



Workgroup OKRs, Session 2

OHDSI Community Call
Feb. 24, 2026 • 11 am ET



Upcoming Community Calls

Date	Topic
Feb. 24	Workgroup OKRs
Mar. 3	New Vocabulary Release
Mar. 10	The Future of Education in OHDSI
Mar. 17	TBA
Mar. 24	OHDSI/OMOP Research Spotlight
Mar. 31	Kickoff to Phenotype April



Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?



OHDSI Shoutouts!



Congratulations to the team of **Andrea Miquel-Dominguez, Eng Hooi Tan, Edward Burn, Antonella Delmestri, Talita Duarte-Salles, Asieh Golozar, Wai Yi Man, Daniel Prieto-Alhambra, Francesc Xavier Avilés-Jurado and Danielle Newby** on the recent publication of **Incidence and survival of head and neck cancers in the United Kingdom 2000-2021** in *Cancer Epidemiology*.

Cancer Epidemiology 101 (2026) 103018

Contents lists available at ScienceDirect

Cancer Epidemiology

journal homepage: www.elsevier.com/locate/canep

ELSEVIER

Incidence and survival of head and neck cancers in the United Kingdom 2000–2021

Andrea Miquel-Dominguez^a, Eng Hooi Tan^{b, c}, Edward Burn^b, Antonella Delmestri^b, Talita Duarte-Salles^{c, d}, Asieh Golozar^{e, f}, Wai Yi Man^b, Daniel Prieto-Alhambra^{b, d, g}, Francesc Xavier Avilés-Jurado^g, Danielle Newby^{b, c}

^a Otorrinolaringología departament, Hospital Joan XXIII de Tarragona, C/ Dr. Mallafrè Guasch, 4, Tarragona 43005, Spain
^b Health Data Sciences (HDS), Translational Sciences, Botnar Research Centre, University of Oxford, Oxford OX3 7LD, UK
^c Fundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (IDIAPJGol), Barcelona, Spain
^d Department of Medical Informatics, Erasmus University Medical Centre, Rotterdam 3015 GD, the Netherlands
^e Nemesis Health, New York, USA
^f OHDSI Center at the Roux Institute, Northeastern University, Boston, MA 02115, USA
^g Head Neck Tumors Unit., Hospital Clínic de Barcelona, Universitat de Barcelona, IDIBAPS, IISPV, Vilarroel 170, Barcelona, Catalonia 08036, Spain

ARTICLE INFO

Keywords:
 Head and neck neoplasms
 Mouth neoplasms
 Oncology
 Incidence
 Cancer survival
 Survival rate
 United Kingdom

ABSTRACT

Background: Understanding the changing burden of head and neck cancers (HNC) is essential to guide public health interventions and inform cancer care strategies.
Methods: We conducted a cohort study using routinely collected primary care data Clinical Practice Research Datalink (CPRD) GOLD from the United Kingdom. Adults aged ≥ 18 years with ≥ 1 year of prior history were included. We estimated crude and age-standardised incidence rates (IRs) and one-, five-, and ten-year survival from 2000 to 2021, stratified by age and calendar year. Findings from CPRD GOLD were compared with primary care data from CPRD Aurum (England only).
Results: There were 12,455 patients with a diagnosis of HNC from CPRD GOLD (69.2 % male; median age 64 years). Crude incidence in GOLD increased from 9.08 (95 % CI: 7.88–10.42) per 100,000 person-years in 2000–15.59 (14.07–17.23) in 2021, with similar trends observed in CPRD Aurum. Age-standardised incidence trends were attenuated overall but remained elevated for oropharyngeal and tongue cancers. Five-year survival improved modestly, from 53.8 % (95 % CI: 51.4–56.3 %) in 2000–2004–58.7 % (56.5–60.9 %) in 2015–2019.
Conclusions: Incidence increases for HNC were attenuated after age standardisation, suggesting a contribution of demographic ageing, although elevations persisted for specific subsites. Small improvements in long term survival highlights more research is needed to improve earlier diagnosis which will lead to better patient outcomes.



OHDSI Shoutouts!



Congratulations to the team of **Blanca Beumer Prieto, Daniel Prieto Arribas, Isabel Moreno-Parro, Berta Sufrate-Vergara, Pilar Fernández-Calle, Roberto Mora Corcovado, Antonio Buño Soto, and Jorge Díaz-Garzón Marco** on the recent publication of **Applying the OMOP common data model to laboratory data: challenges, opportunities, and implications – a use case in biological variation research in *Clinical Chemistry and Laboratory Medicine*.**

DE GRUYTER

Clin Chem Lab Med 2026; aop

Letter to the Editor

Blanca Beumer Prieto, Daniel Prieto Arribas, Isabel Moreno-Parro, Berta Sufrate-Vergara, Pilar Fernández-Calle, Roberto Mora Corcovado, Antonio Buño Soto and Jorge Díaz-Garzón Marco*

Applying the OMOP common data model to laboratory data: challenges, opportunities, and implications – a use case in biological variation research

<https://doi.org/10.1515/ccdm-2025-1668>
Received December 15, 2025; accepted February 10, 2026;
published online February 20, 2026

Keywords: data standardisation; electronic health records; laboratory information systems; real-world data; semantic interoperability

To the Editor,

Clinical laboratory data are central to diagnosis, disease monitoring, and increasingly to real-world evidence generation. However, Laboratory Information Systems (LIS) remain highly heterogeneous, with locally defined structures, codes, and workflows that limit data reuse and comparability across institutions. This lack of standardisation represents a major barrier to multicentre research, particularly for analyses that depend on longitudinal and harmonised measurements, such as indirect estimation of biological variation (BV).

The relevance of addressing this challenge is reinforced by the development of the European Health Data Space (EHDS), which aims to enable primary use for cross-border patient care as well as the secure secondary use of health data for research, innovation, and policy [1]. Without shared data models and common semantics, laboratory data risk becoming a limiting factor rather than an enabler for multisite real-world evidence studies. In this context, clinical laboratories must take an active role in data harmonisation efforts, contributing domain expertise to the design and adoption of interoperable data standards rather than remaining secondary contributors to data integration initiatives.

Common data models (CDMs), such as the Observational Medical Outcomes Partnership (OMOP) CDM, offer a potential solution by enabling reproducible analytics through structural and terminological standardisation across heterogeneous health data sources [2]. When appropriately implemented, OMOP can support federated analyses, cross-database comparability, and scalable real-world evidence generation, including laboratory-driven research questions.



Three Stages of The Journey

Where Have We Been?

Where Are We Now?

Where Are We Going?



Upcoming Workgroup Calls



Date	Time (ET)	Meeting
Wednesday	10 am	Common Data Model
Wednesday	12 pm	Latin America Chapter
Thursday	9 am	Africa Chapter
Thursday	10 am	GIS – Geographic Information System
Friday	11:30 am	Steering

ohdsi.org/upcoming-working-group-calls



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 <i>Dr. Renske Los, Department of Medical Informatics, Erasmus MC</i> <i>Dr. Aniek Markus, Department of Medical Informatics, Erasmus MC</i>	Theatre
9:05	Journey of OHDSI <i>Prof. Peter Rijnbeek, Chair Department of Medical Informatics, Erasmus MC</i>	Theatre
9:30	Collaborator Showcase - part 1 Moderated by <i>Dr. Egill Fridgeirsson, Department of Medical Informatics, Erasmus MC</i>	Theatre
10:00	Speed networking	Theatre
10:15	Coffee Break & posters National Nodes	Queen's Lounge
11:15	Collaborator Showcase - part 2 Moderated by <i>Dr. Egill Fridgeirsson, Department of Medical Informatics, Erasmus MC</i>	Theatre
11:45	Dreaming about the OHDSI journey ahead <i>Dr. Patrick Ryan, Vice President, Observational Health Data Analytics, Johnson & Johnson</i> <i>Dr. Renske Los, Department of Medical Informatics, Erasmus MC</i>	Theatre

12:15	Lunch break & networking & posters/demo's <i>(Early investigator meeting - 13:00-13:45 Queen's Lounge)</i>	La Fontaine & Odyssee Room
13:45	From dreams to reality <i>OHDSI Titan Award winners</i>	Theatre
14:30	Propositions for collaboration from the National Nodes <i>National Node leads</i>	Theatre
14:45	Coffee break & posters/demo's	La Fontaine & Odyssee Room
16:15	The OH Factor <i>To be announced</i>	Theatre
17:00	Closing	Theatre
17:15	Networking reception	Queen's Lounge



2026 Global Symposium

The 2026 OHDSI Global Symposium will return to the Hyatt Regency Hotel in New Brunswick, N.J., on **Oct. 20-22.**





2026 Global Symposium Scientific Review Committee

The deadline to sign up for the Scientific Review Committee is March 1. The first meeting will take place on Thursday, March 5th at 11 am ET.





ATHENA Survey

Athena user survey

Help us understand how to make Athena better

When you submit this form, it will not automatically collect your details like name and email address unless you provide it yourself.

* Required

1. If you are open to follow-up about your feedback, please provide your email address

Enter your answer

2. How do you use athena.ohdsi.org? *

- Search concepts
- Download current version of vocabularies
- Download previous version of vocabularies
- Other

3. What do you use athena.ohdsi.org for? *

- ETL data
- Search concepts to create mappings
- Search concepts for concept sets (value sets, code lists)
- Translate concepts to other languages/find translations
- Use as knowledge graph outside of OMOP CDM
- Other

4. On average, how often do you access athena.ohdsi.org? *

- Every day
- Once a week
- A few times a month
- Once a month
- Once 6 month or less
- Other

5. If there was an Athena API, how would you use it and what would you use it for? *

Enter your answer

6. Anything else you'd like to tell us? *

Enter your answer

You can print a copy of your answer after you submit

Submit



#OHDSISocialShowcase This Week

Tuesday

OMOP Waveform Extension: A Schema for Integrating Physiological Signals and Derived Features into the OMOP CDM

(Jared Houghtaling, Polina Talapova, Brian Gow, Manlik Kwong, Andrew J King, Benjamin Moody, Mike Kriley, Tom Pollard, Andrew E Williams)

OMOP Waveform Extension: A Schema for Integrating Physiological Signals and Derived Features into the OMOP CDM

PRESENTERS: Jared Houghtaling, Polina Talapova

INTRO

- Physiologic waveforms & biosignals (ECG, EEG, ABP, SpO₂, respiratory) carry prognostic signals that structured EHRs miss.
- Mixing waveforms with EHR boosts prediction and phenotyping power (multimodal > unimodal).
- OMOP v5.4 has no dedicated structure for waveforms; most sites silo files off-platform.
- We built an OMOP-aligned extension so multi-site research can finally use these data at scale.

METHODS

Design principles: OMOP conventions (PK/FK, domains, concept IDs), minimal new tables, external storage for raw signals. Inspired by OMOP Imaging extension patterns.

The 4 tables:

1. **WAVEFORM_OCCURRENCE** - the acquisition session (person, visit, start/end).
2. **WAVEFORM_REGISTRY** - file/object index (format, path/URLs, hashes).
3. **WAVEFORM_CHANNEL_METADATA** - per-channel facts (lead/type, sampling rate, units, gain).
4. **WAVEFORM_FEATURE** - derived metrics (numeric/categorical/interval) + algorithm provenance and channel link.

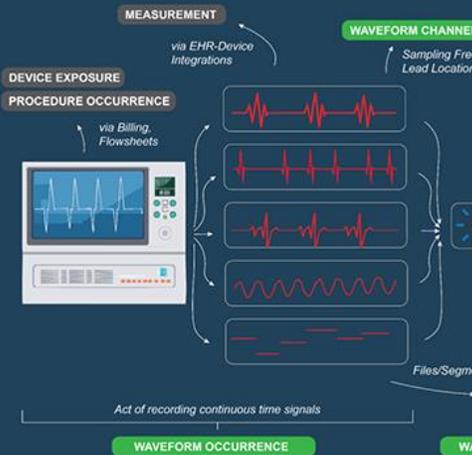
What we mapped: exemplar ICU/telemetry recordings (MIMIC & partner data) into the extension; iterated schema + vocab.

RESULTS

- The schema captured diverse signals (ECG, pleth, ABP, respiratory) and multi-channel structure without info loss.
- Channel metadata preserved what analysis needs: type, rate, units, gain, method.
- Features (e.g. heart rate, QT/QTc, HRV, EEG burst-suppression ratio) were stored with standard concept IDs and timestamps - all queryable alongside routine OMOP measurements.
- Raw files remained outside the CDM, yet traceable via registry references and hashes.
- Current OHDSI tools don't natively render these tables (yet), but SQL/Atlas recipes already let cohorts pull waveform-derived features today.

From monitor beeps and squiggles to a symphony OMOP can conduct

Four tables turn raw signals, channels, and features into score-ready parts - so cohorts hear the whole patient, not just the highlights



Take a picture to download the full paper

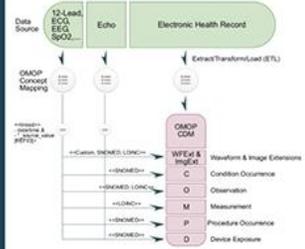


Figure 1. Signal life cycle. Signals (ECG/EEG/ABP/SpO₂, echo) go through ETL into OMOP CDM. WAVEFORM_OCCURRENCE captures the recording session (who/when/visit). WAVEFORM_REGISTRY indexes the file/object and path/hash.





#OHDSISocialShowcase This Week

Wednesday

Improving VSAC to OMOP Mapping Using LLM Assisted Curation

(Robert Barrett, Star Liu, Kyle Zollo-Venecek, Benjamin Riesser, Benjamin Martin)



Improving VSAC to OMOP Mapping Using LLM Assisted Curation

Robert B Barrett^a, Star Liu^a; Kyle Zollo-Venecek^b; Benjamin Riesser, MPA^c; Benjamin Martin, PhD^a

^aBiomedical Informatics and Data Science, Johns Hopkins University, Baltimore, MD, USA

^bCTSI, Tufts University School of Medicine, Boston, MA

^cImproving Health Outcomes, American Medical Association, Greenville, SC

Background

The National Library of Medicine (NLM) Value Set Authority Center (VSAC) provides a crowd-sourced repository of standardized representations of clinical concepts, commonly used in quality measures, decision support tools, and for interoperability. Yet, VSAC value sets are not without error and have demonstrated significant variation in quality measure outcomes when addressed. These errors typically manifest through missing, extraneous, or misconstrued codes. Separately, cross-terminology evaluations (ICD-10 to SNOMED CT) showed divergence in several domains².

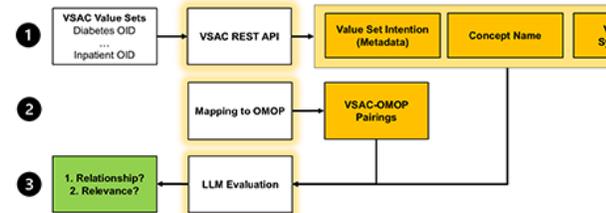
Recent studies demonstrated that large language models (LLMs) can detect or correct such mapping errors with high validity. GPT-4-based methods achieved a 96% valid-term match rate when screening local terminology against SNOMED CT³, and an additional study using LLMs for data-exchange tasks highlighted their utility for harmonizing structured clinical data⁴.

This study used OpenAI's GPT-4o to map six commonly used VSAC value sets and evaluate the relationship between VSAC-OMOP pairings, as well as the concept relevance to Value Set purpose.

Methods

To translate VSAC value sets into validated OMOP concepts, we executed a three-phase pipeline.

Figure 1. VSAC value sets to OMOP concepts mapping pipeline.



The filtering LLM agent was instructed to classify the mappings based on concept relevance and equivalence below:

OMOP Concept to VSAC Concept Relationship	Relevance to Value Set Purpose
Equivalent: Clinically identical concepts, allowing minor wording differences or updated coding definitions if the core meaning is preserved.	In scope: Mapping is consistent with the purpose of the value set.
Narrower: OMOP concept is more exclusive	Out of scope: Mapping is irrelevant to the purpose of the value set.
Broader: OMOP concept is more inclusive	
Unrelated: Concepts are clinically different.	

Four informaticists independently annotated 283 VSAC-to-OMOP concept relationships and relevance. Ground truth was derived via majority voting. Performance was evaluated using precision, recall, F1-score, accuracy, and Cohen's κ to compare GPT-4o results against human consensus.

Contact: rbarrel6@jh.edu

Results

GPT-4o was assessed on 283 mappings across 6 VSAC value sets, covering cardiovascular, endocrine, renal, and administrative domains. For concept relationship classification, GPT-4o achieved high performance (F1-score = 0.968, Cohen's κ = 0.918), with precision of 0.972 and recall of 0.965 (TP: 273, FP: 8, TN: N/A, FN: 10). Value set relevance performance was lower (F1-score = 0.500, κ = 0.493), primarily due to the rarity of unrelated (n=4) concepts (TP: 277, TN: 2, FP: 2, FN: 2).

Value Set	VSAC Concept	OMOP Concept	Disagreement Type	LLM Perspective
Dialysis Services	Hemoperfusion (eg, with activated charcoal or resin)	Hemoperfusion	Technical vs Clinical Scope	Exclude (out-of-scope): "not a dialysis service"
Diabetes	Type 1 diabetes mellitus with periodontal disease	Periodontal disease	Composite-Component Granularity	Narrower relationship; exclude: missing diabetes qualifier
Diabetes	Other specified diabetes mellitus with periodontal disease	Periodontal disease	Composite-Component Granularity	Narrower relationship; exclude: missing diabetes qualifier





#OHDSISocialShowcase This Week

Thursday

Evaluating the effectiveness of using Large Language Models for the development of concept sets

(Joel Swerdel, Dmytro Dymshyts, Anna Ostropolets, Azza Shoaibi, Patrick Ryan, Martijn Schuemie)

Presenter: Joel Swerdel

BACKGROUND

- When developing phenotype definitions for health conditions, one of the first steps is to develop the list of codes used to determine the phenotype.
- In OHDSI, health condition standard codes are usually SNOMED concepts mapped from diagnosis source codes (such as ICD-10-CM) used in the native data.
- Selected concepts are then grouped into concept sets
- Our current method for creating concept sets for phenotype development is inefficient, subjective, and inconsistent.
- Objective: to evaluate the use of Large Language Models (LLMs) for the selection of appropriate condition codes for concept sets used in phenotype algorithms.

METHODS

Process steps:

- Define the clinical idea
 - Create a list of the concept(s) and descendants based on the clinical idea
 - e.g., clinical idea = "Right heart failure" starting concept = "Right ventricular failure", SNOMED concept ID: 4273632
 - Add in all the recommended concepts from PHOEBE
 - Pass the list to LLM for analysis and process results
 - Create a new list of the concepts selected by the LLM and their descendants
 - Pass the new list to the LLM for analysis and process results
 - Create a final list of all the selected concepts
 - Put the concept set into ATLAS
- We tested the LLM process on 15 health conditions: acute myocardial infarction, ischemic stroke, hemorrhagic stroke, plaque psoriasis, nausea, COVID-19, pneumonia, obesity, angina, bronchitis, diarrhea, hypotension, insomnia, pyrexia-fever, and syncope.
 - For each condition, we created human-adjudicated concept sets developed through a collaboration of at least two researchers knowledgeable in concept set development.
 - These concept sets were the "reference standard" to be used for comparison with the LLM-generated concept sets.
 - For this study we used the licensed Johnson and Johnson version of the OpenAI LLM, (OpenAI Model GPT-4o, trained through October, 2023).
 - Procedural calls to the application programming interface (API) for the LLM were made using the R platform.

Evaluation:

Concept Evaluation:

- The number of common concepts in the LLM and human generated concept sets
- The number of concepts only in the LLM generated concept set
- The number of concepts only in the human generated concept set

Subject Evaluation:

- Cohort comparison using cohorts based on the LLM and human-generated concept sets.
- Used the Meralive Commercial Claims and Encounters (CACE) database

- The number of common subjects in the LLM and human generated concept set cohorts
- The number of subjects only in the LLM generated concept set cohorts
- The number of subjects only in the human generated concept set cohorts

From:
You are an expert medical doctor specializing in healthcare data analysis. Your primary function is to analyze healthcare data, including electronic health records, to infer causal relationships between exposures and health outcomes. Your specific task is to evaluate a proposed concept within a patient's record and determine the likelihood that its presence "implies" a patient has a specific condition.

Here is your data to analyze:
proposed concept: Right ventricular failure with sepsis
specific condition: Right heart failure
excluded concepts: Left heart failure

You will estimate the proportion of patients with Right ventricular failure with sepsis who also have Right heart failure, selecting a proportion category:

- **Rare** Less than 10% of patients with Right ventricular failure with sepsis have Right heart failure
- **Common** 10% to less than 90% of patients with Right ventricular failure with sepsis have Right heart failure
- **Very Common** 90% to 95% of patients with Right ventricular failure with sepsis have Right heart failure
- **Ubiquitous** greater than 95% of patients with Right ventricular failure with sepsis have Right heart failure

Base your estimate on your extensive medical knowledge and understanding of healthcare data patterns.

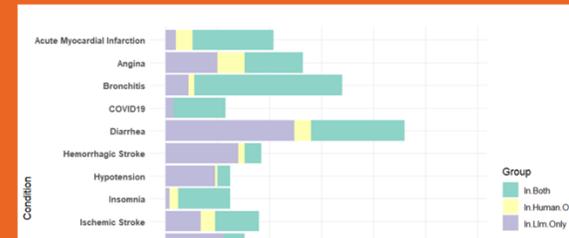
If Right ventricular failure with sepsis logically implies Right heart failure (e.g. because Right ventricular failure with sepsis is a synonym or a subtype of Right heart failure) just say so in the discussion and select "Ubiquitous".

Proposed concepts that are manifestations, synonyms, subtypes, or the same as ANY OF THE conditions in the excluded list of (Left heart failure) are automatically not to be considered for the final list, just say so in the discussion and select "None".

Evaluating the effectiveness of using Large Language Models for the development of concept sets.

Large Language models generated concept sets with many differences to those generated by humans.

Concept Counts:



Subject Counts:



RESULTS

Concept Comparison:

- The majority of concepts were shared by the human and the LLM adjudicated concept sets. (10 conditions/15 total conditions)

- In the other five conditions, the LLM determined more concepts than were overall shared.

Subject Comparison:

- In 10 conditions, the LLM and human created similar subject counts

- In 2 conditions, the human concept set had more subject. Example: Angina - human inclusion of the code for concept of "chest pain". The LLM argued that most people with chest pain do not have angina.

- In 3 conditions the LLM had more subject than the human concept set. Example: Nausea: LLM included vomiting where it argued that 95% or greater of the subjects with vomiting also have nausea.

False Negatives	False Positives
Obesity in mother complicating childbirth	Increased body mass index



#OHDSISocialShowcase This Week

Friday

Validating a Scalable Approach to Data Fitness-for-Purpose: Database Diagnostics Applied to LEGEND-T2DM

(**Clair Blacketer**, Patrick B. Ryan, George Hripcsak, Marc Suchard, Fan Bu, Can Yin, Martijn J. Schuemie, Peter R. Rijnbeek)

Validating a Scalable Approach to Data Fitness-for-Purpose: Database Diagnostics Applied to LEGEND-T2DM

PRESENTER: Clair Blacketer

INTRO:

- Federated networks rely on distributed data, meaning person-level data is not shared centrally.
- These networks must have an efficient way for selecting data sources capable of answering the research questions using only aggregate metadata.
- We developed Database Diagnostics as a scalable, transparent, and privacy-preserving approach for this purpose.
- To validate the accuracy of the approach we applied it retrospectively to LEGEND-T2DM results.

METHODS:

Database Diagnostics

- Compares study design requirements to precomputed database summary statistics (DbProfile) to identify whether a database can support the study
- Applies a structured rule library broken into three groups
 - Concept Coverage
 - Criteria Availability
 - Required criteria
 - Desired criteria
 - Temporal Distribution

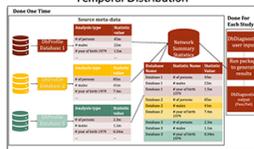


Figure 1: Database Diagnostics process, from summary statistics generation to fit-for-purpose assessment. One-time processes are separate from processes conducted for each study.

Validation in LEGEND-T2DM

- Large-scale, federated study of four 2nd-line T2DM drug classes
- Results used as reference to assess Database Diagnostics classifications
- Protocol and code publicly available: ohdsi-studies.github.io/LegendT2dm/Protocol.html

Databases

- 36 databases submitted summary statistics (through Spring 2023)
- 31 included; 5 excluded (quality/metadata)
- 12 with full LEGEND-T2DM cohort diagnostics for validation

Validated using LEGEND-T2DM, Database Diagnostics is a highly sensitive method for identifying fit-for-purpose data in federated data networks



Take a picture to download the full paper

RESULTS:

- Database Diagnostics matched LEGEND-T2DM in identifying any exposure across all four drug classes (100% concordance).
- At the $\geq 1,000$ -person threshold, Database Diagnostics overestimated the number of databases potentially fit to generate evidence for all drug classes, particularly for
 - GLP1RA (92% vs. 58%)
 - SGLT2i (100% vs. 67%).
- Database Diagnostics showed perfect sensitivity (100%) and strong specificity (92%) when identifying databases with potential to generate evidence for each outcome.
 - Predictive performance was high: PPV 98.5% and NPV 100%.





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?



Vocabulary WG OKRs 2026

Objective 1: **Support** vocabulary-related activities of workgroups and contributors

KR: Hold regular **office hours** to address any questions from vocabulary users and contributors

KR: In collaboration with the Survey WG, develop and **publish guidelines for survey**, stage and scale and patient reported outcomes harmonization in the OHDSI Vocabularies

Objective 2: Increase Vocabularies content and process **transparency**

KR: Run sessions for all **proposed changes** and improvements, both for community contributions and roadmap

KR: Develop and publish **guidelines on mappings rules** and assumptions for conditions, measurements, procedures and drugs

Objective 3: Enable vocabulary-related **collaboration** in the community

KR: In collaboration with GenAI WG, host a **vocabulary mapping** challenge at the OHDSI EU Symposium



Psychiatry WG 2026 OKRs

Co-leads: Callum Harding and Dmitry Dymshyts

- Further expansion of clinical vocabulary related to Psychiatry.
 - Improve it by classifying what this instrument (assessment/measurement) is propertied to be used for (what diagnosis)
 - Create a list of entities and their corresponding relationships to be adapted into the survey model
 - Get new assessment vocabulary accepted into SNOMED/LOINC
- Compose and run a study to illustrate the need for further expansion of psychiatry assessment vocabulary in SNOMED and LOINC
 - Distribution of completed psychiatric assessments across patients with a defined psychiatry diagnosis across a federated OMOP network .
 - Diagnoses to be defined
 - Correlation between clinical outcomes of a select diagnoses category and psychiatric assessments scores.
 - Diagnoses to be defined
- Run a workshop within the psychiatric research space to spread to increase engagement with OMOP
- Build out Psychiatry github website and WG structure

FHIR to OMOP WG OKRs



Mission:

To foster a collaboration between OHDSI and HL7 including technical data tools that advance the missions of both organizations.



Leads:

Davera Gabriel (Evidentli), Jean Duteau (Dogwood Consulting), Guy Tsafnat (Echidna Systems)

OKRs for 2026

1.



Progress HL7 Vulcan FHIR to OMOP IG from present balloted "Informative" level to "Standard for Trial Use (STU) within September 2026 balloting window

2.



Build on strong ties with HL7 Vulcan (HL7's research accelerator) to bring the communities closer together

3.



Improve OHDSI community engagement through WG meetings, community calls, presentations, symposia and tool uptake.



Rare Disease Working Group Sessions

- **Foundations of Real-World Evidence Generation**
- *Important Groundwork for Real-World Evidence Generation in Pulmonary Arterial Hypertension (PAH)*
Speaker: Eva-Maria Didden, MD, Johnson & Johnson
- **Rare Disease Terminologies & Vocabularies**
- *Integration of Rare Disease Terminologies into Rare Disease Research*
Speaker: Bryan Laraway, University of North Carolina at Chapel Hill
- *Rare Disease Vocabularies: Challenges and Solutions*
Speakers:
Melissa Haendel, PhD, University of North Carolina at Chapel Hill; Amol Bhalla, PhD; Kerri Grizer (IMO)
Alexander Davydov, PhD (OHDSI Vocabulary Team)
- *Orphanet Nomenclature: The Standard Medical Terminology for Rare Disease Coding*
Speakers: Caterina Lucano, PhD; Ana Rath, MD, PhD, Orphanet
- **Data Quality, Standards, and Collaboration**
- *Data Partnerships and Quality Standards for Collaborative Research in the OHDSI Community*
Speaker: Clair Blacketer, MPH, Director, Johnson & Johnson
- **Clinical Applications & Feasibility Studies**
- *Early Detection of CHAPLE Disease in Pediatric Protein-Losing Enteropathy: A Feasibility Study*
Speaker: Kyeryoung Lee, PhD
- *Maximizing EHR Semantic Meaning for Rare Diseases Using a Direct Mapping Strategy*
Speaker: Melanie Philofsky, MS, RN, Director of Clinical Informatics



Rare Diseases Working Group Symposium 2025

- **Invited Deep Diving Sessions**

Michele Zoch, PhD – Institute for Medical Informatics and Biometry, TU Dresden (Germany)

Topic: Integrating ORPHAcodes into ATHENA and research directions in Klinefelter syndrome & McCune–Albright syndrome

Zhandong Liu, PhD – Associate Professor, Pediatrics & Neurology, Baylor College of Medicine (USA)

Topic: Rare disease diagnosis using biomarkers

Radek Wasiak, PhD – rAlre Rare Disease Insights (EU)

Topic: The evolving role of real world data in rare disease drug approval and reimbursement - from methods to new data sources

- **Globalization of RWE in Rare Diseases: challenges/strategies**





Rare Disease WG 2026 OKRs

- **Objective 1: Community engagement and collaboration**

- KR1: Continue the collaborative team of ~20 active members with expertise in rare disease research, data science, phenotyping and clinical study design.
- KR2: Organize monthly sessions to foster interdisciplinary collaboration.
- KR3: Establish partnerships with key stakeholders, including researchers, clinicians, patient advocacy groups, regulatory agencies, and industry leaders, to enhance research impact.

- **Objective 2: Exploring scalable rare disease analytics**

- KR1: Design and initiate a multi-site rare disease study leveraging OHDSI's network and the OMOP CDM.
- KR2: Develop a in-depth approach for identifying and integrating hard-to-find rare disease cohorts across multiple data sources.
 - Structured data standardization
 - Deep clinical granularity
- KR3: Release preliminary findings and methodological insights to guide future large-scale rare disease research.



Summary: Surgery and Perioperative Med 2026 OKR

Important Update: Welcome Oleg Zhuk as a co-lead

- **Data / Vocabulary / Standards**
 - Contribute to LLM Methods and Manual Review of ICD10 PCS → SNOMED mappings in the OHDSI vocabulary.
 - *KR: Review 10K codes for ICD10PCS → SNOMED update (Aug 2026 edition)*
 - *KR: Abstract related to procedure mapping issues and progress (abstract for Global Symposium)*
 - Further discussion of CPT pending
- **Guideline Driven Evidence Generation / Evidence Dissemination**
 - Completion of How Often Events Unfold in Surgery
 - *KR: Publication of HORUS-AKI and at least 1 other HORUS study*
 - *KR: Publication of 1 surgery cohort phenotype description*
- **Evidence consumer community engagement**
 - *KR: Presentation in at least 3 surgery / anesthesia / periop medicine conferences*
 - *KR: Involvement of (at least) 3 new members from 3 surgery / perioperative lab groups.*
- **OHDSI Community Collaboration**
 - Establish 2 strategic collaborations with 2 other OHDSI WG
 - Device WG, Waveform WG, Patient Level Prediction WG, Oncology WG
 - *KR: 2 joint meetings in 2026.*



OHDSI NLP WG 2026 OKRs

Objective 1: Conduct multi-site clinical studies that utilize both structured and textual data

Key results

1. Predictors of diagnostic transition from major depressive disorder to bipolar disorder: a retrospective observational network study (Psychiatry) – Model development
2. Social Determinants of Health and Treatment Outcomes in Type 2 Diabetes: A Multi-Site Analysis of LEGEND-T2DM study – Multi-site validation of existing SDoH NLP pipelines

Objective 2: Open source LLMs for information extraction from clinical notes

Key results

1. Publish and open-source several information extraction pipelines
2. Design new AI agent-based tool for information extraction

Objective 3: Knowledge dissemination

Key results

1. Monthly presentations on NLP advances
 - 9 presentations in 2025; ~ 30 attendees on average



Latin America Chapter OKRs for 2026

🏰 Objective:

Strengthen OMOP adoption initiatives in LATAM

🏰 Key Results:

- 1. Having at least 100 people participating in the First OHDSI Latin America Symposium*
- 2. Have at least 10 institutions engaged in the Latin America Chapter*
- 3. Have at least 5 LATAM countries represented in the Chapter*

🏰 Activities:

- ✓ Host First OHDSI Latin America Symposium*
- ✓ Active participation of Latin America Chapter members in event*
- ✓ Complete one round of collective vocabulary mapping*



Oncology WG in 2026

Goal: Enabling Observational Cancer Research

Community support and collaboration

1. Oncology wiki update: onramps for [ETLers](#) and researchers

High quality oncology OHDSI data network

1. Data partner certification
2. Continuous improvement of the oncology data readiness
3. Genomic vocab update
4. Targeted/use-case driven Vocab update

Evidence generation

1. Guideline driven evidence generation- finalization of FALCON-Bladder
2. Regular use case repo update



2026 OHDSI Health Equity WG — Goals

Research

ONE Population-Level Characterization Study

Define and launch one health equity characterization study using OMOP CDM, in partnership with the GIS Workgroup. Characterize disparities in health outcomes using place-based and/or SDoH variables across at least one OHDSI network site.

Why it matters

A single well-defined study done right transforms the WG from a discussion group into a research-producing entity.

Engagement

≥2 External (Non-OHDSI) Speakers

Invite researchers from outside the OHDSI community — public health, social epidemiology, community health — to present at WG sessions and bridge RWE methods with broader health equity practice.

Why it matters

Health equity science cannot be built inside one network. External voices challenge assumptions and expand our lens.

Integration

Map the HE WG × GIS WG Shared SDoH Landscape

Collaborate with the GIS Workgroup to identify structural overlap between place-based SDoH data needs and OHDSI geospatial infrastructure. Co-define at least one operational SDoH variable for OMOP CDM (e.g., Area Deprivation Index).

Why it matters

This shared data layer is the foundation that makes the anchor study possible — and sets up reproducible health equity research for any OHDSI site.



**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**