

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 #37

イブニング カンファレンス(第37回)

2022.12.27



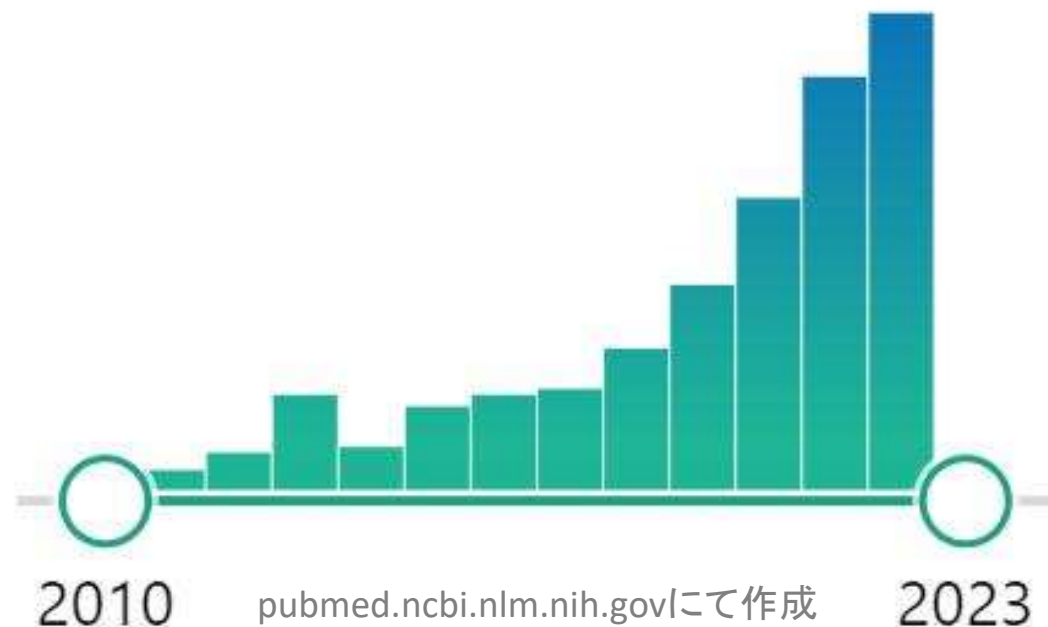
## 本日の内容

- OHDSI関連論文紹介
- 今月のOHDSI Community Call
- OHDSI冊子 “Our Journey 2022版”



# OHDSI関連論文

Pubmedで“OHDSI or OMOP”を検索



全期間累計：11月307本→12月310本



医薬品分子構造から副作用を予測する研究

> [Int J Mol Sci.](#) 2022 Dec 19;23(24):16216. doi: 10.3390/ijms232416216.

## iADRGSE: A Graph-Embedding and Self-Attention Encoding for Identifying Adverse Drug Reaction in the Earlier Phase of Drug Development

[Xiang Cheng](#)<sup>1</sup>, [Meiling Cheng](#)<sup>1</sup>, [Liyi Yu](#)<sup>1</sup>, [Xuan Xiao](#)<sup>1</sup>

Affiliations — collapse

### Affiliation

<sup>1</sup> Department of Computer, Jingdezhen Ceramic University, Jingdezhen 333403, China.

PMID: 36555858 DOI: [10.3390/ijms232416216](#)

## Abstract

Adverse drug reactions (ADRs) are a major issue to be addressed by the pharmaceutical industry. Early and accurate detection of potential ADRs contributes to enhancing drug safety and reducing financial expenses. The majority of the approaches that have been employed to identify ADRs are limited to determining whether a drug exhibits an ADR, rather than identifying the exact type of ADR. By introducing the "multi-level feature-fusion deep-learning model", a new predictor, called iADRGSE, has been developed, which can be used to identify adverse drug reactions at the early stage of drug discovery. iADRGSE integrates a self-attentive module and a graph-network module that can extract one-dimensional sub-structure sequence information and two-dimensional chemical-structure graph information of drug molecules. As a demonstration, cross-validation and independent testing were performed with iADRGSE on a dataset of ADRs classified into 27 categories, based on SOC (system organ classification). In addition, experiments comparing iADRGSE with approaches such as NPF were conducted on the OMOP dataset, using the jackknife test method. Experiments show that iADRGSE was superior to existing state-of-the-art predictors.

## 概要

医薬品副作用（ADR）は、製薬業界が取り組むべき大きな課題である。潜在的なADRを早期かつ正確に検出することは、医薬品の安全性を高め、経済的負担を軽減することに貢献する。しかし、これまでのADR検出手法の多くは、ADRの種類を特定するのではなく、ADRが発現しているかどうかを判定するものであった。iADRGSEは、自己学習モジュールとグラフネットワークモジュールを統合し、薬剤分子の1次元部分構造配列情報と2次元化学構造グラフ情報を抽出することができます。実証実験として、SOC（system organ classification）に基づき27のカテゴリに分類されたADRのデータセットに対して、iADRGSEを用いたクロスバリデーションと独立検定を実施した。また、OMOPデータセットにおいて、ジャックナイフテスト法を用いて、iADRGSEとNPFなどのアプローチとの比較実験を行った。実験の結果、iADRGSEは既存の最先端予測器よりも優れていることが確認された。

## 解説

医薬品設計の初期段階では医薬品の化学構造以外の情報はない。医薬品分子構造からADRを予測する研究のひとつ。ADRECS副作用データベースからADRを27カテゴリに分類し、マルチラベル分類問題として深層学習を行ったモデルを構築した。他の方法による予測と比較実験を実施し、比較優位な結果を得た。

※“OMOP”は、かつてのOMOPプロジェクトで整備した副作用データベースとして参照されていて、OMOP CDMとは関係なかった。



## OHDSIの総説論文

Review > [Korean Circ J. 2022 Dec;52\(12\):853-864. doi: 10.4070/kcj.2022.0294.](#)

# Establishment of an International Evidence Sharing Network Through Common Data Model for Cardiovascular Research

Seng Chan You <sup>1 2</sup>, Seongwon Lee <sup>3</sup>, Byungjin Choi <sup>3 4</sup>, Rae Woong Park <sup>3 5</sup>

Affiliations — collapse

### Affiliations

- 1 Department of Biomedical Systems Informatics, Yonsei University College of Medicine, Seoul, Korea.
- 2 Institute for Innovation in Digital Healthcare, Yonsei University, Seoul, Korea.
- 3 Department of Biomedical Informatics, Ajou University School of Medicine, Suwon, Korea.
- 4 Department of Biomedical Sciences, Ajou University Graduate School of Medicine, Suwon, Korea.
- 5 Department of Biomedical Sciences, Ajou University Graduate School of Medicine, Suwon, Korea. [veritas@ajou.ac.kr](mailto:veritas@ajou.ac.kr).

## Abstract

A retrospective observational study is one of the most widely used research methods in medicine. However, evidence postulated from a single data source likely contains biases such as selection bias, information bias, and confounding bias. Acquiring enough data from multiple institutions is one of the most effective methods to overcome the limitations. However, acquiring data from multiple institutions from many countries requires enormous effort because of financial, technical, ethical, and legal issues as well as standardization of data structure and semantics. The Observational Health Data Sciences and Informatics (OHDSI) research network standardized 928 million unique records or 12% of the world's population into a common structure and meaning and established a research network of 453 data partners from 41 countries around the world. OHDSI is a distributed research network wherein researchers do not own or directly share data but only analyzed results. However, sharing evidence without sharing data is difficult to understand. In this review, we will look at the basic principles of OHDSI, common data model, distributed research networks, and some representative studies in the cardiovascular field using the network. This paper also briefly introduces a Korean distributed research network named FeederNet.

## 概要

後方視的観察研究は、医学の分野で最も広く用いられている研究手法の一つである。しかし、単一のデータソースから推定されるエビデンスは、選択バイアス、情報バイアス、交絡バイアスなどのバイアスを含んでいる可能性が高い。この限界を克服するためには、複数の施設から十分なデータを取得することが最も有効な方法の一つである。しかし、多くの国の複数の機関からデータを取得するには、データ構造やセマンティクスの標準化だけでなく、財政的、技術的、倫理的、法的な問題から膨大な労力を必要とする。

OHDSI 研究ネットワークは、世界人口の12%にあたる9億2800万件のユニークレコードを共通の構造と意味に標準化し、世界41カ国の453のデータパートナーからなる研究ネットワークを確立した。

OHDSIは、研究者がデータを所有したり直接共有したりせず、分析結果のみを共有する分散型研究ネットワークである。しかし、データの共有なしにエビデンスを共有することは理解しがたい。

このレビューでは、OHDSIの基本原則、共通データモデル、分散型研究ネットワーク、そしてこのネットワークを利用した循環器分野の代表的な研究を見ていくことにする。また、FeederNet という韓国の分散型研究ネットワークについても簡単に紹介する。





## DoD病院とVA病院のOMOPデータ連携

> [Mil Med.](#) 2022 Dec 3;usac352. doi: 10.1093/milmed/usac352. Online ahead of print.

### Illustration of Continuous Enrollment and Beneficiary Categorization in DoD and VA Infrastructure for Clinical Intelligence

Veronika Pav<sup>1 2</sup>, Andrew Burns<sup>3</sup>, Courtney Colahan<sup>1</sup>, Brian Robison<sup>4 5</sup>, Jacob Kean<sup>4 5</sup>, Scott DuVall<sup>4 6</sup>

Affiliations – collapse

#### Affiliations

- <sup>1</sup> Kennell and Associates Inc., Falls Church, VA 22042, USA.
- <sup>2</sup> Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205, USA.
- <sup>3</sup> Alqimi Technology Solutions, Inc., Rockville, MD 20850, USA.
- <sup>4</sup> VA Informatics and Computing Infrastructure, VA Salt Lake City Health Care System, Salt Lake City, UT 84148, USA.
- <sup>5</sup> Department of Population Health Sciences, University of Utah School of Medicine, Salt Lake City, UT 84132, USA.
- <sup>6</sup> Department of Internal Medicine, University of Utah School of Medicine, Salt Lake City, UT 84132, USA.



## Abstract

**Introduction:** The DoD and VA Infrastructure for Clinical Intelligence (DaVINCI) data-sharing initiative has bridged the gap between DoD and VA data. DaVINCI utilizes the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) to map DoD and VA-specific health care codes to a standardized terminology. Although OMOP CDM provides a standardized longitudinal view of health care concepts, it fails in capturing multiple and changing relationships beneficiaries have with DoD and VA as it has a static (vs. yearly) person characteristic table. Furthermore, DoD and VA utilize different policies and terminology to identify their respective beneficiaries, which makes it difficult to track patients longitudinally. The primary purpose of this report is to provide a methodology for categorizing beneficiaries and creating continuous longitudinal patient records to maximize the use of the joint DoD and VA data in DaVINCI.

**Materials and methods:** For calendar year 2000-2020, we combined DoD and VA OMOP CDM and source databases to uniquely categorize beneficiaries into the following hierarchical groups: Active Duty, Guard, and Reserve Service Members (ADSMs); Separates; Retirees; Veterans; and Deceased. Once the cohorts were identified, we examined calendar year 2020 health care utilization data using the OMOP VISIT\_OCCURRENCE, DRUG\_EXPOSURE, MEASUREMENT, and PROCEDURE tables. We also used the Defense Enrollment and Eligibility Reporting System source table to derive enrollment periods for DoD beneficiaries. As VA does not have enrollment plans, we utilized the VA's priority groups (1-5) in the SPATIENT source table to crosswalk the DoD's enrollment concept to the VA. We then assessed lengths of continuous enrollments in DoD and VA and the impact of appending the longitudinal records together.

**Results:** The majority of the ADSMs utilized the DoD system, but about 60,557 (3%) were seen in the VA for varied types of care. The market share of care provided to ADSMs by the VA varied by specialty and location. For Retirees, the split between DoD (1,625,874 [75%]) and VA (895,992 [41%]) health care utilization was more significant. The value added for utilizing DaVINCI in longitudinal studies was the highest for researchers normally limited to DoD data only. For beneficiaries who had 5 years of continuous enrollment, DaVINCI increased the potential study population by over 202% compared to using DoD data alone and by over 14% compared to VA data alone. Among beneficiaries with 20 years of continuous enrollment, DaVINCI increased the potential study population by over 133% compared to DoD data and by nearly 39% compared to VA data.

**Conclusions:** DaVINCI has successfully combined DoD and VA data and utilized OMOP CDM to standardize health care concepts. However, to fully maximize the potential of DaVINCI's DoD and VA OMOP databases, researchers must uniquely categorize the DaVINCI cohort and build longitudinal patient records across DoD and VA. Because of the low other health insurance rates among DoD enrollees and their choice to enroll to a DoD Primary Care Manager, we believe this population to be the least censored in the DoD. Applying a similar concept through VA's priority groups (1-5) would enable researchers to follow ADSMs as they transition from the military.

## 概要

**はじめに:** DoD and VA Infrastructure for Clinical Intelligence (DaVINCI) データ共有構想は、DoD と VA のデータ間のギャップを埋めるものである。DaVINCIは、OMOP共通データモデル (CDM) を利用して、DoDとVA固有の医療コードを標準化された用語に対応づけている。OMOP CDM は医療概念の標準的な縦断的ビューを提供するが、受益者が DoD および VA と持つ複数の変化する関係を把握することはできない。さらに、DoD と VA は、それぞれの受益者を識別するために異なる方針と用語を使用しており、患者を縦断的に追跡することが困難である。本報告の主な目的は、DaVINCIにおけるDoDとVAの合同データを最大限に活用するために、受益者を分類し、継続的な縦断的患者記録を作成するための方法論を提供することである。

**材料と方法:** 2000 年から 2020 年の暦年について、DoD と VA の OMOP CDM とソースデータベースを組み合わせ、受益者を以下の階層的なグループに一意に分類した。現役兵 (ADSM)、予備役兵 (Separatees)、退職者 (Retirees)、退役軍人 (Veterans)、死亡者 (Deceased) である。コホートを特定した後、OMOP VISIT\_OCCURRENCE, DRUG\_EXPOSURE, MEASUREMENT, PROCEDURE 表を用いて、2020暦年の医療利用データを調査した。また、Defense Enrollment and Eligibility Reporting System ソーステーブルを使用して、DoD 受益者の登録期間を導出した。VA には登録計画がないため、SPATIENT ソーステーブルの VA の優先順位グループ (1-5) を利用し、DoD の登録概念と VA を相互参照した。次に、DoDとVAにおける継続的な登録の長さ、縦断的な記録を併記することの影響を評価した。

**結果:** ADSMの大多数は国防総省のシステムを利用していたが、約60,557人 (3%) がさまざまな種類のケアのためにVAを受診していた。VAがADSMに提供するケアの市場シェアは、専門分野と場所によって異なる。退職者については、DoD (1,625,874人 [75%]) とVA (895,992人 [41%]) の医療利用がより大きく分かれた。縦断的研究におけるDaVINCI活用の付加価値は、通常DoDデータのみで限定される研究者にとって最も高いものであった。5年間継続して登録している受益者については、DoDのデータのみを使用した場合と比較して、DaVINCIは研究対象者を202%以上、VAのデータのみを使用した場合と比較して、14%以上増加させることができた。20年間継続して登録されている受益者では、DaVINCIはDoDデータと比較して133%以上、VAデータと比較して約39%、潜在的な研究対象者を増加させた。

**結論:** DaVINCIはDoDとVAのデータを統合し、OMOP CDMを活用して医療コンセプトを標準化することに成功した。しかし、DaVINCIのDoDおよびVA OMOPデータベースの可能性を十分に引き出すためには、研究者はDaVINCIコホートを独自に分類し、DoDとVAを横断する長期的な患者記録を構築する必要がある。国防総省の登録者は他の健康保険料率が低く、国防総省のプライマリーケアマネージャーへの登録を選択しているため、この集団は国防総省の中で最も検閲が緩いと考えられる。VA の優先順位グループ (1-5) を通して同様の概念を適用すれば、研究者は ADSM が軍から移行する際に追跡調査を行うことができるだろう。

VA 退役軍人局: 1,293の医療施設で900万人以上の退役軍人にケアを提供。米国最大の病院系列  
DoD 国防総省: 700以上の軍医療施設と民間契約で年間約960万人に医療を提供  
ADSM: 現役/予備役の軍籍者



# 今月のCommunity Call

## ●APAC Community Call テーマ

Dec. 1 APAC Symposium Recap

Dec. 15 APAC Year In Review

## ●Global Community Call テーマ

Nov. 29 Workgroup Updates

Medical Devices, Clinical Trials, Patient-Level Prediction, Psychiatry

Dec. 6 Recent OHDSI Publications

- Comparative risk of thrombosis with thrombocytopenia syndrome or thromboembolic events associated with different covid-19 vaccines
- Transforming and evaluating the UK Biobank to the OMOP Common Data Model for COVID-19 research and beyond
- Adjusting for indirectly measured confounding using large-scale propensity score
- PheValuator 2.0: Methodological improvements for the PheValuator approach to semi-automated phenotype algorithm evaluation

Dec. 13 2022 In Review

Dec. 20 Holiday-Themed Final Call of 2022



## 2022 OKR



**Objective 1:** Enable the device standardization efforts to be interoperable with OMOP to support large scale device data analysis

### 2022 Key Results:

1. Survey the current practice within OHDSI community for device specific data and device generated data for real world evidence generation
2. Define methods to assess and harmonize standards to be included in OMOP vocabulary for device data

Objective 2 in dev:

## Device areas

1. General surgery
2. Cardiovascular devices
3. Orthopedics (Spine)
4. Orthopedics (hips and knees)
5. Neurological devices ( aneurysm, stroke)
6. Vision, lens
7. Breast implant
8. Robotic devices
9. Oncology surgery
10. Mesh for hernia repair

## Interest topics

Is there a need to extend the CDM or Vocabulary?

# Workgroup: Clinical Trials (Tom Walpole)

## Workgroup Objectives

Transform CDISC SDTM Formatted Clinical Trials Data to OMOP CDM

- Propose Conventions
- Generate Mappings
- Publish Guidelines

## Challenges

- Data are currently collected according to CDISC SDTM Standard and require ETL definitions/mappings to convert into OMOP CDM
- Privacy concerns make it difficult to access data for testing
- Clinical Trials data are specialized type of medical record often requiring custom vocabularies

## Our Approach

- Document Proposed Conventions
  - <https://github.com/OHDSI/ClinicalTrialsWGETL/wiki>
- Collect a set of SDTM formatted source data that adequately represents a breadth of data elements
  - Across different Phases and Therapeutic Areas
- Data Sources
  - PhUSE
  - Vivli
  - CPath
  - TrialTwin
- Produce ETL definitions and field mappings for source data
  - <https://github.com/OHDSI/ClinicalTrialsWGETL/tree/master/docs>
- Proof of Concept
  - Perform data transformations using definitions & mappings in a protected environment



## Current research projects

- Improving model transportability (calibration)
- Interpreting external validation
- Empirical evaluation into best practices for patient-level prediction
- Model updating
- Counter-factual prediction
- Comparison of deep learning methods
- Comparison of regularization methods





## Workgroup: Psychiatry (Dmytry Dymshyts)

- Standards: Vocabulary development – initial phase, done not systematically

Initial set of around 280 scales added to the OMOP vocabulary

- Standards: Vocabulary development – second phase: comprehensive and systematic approach
- Standards: explore questionnaires in a RW data

---

### Ongoing Study

- [Predicting Adverse Drug Events during Methylphenidate Treatment in ADHD : International Collaborative Network Study](#)

### Previously run studies

- *The risk of schizophrenia in ACE inhibitor users compared with users of other antihypertensives*
- Covid-19 pandemic impacts on mental health Related conditions Via multi-database network: a Longitudinal Observational study (CERVELLO)



# Our Journey

Where The OHDSI Community Has Been  
And Where We Are Going

*2022 edition*



# OHDSI

OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS