

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 Meeting #72

2025年11月 イブニング・カンファレンス

2025.11.28



本日の内容

■ OHDSI 論文の紹介

■ OHDSI global/APAC から

- Global Symposium参加報告（山下先生／九州大学）

■ 話題

■ OMOP学会発表アンコール2題

(1) An Observational Study on Obesity Disease in Japan Using Payer Claims Data- Comparison between OMOP CDM and Native data-（岳野さん／IQVIA）

(2) OMOPデータを用いた日本の結核患者の治療実態調査-海外の結核低まん延国との比較研究-（伊澤さん／IQVIA）

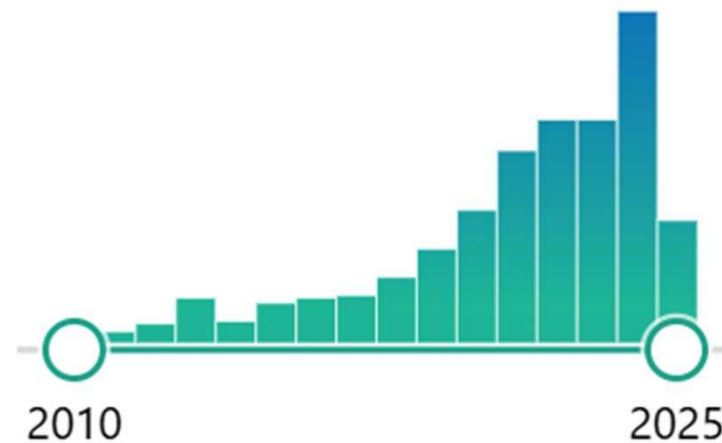


OHDSI 論文の紹介



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: 595報(2025年9月4日) → 619報(2025年11月27日)
- 2024年は約126報



(積み残し論文)

1. Yue Y, Khanal A, Lyu T, Weissman S, Liang C. EHR Phenotyping Methods for Measuring Treatment Adherence Among People Living With HIV in All of Us: Towards Disparities and Inequalities in HIV Care Continuum. 2024;
2. Zhou X, Dhingra LS, Aminorroaya A, Adejumo P, Khera R. A Novel Sentence Transformer-based Natural Language Processing Approach for Schema Mapping of Electronic Health Records to the OMOP Common Data Model. 2024;
3. Biggin F, White LM, Ashcroft Q, Howcroft T, Chandrabalan VV, Emsley H, ほか. Density of routinely collected neurology data depends on patient visit type: an investigation using the observational medical outcomes partnership common data model. BMJ Neurol Open. 2025年;7(2):e001202.
4. Chalabianloo N, Abdullah SS, Omrani MA, Jafari A, Sedig K, Muanda FT. Enhancing adverse drug reaction data quality in Canada: A high-precision pipeline for medication name standardization and enrichment. PLoS One. 2025年;20(9):e0331940.
5. Layton D, Hester L, Golozar A. Editorial: External control arms for single-arm studies: methodological considerations and applications. Front Drug Saf Regul. 2025年;5:1579171.
6. Xie J, Du M, Guo Y, Barboza C, Brash JT, Delmestri A, ほか. Trends in prescription opioid use in Europe: A DARWIN EU(®) multinational cohort study including seven European countries. Front Pharmacol. 2025年;16:1608051.
7. Pan Y, Manuel W, Abeyasinghe R, Zheng J, Davydov A, Yang Q, ほか. Mapping and Harmonization of CVX vaccine terms to the Vaccine Ontology. United States; 2025.
8. Aoyagi Y, Baba M, Terao S, Ikeda Y, Nomura K, Sato A. Feasibility of Converting EMR Data to OMOP CDM and Utilizing OHDSI Analysis Tools in Japan. Stud Health Technol Inform. 2025年8月;329:1946–7.
9. Blacketer C, Schuemie MJ, Moinat M, Voss EA, Camprubi M, Rijnbeek PR, ほか. Advancing Real-World Evidence Through a Federated Health Data Network

OMOPの論文ではなかった>

後日、発表いただきたい>

- (EHDEN): Descriptive Study. J Med Internet Res. 2025年8月;27:e74119.
10. Hu Y, Hong N, Li Y, Peng X, Chen Y, Xu H. PheCatcher: Leveraging LLM-Generated Synthetic Data for Automated Phenotype Definition Extraction from Biomedical Literature. Stud Health Technol Inform. 2025年8月;329:718–22.
11. Jo SB, Ahn ST, Joo HJ, Kim JW, Oh MM. Carbapenem Resistance and ESBL-Producing Enterobacteriaceae in Patients with Urological Infections from 2012 to 2021 in Three Korean Hospitals. Diagnostics (Basel). 2025年8月;15(16).
12. Lee YA, Lu Y, Bian J, Guo J, He X. Transforming the Medicare Claims Data to the OMOP Common Data Model. Stud Health Technol Inform. 2025年8月;329:1570–1.
13. Martin B, Kelly W, Morgan-Cooper H, Falconer T, Park E, Desai P, ほか. Identification of Adult Dermatomyositis Patients Using Real-World Data Sources. Arthritis Care Res (Hoboken). 2025年8月;
14. Ramsey-Hardy C, Skanderson M, Tate JP, Justice AC, Marconi VC, Alcorn C, ほか. Comparison of 2 electronic health record data extraction methods for laboratory tests used in the Veterans Aging Cohort Study Index. JAMIA Open. 2025年8月;8(4):ooaf074.
15. Burkard T, Camprubi M, Prieto-Alhambra D, Rijnbeek P, Pineda Moncusi M. Best practices to design, plan, and execute large-scale federated analyses - key learnings and suggestions from a study comprising 52 databases. Appl Clin Inform. 2025年9月;
16. Kafatos G, Maskell J, Archangelidi O, Neasham D. Validation of the transformed clinical practice research datalink (CPRD) GOLD and aurum data into the OMOP common data model. Health Informatics J. 2025年9月;31(3):14604582251381270.

今回の対象論文:黄色ハイライト



(積み残し論文、続き)

19. Katsch F, Mészáros Á, Héja T, Hussein R, Duftschmid G. Semiautomatic mapping of a national drug terminology to standardised OMOP drug concepts using publicly available supplementary information. *BMC Med Res Methodol*. 2025 年9月;25(1):213.
20. Neumann D, Gebler R, Kiederle J, Beck J, Aubele F, Struebing A, ほか. Development and Implementation of an Open, Modular, and Participatory Toolchain for Distributed IT Development in Healthcare Research - Lessons Learned. *Stud Health Technol Inform*. 2025 年9月;331:378–85.
21. Tomasik R, Konar S, Eklund N, Engels C, Dudova Z, Kacova R, ほか. Definitions to data flow: Operationalizing MIABIS in HL7 FHIR. *J Biomed Inform*. 2025 年9月;104919.
22. Abedian S, Yesakov E, Ostrovskiy S, Hussein R. Integrating Garmin Wearable Data into FHIR-Based Health Systems for Improved Interoperability. *Stud Health Technol Inform*. 2025 年10月;332:185–9.
23. Paradinha R, Barros V, Almeida JR, Oliveira JL. A Semantic-Driven for Cohort Data Harmonisation into OMOP CDM Schema. *Stud Health Technol Inform*. 2025 年10月;332:190–4.
24. Rajwa P, Borkowetz A, Abbott T, Alberti A, Beyer K, Bjartell A, ほか. Observational Health Data Analysis of the Cardiovascular Adverse Events of Systemic Treatment in Patients with Metastatic Hormone-sensitive Prostate Cancer: Big Data Analytics Using the PIONEER Platform. *Eur Urol Focus*. 2025 年10月;S2405-4569(25)00251-2.

(11/27/2025 新出論文)

1. Cuyàs B, Alvarado-Tapias E, Tan EH, Golozar A, Duarte-Salles T, Delmestri A, et al. Trends in incidence, prevalence, and survival of primary liver cancer in the United Kingdom (2000-2021). *Eur J Public Health*. 2025 Nov;ckaf153.
2. Badri P, Hernández I, Long J, Amin M, Friesen R. Chronic orofacial pain and psychological distress: findings from a multidisciplinary university clinic. *J Oral Facial Pain Headache*. 2025 Sept;39(3):152–62.

3. Cai CX, Nishimura A, Baxter S, Goetz K, Hribar M, Toy B, et al. Semaglutide and diabetic retinopathy: an OHDSI network study. *BMJ Open Diabetes Res Care*. 2025 Nov;13(6).
4. Bellas L, Català M, Burn E, Guo Y, Du M, Verhamme K, et al. Secular Trends in the Use of Valproate-Containing Medicines in Women of Childbearing Age in Europe: A Multinational DARWIN EU Network Study. *Pharmacoepidemiol Drug Saf*. 2025 Oct;34(10):e70232.
5. Lee KH, Jang S, Kim GJ, Park S, Kim D, Kwon OJ, et al. Large Language Models for Automating Clinical Trial Criteria Conversion to Observational Medical Outcomes Partnership Common Data Model Queries: Validation and Evaluation Study. *JMIR Med Inform*. 2025 Oct;13:e71252.
6. Lim HC, Wong H, Philip R, Van Der Vegt A, Choo KKR, Pole JD, et al. Streamlining electronic medical record data extraction and validation in digital hospitals: A systematic review to identify optimal approaches and methods. *Learn Health Syst*. 2025 Oct;9(4):e70024.
7. Razzaghi H, Dickinson K, Wieand K, Boss S, Weidlich H, Huang Y, et al. A multifaceted approach to advancing data quality and fitness standards in multi-institutional networks. *J Am Med Inform Assoc*. 2025 Oct;ocaf181.
8. Jung G, Lee J, Gho SM, Han Y, Choi B, Cho JW, et al. Predicting oxcarbazepine-induced hyponatremia in adult epilepsy patients: A multicenter machine learning analysis using real-world CDM data. *Seizure*. 2025 Oct;133:167–74.
9. Datzmann T, Lang C, Tesch F, Spoden M, Dröge P, Ehm F, et al. Evaluation of hybrid stroke quality indicators by integrating NIHSS and claims data for improved outcome prediction. *Sci Rep*. 2025 Nov;15(1):38994.
10. Rappoport N, Livne G, Perry Cohen N, Makover N, Eshel-Geva H, Kapach H, et al. Kineret: Israel's Largest Hospital Network Transformed into the OMOP common data model for collaborative research. *PLoS One*. 2025;20(10):e0334848.



11. Floyd SB, Mills A, Woloff J, Lowson C, Hilton C, Oeffinger D, et al. Development and validation of a computable phenotype for adolescent idiopathic scoliosis. *Learn Health Syst.* 2025 Oct;9(4):e70018.
12. Adams MCB, Hurley RW, Bartels K, Perkins ML, Hudson C, Topaloglu U, et al. Extending the Observational Medical Outcomes Partnership (OMOP) Common Data Model for Critical Care Medicine: A Framework for Standardizing Complex ICU Data Using the Society of Critical Care Medicine's Critical Care Data Dictionary (C2D2). *Crit Care Med.* 2025 Nov;
13. Ompad G, Cesta CE, Cohen JM, Leinonen MK, Taipale H, Li H, et al. Current Use of Common Data Models in the Nordic Countries. *Pharmacoepidemiol Drug Saf.* 2025 Nov;34(11):e70242.
14. Guerrero P, Ernebjerg M, Holst T, Weese D, DiBello H, Ibing S, et al. The AIR·MS data platform for artificial intelligence in healthcare. *JAMIA Open.* 2025 Dec;8(6):ooaf145.
15. Finster M, Wenzel M, Taghizadeh E. Common data models and data standards for tabular health data: a systematic review. *BMC Med Inform Decis Mak.* 2025 Nov;25(1):422.



全米HIV感染者における治療遵守度測定のための電子健康記録フェノタイピング手法: HIVケア継続における格差と不平等に向けて

論文1

[EHR Phenotyping Methods for Measuring Treatment Adherence Among People Living With HIV in All of Us: Towards Disparities and Inequalities in HIV Care Continuum]

Yuanzhen Yue, MS 1, Ashok Khanal, Bpharm 1, Tianchu Lyu, MPH 1, Sharon Weissman, MD 2,3, Chen Liang, PhD 1,4,5

1. University of South Carolina Arnold School of Public Health, Columbia, South Carolina, USA
2. University of South Carolina School of Medicine, Columbia, South Carolina, USA
3. Prisma Health, Columbia, South Carolina, USA
4. University of Washington, School of Medicine, Department of Biomedical Informatics and Medical Education, Seattle, Washington
5. National Institutes of Health, Bethesda, Washington DC

1. 背景

- ART の進歩により HIV 感染者の寿命と QOL は大幅に改善した。
- しかし 米国の HIV 感染者のうちケア継続できているのは約 50% と低い。

(1.背景)

- アドヒアランス測定には自己申告、ピルカウント、電子モニタリング、薬物濃度測定があるが、いずれも完全な方法ではない。大規模 EHR の活用が期待される一方、データ欠損・重複・語彙の粗さなど多くの課題が存在する。



全米HIV感染者における治療遵守度測定のための電子健康記録フェノタイピング手法: HIVケア継続における格差と不平等に向けて

論文1

[EHR Phenotyping Methods for Measuring Treatment Adherence Among People Living With HIV in All of Us: Towards Disparities and Inequalities in HIV Care Continuum]

2. 研究目的

- All of Us (OMOP Common Data Model 準拠)を用いて、HIV 治療アドヒアランスを評価するための 8 つのフェノタイピング指標(metrics)を開発。
- さらにアドヒアランスの格差(人種・年齢・性別など)を探索的に分析。

3. データソース

- All of Us (2018年～、50万超の参加者、80%以上が underrepresented groups)。
- Controlled Tier(EHR、ウェアラブル、調査回答等)を活用。

4. 開発した 8 つの指標

- Metric 1:診断済みだが ART 未開始HIV 診断後に ART 処方記録が存在しない者。
- Metric 2:ケアへの定着(Retention)12 カ月内に HIV 関連受診が 2 回以上(最大 6 カ月の間隔を許容)。
- Metric 3:MPR(Medication Possession Ratio) ART 所持日数 / 期間日数 ×100%。90%以上をアドヒアランス良好と定義。6カ月、1年、2年で測定。
- Metric 4:ウイルス学的失敗(Virologic Failure)ART 開始 36 週間後に、200 copies/mL 以上のウイルス量が連続 2 回。



全米HIV感染者における治療遵守度測定のための電子健康記録フェノタイピング手法: HIVケア継続における格差と不平等に向けて

論文1

[EHR Phenotyping Methods for Measuring Treatment Adherence Among People Living With HIV in All of Us: Towards Disparities and Inequalities in HIV Care Continuum]

(4. 開発した 8 つの指標)

- Metric 5: CD4 に基づくケアへの後期参入(Late Entry) ART 開始前の $CD4 < 200/mm^3$ または $CD4/CD8$ 比 < 0.5 。
- Metric 6: ART 血中濃度 EHR にデータがなく計算不能。
- Metric 7: ケアからの離脱(Self-report)「現在 HIV 診療を受けているか」「治療薬を受けているか」の回答が "No"。
- Metric 8: 治療アドヒアランス不良(臨床メモ)診療記録の観察欄に非アドヒアランスの記載。

5. 結果概要

- Metric 6 はデータ欠如、Metric 7 は統計モデルが収束せず未報告。
- 6つの指標が実用可能。
- 各指標で抽出された患者数が示されている。

主な格差の分析結果

- 黒人(Black/African American)は ART 未開始の可能性が高い($OR=3.13$)。
- ART 開始後の ケア離脱は男性・50-64歳・黒人で多い。
- 男性は MPR アドヒアランスが高い傾向。高齢者(≥ 50 歳)はウイルス学的失敗が少ない。
- 黒人は非アドヒアランスのリスクが高い($OR=5.80$)。



全米HIV感染者における治療遵守度測定のための電子健康記録フェノタイピング手法: HIVケア継続における格差と不平等に向けて

論文1

[EHR Phenotyping Methods for Measuring Treatment Adherence Among People Living With HIV in All of Us: Towards Disparities and Inequalities in HIV Care Continuum]

6. 考察

- EHR 正規化データを利用した HIV アドヒアランスの近似測定は実現可能である。
- All of Us の多様な人口構成は、健康格差研究の基盤として有用。
- 本研究の結果は、従来研究と一致し、指標の妥当性を支持。
- EHR ベースのアドヒアランス測定は、低コストで大規模スクリーニングに適している。

7. 限界と今後の課題

- EHR の欠損、クラス不均衡、主観的記録によるバイアスの可能性。
- Metric 3, 7, 8 には主観性が残る。他データセットでの再検証が必要。

8. 結論

All of Us データを用いて HIV 治療アドヒアランス指標を開発し、健康格差(特に人種格差)の可視化に成功。今後は EHR に基づく体系的な HIV ケア格差研究に応用可能。



電子カルテを OMOP 共通データモデルへスキーママッピングするための Sentence Transformer を用いた新規自然言語処理アプローチ

[A Novel Sentence Transformer-based Natural Language Processing Approach for Schema Mapping of Electronic Health Records to the OMOP Common Data Model]

Xinyu Zhou BS 1, Lovedeep Singh Dhingra MBBS 2, Arya Aminorroaya MD, MPH 2, Philip Adejumo BS 2, Rohan Khera MD, MS 1,2,3,4

1. Department of Biostatistics, Yale School of Public Health, New Haven, CT, USA;
2. Yale School of Medicine, New Haven, CT, USA;
3. Biomedical Informatics and Data Science, Yale School of Medicine, New Haven, CT, USA;
4. Center for Outcomes Research and Evaluation, Yale-New Haven Hospital, New Haven, CT, USA

研究背景

- EHR データを OMOP CDM のような共通データ標準へマッピングすることは、多施設研究に必要不可欠。
- 特に薬剤データは表記揺れや用量・頻度の多様性が多く、マッピングが難しい。
- 既存手法(Usagi、TOKI)は限界があり、大規模な教師データを必要とする。



電子カルテを OMOP 共通データモデルへスキーママッピングするための Sentence Transformer を用いた新規自然言語処理アプローチ

[A Novel Sentence Transformer-based Natural Language Processing Approach for Schema Mapping of Electronic Health Records to the OMOP Common Data Model]

方法

- 公開ボキャブラリ(Athena)および医療略語データベース (Metainventor)を利用し、Sentence Transformer を事前学習・微調整。
- mpnet、BioLORD、GatorTron を基盤に複数の派生モデルを作成。
- Yale New Haven Health System(YNHHS)の薬剤データ 39,441 種を評価対象とした。類似度スコア(コサイン類似度)に基づき OMOP 標準概念を自動推定。
- LLM(SFR-Embedding-Mistral)、Usagi、RapidFuzz などと比較。

主な結果

- 最頻薬剤200種のマッピング:
 - mpnet-drug が最高精度:96.5%
 - Usagi(90.0%)、SFR-Mistral(89.5%)を上回る。
- ランダム抽出薬剤200種:
 - mpnet-drug:83.0%
 - 他の Transformer ベースの drug モデルも既存手法を上回った。
- 文字列一致(RapidFuzz)は 7.5% と非常に低精度。
- 全領域のマッピングで学習したモデル(mpnet-all など)は drug 専用モデルに劣った。



電子カルテを OMOP 共通データモデルへスキーママッピングするための Sentence Transformer を用いた新規自然言語処理アプローチ

論文2

[A Novel Sentence Transformer-based Natural Language Processing Approach for Schema Mapping of Electronic Health Records to the OMOP Common Data Model]

考察

- 公開データのみで高精度を実現しており、機密データを使わない点が利点。
- 薬剤領域は特に複雑で既存手法の限界が顕著であり、Transformer による改善効果大きい。他施設への汎用性も期待できる。
- 今後は他概念領域(手技・診断名)への応用も可能。

結論

- Transformer ベースの Sentence Transformer モデルは、EHR 薬剤データを OMOP CDM 標準概念へ高精度に自動マッピング可能。
- EHR → OMOP CDM 変換の自動化に大きく貢献する技術である。



日常的に収集される神経学データの密度が患者来院タイプによって異なる： 観察医療アウトカムパートナーシップ共通データモデルを用いた調査

[Density of routinely collected neurology data depends on patient visit type: an investigation using the observational medical outcomes partnership common data model]

Fran Biggin 1,2, Laura M White 2, Quinta Ashcroft 2, Timothy Howcroft 2, Vishnu Vardhan Chandrabalan 2, Hedley Emsley 1,2, Jo Knight 1,2

1. Lancaster University Faculty of Health and Medicine, Lancaster, UK
2. Lancashire Teaching Hospitals NHS Foundation Trust, Preston, UK

1. 背景

- OMOP Common Data Model(OMOP CDM)は医療データを標準化して統合解析を可能にする枠組みで、NHSの Secure Data Environment(SDE)にも採用されている。
- 英国では外来(Outpatient; OP)の診断コードが必須ではなく、外来診療データが極めて少ないことが以前から指摘されている。
- 神経内科では慢性疾患が多く外来フォローが中心となるため、外来データの不備は研究と臨床の双方に大きな問題となる。

2. 方法

- Lancashire Teaching Hospitals NHS Foundation Trust (LTHFT)のデータを使用
- 2022/4/1～2023/3/31 に 神経内科外来を1回以上受診した 23,862人 をコホートに設定。
- これら患者の 2021/4～2024/3 の全受診データ(外来、入院、救急)を OMOP CDM 形式で抽出。
- 受診タイプに応じて、以下4つのサブコホートに分類：
外来のみ(OP only)、外来+入院(OP + IP)
外来+救急(OP + ED)、外来+入院+救急(OP + IP + ED)



日常的に収集される神経学データの密度が患者来院タイプによって異なる： 観察医療アウトカムパートナーシップ共通データモデルを用いた調査

[Density of routinely collected neurology data depends on patient visit type: an investigation using the observational medical outcomes partnership common data model]

3. 結果

3.1 データ量の大きな偏り

- 外来のみ群(全体の64%)では、全データの1.3%しか記録されていなかった。
- 外来+入院+救急群(18%)が全レコードの80%以上を占めるほど偏りが大きかった。

3.2 記録の種類(条件・薬剤・検査など)の不足

- 外来のみ群では 診断(condition)記録はほぼ“0%”。
- 薬剤記録もほぼ存在せず、臨床医は患者の過去診断・処方情報をデータから把握できない。

3.3 診療1日あたりの記録数の差

- 入院+救急群(サブコホート4)は 1日あたり中央値 29.9件の測定記録(measurement)を有する。
- 一方、外来のみ群は 中央値 0~0.27件 と極めて少なかった。



日常的に収集される神経学データの密度が患者来院タイプによって異なる： 観察医療アウトカムパートナーシップ共通データモデルを用いた調査

[Density of routinely collected neurology data depends on patient visit type: an investigation using the observational medical outcomes partnership common data model]

4. 考察

- 外来診療の診断コードが欠落していることが最大の問題点であり、研究では外来患者の病態や薬剤情報を捕捉できない。
- その結果、データに基づく研究は「入院や救急を必要とした重症者のみ」を偏って抽出することになり、疾患の全体像(軽症～重症の幅)を正しく捉えられない。
- 不必要に大量のデータ(特に入院・救急での測定値)が蓄積されている一方、外来データは不足しているという非対称性も問題。
- 外来診断の標準的なデータ収集が行われなければ、SDEやOMOP CDMの研究価値が制限され続ける。

5. 結論

- 外来患者のデータは極めて不十分であり、入院や救急患者と比べて著しく情報量が少ない。
- とくに 診断(condition)・薬剤(drug)データの欠落が重大で、臨床現場・研究の両面に支障を来す。
- OMOP CDM による可視化によって、このデータ格差が明確に示され、外来診療のデータ収集改善の必要性が強調された。



カナダの有害事象報告における医薬品名の標準化とデータ品質向上のための高精度パイプライン

[Enhancing adverse drug reaction data quality in Canada: A high-precision pipeline for medication name standardization and enrichment]

Niaz Chalabianloo 1,2,3, Sheikh S. Abdullah 2,3,4,5, Mohammad Ali Omrani 1, Atefeh Jafari 3,6, Kamran Sedig 2,7, Flory Tsobo Muanda 1,3,6,8

1. Department of Physiology and Pharmacology, Western University, London, Ontario, Canada
2. Department of Computer Science, Western University, London, Ontario, Canada
3. ICES Western, London, Ontario, Canada
4. Department of Computer Science, MacEwan University, Edmonton, Alberta, Canada
5. London Health Sciences Centre Research Institute, London, Ontario, Canada
6. Department of Epidemiology and Biostatistics, Western University, London, Ontario, Canada
7. Faculty of Information and Media Studies, Western University, London, Ontario, Canada
8. Lawson Health Research Institute, London Health Sciences Centre, London, Ontario, Canada

背景

- カナダの Canada Vigilance データベースでは、医薬品名のばらつき(ブランド名、一般名、略語、綴りの揺れ)が深刻で、36.8% の名称が有効成分にマッピングされていないという品質問題が存在する。
- 名称の不統一により、有害事象シグナルが分散し、統計的検出力が低下している。

目的

- 医薬品名の 標準化と情報付加(有効成分、ATC 分類) を自動化・高精度で行うパイプラインを構築する。
- RxNorm と OHDSI 語彙を活用し、カナダの ADR データを国際基準に整合させる。



カナダの有害事象報告における医薬品名の標準化とデータ品質向上のための高精度パイプライン

[Enhancing adverse drug reaction data quality in Canada: A high-precision pipeline for medication name standardization and enrichment]

方法

- 46,585 種類の薬剤名を対象に、9 段階の改良を行いながらパイプラインを開発。
- 主なステップ:
 - ・ 前処理(表記揺れ除去、用語正規化、不要語削除、フランス語対応)
 - ・ RxNorm approximate matching API による近似一致
 - ・ OHDSI Athena を fallback として連携
 - ・ ATC コードや薬理クラス情報を RxNav API で追加
 - ・ 特殊ケース(国際ブランド、複合製剤など)をルール化して補正

(方法)

- 全処理を高速化するためにキャッシュ機構を導入。
- 二名の専門家(薬剤師)による blinded validation を実施(200 ケース)。



カナダの有害事象報告における医薬品名の標準化とデータ品質向上のための高精度パイプライン

[Enhancing adverse drug reaction data quality in Canada: A high-precision pipeline for medication name standardization and enrichment]

結果

- 最終マッチ率: 94.5% (44,034 / 46,585 名称)
- 専門家による検証:
Precision: 98.02%, Specificity: 97.22%
Recall: 77.34%
- マッチ成功した名称のうち:
98.6% で有効成分情報を取得, 73.8% で ATC コードを付与
- 不一致となった名称の多くは以下:
サプリメント・自然健康食品, 医療機器, 海外専売ブランド名, 不明瞭な記述

(結果)

- 標準化により、有害事象シグナルが明確化:
 - ・ メサラミン(mesalamine-15ブランド名)と無力症(asthenia) (signal of disproportionate reporting (SDR))
 - ・ ヒドロクロロチアジド(“HCTZ”, “HydroDIURIL”等)と紅斑(erythema)
→ 個別ブランド名では検出できなかったシグナルが、統合後に有意となった。



カナダの有害事象報告における医薬品名の標準化と データ品質向上のための高精度パイプライン

[Enhancing adverse drug reaction data quality in Canada: A high-precision pipeline for medication name standardization and enrichment]

考察

- 医薬品名の異質性は、安全性シグナルの検出力を大幅に低下させる。
- 提案パイプラインは高精度で誤マッチが非常に少なく、薬剤安全性監視の信頼性向上に寄与する。
- 国際基準(RxNorm・ATC)により、FAERS や VigiBase など他国データとの統合解析が容易になる。

限界

- RxNorm に収載されていない自然健康食品・国際ブランド名は対応が難しい。
- 多言語対応(特にフランス語)は限定的。
- 完全自動化では一部の複雑な用語を取りこぼす可能性。
- validation サンプルは200件と小規模。



カナダの有害事象報告における医薬品名の標準化と データ品質向上のための高精度パイプライン

[Enhancing adverse drug reaction data quality in Canada: A high-precision pipeline for medication name standardization and enrichment]

結論

- 薬剤名の標準化は ADR データの品質向上に極めて重要であり、本パイプラインは 高精度・高再現性の実用的手法を提供する。
- カナダ ADR データにおける医薬品名の大規模な不統一 (46,585 名称・36.8% 未マッピング) をほぼ解消し、安全性シグナル検出の感度向上と国際的な相互運用性向上を達成した。



カナダの有害事象報告における医薬品名の標準化とデータ品質向上のための高精度パイプライン

[Enhancing adverse drug reaction data quality in Canada: A high-precision pipeline for medication name standardization and enrichment]

前処理

- 前処理では、薬名から「mg」「tablet」などの不要語や略語を整理して核心となる薬の概念のみを残す
- 英仏バイリンガル表記にも対応するために特定のフランス語用語やブランド名の置換ルールを追加
- 自動処理では誤りやすい複雑な薬名については、専門家が確認した特例マッピングによって確実に補正

マッチング・標準化

- 本手法では、まず過去の結果をキャッシュから確認
- 見つからなければRxNorm APIで最も適切な標準概念(RxCUI; RxNorm Concept Unique Identifier)を取得
- 失敗した場合のみOHDSI Athenaを用いて代替名を探索したうえで再度RxNormに照会
- これらの手順を経ても標準化できない場合は未マッチとして扱う

外部データエンリッチメント

- RxCUI が確定した後に RxNav API を用いて有効成分、標準化一般名、薬効分類、ATC コードなどの追加情報を取得
- 確実に RxNorm 概念に基づいた詳細データで記録を補強
- 各薬剤が標準的かつ一貫した薬理学的情報と結びつけられる構造



外部対照群を用いた単群試験：方法論的考察と応用(Editorial)

論文5

[Editorial: External control arms for single-arm studies: methodological considerations and applications]

Deborah Layton 1, Laura Hester 2 and Asieh Golozar 3

1. Lane, Clark and Peacock (LCP) LLP, London, United Kingdom
2. Johnson & Johnson, Horsham, PA, United States,
3. Nemesis Health, Observational Health Data Sciences and Analytics (OHDSI), New York, NY, United States

1. はじめに(Introduction)

- 外部比較群(External Comparator:EC)は、臨床試験外の患者データを用いて、試験結果の比較や文脈づけを行うために用いられる。
- データ源には電子カルテ、疾患レジストリ、他の臨床試験など多様なリアルワールドデータ(RWD)が含まれる。
- がんや希少疾患領域では、ランダム化比較試験(RCT)が困難な場合、EC の活用が増加している。

(1. はじめに(Introduction))

- 規制当局や HTA(医療技術評価)での活用が増えており、バイアスを最小限にした高品質な方法論の必要性が高まっている。
- 本特集号では EC 研究の透明性と堅牢性を高めるための最新の方法論を扱った 5 つの論文を紹介している。



外部対照群を用いた単群試験: 方法論的考察と応用(Editorial)

論文5

[Editorial: External control arms for single-arm studies: methodological considerations and applications]

2. 用語標準化と概念枠組み(Standardizing nomenclature and conceptual frameworks)

- EC 研究の広がりに伴い、明確で一貫した用語の使用が不可欠。
- Rippin らは “Externally Controlled Trial” という表現は、EC が事前にプロトコルで規定されていない限り誤解を生むと指摘。
- External Comparator Cohort(ECC) や External Comparator(EC) といった用語を推奨。
- EC は RCT の対照群とは根本的に異なる集団であり、別のデータ収集方法を使用するため、RCT の「アーム」という概念は適切でない。
- 研究者・規制当局が現実的な期待値を持ち、観察研究として EC を正しく扱うためには用語の明確化が重要。



外部対照群を用いた単群試験: 方法論的考察と応用(Editorial)

論文5

[Editorial: External control arms for single-arm studies: methodological considerations and applications]

3. Estimand の役割(Estimand Framework の適用)

- Rippin による別の論考では、ICH E9(R1)付録の estimand framework を EC 研究に適用することを提案。
- 5 つの要素(治療条件、集団、評価項目、介入後イベントの扱い、集団レベルの要約)に加えて、
 - ・ ベースライン定義
 - ・ 推定量(marginal estimator)
 - ・ データの完全性など EC 特有の追加要素を考慮する必要があると指摘。
- RWD の不完全性やバイアスがあるため、明確な estimand の設定は EC 研究において特に重要。

4. ターゲット・トライアル・エミュレーション(TTE)

- Arnold らは、RCT と EC 研究のギャップを埋めるため Target Trial Emulation(TTE) を推奨。
- 理想的な RCT の設計(対象基準、治療戦略、アウトカム定義など)を定義し、それを RWD で模倣する手法。
- RWD の質や完全性が低い場合、推定にバイアスが生じるという限界を指摘。
- それにもかかわらず、透明性が高く構造化されたフレームワークであるため、FDA・EMA・MHRA など規制当局の信頼性向上に寄与。



外部対照群を用いた単群試験: 方法論的考察と応用(Editorial)

論文5

[Editorial: External control arms for single-arm studies: methodological considerations and applications]

5. ハイブリッド自然歴研究(Hybrid Natural History Studies)

- Ugoji らは、希少疾患領域で増えている ハイブリッド型自然歴研究(NHS) を解説(回顧的+前向き組み合わせ)。
- 単独の回顧的／前向き EC よりも柔軟性があり、希少疾患での EC 作成に有用。
- 設計・実装には追加の考慮点(疾患特性、運用上の制約など)が必要。
- 文献が少ない領域であり、規制提出を念頭に置いた重要なガイダンスを提供。



外部対照群を用いた単群試験: 方法論的考察と応用(Editorial)

論文5

[Editorial: External control arms for single-arm studies: methodological considerations and applications]

6. バイアスの課題と対策(Bias in EC studies)

- Ackerman らは、特に腫瘍領域での 測定誤差が PFS 推定に与える影響 を解析。
- 主なバイアス:
 - ・ 誤分類バイアス(Misclassification Bias): 進行イベントの誤った分類
 - ・ サーベイランスバイアス(Surveillance Bias): RWD の評価頻度が試験よりも不規則
- サーベイランスバイアス単独の影響は小さいが、誤分類と組み合わせると大きなバイアスを生む。
- 表面的に PFS が一致していても、個々のデータ品質の問題によりバイアスが残存する可能性が高い。

(6. バイアスの課題と対策(Bias in EC studies))

- 推奨:
 - ・ エンドポイント推定アルゴリズムの改善
 - ・ バイアスを定量化するシミュレーション結果
 - ・ 解釈時の文脈化と透明性



外部対照群を用いた単群試験：方法論的考察と応用(Editorial)

[Editorial: External control arms for single-arm studies: methodological considerations and applications]

7. 結論(Conclusion)

- EC 研究は規制当局や HTA において重要性が増しており、方法論の洗練が求められる。
- 用語標準化、TTE、ハイブリッド自然歴研究、バイアス低減などが信頼性向上の鍵。
- 特にがん・希少疾患での EC 需要が増えており、質の高いデータ、透明性、観察研究としての厳密性が今後の成功に不可欠。
- 規制当局・産業界・研究者の継続的協調が EC 活用の発展に重要。



欧州における処方オピオイド使用の動向:7カ国を対象とした DARWIN EU® 多国間コホート研究

[Trends in prescription opioid use in Europe: A DARWIN EU® multinational cohort study including seven European countries]

Junqing Xie 1, Mike Du1, Yuchen Guo 1, Cesar Barboza 2, James T. Brash 3, Antonella Delmestri 1, Talita Duarte-Salles 2,4, Jasmine Gratton 3, Romain Griffier 5, Raivo Kolde 6, Wai Yi Man 1, Núria Mercadé-Besora 4, Marek Oja 6, Sarah Seager 3, Katia Verhamme 2, Dina Vojinovic 7, Edward Burn 1, Daniel Prieto-Alhambra 1,2, Martí Català 1 and Annika M. Jödicke 1

1. Pharmaco- and Device Epidemiology Group, Centre of Statistics in Medicine, NDORMS, University of Oxford, England, United Kingdom
2. Department of Medical Informatics, Erasmus Medical Center, Rotterdam, Netherlands
3. IQVIA, Real World Solution, Brighton, United Kingdom
4. Fundació Institut Universitari per a la recerca a l' Atenció Primària de Salut Jordi Gol i Gurina (IDIAPJGol), Barcelona, Spain
5. Public Health Department, Medical Information Service, Bordeaux University Hospital, Bordeaux, France,
6. Institute of Computer Science, University of Tartu, Tartu, Estonia, 7IQVIA, Real World Solutions, Amsterdam, Netherlands



欧州における処方オピオイド使用の動向:7カ国を対象とした DARWIN EU® 多国間コホート研究

[Trends in prescription opioid use in Europe: A DARWIN EU® multinational cohort study including seven European countries]

1. 背景

- 北米では長年にわたりオピオイド危機が深刻な公衆衛生問題となっており、過量摂取・死亡が増加。
- 欧州でも同様の懸念が高まっており、処方オピオイド使用の動向把握が必要。
- 既存研究は手法がまちまちで国間比較が困難だったため、標準化された方法による調査が求められていた。

2. 目的

- 欧州7カ国の医療／データベースを用い、2012～2022年の処方オピオイド使用の発生率(incidence)と有病率(prevalence)の推移を評価。

(2. 目的)

- 新規オピオイド使用者の特徴・推定される処方適応・治療期間を明らかにする。

3. 方法

- UK・スペイン・オランダのプライマリケア、ベルギー・ドイツの外来データ、フランスの病院データ、エストニアのバイオバンク請求データを対象。
- すべて OMOP CDM に標準化。
- 2012～2022年の新規オピオイド使用を抽出(12か月の洗い出し期間)。
- 弱オピオイド(例:codeine, tramadol)と強オピオイド(例:morphine, fentanyl, oxycodone)に分類。



欧州における処方オピオイド使用の動向:7カ国を対象とした DARWIN EU® 多国間コホート研究

[Trends in prescription opioid use in Europe: A DARWIN EU® multinational cohort study including seven European countries]

4. 主な結果

4.1 オピオイド使用の全体動向

- オピオイド使用の有病率は高止まりし、国によっては増加(例:EBB +113%)。
- 新規処方の発生率は多くの国で減少(例:UK CPRD -50.7%)。
- 2020-2021年はCOVID-19の影響により処方が一時的に低下。

4.2 年齢別の特徴

- 高齢層で最も使用率が高い(例:UK、81歳以上で約31%が使用)。
- 10歳以下は極めて低い(0.1%)。

4.3 オピオイドの種類別動向

- 減少傾向:codeine, tramadol(多くの国で低下)
- 増加傾向:oxydone, morphine, fentanyl(多くの国で上昇)
- 強オピオイドの使用増加が特に顕著であり、公衆衛生上のリスクが懸念される。

4.4 新規使用者の特徴

- 新規使用者は合計 6,196,266人。
- 女性の割合が高い(多くの国で60%前後)。
- 中位年齢50~57歳。
- 治療期間は設定によって大きく変動:
 - ・ 病院:中央値2日
 - ・ プライマリケア:9~20日
 - ・ 強オピオイド:より長期化する傾向



欧州における処方オピオイド使用の動向:7カ国を対象とした DARWIN EU® 多国間コホート研究

[Trends in prescription opioid use in Europe: A DARWIN EU® multinational cohort study including seven European countries]

4.5 推定される処方適応(診断)

- 呼吸器系疾患(咳・上気道感染)
- 疼痛(腰痛、膝痛、脊椎痛など)
- 特に codeine では「咳」関連が目立つ(例:ベルギーで26.9%)。
- 病院では「術後合併症」など急性痛も多い。

5. 考察

- 新規処方は減少しても、使用有病率が高止まりしている背景には「継続使用者」の存在が示唆される。
- 強オピオイドの使用増加は、依存・過量摂取のリスク上昇につながる可能性。
- 高齢者での使用増加は特に注意が必要。
- COVID-19による医療アクセス低下が処方に影響した。

6. 限界

- 診療設定をまたぐデータ連結が不十分な国があり、過小推計の可能性。
- 実際の服薬は確認できず、服薬期間は処方記録に基づく推定。適応の正確な記録がないため、診断は「代理指標」として扱われた。

7. 結論

- 欧州では新規オピオイド処方は減少している一方、有病率は減らず高齢層中心に使用が続いている。
- 特に**強オピオイド(fentanyl, oxycodone, morphine)**の使用増加が顕著。今後は処方の妥当性評価、適切な使用基準の整備、継続的監視が必要。
- ポストパンデミック期の処方動向追跡が重要である。



CVX ワクチン用語の Vaccine Ontology(VO)への マッピングと調和に関する研究

[Mapping and Harmonization of CVX vaccine terms to the Vaccine Ontology]

Yuanyi Pan 1, Warren Manuel 2, Rashmie Abeysinghe 3, Jie Zheng 1, Alexander Davydov 4, Qi Yang 5, Asiyah Yu Lin 6, Licong Cui 2, Yongqun Oliver He 1

1. University of Michigan Medical School, Ann Arbor, MI, USA
2. McWilliams School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX, USA
3. Department of Neurology, The University of Texas Health Science Center at Houston, Houston, TX, USA;
4. Observational Health Data Sciences and Informatics, New York City, NY, USA;
5. IQVIA, Inc., Durham, NC, USA;
6. Axle Research and Technology, Rockville, MD, USA



CVX ワクチン用語の Vaccine Ontology(VO)へのマッピングと調和に関する研究

[Mapping and Harmonization of CVX vaccine terms to the Vaccine Ontology]

1. 背景

- 多様なワクチンが開発・利用される中で、ワクチン情報の標準化は不可欠。
- OMOP Common Data Model(CDM)は CVX や RxNorm を利用して EHR データを標準化するが、これらの語彙は十分な意味的階層構造や関係を持たず、ワクチン分類に限界がある。
- ワクチン領域の参照オントロジーである Vaccine Ontology(VO) を用いて、意味的关系を補完する取り組みを実施。

2. 目的

- CVX(CDC Vaccine Administered code)273語を VO にマッピングし、両者の調和を図ること。
- 半自動化手法(埋め込みベクトル、語レベル類似度、LLM 補正)と手動マッピングを組み合わせることで効率化を図る。



CVX ワクチン用語の Vaccine Ontology (VO) へのマッピングと調和に関する研究

[Mapping and Harmonization of CVX vaccine terms to the Vaccine Ontology]

3. 方法

3.1 CVX 用語の収集と分類

- 273 CVX 用語を疾患・病原体単位で体系的に分類。
- ウイルスワクチン(161)、細菌ワクチン(43)、寄生虫ワクチン(2)、混合ワクチン(33)、癌ワクチン(1)、免疫グロブリン等の“パッシブワクチン”(24)と整理。

3.2 手動マッピング

- VO 既存階層を見直し、必要に応じて新規 VO 用語を追加。
- 1対1の厳密なマッピングを原則とし、VO に存在しない場合は新規作成。
- “AS03 adjuvant” などワクチン以外のコードは VO 内の適切なクラスへ配置。

3.3 半自動マッピング

1. 手法埋め込みベース類似度 (Sentence embedding)
 2. 語レベル類似度 (Jaccard similarity)
 3. 大規模言語モデル (GPT-4o) を用いた精緻化
- それぞれ上位候補を抽出し、正しいマッピングの可能性を提示。



CVX ワクチン用語の Vaccine Ontology(VO)へのマッピングと調和に関する研究

[Mapping and Harmonization of CVX vaccine terms to the Vaccine Ontology]

4. 結果

4.1 CVX の体系化

- CVX の 273語すべてが VO にマッピング可能に。

4.2 半自動手法の性能

- 2022 年版 VO での検証では語レベル類似度 (Hybrid approach 1) が最も高精度(85.55%)
- 2024 年版 VO では精度がさらに向上(87.07%)。

4.3 パッシブワクチン(Passive vaccine)の追加

- CVX には 24件の免疫グロブリン／抗毒素／抗体製剤が含まれる。
- これらは能動免疫ではなく「受動免疫」を誘導するため、VO に “passive vaccine” クラスを新設し、24語を階層的に追加。

4.4 CVX-VO OWL の生成

- CVX に対応する VO 用語のみ抽出した cvx-vo.owl を公開。
- CVX に対する階層構造や論理関係を含み、アプリケーション利用が容易。

4.5 利用例

- Description Logic(DL)クエリにより「H1N1 ワクチンの抽出」などの群検索が容易に。
- VO 全体を検索することで、CVX に含まれないブランド名ワクチンも抽出可能。



CVX ワクチン用語の Vaccine Ontology(VO)への マッピングと調和に関する研究

[Mapping and Harmonization of CVX vaccine terms to the Vaccine Ontology]

5. 考察

- 手動+半自動のハイブリッド方式により、高品質で効率的なマッピングを達成。
- 将来的には RxNorm・SNOMED CT など OMOP の他語彙との統合にも拡張予定。
- “passive vaccine” の定義と階層化により、免疫グロブリン等の扱いが大きく改善。
- VO の拡張により、EHR 解析・ワクチン研究の精密なデータ統合が可能に。

6. 結論

- CVX 273語すべてを VO にマッピングし、語彙の意味的統合を達成。
- 半自動化手法はマッピングプロセスを大幅に支援し、将来の大型語彙統合への応用が可能。VO の強化により、OMOP CDM におけるワクチンデータの精度・検索性・一貫性が向上。



EHDENにおけるリアルワールドエビデンス創出の推進:記述的研究

[Advancing Real-World Evidence Through a Federated Health Data Network (EHDEN): Descriptive Study]

Clair Blacketer 1,2,3, MPH; Martijn J Schuemie 1,3,4, PhD, MS; Maxim Moinat 1,2, MS; Erica A Voss 1,3, PhD, MPH; Montse Camprubi 1,5, MS; Peter R Rijnbeek 1,2, PhD, MS; Patrick B Ryan 1,3,6, PhD, MS

1. OHDSI Collaborators, New York, NY, United States
2. Department of Medical Informatics, Erasmus MC, Rotterdam, The Netherlands
3. Johnson & Johnson (United States), Raritan, NJ, United States
4. Department of Biostatistics, University of California, Los Angeles, Los Angeles, CA, United States
5. Synapse (Spain), Madrid, Spain
6. Department of Biomedical Informatics, Columbia University, New York, NY, United States

1. 背景(Background)

- 欧州では国ごとの医療制度やデータ標準が異なるため、リアルワールドデータ(RWD)の統合・比較が困難。
- EHDEN(European Health Data & Evidence Network)は、欧州全域のRWDを OMOP CDM に標準化し、フェデレーション型解析を可能にするために設立された。

(1. 背景(Background))

- プライバシー保護のため、データは各データパートナーが保持し、解析結果のみ(集計値)を共有する仕組み。



EHDENにおけるリアルワールドエビデンス創出の推進:記述的研究

論文9

[Advancing Real-World Evidence Through a Federated Health Data Network (EHDEN): Descriptive Study]

2. 目的(Objective)

- 2024年9月時点でEHDENに参加する 210のデータソースの特性を記述し、研究や規制利用に向けた透明性を高める。

3. 方法(Methods)

- 欧州のデータパートナーを公募し、採択後に OMOP CDMへの変換(ETL)とデータ品質評価を実施。
- 64の中小企業(SME)がDPを支援し、標準化を支える体制を構築。
- データ品質は Data Quality Dashboard(DQD) や CDM Inspection Report により評価。各データソースは、国、対象者数、医療提供レベル、収集方法などのメタデータをEHDENポータルに登録。

4. 結果(Results)

4.1 データソースの概要

- 全210データソース、30カ国から提供。
- 国別上位:イタリア(13%)、英国(12.5%)、スペイン(11.5%)。
- 平均対象者数:約214万人、中央値:約45万人。

4.2 医療レベル・対象者の選択基準

- 二次医療のみ:46.7%(98件)
- 混合(一次+二次):42.4%(89件)
- 一次医療のみ:11%(23件)
- 対象者の取り込み基準:
 - ・ 医療受診:55.7%
 - ・ 疾患特異的:32.9%
 - ・ 人口ベース:11.4%



EHDENにおけるリアルワールドエビデンス創出の推進:記述的研究

論文9

[Advancing Real-World Evidence Through a Federated Health Data Network (EHDEN): Descriptive Study]

4.3 データ収集方法

- EHR(電子カルテ):74.7%(157件)
 - ・ うち85件がEHRのみで構成。
- 検査データ:29.5%(62件)(単独は1件のみ)
- その他:請求データ、症例報告書、調査、NLP、死亡票など、多様な組み合わせが存在。

5. 考察(Discussion)

- 欧州のRWDは多様であり、多数のデータソースを跨いだ解析が不可欠。
- EHDENは過去のEU-ADR・EMIFプロジェクトの課題を改善し、SMEによる技術支援でスケール可能な仕組みを確立。
- 大規模国際共同研究(例:COVID-19関連のAESI発生率研究)や規制機関(DARWIN EU)などで実際に活用されている。
- 今後はデータパートナーの継続支援、さらなるデータ拡充、研究の実装強化が課題。



EHDENにおけるリアルワールドエビデンス創出の推進:記述的研究

論文9

[Advancing Real-World Evidence Through a Federated Health Data Network (EHDEN): Descriptive Study]

6. 結論(Conclusions)

- EHDENは欧州最大のフェデレーション型RWDネットワークとなり、標準化・品質管理されたデータに基づく研究を実現。
- 今後はEHDEN Foundation が中心となり、より高品質なエビデンス創出と政策・規制での活用を推進していく。



PheCatcher: LLM が生成した合成データを活用した、生物医学文献からの自動フェノタイプ定義抽出

論文10

[PheCatcher: Leveraging LLM-Generated Synthetic Data for Automated Phenotype Definition Extraction from Biomedical Literature]

Yan Hu, Na Hong, Yiming Li, Xueqing Peng, Yong Chen, Hua Xu

1. McWilliams School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX, USA
2. Department of Biomedical Informatics and Data Science, Yale School of Medicine, Yale University, New Haven, CT, USA
3. Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA



PheCatcher: LLM が生成した合成データを活用した、生物医学文献からの自動フェノタイプ定義抽出

[PheCatcher: Leveraging LLM-Generated Synthetic Data for Automated Phenotype Definition Extraction from Biomedical Literature]

- 要旨 (Abstract)
- フェノタイプ定義は、精密医療および個別化医療の発展に不可欠である。
- PheKB や OHDSI ライブラリといったフェノタイプ知識ベースは存在するものの、依然として多くが手作業に依存している。
- 本研究では、生物医学文献からフェノタイプと標準化コードを自動抽出するために、BiomedBERT ベースの固有表現抽出 (NER) と関係抽出 (RE) を統合した自動パイプライン PheCatcher を提案する。
- 人手アノテーションを補完するために GPT-4 を用いて合成データを生成し、モデル性能を向上させた。
- その結果、NER モデルの「フェノタイプ」エンティティに対する F1 スコアは 0.616 から 0.800 に向上し、RE モデルは 0.901 の F1 スコアを達成した。さらに、このパイプラインを PubMed Central (PMC) の論文に適用したところ、173,283 件のフェノタイプ定義を抽出することに成功し、すでに公開されている。
- 本研究は、情報抽出 (IE) における合成データの有用性を示すとともに、合成データを活用して IE システム全体を構築できることを示した初めての証拠を提供するものである。



Global/APACの動き



10月の OHDSI Global/APAC

●Global Community Call テーマ

- Nov 4 OHDSI 2025 Best Community Contribution Honorees
- Nov 11 OHDSI 2025 Software Demos
- Nov 18 DARWIN EU® Initiative Update
- Nov 25 Early-Stage Researcher Spotlight

●APAC Call テーマ <https://ohdsi.org/apac/>

- Nov 6 Generative AI
OHDSI News
Showcase an OHDSI use case of using generative AI for phenotyping by Song Gyeol
Generative AI-based OMOP mapping with HDR UK's Lettuce tool by HDR UK team
Explorations of the Lettuce Generative AI tool for OMOP mapping by A*STAR Data Management Group
- Nov 20 Community Call:



Call for Volunteers

- APAC Communications Team