

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 #42

イブニング カンファレンス(第42回)

2023.5.29



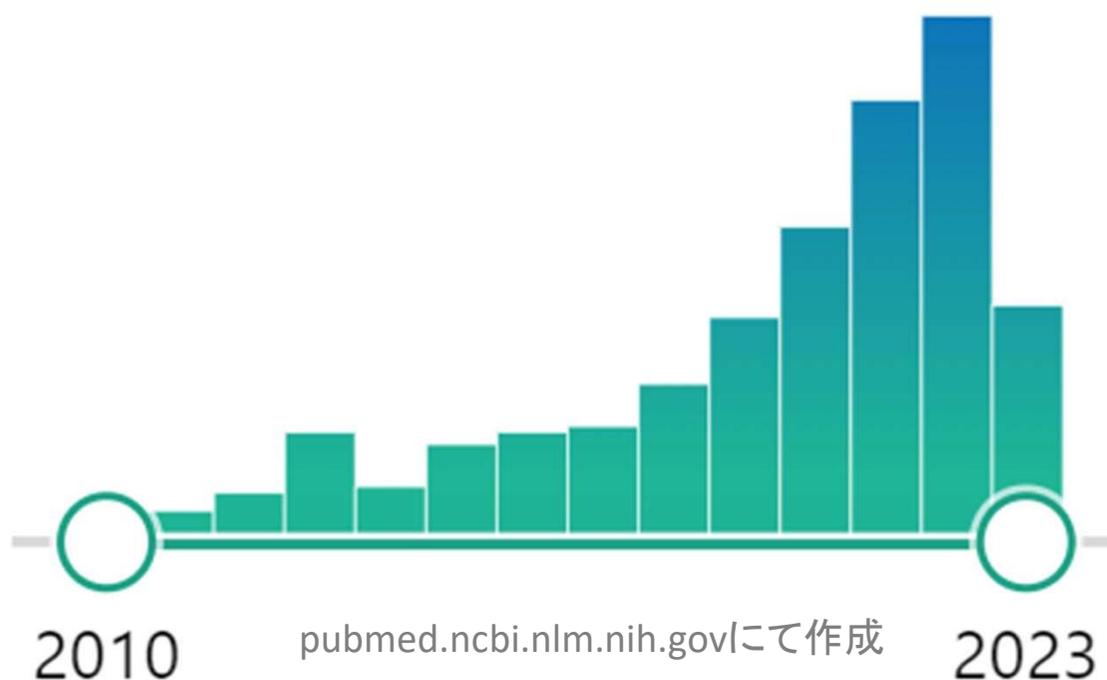
本日の内容

- OHDSI関連論文紹介
- 今月のOHDSI Community Call
- SOS Challenge Tutorial まとめ



OHDSI関連論文

Pubmedで“OHDSI or OMOP”を検索



全期間累計：4月323本→4月340本

- **Clinical encounter heterogeneity and methods for resolving in networked EHR data: a study from N3C and RECOVER programs.** J Am Med Inform Assoc. 2023 May 19;30(6):1125-1136. PMID: 37087110.
- **Phenotyping in distributed data networks: selecting the right codes for the right patients.** AMIA Annu Symp Proc. 2023 Apr 29;2022:826-835. PMID: 37128407.
- **Hierarchy-aware Adverse Reaction Embeddings for Signal Detection.** AMIA Annu Symp Proc. 2023 Apr 29;2022:596-605. PMID: 37128452.
- **Feasibility of Linking Area Deprivation Index Data to the OMOP Common Data Model.** AMIA Annu Symp Proc. 2023 Apr 29;2022:587-595. PMID: 37128466.
- **How to Improve Cancer Patients Enrollment in Clinical Trials From Real-Life Databases Using the Observational Medical Outcomes Partnership Oncology Extension: Results of the PENELOPE Initiative in Urologic Cancers.** JCO Clin Cancer Inform. 2023 May;7:e2200179. PMID: 37167578.
- **Transforming Documents of the Austrian Nationwide EHR System into the OMOP CDM.** Stud Health Technol Inform. 2023 May 2;301:54-59. PMID: 37172152.
- **Standardized Database of 12-Lead Electrocardiograms with a Common Standard for the Promotion of Cardiovascular Research: KURIAS-ECG.** Healthc Inform Res. 2023 Apr;29(2):132-144. Epub 2023 Apr 30. PMID: 37190737.
- **Feasibility Study of Federated Learning on the Distributed Research Network of OMOP Common Data Model.** Healthc Inform Res. 2023 Apr;29(2):168-173. Epub 2023 Apr 30. PMID: 37190741.
- **Fitness for Use of Anatomical Therapeutic Classification for Real World Data Research.** Stud Health Technol Inform. 2023 May 18;302:711-715. PMID: 37203475.
- **Towards ETL Processes to OMOP CDM Using Metadata and Modularization.** Stud Health Technol Inform. 2023 May 18;302:751-752. PMID: 37203486.
- **Modeling Patient Treatment Trajectories Using Markov Chains for Cost Analysis.** Stud Health Technol Inform. 2023 May 18;302:755-756. PMID: 37203488.
- **Converting HL7 CDA Based Nationwide Austrian Medication Data to OMOP CDM.** Stud Health Technol Inform. 2023 May 18;302:899-900. PMID: 37203528.
- **German Claims Data for Real-World Research: Content Coverage Evaluation in OMOP CDM.** Stud Health Technol Inform. 2023 May 18;302:3-7. PMID: 37203598.
- **Data Quality Assessment for Observational Medical Outcomes Partnership Common Data Model of Multi-Center.** Stud Health Technol Inform. 2023 May 18;302:322-326. PMID: 37203671.
- **SEnDAE: A resource for expanding research into social and environmental determinants of health.** Comput Methods Programs Biomed. 2023 Apr 8;238:107542. PMID: 37224727
- **Ontologizing health systems data at scale: making translational discovery a reality.** NPJ Digit Med. 2023 May 19;6(1):89. PMID: 37208468.
- **RD-MON - Building a Rare Disease Monitor to Enhance Awareness for Patients with Rare Diseases in Intensive Care.** Stud Health Technol Inform. 2023 May 18;302:358-359. PMID: 37203683.



受診歴の同質化

> [J Am Med Inform Assoc. 2023 May 19;30\(6\):1125-1136. doi: 10.1093/jamia/ocad057.](#)

Clinical encounter heterogeneity and methods for resolving in networked EHR data: a study from N3C and RECOVER programs

Peter Leese¹, Adit Anand², Andrew Girvin³, Amin Manna³, Saaya Patel², Yun Jae Yoo², Rachel Wong², Melissa Haendel⁴, Christopher G Chute⁵, Tellen Bennett⁶, Janos Hajagos², Emily Pfaff⁷, Richard Moffitt^{2 8 9}

PMID: 37087110 PMCID: PMC10198518

Affiliations

- 1 NC TraCS Institute, UNC-School of Medicine, Chapel Hill, North Carolina, USA.
- 2 Department of Biomedical Informatics, Stony Brook University, Stony Brook, New York, USA.
- 3 Palantir Technologies, Denver, Colorado, USA.
- 4 Department of Biomedical Informatics, University of Colorado Anschutz Medical Campus, Denver, Colorado, USA.
- 5 Schools of Medicine, Public Health, and Nursing, Johns Hopkins University, Baltimore, Maryland, USA.
- 6 Department of Pediatrics, University of Colorado Anschutz Medical Campus, Denver, Colorado, USA.
- 7 Department of Medicine, UNC Chapel Hill, Chapel Hill, North Carolina, USA.
- 8 Department of Biomedical Informatics, Emory University, Atlanta, Georgia, USA.
- 9 Department of Hematology and Medical Oncology, Emory University, Atlanta, Georgia, USA.

Abstract

Objective: Clinical encounter data are heterogeneous and vary greatly from institution to institution. These problems of variance affect interpretability and usability of clinical encounter data for analysis. These problems are magnified when multisite electronic health record (EHR) data are networked together. This article presents a novel, generalizable method for resolving encounter heterogeneity for analysis by combining related atomic encounters into composite "macrovisits."

Materials and methods: Encounters were composed of data from 75 partner sites harmonized to a common data model as part of the NIH Researching COVID to Enhance Recovery Initiative, a project of the National Covid Cohort Collaborative. Summary statistics were computed for overall and site-level data to assess issues and identify modifications. Two algorithms were developed to refine atomic encounters into cleaner, analyzable longitudinal clinical visits.

Results: Atomic inpatient encounters data were found to be widely disparate between sites in terms of length-of-stay (LOS) and numbers of OMOP CDM measurements per encounter. After aggregating encounters to macrovisits, LOS and measurement variance decreased. A subsequent algorithm to identify hospitalized macrovisits further reduced data variability.

Discussion: Encounters are a complex and heterogeneous component of EHR data and native data issues are not addressed by existing methods. These types of complex and poorly studied issues contribute to the difficulty of deriving value from EHR data, and these types of foundational, large-scale explorations, and developments are necessary to realize the full potential of modern real-world data.

Conclusion: This article presents method developments to manipulate and resolve EHR encounter data issues in a generalizable way as a foundation for future research and analysis.

概要

目的: 臨床のエンカウンター（受診）データは不均質であり、施設によって大きく異なる。このような差異の問題は、臨床受診データの解釈可能性や分析への利用可能性に影響を与える。これらの問題は、多施設の電子カルテ（EHR）データがネットワークで結ばれている場合に拡大される。本稿では、関連する個々の受診を複合的な“マクロビジット”に組み合わせることによって、分析のための受診の異質性を解決する、一般化可能な新しい方法を紹介します。

材料と方法: エンカウンターは、National Covid Cohort Collaborative(N3C)のプロジェクトであるNIH Researching COVID to Enhance Recovery Initiativeの一環として、共通のデータモデルに調和された75のパートナーサイトからのデータで構成されていた。問題点を評価し、修正を特定するために、全体およびサイトレベルのデータについて要約統計が計算された。2つのアルゴリズムを開発し、個々の受診をよりクリーンで分析可能な縦断的な臨床受診に改良した。

結果: 個々の受診のデータは、滞在期間（LOS）および1回あたりのOMOP CDM測定数において、サイト間で大きな格差があることが判明した。マクロビジットに集計したところ、LOSと測定値のばらつきは減少した。さらに、入院中のマクロビジットを特定するアルゴリズムを導入することで、データのばらつきがさらに小さくなった。

考察: エンカウンターはEHRデータの複雑で異質な構成要素であり、ネイティブデータの問題は既存の手法では対処できないものである。このような種類の複雑で研究が不十分な問題は、EHRデータから価値を引き出すことの難しさの一因であり、現代の実世界データの可能性を最大限に実現するためには、このような種類の基礎的で大規模な探索と開発が必要である。

結論: 本稿では、今後の研究・分析の基礎として、EHRエンカウンターデータの問題を一般化可能な方法で操作・解決するための手法開発を紹介する。



PHOEBE
コンセプトセットを作るツール

› [AMIA Annu Symp Proc. 2023 Apr 29;2022:826-835. eCollection 2022.](#)

Phenotyping in distributed data networks: selecting the right codes for the right patients

[Anna Ostroplets](#)¹, [Patrick Ryan](#)^{1 2}, [George Hripcsak](#)^{1 3}

Affiliations – collapse

Affiliations

- 1 Columbia University, New York, NY, USA.
- 2 Janssen Research and Development, Titusville, NJ.
- 3 New York-Presbyterian Hospital, New York, NY.

PMID: 37128407 PMCID: [PMC10148372](#)

Abstract

Observational data can be used to conduct drug surveillance and effectiveness studies, investigate treatment pathways, and predict patient outcomes. Such studies require developing executable algorithms to find patients of interest or phenotype algorithms. Creating reliable and comprehensive phenotype algorithms in data networks is especially hard as differences in patient representation and data heterogeneity must be considered. In this paper, we discuss a process for creating a comprehensive concept set and a recommender system we built to facilitate it. PHenotype Observed Entity Baseline Endorsements (PHOEBE) uses the data on code utilization across 22 electronic health record and claims datasets mapped to the Observational Health Data Sciences and Informatics (OHDSI) Common Data Model from the 6 countries to recommend semantically and lexically similar codes. Coupled with Cohort Diagnostics, it is now used in major network OHDSI studies. When used to create patient cohorts, PHOEBE identifies more patients and captures them earlier in the course of the disease.

概要

観察データは、薬剤の監視や有効性の研究、治療経路の調査、患者の転帰の予測などに利用することができます。このような研究には、対象となる患者を見つけるための実行可能なアルゴリズム、すなわち Phenotypingアルゴリズムの開発が必要である。データネットワークにおける信頼性の高い包括的な Phenotypingアルゴリズムの作成は、患者の表現の違いやデータの異質性を考慮する必要があるため、特に困難である。本論文では、包括的なコンセプトセットを作成するためのプロセスと、それを促進するために構築したレコメンダーシステムについて説明する。PHenotype Observed Entity Baseline Endorsements (PHOEBE) は、6カ国の OHDSI 共通データモデルにマッピングされた22の電子健康記録と請求データセットにわたるコード使用に関するデータを使用して、意味的および語彙的に類似したコードを推薦する。Cohort Diagnosticsと組み合わせることで、現在では主要なネットワークOHDSI研究において使用されています。患者コホートの作成に使用すると、PHOEBEはより多くの患者を特定し、病気の経過の早い段階で捕捉することができます。



安全性シグナルを特定する自動化手法

> AMIA Annu Symp Proc. 2023 Apr 29;2022:596-605. eCollection 2022.

Hierarchy-aware Adverse Reaction Embeddings for Signal Detection

Venkata Joopudi¹, Bharath Dandala¹, Ching-Huei Tsou¹, Jennifer J Liang¹

Affiliations – collapse

Affiliation

¹ IBM Research, Yorktown Heights, NY.

PMID: 37128452 PMCID: [PMC10148317](#)

Abstract

Post-market drug surveillance monitors new and evolving treatments for their effectiveness and safety in real-world conditions. A large amount of drug safety surveillance data is captured by spontaneous reporting systems such as the FAERS. Developing automated methods to identify actionable safety signals from these databases is an active area of research. In this paper, we propose two novel network representation learning methods (HARE and T-HARE) for signal detection that jointly utilize association information between drugs and medical outcomes from the FAERS and ancestral information in medical ontologies. We evaluate these methods using two publicly available reference datasets, EU-ADR and OMOP corpus. Experimental results showed that the proposed methods significantly outperformed standard methodologies based on disproportionality metrics and the existing state-of-the-art aer2vec method with statistically significant improvements on both EU-ADR and OMOP datasets. Through quantitative and qualitative analysis, we demonstrate the potential of the proposed methods for effective signal detection.

概要

市販後医薬品安全性監視は、新しい治療法や進化する治療法について、その有効性と安全性を実環境の中で監視するものです。FAERSのような自発的な報告システムによって、大量の医薬品安全性監視データが取得されています。これらのデータベースから実用的な安全性シグナルを特定する自動化手法を開発することは、活発な研究分野である。本論文では、シグナル検出のための2つの新しいネットワーク表現学習法（HAREとT-HARE）を提案し、FAERSから得られる薬剤と医療結果の関連情報および医療オントロジーの祖先情報を共同で利用する。これらの手法を、EU-ADRとOMOPコーパスという2つの公開された参照データセットを用いて評価した。実験の結果、提案手法は、EU-ADRとOMOPの両データセットにおいて、不均衡度メトリクスに基づく標準的な手法と既存の最先端手法であるaer2vec手法を統計的に有意に改善することが示された。定量的・定性的な分析を通じて、効果的な信号検出のための提案手法の可能性を実証することができた。

OMOPコーパス：有害事情データベース
(OMOP CDMのことではない)



OMOPと地理的情報とのリンク

> [AMIA Annu Symp Proc. 2023 Apr 29;2022:587-595. eCollection 2022.](#)

Feasibility of Linking Area Deprivation Index Data to the OMOP Common Data Model

Xinzhuo Jiang¹, Maura A Beaton¹, Jake Gillberg², Andrew Williams², Karthik Natarajan¹

Affiliations – collapse

Affiliations

- 1 Columbia University Irvine Medical Center, New York, NY.
- 2 Tufts Medical Center, Boston, MA.

PMID: 37128466 PMCID: [PMC10148322](#)

Abstract

Linking Area Deprivation Index (ADI) scores to observational data offers the opportunity to characterize healthcare treatment and outcomes that are driven by socioeconomic deprivation. The current study aims to assess the feasibility of creating an analysis package to link ADI rankings to multiple patient-level EHR datasets transformed into the OMOP CDM. Patients within two cancer cohorts (breast cancer and multiple myeloma) were identified within two OMOP datasets and their records were linked with ADI scores using address information in the OMOP location table. With ADI linked to patient addresses, we generated visualizations showing the geographic distribution of each cohort based on ADI scores. Additionally, further assessment showed that over 89% of patient addresses could successfully be linked with ADI rankings. In conducting this assessment, we have demonstrated that developing a package to link ADI scores with multiple OMOP datasets is feasible.

概要

地理的剥奪指数（ADI）スコアを観察データにリンクさせることで、社会経済的剥奪に起因する医療治療と転帰を特徴付ける機会が得られる。本研究では、OMOP CDMに変換された複数の患者レベルのEHRデータセットにADIランキングをリンクさせる解析パッケージを作成することの実現可能性を評価することを目的としています。2つのがんコホート（乳がんと多発性骨髄腫）の患者を2つのOMOPデータセットで特定し、その記録をOMOPロケーションテーブルの住所情報を使ってADIスコアとリンクさせた。ADIを患者の住所にリンクさせることで、ADIスコアに基づく各コホートの地理的分布を示すビジュアライゼーションを作成しました。さらに、さらなる評価により、患者の住所の89%以上がADIランキングとうまくリンクできることがわかりました。この評価の実施により、ADIスコアと複数のOMOPデータセットをリンクさせるパッケージの開発が可能であることが実証されました。

※ADIスコア：地域の社会経済的状況を多面的に評価する指数



> JCO Clin Cancer Inform. 2023 May;7:e2200179. doi: 10.1200/CCI.22.00179.

How to Improve Cancer Patients ENrollment in Clinical Trials From rEal-Life Databases Using the Observational Medical Outcomes Partnership Oncology Extension: Results of the PENELOPE Initiative in Urologic Cancers

Emmanuelle Kempf^{1 2}, Morgan Vaterkowski^{3 4}, Damien Leprovost^{1 3}, Nicolas Griffon^{1 3}, David Ouagne³, Stéphane Breant³, Patricia Serre³, Alexandre Mouchet³, Bastien Rance⁵, Gilles Chatellier⁵, Ali Bellamine³, Marie Frank⁶, Julien Guerin⁷, Xavier Tannier¹, Alain Livartowski⁷, Martin Hilka³, Christel Daniel^{1 3}

PMID: 37167578

Affiliations

- 1 Sorbonne Université, Inserm, Université Sorbonne Paris Nord, Laboratoire d'Informatique Médicale et d'Ingénierie des Connaissances pour la e-Santé, LIMICS, Paris, France.
- 2 Department of Medical Oncology, Assistance Publique Hôpitaux de Paris, Henri Mondor Teaching Hospital, Créteil, France.
- 3 Innovation and Data, Paris, IT Department, Assistance Publique Hôpitaux de Paris, Paris, France.
- 4 EPITA School of Engineering and Computer Science, Paris, France.
- 5 Department of Medical Informatics, Assistance Publique Hôpitaux de Paris, Centre-Université de Paris (APHP-CUP), Université de Paris, Paris, France.
- 6 Department of Medical Information, Paris Saclay Teaching Hospital, Assistance Publique Hôpitaux de Paris, Paris, France.
- 7 IT Department, Curie Hospital, Paris, France.

Abstract

Purpose: To compare the computability of Observational Medical Outcomes Partnership (OMOP)-based queries related to prescreening of patients using two versions of the OMOP common data model (CDM; v5.3 and v5.4) and to assess the performance of the Greater Paris University Hospital (APHP) prescreening tool.

Materials and methods: We identified the prescreening information items being relevant for prescreening of patients with cancer. We randomly selected 15 academic and industry-sponsored urology phase I-IV clinical trials (CTs) launched at APHP between 2016 and 2021. The computability of the related prescreening criteria (PC) was defined by their translation rate in OMOP-compliant queries and by their execution rate on the APHP clinical data warehouse (CDW) containing data of 205,977 patients with cancer. The overall performance of the prescreening tool was assessed by the rate of true- and false-positive cases of three randomly selected CTs.

Results: We defined a list of 15 minimal information items being relevant for patients' prescreening. We identified 83 PC of the 534 eligibility criteria from the 15 CTs. We translated 33 and 62 PC in queries on the basis of OMOP CDM v5.3 and v5.4, respectively (translation rates of 40% and 75%, respectively). Of the 33 PC translated in the v5.3 of the OMOP CDM, 19 could be executed on the APHP CDW (execution rate of 58%). Of 83 PC, the computability rate on the APHP CDW reached 23%. On the basis of three CTs, we identified 17, 32, and 63 patients as being potentially eligible for inclusion in those CTs, resulting in positive predictive values of 53%, 41%, and 21%, respectively.

Conclusion: We showed that PC could be formalized according to the OMOP CDM and that the oncology extension increased their translation rate through better representation of cancer natural history.

概要

目的: OMOP共通データモデル (CDM ; v5.3およびv5.4) の2つのバージョンを用いて、患者の事前スクリーニングに関連するOMOPベースのクエリの計算可能性を比較し、Greater Paris University Hospital (APHP) 事前スクリーニングツールの性能を評価することである。

材料と方法: がん患者を対象としたプレスクリーニングに関連するプレスクリーニング情報項目を特定した。2016年から2021年の間にAPHPで開始された学術および企業主催の泌尿器科第I-IV相臨床試験 (CT) 15件をランダムに選択しました。関連するプレスクリーニング基準 (PC) の計算可能性は、OMOP準拠のクエリにおける翻訳率と、がん患者205,977人のデータを含むAPHP臨床データウェアハウス (CDW) 上での実行率によって定義されました。プレスクリーニングツールの総合的な性能は、無作為に選んだ3つのCTの真陽性と偽陽性の割合によって評価された。

結果: 患者さんのプレスクリーニングに関連する15の最小限の情報項目を定義した。15件のCTから、534の適格性基準のうち83のPCを特定した。OMOP CDM v5.3およびv5.4に基づいて、それぞれ33および62のPCを翻訳した (翻訳率はそれぞれ40%および75%) 。OMOP CDM v5.3で翻訳した33個のPCのうち、19個がAPHP CDWで実行できた (実行率58%) 。83個のPCのうち、APHP CDWでの計算可能率は23%に達した。3つのCTをもとに、そのCTに組み入れられる可能性のある患者を17人、32人、63人特定した結果、陽性的中率はそれぞれ53%、41%、21%であった。

結論: PCは、OMOP CDMに従って形式化することができ、癌の自然史をよりよく表現することによって、オンコロジーエクステンションの翻訳率が向上することが示された。



オーストリアEHRのOMOP化

> Stud Health Technol Inform. 2023 May 2;301:54-59. doi: 10.3233/SHTI230011.

Transforming Documents of the Austrian Nationwide EHR System into the OMOP CDM

Raffael Lukas Korntheuer¹, Florian Katsch^{1 2}, Georg Duftschmid¹

Affiliations – collapse

Affiliations

- 1 Section of Medical Information Management, Center for Medical Statistics, Informatics and Intelligent Systems, Medical University of Vienna, Vienna, Austria.
- 2 Ludwig Boltzmann Institute for Digital Health and Prevention, Salzburg, Austria.

PMID: 37172152 DOI: [10.3233/SHTI230011](https://doi.org/10.3233/SHTI230011)

Abstract

The Austrian nationwide EHR system ELGA can contribute valuable data for research due to its high volume of data and broad population coverage. In order to be applicable in international research projects, transformation to a standardized, research-oriented data model such as the OMOP common data model is essential. In this paper we describe our experience with the corresponding transformation task. Using Python scripts, we implemented a prototypical process that extracts, transforms, maps, and loads fully structured sections of ELGA documents to an OMOP database.

概要

オーストリアの全国的なEHRシステムであるELGAは、そのデータ量の多さと幅広い人口をカバーしていることから、研究にとって貴重なデータを提供することができます。国際的な研究プロジェクトに適用するためには、OMOP共通データモデルのような標準化された研究指向のデータモデルへの変換が不可欠である。本論文では、対応する変換タスクに関する私たちの経験について述べる。Pythonスクリプトを使用して、ELGA文書の完全な構造化セクションを抽出、変換、マッピングし、OMOPデータベースにロードするプロトタイプのプロセスを実装しました。

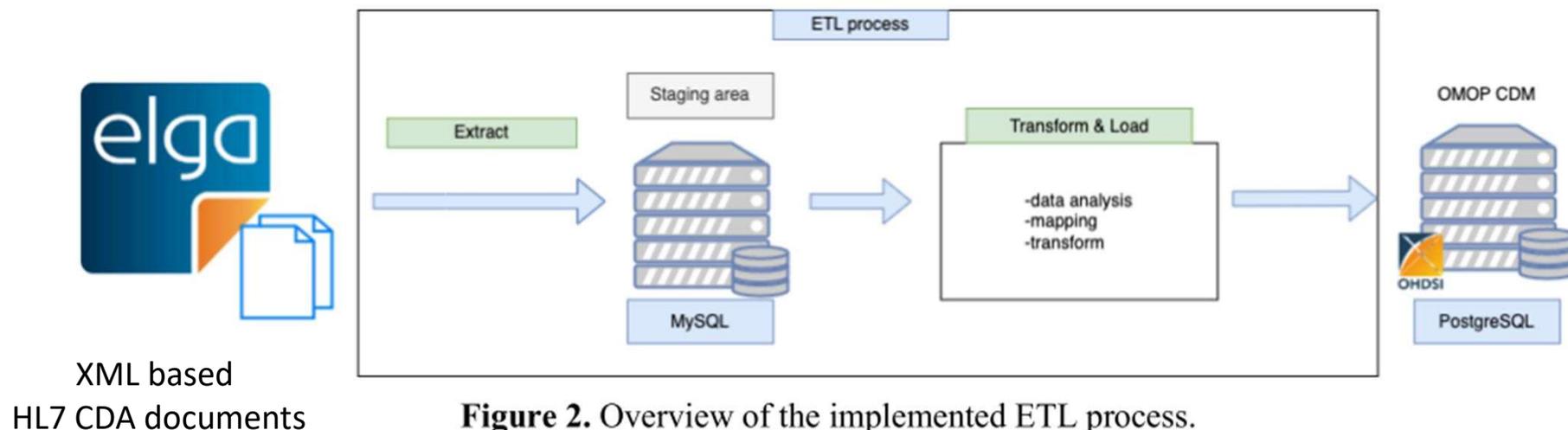


Figure 2. Overview of the implemented ETL process.



標準化された心電図データベースを構築

➤ [Healthc Inform Res. 2023 Apr;29\(2\):132-144. doi: 10.4258/hir.2023.29.2.132. Epub 2023 Apr 30.](#)

Standardized Database of 12-Lead Electrocardiograms with a Common Standard for the Promotion of Cardiovascular Research: KURIAS-ECG

Hakje Yoo¹, Yunjin Yum², Soo Wan Park¹, Jeong Moon Lee¹, Moonjoung Jang¹,
Yoojoong Kim³, Jong-Ho Kim⁴, Hyun-Joon Park⁵, Kap Su Han⁵, Jae Hyoung Park⁴,
Hyung Joon Joo^{1 4 6}

PMID: 37190737

Affiliations

- 1 Korea University Research Institute for Medical Bigdata Science, Korea University College of Medicine, Seoul, Korea.
- 2 Department of Biostatistics, Korea University College of Medicine, Seoul, Korea.
- 3 School of Computer Science and Information Engineering, The Catholic University of Korea, Bucheon, Korea.
- 4 Department of Cardiology, Cardiovascular Center, Korea University College of Medicine, Seoul, Korea.
- 5 Korea University Research Institute for Healthcare Service Innovation, Korea University College of Medicine, Seoul, Korea.
- 6 Department of Medical Informatics, Korea University College of Medicine, Seoul, Korea.

Abstract

Objectives: Electrocardiography (ECG)-based diagnosis by experts cannot maintain uniform quality because individual differences may occur. Previous public databases can be used for clinical studies, but there is no common standard that would allow databases to be combined. For this reason, it is difficult to conduct research that derives results by combining databases. Recent commercial ECG machines offer diagnoses similar to those of a physician. Therefore, the purpose of this study was to construct a standardized ECG database using computerized diagnoses.

Methods: The constructed database was standardized using Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) and Observational Medical Outcomes Partnership-common data model (OMOP-CDM), and data were then categorized into 10 groups based on the Minnesota classification. In addition, to extract high-quality waveforms, poor-quality ECGs were removed, and database bias was minimized by extracting at least 2,000 cases for each group. To check database quality, the difference in baseline displacement according to whether poor ECGs were removed was analyzed, and the usefulness of the database was verified with seven classification models using waveforms.

Results: The standardized KURIAS-ECG database consists of high-quality ECGs from 13,862 patients, with about 20,000 data points, making it possible to obtain more than 2,000 for each Minnesota classification. An artificial intelligence classification model using the data extracted through SNOMED-CT showed an average accuracy of 88.03%.

Conclusions: The KURIAS-ECG database contains standardized ECG data extracted from various machines. The proposed protocol should promote cardiovascular disease research using big data and artificial intelligence.

概要

目的: 心電図（ECG）を用いた専門家による診断では、個人差が生じるため、均一な品質を保つことができない。臨床研究においては、過去の公開データベースを利用することができるが、データベースを組み合わせるための共通規格はない。そのため、データベースを組み合わせる結果を導き出すような研究を行うことは難しい。最近の市販の心電図装置では、医師と同様の診断が可能である。そこで、本研究では、コンピュータによる診断を用いて標準的な心電図データベースを構築することを目的とした。

方法: 構築したデータベースは、SNOMED CTとOMOP-CDMを用いて標準化し、ミネソタ分類に基づいて10群に分類した。また、高品質な波形を抽出するために、品質の悪い心電図は削除し、各グループで少なくとも2,000例を抽出することでデータベースの偏りを最小化した。データベースの品質を確認するため、質の悪い心電図を除去したかどうかによるベースライン変位の違いを分析し、波形を用いた7つの分類モデルでデータベースの有用性を検証した。

結果: 標準化されたKURIAS-ECGデータベースは、13,862人の患者の高品質な心電図からなり、約20,000点のデータから、各ミネソタ分類に対して2,000以上のデータを得ることが可能であった。SNOMED-CTにより抽出されたデータを用いた人工知能による分類モデルでは、平均88.03%の精度を示した。

結論: KURIAS-ECGデータベースは、様々な機械から抽出された標準化されたECGデータを含んでいる。提案されたプロトコルは、ビッグデータと人工知能を用いた心血管疾患研究を促進するはずである。



OMOP CDM で連合学習

> [Healthc Inform Res.](#) 2023 Apr;29(2):168-173. doi: 10.4258/hir.2023.29.2.168. Epub 2023 Apr 30.

Feasibility Study of Federated Learning on the Distributed Research Network of OMOP Common Data Model

PMID: 37190741

Geun Hyeong Lee^{1 2}, Jonggul Park³, Jihyeong Kim³, Yeesuk Kim⁴, Byungjin Choi⁵,
Rae Woong Park^{5 6}, Sang Youl Rhee^{7 8}, Soo-Yong Shin^{1 2 9}

Affiliations

- 1 Department of Digital Health, SAIHST, Sungkyunkwan University, Seoul, Korea.
- 2 Department of Intelligent Precision Healthcare Convergence, Sungkyunkwan University, Suwon, Korea.
- 3 Evidnet Inc., Seongnam, Korea.
- 4 Department of Orthopedic Surgery, Hanyang University College of Medicine, Seoul, Korea.
- 5 Department of Biomedical Informatics, Ajou University School of Medicine, Suwon, Korea.
- 6 Department of Biomedical Sciences, Graduate School of Medicine, Ajou University, Suwon, Korea.
- 7 Department of Endocrinology and Metabolism, College of Medicine, Kyung Hee University, Seoul, Korea.
- 8 Center for Digital Health, Medical Science Research Institute, Kyung Hee University Medical Center, Seoul, Korea.
- 9 Center for Research Resource Standardization, Samsung Medical Center, Seoul, Korea.

Abstract

Objectives: Since protecting patients' privacy is a major concern in clinical research, there has been a growing need for privacy-preserving data analysis platforms. For this purpose, a federated learning (FL) method based on the Observational Medical Outcomes Partnership (OMOP) common data model (CDM) was implemented, and its feasibility was demonstrated.

Methods: We implemented an FL platform on FeederNet, which is a distributed clinical data analysis platform based on the OMOP CDM in Korea. We trained it through an artificial neural network (ANN) using data from patients who received steroid prescriptions or injections, with the aim of predicting the occurrence of side effects depending on the prescribed dose. The ANN was trained using the FL platform with the OMOP CDMs of Kyung Hee University Medical Center (KHMC) and Ajou University Hospital (AUH).

Results: The area under the receiver operating characteristic curves (AUROCs) for predicting bone fracture, osteonecrosis, and osteoporosis using only data from each hospital were 0.8426, 0.6920, and 0.7727 for KHMC and 0.7891, 0.7049, and 0.7544 for AUH, respectively. In contrast, when using FL, the corresponding AUROCs were 0.8260, 0.7001, and 0.7928 for KHMC and 0.7912, 0.8076, and 0.7441 for AUH, respectively. In particular, FL led to a 14% improvement in performance for osteonecrosis at AUH.

Conclusions: FL can be performed with the OMOP CDM, and FL often shows better performance than using only a single institution's data. Therefore, research using OMOP CDM has been expanded from statistical analysis to machine learning so that researchers can conduct more diverse research.

概要

目的: 臨床研究において患者のプライバシー保護は大きな関心事であるため、プライバシーを保護したデータ解析基盤の必要性が高まっている。このため、OMOP CDMに基づく連合学習 (FL) 手法を実装し、その実現可能性を実証した。

方法: 韓国のOMOP CDMに基づく分散型臨床データ解析プラットフォームであるFeederNet上に、FLプラットフォームを実装した。ステロイドの処方や注射を受けた患者のデータを用いて、処方量に応じた副作用の発生を予測することを目的に、人工ニューラルネットワーク (ANN) を用いて学習させた。ANNの学習は、キョンヒ大学医療センター (KHMC) とアジュ大学病院 (AUH) のOMOP CDMを用い、FLプラットフォームで行った。

結果: 各病院のデータのみを用いた骨折、骨壊死、骨粗鬆症の予測における受信者動作特性曲線下面積 (AUROC) は、KHMCではそれぞれ0.8426, 0.6920, 0.7727, AUHでは0.7891, 0.7049, 0.7544 となった。一方、FLを用いた場合のAUROCは、KHMCで0.8260、0.7001、0.7928、AUHで0.7912、0.8076、0.7441となった。特に、AUHではFLにより骨壊死に対する成績が14%向上した。

結論: FLはOMOP CDMで実施可能であり、FLは単一施設のデータのみを使用するよりも優れた性能を示すことが多い。そのため、OMOP CDMを用いた研究は、統計解析から機械学習へと広がり、研究者がより多様な研究を行うことができるようになった。



ATCコード薬剤オーダー情報のRxNorm化

Observational Study > Stud Health Technol Inform. 2023 May 18;302:711-715.
doi: 10.3233/SHTI230245.

Fitness for Use of Anatomical Therapeutic Classification for Real World Data Research

Ines Reinecke ¹, Elisa Henke ¹, Yuan Peng ¹, Martin Sedlmayr ¹, Franziska Bathelt ¹

Affiliations – collapse

Affiliation

¹ Carl Gustav Carus Faculty of Medicine, Center for Medical Informatics, Institute for Medical Informatics and Biometry, Technische Universität Dresden, Dresden, Germany.

PMID: 37203475 DOI: [10.3233/SHTI230245](https://doi.org/10.3233/SHTI230245)

Abstract

Introduction: Real-world data (RWD) is gaining importance in research. For instance, the European Medicines Agency (EMA) is currently in the process of establishing a cross-national research network that utilizes RWD for research. However, data harmonization across countries must be carefully considered to avoid misclassification and bias.

Objectives: This paper aims to investigate the extent to which a correct assignment of RxNorm ingredients is possible for medication orders that include only ATC codes.

Methods: In this study, we analyzed 1,506,059 medication orders from the University Hospital Dresden (UKD) and merged them with the ATC vocabulary in the Observational Medical Outcomes Partnership (OMOP) including relevant relationship mappings to RxNorm.

Results: We identified 70.25% of all medication orders were single ingredients with direct mapping to RxNorm. However, we also identified a significant complexity in mappings for the other medication orders that was visualized in an interactive scatterplot.

Discussion: The majority of medication orders under observation (70.25%) are single ingredients and can be standardized to RxNorm, combination drugs pose a challenge due to the different approaches of ingredient assignments in ATC and RxNorm. The provided visualization can help research teams gain a better understanding of problematic data and further investigate identified issues.

概要

はじめに: 研究においてリアルワールドデータ (RWD) の重要性が高まっています。例えば、欧州医薬品庁 (EMA) は現在、RWDを研究に活用する国横断的な研究ネットワークの構築を進めています。しかし、国を超えてのデータの調和は、誤分類やバイアスを避けるために慎重に検討する必要があります。

目的: 本稿では、ATCコードのみを含む薬剤オーダーについて、RxNorm成分の正しい割り当てがどの程度可能であることを調査することを目的とする。

方法: 本研究では、ドレスデン大学病院 (UKD) の1,506,059件の薬剤オーダーを分析し、RxNormとの関連マッピングを含むOMOPのATC語彙と統合した。

結果: 全薬剤オーダーの70.25%が単一成分であり、RxNormに直接マッピングされていることが確認された。しかし、それ以外の薬については、マッピングが著しく複雑であることが確認され、インタラクティブな散布図で可視化された。

考察: 観察下の薬剤オーダーの大部分 (70.25%) は単一成分であり、RxNormに標準化することができる。配合剤は、ATCとRxNormの成分割り当てのアプローチが異なるため、難題となっている。この可視化は、研究チームが問題のあるデータをよりよく理解し、特定された問題をさらに調査するのに役立つ。



メタデータ駆動型ETLプロセス のコンセプト

> Stud Health Technol Inform. 2023 May 18;302:751-752. doi: 10.3233/SHTI230256.

Towards ETL Processes to OMOP CDM Using Metadata and Modularization

Yuan Peng ¹, Elisa Henke ¹, Martin Sedlmayr ¹, Franziska Bathelt ¹

Affiliations – collapse

Affiliation

¹ Institute for Medical Informatics and Biometry, Carl Gustav Carus Faculty of Medicine, Technische Universität Dresden, Dresden, Germany.

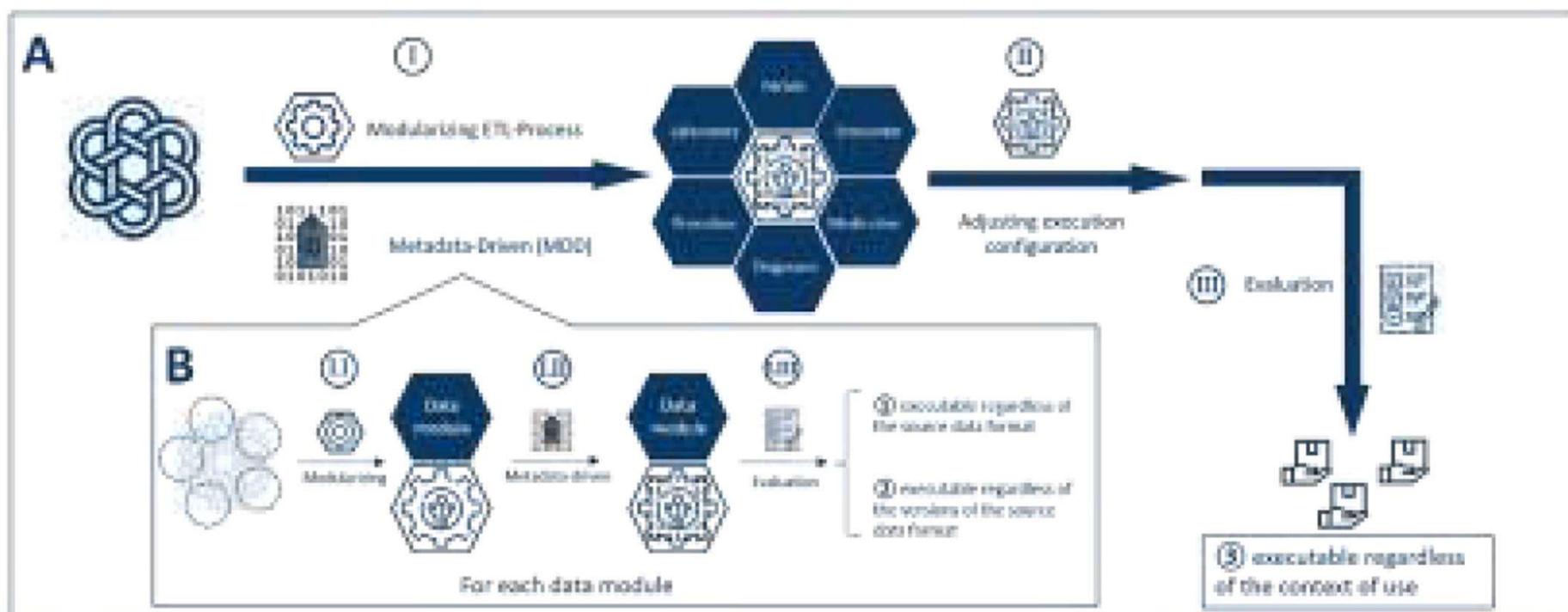
PMID: 37203486 DOI: [10.3233/SHTI230256](https://doi.org/10.3233/SHTI230256)

Abstract

OMOP common data model (CDM) is designed for analyzing large clinical data and building cohorts for medical research, which requires Extract-Transform-Load processes (ETL) of local heterogeneous medical data. We present a concept for developing and evaluating a modularized metadata-driven ETL process, which can transform data into OMOP CDM regardless of 1) the source data format, 2) its versions and 3) context of use.

概要

OMOP CDMは、大規模な臨床データの解析や医学研究のためのコホート構築のために設計されており、ローカルな異種医療データの抽出-変換-ロードプロセス（ETL）が必要となる。我々は、1) ソースデータ形式、2) そのバージョン、3) 使用のコンテキストに関係なく、データをOMOP CDMに変換できる、モジュール化されたメタデータ駆動型ETLプロセスを開発し評価するコンセプトを提示する。





今月のCommunity Call

- APAC Call テーマ

May. 18 Community Call: Training Session #4

A*STAR Data Vault

a research platform for the Singapore ecosystem and beyond

- Global Community Call テーマ

May 2 SOS Challenge Week 6: Network Execution

May 9 SOS Challenge Week 7: Study Diagnostics

May 16 SOS Challenge Week 8: Evidence Synthesis

May 23 SOS Challenge Week 9: Interpreting the Results



2023 European Symposium

<https://www.ohdsi-europe.org/index.php/symposium-2023>

日程：7月 1-3日（メインシンポジウムは7月3日）

場所：1 - 2日 Erasmus University Medical Center
3日 Steam Ship Rotterdam





2023 APAC Symposium

登録受付中！

<https://www.ohdsi.org/2023apacsymposium/>

日程：7月 13-14日 MEDINFO 2023が 7/8-12に開催

場所： University of New South Wales, Sydney, Australia

13日 **Main Conference**

Session 1: OHDSI Global

Session 2: Research

Session 3: OHDSI APAC



登録サイトより

14日 **Tutorials**

ETL、Characterization、

Population-Level Estimation、 Patient-Level Prediction



Save Our Sisyphus Challenge

<https://ohdsi.org/sos-challenge/>

Tutorial 全9回

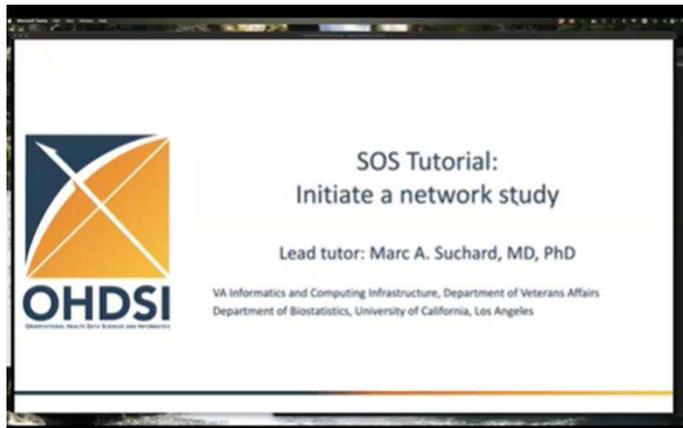
Date	Times	Topic
Mar. 28	11 am / 7 pm ET	SOS Week 1 Tutorial: Initiating A Network Study
Apr. 4	11 am / 7 pm ET	SOS Week 2 Tutorial: Data Diagnostics
Apr. 11	11 am / 7 pm ET	SOS Week 3 Tutorial: Phenotype Development
Apr. 18	11 am / 7 pm ET	SOS Week 4 Tutorial: Phenotype Evaluation
Apr. 25	11 am / 7 pm ET	SOS Week 5 Tutorial: Creating Analysis Specifications
May 2	11 am / 7 pm ET	SOS Week 6 Tutorial: Network Execution
May 9	11 am / 7 pm ET	SOS Week 7 Tutorial: Study Diagnostics
May 16	11 am / 7 pm ET	SOS Week 8 Tutorial: Evidence Synthesis
May 23	11 am / 7 pm ET	SOS Week 9 Tutorial: Interpreting The Results

一気に紹介します



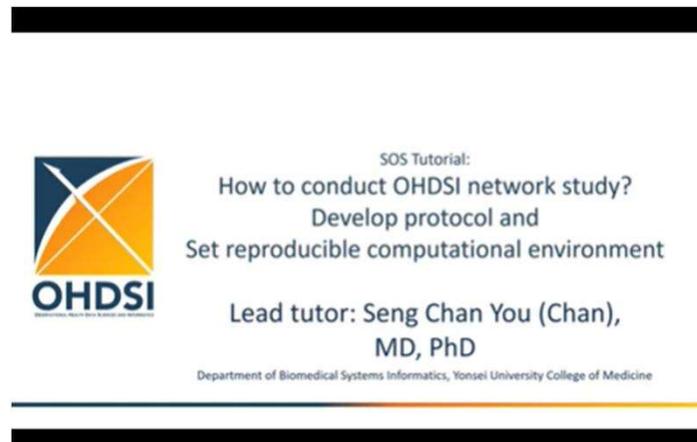
Week 1: Initiating A Network Study

Session 1: Marc Suchard



この動画では、ネットワーク研究を開始するための手順について説明されています。まず、OHDSI Rツールスタックと必要なソフトウェアについて説明があり、GitHubを使用してオープンサイエンスを共有する方法についても解説されています。また、プロトコルの起草やマニュスクリプトの設定についても触られています。具体的な手順としては、GitHubでリポジトリをリクエストし、ローカルでコピーを作成し、プロトコルの起草をMarkdown形式で行うことが挙げられます。重要なポイントとしては、コードの重複を避けるためにMarkdown形式で起草することが大切であることが強調されています。また、GitHubを使用することで、バージョン管理ができるため、誤りを取り消したり、変更を追跡したりすることができます。

Session 2: Seng Chan You



この動画では、SOS Challengeにおけるネットワーク研究の開始方法について紹介されています。まず、OHDSIとは何か、その目的や特徴について説明されています。OHDSIは、マップCDMを使用した大規模なヘルスデータベースのネットワークに基づくオフィスデータアナリティクスソリューションを提供する国際的なコラボレーションコンソーシアムであり、健康と疾患に関する包括的な理解を生み出すことを目指しています。

次に、ネットワーク研究を行うためのプロトコルの開発方法や、再現可能な計算環境の設定方法について説明されています。具体的には、GitHubを使用してバージョン管理を行い、エンドツーエンドの解析プログラムを開発することが重要であると説明されています。また、Dockerを使用することで、アプリケーションの開発、テスト、本番環境での実行を簡単に行うことができます。

最後に、具体的な研究の例として、フルオロキノロン系抗生物質の処方による大動脈瘤や解離性大動脈瘤のリスクについての研究が紹介されています。この研究では、GitHubを使用してプロトコルの開発や解析プログラムの作成を行い、再現可能な計算環境を構築することで、高品質な研究成果を得ることが目指されています。



Network Studyを実施するステップ

By ChatGPT

OHDSIネットワーク研究を面白く実施するためのステップバイステップガイド：

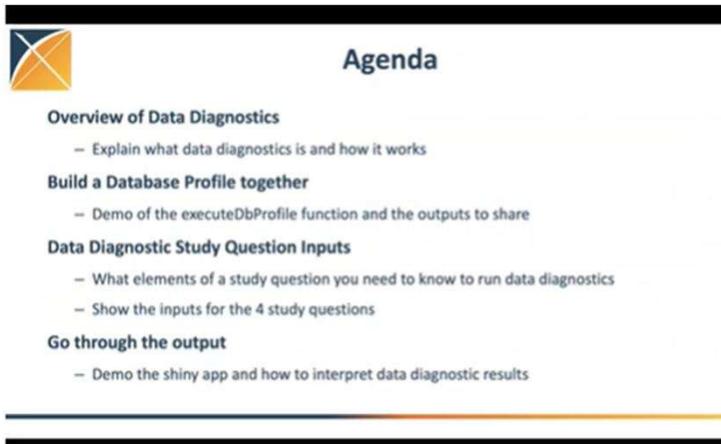
- 1. OHDSIコミュニティに参加する**：活気溢れるOHDSIコミュニティに参加し、専門家とのつながりを持ち、革新的な研究に協力しましょう。
- 2. 興味深い研究テーマの選択**：好奇心を引き起こし、独自の視点で探求できる魅力的な研究テーマを選びましょう。
- 3. 協力的なチームの構築**：異なる専門分野の専門家からなる多様なチームを編成し、幅広い視点とスキルを研究にもたらします。
- 4. 独自の研究手法の設計**：OHDSIネットワーク内の非伝統的なデータソースを考慮し、独創的な研究設計を作成しましょう。
- 5. OHDSI CDMを使用したデータの統合**：OHDSI共通データモデル（CDM）を使用して、異なるデータソースのデータをシームレスに統合しましょう。
- 6. OHDSIのツールと分析の活用**：AtlasやAchillesなど、OHDSIが提供する強力なツールや分析手法を活用してデータを効果的に分析・視覚化します。
- 7. 魅力的なストーリーテリング**：魅力的な視覚的表現やストーリーテリング手法、現実世界の例を活用して、研究結果に関する物語性を作り上げ、聴衆を魅了しましょう。
- 8. 共有と協力**：研究結果を出版物やプレゼンテーション、協力関係を通じて科学コミュニティと共有し、知識交換を促進し、他の人々にインスピレーションを与えましょう。
- 9. 創造性と楽しみを取り入れる**：インタラクティブなデータ可視化やゲーミフィケーションの要素を活用するなど、研究に創造性と楽しみを取り入れましょう。
- 10. 情熱を持ち続け、学び続ける**：OHDSIの研究への情熱を持ち続け、常に新しい手法やベストプラクティスを学び、探求し続けましょう。

OHDSIネットワーク研究の実施には、共同作業、OHDSI標準とガイドラインの遵守、オープンサイエンスの原則へのコミットメントが必要であることを忘れないでください。OHDSIコミュニティに参加することで、研究のプロセスを通じて貴重なサポートとガイダンスを提供することができます。



Week 2: Data Diagnostics

Session 1: Clair Blacketer, Mui Van Zandt, Sarah Seager

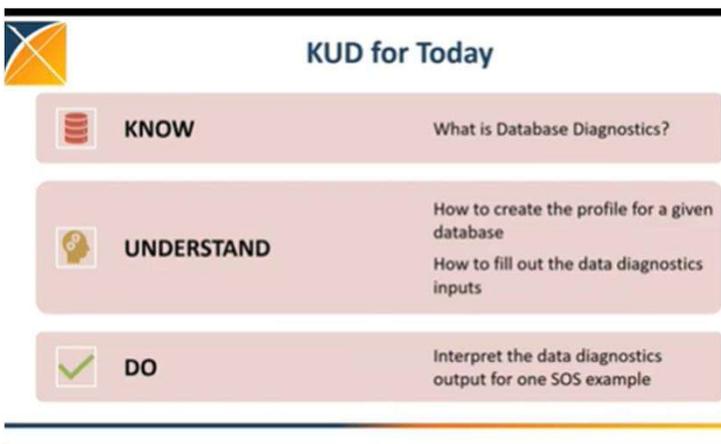


Agenda

- Overview of Data Diagnostics**
 - Explain what data diagnostics is and how it works
- Build a Database Profile together**
 - Demo of the executeDbProfile function and the outputs to share
- Data Diagnostic Study Question Inputs**
 - What elements of a study question you need to know to run data diagnostics
 - Show the inputs for the 4 study questions
- Go through the output**
 - Demo the shiny app and how to interpret data diagnostic results

このビデオでは、データ診断とは何かについて説明されています。データ診断は、集計された要約統計情報のみを使用して、研究問題に対する要素を持つデータベースを特定することができる軽量でシンプルなRパッケージです。データ診断を実行するためには、まず各データベースにプロファイルを生成する必要があります。プロファイルには、性別、出生年、状態概念IDなどの要約統計情報が含まれます。これらの情報は、OHDSI連携センターに送信され、特定の研究問題に対する各データベースの適切性が判断されます。このプロセスにより、ネットワーク研究のオーバーヘッドが削減され、ネットワーク全体の理解が深まります。プロファイルを生成する方法について、Rスタジオを使用して詳しく説明されています。

Session 2: Mui Van Zandt/Jing Li



KUD for Today

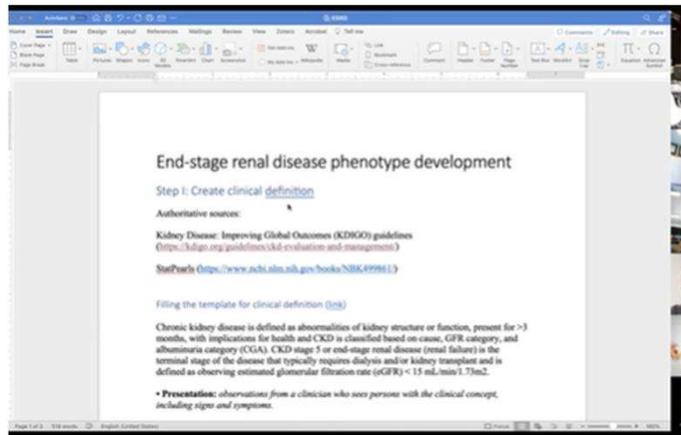
 KNOW	What is Database Diagnostics?
 UNDERSTAND	How to create the profile for a given database How to fill out the data diagnostics inputs
 DO	Interpret the data diagnostics output for one SOS example

データ診断とは何か、なぜ重要なのか、影響とネットワーク研究の方法に焦点を当てます。フルオロキノロン研究を例に、チュートリアルを行います。4つの主要なトピックがあります。まず、データ診断の概要、次にデータベースプロファイルの作成、そしてデータ診断ツールに研究質問を入力する方法、最後にRシャイニーアプリで結果を確認します。データ診断は、集計サマリー統計を使用して、研究質問に回答するために必要な要素を持つデータベースを特定するためのパッケージです。集計サマリー統計を使用して、患者レベルのレコードを見るのではなく、集約データを中心に使用します。集計されたデータは、中央のOHDSIセンターに保存されます。データベースプロファイルを作成する前に、GitHubレポジトリからデータ診断の情報をダウンロードしてください。データ診断ツールを使用して、コホートのパラメータに影響を与えるための質問を入力し、メタデータテーブルに対して実行します。結果は、どのデータベースが特定の研究質問に最適かを示すシャイニーレポートを作成します。



Week 3: Phenotype Development

Session 1a: Anna Ostropolets – Outcome Design



この動画では、エンドステージ腎臓病の表現型を開発するプロセスが説明されています。最初のステップは、PubMedのガイドラインやリソースを使用して臨床的な定義を作成することです。定義には、3か月以上の腎臓機能の障害、透析または腎移植が必要で、推定糸球体濾過率が1平方メートルあたり15ミリリットル未満が含まれます。次に、以前のエンドステージ腎臓病に関する研究を調べ、診断コード、測定、手順、および観察に関連するコンセプトセットを作成します。その後、これらのコンセプトセットを使用して、疾患に関連する診断コード、測定、手順を含むコホート定義の包含基準を作成する方法を示します。手順に基づくコンセプトセットにすべての関連語彙を含め、慢性的な状態を確認するために複数のイベントを必要とすることの重要性を強調しています。

Session 1b: Asieh Golozar – Exposure Design



antiVEGF exposure cohort design

このビデオでは、白内障の治療における異なる抗血管内皮増殖因子治療法のリスクを比較するための特定の cohorts の作成方法について説明しています。対象となる人口は、視覚障害の診断を受けた18歳以上の患者で、少なくとも3回の月間抗血管内皮増殖因子注射を受けている人々です。スピーカーは、より良い季節的曝露 cohorts の作成方法を説明し、仮想的な患者の旅を描きます。彼らは、最初のベバシズマブ注射、2回目の注射、そして患者がリスクにさらされる第3回目の注射を特定する必要があると説明します。また、365日間の前向き観察、以前の抗血管内皮増殖因子の暴露がないこと、および視覚障害の診断が含まれる参加基準について説明します。スピーカーはAtlasツールを使用してcohortsを定義し、cohortsを評価する方法について説明します。



Week 3: Phenotype Development

Session 2a: Christian Reich – Exposure Design

How do you build exposures

- Step 1: Ingredients
- Step 2: modifications
 - Route of administration
 - Combinations
 - Dosing

Clinical drug forms
Clinical drug components

このビデオでは、フルオロキノロン抗生物質の使用後に大動脈瘤または解離のリスクがあるかどうかの研究質問のフェノタイプの開発について話し合われています。彼らは、キノリン暴露後の大動脈イベントの発生と時間を特徴付け、キノロンと他の抗生物質の比較的安全性を推定し、フルオロキノロン抗生物質に暴露した後に患者が大動脈瘤または解離を経験する可能性を予測するリスク予測モデルを開発することを目的としています。彼らは、尿路感染症をキノロン抗生物質の主な使用法として使用することを提案しています。その後、暴露を定義するプロセスに焦点を当て、薬剤の成分と修正（投与経路や投与量など）を特定することが含まれます。彼らは、Atlasと成分および臨床薬剤フォームのコンセプトセットを使用して、フルオロキノロンのコードを定義する方法を示しています。

Session 2b: Gowtham Rao – Outcome Design

Phenotype
A phenotype is a specification of an observable, potentially changing state of an organism

Cohort Definition 1
Cohort Definition 2
Cohort Definition 3
Cohort definition = the specification for how to identify a cohort

A codeset is NOT a cohort...
...logic for how to use the codes in criteria is required

Cohort = a set of persons who satisfy one or more inclusion criteria for a duration of time

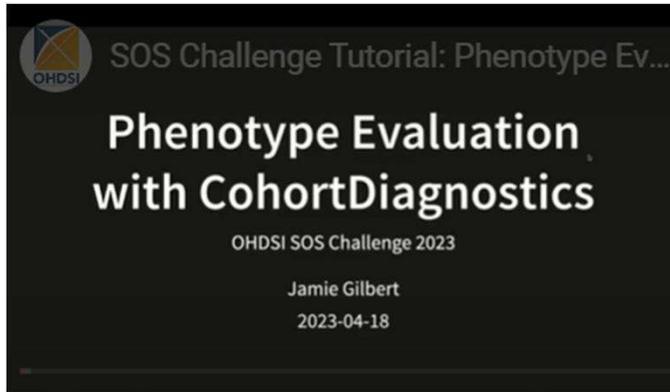
OHDSI | www.ohdsi.org | #JoinTheJourney | linkedin.com/company/ohdsi

このビデオでは、経済学における観測可能な潜在的に変化する状態の仕様であるフェノタイプの開発プロセスについて話しています。最初のステップは、フェノタイプが何であるかを理解し、コホートを特定する方法の仕様を構築することです。これは、通常、Atlasと呼ばれるOHDSIツールによってレンダリングされたSQLコードによって行われます。コホートとは、一定期間にわたって1つ以上の包含基準を満たす人々の集合です。フェノタイプは1つ以上のコード定義で表される場合があり、各コード定義はコホートに具現化されます。その後、スピーカーはフェノタイプ開発評価のプロセスについて説明し、臨床的な記述と文献ソースからの事前知識を使用して、ターゲットフェノタイプの共有理解を開発することが重要であると述べています。スピーカーはまた、フェノタイプの開発における文献レビューの重要性について説明し、iota dissectionおよびiota aneurysmのフェノタイプを評価した研究の例を提供しています。



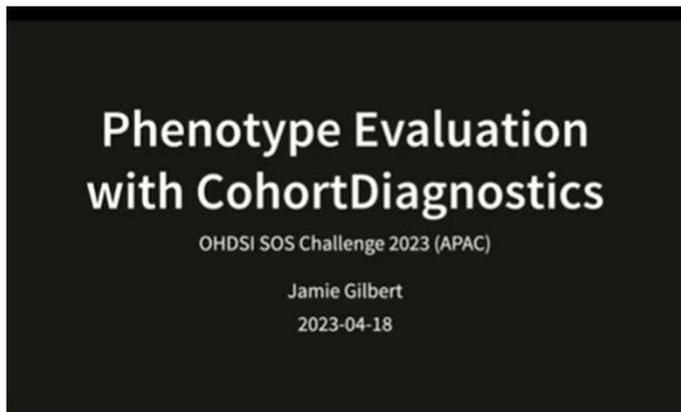
Week 4: Phenotype Evaluation

Session 1: James Gilbert, Azza Shoaibi



このビデオでは、JJのパトリックのチームのジェイミーが、シャイニーツールを使用してコホート診断を実行し、コホートを理解する方法を説明しています。目的は、パッケージに慣れ親しむことで、フェノタイプの開発を反復的に行う方法を説明することです。最初のステップは、Hadesのセットアップガイドに従ってRとR Studioをセットアップすることです。次に、コホート診断を実行するためのパッケージをセットアップする必要があります。次に、Atlasから研究定義を取得し、識別子を見つける必要があります。その後、興味のあるコホートIDのベクトルを作成し、OHDSI Web APIを使用してコホート定義セットをエクスポートする必要があります。コホートを表示した後、データベースコネクタの接続詳細を設定し、コホートジェネレータパッケージを使用してコホートを生成する必要があります。次に、コホート診断を実行し、シャイニーアプリのデータを生成する必要があります。最後に、結果データベースを作成し、シャイニーアプリを起動する必要があります。詳細な読み物やトラブルシューティングのリソースが提供されています。

Session 2: James Gilbert, Evan Minty



このビデオでは、SOS ChallengeのPhenotype Evaluationについて説明されています。前回のビデオで、ChristianとGothamが曝露のための概念セットの作成方法と、aortic aneurysmのためのコホートの作成方法を説明しました。今回は、JPEGがコホート診断を実行する方法と、aortic aneurysmフェノタイプの評価プロセスをイラストレーションする具体的な例を説明し、Evanが公開された5つのデータベースで実行されたコア診断の結果を臨床的にレビューする方法を説明します。高レベルの目標は、このコールに参加しているすべての人が自分自身でフェノタイプを評価するために、コホート診断を実行する方法に精通していることです。新しいデータソースがフルオロキノロンの研究に参加することに関連して、フェノタイプを評価することが特に重要であることがわかっています。ビデオでは、コホートジェネレータ、コホート診断、またはOHDSI Web APIなどの必要なパッケージをインストールする方法、およびコホートを取得して表示する方法について説明されています。また、Phenotype Evaluationに関するリソースについても言及されています。



Week 5: Analysis Design

Session 1: Patrick Ryan

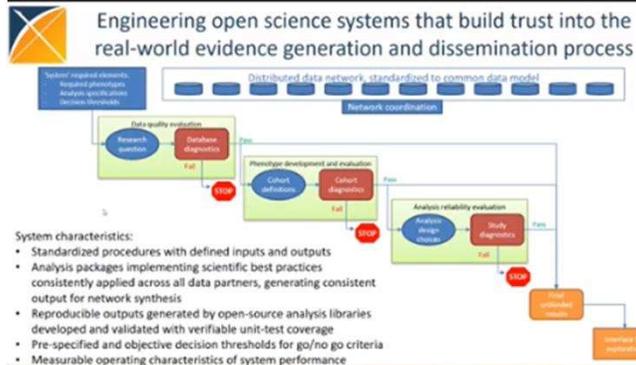


Sisyphus Challenge Week 5: Standardized analysis design

Patrick Ryan, PhD
Vice President, Observational Health Data Analytics, Johnson & Johnson
Adjunct Assistant Professor of Biomedical Informatics at Columbia University

このビデオでは、SOS Challengeというプロジェクトについて説明されています。このプロジェクトでは、信頼性の高いエビデンスを得るために、共同で研究に必要な要素を確立し、データ診断、フェノタイプ開発と評価、コホート診断、分析実行、研究診断のステップを踏んでいます。このビデオでは、データ診断からフェノタイプ開発と評価、そして現在分析設計のステップに進んでいることが説明されています。また、このプロジェクトに参加するデータパートナーの役割や、標準化された分析ツールの紹介も行われています。最後に、このプロジェクトがCindyのリサーチクエストから始まり、結果の解釈と論文の準備まで彼女が関わっていることが強調されています。

Session 2: Patrick Ryan



このビデオでは、Patrick Ryan氏がSisyphusチャレンジの分析設計についてのチュートリアルを提供しています。彼は、これまでの進捗状況と、信頼性の高いエビデンスを生成するためのオープンサイエンスプロセスを共同で設計するビジョンについて、簡単に振り返ります。そして、Sisyphusチャレンジの4つの臨床研究質問と、各研究の可能性のあるデータソースを特定するためにデータベースプロファイルツールを使用して進捗状況について説明します。次に、分析設計フェーズに移り、ストラテジストパイプライン内で現在利用可能な標準化された解析をFluoroquinolonesと大動脈瘤のリスク増加の研究に使用する方法を説明します。最後に、結果の解釈の重要性と、9週目の普及計画について強調します。



Week 6: Network Execution

Session 1: Jenna Reps, Jack Brewster



Sisyphus Challenge Week 6:
Strategus execution

Jenna/Jack
Sena/Chungsoo

このビデオでは、SOS Challenge Studyの実行方法について説明されています。まず、前回までの作業の概要を説明し、Json仕様を作成し、Strategistパッケージに接続して実行する方法を説明しています。Strategistパッケージは、実行するためのパッケージを1つにまとめ、実行環境を管理することで、再現性を確保しています。また、実行するためには、AIM CDMにデータをマッピングし、R、RStudio、Javaなどのソフトウェアをインストールする必要があります。実行手順は、プロジェクトのダウンロード、設定、実行、CSVファイルの確認、そして共有の5つのステップで構成されています。Anthony Cerner氏が開発したStrategistパッケージは、モジュール化されており、様々な解析方法に対応しています。もし15分以上かかる場合は、Anthony Cerner氏が開催するオフィスアワーに参加することをお勧めします。

Session 2: Anthony Sena, Chungsoo Kim

Sisyphus Challenge Week 6: Strategus execution



Anthony Sena
Director, Observational Health Data Analytics,
Janssen Research & Development



Chungsoo Kim
PhD Candidate, Medical Informatics,
Ajou University

このビデオでは、ストラテジストを使用してネットワークスタディを実行する方法が説明されています。前の5週間の内容を振り返り、リアルワールドエビデンスを生成するためのスキーマについて説明しました。今回は、フルオロキノロンスタディに焦点を当て、データベース診断を使用してデータベースを評価し、コホートがスタディに使用できるかどうかを決定する方法について説明しました。ストラテジストパイプラインの概念について説明し、Jsonドキュメントで表される分析設計選択肢をストラテジストに入力し、CSVファイルとデータモデルを生成する方法について説明しました。ストラテジストモジュールについて説明し、コホート生成、コホートインシデンスレートの計算、コホート診断の実行、集団レベル効果推定、患者レベル予測などの機能があることを説明しました。ストラテジストの実行に必要な前提条件について説明し、プロジェクトのダウンロード、ローカル設定の構成、スタディの実行、CSVファイルの確認、結果の共有方法について説明しました。



Week 7: Study Diagnostics

Session 1: Fan Bu, George Hripcsak



SOS Challenge Week 7:
Study Diagnostics
Anti-VEGF and kidney failure
Fan Bu and George Hripcsak

このビデオでは、スピーカーはAnti-VEGFと腎不全研究を例に挙げ、研究診断について説明しています。スピーカーは、客観的な診断セットを通じて分析の信頼性を評価することが重要であると説明しています。スピーカーは、統計的パワー、経験的等価体、共変量バランス、一般化可能性など、研究診断の概念的な部分を説明しています。スピーカーは、それぞれのコンポーネントの良い例と悪い例を示し、シャイニーアプリを使用して分析の信頼性をチェックする方法を実演しています。スピーカーは、研究結果を広めて解釈する前に診断を実施することが重要であると強調し、結果がチェリーピックされていないこと、研究者のバイアスが低減されていることを確認する必要があると述べています。

Session 2: Mitchell Conover, Nicole Pratt



SOS Challenge Week 7:
Study Diagnostics
Fluoroquinolones & Aortic Aneurisms / Aortic Dissection
Nicole Pratt & Mitch Conover

* Slides prepared by Fan Bu and George Hripcsak

このビデオでは、SOS Challengeの研究診断について、Mitchell ConoverとNicole Prattが説明しています。まず、概念的なアイデアとアプリのスクリーンショットについて説明し、その後、ライブアプリケーションを使用して初期結果を確認します。質問はチャットに投稿し、参加者はデータを提供するように奨励されます。研究診断は、研究デザインの3つのドメイン（特性、比較コホート、自己制御ケースシリーズ）に関連しています。診断には、統計的な力、比較コンパレータの類似性、一定のバイアス、時間変動的な混合、残留バイアス、ヘテロジェネティ、患者レベルの予測などが含まれます。また、研究前に診断ルールを事前に指定することで、投稿者バイアスを減らすことができます。最後に、実際の研究例を使用して、最小検出相対リスクと実証的均衡について説明しました。



Week 8: Evidence Synthesis

Session 1: Yong Chen, Martijn Schuemie

Solution – likelihood profiling

Data sites share actual shape of likelihood, instead of just the hazard ratio + confidence interval

Key ideas

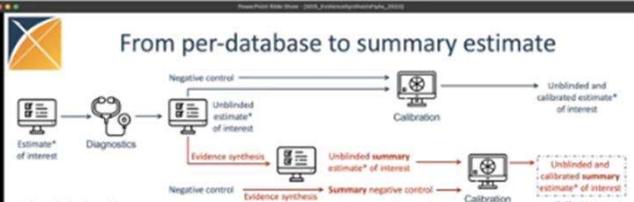
- directly approximate the shape of the per-site log likelihood function
- communicate the parameters that summarize the shape of likelihood function from each site



このビデオでは、Evidence Synthesisというパッケージについて説明されています。このパッケージは、プライバシーと規制上の理由から、異なるデータサイトで収集されたデータを共有することができない場合に使用されます。このパッケージは、異なるデータサイトで収集されたデータを分析するための方法を提供します。このパッケージでは、likelihood profilingという手法が使用されます。この手法は、異なるデータサイトで収集されたデータに基づいて、カウンターファクトを推定するために使用されます。この手法では、likelihood functionを共有することができます。この手法は、メタ分析に代わる手法として提案されています。このパッケージには、fix effect modelとrunning effects modelの2つのモデルが実装されています。このパッケージは、異なるデータサイトで収集されたデータを分析するための有効な手法として、医療研究者にとって非常に重要なものです。

Session 2: Marc Suchard

From per-database to summary estimate



Two "choices":

1. Meta-analyze calibrated estimates (once)
2. Calibrate meta-analytic estimates (many)
 - **best practice:** non-independent confounding across data-sources

* Estimate here means likelihood profile, which may not be expressed as an estimate

このビデオでは、複数のデータソースからのエビデンスを統合する方法について説明されています。まず、各データソースから得られた情報をプライバシーを保護しながら集約するために、専用の解析コードが開発されます。次に、ハザード比の最大尤度推定値と信頼区間だけでなく、尤度関数の形状を直接近似し、各データソースから得られた尤度関数のグリッド点を使用して、ベイズランダムエフェクトモデルを適用します。最後に、ベイズモデルを使用して、ポピュレーションレベルの平均値と各データソースの分散を推定し、信頼区間を計算します。この方法により、プライバシーを保護しながら、複数のデータソースからより包括的で信頼性の高い推定値を得ることができます。



Week 9: Interpreting the Results

Session 1: Cindy Cai

Wilmer Eye Institute
Johns Hopkins Medicine

Preliminary Results

Discussion rather than Tutorial

この動画では、SOS Challengeにおいて、抗血管内皮増殖因子薬（anti-VEGF）と腎不全の関係について、比較的安全な薬剤であるランダムズマブの使用を推奨するために、人口レベルの効果推定を行った過程が説明されています。9つのデータベースを用いて、約500万人の患者を対象に、ビバシズマブ、フルベルセプト、ランダムズマブの3種類の薬剤の腎不全のリスクを比較しました。データベースの選定、研究設計、プロトコル作成、データ分析、結果の解釈など、多くのステップを経て、ランダムズマブの使用が腎不全リスクを低減することが示されました。この結果は、臨床的な意思決定に役立つものであると考えられます。

Session 2: Jack Janetzki, Seng Chan You

OHDSI SOS Challenge 2023

Is fluoroquinolone use associated with the development of aortic aneurysms and aortic dissections? An international network study

Chief investigators:
Seng Chan You, Seonji Kim, Jung Ho Kim, Jung Ah Lee - Yonsei University
Jack Janetzki, Nicole Pratt - University of South Australia

YONSEI UNIVERSITY
OHDSI
Observational Health Data Science and Informatics

このビデオでは、SOS Challengeのフルオロキノロン研究の結果を解釈する方法が説明されています。まず、過去のセッションの録画がohdsi.orgのSOS Challengeに投稿されていることが紹介され、これらのリソースを活用することが推奨されています。次に、今後のステップについて説明され、データパートナーが分析パッケージを実行し、結果を共有する機会があることが強調されています。その後、研究の背景や目的、研究質問について説明され、9つのデータパートナーからの結果が紹介されました。最後に、様々なグラフや統計データが紹介され、結果の解釈についてのディスカッションが行われました。