

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 #38

イブニングカンファレンス(第38回)

2023.1.31



## 本日の内容

- OHDSI関連論文紹介
- 今月のOHDSI Community Call
- Eunomia由来の日本版疑似データ



# OHDSI関連論文

Pubmedで“OHDSI or OMOP”を検索



全期間累計：12月310本→1月313本



## 非構造化処方箋データの構造化

› [JMIR Med Inform. 2023 Jan 25;11:e40312. doi: 10.2196/40312.](#)

# Assessment and Improvement of Drug Data Structuredness From Electronic Health Records: Algorithm Development and Validation

Ines Reinecke<sup>1</sup>, Joscha Siebel<sup>1</sup>, Saskia Fuhrmann<sup>2 3</sup>, Andreas Fischer<sup>3</sup>, Martin Sedlmayr<sup>2</sup>, Jens Weidner<sup>1</sup>, Franziska Bathelt<sup>1</sup>

Affiliations – collapse

### Affiliations

- <sup>1</sup> Institute for Medical Informatics and Biometry, Carl Gustav Carus Faculty of Medicine, Technische Universität Dresden, Dresden, Germany.
- <sup>2</sup> Center for Evidence-Based Healthcare, Carl Gustav Carus Faculty of Medicine, Technische Universität Dresden, Dresden, Germany.
- <sup>3</sup> Hospital Pharmacy, University Hospital Carl Gustav Carus, Dresden, Germany.

PMID: 36696159 DOI: [10.2196/40312](#)

## Abstract

**Background:** Digitization offers a multitude of opportunities to gain insights into current diagnostics and therapies from retrospective data. In this context, real-world data and their accessibility are of increasing importance to support unbiased and reliable research on big data. However, routinely collected data are not readily usable for research owing to the unstructured nature of health care systems and a lack of interoperability between these systems. This challenge is evident in drug data.

**Objective:** This study aimed to present an approach that identifies and increases the structuredness of drug data while ensuring standardization according to Anatomical Therapeutic Chemical (ATC) classification.

**Methods:** Our approach was based on available drug prescriptions and a drug catalog and consisted of 4 steps. First, we performed an initial analysis of the structuredness of local drug data to define a point of comparison for the effectiveness of the overall approach. Second, we applied 3 algorithms to unstructured data that translated text into ATC codes based on string comparisons in terms of ingredients and product names and performed similarity comparisons based on Levenshtein distance. Third, we validated the results of the 3 algorithms with expert knowledge based on the 1000 most frequently used prescription texts. Fourth, we performed a final validation to determine the increased degree of structuredness.

**Results:** Initially, 47.73% (n=843,980) of 1,768,153 drug prescriptions were classified as structured. With the application of the 3 algorithms, we were able to increase the degree of structuredness to 85.18% (n=1,506,059) based on the 1000 most frequent medication prescriptions. In this regard, the combination of algorithms 1, 2, and 3 resulted in a correctness level of 100% (with 57,264 ATC codes identified), algorithms 1 and 3 resulted in 99.6% (with 152,404 codes identified), and algorithms 1 and 2 resulted in 95.9% (with 39,472 codes identified).

**Conclusions:** As shown in the first analysis steps of our approach, the availability of a product catalog to select during the documentation process is not sufficient to generate structured data. Our 4-step approach reduces the problems and reliably increases the structuredness automatically. Similarity matching shows promising results, particularly for entries with no connection to a product catalog. However, further enhancement of the correctness of such a similarity matching algorithm needs to be investigated in future work.

## 概要

**背景** デジタル化により、レトロスペクティブなデータから現在の診断や治療法に関する洞察を得る機会が数多く提供されている。このような背景から、ビッグデータを用いた偏りのない信頼性の高い研究を支援するために、実世界のデータとそのアクセスの重要性が高まっている。しかし、日常的に収集されるデータは、医療システムの非構造化やシステム間の相互運用性の欠如により、研究に利用することは容易ではありません。このような課題は、薬物データにおいても顕著である。

**目的** 本研究では、ATC (Anatomical Therapeutic Chemical) 分類に基づく標準化を確保しつつ、医薬品データの構造化を特定・向上させるアプローチを提示することを目的とした。

**方法** 本研究では、入手可能な処方箋と医薬品カタログに基づき、4つのステップから構成されるアプローチを行った。まず、ローカルな薬剤データの構造化を分析し、アプローチ全体の有効性を比較するポイントを定義した。第二に、非構造化データに対して、成分や商品名などの文字列比較によるATCコードへの変換、レーベンシュタイン距離（編集距離）による類似度比較の3つのアルゴリズムを適用しました。第三に、3つのアルゴリズムの結果を、最も頻繁に使用される処方箋テキスト1000件に基づく専門家の知識で検証した。第四に、最終的な検証を行い、構造化の度合いの増加を判定した。

**結果** 当初、1,768,153件の処方箋のうち、47.73% (n=843,980) が構造化された処方箋と分類された。3つのアルゴリズムを適用した結果、最も頻度の高い1000件の薬剤処方箋をもとに、構造化度を85.18% (n=1,506,059) にまで高めることができた。また、アルゴリズム1, 2, 3の組み合わせにより、正答率100% (ATCコード57,264件) , アルゴリズム1, 3により99.6% (同152,404件) , アルゴリズム1, 2により95.9% (同39,472件) を達成した。

**結論** 本アプローチの最初の分析ステップで示したように、文書化プロセスで選択する製品カタログが利用可能なだけでは、構造化データを生成するのに十分ではありません。我々の4ステップのアプローチは、問題を軽減し、確実に構造化度を自動的に増加させる。類似性マッチングは、特に製品カタログと関連性のないエントリに対して有望な結果を示している。しかし、このような類似性マッチングアルゴリズムの正しさのさらなる強化は、今後の研究課題である。



医療機器の実環境評価のための  
OMOP CDM

> [JAMIA Open. 2023 Jan 10;6\(1\):ooac108. doi: 10.1093/jamiaopen/ooac108. eCollection 2023 Apr.](#)

**Integrating real-world data to assess cardiac ablation device outcomes in a multicenter study using the OMOP common data model for regulatory decisions: implementation and evaluation**

Yue Yu <sup>1</sup>, Guoqian Jiang <sup>2</sup>, Eric Brandt <sup>3</sup>, Tom Forsyth <sup>3</sup>, Sanket S Dhruva <sup>4</sup>, Shumin Zhang <sup>5</sup>, Jiajing Chen <sup>3</sup>, Peter A Noseworthy <sup>6</sup>, Amit A Doshi <sup>7</sup>, Kimberly Collison-Farr <sup>3</sup>, Dure Kim <sup>8</sup>, Joseph S Ross <sup>9</sup>, Paul M Coplan <sup>5 10</sup>, Joseph P Drozda Jr <sup>3</sup>

Affiliations + expand

PMID: 36632328 PMID: PMC9831049 DOI: 10.1093/jamiaopen/ooac108

## Abstract

The objective of this study is to describe application of the Observational Medical Outcomes Partnership (OMOP) common data model (CDM) to support medical device real-world evaluation in a National Evaluation System for health Technology Coordinating Center (NESTcc) Test-Case involving 2 healthcare systems, Mercy Health and Mayo Clinic. CDM implementation was coordinated across 2 healthcare systems with multiple hospitals to aggregate both medical device data from supply chain databases and patient outcomes and covariates from electronic health record data. Several data quality assurance (QA) analyses were implemented on the OMOP CDM to validate the data extraction, transformation, and load (ETL) process. OMOP CDM-based data of relevant patient encounters were successfully established to support studies for FDA regulatory submissions. QA analyses verified that the data transformation was robust between data sources and OMOP CDM. Our efforts provided useful insights in real-world data integration using OMOP CDM for medical device evaluation coordinated across multiple healthcare systems.

## 概要

本研究の目的は、医療機器の実環境評価を支援するための OMOP CDM の適用について、Mercy HealthとMayo Clinicという2つの医療機関が関与するNESTcc（National Evaluation System for Health Technology Coordinating Center）テストケースで説明することである。CDMの実装は、複数の病院を持つ2つの医療システム間で調整され、サプライチェーンデータベースからの医療機器データ、電子カルテデータからの患者の転帰と共変量の両方を集約した。データ抽出、変換、ロード（ETL）プロセスを検証するため、いくつかのデータ品質保証（QA）分析が OMOP CDM 上で実施された。OMOP CDM に基づいて関連する患者との面会データを確立し、FDA の規制当局に申請するための研究を支援することに成功した。QA分析により、データソースとOMOP CDMの間でデータ変換が確実に行われることが検証されました。我々の取り組みは、複数の医療システム間で調整された医療機器評価のためのOMOP CDMを用いた実世界のデータ統合に有用な洞察を提供した。





## OMOPをプロセスマイニングに活用する

Observational Study > PLoS One. 2023 Jan 3;18(1):e0279641. doi: 10.1371/journal.pone.0279641.

eCollection 2023.

# Exploring the potential of OMOP common data model for process mining in healthcare

Kangah Park<sup>1</sup>, Minsu Cho<sup>2</sup>, Minseok Song<sup>1</sup>, Sooyoung Yoo<sup>3</sup>, Hyunyoung Baek<sup>3</sup>, Seok Kim<sup>3</sup>, Kidong Kim<sup>4</sup>

Affiliations – collapse

### Affiliations

- 1 Department of Industrial and Management Engineering, Pohang University of Science and Technology (POSTECH), Pohang, South Korea.
- 2 School of Information Convergence, Kwangwoon University, Seoul, South Korea.
- 3 Healthcare ICT Research Center, Office of eHealth Research and Businesses, Seoul National University Bundang Hospital, Seongnam, South Korea.
- 4 Department of Obstetrics and Gynecology, Seoul National University Bundang Hospital, Seongnam, South Korea.

## Abstract

**Background and objective:** Recently, Electronic Health Records (EHR) are increasingly being converted to Common Data Models (CDMs), a database schema designed to provide standardized vocabularies to facilitate collaborative observational research. To date, however, rare attempts exist to leverage CDM data for healthcare process mining, a technique to derive process-related knowledge (e.g., process model) from event logs. This paper presents a method to extract, construct, and analyze event logs from the Observational Medical Outcomes Partnership (OMOP) CDM for process mining and demonstrates CDM-based healthcare process mining with several real-life study cases while answering frequently posed questions in process mining, in the CDM environment.

**Methods:** We propose a method to extract, construct, and analyze event logs from the OMOP CDM for process types including inpatient, outpatient, emergency room processes, and patient journey. Using the proposed method, we extract the retrospective data of several surgical procedure cases (i.e., Total Laparoscopic Hysterectomy (TLH), Total Hip Replacement (THR), Coronary Bypass (CB), Transcatheter Aortic Valve Implantation (TAVI), Pancreaticoduodenectomy (PD)) from the CDM of a Korean tertiary hospital. Patient data are extracted for each of the operations and analyzed using several process mining techniques.

**Results:** Using process mining, the clinical pathways, outpatient process models, emergency room process models, and patient journeys are demonstrated using the extracted logs. The result shows CDM's usability as a novel and valuable data source for healthcare process analysis, yet with a few considerations. We found that CDM should be complemented by different internal and external data sources to address the administrative and operational aspects of healthcare processes, particularly for outpatient and ER process analyses.

**Conclusion:** To the best of our knowledge, we are the first to exploit CDM for healthcare process mining. Specifically, we provide a step-by-step guidance by demonstrating process analysis from locating relevant CDM tables to visualizing results using process mining tools. The proposed method can be widely applicable across different institutions. This work can contribute to bringing a process mining perspective to the existing CDM users in the changing Hospital Information Systems (HIS) environment and also to facilitating CDM-based studies in the process mining research community.

## 概要

**背景と目的:** 近年、電子カルテ（EHR）の共通データモデル（CDM）への変換が進んでいる。CDMは、共同観察研究を促進するために標準化された語彙を提供するデータベーススキーマとして設計されている。しかし、イベントログからプロセスに関する知識（プロセスモデルなど）を抽出する手法である**ヘルスケアプロセスマイニング**にCDMデータを活用する試みは、これまでほとんど行われていない。本論文では、プロセスマイニングのためにOMOP CDMからイベントログを抽出・構築・分析する方法を紹介し、プロセスマイニングでよくある質問に答えながら、いくつかの実際の研究事例を用いてCDMベースの医療プロセスマイニングをCDM環境下で実証しています。

**方法:** OMOP CDMから、入院、外来、救急、患者旅程などのイベントログを抽出・構築し、分析する手法を提案する。提案手法を用いて、韓国の三次病院のCDMから、複数の外科手術症例（腹腔鏡下子宮全摘術（TLH）、股関節全置換術（THR）、冠動脈バイパス術（CB）、経カテーテル大動脈弁移植術（TAVI）、膵臓十二指腸切除術（PD）など）のレトロスペクティブデータを抽出する。各手術について患者データを抽出し、いくつかのプロセスマイニング技術を用いて分析した。

**結果:** プロセスマイニングにより、抽出されたログを用いて、クリニカルパス、外来プロセスモデル、救急室プロセスモデル、患者ジャーニーを実証した。この結果は、CDMが医療プロセス分析のための新規かつ貴重なデータソースとして有用であることを示しているが、一方でいくつかの考慮すべき点もある。特に外来患者や救命救急センターのプロセス分析において、医療プロセスの管理・運用の側面を扱うためには、CDMを様々な内外のデータソースで補完する必要があることが分かった。

**結論:** 我々の知る限り、CDMをヘルスケアプロセスマイニングに活用したのは我々が初めてである。具体的には、関連するCDMテーブルの検索からプロセスマイニングツールを用いた結果の可視化まで、プロセス分析を実演しながらステップバイステップのガイダンスを提供する。提案する手法は、様々な医療機関に広く適用することができる。本研究は、変化する病院情報システム（HIS）環境において、既存のCDMユーザーにプロセスマイニングの視点をもたらすとともに、プロセスマイニング研究コミュニティにおけるCDMベースの研究を促進することに貢献することができる。



## Long COVIDを症状類似性でクラスタ分けした

➤ EBioMedicine. 2023 Jan;87:104413. doi: 10.1016/j.ebiom.2022.104413. Epub 2022 Dec 21.

### Generalisable long COVID subtypes: findings from the NIH N3C and RECOVER programmes

Justin T Reese<sup>1</sup>, Hannah Blau<sup>2</sup>, Elena Casiraghi<sup>3</sup>, Timothy Bergquist<sup>4</sup>, Johanna J Loomba<sup>5</sup>, Tiffany J Callahan<sup>6</sup>, Bryan Laraway<sup>7</sup>, Corneliu Antonescu<sup>8</sup>, Ben Coleman<sup>2</sup>, Michael Gargano<sup>2</sup>, Kenneth J Wilkins<sup>9</sup>, Luca Cappelletti<sup>10</sup>, Tommaso Fontana<sup>10</sup>, Nariman Ammar<sup>11</sup>, Blessy Antony<sup>12</sup>, T M Murali<sup>12</sup>, J Harry Caufield<sup>1</sup>, Guy Karlebach<sup>2</sup>, Julie A McMurry<sup>7</sup>, Andrew Williams<sup>13</sup>, Richard Moffitt<sup>14</sup>, Jineta Banerjee<sup>4</sup>, Anthony E Solomonides<sup>15</sup>, Hannah Davis<sup>16</sup>, Kristin Kostka<sup>17</sup>, Giorgio Valentini<sup>10</sup>, David Sahner<sup>18</sup>, Christopher G Chute<sup>19</sup>, Charisse Madlock-Brown<sup>11</sup>, Melissa A Haendel<sup>7</sup>, Peter N Robinson<sup>20</sup>; N3C Consortium; RECOVER Consortium

Collaborators, Affiliations + expand

PMID: 36563487 PMID: PMC9769411 DOI: 10.1016/j.ebiom.2022.104413

## Abstract

**Background:** Stratification of patients with post-acute sequelae of SARS-CoV-2 infection (PASC, or long COVID) would allow precision clinical management strategies. However, long COVID is incompletely understood and characterised by a wide range of manifestations that are difficult to analyse computationally. Additionally, the generalisability of machine learning classification of COVID-19 clinical outcomes has rarely been tested.

**Methods:** We present a method for computationally modelling PASC phenotype data based on electronic healthcare records (EHRs) and for assessing pairwise phenotypic similarity between patients using semantic similarity. Our approach defines a nonlinear similarity function that maps from a feature space of phenotypic abnormalities to a matrix of pairwise patient similarity that can be clustered using unsupervised machine learning.

**Findings:** We found six clusters of PASC patients, each with distinct profiles of phenotypic abnormalities, including clusters with distinct pulmonary, neuropsychiatric, and cardiovascular abnormalities, and a cluster associated with broad, severe manifestations and increased mortality. There was significant association of cluster membership with a range of pre-existing conditions and measures of severity during acute COVID-19. We assigned new patients from other healthcare centres to clusters by maximum semantic similarity to the original patients, and showed that the clusters were generalisable across different hospital systems. The increased mortality rate originally identified in one cluster was consistently observed in patients assigned to that cluster in other hospital systems.

**Interpretation:** Semantic phenotypic clustering provides a foundation for assigning patients to stratified subgroups for natural history or therapy studies on PASC.

## 概要

**背景** SARS-CoV-2感染後の急性後遺症（PASCまたはlong COVID）を持つ患者の層別化は、精密な臨床管理戦略を可能にするであろう。しかし、long COVIDの理解は不完全であり、計算機による解析が困難な幅広い症状によって特徴づけられる。さらに、COVID-19の臨床転帰の機械学習による分類の一般性は、これまでほとんど検証されていない。

**方法** 我々は、電子カルテ（EHR）に基づくPASC表現型データを計算でモデル化し、意味的類似性を用いて患者間のペアワイズ表現型の類似性を評価する方法を提示する。この手法は、非線形類似性関数を定義し、表現型異常の特徴空間から、教師なし機械学習を用いてクラスタリング可能なペアワイズ患者類似性の行列にマッピングするものである。

**発見** PASC患者には6つのクラスタが存在し、それぞれが異なる表現型異常のプロファイルを有していた。その中には、肺、神経精神、心血管系の異常が異なるクラスタと、広範囲で重度の症状および死亡率の上昇と関連するクラスタが含まれていた。COVID-19の急性期において、クラスタへの加入は、さまざまな既往症や重症度の指標と有意な関連を示した。他の医療センターからの新規患者を、元の患者との最大限の意味的類似性によってクラスタに割り当て、クラスタが異なる病院システム間で一般化可能であることを明らかにした。あるクラスタでもともと確認されていた死亡率の上昇は、他の病院システムでそのクラスタに割り当てられた患者でも一貫して観察された。

**解釈** 意味論的表現型クラスタリングは、PASCの自然史研究または治療研究のために、患者を層別化されたサブグループに割り当てるための基礎を提供するものである。



# 今月のCommunity Call

- APAC Community Call テーマ

- ・ 今年は今月1回開催になります。2月～5月はTutorialを行います。

<https://ohdsi.org/apac2021-2022/>

- Global Community Call テーマ

Jan. 10 Where Can OHDSI Go in 2023?

Jan. 17 OHDSI Speed Dating  
multiple sessions with different breakout rooms

Jan. 24 Collaborations for Strategic Priorities  
3つの重点項目

- Standardized vocabularies
- Standardized data network
- Collaborating on network studies: “The Sisyphus Challenge”



# Eunomia由来の日本版疑似データ

## Eunomiaとは

OHDSIでのテストデータ。

Rのパッケージになっている。

<https://github.com/OHDSI/Eunomia>

独立してCRANに登録されていたが今はなくなって、HADESの一部になった。

```
library(Eunomia)
connectionDetails <- getEunomiaConnectionDetails()
connection <- connect(connectionDetails)
querySql(connection, "SELECT COUNT(*) FROM person;")
# COUNT(*)
#1    2694

getTableNames(connection, databaseSchema = 'main')
disconnect(connection)
```

SQLiteのDB