Collaboration Update (It's Good News!)

OHDSI Community Call
Feb. 10, 2026 • 11 am ET

Three Stages of The Journey

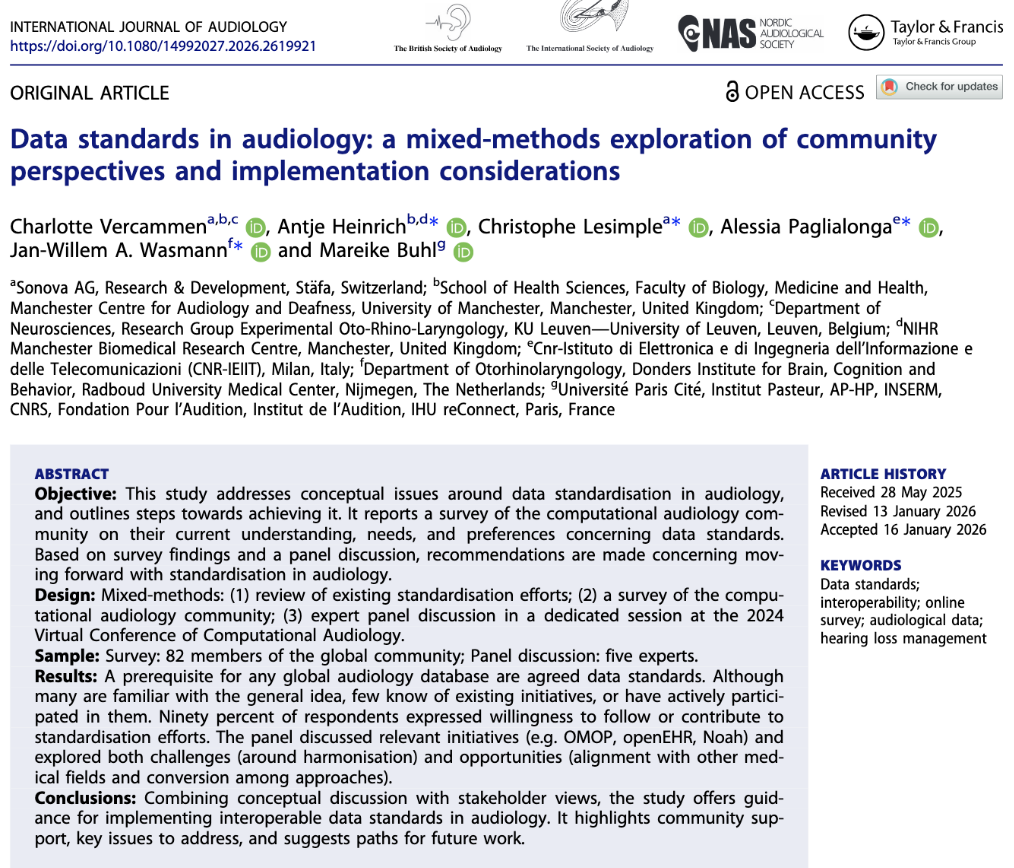
Where Have We Been?

Where Are We Now?

Where Are We Going?

OHDSI Shoutouts!

Congratulations to the team of **Charlotte Vercammen, Antje Heinrich, Christophe Lesimple, Alessia Paglialonga, Jan-Willem A Wasmann, and Mareike Buhl** on the recent publication of **Data standards in audiology: a mixed-methods exploration of community perspectives and implementation consideration** in the *International Journal of Audiology*.



INTERNATIONAL JOURNAL OF AUDIOLOGY
https://doi.org/10.1080/14992027.2026.2619921
The British Society of Audiology
The International Society of Audiology
NAS NORDIC AUDIOLOGICAL SOCIETY
Taylor & Francis
Taylor & Francis Group
ORIGINAL ARTICLE
OPEN ACCESS
Check for updates

Data standards in audiology: a mixed-methods exploration of community perspectives and implementation considerations

Charlotte Vercammen[a,b,c], Antje Heinrich[b,d*], Christophe Lesimple[a*], Alessia Paglialonga[e*], Jan-Willem A. Wasmann[f*] and Mareike Buhl[g]

[a]Sonova AG, Research & Development, Stäfa, Switzerland; [b]School of Health Sciences, Faculty of Biology, Medicine and Health, Manchester Centre for Audiology and Deafness, University of Manchester, Manchester, United Kingdom; [c]Department of Neurosciences, Research Group Experimental Oto-Rhino-Laryngology, KU Leuven—University of Leuven, Leuven, Belgium; [d]NIHR Manchester Biomedical Research Centre, Manchester, United Kingdom; [e]Cnr-Istituto di Elettronica e di Ingegneria dell'Informazione e delle Telecomunicazioni (CNR-IEIIT), Milan, Italy; [f]Department of Otorhinolaryngology, Donders Institute for Brain, Cognition and Behavior, Radboud University Medical Center, Nijmegen, The Netherlands; [g]Université Paris Cité, Institut Pasteur, AP-HP, INSERM, CNRS, Fondation Pour l'Audition, Institut de l'Audition, IHU reConnect, Paris, France

ABSTRACT

Objective: This study addresses conceptual issues around data standardisation in audiology, and outlines steps towards achieving it. It reports a survey of the computational audiology community on their current understanding, needs, and preferences concerning data standards. Based on survey findings and a panel discussion, recommendations are made concerning moving forward with standardisation in audiology.

Design: Mixed-methods: (1) review of existing standardisation efforts; (2) a survey of the computational audiology community; (3) expert panel discussion in a dedicated session at the 2024 Virtual Conference of Computational Audiology.

Sample: Survey: 82 members of the global community; Panel discussion: five experts.

Results: A prerequisite for any global audiology database are agreed data standards. Although many are familiar with the general idea, few know of existing initiatives, or have actively participated in them. Ninety percent of respondents expressed willingness to follow or contribute to standardisation efforts. The panel discussed relevant initiatives (e.g. OMOP, openEHR, Noah) and explored both challenges (around harmonisation) and opportunities (alignment with other medical fields and conversion among approaches).

Conclusions: Combining conceptual discussion with stakeholder views, the study offers guidance for implementing interoperable data standards in audiology. It highlights community support, key issues to address, and suggests paths for future work.

ARTICLE HISTORY
Received 28 May 2025
Revised 13 January 2026
Accepted 16 January 2026

KEYWORDS
Data standards; interoperability; online survey; audiological data; hearing loss management

OHDSI Shoutouts!

Congratulations to the team of **Hanieh Razzaghi, Kimberley Dickinson, Kaleigh Wieand, Samuel Boss, Hunter Weidlich, Yungui Huang, Keith Morse, Sujan Kumar Mutyala, Jyothi Priya Alekapatti Nandagopal, Karthik Viswanathan, Christopher B Forrest, and L Charles Bailey** on the recent publication of **A multifaceted approach to advancing data quality and fitness standards in multi-institutional networks** in *JAMIA*.

Journal of the American Medical Informatics Association, 2026, **33(2)**, 371–382
https://doi.org/10.1093/jamia/ocaf181
Advance access publication 23 October 2025
Research and Applications

AMIA INFORMATICS PROFESSIONALS. LEADING THE WAY.
OXFORD

Research and Applications

A multifaceted approach to advancing data quality and fitness standards in multi-institutional networks

Hanieh Razzaghi, PhD, MPH*[1,2,3], Kimberley Dickinson, MS[1,2], Kaleigh Wieand, BS[1,2], Samuel Boss, BS[1,2], Hunter Weidlich, BS[1,2], Yungui Huang, PhD[4], Keith Morse, MD[5], Sujan Kumar Mutyala, MS[6], Jyothi Priya Alekapatti Nandagopal, MS[7], Karthik Viswanathan, MS[8], Christopher B. Forrest, MD, PhD[1,2,3,9], L. Charles Bailey, MD, PhD[1,2,3,9]

[1]Applied Clinical Research Center, Children's Hospital of Philadelphia, Philadelphia, PA 19146, United States, [2]Department of Pediatrics, Children's Hospital of Philadelphia, Philadelphia, PA 19104, United States, [3]Department of Biomedical and Health Informatics, Children's Hospital of Philadelphia, Philadelphia, PA 19146, United States, [4]IT Research and Innovation, Nationwide Children's Hospital, Columbus, OH 43205, United States, [5]Department of Pediatrics, Stanford University School of Medicine, Palo Alto, CA 94305, United States, [6]Information Services, Texas Children's Hospital, Houston, TX 77030, United States, [7]IS Research Data & Technology Services, Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45229, United States, [8]Biomedical Research Informatics Center, Nemours Children's Hospital, Wilmington, DE 19803, United States, [9]Department of Pediatrics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA 19104, United States

*Corresponding author: Hanieh Razzaghi, PhD, MPH, Roberts Center for Pediatric Research 11361, Children's Hospital of Philadelphia, 2716 South Street, Philadelphia, PA 19146, United States (razzaghih@chop.edu)

Abstract

Objective: To construct a data quality (DQ) system that incorporates combinations of methods to evaluate data characteristics and analytic fitness across research questions for multiple uses.

Materials and Methods: Drawing from experience of other data quality programs, network data extraction needs, and recurring study requirements, we developed 5 standards to guide development of a modular, multifaceted data quality system. These included annotation and documentation, ability to measure research readiness, reproducibility across networks, flexibility for the user, and interpretability to research and project teams. Implementation of checks based on these principles focused on reusability and interactive visualization of results.

Results: We identified 10 check types producing over 444 check applications and deployed them in 2 multi-institutional networks. Check types span structural conformance to a data model, utility for common research needs, and study-specific customization. All check types are customizable without dependencies between them. A dashboard visualizes results, permitting adjustments based on number of data sources, need for source masking, and the user's focus. All components can be applied as written to any data source using OMOP and are readily modified for other data models.

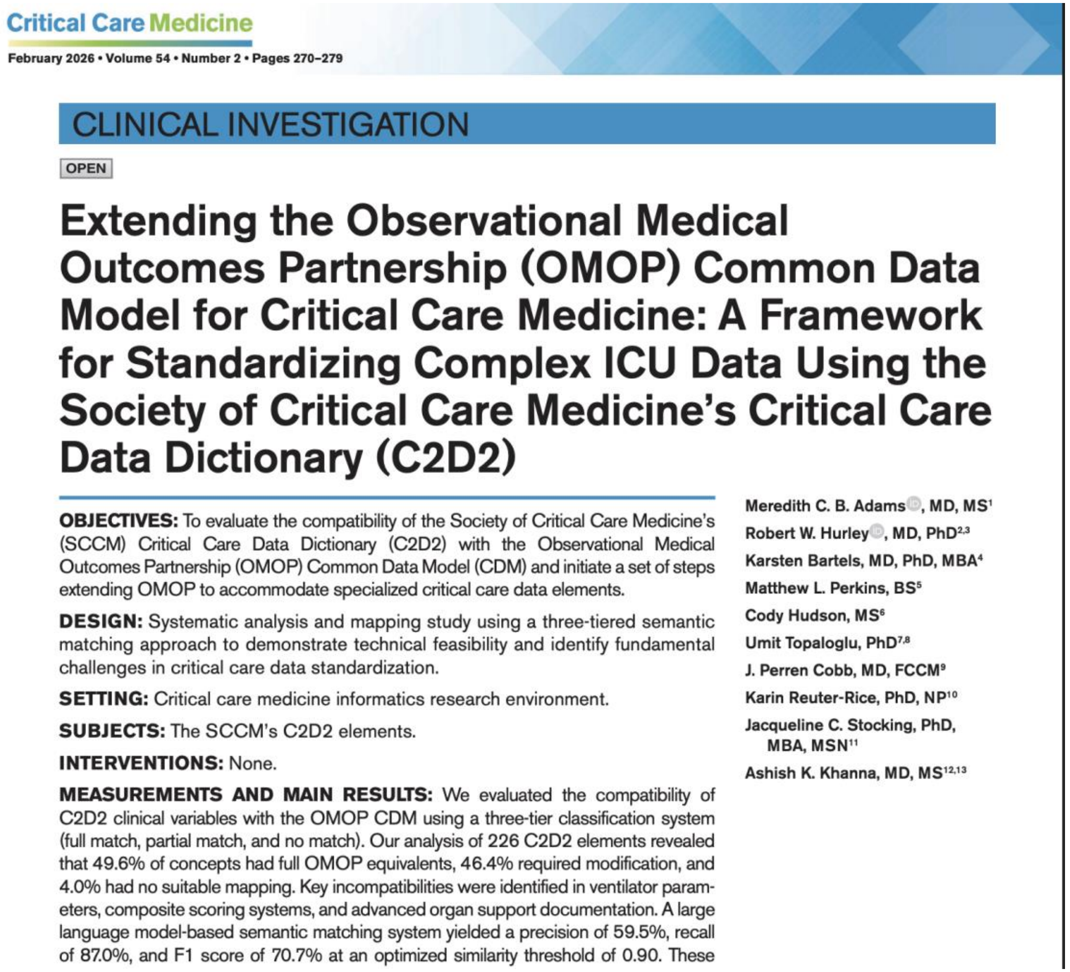
Discussion: We have extended previous work through our novel and multifaceted approach to data quality assessment, addressing needs in both network data improvement and research usage. We developed a capable and deployable system rather than tailoring to specific use cases.

Conclusion: Our novel DQ assessment system provides essential components for future standardization and collaboration to improve fitness of clinical data for intended use.

Key words: data quality; electronic health records; clinical research networks; multi-institutional research; learning health systems.

OHDSI Shoutouts!

Congratulations to the team of **Meredith C B Adams, Robert W Hurley, Karsten Bartels, Matthew L Perkins, Cody Hudson, Umit Topaloglu, J Perren Cobb, Karin Reuter-Rice, Jacqueline C Stocking, Ashish K Khanna** on the recent publication of **Extending the Observational Medical Outcomes Partnership (OMOP) Common Data Model for Critical Care Medicine: A Framework for Standardizing Complex ICU Data Using the Society of Critical Care Medicine's Critical Care Data Dictionary (C2D2)** in *Critical Care Medicine*.


Critical Care Medicine
February 2026 • Volume 54 • Number 2 • Pages 270–279

CLINICAL INVESTIGATION

OPEN

Extending the Observational Medical Outcomes Partnership (OMOP) Common Data Model for Critical Care Medicine: A Framework for Standardizing Complex ICU Data Using the Society of Critical Care Medicine's Critical Care Data Dictionary (C2D2)

Meredith C. B. Adams, MD, MS[1]
Robert W. Hurley, MD, PhD[2,3]
Karsten Bartels, MD, PhD, MBA[4]
Matthew L. Perkins, BS[5]
Cody Hudson, MS[6]
Umit Topaloglu, PhD[7,8]
J. Perren Cobb, MD, FCCM[9]
Karin Reuter-Rice, PhD, NP[10]
Jacqueline C. Stocking, PhD, MBA, MSN[11]
Ashish K. Khanna, MD, MS[12,13]

OBJECTIVES: To evaluate the compatibility of the Society of Critical Care Medicine's (SCCM) Critical Care Data Dictionary (C2D2) with the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) and initiate a set of steps extending OMOP to accommodate specialized critical care data elements.

DESIGN: Systematic analysis and mapping study using a three-tiered semantic matching approach to demonstrate technical feasibility and identify fundamental challenges in critical care data standardization.

SETTING: Critical care medicine informatics research environment.

SUBJECTS: The SCCM's C2D2 elements.

INTERVENTIONS: None.

MEASUREMENTS AND MAIN RESULTS: We evaluated the compatibility of C2D2 clinical variables with the OMOP CDM using a three-tier classification system (full match, partial match, and no match). Our analysis of 226 C2D2 elements revealed that 49.6% of concepts had full OMOP equivalents, 46.4% required modification, and 4.0% had no suitable mapping. Key incompatibilities were identified in ventilator parameters, composite scoring systems, and advanced organ support documentation. A large language model-based semantic matching system yielded a precision of 59.5%, recall of 87.0%, and F1 score of 70.7% at an optimized similarity threshold of 0.90. These

OHDSI Shoutouts!

Congratulations to the team of **Md Fantacher Islam, Molly Douglas, Jarrod Mosier and Vignesh Subbian** on the recent publication of **Standardizing Data Elements for Implementation of ICU Liberation Bundle** in *Applied Clinical Informatics*.

Accepted Manuscript

Submission Date: 2025-07-28
Accepted Date: 2026-02-02
Accepted Manuscript online: 2026-02-03

Applied Clinical Informatics

Standardizing Data Elements for Implementation of ICU Liberation Bundle

Md Fantacher Islam, Molly Douglas, Jarrod Mosier, Vignesh Subbian.

Affiliations below.

DOI: 10.1055/a-2802-7458

Please cite this article as: Islam M, Douglas M, Mosier J et al. Standardizing Data Elements for Implementation of ICU Liberation Bundle. ACI 2026. doi: 10.1055/a-2802-7458

Conflict of Interest: The authors declare that they have no conflict of interest.

Abstract:
Background and Significance: Getting patients out of intensive care units (ICUs) is a major goal for acute care clinicians, as prolonged stays increase the risk of complications and strain critical resources such as staff, equipment, and beds. The ICU Liberation bundle or the ABCDEF (A-F) care bundle is an evidence-based framework for improving outcomes in critically ill patients by addressing pain, sedation, delirium, mobility, and family engagement. However, variability in documentation and lack of standardized data elements hinder effective implementation and evaluation of adherence to bundle components.
Objectives: This study aims to characterize data elements of the A-F liberation bundle using a large, single-center critical care database and to develop standardized bundle cards that map bundle components to controlled vocabularies.
Methods: We conducted a retrospective analysis of data elements related to A-F bundle using the MIMIC-IV database. Clinical concepts were mapped to standardized vocabularies and aligned with the OMOP common data model. Bundle cards were developed for each component to provide structured, accessible documentation of assessment tools, adherence criteria, and terminology mappings.
Results: Pain assessments were documented in over 11,000 patients, with a median of 23 assessments per day. Sedation levels for nearly 59,000 patients were evaluated, with 37.7% meeting Society of Critical Care Medicine (SCCM) adherence criteria. Delirium assessments followed standardized protocols incorporating RASS and CAM-ICU scores. Components E and F lacked formal compliance specifications; bundle cards for these components identified key activities and highlighted gaps in standardized vocabularies. Adherence analyses revealed variability likely due to non-standardized documentation practices.
Conclusion: We developed and validated six ICU Liberation Bundle cards that map bundle components to standardized vocabularies and common data models, enabling retrospective adherence evaluation in real-world data. These information resources promote consistent documentation, support interoperability, and provide a foundation for prospective monitoring to enhance bundle implementation in critical care.

Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?

2026 Global Symposium

2026 OHDSI Global Symposium Call for Plenary Sessions

Symposium plenaries provide opportunities to share innovative, community-developed content to empower researchers to generate reliable real-world evidence. The community is currently seeking proposals for our #OHDSI2026 plenaries. These sessions will be 60 minutes in duration and must touch on at least two of following pillars of our community:

- Open community data standards
- Methodological research
- Open-source development
- Clinical applications

Plenary sessions must also involve three or more on-stage participants across at least two organizations. Sessions may include a combination of keynote talks, panel discussions, interactive activities, and more. We strongly encourage using multiple formats and synthesizing completed research, current perspectives and future calls-to-action to maximize community engagement.

The deadline for proposal submissions is January 30, 2026. Please use the link below to submit your proposal by answering the following questions:

- Name(s) of plenary session organizers:
- Your email address(es):
- Short (2,500 character max) description / abstract of your proposed session:
- Which pillars are you targeting:
- One sentence "pitch" of your session to excite the community:
- Names and roles of individuals who have tentatively agreed to participate in your session:

Deadline to submit proposals for #OHDSI2026 plenaries or tutorials is Feb. 20, 2026!

Who We Are | Updates & News | Standards | Software Tools | Network Studies | Community Forums | Education | New To OHDSI?
Community Calls | Past Events | Workgroups | Tutorials | 2025 'Our Journey' Annual Report | Current Events | Support & Sponsorship
2025 Global Symposium | 2025 APAC Symposium | 2026 Global Symposium | Github | YouTube | X/Twitter | LinkedIn | Newsletters

2026 OHDSI Global Symposium

Oct. 20-22 • New Brunswick, N.J. • Hyatt Regency Hotel

2026 OHDSI Global Symposium Call for Tutorials

Tutorial sessions aim to deliver educational content, led by community members who wish to train our global collaborators on scientific, technical, and other skills that can support advancing OHDSI's mission and the effective use of real-world data and the generation and dissemination of reliable real-world evidence. Examples of prior tutorials offered are provided here: https://www.ohdsi.org/tutorials.

Tutorial sessions are 4 hours in duration. Registrants for your tutorial will be requested to pay a registration fee. The fees will be used to offset the costs of the symposium and other OHDSI expenses. Sessions may include a combination of talks, interactive activities, and more. We strongly encourage using multiple formats to maximize community engagement. Your session must include at least three people from at least two different organizations.

The deadline for tutorial proposal submissions is January 30, 2026. Please use the link below to submit your proposal by answering the following questions:

- Name(s) of tutorial session organizers:
- Your email address(es):
- Short (2,500 character) description / abstract of your proposed session:
- Names and roles of individuals who have tentatively agreed to participate in your session:

2026 Europe Symposium

The 2026 OHDSI Europe Symposium returns to Rotterdam next year and will be held **April 18-20.**

Registration is open on the **OHDSI & OHDSI Europe** web sites.

Time	Symposium Agenda - Monday April 20, 2026	Location
8:00	*Registration and Coffee*	Queen's Lounge
9:00	**Welcome to OHDSI Europe** Dr. Renske Los, Department of Medical Informatics, *Erasmus MC* Dr. Aniek Markus, Department of Medical Informatics, *Erasmus MC*	Theatre
9:05	**Journey of OHDSI** Prof. Peter Rijnbeek, Chair Department of Medical Informatics, *Erasmus MC*	Theatre
9:30	**Collaborator Showcase - part 1** Moderated by Dr. Egill Fridgeirsson, Department of Medical Informatics, *Erasmus MC*	Theatre
10:00	**Speed networking**	Theatre
10:15	*Coffee Break & posters National Nodes*	Queen's Lounge
11:15	**Collaborator Showcase - part 2** Moderated by Dr. Egill Fridgeirsson, Department of Medical Informatics, *Erasmus MC*	Theatre
11:45	**Dreaming about the OHDSI journey ahead** Dr. Patrick Ryan, Vice President, Observational Health Data Analytics, *Johnson & Johnson* Dr. Renske Los, Department of Medical Informatics, *Erasmus MC*	Theatre
12:15	*Lunch break & networking & posters/demo's* *(Early investigator meeting - 13:00-13:45 Queen's Lounge)*	La Fontaine & Odyssee Room
13:45	**From dreams to reality** OHDSI Titan Award winners	Theatre
14:30	**Propositions for collaboration from the National Nodes** National Node leads	Theatre
14:45	*Coffee break & posters/demo's*	La Fontaine & Odyssee Room
16:15	**The OH Factor** To be announced	Theatre
17:00	**Closing**	Theatre
17:15	*Networking reception*	Queen's Lounge

Columbia DBMI Summer School

The 2026 Summer School in Observational Health Data Science & Informatics, AI, and Real World Evidence

COLUMBIA | COLUMBIA UNIVERSITY DEPARTMENT OF BIOMEDICAL INFORMATICS

OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

June 22–26, 2026, Columbia Biomedical Informatics

The Columbia OHDSI Summer School provides health professionals, researchers, and industry practitioners with an immersive, hands-on training to working with real-world health data and generating real-world evidence (RWE). Participants will explore the types of healthcare data captured during routine clinical care—such as electronic health records and administrative claims—and learn how to standardize these data using the OMOP Common Data Model to support collaborative, distributed research as part of a data network.

Over the course of the week, participants will engage with three real-world analytic use cases:

- **Clinical characterization** – using descriptive epidemiology to study disease natural history and treatment patterns
- **Population-level estimation** – applying causal inference to assess drug safety and comparative effectiveness
- **Patient-level prediction** – leveraging machine learning for early disease detection and precision medicine

Participants will be guided through the full RWE study lifecycle: from designing observational studies tailored to each use case, to applying open-source tools form the OHDSI community, and executing analyses across real-world data sources.

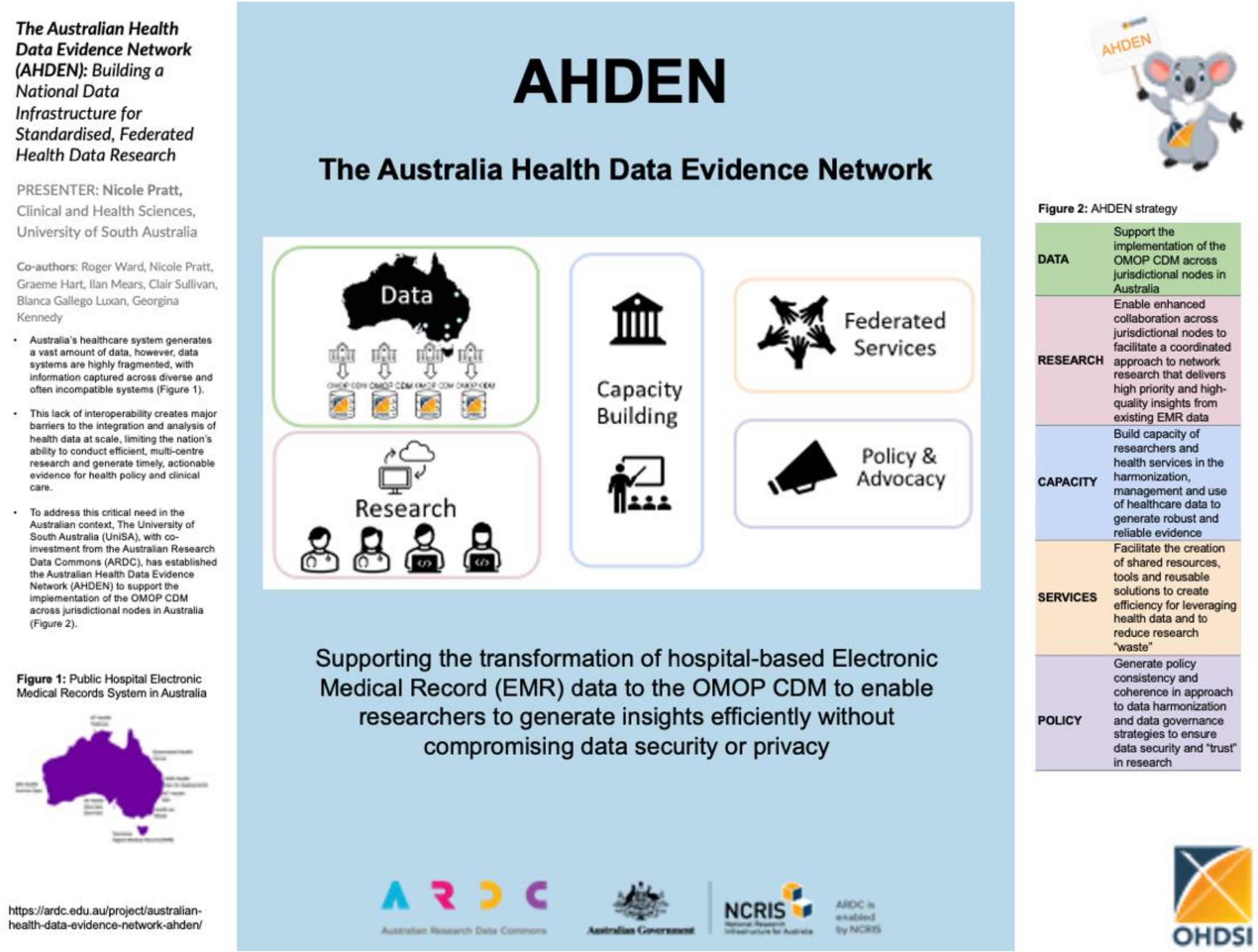
The curriculum combines foundational lectures on analytical methods with hands-on, interactive, faculty-led group exercises. In addition, participants will have dedicated time to develop and advance their own study concepts with personalized feedback and mentoring.

#OHDSISocialShowcase This Week

Monday

Australian Health Data Evidence Network (AHDEN): Building a National Data Infrastructure for Standardised, Federated Health Data Research

(Roger Ward, Nicole Pratt, Graeme Hart, Ilan Meyers, Clair Sullivan, Blanca Gallego Luxan, Georgina Kennedy)



The Australian Health Data Evidence Network (AHDEN): Building a National Data Infrastructure for Standardised, Federated Health Data Research

PRESENTER: Nicole Pratt, Clinical and Health Sciences, University of South Australia

Co-authors: Roger Ward, Nicole Pratt, Graeme Hart, Ilan Mears, Clair Sullivan, Blanca Gallego Luxan, Georgina Kennedy

- Australia's healthcare system generates a vast amount of data, however, data systems are highly fragmented, with information captured across diverse and often incompatible systems (Figure 1).
- This lack of interoperability creates major barriers to the integration and analysis of health data at scale, limiting the nation's ability to conduct efficient, multi-centre research and generate timely, actionable evidence for health policy and clinical care.
- To address this critical need in the Australian context, The University of South Australia (UniSA), with co-investment from the Australian Research Data Commons (ARDC), has established the Australian Health Data Evidence Network (AHDEN) to support the implementation of the OMOP CDM across jurisdictional nodes in Australia (Figure 2).

Figure 1: Public Hospital Electronic Medical Records System in Australia

https://ardc.edu.au/project/australian-health-data-evidence-network-ahden/

AHDEN

The Australia Health Data Evidence Network

Supporting the transformation of hospital-based Electronic Medical Record (EMR) data to the OMOP CDM to enable researchers to generate insights efficiently without compromising data security or privacy

Figure 2: AHDEN strategy

DATA	Support the implementation of the OMOP CDM across jurisdictional nodes in Australia
RESEARCH	Enable enhanced collaboration across jurisdictional nodes to facilitate a coordinated approach to network research that delivers high priority and high-quality insights from existing EMR data
CAPACITY	Build capacity of researchers and health services in the harmonization, management and use of healthcare data to generate robust and reliable evidence
SERVICES	Facilitate the creation of shared resources, tools and reusable solutions to create efficiency for leveraging health data and to reduce research "waste"
POLICY	Generate policy consistency and coherence in approach to data harmonization and data governance strategies to ensure data security and "trust" in research

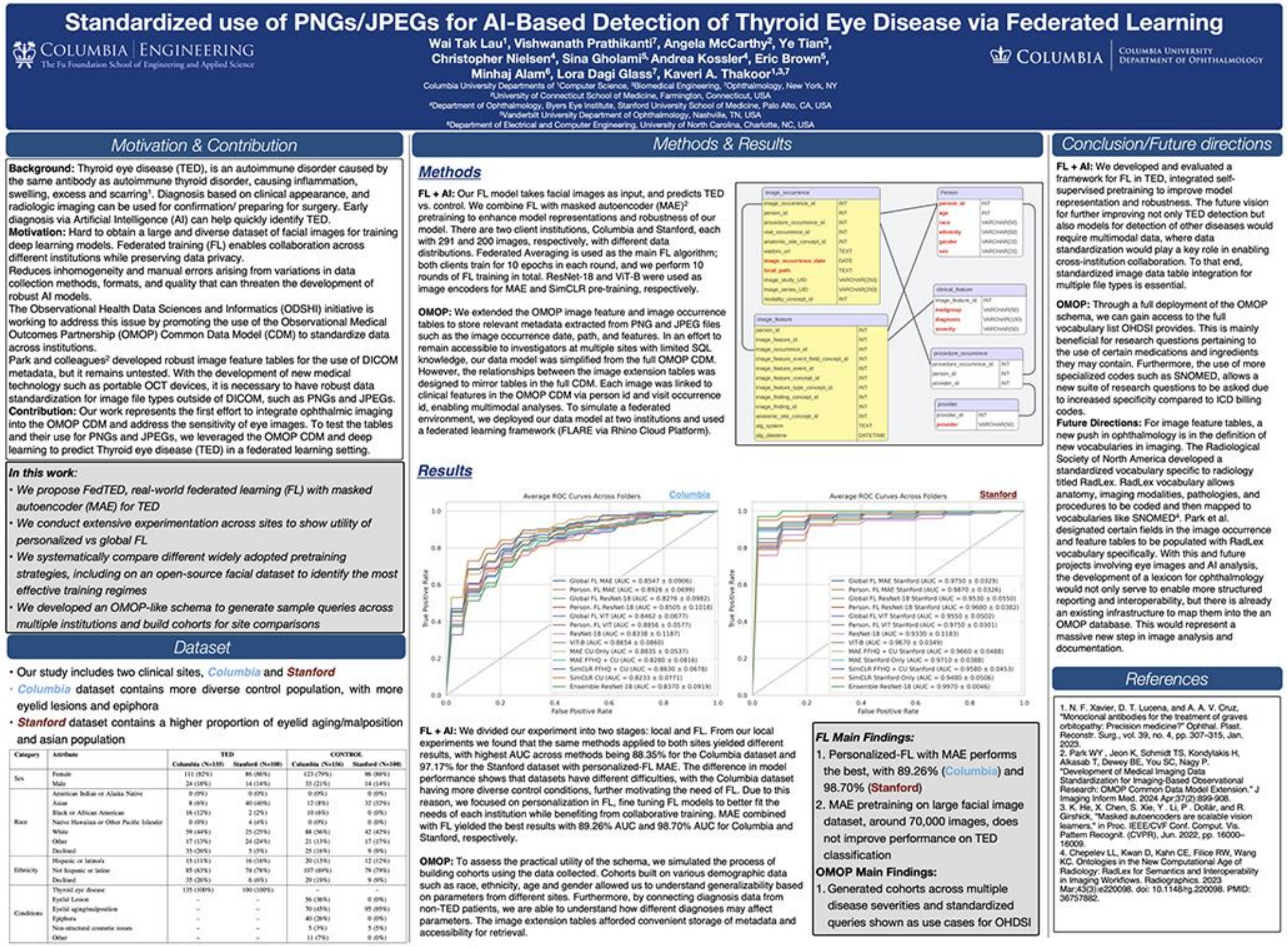
OHDSI

#OHDSISocialShowcase This Week

Tuesday

Standardized use of PNGs/JPEGs for AI-Based Detection of Thyroid Eye Disease via Federated Learning

(Michael Lau, Vishwanath Prathikanti, Angela McCarthy, Ye Tian, Christopher Nielsen, Sina Gholami, Andrea Kossler, Eric Brown, Minhaj Alam, Lora Dagi Glass, Kaveri A. Thakoor)



#OHDSISocialShowcase This Week

Wednesday

Transforming Breast and Cervical Cancer Screening Data into the OMOP CDM: Early Implementation Insights from Senegal

(Ousmane Diop, Rachel Odhiambo, Abdoulaye Samba Diallo, Ousmane Diouf, Bakary Dembo Diatta, Mamadou Lamine Cissé, Fatou Mbaye, Yacine Amet Dia, Mame Sokhna Gueye, Aminata Dia, Abdou Padane, Nafissatou Leye, Seyni Ndiaye, Abdoulaye Leye Sarr, Maryline Aza-Gnandji, Mamadou Ndao, Astou Guèye, Steve Bicko Cygu, Samuel Iddi, Miranda Barasa, Agnes Kiragga , Moussa Sarr, Souleymane Mboup, Aminata Mboup)

OHDSI

Transforming Breast and Cervical Cancer Screening Data into the OMOP CDM: Integrating Clinical and Genomic Insights from Senegal

Ousmane Diop[1], Rachel Odhiambo[3], Mamadou Lamine Cissé[2], Fatou Mbaye[2], Yacine Amet Dia[1], Abdou Padane[1], Nafissatou Leye[1], Steve Bicko Cygu[3], Samuel Iddi[3], Agnes Kiragga[3], Moussa Sarr[1], Souleymane Mboup[1], Aminata Mboup[1]

[1] Institut de Recherche en Santé, de Surveillance Épidémiologique et de Formations (IRESSEF), Dakar, Senegal
[2] Hôpital des Enfants de Diamniadio, Dakar, Senegal
[3] African Population and Health Research Center (APHRC), Nairobi, Kenya

Background

- Breast and cervical cancers remain major health burdens in LMICs, where screening coverage is limited and late-stage diagnoses are common.
- In Senegal alone, recent estimates reported about 2 000 new cervical cancer cases and nearly 1 800 breast cancer cases in 2022, making them the two most frequent cancers among women.
- Annual Pink October campaigns have become essential for raising awareness and providing community-based screening, yet the large volumes of data generated during these initiatives are rarely exploited for research or long-term follow-up.
- The OMOP Common Data Model (CDM), developed by the OHDSI community, offers a standardized framework to harmonize such data and enable reproducible, collaborative research.
- This study represents the first implementation in West Africa mapping cancer screening data into OMOP CDM while also integrating genomic and microbiome results, demonstrating feasibility in a low-resource setting and opening opportunities for future multi-omic cancer research in the region.

Methods

- This study analyzed data from 491 women who participated in a community-based breast and cervical cancer screening campaign jointly organized by IRESSEF and Diamniadio Children's Hospital.
- Dataset included demographics, clinical outcomes, HPV PCR results, lesion status, comorbidities, and metagenomic profiles of the vaginal microbiome and virome.
- Data transformation followed the OHDSI ETL framework using White Rabbit (profiling), Rabbit-in-a-Hat (ETL design), and Usagi/Athena (concept mapping), with implementation in PostgreSQL.
- Variables originally collected in French were translated into English to ensure OMOP compatibility.
- Data were mapped to core OMOP domains (person, observation, condition_occurrence, measurement, procedure_occurrence…), and microbial/viral abundances were captured as quantitative measurements—representing one of the first attempts to integrate multi-omic data into OMOP CDM in West Africa.

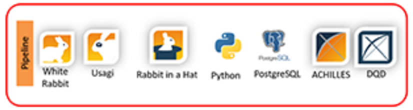


Figure 1 : ETL pipeline

Contact: ousmane.diop@iressef.org

Results

- The transformation achieved ~80% overall mapping to OMOP domains, with full coverage of demographic variables, 85% of clinical conditions, and 75% of laboratory and measurement data.
- Some specific variables, such as PCR-specific tests and genomic data required manual curation due to missing OMOP equivalents.
- Study population (n=491) ranged in age from 14–71 years (mean 35), with frequent comorbidities including diabetes and hypertension.
- Integration of microbiome and genomic data revealed distinct patterns: Gardnerella vaginalis, G. piotii, and G. swidsinskii were enriched among women with precancerous lesions and positive HPV tests, while Lactobacillus variants were more common in women without lesions, suggesting a potential protective role.
- Oncogenic HPV types (e.g., Alphapapillomavirus 7) and co-infections such as Human gammaherpesvirus 4 were also detected, demonstrating OMOP CDM's ability to support integrative multi-omic analyses in this setting.

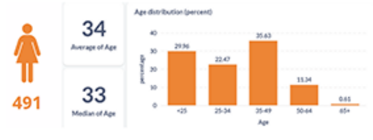


Figure 2 : Summary statistics

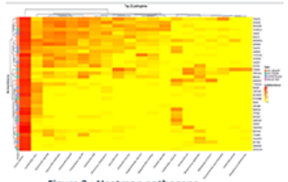


Figure 3 : Heatmap pathogens

Conclusions

This initiative represents the first implementation in West Africa that integrates clinical, demographic, and genomic cancer screening data into the OMOP CDM. By aligning local data with international standards, it lays the foundation for reproducible and comparable analyses across settings and fosters opportunities for collaborative research.

Early insights from the microbiome analysis, particularly the associations between Gardnerella species, Lactobacillus variants, and precancerous cervical lesions, point to potential biomarkers that warrant further investigation. Ultimately, this work illustrates how data standardization can bridge clinical research and bioinformatics in LMICs, enabling the development of more precise, context-specific strategies for cancer prevention and control. By sharing ETL scripts and mappings, this work can serve as a reusable template for other African institutions adopting OMOP CDM.

References

Ajah MI, et al. Exploring the role of cervico-vaginal microbiota as risk factor for cervical cancer in Sub-Saharan Africa: a systematic review and meta-analysis. BMC Women's Health. 2025;25(1):160. DOI:10.1186/s12905-025-03688-1

Musa J, Maiga M, Green SJ, et al. Vaginal microbiome community state types and high-risk human papillomaviruses in cervical precancer and cancer in North-central Nigeria. BMC Cancer. 2023;23:683. DOI:10.1186/s12885-023-11187-5

#OHDSISocialShowcase This Week

Thursday

Thematic Classification of Articles Using Graph Representations

(**Robert Barrett**, Haeun Lee, Paul Nagy)

Thematic Classification of Articles Using Graph Representations

PRESENTER: **Robert Barrett**

INTRO:
- Tracking community-driven impact on scientific literature is difficult
- Keyword-based retrieval of articles are often insufficient for topic identification
- Article retrieval through standard methods (APIs) are frequently limited to abstracts and metadata
- **Our team aimed to identify scientific articles influenced by the OHDSI community, allowing for more flexible search criteria to capture related articles within a broader reach.**

METHODS
1. 704 known OHDSI pre-determined articles and 877 non-OHDSI-related articles retrieved from PubMed, spanning 2010 to 2025 were used for training/testing. An 80:20 test/train split was used for model development
2. Preprocessing strategies evaluated:
   a. Numeric
   b. Removal of numeric tokens
   c. Stemming (Porter algorithm)
   d. Lemmatization (WordNet-based)
   e. Stop-word removal (standard NLTK stopword list)
3. A graph was constructed using the DOI, author name, and journal name as nodes. Directed edges were created for cited articles, authorship, and the journal they were published
4. PageRank, degree, related-author count, and citation overlap were calculated as features, in addition to extracted TF-IDF features from the abstract, title, and keywords
5. XGBoost/Logistic Regression were compared for discriminatory power

JOHNS HOPKINS
SCHOOL of MEDICINE

A system for classifying topic-related articles

Preprocessor	Pipeline	Precision	Recall	F1	Accuracy
Raw	TF-IDF	0.944	0.912	0.928	0.937
Raw	SciBERT	0.918	**0.926**	0.922	0.930
No numbers	TF-IDF	0.944	0.912	0.928	0.937
No numbers	SciBERT	0.915	0.918	0.916	0.925
Stemming	TF-IDF	0.950	0.909	0.929	0.938
Stemming	SciBERT	0.908	0.910	0.909	0.919
Lemma	TF-IDF	0.947	0.908	0.927	0.936
Lemma	SciBERT	0.913	0.922	0.917	0.926
Stopwords	TF-IDF	**0.949**	0.915	**0.931**	**0.940**
Stopwords	SciBERT	0.934	0.925	0.929	0.937

Table 1: Performance comparison of abstract only preprocessing strategies and pipelines

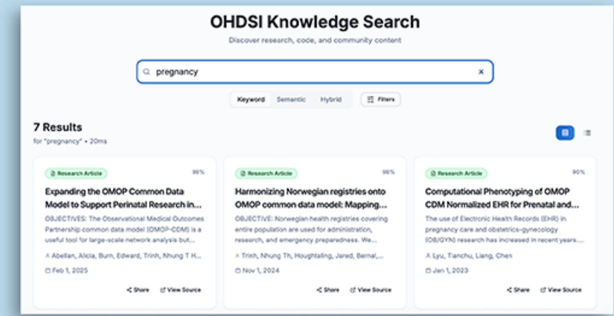


Figure 1: Visual demonstration of article retrieval post-classification

Take a picture to download the full paper

RESULTS
- In evaluating preprocessing strategies, TFIDF + LogisticRegression pipeline yielded the strongest overall scores:
  - Precision: 0.949
  - Recall : 0.915
  - F1 score : 0.931
  - Accuracy : 0.940
- XGBoost model achieved the best performance:
  - Accuracy: 0.953
  - Precision (PPV): 0.986
  - Recall (Sensitivity): 0.920
  - F1-score: 0.952
  - Specificity: 0.987
  - (TN = 148, TP = 138, FP = 2, FN = 12)

Feature importance analysis (mean decrease in impurity) was primarily driven by:
- OHDSI-related-author count
- Citation-overlap
- TF-IDF-derived signals such as "OMOP" and "common data model"

CONCLUSION
- Demonstrates high specificity/sensitivity classification of OHDSI-related articles
- Improved value in classification through graph-derived features
- Vocabulary coverage and preprocessing methods are critical for model success
- Potential for topic agnostic process for classification

ACKNOWLEDGEMENT
This research was supported by the NLM, part of the NIH, under award number T15LM013979

Barrett, Robert; Lee, Haeun; Nagy, Paul

OHDSI

#OHDSISocialShowcase This Week

Friday

Replicating Alzheimer's Research using standardized phenotyping with the OMOP common data model imaging extension

(Gabriel Lucca de Oliveira Salvador, Jen Park, Teri Sippel Schmidt, Blake Dewey, Paul Nagy)

Replicating Alzheimer's Research using standardized phenotyping with the OMOP common data model imaging extension

PRESENTER: Jen Park

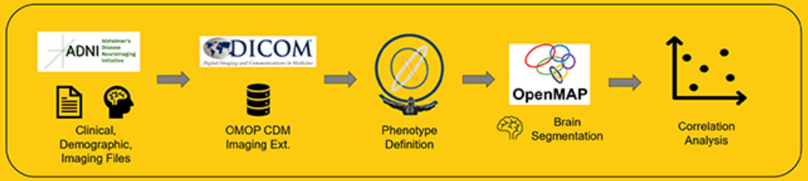
INTRODUCTION:

- We implemented the OMOP Medical Imaging extension (MI-CDM) on the public ADNI research database (phase 4), harmonizing MRI metadata and clinical measures to support imaging-enhanced computable phenotypes.
- This is the first in our knowledge to replicate the published AD study to test reproducibility of imaging research.

METHODS

1. We extracted Digital Imaging and Communications in Medicine (DICOM) metadata around the imaging acquisition parameters for magnetic resonance imaging.
2. This metadata was added to the OMOP CDM for DICOM vocabulary concepts. Demographic and neuropsychiatric data were transformed from the ADNI dataset and organized into OMOP CDM tables.
3. Inclusion/exclusion criteria were established for the patients with T1-weighted brain MRI scans and neuropsychiatric inventory scores. The images meeting this criteria was used to obtain volumetric values of brain regions using the OpenMap-T1 ML segmentation algorithm.
4. We replicated a published ADNI study correlating hippocampal volumes with varying degrees of dementia and AD.

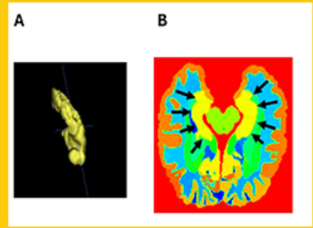
MI-CDM Enables Imaging-enhanced Computable Phenotypes, Advancing Reproducible Alzheimer's Disease Research.



CORRELATION ANALYSIS RESULTS

- We stratified by age, sex, and neurological condition to compare the right and left hippocampal volume from our algorithm and published work.
- Both models found reduced brain volume in AD groups compared to brain volumes with other neurological conditions.

Take a picture to download the full paper



BRAIN SEGMENTATION ALGORITHM COMPARISON

The figures illustrates the differences in segmentation patterns between the original study and OpenMap-T1.

A - Image from the study replicated showing the authors' semi-manual segmentation of one of the hippocampal formations.

B – Segmentation performed by OpenMap-T1 illustrating the head, body, and tail of the hippocampus (black arrows).

RESULTS

- We extracted and loaded 545 studies (4,756 series, 14,816 images) from 289 patients, plus 4,152 demographic records and 88,819 neuropsychiatric scores mapped to the Measurement table.
- By using the DICOM metadata and neuropsychiatric inventory (NPI) on ATLAS, we found patients meeting the criteria of having done volumetric T1 series with NPI survey completed.
- 100% of DICOM series identified from the cohort definition were able to run OpenMAP algorithm.
- The results of brain segmentation algorithm was loaded to the MI-CDM Measurement table, so that the correlation analysis could be done on ATLAS characterization tab by neurological condition.
- Results showed hippocampal volume decline with dementia and age.
- Visual review revealed that OpenMap-T1 segments the entire hippocampus, unlike the original study's focus on the hippocampal head and limited brain areas.

DISCUSSION

- We replicated a published imaging study that used the same research database (ADNI) with our internal brain segmentation algorithm.
- By organizing file-based database into standardized common data model, MI-CDM, we developed a system to replicate imaging research in computable and reproducible manner.
- Future studies should further evaluate reproducibility of imaging algorithms using MI-CDM.

Gabriel Salvador, Jen Park, Teri Schmidt, Blake Dewey, Paul Nagy

JOHNS HOPKINS MEDICINE

OHDSI

Where Are We Going?

Any other announcements of upcoming work, events, deadlines, etc?

Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?

The weekly OHDSI community call is held every Tuesday at 11 am ET.

Everybody is invited!

Links are sent out weekly and available at: ohdsi.org/community-calls-2025