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 Japan evening conference #22

イブニング カンファレンス(第22回)

2021.9.29

JAPAN OHDSI オデッセイ ジャパン
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

OHDSI 2021 Global Symposium

https://www.ohdsi.org/2021-ohdsi-global-symposium/

Day 1: OHDSI's Impact on the COVID-19 Pandemic

Plenary Presentations

The Sept. 14 plenary during the 2021 OHDSI Symposium featured the State of the Community presentation, updates from the global OHDSI network, and then a series of talks about OHDSI's impact on the COVID-19 pandemic.

0:00 – OHDSI State of the Community (George Hripcsak)
20:26 – OHDSI network update: US (Kristin Kostka)
27:47 – OHDSI network update: Europe (Peter Rijnbeek)
36:18 – OHDSI network update: Asia Pacific (Mui Van Zandt)
44:06 – Incidence rate method sensitivity and anchoring (Anna Ostropolets)

Watch the Day 1 Ple

Collaborator Showcase

The Collaborator Showcase received a record number of submissions this year, and following a peer-review process, there were more than 100 submissions of posters, software demos, and lightning talks that comprised the 2021 Collaborator Showcase. The presentations featured OHDSI advancements in **observational data standards and management**, **methodological research**, **open-source analytics development**, and **clinical applications**.

See All Posters, Software Demos and Lightning Talks from The 2021 Collaborator Showcase

Lightning Talks

Observational Data Standards and Management

Day 2: The Journey To Reliable Evidence

Plenary Presentations

The Sept. 15 plenary during the 2021 OHDSI Symposium focused on the journey to reliable evidence. The first portion of the plenary included four presentations, which are time-stamped below.

0:00 – Glucagon-Like Peptide 1 Receptor Agonists and Chronic Lower Respiratory Disease Exacerbations Among Patients With Type 2 Diabetes (Yasser Albogami)
ons From The OHDSI Reproducibility Workshop (Anna
ourney to Reliable Evidence: Reproducibility (Mitchell
ourney to Reliable Evidence: Generalizability (Christian

Watch the Day 2 Plenary Talks Here

Reaction Panel

plenary talks on "The Journey to Reliable Evidence," we to welcome three leaders in the field to engage in a anel. The panelists and moderator are listed below:

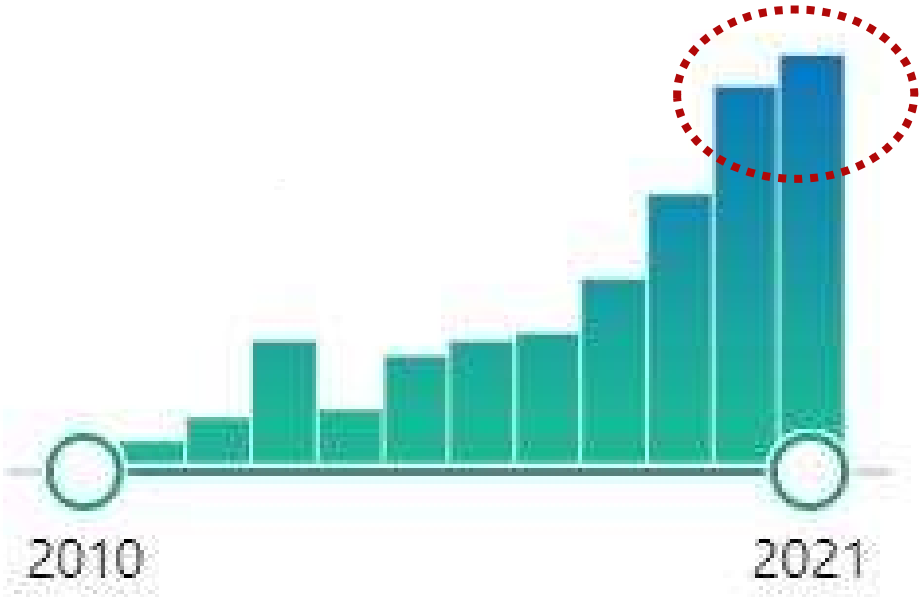
Wang, PhD • Assistant Professor of Medicine at Harvard ol; Associate Epidemiologist in the Division of demiology and Pharmacoeconomics at Brigham and spital

hammari, PhD, MS, RPH • Advisor, Saudi Food and y Executive President; Associate Professor of Pharmacy ior researcher, Medication Safety Research Chair, King ity. VP, ISPE Middle East (Saudi Arabia)

Watch the Day 2 Reaction Panel Here

OHDSI関連論文

Pubmedで”OHDSI or OMOP”を検索



pubmed.ncbi.nlm.nih.govにて作成

全期間累計：8月210本→９月215本

OHDSI論文 新規1

Comput Methods Programs Biomed. 2021 Sep 6;211:106394. doi: 10.1016/j.cmpb.2021.106394.
Online ahead of print.

A standardized analytics pipeline for reliable and rapid development and validation of prediction models using observational health data

Sara Khalid [1], Cynthia Yang [2], Clair Blacketer [3], Talita Duarte-Salles [4], Sergio Fernández-Bertolín [4], Chungsoo Kim [5], Rae Woong Park [6], Jimyung Park [5], Martijn J Schuemie [3], Anthony G Sena [7], Marc A Suchard [8], Seng Chan You [9], Peter R Rijnbeek [2], Jenna M Reps [10]

USA (11)	EUROPE (8)	ASIA-PACIFIC (3)
Columbia University (NY – EHR)	CPRD (UK – EHR)	HIRA (South Korea – Administrative Claims)
Department of Veterans Affairs (National – EHR)	DA Germany (Germany – EHR)	DCMC (South Korea – EHR)
HealthVerity (Claims linked to diagnostic testing)	HM Hospitales (Spain – Hospital Billing)	Nanfang Hospital (China – EMR)
IQVIA Open Claims (National – Administrative Claims)	IPCI (Netherlands – EHR)	
Optum EHR (National – EHR)	LPD France (France – EHR)	
Optum SES (National – EHR linked to Socio-economic data)	LPD Italy (Italy – EHR)	
Premier (National – Hospital Billing)	SIDIAP (Spain – EHR)	
Stanford University (CA – EHR)	SIDIAP-H (Spain – EHR Hospital linkage)	
Tufts University (MA – EHR)		
University of Colorado Anschutz Medical Campus (CO – EHR)		
University of Washington Medicine COVID Research Dataset (WA – EHR)		

Together, OHDSI has studied:
- ***>7.4m** patients tested for SARS-CoV-2*
- ***>1.6m** patients diagnosed or tested positive for COVID-19*
- ***>300k** patients hospitalized with COVID-19*

COVID-19データを持つOHDSIデータネットワーク

22個, 2020年11月

Abstract

Background and objective: As a response to the ongoing COVID-19 pandemic, several prediction models in the existing literature were rapidly developed, with the aim of providing evidence-based guidance. However, none of these COVID-19 prediction models have been found to be reliable. Models are commonly assessed to have a risk of bias, often due to insufficient reporting, use of non-representative data, and lack of large-scale external validation. In this paper, we present the Observational Health Data Sciences and Informatics (OHDSI) analytics pipeline for patient-level prediction modeling as a standardized approach for rapid yet reliable development and validation of prediction models. We demonstrate how our analytics pipeline and open-source software tools can be used to answer important prediction questions while limiting potential causes of bias (e.g., by validating phenotypes, specifying the target population, performing large-scale external validation, and publicly providing all analytical source code).

Methods: We show step-by-step how to implement the analytics pipeline for the question: 'In patients hospitalized with COVID-19, what is the risk of death 0 to 30 days after hospitalization?'. We develop models using six different machine learning methods in a USA claims database containing over 20,000 COVID-19 hospitalizations and externally validate the models using data containing over 45,000 COVID-19 hospitalizations from South Korea, Spain, and the USA.

Results: Our open-source software tools enabled us to efficiently go end-to-end from problem design to reliable Model Development and evaluation. When predicting death in patients hospitalized with COVID-19, AdaBoost, random forest, gradient boosting machine, and decision tree yielded similar or lower internal and external validation discrimination performance compared to L1-regularized logistic regression, whereas the MLP neural network consistently resulted in lower discrimination. L1-regularized logistic regression models were well calibrated.

Conclusion: Our results show that following the OHDSI analytics pipeline for patient-level prediction modelling can enable the rapid development towards reliable prediction models. The OHDSI software tools and pipeline are open source and available to researchers from all around the world.

概要

背景と目的：現在進行中のCOVID-19パンデミックへの対応として、エビデンスに基づくガイダンスを提供することを目的に、既存の文献にあるいくつかの予測モデルが急速に開発された。しかし、これらのCOVID-19予測モデルはいずれも信頼できるものではないことが判明した。一般的にモデルにはバイアスのリスクがあると評価されており、その原因としては、報告が不十分であること、代表性のないデータを使用していること、大規模な外部検証が行われていないことなどが挙げられる。この論文では、患者レベルの予測モデルのためのOHDSI解析パイプラインを、予測モデルの開発と検証を迅速かつ確実に行うための標準的なアプローチとして紹介する。表現型の検証、対象集団の特定、大規模な外部検証の実施、すべての解析ソースコードの公開などにより、バイアスの可能性を抑えつつ、重要な予測問題に答えるために、我々の解析パイプラインとオープンソースのソフトウェアツールをどのように使用できるかを示す。

方法: 「COVID-19で入院した患者において、入院から30日までの死亡リスクは何か」という質問に対する分析パイプラインの実装方法をステップバイステップで示します。20,000件以上のCOVID-19による入院を含む米国の請求データベースにおいて、6種類の機械学習手法を用いてモデルを開発し、韓国、スペイン、米国の45,000件以上のCOVID-19による入院を含むデータを用いてモデルを外部から検証する。

結果: オープンソースのソフトウェアツールを使用することで、問題の設計から信頼性の高いモデルの開発と評価まで、エンドツーエンドで効率的に行うことがでた。COVID-19で入院した患者の死亡を予測する場合、AdaBoost、ランダムフォレスト、勾配ブースティングマシン、決定木は、L1正則化ロジスティック回帰と比較して、内部検証および外部検証の判別性能は同等かそれ以下であったが、MLPニューラルネットワークは一貫して判別性能が低かった。L1正則化ロジスティック回帰モデルはよくキャリブレーションされていた。

結論: 今回の結果は、患者レベルの予測モデリングのためのOHDSI分析パイプラインに従うことで、信頼性の高い予測モデルに向けた迅速な開発が可能になることを示している。OHDSIのソフトウェアツールとパイプラインはオープンソースで、世界中の研究者が利用できる。

OHDSI論文
新規2

Stud Health Technol Inform. 2021 Sep 21;283:95-103. doi: 10.3233/SHTI210546.

The Usage of OHDSI OMOP – A Scoping Review

Ines Reinecke [1], Michéle Zoch [1], Christian Reich [2 3], Martin Sedlmayr [1], Franziska Bathelt [1]

Affiliations – collapse

Affiliations

1 Institute for Medical Informatics and Biometry at Carl Gustav Carus Faculty of Medicine at Technische Universität Dresden, Germany.
2 IQVIA, Cambridge, MA, USA.
3 Observational Health Data Sciences and Informatics (OHDSI), New York, NY, USA.

Abstract

OHDSI, a fast growing open-science research community seeks to enable researchers from around the globe to conduct network studies based on standardized data and vocabularies. There is no comprehensive review of publications about OHDSI's standard: the OMOP Common Data Model and its usage available. In this work we aim to close this gap and provide a summary of existing publications including the analysis of its meta information such as the choice of journals, journal types, countries, as well as an analysis by topics based on a title and abstract screening. Since 2016, the number of publications has been constantly growing and the relevance of the OMOP CDM is increasing in terms of multi-country studies based on observational patient data.

概要

OHDSIは、世界中の研究者が標準化されたデータと語彙に基づいてネットワーク研究を行うことを目的とした、急速に成長しているオープンサイエンス研究コミュニティです。しかし，OHDSIの標準規格であるOMOP Common Data Modelとその使用法に関する包括的な出版物のレビューは存在しない。本研究では、このギャップを埋めることを目的とし、ジャーナルの選択、ジャーナルの種類、国などのメタ情報の分析や、タイトルとアブストラクトのスクリーニングに基づくトピック別の分析を含む、既存の出版物の概要を提供します。2016年以降、論文数は常に増加しており、観察された患者データに基づく多国間の研究という点で、OMOP CDMの関連性は高まっています。

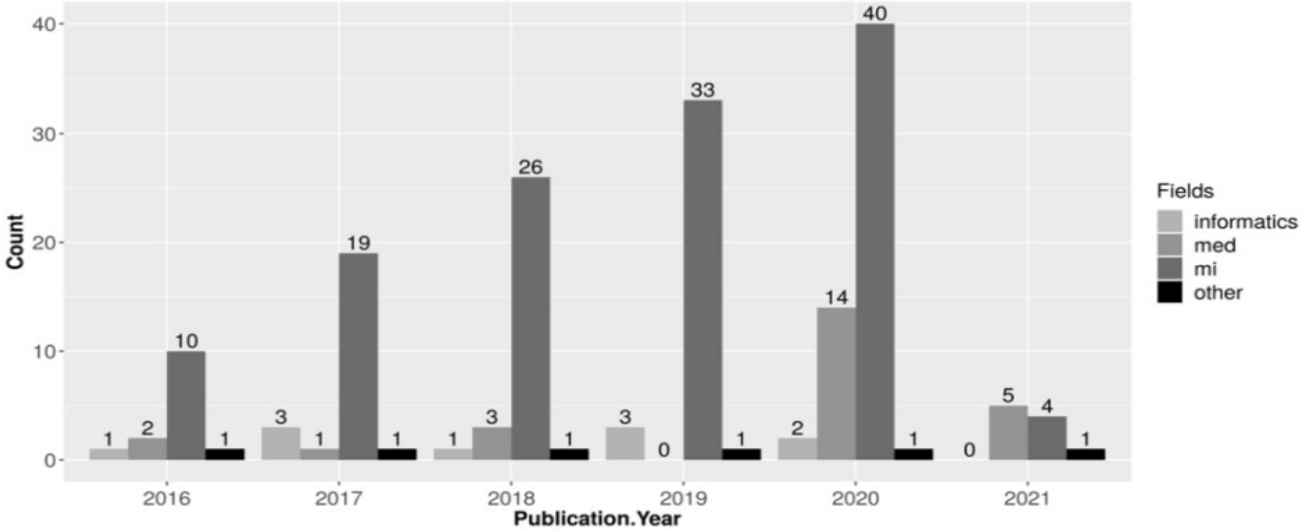


Figure 3: Annual distribution of different journal fields

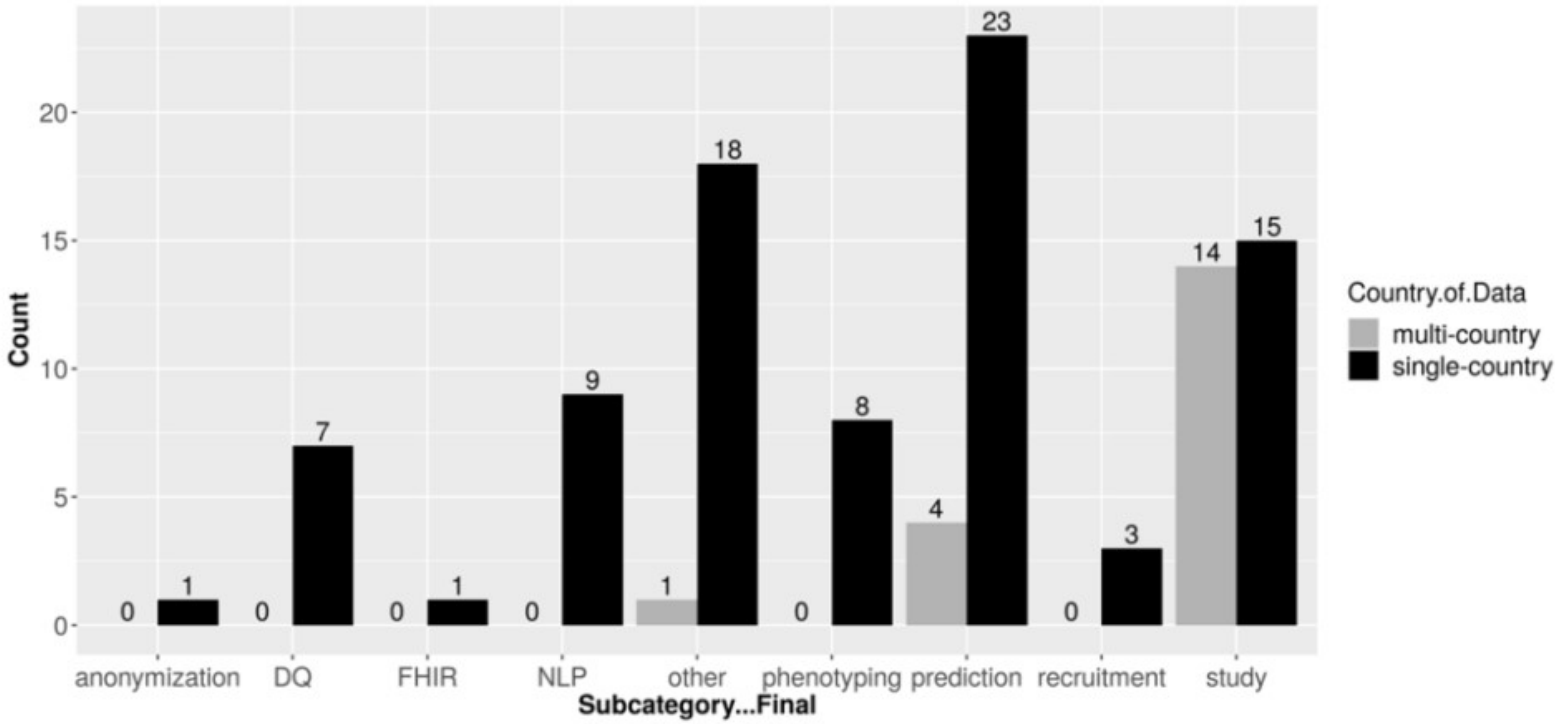


Figure 4: Distribution of Articles by Specific Usage (subcategories) and country of data

OHDSI論文
新規3

Stud Health Technol Inform. 2021 Sep 21;283:86-94. doi: 10.3233/SHTI210545.

Towards the Representation of Genomic Data in HL7 FHIR and OMOP CDM

Yuan Peng [1], Azadeh Nassirian [1], Najia Ahmadi [1], Martin Sedlmayr [1], Franziska Bathelt [1]

Affiliations — collapse

Affiliation

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

Abstract
High throughput sequencing technologies have facilitated an outburst in biological knowledge over the past decades and thus enables improvements in personalized medicine. In order to support (international) medical research with the combination of genomic and clinical patient data, a standardization and harmonization of these data sources is highly desirable. To support this increasing importance of genomic data, we have created semantic mapping from raw genomic data to both FHIR (Fast Healthcare Interoperability Resources) and OMOP (Observational Medical Outcomes Partnership) CDM (Common Data Model) and analyzed the data coverage of both models. For this, we calculated the mapping score for different data categories and the relative data coverage in both FHIR and OMOP CDM. Our results show, that the patients genomic data can be mapped to OMOP CDM directly from VCF (Variant Call Format) file with a coverage of slightly over 50%. However, using FHIR as intermediate representation does not lead to further information loss as the already stored data in FHIR can be further transformed into OMOP CDM format with almost 100% success. Our findings are in favor of extending OMOP CDM with patient genomic data using ETL to enable the researchers to apply different analysis methods including machine learning algorithms on genomic data.

概要
ハイスループットシーケンシング技術は、過去数十年にわたって生物学的知識の増大を促進し、個別化医療の向上を可能にしてきた。ゲノムデータと患者の臨床データを組み合わせて（国際的な）医学研究を支援するためには、これらのデータソースの標準化と調和が強く望まれる。そこで、ゲノム生データをFHIRおよびOMOPのCDMにセマンティックマッピングし、両モデルのデータカバレッジを解析した。そのために、異なるデータカテゴリーに対するマッピングスコアと、FHIRとOMOP CDMの両方における相対的なデータカバー率を算出しました。その結果、患者のゲノムデータをVCF（Variant Call Format）ファイルから直接OMOP CDMにマッピングした場合、カバー率は50%をわずかに超える程度でした。しかし、FHIRを中間表現として使用しても、FHIRにすでに保存されているデータをさらにOMOP CDMフォーマットに変換することは、ほぼ100%成功するため、さらなる情報の損失にはなりません。ETLを用いてOMOP CDMを患者のゲノムデータで拡張することで、研究者がゲノムデータに機械学習アルゴリズムを含む様々な解析方法を適用できるようになることが今回の発見でわかりました。

- VCF data→OMOP CDM, using OMOP G-CDM
- VCF data→OMOP CDM, using standard MEASUREMENT table
- VCF data→FHIR→OMOP CDM, using OMOP G-CDM
- VCF data→FHIR→OMOP CDM, using standard MEASUREMENT table

OHDSI論文
新規4

JAMIA Open. 2021 Feb 4;4(3):ooab001. doi: 10.1093/jamiaopen/ooab001. eCollection 2021 Jul.

Transforming and evaluating electronic health record disease phenotyping algorithms using the OMOP common data model: a case study in heart failure

Vaclav Papez [1 2], Maxim Moinat [3], Stefan Payralbe [3], Folkert W Asselbergs [1 4], R Thomas Lumbers [1 2], Harry Hemingway [1 2], Richard Dobson [1 2 5], Spiros Denaxas [1 2 6]

Affiliations – collapse

Affiliations

1 Institute of Health Informatics, University College London, London, UK.
2 Health Data Research UK, London, UK.
3 The Hyve, Utrecht, Netherlands.
4 Department of Cardiology, Division Heart & Lungs, University Medical Center Utrecht, Utrecht University, Utrecht, the Netherlands.
5 Department of Biostatistics and Health Informatics, Institute of Psychiatry, Psychology and Neuroscience (IoPPN), King's College London, London, UK.
6 The Alan Turing Institute, London, UK.

Abstract

Objective: The aim of the study was to transform a resource of linked electronic health records (EHR) to the OMOP common data model (CDM) and evaluate the process in terms of syntactic and semantic consistency and quality when implementing disease and risk factor phenotyping algorithms.

Materials and methods: Using heart failure (HF) as an exemplar, we represented three national EHR sources (Clinical Practice Research Datalink, Hospital Episode Statistics Admitted Patient Care, Office for National Statistics) into the OMOP CDM 5.2. We compared the original and CDM HF patient population by calculating and presenting descriptive statistics of demographics, related comorbidities, and relevant clinical biomarkers.

Results: We identified a cohort of 502 536 patients with the incident and prevalent HF and converted 1 099 195 384 rows of data from 216 581 914 encounters across three EHR sources to the OMOP CDM. The largest percentage (65%) of unmapped events was related to medication prescriptions in primary care. The average coverage of source vocabularies was >98% with the exception of laboratory tests recorded in primary care. The raw and transformed data were similar in terms of demographics and comorbidities with the largest difference observed being 3.78% in the prevalence of chronic obstructive pulmonary disease (COPD).

Conclusion: Our study demonstrated that the OMOP CDM can successfully be applied to convert EHR linked across multiple healthcare settings and represent phenotyping algorithms spanning multiple sources. Similar to previous research, challenges mapping primary care prescriptions and laboratory measurements still persist and require further work. The use of OMOP CDM in national UK EHR is a valuable research tool that can enable large-scale reproducible observational research.

概要

目的: 本研究の目的は，リンクされた電子健康記録（EHR）のリソースをOMOP共通データモデル（CDM）に変換し，疾患および危険因子の表現型アルゴリズムを実装する際の構文的・意味的な一貫性と品質の観点からプロセスを評価することである。

材料と方法: 心不全（HF）を例に、3つの全国EHRソース（Clinical Practice Research Datalink, Hospital Episode Statistics Admitted Patient Care, Office for National Statistics）をOMOP CDM 5.2に表現した。人口統計、関連する併存疾患、および関連する臨床バイオマーカーの記述統計を計算して発表することにより、オリジナルとCDMのHF患者集団を比較した。

結果: 3つのEHRソースで216,581,914回の受診から得られた1,099,195,384行のデータをOMOP CDMに変換した。マップされていないイベントの最大の割合（65%）は、プライマリ・ケアにおける薬の処方に関するものであった。プライマリケアで記録された臨床検査を除き、ソースボキャブラリの平均カバー率は98%以上であった。生データと変換後のデータは、人口統計と併存疾患の点で類似しており、最大の違いは慢性閉塞性肺疾患（COPD）の有病率が3.78%であった。

結論: 今回の研究では、OMOP CDMが、複数の医療機関に接続されたEHRを変換し、複数のソースにまたがる表現型アルゴリズムを表現するために、うまく適用できることを示した。これまでの研究と同様に、プライマリケアの処方箋と検査測定値をマッピングするという課題はまだ残っており、さらなる研究が必要である。英国のEHRでOMOP CDMを使用することは、大規模で再現性のある観察研究を可能にする貴重な研究ツールとなる。

OHDSI論文
新規5

AMIA Annu Symp Proc. 2021 May 17;2021:394-403. eCollection 2021.

A Comparison between Human and NLP-based Annotation of Clinical Trial Eligibility Criteria Text Using The OMOP Common Data Model

Xinhang Li [1 2], Hao Liu [1 2], Fabrício Kury [1], Chi Yuan [1], Alex Butler [1], Yingcheng Sun [1], Anna Ostropolets [1], Hua Xu [3], Chunhua Weng [1]

Affiliations — collapse

Affiliations

1 Department of Biomedical Informatics, Columbia University, New York, NY, USA.
2 equal-contribution first authors.
3 School of Biomedical Informatics, The University of Texas Health Science Center at Houston, TX, USA.

Abstract

Human annotations are the established gold standard for evaluating natural language processing (NLP) methods. The goals of this study are to quantify and qualify the disagreement between human and NLP. We developed an NLP system for annotating clinical trial eligibility criteria text and constructed a manually annotated corpus, both following the OMOP Common Data Model (CDM). We analyzed the discrepancies between the human and NLP annotations and their causes (e.g., ambiguities in concept categorization and tacit decisions on inclusion of qualifiers and temporal attributes during concept annotation). This study initially reported complexities in clinical trial eligibility criteria text that complicate NLP and the limitations of the OMOP CDM. The disagreement between and human and NLP annotations may be generalizable. We discuss implications for NLP evaluation.

概要

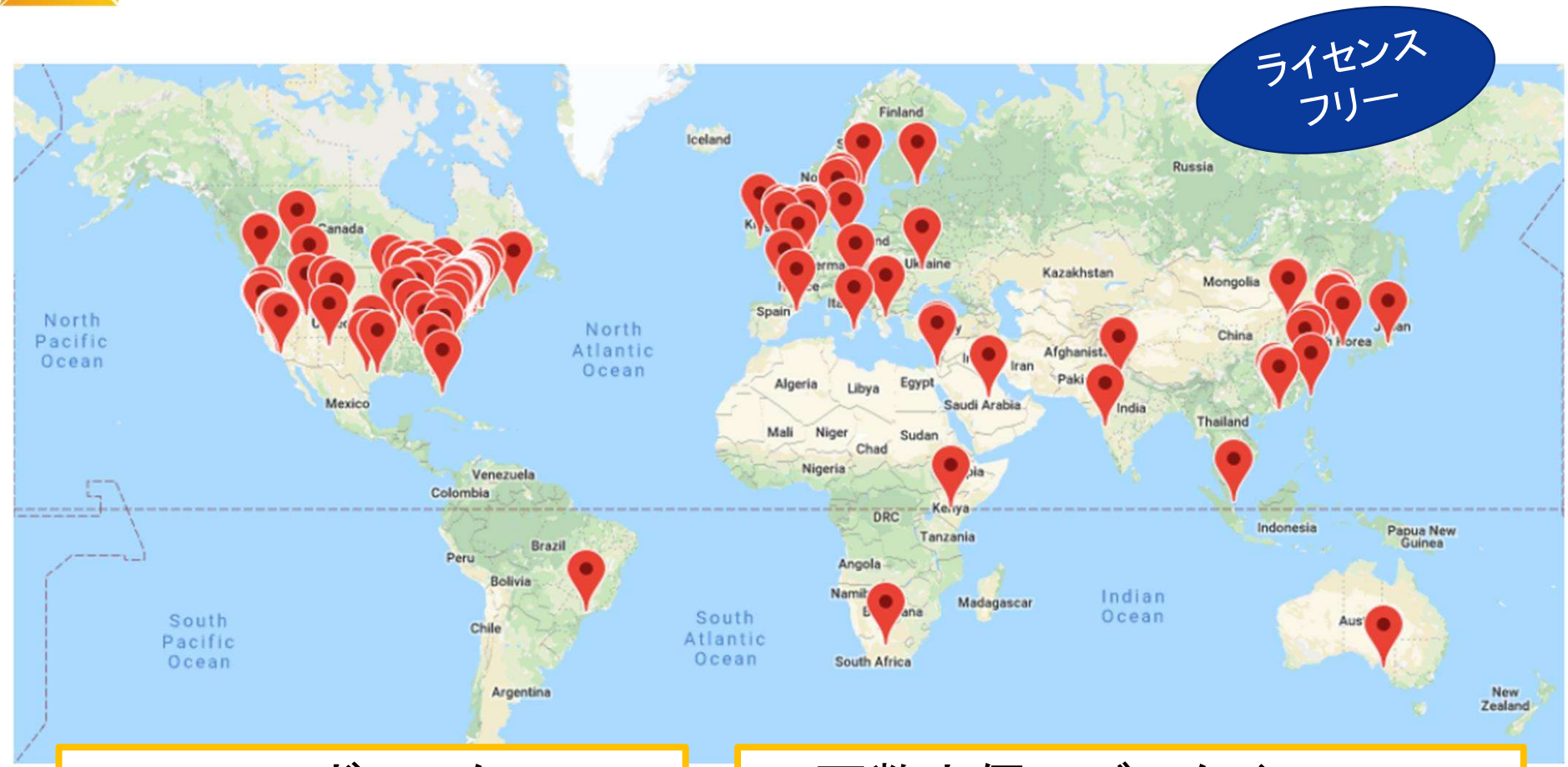
人間によるアノテーションは，自然言語処理（NLP）手法を評価するためのゴールドスタンダードとして確立されている。本研究の目的は，人間とNLPの間の不一致を定量化し，評価することである。本研究では，臨床試験の適格性基準テキストをアノテーションするためのNLPシステムを開発し，OMOP Common Data Model (CDM)に基づいて手動でアノテーションしたコーパスを構築しました。人間とNLPのアノテーションの不一致とその原因（概念の分類における曖昧さや、概念のアノテーション時に修飾語や時間属性を含めることに関する暗黙の決定など）を分析しました。本研究ではまず、NLPを複雑にする臨床試験の適格性基準テキストの複雑さと、OMOP CDMの限界を報告しました。人間のアノテーションとNLPのアノテーションの間の不一致は、一般化できる可能性があります。また、NLP評価への影響についても言及しています。

OMOP概説

名称の説明

- OMOP　オモップ
  - 診療情報を格納する<u>**データモデルの名称**</u>
  - フルネームはOMOP Common Data Model（**OMOP-CDM**）
  - かつてあったOMOPという名称の活動（FDAと米製薬業界の合同活動）にて策定されたものをOHDSIが引き継いで発展させた。由来から今でもOMOP-CDMと呼ぶ。OMOPの英語発音はオモップからオゥモップ。
- OHDSI　オデッセイ
  - OMOP-CDMを策定・活用する<u>**コミュニティの名称**</u>
  - 古代ギリシャの叙事詩**Odyssey**（古代トロイア戦争英雄の家路の苦難辛苦物語。英語表現）にちなむもので、同じ発音ができるようにと作られた短縮形。
  - Odysseyは現代では長い冒険の旅といった意味に転用されている。コミュニティが成立するまでの紆余曲折を表すためOdysseyにちなんだとのこと。
  - “Odyssey”と同じ発音をする趣旨から、日本語読みは「オデッセイ」。英語発音は人/場面により異なりますが、オデシーからオゥデシー、オゥデッシー。

世界に広がるOMOP/OHDSI

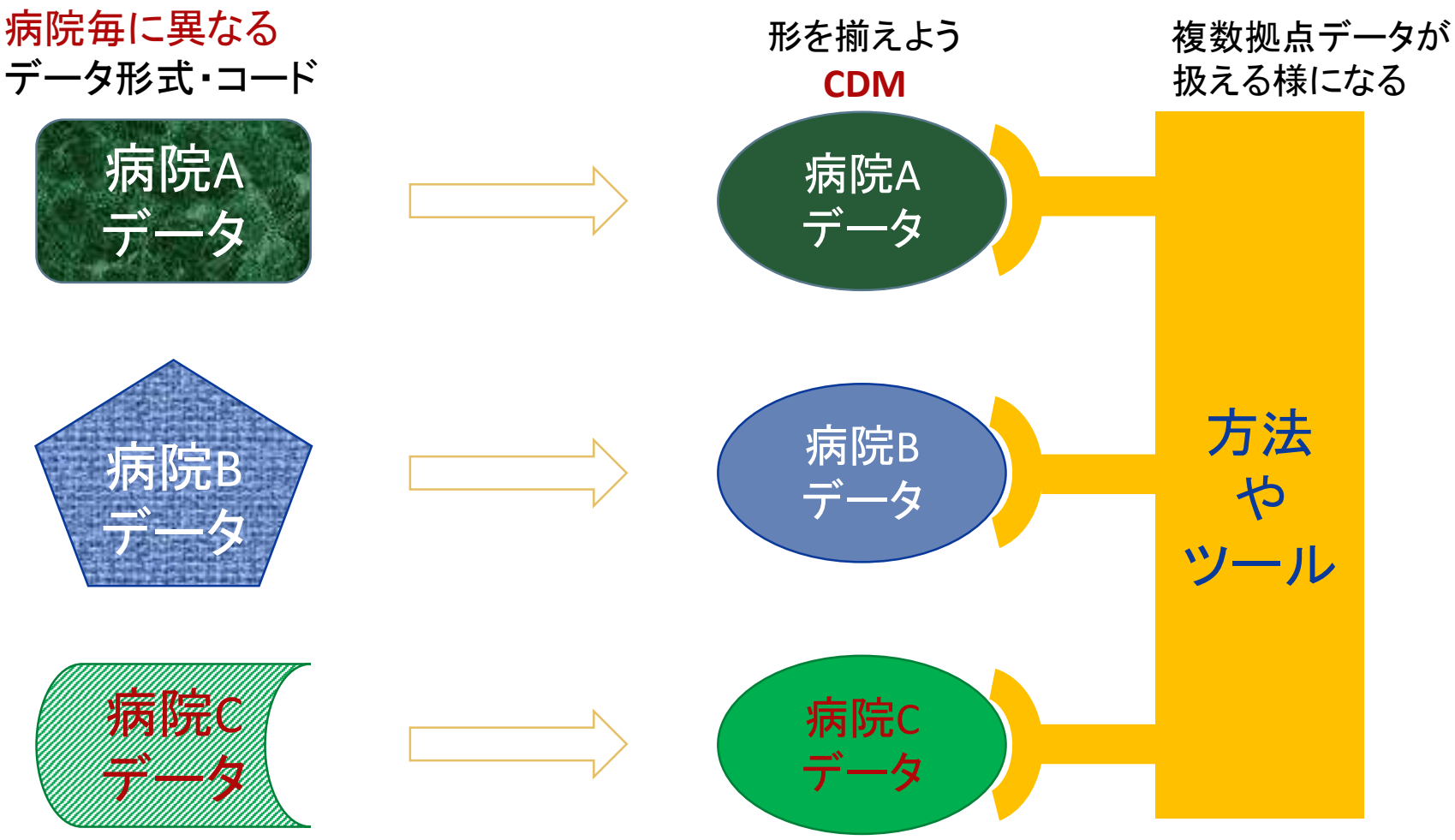


OHDSI コラボレーター
- 学術、企業、政府
- 約30ヵ国、6大陸

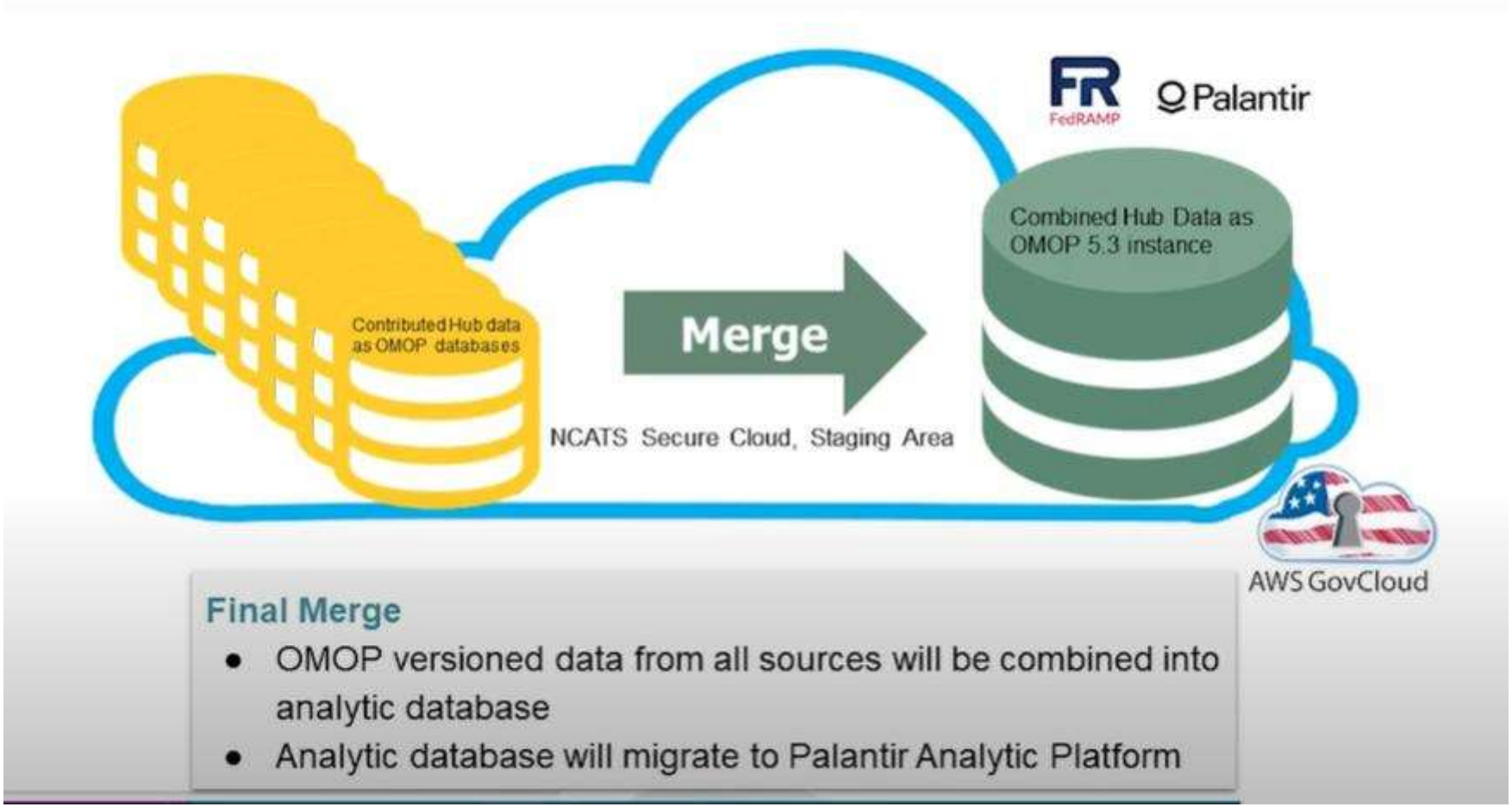
- 百数十個のデータベース
- 重複を除いて推定6億人以上
- 2021/9/16現在、69個の国際共同研究が登録

Common Data Modelとは

- 標準化された(診療情報の)データ形式



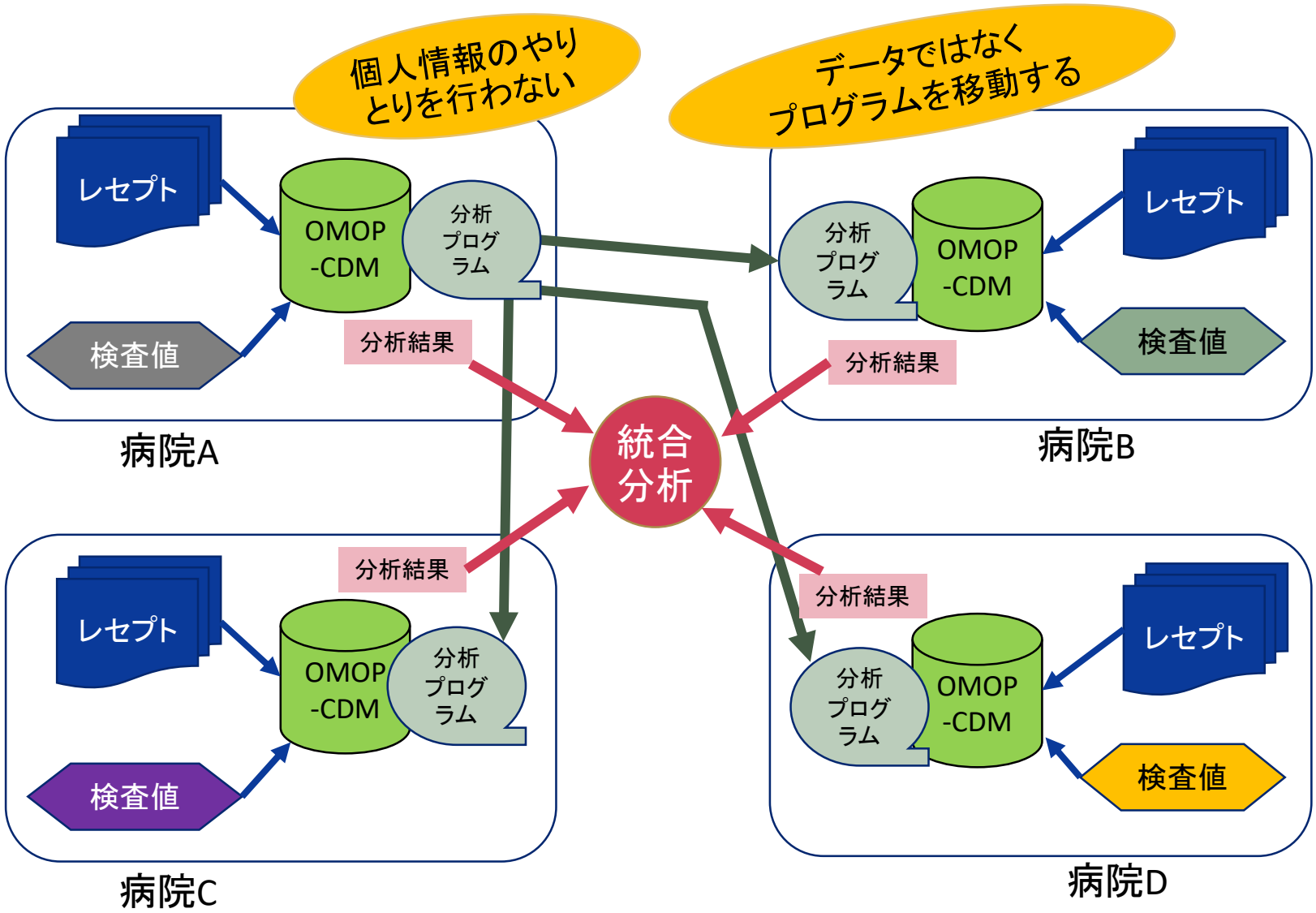
複数拠点連携
同じ形式でデータ統合（N3Cの例）



- 国内でもそれぞれの形式で一般的に行われている方法

N3C: National COVID Cohort Collaborative
https://ncats.nih.gov/n3c/about

複数拠点連携
データ統合はしない方法（OHDSI-ATLAS方式）



OMOP以外のCDMとの違い

- 診療情報CDMはOMOPだけではないが、、
  - 米国だけでも、Sentinel、PCORNet、i2b2等がある
  - 日本では、SS-MIX2、SDM等

国際的なCDMはOMOPのみ

- なぜ国際化が難しいのか

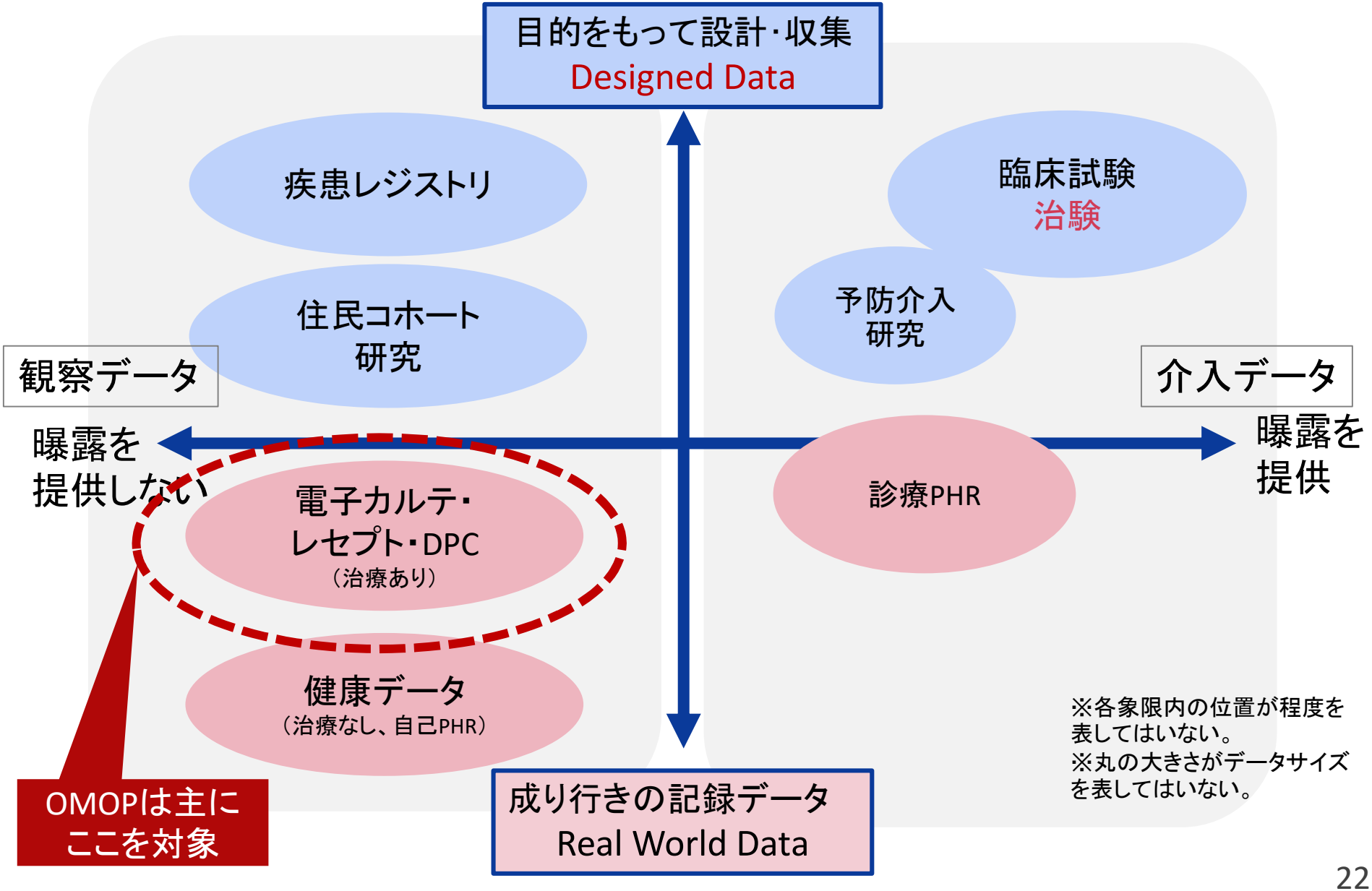
→形式をあわせるだけでは済まないから。

ボキャブラリ（ターミノロジー）が国ごとに異なる

OMOP化するデータ元の種類

- 保険請求データ、管理データ
  - 日本ではレセプトやDPC
  - 現状ではこの種類が最も多い
- 電子カルテ、病院情報システム
  - 検査結果値等の詳細データ
  - 電カル以外のデータ
- 疾患レジストリ（集積するために個々登録する）
  - がんゲノムのレジストリ、バイオバンクなど
- 疫学調査データ
  - All-of-US（米国の100万人全国コホート）
  - National COVID Cohort Collaborative（N3C）
- その他
  - 救急搬送記録、予防接種記録、等

様々な医療・健康データ


目的をもって設計・収集
Designed Data
疾患レジストリ
住民コホート
研究
臨床試験
治験
予防介入
研究
観察データ
介入データ
曝露を
提供しない
曝露を
提供
電子カルテ・
レセプト・DPC
（治療あり）
診療PHR
健康データ
（治療なし、自己PHR）
OMOPは主に
ここを対象
成り行きの記録データ
Real World Data
※各象限内の位置が程度を
表してはいない。
※丸の大きさがデータサイズ
を表してはいない。

OMOP-CDM 3つの構成要素

(1) データ形式

- 約40個のテーブル群
- 分析者に受け入れられやすい2次元の表

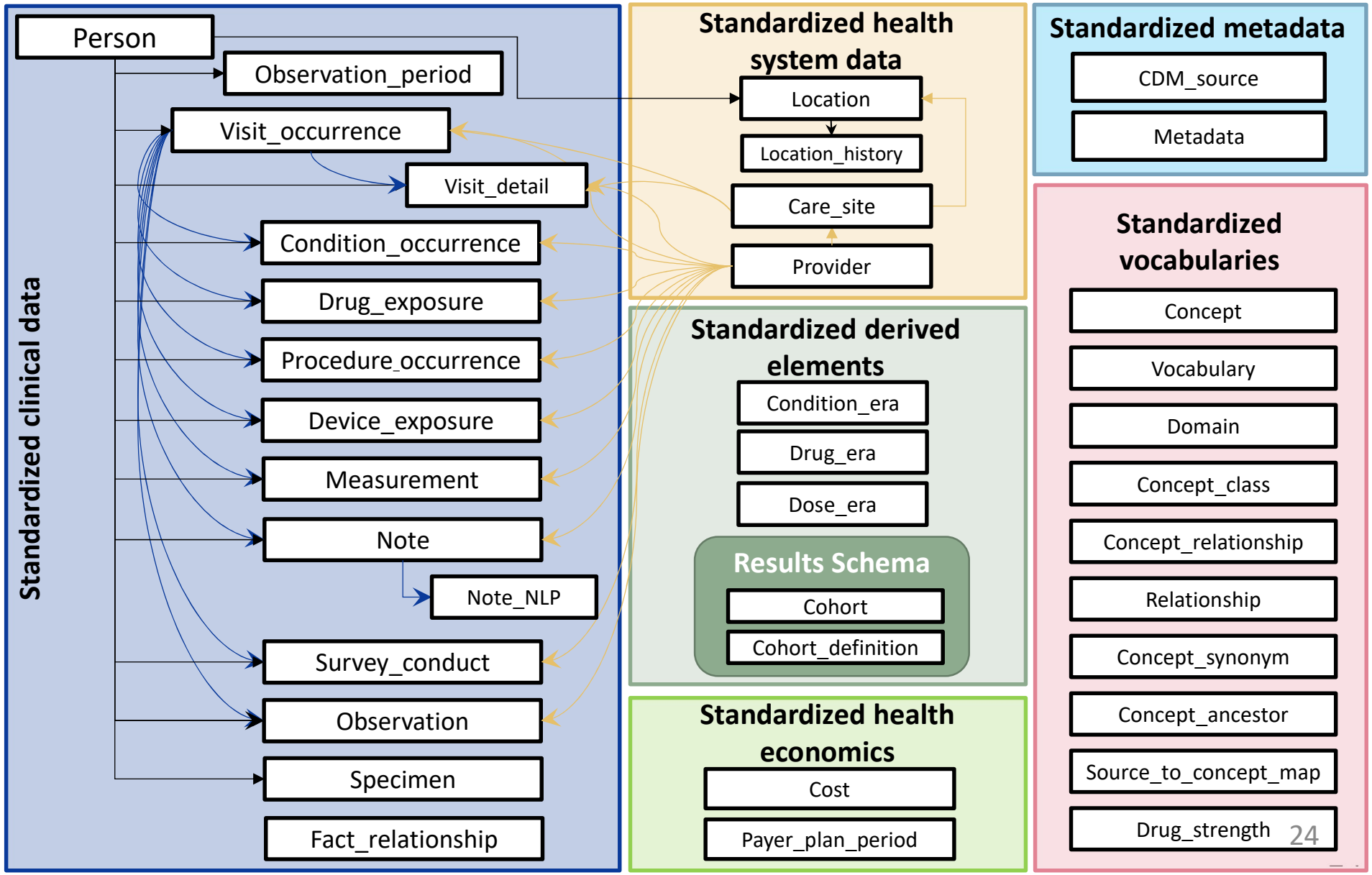
(2) 用語 ボキャブラリ

- 800万語以上の整理された用語
- 国際対応ができるCDMである所以

(3) コミュニティ OHDSI

- 分析系、IT系どちらも
- ODHSIツール javaやRで書かれたツール
- “OHDSI Study” 国際共同研究

OMOP-CDMのテーブル群（ver.6）



具体例（The Book of OHDSIより）

- ローレンさん：子宮内膜症

PERSON Table

1982年3月12日生まれ、36歳、白人

列名	値	(説明)
PERSON_ID	1	PERSONテーブル中のユニーク値
GENDER_CONCEPT_ID	8532	女性を表す「コンセプトID(CID)」
YEAR_OF_BIRTH	1982	
MONTH_OF_BIRTH	3	
DAY_OF_BIRTH	12	
BIRTH_DATETIME	1982-03-12 00:00:00	
DEATH_DATETIME		
RACE_CONCEPT_ID	8527	白人を表す「コンセプトID(CID)」

以下略

コンセプトID
OMOPボキャブラリにおける用語番号

具体例：観察期間テーブル

OBSERVATION_PERIOD Table

外来受診：2010/1/6、2011/1/6、2012/1/6、2013/1/7
救急外来：2013/1/14
入院：2013/1/17～2013/1/24

列名	値	(説明)
OBSERVATION_ PERIOD_ID	1	自動採番されるユニーク値
PERSON_ID	1	PERSON Tableにあるもの
OBSERVATION_PERIOD_ START_DATE	2010-01-06	記録にある最初の受診日
OBSERVATION_PERIOD_ END_DATE	2013-01-24	記録にある最後の受診日
PERIOD_TYPE_ CONCEPT_ID	44814725	記録の種類を表すCID

具体例：受診テーブル

VISIT_OCCURRENCE Table

各受診毎の情報
ここでは入院(2013/1/17～2013/1/24)の例

列名	値	(説明)
VISIT_OCCURRENCE_ ID	514	自動採番されるユニーク値
PERSON_ID	1	PERSON Tableにあるもの
VISIT_CONCEPT_ID	9201	入院を表すコンセプトID(CID)
VISIT_START_DATE	2013-01-17	最初の日(入院日)
VISIT_END_DATE	2013-01-24	最後の日(退院日)
VISIT_TYPE_ CONCEPT_ID	32035	記録由来タイプ(電子カルテ)
VISIT_SOURCE_ VALUE	inpatient	元データの値(標準化されていない)

途中&以下略

具体例：傷病テーブル

CONDITION_OCCURRENCE Table

傷病・疾病・兆候・症状の情報
「月経困難症」

列名	値	(説明)
CONDITION_OCCURRENCE_ ID	964	自動採番されるユニーク値
PERSON_ID	1	PERSON Tableにあるもの
CONDITION_CONCEPT_ID	194696	月経困難症を表すコンセプトID(CID)
CONDITION_START_ DATE	2010-01-06	記録された最初の日
CONDITION_END_ DATE	NULL	最後の日. 情報がない場合が多い
CONDITION_TYPE_ CONCEPT_ID	32020	記録由来タイプ(カルテ中の診断)
VISIT_OCCURRENCE_ ID	509	診断がなされた受診ID

途中&以下略

具体例：医薬品テーブル

DRUG_EXPOSURE Table

医薬品使用情報
アセトアミノフェン325mgを1日2錠、30日分で合計60錠

列名	値	(説明)
DRUG_EXPOSURE_ID	1001	自動採番されるユニーク値
PERSON_ID	1	PERSON Tableにあるもの
DRUG_CONCEPT_ID	1127433	アセトアミノフェン325mgのCID
DRUG_EXPOSURE_ START_DATE	2010-01-06	使用開始日
DRUG_EXPOSURE_ END_DATE	2010-02-05	使用終了日
DRUG_TYPE_ CONCEPT_ID	38000177	記録由来種類(処方箋)
QUANTITY	60	数量
DAYS_SUPPLY	30	日数
ROUTE_CONCEPT_ID	4132161	「経口」のCID

途中&以下略

具体例：医療行為テーブル

PROCEDURE_OCCURRENCE Table

処置・医療行為情報
「骨盤超音波検査」

列名	値	(説明)
PROCEDURE_ OCCURRENCE_ID	1227	自動採番されるユニーク値
PERSON_ID	1	PERSON Tableにあるもの
PROCEDURE_ CONCEPT_ID	4127451	骨盤超音波検査
PROCEDURE_DATE	2013-01-14	実施日
PROCEDURE_TYPE_ CONCEPT_ID	38000275	記録由来種類(電カルオーダー)
MODIFIER_CONCEPT_ ID	0	修飾子
QUANTITY	0	数量(0と1は同じ)
PROVIDER_ID	NULL	実施者(医療者のID)
PROCEDURE_SOURCE_ VALUE	304435002	元データにおける表現/コード

途中&以下略

構成要素2：ボキャブラリ

一般には
・用語
・ターミノロジー
とそのコード

様々な医薬品コード

- **カロナール錠200**（アセトアミノフェン200mg）
  レセプトコード：　620002023
  YJコード：　1141007F1063
  HOT9コード：　100713802
  院内コード：医療機関毎に異なる

- **アセトアミノフェン錠200mg「JG」**
  レセプトコード：　621520803
  YJコード：　1141007F1152
  HOT9コード：　115208103
  院内コード：　医療機関毎に異なる

- **Acetaminophen 200 MG Oral Tablet**
  RxNorm: 430834

ボキャブラリ:ひとつの表にまとめる

傷病

ICD-10

ICD-9

CPT4

SNOEMD

RxNorm

医薬品

LOINC

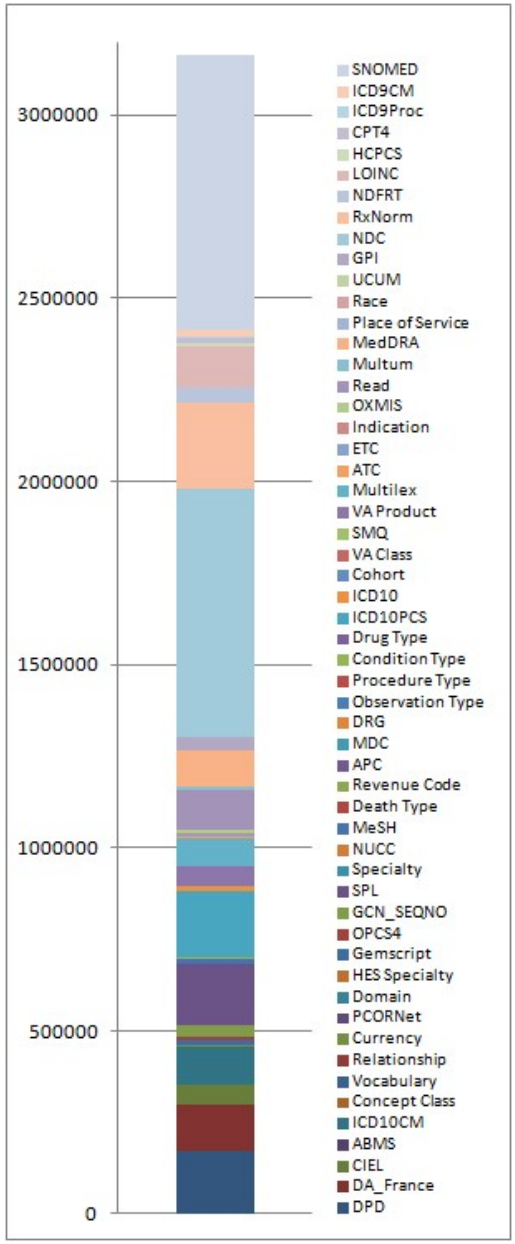
検査

Read

MedDRA

etc.

misc.



各々の用語は「Concept」として扱われる

CONCEPT Table

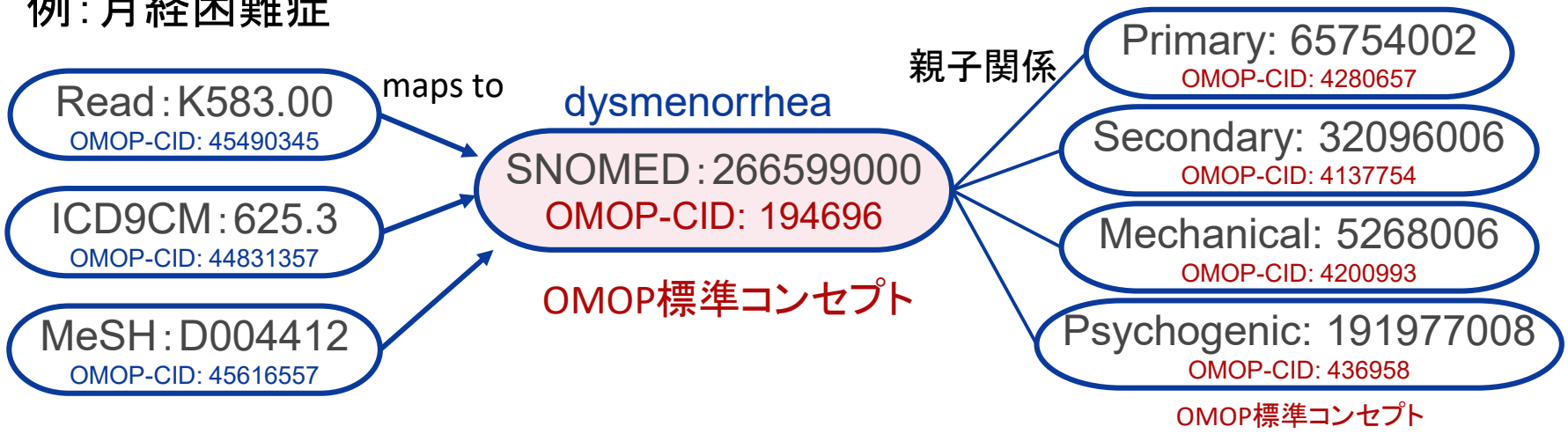
CONCEPT_ID	313217
CONCEPT_NAME	Atrial fibrillation
DOMAIN_ID	Condition
VOCABULARY_ID	SNOMED
CONCEPT_CLASS_ID	Clinical Finding
STANDARD_CONCEPT	S
CONCEPT_CODE	49436004
VALID_START_DATE	01-Jan-1970
VALID_END_DATE	31-Dec-2099
INVALID_REASON	

- For use in CDM
- English description
- Domain
- Vocabulary
- Class in SNOMED
- Concept in data
- Code in SNOMED
- Valid during time interval

OMOPボキャブラリの特長

- 世界の医療用語集を**ひとつに**まとめている。
- 疾患、医薬品、検査等すべての分野を包含。
- 現在800万語以上、それぞれにIDを割当。
- さらに、**用語の関係性**が維持保守されている。

例：月経困難症



- 各国の用語コードをOMOP標準にマッピングできる
- 親子関係により自在な粒度で扱える

構成要素3:コミュニティ

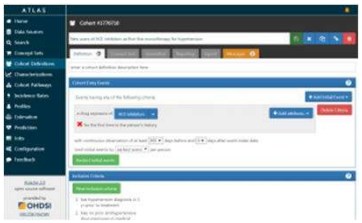
OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

Join the Journey

256 collaborators in 27 different countries over six continents

《Active Working Groups》
- Algorithmic Phenotyping
  (under Phenotype Dev. & Evaluation WG)
- Atlas & WebAPI WG
- CDM and Vocabulary Development WG
- Causal Inference
- China WG
- Clinical Trials WG
- Data Quality
- EHR-Electronic Health Record WG
- GIS-Geographic Information System WG
- HADES-Health Analytics Data-to-Evidence Suite
- Latin America
- Medical Devices WG
- Natural Language Processing WG
- OHDSI Steering Working Group
- Oncology WG
- Patient-Generated Health Data WG
- Patient-Level Prediction WG
- Pharmacovigilance evidence investigation
  (used to be Knowledge Base / Laertes)
- Population-Level Estimation WG
- Psychiatry WG
- UK BioBank
- Women of OHDSI
- Early Stage Researchers

OHDSIツール



ATLAS 分析のためのWebUIツール

デモサイト http://atlas-demo.ohdsi.org/

HADES 分析のためのRライブラリ

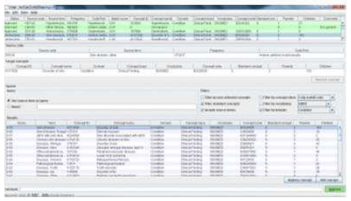
DATA QUALITY DASHBOARD データ品質チェックツール

ACHILLES データのプロファイリングツール

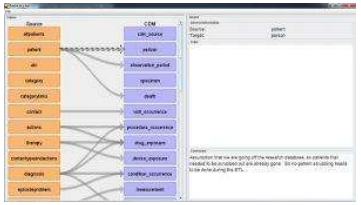
ATHENA ボキャブラリの総本山

https://athena.ohdsi.org/

USAGI 標準コードへのマッピング支援ツール



WHITERABBIT, RABBIT-IN-A-HAT ETL支援ツール



地域コミュニティ

EHDEN
EUROPEAN HEALTH DATA & EVIDENCE NETWORK

1億人以上のデータを目指す

OHDSI JAPAN
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

OHDSI China
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

OHDSI KOREA
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

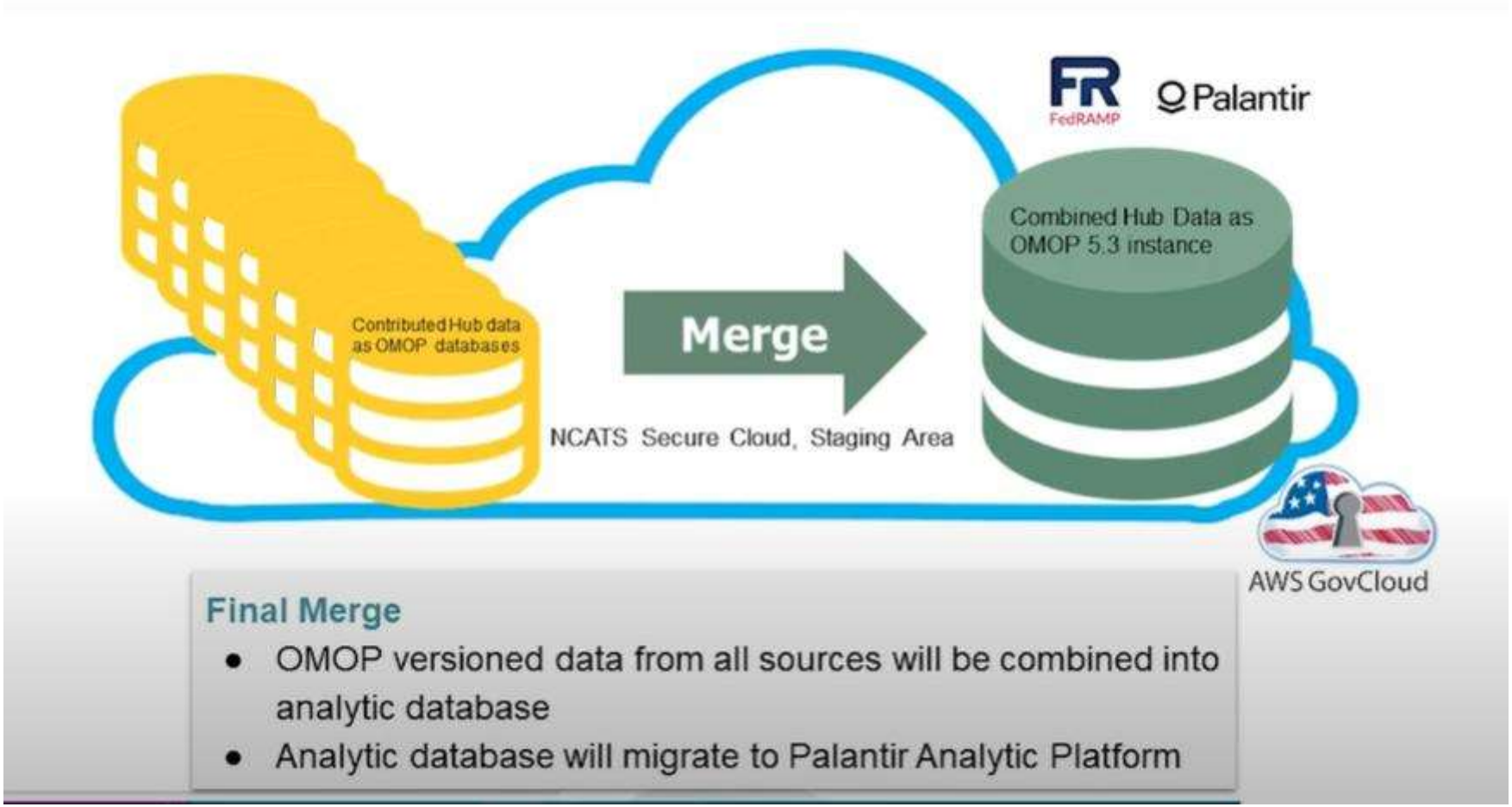
FEEDER NET

そのほかTaiwan, Singapore, Australia

NIH **All of Us** Research Program

100万人の全米住民コホート　臨床データはOMOP-CDMで

N3C: National COVID Cohort Collaborative



- 国内でもそれぞれの形式で一般的に行われている方法

N3C: National COVID Cohort Collaborative
https://ncats.nih.gov/n3c/about

欧州の活動：EHDEN

EHDEN
EUROPEAN HEALTH DATA & EVIDENCE NETWORK

HOME | RESOURCES | EHDEN | OPEN CALLS | NEWS | BLOG | CONTACT

On a mission to become the trusted open science community built for health data research via a European federated network

Federated network

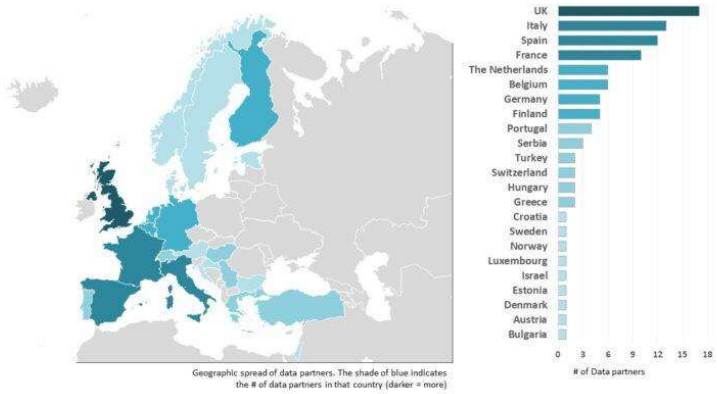
Find out more about the data partners in the EHDEN federated network.

SME Catalogue

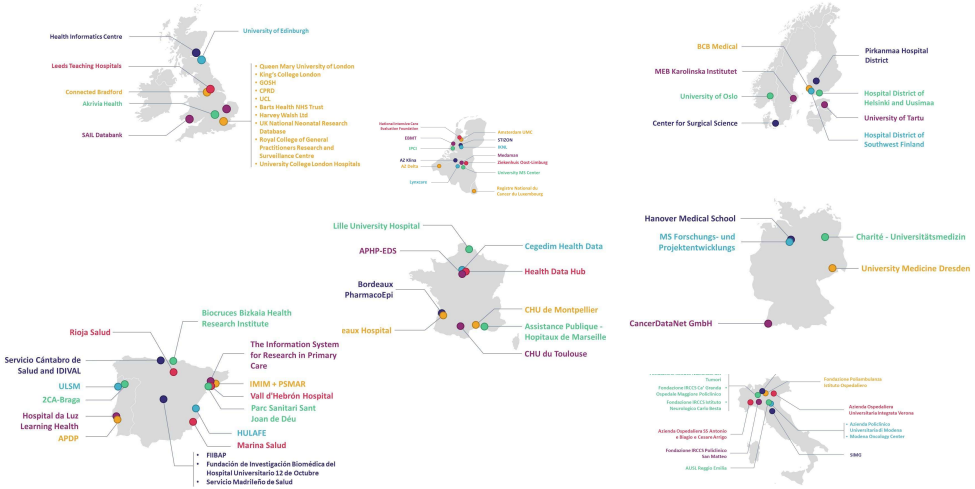
Browse our online catalogue of EHDEN-trained and certified SMEs

EHDEN Academy

Publicly available, free and on-demand training developed by the OHDSI and EHDEN community.



Geographic spread of data partners. The shade of blue indicates the # of data partners in that country (darker = more)



SME Catalogue

Browse our online catalogue of EHDEN-trained and certified SMEs

- OHDSI Software and Tools (24)
- OMOP CDM ETL (28)
- Technical infrastructure services (24)
- OHDSI Training (15)
- OMOP Standardized Vocabularies (23)

Quretec

Country: Estonia

Work Languages: English, Estonian (basic un

Services: OHDSI Software and Tools, OMOP Standardized Vocabularies

Kaiser & Preusse

Country: Germany

Work Languages: English, German

Services: OHDSI Software and Tools, OMOP CDM ETL, Technical infrastructure services, OHDSI Training

SmartMed B.V.

Country: The Netherlands

Work Languages: English, Dutch

EHDEN Academy

JOIN THE JOURNEY

Log in

Tools, Skills & Methods Training for Working with Real World Observational Data in Generating Evidence

EHDEN Academy
Getting Started
A brief introduction to the EHDEN Academy.

EHDEN Academy
EHDEN Foundation
Provides an overview of the EHDEN project including high-level...

Non-professional
Patient Organisations: Introduction to Rea...
A basic modular course on real world health data & research for the public.

Skill
Phenotype Definition, Characterisation an...
Defining phenotypes, characterising and evaluating using OHDSI...

Skill
R for Patient-level Prediction
Guidance on installing the PLP Package and predicting via R.

Skill
Patient-Level Prediction
Patient prediction modelling using OHDSI tools, in particular ATLAS.

Skill
Population-level Effect Estimation
Population-level effect estimation method: comparative cohort...

Tool
Infrastructure
Install and configure the OHDSI infrastructure.

Tool
OHDSI-in-a-Box
Deploy a single instance implementation of OHDSI tools and sample data.

Tool
OMOP CDM and Standardised Vocabularies
The structure of the common data model and its vocabularies.

Tool
Extract, Transform and Load
Map raw observational data to the OMOP CDM.

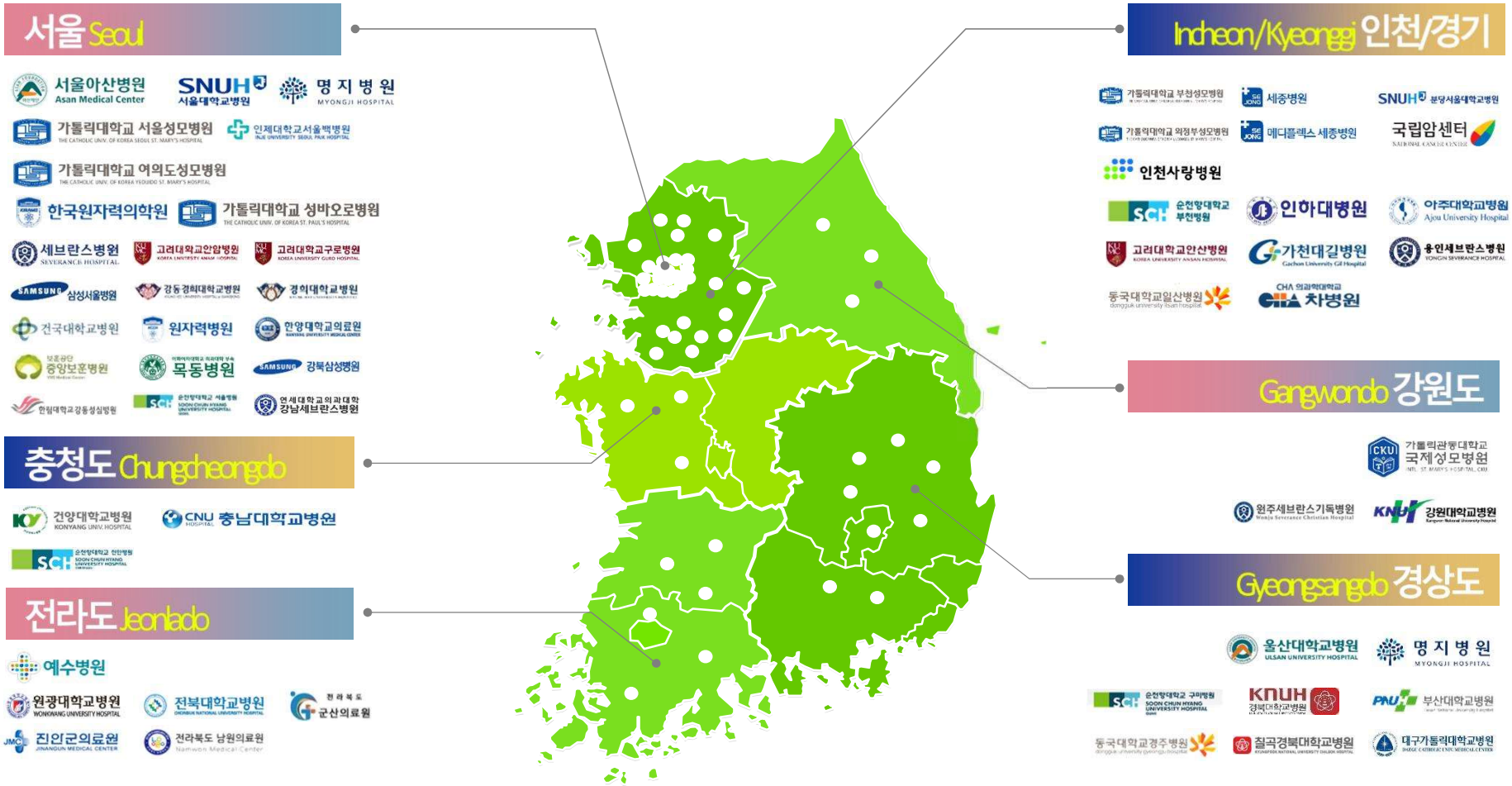
Tool
ETL Learning Pathway: Data Partner & SME Real...
Three of EHDEN's Data Partners and SMEs: insights into COVID-19 ET...

Tool
ATLAS
Design and execute analyses on observational data.

韓国での活動：FEEDER-NET

Data Network of 60+ Hospitals, 98M Patients

70% of Tertiary Teaching Hospitals

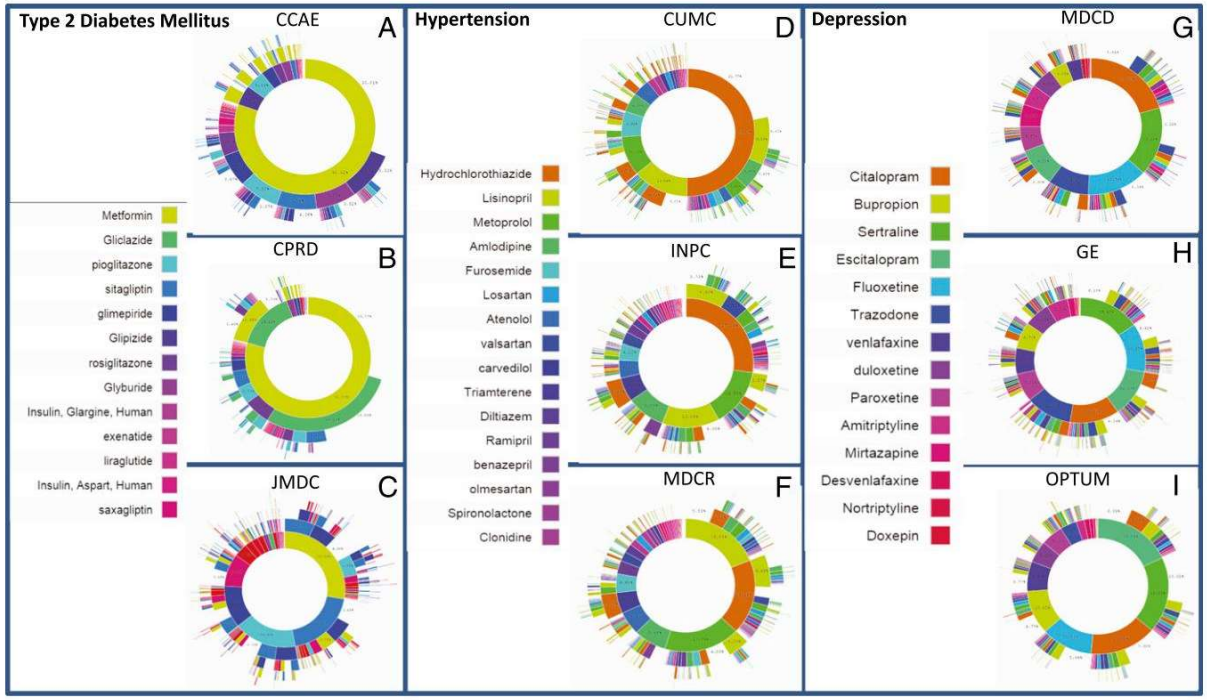


OHDSI国際共同研究の例

CrossMark
click for updates
COLLOQUIUM PAPER

Characterizing treatment pathways at scale using the OHDSI network

George Hripcsak[a,b,c,1], Patrick B. Ryan[c,d], Jon D. Duke[c,e], Nigam H. Shah[c,f], Rae Woong Park[c,g], Vojtech Huser[c,h], Marc A. Suchard[c,i,j,k], Martijn J. Schuemie[c,d], Frank J. DeFalco[c,d], Adler Perotte[a,c], Juan M. Banda[c,f], Christian G. Reich[c,l], Lisa M. Schilling[c,m], Michael E. Matheny[c,n,o], Daniella Meeker[c,p,q], Nicole Pratt[c,r], and David Madigan[c,s]



- 4カ国
- 11データソース
- 患者数合計2.5億人
- T2DM、高血圧、うつ
  治療薬推移実態調査

国際共同研究例(その2)

JAMA Network 掲載 2018/8
3カ国での共同研究

JAMA Network | Open

Original Investigation | Diabetes and Endocrinology

Association of Hemoglobin A_{1c} Levels With Use of Sulfonylureas, Dipeptidyl Peptidase 4 Inhibitors, and Thiazolidinediones in Patients With Type 2 Diabetes Treated With Metformin

Analysis From the Observational Health Data Sciences and Informatics Initiative

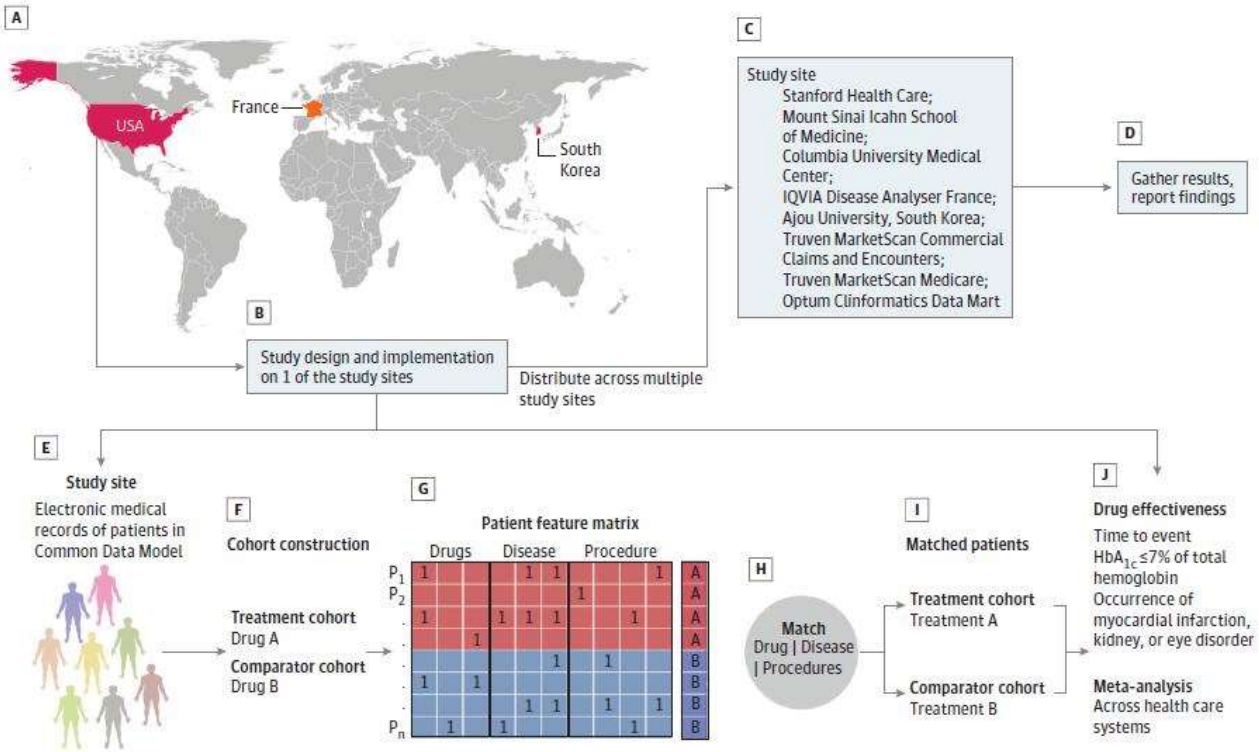
Rohit Vashisht, PhD; Kenneth Jung, PhD; Alejandro Schuler, MS; Juan
Kipp W. Johnson, MD, PhD; Mark M. Shervey, PhD; Hua Xu, PhD; Yon
Anthony Reckard, BS; Christian G. Reich, MD; James Weaver, MPH, M

糖尿病患者
メトホルミン
→ 第2選択
→ SU薬
→ DPP-4阻害剤
→ TZD系薬剤

第2選択薬剤による
下記の違いをみる。

- HbA1c低下
- 心筋梗塞
- 腎臓障害
- 眼障害

Figure 1. Overview of Multinational Cohort Study Design



国際共同研究例（その3）

THE LANCET

Comprehensive comparative effectiveness and safety of first-line antihypertensive drug classes: a systematic, multinational, large-scale analysis

Marc A Suchard, Martijn J Schuemie, Harlan M Krumholz, Seng Chan You, RuiJun Chen, Nicole Pratt, Christian G Reich, Jon Duke, David Madigan, George Hripcsak, Patrick B Ryan

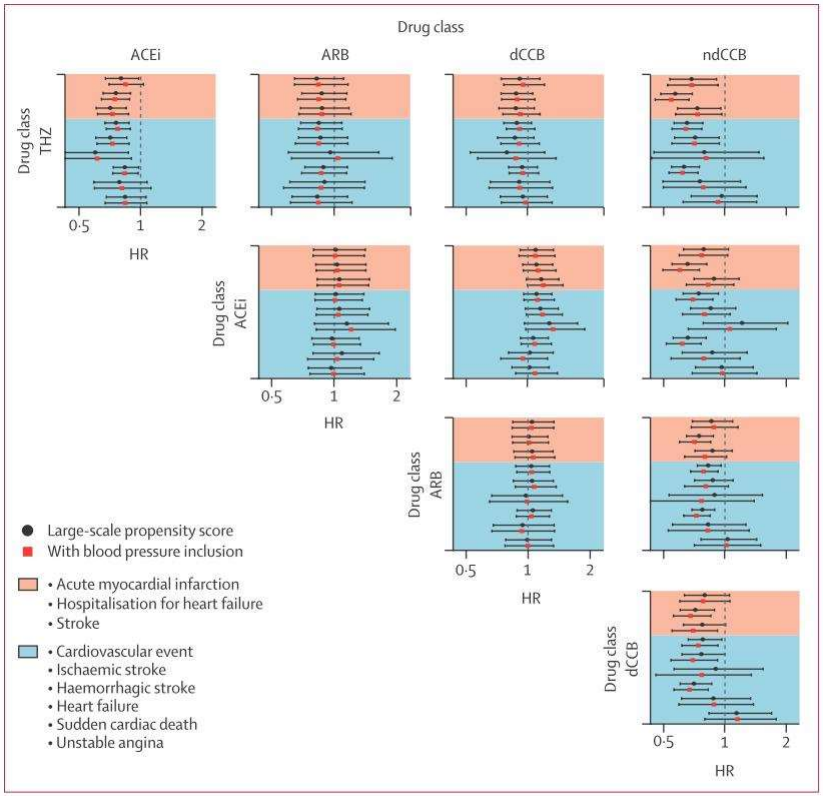
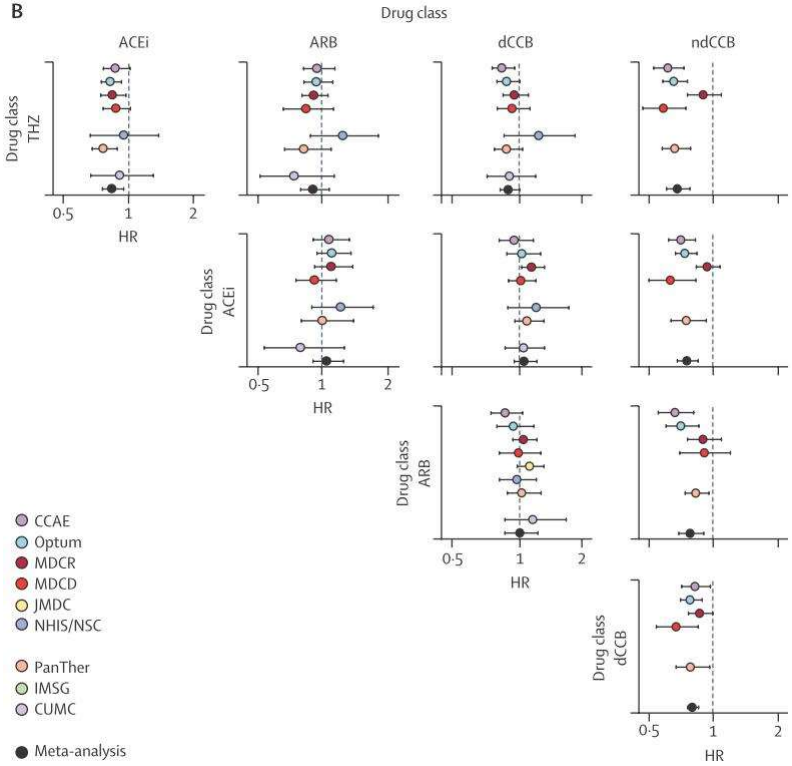


Figure 3: Effectiveness estimates comparing THZ to ACEi, ARB, dCCB, and ndCCB new users using propensity scores with and without baseline blood pressure adjustment in the Optum Pan-Therapeutic database



(B) Cardiovascular event risk estimates