



Tribute to a Titan: Jamie Weaver

OHDSI Community Call
Aug. 19, 2025 • 11 am ET



Upcoming Community Calls

Date	Topic
Aug. 26	Large-Language Model Innovations in OHDSI
Sept. 2	Standardized Vocabulary Summer Refresh Update
Sept. 9	Global Symposium Preview
Sept. 16	TBA
Sept. 23	Recent OHDSI Publications
Sept. 30	OHDSI 2025 Poster Preview Mad Minutes / Symposium Logistics
Oct. 7	No Call – OHDSI Symposium
Oct. 14	Welcome to OHDSI
Oct. 21	Meet the Titans



Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?





OHDSI Shoutouts!



Congratulations to the team of **Benjamin Martin, Will Kelly, Hannah Morgan-Cooper, Thomas Falconer, Elizabeth Park, Priya Desai, David Fiorentino, Lorinda Chung, Sean Yen, Zachary Wang, Didem Saygin, Michael George, Gowtham Rao, Joel Swerdel, Azza Shoaibi, and Christopher Mecoli** on the publication of **Identification of Adult Dermatomyositis Patients Using Real-World Data Sources** in *Arthritis Care & Research*.

Arthritis Care & Research

AN OFFICIAL JOURNAL OF THE AMERICAN COLLEGE OF RHEUMATOLOGY

Original Article | [Full Access](#)

Identification of Adult Dermatomyositis Patients Using Real-World Data Sources

Benjamin Martin, Will Kelly, Hannah Morgan-Cooper, Thomas Falconer, Elizabeth Park, Priya Desai, David Fiorentino, Lorinda Chung, Sean Yen, Zachary Wang, Didem Saygin ... [See all authors](#) ▾

First published: 12 August 2025 | <https://doi.org/10.1002/acr.25625>



Author disclosures are available at <https://onlinelibrary.wiley.com/doi/10.1002/acr.25625>. This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/acr.25625.

PDF TOOLS SHARE

Abstract

Objective

Studying rare diseases like dermatomyositis (DM) in single-center cohorts is challenging due to small sample sizes and limited generalizability. This study develops and evaluates case identification algorithms for DM to enable coordinated analysis across multiple data sources.



Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?





Upcoming Workgroup Calls



Date	Time (ET)	Meeting
Tuesday	12 pm	CDM Vocabulary Subgroup
Tuesday	12 pm	ATLAS/WebAPI
Wednesday	8 am	Psychiatry
Wednesday	11 am	Common Data Model
Wednesday	1 pm	Perinatal and Reproductive Health
Thursday	8 am	Medical Devices
Thursday	8 am	India Community Call
Thursday	9 am	Oncology Vocabulary/Development Subgroup
Thursday	11 am	Themis
Thursday	12 pm	HADES
Friday	9 am	Phenotype Development and Evaluation
Friday	10 am	Transplant
Friday	10 am	GIS – Geographic Information System
Friday	11 am	Clinical Trials
Friday	11:30 am	Steering
Monday	10 am	Healthcare Systems Interest Group
Tuesday	9 am	Data2Evidence



OHDSI 2025 Agenda Posted

Agenda • Tuesday, Oct. 7

Time (ET)	Session/Topic
7:00 am - 8:00 am	Lite Breakfast and Registration, Exhibits
8:00 am - 12:00 pm	Introductory Tutorial: An Introduction to the Journey from Data to Evidence Using OHDSI Vocabulathon 2025
12:00 pm - 1:00 pm	Buffet Lunch for Tutorial Registrants, Exhibits
1:00 pm - 5:00 pm	Advanced Tutorial: Developing and Evaluating Your Extract, Transform, Load (ETL) Process to the OMOP Common Data Model Advanced Tutorial: Using the OHDSI Standardized Vocabularies for Research Advanced Tutorial: Clinical Characterization Applications to Generate Reliable Real-World Evidence Advanced Tutorial: Population-Level Effect Estimation Applications to Generate Reliable Real-World Evidence Advanced Tutorial: Patient-Level Prediction Applications to Generate Reliable Real-World Evidence
5:00 pm - 6:00 pm	Collaborator Showcase Poster Placement
6:00 pm - 8:00 pm	Networking Reception; Collaborator Showcase Preview; Pre-Registration

Agenda • Wednesday, Oct. 8

Time (ET)	Topic
7:00 am - 8:00 am	Lite Breakfast and Registration, Exhibits
7:15 am - 7:45 am	Newcomer Orientation
8:00 am - 9:00 am	State of the Community: Welcome to OHDSI
9:00 am - 9:30 am	Group Networking Activity
9:30 am - 10:15 am	Collaborator Showcase Poster/Software Demo Session #1
10:15 am - 12:00 pm	Plenary
12:00 pm - 1:00 pm	Buffet Lunch, Exhibits
1:00 pm - 2:00 pm	Presentation
2:00 pm - 2:45 pm	Collaborator Showcase Lightning Talk Session #1
2:45 pm - 3:30 pm	Collaborator Showcase Poster/Software Demo Session #2
3:30 pm - 4:15 pm	Collaborator Showcase Poster/Software Demo Session #3
4:15 pm - 5:00 pm	Collaborator Showcase Lightning Talk Session #2
5:00 pm - 6:00 pm	Titan Awards, Wednesday Closing Activity
6:00 pm - 6:15 pm	Group Photo
6:15 pm - onward	Free Time

Agenda • Thursday, Oct. 9

Time (ET)	Meetings
7:00 am - 8:00 am	Lite Breakfast, Exhibits
8:00 am - 10:00 am	Session 1 of Workgroup Activities Featuring: Africa Chapter, APAC Chapter, Medical Imaging, GIS - Geographic Information System, HADES Hackathon, Oncology, Common Data Model, ATLAS/WebAPI, Phenotype Development and Evaluation, Dentistry, and Latin America
10:00 am - 10:30 am	Break, Exhibits
10:30 am - 12:30 pm	Session 2 of Workgroup Activities Featuring: Perinatal and Reproductive Health, Industry, Natural Language Processing, GIS - Geographic Information System, HADES Hackathon, Oncology, Common Data Model, ATLAS/WebAPI, Phenotype Development and Evaluation, Early-Stage Researchers, and Vocabularies
12:30 pm - 1:30 pm	Buffet Lunch and Exhibits
1:30 pm - 3:30 pm	Session 3 of Workgroup Activities Featuring: Surgery and Perioperative Medicine, Rare Diseases, Medical Devices, Psychiatry, HADES Hackathon, Health Equity, Evidence Network Data Partners, Data Bricks User Group, Eyecare and Vision Research, Women of OHDSI, CDM Survey
3:30 pm - 5:00 pm	Workgroup Summary

ohdsi.org/ohdsi2025



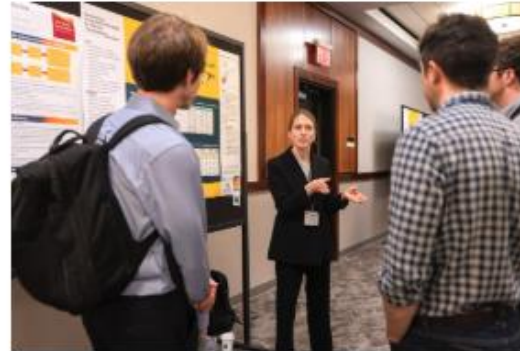
Global Symposium: Oct. 7-9



OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

- Who We Are ▾
- Updates & News ▾
- Standards
- Software Tools ▾
- Network Studies ▾
- Community Forums ▾
- Education ▾
- New To OHDSI? ▾
- Community Calls ▾
- Past Events ▾
- Workgroups ▾
- 2024 'Our Journey' Annual Report
- Current Events ▾
- Support & Sponsorship
- 2025 Global Symposium ▾
- 2025 Africa Symposium
- 2025 APAC Symposium
- Github
- YouTube
- Twitter
- LinkedIn
- Newsletters ▾

- 2025 Global Symposium Homepage
- Register for OHDSI2025
- OHDSI 2025 Agenda
- OHDSI 2025 Collaborator Showcase
- OHDSI 2025 Tutorials



2025 OHDSI Global Symposium

ohdsi.org/ohdsi2025



Titan Award Nominations Are Open

The Titan Awards have been handed out annually since 2018 to recognize OHDSI collaborators (or collaborating institutions) for their contributions towards OHDSI's mission.

Nominations for the 2025 Titan Awards are now open. **Please complete your nominations by our Sept. 9 (8 pm ET) deadline!**

ohdsi.org/titan-awards



Africa Symposium: Nov. 10-12

2025 Global Symposium ▾ 2025 Africa Symposium 2025 APAC Symposium Github YouTube Twitter LinkedIn Newslet

Join Us At The Inaugural OHDSI Africa Symposium

Nov. 10-12, 2025 • Joint Clinical Research Centre (JCRC) & Mestil Hotel Kampala



The inaugural OHDSI Africa Symposium will be held in Kampala at the Joint Clinical Research Centre (JCRC) and Mestil Hotel. Our community is delighted to introduce a new face-to-face opportunity in Africa, where OHDSI is growing at an exciting pace. We hope you will join us for this historical moment.

The first OHDSI Africa symposium will be hosted by JCRC and will begin with a dedicated one-day training course at JCRC, followed by a two-day main conference at Mestil hotel. Below are some important dates for you to save to your calendar:

Collaborator Showcase

- Submissions deadline: September 10
- Submissions review: September 11-30
- Notification of acceptance: October 5

Symposium

- Tutorial: November 10 at JCRC
- Main conference: November 11-12 at Mestil Hotel

Mestil Hotel Accommodations

Booking Code: JCRC

Booking Link: https://direct-book.com/properties/MestilDIRECT?promotion_code=JCRC25

Register Me for the 2025 OHDSI Africa Symposium!

2025 OHDSI Africa Symposium Full Agenda



OHDSI Africa Symposium			
JOINT CLINICAL RESEARCH CENTER Pre Symposium Training			
Monday Nov 10			
Time	Location	Description	Trainer
8:00	JCRC Cafeteria	Registration and Coffee	
9:00	JCRC training Room 1	Fundamentals: CDM and Standardized Vocabularies	
	JCRC training Room 2	Fundamentals: The ETL Process	
10:00	JCRC cafeteria	Coffee/Break	
10:30	JCRC training Room 1	Fundamentals: CDM and Standardized Vocabularies-continuation	
	JCRC training Room 2	Fundamentals: The ETL Process-continuation	
12:45	JCRC cafeteria	Lunch Break	
13:00		Multi-institutional Implementation: Data Science Without Borders	Agnes Kiragga
African Policymaker Perspectives			
11:45		Africa CDC	Bekure Tamirat
12:00		County Ministry of Health Kisumu, Kenya	Gregory Ganda
12:15		African Health Data Space (substitutable, tightly linked to VODAN)	TBD
12:30		Panel Discussion: How Can OHDSI Support Policymakers' Decision Making Process?	National Ministers, Bekure, Ganda, Kiragga, Ryan, Mui; (too many panelists?) Moderated by Alex Aswimee
13:00		Lunch	
Collaborators Showcase			
14:00	Hallways & Booth	Poster Presentations, Sponsor Booths, Open Source Software	Has a cost been established for Sponsor Booths? US\$200-\$500, Francis to firm up figures
16:00		Lightning Talks from 6 selected abstracts, 7 min each	
17:00		Networking Reception	
17:30		Talent Show, Cultural Program or other Entertainment	
Wednesday Nov 12			
8:00	Mestil Hotel Cafe	Registration and Coffee	
Building OHDSI Capacity in Africa			
9:00		Translations of the Book of OHDSI, example Kishwari	Michel Walravens, Pauline Andesco

Symposium DAY 1 Opening of OHDSI Meeting Mestil Hotel			
Tuesday Nov 11			
Time	Location	Description	Speaker
8:00 AM	Mestil Hotel Cafe	Registration and Coffee	Café area
Opening of OHDSI Meeting			
9:00AM	Burinyonyi hall Aswa & Kafu	Opening remarks	Dr. Cissy Kityo, JCRC Executive Director
9:10AM		OHDSI Global Perspective	Mul Van Zandt (APAC Chapter lead and deep involvement globally with IQVIA work)
9:20AM		Official opening of the meeting by the Minister of Health Uganda	Dr. Jane Ruth Aceng
9:30AM		Remarks from MOH Health Informatics Division	Paul Mbaka
9:40		Remarks from Ministry of Sci Tech & Innovation	Kenneth Bagarukaya or Brenda ...
OHDSI Implementations Across Africa			
9:50AM		History of OHDSI Africa, Country representation	Cynthia Sung, Duke-NUS
10:00AM		JCRC's journey with OHDSI - Uganda experience	Francis Kanyike, JCRC
10:15		Implementation in Rwanda @ Rwanda Biomedical Center and network across multiple hospitals	Jean-Claude Semuto or Pacificque Nizeyimana or Lars Halvorsen
10AM		Implementation in Cameroon Duolua General Hospital	Luc Baudoin Fankoua
Coffee / Tea Break			
10AM		Harmonising Medication Data to the OHDSI Vocabulary	Adam Bouras in Morocco
10AM		Harmonising Mental Health Data	
9:15		Standardizing Terminology Unique to the African Context	Andy Kanter (backup - Lars Halvorsen)
9:30		BRIDGE Training Network	Marc Twagirimukiza or Pascal Coorveits
9:45		Keynote Address: Generating reliable evidence to accelerate insights into population health and disease management	Patrick Ryan VP, Observational Health Data Analytics, J&J
10:15		Break	
		Connections with Other African Data Science Organizations	
TBD		APCC	Kobus Herbst/Agnes Kiragga
10:45		OpenMRS	Jayasanka Weerasinghe
11:00		DS-I Africa	Francis Agamah or Katherine Johnston from DS-I Africa Coordinating Center at Univ Cape Town
11:15		HELINA	Steven Waynee or Carolyn Kamanska
11:30		VODAN	Mouhamad Mpezamihigo, Vice Chancellor Equator University of Science & Technology Uganda, VODAN Board Member
12:00		Panel Discussion: Generating Synergy Among African Data Science and Informatics Communities	Jayasanka, Agamah, Waynee or Kamanska, Mpezamihigo, Moderated by Agnes Kiragga
12:30		Lunch	

ohdsi.org/africa2025



APAC Symposium: Dec. 6-7

The 2025 OHDSI APAC Symposium will be held Dec. 6-7 in Shanghai, China at the Shanghai Jiao Tong University. It will feature a 1-day tutorial and a 1-day main conference. Here are some important dates for you to save to your calendar:

Collaborator Showcase

- Submissions deadline: September 7
- Submissions review: September 8 – October 9
- Notification of acceptance: October 17



ohdsi.org/apac2025

SURVEY DATA AND THE OMOP CDM: LANDSCAPE ASSESSMENT

- <https://forms.gle/f18ufspAFT3jSYrk6>
- Open through August 31, 2025



#OHDSISocialShowcase This Week

Monday

Insights from the first OHDSI Germany Study-a-thon

(**Michele Zoch**, Martin Koch, Mario Menk, Corinna Grasemann, Mirko Gruhl, Ines Reinecke)

Study-a-thons boost collaboration and strengthen both methodological alignment and understanding of observational studies

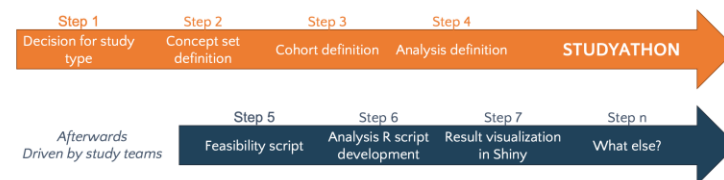
Insights from the first OHDSI Germany Study-a-thon

Background: Germany has established an active OHDSI National Node to align German healthcare data with the OMOP CDM. Initial efforts focused on ETL development, terminology mapping and adapting the OHDSI tools to national legal and technical requirements. A multi-stakeholder group was formed to promote collaboration and methodological alignment. To move from infrastructure to application, OHDSI Germany hosted its first Study-a-thon, offering hands-on experience with OHDSI tools and observational study design.

Methods: The Study-a-thon adapted the Hughes et al. (2022) model to address specific challenges in the German context (see table).

Challenges	Study-a-thon Solutions
Incomplete OMOP conversions	Hands-on, exploratory training
Lack of available data	Synthetic "playground" in ATLAS
Heterogeneous participant expertise	Focus on fundamentals: cohort building, design, tools
Need for national and international compliance	Alignment with the framework of the German Medical Informatics Initiative

Results



Limitation: The first OHDSI Germany Study-a-thon proved that intensive, cross-sector collaboration accelerates open science and builds methodological capacity. It bridged the gap between technical readiness and application. Structured follow-up will be key to sustaining its impact.



Michele Zoch (zoch@ohdsi.org), Martin Koch, Mario Menk, Corinna Grasemann, Mirko Gruhl, Ines Reinecke



#OHDSISocialShowcase This Week

Tuesday

Mapping Source Data to the OMOP Common Data Model for Telehealth in Chronic Heart Failure: Process, Challenges and Lessons Learned

(**Karl Kreiner**, Aaron Lauschensky, Martin Baumgartner, Dieter Hayn, Günter Schreier)

OMOP Registry of Telehealth Data for Over 5,600 Chronic Heart Failure Patients

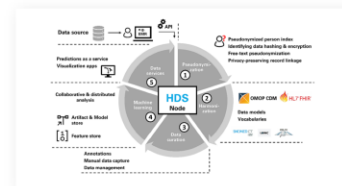
Title: Mapping Source Data to the OMOP Common Data Model for Telehealth in Chronic Heart Failure: Process, Challenges and Lessons Learned

Background: Herzmobil Tirol is an Austrian telehealth-supported Chronic Heart Failure (CHF) programme shown to reduce mortality and costs. As part of d4Health project, we mapped data from HMT, hospital systems, labs, and ERP software to the OMOP CDM and built a smart federated CHF registry for scalable, reproducible analysis.

Result 1: Tables and CDM mappings used

Table	Number of Records	Description	Ratio of standard concepts
Person	1,005	Record linked patients; person, name, address, and a pseudonym linked to EUHED service.	100%
Measurement	> 3.4 Mio	Laboratory data and self-reported biomarker data	91.72%
Observation	24,510	Personal visit, death	100%
ConditionOccurrence	147,413	Discharge diagnosis	99.26%
VisitOccurrence	10,171	In-patient, out-patient and homehealth visits	100%
Note	10,050	Pseudonymized clinical notes from the telehealth service	100%
DeviceExposure	> 1.8 Mio	Meta-data regarding devices used in the telehealth service	66%
DrugExposure	10,145	Documented drug prescription in the telehealth service	See discussion

Result 2: Health Data Nodes (HDS) – the build block of the CHF registry.



Methods

- 1 We mapped data from telehealth, hospital, lab, cost, and mortality registries to the OMOP CDM using automated ETL processes and **privacy-preserving record linkage**. Key variables were defined with clinical experts, and **free-text notes were pseudonymized** to protect patient privacy.
- 2 We implemented a Python-based software package called **"Health Data Nodes"** bundling an ETL framework for data ingestion, tools for data analysis and automatic data pseudonymization.

Limitation: Some data, like treatment phases, telehealth interactions, and regional drug details, were difficult to represent in the OMOP model. Despite these challenges, we plan to expand to other provinces and include new data using the federated Health Data Nodes.



Karl Kreiner¹, Aaron Lauschensky¹, Martin Baumgartner¹, Dieter Hayn¹, Günter Schreier¹
¹AIT Austrian Institute of Technology GmbH





#OHDSISocialShowcase This Week



National Institute for Public Health
and the Environment
Ministry of Health, Welfare and Sport

Wednesday

Achieving consensus on a OMOP CDM used for federative automated surveillance in healthcare associated infections: A Delphi study

(Sophie van Wingerden, Stephanie van Rooden)

Achieving consensus on an OMOP CDM for federated automated surveillance of healthcare associated infections: A Delphi study

Authors: S.H. van Wingerden¹, W.S. Brijs¹, R. Cornet¹, S. Haitjema¹, J. Herderschee¹, A.F. voor in't Holt¹, A.M. Kaiser¹, J. Kersten¹, P.P.A. Lastrade¹, J.J.W. Ploegmakers¹, A.E. Smilder¹, H.R.A. Streefkerk^{1,2}, S.R. van der Voort¹, S.M. van Rooden¹
¹ RIVM, National Institute for Public Health and the Environment, Center for Infectious Disease Control, ² St. Antonius Ziekenhuis Nieuwegein, ³ Amsterdam UMC, ⁴ UMC Utrecht, ⁵ OLVG Lab BV, Amsterdam, ⁶ Erasmus MC, ⁷ Franciscus Gasthuis & Vlieland Hospital, ⁸ Viecuri C, ⁹ UMC Groningen, ¹⁰ Meander MC Collaboration on Infectious Diseases.
Contact: sophie.vanwingerden@rivm.nl

Background

- The RIVM Center for Infectious Disease Control (CIb) is commissioned to monitor and prevent infectious diseases, including hospital-acquired infections (HAI).
- The classical manual surveillance method is error-prone and resource-intensive.
- In automated HAI surveillance algorithms are applied to routine care data from electronic health records (EHRs) to automatically detect infections according to surveillance definitions.
- The federated automated surveillance approach can optimize implementation of automated surveillance (figure 1), where algorithms and computing services are developed centrally and applied locally (in the hospitals) and only (aggregated) outcomes are collected centrally.
- Ideally FAIR data at the source (within hospitals) are being reused in a federated network. However, because of heterogeneity of the EHR data and systems data harmonization of routine care data is inevitable. Which is in line with the European Health Data Space (EHDS).

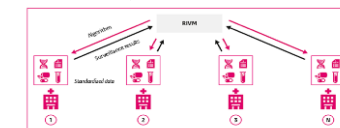


Figure 1: Schematic representation of federated automated surveillance. As a first step in the development of federated automated surveillance, this study aimed to assess whether consensus of clinical data representation within the EHR systems to a Common Data Model (CDM) (figure 2) could be reached.



Figure 2: Schematic representation of federated automated surveillance. As a first step in the development of federated automated surveillance, this study aimed to assess whether consensus of clinical data representation within the EHR systems to a Common Data Model (CDM) (figure 2) could be reached.

Methods

- Establishment of a Working group: Domain experts from 9 Dutch hospitals with multidisciplinary specialties: medical microbiologists, orthopedic surgeons, infection preventionists and IT specialists.
- Use case: semi-automated surveillance of surgical site infections after hip and knee replacement (PAS-ORTHO). For this surveillance outcome predefined data specifications are being used in a validated algorithm (table 1 and left panel figure 2).

Domains	Number of Variables	Variables
Demographics	2	Age, Sex, Date of Birth
Measurements	2	Weight, Height, BMI
Procedures	5	Procedure, Date of Surgery, Date of Discharge
Location	5	Department, Ward, Room, Room Number, Room Type
Medication	5	Medication, Date of Medication, Medication Type, Medication Dose
Diagnosis	5	Diagnosis, Date of Diagnosis, Diagnosis Type, Diagnosis Code
Total	32	

Table 1: Data specifications PAS-ORTHO surveillance

- A Delphi study: consisting of six rounds and aiming to reach a 75% consensus on mapping of the PAS-ORTHO data specifications to standard terminologies and ontologies. And achieve agreement to which source data these data specifications for the surveillance refer to.

Delphi Round	Information provided	Outcome Working group
1. Mapping data specifications	Delphi specification of PAS-ORTHO (mapped to a terminology of microbiology standards or ontologies)	Agreement on proposed mappings
2. Review round 1	Round 1 including responses from working group	Agreement on proposed mapping with inclusion input from round 1
3. Metadata of source data	Interpretation of the context on how the information is registered	Consensus of metadata of the original source data from EHR system
4. Comparability source data and mappings	Overview of the source data metadata and mappings	Evaluation of comparability of source data mappings of the CDM
5. Review round 5 and mapping	Round 5 including responses from working group	Finalization of comparability of source data mappings of the CDM and final mapping of the source data
6. Discussion meeting	Overview all of Delphi rounds and CDM decision making and output	Reaching consensus on data elements without agreement

Table 2: Overview Delphi rounds

- The rounds (table 2) were sent out in Excel files where experts could review the data elements and assess the level of agreement with the proposed translation through a multiple choice dropdown function (figure 3).



Figure 3: Method in Excel for determining agreement

Results

- Delphi rounds: The response rate was high and the working group finalized all rounds (table 2). A consensus of 55.6% was reached within the first five rounds. Most disagreement was around the microbiology items, due to differences in lab and specimen protocols.
- After a discussion meeting a final consensus percentage of 81.5% has been reached. Items that had not been agreed upon will be discussed with Nistiz (competence center for digital information management in healthcare).
- Source data metadata: for understanding the context of the registration of the source data, specific metadata (e.g. system, time, location, use of codes or standards) were collected. In certain cases finding the right person with knowledge on the source data registration appeared to be difficult (figure 4 part B).
- Consensus document: next to the CDM itself also the process of the decision making to the CDM will be described. This document enhances the reproducibility of the CDM and can be used for further research and other surveillance outcomes.

Discussion

First draft of the OMOP CDM based on the consensus of the mapping of the PAS-ORTHO data specifications and EHR data is expected in July 2021 (figure 2). The CDM will be tested by means of a connectathon. This event evaluates the whole pipeline from data extraction from EHR, to transformation to the CDM and running the algorithm for automated surveillance. During the connectathon the validity and comparability of routine care data transformed to a common data model (CDM) will be tested, and of surveillance outcomes based on automated surveillance where an algorithm is applied on transformed data.

Conclusions

- Reaching consensus on a CDM with domain experts is an important first step to harmonize the data in a valid and comparable manner for surveillance and research purposes.
- For in-depth information of the meaning of all data elements in the CDM, the link to the relevant source data has to be clear. Obtaining this metadata from the participating hospitals appeared time consuming and complicated by ambiguity in translation of the original source data and involving data experts within the hospital.
- Further assessment of quality of transformed and harmonized CDM data will be validated during a connectathon.



@OHDSI

www.ohdsi.org

#JoinTheJourney



ohdsi



#OHDSISocialShowcase This Week

Thursday

Valuing Clinical Text: From information extraction to predictive modeling

(Tom Seinen, Jan Kors, Peter Rijnbeek)

Extracting insights from clinical text
and quantifying its added value

Valuing Clinical Text: From information extraction to predictive modeling

Background • The adoption of electronic health records (EHRs) enables large-scale research using structured data, but critical information remains locked in unstructured clinical text, especially in non-English languages like Dutch. Our work, bundled in my thesis, addresses methodological and language barriers in extracting and using Dutch clinical narratives for diagnostic and prognostic predictive modeling.

Methods

Information Extraction • We validated Dutch concept extraction tools using translated clinical corpora and measured the overlap and unique insights between structured and unstructured data in Dutch GP records.

Clinical Prediction Models • We systematically reviewed models using clinical text and compared more than 80 prognostic prediction models trained on text and/or structured data from Dutch EHRs. Furthermore, we evaluated how text can refine unspecific clinical codes.

Results

- We demonstrated strong concept extraction performance on Dutch clinical text and quantitatively showed that unstructured text provides complementary information to structured data.
- Integrating unstructured text with structured data enhanced predictive accuracy across several tasks. Leveraging insights from clinical text also improved the specificity and quality of ambiguously coded structured data.

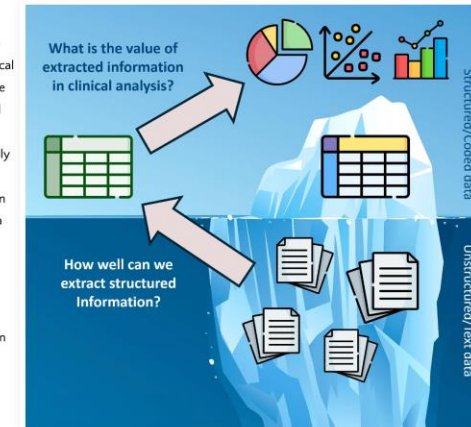


Diagram illustrating the process of extracting structured, coded data from unstructured clinical text. This transformation enhances the value of clinical narratives for clinical analysis, such as predictive modeling. The top half shows structured data's directly available for clinical analysis, while the bottom half represents the bulk of data residing in unstructured formats.



Conclusion: Integrating clinical texts into large-scale research enhances prediction models and data quality. Ongoing work on multilingual NLP, interpretable models, and external validation is crucial to deliver actionable insights and improve patient outcomes across healthcare systems.

Tom M Seinen
Department of Medical Informatics, Erasmus MC, The Netherlands
t.seinen@erasmusmc.nl





#OHDSISocialShowcase This Week

Friday

Loss function influence on hyperparameter optimization for observational healthcare prediction models

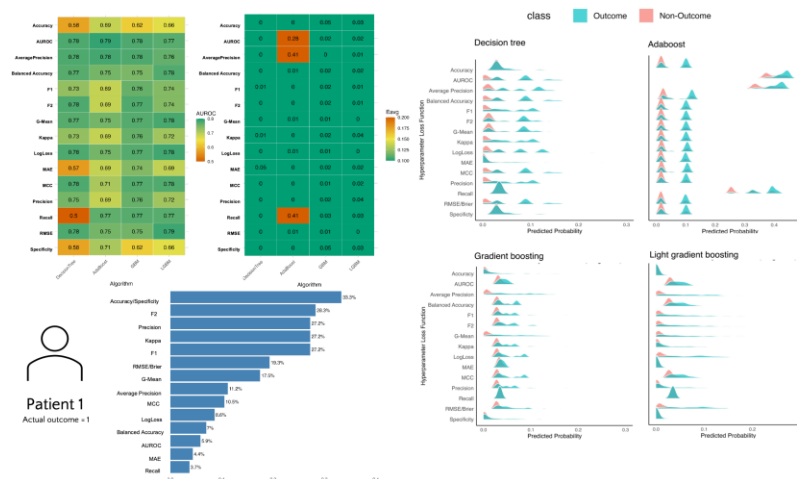
(Fleur Vereijken, Jenna Reys, Peter Rijnbeek, Ross Williams)

The choice of hyperparameter optimization makes a difference

Hyperparameter loss function influence on observational healthcare prediction models

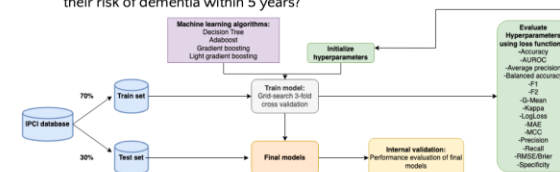
Background: Hyperparameters significantly impact model development by influencing model complexity, behavior, and training efficiency. Achieving "optimal" model performance requires careful tuning of these hyperparameters. How can we determine which hyperparameter combination is truly 'optimal'? This research evaluates how different loss functions used in hyperparameter optimization affect performance in observational healthcare models.

Results



Methods

Prediction task: For patients (55-84 y) with their first outpatient visit in 2014, what is their risk of dementia within 5 years?





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**