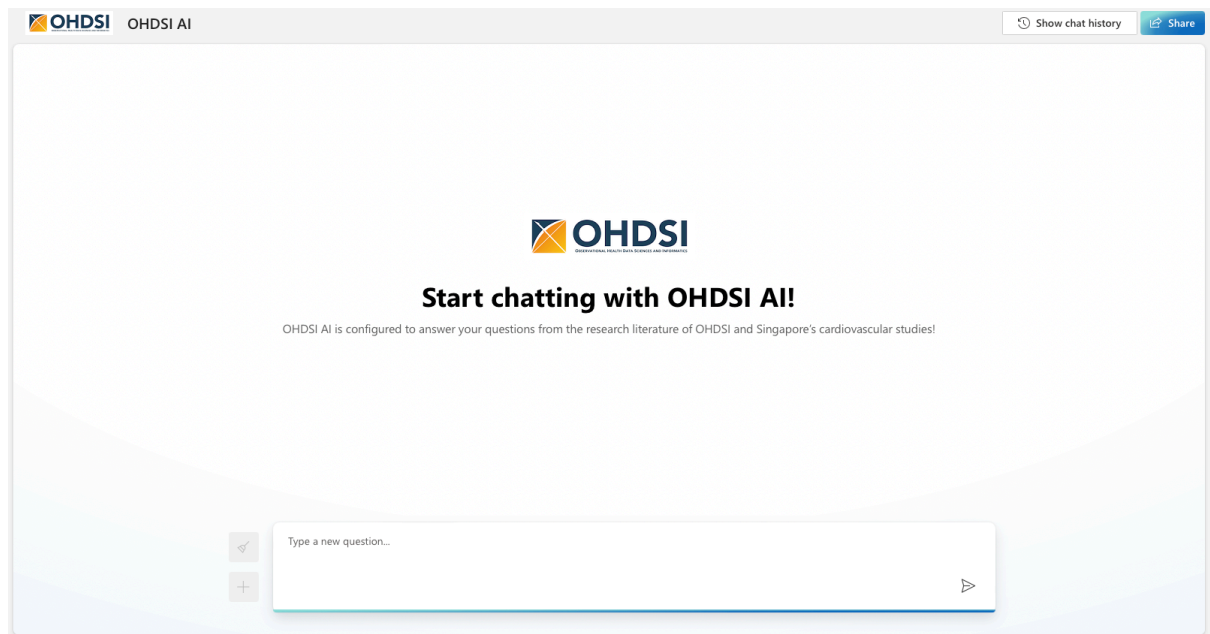


Figure 2. OHDSI AI Web Application

The Azure OpenAI Service⁷ provides access to OpenAI's large language models while maintaining data security as the data remains encrypted privately within the Azure tenant environment. OpenAI's latest flagship generative AI model, GPT-4o model was deployed in A*STAR Commercial Cloud (Azure). The GPT-4o model was ingested with the open access articles from the OHDSI research publications⁸ and Singapore's cardiovascular research publications from ATTRaCT study⁹. The Book of OHDSI¹⁰ was used as the grounding data content to enhance the GPT-4o model's comprehension of OHDSI contextual knowledge.

The OHDSI AI was developed using a Retrieval Augmented Generation (RAG) solution with Azure AI Search⁷, providing indexing and query capabilities. RAG is an architecture that augments the capabilities of a large language model by adding an information retrieval system. The user prompt is sent to Azure AI Search to find relevant information. The top ranked search results are sent to the

GPT-4o model. The prompt completions are generated using natural language understanding and reason capabilities of the large language model. The web application to interact with OHDSI AI was developed to provide the end user experience with question-answering and summarization tasks.

Results

The OHDSI AI performs fast and flexible query parsing and matching for health knowledge translation, from the OHDSI research publications and Singapore's cardiovascular research publications. OHDSI AI is capable of generating scientific hypotheses for impactful research outcomes. A sample chat with the prompt/completion of OHDSI AI has been provided in this section. Figure 3 illustrates the summarization capability of OHDSI AI for a cardiovascular-related OHDSI publication.

[Sample chat with the prompt/completion of OHDSI AI]

User: I'm planning for an OHDSI network study on the effectiveness of beta blockers for heart failure patients. Please refer to the OHDSI publications and suggest top 3 research ideas. Build upon the findings from Singapore's ATTRaCT cardiovascular research study.

OHDSI AI: Certainly! Leveraging insights from both OHDSI publications and the findings from Singapore's ATTRaCT cardiovascular research study can provide a robust foundation for your study on the effectiveness of beta blockers in heart failure patients. Here are the top three research ideas:

1. Comparative Effectiveness of Different Beta Blockers in Asian Heart Failure Patients

Idea: Conduct a large-scale comparative effectiveness study of different beta blockers (e.g., metoprolol, carvedilol, bisoprolol) specifically in Asian heart failure patients. This study can help determine which beta blocker is most effective in improving patient outcomes in this population.

Rationale: The ATTRaCT study highlights the unique characteristics and risk factors of heart failure in Asian populations. By focusing on this demographic, the study can provide tailored insights that are directly applicable to clinical practice in Asia.

OHDSI Publication:

- **Hripcsak, G., et al. (2015).** Observational Health Data Sciences and Informatics (OHDSI): opportunities for observational researchers. *Studies in Health Technology and Informatics*, 216, 574-578.

ATTRaCT Findings:

- The ATTRaCT study identifies hypertension, diabetes, and coronary artery disease as prevalent risk factors in Asian heart failure patients. Understanding how different beta blockers perform in this context can help optimize treatment strategies.

2. Longitudinal Outcomes of Beta Blocker Therapy in Hypertensive Heart Failure Patients

Idea: Conduct a longitudinal study to evaluate the long-term outcomes of beta blocker therapy in heart failure patients with a history of hypertension. Assess the impact of continuous versus intermittent use of beta blockers on patient outcomes such as survival rates, hospital readmissions, and quality of life.

Rationale: Hypertension is a major risk factor for heart failure, as highlighted by the ATTRaCT study. Understanding the long-term effectiveness of beta blockers in hypertensive heart failure patients can provide valuable insights for managing this high-risk group.

OHDSI Publication:

- **Suchard, M. A., et al. (2019).** Comprehensive comparative effectiveness and safety of first-line antihypertensive drug classes: a systematic, multinational, large-scale analysis. *The Lancet*, 394(10211), 1816-1826.

ATTRaCT Findings:

- The ATTRaCT study emphasizes the high prevalence of hypertension among heart failure patients in Asia. This study can help determine the long-term benefits and risks of beta blocker therapy in this specific subgroup.

3. Machine Learning for Predictive Modeling of Beta Blocker Response in Diabetic Heart Failure Patients

Idea: Apply machine learning techniques to develop predictive models that identify heart failure patients with diabetes who are most likely to benefit from beta blocker therapy. Use OHDSI's extensive datasets to train and validate these models.

Rationale: Diabetes is a significant risk factor for heart failure, as identified in the ATTRaCT study. Predictive modeling can help personalize treatment plans, ensuring that diabetic heart failure patients receive the most effective therapies based on their individual characteristics and risk profiles.

OHDSI Publication:

- **Reps, J. M., et al. (2018).** Design and implementation of a standardized framework to generate and evaluate patient-level prediction models using observational healthcare data. *Journal of the American Medical Informatics Association*, 25(8), 969-975.

ATTRaCT Findings:

- The ATTRaCT study reports a high prevalence of diabetes among heart failure patients in Asia. Developing predictive models for this subgroup can help optimize treatment and improve outcomes.

Summarise the OHDSI publication "Comprehensive comparative effectiveness and safety of first-line antihypertensive drug classes: a systematic, multinational, large-scale analysis". Please include the major findings from this study.

The OHDSI (Observational Health Data Sciences and Informatics) publication titled "Comprehensive comparative effectiveness and safety of first-line antihypertensive drug classes: a systematic, multinational, large-scale analysis" presents a large-scale study that evaluates the effectiveness and safety of different first-line antihypertensive drug classes. The study leverages a vast amount of health data from multiple countries to provide a comprehensive comparison.

Major Findings:

1. **Effectiveness:** The study found that thiazide or thiazide-like diuretics were generally more effective in preventing cardiovascular events compared to other first-line antihypertensive drug classes, such as ACE inhibitors, angiotensin II receptor blockers (ARBs), calcium channel blockers, and beta-blockers.
2. **Safety:** Thiazide diuretics also showed a favorable safety profile, with fewer adverse events compared to other drug classes. However, the study noted some specific risks associated with each drug class, such as electrolyte imbalances with thiazides and cough with ACE inhibitors.
3. **Consistency Across Populations:** The findings were consistent across different populations and healthcare settings, reinforcing the robustness of the results.
4. **Clinical Implications:** The results suggest that thiazide or thiazide-like diuretics should be considered as a preferred first-line treatment for hypertension due to their superior balance of effectiveness and safety.

This study provides valuable insights for clinicians in making informed decisions about antihypertensive therapy, emphasizing the importance of considering both the effectiveness and safety profiles of medications.

AI-generated content may be incorrect

Figure 3. Summarization of a Cardiovascular-related OHDSI Publication by OHDSI AI

Conclusions

The OHDSI AI excels at communicating the results of scientific studies and understanding the epidemiology of cardiovascular diseases. Generative AI for scientific research is capable of automated hypothesis generation as shown by the advanced reasoning capabilities of OHDSI AI. The OHDSI AI lays the foundations for developing a cross-domain knowledge translation system across APAC and global OHDSI data network, revolutionising the communication of scientific health evidence in an open science system.

References/Citations

1. Observational Health Data Sciences and Informatics. The Book of OHDSI [Internet]. [place unknown: Observational Health Data Sciences and Informatics]; 2021 [cited 2024 June 21] Available from: <https://ohdsi.github.io/TheBookOfOhdsi/>.
2. Observational Health Data Sciences and Informatics. APAC Community Calls. August 17, 2023 - Symposium Reviews, India Chapter Update & More [Internet]. [place unknown: Observational Health Data Sciences and Informatics]; [cited 2024 June 21]. Available from: <https://ohdsi.org/apac/>.
3. Ho C, Ang LT, Ng M, Png H, Tan SL, Ye E, et al. GUSTO Data Vault: Working Towards OMOP Data Standardisation. Poster session presented at: 2022 APAC OHDSI Symposium; 2022 Nov 13; Taipei, Taiwan.
4. Ho C, Ang LT, Ng M, Png H, Tan SL, Ye E, et al. GUSTO Data Vault: Laying the foundations for an open science system with OMOP Data Catalogue. Lightning Talk presented at: 2023 OHDSI APAC Symposium; 2023 July 13; Sydney, Australia.
5. Ho C, Ang LT, Ng M, Png H, Tan SL, Ye E, et al. GUSTO Data Vault: Laying the foundations for an open science system with OMOP Data Catalogue. Software Demonstration presented at: 2023 OHDSI Global Symposium; 2023 October 20; New Jersey, USA.

6. Agency for Science, Technology and Research. OHDSI AI [Internet]. Singapore: Agency for Science, Technology and Research; [cited 2024 June 21]. Available from: <https://cvd-datavault.org/>.
7. Microsoft. Azure OpenAI Service documentation [Internet]. USA: Microsoft; [cited 2024 June 21]. Available from: <https://learn.microsoft.com/en-us/azure/ai-services/openai/>.
8. Observational Health Data Sciences and Informatics. Publications [Internet]. [place unknown: Observational Health Data Sciences and Informatics]; [cited 2024 June 21]. Available from: <https://www.ohdsi.org/publications/>.
9. Agency for Science, Technology and Research. ATTRaCT Publications & Press Releases [Internet]. Singapore: Agency for Science, Technology and Research; [cited 2024 June 21]. Available from: <https://www.a-star.edu.sg/attract/publications-press-releases>.