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

OHDSI Japan evening conference #49

イブニング カンファレンス(第49回)

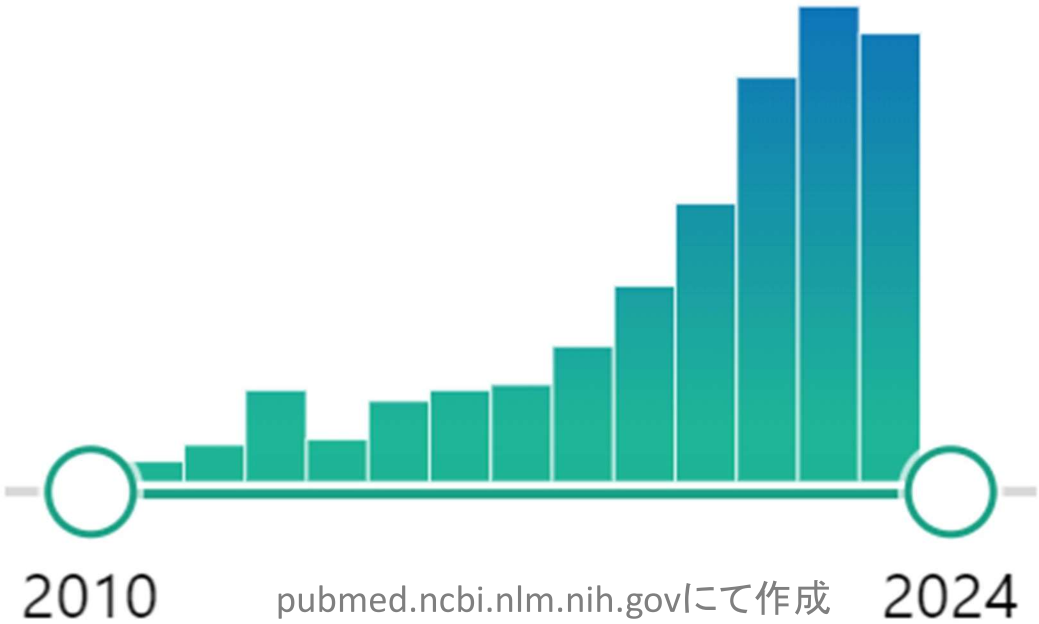
2023.12.28

本日の内容

- OHDSI関連論文紹介
- OHDSI Global 10周年
- ATLASを触ってみよう Hands-on

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：１１月374本→12月386本

- 検索に漏れているものがあるため、実際は累計500本を超えている。
- 年間では100本ペース。

OHDSI論文
新規1

より優れた安全性サーベイランス

> Stat Med. 2024 Jan 30;43(2):395-418. doi: 10.1002/sim.9968. Epub 2023 Nov 27.

Bayesian safety surveillance with adaptive bias correction

Fan Bu [1] [2], Martijn J Schuemie [1] [3], Akihiko Nishimura [4], Louisa H Smith [5] [6], Kristin Kostka [6], Thomas Falconer [7], Jody-Ann McLeggon [7], Patrick B Ryan [3], George Hripcsak [7], Marc A Suchard [1]

PMID: 38010062

Affiliations

1 Department of Biostatistics, University of California, Los Angeles, California, USA.
2 Department of Biostatistics, University of Michigan-Ann Arbor, Ann Arbor, Michigan, USA.
3 Janssen Research and Development, Raritan, New Jersey, USA.
4 Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland, USA.
5 Department of Health Sciences, Northeastern University, Portland, Maine, USA.
6 The OHDSI Center at the Roux Institute, Northeastern University, Portland, Maine, USA.
7 Department of Biomedical Informatics, Columbia University, New York, New York, USA.

Abstract

Postmarket safety surveillance is an integral part of mass vaccination programs. Typically relying on sequential analysis of real-world health data as they accrue, safety surveillance is challenged by sequential multiple testing and by biases induced by residual confounding in observational data. The current standard approach based on the maximized sequential probability ratio test (MaxSPRT) fails to satisfactorily address these practical challenges and it remains a rigid framework that requires prespecification of the surveillance schedule. We develop an alternative Bayesian surveillance procedure that addresses both aforementioned challenges using a more flexible framework. To mitigate bias, we jointly analyze a large set of negative control outcomes that are adverse events with no known association with the vaccines in order to inform an empirical bias distribution, which we then incorporate into estimating the effect of vaccine exposure on the adverse event of interest through a Bayesian hierarchical model. To address multiple testing and improve on flexibility, at each analysis timepoint, we update a posterior probability in favor of the alternative hypothesis that vaccination induces higher risks of adverse events, and then use it for sequential detection of safety signals. Through an empirical evaluation using six US observational healthcare databases covering more than 360 million patients, we benchmark the proposed procedure against MaxSPRT on testing errors and estimation accuracy, under two epidemiological designs, the historical comparator and the self-controlled case series. We demonstrate that our procedure substantially reduces Type 1 error rates, maintains high statistical power and fast signal detection, and provides considerably more accurate estimation than MaxSPRT. Given the extensiveness of the empirical study which yields more than 7 million sets of results, we present all results in a public R ShinyApp. As an effort to promote open science, we provide full implementation of our method in the open-source R package EvidenceSynthesis.

概要

市販後安全性サーベイランスは、集団予防接種プログラムに不可欠な要素である。一般的に、実世界の健康データを逐次的に分析することに依存しているが、安全性サーベイランスには、**逐次的な多重検定**や**観察データにおける残留交絡**によって誘発されるバイアスがつきまとうという課題がある。最大化逐次確率比検定（MaxSPRT）に基づく現在の標準的なアプローチは、このような現実的な課題に満足に対処できず、サーベイランススケジュールの事前指定を必要とする硬直的な枠組みである。我々は、より柔軟な枠組みを用いて、前述の2つの課題に対処する代替的なベイズ監視手順を開発する。バイアスを軽減するために、ワクチンとの関連性が知られていない有害事象である大規模なネガティブコントロールによる結果を共同で解析し、経験的バイアス分布を導き出し、それをベイズ階層モデルを通して、対象有害事象に対するワクチン曝露の影響を推定する。多重検定に対応し、柔軟性を向上させるために、各分析時点において、ワクチン接種が有害事象のより高いリスクを誘発するという対立仮説を支持する事後確率を更新し、安全性シグナルの逐次検出に使用する。3億6,000万人以上の患者をカバーする6つの米国の観察的医療データベースを用いた実証的評価を通じて、我々は、2つの疫学的デザイン（歴史的比較対象および自己対照症例シリーズ）の下で、検定誤差および推定精度について、MaxSPRTに対する提案手順をベンチマークする。その結果、本手法はMaxSPRTと比較して、タイプ1エラー率を大幅に減少させ、高い統計的検出力と高速なシグナル検出を維持し、かなり正確な推定が可能であることが示された。700万セット以上の結果が得られた実証研究の広範さを考慮し、すべての結果を公開のR ShinyAppで提示する。オープンサイエンスを促進する努力として、我々はオープンソースのRパッケージEvidenceSynthesisで我々の手法の完全な実装を提供する。

OHDSI論文
新規2

眼科領域での OMOP コンセプト

> Ophthalmol Sci. 2023 Aug 25;3(4):100391. doi: 10.1016/j.xops.2023.100391. eCollection 2023 Dec.

Advancing Toward a Common Data Model in Ophthalmology: Gap Analysis of General Eye Examination Concepts to Standard Observational Medical Outcomes Partnership (OMOP) Concepts

Cindy X Cai [1], William Halfpenny [2], Michael V Boland [3], Harold P Lehmann [4] [5], Michelle Hribar [6] [7] [8], Kerry E Goetz [6], Sally L Baxter [2] [9]

Affiliations

1 Wilmer Eye Institute, Johns Hopkins School of Medicine, Baltimore, Maryland.
2 Division of Biomedical Informatics, Department of Medicine, University of California San Diego, La Jolla, California.
3 Massachusetts Eye and Ear, Harvard Medical School, Boston, Massachusetts.
4 Division of Health Sciences Informatics, Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland.
5 Biomedical Informatics and Data Science, Division of General Internal Medicine, Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland.
6 Office of Data Science and Health Informatics, National Eye Institute, National Institute of Health, Bethesda, Maryland.
7 Department of Ophthalmology, Casey Eye Institute, Portland, Oregon.
8 Department of Medical Informatics and Clinical Epidemiology, Oregon Health & Science University, Portland, Oregon.
9 Division of Ophthalmology Informatics and Data Science, Viterbi Family Department of Ophthalmology and Shiley Eye Institute, University of California San Diego, La Jolla, California.

PMID: 38025162

Abstract

Purpose: Evaluate the degree of concept coverage of the general eye examination in one widely used electronic health record (EHR) system using the Observational Health Data Sciences and Informatics Observational Medical Outcomes Partnership (OMOP) common data model (CDM).
Design: Study of data elements.
Participants: Not applicable.

Methods: Data elements (field names and predefined entry values) from the general eye examination in the Epic foundation system were mapped to OMOP concepts and analyzed. Each mapping was given a Health Level 7 equivalence designation-*equal* when the OMOP concept had the same meaning as the source EHR concept, *wider* when it was missing information, *narrower* when it was overly specific, and *unmatched* when there was no match. Initial mappings were reviewed by 2 graders. Intergrader agreement for equivalence designation was calculated using Cohen's kappa. Agreement on the mapped OMOP concept was calculated as a percentage of total mappable concepts. Discrepancies were discussed and a final consensus created. Quantitative analysis was performed on *wider* and *unmatched* concepts.
Main outcome measures: Gaps in OMOP concept coverage of EHR elements and intergrader agreement of mapped OMOP concepts.
Results: A total of 698 data elements (210 fields, 488 values) from the EHR were analyzed. The intergrader kappa on the equivalence designation was 0.88 (standard error 0.03, $P < 0.001$). There was a 96% agreement on the mapped OMOP concept. In the final consensus mapping, 25% (1% fields, 31% values) of the EHR to OMOP concept mappings were considered *equal*, 50% (27% fields, 60% values) *wider*, 4% (8% fields, 2% values) *narrower*, and 21% (52% fields, 8% values) *unmatched*. Of the *wider* mapped elements, 46% were missing the laterality specification, 24% had other missing attributes, and 30% had both issues. *Wider* and *unmatched* EHR elements could be found in all areas of the general eye examination.
Conclusions: Most data elements in the general eye examination could not be represented precisely using the OMOP CDM. Our work suggests multiple ways to improve the incorporation of important ophthalmology concepts in OMOP, including adding laterality to existing concepts. There exists a strong need to improve the coverage of ophthalmic concepts in source vocabularies so that the OMOP CDM can better accommodate vision research.

概要
目的: OMOP CDMを用いて、広く使用されている1つの電子カルテ（EHR）システムにおける一般的な眼科検診の概念カバーの程度を評価する。
デザイン: データ要素の研究。**参加者:** 該当者なし。

方法: Epic基盤システムの一般眼科検診のデータ要素（フィールド名と定義済みの入力値）をOMOP概念にマッピングし、分析した。OMOP コンセプトがソース EHR コンセプトと同じ意味を持っている場合は同等、情報が不足している場合はより広い範囲、特定されすぎている場合はより狭い範囲、一致するものがない場合は不一致とした。最初のマッピングは2人の採点者によって検討された。同等性の指定に関する評定者間の一致は、コーエンのカッパを用いて計算された。マッピングされたOMOP概念に関する一致度は、マッピング可能な全概念に対するパーセンテージとして計算された。不一致は議論され、最終的なコンセンサスが形成された。より広い概念と一致しない概念について定量的解析が行われた。

主要評価項目: EHR 要素の OMOP 概念適用範囲におけるギャップ、およびマッピングされた OMOP 概念の評定者間合意。

結果: EHRの合計698のデータ要素（210フィールド、488値）が解析された。同等性指定に関する評定者間のカッパは0.88（標準誤差0.03、P < 0.001）であった。マッピングされたOMOP概念の一致率は96%であった。最終的なコンセンサスマッピングでは、EHRからOMOP概念へのマッピングの25%（1%のフィールド、31%の値）が等しく、50%（27%のフィールド、60%の値）がより広く、4%（8%のフィールド、2%の値）がより狭く、21%（52%のフィールド、8%の値）が一致しなかった。より広くマッピングされた要素のうち、46%はラテラリティ指定が欠落しており、24%はその他の属性が欠落しており、30%は両方の問題を抱えていた。より広範なEHR要素と一致しないEHR要素は、一般的な眼科検診の全領域で見つけることができた。

結論: 一般眼科検査におけるほとんどのデータ要素は、OMOP CDMを使用して正確に表現することが**できなかった**。われわれの研究は、既存の概念にラテラリティを追加するなど、OMOPに重要な眼科概念を組み込むことを改善する複数の方法を示唆している。OMOP CDM が視覚研究によりよく対応できるように、ソース・ボキャブラリーにおける眼科概念のカバレッジを改善する必要性が強く存在する。

OHDSI論文 新規3

Azureでのデータ変換

> Heliyon. 2023 Nov 2;9(11):e21586. doi: 10.1016/j.heliyon.2023.e21586. eCollection 2023 Nov.

Development and validation of the SickKids Enterprise-wide Data in Azure Repository (SEDAR)

Lin Lawrence Guo [1], Maryann Calligan [1], Emily Vettese [1], Sadie Cook [1], George Gagnidze [2], Oscar Han [2], Jiro Inoue [1], Joshua Lemmon [1], Johnson Li [2], Medhat Roshdi [2], Bohdan Sadovy [2], Steven Wallace [2], Lillian Sung [1] [3]

Affiliations − collapse

Affiliations

1 Program in Child Health Evaluative Sciences, The Hospital for Sick Children, Toronto, Canada.
2 Information Management Technology, The Hospital for Sick Children, Toronto, Canada.
3 Division of Haematology/Oncology, The Hospital for Sick Children, Toronto, Canada.

PMID: 38027579 PMCID: PMC10661187 DOI: 10.1016/j.heliyon.2023.e21586

Abstract

Objectives: To describe the processes developed by The Hospital for Sick Children (SickKids) to enable utilization of electronic health record (EHR) data by creating sequentially transformed schemas for use across multiple user types.
Methods: We used Microsoft Azure as the cloud service provider and named this effort the SickKids Enterprise-wide Data in Azure Repository (SEDAR). Epic Clarity data from on-premises was copied to a virtual network in Microsoft Azure. Three sequential schemas were developed. The Filtered Schema added a filter to retain only SickKids and valid patients. The Curated Schema created a data structure that was easier to navigate and query. Each table contained a logical unit such as patients, hospital encounters or laboratory tests. Data validation of randomly sampled observations in the Curated Schema was performed. The SK-OMOP Schema was designed to facilitate research and machine learning. Two individuals mapped medical elements to standard Observational Medical Outcomes Partnership (OMOP) concepts.
Results: A copy of Clarity data was transferred to Microsoft Azure and updated each night using log shipping. The Filtered Schema and Curated Schema were implemented as stored procedures and executed each night with incremental updates or full loads. Data validation required up to 16 iterations for each Curated Schema table. OMOP concept mapping achieved at least 80 % coverage for each SK-OMOP table.
Conclusions: We described our experience in creating three sequential schemas to address different EHR data access requirements. Future work should consider replicating this approach at other institutions to determine whether approaches are generalizable.

概要
目的: The Hospital for Sick Children（SickKids）が開発した、複数のユーザータイプで使用できるように順次変換されたスキーマを作成することで、電子カルテ（EHR）データの利用を可能にするプロセスを説明すること。

方法: クラウド・サービス・プロバイダーとして Microsoft Azure を使用し、この取り組みを SickKids Enterprise-wide Data in Azure Repository (SEDAR) と名付けました。オンプレミスの Epic Clarity データを Microsoft Azure の仮想ネットワークにコピーしました。3つの連続したスキーマが開発されました。Filtered Schema は、SickKids と有効な患者のみを保持するフィルタを追加しました。Curated Schemaは、ナビゲートとクエリが容易なデータ構造を作成した。各テーブルには、患者、病院受診、臨床検査などの論理的な単位が含まれていました。キュレーション・スキーマで無作為にサンプリングしたオブザベーションのデータ検証を実施した。SK-OMOP スキーマは、研究および機械学習を容易にするために設計された。2人の担当者が、医療要素を標準的なOMOP概念にマッピングした。

結果: Clarityデータのコピーが Microsoft Azureに転送され、ログシッピングを使用して毎晩更新されました。Filtered SchemaとCurated Schemaはストアドプロシージャとして実装され、増分更新またはフルロードで毎晩実行されました。データの検証には、各キュレーション・スキーマ・テーブルで最大16回の反復が必要だった。OMOPコンセプトマッピングは、各SK-OMOPテーブルで少なくとも80%のカバレッジを達成した。

結論: 異なるEHRデータアクセス要件に対応するために3つの連続スキーマを作成した経験について述べた。今後の研究では、このアプローチを他の機関でも再現し、アプローチが一般化可能かどうかを判断することを検討すべきである。

OHDSI論文
新規4

Phenotypingのロバスト度

> JAMIA Open. 2023 Nov 21;6(4):ooad096. doi: 10.1093/jamiaopen/ooad096. eCollection 2023 Dec.

Evaluating the impact of alternative phenotype definitions on incidence rates across a global data network

Rupa Makadia [1, 2], Azza Shoaibi [1, 2], Gowtham A Rao [1, 2], Anna Ostropolets [1, 3], Peter R Rijnbeek [1, 4], Erica A Voss [1, 2], Talita Duarte-Salles [1, 5], Juan Manuel Ramírez-Anguita [6], Miguel A Mayer [7], Filip Maljković [8], Spiros Denaxas [9, 10], Fredrik Nyberg [11], Vaclav Papez [9], Anthony G Sena [1, 2, 4], Thamir M Alshammari [1, 12], Lana Y H Lai [1, 13], Kevin Haynes [2], Marc A Suchard [1, 14], George Hripcsak [1, 3], Patrick B Ryan [1, 2, 3]

- OHDSI Collaborators, Observational Health Data Sciences and Informatics (OHDSI), New York, NY 10027, United States.
- [2]Global Epidemiology, Janssen Pharmaceutical Research and Development, LLC, Titusville, NJ 08560, United States.
- [3]Department of Biomedical Informatics, Columbia University Irving Medical Center, New York, NY 10027, United States.
- [4]Department of Medical Informatics, Erasmus University Medical Center, Rotterdam, 3000 CA, The Netherlands.
- [5]Fundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (IDIAPJGol), Barcelona, 08007, Spain.
- [6]Research Programme on Biomedical Informatics (GRIB), Hospital del Mar Medical Research Institute (IMIM), Barcelona, 08003, Spain.
- [7]Management Control Department, Parc de Salut Mar (PSMAR), Barcelona, 08007, Spain.
- [8]Research and Development, Heliant d.o.o, Belgrade, 11000, Serbia.
- [9]Institute of Health Informatics, University College London, London, NW1 2DA, United Kingdom.
- [10]British Heart Foundation Data Science Centre, HDR, London, NW1 2DA, United Kingdom.
- [11]School of Public Health and Community Medicine, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, 40530, Sweden.
- [12]College of Pharmacy, Prince Sattam Bin Abdulaziz University, Riyadh, 11942, Saudi Arabia.
- [13]Division of Informatics, Imaging and Data Sciences, University of Manchester, Manchester, M13 9PL, United Kingdom.
- [14]Department of Biostatistics, University of California, Los Angeles, Los Angeles, CA 90025, United States.

Abstract

Objective: Developing accurate phenotype definitions is critical in obtaining reliable and reproducible background rates in safety research. This study aims to illustrate the differences in background incidence rates by comparing definitions for a given outcome.

Materials and methods: We used 16 data sources to systematically generate and evaluate outcomes for 13 adverse events and their overall background rates. We examined the effect of different modifications (inpatient setting, standardization of code set, and code set changes) to the computable phenotype on background incidence rates.

Results: Rate ratios (RRs) of the incidence rates from each computable phenotype definition varied across outcomes, with inpatient restriction showing the highest variation from 1 to 11.93. Standardization of code set RRs ranges from 1 to 1.64, and code set changes range from 1 to 2.52.

Discussion: The modification that has the highest impact is requiring inpatient place of service, leading to at least a 2-fold higher incidence rate in the base definition. Standardization showed almost no change when using source code variations. The strength of the effect in the inpatient restriction is highly dependent on the outcome. Changing definitions from broad to narrow showed the most variability by age/gender/database across phenotypes and less than a 2-fold increase in rate compared to the base definition.

Conclusion: Characterization of outcomes across a network of databases yields insights into sensitivity and specificity trade-offs when definitions are altered. Outcomes should be thoroughly evaluated prior to use for background rates for their plausibility for use across a global network.

概要
目的: 安全性研究において、正確な表現型を定義することは、信頼性が高く再現性のあるバックグランド発生率を得るために重要である。本研究の目的は、ある結果に対する定義を比較することにより、バックグラウンド発生率の違いを明らかにすることである。

材料と方法: 16のデータソースを用いて、13の有害事象のアウトカムとその全体のバックグラウンド率を系統的に作成し、評価した。計算可能な表現型に対するさまざまな修正（入院設定、コードセットの標準化、コードセットの変更）がバックグラウンド発生率に及ぼす影響を検討した。

結果: 各計算可能な表現型の定義による発生率の率比（RR）は転帰によって異なり、入院制限では1～11.93と最も大きなばらつきを示した。コードセットの標準化RRは1から1.64の範囲であり、コードセットの変更は1から2.52の範囲であった。

考察: 最も影響が大きい変更は、入院中のサービス提供場所を必要とするもので、基本定義では少なくとも2倍の発生率になる。標準化では、ソースコードのバリエーションを使用してもほとんど変化は見られなかった。入院制限における効果の強さは、アウトカムに大きく依存する。定義を広義から狭義に変更した場合、表現型間で年齢／性別／データベースによるばらつきが最も大きく、基本定義と比較して発生率の増加は2倍未満であった。

結論: データベースのネットワーク全体にわたる転帰の特徴づけにより、定義を変更した場合の感度と特異度のトレードオフに関する洞察が得られる。アウトカムをバックグラウンド率に使用する前に、グローバルネットワーク全体で使用する妥当性を徹底的に評価すべきである。

OHDSI論文
新規5

様々なデータモデル間の変換

> J Biomed Inform. 2023 Nov 29:149:104558. doi: 10.1016/j.jbi.2023.104558. Online ahead of print.

Convert-Pheno: A software toolkit for the interconversion of standard data models for phenotypic data

Manuel Rueda [1], Ivo C Leist [2], Ivo G Gut [2]

Affiliations – collapse

Affiliations

1 Centro Nacional de Análisis Genómico, C/Baldiri Reixac 4, 08028 Barcelona, Spain; Universitat de Barcelona (UB), Barcelona, Spain. Electronic address: manuel.rueda@cnag.eu.

2 Centro Nacional de Análisis Genómico, C/Baldiri Reixac 4, 08028 Barcelona, Spain; Universitat de Barcelona (UB), Barcelona, Spain.

PMID: 38035971 DOI: 10.1016/j.jbi.2023.104558

Abstract

Efficient sharing and integration of phenotypic data is crucial for advancing biomedical research and enhancing patient outcomes in precision medicine and public health. To achieve this, the health data community has developed standards to promote the harmonization of variable names and values. However, the use of diverse standards across different research centers can hinder progress. Here we present Convert-Pheno, an open-source software toolkit that enables the interconversion of common data models for phenotypic data such as Beacon v2 Models, CDISC-ODM, OMOP-CDM, Phenopackets v2, and REDCap. Along with the software, we have created a detailed documentation that includes information on deployment and installation.

概要

表現型データの効率的な共有と統合は、生物医学研究を推進し、精密医療と公衆衛生における患者の転帰を向上させるために極めて重要である。これを達成するために、ヘルスデータコミュニティは変数名と値の調和を促進する標準を開発した。しかし、異なる研究センター間で多様な標準を使用することは、研究の進展を妨げる可能性がある。ここでは、Beacon v2 Models、CDISC-ODM、OMOP-CDM、Phenopackets v2、REDCapなどの表現型データの共通データモデルの相互変換を可能にするオープンソースのソフトウェアツールキットであるConvert-Phenoを紹介する。ソフトウェアとともに、展開とインストールに関する情報を含む詳細なドキュメントを作成した。

OHDSI論文
新規6

OMOP CDMからFHIR取り出し

› medRxiv. 2023 Nov 22:2023.08.09.23293900. doi: 10.1101/2023.08.09.23293900. Preprint

MENDS-on-FHIR: Leveraging the OMOP common data model and FHIR standards for national chronic disease surveillance

Shahim Essaid [1], Jeff Andre [2], Ian M Brooks [1] [3], Katherine H Hohman [4], Madelyne Hull [3], Sandra L Jackson [5], Michael G Kahn [1] [3], Emily M Kraus [6] [7], Neha Mandadi [1] [3], Amanda K Martinez [4], Joyce Y Mui [1] [3], Bob Zambarano [2], Andrey Soares [8]

PMID: 38045364

Affiliations

1. Department of Biomedical Informatics, University of Colorado Anschutz Medical Campus, Denver CO.
2. Commonwealth Informatics Inc, Waltham MA.
3. Health Data Compass, University of Colorado Anschutz Medical Campus, Denver CO.
4. National Association of Chronic Disease Directors (NACDD), Decatur GA.
5. National Center for Chronic Disease Prevention and Health Promotion, Centers for Disease Control and Prevention (CDC), Atlanta GA.
6. Kraushold Consulting, Denver CO.
7. Public Health Informatics Institute, Decatur, GA.
8. Department of Medicine, University of Colorado Anschutz Medical Campus, Denver CO.

Abstract

Objective: The Multi-State EHR-Based Network for Disease Surveillance (MENDS) is a population-based chronic disease surveillance distributed data network that uses institution-specific extraction-transformation-load (ETL) routines. MENDS-on-FHIR examined using Health Language Seven's Fast Healthcare Interoperability Resources (HL7® FHIR®) and US Core Implementation Guide (US Core IG) compliant resources derived from the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) to create a standards-based ETL pipeline.

Materials and methods: The input data source was a research data warehouse containing clinical and administrative data in OMOP CDM Version 5.3 format. OMOP-to-FHIR transformations, using a unique JavaScript Object Notation (JSON)-to-JSON transformation language called Whistle, created FHIR R4 V4.0.1/US Core IG V4.0.0 conformant resources that were stored in a local FHIR server. A REST-based Bulk FHIR $export request extracted FHIR resources to populate a local MENDS database.

Results: Eleven OMOP tables were used to create 10 FHIR/US Core compliant resource types. A total of 1.13 trillion resources were extracted and inserted into the MENDS repository. A very low rate of non-compliant resources was observed.

Discussion: OMOP-to-FHIR transformation results passed validation with less than a 1% non-compliance rate. These standards-compliant FHIR resources provided standardized data elements required by the MENDS surveillance use case. The Bulk FHIR application programming interface (API) enabled population-level data exchange using interoperable FHIR resources. The OMOP-to-FHIR transformation pipeline creates a FHIR interface for accessing OMOP data.

Conclusion: MENDS-on-FHIR successfully replaced custom ETL with standards-based interoperable FHIR resources using Bulk FHIR. The OMOP-to-FHIR transformations provide an alternative mechanism for sharing OMOP data.

概要
目的: Multi-State EHR-Based Network for Disease Surveillance（MENDS）は、施設固有の抽出-変換-ロード（ETL）ルーチンを使用する集団ベースの慢性疾患サーベイランス分散データネットワークである。MENDS-on-FHIR では、 HL7® FHIR® と、 US Core Implementation Guide (US Core IG) 準拠のリソースを使用して、 OMOP CDM から標準に基づく ETL パイプラインを作成することを検討した。

材料と方法: 入力データソースは、OMOP CDM Version 5.3形式の臨床データおよび管理データを含む研究データウェアハウスであった。Whistleと呼ばれる独自のJavaScript Object Notation（JSON）-to-JSON変換言語を使用したOMOP-to-FHIR変換により、FHIR R4 V4.0.1/US Core IG V4.0.0適合リソースが作成され、ローカルのFHIRサーバーに保存された。RESTベースの**Bulk FHIR** $exportリクエストは、ローカルのMENDSデータベースに入力するFHIRリソースを抽出した。

結果: 10のFHIR/US Core準拠リソースタイプを作成するために11のOMOPテーブルが使用された。合計1兆1300億のリソースが抽出され、MENDSリポジトリに挿入された。非準拠リソースの割合は非常に低かった。

考察: OMOPからFHIRへの変換結果は、1%未満の非準拠率で検証に合格しました。これらの標準準拠のFHIRリソースは、MENDSサーベイランスのユースケースで必要とされる標準化されたデータ要素を提供した。一括 FHIR アプリケーション・プログラミング・インターフェース（API）は、相互運用可能な FHIR リソースを使用した集団レベルのデータ交換を可能にした。OMOP-to-FHIR変換パイプラインは、OMOPデータにアクセスするためのFHIRインターフェースを作成する。

結論: MENDS-on-FHIRは、Bulk FHIRを使用して、カスタムETLを標準ベースの相互運用可能なFHIRリソースに置き換えることに成功した。OMOP-to-FHIR変換は、OMOPデータを共有するための代替メカニズムを提供する。

来月へ

- Conversion of CPRD AURUM Data into the OMOP Common Data Model.

Inform Med Unlocked. 2023;43:101407. doi: 10.1016/j.imu.2023.101407. Epub 2023 Nov 10.
PMID: 38046363

- Transforming Estonian health data to the Observational Medical Outcomes Partnership (OMOP) Common Data Model: lessons learned.

JAMIA Open. 2023 Dec 5;6(4):ooad100. doi: 10.1093/jamiaopen/ooad100. eCollection 2023 Dec.
PMID: 38058679

- MI-Common Data Model: Extending Observational Medical Outcomes Partnership-Common Data Model (OMOP-CDM) for Registering Medical Imaging Metadata and Subsequent Curation Processes.

JCO Clin Cancer Inform. 2023 Sep;7:e2300101. doi: 10.1200/CCI.23.00101.
PMID: 38061012

- Patient-Centered Economic Burden of Exudative Age-Related Macular Degeneration: Retrospective Cohort Study.

JMIR Public Health Surveill. 2023 Dec 8;9:e49852. doi: 10.2196/49852.
PMID: 38064251

- Implementation of inclusion and exclusion criteria in clinical studies in OHDSI ATLAS software.

Sci Rep. 2023 Dec 18;13(1):22457. doi: 10.1038/s41598-023-49560-w.
PMID: 38105303

- From data strategy to implementation to advance cancer research and cancer care: A French comprehensive cancer center experience.

PLOS Digit Health. 2023 Dec 19;2(12):e0000415. doi: 10.1371/journal.pdig.0000415. eCollection 2023 Dec.
PMID: 38113207

今月のOHDSI Global

●APAC Call テーマ

Dec. 14　APAC 2023 Recap/Chapter Year-End Updates

●Global Community Call テーマ

Nov. 28　OHDSI Coordinating Center

Dec. 5　Recent OHDSI Publications

Dec. 12　Happy 10th Birthday, OHDSI - Decade & Year in Review
OHDSI 10周年！

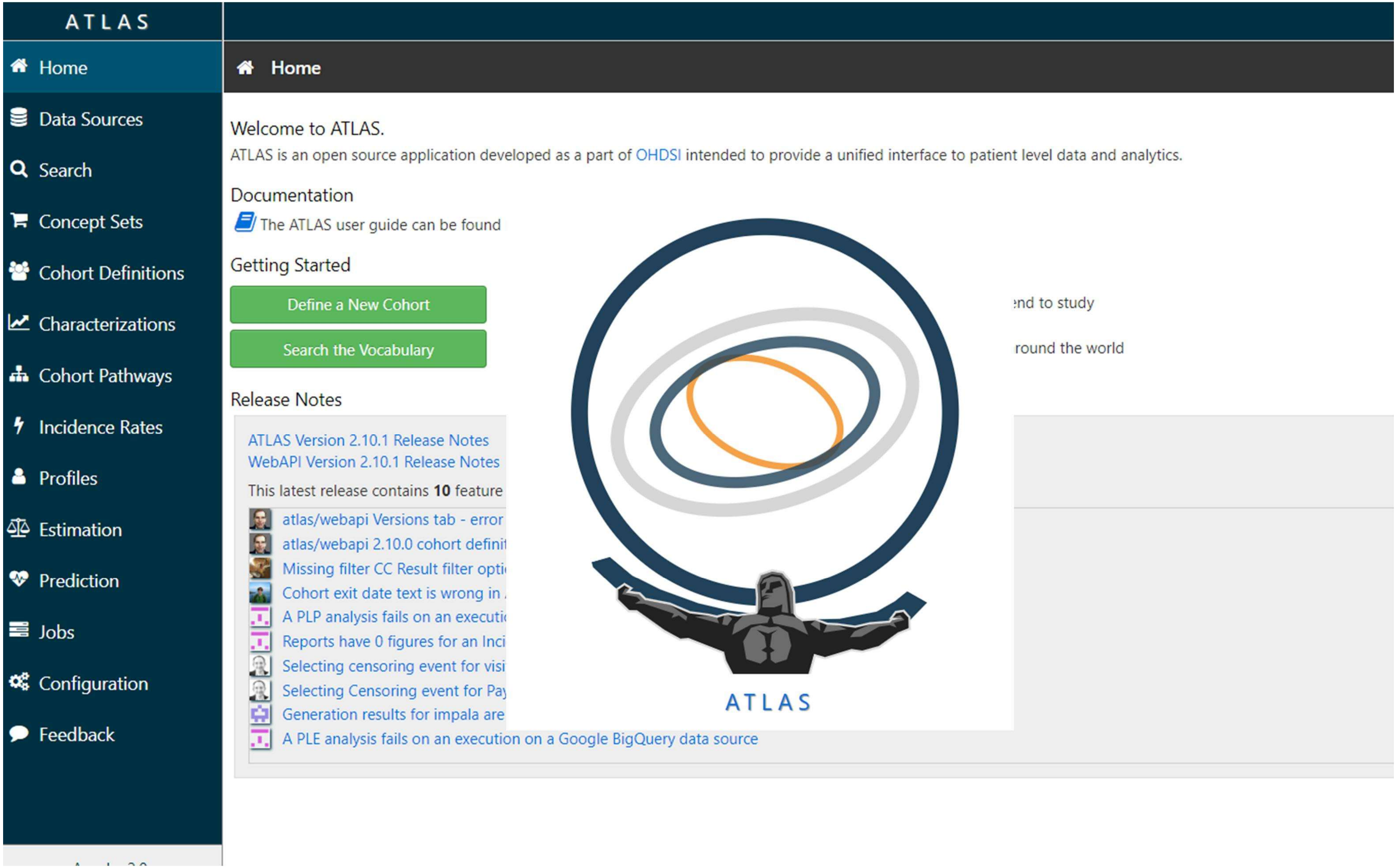
Dec. 19　Holiday-Themed Goodbye to 2023!

Dec. 12: Happy 10th Birthday to OHDSI

@OHDSI　www.ohdsi.org　#JoinTheJourney　ohdsi

2013年12月16日、George Hripcsakが中心となってOHDSIコミュニティが正式に結成された。年が明けてすぐ、コロンビア大学の生物医学情報学科内で最初の対面会議が開かれた。そこからどのようにして、3,800人以上の協力者を擁するグローバル・コミュニティになったのだろうか？12月12日のコミュニティ・コールでは、OHDSIの10年を振り返るとともに、2023年の年頭所感を簡単に述べた。

ATLASを触ってみよう



データセット説明

SynPUFデータセット

CMS.gov
Centers for Medicare & Medicaid Services

Search CMS | Search

Medicare | Medicaid/CHIP | Medicare-Medicaid Coordination | Private Insurance | Innovation Center | Regulations & Guidance | Research, Statistics, Data & Systems | Outreach & Education

Home › Research, Statistics, Data & Systems › Medicare Claims Synthetic Public Use Files (SynPUFs) › CMS 2008-2010 Data Entrepreneurs' Synthetic Public Use File (DE-SynPUF)

CMS 2008-2010 Data Entrepreneurs' Synthetic Public Use File (DE-SynPUF)

DE1.0 Sample 1

DE1.0 Sample 2

CMS 2008-2010 Data Entrepreneurs' Synthetic Public Use File (DE-SynPUF)

CMS 2008-2010 Data Entrepreneurs' Synthetic Public Use File (DE-SynPUF)

The DE-SynPUF was created with the goal of providing a realistic set of claims data in the public domain while providing the very highest degree of protection to the Medicare beneficiaries' protected health information. The purposes of the DE-SynPUF are to:

DE-SynPUF	Unit of record	Number of Records 2008	Number of Records 2009	Number of Records 2010
Beneficiary Summary	Beneficiary	2,326,856	2,291,320	2,255,098
Inpatient Claims	claim	547,800	504,941	280,081
Outpatient Claims	claim	5,673,808	6,519,340	3,633,839
Carrier Claims	claim	34,276,324	37,304,993	23,282,135
Prescription Drug Events (PDE)	event	39,927,827	43,379,293	27,778,849

⇒5%で
11万人

https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/SynPUFs/DE_Syn_PUF

Syntheaというものもある

今回は使わない

https://synthea.mitre.org/

SyntheticMass SYNTHEA

Downloads

- SyntheticMass Data, Version 2 (24 May, 2017): 21GB. FHIR 3.0.1, CSV, C-CDA
- SyntheticMass Data, Version 1 (27 Feb, 2017): 28GB. FHIR 1.8.0, CSV, C-CDA

In addition, sample files containing 1,000 patient records in multiple formats are available:

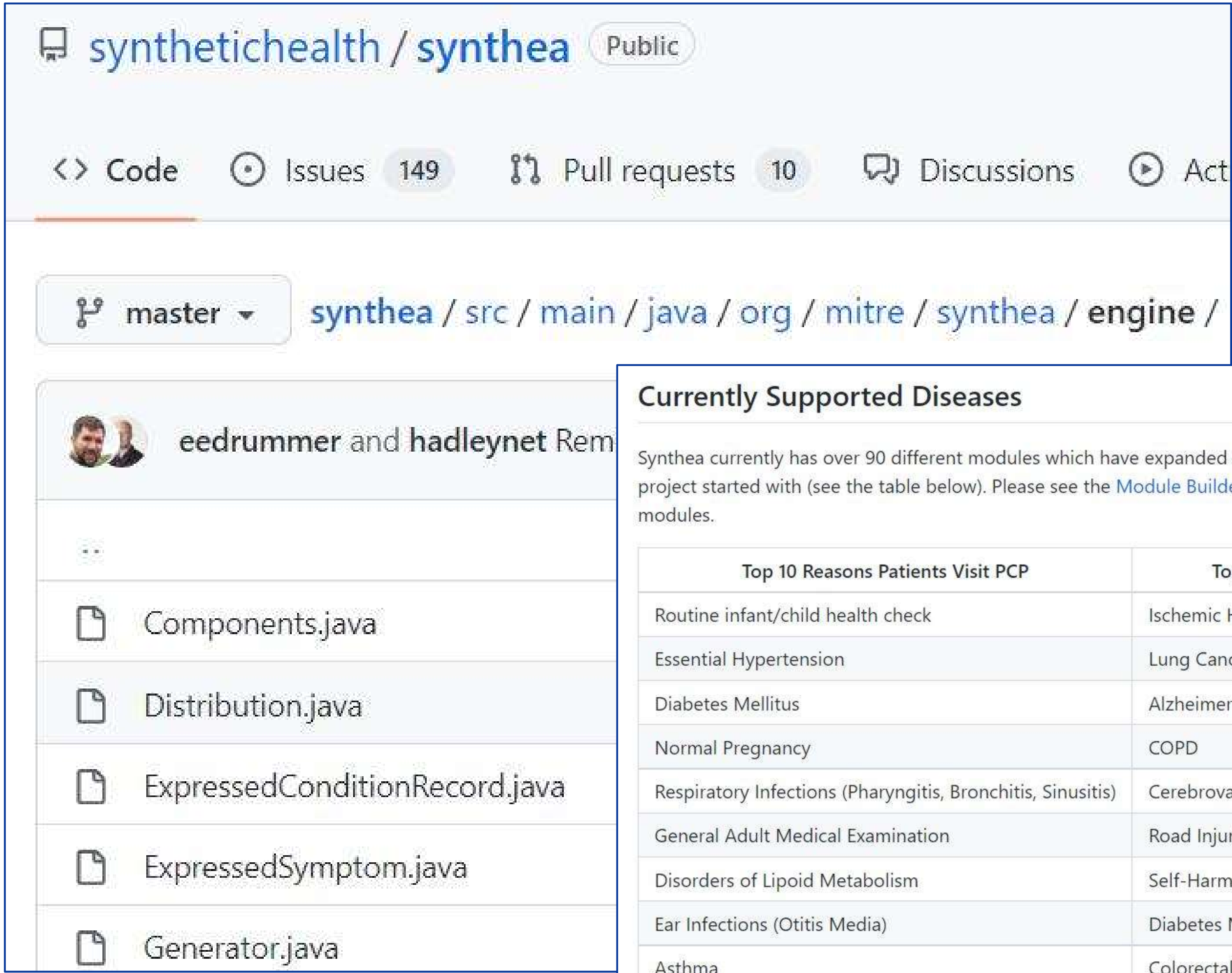
- 1K Sample Synthetic Patient Records, FHIR R4 | [mirror]: 81 MB
- 1K Sample Synthetic Patient Records, FHIR STU3 | [mirror]: 83 MB
- 1K Sample Synthetic Patient Records, FHIR DSTU2 | [mirror]: 46 MB
- 1K Sample Synthetic Patient Records, C-CDA | [mirror]: 47 MB
- 1K Sample Synthetic Patient Records, CSV | [mirror]: 9 MB

Specialized Data Sets

COVID-19

- COVID-19 10K, CSV | [mirror]: 54 MB
    - Ten thousand synthetic patients records with COVID-19 in the CSV format.
- COVID-19 100K, CSV: 512 MB
    - One hundred thousand synthetic patients records with COVID-19 in the CSV format.

患者データを生成できる



Currently Supported Diseases

Synthea currently has over 90 different modules which have expanded beyond the original "*Two Top Tens*" that the project started with (see the table below). Please see the Module Builder to browse the currently supported disease modules.

Top 10 Reasons Patients Visit PCP	Top 10 Years of Life Lost
Routine infant/child health check	Ischemic Heart Disease
Essential Hypertension	Lung Cancer
Diabetes Mellitus	Alzheimer's Disease
Normal Pregnancy	COPD
Respiratory Infections (Pharyngitis, Bronchitis, Sinusitis)	Cerebrovascular Disease
General Adult Medical Examination	Road Injuries
Disorders of Lipoid Metabolism	Self-Harm
Ear Infections (Otitis Media)	Diabetes Mellitus
Asthma	Colorectal Cancer
Urinary Tract Infections	Drug Use Disorders (limited to Opioids)

データセットを見てみましょう

CMS DE-SynPUF

https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/SynPUFs/DE_Syn_PUF

DE1.0 Sample 1 2008 Beneficiary Summary File (ZIP)
DE1.0 Sample 1 2008-2010 Carrier Claims 1
DE1.0 Sample 1 2008-2010 Carrier Claims 2
DE1.0 Sample 1 2008-2010 Inpatient Claims (ZIP)
DE1.0 Sample 1 2008-2010 Outpatient Claims (ZIP)
DE1.0 Sample 1 2008-2010 Prescription Drug Events
DE1.0 Sample 1 2009 Beneficiary Summary File (ZIP)
DE1.0 Sample 1 2010 Beneficiary Summary File (ZIP)

Synthea

https://synthea.mitre.org/

小さなサンプルの
CSV形式のファイルをダウンロードしてみる

- 1K Sample Synthetic Patient Records, CSV | [mirror]: 9 MB

ATLASを見てみましょう

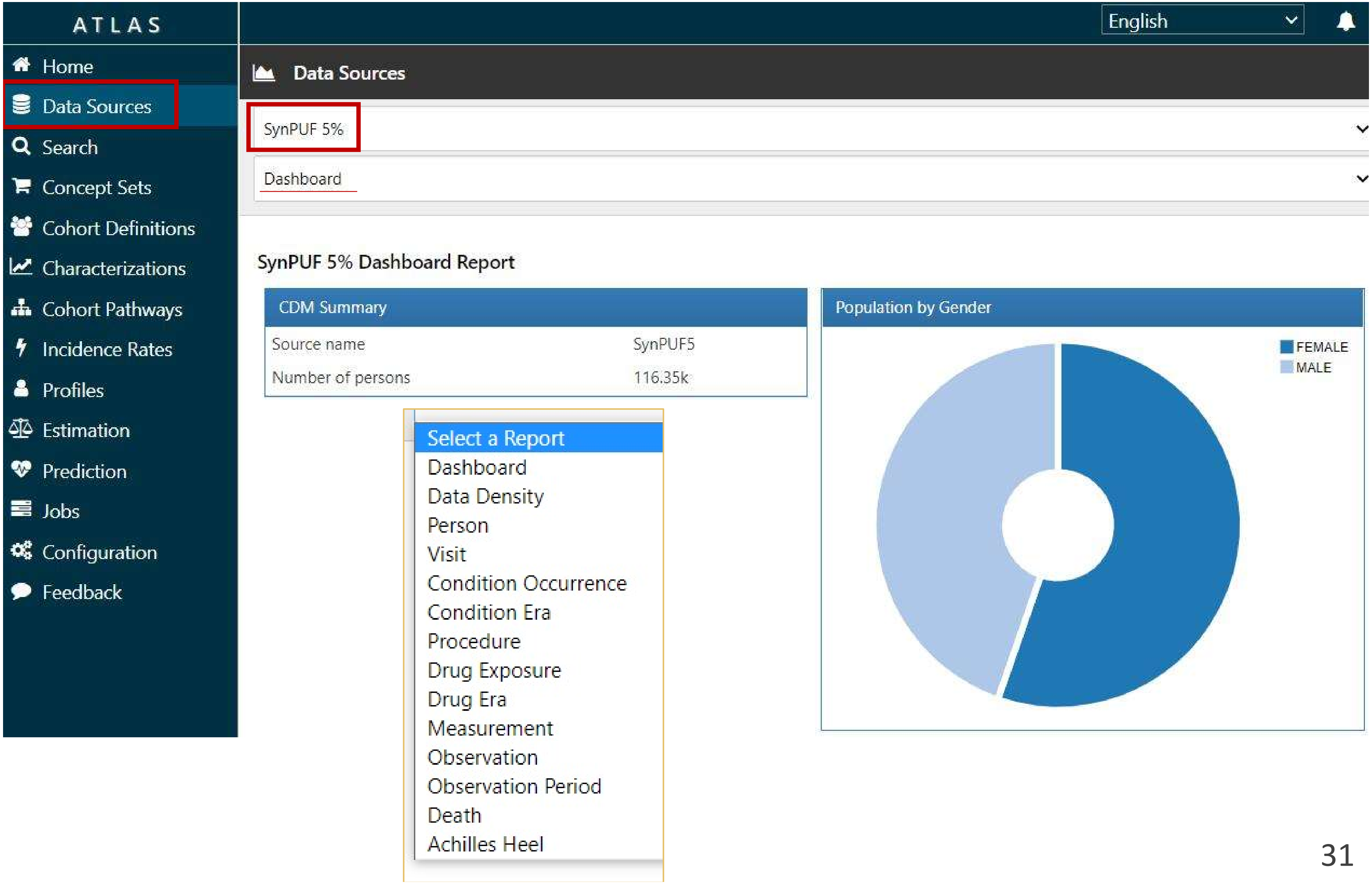
ATLAS は、OHDSI コミュニティによって開発された、無料で公開されている Web ベースのオープンソース ソフトウェア アプリケーションであり、患者レベルの観察データから実世界の証拠を生成するための観察分析の設計と実行をサポートします。

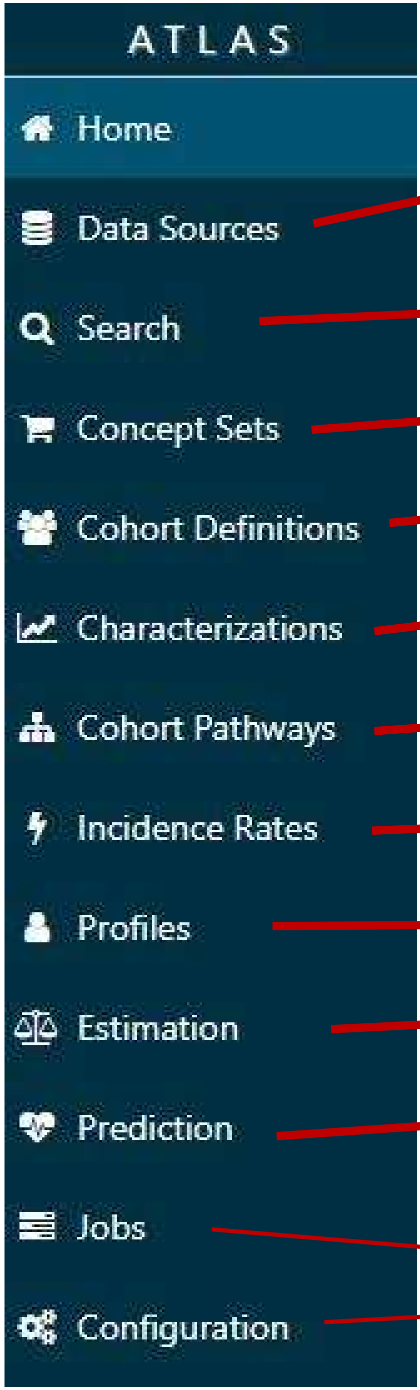
Atlas は、OMOP Common Data Model V5に標準化された 1 つまたは複数の観測データベースにわたって分析を実行するために、機関内にローカルにインストールできるオープンな科学分析プラットフォームであり、同じオープン サイエンス コミュニティの基準とツールを採用しているOHDSI コミュニティの他の組織との分析設計の交換を容易にします。

デモサイト　https://atlas-demo.ohdsi.org/
GitHUB　https://github.com/OHDSI/Atlas/wiki

YouTube
ATLAS Overview
https://www.youtube.com/watch?v=dr9FhEkf04o

Data Sources ⇒ “SynPUF 5%”を選ぶ ⇒ Dashboard





どのようなデータが入っているか、想定と異なるデータになってないかを確認する。

ボキャブラリからコンセプトを検索する

コンセプトのセットを形成する

コホート（患者群）を定義する

コホートの特徴を集計する

コホートの患者の状態推移を見える化する

罹患率を算出する

個々の患者の情報を見る

推計統計分析を行う

機械学習による予測を行う

実行中の処理リスト

ATLAS/WebAPI上の諸元

Concept

Concept Set

Cohort

Concept、 Concept Set、 Cohort

自分で命名

- Diabetes Mellitus
  - Type 1 diabetes
  - Type 2 diabetes
- Hypertension
- Thiazide
  - Hydrochlorothiazide
  - Methyclothiazide
  - Polythiazide
- Metformin
- Glyburide / Metformin
- Linagliptin / Metformin
- Sulfonylurea
  - Glyburide
  - Glipizide
  - Tolazamide

HypertensionSet

- Hypertension

ThiazideSet

- Hydrochlorothiazide
- Methyclothiazide
- Polythiazide

DMdrugSet

- Glyburide
- Glipizide
- Tolazamide
- Metformin
- Glyburide / Metformin
- Linagliptin / Metformin

Cohort-A

- HypertensionSet
- ThiazideSet
  30日以上使用
- 収縮期血圧 > 150mmHg

Cohort-B

- T2DM-Set
- MetforminSet
  30日以上使用
- 持続血糖測定器使用Set

OMOPで登場する、領域を表すことば

Visit
　受診（外来、入院）

Condition
　患者状態（疾患・傷病、症候、症状）

Drug
　医薬品使用

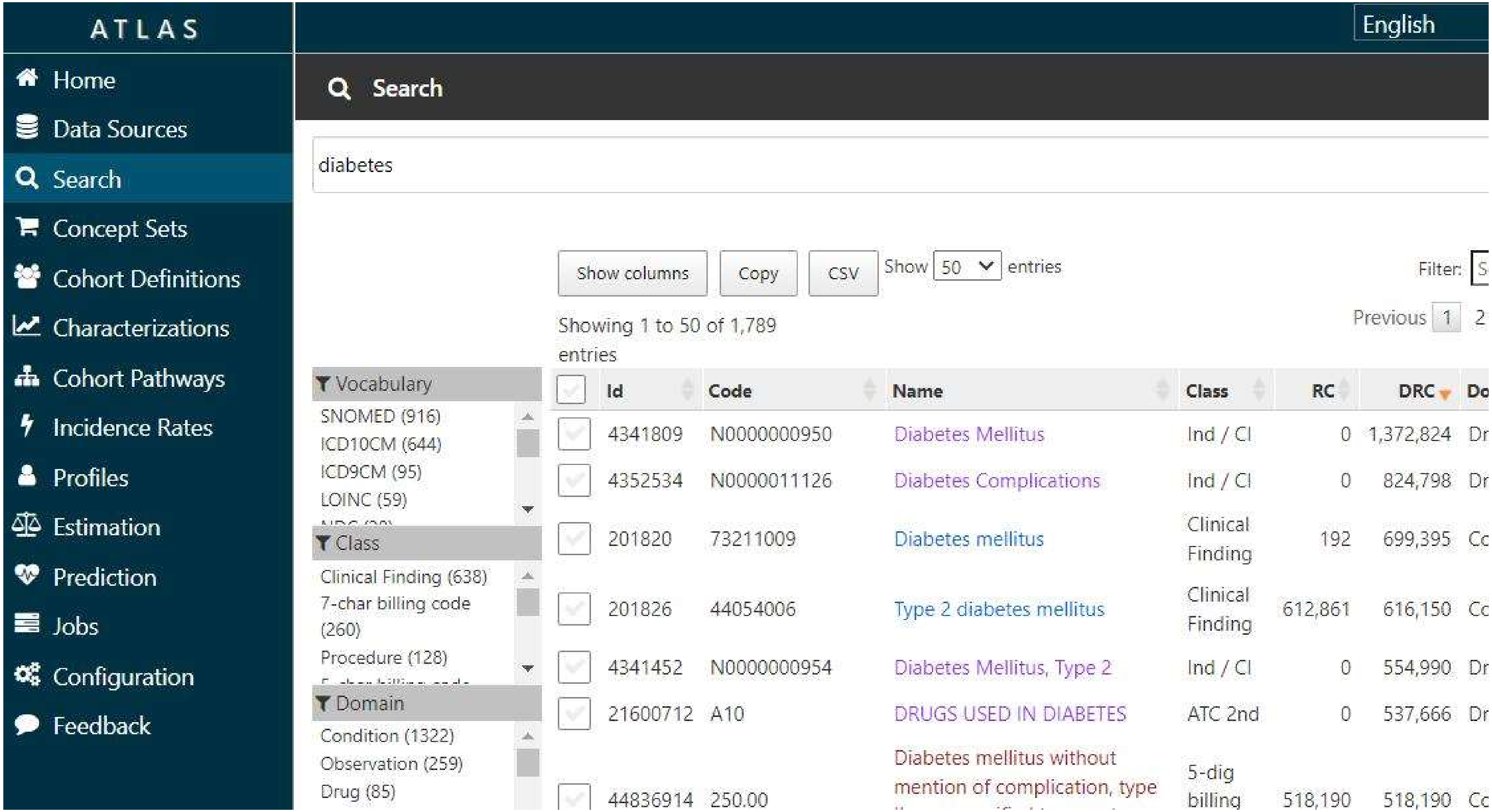
Measurement
　検査値

Observation
　(その他の)得られたデータ

Procedure
　処置、手術

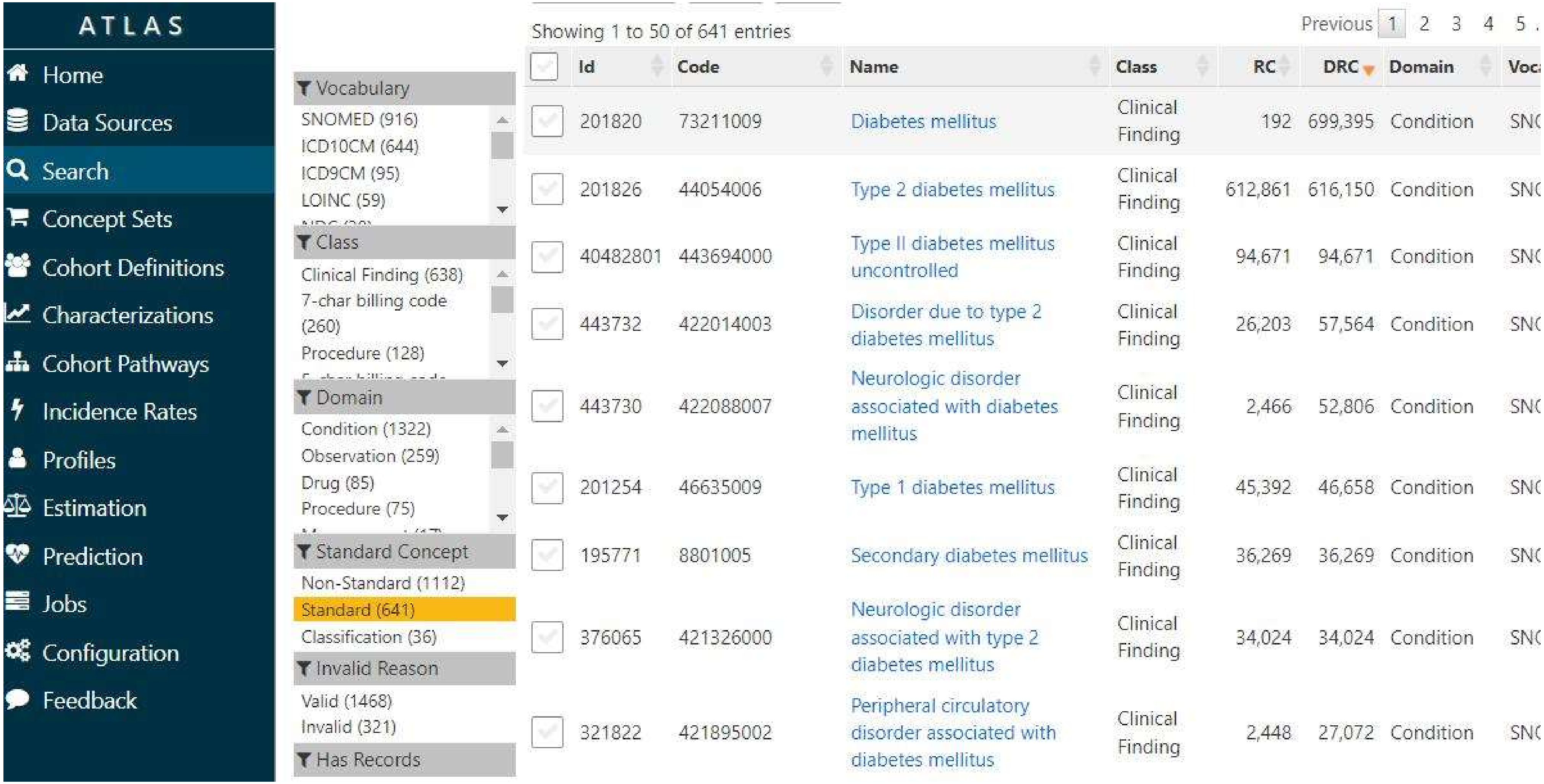
Device
　医療機器使用

慣らし運転：コンセプト検索例　　“diabetes”



“metformin”なども試してみましょう。

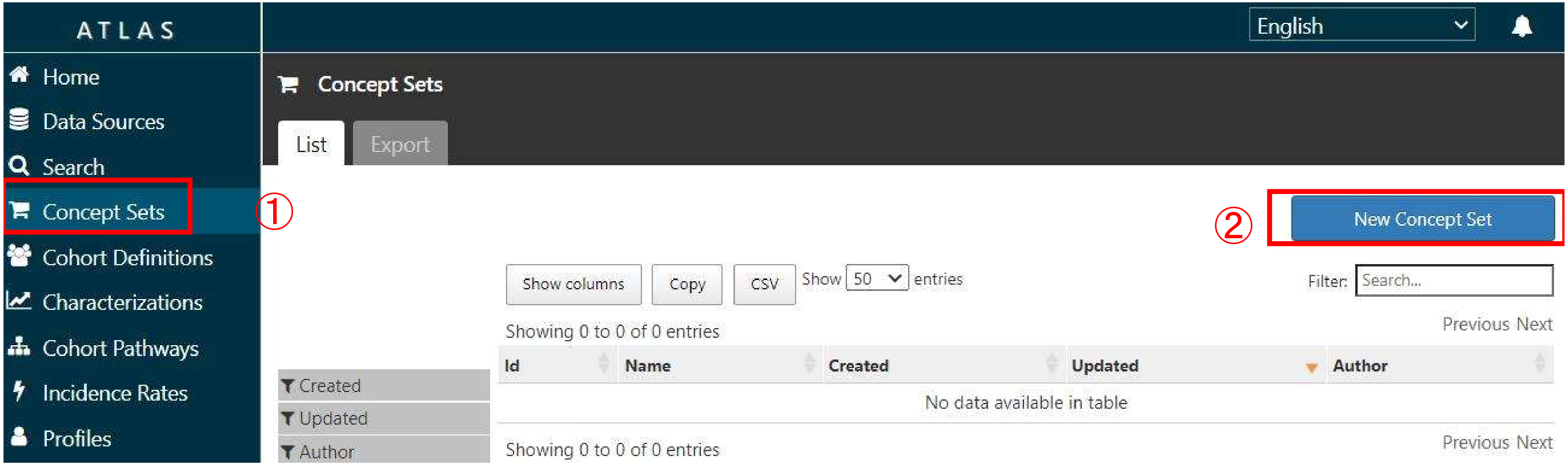
慣らし運転："Standard"をクリックする

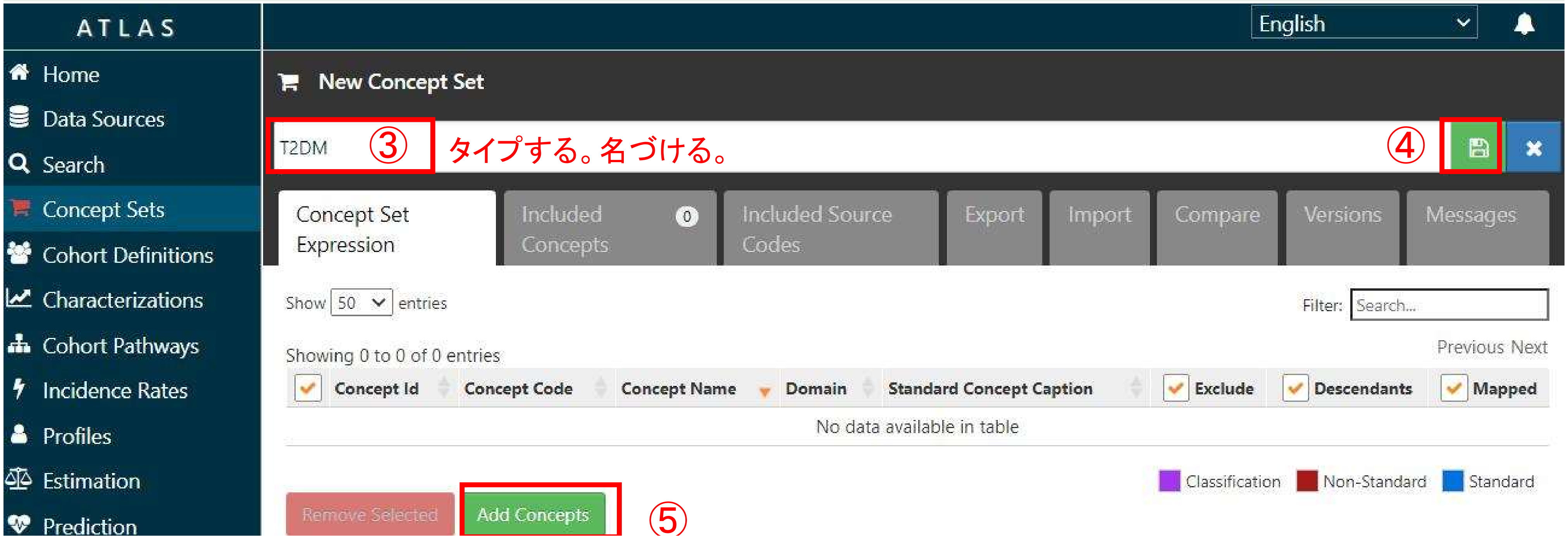


"metformin"なども検索してみましょう。

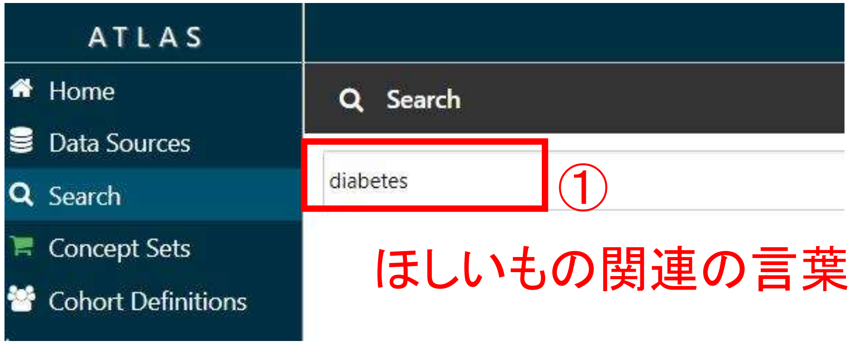
ステップ1：Concept Setを作る

T2DMの患者を示すもの





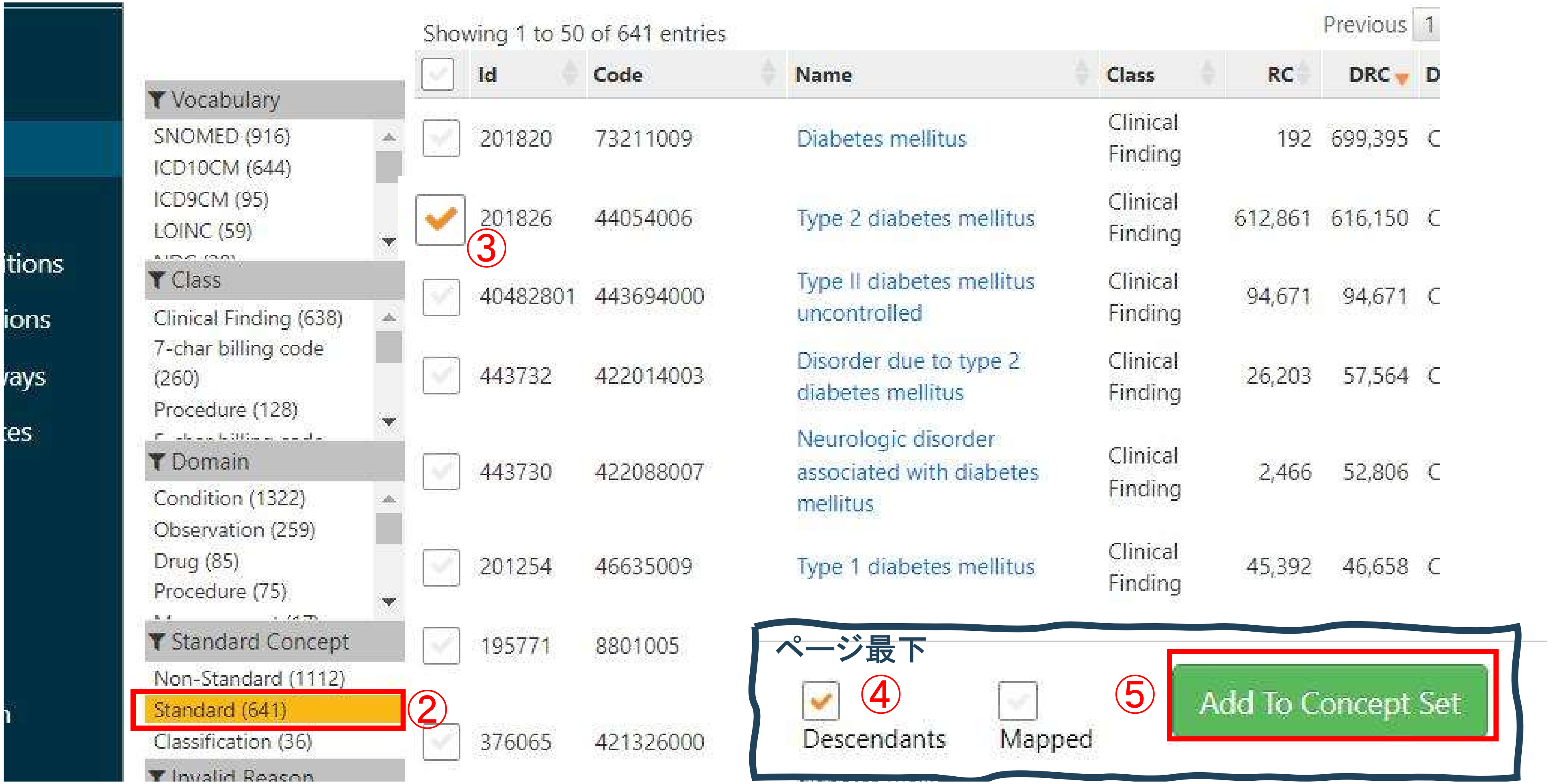
ステップ1：Concept Setを作る：conceptを追加する



①

緑になってる

ほしいもの関連の言葉を入れて検索



②

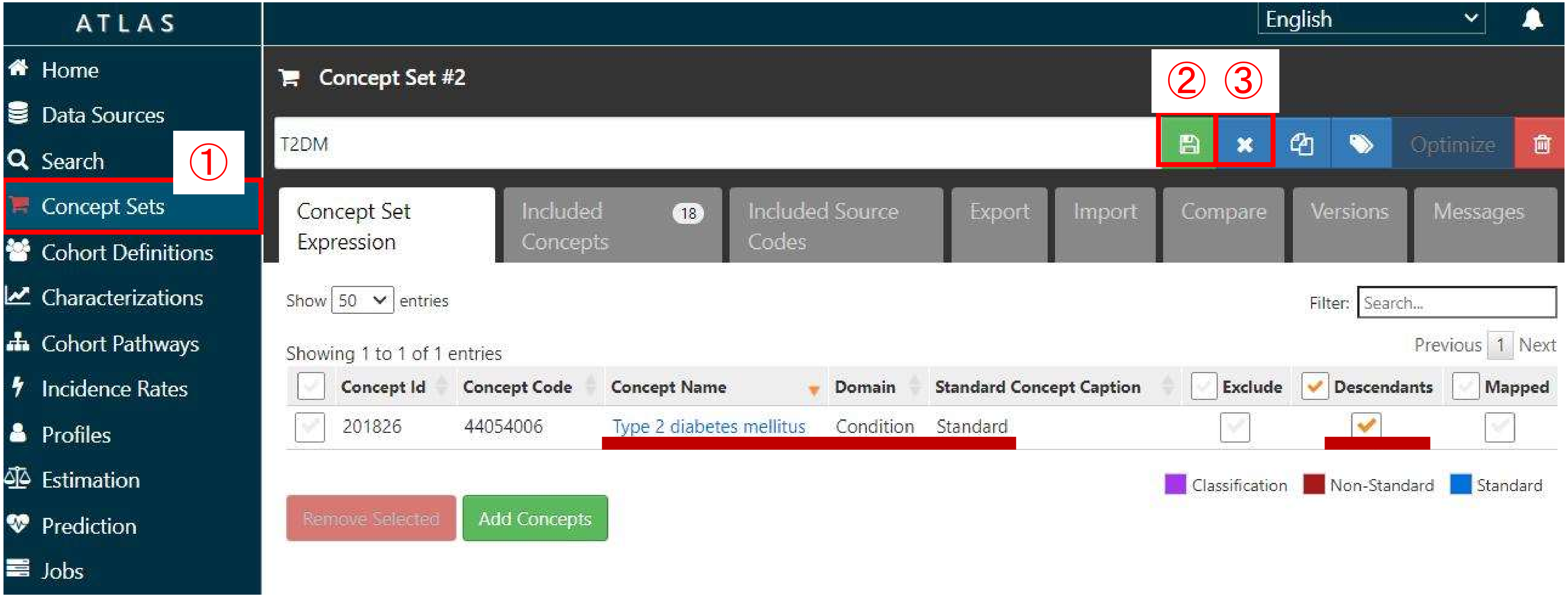
③

ページ最下

④

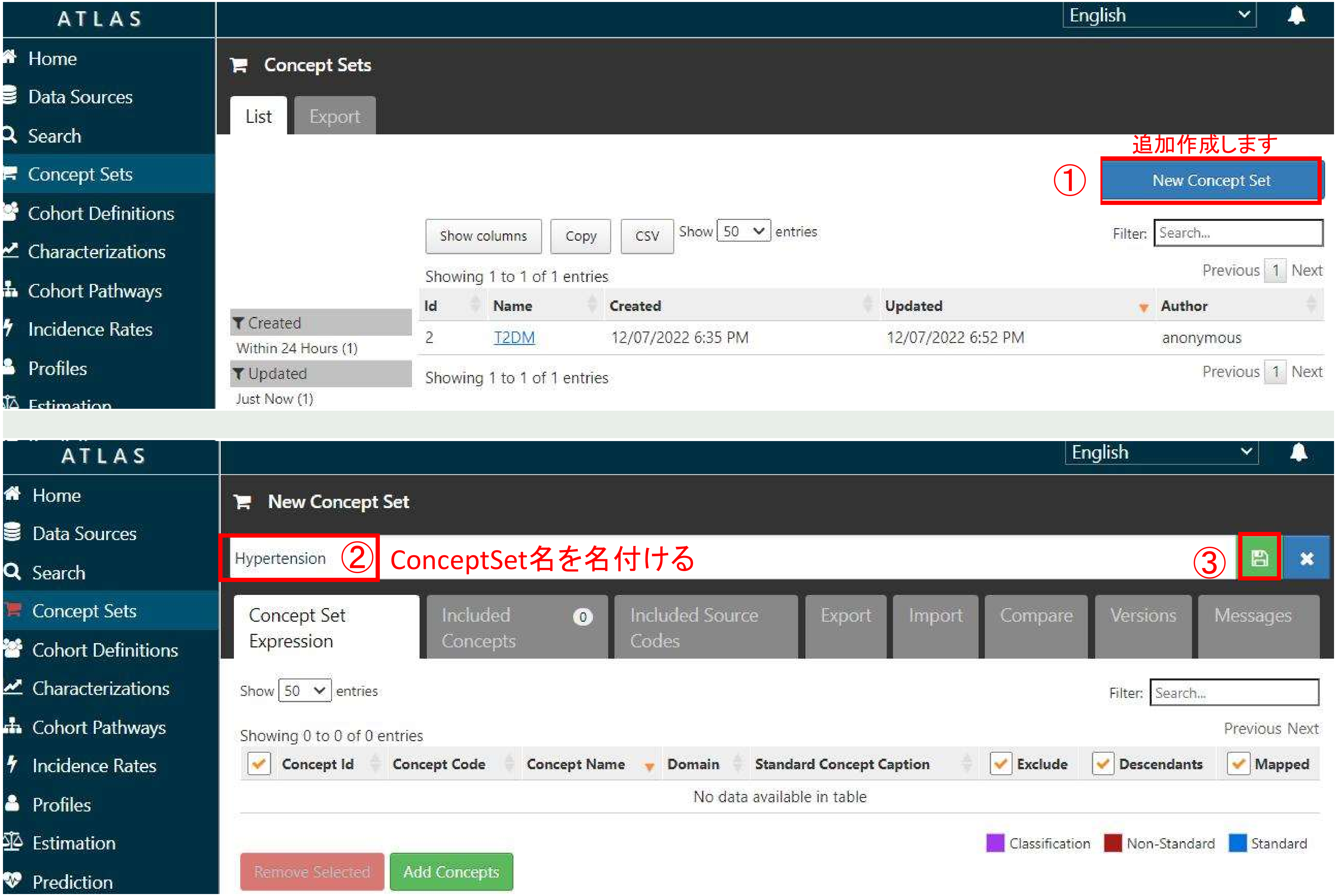
⑤

ステップ1：Concept Setを作る：確認



上記のようになっているか確認して、
②、③をクリック

ステップ1：Concept Setを作る-その２　　hypertension

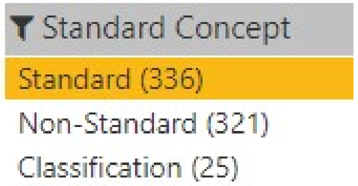


ステップ1：Concept Setを作る-その2

同様に"hypertension"で繰り返す。

1. Add Conceptsする

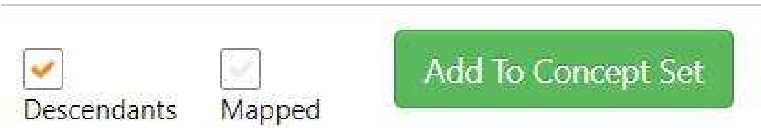
2. "hypertension"を検索



3. Standardを選ぶ

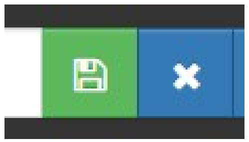
4. Malignant essential hypertensionを選ぶ

5. Descendantsにチェックして、Add To Concept Setする

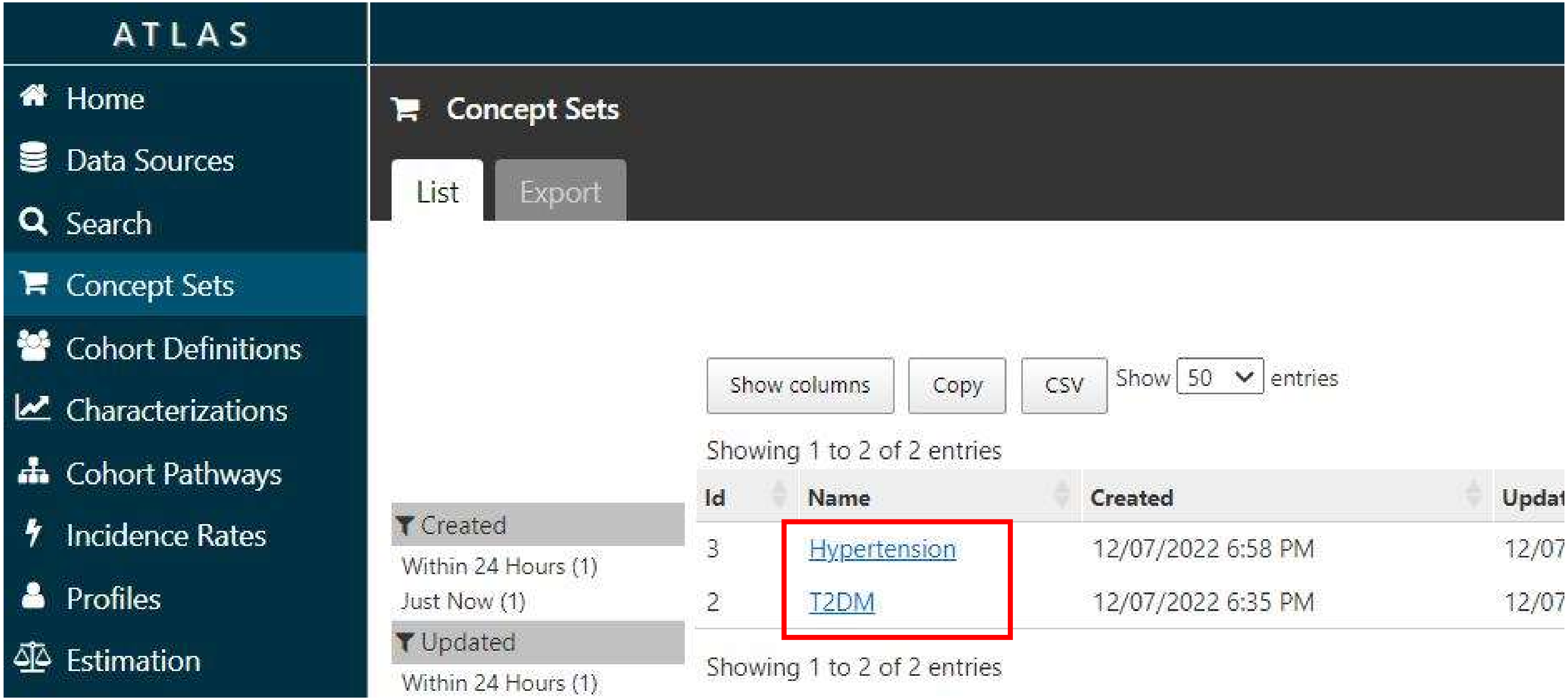


6. メインメニュー"Concept Sets"

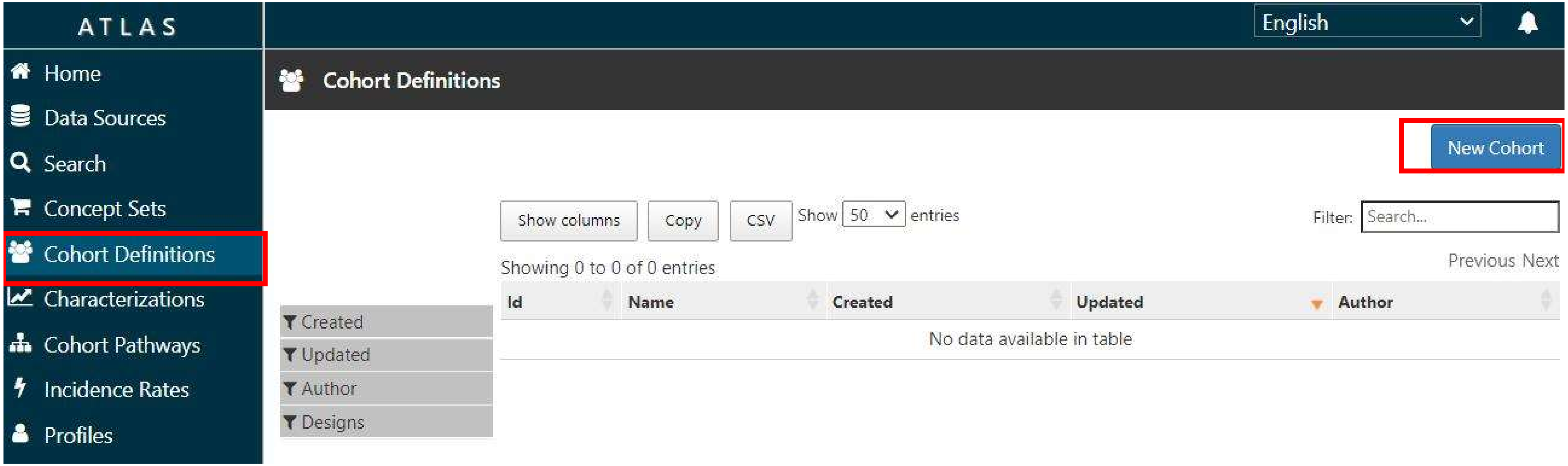
7. 保存して、Close

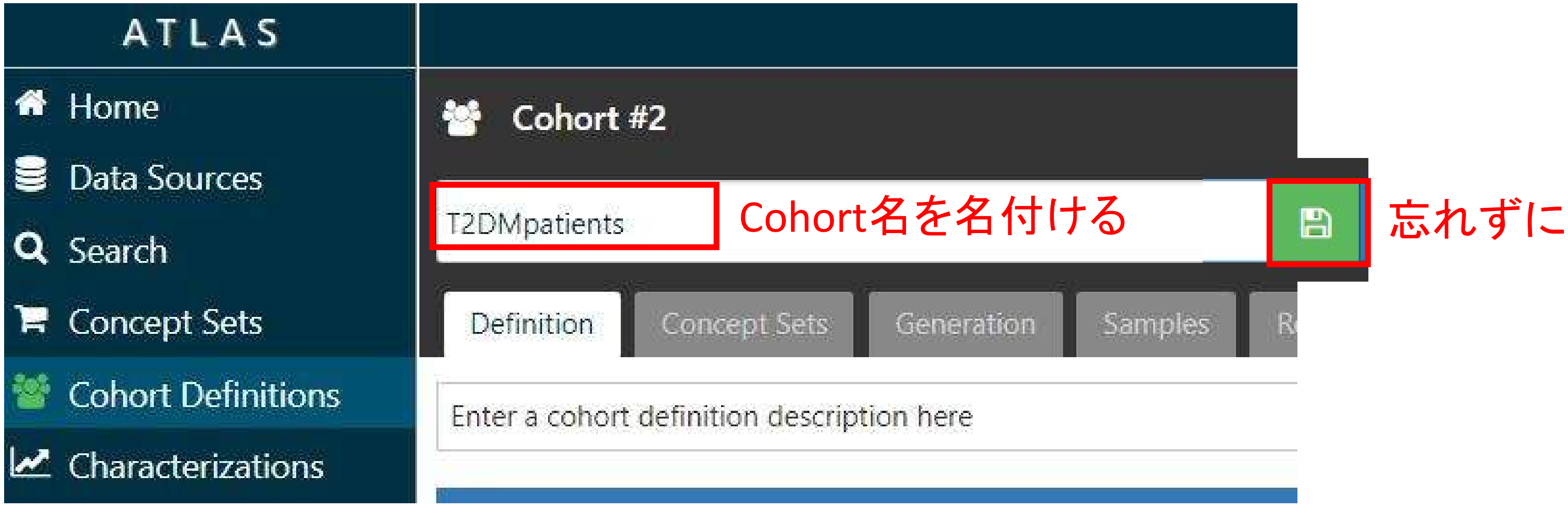


ステップ1：完了
２つのConceptSetができた！



ステップ２：Cohortを作る





ATLASで
簡易分析

Characterization　特徴分析

- T2DM-chara
  Cohort definition: T2DM-cohort
  Feature analyses: Demographics Index Year, etc.

他のcohortについてもやってみましょう
Hypertension-cohort
ACEI-cohort

Chort Pathways 患者群推移

- DMdrugPath

  Target Cohorts: T2DM-cohort

  Event Cohorts: Metformin-cohort, SUagent-cohort

Hypertensionについてもやってみましょう

Incident Rates　罹患率

試しに

Target Cohorts: ACEI-cohort

Outcome Cohorts: AMI-cohort

他の組み合わせもやってみましょう

Profiles 個々の患者情報

DataSource: SynPUF5
PatientID=1
PatientID=100

適宜PatientIDを入れてみましょう

ATLASで
分析設計

複雑な時間がかかる分析は2段階で

分析設計
ツール

ATLAS

ATLAS

JSON
ファイル

設計
パラメータ

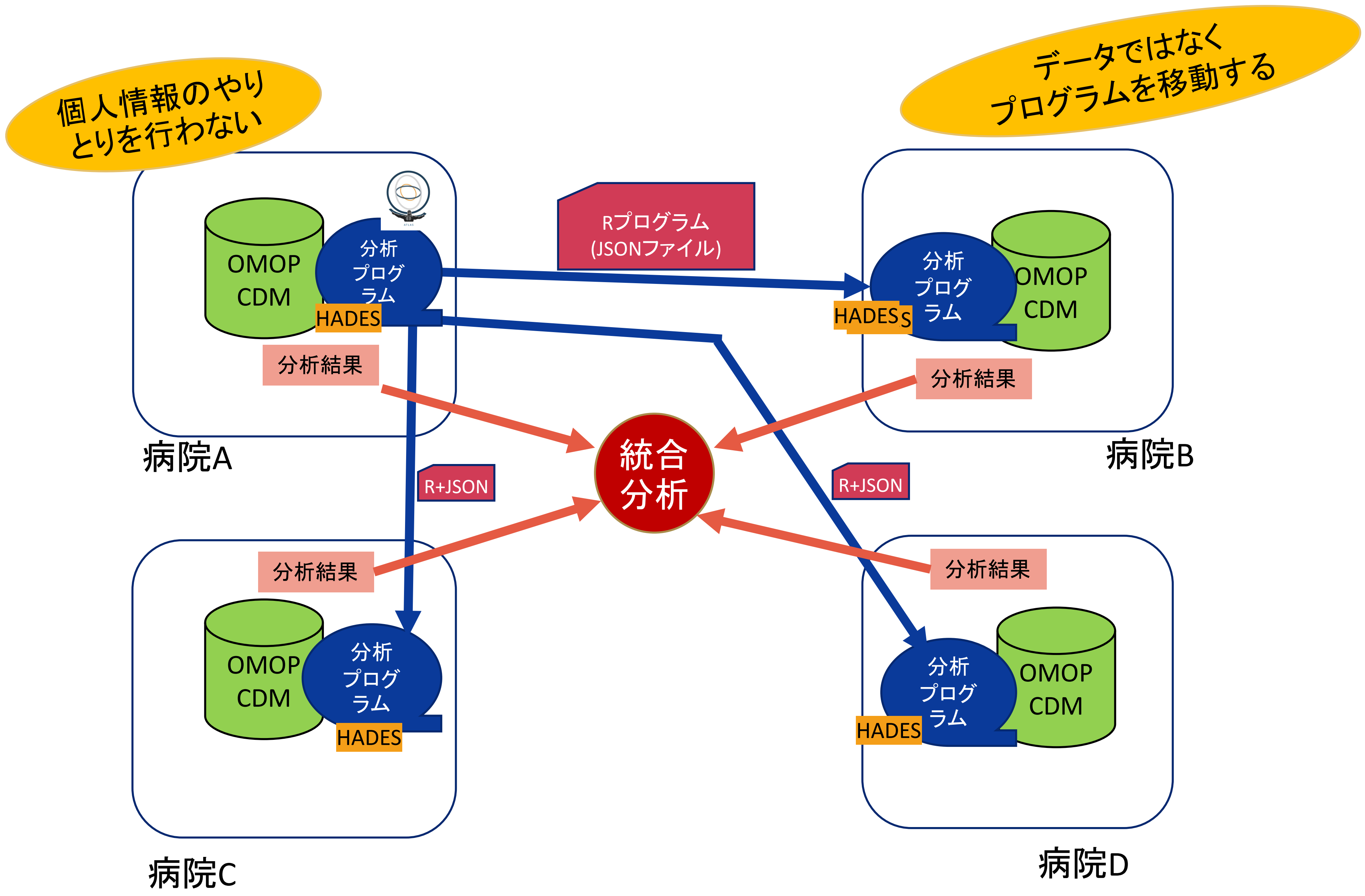
分析実行
ライブラリ

HADES
HEALTH ANALYTICS DATA-TO-EVIDENCE SUITE

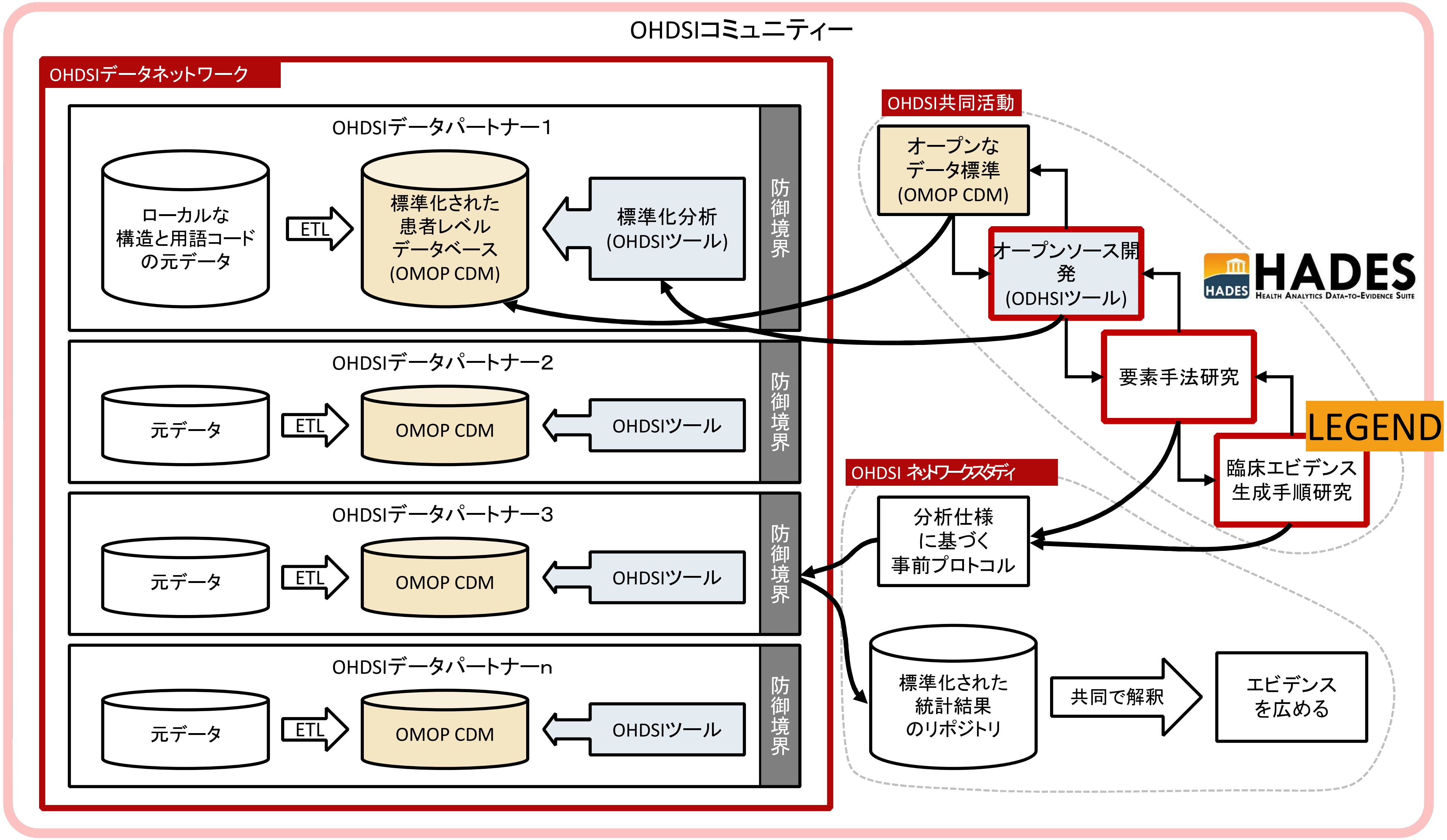
R Studio

HADES

連携分析 Federation Analysis



OHDSIが行っていること



HADES

HEALTH ANALYTICS DATA-TO-EVIDENCE SUITE

Population-level estimation

CohortMethod
New-user cohort studies using large-scale regression for propensity and outcome models.
Learn more...

SelfControlledCaseSeries
Self-Controlled Case Series analysis using few or many predictors, includes splines for age and seasonality.
Learn more...

SelfControlledCohort
A self-controlled cohort design, where time preceding exposure is used as control.
Learn more...

EvidenceSynthesis
Routines for combining causal effect estimates and study diagnostics across multiple data sites in a distributed study.
Learn more...

Patient-level prediction

PatientLevelPrediction
Build and evaluate predictive models for user-specified outcomes, using a wide array of machine learning algorithms.
Learn more...

DeepPatientLevelPrediction
Performing patient level prediction using deep learning
Learn more...

EnsemblePatientLevelPrediction
Building and validating ensemble patient-level predictive models.
Learn more...

HADES

HEALTH ANALYTICS DATA-TO-EVIDENCE SUITE

Cohort construction and evaluation

Capr
Develop and manipulate complex cohort definitions in R
Learn more...

CirceR
An R wrapper for Circe, a library for creating cohort definitions, expressing them as JSON, SQL, or Markdown.
Learn more...

CohortGenerator
Instantiating cohorts in a database based on a set of cohort definitions.
Learn more...

PhenotypeLibrary
The OHDSI Phenotype Library: a collection of community-maintained pre-defined cohorts.
Learn more...

CohortDiagnostics
Generate a wide set of diagnostics to evaluate cohort definitions against databases in the CDM.
Learn more...

PheValuator
Semi-automated evaluation of cohorts, producing metrics suchs as sensitivity, specificity, and positive and negative preditive value.
Learn more...

CohortExplorer
Visually explore all individual-level data of patients in a cohort
Learn more...

Evidence Quality

EmpiricalCalibration
Use negative control exposure-outcome pairs to profile and calibrate a particular analysis design.
Learn more...

MethodEvaluation
Use real data and established reference sets as well as simulations injected in real data to evaluate the performance of methods.
Learn more...

Supporting packages

Andromeda
Storing very large data objects on a local drive, while still making it possible to manipulate the data in an efficient manner.
Learn more...

BigKnn
A large scale k-nearest neighbor classifier using the Lucene search engine.
Learn more...

Cyclops
Highly efficient implementation of regularized logistic, Poisson and Cox regression.
Learn more...

DatabaseConnector
Connect directly to a wide range of database platforms, including SQL Server, Oracle, and PostgreSQL.
Learn more...

Eunomia
A standard CDM dataset for testing and demonstration purposes that runs on an embedded SQLite database.
Learn more...

FeatureExtraction
Automatically extract large sets of features for user-specified cohorts using data in the CDM.
Learn more...

Hydra
Hydrating package skeletons into executable R study packages based on specifications in JSON format.
Learn more...

IterativeHardThresholding
Performing L0-based regressions using Cyclops
Learn more...

OhdsiSharing
Securely sharing (large) files between OHDSI collaborators.
Learn more...

ParallelLogger
Support for parallel computation with logging to console, disk, or e-mail.
Learn more...

ROhdsiWebApi
Interact with OHDSI WebAPI web services.
Learn more...

SqlRender
Generate SQL on the fly for the various SQL dialects.
Learn more...

LEGEND in Principle

1. LEGEND will generate evidence at a large scale.
2. Dissemination of the evidence will not depend on the estimated effects.
3. LEGEND will generate evidence using a prespecified analysis design.
4. LEGEND will generate evidence by consistently applying a systematic process across all research questions.
5. LEGEND will generate evidence using best practices.
6. LEGEND will include empirical evaluation through the use of control questions.
7. LEGEND will generate evidence using open-source software that is freely available to all.
8. LEGEND will not be used to evaluate new methods.
9. LEGEND will generate evidence across a network of multiple databases.
10. LEGEND will maintain data confidentiality; patient-level data will not be shared between sites in the network.

Aim

Avoids publication bias, Enhances transparency, Avoids P hacking, Minimizes bias, Allows replication, Improves interpretability, Enhances generalizability, Uncovers potential between-site heterogeneity, Privacy.

エビデンスの信頼性

Desired attribute	Question	Researcher	Data	Analysis		Result
Repeatable	Identical	Identical	Identical	Identical	=	Identical
Reproducible	Identical	Different	Identical	Identical	=	Identical
Replicable	Identical	Same or different	Similar	Identical	=	Similar
Generalizable	Identical	Same or different	Different	Identical	=	Similar
Robust	Identical	Same or different	Same or different	Different	=	Similar
Calibrated	Similar (controls)	Identical	Identical	Identical	=	Statistically consistent

Figure 14.1, The Book of OHDSI

NegativeControlのConcept Setを作りましょう

Concept Set

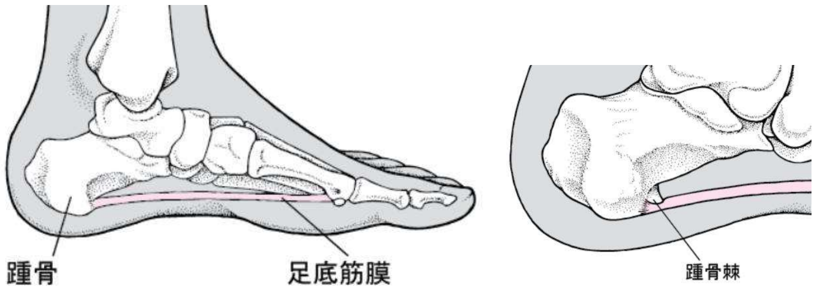
名称：NegControl

下層は含まない

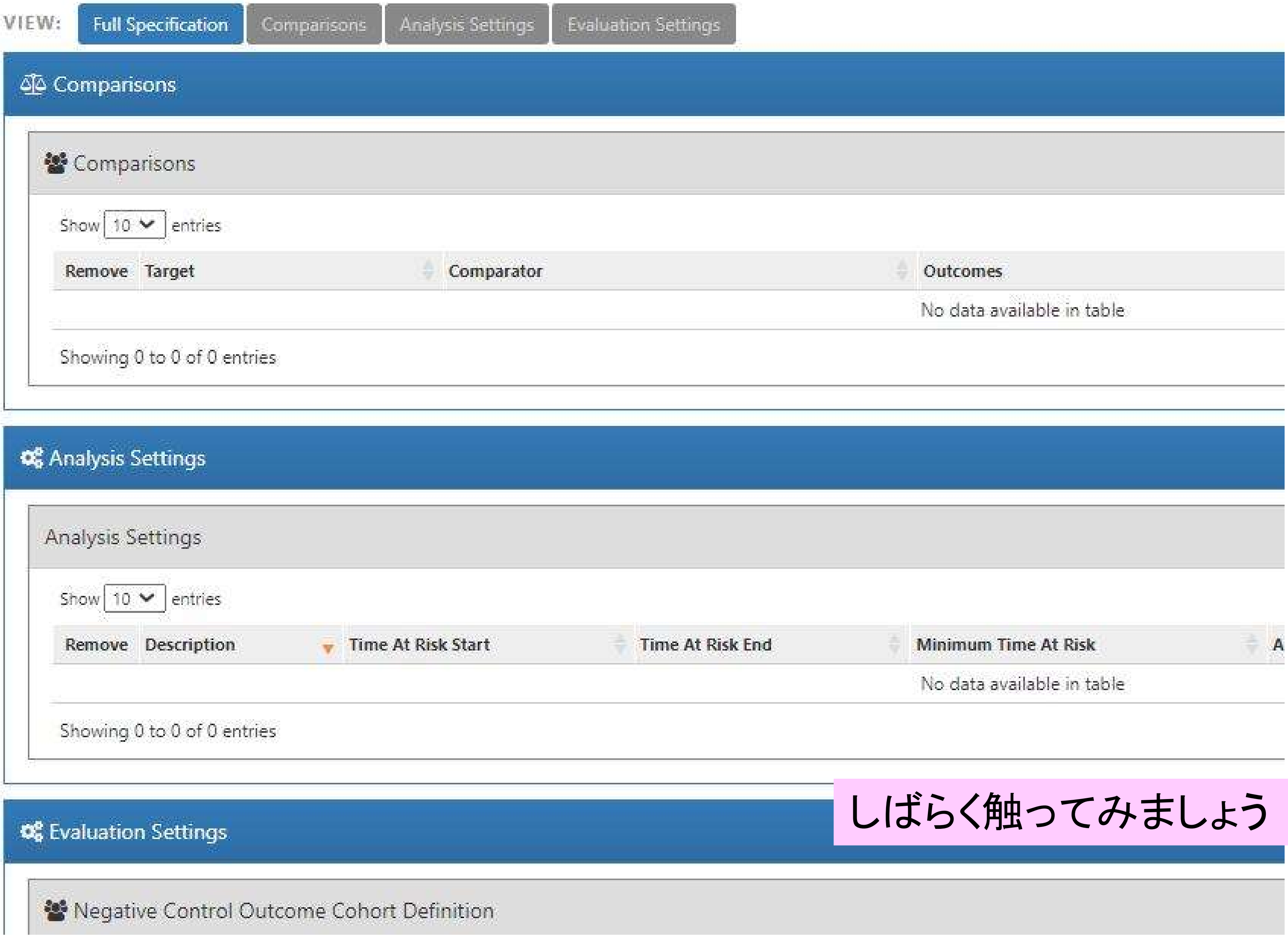
Showing 1 to 5 of 5 entries

	Concept Id	Concept Code	Concept Name	Domain	Standard Conce
	72748	74779009	Strain of rotator cuff capsule	Condition	Standard
	76786	63643000	Derangement of knee	Condition	Standard
	73560	55260003	Calcaneal spur	Condition	Standard
	73241	197210001	Anal and rectal polyp	Condition	Standard
	75911	65358001	Acquired hallux valgus	Condition	Standard

回旋筋腱板（肩関節包）の歪み
膝関節症
踵骨棘（しょうこつきょく）
肛門直腸ポリープ
後天性外反母趾

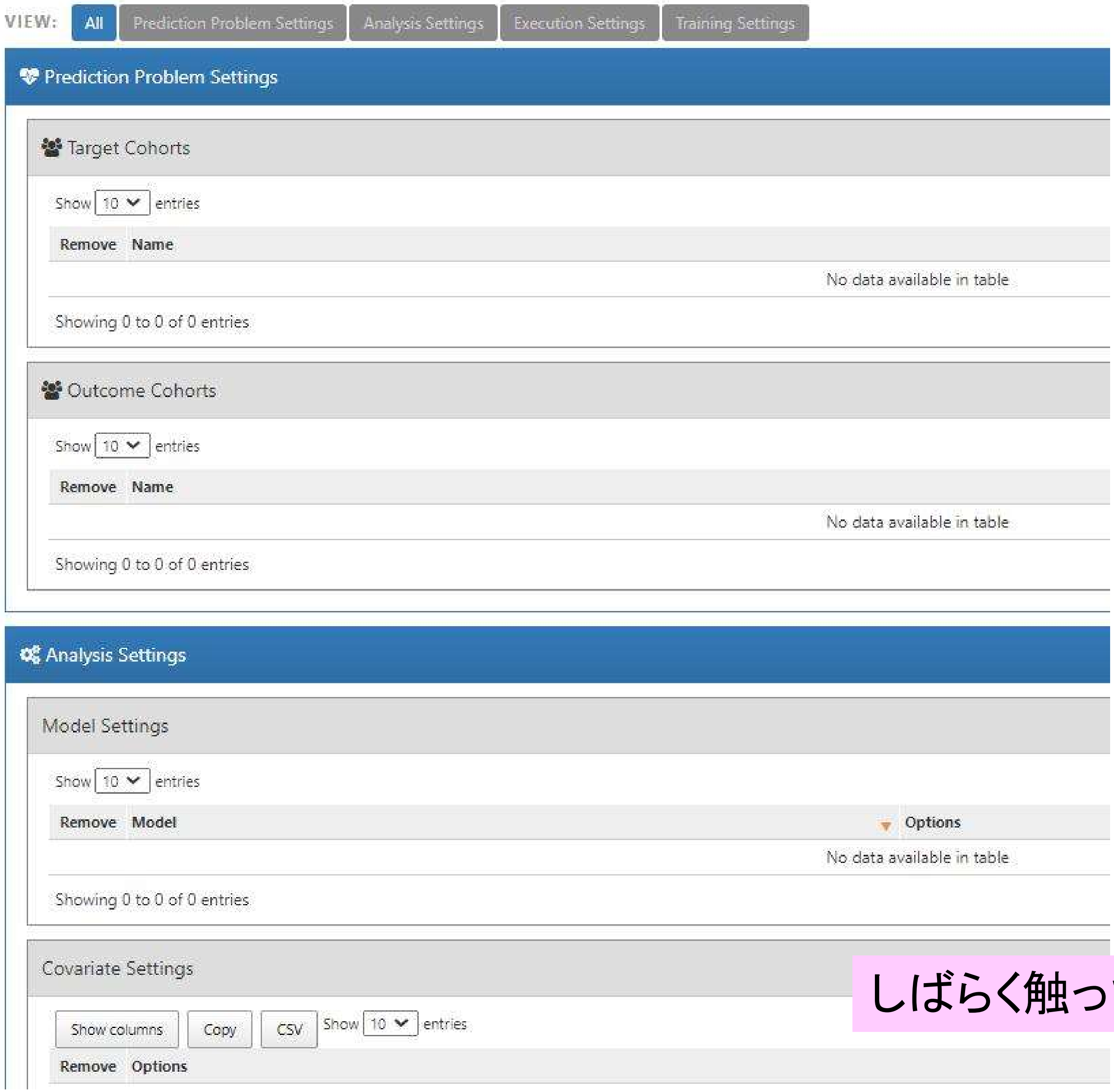


Estimation (Population Level Estimation)



しばらく触ってみましょう

Prediction (Patient Level Prediction)



しばらく触ってみましょう

さいごに

学習サイトの紹介

https://academy.ehden.eu/

EHDEN Academy
EUROPEAN HEALTH DATA & EVIDENCE NETWORK

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

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

Skill
Open Science & FAIR Principles

Skill
Introduction to Data Quality
Introduction to data quality and data quality dashboards.

Skill
Phenotype Definition, Characterisation...
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

Introduction to Usagi & Code Mappings for an...

Introduction to the Usagi tool, importing codes, review and output the co...

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 Re...

Three of EHDEN's Data Partners and SMEs: insights into COVID-19 E...

Tool

ATLAS

Design and execute analyses on observational data.

Methods

Health Technology Assessment

Non-professional

Patient Organisations: Introduction to Re...

A basic modular course on real world health data & research for the public.