

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 evening conference #51

イブニングカンファレンス(第51回)

2024.2.29



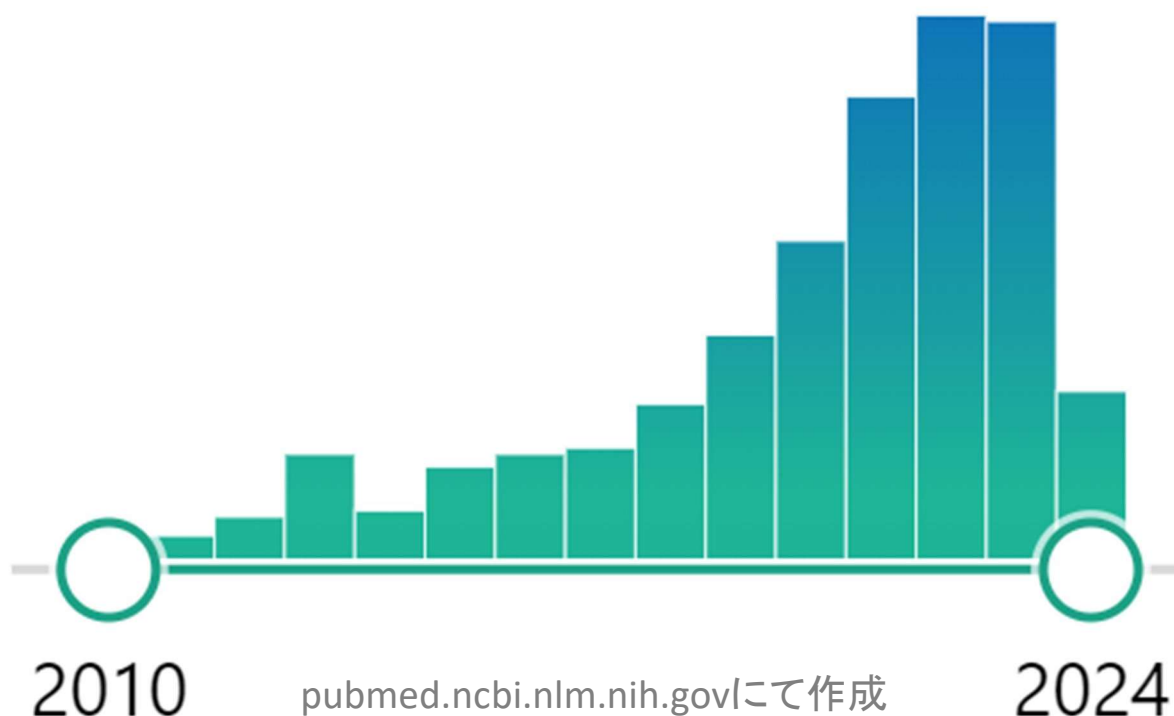
## 本日の内容

- OHDSI関連論文紹介
- OHDSI Global の話題から
- Study-a-thonについて



# OHDSI関連論文

Pubmedで“OHDSI or OMOP”を検索



全期間累計：1月399本→2月407本

- 検索に漏れているものがあるため、実際は累計500本を超えている。
- 年間では100本ペース。



## OHDSI PLPをmlr3と比較

> [Sci Rep. 2024 Jan 27;14\(1\):2287. doi: 10.1038/s41598-024-52723-y.](#)

# A comparative patient-level prediction study in OMOP CDM: applicative potential and insights from synthetic data

Najia Ahmadi <sup># 1</sup>, Quang Vu Nguyen <sup># 2</sup>, Martin Sedlmayr <sup>2</sup>, Markus Wolfien <sup>2 3</sup>

Affiliations – collapse

### Affiliations

- 1 Institute for Medical Informatics and Biometry, Faculty of Medicine Carl Gustav Carus, Technische Universität Dresden, 01307, Dresden, Germany. [najia.ahmadi@tu-dresden.de](mailto:najia.ahmadi@tu-dresden.de).
  - 2 Institute for Medical Informatics and Biometry, Faculty of Medicine Carl Gustav Carus, Technische Universität Dresden, 01307, Dresden, Germany.
  - 3 Center for Scalable Data Analytics and Artificial Intelligence (ScaDS.AI), Dresden, Germany.
- # Contributed equally.

## Abstract

The emergence of collaborations, which standardize and combine multiple clinical databases across different regions, provide a wealthy source of data, which is fundamental for clinical prediction models, such as patient-level predictions. With the aid of such large data pools, researchers are able to develop clinical prediction models for improved disease classification, risk assessment, and beyond. To fully utilize this potential, Machine Learning (ML) methods are commonly required to process these large amounts of data on disease-specific patient cohorts. As a consequence, the Observational Health Data Sciences and Informatics (OHDSI) collaborative develops a framework to facilitate the application of ML models for these standardized patient datasets by using the Observational Medical Outcomes Partnership (OMOP) common data model (CDM). In this study, we compare the feasibility of current web-based OHDSI approaches, namely ATLAS and "Patient-level Prediction" (PLP), against a native solution (R based) to conduct such ML-based patient-level prediction analyses in OMOP. This will enable potential users to select the most suitable approach for their investigation. Each of the applied ML solutions was individually utilized to solve the same patient-level prediction task. Both approaches went through an exemplary benchmarking analysis to assess the weaknesses and strengths of the PLP R-Package. In this work, the performance of this package was subsequently compared versus the commonly used native R-package called Machine Learning in R 3 (mlr3), and its sub-packages. The approaches were evaluated on performance, execution time, and ease of model implementation. The results show that the PLP package has shorter execution times, which indicates great scalability, as well as intuitive code implementation, and numerous possibilities for visualization. However, limitations in comparison to native packages were depicted in the implementation of specific ML classifiers (e.g., Lasso), which may result in a decreased performance for real-world prediction problems. The findings here contribute to the overall effort of developing ML-based prediction models on a clinical scale and provide a snapshot for future studies that explicitly aim to develop patient-level prediction models in OMOP CDM.

## 要旨

異なる地域にまたがる複数の臨床データベースを標準化し統合する共同研究の出現は、患者レベルの予測など臨床予測モデルの基礎となる豊富なデータ源を提供する。このような大規模なデータプールを利用することで、研究者は、疾患の分類やリスク評価、さらにはそれ以上の改善を目的とした臨床予測モデルを開発することができる。この可能性を十分に活用するために、機械学習（ML）手法は、一般的に、疾患特異的な患者コホートに関するこれらの大量のデータを処理するために必要とされる。そのため、OHDSI共同研究体は、OMOP CDM を用いて、これらの標準化された患者データセットに対するMLモデルの適用を容易にするフレームワークを開発している。

本研究では、OMOPにおけるMLベースの患者レベル予測分析を行うためのネイティブソリューション（Rベース）に対して、現在のウェブベースのOHDSIアプローチ、すなわちATLASと "Patient-level Prediction"（PLP）の実現可能性を比較する。これにより、潜在的なユーザーは、自分の調査に最も適したアプローチを選択できるようになる。

適用されたMLソリューションはそれぞれ、同じ患者レベル予測タスクを解くために利用された。どちらのアプローチも、PLP R-Packageの弱点と長所を評価するために、模範的なベンチマーク分析を経た。この研究では、このパッケージの性能は、その後、Machine Learning in R 3 (mlr3)と呼ばれる一般的に使用されているネイティブRパッケージとそのサブパッケージと比較された。それぞれのアプローチは、性能、実行時間、モデル実装の容易さで評価された。その結果、PLPパッケージの方が実行時間が短く、スケーラビリティが高いこと、直感的なコード実装、可視化の可能性が高いことが示された。しかし、特定のML分類器（例：Lasso）の実装において、ネイティブパッケージと比較した場合の限界が示され、これは実世界の予測問題に対する性能低下をもたらす可能性がある。ここで得られた知見は、臨床規模でのMLベースの予測モデル開発の全体的な取り組みに貢献し、OMOP CDMで患者レベルの予測モデルを開発することを明確に目指した将来の研究のためのスナップショットを提供する。



## 画像CDMつづき

> *J Imaging Inform Med.* 2024 Feb 5. doi: 10.1007/s10278-024-00982-6. Online ahead of print.

# Development of Medical Imaging Data Standardization for Imaging-Based Observational Research: OMOP Common Data Model Extension

Woo Yeon Park <sup># 1</sup>, Kyulee Jeon <sup># 2 3</sup>, Teri Sippel Schmidt <sup>4</sup>, Haridimos Kondylakis <sup>5</sup>,  
Tarik Alkasab <sup>6</sup>, Blake E Dewey <sup>7</sup>, Seng Chan You <sup>2 3</sup>, Paul Nagy <sup>4</sup>

PMID: 38315345

### Affiliations

- 1 Biomedical Informatics and Data Science, Johns Hopkins University, 855 N Wolfe St, Rangos 616, Baltimore, MD, USA. wpark11@jhu.edu.
- 2 Department of Biomedical Systems Informatics, Yonsei University College of Medicine, Seoul, Korea.
- 3 Institute for Innovation in Digital Healthcare, Yonsei University, Seoul, Korea.
- 4 Biomedical Informatics and Data Science, Johns Hopkins University, 855 N Wolfe St, Rangos 616, Baltimore, MD, USA.
- 5 Institute of Computer Science, Foundation of Research & Technology-Hellas (FORTH), Heraklion, Greece.
- 6 Department of Radiology, Massachusetts General Hospital, Boston, MA, USA.
- 7 Department of Neurology, Johns Hopkins University, Baltimore, MD, USA.



## Abstract

The rapid growth of artificial intelligence (AI) and deep learning techniques require access to large inter-institutional cohorts of data to enable the development of robust models, e.g., targeting the identification of disease biomarkers and quantifying disease progression and treatment efficacy. The Observational Medical Outcomes Partnership Common Data Model (OMOP CDM) has been designed to accommodate a harmonized representation of observational healthcare data. This study proposes the Medical Imaging CDM (MI-CDM) extension, adding two new tables and two vocabularies to the OMOP CDM to address the structural and semantic requirements to support imaging research. The tables provide the capabilities of linking DICOM data sources as well as tracking the provenance of imaging features derived from those images. The implementation of the extension enables phenotype definitions using imaging features and expanding standardized computable imaging biomarkers. This proposal offers a comprehensive and unified approach for conducting imaging research and outcome studies utilizing imaging features.

## 概要

人工知能（AI）やディープラーニング技術の急速な発展により、疾患バイオマーカーの同定や疾患の進行や治療効果の定量化など、ロバストなモデルの開発を可能にする大規模な施設間コホートデータへのアクセスが必要とされている。OMOP CDM は、観察医療データの調和された表現に対応するように設計されている。本研究ではMedical Imaging CDM (MI-CDM)の拡張を提案し、OMOP CDMに2つの新しいテーブルと2つの語彙を追加することで、画像研究をサポートするための構造的・意味的要件に対応する。これらのテーブルは、DICOMデータソースをリンクする機能と、それらの画像から得られた画像特徴の出所を追跡する機能を提供する。この拡張機能の実装により、画像特徴を用いた表現型の定義が可能となり、標準化された計算可能な画像バイオマーカーの拡張が可能となる。本提案は、画像特徴を利用した画像研究およびアウトカム研究を実施するための包括的かつ統一的なアプローチを提供する。





## MedDRAとICDのマッピング

› [BMC Med Inform Decis Mak.](#) 2024 Feb 7;23(Suppl 4):299. doi: 10.1186/s12911-023-02375-1.

# Evaluating MedDRA-to-ICD terminology mappings

Xinyuan Zhang<sup>1</sup>, Yixue Feng<sup>2</sup>, Fang Li<sup>1</sup>, Jin Ding<sup>1</sup>, Danyal Tahseen<sup>3</sup>, Ezekiel Hinojosa<sup>3</sup>,  
Yong Chen<sup>4</sup>, Cui Tao<sup>5 6</sup>

Affiliations – collapse

### Affiliations

- 1 McWilliam School of Biomedical Informatics, University of Texas Health Science Center at Houston, Houston, TX, USA.
- 2 School of Engineering and Applied Science, University of Pennsylvania, Philadelphia, PA, USA.
- 3 McGovern Medical School, University of Texas Health Science Center at Houston, Houston, TX, USA.
- 4 The Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA.
- 5 McWilliam School of Biomedical Informatics, University of Texas Health Science Center at Houston, Houston, TX, USA. Cui.Tao@uth.tmc.edu.
- 6 Department of Artificial Intelligence and Informatics, Mayo Clinic, Jacksonville, FL, USA. Cui.Tao@uth.tmc.edu.

## Abstract

**Background:** In this era of big data, data harmonization is an important step to ensure reproducible, scalable, and collaborative research. Thus, terminology mapping is a necessary step to harmonize heterogeneous data. Take the Medical Dictionary for Regulatory Activities (MedDRA) and International Classification of Diseases (ICD) for example, the mapping between them is essential for drug safety and pharmacovigilance research. Our main objective is to provide a quantitative and qualitative analysis of the mapping status between MedDRA and ICD. We focus on evaluating the current mapping status between MedDRA and ICD through the Unified Medical Language System (UMLS) and Observational Medical Outcomes Partnership Common Data Model (OMOP CDM). We summarized the current mapping statistics and evaluated the quality of the current MedDRA-ICD mapping; for unmapped terms, we used our self-developed algorithm to rank the best possible mapping candidates for additional mapping coverage.

**Results:** The identified MedDRA-ICD mapped pairs cover 27.23% of the overall MedDRA preferred terms (PT). The systematic quality analysis demonstrated that, among the mapped pairs provided by UMLS, only 51.44% are considered an exact match. For the 2400 sampled unmapped terms, 56 of the 2400 MedDRA Preferred Terms (PT) could have exact match terms from ICD.

**Conclusion:** Some of the mapped pairs between MedDRA and ICD are not exact matches due to differences in granularity and focus. For 72% of the unmapped PT terms, the identified exact match pairs illustrate the possibility of identifying additional mapped pairs. Referring to its own mapping standard, some of the unmapped terms should qualify for the expansion of MedDRA to ICD mapping in UMLS.

## 要旨

**背景:** ビッグデータの時代において、データの調和は、再現性、拡張性、共同研究を保証するための重要なステップである。したがって、用語のマッピングは異種データを調和させるために必要なステップである。例えば、Medical Dictionary for Regulatory Activities (MedDRA)とInternational Classification of Diseases (ICD)を例にとると、これらの間のマッピングは、医薬品の安全性とファーマコビジランス研究にとって不可欠である。我々の主な目的は、MedDRAとICDのマッピング状況を定量的・定性的に分析することである。Unified Medical Language System (UMLS)とObservational Medical Outcomes Partnership Common Data Model (OMOP CDM)を用いて、MedDRAとICDの現在のマッピング状況を評価する。マッピングされていない用語については、独自に開発したアルゴリズムを用いて、マッピング候補をランク付けし、マッピング範囲の追加を行った。

**結果:** マッピングされた MedDRA-ICD のペアは、MedDRA の優先用語 (PT) 全体の 27.23%をカバーしている。系統的な品質分析によると、UMLSから提供されたマッピングペアのうち、完全一致と考えられるのは51.44%のみであった。2400の未マッピング用語のうち、56のMedDRA優先用語 (PT) がICDと完全一致する可能性がある。

**結論:** MedDRA と ICD 間でマップされたペアの一部は、粒度と焦点の違いにより完全一致しない。マッピングされていない PT 用語の 72%について、完全一致のペアが特定されたことから、さらにマッピングされたペアが特定される可能性が示された。UMLS のマッピング標準を参照すると、未マッピングの用語の一部は、MedDRA から UMLS の ICD へのマッピングの拡張に適格である。



## 汎アフリカのデータハブ

Review > [Front Digit Health. 2024 Jan 29;6:1329630. doi: 10.3389/fdgth.2024.1329630.](#)  
eCollection 2024.

# INSPIRE datahub: a pan-African integrated suite of services for harmonising longitudinal population health data using OHDSI tools

Tathagata Bhattacharjee<sup>1</sup>, Sylvia Kiwuwa-Muyingo<sup>2</sup>, Chifundo Kanjala<sup>1 3</sup>, Molulaqhoa L Maoyi<sup>4</sup>, David Amadi<sup>1</sup>, Michael Ochola<sup>2</sup>, Damazo Kadengye<sup>2 5</sup>, Arofan Gregory<sup>6</sup>, Agnes Kiragga<sup>2</sup>, Amelia Taylor<sup>7</sup>, Jay Greenfield<sup>5</sup>, Emma Slaymaker<sup>1</sup>, Jim Todd<sup>1</sup>; INSPIRE Network

### Affiliations

- 1 Department of Population Health, Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, University of London, London, United Kingdom.
- 2 African Population and Health Research Center (APHRC), Nairobi, Kenya.
- 3 UNICEF (Malawi), Lilongwe, Malawi.
- 4 South African Population Research Infrastructure Network (SAPRIN), South African Medical Research Council, Durban, South Africa.
- 5 Department of Economics and Statistics, Kabale University, Kabale, Uganda.
- 6 Committee on Data of the International Science Council (CODATA), Paris, France.
- 7 Malawi University of Business and Applied Sciences, Blantyre, Malawi.

## Abstract

**Introduction:** Population health data integration remains a critical challenge in low- and middle-income countries (LMIC), hindering the generation of actionable insights to inform policy and decision-making. This paper proposes a pan-African, Findable, Accessible, Interoperable, and Reusable (FAIR) research architecture and infrastructure named the INSPIRE datahub. This cloud-based Platform-as-a-Service (PaaS) and on-premises setup aims to enhance the discovery, integration, and analysis of clinical, population-based surveys, and other health data sources.

**Methods:** The INSPIRE datahub, part of the Implementation Network for Sharing Population Information from Research Entities (INSPIRE), employs the Observational Health Data Sciences and Informatics (OHDSI) open-source stack of tools and the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) to harmonise data from African longitudinal population studies. Operating on Microsoft Azure and Amazon Web Services cloud platforms, and on on-premises servers, the architecture offers adaptability and scalability for other cloud providers and technology infrastructure. The OHDSI-based tools enable a comprehensive suite of services for data pipeline development, profiling, mapping, extraction, transformation, loading, documentation, anonymization, and analysis.

**Results:** The INSPIRE datahub's "On-ramp" services facilitate the integration of data and metadata from diverse sources into the OMOP CDM. The datahub supports the implementation of OMOP CDM across data producers, harmonizing source data semantically with standard vocabularies and structurally conforming to OMOP table structures. Leveraging OHDSI tools, the datahub performs quality assessment and analysis of the transformed data. It ensures FAIR data by establishing metadata flows, capturing provenance throughout the ETL processes, and providing accessible metadata for potential users. The ETL provenance is documented in a machine- and human-readable Implementation Guide (IG), enhancing transparency and usability.

**Conclusion:** The pan-African INSPIRE datahub presents a scalable and systematic solution for integrating health data in LMICs. By adhering to FAIR principles and leveraging established standards like OMOP CDM, this architecture addresses the current gap in generating evidence to support policy and decision-making for improving the well-being of LMIC populations. The federated research network provisions allow data producers to maintain control over their data, fostering collaboration while respecting data privacy and security concerns. A use-case demonstrated the pipeline using OHDSI and other open-source tools.

## 要旨

**はじめに:** 中低所得国（LMIC）では、集団健康データの統合が依然として重要な課題であり、政策や意思決定に役立つ実用的な知見の創出を妨げている。本稿では、汎アフリカ的で、検索可能で、アクセス可能で、相互運用可能で、再利用可能な（FAIR）研究アーキテクチャとインフラストラクチャであるINSPIRE datahubを提案する。このクラウドベースのPlatform-as-a-Service（PaaS）とオンプレミスのセットアップは、臨床データ、集団ベースの調査、その他の健康データソースの発見、統合、分析を強化することを目的としている。

**方法:** INSPIRE datahubは、Implementation Network for Sharing Population Information from Research Entities (INSPIRE)の一部であり、OHDSIオープンソースツールスタックとOMOP共通データモデル(CDM)を採用し、アフリカの縦断的集団研究のデータを調和させている。Microsoft AzureとAmazon Web Servicesのクラウドプラットフォーム、およびオンプレミスのサーバー上で動作するこのアーキテクチャは、他のクラウドプロバイダーや技術インフラへの適応性と拡張性を提供する。OHDSIベースのツールは、データパイプラインの開発、プロファイリング、マッピング、抽出、変換、ロード、ドキュメンテーション、匿名化、分析のための包括的な一連のサービスを可能にする。

**結果:** INSPIRE datahubの "On-ramp "サービスは、多様なソースからのデータとメタデータのOMOP CDMへの統合を促進する。データハブは、ソースデータを標準的な語彙で意味的に調和させ、構造的にOMOPテーブル構造に適合させることで、データ作成者間でのOMOP CDMの実装をサポートする。OHDSIツールを活用し、datahubは変換されたデータの品質評価と分析を行う。メタデータのフローを確立し、ETL プロセス全体を通して実績情報を取得し、潜在的なユーザがアクセス可能なメタデータを提供することで、FAIR データを保証する。ETLの実績は、機械的・人間的に読み取り可能な実装ガイド（IG）に文書化され、透明性とユーザビリティを向上させる。

**結論:** 汎アフリカINSPIREデータハブは、LMICsの保健データを統合するためのスケーラブルで体系的なソリューションを提示している。FAIRの原則を遵守し、OMOP CDMのような確立された標準を活用することで、このアーキテクチャは、LMICの人々の幸福を向上させるための政策や意思決定を支援するエビデンスの生成における現在のギャップに対処している。連携研究ネットワークの規定によって、データ作成者はデータの管理を維持することができ、データのプライバシーとセキュリティの懸念を尊重しながら、共同研究を促進することができる。OHDSIと他のオープンソースツールを使用したパイプラインの使用例が示された。





## Transformerモデルによる予後予測

› [J Am Med Inform Assoc. 2024 Feb 13:ocae025. doi: 10.1093/jamia/ocae025. Online ahead of print.](#)

# Transformer-based time-to-event prediction for chronic kidney disease deterioration

Moshe Zisser<sup>1</sup>, Dvir Aran<sup>2 3</sup>

Affiliations – collapse

### Affiliations

- 1 Faculty of Data and Decision Sciences, Technion-Israel Institute of Technology, Haifa, 3200003, Israel.
- 2 Faculty of Biology, Technion-Israel Institute of Technology, Haifa, 3200003, Israel.
- 3 The Taub Faculty of Computer Science, Technion-Israel Institute of Technology, Haifa, 3200003, Israel.

PMID: 38349850 DOI: [10.1093/jamia/ocae025](#)



## Abstract

**Objective:** Deep-learning techniques, particularly the Transformer model, have shown great potential in enhancing the prediction performance of longitudinal health records. Previous methods focused on fixed-time risk prediction, however, time-to-event prediction is often more appropriate for clinical scenarios. Here, we present STRAFE, a generalizable survival analysis Transformer-based architecture for electronic health records.

**Materials and methods:** The input for STRAFE is a sequence of visits with SNOMED-CT codes in OMOP-CDM format. A Transformer-based architecture was developed to calculate probabilities of the occurrence of the event in each of 48 months. Performance was evaluated using a real-world claims dataset of over 130 000 individuals with stage 3 chronic kidney disease (CKD).

**Results:** STRAFE showed improved mean absolute error (MAE) compared to other time-to-event algorithms in predicting the time to deterioration to stage 5 CKD. Additionally, STRAFE showed an improved area under the receiver operating curve compared to binary outcome algorithms. We show that STRAFE predictions can improve the positive predictive value of high-risk patients by 3-fold. Finally, we suggest a novel visualization approach to predictions on a per-patient basis.

**Discussion:** Time-to-event predictions are the most appropriate approach for clinical predictions. Our deep-learning algorithm outperformed not only other time-to-event prediction algorithms but also fixed-time algorithms, possibly due to its ability to train on censored data. We demonstrated possible clinical usage by identifying the highest-risk patients.

**Conclusions:** The ability to accurately identify patients at high risk and prioritize their needs can result in improved health outcomes, reduced costs, and more efficient use of resources.

## 要旨

**目的:** ディープラーニング技術、特にTransformerモデルは、縦断的な健康記録の予測性能を高める上で大きな可能性を示している。これまでの手法では、固定時間のリスク予測に焦点が当てられていたが、臨床シナリオにおいては、時間対イベント予測がより適切であることが多い。ここでは、電子カルテのための一般化可能な生存分析TransformerベースのアーキテクチャであるSTRAFEを紹介する。

**材料と方法:** STRAFEの入力は、OMOP-CDM形式のSNOMED-CTコードを持つ一連の診察である。Transformerベースのアーキテクチャを開発し、48カ月の各月におけるイベントの発生確率を計算した。ステージ3の慢性腎臓病（CKD）患者130,000人以上からなる実世界の請求データセットを用いて性能を評価した。

**結果:** STRAFEは、ステージ5のCKDに悪化するまでの予測において、他のtime-to-eventアルゴリズムと比較して、平均絶対誤差（MAE）の改善を示した。さらに、STRAFEは二値転帰アルゴリズムと比較して、受信者動作曲線下面積の改善を示した。我々は、STRAFE予測により、高リスク患者の陽性的中率を3倍改善できることを示す。最後に、患者ごとの予測に対する新しい可視化アプローチを提案する。

**考察:** time-to-event予測は、臨床予測に最も適したアプローチである。我々のディープラーニングアルゴリズムは、他のtime-to-event予測アルゴリズムだけでなく、固定時間アルゴリズムも上回った。我々は、最もリスクの高い患者を同定することにより、臨床的利用の可能性を示した。

**結論:** 高リスクの患者を正確に同定し、そのニーズに優先順位をつける能力は、健康アウトカムの改善、コストの削減、資源の効率的な利用につながる可能性がある。



## 臨床試験リクルート支援システム

> [Trials](#). 2024 Feb 16;25(1):125. doi: 10.1186/s13063-024-07918-z.

# Effectiveness of IT-supported patient recruitment: study protocol for an interrupted time series study at ten German university hospitals

Martin Boeker <sup>1 2</sup>, Daniela Zöller <sup>1</sup>, Romina Blasini <sup>3</sup>, Philipp Macho <sup>4</sup>, Sven Helfer <sup>5</sup>,  
Max Behrens <sup>1</sup>, Hans-Ulrich Prokosch <sup>6</sup>, Christian Gulden <sup>7</sup>

PMID: 38365848 PMCID: PMC10870691

### Affiliations

- <sup>1</sup> Institute of Medical Biometry and Statistics, Medical Faculty and Medical Center, University of Freiburg, Freiburg im Breisgau, Germany.
- <sup>2</sup> Chair of Medical Informatics, Institute of Artificial Intelligence and Informatics in Medicine, Klinikum rechts der Isar, School of Medicine and Health, Technical University of Munich, Munich, Germany.
- <sup>3</sup> Institute of Medical Informatics, Justus-Liebig-University Gießen, Gießen, Germany.
- <sup>4</sup> Institute of Medical Biostatistics, Epidemiology and Informatics (IMBEI), Mainz University Medical Center, Mainz, Germany.
- <sup>5</sup> Department of Pediatrics, Faculty of Medicine and University Hospital Carl Gustav Carus, Technische Universität Dresden, Dresden, Germany.
- <sup>6</sup> Chair of Medical Informatics, Department of Medical Informatics, Biometrics and Epidemiology, Friedrich-Alexander Universität Erlangen-Nürnberg, Erlangen, Germany.
- <sup>7</sup> Chair of Medical Informatics, Department of Medical Informatics, Biometrics and Epidemiology, Friedrich-Alexander Universität Erlangen-Nürnberg, Erlangen, Germany.  
christian.gulden@fau.de.

## Abstract

**Background:** As part of the German Medical Informatics Initiative, the MIRACUM project establishes data integration centers across ten German university hospitals. The embedded MIRACUM Use Case "Alerting in Care - IT Support for Patient Recruitment", aims to support the recruitment into clinical trials by automatically querying the repositories for patients satisfying eligibility criteria and presenting them as screening candidates. The objective of this study is to investigate whether the developed recruitment tool has a positive effect on study recruitment within a multi-center environment by increasing the number of participants. Its secondary objective is the measurement of organizational burden and user satisfaction of the provided IT solution.

**Methods:** The study uses an Interrupted Time Series Design with a duration of 15 months. All trials start in the control phase of randomized length with regular recruitment and change to the intervention phase with additional IT support. The intervention consists of the application of a recruitment-support system which uses patient data collected in general care for screening according to specific criteria. The inclusion and exclusion criteria of all selected trials are translated into a machine-readable format using the OHDSI ATLAS tool. All patient data from the data integration centers is regularly checked against these criteria. The primary outcome is the number of participants recruited per trial and week standardized by the targeted number of participants per week and the expected recruitment duration of the specific trial. Secondary outcomes are usability, usefulness, and efficacy of the recruitment support. Sample size calculation based on simple parallel group assumption can demonstrate an effect size of  $d=0.57$  on a significance level of 5% and a power of 80% with a total number of 100 trials (10 per site). Data describing the included trials and the recruitment process is collected at each site. The primary analysis will be conducted using linear mixed models with the actual recruitment number per week and trial standardized by the expected recruitment number per week and trial as the dependent variable.

**Discussion:** The application of an IT-supported recruitment solution developed in the MIRACUM consortium leads to an increased number of recruited participants in studies at German university hospitals. It supports employees engaged in the recruitment of trial participants and is easy to integrate in their daily work.

## 要旨

**背景:** German Medical Informatics Initiativeの一環として、MIRACUMプロジェクトはドイツの10の大学病院にデータ統合センターを設立した。組み込まれたMIRACUMユースケース "Alerting in Care - IT Support for Patient Recruitment" は、適格基準を満たす患者を自動的にリポジトリに照会し、スクリーニング候補として提示することで、臨床試験へのリクルートをサポートすることを目的としている。本研究の目的は、開発されたリクルートメントツールが、参加者数を増加させることにより、多施設環境における試験のリクルートにプラスの効果をもたらすかどうかを調査することである。副次的な目的は、提供されたITソリューションの組織的負担とユーザー満足度を測定することである。

**方法:** 本研究では、15ヵ月間の中断時系列計画 (Interrupted Time Series Design) を採用している。すべての試験は、定期的なリクルートを伴う無作為化期間の対照相から開始され、追加のIT支援を伴う介入相に変更される。介入は、一般診療で収集された患者データを用いて特定の基準に従ってスクリーニングを行うリクルート支援システムの適用からなる。選択されたすべての試験の組み入れ基準および除外基準は、OHDSI ATLASツールを用いて機械可読形式に変換される。データ統合センターからの全患者データは定期的にこれらの基準と照合される。主要アウトカムは、各試験および週ごとに募集された参加者数で、週ごとの目標参加者数と特定の試験の予想募集期間によって標準化されている。副次的アウトカムは、リクルート支援の使いやすさ、有用性、有効性である。単純な並行群仮定に基づくサンプルサイズの計算では、有意水準5%で $d=0.57$ の効果量を示すことができ、総試験数100 (1施設あたり10) で検出力80%であった。組み入れられた試験と募集プロセスに関するデータは、各施設で収集される。一次解析は、従属変数として、週および試験ごとの実際の採用数を週および試験ごとの予想採用数で標準化した線形混合モデルを用いて実施する。

**考察:** MIRACUMコンソーシアムで開発されたIT支援型リクルートソリューションの適用により、ドイツの大学病院における研究において、リクルートされた参加者数が増加した。このソリューションは、試験参加者の募集に携わる従業員をサポートし、日常業務に簡単に組み込むことができる。



## OMOP CDM 概説

Review > [BMJ Health Care Inform. 2024 Feb 21;31\(1\):e100953. doi: 10.1136/bmjhci-2023-100953.](#)

# Seamless EMR data access: Integrated governance, digital health and the OMOP-CDM

Christine Mary Hallinan<sup>1</sup>, Roger Ward<sup>2</sup>, Graeme K Hart<sup>3</sup>, Clair Sullivan<sup>4</sup>, Nicole Pratt<sup>5</sup>, Ashley P Ng<sup>6 7</sup>, Daniel Capurro<sup>3 8</sup>, Anton Van Der Vegt<sup>9</sup>, Siaw-Teng Liaw<sup>10</sup>, Oliver Daly<sup>3</sup>, Blanca Gallego Luxan<sup>11</sup>, David Bunker<sup>9</sup>, Douglas Boyle<sup>1</sup>

PMID: 38387992 PMCID: [PMC10882353](#)

### Affiliations

<sup>1</sup>Health and Biomedical Informatics Centre, Research Information Technology Unit (HaBIC R2), Department of General Practice and Primary Care, The University of Melbourne Faculty of Medicine Dentistry and Health Sciences, Melbourne, Victoria, Australia [hallinan@unimelb.edu.au](mailto:hallinan@unimelb.edu.au) [dboyle@unimelb.edu.au](mailto:dboyle@unimelb.edu.au).

<sup>2</sup>Health and Biomedical Informatics Centre, Research Information Technology Unit (HaBIC R2), Department of General Practice and Primary Care, The University of Melbourne Faculty of Medicine Dentistry and Health Sciences, Melbourne, Victoria, Australia.

<sup>3</sup>School of Computing and Information Systems, Faculty of Engineering and Information Technology, Centre for the Digital Transformation of Health, The University of Melbourne Faculty of Medicine Dentistry and Health Sciences, Melbourne, Victoria, Australia.

<sup>4</sup>Queensland Digital Health Centre (QDHeC), Centre for Health Services Research, The University of Queensland Faculty of Medicine, Woolloongabba, Queensland, Australia.

<sup>5</sup>Quality Use of Medicines and Pharmacy Research Centre, Clinical and Health Sciences, University of South Australia, Adelaide, South Australia, Australia.

<sup>6</sup>Clinical Haematology Department, The Royal Melbourne Hospital, Peter MacCallum Cancer Centre, Melbourne, Victoria, Australia.

<sup>7</sup>Sir Peter MacCallum Department of Oncology, The University of Melbourne Faculty of Medicine Dentistry and Health Sciences, Melbourne, Victoria, Australia.

<sup>8</sup>Department of General Medicine, The Royal Melbourne Hospital, Parkville, Victoria, Australia.

<sup>9</sup>Queensland Digital Health Centre (QDHeC), Centre for Health Services Research, The University of Queensland Faculty of Medicine, Herston, Queensland, Australia.

<sup>10</sup>School of Population Health, UNSW, Sydney, New South Wales, Australia.

<sup>11</sup>Centre for Big Data Research in Health (CBDRH), UNSW, Sydney, New South Wales, Australia.

## Abstract

**Objectives** In this overview, we describe the Observational Medical Outcomes Partnership Common Data Model (OMOP-CDM), the established governance processes employed in EMR data repositories, and demonstrate how OMOP transformed data provides a lever for more efficient and secure access to electronic medical record (EMR) data by health service providers and researchers. **Methods** Through pseudonymisation and common data quality assessments, the OMOP-CDM provides a robust framework for converting complex EMR data into a standardised format. This allows for the creation of shared end-to-end analysis packages without the need for direct data exchange, thereby enhancing data security and privacy. By securely sharing de-identified and aggregated data and conducting analyses across multiple OMOP-converted databases, patient-level data is securely firewalled within its respective local site. **Results** By simplifying data management processes and governance, and through the promotion of interoperability, the OMOP-CDM supports a wide range of clinical, epidemiological, and translational research projects, as well as health service operational reporting. **Discussion** Adoption of the OMOP-CDM internationally and locally enables conversion of vast amounts of complex, and heterogeneous EMR data into a standardised structured data model, simplifies governance processes, and facilitates rapid repeatable cross-institution analysis through shared end-to-end analysis packages, without the sharing of data. **Conclusion** The adoption of the OMOP-CDM has the potential to transform health data analytics by providing a common platform for analysing EMR data across diverse healthcare settings.

## 要旨

**目的** この総論では、OMOP-CDMと、EMRデータリポジトリで採用されている確立されたガバナンスプロセスについて説明し、OMOPで変換されたデータが、医療サービス提供者や研究者が電子カルテデータをより効率的かつ安全に利用するためのテコとなることを示す。**方法** 仮名化と共通のデータ品質評価を通じて、OMOP-CDMは複雑なEMRデータを標準化された形式に変換するための強固な枠組みを提供する。これにより、データを直接交換することなく、エンド・ツー・エンドの共有分析パッケージを作成することが可能となり、データの安全性とプライバシーが強化される。非識別・集計データを安全に共有し、複数のOMOP変換データベースを横断して分析を行うことで、患者レベルのデータはそれぞれのローカルサイト内で安全にファイアウォールされる。結果 データ管理プロセスとガバナンスを簡素化し、相互運用性を促進することで、OMOP-CDM は臨床、疫学、トランスレーショナルリサーチプロジェクトや医療サービスの業務報告など、幅広い分野をサポートする。**考察** OMOP-CDMを国際的、地域的に採用することで、複雑で異種混合の膨大な量のEMRデータを標準化された構造化データモデルに変換し、ガバナンスプロセスを簡素化し、データを共有することなく、共有されたエンドツーエンドの解析パッケージを通じて、迅速な反復可能な機関横断的解析を容易にすることができる。**結論** OMOP-CDMの採用は、多様な医療環境にまたがるEMRデータを分析するための共通プラットフォームを提供することで、医療データ分析を変革する可能性を秘めている。





# 今月のOHDSI Global

## ● APAC Call テーマ

### Feb. 1 Scientific Forum

- Results of survey on support areas
- Plans for 2024 Q1
- APAC study updates

### Feb. 15 Community Call: New Workgroup Introductions

- Industry (Mui Van Zandt)
- Evidence Translation (Nicole Pratt)



# 今月のOHDSI Global

## ● Global Community Call テーマ

Jan. 30 Introduction to **Phenotype Phebruary**

Feb. 6 **Workgroup OKRs/Phenotype Phebruary Update #1**

Methods Research、HADES、周産期および生殖医療、レジストリ、ステアリンググループ

Feb. 13 Workgroup OKRs/Phenotype Phebruary Update #2

共通データモデル、ネットワークデータ品質、アジア太平洋 (APAC)、産業、眼科と視覚研究、外科・周術期医学

Feb. 20 Workgroup OKRs/Phenotype Phebruary Update #3

Themis、ヘルスケアシステム、生成 AI および基礎モデル、腫瘍学、ワクチン語彙、患者レベル予測 (PLP)、ATLAS、オープンソースコミュニティ、精神医学、自然言語処理 (NLP)

Feb. 27 Workgroup OKRs/Phenotype Phebruary Update #4

FHIR+OMOP、健康の公平性、アフリカ支部、電子動物健康記録、CDM 語彙、表現型の開発と評価、歯科、医用画像、医療機器、GIS



# Study-a-thon

Study-a-thonとは

特定のトピックについて、医学的エビデンスを創出するための数日間にわたる集中的なワークショップ。

(参考サイト例)

<https://www.ehden.eu/study-a-thon/>

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9670330/>

<https://www.thehyve.nl/articles/studyathon-multidisciplinary-collaboration-in-biomedical-open-science>

## Who attends a study-a-thon?

Researchers  
Epidemiologists  
Data Scientists  
Clinicians  
Data Partners  
Patient Organisations  
Post-Doc's  
Students

### Critical

Those skilled and experienced in OHDSI tools, methods and skills to work with CDM-mapped datasets

## Preparation

### What's the question?

What does the literature say?

### Do you have the data (mapped) available to conduct your study?

Organising:

Leadership  
Objectives  
Project Management  
Agreements/ approval  
Resources  
Communications  
Sub-teams  
Data Partner

## Build

### Minimum of 2-3 months in advance:

Leadership and organising  
Project plan & agreements  
Engagement with Data Partners  
Engagement with Clinicians  
Organising the study-a-thon event (virtual or live)  
Communications  
Building overall study-a-thon group & methods development

## Fuel

### A study-a-thon can only be successful if you have:

Representative data curated / mapped in time for the study event  
Clinical input into the research question, methods  
Literature search and review  
Approvals and agreements

## Drive

### During a study-a-thon, some pivotal points:

Project managed itinerary with milestones  
A cadence of team / sub-team meetings  
Data Partner checks pre-event  
Communications  
Cascading work schedule

## Arrive

What was the pre-agree aim of the study-a-thon?

Abstract, manuscripts, answering the clinical questions

Ensuring post-event plan for study completion

Communications

Impact on therapeutic area?

## Post-event

Ensuring all work will be completed

Engagement with Data Partners, also additional package runs

Abstract / manuscript completion

Communications

Research plan milestones

**What did we learn?**