

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 #32

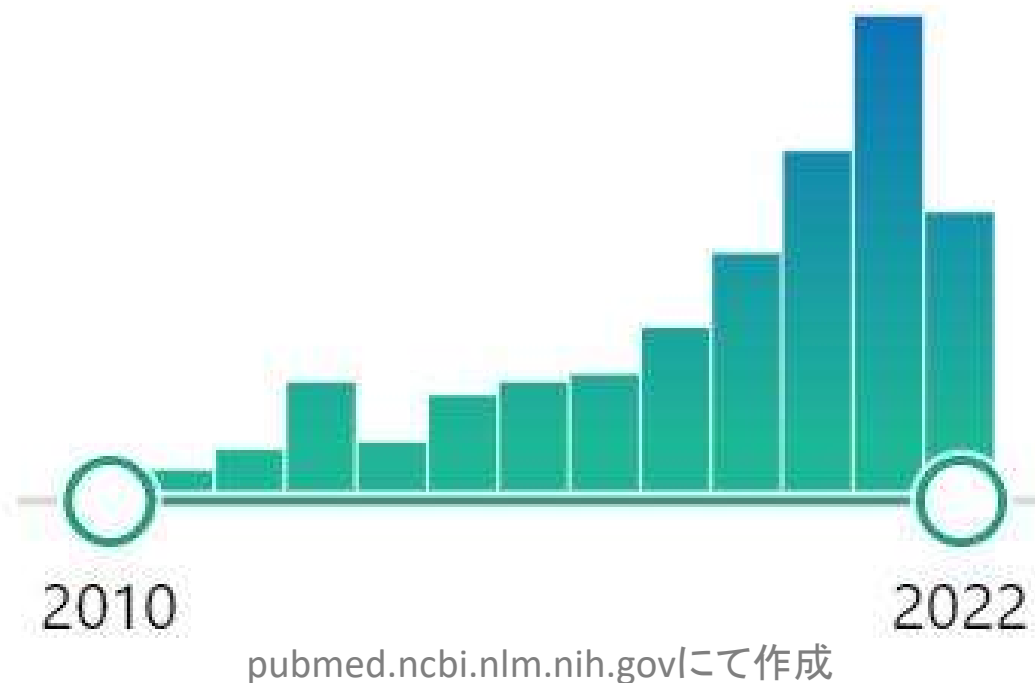
イブニング カンファレンス(第32回)

2022.7.29



OHDSI関連論文

Pubmedで“OHDSI or OMOP”を検索



全期間累計：6月269本→7月276本



COVID-19関連情報を抽出するNLPアプリを開発中

› Stud Health Technol Inform. 2022 Jun 6;290:1062-1063. doi: 10.3233/SHTI220268.

A Natural Language Processing Tool Offering Data Extraction for COVID-19 Related Information (DECOVRI)

Paul M Heider¹, Ronak M Pipaliya², Stéphane M Meystre¹

Affiliations – collapse

Affiliations

¹ Biomedical Informatics Center, Medical University of South Carolina, Charleston, SC, USA.

² College of Medicine, Medical University of South Carolina, Charleston, SC, USA.

PMID: 35673206 DOI: 10.3233/SHTI220268

Abstract

A new natural language processing (NLP) application for COVID-19 related information extraction from clinical text notes is being developed as part of our pandemic response efforts. This NLP application called DECOVRI (Data Extraction for COVID-19 Related Information) will be released as a free and open source tool to convert unstructured notes into structured data within an OMOP CDM-based ecosystem. The DECOVRI prototype is being continuously improved and will be released early (beta) and in a full version.

概要

パンデミック対策の一環として、臨床テキストノートからCOVID-19関連情報を抽出するための新しい自然言語処理（NLP）アプリケーションを開発中。DECOVRI（Data Extraction for COVID-19 Related Information）と呼ばれるこのNLPアプリケーションは、OMOP CDMベースのエコシステム内で非構造化メモを構造化データに変換するための無料のオープンソースツールとして公開される予定である。DECOVRIのプロトタイプは継続的に改良されており、早期（ベータ）および完全版としてリリースされる予定になっている。



LEGEND-T2DMの解説記事

➤ [BMJ Open. 2022 Jun 9;12\(6\):e057977. doi: 10.1136/bmjopen-2021-057977.](#)

Large-scale evidence generation and evaluation across a network of databases for type 2 diabetes mellitus (LEGEND-T2DM): a protocol for a series of multinational, real-world comparative cardiovascular effectiveness and safety studies

Rohan Khera ^{1 2}, Martijn J Schuemie ^{3 4}, Yuan Lu ^{1 2}, Anna Ostropolets ⁵, RuiJun Chen ⁶, George Hripcsak ^{5 7}, Patrick B Ryan ^{3 5}, Harlan M Krumholz ^{1 2}, Marc A Suchard ^{8 9 10 11}

Affiliations + expand

PMID: 35680274 PMCID: PMC9185490 DOI: 10.1136/bmjopen-2021-057977

LEGEND-T2DMを知りたいときに！

Abstract

Introduction: Therapeutic options for type 2 diabetes mellitus (T2DM) have expanded over the last decade with the emergence of cardioprotective novel agents, but without such data for older drugs, leaving a critical gap in our understanding of the relative effects of T2DM agents on cardiovascular risk.

Methods and analysis: The large-scale evidence generations across a network of databases for T2DM (LEGEND-T2DM) initiative is a series of systematic, large-scale, multinational, real-world comparative cardiovascular effectiveness and safety studies of all four major second-line anti-hyperglycaemic agents, including sodium-glucose co-transporter-2 inhibitor, glucagon-like peptide-1 receptor agonist, dipeptidyl peptidase-4 inhibitor and sulfonylureas. LEGEND-T2DM will leverage the Observational Health Data Sciences and Informatics (OHDSI) community that provides access to a global network of administrative claims and electronic health record data sources, representing 190 million patients in the USA and about 50 million internationally. LEGEND-T2DM will identify all adult, patients with T2DM who newly initiate a traditionally second-line T2DM agent. Using an active comparator, new-user cohort design, LEGEND-T2DM will execute all pairwise class-versus-class and drug-versus-drug comparisons in each data source, producing extensive study diagnostics that assess reliability and generalisability through cohort balance and equipoise to examine the relative risk of cardiovascular and safety outcomes. The primary cardiovascular outcomes include a composite of major adverse cardiovascular events and a series of safety outcomes. The study will pursue data-driven, large-scale propensity adjustment for measured confounding, a large set of negative control outcome experiments to address unmeasured and systematic bias.

Ethics and dissemination: The study ensures data safety through a federated analytic approach and follows research best practices, including prespecification and full disclosure of results. LEGEND-T2DM is dedicated to open science and transparency and will publicly share all analytic code from reproducible cohort definitions through turn-key software, enabling other research groups to leverage our methods, data and results to verify and extend our findings.

概要

はじめに

2型糖尿病（T2DM）に対する治療の選択肢は、心保護作用のある新規薬剤の出現により過去10年間で拡大したが、旧来の薬剤についてはそのようなデータがなく、T2DM治療薬の心血管リスクに対する相対的効果に関する我々の理解には重大なギャップが残されている。

方法と解析

- LEGEND-T2DM（2型糖尿病に関するデータベースのネットワークによる大規模なエビデンス生成）は、一連の体系的で大規模な多国間の実環境における、4つの主要な第二選択抗高血糖薬（SGLT2阻害薬、GLP-1受容体作動薬、DPP-4阻害薬、SU剤）の心血管系に対する有効性と安全性の比較調査である。
- LEGEND-T2DMは、OHDSIコミュニティを活用することで、米国で1億9千万人、海外で約5千万人の患者を含む請求データおよび電子カルテのグローバルネットワークへのアクセスを提供するものだ。
- LEGEND-T2DMは、従来の2次治療薬であるT2DM治療薬を新たに開始したすべての成人T2DM患者を対象とする。LEGEND-T2DMは、アクティブコンパレータ、新規使用者コホートデザインを用いて、各データソースにおけるクラス対クラス、薬剤対薬剤のペア比較をすべて実施し、コホートバランスと等質性による信頼性と一般性の評価、心血管アウトカムと安全性の相対リスクの検証を行う大規模調査法を作成する。
- 主要な心血管系アウトカムには、主要な心血管系有害事象の複合と一連の安全性アウトカムが含まれる。本調査では、測定された交絡に対するデータドリブンな大規模な傾向調整、測定されない偏りや系統的な偏りに対処するための大規模な陰性対照アウトカム試験が追求される。

倫理と普及

本研究は、統合された解析手法によりデータの安全性を確保し、事前指定の実施や結果の完全な開示など、研究のベストプラクティスに従ったものである。LEGEND-T2DMはオープンサイエンスと透明性を重視し、ターンキーソフトウェアによる再現可能なコホート定義から全ての解析コードを公開し、他の研究グループが我々の方法、データ、結果を活用して我々の発見を検証し、拡張することを可能にする。



非構造化文書から甲状腺癌診断とその病期
を取り出しOMOP化した。

Multicenter Study > Appl Clin Inform. 2022 May;13(3):521-531. doi: 10.1055/s-0042-1748144.

Epub 2022 Jun 15.

Transforming Thyroid Cancer Diagnosis and Staging Information from Unstructured Reports to the Observational Medical Outcome Partnership Common Data Model

Sooyoung Yoo¹, Eunsil Yoon¹, Dachung Boo¹, Borham Kim¹, Seok Kim¹, Jin Chul Paeng²,
le Ryung Yoo³, In Young Choi^{4 5}, Kwangsoo Kim⁶, Hyun Gee Ryoo^{7 8}, Sun Jung Lee^{4 5},
Eunhye Song⁹, Young-Hwan Joo¹⁰, Junmo Kim¹¹, Ho-Young Lee^{1 2}

Affiliations + expand

PMID: 35705182 PMCID: PMC9200482 DOI: 10.1055/s-0042-1748144

Abstract

Background: Cancer staging information is an essential component of cancer research. However, the information is primarily stored as either a full or semistructured free-text clinical document which is limiting the data use. By transforming the cancer-specific data to the Observational Medical Outcome Partnership Common Data Model (OMOP CDM), the information can contribute to establish multicenter observational cancer studies. To the best of our knowledge, there have been no studies on OMOP CDM transformation and natural language processing (NLP) for thyroid cancer to date.

Objective: We aimed to demonstrate the applicability of the OMOP CDM oncology extension module for thyroid cancer diagnosis and cancer stage information by processing free-text medical reports.

Methods: Thyroid cancer diagnosis and stage-related modifiers were extracted with rule-based NLP from 63,795 thyroid cancer pathology reports and 56,239 Iodine whole-body scan reports from three medical institutions in the Observational Health Data Sciences and Informatics data network. The data were converted into the OMOP CDM v6.0 according to the OMOP CDM oncology extension module. The cancer staging group was derived and populated using the transformed CDM data.

Results: The extracted thyroid cancer data were completely converted into the OMOP CDM. The distributions of histopathological types of thyroid cancer were approximately 95.3 to 98.8% of papillary carcinoma, 0.9 to 3.7% of follicular carcinoma, 0.04 to 0.54% of adenocarcinoma, 0.17 to 0.81% of medullary carcinoma, and 0 to 0.3% of anaplastic carcinoma. Regarding cancer staging, stage-I thyroid cancer accounted for 55 to 64% of the cases, while stage III accounted for 24 to 26% of the cases. Stage-II and -IV thyroid cancers were detected at a low rate of 2 to 6%.

Conclusion: As a first study on OMOP CDM transformation and NLP for thyroid cancer, this study will help other institutions to standardize thyroid cancer-specific data for retrospective observational research and participate in multicenter studies..

背景: 癌の病期情報は、癌研究に不可欠な要素である。しかし、その情報は、主に完全または半構造化された自由形式の臨床文書として保存されており、データの活用に制限がある。このようながん特有のデータをOMOP CDMに変換することで、多施設共同観察研究の確立に貢献することができる。これまで、甲状腺がんのOMOP CDM変換と自然言語処理（NLP）に関する研究は行われていないため、本研究では、甲状腺がんのOMOP CDM変換と自然言語処理（NLP）に関する研究を行う。

目的: OMOP CDM oncology 拡張モジュールの甲状腺がん診断とがんステージ情報への適用性を、フリーテキストの診療報告書进行处理することで実証することを目的とした。

方法: OHDSIデータネットワークの3医療機関の甲状腺がん病理診断報告書63,795件とヨード全身検査報告書56,239件からルールベースNLPで甲状腺がん診断とステージ関連修飾子を抽出した。データはOMOP CDM oncology extension moduleに従い、OMOP CDM v6.0に変換された。変換されたCDMデータを用いて、癌の病期分類が導き出され、入力された。

結果: 抽出された甲状腺癌のデータは、OMOP CDMに完全に変換された。甲状腺癌の病理組織型分布は、乳頭癌95.3～98.8%、濾胞癌0.9～3.7%、腺癌0.04～0.54%、髄様癌0.17～0.81%、退形成癌0～0.3%程度であり、病理学的に有意な差が認められた。癌の病期分類では、甲状腺癌はI期が55～64%、III期が24～26%であった。II期およびIV期の甲状腺がんは2～6%と低い発見率であった。

結論: 甲状腺癌のOMOP CDM変換とNLPに関する最初の研究として、本研究は、他の施設が甲状腺癌特有のデータを標準化し、後ろ向き観察研究、多施設研究に参加するのに役立つと思われる。



ATLASコホート定義を各OMOP-CDMサイトで実行し
人数を返す機能をドイツの既存コンソーシアム内
で作ってみた。

> [Stud Health Technol Inform.](#) 2022 Jun 29;295:515-516. doi: 10.3233/SHTI220778.

An OHDSI ATLAS Extension to Support Feasibility Requests in a Research Network

Ines Reinecke ¹, Mirko Gruhl ¹, Martin Pinnau ², Fatma Betül Altun ², Michael Folz ²,
Michèle Zoch ¹, Franziska Bathelt ¹, Martin Sedlmayr ¹

Affiliations – collapse

Affiliations

- ¹ Carl Gustav Carus Faculty of Medicine, Center for Medical Informatics, Institute for Medical Informatics and Biometry, Technische Universität Dresden, Dresden, Germany.
- ² Institute of Medical Informatics, Goethe University Frankfurt, University Hospital Frankfurt, Frankfurt am Main, Germany.

PMID: 35773924 DOI: 10.3233/SHTI220778

Abstract

Checking the feasibility of real-world data to answer a certain research question is crucial especially in a multi-site research network. In this work we present an extension of the ATLAS user interface for the OMOP common data model that integrates into an existing national feasibility network and thus foster capabilities for future participation in international research studies.

概要

ある研究課題に答えるために実世界のデータの実現可能性をチェックすることは、特に複数サイトの研究ネットワークにおいて重要である。この研究では、ATLASのユーザーインターフェースを拡張し、既存の国内フィージビリティネットワークに統合することで、将来の国際研究への参加に必要な能力を育成することを提案する。

- OHDSI ATLASの機能は、実現可能性リクエストのサイト間データ共有をサポートする既存のアーキテクチャに基づき、このアーキテクチャ内でコホート定義を作成・配布する中央コンポーネントとして機能するよう拡張される必要がありました。
- まず、中央プロジェクト提案モジュールProSkiveによって追加された情報を抽出し、一意の識別子を含むSQLスクリプトとしてエクスポートされたコホート定義を転送するために必要な中央ブローカーのアドレスを特定するためにATLASが拡張されました。
- 次に、ATLASのユーザーインターフェースにエクスポート機能を作成し、コホート定義を実行可能なSQLステートメントとして集中ブローカーインスタンスに自動的に送信しました。
- 基本クライアント認証 中央の ATLAS インスタンスとブローカー間の信頼できる通信を保証する、基本的なクライアント認証のセキュリティも実装されました。
- そのため、各 MIRACUM サイトのConnector Component Federated Search (CCFS) コンポーネントは、ブローカーから SQL 文を取得することができます。
- OMOPデータベースに対してローカルでSQL文を実行し、ネットワーク上の各サイトから利用可能なレコードの数を返信します。



季節性変動検出を各種手法で試した

Observational Study > BMC Med Res Methodol. 2022 Jul 2;22(1):182.

doi: 10.1186/s12874-022-01652-3.

Empirical assessment of alternative methods for identifying seasonality in observational healthcare data

Anthony Molinaro ¹, Frank DeFalco ²

Affiliations – collapse

Affiliations

¹ Janssen Research and Development, 1125 Trenton Harbourton Rd, Titusville, NJ, 08560, USA.
amoln19@its.jnj.com.

² Janssen Research and Development, 1125 Trenton Harbourton Rd, Titusville, NJ, 08560, USA.

PMID: 35780114 PMCID: PMC9250712 DOI: 10.1186/s12874-022-01652-3

Background: Seasonality classification is a well-known and important part of time series analysis. Understanding the seasonality of a biological event can contribute to an improved understanding of its causes and help guide appropriate responses. Observational data, however, are not comprised of biological events, but timestamped diagnosis codes the combination of which (along with additional requirements) are used as proxies for biological events. As there exist different methods for determining the seasonality of a time series, it is necessary to know if these methods exhibit concordance. In this study we seek to determine the concordance of these methods by applying them to time series derived from diagnosis codes in observational data residing in databases that vary in size, type, and provenance.

Methods: We compared 8 methods for determining the seasonality of a time series at three levels of significance (0.01, 0.05, and 0.1), against 10 observational health databases. We evaluated 61,467 time series at each level of significance, totaling 184,401 evaluations.

Results: Across all databases and levels of significance, concordance ranged from 20.2 to 40.2%. Across all databases and levels of significance, the proportion of time series classified seasonal ranged from 4.9 to 88.3%. For each database and level of significance, we computed the difference between the maximum and minimum proportion of time series classified seasonal by all methods. The median within-database difference was 54.8, 34.7, and 39.8%, for $p < 0.01$, 0.05, and 0.1, respectively.

Conclusion: Methods of binary seasonality classification when applied to time series derived from diagnosis codes in observational health data produce inconsistent results. The methods exhibit considerable discord within all databases, implying that the discord is a result of the difference between the methods themselves and not due to the choice of database. The results indicate that researchers relying on automated methods to assess the seasonality of time series derived from diagnosis codes in observational data should be aware that the methods are not interchangeable and thus the choice of method can affect the generalizability of their work. Seasonality determination is highly dependent on the method chosen.

背景: 季節性の分類は、時系列解析の重要な部分としてよく知られている。生物学的事象の季節性を理解することは、その原因に対する理解を深め、適切な対応を導くのに役立つ。しかし、観測データは生物学的事象から構成されているのではなく、タイムスタンプされた診断コードの組み合わせ（追加要件あり）が生物学的事象のプロキシとして使用されている。時系列の季節性を決定する方法は様々であり、これらの方法が一致するかどうかを知ることが必要である。本研究では、サイズ、タイプ、出所が異なるデータベースに保存されている観測データの診断コードに由来する時系列にこれらの方法を適用することで、これらの方法の一致性を明らかにすることを目的とする。

方法: 時系列の季節性を判定する8つの方法を、3つの有意水準（0.01、0.05、0.1）で、10の観測的健康データベースに対して比較した。各有意水準で61,467の時系列を評価し、合計184,401の評価を行った。

結果: すべてのデータベースと有意水準において、一致度は20.2%から40.2%の範囲であった。すべてのデータベースと有意水準において、季節性と判定された時系列の割合は、4.9%から88.3%の範囲であった。各データベースと有意水準について、すべての手法で季節性があると分類された時系列の割合の最大値と最小値の差を計算した。データベース内の差の中央値は、 $p < 0.01$ 、0.05、0.1において、それぞれ54.8、34.7、39.8%であった。

結論: 観察的健康データの診断コードに由来する時系列にバイナリ季節性分類の方法を適用した場合、一貫性のない結果をもたらす。この結果は、データベースの選択によるものではなく、手法の違いによるものであることを示唆している。この結果は、観察データの診断コードから得られる時系列の季節性を評価するために自動化された手法に依存している研究者は、その手法が互換性がなく、したがって手法の選択が研究の一般化可能性に影響することを認識すべきであることを示している。季節性の判定は、選択した手法に大きく依存する。



OMOP ボキャブラリのコンセプトマッピングを利用して国内の異コードDBから共通概念を特定した。

➤ [JMIR Form Res. 2022 May 17. doi: 10.2196/37821. Online ahead of print.](#)

Methodological issues for using a common data model (CDM) of COVID-19 vaccine uptake and important adverse events of interest (AEIs): the Data and Connectivity COVID-19 Vaccines Pharmacovigilance (DaC-VaP) United Kingdom feasibility study

Gayathri Delanerolle ¹, Robert Williams ¹, Ana Stipancic ², Rachel Byford ¹, Anna Forbes ¹, Sneha Anand ¹, Declan Bradley ³, Ruby Tsang ¹, Siobhán Murphy ³, Ashley Akbari ⁴, Stuart Bedston ⁴, Ronan Lyons ⁴, Rhiannon Owen ⁴, Jillian Beggs ⁵, Antony Chuter ⁵, Dominique Balharry ⁵, Mark Joy ¹, Aziz Sheikh ⁵, F D Richard Hobbs ¹, Simon de Lusignan ⁶

Affiliations + expand

PMID: 35786634 DOI: 10.2196/37821

Background: The Data and Connectivity COVID-19 Vaccines Pharmacovigilance (DaC-VaP) UK-wide collaboration was created to monitor vaccine uptake and effectiveness and provide pharmacovigilance using routine clinical and administrative data. To monitor these, pooled analyses may be needed. However, variation in terminologies present a barrier as, England uses the Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT), while the rest of the UK uses the Readv2 terminology in primary care. The availability of data sources is not uniform across the UK.

Objective: To use the concept mappings in the Observational Medical Outcomes Partnership (OMOP) common data model (CDM) to identify common concepts recorded and to report these in a repeated cross-sectional study. We plan to do this for vaccine coverage and two adverse events of interest (AEIs), cerebral venous sinus thrombosis (CVST) and anaphylaxis. We identified concept mappings to SNOMED CT, Readv2, the World Health Organisation's International Classification of Disease version 10 (ICD-10) terminology and the UK's Dictionary of Medicines and Devices (dm+d).

Methods: Exposures and outcomes of interest to DaC-VaP for pharmacovigilance studies were selected. Mappings of these variables to different terminologies used across the UK devolved nations health services were identified from the Observational Health Data Sciences and Informatics (OHDSI) Athena online browser. Lead analysts from each nation then confirm or add to the mappings identified. These mappings will then be used to report AEIs in a common format. We will report rates for windows of 0-2 days and 3-28 days post-vaccine every 28 days.

Results: We list the mappings between Read v2, SNOMED CT, ICD-10 and dm+d. For vaccine exposure, we found clear mapping from OMOP to our clinical terminologies, though dm+d had codes not listed by OMOP at the time of searching. We found a list of CVST and anaphylaxis codes. For CVST we had to use a broader cerebral venous thrombosis conceptual approach to include Read v2. We identified 56 SNOMED clinical terms from which we selected 47, and 15 Read v2 codes. For anaphylaxis, our refined search identified 60 SNOMED codes and 9 from Read v2, from which we selected 10 and 4 clinical terms to include in our repeated cross-sectional studies.

Conclusions: This approach enables the use of mappings to different terminologies within the OMOP CDM without the need to catalogue an entire database. However, Read v2 has less granular concepts than some terminologies such as SNOMED CT. Additionally, the OMOP CDM cannot compensate for limitations in the clinical coding system. Neither Read v2 nor ICD-10 are sufficiently granular to enable CVST to be specifically flagged. Hence, any pooled analysis will have to be at the less specific level of cerebrovascular venous thrombosis. Overall the mappings within this CDM are useful, and our method could be used for rapid collaborations where there are only a limited number of concepts to pool.

背景: Data and Connectivity COVID-19 Vaccines Pharmacovigilance (DaC-VaP) UK-wide collaborationは、ワクチンの摂取と効果をモニターし、日常的な臨床・行政データを用いてファーマコビジランスを提供するために創設されたものである。これらのモニタリングには、プール解析が必要であろう。しかし、イングランドではSNOMED CTを使用しているが、イギリスの他の地域ではプライマリーケアでRead v2用語を使っているため、用語の違いが障壁になっている。また、データソースの利用可能性も英国内で一様ではない。

目的: OMOP CDMのコンセプトマッピングを利用して、記録された共通概念を特定し、これを反復横断研究で報告すること。我々は、ワクチン接種率と、脳静脈洞血栓症（CVST）とアナフィラキシーという2つの関心のある有害事象（AEI）についてこれを行う予定である。SNOMED CT、Read v2、世界保健機関（WHO）の国際疾病分類第10版（ICD-10）用語、英国の医薬品・機器辞書（dm+d）に対するコンセプトマッピングを確認した。

方法: ファーマコビジランス研究のためにDaC-VaPが関心を持つ曝露と転帰を選択した。これらの変数と連合各国の医療サービスで使用されるさまざまな用語のマッピングは、OHDSI Athenaオンラインブラウザから確認した。その後、各国の主任分析者が、特定されたマッピングを確認または追加する。これらのマッピングは、共通のフォーマットでAEIを報告するために使用されます。我々は、ワクチン後0-2日および3-28日のウィンドウのレートを28日ごとに報告する予定である。

結果: Read v2、SNOMED CT、ICD-10、dm+dのマッピングをリストアップした。ワクチン曝露については、OMOPから我々の臨床用語への明確なマッピングが見つかったが、dm+dには検索時にOMOPでリストアップされていないコードがあった。CVSTとアナフィラキシーのコードはリストアップされていた。CVSTについては、Read v2を含めるために、より広範な脳静脈血栓症の概念的なアプローチを用いる必要があった。56のSNOMED臨床用語を確認し、その中から47、Read v2コード15を選択しました。アナフィラキシーについては、精緻な検索により60のSNOMEDコードとRead v2の9つのコードを特定し、そこから10と4の臨床用語を選択して、反復横断的研究に含めることができました。

結論: このアプローチにより、データベース全体を目録化することなく、OMOP CDM内の異なる専門用語へのマッピングを使用することができる。しかし、Read v2はSNOMED CTのようないくつかの用語集に比べて、概念の粒度が大きい。さらに、OMOP CDMは臨床コーディングシステムの制限を補うことができない。Read v2もICD-10も、CVSTに特異的なフラグを付けるには粒度が不十分である。従って、プール解析は、脳血管性静脈血栓症というあまり特定のでないレベルで行わなければならない。全体として、このCDMのマッピングは有用であり、我々の方法は、プールする概念の数が限られている場合、迅速な共同作業に使用することができる。



再識別リスクに関するアンケート調査

Observational Study > J Korean Med Sci. 2022 Jul 4;37(26):e205. doi: 10.3346/jkms.2022.37.e205.

Perceived Risk of Re-Identification in OMOP-CDM Database: A Cross-Sectional Survey

Yae Won Tak ^{# 1}, Seng Chan You ^{# 2}, Jeong Hyun Han ¹, Soon-Seok Kim ³, Gi-Tae Kim ⁴, Yura Lee ⁵

Affiliations – collapse

Affiliations

- 1 Department of Information Medicine, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Korea.
 - 2 Department of Biomedical Systems Informatics, Yonsei University College of Medicine, Seoul, Korea.
 - 3 Department of Big Data Science, Halla University, Wonju, Korea.
 - 4 UPS Data Corporation, Seoul, Korea.
 - 5 Department of Information Medicine, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Korea. haepary@naver.com.
- # Contributed equally.

Background: The advancement of information technology has immensely increased the quality and volume of health data. This has led to an increase in observational study, as well as to the threat of privacy invasion. Recently, a distributed research network based on the common data model (CDM) has emerged, enabling collaborative international medical research without sharing patient-level data. Although the CDM database for each institution is built inside a firewall, the risk of re-identification requires management. Hence, this study aims to elucidate the perceptions CDM users have towards CDM and risk management for re-identification.

Methods: The survey, targeted to answer specific in-depth questions on CDM, was conducted from October to November 2020. We targeted well-experienced researchers who actively use CDM. Basic statistics (total number and percent) were computed for all covariates.

Results: There were 33 valid respondents. Of these, 43.8% suggested additional anonymization was unnecessary beyond, "minimum cell count" policy, which obscures a cell with a value lower than certain number (usually 5) in shared results to minimize the liability of re-identification due to rare conditions. During extract-transform-load processes, 81.8% of respondents assumed structured data is under control from the risk of re-identification. However, respondents noted that date of birth and death were highly re-identifiable information. The majority of respondents (n = 22, 66.7%) conceded the possibility of identifier-contained unstructured data in the *NOTE* table.

Conclusion: Overall, CDM users generally attributed high reliability for privacy protection to the intrinsic nature of CDM. There was little demand for additional de-identification methods. However, unstructured data in the CDM were suspected to have risks. The necessity for a coordinating consortium to define and manage the re-identification risk of CDM was urged.

背景: 情報技術の進歩により、健康データの質と量は膨大になった。そのため、観察研究が増加するとともに、プライバシー侵害の脅威も生じている。近年、共通データモデル (CDM) に基づく分散型研究ネットワークが登場し、患者レベルのデータを共有することなく国際共同医学研究を行うことが可能となった。各機関のCDMデータベースはファイアウォール内に構築されているが、再識別のリスクは管理が必要である。そこで本研究では、CDM利用者がCDMと再識別のリスク管理に対してどのような認識を持っているかを明らかにすることを目的とする。

方法: CDMに関する具体的な深層的な質問に答えることを目的とした調査を、2020年10月から11月にかけて実施した。CDMを積極的に利用している経験豊富な研究者を対象とした。すべての共変量について基本統計量 (総数および割合) を算出した。

結果: 有効回答者は33名であった。このうち、43.8%が「セル最小数」ポリシー以上の匿名化は不要と回答している。ETLプロセスにおいて、81.8%の回答者が、構造化データは再識別のリスクから管理されていると想定している。しかし、回答者は、生年月日と死亡が再識別可能性の高い情報であることを指摘している。大多数の回答者 (n = 22, 66.7%) は、NOTEテーブルに識別子を含む非構造化データが含まれる可能性を認めている。

結論: 全体として、CDM利用者は、プライバシー保護に対する高い信頼性をCDMの本質的な性質に起因すると考えている。また、追加の非識別化手法に対する要望は少なかった。しかし、CDM の非構造化データにはリスクがあることが疑われた。CDMの再識別リスクを定義し、管理するための調整用コンソーシアムの必要性が強く要請された。



トラマドールとコデインでの
股関節骨折のリスク比較

Observational Study > Drug Saf. 2022 Jul;45(7):791-807. doi: 10.1007/s40264-022-01198-9.

Epub 2022 Jul 9.

Hip Fracture Risk After Treatment with Tramadol or Codeine: An Observational Study

Erica A Voss^{1 2 3}, Saberi Rana Ali⁴, Arun Singh⁴, Peter R Rijnbeek^{5 6}, Martijn J Schuemie^{4 5}, Daniel Fife⁴

Affiliations + expand

PMID: 35810265 PMCID: PMC9296392 DOI: 10.1007/s40264-022-01198-9

Introduction: Hip fractures among older people are a major public health issue, which can impact quality of life and increase mortality within the year after they occur. A recent observational study found an increased risk of hip fracture in subjects who were new users of tramadol compared with codeine. These drugs have somewhat different indications. Tramadol is indicated for moderate to severe pain and can be used for an extended period; codeine is indicated for mild to moderate pain and cough suppression.

Objective: In this observational study, we compared the risk of hip fracture in new users of tramadol or codeine, using multiple databases and analytical methods.

Methods: Using data from the Clinical Practice Research Datalink and three US claims databases, we compared the risk of hip fracture after exposure to tramadol or codeine in subjects aged 50-89 years. To ensure comparability, large-scale propensity scores were used to adjust for confounding.

Results: We observed a calibrated hazard ratio of 1.10 (95% calibrated confidence interval 0.99-1.21) in the Clinical Practice Research Datalink database, and a pooled estimate across the US databases yielded a calibrated hazard ratio of 1.06 (95% calibrated confidence interval 0.97-1.16).

Conclusions: Our results did not demonstrate a statistically significant difference between subjects treated for pain with tramadol compared with codeine for the outcome of hip fracture risk.

はじめに: 高齢者の股関節骨折は公衆衛生上の大きな問題であり、QOLに影響を与え、発生後1年以内の死亡率を増加させる可能性がある。最近の観察研究では、トラマドールを初めて使用する被験者では、コデインと比較して股関節骨折のリスクが高いことが明らかになった。これらの薬剤は、適応症がやや異なる。トラマドールは中等度から重度の疼痛に適応があり、長期間使用することができる。コデインは軽度から中等度の疼痛および咳の抑制に適応がある。

目的: この観察研究では、複数のデータベースと分析方法を用いて、トラマドールまたはコデインの新規使用者における股関節骨折のリスクを比較した。

方法: Clinical Practice Research Datalinkと米国の3つの請求データベースのデータを用いて、50～89歳の被験者を対象に、トラマドールまたはコデインに曝露した後の股関節骨折のリスクを比較検討した。比較可能性を確保するため、交絡を調整するために大規模傾向スコアを用いた。

結果: Clinical Practice Research Datalinkデータベースでは校正済みハザード比1.10（95%校正済み信頼区間0.99-1.21）を観察し、米国のデータベースでプールした推定では校正済みハザード比1.06（95%校正済み信頼区間0.97-1.16）であった。

結論: 股関節骨折のリスクというアウトカムについて、コデインと比較してトラマドールで疼痛治療を受けた被験者の間に統計的に有意な差は示されなかった。



健康の社会的決定要因データを
標準の語彙にマッピングするための実用的なアプローチ

> AMIA Annu Symp Proc. 2022 May 23;2022:396-405. eCollection 2022.

Advancing Interoperability of Patient-level Social Determinants of Health Data to Support COVID-19 Research

Jimmy Phuong^{1 2}, Stephanie Hong³, Matvey B Palchuk⁴, Juan Espinoza⁵, Daniella Meeker⁶, David A Dorr⁷, Galina Lozinski⁸, Charisse Madlock-Brown⁹, William G Adams⁸

Affiliations + expand

PMID: 35854720 PMID: PMC9285174

Including social determinants of health (SDoH) data in health outcomes research is essential for studying the sources of healthcare disparities and developing strategies to mitigate stressors. In this report, we describe a pragmatic design and approach to explore the encoding needs for transmitting SDoH screening tool responses from a large safety-net hospital into the National Covid Cohort Collaborative (N3C) OMOP dataset. We provide a stepwise account of designing data mapping and ingestion for patient-level SDoH and summarize the results of screening. Our approach demonstrates that sharing of these important data - typically stored as non-standard, EHR vendor specific codes - is feasible. As SDoH screening gains broader use nationally, the approach described in this paper could be used for other screening instruments and improve the interoperability of these important data.

健康アウトカム研究に健康の社会的決定要因（SDoH）データを含めることは、医療格差の原因を研究し、ストレス要因を軽減するための戦略を開発するために不可欠である。本報告では、大規模なセーフティネット病院からSDoHスクリーニングツールの回答をNational Covid Cohort Collaborative (N3C) OMOPデータセットに転送するためのエンコーディングニーズを調査するための実用的なデザインとアプローチについて説明する。

我々は、患者レベルのSDoHのためのデータマッピングと取り込みを設計するための段階的な説明を提供し、スクリーニングの結果を要約している。私たちのアプローチは、通常、非標準のEHRベンダー固有のコードとして保存されているこれらの重要なデータの共有が可能であることを実証している。SDoHスクリーニングが全国的に広く使用されるようになると、この論文で説明したアプローチは、他のスクリーニング機器にも使用され、これらの重要なデータの相互運用性を向上させることができるようになる。



最近のGlobal OHDSI話題から

■ OHDSI Global Symposium

・ 10月14-16日

Oct. 14 – Main Conference

Oct. 15 – Tutorial

Oct. 15-16 – Workgroup Activities

■ OHDSI APAC Symposium

11月12-13日に台北で。

(現在のプログラム案)

12日 プレ日 チュートリアルday

13日 本開催 9:00-17:00

Nov 13

Opening Remarks

Group Photo

Key note - OHDSI Global Presentation

OHDSI APAC Intro

Researches in OHDSI APAC

Researches using Taiwan national data

Researches using TMUCRD data

Lunch & ePoster Presentation

Panel - Standardization & common data models

APAC Regional adaption to standardization

Poster session

Closing remarks



最近のGlobal OHDSI話題から

7月のCommunity Call

●APAC Community Call テーマ

June 30 European Symposium Recap
Regional Chapter Mid-Year Updates (Japan, Korea, China)

July 14 APAC Study Quarterly Updates, Part 1
(1) Characterization of Health by OHDSI AP chapter to identify Temporal Effect of the Pandemic. (2)Data quality of OHDSI APAC: CDM Inspection study.

●Global Community Call テーマ

July 26 CDM Update Process

July 19 Workgroup Updates
(Early-Stage Researchers, NLP, Phenotype Evaluation & Development, HADES)

July 12 New Adopter/Community Member Introductions

June 28 European Symposium Recap



疑似データ作りました

Synthea生成データをもとに造った日本風データ

GitHubに掲載

<https://github.com/ohdsi-japan/sample-data-ja>

allergies.csv	アレルギー情報
conditions.csv	傷病名
encounters.csv	受診
imaging_studies.csv	画像検査
medications.csv	医薬品
observations.csv	検体検査
patients.csv	患者基本情報
procedures.csv	処置
supplies.csv	器材

とりあえずミニ版で115人分



日本疑似データ 三二版

中はこんなデータ#1/3

allergies.txt

レセプト傷病名コード

START	STOP	PATIENT	ENCOUNTER	CODE	DESCRIPTION	ICD10	CATEGORY
2013-05-08		pt023	en0162	850458	真菌アレルギー	T784	環境
2013-05-08		pt023	en0162	8848611	ダニアレルギー	T784	環境
2013-05-08		pt023	en0162	8850550	動物アレルギー	T784	環境

conditions.txt

レセプト傷病名コード

START	STOP	PATIENT	ENCOUNTER	CODE	DESCRIPTION	ICD10
2012-08-04	2012-09-24	pt034	en5129	3829010	中耳炎	H669
2012-10-18	2012-10-18	pt034	en4579	2050003	急性骨髄性白血病	C920
2012-10-18	2012-10-18	pt034	en2187	8842350	発熱性好中球減少症	D70

encounters.txt

Id	START	STOP	PATIENT	ENCOUNTER	CLASS
en5129	2012-08-04	2012-08-04	pt034		外来
en4579	2012-09-24	2012-09-24	pt034		健診
en2187	2012-10-18	2012-10-19	pt034		入院



日本疑似データ 三二版

中はこんなデータ#2/3

imaging_studies.txt

Id	DATE	PATIENT	ENCOUNTER	SERIES_UID	BODYSITE	MODALITY_DESCRIPTION
im30	2020-03-29	pt063	en2599	1.2.840.999	上腕	デジタルX線撮影
im08	2013-11-04	pt105	en2704	1.2.840.999	骨盤	デジタルX線撮影
im56	2019-05-03	pt054	en2546	1.2.840.999	鎖骨部	デジタルX線撮影

medications.txt

厚労省薬価基準コードの上9桁

START	STOP	PATIENT	ENCOUNTER	CODE	DESCRIPTION	DISPENSES
2012-08-04	2012-08-18	pt034	en5129	6131001M2	アモキシシリンカプセル250mg	1
2012-08-04	2012-08-18	pt034	en5129	1141007R1	アセトアミノフェンDS小児用20%	1
2012-12-04	2012-12-12	pt034	en4622	6131001M1	アモキシシリンカプセル125mg	1

observations.txt

LOINCコード

DATE	PATIENT	ENCOUNTER	CODE	CATEGORY	DESCRIPTION	VALUE	UNITS
2012-09-24	pt034	en4579	8302-2	バイタル	身長	95.9	cm
2012-09-24	pt034	en4579	29463-7	バイタル	体重	15.3	kg
2012-09-24	pt034	en4579	39156-5	バイタル	BMI	16.7	kg/m2



日本疑似データ 三二版

中はこんなデータ#3/3

patients.txt

ランダム名+数字

Id	BIRTHDATE	DEATHDATE	SEI	MEI	GENDER	PREFECTURE	CITY	ZIP
pt034	2009-10-19		森山6	功935	M	東京都	昭島市	196-0015
pt030	1978-12-06		千葉3	友雛595	F	東京都	国立市	186-0001
pt027	2021-03-25		谷2	由紀305	F	東京都	港区	108-0014

procedures.txt

START	STOP	PATIENT	ENCOUNTER	DESCRIPTION
2012-09-24	2012-09-24	pt034	en4579	医薬品確認調整
2012-10-18	2012-10-18	pt034	en2187	化学療法
2013-03-25	2013-03-25	pt034	en1004	医薬品確認調整

supplies.txt

DATE	PATIENT	ENCOUNTER	DESCRIPTION	QUANTITY
2021-04-06	pt051	en4504	Disposable Medical Mask	2
2021-04-06	pt051	en4504	Nitrile Glove (Powder-free/Antibacterial)	24
2021-04-06	pt051	en4504	Protective Gown (Single-use)	12



何が大変だったか

やはり用語マッピング（日本化）

- ひとつひとつ確認が必要
病名・医薬品・検査 それぞれ百数十個といえども..
- 日本にない医薬品をどうするか
→今回は疑似データなので
類似した別の医薬品に置き換えた。
- 日本のコード体系と粒度が一致しない検査をどうするか
→今回はあきらめてLOINC粒度のまま。
- 日本にない診療概念をどうするか（医療制度が異なる）
→日本にはないデータ種類は削除した。
- 用語コードそのものは問題にならなかった。
→コードを使うところではない変換が必要なため
(意味の読み取りが必要)