

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 #30

イブニング カンファレンス(第30回)

2022.5.30



# 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：4月250本→5月262本



社会的決定要因を組み込むことで、  
予定外再入院予測の精度が向上した

➤ J Am Med Inform Assoc. 2022 May 2;ocac060. doi: 10.1093/jamia/ocac060. Online ahead of print.

## Inclusion of social determinants of health improves sepsis readmission prediction models

Fatemeh Amrollahi <sup>1</sup>, Supreeth P Shashikumar <sup>1</sup>, Angela Meier <sup>2</sup>, Lucila Ohno-Machado <sup>1</sup>,  
Shamim Nemati <sup>1</sup>, Gabriel Wardi <sup>2 3</sup>

Affiliations — collapse

### Affiliations

- 1 Division of Biomedical Informatics, University of California San Diego, San Diego, California, USA.
- 2 Division of Pulmonary, Critical Care and Sleep Medicine, University of California San Diego, San Diego, California, USA.
- 3 Department of Emergency Medicine, University of California San Diego, San Diego, California, USA.

PMID: 35511233 DOI: 10.1093/jamia/ocac060

## Abstract

**Objective:** Sepsis has a high rate of 30-day unplanned readmissions. Predictive modeling has been suggested as a tool to identify high-risk patients. However, existing sepsis readmission models have low predictive value and most predictive factors in such models are not actionable.

**Materials and methods:** Data from patients enrolled in the AllofUs Research Program cohort from 35 hospitals were used to develop a multicenter validated sepsis-related unplanned readmission model that incorporates clinical and social determinants of health (SDH) to predict 30-day unplanned readmissions. Sepsis cases were identified using concepts represented in the Observational Medical Outcomes Partnership. The dataset included over 60 clinical/laboratory features and over 100 SDH features.

**Results:** Incorporation of SDH factors into our model of clinical and demographic features improves model area under the receiver operating characteristic curve (AUC) significantly (from 0.75 to 0.80;  $P < .001$ ). Model-agnostic interpretability techniques revealed demographics, economic stability, and delay in getting medical care as important SDH predictive features of unplanned hospital readmissions.

**Discussion:** This work represents one of the largest studies of sepsis readmissions using objective clinical data to date (8935 septic index encounters). SDH are important to determine which sepsis patients are more likely to have an unplanned 30-day readmission. The AllofUS dataset provides granular data from a diverse set of individuals, making this model potentially more generalizable than prior models.

**Conclusion:** Use of SDH improves predictive performance of a model to identify which sepsis patients are at high risk of an unplanned 30-day readmission.

## 概要

**目的:** 敗血症は 30 日間の予定外再入院率が高い。高リスクの患者を特定するためのツールとして予測モデルが提案されている。しかし、既存の敗血症再入院モデルは予測値が低く、そのようなモデルにおける予測因子の多くは実用的でない。

**材料と方法:** 35病院のAllofUs Research Programコホートに登録された患者のデータを用いて、30日の計画外再入院を予測するために、臨床的および社会的決定要因（SDH）を組み込んだ多施設で検証済みの敗血症関連計画外再入院モデルを開発した。敗血症の症例は、OMOPで表される概念を用いて同定された。データセットには60以上の臨床的/実験的特徴と100以上のSDH特徴が含まれていた。

**結果:** SDH因子を臨床的・人口統計的特徴のモデルに組み込むことで、モデルの受信者動作特性曲線下面積（ROC-AUC）が有意に向上した（0.75から0.80に。 $P < 0.001$ ）。モデル診断的解釈可能技法により、人口統計学、経済的安定性、診療の遅れが計画外再入院の重要なSDH予測特徴であることが明らかになった。

**考察:** 本研究は、客観的な臨床データを用いた敗血症の再入院に関するこれまでで最大規模の研究である（8935人の敗血症の受診）。SDHは、どの敗血症患者が予定外の30日再入院をする可能性が高いかを判断するのに重要である。AllofUSデータセットは、多様な個人からのきめ細かいデータを提供しており、このモデルは先行モデルよりも一般化できる可能性がある。

**結論:** SDHの使用により、どの敗血症患者が予定外の30日後再入院のリスクが高いかを特定するモデルの予測性能が向上した。



## NSAID使用はCOVID重症化と関連しない (N3C研究)

Multicenter Study > Virol J. 2022 May 15;19(1):84. doi: 10.1186/s12985-022-01813-2.

### NSAID use and clinical outcomes in COVID-19 patients: a 38-center retrospective cohort study

Justin T Reese<sup>1</sup>, Ben Coleman<sup>2 3</sup>, Lauren Chan<sup>4</sup>, Hannah Blau<sup>2</sup>, Tiffany J Callahan<sup>5 6</sup>, Luca Cappelletti<sup>7</sup>, Tommaso Fontana<sup>7</sup>, Katie R Bradwell<sup>8</sup>, Nomi L Harris<sup>9</sup>, Elena Casiraghi<sup>7 10</sup>, Giorgio Valentini<sup>7 10</sup>, Guy Karlebach<sup>2</sup>, Rachel Deer<sup>11</sup>, Julie A McMurry<sup>6</sup>, Melissa A Haendel<sup>6</sup>, Christopher G Chute<sup>12</sup>, Emily Pfaff<sup>13</sup>, Richard Moffitt<sup>14</sup>, Heidi Spratt<sup>11</sup>, Jasvinder A Singh<sup>15 16</sup>, Christopher J Mungall<sup>9</sup>, Andrew E Williams<sup>17 18 19</sup>, Peter N Robinson<sup>20 21</sup>

Affiliations + expand

PMID: 35570298 PMCID: PMC9107579 DOI: 10.1186/s12985-022-01813-2

## Abstract

**Background:** Non-steroidal anti-inflammatory drugs (NSAIDs) are commonly used to reduce pain, fever, and inflammation but have been associated with complications in community-acquired pneumonia. Observations shortly after the start of the COVID-19 pandemic in 2020 suggested that ibuprofen was associated with an increased risk of adverse events in COVID-19 patients, but subsequent observational studies failed to demonstrate increased risk and in one case showed reduced risk associated with NSAID use.

**Methods:** A 38-center retrospective cohort study was performed that leveraged the harmonized, high-granularity electronic health record data of the National COVID Cohort Collaborative. A propensity-matched cohort of 19,746 COVID-19 inpatients was constructed by matching cases (treated with NSAIDs at the time of admission) and 19,746 controls (not treated) from 857,061 patients with COVID-19 available for analysis. The primary outcome of interest was COVID-19 severity in hospitalized patients, which was classified as: moderate, severe, or mortality/hospice. Secondary outcomes were acute kidney injury (AKI), extracorporeal membrane oxygenation (ECMO), invasive ventilation, and all-cause mortality at any time following COVID-19 diagnosis.

**Results:** Logistic regression showed that NSAID use was not associated with increased COVID-19 severity (OR: 0.57 95% CI: 0.53-0.61). Analysis of secondary outcomes using logistic regression showed that NSAID use was not associated with increased risk of all-cause mortality (OR 0.51 95% CI: 0.47-0.56), invasive ventilation (OR: 0.59 95% CI: 0.55-0.64), AKI (OR: 0.67 95% CI: 0.63-0.72), or ECMO (OR: 0.51 95% CI: 0.36-0.7). In contrast, the odds ratios indicate reduced risk of these outcomes, but our quantitative bias analysis showed E-values of between 1.9 and 3.3 for these associations, indicating that comparatively weak or moderate confounder associations could explain away the observed associations.

**Conclusions:** Study interpretation is limited by the observational design. Recording of NSAID use may have been incomplete. Our study demonstrates that NSAID use is not associated with increased COVID-19 severity, all-cause mortality, invasive ventilation, AKI, or ECMO in COVID-19 inpatients. A conservative interpretation in light of the quantitative bias analysis is that there is no evidence that NSAID use is associated with risk of increased severity or the other measured outcomes. Our results confirm and extend analogous findings in previous observational studies using a large cohort of patients drawn from 38 centers in a nationally representative multicenter database.

## 概要

**背景:** 非ステロイド性抗炎症薬（NSAIDs）は、痛み、発熱、炎症を抑えるために一般的に使用されているが、市中肺炎の合併症との関連が指摘されている。2020年のCOVID-19パンデミック開始直後の観察では、イブプロフェンがCOVID-19患者の有害事象のリスク上昇と関連することが示唆されたが、その後の観察研究ではリスク上昇は示されず、NSAID使用によるリスク低下が示されたケースもあった。

**方法:** 38施設のレトロスペクティブコホート研究を実施し、National COVID Cohort Collaborative (N3C)の整合性のとれた、粒度の高い電子カルテデータを活用した。解析可能なCOVID-19患者857,061例から、ケース（入院時にNSAIDsによる治療を受けた）と対照19,746例（治療を受けていない）をマッチングし、傾向をマッチさせたCOVID-19入院患者のコホートを構築した。主要評価項目は、入院患者におけるCOVID-19の重症度で、中等度、重度、死亡／ホスピスに分類された。副次的アウトカムは、COVID-19診断後の任意の時点における急性腎障害（AKI）、体外式膜酸素化（ECMO）、人工呼吸器装着、全死因死亡率であった。

**結果:** ロジスティック回帰分析の結果、NSAIDの使用はCOVID-19の重症度上昇と関連しなかった（OR:0.57 95%CI:0.53-0.61）。ロジスティック回帰を用いた副次的アウトカムの解析では、NSAIDの使用は全死亡（OR:0.51 95%CI:0.47-0.56）、侵襲的換気（OR:0.59 95%CI:0.55-0.64）、AKI（OR:0.67 95%CI:0.63-0.72）またはECMO（OR:0.51 95%CI:0.36-0.7）リスクの上昇と関連がないことが明らかにされた。一方、オッズ比はこれらの転帰のリスク低下を示しているが、定量的バイアス分析では、これらの関連についてE値が1.9～3.3となり、比較的弱いまたは中程度の交絡因子でこの関連を説明できそうなことが示された。

**結論:** 研究の解釈は、観察研究デザインによって制限される。NSAIDsの使用に関する記録は不完全であった可能性がある。本研究は、NSAIDの使用がCOVID-19入院患者のCOVID-19重症度、全死亡、人工呼吸器装着、AKI、ECMOの増加と関連しないことを実証している。定量的バイアス分析に照らして保守的に解釈すると、NSAIDの使用が重症度や他の測定結果の増加リスクと関連するという証拠はない、ということである。今回の結果は、全国を代表する多施設データベースの38施設から抽出された大規模な患者コホートを用いた過去の観察研究での類似の知見を確認し、さらに拡張したものである。





甲状腺疾患におけるLT3療法の長期安全性の検討

> [Thyroid](#). 2022 May 16. doi: 10.1089/thy.2021.0634. Online ahead of print.

**Heart failure and stroke risks in users of liothyronine with or without levothyroxine compared to levothyroxine alone: A propensity score matched analysis**

Wook Yi <sup>1</sup>, Bo Hyun Kim <sup>2 3</sup>, Mijin Kim <sup>4 5</sup>, Jinmi Kim <sup>6</sup>, Myungsoo Im <sup>7</sup>, Soree Ryang <sup>8</sup>, Eun Heui Kim <sup>9</sup>, Yun Kyung Jeon <sup>10</sup>, Sang Soo Kim <sup>11</sup>, In Joo Kim <sup>12</sup>

Affiliations + expand

PMID: 35570696 DOI: 10.1089/thy.2021.0634

## Abstract

**Background:** Combination therapy with liothyronine (LT3) and levothyroxine (LT4) is used in patients with persistent symptoms despite being administered an adequate dose of LT4. LT3 may also be used in some thyroid cancer patients preparing for radioactive iodine therapy. However, there is a controversy regarding the safety of LT3 use, and there has been no definite evidence of long-term safety of LT3 therapy in Asian populations. The aim of this study was to examine the long-term safety of LT3 therapy using the Common Data Model (CDM).

**Methods:** We conducted a retrospective multicenter study across four hospital databases encoded in the Observational Medical Outcomes Partnership (OMOP) CDM. LT3 users were defined as those who received an LT3 prescription for at least 90 days (with or without LT4), and their safety outcomes were compared with those in LT4-only users after 1:4 propensity score matching. Safety outcomes included the incidences of osteoporosis, cardiovascular disease, cancer, anxiety disorder, and mood disorder.

**Results:** We identified 1,434 LT3 users and 3,908 LT4-only users. There was a statistically significant difference in the incidence rate of safety outcomes between LT3 users and LT4-only users. The risks of heart failure (incidence rate ratio [IRR] = 1.664, 95% confidence interval [CI] 1.002-2.764,  $p = 0.049$ ) and stroke (IRR = 1.757, 95% CI 1.073-2.877,  $p = 0.025$ ) were higher in LT3 users than in LT4-only users. When subgroup analysis was performed according to the presence/absence of thyroid cancer history and duration of thyroid hormone replacement, the risk of heart failure was higher in LT3 users with a history of thyroid cancer and those who underwent  $\geq 52$  weeks of LT3 therapy. Additionally, the risk of stroke was higher in LT3 users without thyroid cancer history and those who underwent  $\geq 52$  weeks of LT3 therapy.

**Conclusions:** The use of LT3 was associated with increased incidence of heart failure and stroke in patients with a longer duration of LT3 use and history of thyroid cancer. Therefore, clinicians should consider the risk of heart failure and stroke in thyroid cancer patients with long-term use of LT3. These findings require confirmation in other populations.

## 概要

**背景:** リオチロニン（LT3）とレボチロキシン（LT4）の併用療法は、十分な量のLT4を投与しても症状が持続する患者に用いられる。また、放射性ヨウ素治療の準備をしている甲状腺がん患者にもLT3が使用されることがある。しかし、LT3使用の安全性については議論があり、アジア人集団におけるLT3治療の長期安全性に関する明確なエビデンスは得られていない。本研究の目的は、CDMを用いてLT3療法の長期安全性を検討することである。

**方法:** 我々は、OMOP CDMにコード化された4つの病院データベースにおいて、多施設共同レトロスペクティブ研究を実施した。LT3使用者は、90日以上LT3処方（LT4併用または非併用）を受けた者と定義し、1：4の傾向スコアマッチング後、LT4のみの使用者と安全性転帰を比較した。安全性の評価項目は、骨粗鬆症、心血管疾患、癌、不安障害、気分障害の発生率であった。

**結果:** LT3投与者1,434名、LT4投与者3,908名を確認した。LT3使用者とLT4のみの使用者の間で、安全性アウトカムの発生率に統計的に有意な差があった。心不全（発生率比 [IRR] = 1.664、95%信頼区間 [CI] 1.002-2.764、 $p = 0.049$ ）および脳卒中（IRR = 1.757、95%CI 1.073-2.877、 $p = 0.025$ ）のリスクはLT4のみのユーザーよりLT3ユーザーが高いことが示された。甲状腺がん歴の有無、甲状腺ホルモン補充期間によってサブグループ解析を行ったところ、甲状腺がん歴のあるLT3使用者とLT3療法を52週間以上行った使用者で心不全リスクが高くなった。また、脳卒中のリスクは、甲状腺癌の既往がないLT3使用者と52週以上のLT3療法を受けた患者で高かった。

**結論:** LT3使用期間が長く、甲状腺癌の既往がある患者では、LT3の使用は心不全および脳卒中の発生率の上昇と関連していた。したがって、臨床医は甲状腺癌患者におけるLT3の長期使用による心不全と脳卒中のリスクを考慮する必要がある。これらの知見は、他の集団での確認が必要である。





➤ JAMIA Open. 2022 Mar 16;5(1):ooac021. doi: 10.1093/jamiaopen/ooac021. eCollection 2022 Apr.

## Trajectories: a framework for detecting temporal clinical event sequences from health data standardized to the Observational Medical Outcomes Partnership (OMOP) Common Data Model

Kadri Künnapu<sup>1</sup>, Solomon Ioannou<sup>2</sup>, Kadri Ligi<sup>1 3</sup>, Raivo Kolde<sup>3</sup>, Sven Laur<sup>1 3</sup>,  
Jaak Vilo<sup>1 3 4</sup>, Peter R Rijnbeek<sup>2</sup>, Sulev Reisberg<sup>1 3 4</sup>

Affiliations

### Affiliations

- 1 STACC, Tartu, Estonia.
- 2 Department of Medical Informatics, Erasmus University Medical Center, Rotterdam, the Netherlands.
- 3 Institute of Computer Science, University of Tartu, Tartu, Estonia.
- 4 Quretec, Tartu, Estonia.

## Abstract

**Objective:** To develop a framework for identifying temporal clinical event trajectories from Observational Medical Outcomes Partnership-formatted observational healthcare data.

**Materials and methods:** A 4-step framework based on significant temporal event pair detection is described and implemented as an open-source R package. It is used on a population-based Estonian dataset to first replicate a large Danish population-based study and second, to conduct a disease trajectory detection study for type 2 diabetes patients in the Estonian and Dutch databases as an example.

**Results:** As a proof of concept, we apply the methods in the Estonian database and provide a detailed breakdown of our findings. All Estonian population-based event pairs are shown. We compare the event pairs identified from Estonia to Danish and Dutch data and discuss the causes of the differences. The overlap in the results was only 2.4%, which highlights the need for running similar studies in different populations.

**Conclusions:** For the first time, there is a complete software package for detecting disease trajectories in health data.

## 概要

**目的:** OMOP 形式の観察医療データから時間的な臨床イベントの軌跡を特定するための枠組みを開発すること。

**材料と方法:** 有意な時間的イベントのペア検出に基づく4ステップのフレームワークを説明し、オープンソースのRパッケージとして実装する。このフレームワークをエストニアの人口ベースのデータセットで使い、まずデンマークの大規模な人口ベースの研究を再現し、次にエストニアとオランダのデータベースを例として2型糖尿病患者の疾患軌道検出の研究を行った。

**結果:** 概念実証として、エストニアのデータベースで本手法を適用し、その結果の詳細な内訳を示す。エストニアの人口ベースのイベントペアはすべて表示されます。エストニアから特定されたイベントペアをデンマークやオランダのデータと比較し、その違いの原因について考察しています。結果の重複はわずか2.4%であり、異なる集団で同様の研究を実施する必要性が浮き彫りになった。

**結論:** 健康データにおける疾病の軌跡を検出するための完全なソフトウェアパッケージが初めて登場した。



➤ [Healthc Inform Res.](#) 2022 Apr;28(2):112-122. doi: 10.4258/hir.2022.28.2.112. Epub 2022 Apr 30.

## Applying the OMOP Common Data Model to Facilitate Benefit-Risk Assessments of Medicinal Products Using Real-World Data from Singapore and South Korea

Hui Xing Tan <sup>1</sup>, Desmond Chun Hwee Teo <sup>1</sup>, Dongyun Lee <sup>2</sup>, Chungsoo Kim <sup>3</sup>, Jing Wei Neo <sup>1</sup>, Cynthia Sung <sup>1 4</sup>, Haroun Chahed <sup>1</sup>, Pei San Ang <sup>1</sup>, Doreen Su Yin Tan <sup>5</sup>, Rae Woong Park <sup>2 3</sup>, Sreemanee Raaj Dorajoo <sup>1</sup>

Affiliations — collapse

### Affiliations

- 1 Vigilance & Compliance Branch, Health Products Regulation Group, Health Sciences Authority, Singapore.
- 2 Department of Biomedical Informatics, Ajou University School of Medicine, Suwon, Korea.
- 3 Department of Biomedical Sciences, Graduate School of Medicine, Ajou University, Suwon, Korea.
- 4 Health Services and Systems Research, Duke-NUS Medical School, Singapore.
- 5 Department of Pharmacy, Khoo Teck Puat Hospital, Singapore.

## Abstract

**Objectives:** The aim of this study was to characterize the benefits of converting Electronic Medical Records (EMRs) to a common data model (CDM) and to assess the potential of CDM-converted data to rapidly generate insights for benefit-risk assessments in post-market regulatory evaluation and decisions.

**Methods:** EMRs from January 2013 to December 2016 were mapped onto the Observational Medical Outcomes Partnership-CDM (OMOP-CDM) schema. Vocabulary mappings were applied to convert source data values into OMOP-CDM-endorsed terminologies. Existing analytic codes used in a prior OMOP-CDM drug utilization study were modified to conduct an illustrative analysis of oral anticoagulants used for atrial fibrillation in Singapore and South Korea, resembling a typical benefit-risk assessment. A novel visualization is proposed to represent the comparative effectiveness, safety and utilization of the drugs.

**Results:** Over 90% of records were mapped onto the OMOP-CDM. The CDM data structures and analytic code templates simplified the querying of data for the analysis. In total, 2,419 patients from Singapore and South Korea fulfilled the study criteria, the majority of whom were warfarin users. After 3 months of follow-up, differences in cumulative incidence of bleeding and thromboembolic events were observable via the proposed visualization, surfacing insights as to the agent of preference in a given clinical setting, which may meaningfully inform regulatory decision-making.

**Conclusions:** While the structure of the OMOP-CDM and its accessory tools facilitate real-world data analysis, extending them to fulfil regulatory analytic purposes in the post-market setting, such as benefit-risk assessments, may require layering on additional analytic tools and visualization techniques.

## 概要

**目的:** 電子カルテを共通データモデル (CDM) に変換することの利点を明らかにし、CDMに変換されたデータが市販後の薬事評価・決定におけるベネフィット・リスク評価のための洞察を迅速に生成する可能性を評価することである。

**方法:** 2013年1月から2016年12月までのEMRを OMOP-CDMスキーマにマッピングした。用語マッピングを適用し、ソースデータの値を OMOP-CDM 推奨の用語に変換した。OMOP-CDMの薬剤利用調査で使用された既存の分析コードを修正し、シンガポールと韓国で心房細動に使用される経口抗凝固剤について、典型的なベネフィット・リスク評価に類似した分析を実施した。薬剤の有効性、安全性、使用率の比較を表現するために、新しい視覚化を提案した。

**結果:** 90%以上のレコードがOMOP-CDMにマップされた。CDMのデータ構造と解析コードのテンプレートにより、解析のためのデータクエリーが簡素化された。シンガポールと韓国から合計2,419人の患者が研究基準を満たし、そのほとんどがワルファリン使用者であった。3ヶ月の追跡調査後、出血および血栓塞栓イベントの累積発生率の違いは、提案した可視化によって観察することができ、特定の臨床環境における優先薬剤に関する洞察を浮上させ、規制当局の意思決定に有意義な情報を提供することができる。

**結論:** OMOP-CDMの構造とその付属ツールは実世界のデータ解析を容易にするが、市販後における規制当局の解析目的、例えばベネフィット・リスク評価を満たすために拡張するには、解析ツールや可視化技術をさらに追加する必要があるかもしれない。





Review > J Hepatol. 2022 Jun;76(6):1318-1329. doi: 10.1016/j.jhep.2022.03.003.

## "Beyond MELD" – Emerging strategies and technologies for improving mortality prediction, organ allocation and outcomes in liver transplantation

Jin Ge <sup>1</sup>, W Ray Kim <sup>2</sup>, Jennifer C Lai <sup>1</sup>, Allison J Kwong <sup>3</sup>

Affiliations – collapse

### Affiliations

- 1 Division of Gastroenterology and Hepatology, Department of Medicine, University of California - San Francisco, San Francisco, CA, USA.
- 2 Division of Gastroenterology and Hepatology, Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA. Electronic address: wrkim@stanford.edu.
- 3 Division of Gastroenterology and Hepatology, Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA.

PMID: 35589253 DOI: 10.1016/j.jhep.2022.03.003

## Abstract

In this review article, we discuss the model for end-stage liver disease (MELD) score and its dual purpose in general and transplant hepatology. As the landscape of liver disease and transplantation has evolved considerably since the advent of the MELD score, we summarise emerging concepts, methodologies, and technologies that may improve mortality prognostication in the future. Finally, we explore how these novel concepts and technologies may be incorporated into clinical practice.

## 概要

この総説では、末期肝疾患モデル（MELD）スコアと、一般肝臓病学および移植肝臓病学におけるその2つの目的について考察している。MELDスコアの登場以来、肝疾患と移植の状況は大きく変化しており、将来的に死亡予後を改善する可能性のある新しい概念、方法論、技術について要約している。最後に、これらの新しい概念や技術をどのように臨床に取り入れることができるかを検討する。

肝移植待機リスト中の対象患者優先順位に、肝状態をあらわす「MELDスコア」が20年前から使われているが、もっと今風な色々ができるはずだという考察。

ひとつとして電カルデータからの機械学習でより正確で包括的かつリアルタイムの死亡予測ができる。N3Cを例に標準化され大規模に活用できるOMOPが引き合いに出されている。

MELDスコア＝

$$\begin{aligned} & 9.57 * \ln(\text{血清クレアチニン値 mg/dl}) \\ & + 3.78 * \ln(\text{血清総ビリルビン値 mg/dl}) \\ & + 11.20 * \ln(\text{PT-INR (血液凝固能)}) \\ & + 6.43 \end{aligned}$$

※小数点第1位を四捨五入した整数とする。





機械学習でlong COVID候補患者を  
抽出する。

➤ [Lancet Digit Health](#). 2022 May 16;S2589-7500(22)00048-6. doi: 10.1016/S2589-7500(22)00048-6.  
Online ahead of print.

## Identifying who has long COVID in the USA: a machine learning approach using N3C data

Emily R Pfaff <sup>1</sup>, Andrew T Girvin <sup>2</sup>, Tellen D Bennett <sup>3</sup>, Abhishek Bhatia <sup>4</sup>, Ian M Brooks <sup>5</sup>,  
Rachel R Deer <sup>6</sup>, Jonathan P Dekermanjian <sup>7</sup>, Sarah Elizabeth Jolley <sup>8</sup>, Michael G Kahn <sup>9</sup>,  
Kristin Kostka <sup>10</sup>, Julie A McMurry <sup>11</sup>, Richard Moffitt <sup>12</sup>, Anita Walden <sup>11</sup>, Christopher G Chute <sup>13</sup>,  
Melissa A Haendel <sup>11</sup>, N3C Consortium

Collaborators, Affiliations + expand

PMID: 35589549 PMCID: [PMC9110014](#) DOI: [10.1016/S2589-7500\(22\)00048-6](#)

## Abstract

**Background:** Post-acute sequelae of SARS-CoV-2 infection, known as long COVID, have severely affected recovery from the COVID-19 pandemic for patients and society alike. Long COVID is characterised by evolving, heterogeneous symptoms, making it challenging to derive an unambiguous definition. Studies of electronic health records are a crucial element of the US National Institutes of Health's RECOVER Initiative, which is addressing the urgent need to understand long COVID, identify treatments, and accurately identify who has it-the latter is the aim of this study.

**Methods:** Using the National COVID Cohort Collaborative's (N3C) electronic health record repository, we developed XGBoost machine learning models to identify potential patients with long COVID. We defined our base population ( $n=1\,793\,604$ ) as any non-deceased adult patient (age  $\geq 18$  years) with either an International Classification of Diseases-10-Clinical Modification COVID-19 diagnosis code (U07.1) from an inpatient or emergency visit, or a positive SARS-CoV-2 PCR or antigen test, and for whom at least 90 days have passed since COVID-19 index date. We examined demographics, health-care utilisation, diagnoses, and medications for 97 995 adults with COVID-19. We used data on these features and 597 patients from a long COVID clinic to train three machine learning models to identify potential long COVID among all patients with COVID-19, patients hospitalised with COVID-19, and patients who had COVID-19 but were not hospitalised. Feature importance was determined via Shapley values. We further validated the models on data from a fourth site.

**Findings:** Our models identified, with high accuracy, patients who potentially have long COVID, achieving areas under the receiver operator characteristic curve of 0.92 (all patients), 0.90 (hospitalised), and 0.85 (non-hospitalised). Important features, as defined by Shapley values, include rate of health-care utilisation, patient age, dyspnoea, and other diagnosis and medication information available within the electronic health record.

**Interpretation:** Patients identified by our models as potentially having long COVID can be interpreted as patients warranting care at a specialty clinic for long COVID, which is an essential proxy for long COVID diagnosis as its definition continues to evolve. We also achieve the urgent goal of identifying potential long COVID in patients for clinical trials. As more data sources are identified, our models can be retrained and tuned based on the needs of individual studies.

## 概要

**背景:** SARS-CoV-2感染による急性期以降の後遺症は、long COVIDとして知られ、COVID-19パンデミックからの回復に、患者や社会に深刻な影響を及ぼしている。long COVIDは、進化する不均一な症状によって特徴付けられ、明確な定義を導き出すことは困難である。電子カルテの研究は、米国国立衛生研究所のRECOVERイニシアチブの重要な要素であり、long COVIDの理解、治療法の特定、および患者の正確な特定が急務であるため、本研究では後者を目的とした。

**方法:** National COVID Cohort Collaborative (N3C) の電子カルテリポジトリを使用して、long COVIDを持つ潜在的な患者を特定するためにXGBoost機械学習モデルを開発した。基本集団 ( $n=1,793,604$ ) を、入院または救急受診によるICD-10-CMのCOVID-19診断コード (U07.1) 、またはSARS-CoV-2 PCRもしくは抗原検査が陽性で、COVID-19インデックス日から少なくとも90日経過した非死亡成人患者 (18歳以上) と定義した。COVID-19を発症した成人97,995人の人口統計学、医療利用、診断、投薬について検討した。これらの特徴に関するデータと、あるlong COVIDクリニックの患者597人を用いて、COVID-19の全患者、COVID-19で入院した患者、COVID-19だが入院していない患者からlong COVIDの可能性を特定する3つの機械学習モデルを学習させた。特徴の重要度はシャープレイ値によって決定された。さらに、4番目の施設のデータでモデルの検証を行った。

**結果:** 我々のモデルは、long COVIDの可能性のある患者を高い精度で同定し、ROC-AUCは0.92 (全患者) 、 0.90 (入院患者) 、 0.85 (非入院患者) であった。シャープレイ値で定義される重要な特徴には、ヘルスケア利用率、患者年齢、呼吸困難、電子カルテ内で利用可能な他の診断および投薬情報が含まれる。

**解釈:** 我々のモデルによってlong COVIDの可能性があると同定された患者は、long COVIDの専門クリニックでのケアが妥当な患者と解釈することができる。定義が進化し続けているため表現が困難なlong COVID診断に必須の代理ポイントとなる。また、臨床試験の対象となるlong COVID候補患者を特定するという緊急の目標も達成される。より多くのデータソースが特定されれば、我々のモデルは個々の研究のニーズに基づいて再トレーニングやチューニングを行うことができる。



# 最近のGlobal OHDSI話題から

## ■ 2022年のOHDSI Symposium開催予定

※今年からリアル開催に戻った

- ・ 6月 24-46日 [European Symposium](#)
- ・ 10月14-16日 [Global Symposium](#) 有料化された
- ・ APAC Symposiumは11月12-13日に台北で開催予定 有料

左の2つは  
登録受付中

## ■ OHDSI 1<sup>st</sup> DevCon April 22, 2022.

<https://www.ohdsi.org/devcon2022/>

### Workshops

8 am: ATLAS (**Anthony Sena**); HADES Introduction (**Adam Black**),

8:30 am: WebAPI (**Anthony Sena**); CohortDiagnostics (**James Gilbert**)

9 am: WhiteRabbit/Rabbit in a Hat (**Maxim Moinat**); Patient-Level Prediction (**Jenna Reys**)

9:30 am: Data Quality Dashboard (**Clair Blacketer**); Cyclops (**Marc Suchard**)

Then keynote talk, panel discussion, etc.



# 最近のGlobal OHDSI話題から

## 5月のCommunity Call

### ●APAC Community Call テーマ

May 5 Cohort Diagnostic Tool (Covid Phenotyping)

May 19 Mini Tutorial#1

OHDSI.org  
に録画あり

### ●Global Community Call テーマ

May 3 DARWIN EU Initiative

May 10 Mother's Day-themed conversations

- Clinical Research Opportunities: Brainstorm on maternal health.
- How can OHDSI support collaborator mothers to advance their personal and professional development?
- Recommended best practices for modeling pregnancy episodes and mother-child linkage.

May 17 OHDSI Debates

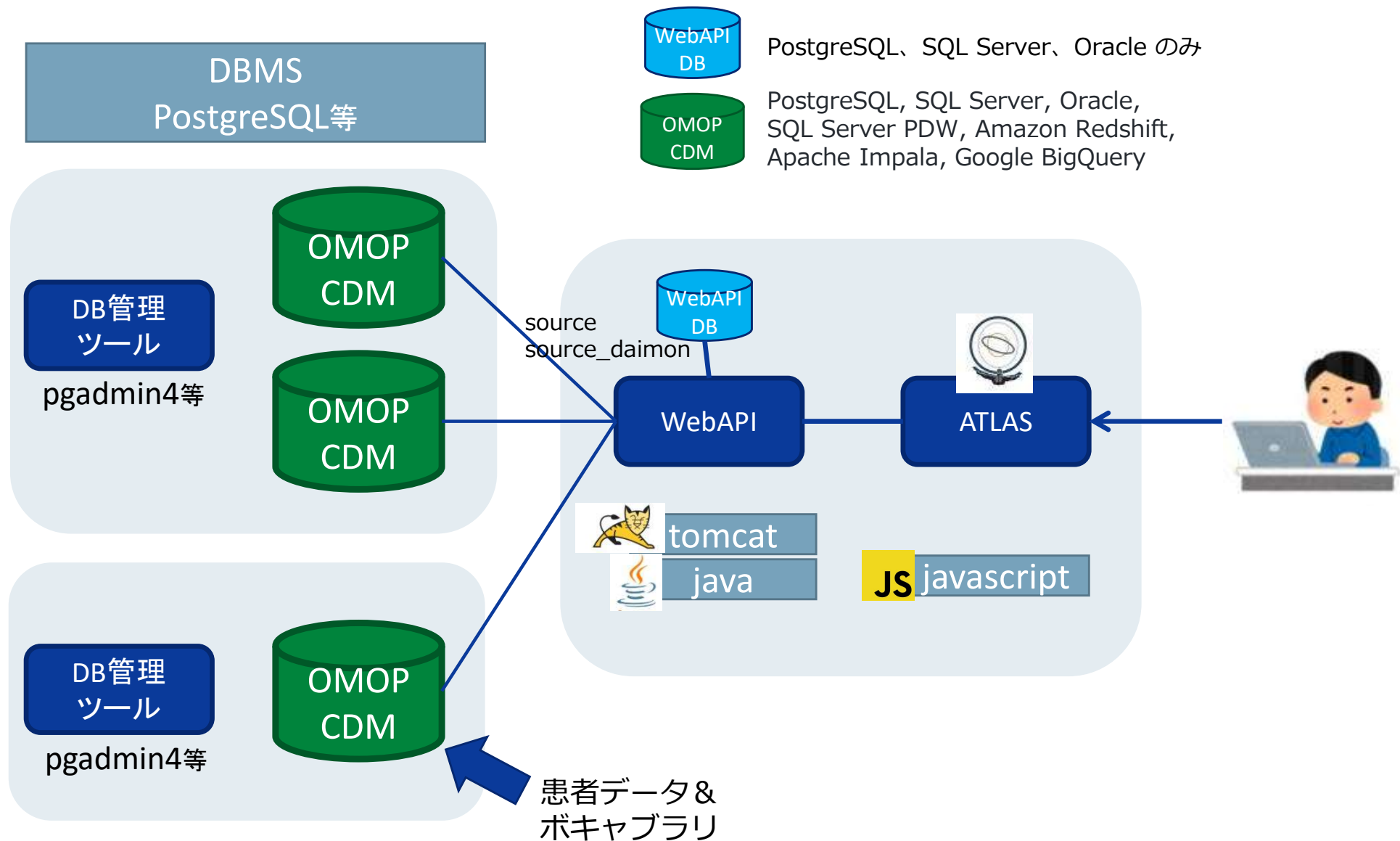
- Phenotype Development: One-size-fits-all vs. Tailored-per-databases
- Study Diagnostics: Nice-to-have vs. Essential requirements

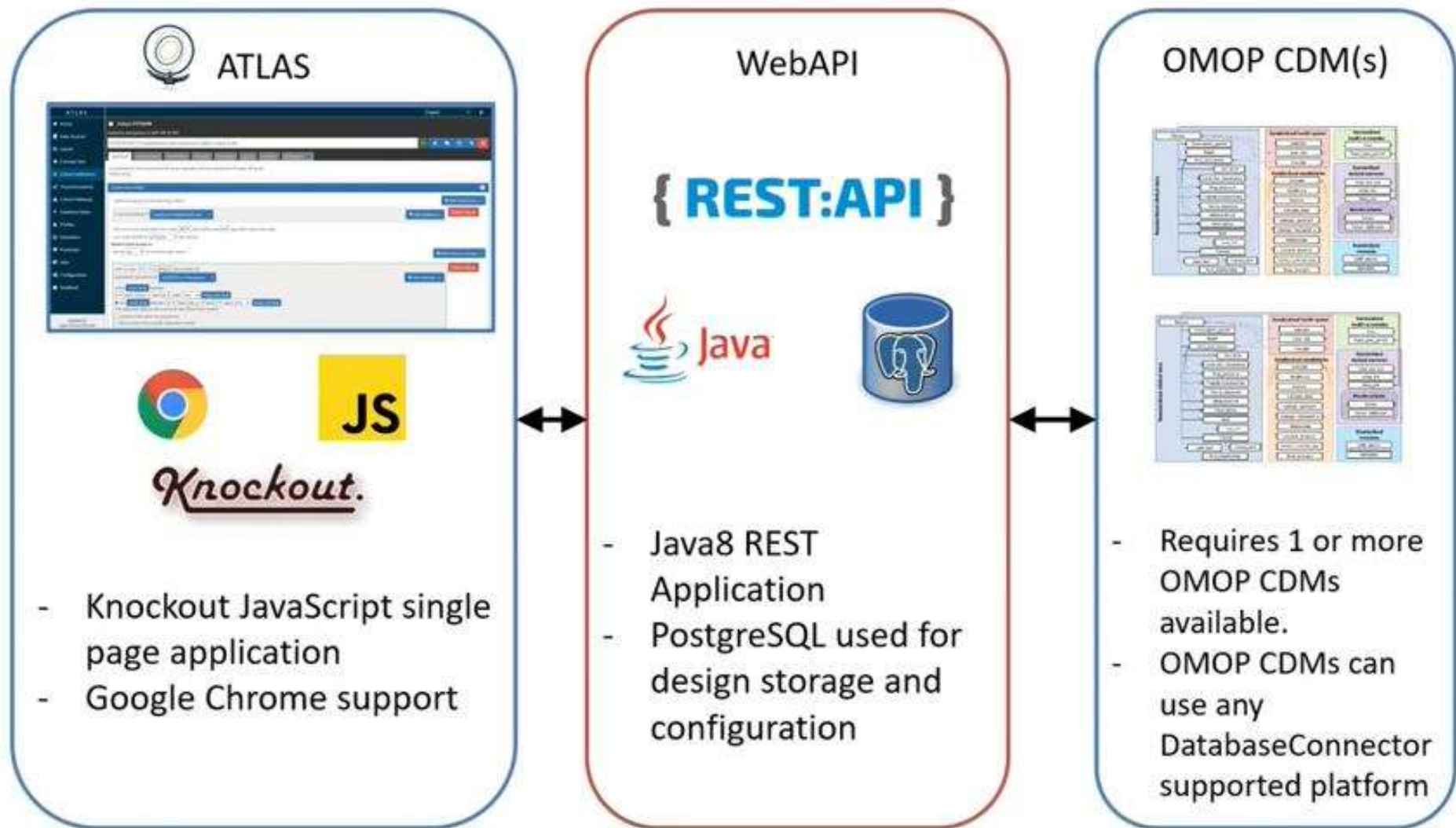
May 24 OHDSI Open Studies

- IBD characterization.
- Characterization of Health by OHDSI Asia-Pacific chapter to identify Temporal Effect of the Pandemic (CHAPTER) Study.
- Applying the Decentralized Generalized Linear Mixed Effects Model (dGEM) for Hospital Profiling of COVID-19 Mortality Data across OHDSI Network.
- Real world safety of treatments for multiple sclerosis.
- Comparison of mortality, morbidities & healthcare resources utilization between patients with and without a diagnosis of COVID-19.
- Quality assessment of CDM databases across the OHDSI-AP network.



# Inside OMOP #3 WebAPI/ATLAS









# ATLAS Technology

- Knockout JavaScript – Model-View-ViewModel (MVVM) Pattern
- Learning resources:
  - Knockout Tutorial: <http://learn.knockoutjs.com/>
  - What is MVVM: <https://www.youtube.com/watch?v=fo6rvTP9kkc>
  - JavaScript ES6: [https://www.w3schools.com/js/js\\_es6.asp](https://www.w3schools.com/js/js_es6.asp)
  - RequireJS: <https://requirejs.org/>



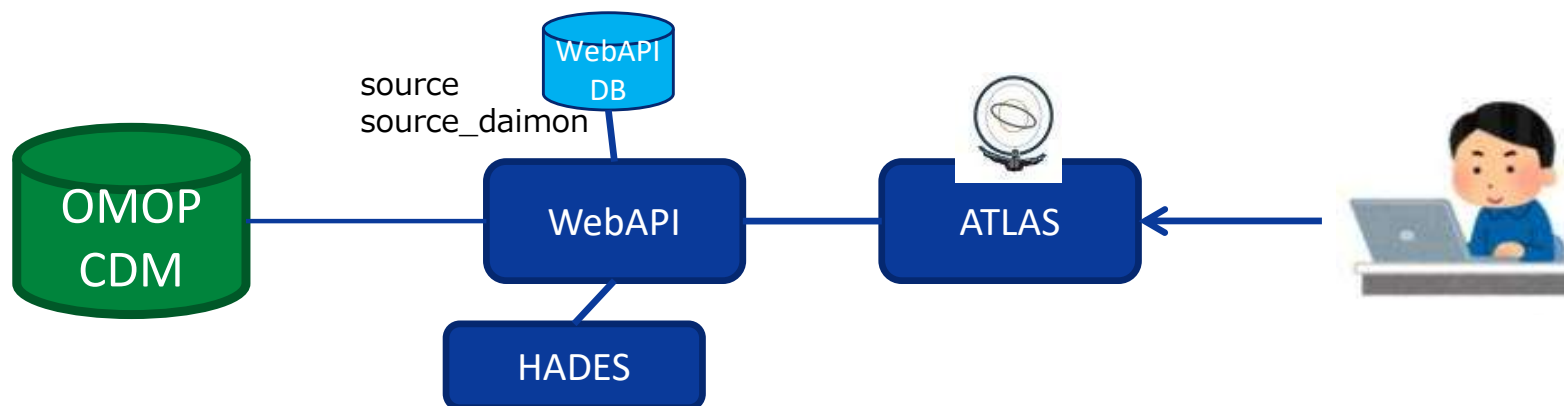
# WebAPI Technology

- Java8 Spring Application
- Learning resources:
  - Maven: <https://maven.apache.org/pom.html>
  - Spring: <https://spring.io/guides/gs/rest-service/>
  - Hibernate: <https://www.baeldung.com/spring-boot-hibernate>
  - Flyway: <https://flywaydb.org/documentation/tutorials/repeatable>



# WebAPI – Connection to HADES

- Clue: check WebAPI's pom.xml
- HADES Libraries
  - SqlRender – used to translate SQL from OHDSI-SQL to database specific dialect
  - FeatureExtraction – used by the characterization module to construct aggregate summary statistics for cohorts
  - Hydra – used by estimation/prediction to produce an R study package based on JSON input specification





# source, source\_daimonテーブル

※CDMセットが複数あるときは、  
sourceは複数行になり、  
source\_daimonはsource\_idで区別する。

## ■ source

カラム名	値
source_id	12
source_name	SynPUF 5% ←ATLASで表示される
source_key	SynPUF5 ←ATLASで使われる
source_connection	jdbc:postgresql://127.0.0.1:5432/cdmSynPUF5?user=xxxxx&password=xxxxx ←接続表現
source_dialect	postgresql ←DBMSの種類

## ■ source\_daimon

カラム名	source_daimon_id	source_id	daimon_type	table_qualifier	priority
行	100	12	0	cdm	0
行	101	12	1	cdm	1
行	102	12	2	results	1
行	103	12	5	temp	0

↑  
一応シーケンシャル

0:cdm  
1:vocabulary  
2:results  
5:temp

↑  
schema実体名