

OHDSI内では、実名での活動になります。  
Zoom参加時も「名前は実氏名で」お願いします。



**OHDSI**  
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

オデッセイ  
ジャパン

OHDSI Japan Meeting #77

2026年4月 イブニング・カンファレンス

2026.4.27

---



## 本日の内容

- OHDSI 論文の紹介
- OHDSI global/APAC から
- 話題 MEDIS標準病名マスタにおけるICD-10経由のボキャブラリマッピング手法と実装（井上様／株式会社 Yuimedi）

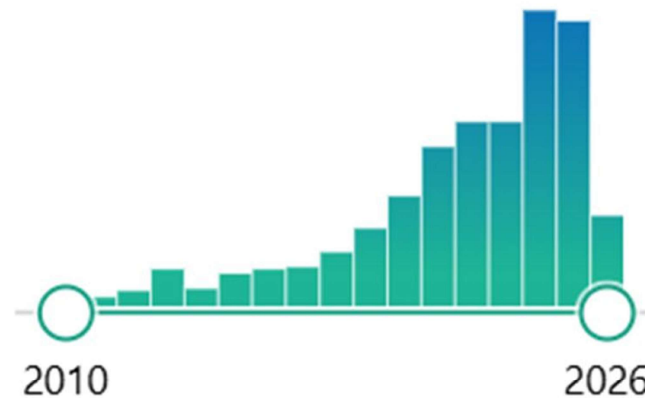


# OHDSI 論文の紹介



## OHDSI関連論文

- Pubmedで”OHDSI or OMOP”を検索
- 



pubmed.ncbi.nlm.nih.govにて作成

- 全期間累計: 638報(2026年2月19日) →666報(2026年4月15日)
- 2024年は約126報



### (積み残し論文)

1. Lee YA, Lu Y, Bian J, Guo JS, He X. Harmonizing Medicare Claims Data with OMOP: A Validated ETL Pipeline. AMIA Annu Symp Proc. 2024;2024:715–23. PubMed PMID: 41726451; PubMed Central PMCID: PMC12919608.
2. Hunt NB, Souverein P, Bazelier M, Barclay N, Delmestri A, Sturkenboom M, et al. Implementation of OMOP and ConcePTION Common Data Models in CPRD GOLD: Risk of Bleeding and Cardiovascular Outcomes From Anticoagulant Use. Clin Pharmacol Ther. 2026 Mar. doi:[10.1002/cpt.70242](https://doi.org/10.1002/cpt.70242) PubMed PMID: 41793101.
3. Kim B, Song W, Yoon E, Kim S, Lee HY, Kim JH, et al. Transforming unstructured breast cancer pathology reports into the Observational Medical Outcomes Partnership Common Data Model. BMC Med Inform Decis Mak. 2026 Feb. doi:[10.1186/s12911-026-03375-7](https://doi.org/10.1186/s12911-026-03375-7) PubMed PMID: 41764468.
4. Jung H, Yoo S, Kim S, Chung J, Lee HY. Transforming nursing documentation data into the Observational Medical Outcomes Partners common data model. Int J Med Inform. 2026 Feb;213:106358. doi:[10.1016/j.ijmedinf.2026.106358](https://doi.org/10.1016/j.ijmedinf.2026.106358) PubMed PMID: 41734431.
5. El Mekkaoui K, Mukherjee C, Santana C, Silva H, Quazi A, Balabhadruni M, et al. The impact of the Social Exposome in Cardiovascular Health and Disease. J Precis Med (Amst). 2025 Oct;3:100015. doi:[10.1016/j.premed.2025.100015](https://doi.org/10.1016/j.premed.2025.100015) PubMed PMID: 41847122; PubMed Central PMCID: PMC12990328.
6. Chen HY, Ostropolets A, Weng C, Hripcsak G. Knowledge Engineering for Medical Vocabularies Using Large Language Models. AMIA Annu Symp Proc. 2024;2024:248–56. PubMed PMID: 41726449; PubMed Central PMCID: PMC12919453.
7. Tamana S, Yiangou K, Orphanou K, Chatzimatthaiou S, Kountouris P, Cremonesi F. FAIR data gaps and collaboration willingness among hemoglobinopathy research centers. Sci Data. 2026 Mar. doi:[10.1038/s41597-026-06950-9](https://doi.org/10.1038/s41597-026-06950-9) PubMed PMID: 41775724.
8. Cooper LN, Vadsariya A, Varghese M, Nayee B, Moon J, Katterapalli C, et al. LEVERAGING EPIC'S NATIVE ETL INFRASTRUCTURE FOR OMOP CDM

IMPLEMENTATION: A COLLABORATIVE EXPERIENCE. AMIA Annu Symp Proc. 2024;2024:287–92. PubMed PMID: 41726461; PubMed Central PMCID: PMC12919598.

9. Newbury A, Jiang X, Natarajan K, Gürsoy G. Cross Biobank Comparison of Phenomic Profiles. AMIA Annu Symp Proc. 2024;2024:939–48. PubMed PMID: 41726399; PubMed Central PMCID: PMC12919452.



(新出論文)

10. Lin JB, Mataraso SJ, Chadha M, Velez G, Mruthyunjaya P, Aghaeepour N, et al. Multimodal AI fuses proteomic and EHR data for rational prioritization of protein biomarkers in diabetic retinopathy. medRxiv : the preprint server for health sciences. United States; 2026. p. 2026.02.23.26346903.

doi:[10.64898/2026.02.23.26346903](https://doi.org/10.64898/2026.02.23.26346903) PubMed PMID: 41918763; PubMed Central PMCID: PMC13034698.

11. Kim JH, Seo Y, Shin SY, Kim EJ, Han KS, Joo HJ. Temporal Trends and Clinical Implications of Cardiac Troponin Testing in Emergency Departments: A Multicenter Retrospective Study. J Clin Med. 2026 Mar;15(6).

doi:[10.3390/jcm15062426](https://doi.org/10.3390/jcm15062426) PubMed PMID: 41899348; PubMed Central PMCID: PMC13026465.

12. Muylle KM, Decoster L, Verleyen D, Germonpre PR, Masuy I, Verheyden B, et al. Real-world demographics, clinical characteristics, and testing and treatment patterns in patients with non-small-cell lung cancer and extensive-stage small-cell lung cancer in Belgium: the AIBED study. ESMO Real World Data Digit Oncol. 2026 Mar;11:100668.

doi:[10.1016/j.esmorw.2025.100668](https://doi.org/10.1016/j.esmorw.2025.100668) PubMed PMID: 41930294; PubMed Central PMCID: PMC13040885.

13. Bhattacharjee T, Mugotitsa B, Ochola M, Momanyi R, Andeso P, Amadi D, et al. Migrating longitudinal African mental health data from staging to the OMOP common data model within the INSPIRE network datahub. Front Psychiatry. 2026;17:1751529. doi:[10.3389/fpsyt.2026.1751529](https://doi.org/10.3389/fpsyt.2026.1751529) PubMed PMID: 41877886; PubMed Central PMCID: PMC13006644.

14. Du M, Prats-Urbe A, Mercadé-Besora N, Lopez-Guell K, Guo Y, Alcalde-Herraiz M, et al. CohortCharacteristics: an R package for population characterisation in observational studies using the OMOP common data model. Eur J Epidemiol. 2026 Apr. doi:[10.1007/s10654-025-01352-4](https://doi.org/10.1007/s10654-025-01352-4) PubMed PMID: 41933108.

15. Song S, Seo H, Kim IS, Kim M, Hee LY, Kim JE, et al. A Multicenter Propensity Score-Matched Cohort Study of Preoperative Antiplatelet Therapy and Postoperative Outcomes in Elderly Surgical Patients. Medicina

(Kaunas). 2026 Mar;62(3). doi:[10.3390/medicina62030521](https://doi.org/10.3390/medicina62030521) PubMed PMID: 41901602; PubMed Central PMCID: PMC13027806.

16. Lambert A, Castagne C, Pau D, Chmiel J, Labarga A, Boernert E, et al. Evaluating the statistical utility and information loss in the transformation of a real-world oncology database from CDISC-SDTM to OMOP-CDM. ESMO Real World Data Digit Oncol. 2026 Mar;11:100655.

doi:[10.1016/j.esmorw.2025.100655](https://doi.org/10.1016/j.esmorw.2025.100655) PubMed PMID: 41930302; PubMed Central PMCID: PMC13040891.

17. Marfoggia A, Arcobelli VA, Moscato S, La Mattina AA, Mellone S, Carbonaro A. Challenges of health data standard adoption and usage: a systematic review. J Biomed Inform. 2026 Mar;178:105022.

doi:[10.1016/j.jbi.2026.105022](https://doi.org/10.1016/j.jbi.2026.105022) PubMed PMID: 41871756.



# MedicareクレームデータとOMOPの統合:検証済みETLパイプライン

Harmonizing Medicare Claims Data with OMOP: A Validated ETL Pipeline. AMIA Annu Symp Proc. 2024;2024:715-23. PubMed PMID: 41726451; PubMed Central PMCID: PMC12919608.

[Yao An Lee](#)<sup>1</sup>, [Ying Lu](#)<sup>1</sup>, [Jiang Bian](#)<sup>2,3</sup>, [Jingchuan Serena Guo](#)<sup>1</sup>, [Xing He](#)<sup>2,3</sup>

1. Department of Pharmaceutical Outcomes & Policy, College of Pharmacy, University of Florida, Gainesville, Florida, USA
2. Department of Biostatistics and Health Data Science, School of Medicine, Indiana University, Indianapolis, Indiana, USA
3. Center for Biomedical Informatics, Regenstrief Institute, Indianapolis, Indiana, USA



## MedicareクレームデータとOMOPの統合:検証済みETLパイプライン

- 研究背景
  - リアルワールドデータ(RWD)の活用拡大
  - EHRとレセプトデータの統合の必要性
  - 異なるデータモデル・用語体系の課題
- データソースの特徴
  - Medicare LDS (Limited Data Set): 医療利用・費用の網羅的データ
  - EHR: 臨床詳細(検査値・診療記録)
  - 両者は相補的關係
- 研究目的
  - LDSをOMOP CDM v5.3へ標準化
  - 相互運用性の向上
  - 多施設共同研究への応用
- OMOP CDMの概要
  - OHDSIが開発した標準データモデル
  - 主要テーブル: Person, Condition, Drug等
  - 標準語彙(SNOMED, RxNorm, LOINC)を使用



## MedicareクレームデータとOMOPの統合:検証済みETLパイプライン

- ETL設計
  - PythonベースのETLパイプライン
  - WhiteRabbit:データプロファイリング
  - Rabbit-in-a-Hat:マッピング設計
- 変換ロジック
  - Direct mapping(値の直接変換)
  - Rule-based transformation(形式変換)
  - Concept mapping(標準概念へのマッピング)
- データマッピング
  - Medicare LDS → OMOP 15テーブル
  - 診断:ICD → SNOMED
  - 薬剤:NDC → RxNorm
- データ品質評価
  - OMOP Data Quality Dashboard使用
  - 適合性・完全性・妥当性を評価
  - 約99%のチェックに合格



## MedicareクレームデータとOMOPの統合:検証済みETLパイプライン

- 結果(臨床的整合性)
  - 患者数、性別、人種分布一致
  - 主要疾患(糖尿病、心不全等)一致
  - 高いデータ忠実性
- 課題と限界
  - 未マッピングコードの存在
  - 構造的に変換不可能なデータ
  - 臨床詳細(検査値等)の不足
- 結論
  - 高品質なETL基盤を確立
  - 医療データの標準化に貢献
  - RWD研究の基盤として有用



## CPRD GOLDにおけるOMOPおよびConcePTION共通データモデルの実装: 抗凝固薬使用による出血および心血管アウトカムリスク

Implementation of OMOP and ConcePTION Common Data Models in CPRD GOLD: Risk of Bleeding and Cardiovascular Outcomes From Anticoagulant Use. Clin Pharmacol Ther. 2026 Mar. doi:[10.1002/cpt.70242](https://doi.org/10.1002/cpt.70242) PubMed PMID: 41793101.

[Nicholas B. Hunt](#)<sup>1</sup>, [Patrick Souverein](#)<sup>1</sup>, [Marloes Bazelier](#)<sup>1</sup>, [Nicola Barclay](#)<sup>2</sup>, [Antonella Delmestri](#)<sup>2</sup>, [Miriam Sturkenboom](#)<sup>3</sup>, [Daniel Prieto-Alhambra](#)<sup>2</sup>, [Helga Gardarsdottir](#)<sup>1,4,5</sup>, [Olaf Klungel](#)<sup>1</sup>

1. Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Utrecht, The Netherlands;
2. Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK;
3. Department of Data Science & Biostatistics, Julius Global Health, University Medical Center Utrecht, Utrecht, The Netherlands;
4. Department of Clinical Pharmacy, University Medical Centre Utrecht, Utrecht, The Netherlands;
5. Department of Pharmaceutical Sciences, School of Health Sciences, University of Iceland, Reykjavik, Iceland.



## CPRD GOLDにおけるOMOPおよびConcePTION共通データモデルの実装:抗凝固薬使用による出血および心血管アウトカムリスク

### • 背景

- 複数の医療データベースを統合するために共通データモデル(CDM)が使用される。
- しかし、異なるCDMの選択が研究結果に影響する可能性がある。

### • 目的

- 同一データ(CPRD GOLD)をOMOPとCDMConcePTION CDMに変換し、結果の違いを比較

### • 方法

- 対象:非弁膜症性心房細動患者
- 比較:
  - DOAC(直接経口抗凝固薬)
  - VKA(ビタミンK拮抗薬)
- 評価項目:
  - 大出血
  - 脳卒中
  - 心血管疾患(CVD)
  - 股関節骨折(ネガティブコントロール)

- ConcePTION CDM: 構文的ハーモナイゼーション(syntactic harmonization)のみが求められる
- OMOP CDM: 構文的ハーモナイゼーションに加えて、意味的ハーモナイゼーション(semantic harmonization:元のコード体系を標準化語彙へマッピングすること)の両方が必要とされる



## CPRD GOLDにおけるOMOPおよびConcePTION共通データモデルの実装:抗凝固薬使用による出血および心血管アウトカムのリスク

### • 主な結果

- 心血管疾患(CVD)
  - OMOP: リスク低下(HR 0.82)
  - ConcePTION: 差なし(HR 0.99)
- 脳卒中
  - 両CDMでほぼ同等(わずかに増加)
- 大出血
  - 両者とも有意差なし

CDMにより結果が変わることが確認された。

### • なぜ差が出るのか

- 主な原因: コホート定義の違い
- 疾患定義(フェノタイピング)の違い
- 薬剤曝露期間の推定方法の違い
- データマッピング時の補完処理(imputation)

### • 結論

- CDMの違いは研究結果に実質的な影響を与える
- 特に以下に注意が必要:
  - データマッピング方法
  - 分析設計
  - 変数定義

CDMは単なる形式ではなく、結果を左右する重要要因



## 非構造化乳がん病理レポートをOMOP共通データモデルへの変換

Transforming unstructured breast cancer pathology reports into the Observational Medical Outcomes Partnership Common Data Model. BMC Med Inform Decis Mak. 2026 Feb. doi:[10.1186/s12911-026-03375-7](https://doi.org/10.1186/s12911-026-03375-7) PubMed PMID: 41764468.

[Borham Kim](#)<sup>1</sup>, [Wong Eun Song](#)<sup>1,7</sup>, [Eunsil Yoon](#), [Seok Kim](#)<sup>1</sup>, [Ho-Young Lee](#)<sup>1,2</sup>, [Jee Hyun Kim](#)<sup>3</sup>, [Koung Jin Suh](#)<sup>3</sup>, [Kwang-Il Kim](#)<sup>3</sup>, [So Yeon Park](#)<sup>4</sup>, [Eun-Kyu Kim](#)<sup>5</sup>, [Se Hyun Kim](#)<sup>3</sup>, [Seonghae Yoon](#)<sup>6</sup> & [Sooyoung Yoo](#)<sup>1</sup>

1. Office of eHealth Research and Business, Seoul National University Bundang Hospital, Seongnam, Republic of Korea
2. Department of Nuclear Medicine, Seoul National University College of Medicine, Seoul, Republic of Korea
3. Department of Internal Medicine, Seoul National University Bundang Hospital, Seoul National University College of Medicine, Seongnam, Republic of Korea
4. Department of Pathology, Seoul National University Bundang Hospital, Seoul National University College of Medicine, Seongnam, Republic of Korea
5. Department of Surgery, Seoul National University College of Medicine, Breast Care Center, Seoul National University Bundang Hospital, Seongnam, Republic of Korea
6. Department of Clinical Pharmacology and Therapeutics, Seoul National University Bundang Hospital, Seoul National University College of Medicine, Seongnam, Republic of Korea
7. Department of Health Science and Technology, Graduate School of Convergence Science and Technology, Seoul National University, Seoul, Republic of Korea



## 非構造化乳がん病理レポートをOMOP共通データモデルへ変換

### • 背景

- 乳がんの病理レポートは治療方針決定に重要な情報を含む。
- しかし、多くが自由記述(非構造化)であり、データ活用が困難
- NLP(自然言語処理)による情報抽出が必要

### • 目的

- 病理レポートから抽出したデータをOMOP-CDMに標準化
- 臨床研究への活用可能性を検証

### • 方法

- 対象:乳がん患者10,730人、31,135レポート
- 手法:
  - ルールベースNLP
  - テキスト前処理・分割・エンティティ抽出
  - OMOP標準語彙へマッピング

### • 結果

- 抽出精度:F1スコア  $\approx$  0.998(非常に高精度)
- OMOP-CDMへ統合(複数テーブルに格納)
- データ品質:96%



## 非構造化乳がん病理レポートをOMOP共通データモデルへ変換

- 有用性検証(予測モデル)
  - 5年死亡予測モデルを構築
  - AUROC = 0.840(高い予測性能)
  - 病理データが予測に大きく寄与
- 意義
  - 病理 + 臨床データの統合分析が可能
  - 多施設共同研究
  - リアルワールドデータ活用に貢献
- 限界
  - ルールベースNLPの柔軟性制限
  - 英語中心(多言語対応不足)
  - 施設間フォーマット差への依存
- 結論
  - NLP + OMOP CDMにより病理データの再利用が可能
  - がん研究の基盤強化に寄与



## 看護記録データをOMOP共通データモデルへ変換する研究

Transforming nursing documentation data into the Observational Medical Outcomes Partners common data model. Int J Med Inform. 2026 Feb;213:106358.  
doi:[10.1016/j.ijmedinf.2026.106358](https://doi.org/10.1016/j.ijmedinf.2026.106358) PubMed PMID: 41734431.

Hyesil Jung <sup>a</sup> , Sooyoung Yoo <sup>b</sup> , Seok Kim <sup>b</sup> , Jeehae Chung <sup>c</sup> , Ho-Young Lee <sup>b</sup>

- a. School of Nursing, Inha University, South Korea
- b. Office of eHealth Research and Business, Healthcare Innovation Park, Seoul National University Bundang Hospital, South Korea
- c. College of Nursing, The Catholic University of Korea, South Korea



## 看護記録データをOMOP共通データモデルへ変換する研究

### • 背景

- 電子カルテ(EHR)は研究に有用だが、看護記録は活用が進んでいない
- 理由:
  - 自由記述(非構造データ)
  - データ品質のばらつき

### • 目的

- 看護記録を OMOP CDM形式へ標準化
- 変換標準化データの有用性をコホート生成で検証

### • 方法

- 約400万件の記録から4006種類の看護記述を抽出
- SNOMED CTで標準化
- OMOP CDMに格納「悪心(nausea)」
- 患者コホートを作成し比較

### • 結果


- ① データ標準化
  - 98.9%がSNOMED CTに対応
  - 約2億件の看護記録を変換
- ② コホート比較
  - 看護記録ベース:214,830件
  - 診断ベース:12,381件

約10倍以上の患者を抽出可能



## 看護記録データをOMOP共通データモデルへ変換する研究

- 考察
  - 医師の診断は主要疾患中心
  - 看護記録は症状(例:悪心)を継続的に記録
    - より網羅的な患者把握が可能
- 意義
  - 看護データの標準化により多施設研究が可能
  - 患者状態の詳細分析が可能
- 限界
  - 自由記述データ未対応
  - 単一施設データ
- 結論
  - 看護記録のCDM変換は有効
  - 今後は多施設展開が必要




## 心血管健康と疾患における社会エクスポソームの影響

The impact of the Social Exposome in Cardiovascular Health and Disease. J Precis Med (Amst). 2025 Oct;3:100015. doi:[10.1016/j.premed.2025.100015](https://doi.org/10.1016/j.premed.2025.100015) PubMed PMID: 41847122; PubMed Central PMCID: PMC12990328.


[Khaoula El Mekkaoui](#)<sup>a,b,c</sup> · [Chitra Mukherjee](#)<sup>a</sup> · [Clodomir Santana](#)<sup>a</sup> · [Hitalo Silva](#)<sup>a</sup> · [Arnib Quazi](#)<sup>a</sup> · [Meghana Balabhadruni](#)<sup>a</sup> · [Aditya Ballal](#)<sup>d</sup> · [Shamika V. Gavaskar](#)<sup>a</sup> · [Kea T. Turqueza](#)<sup>a</sup> · [Leslie Molina](#)<sup>a</sup> · [Pablo E. Acevedo](#)<sup>a</sup> · [Kay Moua](#)<sup>a</sup> · [Omar Dawar](#)<sup>a</sup> · [Federico Garcia](#)<sup>a</sup> · [Mohammad S. Soroya](#)<sup>a</sup> · [Xiao-Dong Zhang](#)<sup>a</sup> · [Ezra A. Amsterdam](#)<sup>a</sup> · [Chao-Yin Chen](#)<sup>d</sup> · [Padmini Sirish](#)<sup>a</sup> · [Leighton T. Izu](#)<sup>d</sup> · [Ronaldo Menezes](#)<sup>e</sup> · [Nipavan Chiamvimonvat](#)<sup>a,f</sup> · [Vladimir Filkov](#)<sup>g</sup> · [Martin Cadeiras](#)<sup>a</sup> · [David A. Liem](#)<sup>a</sup>

- a. Department of Medicine, Division of Cardiovascular Disease, University of California, Davis, USA
- b. VTT Technical Research Centre of Finland, Espoo, Finland
- c. Department of Computer Science, Aalto University, Espoo, Finland
- d. Department of Pharmacology, School of Medicine, University of California, Davis, USA
- e. Department of Computer Science, University of Exeter, Exeter, UK
- f. Department of Basic Medical Sciences and Translational Cardiovascular Research Center, University of Arizona, College of Medicine, Phoenix, USA
- g. Department of Computer Science, University of California, Davis, USA



## 心血管健康と疾患における社会エクスポソームの影響

- 研究の背景
  - 心血管疾患(CVD)は世界的な主要死因
  - 遺伝要因だけではリスクの半分しか説明できない
  - 環境・社会・生活習慣の影響が極めて重要
- エクスポソーム(Exposome)とは
  - 人が生涯で受けるすべての環境曝露の総体
  - 以下の2つで構成:
    - 外部:環境・社会・生活
    - 内部:炎症・代謝・ストレス反応遺伝(ゲノム)を補完する概念
- 社会エクスポソーム(重要ポイント)
  - 社会的健康決定要因(SDoH):
    - 医療アクセス
    - 教育
    - 経済状況
    - 社会的つながり
    - 住環境
  - 健康格差の本質的原因



## 心血管健康と疾患における社会エクスポソームの影響

- データ統合の新アプローチ
  - 電子カルテ(EHR) + 公的データ(国勢調査など)
  - GIS(地理情報)活用
  - OMOPデータモデルで統合
  - 個別化リスク予測が可能
- 提案フレームワーク(論文の核心)
  - データ標準化(EHR + 環境データ)AI・機械学習
  - 解析フェデレーテッドラーニング(プライバシー保護)
  - 精密医療(Precision Medicine)の実現
- 課題
  - 曝露データの複雑性・多次元性
  - データ標準化の不足
  - プライバシー問題
  - 個人レベル測定の難しさ
- 結論
  - 心血管疾患は「社会×環境×生物」の統合問題
  - 社会エクスポソームの理解が鍵
  - 個別化医療・予防戦略の高度化に貢献



## 大規模言語モデルを用いた医療用語の知識工学

Knowledge Engineering for Medical Vocabularies Using Large Language Models. AMIA Annu Symp Proc. 2024;2024:248–56. PubMed PMID: 41726449; PubMed Central PMCID: PMC12919453.

[Hsin Yi Chen](#)<sup>1</sup>, [Anna Ostropolets](#)<sup>1,2</sup>, [Chunhua Weng](#)<sup>1</sup>, [George Hripcsak](#)<sup>1</sup>

1. Department of Biomedical Informatics, Columbia University, New York, NY, USA
2. Observational Health Data Analytics, Johnson & Johnson, Titusville, New Jersey, USA

### • 背景

- 医療用語体系(SNOMED CT、ICDなど)は医療データ分析に不可欠
- しかし、作成・維持には、専門家の長期関与、高コストが必要

### • 研究目的

- LLM(GPT-4o)を用いて医療用語管理の自動化可能性を検証
- 特に以下の3タスクに注目:
  - 類似語判定(同義語)
  - 包含関係(上位・下位)
  - グルーピング



## 大規模言語モデルを用いた医療用語の知識工学

### • 方法

- データ: 心血管系1,533用語 (SNOMED CT)
- 手法:
  - テキスト埋め込み(embedding)
  - コサイン類似度
  - ChatGPTによる補正
- 評価: OHDSI標準語彙を基準

### • 主な結果

- ① 類似語判定
  - Precision: 0.78
  - Recall: 0.41
    - 精度は高いが見落としが多い
- ② 包含関係(階層構造)
  - Precision: 0.74
  - Recall: 0.08
    - 階層の網羅性が極めて低い
- ③ グループング
  - Precision: 0.78
  - Recall: 0.52
    - 中程度の性能



## 大規模言語モデルを用いた医療用語の知識工学

- 考察

- LLMの強み

- 高い精度(誤分類が少ない)
    - 医療概念理解が可能

- 課題

- Recallが低い(網羅性不足)
    - ハルシネーション
    - 長文入力で性能低下(“lost in the middle”現象)

- 結論

- LLMは医療用語管理に有望

- ただし実用には以下が必要:

- Recall改善
    - 精度向上
    - スケーラビリティ検証



## FAIRデータギャップとヘモグロビン病研究センター間の協力意欲

FAIR data gaps and collaboration willingness among hemoglobinopathy research centers. Sci Data. 2026 Mar. doi:[10.1038/s41597-026-06950-9](https://doi.org/10.1038/s41597-026-06950-9) PubMed PMID: 41775724.

[Stella Tamana](#)<sup>1</sup>, [Kristia Yiangou](#)<sup>2</sup>, [Kalia Orphanou](#)<sup>1</sup>, [Sotiroula Chatzimatthaiou](#)<sup>1</sup>, [Petros Kountouris](#)<sup>1</sup>  
& [Francesco Cremonesi](#)<sup>3</sup>

1. Department of Blood Disorder Genetics and Thalassemia, The Cyprus Institute of Neurology and Genetics, Nicosia, Cyprus
2. Biostatistics Unit, The Cyprus Institute of Neurology and Genetics, Nicosia, Cyprus
3. Centre Inria de l' Université Côte d' Azur, Sophia Antipolis, France



## FAIRデータギャップとヘモグロビン病研究センター間の協力意欲

- 研究の背景

- サラセミアや鎌状赤血球症などのヘモグロビン異常症は世界的に重要な遺伝性疾患
- 多施設研究にはデータの標準化・共有が不可欠
- しかし、現状はFAIR原則(Findable, Accessible, Interoperable, Reusable)に十分準拠していない

- 研究目的

- 各研究施設のデータ管理状況の把握
- メタデータ・標準規格の使用状況の評価
- 共同研究(特にデータ共有)への意欲の調査



## FAIRデータギャップとヘモグロビン病研究センター間の協力意欲

### • 方法

- HELIOSネットワークでアンケート調査
- 対象:22カ国・44施設
- 評価項目:
  - データ種類
  - メタデータ
  - 標準化
  - データ共有意欲

### • 主な結果

- ① データの種類
  - 基本データ(人口情報・検査・遺伝子)は豊富
  - 画像・オミクスなど高度データは不足
- ② FAIR対応状況
  - メタデータ整備:50%
  - オントロジー使用:20%
  - OMOP/CDISCなど標準モデル:未使用
- ③ データ共有意欲
  - 多施設研究参加意欲:95%
  - 分散型(Federated):86%
  - 集中型:68%



## FAIRデータギャップとヘモグロビン病研究センター間の協力意欲

- 課題
  - データが分散・非標準化
  - 相互運用性が低い
  - 匿名化・ガバナンス不足
- 結論
  - データ標準化・基盤整備が急務
  - 国際的な協調とFAIR対応の推進が必要
- 考察
  - FAIR準拠は不十分だが、共同研究の意欲は非常に高い
  - 分散型データ共有が現実的な解決策



## EpicのネイティブETL基盤を活用したOMOP CDM実装： 協働的な取り組み

LEVERAGING EPIC'S NATIVE ETL INFRASTRUCTURE FOR OMOP CDM IMPLEMENTATION: A COLLABORATIVE EXPERIENCE. AMIA Annu Symp Proc. 2024;2024:287-92. PubMed PMID: 41726461; PubMed Central PMCID: PMC12919598.

[Lauren N Cooper](#)<sup>1</sup>, [Aamirah Vadsariya](#)<sup>1</sup>, [Mereeja Varghese](#)<sup>1</sup>, [Bhavini Nayee](#)<sup>1</sup>, [Jessica Moon](#)<sup>1</sup>, [Chaitanya Katterapalli](#)<sup>2</sup>, [Clark Walker](#)<sup>2</sup>, [Chris Gonzalez](#)<sup>2</sup>, [Sonam Sohal](#)<sup>2</sup>, [Christoph U Lehmann](#)<sup>1</sup>, [Ferdinand Velasco](#)<sup>2</sup>, [Mujeeb Basit](#)<sup>1</sup>, [DuWayne Willett](#)<sup>1</sup>


1. University of Texas Southwestern Medical Center, Dallas, Texas, USA
2. Texas Health Resources, Arlington, Texas, USA

### • 研究背景

- 電子カルテ(EHR)には膨大なデータが存在するが、施設間での活用は困難
- 共同研究にはデータ標準化が不可欠
- OMOP CDMによりデータを標準化し、OHDSIネットワークで共有可能

### • 目的

- Epicの既存ETL機能を活用してOMOP CDMを構築
- 新規開発を避け、コストと工数を削減
- 医療機関間の共同研究を促進




## EpicのネイティブETL基盤を活用したOMOP CDM実装: 協働的な取り組み

### • 方法

- Epic Caboodleをデータ基盤として使用
- 25のOMOPテーブルへEHRデータを変換
- OHDSIツール(WhiteRabbit, Usagiなど)を活用
- ETL処理:
  - THR:週次(7~8時間)
  - UTSW:日次(5~6時間)

### • 結果

- データ品質:
  - THR (Texas Health Resources): 97%
  - UTSW (Univ Texas Southwestern):98%
- 数百万~数億件規模のデータ処理に成功
- OMOP環境でのクエリ実行も有効性確認



## EpicのネイティブETL基盤を活用したOMOP CDM実装： 協働的な取り組み

- 考察


- 既存ETL活用により：
  - 開発コスト削減
  - 保守負担軽減
  - データ品質向上
- 異なる施設間でのデータ統合が容易に

- 課題

- 大規模データによる処理時間増加
- DB容量不足(TempDB問題)
- コードマッピングの複雑性

- 結論

- Epic ETLの活用は実用的で効率的
- OMOP導入のハードルを低減
- 将来的にはAI・機械学習との連携も期待



## バイオバンク間におけるフェノームプロファイルの比較

Cross Biobank Comparison of Phenomic Profiles. AMIA Annu Symp Proc. 2024;2024:939-48.  
PubMed PMID: 41726399; PubMed Central PMCID: PMC12919452.

[Abigail Newbury](#)<sup>1,2</sup>, [Xinzhao Jiang](#)<sup>1</sup>, [Karthik Natarajan](#)<sup>1</sup>, [Gamze Gürsoy](#)<sup>1,2,3</sup>


1. 1Department of Biomedical Informatics, Columbia University, New York City, NY, US
2. 2New York Genome Center, New York City, NY, US
3. 3Department of Computer Science, Columbia University, New York City, NY, US

### • 研究概要

- 米国「All of Us (AoU)」と英国「UK Biobank (UKBB)」という大規模バイオバンクを比較
- 電子カルテ (EHR) とフェノタイプ定義を用いて423疾患コホートを分析

### • 目的

- バイオバンク間の違いを明確化
- 以下を比較:
  - 有病率
  - 人口統計 (年齢・人種など)
  - 社会的決定要因 (SDOH)
  - 遺伝子関連 (GWAS)




## バイオバンク間におけるフェノームプロファイルの比較

### • 方法

- OHDSIフェノタイプライブラリを使用
- AoU用にクエリ変換ツール「Atlas2AoU」を開発
- 3疾患を詳細解析:
  - 2型糖尿病(T2D)
  - COPD
  - 急性心筋梗塞(MI)

### • 主な結果

- ① 有病率
  - AoUは約80%の疾患でUKBBより有病率が高い
  - UKBBは比較的「健康な集団」
- ② 疾患別(T2D, COPD, MI)
  - 地域分布は既知の疫学と一致
  - 例:
    - COPD → 喫煙率の高い地域で高頻度
    - MI → 率の高い地肥満域と一致



## バイオバンク間におけるフェノームプロファイルの比較

### • 主な結果

#### ③ 社会的決定要因(SDOH)

両データで有意差あり:

- 喫煙量
- 収入
- 教育
- 飲酒頻度
- 住宅所有

→ 集団背景の違いが大きい

#### ④ 遺伝子解析(GWAS)

- UKBBの方が有意遺伝子を多く検出
  - AoUはサンプル数が多いにも関わらず少ない
- 原因:
- 人口構造の違い
  - 交絡因子の影響

### • 結論

- AoUは疾病負荷が高い集団
- バイオバンク間比較では以下が重要:
  - 社会的背景
  - データ収集方法
  - 表現型定義の違い
- 解釈には注意が必要



## マルチモーダルAIは、プロテオミクスデータと電子カルテ(EHR)データを統合し、糖尿病性網膜症におけるタンパク質バイオマーカーの合理的な優先順位付け

Multimodal AI fuses proteomic and EHR data for rational prioritization of protein biomarkers in diabetic retinopathy. medRxiv: the preprint server for health sciences. United States; 2026. p. 2026.02.23.26346903. doi:[10.64898/2026.02.23.26346903](https://doi.org/10.64898/2026.02.23.26346903) PubMed PMID: 41918763; PubMed Central PMCID: PMC13034698.

Jonathan B. Lin, MD, PhD <sup>1</sup>, Samson J. Mataraso, PhD <sup>2,3,4</sup>, Madhumeeta Chadha, PhD <sup>1</sup>, Gabriel Velez, MD, PhD <sup>1</sup>, Prithvi Mruthyunjaya, MD, MHS <sup>1</sup>, Nima Aghaeepour, PhD <sup>2,3,4</sup>, & Vinit B. Mahajan MD, PhD <sup>1</sup>

1. Molecular Surgery Laboratory, Department of Ophthalmology, Stanford University, Palo Alto, CA 94304, USA
2. Department of Anesthesiology, Perioperative and Pain Medicine, Stanford University School of Medicine, Stanford, CA 94305, USA
3. Department of Pediatrics, Stanford University School of Medicine, Stanford, CA 94305, USA
4. Department of Biomedical Data Science, Stanford University School of Medicine, Stanford, CA 94305, USA



マルチモーダルAIは、プロテオミクスデータと電子カルテ(EHR)データを統合し、糖尿病性網膜症におけるタンパク質バイオマーカーの合理的な優先順位付け

- 背景

- 糖尿病網膜症(DR)は世界的な失明原因
- 既存治療(VEGFなど)はすべての患者に有効ではない
- 新しい分子メカニズムの理解が必要

- 課題

- プロテオミクス研究はサンプル数が少ない
- 数十~数百の候補タンパク質から重要なものを選ぶのが困難

- アプローチ

- 約32万人のEHRデータ + 101例のプロテオミクス
- COMET(Clinical and Omics Multi-Modal Analysis Enhanced with Transfer Learning; 転移学習型マルチモーダルAI)を開発
  - EHRで事前学習
  - その後プロテオミクスと統合



マルチモーダルAIは、プロテオミクスデータと電子カルテ(EHR)データを統合し、糖尿病性網膜症におけるタンパク質バイオマーカーの合理的な優先順位付け

- 主な結果

- ① モデル性能
  - AUROC 0.98(最高性能)
  - 単独モデルより大幅に改善
- ② 重要バイオマーカー
  - SERPINE1
  - QPCT
  - AKR1C2
  - IL2RB
  - SRSF6
- ③ 生物学的示唆
  - 炎症・免疫・代謝・血管機能に関与
  - DRの病態に深く関連

- 重要なポイント

- EHRと分子データの統合により
  - 臨床的に意味のあるタンパク質を優先抽出可能
- 小規模オミクスでも有効

- 結論

- マルチモーダルAIは
  - バイオマーカー探索に有用
  - 新規治療標的発見に有用



# Global/APACの動き



# 4月の OHDSI Global/APAC

- Global Community Call テーマ [Community Calls 2026 – OHDSI](#)

Apr. 7      Phenotype April Week 1: Phenotype Development - Live Build of Cohorts (Rao)

Apr. 14     Phenotype April Week 2: Phenotype Evaluation Using KEEPER (Ryan)

- APAC Call テーマ <https://ohdsi.org/apac/>

Apr 2      Scientific Forum: Working with What We Have: Challenges for Running OMOP-Based Network Studies in Asia-Pacific

APAC investigators and global OHDSI leaders investigate the structural and methodological characteristics of Asia-Pacific databases limiting participation in global network studies.

Apr 16     Community Call

- 2026 APAC Study Proposal



- Global Symposium 2025 Tutorial Videos
  - [Tutorials – OHDSI](#)
- ボキャブラリ 2025冬のリリース
  - [Releases · OHDSI/Vocabulary-v5.0](#)
- The OMOP School (OMOP 4 Sweden)
  - A 3+1 Day OMOP CDM Bootcamp with hands-on training and workshop that turns your data harmonization vision into reality.
  - May 26th – 28th + May 29th (optional extra day)



| Time  | Symposium Agenda - Monday April 20, 2026   | Location       |
|-------|--|----------------|
| 8:00  | <b>Registration and Coffee</b>   | Queen's Lounge |
| 9:00  | <b>Welcome to OHDSI Europe</b><br>Dr. Renske Los, Department of Medical Informatics, Erasmus MC<br>Dr. Aniek Markus, Department of Medical Informatics, Erasmus MC   | Theatre        |
| 9:05  | <b>Journey of OHDSI</b><br>Prof. Peter Rijnbeek, Chair Department of Medical Informatics, Erasmus MC   | Theatre        |
| 9:30  | <b>Collaborator Showcase - part 1</b><br>Moderated by Dr. Egill Fridgeirsson, Department of Medical Informatics, Erasmus MC  | Theatre        |
| 10:00 | <b>Speed networking</b>  | Theatre        |
| 10:15 | <b>Coffee Break &amp; posters National Nodes</b>   | Queen's Lounge |
| 11:15 | <b>Collaborator Showcase - part 2</b><br>Moderated by Dr. Egill Fridgeirsson, Department of Medical Informatics, Erasmus MC  | Theatre        |
| 11:45 | <b>Dreaming about the OHDSI journey ahead</b><br>Dr. Patrick Ryan, Vice President, Observational Health Data Analytics, Johnson & Johnson<br>Dr. Renske Los, Department of Medical Informatics, Erasmus MC | Theatre        |

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



# European Symposium: ATLAS 3.0:次世代コホート設計ツールへ

## ポイント

- ATLASがフルリニューアル
- UI/UX大幅改善
- 既存定義との後方互換性を維持
- プラグイン型アーキテクチャを採用
- ATLAS 2.xとの双方向変換が可能

## 示唆

- 日本での教育・普及の加速に直結

### Atlas 3.0: Building the Next Generation of OHDSI ATLAS

#### 12:15 - 13:45

As the OHDSI community expands and use cases grow, the original ATLAS platform presents a steep learning curve for new users. This demonstration introduces Atlas 3.0, a ground-up reimplemention designed to make cohort building more intuitive, while preserving full backward compatibility with existing definitions and WebAPI infrastructure.

Built on a modern tech stack, Atlas 3.0 lowers the entry barrier through an improved user experience, contextual guidance, and progressive disclosure. The demonstration will showcase the redesigned interface, functional parity for core workflows, a new modular plugin architecture, and a bidirectional conversion service enabling seamless migration from Atlas 2.x.

by *Peter Hoffmann*



# European Symposium: 自然言語でコホート作成

## ポイント

- 自然言語からコホート定義を自動生成
- Phenotype Library検索・テンプレート取得も自動化
- JSON構造を自動生成
- MCP (Model Context Protocol) ベースの設計

## 示唆

- 「ATLAS操作」から「対話型設計」へシフト

### Conversational Cohort Management in Data2Evidence via Model Context Protocol

#### 14:45 - 16:15

Creating valid cohort definitions in OHDSI research traditionally requires deep technical knowledge of OMOP concepts. This demonstration showcases a novel, conversational interface integrated into the open-source Data2Evidence platform. Powered by Large Language Models (LLMs) and the Model Context Protocol (MCP), our chat assistant allows researchers to construct and manage cohorts using simple natural language.

The system autonomously handles multi-step workflows - searching the OHDSI Phenotype Library, retrieving templates, and generating validated JSON structures. Attendees will see demonstrations of AI-driven cohort creation, illustrating how this approach empowers domain experts to define study populations intuitively.

*by Peter Hoffmann*



## European Symposium: 環境構築ゼロでOHDSI実行環境

### ポイント

- GitHub Codespacesで即起動
  - ATLAS / WebAPI / HADES / DB を内包
- ブラウザだけで利用可能
- DevContainerで環境完全再現
- 数分で解析開始可能

### 示唆

- 「Time-to-Science」短縮

#### From Fork to Findings: Reducing "Time-to-Science" with a One-Click OHDSI Sandbox via GitHub Codespaces

**14:45 - 16:15**

The deployment of the OHDSI analytical stack often involves significant operational overhead, complex dependency management, and local infrastructure requirements. To address these barriers, this demonstration presents a cloud-native deployment approach using GitHub Codespaces that encapsulates ATLAS, WebAPI, the HADES R package suite, and a PostgreSQL-backed OMOP database within a browser-accessible environment.

The architecture leverages the Development Container (DevContainer) specification to automate multi-service environment setup without requiring local software installation or manual network configuration. Attendees will instantiate an isolated analytical workspace directly from a template repository and observe automated service initialization and port forwarding in action, enabling immediate access to ATLAS and HADES for cohort definition and analytical package execution within minutes of environment launch.

*by Renata Silva*



# 2026 OHDSI Global Symposium

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

**Deadline Feb. 20, 2026**

- **2026 OHDSI Global Symposium Call for Plenary Sessions**
  - Open community data standards
  - Methodological research
  - Open-source development
  - Clinical applications
- **2026 OHDSI Global Symposium Call for Tutorials**

**Deadline..**

- **Collaborator Showcase**