

— — OHDSI

Observational Health Data Sciences and Informatics (OHDSI)

2025-03-10

Contents

| | |
|------------------------------------|----|
| 序章 | ix |
| この本の目標 | ix |
| 本書の構成 | ix |
| 貢献者 | ix |
| ソフトウェアのバージョン | x |
| ライセンス | x |
| 本書が作成された方法 | x |
| I OHDSI コミュニティ | 1 |
| 1 OHDSI コミュニティ | 3 |
| 1.1 データからエビデンスへの旅 | 3 |
| 1.2 観察医療アウトカムパートナーシップ (OMOP) | 3 |
| 1.3 オープンサイエンスの協働組織としてのOHDSI | 4 |
| 1.4 OHDSIの進展 | 5 |
| 1.5 OHDSIにおける協力 | 6 |
| 1.6 まとめ | 6 |
| 2 どこから始めようか | 7 |
| 2.1 旅に参加しよう | 7 |
| 2.2 どこにフィットするか | 11 |
| 2.3 まとめ | 11 |
| 3 オープンサイエンス | 13 |
| 3.1 オープンサイエンス | 13 |
| 3.2 オープンサイエンスの実践: the Study-a-Thon | 14 |
| 3.3 オープンスタンダード | 14 |
| 3.4 オープンソース | 14 |
| 3.5 オープンデータ | 14 |
| 3.6 オープンディスコース | 14 |
| 3.7 OHDSIとFAIRガイディングプリンシブルズ | 15 |

| | |
|------------------------------|----|
| II 共通データモデル | 17 |
| 4 共通データモデル | 19 |
| 4.1 デザインの原則 | 19 |
| 4.2 データモデルの規約 | 21 |
| 4.3 CDM標準化テーブル | 23 |
| 4.4 追加情報 | 32 |
| 4.5 まとめ | 32 |
| 4.6 演習 | 32 |
| 5 標準化ボキャブラリ | 35 |
| 5.1 なぜボキャブラリが必要で、なぜ標準化が必要なのか | 35 |
| 5.2 コンセプト | 37 |
| 5.3 関係 | 42 |
| 5.4 階層 | 44 |
| 5.5 内部参照テーブル | 45 |
| 5.6 特別な状況 | 45 |
| 5.7 まとめ | 46 |
| 5.8 演習 | 46 |
| 6 ETL（抽出-変換-読込） | 47 |
| 6.1 はじめに | 47 |
| 6.2 ステップ1: ETLのデザイン | 47 |
| 6.3 ステップ2: コードマッピングの作成 | 55 |
| 6.4 ステップ3: ETLの実装 | 60 |
| 6.5 ステップ4: 品質管理 | 61 |
| 6.6 ETLの規約とTHEMIS | 61 |
| 6.7 CDMおよびETLのメンテナンス | 61 |
| 6.8 ETLに関する最終的な考察 | 62 |
| 6.9 まとめ | 62 |
| 6.10 演習 | 62 |
| III データ解析 | 65 |
| 7 データ解析の使用例 | 67 |
| 7.1 特性評価 | 67 |
| 7.2 集団レベルの推定 | 68 |
| 7.3 患者レベルの予測 | 69 |
| 7.4 高血圧症におけるユースケース | 69 |
| 7.5 観察研究の限界 | 70 |
| 7.6 まとめ | 70 |
| 7.7 演習 | 71 |
| 8 OHDSI 分析ツール | 73 |

| | |
|-------------------------------------|-----|
| 8.1 分析の実装 | 73 |
| 8.2 分析戦略 | 74 |
| 8.3 ATLAS | 74 |
| 8.4 Methods Library | 76 |
| 8.5 展開戦略 | 82 |
| 8.6 まとめ | 83 |
| | |
| 9 SQLとR | 85 |
| 9.1 SqlRender | 86 |
| 9.2 DatabaseConnector | 92 |
| 9.3 CDMへのクエリ | 95 |
| 9.4 クエリ実行時にボキャブラリを使用する | 98 |
| 9.5 QueryLibrary | 99 |
| 9.6 簡単な研究のデザイン | 99 |
| 9.7 SQLとRを使用した研究の実施 | 100 |
| 9.8 まとめ | 105 |
| 9.9 演習 | 105 |
| | |
| 10 コホートの定義 | 107 |
| 10.1 コホートとは? | 107 |
| 10.2 ルールベースのコホート定義 | 108 |
| 10.3 コンセプトセット | 109 |
| 10.4 確率的コホート定義 | 109 |
| 10.5 コホート定義の妥当性 | 110 |
| 10.6 高血圧のコホート定義 | 110 |
| 10.7 ATLASを用いたコホートの実装 | 110 |
| 10.8 SQLを使用したコホートの実装 | 118 |
| 10.9 要約 | 125 |
| 10.1 演習 | 125 |
| | |
| 11 特性評価 | 127 |
| 11.1 データベースレベルの特性評価 | 127 |
| 11.2 コホート特性評価 | 128 |
| 11.3 治療経路 | 128 |
| 11.4 発生率 | 128 |
| 11.5 高血圧症患者の特性評価 | 129 |
| 11.6 ATLASにおけるデータベースの特性評価 | 129 |
| 11.7 ATLASにおけるコホート特性分析 | 132 |
| 11.8 Rでのコホートの特性評価 | 138 |
| 11.9 ATLASにおけるコホート経路分析 | 141 |
| 11.10 ATLASにおける発生率分析 | 144 |
| 11.1 まとめ | 147 |
| 11.1 演習 | 148 |
| | |
| 12 集団レベルの推定 | 149 |

| | |
|--|-----|
| 12.1 コホートメソッド設計 | 149 |
| 12.2 自己対照コホートデザイン | 151 |
| 12.3 症例対照デザイン | 152 |
| 12.4 ケース・クロスオーバーデザイン | 153 |
| 12.5 自己対照症例シリーズデザイン | 153 |
| 12.6 高血圧研究のデザイン | 154 |
| 12.7 ATLASを使用した研究の実施 | 155 |
| 12.8 Rを使用した研究の実施 | 165 |
| 12.9 研究の結果 | 171 |
| 12.1 まとめ | 175 |
| 12.1 演習 | 176 |
| | |
| 13 患者レベル予測 | 177 |
| 13.1 予測課題 | 177 |
| 13.2 データ抽出 | 178 |
| 13.3 モデルの適合 | 179 |
| 13.4 予測モデルの評価 | 183 |
| 13.5 患者レベル予測研究のデザイン | 185 |
| 13.6 ATLASでの研究の実装 | 187 |
| 13.7 Rでの研究実施 | 197 |
| 13.8 アウトカム普及 | 202 |
| 13.9 患者レベル予測機能の追加 | 210 |
| 13.1 まとめ | 210 |
| 13.1 演習 | 211 |
| | |
| IV エビデンスの質 | 213 |
| | |
| 14 エビデンスの質 | 215 |
| 14.1 信頼できるエビデンスの属性 | 215 |
| 14.2 エビデンスの質の理解 | 216 |
| 14.3 エビデンスの質の伝達 | 216 |
| 14.4 まとめ | 216 |
| | |
| 15 データ品質 | 219 |
| 15.1 データ品質問題の原因 | 219 |
| 15.2 一般的なデータ品質 | 220 |
| 15.3 研究特有のチェック | 223 |
| 15.4 実践におけるACHILLES | 223 |
| 15.5 Data Quality Dashboardの実践 | 225 |
| 15.6 特定の研究チェックの実践 | 228 |
| 15.7 まとめ | 230 |
| 15.8 演習 | 230 |
| | |
| 16 臨床的妥当性 | 231 |

| | |
|---|-----|
| 16.1 医療データベースの特性 | 231 |
| 16.2 コホートバリデーション | 231 |
| 16.3 ソースレコード検証 | 232 |
| 16.4 PheEvaluator | 233 |
| 16.5 エビデンスの一般化可能性 | 240 |
| 16.6 まとめ | 240 |
| | |
| 17 ソフトウェアの妥当性 | 243 |
| 17.1 研究コードの妥当性 | 243 |
| 17.2 Methods Libraryのソフトウェア開発プロセス | 244 |
| 17.3 Methods Libraryのテスト | 246 |
| 17.4 まとめ | 246 |
| | |
| 18 方法の妥当性 | 247 |
| 18.1 デザイン特有の診断 | 247 |
| 18.2 推定のための診断 | 248 |
| 18.3 実践におけるメソッド検証 | 252 |
| 18.4 OHDSIメソッド評価ベンチマーク | 258 |
| 18.5 まとめ | 258 |
| | |
| V OHDSI研究 | 261 |
| | |
| 19 研究の段階 | 263 |
| 19.1 一般的なベストプラクティスガイドライン | 263 |
| 19.2 詳細な研究手順 | 264 |
| 19.3 まとめ | 267 |
| | |
| 20 OHDSI ネットワーク研究 | 269 |
| 20.1 OHDSI 研究ネットワークとして | 269 |
| 20.2 OHDSIネットワーク研究 | 269 |
| 20.3 OHDSIネットワーク研究の実行 | 271 |
| 20.4 展望: ネットワーク研究の自動化を利用する | 272 |
| 20.5 OHDSIネットワーク研究のベストプラクティス | 273 |
| 20.6 まとめ | 273 |
| | |
| Appendix | 274 |
| | |
| A 用語集 | 275 |
| | |
| B コホート定義 | 279 |
| B.1 ACE阻害薬 | 279 |
| B.2 ACE阻害薬単剤療法新規ユーザー | 280 |
| B.3 急性心筋梗塞 (AMI) | 283 |
| B.4 血管性浮腫 | 284 |
| B.5 サイアザイド様利尿薬単剤療法の新規ユーザー使用者 | 285 |

| | |
|--|-----|
| B.6 高血圧のための第一選択治療を開始する患者 | 288 |
| B.7 追跡期間が3年以上ある高血圧のための第一選択治療を開始する患者 | 291 |
| B.8 ACE阻害薬の使用 | 291 |
| B.9 アンジオテンシン受容体拮抗薬（ARB）の使用 | 292 |
| B.10 サイアザイドおよびサイアザイド様利尿薬の使用 | 292 |
| B.11 ジヒドロピリジン系カルシウムチャネル遮断薬（DCCB）の使用 | 293 |
| B.12 非ジヒドロピリジン系カルシウムチャネル遮断薬（NDCCB）の使用 | 293 |
| B.13 ベータ遮断薬使用 | 293 |
| B.14 ループ利尿薬使用 | 294 |
| B.15 カリウム保持性利尿薬使用 | 294 |
| B.16 アルファ1遮断薬使用 | 295 |
| C ネガティブコントロール | 297 |
| C.1 ACE阻害薬とサイアザイド・サイアザイド様利尿薬 | 297 |
| D プロトコルテンプレート | 301 |
| E 回答例 | 303 |
| E.1 共通データモデル | 303 |
| E.2 標準化ボキャブラリ | 307 |
| E.3 ETL (Extract-Transform-Load) | 307 |
| E.4 データ分析のユースケース | 308 |
| E.5 SQLとR | 308 |
| E.6 コホートの定義 | 310 |
| E.7 特性評価 | 315 |
| E.8 集団レベルの推定 | 322 |
| E.9 患者レベルの予測 | 326 |
| E.10 データ品質 | 328 |
| E.11 | 329 |
| Bibliography | 331 |
| Index | 343 |

これは、OHDSI コラボレーションについての本です。この本は、OHDSI コミュニティに <http://book.ohdsi.org> から利用でき、常に最新バージョンを表示します。物理的なコピーで原価価格で入手可能です。

この本は、OHDSI の中心的な知識リポジトリとなることを目的としており、OHDSI コミュニティ標準、および OHDSI ツールについて説明します。本書は、OHDSI の初心者とベテラン開発者を対象としたものです。

この本は5つの主要な部に分かれています：

- I. OHDSI コミュニティ
- II. 統一されたデータ表現
- III. データ分析
- IV. エビデンスの質
- V. OHDSI 研究

各部には複数の章があり、各章は次の順序に従います：導入、理論、実践、要約、演習。

各章には1名または複数の章の著者がリストされています。これらは章の執筆を主導した人です。

| | | |
|-------------------|------------------|----------------------|
| Hamed Abedtash | Mustafa Ascha | Mark Beno |
| Clair Blacketer | David Blatt | Brian Christian |
| Gino Cloft | Frank DeFalco | Sara Dempster |
| Jon Duke | Sergio Eslava | Clark Evans |
| Thomas Falconer | George Hripcak | Vojtech Huser |
| Mark Khayter | Greg Klebanov | Kristin Kostka |
| Bob Lanese | Wanda Lattimore | Chun Li |
| David Madigan | Sindhoosha Malay | Harry Menegay |
| Akihiko Nishimura | Ellen Palmer | Nirav Patil |
| Jose Posada | Nicole Pratt | Dani Prieto-Alhambra |
| Christian Reich | Jenna Reps | Peter Rijnbeek |
| Patrick Ryan | Craig Sachson | Izzy Saridakis |
| Paola Saroufim | Martijn Schuemie | Sarah Seager |
| Anthony Sena | Sunah Song | Matt Spotnitz |
| Marc Suchard | Joel Swerdel | Devin Tian |
| Don Torok | Kees van Bochove | Mui Van Zandt |
| Erica Voss | Kristin Waite | Mike Warfe |
| Jamie Weaver | James Wiggins | Andrew Williams |
| Seng Chan You | | |

この本の大部分はOHDSIのオープンソースソフトウェアについてであり、このソフトウェアは

- ACHILLES: バージョン 1.6.6
- ATLAS: バージョン 2.7.3
- EUNOMIA: バージョン 1.0.0
- 方法ライブラリパッケージ: 表 1を参照

この本は Creative Commons Zero v1.0 Universal license に基づいてライセンスされています。



この本は RMarkdown を使用して bookdown パッケージで書かれています。オンラインバージョンは <https://github.com/OHDSI/TheBookOfOhdsiInJapanese/> から自動的に再構築され、継続的に “travis” によって管理されます。定期的に本の状態のスナップショットが取得され、「版 (」

Table 1: 本書で使用されているMethods Libraryのパッケージのバージョン

| パッケージ | バージョン |
|--------------------------|-------|
| CaseControl | 1.6.0 |
| CaseCrossover | 1.1.0 |
| CohortMethod | 3.1.0 |
| Cyclops | 2.0.2 |
| DatabaseConnector | 2.4.1 |
| EmpiricalCalibration | 2.0.0 |
| EvidenceSynthesis | 0.0.4 |
| FeatureExtraction | 2.2.4 |
| MethodEvaluation | 1.1.0 |
| ParallelLogger | 1.1.0 |
| PatientLevelPrediction | 3.0.6 |
| SelfControlledCaseSeries | 1.4.0 |
| SelfControlledCohort | 1.5.0 |
| SqlRender | 1.6.2 |

Part I

OHDSI

Chapter 1

OHDSI

著者：Patrick Ryan & George Hripcsak

集まることは始まりであり、共にいることは進歩であり、共に働くことが成功である
ヘンリー・フォード

1.1

世界中のあらゆる医療現場、大学の医療センターや診療所、規制当局や医療製品メーカー

10年以上もの間、多くの人々が「患者と医療従事者が協力して医療を選択するための最善」(Olsen et al., 2007)。この大志の主たる要素は、日常診療の過程で収集された患者レベル(Olsen et al., 2007)。多くの分野で目覚ましい進歩が遂げられている一方で、私たちはこ

なぜでしょうか？その理由の一つとして、患者レベルのデータから信頼性の高いエビデンス

ソースシステムには、さまざまな患者レベルのデータを収集するさまざまなタイプの観察出発点（ソースデータ）と目的の目的地（エビデンス）とは別に、この課題はそのプロセスまた、観察データネットワークで得られた結果と他の情報源からのエビデンスを統合し、

1.2 OMOP

観察研究におけるコラボレーションの顕著な例として、Observational Medical Outcomes Partnership (OMOP) が挙げられます。OMOPは官民パートナーシップで(Stang et al., 2010)。OMOPは、多様なステークホルダーによるガバナンス体制を確立し

チームは集中型環境と分散型研究ネットワークの両方で、異なる観察データベースにまたがり(Overhage et al., 2012)。OMOPの実験により、異なる医療現場から得られた異なるデータ

OMOPは設立当初からオープンサイエンスのアプローチを採用し、研究デザイン、データモデル、CDMは、医療介入や医療制度政策の比較効果など、より広範な分析事例をサポートする

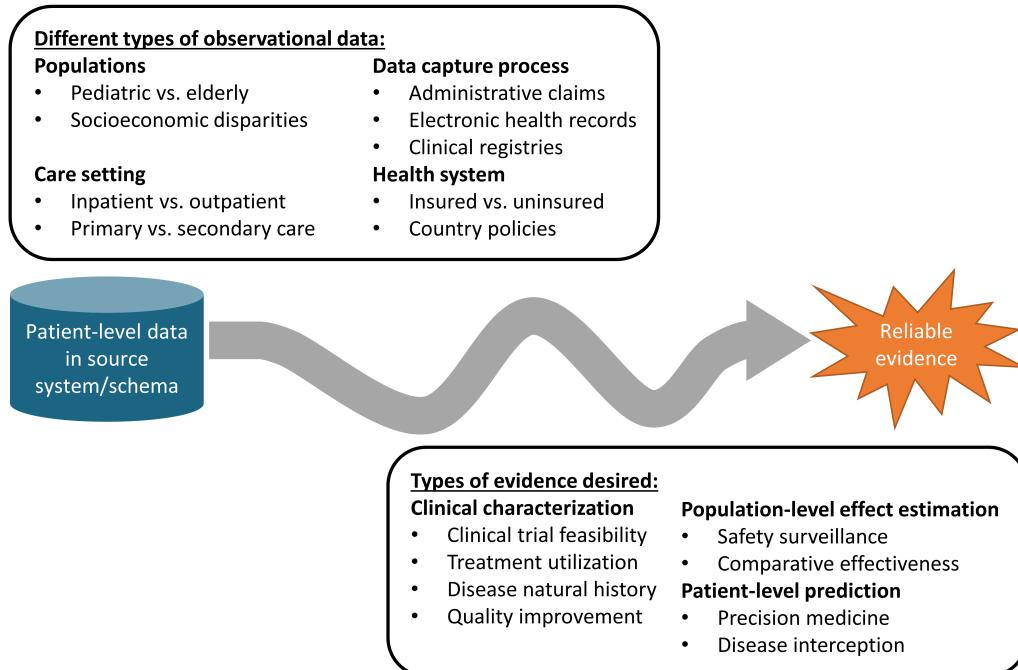


Figure 1.1: データからエビデンスへの旅

OMOPは、大規模な実証実験の完了(Ryan et al., 2012, 2013b)、方法論の革新(Schuemie et al., 2014)、安全性に関する意思決定のための観察データの適切な利用に役立つ知識の創出(et al., 2013b,a)に成功しましたが、OMOPの遺産は、オープンサイエンスの原則を早期に採用

OMOPプロジェクトが完了し、FDAのアクティブサーベイランス活動に情報を提供するための

- ・ オープンコミュニティのデータ標準、標準化ボキャブラリ、ETL（抽出-変換-読込）規約の確立に向けたコラボレーション。これにより、基礎となるデータ品質
- ・ 医薬品の安全性に留まらず、臨床的特性、集団レベルの推定、患者レベルの予測など、
- ・ コミュニティ全体で関心のある重要な健康問題に対処する臨床応用に関するコラボレーション

このような洞察から、OHDSIは誕生しました。

1.3 OHDSI

Observational Health Data Sciences and Informatics (OHDSI、発音は「オデッセイ」) は、 et al., 2015)。OHDSIは、観察医療データの適切な利用に関する科学的ベストプラクティスを確

1.3.1

健康に関する意思決定とケアを向上させるエビデンスを協力して生成することにより、ニ

1.3.2

観察研究によって健康と疾病に関する包括的な理解が得られる世界。

1.3.3

- 革新性: 観察研究は、革新的な恩恵を得ることができる分野です。我々の仕事においても、常に新しい発見や洞察をもたらす可能性があります。
- 再現性: 正確で再現可能な、適切に調整されたエビデンスが健康の改善に不可欠です。他の研究者や医療従事者によって再現可能であることが確認できます。
- コミュニティ: 患者、医療従事者、研究者、そして私たちの活動に賛同する方など、多様な人々が共同で取り組むことで、より広範な解決策を実現できます。
- コラボレーション: 私たちは協力して、コミュニティの参加者の現実的なニーズを理解し、それを実現するための戦略を立てます。
- オープン性: 私たちは私たちが生み出す方法、ツール、生成されたエビデンスなどを公開することで、透明性と信頼性を確保します。
- 有益性: コミュニティ内の個人や組織の権利を常に保護するよう努めています。

1.4 OHDSI

OHDSIは2014年の発足以来、学術界、医療製品業界、規制当局、政府、保険者、技術提携者など、多様な組織による協力によって成長しています。OHDSIの協力者マップ（図1.2）は、国際的なコミュニティの広さと多様性を示しています。



Figure 1.2: 2019年8月現在のOHDSI協力者の地図

¹<https://www.ohdsi.org/who-we-are/collaborators/>

2019年8月現在、OHDSIは20か国以上から100以上の異なる医療データベースのデータネット²を維持するOHDSIが維持するオープンコミュニティデータ標準を用いた分散型ネットワークです。

OHDSIの開発者コミュニティは、OMOP CDMを基盤として、以下の3つのユースケースをサポートします。1) 疾病の自然史、治療実態、品質向上のための臨床的特性評価；2) 医療製品の安全性監視と比較評価；3) 精密医療や疾病予防のための機械学習アルゴリズムを適用する患者レベルの予測。OHDSIの開発者コミュニティは、CDMの採用、データの品質評価、OHDSIネットワーク研究の促進を支援するアプリケーションを開発しています。

OHDSIのオープンサイエンスコミュニティアプローチとオープンソースツールにより、観察研究の論文がAcademy of Scienceに掲載され、2億5000万人以上の患者データを対象とした11のデータソース (Hripcsak et al., 2016)。OHDSIは交絡因子調整のための新しい統計的手法 (Tian et al., 2018) や因果推論のための観察的エビデンスの妥当性評価 (Schuemie et al., 2018a) など、複数の分野でこれらのアプローチを適用しています。てんかん (Duke et al., 2017) から第二選択の糖尿病治療薬の比較効果 (Vashisht et al., 2018) や、うつ病治療の安全性比較に関する大規模な集団レベルの効果推定研究 (Schuemie et al., 2018b) に至るまで、さまざまな分野で適用されています。OHDSIコミュニティ (Reps et al., 2018)、さまざまな治療領域で適用されています (Johnston et al., 2019; Cepeda et al., 2018; Reps et al., 2019)。

1.5 OHDSI

OHDSIはエビデンスを生成するためのコラボレーションを促進することを目的としたコミュニティです。2章（「どこから始めようか」）を参照し、参加方法をご確認ください。

1.6



- OHDSIのミッションは、健康に関する意思決定とケアを向上させるエビデンスを協力して生成します。
- 私たちのビジョンは、観察研究が健康と疾患に関する包括的な理解をもたらす世界を実現することです。
- OHDSIの協力者は、オープンコミュニティのデータ標準、方法論的研究、オープンソースツールを開発・貢献します。

²<https://github.com/OHDSI>

Chapter 2

著者 : Hamed Abedtash & Kristin Kostka

「千里の道も一歩から」 - 老子

OHDSIコミュニティは、学術界、産業界、政府機関といった多くの利害関係者で構成され

2.1

OHDSIには、患者、医療専門家、研究者、あるいは私たちの活動に賛同する人など、誰も

2.1.1 OHDSI

OHDSIフォーラム¹は、OHDSIコミュニティのコラボレーターが投稿メッセージの形で会

OHDSIフォーラムには、次のようなコンテンツのカテゴリがあります：

- ・一般: OHDSIコミュニティに関する一般的なディスカッションと参加方法
- ・実装者: 共通データモデルとOHDSI分析フレームワークをローカル環境に実装する方
- ・開発者: OHDSIアプリケーションや他のOMOP CDMを活用するツールのオープンソース
- ・研究者: CDMベースの研究に関するディスカッション（エビデンス生成、共同研究）
- ・CDMビルダー: 要件、ボキャブラリ、技術的側面を含む進行中のCDM開発に関する議論
- ・語彙ユーザー: ボキャブラリコンテンツに関するディスカッション
- ・地域支部（例：韓国、中国、ヨーロッパ）: OMOP実装やOHDSIコミュニティ活動

自分のトピックを投稿するには、アカウントにサインアップする必要があります。フォームへご登録ください
to OHDSI! - Please introduce yourself (OHDSIへようこそ！-自己紹介をお願いします)
自己紹介と自分の仕事について簡単に教えてください、2) コミュニティでどのように貢献

¹<https://forums.ohdsi.org>

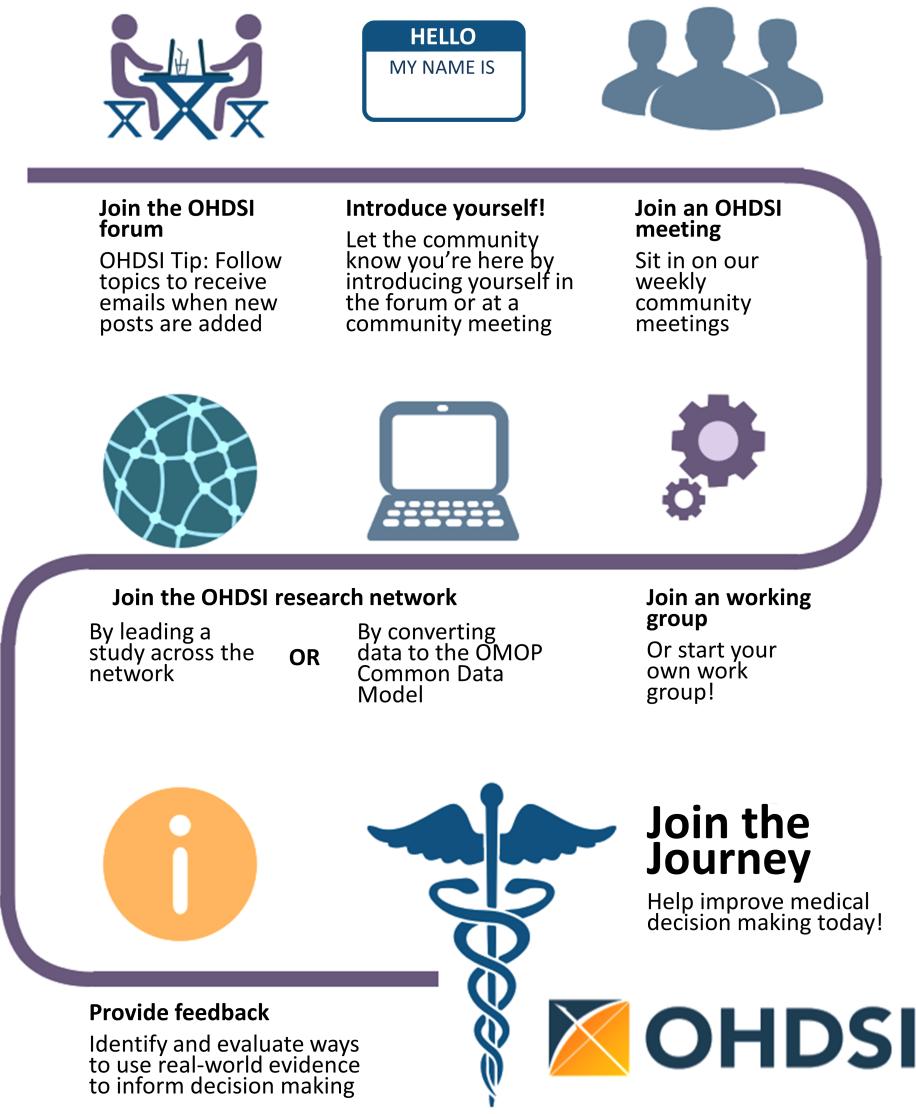


Figure 2.1: 旅に参加しよう — OHDSIのコラボレーターになるには



トピックを選択して「ウォッチ」することができます。ウォッチしているトピック

2.1.2 OHDSI

OHDSIは、コラボレーター同士が学び合い、将来の協力を促進する機会を提供するため、

OHDSIシンポジウムは、米国、ヨーロッパ、アジアで毎年開催される学術会議で、コラボレーターと共同研究者がOHDSI共同研究者がコースの講師となって教えるもので、コミュニティの発展に貢献します。

OHDSIコラボレーターの対面イベントは、通常は共通の関心事の問題に焦点を当てた小規模なa-thonイベントを主催してきました。この複数日にわたるセッションの目的は、適切な範囲でOHDSIコミュニティのパワーをもっと学びましょう。過去のシンポジウム、対面会議を確認して、

2.1.3 OHDSI

OHDSIコミュニティ コールは、OHDSI コミュニティ内で進行中の活動にスポットライトを当てるための定期的な会議です。

OHDSIコラボレーターは、毎週の電話会議への参加を歓迎され、コミュニティディスカウントや情報交換の場として利用されています。

OHDSI コミュニティの新規参加者として、OHDSIネットワーク全体で何が起こっているかを理解するための定期的な会議です。

2.1.4 OHDSI

OHDSIには、ワークグループチームが主導するさまざまな進行中のプロジェクトがあります。各チームがあり、プロジェクトの目的、目標、コミュニティに提供される成果物を決定します。プロジェクトは、Wikiで管理されています。

表2.1は、アクティブなOHDSI作業グループのクイックリファレンスです。是非コールに登録して、最新情報を確認してください。

Table 2.1: 注目すべきOHDSI作業グループ

| ワークグループ名 | 目的 | 対象参加者 |
|----------------|--|---------------------------|
| Atlas & WebAPI | AtlasとWebAPIは、OMOP共通データモデルを基礎としたWebAPI構造を定義するための仕組みです。 | & JavaScript ソフトウェア開発者 |
| CDM & ポキヤブライ | 臨床患者データに適用されるOMOP共通データ標準に基づいた標準化された分析をサポートするための仕組みです。 | |
| ゲノム解析 | OMOP CDMを拡張して、患者のゲノムデータを組み込みます。グループ化された標準化された分析をサポートするための仕組みです。 | すべての人が参加可能 |
| 集団レベルの推定 | 互換スキーマを定義します。 | |
| 自然言語処理 | 正確で信頼性が高く、再現性のある集団参加可能な効果推定についての研究を行います。 | |
| 患者レベルの予測 | OHDSI傘下の観察研究で、AIによる未だ参画していないが使用を促進する複数の対象とするアウトカムによるのが参加可能、対象とする | |

| ワークグループ名 | 目的 | 対象参加者 |
|--|--|--|
| ゴールドスタンダード表現型OHDSIコミュニティのメンバーやタブのキーテレホン検証され検証は FHIR ワークグループ | OHDSI | 相互運用性に関心のあるすべての人が参 FHIR統合のロードマップを確立し、OHDSIベースの観察研究のため |
| GIS | OMOP CD- | 健康関連の地理属性に興味のあるすべて |
| 臨床試験 | Mを拡張し、患者の環境曝露の履歴をその臨床フェノタイプと関連 OHDSIプラットフォームと臨床試験に興味のある面で試験を実施する OHDSI | |
| THEMIS | ツールの更新の推進を支援します。 THEMIS の目的は、OMOP CDM 規則を超える標準規則を開発し、各 OMOP サイトで設計された ETL (抽出-変換-読み込み) | ETL標準化に関心のあるすべての人が参 |
| メタデータ & 注釈 | プロトコルが最高品質で、再現可能かつ効率的であることを保証す 私たちの目標は、人間と機械が作成者が参加可能な注釈を 共通データモデル | |
| 患者生成医療データ (PGHD) | に保存するための標準プロセスを定義し、研究者が観察データセッ この ワーキンググループの目標は、スマートフォン/アプリ/ウェアラブル デバイスから生成される PGHD の ETL | すべての人が参加可能 |
| OHDSI女性グループ 運営委員会 | 規則、臨床データとの統合プロセス、分析プロセスを開発すること OHDSIコミュニティ内の女性がミーティングに参加する技術や人材 OHDSIのすべての活動とイニシアチブがテクノロジーを通じてコミュニティの の使命、ビジョン、価値観を維持します。さらに、このグループは の将来の方向性についてガイダンスを提供することで、コロンビア OHDSI 調整センターの諮問グループとして機能します。 | |

2.1.5 OHDSI

OHDSI 地域支部は、地理的な地域に所在し、地域特有の問題に対処するため、ローカルネットワークイベントや会議を開催したいと考えている OHDSI コラボレーターのグループです。OHDSI 地域支部を設立したい場合は、OHDSI Web サイトで説明されている OHDSI 地域支部のプロセスに従って設立できます。

²<https://www.ohdsi-europe.org/>

³<https://forums.ohdsi.org/c/For-collaborators-wishing-to-communicate-in-Korean>

⁴<https://ohdsichina.org/>

2.1.6 OHDSI

OHDSIのコラボレーターの多くは、データをOMOP共通データモデルに変換することにCDMとボキャブラリに関するチュートリアル、変換を支援する無料で利用可能なツール、

2.2

ここまで読んで、あなたは「私はOHDSIコミュニティのどこに属しているのだろう？」

私は臨床研究者で、研究を始めたいと思っています。特定の質問に答えるためにOHDSIリサーチネットワークを使用したい臨床研究者であるなら、たとえば論文を発表しま

私はOHDSIコミュニティが発信する情報を読んで利用したいと思っています。

患者、臨床医、医療の専門家のいずれであっても、OHDSIは健康アウトカムをよりよく理解するためのツールを提供します。OHDSIがどのようなエビデンスを生成したか、または生成中であるかを知るためにふるい

私は医療のリーダーとして働いています。データ所有者、またはその代表者であるかもしれません。CDMとOHDSI分析ツールの有用性を評価しています。組織の管理者/リーダーとして、また個人として、CDMがあなたのユースケースにどのように役立つかを知りたいと思っているかもしれません。第7章（データ分析のユースケース）を読むと、OMOP CDMやOHDSI分析ツールで実現できることを学ぶことができます。

私はデータベース管理者で、私の機関のデータをETLまたはOMOP CDMに変換したいと考えています。データを「OMOP」することは、斬新で価値のある取り組みです。OMOP CDMの実装を成功に導く支援となる知識がコミュニティには豊富にあります。遠慮しないで質問してください。

私はバイオ統計学者かつ、またはメソッドの開発者で、OHDSIツールスタックへの貢献に興味があります。Renderパッケージの問題であれば、OHDSI/SqlRenderのGitHubリポジトリに提出してみてください。

私はソフトウェア開発者で、OHDSIツールスタックを補完するツールの構築に関心があります。OHDSIコミュニティへようこそ！OHDSIのミッションの一環として、私たちのツールはApache License 2.0で公開されています。

私はコンサルタントで、OHDSIコミュニティに助言したいと考えています。OHDSIコミュニティへようこそ！あなたの専門知識は貴重であり、高く評価されています。必要な情報があれば、OHDSIの対面ミーティングで専門知識を提供して貢献することを検討ください。

私は学生で、OHDSIについてもっと学びたいと思っています。あなたは正しい場所にいるかもしれません。OHDSIのコミュニティは、データ科学、統計学、医療情報学などの多様な背景を持つ人々によって構成されています。

2.3



- OHDSIコミュニティに参加するのは、挨拶するのと同じくらい簡単です。OHDSIのウェブサイトで登録できます。
- 研究やETLに関する質問をOHDSIフォーラムに投稿ください。

Chapter 3

著者 : Kees van Bochove

OHDSIコミュニティの発足当初から、オープンソースソフトウェアの利用、すべての会議記録の公開、また、プライバシーへの配慮が非常に重要であり、通常は正当な理由から公開されない困難な問題について触れてきます。

3.1

「オープンサイエンス」という用語は1990年代から使われていましたが実際に注目を集めました(Wikipedia, 2019a)。ではこれを「科学的研究（出版物、データ、物理的サンプル、ソフトウェア）として称賛しました。実際、この章で詳しく見ていくように、オープンサイエンスの実践

オープンサイエンスまたは「サイエンス2.0」のアプローチ(Wikipedia, 2019b)は、現在の科学的手法における多くの認識された問題に対処することを意味します。hacking、その他の同様の統計的問題に直面し、再現は困難であるという懸念が高まっています³には、この問題の例がいくつか紹介されています。ある分野の論文に系統的な検証を適応(Wikiquote, 2019)。著者らは、ランダム化デザインの不備による統計的有意性について述べた(Allison et al., 2016)。同じ論文集の別の論文では、物理学の経験を例に挙げ、完全な再現性(Chen et al., 2018)。

OHDSIコミュニティはこれらの課題に対して独自の方法で取り組んでおり、大規模な医療データ(Schuemie et al. (2018b))によると、現在のパラダイムは「信頼性が不明な独自の研究データベース」を「標準化されたデータモデル」にマッピングする医療データソースのネットワーク、誰もが利用できる形で公開されている疾患発生状況などの大規模なベースラインデータの組み合わせによって構成される

¹<https://ohdsi.github.io/TheBookOfOhdsi/OpenScience.html#fn17>

²<https://www.ehden.eu/webinars/>

³<https://www.nature.com/collections/prbfkwmwvz>

3.2 : the Study-a-Thon

コミュニティにおける最近の動きとして、「study-a-thon」の出現が挙げられます。これは、a-thonの時間の多くは、統計的アプローチ（第2章参照）、データソースの適合性、インタラクティブな開発環境を用いたデータ分析などです。⁴ a-thonの場合は、さまざまな人工膝関節置換術の術後の有害作用の研究に焦点が当てられ、study-a-thonの期間中にOHDSIフォーラムとツールを使用してインタラクティブに結果が発表されました。⁵ 第20章（OHDSIの開発）⁶ および第21章（OHDSIの利用）⁷ に詳しく説明されています。ATLASなどのOHDSIツールは、コホート定義の迅速な作成、交換、議論、テストを行なうためのツールです。

3.3

OHDSIコミュニティで維持されている非常に重要なコミュニティリソースは、OMOP共通データモデル（第4章参照）と関連する標準ボキャブラリ（第5章参照）です。このモデル自体は観察医療データモデル（第7章参照）。しかし、世界中のさまざまなコーディングシステム、医療パラダイム、さまざまなデータモデルが標準化されています。OHDSIは、OHDSIの標準化されたコーディングシステム（Garza et al., 2016）。

3.4

OHDSIコミュニティが提供するもう一つの重要なリソースはオープンソースのプログラムです（第6章参照）、広く使用されている統計手法の強力なスイートを含むOHDSIメソッドライブラリ（第8章参照）などがあります。オープンサイエンスの観点から、最も重要なリソースの一つは、OHDSIの実行コード（第20章参照）です。これらのプログラムは、GitHubを介して公開されています。OHDSIの実行コードは、OHDSIのデータモデル（第17章）を用いてデータを処理するためのスクリプトやプログラムです。

3.5

医療データはプライバシーセンシティブな性質を持つため、完全にオープンで包括的な患者レコードを公開することは困難です。しかし、OHDSIは、OHDSIのデータモデル（第6章）⁸ またはOHDSIの実行コード（第20章）⁹ を活用して、データをOMOPにマッピングした利用可能なCDMの間のマッピングを透明化するため、データソースがOHDSI ETLまたは「マッピング」ツールによってOHDSIのデータモデルにマッピングされることがあります。

3.6

オープンスタンダード、オープンソース、オープンデータは素晴らしい資産ですが、それだけではありません。

⁴<https://youtu.be/X5yu0JoL6xs>

⁵<https://www.ema.europa.eu/en/events/common-data-model-europe-why-which-how>

⁶<https://forums.ohdsi.org>

⁷<https://www.ohdsi.org/web/wiki>

⁸<https://www.ohdsi.org/web/wiki/doku.php?id=projects:overview>

⁹<https://github.com/ohdsi>

コミュニティ全体でオープンな姿勢とオープンな議論が奨励されており、この本もまさにwiki、コミュニティコール、GitHubリポジトリによって促進されたオープンなプロセスのコラボレーターなしには、プロセスやツールは空虚な殻にすぎないことを強調しておくれ¹⁰。1で議論したように、コラボレーションとオープンサイエンスを通じて健康を改善すると

3.7 OHDSI FAIR

3.7.1

この章の最後の段落では、Wilkinson et al. (2016) で発表されたFAIR原則でOHDSIコミュニティとツールの現状を概観します。

3.7.2

OMOPにマッピングされ、分析に用いられる医療データベースは、科学的観点から、将来(Oliveira et al., 2019)。このアプローチは、IMI EHDENプロジェクトでさらに発展していく

3.7.3

OMOPマッピングされたデータのオープンプロトコルを介したアクセスは、通常、OMOP CDMと組み合わせたSQLインターフェースを通じて実現され、OMOPデータへのアクセス。EHDENのようなプロジェクトの活発な研究テーマであり、運営目標でもあります。しか

3.7.4

相互運用性は、OMOPデータモデルとOHDSIツールの強みであるといえるでしょう。エビデンス生成ツールとしてFHIR、HL7 CIMI、openEHRなどの医療業務における相互運用性標準規格との整合により、Athenaは重要なツールです。これらのツールは、他の利用可能な医療用コードシステム。

3.7.5

再利用に関するFAIR原則は、データライセンス、データの由来（データの発生経緯の明確化）とデータの再現性である。データライセンスは、特に管轄区域をまたぐ場合、複雑なトピックであり、本書で詳しく説明する。ETLツールは現在、この情報を自動的に生成していませんが、データ品質作業グループやCDMにこのメタデータを体系的に添付する方法を検討することは、メタデータ作業部会の議題



- OHDSIコミュニティは、医療におけるエビデンス生成の相互運用性と再現性を重視している。
- また、単一の研究と単一の推定値による医療研究から、大規模な体系的なエビデンス生成を目指している。

¹⁰<https://github.com/OHDSI/TheBookOfOhdsi>

¹¹<https://emif-catalogue.eu>

Part II

Chapter 4

著者: Clair Blacketer

観察データは、患者が医療を受ける際に起こる出来事を示すものです。このデータは世界
1) 直接的に研究を支援するため（よくあるのは調査データや登録データの形で）、2)
医療の提供をサポートするため（いわゆるEHR - 電子的健康記録）、3)
医療の費用を管理するため（いわゆる保険請求データ）。この3つの目的はすべて臨床研

なぜ観察医療データに共通データモデルが必要なのでしょうか？

それぞれの主要なニーズに応じて、観察データベースが臨床イベントをすべて均等に捉え
この標準を提供するのが共通データモデル（CDM）です。標準化された内容（第5章参照）
CDMのすべてのテーブルの概要は、図 4.1に示されています。

4.1

CDMは、以下の目的のために最適化されています。

- ・特定の医療介入（薬物曝露、処置（プロシージャー）、医療政策の変更など）やア
- ・人口統計情報、疾患の自然経過、医療提供、利用と費用、併存疾患、治療や治療の
- ・個々の患者でアウトカムが発生する可能性を予測する——第13章参照。
- ・これらの介入が集団に及ぼす影響を推定する——第12章参照。

この目標を達成するために、CDMの開発は以下のデザイン要素に従います：

- ・目的適合性: CDMは、医療従事者または保険者の運用ニーズを満たす目的ではなく、
- ・データ保護: 名前や正確な誕生日など、患者の身元や保護を危うくする可能性のある
- ・ドメインの設計: ドメインは、各レコードに最低限、個人の識別情報と日付が記録さ

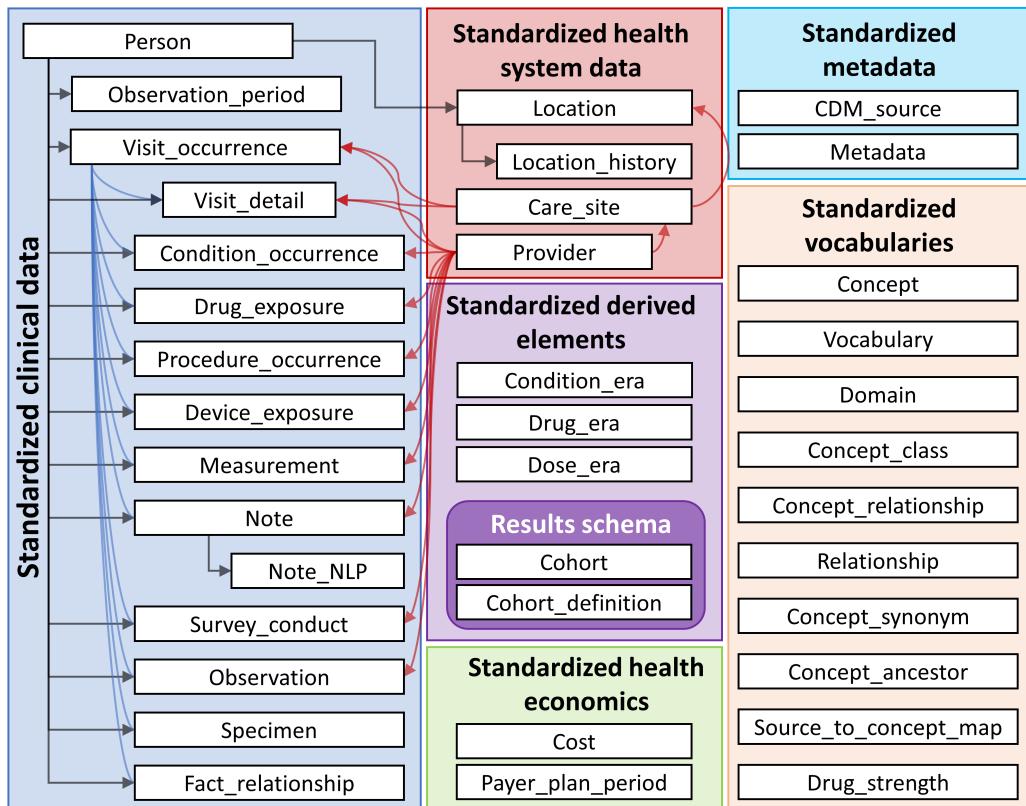


Figure 4.1: CDMバージョン6.0のすべてのテーブルの概要。テーブル間のすべての関係が示されています。

- ・ ドメインの根拠: ドメインは、分析のユースケースがある場合（たとえばコンディショナリレーションシップモデルで特定され、個別に定義されます。他のデータはすべて属性-値構造のオブザベーション（観察）テーブルに保持できます。
- ・ 標準化ボキャブラリ: それらの記録の内容を標準化するために、CDMは、標準的な既存のボキャブラリの再利用: 可能な場合、国立医学図書館、退役軍人省、疾病予防・ソースコードの保持: すべてのコードが標準化ボキャブラリにマッピングされている。
- ・ 技術の中立性: CDMは特定のテクノロジーを必要としません。Oracle、SQL Serverなどのあらゆるリレーショナルデータベース、またはSAS分析データセット。
- ・ スケーラビリティ: CDMは、データベースに含まれる何億人もの人々や何十億件もの後方互換性: これまでのCDMからの変更はすべてgithubリポジトリ(<https://github.com/HL7/CDM>)に記載されています。

4.2

CDMでは、暗黙的および明示的な規約が数多く採用されています。CDMに対応するメソッド

4.2.1

CDMは「人中心」のモデルと見なされており、すべての臨床イベントのテーブルがPERSON

4.2.2

スキーマ（または一部のシステムではデータベースユーザー）により、読み取り専用テーブル

4.2.3

CDMはプラットフォームに依存しません。データ型はANSI SQLデータ型（VARCHAR、INT

注意: データモデル自体はプラットフォームに依存しませんが、それに対応するために構成

4.2.4

異なる性質のイベントはドメインに整理されています。これらのイベントはドメイン固有（5.2.3 参照）。各標準コンセプトには一意のドメイン割り当てがあり、どのテーブルに記載

Table 4.1: 各ドメインに属する標準コンセプトの数

| コンセプト数 | ドメインID | コンセプト数 | ドメインID |
|---------|-----------|--------|------------|
| 1731378 | 薬剤 (Drug) | 183 | 経路 (Route) |

| コンセプト数 | ドメインID | コンセプト数 | ドメインID |
|--------|-----------------------------------|--------|-------------------------------------|
| 477597 | デバイス (Device) | 180 | 通貨 (Currency) |
| 257000 | プロシージャー (Procedure) | 158 | 支払者 (Payer) |
| 163807 | コンディション (Condition) | 123 | ビジット (受診期間) (Visit) |
| 145898 | オブザベーション (Observation) | | 費用 (Cost) |
| 89645 | メジャーメント (測定) (Measurement) | | 人種 (Race) |
| 33759 | 特定の解剖学的部位 (Spec Anatomic Site) | 13 | プランの中断理由 (Plan Stop Reason) |
| 17302 | 測定値 (Meas Value) | 11 | プラン (Plan) |
| 1799 | 試料 (Specimen) | 6 | エピソード (Episode) |
| 1215 | 医療従事者専門 (Provider Specialty) | 6 | スポンサー (Sponsor) |
| 1046 | 単位 (Unit) | 5 | 測定値符号 (Meas Value Operator) |
| 944 | メタデータ (Metadata) | 3 | 特定の疾患ステータス (Spec Disease Status) |
| 538 | 収益コード (Revenue Code) | 2 | 性別 (Gender) |
| 336 | タイプコンセプト (Type Concept) | 2 | 民族性 (Ethnicity) |
| 194 | 関係性 (Relationship) | 1 | オブザベーションタイプ (Observation Type) |

4.2.5

CDMデータテーブル内の各レコードのコンテンツは完全に正規化され、コンセプトを通じて表 CONCEPTテーブルのレコードには、各コンセプトの詳細情報（名前、ドメイン、クラスなど 5章を参照）。

4.2.6

すべてのテーブルの変数名は1つの規約に従います。

Table 4.2: フィールド名の規約

| 記法 | 説明 |
|---------------------------|---|
| [Event]_ID | 各レコードの固有の識別子で、イベントテーブル |
| [Event]_CONCEPT_ID | CONCEPT参照テーブルの標準コンセプトレコード = 31967には- |
| [Event]_SOURCE_CONCEPT_ID | NOMEDコンセプトの「吐き気」の参照値が含まれる CONCEPT参照テーブルのレコードへの外部キー = |
| [Event]_TYPE_CONCEPT_ID | 45431665は「吐き気」というコンセプトを、F |
| [Event]_SOURCE_VALUE | CTコンセプト31967です。標準分析アプリケーション |
| | ソース情報の出所を示す標準化ボキャブラリ内に ソースデータでこのイベントがどのように表現される 9コード787.02に対応する「78702」のレコード |

4.2.7

多くのテーブルには、ソース値、ソースコンセプト、標準コンセプトとして、複数の場所で

- ソース値は、ソースデータにおけるイベントレコードのオリジナル表現です。これ
- コンセプトは、CDM特有のエンティティであり、臨床事実の意味を標準化します。
- ソースコンセプトは、ソースで使用されるコードを表すコンセプトです。ソースコンセプト
- 標準コンセプトは、ソースで使用されるコードシステムとは無関係に、すべてのデ

ソース値は、便宜上、品質保証 (QA) の目的でのみ提供されます。ソース値には、特定のコンディション「肺結核」 (TB) のICD9CMコードは011で示されています。

文脈がなければ、コード011は、UB04ボキャブラリでは「病院入院患者（メディケアパトロニー）」と表示されます。この場合、ソースと標準の両方のコンセプトIDが役立ちます。ICD9CMの011を示すTBのソースコンセプトは、図4.3に示されているように、「非標準から標準へのマッピング」ボキャブラリーから Standard Concept 253954にマップされます。この同じマッピングコードなどにも存在するため、SNOMED 標準コンセプトを参照するあらゆる検索は、サ

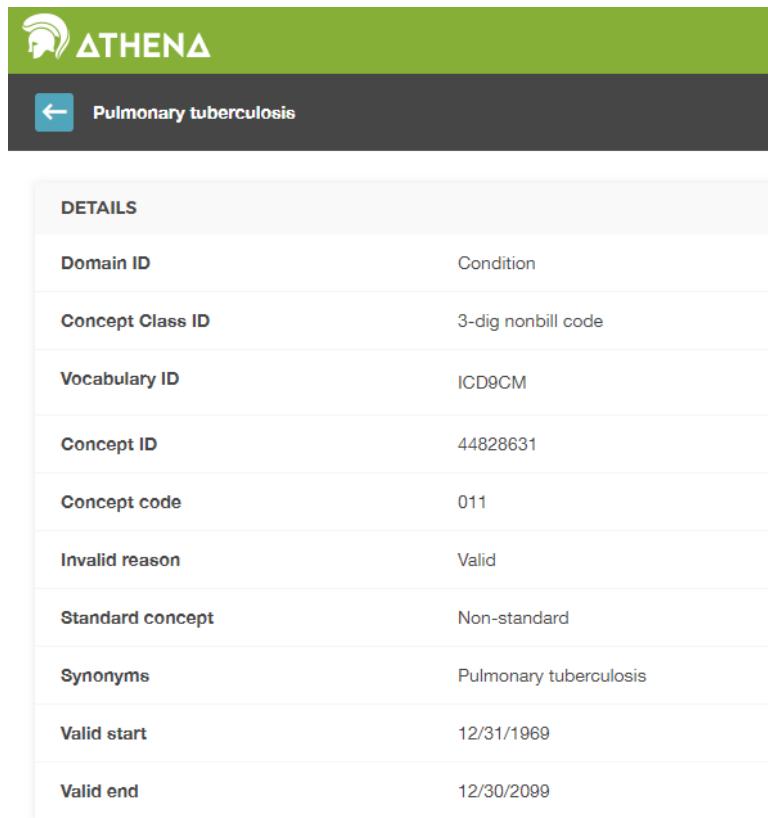
標準コンセプトとソースコンセプトとの関係の例を表4.7に示します。

4.3 CDM

CDMには16の臨床イベントテーブル、10のボキャブラリテーブル、2つのメタデータテーブルがあります。これらのテーブルはCDM Wikiで完全に指定されています¹。

これらのテーブルが実際にどのように使用されるかを説明するために、本章の残りの部分

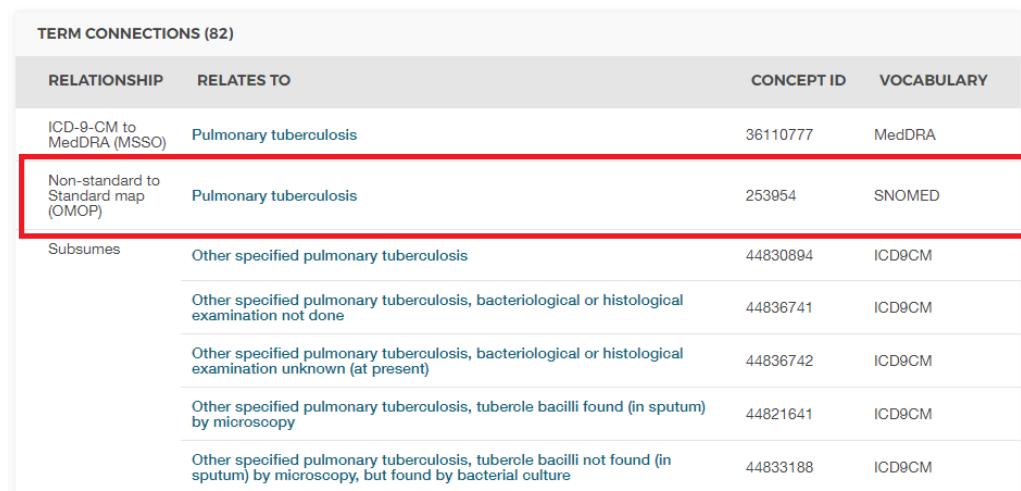
¹<https://github.com/OHDSI/CommonDataModel/wiki>



The screenshot shows the ATHENA interface with the title "Pulmonary tuberculosis". Below it is a table titled "DETAILS" containing the following information:

| | |
|------------------|------------------------|
| Domain ID | Condition |
| Concept Class ID | 3-dig nonbill code |
| Vocabulary ID | ICD9CM |
| Concept ID | 44828631 |
| Concept code | 011 |
| Invalid reason | Valid |
| Standard concept | Non-standard |
| Synonyms | Pulmonary tuberculosis |
| Valid start | 12/31/1969 |
| Valid end | 12/30/2099 |

Figure 4.2: 肺結核のICD9CMコード



The screenshot shows the "TERM CONNECTIONS (82)" section of the ATHENA interface. It lists connections between ICD-9-CM and SNOMED terms. One connection is highlighted with a red border:

| RELATIONSHIP | RELATES TO | CONCEPT ID | VOCABULARY |
|-------------------------------------|--|------------|------------|
| ICD-9-CM to MedDRA (MSSO) | Pulmonary tuberculosis | 36110777 | MedDRA |
| Non-standard to Standard map (OMOP) | Pulmonary tuberculosis | 253954 | SNOMED |
| Subsumes | Other specified pulmonary tuberculosis | 44830894 | ICD9CM |
| | Other specified pulmonary tuberculosis, bacteriological or histological examination not done | 44836741 | ICD9CM |
| | Other specified pulmonary tuberculosis, bacteriological or histological examination unknown (at present) | 44836742 | ICD9CM |
| | Other specified pulmonary tuberculosis, tubercle bacilli found (in sputum) by microscopy | 44821641 | ICD9CM |
| | Other specified pulmonary tuberculosis, tubercle bacilli not found (in sputum) by microscopy, but found by bacterial culture | 44833188 | ICD9CM |

Figure 4.3: 肺結核のSNOMEDコード

4.3.1 :

子宮内膜症は、通常女性の子宮内膜にある細胞が体の他の場所に生じる痛みを伴う状態です。



この痛みを伴う旅のすべての段階で、どれほど痛みを感じているかを皆に納得させ
Laurenは何年も子宮内膜症の症状に悩まされてきましたが、診断を受けるまでには卵巣囊腫
[//endometriosis-uk.org/laurens-story](http://endometriosis-uk.org/laurens-story) をご覧ください。

4.3.2 PERSON

Laurenについてわかっていること

- ・ 彼女は36歳の女性です
- ・ 彼女の誕生日は1982年3月12日です
- ・ 彼女は白人です
- ・ 彼女はイギリス人です

これを踏まえると、彼女のPERSONテーブルは次のようにになります：

Table 4.3: PERSONテーブル

| 列名 | 値 | 説明 |
|----------------------|---------------------|--|
| PERSON_ID | 1 | PERSON_IDはソースから直接である。 |
| GENDER_CONCEPT_ID | 8532 | 女性性別を参照するコンセプトID |
| YEAR_OF_BIRTH | 1982 | |
| MONTH_OF_BIRTH | 3 | |
| DAY_OF_BIRTH | 12 | |
| BIRTH_DATETIME | 1982-03-12 00:00:00 | 時間が不明の場合は真夜中が使用される。 |
| DEATH_DATETIME | | |
| RACE_CONCEPT_ID | 8527 | 白人を示すコンセプトIDは8527である。 |
| ETHNICITY_CONCEPT_ID | 38003564 | これはヒスパニック系の人々を他の人に格納されます。米国以外では使用されない。 |
| ID | | 彼女の住所は不明です。 |
| LOCATION_ID | | 彼女のプライマリケア医療従事者 |
| PROVIDER_ID | | 彼女の主な医療施設は不明です。 |
| CARE_SITE | | |

| 列名 | 値 | 説明 |
|-----------------------------|---------|--|
| PERSON_SOURCE_VALUE | 1 | 通常、これはソースデータでの彼女の読み取り |
| GENDER_SOURCE_VALUE | F | ソースに表示されている性別値がここに格納 |
| GENDER_SOURCE_CONCEPT_ID | 0 | ソースの性別値がOHDSI がサポートするコーディングスキームで F」であり、PCORNet ボキャブラリコンセプトに記載されてい |
| RACE_SOURCE_VALUE | white | ソースに表示されてい人種値がここに格納 |
| RACE_SOURCE_CONCEPT_ID | 0 | 同様にGENDER_SOURCE_CONCEPT_ID |
| ETHNICITY_SOURCE_VALUE | english | ソースに表示されている民族値がここに格納 |
| ETHNICITY_SOURCE_CONCEPT_ID | 0 | 同様にGENDER_SOURCE_CONCEPT_ID |

4.3.3 OBSERVATION_PERIOD

OBSERVATION_PERIODテーブルは、妥当な感度と特異度が期待されるソースシステムにおいて

Laurenの観察期間はどのように定義されているのですか？

Laurenの情報が表4.4に示されているように電子健康記録システムに記録されているとします。

Table 4.4: Laurenのヘルスケアエンカウンター

| エンカウンターID | 開始日 | 終了日 | タイプ |
|-----------|------------|------------|------|
| 70 | 2010-01-06 | 2010-01-06 | 外来患者 |
| 80 | 2011-01-06 | 2011-01-06 | 外来患者 |
| 90 | 2012-01-06 | 2012-01-06 | 外来患者 |
| 100 | 2013-01-07 | 2013-01-07 | 外来患者 |
| 101 | 2013-01-14 | 2013-01-14 | 歩行可能 |
| 102 | 2013-01-17 | 2013-01-24 | 入院患者 |

エンカウンターレコードに基づいて彼女のOBSERVATION_PERIODテーブルは次のようになる

Table 4.5: OBSERVATION_PERIODテーブル

| 列名 | 値 | 説明 |
|-----------------------|------------|----------------------------------|
| OBSERVATION_PERIOD_ID | 1 | これは通常、自動生成された値です。 |
| PERSON_ID | 1 | これはPERSONテーブルでLauraのIDです。 |
| OBSERVATION_PERIOD_ | 2010-01-06 | これは記録上、彼女の最初のエンカウンターの日付です。 |
| START_DATE | | |
| OBSERVATION_PERIOD_ | 2013-01-24 | これは記録上、彼女の最後のエンカウンターの日付です。 |
| END_DATE | | |
| PERIOD_TYPE_ | 44814725 | 「Obs Period Type」コンセプトクラスを持つボーラー |
| CONCEPT_ID | | |

4.3.4 VISIT_OCCURRENCE

VISIT_OCCURRENCEテーブルには、患者が医療システムを利用した際の情報が格納されています。

Laurenのエンカウンターが受診期間としてどのように表現されるか？

例として、ビジットのエンカウンターをVISIT_OCCURRENCEテーブルで表現しましょう。

Table 4.6: VISIT_OCCURRENCEテーブル。

| 列名 | 値 | 説明 |
|---------------------------------|---------------------|--------------------------------|
| VISIT_OCCURRENCE_ID | 514 | これは通常、自動生成された値です。 |
| PERSON_ID | 1 | これはPERSONテーブルでLaurenのIDです。 |
| VISIT_CONCEPT_ID | 9201 | 入院ビジットを参照するキーは9201です。 |
| VISIT_START_DATE | 2013-01-17 | ビジットの開始日です。 |
| VISIT_START_DATETIME | 2013-01-17 00:00:00 | ビジットの日付と時間です。時間部分は00:00:00です。 |
| VISIT_END_DATE | 2013-01-24 | ビジットの終了日です。これは1月24日です。 |
| VISIT_END_DATETIME | 2013-01-24 00:00:00 | ビジットの終了日と時間です。時間部分は00:00:00です。 |
| VISIT_TYPE_ | 32034 | ビジットレコードの出所を示すコードです。 |
| CONCEPT_ID | | |
| PROVIDER_ID | NULL | エンカウンターレコードに医療従事者のIDがNULLです。 |
| CARE_SITE_ID | NULL | エンカウンターレコードに関連する施設IDがNULLです。 |
| VISIT_SOURCE_VALUE | 入院 | ビジット値はソースでどのように表されるかを示す値です。 |
| VISIT_SOURCE_CONCEPT_ID | NULL | ビジット値は、ソースがOHDSIに登録された概念IDです。 |
| ADMITTED_FROM_CONCEPT_ID | NULL | 既知の場合、患者が入院した場所を示すソースがNULLです。 |
| ADMITTED_FROM_SOURCE_CONCEPT_ID | NULL | 患者が入院した元の場所を表すソースがNULLです。 |

| 列名 | 値 | 説明 |
|---------------------------|------|---|
| DISCHARGE_TO_CONCEPT_ID | NULL | 既知の場合、患者が退院した先の場所を表すソース値が含まれます。8615「介護付き生活施設」となります。 |
| DISCHARGE_TO_SOURCE_VALUE | NULL | 患者が退院した場所を表すソース値が含まれます。 |
| PRECEDING_VIS | NULL | 現在のビジットの直前のビジットを示します。 |

- 患者は、入院患者の場合によくあるように、1回の来院中に複数の医療従事者とやりとりwikiを参照ください。

4.3.5 CONDITION_OCCURRENCE

CONDITION_OCCURRENCEテーブルのレコードは、医療従事者によって観察された、または

Laurenの状態は何ですか？

彼女の記録を再確認すると、次のように述べられています。：

約3年前、それまでにも痛かった生理痛がますますひどくなっていることに気づきました。月経痛（月経困難症）のSNOMEDコードは266599000です。表4.7は、それがCONDITION_

Table 4.7: CONDITION_OCCURRENCEテーブル

| 列名 | 値 | 説明 |
|-----------------------------|---------------------|--|
| CONDITION_OCCURRENCE_ID | 964 | これは通常、自動生成された値で、各レコードに1つずつ割り当てられます。 |
| PERSON_ID | 1 | これは、PERSONテーブルのLauraのレコードIDです。 |
| CONDITION_CONCEPT_ID | 194696 | これは、SNOMEDコード266599000を表す外部概念IDです。 |
| CONDITION_START_DATE | 2010-01-06 | コンディションが記録された日付です。 |
| CONDITION_START_DATETIME | 2010-01-06 00:00:00 | コンディションが記録された日時です。 |
| CONDITION_END_DATE | NULL | コンディションが終了したと見なされる日付です。 |
| CONDITION_END_DATETIME | NULL | 既知の場合、コンディションが終了した日時です。 |
| CONDITION_TYPE_CONCEPT_ID | 32020 | この列は、レコードの由来に関する情報です。32020「EHRエンカウンター診断」）というコンセプトIDです。 |
| CONDITION_STATUS_CONCEPT_ID | 4203942 | これが分かると、状況と理由がわかります。 |

| 列名 | 値 | 説明 |
|-------------------------------|-----------|----------------------|
| STOP_REASON | NULL | 既知の場合、ソースデータに示され |
| PROVIDER_ID | NULL | コンディションレコードに診断を付 |
| VISIT_OCCURRENCE_ID | 509 | IDがこのフィールドに入ります。これ |
| CONDITION_SOURCE_VALUE | 266599000 | コンディションが診断されたビジ |
| CONDITION_SOURCE_CONCEPT_ID | 194696 | これはコンディションを表す元の |
| CONDITION_STATUS_SOURCE_VALUE | 0 | ソースからのコンディションの値か |
| CONDITION_STATUS_SOURCE_VALUE | 0 | OHDSIで認識されるボキャブラリを使用 |
| CONDITION_STATUS_SOURCE_VALUE | 0 | がここに入ります。月経困難症の |
| CONDITION_STATUS_SOURCE_VALUE | 0 | コードなので、そのコードを表す |
| CONDITION_STATUS_SOURCE_VALUE | 194696 | 194696 |
| CONDITION_STATUS_SOURCE_VALUE | 0 | です。この場合、CONDITION_CO |
| CONDITION_STATUS_SOURCE_VALUE | 0 | フィールドと同じ値になります。 |
| CONDITION_STATUS_SOURCE_VALUE | 0 | もしソースからのコンディション |
| CONDITION_STATUS_SOURCE_VALUE | 0 | OHDSI |
| CONDITION_STATUS_SOURCE_VALUE | 0 | がサポートするコード化スキーム |

4.3.6 DRUG_EXPOSURE

DRUG_EXPOSUREテーブルは、患者の体内への薬剤の意図的使用または実際の導入に関する情報

Laurenの薬物への曝露はどのように表現されますか？

月経困難症の痛みを改善するために、Laurenは2010年01月06日の来院時に、375mgのアセトアミノフェンを服用しました。

Table 4.8: DRUG_EXPOSUREテーブル

| 列名 | 値 | 説明 |
|------------------------------|---------------------|----------------------|
| DRUG_EXPOSURE_ID | 1001 | 通常、各レコードの一意な識別子 |
| PERSON_ID | 1 | PERSONテーブルのLaurenのレコ |
| DRUG_CONCEPT_ID | 1127433 | 医薬品のコンセプト。アセトアミ |
| DRUG_EXPOSURE_START_DATE | 2010-01-06 | NDC コードは RxNorm |
| DRUG_EXPOSURE_START_DATETIME | 2010-01-06 00:00:00 | コード 313782 |
| DRUG_EXPOSURE_START_DATETIME | 2010-01-06 00:00:00 | に対応し、コンセプト |
| DRUG_EXPOSURE_START_DATETIME | 2010-01-06 00:00:00 | 1127433を表します。 |
| DRUG_EXPOSURE_START_DATETIME | 2010-01-06 00:00:00 | 薬剤曝露の開始日。 |
| DRUG_EXPOSURE_START_DATETIME | 2010-01-06 00:00:00 | 薬剤曝露の開始日時。時間が不明 |

| 列名 | 値 | 説明 |
|----------------------------|---------------------|--|
| DRUG_EXPOSURE_END_DATE | 2010-02-05 | 薬剤曝露の終了日。様々な情報源によつて異なる形式で記録される場合がある。 |
| DRUG_EXPOSURE_END_DATETIME | 2010-02-05 00:00:00 | 薬剤曝露の終了日時。DRUG_EXPOSURE_END_DATE と併用する。 |
| VERBATIM_END_DATE | NULL | 情報源が実際の終了日を明確に記録していない場合の代替。 |
| DRUG_TYPE_CONCEPT_ID | 38000177 | この欄は、記録の出所に関する情報（保有元）である。 |
| STOP_REASON | NULL | 薬剤の投与が中止された理由。理由には多くの国で処方システムの一部となっています。 |
| REFILLS | NULL | リフィルからの始まります。Lauren のアセトアミノフェンの場合、リフィル NULL です。 |
| QUANTITY | 60 | 最初の処方箋または調剤記録に記録された処方された薬の量。 |
| DAYS_SUPPLY | 30 | 元の処方箋または調剤記録に記録されることは CDM ではまだ標準化されておらず、逐語的に記述されることがあります。 |
| SIG | NULL | このコンセプトは、患者が曝露された薬 4132161 (“Oral (経口)”) が使用されています。 |
| ROUTE_CONCEPT_ID | 4132161 | 製造業者から医薬品の特定の数量または薬剤レコードに処方プロバイダがリストされる場合。 |
| LOT_NUMBER | NULL | VISIT_OCCURRENCE テーブルへの外部キー。 |
| PROVIDER_ID | NULL | A foreign key to the VISIT_DETAIL table during which the Drug was prescribed. |
| VISIT_OCCURRENCE_ID | 509 | ソース・データに表示される医薬品のソース。 |
| VISIT_DETAIL_ID | NULL | 薬剤のソースデータでの値を表すコンセプト。 |
| DRUG_SOURCE_VALUE | 69842087651 | 750264 ND-CODEで” Acetaminophen 325 MG Oral Tablet (アセトアミノフェン 325 MG 経口錠) “を表します。 |
| DRUG_SOURCE_CONCEPT_ID | 750264 | |

| 列名 | 値 | 説明 |
|--------------------|------|-----------------|
| ROUTE_SOURCE_VALUE | NULL | 情報源に詳述されている投与経路 |

4.3.7 PROCEDURE_OCCURRENCE

PROCEDURE_OCCURRENCEテーブルには、医療従事者が診断または治療目的で患者に実施したプロシージャーを記録するための情報が格納されています。

- 医療保険請求データには、実施されたプロシージャーを含む、提供された医療サービス。
- オーダーとしてプロシージャーを取り込む電子カルテ。

Laurenはどのプロシージャーを受けたか?

彼女の記述から、2013-01-14に左卵巣の超音波検査を受け、4x5cmの囊胞があることが分かります。

Table 4.9: PROCEDURE_OCCURRENCEテーブル

| Column name | Value | Explanation |
|---------------------------|---------------------|---|
| PROCEDURE_OCCURRENCE_ID | 1277 | これは通常、各レコードの一意な識別子です。 |
| PERSON_ID | 1 | これはPERSONテーブルのローラーIDです。 |
| PROCEDURE_CONCEPT_ID | 4127451 | 骨盤超音波検査のSNOMEDプロシージャーの概念IDです。 |
| PROCEDURE_DATE | 2013-01-14 | プロシージャーが実施された日付。 |
| PROCEDURE_DATETIME | 2013-01-14 00:00:00 | プロシージャーが行われた日時。時間部分は00:00:00です。 |
| PROCEDURE_TYPE_CONCEPT_ID | 38000275 | このカラムはプロシージャーの記録IDです。38000275 ('EHR order list entry') が使用されています。 |
| MODIFIER_CONCEPT_ID | 0 | これは手技の修飾子を表すコンセプトIDです。例えば、CPT4のプロシージャーが両側で行われる場合のID 42739579 ('両側プロシージャー') が使用されます。 |
| QUANTITY | 0 | オーダーされた、または実施された量です。 |
| PROVIDER_ID | NULL | ProcedureレコードにProviderがない場合にのみNULLです。 |
| VISIT_OCCURRENCE_ID | 740 | 情報がある場合には、これはProcedureレコードのVISIT_occurrence_idです。 |
| VISIT_DETAIL_ID | NULL | 情報がある場合、プロシージャーテーブルからVISIT_detail_idとして取得）。 |

| Column name | Value | Explanation |
|------------------------------|-----------|------------------------------------|
| PROCEDURE_SOURCE_ VALUE | 304435002 | ソース・データに表示されているプロシージャーのソースデータの値を表す |
| PROCEDURE_SOURCE_ CONCEPT_ID | 4127451 | ソース・データに表示される修飾子のソースデータの値を表す |
| MODIFIER_SOURCE_ VALUE | NULL | ソース・データに表示される修飾子のソースデータの値を表す |

4.4

本章では、CDMに用意されている表の一部のみを取り上げ、データの表現方法の例として紹介します。より詳しい情報については、[ウィキサイト²](#)をご覧ください。

4.5



- CDMは広範囲の観察研究活動をサポートするように設計されています。
- CDMは人中心のモデルです。
- CDMはデータの構造を標準化するだけでなく、標準化ボキャブラリを通じてコンテキストを明確にします。
- 完全な追跡可能性を確保するために、ソースコードはCDMで管理されています。

4.6

前提条件

これらの最初の練習問題のために、以前に議論されたCDMテーブルを確認する必要があります。A

演習 4.1. ジョンは1974年8月4日生まれのアフリカ系アメリカ人男性です。この情報をエンコードしてみましょう。

演習 4.2. ジョンは2015年1月1日に現在の保険に加入しました。彼の保険データは2019年7月現在のもので、以下の情報を用いてください。

演習 4.3. ジョンは2019年5月1日にイブプロフェン200 MG経口錠剤（NDCコード：761680-001）を購入しました。この情報をエンコードしてみましょう。

²<https://github.com/OHDSI/CommonDataModel/wiki>

³<http://athena.ohdsi.org/>

⁴<http://atlas-demo.ohdsi.org>

前提条件

最後の3つの課題には、セクション 8.4.5 で説明されているようにR、R-Studio、およびJavaがインストールされていることが前提となります。また、SqlRender

```
install.packages(c("SqlRender", "DatabaseConnector", "remotes"))
remotes::install_github("ohdsi/Eunomia", ref = "v1.0.0")
```

Eunomiaパッケージは、ローカルのRセッション内で実行されるCDM内のシミュレートさ

```
connectionDetails <- Eunomia::getEunomiaConnectionDetails()
```

CDMデータベーススキーマは「main」です。これはCONDITION_OCCURRENCEテーブ

```
library(DatabaseConnector)
connection <- connect(connectionDetails)
sql <- "SELECT *
FROM @cdm.condition_occurrence
LIMIT 1;"
result <- renderTranslateQuerySql(connection, sql, cdm = "main")
```

演習 4.4. SQLとRを使用して、「消化管出血」（コンセプトID192671）のすべてのレコ

演習 4.5. SQLとRを使用して、ソースコードを使用して「消化管出血」のすべてのレコー
10を使用しており、関連するICD-10コードは「K92.2」です。

演習 4.6. SQLとRを使用して、PERSON_ID 61の人物の観察期間を取得してください。

提案される答えは付録 E.1にあります。

Chapter 5

著者: Christian Reich & Anna Ostropolets

OMOP標準化ボキャブラリは、単に「ボキャブラリ」と呼ばれることが多く、データの内
本章では、まず標準化ボキャブラリの主な原則、その構成要素、関連する規則、慣例、お

5.1

医学ボキャブラリの歴史は、中世のロンドンでペストやその他の疾患の流行を管理するた
(“Bill of Mortality”) に遡ります (図 5.1 参照)。

それ以来、分類の規模と複雑性は大幅に拡大し、医療の他の側面、例えばプロシージャー
“International Classification of Disease (ICD)” を作成しています。各国の政府は、ICD

その結果、各国、各地域、医療制度、医療機関は、それぞれ独自の分類法を持つ傾向にあ
Nomenclature of Medicine (SNOMED) や、Logical Observation Identifiers
Names and Codes (LOINC) などの幅広い標準の作成を開始しました。米国では、Health
IT Standards Committee (HITAC) が、SNOMED、LOINC、および薬剤用ボキャブラリで
Coordinator for Health IT (ONC) に推奨しています。

OHDSIは、観察研究のためのグローバルスタンダードであるOMOP CD-
Mを開発しました。OMOP標準化ボキャブラリは、CDMの一部として、主に次の2つの目

- ・ コミュニティで使用されるすべてのボキャブラリの共通リポジトリ
- ・ 研究使用のための標準化とマッピング

標準化ボキャブラリはコミュニティに無料で提供されており、OMOP CD-
Mインスタンスでは必須の参照テーブルとして使用する必要があります。

1660.

A General BILL for this present Year,

Ending the 11th Day of December 1660.

According to the Report made to the King's most excellent Majesty,
By the Company of Parish Clerks of LONDON, &c.

DISEASES and CASUALTIES.

| | | | | | | |
|---|-----------------------------|---|--------------------------|---|------------------|----|
| A | Bortive and Stillborn — — — | 421 | Flox and Small Pox — — — | 1523 | Palsey — — — — — | 17 |
| Aged — — — | 909 | Found dead in the Streets, { | 2 | Plague — — — — — | 36 | |
| Ague and Fever — — — | 2303 | Fields, &c. — — — — — | 2 | Plurify — — — — — | 12 | |
| Apoplexy and Suddenly — — — | 91 | French Pox — — — — — | 51 | Quinny and fore Throat — — — | 21 | |
| Blafted and Planet — — — | 3 | Gout — — — — — | 4 | Rickets — — — — — | 441 | |
| Bleeding and bloody Ifsue — — — | 7 | Grief — — — — — | 13 | Rising of the Lights — — — | 210 | |
| Bloody Flux, Scowring, and } Flux — — — — — | 346 | Griping in the Guts — — — | 253 | Rupture — — — — — | 12 | |
| Burnt and Scalded — — — | 6 | Hanged and made away them- } selves — — — — — | 11 | Scurvy — — — — — | 82 | |
| Cancer, Gangrene and Fistula — — — | 63 | Head-ach and Headmouldshot — — — | 35 | Shot — — — — — | 7 | |
| Canker, fore Mouth and Thrush — — — | 73 | Jaundies — — — — — | 102 | Shingles — — — — — | 1 | |
| Childbed — — — | 226 | Imposthume — — — — — | 105 | Sores, Ulcers, broken and } bruised Limbs — — — — — | 61 | |
| Chrisomes and Infants — — — | 858 | Killed by feveral Accidents — — — | 55 | Spleen — — — — — | 7 | |
| Cold, Cough and Hiccough — — — | 33 | King's Evil — — — — — | 28 | Spotted Fever and Purples — — — | 368 | |
| Colick and Wind — — — | 116 | Lethargy — — — — — | 6 | Starved — — — — — | 7 | |
| Consumption and Tisfick — — — | 2982 | Livergrown — — — — — | 8 | Strangury — — — — — | 22 | |
| Convulsion — — — | 742 | Lunatick and Frenzy — — — — — | 14 | Stopping of the Stomach — — — | 186 | |
| Cut of the Stone and Stone — — — | 46 | Megrims — — — — — | 5 | Surfeit — — — — — | 202 | |
| Dropfy and Tympany — — — | 646 | Measles — — — — — | 6 | Swine Pox — — — — — | 2 | |
| Drowned — — — | 57 | Mother — — — — — | 1 | Teeth and Worms — — — — — | 839 | |
| Executed — — — | 7 | Murthered — — — — — | 7 | Vomiting — — — — — | 8 | |
| Falling Sickness — — — | 4 | Overlaid and Starved at Nurfe — — — — — | 46 | Wen — — — — — | 1 | |

Figure 5.1: 1660年のロンドン死亡報告書には、その時代に知られていた62の疾患の分類シス

5.1.1

標準化ボキャブラリのすべてのボキャブラリは、共通の形式に統合されています。これはCDMワークグループの一部であるOHDSIボキャブラリチームによって構築と運営がなさ
GitHubページ³に投稿して、私たちのリソースを改善するのにご協力ください。

5.1.2

標準化ボキャブラリを得るために、自分でPallasを実行する必要はありません。代わりに

OMOP CDMのボキャブラリをすべて選んで、標準化ボキャブラリテーブルのすべてを含
5.2.6 参照）と非常に一般的な使用法は事前に選択されています。提供元データで使用さ

5.1.3 :

OHDSIは一般に、既存のボキャブラリを採用することを優先します。なぜなら、1) 多く
5.2.10 参照）。現在、OHDSIはタイプコンセプト（例：コンディションタイプコンセプ
Extensionボキャブラリです（セクション 5.6.9 参照）。

5.2

OMOP CDMの臨床イベントはすべてコンセプトとして表現されます。これらはデータレ
5.2 を参照）。

このシステムは包括的であることを意味し、患者の医療体験に関連するすべてのイベント

5.2.1 ID

各コンセプトにはプライマリキーとして使用されるコンセプトIDが割り当てられます。こ

5.2.2

各コンセプトには1つの名称が割り当てられます。名称は常に英語表記です。名称はボキ

5.2.3

各コンセプトにはDOMAIN_IDフィールドにドメインが割り当てられています。これは数
5.2.6 参照）は常に単一のドメインが割り当てられます。ドメインは、臨床イベントやイ

¹<https://github.com/OHDSI/Vocabulary-v5.0>

²<https://forums.ohdsi.org>

³<https://github.com/OHDSI/CommonDataModel/issues>

⁴<http://athena.ohdsi.org>

| | | |
|------------------|---------------------|------------------------------------|
| CONCEPT_ID | 313217 | Primary key |
| CONCEPT_NAME | Atrial fibrillation | English description |
| DOMAIN_ID | Condition | Domain |
| VOCABULARY_ID | SNOMED | Vocabulary |
| CONCEPT_CLASS_ID | Clinical Finding | Class in vocabulary |
| STANDARD_CONCEPT | S | Standard, Source of Classification |
| CONCEPT_CODE | 49436004 | Code in vocabulary |
| VALID_START_DATE | 01-Jan-1970 | |
| VALID_END_DATE | 31-Dec-2099 | |
| INVALID_REASON | | Valid during time interval |

Figure 5.2: OMOP CDMにおける標準化ボキャブラリコンセプトの標準的な表現。提示されて

ドメインの割り当ては、Pallasに示されているヒューリスティックな手法を使用してボキャブ（5.3 参照）。

ドメインのヒューリスティックは、ドメインの定義に従います。これらの定義はCDMのテーブル（4 参照）。ヒューリスティックは完全ではなく、グレーゾーンも存在します（セクション 5.6 「特別な状況」 参照）。ドメインが誤って割り当てられているコンセプトがある場合、フ

5.2.4

各ボキャブラリには短い大文字小文字区別のない一意の英数字IDが割り当てられており、通常9-CMのボキャブラリIDは「ICD9CM」です。現在、OHDSIでサポートされているボキャブラリ

5.2.5

一部のボキャブラリでは、大文字と小文字を区別する固有の英数字IDによって表されるコード

Table 5.1: コンセプトクラスにおける水平および垂直のサブ分類原則を持つボキャブラリ

| コンセプトクラスの区分原則 | ボキャブラリ |
|---------------|------------------------------------|
| 水平 | すべての薬剤ボキャブラリ、ATC、CDT、Episode、 |
| 垂直 | CIEL、HES専門、ICDO3、MeSH、NAACCR、NDFFR |
| 混在 | CPT4、ISBT、LOINC |
| なし | APC、すべてのタイプコンセプト、民族性、OXMIS、 |

水平コンセプトクラスにより、特定の階層レベルを決定することができます。たとえば、医薬品

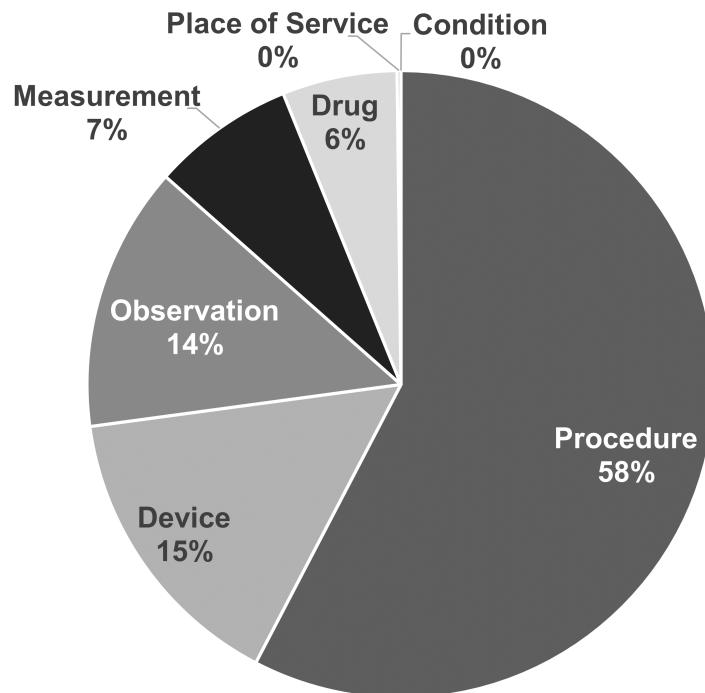


Figure 5.3: プロシージャーボキャブラリCPT4およびHCPCSにおけるドメインの割り当て

5.2.6

各臨床イベントを表す1つのコンセプトが標準として指定されます。例えば、MESHコード

5.2.7

非標準コンセプトは臨床イベントを表現するためには使用されませんが、標準化ボキャブラリ（5.3.1参照）。非標準コンセプトにはSTANDARD_CONCEPTフィールドに値がありません。

5.2.8

これらのコンセプトは標準ではなく、したがってデータを表現するためには使用されません（5.1.2 参照）では、標準のSNOMEDコンセプト「心房細動」を取得します（CONCEPT_ATTRIBUTES 5.4 を参照） - 図 5.4 を参照。

標準、非標準、分類のコンセプトの選択は、通常各ドメインごとにボキャブラリレベルで（5.2.10 参照）し、異なるボキャブラリから同じ意味を持つ複数のコンセプトが競合する場合つまり、「標準ボキャブラリ」というものは存在しません。例については表 5.2 を参照ください。

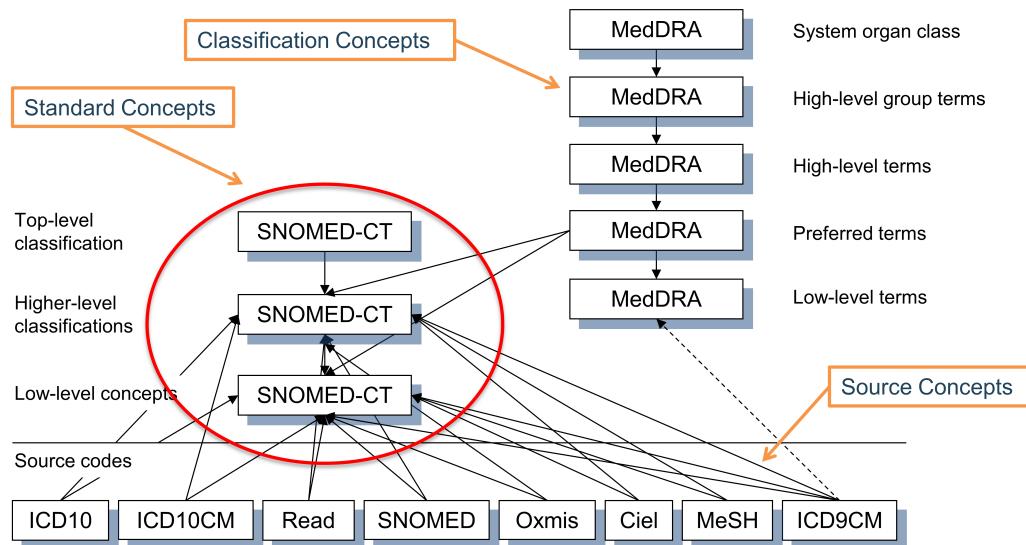


Figure 5.4: コンディションドメインにおける標準、非標準ソースおよび分類コンセプトとその割り当て

Table 5.2: 標準/非標準/分類コンセプトの割り当てに利用するボキャブラリのリスト

| ドメイン | 標準コンセプトのためのボキャブラリ | 分類のためのボキャブラリ | ボキャブラリ |
|-------------|--|--|---------|
| コンディション | SNOMED, ICDO3 | SNOMED Veterinary | MedDRA |
| プロシージャー | SNOMED, CPT4, HCPCS, ICD10PCS, ICD9Proc, OPCS4 | SNOMED Veterinary, HemOnc, NAACCR | 現時点ではなし |
| メジャーメント（測定） | SNOMED, LOINC | SNOMED Veterinary, NAACCR, CPT4, HCPCS, OPCS4, PPI | 現時点ではなし |
| 薬剤 | RxNorm, RxNorm Extension, CVX | HCPCS, CPT4, HemOnc, NAACCR | ATC |
| デバイス | SNOMED | 他のボキャブラリ、 他のボキャブラリ | 現時点ではない |
| オブザベーション | SNOMED | 他のボキャブラリ | 現時点ではなし |
| ビジット | CMS Place of Service, ABMT, NUCC | SNOMED, HCPCS, CPT4, UB04 | 現時点ではなし |

5.2.9

コンセプトコードはソースボキャブラリで使用される識別子です。たとえば、ICD9CMまたは5.3 参照）。

Table 5.3: 同じコンセプトコード1001を持つが、異なるボキャブラリ、ドメインID、コンセプトクラス

| コンセプトID | コンセプトコード | コンセプト名 | ドメインID | ボキャブラリ | コンセプトクラス |
|----------|----------|---|---------|-----------|----------|
| 35803438 | 1001 | 顆粒球コロニ ² 薬刺激因子 | HemOnc | コンポーネントクラ | |
| 35942070 | 1001 | AJCC TNM | メジャー・メン | NAACCR | NAACCR変数 |
| | | Clin T | | | |
| 1036059 | 1001 | アンチビリン薬剤 | RxNorm | 成分 | |
| 38003544 | 1001 | レジデンシヤ収益コード - 精神科 | 収益コード | 収益コード | |
| 43228317 | 1001 | アセプロメタ ² 薬剤マレイン酸塩M | 酸塩M | 成分 | |
| 45417187 | 1001 | ブロムフェニ ² 薬剤ンマレイ M酸塩m 10 mg/ml注射用溶液 | Multum | | |
| 45912144 | 1001 | 血清 | 標本 | CIEL | 標本 |

5.2.10

ボキャブラリは、固定されたコードセットを持つ恒久的なコーパスであることはまれです。CDMは、患者の経時的データをサポートするモデルであり、過去に使用されていたが現

- アクティブまたは新しいコンセプト
 - 説明: 使用中のコンセプト。
 - VALID_START_DATE: コンセプトの生成日。不明の場合はボキャブラリへの取
1-1。
 - VALID_END_DATE: 「将来、定義されていない時点で無効になる可能性がある」。
 - INVALID_REASON: NULL
- 非推奨のコンセプトで後継なし
 - 説明: 非アクティブであり、標準として使用することはできない（セクション
5.2.6 参照）。
 - VALID_START_DATE: コンセプトの生成日。不明の場合はボキャブラリへの取
1-1。
 - VALID_END_DATE: 過去の廃止日。不明の場合はボキャブラリ内のコンセプト
 - INVALID_REASON: “D”
- 後継コンセプトとともにアップグレードされたコンセプト
 - 説明: コンセプトは非アクティブだが、後継コンセプトが定義されています。這
1-1。
 - VALID_START_DATE: コンセプトの生成日。不明の場合はボキャブラリへの取
1-1。
 - VALID_END_DATE: アップグレードが行われた過去の年月日。不明の場合は、
 - INVALID_REASON: “U”
- 別の新しいコンセプトで再利用されたコード
 - 説明: 非推奨のコンセプトコードが、新しいコンセプトで再利用されました。

- VALID_START_DATE: コンセプトの生成日。不明の場合はボキャブラリへの取り込み日
- VALID_END_DATE: 非推奨であることを示す過去の日、またはそれがわからない場合
- INVALID_REASON: “R”

一般に、コンセプトコードは再利用されません。しかし、特にHCPCS、NDC、DRGなど、この値は一意です。これらの再使用されるコンセプトコードは、INVALID_REASONフィールドに

5.3

任意の2つのコンセプトは、そのドメインやボキャブラリーが同じであるかどうかに関係なく、「Maps to」関係には反対の関係「Mapped from」があります。

CONCEPT_RELATIONSHIPテーブルのレコードには、ライフサイクルフィールドRELATIONSHIP_TYPEが付いています。

5.3.1

これらの関係は、非標準のコンセプトから標準コンセプトへの変換を提供し、2つの関係IDペア（Table 5.4 を参照）。

Table 5.4: マッピング関係の種類

| 関係IDペア | 目的 |
|---------------------------------------|--|
| “Maps to” と “Mapped from” | 標準コンセプトはそれ自身にマッピングされ、非標準コンセプトがマッピングされる |
| “Maps to value” と “Value mapped from” | MEASUREMENTとOBSERVATIONテーブルのVALUE |

これらのマッピング関係の目的は、同等のコンセプト間の相互参照を可能にし、臨床イベントCDMでどのように表現されるかを統一することです。これは標準化ボキャブラリの主要な成果物です。

「同等のコンセプト」とは、同じ意味を持ち、さらに重要なことには、階層下位のコンセプトW61.51「ガチョウに噛まれる」は、標準のコンディションコンセプトとして使用されるSNOMEDCT 217716004「鳥に突かれる」にマッピングされ、コンテキストとしての鳥がガチョウである。

一部のマッピングでは、ソースコンセプトが複数の標準コンセプトにリンクされます。たとえば、OMOP CDMテーブルの値を分割することを目的としています。これは次の状況で発生します：

- 検査と結果の値からなるメジャーメント
- 本人または家族の病歴
- 物質に対するアレルギー
- 予防接種の必要性

このような状況では、ソースコンセプトは属性（テストまたは履歴）と値（テスト結果またはオブザーベーション）関係はこのソースを属性コンセプトにマッピングし、「Maps to value」は値コンセプトにマッピングします。Figure 5.5を参照ください。

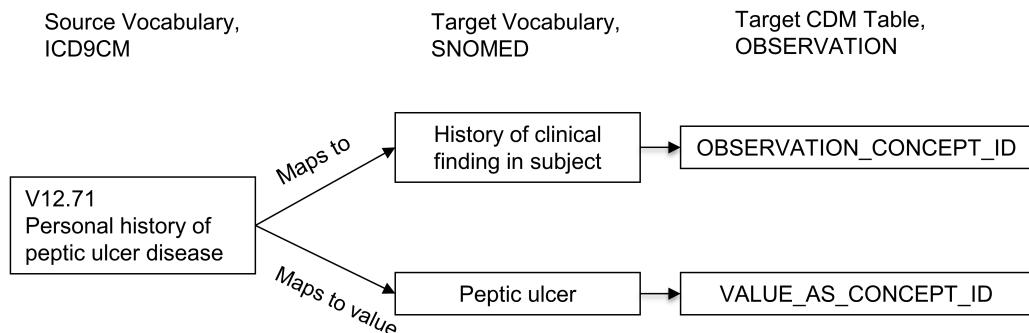


Figure 5.5: ソースコンセプトと標準コンセプト間の一対多のマッピング。事前に組み合った「Maps to」関係はメジャーメントまたはオブザベーションのドメインのコンセプトにマッピングできます。「Maps to value」コンセプトにはドメインの制限はありません。

コンセプトのマッピングは、無料で提供され、ネットワーク研究を行うコミュニティの取扱いです。マッピング規則の詳細な説明は、OHDSI Wikiで見つけることができます⁵。

5.3.2

階層関係は、「Is a」 - 「Subsumes」関係によって定義されます。階層関係は、子コンセプトが親コンセプトを「Subsumes」する関係です。たとえば、ICD-10「心房細動」は、SNOMED「心房性不整脈」と「Is a」関係で関連しています。両コンセプトは、不整脈の種類（一方では細動と定義されています）を示す概念です。ICD-10「心房細動」はSNOMED「心房性不整脈」と「Is a」関係で関連しています。両コンセプトは、不整脈の種類（一方では細動と定義されています）を示す概念です。

5.3.3

これらの関係は通常、「ボキャブラリ A - ボキャブラリ B は同等」というタイプであり、「ボキャブラリ A - RxNorm と同等」）は常に「Maps to」関係によって複製されます。

5.3.4

内部ボキャブラリ間の関係は通常、ボキャブラリの提供者によって提示されます。OHDSI Wikiの個々のボキャブラリの個々のボキャブラリー文書に完全な説明が記載されています。

これらの多くは、臨床イベント間の関係を定義しており、情報検索に使用することができます。たとえば、「site_of (部位の検索)」関係に従うことで検索することができます（表 5.5 を参照）。

⁵<https://www.ohdsi.org/web/wiki/doku.php?id=documentation:vocabulary:mapping>

⁶<https://www.ohdsi.org/web/wiki/doku.php?id=documentation:vocabulary>

Table 5.5: 尿道の「Finding site of」関係で、すべてこの解剖学的構造に位置する状態を示しています。

| CONCEPT_ID_1 | CONCEPT_ID_2 |
|------------------------|--------------------|
| 4000504 “Urethra part” | 36713433 “部分的尿道重複” |
| 4000504 “Urethra part” | 433583 “下部尿道裂孔” |
| 4000504 “Urethra part” | 443533 “男性下部尿道裂孔” |
| 4000504 “Urethra part” | 4005956 “女性下部尿道裂孔” |

これらの関係の質と網羅性は、元のボキャブラリーの質によって異なります。一般に、SNOMED CT

5.4

ドメイン内では、標準および分類コンセプトは階層構造に整理され、CONCEPT_ANCESTOR + CONCEPT_DESCENDANTテーブルは、階層関係を通じてつながっているすべてのコンセプトを網羅する “Is a” - “Subsumes” のペア（図 5.6 参照）であり、ボキャブラリ間の階層を結びつけるそ

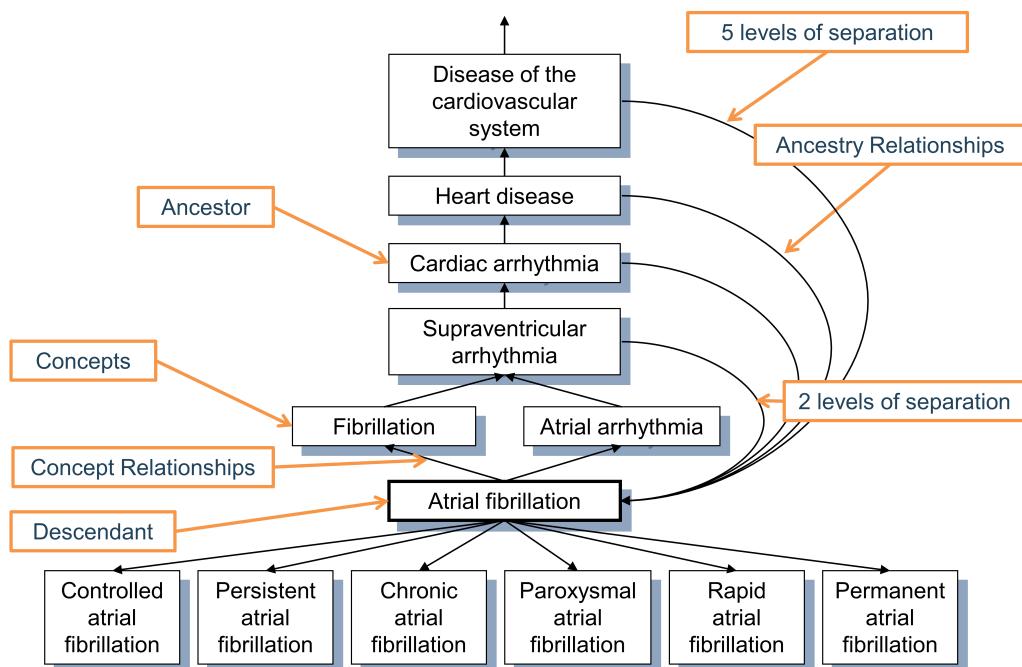


Figure 5.6: 「心房細動」という条件の階層。第一度の先祖関係は「Is a」と「Subsumes」関係によって定義され、それより高次の関係はすべて推論され、CONCEPT_RELATIONSHIPテーブルに記録される。

祖先の度合い、つまり祖先と子孫の間のステップ数は、MIN_LEVELS_OF_SEPARATIONおよびMAX_LEVELS_OF_SEPARATIONによって定義される。現時点では、高品質で包括的な階層は薬剤とコンディションの2つのドメインにのみ存在しま

5.5

DOMAIN_ID、VOCABULARY_ID、CONCEPT_CLASS_ID（すべてCONCEPTレコード内 DOMAIN、VOCABULARY、CONCEPT_CLASS、RELATIONSHIPで定義されており、*_ID フィールドを主キーとして、より詳細な *_NAME フィールドと、CONCEPT テーブルへの参照を持つ *_CONCEPT_ID フィールドを含んでいます。CONCEPT テーブルには、参照テーブルのレコードそれぞれに対応するコンセプトが含まれています。また、VOCABULARYテーブルには、オリジナルのボキャブラリソースとバージョンを参

5.6

5.6.1

OMOP CDMと標準化ボキャブラリにおける性別は、出生時の生物学的性別を意味します。このようなケースは、OBSERVATION テーブルのレコードでカバーする必要があります。

5.6.2

これらは米国政府の定義に従います。民族はヒスパニック系または非ヒスパニック系の区

5.6.3 OMOP

ICD-9やICD-10などの一般に使用されているコーディング体系は、適切な診断評価に基づくコンディションドメインは、このセマンティックスペースと完全に一致するものではありません。例えば、コンディションには診断が下される前に記録される徴候や症状も含まれます。ま

5.6.4

同様に、HCPCSやCPT4のようなコーディングシステムは医療プロシージャーのリストで

5.6.5

医療機器のコンセプトには、標準コンセプトのソースとして使用できる標準化されたコ

5.6.6

ビジットのコンセプトは、医療受診の性質を定義します。多くのソースシステムでは、こ

5.6.7

医療従事者は、医療従事者ドメインで定義されます。これには、医師や看護師などの医療

5.6.8

標準化ボキャブラリは包括的に医療のあらゆる側面をカバーしています。しかし、一部の治療

5.6.9

薬剤ドメインの多くのコンセプトは、米国国立医学図書館が作成した公的に利用可能なボキャブラリExtensionというボキャブラリに追加されます。

5.6.10 NULL

多くのボキャブラリには、情報の欠如に関するコードが含まれています。例えば、5つの性別

5.7



- すべてのイベントと管理上の事実は、OMOP標準化ボキャブラリでコンセプト、コード化されています。
- これらのほとんどは既存のコーディングスキームやボキャブラリから採用されています。
- すべてのコンセプトにはドメインが割り当てられ、そのコンセプトが表す事象がCDER上に定義されています。
- 異なるボキャブラリにおける同等の意味を持つコンセプトは、そのうちの1つにマッピングされています。
- マッピングは「Maps to」および「Maps to value」というコンセプト関係を通じて行われます。
- 分類コンセプトという追加のコンセプトクラスがあり、これらは非標準ですが、ソースコードで明確に示されています。
- コンセプトには時間の経過とともにライフサイクルがあります。
- ドメイン内のコンセプトは階層に整理されています。階層の質はドメインごとに異なります。
- 間違いや不正確さを発見した場合は、コミュニティに積極的に参加することを強く推奨します。

5.8

前提条件

最初の演習では、標準化ボキャブラリのコンセプトを検索する必要があります。これはATHENA⁷で行います。

演習 5.1. “消化管出血” の標準コンセプトIDは何ですか？

演習 5.2. “消化管出血” の標準コンセプトに対応するICD-10CMコードは何ですか？この標準化されたICD-10CMコードはどれですか？

演習 5.3. “消化管出血” の標準コンセプトに相当するMedDRAの優先用語は何ですか？

回答例は付録 E.2を参照のこと。

⁷<http://athena.ohdsi.org/>

⁸<http://atlas-demo.ohdsi.org>

Chapter 6

ETL - -

著者: Clair Blacketer & Erica Voss

6.1

ネイティブ/生データからOMOP共通データモデル（CDM）を作成するには、ETL（抽出変換-読込）プロセスを作成する必要があります。このプロセスでは、データをCDMに再

ETLの作成は通常、大規模な取り組みとなります。長年にわたり、私たちは以下の4つの手順を実行してきました。

1. データの専門家とCDMの専門家が共同でETLをデザインする。
2. 医学的知識を持つ人がコードのマッピングをする。
3. 技術者がETLを実装する。
4. 全員が品質管理に関与する。

本章では、これらのステップをそれぞれ詳しく説明します。OHDSIコミュニティでは、このプロセスを「ETLの4つの手順」と呼んでいます。

6.2 1: ETL

ETLのデザインと実装を明確に区別することが重要です。ETLのデザインにはソースデータベースやターゲットデータベースの構造を理解する必要があります。

ETLデザインプロセスを支援するために密接に統合された2つのツールが開発されました。RabbitとRabbit-in-a-Hatです。

6.2.1 White Rabbit

データベースでETLプロセスを開始するには、データ（テーブル、フィールド、内容など）を読み込む必要があります。White Rabbitは、縦断的な医療データベースのETLをOMOP

CDM用に準備するためのソフトウェアツールです。White Rabbitはデータをスキャンし、ETL

範囲と目的

White Rabbitの主な機能は、ソースデータをスキャンし、テーブル、フィールド、フィールド Server、Oracle、PostgreSQL、Microsoft APS、Microsoft Access、Amazon Redshift) に保存される必要があります。スキャンによって、例えばRabbit-In-a-Hatツールと併用して、ETLデザイン時に参照用として使用できるレポートが作成されます。Rabbitは標準的なデータプロファイリングツールとは異なり、生成された出力データファイル

プロセス概要

ソフトウェアを使用してソースデータをスキャンする一般的な手順は以下の通りです：

1. 作業フォルダを設定します。作業フォルダは、結果がエクスポートされるローカルのディレクトリです。
2. ソースデータベースまたはCSVテキストファイルに接続し、接続をテストします。
3. スキャン対象のテーブルを選択し、テーブルをスキャンします。
4. White Rabbitがソースデータに関する情報をエクスポートします。

作業フォルダの設定

White Rabbitアプリケーションをダウンロードしてインストールした後、まず最初に作業フォルダを設定します。White Rabbitが作成するすべてのファイルはこのローカルフォルダにエクスポートされます。図6.1「Scan Folder (フォルダの選択)」ボタンを使用して、スキャン文書を保存するローカル環境をナビゲートできます。

データベースへの接続

White Rabbitは区切りテキストファイルとさまざまなデータベースプラットフォームをサポートしています。

データベース内のテーブルをスキャン

データベースに接続後、含まれるテーブルをスキャンできます。スキャンによりETLのデザインが完了します。図6.2に示されたスキャンタブを使用して、「Add」(Ctrl+マウスクリック)をクリックして選択肢を確認し、「Select all in DB」ボタンをクリックできます。

スキャンにはいくつかの設定オプションもあります：

- 「フィールド値をスキャン」をチェックすると、列に表示される値を調査したいことをWhite Rabbitに通知します。
- 「最小セル数」はフィールド値のスキャン時のオプションです。デフォルトでは5に設定されています。
- 「テーブルあたりの行数」はフィールド値のスキャン時のオプションです。デフォルトでは100,000行に設定されています。

すべての設定が完了したら、「テーブルをスキャン」ボタンをクリックします。スキャンが完了すると、スキャン結果が表示されます。

¹<https://github.com/OHDSI/WhiteRabbit>.

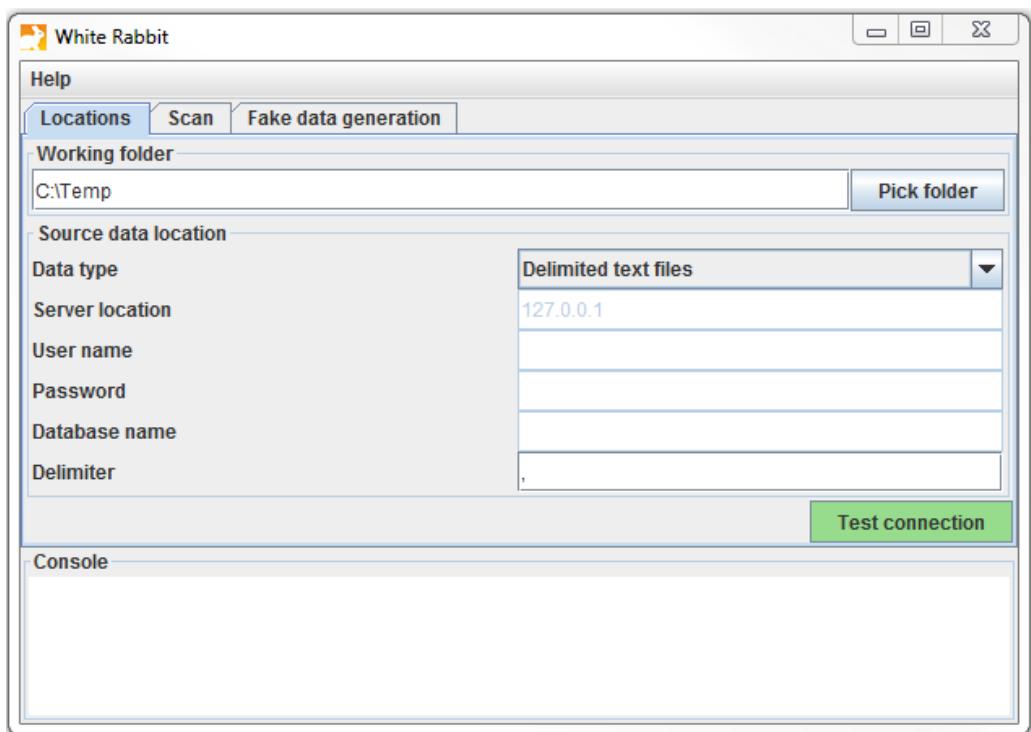


Figure 6.1: White Rabbit アプリケーションの作業フォルダを指定するための「Pick Folder」ボタン

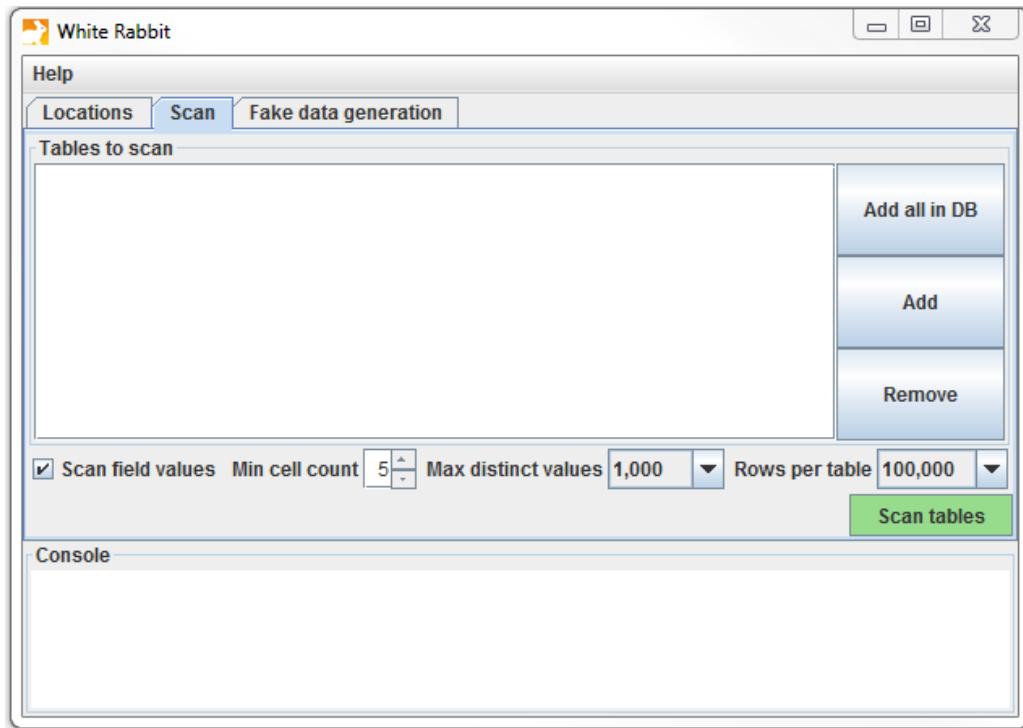


Figure 6.2: White Rabbit スキャンタブ

スキャンレポートの解釈

スキャンが完了すると、選択したフォルダにスキャンしたテーブルごとにタブが設けられたExcelファイルが生成されます。また、概要タブも用意されています。概要タブには、スキャンしたすべてのテーブル、各テーブルの構造、各フィールドの値の頻度などの情報が表示されます。

各テーブルのタブには、各フィールド、各フィールド内の値、各値の頻度が示されます。各ソースデータの構造を理解するのに強力であり、存在するものを強調して表示します。例として、Rabbitは1を男性、2を女性として定義することなく、データホルダーが通常、ソースシステムによって定義される性別コードを表示します。

6.2.2 Rabbit-In-a-Hat

White Rabbitスキャンを手にすると、ソースデータの全体像やCDMの仕様が掴めます。次に、Rabbitソフトウェアと共に提供されるRabbit-in-a-Hatツールは、これらの分野の専門家チームがデータを分析するためのツールです。最初のラウンドでは、テーブル間のマッピングを共同で実行します。

範囲と目的

Rabbit-In-a-HatはWhite Rabbitのスキャン文書を読み取り、表示するように設計されています。Rabbitはソースデータに関する情報を生成し、Rabbit-In-a-Hatはその情報を使用して、グラフやマップなどを生成します。

| | A | B | C | D | E | F | G |
|----|---------------|-------------------|---------|------------|--------|----------------|----------------|
| 1 | Table | Field | Type | Max length | N rows | N rows checked | Fraction empty |
| 2 | dbo.allergies | start | date | 10 | 3184 | 3184 | 0 |
| 3 | dbo.allergies | stop | date | 10 | 3184 | 3184 | 0.725188442 |
| 4 | dbo.allergies | patient | varchar | 36 | 3184 | 3184 | 0 |
| 5 | dbo.allergies | encounter | varchar | 36 | 3184 | 3184 | 0 |
| 6 | dbo.allergies | code | varchar | 9 | 3184 | 3184 | 0 |
| 7 | dbo.allergies | description | varchar | 24 | 3184 | 3184 | 0 |
| 8 | | | | | | | |
| 9 | dbo.careplans | id | varchar | 36 | 30199 | 30199 | 0 |
| 10 | dbo.careplans | start | date | 10 | 30199 | 30199 | 0 |
| 11 | dbo.careplans | stop | date | 10 | 30199 | 30199 | 0.057849598 |
| 12 | dbo.careplans | patient | varchar | 36 | 30199 | 30199 | 0 |
| 13 | dbo.careplans | encounter | varchar | 36 | 30199 | 30199 | 0 |
| 14 | dbo.careplans | code | varchar | 15 | 30199 | 30199 | 0 |
| 15 | dbo.careplans | description | varchar | 62 | 30199 | 30199 | 0 |
| 16 | dbo.careplans | reasoncode | varchar | 9 | 30199 | 30199 | 0.050796384 |
| 17 | dbo.careplans | reasondescription | varchar | 56 | 30199 | 30199 | 0.050796384 |
| 18 | | | | | | | |

◀ ▶ Overview dbo.allergies dbo.carepla ... ⊕ ⋮ ◀ ▶

Figure 6.3: スキャンレポートのサンプル概要タブ

| | A | B |
|---|-------------------|-----------|
| 1 | Sex | Frequency |
| 2 | 2 | 61491 |
| 3 | 1 | 35401 |
| 4 | List truncated... | |

Figure 6.4: 単一列のサンプル値

In-a-HatはETLプロセスのドキュメントを生成しますが、ETLを作成するコードは生成しません。

プロセス概要

このソフトウェアを使用してETLのドキュメントを生成する一般的な手順は以下の通りです：

1. White Rabbitのスキャン結果を完了させます。
2. スキャン結果を開くと、インターフェースにソーステーブルとCDMテーブルが表示されます。
3. ソーステーブルが対応するCDMテーブルに情報を提供する場合は、ソーステーブルをCDMテーブルにマッピングします。
4. 各ソーステーブルからCDMテーブルへの接続について、ソース列とCDM列の詳細でさらに詳しく説明します。
5. Rabbit-In-a-Hatの作業内容を保存し、MS Word文書にエクスポートします。

ETLロジックの記述

White RabbitスキャンレポートをRabbit-In-a-Hatで開くと、ソースデータをOMOP CDMに変換する方法のロジックの設計と記述する準備が整ったことになります。次のセクションでは、ETLの一般的なフローと、最初にマッピングするテーブルについて説明します。

ETLの一般的なフロー

CDMは人を中心としたモデルであるため、PERSONテーブルのマッピングを最初に始めることになります。

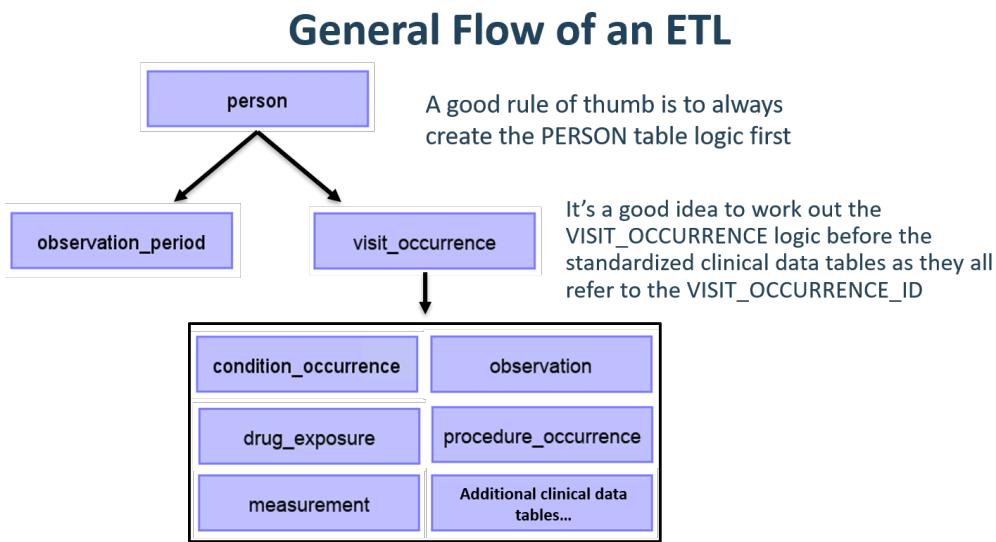


Figure 6.5: ETLの一般的なフローと、最初にマッピングするテーブル

CDM変換中に中間テーブルの作成が必要になることがあります。これは、イベントに正しくマッピングするためです。

²Synthea™は実際の患者をモデル化することを目的とした患者ジェネレーターです。データはアプリケーション://github.com/synthetichealth/synthea/wikiをご覧ください。

マッピング例 : Personテーブル

Syntheaデータ構造にはpatientsテーブルに20のカラムがありますが、図6.6に示されているように、すべてがPERSONテーブルを埋めるために必要というわけではありません。PERSONテーブルで使用されていないSynthea patientsテーブルのデータポイントの多く

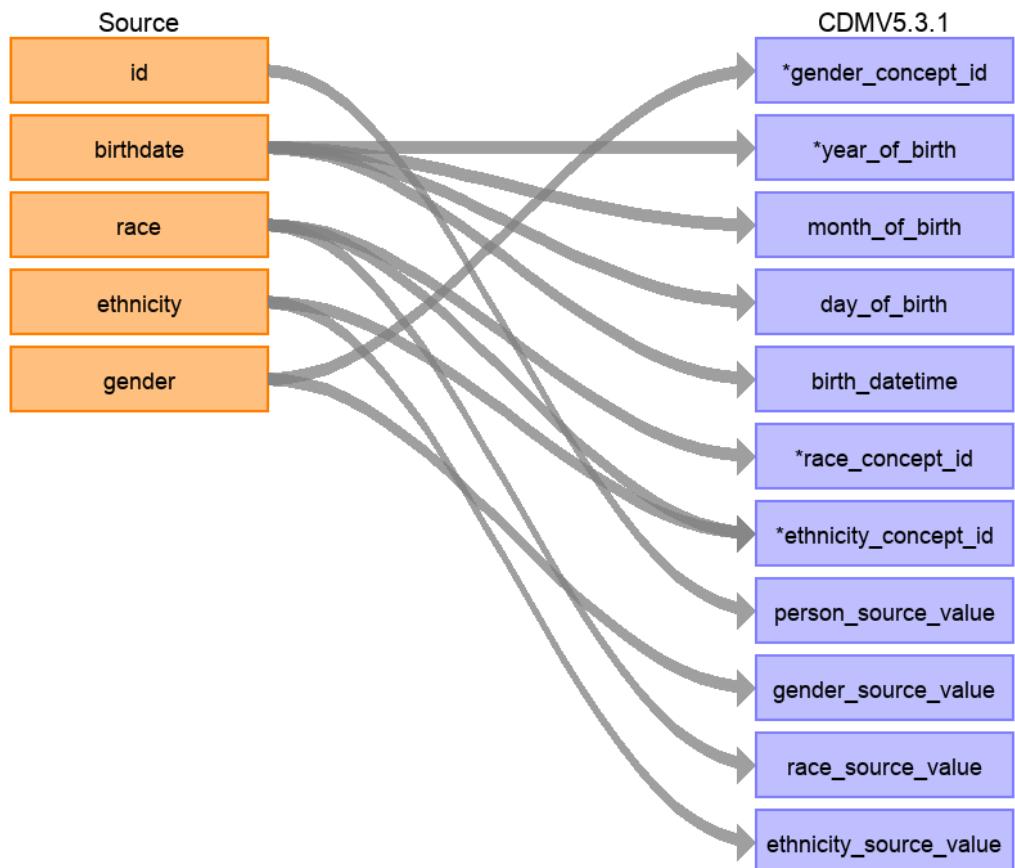


Figure 6.6: Synthea PatientsテーブルからCDM PERSONテーブルへのマッピング

表6.1には、Synthea patientsテーブルをCDM PERSONテーブルに変換するために適用されるField』（変換先フィールド）は、CDMのどこにデータがマッピングされるかを示しています。field』（変換元フィールド）では、CDMカラムにデータを入力するのに使用されるソースと、&comments』（ロジックとコメント）カラムには、ロジックの説明が記載されています。

Table 6.1: Synthea PatientsテーブルをCDM PERSON-テーブルに変換するためのETLロジック

| 目的フィールド | ソースフィールド | ロジックとコメント |
|--------------------------|----------------|---|
| PERSON_ID | | 自動生成。 PERSON_IDは実装時に生成されます。これ |
| GENDER_CONCEPT_ID | gender | ソースからのidフィールドは、その値を性別が「M」の場合、GENDER_CONC |
| YEAR_OF_BIRTH | birthdate | 生年月日から年を取得します。 |
| MONTH_OF_BIRTH | birthdate | 生年月日から月を取得します。 |
| DAY_OF_BIRTH | birthdate | 生年月日から日を取得します。 |
| BIRTH_DATETIME | birthdate | 0時を00:00:00とします。ここでは、 |
| RACE_CONCEPT_ID | race | race = 'WHITE' の場合は8527、 race = 'BLACK' の場合は8516、 race = 'ASIAN' の場合は8515、それ以外の場 |
| ETHNICITY_CONCEPT_ID | race ethnicity | race = 'HIS-PANIC'、または民族が ('CENTRAL_AMERICAN'、 'DOMINICAN' の場合、38003563 と設定し、それ以外の場合は0 に設定します。これは、複数のソース列にどのように影響するかを示す良い例では、民族は ヒスパニックまたは非ヒスパニックとして表されるため、ソース列 race とソース列ethnicityの両方の値がこの値を決定します。 |
| LOCATION_ID | | |
| PROVIDER_ID | | |
| CARE_SITE_ID | | |
| PERSON_SOURCE_VALUE | id | |
| GENDER_SOURCE_VALUE | gender | |
| GENDER_SOURCE_CONCEPT_ID | | |
| RACE_SOURCE_VALUE | race | |

| 目的フィールド | ソースフィールド | ロジックとコメント |
|-------------------|-----------|-----------|
| RACE_SOURCE_ | | |
| CONCEPT_ID | | |
| ETHNICITY_ | ethnicity | |
| SOURCE_VALUE | | |
| ETHNICITY_ | | |
| SOURCE_CONCEPT_ID | | |

SyntheaデータセットがCDMにどのようにマッピングされたかについては、仕様書全文を参照してください。

6.3 2:

OMOPボキャブラリには、常にソースコードが追加されています。これは、CDMにデータ（例：ICD10CMコード）から標準コンセプト（例：SNOMEDコード）へのマッピングを抽出するための「Maps to」関係を示すCONCEPT_RELATIONSHIPテーブルのレコードを使用できます。例として、ICD10CMコード「I21」（「急性心筋梗塞」）の標準コンセプトIDを特定するには、次のSQLクエリを使用します。

```
SELECT concept_id_2 AS standard_concept_id
FROM concept_relationship
INNER JOIN concept AS source_concept
  ON concept_id = concept_id_1
WHERE concept_code = 'I21'
  AND vocabulary_id = 'ICD10CM'
  AND relationship_id = 'Maps to';
```

| STANDARD_CONCEPT_ID |
|---------------------|
| 312327 |

残念ながら、ソースデータがボキャブラリに含まれていないコーディングシステムを使用する場合は、以下の手順でマッピングを作成できます。

- 最も頻繁に使用されるコードに焦点を当てる。使用されることのないコードや使用頻度の低いコードは除外する。
- 可能な限り既存の情報を活用しましょう。例えば、多くの国の医薬品コードはATCコードやUsagiコードで表されています。
- Usagiを使用しましょう。

6.3.1 Usagi

Usagiはコードマッピングを手動で作成するプロセスを支援するツールです。コードの記述

³<https://ohdsi.github.io/ETL-Synthea/>

⁴<https://translate.google.com/>

⁵<https://github.com/OHDSI/Usagi>

範囲と目的

マッピングが必要なソースコードはUsagiに読み込みます（コードが英語でない場合は追加の翻訳が必要）。

プロセス概要

このソフトウェアを使用する一般的な手順は次のとおりです：

1. マッピングしたいソースシステムからソースコードを読み込みます。
2. Usagiは用語の類似性アプローチを実行してソースコードをボキャブラリコンセプトにマップします。
3. Usagiのインターフェースを利用して、自動提案の正しさを確認し、必要に応じて改善します。
4. ボキャブラリのSOURCE_TO_CONCEPT_MAPにマッピングをエクスポートします。

ソースコードをUsagiにインポート

ソースコードをCSVまたはExcel (.xlsx) ファイルにエクスポートします。これには、ソースコードとその説明が含まれます。

注意事項: ソースコードの抽出はドメインごとに分けて行い、1つの大きなファイルにまとめないでください。

ソースコードはFile → Import codesメニューからUsagiに読み込まれます。ここでは「Import codes …」が表示されます（図6.7）。この図では、ソースコードの用語はオランダ語で、英語の説明が付いています。

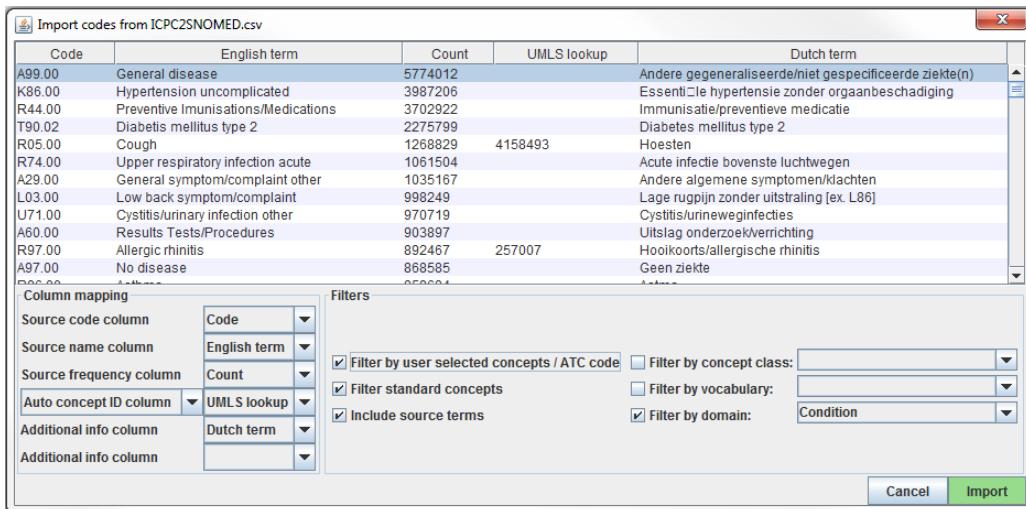


Figure 6.7: Usagiコード入力画面

「Column mapping」セクション（左下）では、インポートしたテーブルをUsagiに対してどのようにマッピングするかを設定します。各列の「Additional info」列をソースコードとボキャブラリコンセプトコードを関連付けるための情報として使用します。

最後に、「Filters」セクション（右下）では、Usagiがマッピングする際の制限をいくつか設定できます。「Filter by automatically selected concepts / ATC code」や「Filter by standard concepts」オプションをオフにすると、分類コンセプトも考慮されます。各フィルターオプションの説明を参考してください。

特別なフィルターとして「Filter by automatically selected concepts / ATC code」があります。検索を制限する情報がある場合、CONCEPT_IDSのリストまたはATCコードを指定された列に記述することで、検索を制限することができます（セミコロン区切りで複数のIDを指定できます）。

1. Column mappingセクションで、“Auto concept ID column”から“ATC column”に切り替えます。
2. Column mappingセクションで、ATCコードを含む列を“ATC column”として選択します。
3. フィルターセクションで「ユーザーが選択したコンセプト/ATCコードによるフィルタリング」をONにします。

ATCコード以外の情報源を使用して制限することもできます。上図の例では、UMLSから“concept ID column”を使用する必要があります。

すべての設定が完了したら、“Import”ボタンをクリックしてファイルをインポートします。

ソースコードからボキャブラリへのコンセプトマップの確認

ソースコードの入力ファイルをインポートすると、マッピング処理が始まります。図6.8では、Usagiの画面は、コンセプトテーブル、選択されているマッピング・セクション検索を実行する場所の3つの主要な部分で構成されていることがわかります。いずれのテ

The screenshot shows the Usagi software interface with three main sections:

- Overview table:** A table showing the results of the mapping process. It includes columns for Status, Source code, Source term, Frequency, Match score, Concept ID, Concept name, Domain, Concept class, Vocabulary, Concept code, and Standard concept. A red box highlights the first few rows.
- Selected mapping:** A table showing the selected mappings. It includes columns for Source code, Source term, Frequency, and various concept details. A red box highlights the first row.
- Search facility:** A search interface with fields for Query, Filters, and Results. It includes checkboxes for "Use source term as query", "Filter by user selected concepts / ATC code", "Filter by concept", "Filter standard concepts", "Filter by vocabulary", "Include source terms", and "Filter by domain". The results table below shows a list of terms and their mappings. A red box highlights the search filters.

Figure 6.8: Usagiでのソースコード入力画面

提案されたマッピングの承認

「コンセプトテーブル」には、ソース・コードとコンセプトの現在のマッピングが表示されます。図6.8の例では、ユーザーがドメインをコンディションに限定しているため、オランダ語の説明は、ソースコードの説明とコンセプト名および同義語を比較して、最適な一致を見つけています。“Include source terms”（ソース用語を含める）を選択していたため、Usagiは、特定のコンセプトにマッピングされるボキャブラリ内のすべてのソースコンセプトのマッピングできない場合は、CONCEPT_ID = 0 にマッピングされます。

ソース・コードを関連する標準ボキャブラリにマッピングする際には、コーディング・システムTable”（概要テーブル）のコードごとに作業して、Usagiが提案したマッピングを受け入れます。@ref:fig:usagiOverviewでは、オランダ語の“Hoesten”は英語の“Cough”（咳嗽）に翻訳されています。Usagiは“Cough”を使って、“4158493-C/O-cough”というボキャブラリコンセプトにマッピングしました。このマッチしたペアのマッチ（承認）”ボタンを押すことで、このマッピングを承認することができます。

新しいマッピングの検索

Usagiがマップを提案した場合、ユーザはより良いマッピングを見つけるか、マップをコンセプトID = 0)に設定する必要があります。図6.8の例では、オランダ語の「Hoesten」を「Cough」と訳しています。Usagiの提案は、UMLSから自動的に導出されたマッピングで識別されます。手動の検索ボックスを使用する場合、Usagiはあいまい検索を使用し、ANDやORのような論理操作を続けると、より適切なマッピングを見つけるために「Cough」という検索語を使用します。これらの検索条件を適用すると、「254761-Cough」が見つかり、これがオランダ語のコードと「Source Code (選択されたソースコード)」セクションの更新後に表示される「Replace concept (コンセプトを置換する)」ボタンを押し、「Approve (承認)」ボタンを押します。また、「Add concept (コンセプトの追加)」ボタンもあり、複数の標準化ボキャブラリのコンセプトを1つのソースコードにマッピングすることができます。

コンセプト情報

マッピングする適切なコンセプトを探す場合、コンセプトの「社会性」を考慮することが重要です。Shift + Cキーを押すか、上部メニューバーのview (閲覧) -> Concept information (コンセプト情報)を選択することで、より多くの情報を表示することもできます。

図6.9は、コンセプト情報パネルを示しています。このパネルには、コンセプトに関する一般的な情報が表示されています。すべてのコードがチェックされるまで、コードごとにこのプロセスを繰り返します。画面上部の「マッピングにコメントを追加することもでき、マッピングがどのように決定されたのかを記録することができます。

ベスト・プラクティス

- コーディングスキームの経験がある人に参加してもらってください。
- 列名をクリックすると、「コンセプトテーブル」の列を並べ替えることができます。“Match Score (マッチングスコア)”で並べ替えると、Usagiが最も信頼するコードを最初に確認できます。“Frequency (頻度)”での並べ替えも重要です。頻度に利用されるコードとそうでないコードに重点的に注目します。
- 場合によっては、CONCEPT_ID = 0にマッピングしても問題ありませんが、一部のコードは必ずしもこの値を取る必要があります。
- コンセプトの文脈、特にその親と子を考慮することが重要です。

Usagiで作成されたマッピングのエクスポート

USAGI内でマッピングしたら、それをエクスポートしてボキャブラリ SOURCE_TO_CONCEPT テーブルに追加するのが、次に進むための最良の方法です。

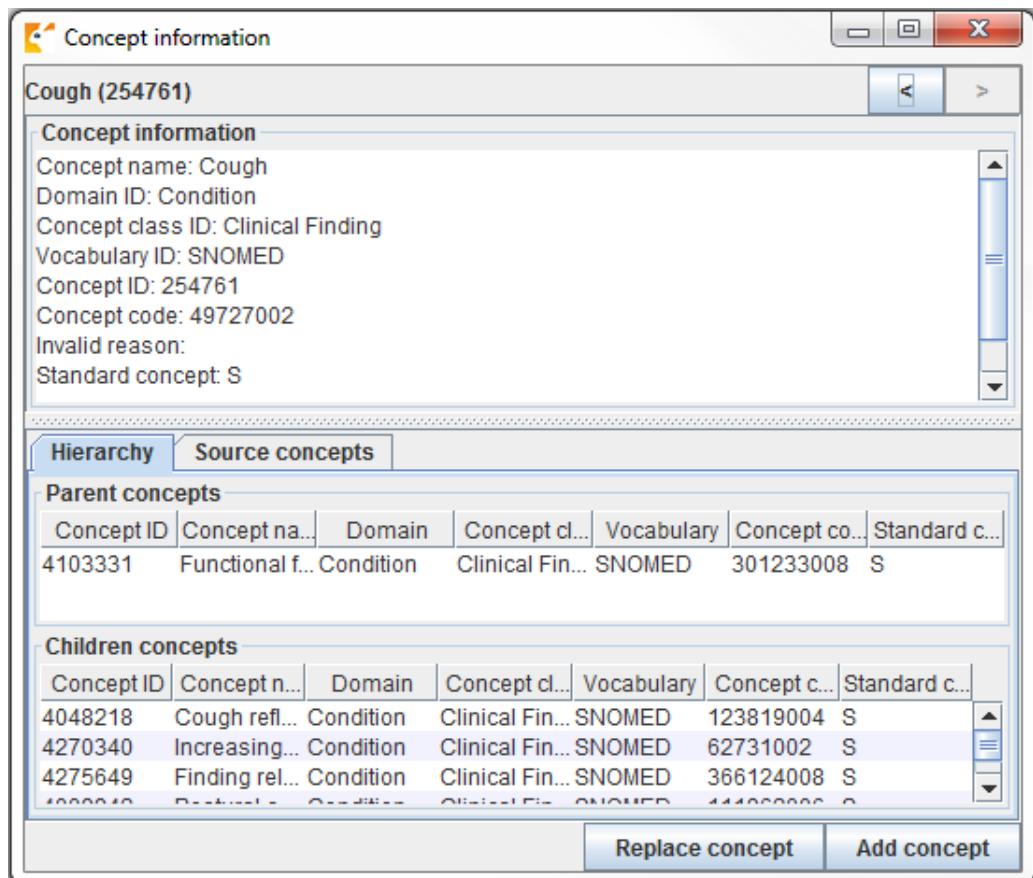


Figure 6.9: Usagi コンセプト情報パネル

マッピングをエクスポートするには、File(ファイル)->Export source_to_concept_map(ソースからコンセプトへのマッピングをエクセルポート)を選択します。どのSOURCE_VOCABULARY_ID(識別子)を入力ください。UsagiはこのIDをSOURCE_VOCABULARY_IDとして使用し、SOURC

SOURCE_VOCABULARY_IDを選択後、出力したCSVに名前を付けて保存します。出力したCSVは「承認」ステータスのマッピングのみがCSVファイルにエクスポートされることに留意ください。

Usagiで作成されたマッピングの更新

多くの場合、マッピングは一度だけの作業ではありません。データが更新されると、新しいソース・コードのセットが更新された場合は、次の手順で更新できます。

1. 新しいソースコードファイルをインポートします。
2. File(ファイル)->Apply previous mapping(以前のマッピングを適用する)を選択します。
3. 古いマッピングから引き継いだ承認済みのマッピングを継承していないコードを特定します。

ボキャブラリが更新された場合は、以下の手順に従います：

1. Athena から新しいボキャブラリファイルをダウンロードします。
2. Usagi インデックスを再構築します (Help (ヘルプ) -> Rebuild index (インデックスを再構築する))。
3. マッピング・ファイルを開きます。
4. 新しいバージョンのボキャブラリで標準コンセプトでなくなったコンセプトにマッピング

6.4 3: ETL

デザインとコードマッピングを完了すると、ETLプロセスをソフトウェアで実装することができます(6.7部を参照ください)。

実装の具体的な内容は施設ごとに異なり、インフラストラクチャ、データベースの規模、ETLツール

ETLの例をいくつか挙げます(複雑さの順に記載)：

- ETL-Synthea - Syntheaデータベースを変換するために書かれたSQLビルダー
 - <https://github.com/OHDSI/etl-synthea>
- ETL-CDMBuilder - 複数のデータベースを変換するためにデザインされた.NETアプリケーション
 - <https://github.com/OHDSI/etl-cdmbuilder>
- ETL-LambdaBuilder - AWS lambda機能を使用するビルダー
 - <https://github.com/OHDSI/etl-lambdabuilder>

複数回の独立した試みの後、ユーザーフレンドリーな”究極の”ETLツールの開発を断念しました。技術担当者が実装を開始する準備ができたら、ETLデザイン文書を彼らと共有するべきです。

6.5 4:

抽出、変換、読込のプロセスでは品質管理は反復的なプロセスとなります。典型的なパターンは、ソースデータの実装->ロジックの実装->ロジックのテスト->ロジックの修正・記述です。CDMをテストする方法

- ETL設計文書、コンピュータコード、およびコードマッピングのレビュー
 - コンピュータコードにおける最大の課題は、ネイティブデータのソースコード
- ソースデータとターゲットデータのサンプルに関する情報を手動で比較します。
 - 理想的には、多数のユニークレコードを持つ人物のデータを1件ずつ確認する。
- ソースデータとターゲットデータの全体的なカウントを比較します。
 - 特定の問題にどのように対処するかによって、カウントに期待される多少の差異を持つ人々を削除することを選択しています。なぜなら、そのような人々における来院は、ネイティブデータにおける来院またはビジットとは異なる方
 - CDMデータの全体的なカウントを比較する際には、これらの相違を考慮し、子
- ソースデータで既済の研究をCDMバージョンで再現します。
 - これはソースデータとCDMバージョンとの間の主な相違点を理解するのに適しています。
- ETLで対処すべきソースデータのパターンを再現するユニットテストを作成します。
 - ユニットテストは、ETL変換の品質と精度を評価する際に非常に便利です。通常

以上がETLの観点から品質管理にアプローチするハイレベルの方法です。OHDSIコミュニティの15章を参照ください。

6.6 ETL THEMIS

データをCDMに変換するグループが増えるにつれ、特定の状況でETLがどのように対処するか

OHDSIコミュニティは、一貫性を向上させるために慣行を文書化し始めました。OHDSI Wikiで参照できます⁶。各CDMテーブルには、ETLを設計する際に参照できる独自の慣行

すべてのデータシナリオを文書化し、発生した場合に何をするかをドキュメント化するOHDSI Wikiに文書化するメンバーで構成されています。THEMISは、古代ギリシャの神々を司る

6.7 CDM ETL

ETLをデザインし、マッピングを作成し、ETLを実装し、品質管理措置を構築することは、医療データソースは常に変化し続けることがよくあります。新しいデータが利用可能になると、バグが見つかった場合、それに対処する必要があります。ただし、すべてのバグが同じ重要なOMOPボキャブラリもまた、ソースデータと同様に常に変化しています。実際、ボキャブラリ

⁶<https://github.com/OHDSI/CommonDataModel/wiki>

⁷<https://github.com/OHDSI/Themis>

⁸<http://forums.ohdsi.org/>

CDMまたはETLのメンテナンスが必要となる最後の要因は、共通データモデル自体が更新され

6.8 ETL

ETLプロセスが異なる理由は数多くありますが、その主な理由のひとつは、私たちがすべてユ

- 80/20のルール。ソースコードを手動でコンセプトセットにマッピングするのにあまり時間はないからです。
- これだけで、まずスタートを切ることができます。残りのコードについては、ユースケーラブルなアプローチを採用する必要があります。
- 研究の品質に見合わないデータが失われることを恐れる必要はありません。これらのレコードはOHDSI CDMに統合されるべきです。
- CDMはメンテナンスが必要です。ETLが完了したからといって、二度と触らないといふことはありません。
- OHDSI CDMの開始、データベースの変換、分析ツールの実行のサポートが必要な場合は、ETLプロセスが複数回実行されることがあります。

6.9



- ETLにアプローチするための一般的に合意されたプロセスが存在します
 - * データ専門家とCDM専門家が協力してETLを設計する
 - * 医療知識を持つ人がコードマッピングを作成する
 - * 技術者がETLを実装する
 - * すべての関係者が品質管理に関与する
- OHDSIコミュニティはこれらのステップを促進するためにツールを開発しており、
- 参考にできる多くのETL例や合意された慣行があります

6.10

演習 6.1. ETLプロセスのステップを正しい順序に並べてください：

- A) データ専門家とCDM専門家が協力してETLを設計する
- B) 技術者がETLを実装する
- C) 医療知識を持つ人がコードマッピングを作成する
- D) すべての関係者が品質管理に関与する

演習 6.2. 選択したOHDSIリソースを使用して、表 6.3 に示すPERSON-レコードに関する4つの問題点を指摘してください（表はスペースのため省略されています）

Table 6.3: PERSONテーブル

| 列 | 値 |
|-----------|----------|
| PERSON_ID | A123B456 |

⁹<https://forums.ohdsi.org/c/implementers>

| 列 | 値 |
|------------------------|----------|
| GENDER_CONCEPT_ID | 8532 |
| YEAR_OF_BIRTH | NULL |
| MONTH_OF_BIRTH | NULL |
| DAY_OF_BIRTH | NULL |
| RACE_CONCEPT_ID | 0 |
| ETHNICITY_CONCEPT_ID | 8527 |
| PERSON_SOURCE_VALUE | A123B456 |
| GENDER_SOURCE_VALUE | F |
| RACE_SOURCE_VALUE | WHITE |
| ETHNICITY_SOURCE_VALUE | 提供されていない |

演習 6.3. VISIT_OCCURRENCEレコードを生成してみましょう。以下はSyntheaのためにPATIENT、START、ENDのデータを昇順で並べ替えます。その後、PERSON_IDごとに、

- MIN(START)をVISIT_START_DATEとして設定
- MAX(END)をVISIT_END_DATEとして設定
- “IP”をPLACE_OF_SERVICE_SOURCE_VALUEとして設定

ソースデータに図 6.10 に示されるようなビットセットがある場合、CDMで生成される

| Data Output Explain Messages Notifications Query History | | | | | |
|--|--------------------------------|---------------|--------------|-------------------------------------|--|
| | id character varying (1000) | start date | stop date | patient character varying (1000) | encounterclass character varying (1000) |
| 1 | 12 | 2004-09-26 | 2004-09-27 | 11 | inpatient |
| 2 | 13 | 2004-09-27 | 2004-09-30 | 11 | inpatient |

Figure 6.10: 例のソースデータ。

提案された回答は付録 E.3 を参照ください。

Part III

Chapter 7

著者: David Madigan

OHDSI共同研究は、通常、請求データベースや電子カルテデータベースなどの形式で、
が重点的に取り組むユースケースは、主に以下の3つのカテゴリーに分類されます。

- 特性評価
- 集団レベルの推定
- 患者レベルの予測

以下でこれらについて詳しく説明します。すべてのユースケースにおいて、生成されるエビデンスの質については、エビデンスの質に関する本の（第14章-第18章）で詳しく説明されています。

7.1

特性評価は次のような質問に答えようとします

かれらに何が起こったのか？

データを用いて、コホートやデータベース全体の集団の特性、医療行為、経時的な変化などを分析するための質問です。
データが答えを提供できる問い合わせには次のような例があります：

- 新たに心房細動と診断された患者のうち、何人がワルファリンの処方を受けたのか？
- 人口股関節置換術を受けた患者の平均年齢は？
- 65歳以上の患者の肺炎の発生率は？

典型的な特性評価は以下のように定式化されます：

- 何人の患者が…？
- どのくらいの頻度で…？
- 患者の何パーセントが…？
- 検査値の分布はどのようにになっているか…？
- の患者のHbA1c値は…？

- ・ の患者の検査値は…？
- ・ の患者の曝露期間の中央値は…？
- ・ 経時的な傾向は？
- ・ これらの患者が使用している他の薬剤は何か？
- ・併用療法は？
- ・ の症例が十分にあるか？
- ・ Xを研究は可能か？
- ・ の人口統計は？
- ・ のリスク要因は？（特定のリスク要因を識別する場合、予測ではなく推定）
- ・ の予測因子は？

そして期待されるアプローチは以下の通りです：

- ・ カウントまたはパーセンテージ
- ・ 平均
- ・ 記述統計
- ・ 発生率
- ・ 有病率
- ・ コホート
- ・ ルールベースの表現型
- ・ 薬剤利用
- ・ 疾患の自然経過
- ・ 服薬アドヒアランス
- ・ 併存疾患のプロファイル
- ・ 治療経路
- ・ 治療方針

7.2

限定的ではありますが、データは医療介入の効果に関する因果推論を裏付けることができ、

因果効果とは何か？

私たちは因果効果を理解することで、行動の結果を理解したいと考えています。例えば、あるデータは、次のような問い合わせに対する答えを提供することができます：

- ・ 新たに心房細動と診断された患者において、治療開始後最初の1年間に、ワルファリンは
- ・ メトホルミンの下痢に対する因果効果は年齢によって異なるか？

典型的な集団レベルの効果推定の問い合わせは次のように定式化されます：

- ・ …の効果は？
- ・ 介入を行った場合、どうなるのか？
- ・ どちらの治療がより効果的か？
- ・ Yに対するXのリスクは？
- ・ のイベント発生までの時間は？

そして、期待されるアウトプットは以下の通りです：

- 相対リスク
- ハザード比
- オッズ比
- 平均治療効果
- 因果効果
- 関連
- 相関
- 安全性監視
- 比較効果

7.3

データベースに収集された患者の医療履歴に基づいて、将来の健康イベントに関する患者

私には何が起こるのか？

データは、以下のような質問に対する答えを提供することができます：

- 新たに重度うつ病と診断された特定の患者について、診断後1年以内に自殺を図る確率
- 新たに心房細動と診断され、ワルファリンによる治療開始後1年以内に虚血性脳卒中

典型的な患者レベルの予測に関する問い合わせ次のように定式化されます：

- この患者が…になる可能性はどの程度か？
- 誰が…の候補となるのか？

そして、期待されるアウトプットは以下の通りです：

- 個人の確率
- 予測モデル
- 高リスク/低リスクグループ
- 確率的な表現型

集団レベルの推定と患者レベルの予測はある程度重複することがあります。例えば、予測



人々は予測モデルを因果モデルとして誤って解釈する傾向があります。しかし、予

7.4

あなたは、高血圧症の第一選択治療として急性心筋梗塞や血管性浮腫果に対するACE阻害

7.4.1

急性心筋梗塞は高血圧症患者に起こりうる心血管系の合併症であり、高血圧症に対する有効な10章を参照）。曝露集団のベースライン特性（人口統計学的特性、併存疾患、併用薬など）を11章を参照）の解析を実行します。この曝露集団における特定のアウトカムの発生率を推定す

7.4.2

集団レベルの効果推定研究（第12章を参照）は、急性心筋梗塞と血管性浮腫のアウトカムに対

7.4.3

曝露の因果効果とは独立して、アウトカムのリスクが最も高い患者を特定しようすることに13章を参照）。ここでは、ACE阻害薬の新規ユーザーの中で、治療開始後1年内に急性心筋

7.5

OHDSIデータベースでは答えを出せない重要な医療問題は数多くあります。以下はその例です

- ・ プラセボと比較した介入の因果効果。治療と非治療の比較は可能であっても、プラセボ比較の結果を除外する。
- ・ 市販薬に関するもの。
- ・ 多くのアウトカムやその他の変数は、ほとんど記録されていないか、まばらにしか記録されていない。
- ・ 患者は体調が悪いときにしか医療システムを利用しない傾向があるため、治療の有益性を評価する。

7.5.1

OHDSIデータベースに記録された臨床データは、臨床の現実と乖離している可能性があります。15章や第16章では、これらの問題について説明しており、ベストプラクティスでは、このよ Fuller (2009) を参照ください。

7.5.2

OHDSIデータベースにおける欠損は微妙な課題を呈します。データベースに記録されるべき健 et al. (2017) はこのトピックに関する有用な入門書を提供しています。

7.6



- 観察研究では、3つの大きなユースケースのカテゴリーを区別します。
- 特性評価は「彼らに何が起ったか？」という問い合わせに答えようとします。
- 集団レベルの推定は「因果効果は何か？」という問い合わせに答えようとします。
- 患者レベルの予測は「私には何が起こるのか？」という問い合わせに答えようとします。

- 予測モデルは因果モデルではありません。強い予測因子に介入してもアウトカムが変わらないことがあります。
- 観察医療データでは答えられない問い合わせもあります。

7.7

演習 7.1. これらの質問はどのユースケースのカテゴリーに属しますか？

1. 最近非ステロイド性抗炎症薬（NSAID）を投与された患者における消化管（GI）出血のリスクを評価する。
2. ベースラインの特性に基づいて、特定の患者が今後 1 年間に GI 出血を経験する確率を予測する。
3. セレコキシブと比較してジクロフェナクによる GI 出血のリスク増加を推定する。

演習 7.2. ジクロフェナクによる GI 出血のリスクがプラセボ（偽薬）に比べてどの程度高いですか？

推薦される解答は、付録 E.4 を参照ください。

Chapter 8

OHDSI

著者: Martijn Schuemie & Frank DeFalco

OHDSIは、患者レベルの観察データに対するさまざまなデータ分析のユースケースをサポートするオープンソースのフレームワークです。本章では、最初に分析を実装するさまざまな方法を説明し、分析で採用できる戦略について議論します。

8.1

図 8.1 は、CDMを使用してデータベースに対して研究を実装するために選択できるさまざま

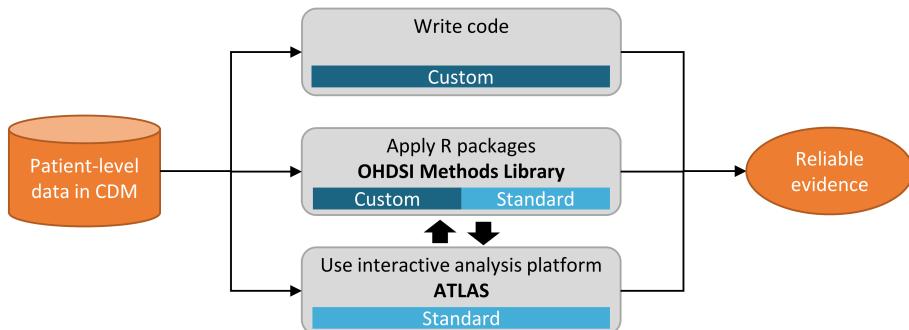


Figure 8.1: CDMのデータに対する分析を実装するさまざまな方法

研究を実装するための主なアプローチは3つあります。最初の方法は、OHDSIが提供する

2番目の方法は、Rで分析を開発し、OHDSI Methods Libraryのパッケージを利用する方法です。9章で詳しく説明していますが、PostgreSQL、SQL Server、OracleなどのさまざまなデータベースでOHDSI Methods Libraryのコンポーネントを再利用することで、完全にカスタムコードを使用する場合よりも簡単に分析を実装できます。

3番目の方法は、プログラマーでなくても幅広く分析を効率的に実行できるウェブベース

Librariesを使用しますが、分析をデザインするための単純なグラフィカルインターフェイスをLibraryで利用可能なすべてのオプションをサポートしているわけではありません。大半の研究

ATLASとMethods Libraryは独立したものではありません。ATLASで呼び出される複雑な分析のLibraryのパッケージへの呼び出しを通じて実行されます。同様に、Methods Libraryで使用されるコホートは、多くの場合、ATLASでデザインされています。

8.2

CDMに対する分析を実装するための戦略に加え、例えばカスタムコーディングやMethods Libraryで提供される標準分析コードの利用など、エビデンスを生成するための分析技術を使用8.2 は、OHDSIで採用されている3つの戦略を示しています。

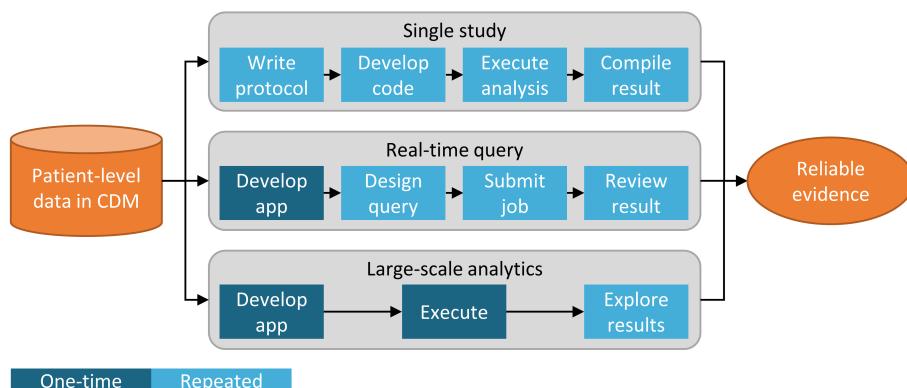


Figure 8.2: (臨床の)問い合わせに対するエビデンスを生成するための戦略

最初の戦略では、各分析を個別の研究として扱います。分析はプロトコルで事前に規定し、(et al., 2017)。ここでは、まずプロトコルが作成され、OHDSI Methods Libraryを使用した分析コードが開発され、OHDSIネットワーク全体で実行され、結果がまとめられます。

第二の戦略では、リアルタイムまたはほぼリアルタイムで特定のクラスの問い合わせに答えられるアプリケーションが生成されます。

第三の戦略では、同様に問い合わせに焦点を当てますが、その問い合わせのクラス内のすべてのエビデンス(et al., 2018b)。この研究では、すべてのうつ病治療が、4つの大規模な観察研究データベース¹で利用できます。

8.3 ATLAS

ATLASは、OHDSI コミュニティが開発した、標準化された患者レベルの観察データをCDM 形式で分析する設計と実行を支援する、無料で公開されているウェブベースのツールです。OHDSI WebAPI と組み合わせてウェブアプリケーションとして展開され、通常は

¹<http://data.ohdsi.org/SystematicEvidence/>

Apache Tomcat 上でホストされます。リアルタイム分析を行うには、CDM 内の患者レベルデータへのアクセスが必要であるため、通常は組織のファイアウォールの外側に配置されます。

The screenshot shows the ATLAS web application interface. On the left is a dark sidebar with various navigation links: Home, Data Sources, Search, Concept Sets, Cohort Definitions (which is currently selected), Characterizations, Cohort Pathways, Incidence Rates, Profiles, Estimation, Prediction, Jobs, Configuration, and Feedback. Below this sidebar, there's a Apache 2.0 open source software notice and a logo for OHDSI with the text "join the journey". The main content area has a header "Cohort #1770710" and a sub-header "New users of ACE inhibitors as first-line monotherapy for hypertension". Below this are tabs for Definition, Concept Sets, Generation, Reporting, Export, and Messages. A text input field says "enter a cohort definition description here". The main panel is titled "Cohort Entry Events" and contains a form for defining events. It includes fields for "Events having any of the following criteria:" (with a dropdown for "a drug exposure of ACE inhibitors" and a red "X" button for "for the first time in the person's history"), "with continuous observation of at least [365] days before and [0] days after event index date", and "Limit initial events to: earliest event per person". There are buttons for "+ Add Initial Event", "+ Add attribute...", and "Delete Criteria". Below this is another section titled "Inclusion Criteria" with a "New inclusion criteria" button and a list of two items: "1. has hypertension diagnosis in 1 yr prior to treatment" and "2. Has no prior antihypertensive drug exposures in medical".

Figure 8.3: ATLASユーザインターフェース

図 8.3 に ATLAS のスクリーンショットを示します。左側には ATLAS の様々な機能を示すナビゲーションメニューがあります。データソース データソースは、ATLAS プラットフォームに構成された各データソースの一覧です。ボキャブラリ検索 ATLAS は OMOP 標準化ボキャブラリを検索し、これらのボキャブラリを統合できます。コンセプトセット コンセプトセットは、標準化された分析全体で使用するコンセプトの定義です。コホート定義 コホート定義は、一定期間内に 1 つまたは複数の基準を満たす人のセットです。特性評価 特性評価は、定義された 1 つまたは複数のコホートを調査し、これらの患者集の特徴を分析します。コホート経路 コホート経路は、1 つまたは複数の集団内で発生する臨床イベントのシリーズです。発生率 発生率は、対象集団内のアウトカムの発生率を推定するためのツールです。このツールは、過去のデータから将来の発生率を予測します。プロファイル プロファイルは、個々の患者の縦断的観察データを調査し、特定の個人に関する情報を収集します。集団レベル推定 推定は、比較コホートデザインを使用して集団レベルの効果推定研究を行います。患者レベルの予測 予測は、機械学習アルゴリズムを適用して、患者レベルの予測分析を行います。ジョブ ジョブメニュー項目を選択して、WebAPI を通じて実行されているプロセスの状況を確認できます。構成 構成メニュー項目を選択して、ソース構成セクションに構成されたデータソースを管理できます。フィードバック フィードバックリンクをクリックすると、ATLAS の課題ログにアクセスできます。

8.3.1

ATLASとWebAPIは、プラットフォーム全体で機能やデータソースへのアクセスを制御するためShiroライブラリを活用して構築されています。セキュリティシステムの詳細は、オンラインリポジトリ²。

8.3.2

ATLASのドキュメントは、ATLAS GitHubリポジトリのwikiでオンラインで確認できます³。このwikiには、さまざまなアプリケーション機能に関する情報や、オンラインビデオチュートリアルがあります。

8.3.3

ATLASのインストールは、OHDSI WebAPIと組み合わせて行います。各コンポーネントのインストールGitHubリポジトリのセットアップガイド⁴とWebAPI GitHubリポジトリのインストールガイド⁵。

8.4 Methods Library

OHDSI Methods libraryは、図 8.4 に示されているオープンソースのRパッケージのコレクションです。

これらのパッケージは、CDM内のデータから始まり、推定値やそれを裏付ける統計、図表を生成するための高度な標準化分析を提供することもできます。Methods libraryは、透明性、再現性、効率性を重視しています。

Methods libraryはすでに多くの公表された臨床研究 (Boland et al., 2017; Duke et al., 2017; Ramcharran et al., 2017; Weinstein et al., 2017; Wang et al., 2017; Ryan et al., 2017, 2018; Vashisht et al., 2018; Yuan et al., 2018; Johnston et al., 2019) で使用されており、方法論の研究にも利用されています (Schuemie et al., 2014, 2016; Reps et al., 2018; Tian et al., 2018; Schuemie et al., 2018a,b; Reps et al., 2019)。Methods library内のメソッドの実装の妥当性については第17章で説明されています。

8.4.1

すべてのパッケージで組み込まれている重要な機能の一つは、多くの分析を効率的に実行できることです。この計算効率により、大規模な分析が可能になり、多くの質問に一度に回答することができます。このように、経験則に基づくキャリブレーションを行うことも不可欠です。

²<https://github.com/OHDSI/WebAPI/wiki/Security-Configuration>

³<https://github.com/OHDSI/ATLAS/wiki>

⁴<https://github.com/OHDSI/Atlas/wiki/Atlas-Setup-Guide>

⁵<https://github.com/OHDSI/WebAPI/wiki/WebAPI-Installation-Guide>

| | | | |
|-----------------------------------|---|---|---|
| Prediction and estimation methods | Cohort Method New-user cohort studies using large-scale regression for propensity and outcome models | Self-Controlled Case Series Self-Controlled Case Series analysis using few or many predictors, includes splines for age and seasonality. | Self-Controlled Cohort A self-controlled cohort design, where time preceding exposure is used as control. |
| | Patient Level Prediction Build and evaluate predictive models for user-specified outcomes, using a wide array of machine learning algorithms. | Case-control Case-control studies, matching controls on age, gender, provider, and visit date. Allows nesting of the study in another cohort. | Case-crossover Case-crossover design including the option to adjust for time-trends in exposures (so-called case-time-control). |
| Method characterization | Empirical Calibration Use negative control exposure-outcome pairs to profile and calibrate a particular analysis design. | Method Evaluation Use real data and established reference sets as well as simulations injected in real data to evaluate the performance of methods. | Evidence Synthesis Combining study diagnostics and results across multiple sites. |
| Supporting packages | Database Connector Connect directly to a wide range of database platforms, including SQL Server, Oracle, and PostgreSQL. | Sql Render Generate SQL on the fly for the various SQL dialects. | Cyclops Highly efficient implementation of regularized logistic, Poisson and Cox regression. |
| | ParallelLogger Support for parallel computation with logging to console, disk, or e-mail. | Feature Extraction Automatically extract large sets of features for user-specified cohorts using data in the CDM. | |

Figure 8.4: OHDSI Methods Libraryのパッケージ

8.4.2

Methods libraryは、非常に大規模なデータベースに対しても大量のデータを含む計算を実行できます。

- 大部分のデータ操作はデータベースサーバー上で実行されます。分析は通常、データベース libraryはSqlRenderやDatabaseConnectorパッケージを介して関連データの前処理や抽出を行います。
- 大量のローカルデータオブジェクトはメモリ効率の良い方法で保存されます。ローカル libraryはffパッケージを使用して大規模データオブジェクトを保存、処理します。これにより、データを効率的に読み込むことができます。
- 必要に応じて高性能コンピューティングが適用されます。例えば、Cyclopsパッケージはデータ library全体で使用される非常に効率的な回帰エンジンを実装しており、これにより通常は複数日かかる計算を数秒で実行することができる場合があります。

8.4.3

Rはパッケージを文書化するための標準的な方法を提供しています。各パッケージには、パッケージ Libraryのウェブサイト⁶、パッケージのGitHubリポジトリ、CRANで利用できます。さらに、パッケージマニュアルに加えて、多くのパッケージはビネットが提供されています。ビネット Libraryのウェブサイト、パッケージのGitHubリポジトリ、CRANで入手可能なパッケージはCohortMethodパッケージです。

8.4.4

システム要件を検討する際には、二つのコンピューティング環境が関連してきます：データベースサーバーと分析ワークステーション。

データベースサーバーはCDM形式の観察医療データを保持する必要があります。Methods libraryは、従来のデータベースシステム（PostgreSQL、Microsoft SQL Server、Oracle）、パラレルデータウェアハウス（Microsoft APS、IBM Netezza、Amazon Redshift）に加えビッグデータプラットフォーム（Impala経由でのHadoop、BigQuery）など、幅広いデータベース管理システムをサポートしています。

分析ワークステーションは、Methods libraryがインストールされ実行される場所です。これがWindows Serverが実行されるリモートサーバーかに関わらず、Rがインストールされている必要があります。OHDSI RパッケージではJavaがインストールされている必要があります。分析ワークステーションはデータベースサーバーと連携するための接続情報を保持する必要があります。

8.4.5

OHDSI Rパッケージを実行するために必要な環境をインストールするための手順は次の通りです。

- Rは統計的コンピューティング環境です。基本的なユーザインターフェースとして主にWindows Serverが実行されるリモートサーバーかに関わらず、Rがインストールされている必要があります。
- Rtoolsは、WindowsでRパッケージをソースからビルドする際に必要なプログラマー式で開発環境です。
- RStudioは、Rを使いやさしくするIDE（統合開発環境）です。コードエディタ、デバッグ、実行環境などを統合して提供します。
- Javaは、OHDSI Rパッケージの一部のコンポーネント、例えばデータベースへの接続に必要なものです。

⁶<https://ohdsi.github.io/MethodsLibrary>

⁷<https://ohdsi.github.io/CohortMethod/articles/MultipleAnalyses.html>

以下では、Windows環境でのそれぞれのインストール方法を説明します。



Windowsでは、RとJavaはどちらも32ビットと64ビットのアーキテクチャがあります。

Rのインストール

1. <https://cran.r-project.org/> で、図 8.5 に示されるように「Download R for Windows」、「base」の順にクリックし、ダウンロードしてください。



[CRAN](#)
[Mirrors](#)
[What's new?](#)
[Task Views](#)
[Search](#)

R-3.6.1 for Windows (32/64 bit)

[Download R 3.6.1 for Windows](#) (81 megabytes, 32/64 bit)

[Installation and other instructions](#)
[New features in this version](#)

If you want to double-check that the package you have downloaded matches the package distributed by CRAN, you can compare the [md5sum](#) of the .exe to the [fingerprint](#) on the master server. You will need a version of md5sum for windows: both [graphical](#) and [command line versions](#) are available.

Figure 8.5: CRANからのRのダウンロード

2. ダウンロードが完了したら、インストーラを実行します。2つの例外を除いて、すばやく、プログラムファイルにはインストールしない方が良いでしょう。代わりに、図 8.6 のように、CドライブのサブフォルダとしてRを作成します。次に、RとJavaの図 8.7 のように32ビットアーキテクチャを無効にします。

完了すると、スタートメニューからRを選択できるようになります。

Rtoolsのインストール

1. <https://cran.r-project.org/> にアクセスし、「Download R for Windows」をクリックし、次に「Rtools」をクリックして、最新版のRtoolsをダウンロードします。
2. ダウンロードが完了後、インストーラを実行します。すべてデフォルトのオプション

RStudioのインストール

1. <https://www.rstudio.com/> にアクセスし、「Download RStudio」またはRStudioの8.8に示されるようにWindows用のインストーラをダウンロードします。
2. ダウンロードが完了後、インストーラを実行します。すべてデフォルトのオプション

Javaのインストール

1. <https://java.com/en/download/manual.jsp> にアクセスし、図 8.9に示されるように、Windows64ビット版のインストーラを選択します。32ビット

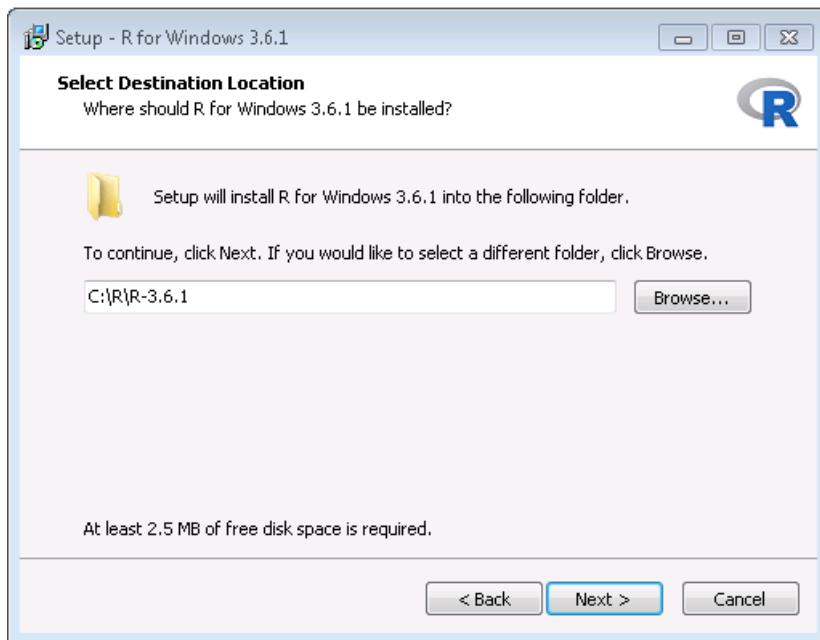


Figure 8.6: Rフォルダの設定

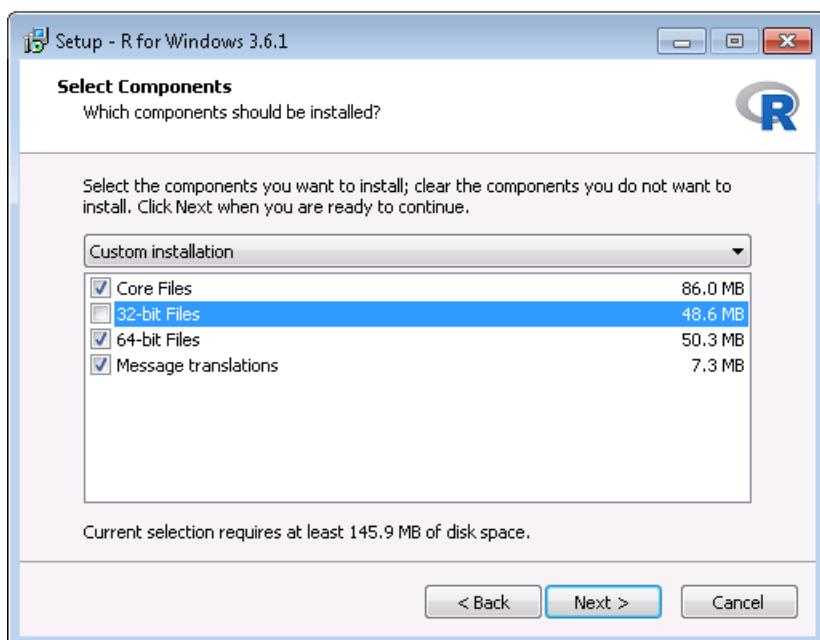


Figure 8.7: 32ビットバージョンのRを無効化

Installers for Supported Platforms

| Installers | Size | Date | MD5 |
|--|----------|------------|-----------|
| RStudio 1.2.1335 - Windows 7+ (64-bit) | 126.9 MB | 2019-04-08 | d0e2470f1 |
| RStudio 1.2.1335 - Mac OS X 10.12+ (64-bit) | 121.1 MB | 2019-04-08 | 6c570b0e2 |
| RStudio 1.2.1335 - Ubuntu 14/Debian 8 (64-bit) | 92.2 MB | 2019-04-08 | c1b07d051 |

Figure 8.8: RStudioのダウンロード

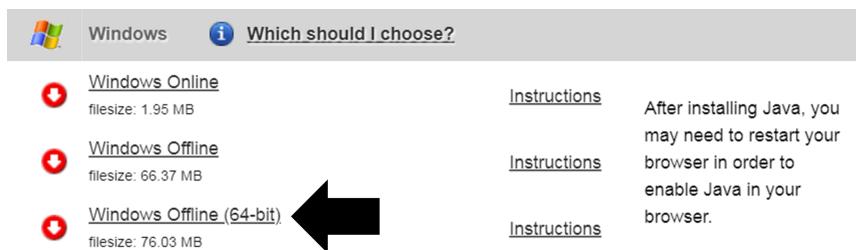


Figure 8.9: Javaのダウンロード

2. ダウンロード後、インストーラを実行します。

インストールの確認

これで準備は整ったはずですが、念のため確認しておきましょう。Rを起動し、下記のよ

```
install.packages("SqlRender")
library(SqlRender)
translate("SELECT TOP 10 * FROM person;", "postgresql")
```

```
## [1] "SELECT * FROM person LIMIT 10;"
## attr(,"sqlDialect")
## [1] "postgresql"
```

この関数はJavaを使用するので、すべてがうまくいけば、RとJavaの両方が正しくインス

もう一つのテストは、ソースパッケージがビルドできるかどうかを確認することです。GitHubリポジトリからCohortMethodパッケージをインストールします：

```
install.packages("drat")
drat::addRepo("OHDSI")
install.packages("CohortMethod")
```

8.5

ATLASやMethods Libraryを含むOHDSIツールスタック全体を組織内で展開することは、非常に複雑な手順が必要です。そのため、OHDSIツールはAWSのWeb Services (AWS)で提供されています。

8.5.1 Broadsea

Broadsea⁸はDockerコンテナ技術⁹を使用しています。OHDSIツールは依存関係とともに、Dockerイメージとして用意されています。Windows、MacOS、Linuxなどのほとんどのオペレーティングシステムで利用可能です。Broadsea Dockerイメージには、Methods libraryやATLASを含む主なOHDSIツールが含まれています。

8.5.2 Amazon AWS

Amazonは、AWSクラウドコンピューティング環境でボタンをクリックするだけでインスタンスを起動するOHDSI-in-a-Box¹⁰とOHDSIonAWS¹¹です。

OHDSI-in-a-Boxは特に学習環境として作成されたものであり、OHDSIコミュニティが提供するツール（OMOP CDMマッピング、ETLツールなど）も、OHDSI-in-a-Boxに含まれています。OHDSI-in-a-Boxのアーキテクチャは、図 8.10 に示されています。

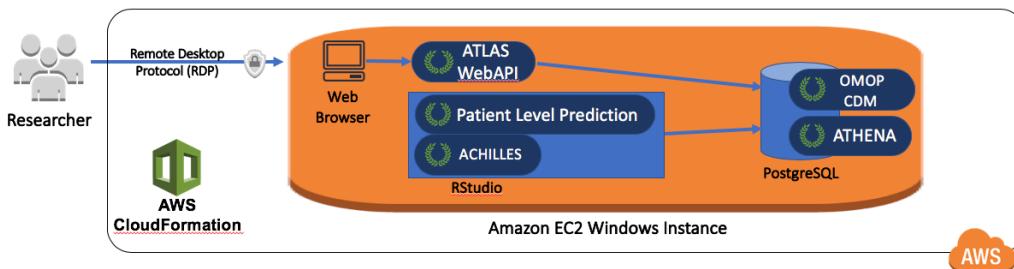


Figure 8.10: OHDSI-in-a-BoxのAmazon Web Servicesアーキテクチャ

OHDSIonAWSは、企業向け、マルチユーザー対応、拡張性や耐障害性に優れたOHDSI環境の構築を容易にします。複数のサンプルデータセットが含まれており、組織の実際のヘルスケアデータを自動的にロードできます。データはAmazon Redshiftデータベースプラットフォームに配置され、OHDSIツールによって分析されます。ATLASの中間結果はPostgreSQLデータベースに保存されます。ユーザーはフロントエンドで、(Apache Serverを活用) を通じてATLASやRStudioにアクセスできます。RStudioにはOHDSI Methods Libraryがすでにインストールされており、データベースへの接続に使用できます。OHDSIonAWSは、OHDSIツールを簡単に組織内に展開するための強力なツールです。

⁸<https://github.com/OHDSI/Broadsea>

⁹<https://www.docker.com/>

¹⁰<https://github.com/OHDSI/OHDSI-in-a-Box>

¹¹<https://github.com/OHDSI/OHDSIonAWS>

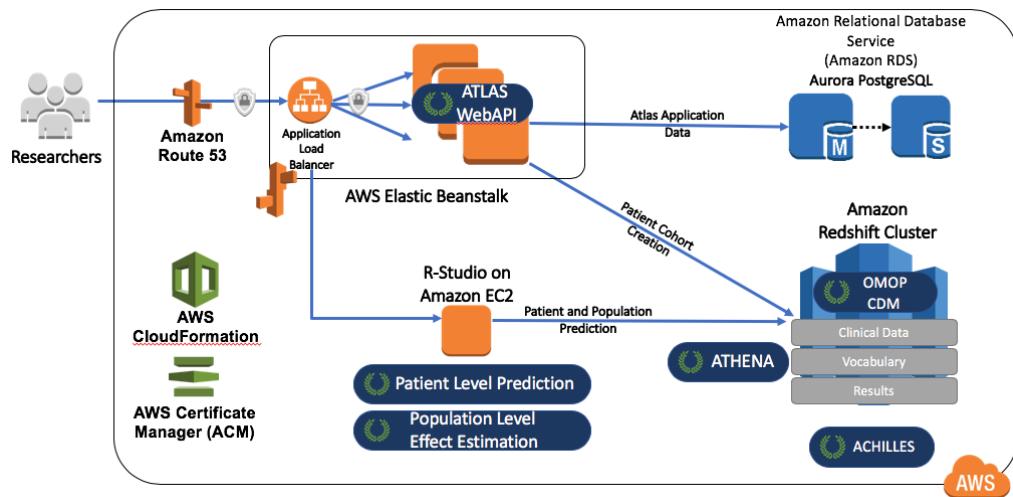


Figure 8.11: OHDSI on AWSのAmazon Web Servicesアーキテクチャ

8.6



- CDM内のデータに対して分析を行うには
 - * カスタムコードを作成する
 - * OHDSI Methods LibraryのRパッケージを使用したコードを作成する
 - * インタラクティブな分析プラットフォームATLASを使用する
- OHDSIツールはさまざまな分析戦略を用いています
 - * 単一研究
 - * リアルタイムクエリ
 - * 大規模アナリティクス
- OHDSIアナリティクスツールのほとんどは、以下に組み込まれています
 - * インタラクティブな分析プラットフォームATLAS
 - * OHDSI Methods LibraryのRパッケージ
- OHDSIツールの展開を容易にするいくつかの戦略があります。

Chapter 9

SQL R

著者: Martijn Schuemie & Peter Rijnbeek

共通データモデル（CDM）はリレーションナルデータベースモデルです（すべてのデータは関連付け可能）。SQL Serverなどのソフトウェアプラットフォームを使用してリレーションナルデータベースを操作できます。OHDSI LibraryなどのさまざまなOHDSIツールは、バックグラウンドでデータベースにクエリを投げます。これらのツールは多くの場合、ユーザーがデータを適切に分析できるよう、ガイドするように設計されています。

リレーションナルデータベースをクエリする標準的な言語はSQL（Structured Query Language）で、クエリやデータ変更に使用できます。SQLの基本コマンドは確かに似ています。例えば、SQL Server上のPERSONテーブルの最初の10行を取得するには、次のように入力します：

```
SELECT TOP 10 * FROM person;
```

一方、PostgreSQLでは同じクエリは次のようにになります：

```
SELECT * FROM person LIMIT 10;
```

OHDSIでは、プラットフォーム固有の表現に依存しないことを望んでいます。すなわち、OHDSI SQL - は主にSQL Server SQL表現のサブセットです。本章で例示するSQL文はすべてOHDSI SQLで書かれています。

各データベースプラットフォームには、SQLを使用したデータベースのクエリのための独自の構文があります。

そのため、CDMに準拠したデータベースに対してOHDSIツールを使用せずにクエリを実行するには、OHDSI SQLを学ぶ必要があります。

本章では、読者がSQLの基本的な理解をしていることを前提としています。まず、SqlRenderでは、CDMにクエリを出すためのSQL（この場合OHDSI SQL）を使用する方法を説明します。

9.1 SqlRender

SqlRender パッケージは CRAN (Comprehensive R Archive Network) で入手可能であり、以

```
install.packages("SqlRender")
```

SqlRenderは、従来のデータベースシステム（PostgreSQL、Microsoft SQL Server、SQLite、Oracle）や並列データウェアハウス（Microsoft APS、IBM Netezza、Amazon Redshift）に加え、ビッグデータプラットフォーム（Hadoop から Impala、Google BigQuery）など、幅広い技術プラットフォームをサポートしています。

9.1.1 SQL

パッケージの機能のひとつは、SQLのパラメータ化をサポートすることです。しばしば、いくつかのパラメータに基づいて、SQLの小さなバリエーションを生成する必要が

パラメータ値の置換

◎文字を使用して、レンダリング時に実際のパラメータ値と置換する必要があるパラメータ名 a という変数がSQLで言及されています。render 関数の呼び出しでは、このパラメータの値が

```
sql <- "SELECT * FROM concept WHERE concept_id = @a;"  
render(sql, a = 123)
```

```
## [1] "SELECT * FROM concept WHERE concept_id = 123;"
```

ほとんどのデータベース管理システムが提供するパラメータ化とは異なり、テーブル名やフィ

```
sql <- "SELECT * FROM @x WHERE person_id = @a;"  
render(sql, x = "observation", a = 123)
```

```
## [1] "SELECT * FROM observation WHERE person_id = 123;"
```

パラメータ値は、数値、文字列、ブーリアン変数、ベクトル（カンマ区切りのリストに変換さ

```
sql <- "SELECT * FROM concept WHERE concept_id IN (@a);"  
render(sql, a = c(123, 234, 345))
```

```
## [1] "SELECT * FROM concept WHERE concept_id IN (123,234,345);"
```

If-Then-Else

時には、1つまたは複数のパラメータの値に基づいてコードブロックをオンまたはオフにする
? {if true} : {if false} 構文を使用して行います。condition が true

または 1 の場合、if true ブロックが使用され、それ以外の場合は if false ブロックが（存在する場合）表示されます。

```
sql <- "SELECT * FROM cohort {@x} ? {WHERE subject_id = 1}"
render(sql, x = FALSE)
```

```
## [1] "SELECT * FROM cohort "
```

```
render(sql, x = TRUE)
```

```
## [1] "SELECT * FROM cohort WHERE subject_id = 1"
```

簡単な比較もサポートされています：

```
sql <- "SELECT * FROM cohort {@x == 1} ? {WHERE subject_id = 1};"
render(sql, x = 1)
```

```
## [1] "SELECT * FROM cohort WHERE subject_id = 1;"
```

```
render(sql, x = 2)
```

```
## [1] "SELECT * FROM cohort ;"
```

IN 演算子もサポートされています：

```
sql <- "SELECT * FROM cohort {@x IN (1,2,3)} ? {WHERE subject_id = 1};"
render(sql, x = 2)
```

```
## [1] "SELECT * FROM cohort WHERE subject_id = 1;"
```

9.1.2 SQL

SqlRender パッケージのもう一つの機能は、OHDSI SQL から他のSQL表現へ変換することです。

```
sql <- "SELECT TOP 10 * FROM person;"
translate(sql, targetDialect = "postgresql")
```

```
## [1] "SELECT * FROM person LIMIT 10;"
```

```
## attr(,"sqlDialect")
```

```
## [1] "postgresql"
```

targetDialect パラメータには次の値が設定可能です：“oracle”，“postgresql”，“pdw”，“redshift”，“impala”，“netezza”，“bigquery”，“sqlite”，“sql server”。



SQL関数や構文を適切に変換できる範囲には限界があります。その理由は、パッケージはSQLが独自の新しいSQL方言として開発された主な理由です。しかし、可能な限り、車輪Serverの構文に従うようにしています。

最大限の努力を尽くしても、サポートされているすべてのプラットフォーム上でエラーなく実SQLを記載するには、考慮すべき点がいくつかあります。以下では、これらの考慮事項について

Translateがサポートする関数と構造

これらのSQL Server関数はテスト済であり、各表現への正確な変換が確認されています：

Table 9.1: “translate (翻訳)によりサポートされる関数と構造

| 関数 | 関数 | 関数 |
|-------------------|------------|------------|
| ABS | EXP | RAND |
| ACOS | FLOOR | RANK |
| ASIN | GETDATE | RIGHT |
| ATAN | HASHBYTES* | ROUND |
| AVG | ISNULL | ROW_NUMBER |
| CAST | ISNUMERIC | RTRIM |
| CEILING | LEFT | SIN |
| CHARINDEX | LEN | SQRT |
| CONCAT | LOG | SQUARE |
| COS | LOG10 | STDEV |
| COUNT | LOWER | SUM |
| COUNT_BIG | LTRIM | TAN |
| DATEADD | MAX | UPPER |
| DATEDIFF | MIN | VAR |
| DATEFROMPARTS | MONTH | YEAR |
| DATETIMEFROMPARTS | NEWID | |
| DAY | PI | |
| EOMONTH | POWER | |

* Oracleでは特別な権限が必要です。SQLiteでは同等のものはありません。

同様に、多くのSQL構文構造がサポートされています。以下は、正確に翻訳されることが確認

```
-- Simple selects:
SELECT * FROM table;

-- Selects with joins:
SELECT * FROM table_1 INNER JOIN table_2 ON a = b;
```

```
-- Nested queries:  
SELECT * FROM (SELECT * FROM table_1) tmp WHERE a = b;  
  
-- Limiting to top rows:  
SELECT TOP 10 * FROM table;  
  
-- Selecting into a new table:  
SELECT * INTO new_table FROM table;  
  
-- Creating tables:  
CREATE TABLE table (field INT);  
  
-- Inserting verbatim values:  
INSERT INTO other_table (field_1) VALUES (1);  
  
-- Inserting from SELECT:  
INSERT INTO other_table (field_1) SELECT value FROM table;  
  
-- Simple drop commands:  
DROP TABLE table;  
  
-- Drop table if it exists:  
IF OBJECT_ID('ACHILLES_analysis', 'U') IS NOT NULL  
    DROP TABLE ACHILLES_analysis;  
  
-- Drop temp table if it exists:  
IF OBJECT_ID('tempdb..#cohorts', 'U') IS NOT NULL  
    DROP TABLE #cohorts;  
  
-- Common table expressions:  
WITH cte AS (SELECT * FROM table) SELECT * FROM cte;  
  
-- OVER clauses:  
SELECT ROW_NUMBER() OVER (PARTITION BY a ORDER BY b)  
    AS "Row Number" FROM table;  
  
-- CASE WHEN clauses:  
SELECT CASE WHEN a=1 THEN a ELSE 0 END AS value FROM table;  
  
-- UNIONs:  
SELECT * FROM a UNION SELECT * FROM b;  
  
-- INTERSECTIONS:  
SELECT * FROM a INTERSECT SELECT * FROM b;  
  
-- EXCEPT:  
SELECT * FROM a EXCEPT SELECT * FROM b;
```

文字列の連結

文字列の連結は、SQL Serverが他の言語よりも特異ではない領域の1つです。SQL Serverでは、`SELECT first_name + ' ' + last_name AS full_name FROM table`と書きますが、これは PostgreSQL と Oracle では `SELECT first_name || ' ' || last_name AS full_name FROM table`でなければなりません。SqlRenderは、連結用のオペレーターを自動的に付加する機能を持っています。そのため、`SELECT last_name + CAST(age AS VARCHAR(3)) AS full_name FROM table`も正しく翻訳されます。曖昧さを避けるために、2つ以上の文字列を連結する場合関数を使用するのが最善の方法です。

テーブルエイリアスとASキーワード

多くのSQL表現ではテーブルエイリアスを定義する際に AS キーワードを使用できますが、MySQL、SQL Server、PostgreSQL、Redshiftなどでは問題なく動作します：

```
-- Using AS keyword
SELECT *
FROM my_table AS table_1
INNER JOIN (
    SELECT * FROM other_table
) AS table_2
ON table_1.person_id = table_2.person_id;

-- Not using AS keyword
SELECT *
FROM my_table table_1
INNER JOIN (
    SELECT * FROM other_table
) table_2
ON table_1.person_id = table_2.person_id;
```

しかし、Oracleでは AS キーワードを使用するとエラーが発生します。上記の例では、最初のクエリで AS キーワードを使用しないことを推奨します。（注：SqlRenderではOracleがASの使用を許可していないテーブルエイリアスとOracleがASの使用を要求しているフィールドエイリアスを混在する場合に問題が発生する）

テンポテーブル

テンポテーブルは中間結果を保存するのに非常に有用であり、正しく使用するとクエリのパフォーマンスを大幅に向上させることができます。

- 異なるユーザーからのテーブルが競合しないように、テーブル名にランダムな文字列を追加する
- テンポラリテーブルが作成されるスキーマをユーザーが指定できるようにします。

例えば：

```
sql <- "SELECT * FROM #children;"
translate(sql, targetDialect = "oracle", oracleTempSchema = "temp_schema")
```

```
## Warning: The 'oracleTempSchema' argument is deprecated. Use 'tempEmulation'.
## This warning is displayed once every 8 hours.

## [1] "SELECT * FROM temp_schema.eqbk52o3children ;"
## attr(,"sqlDialect")
## [1] "oracle"
```

ユーザーは `temp_schema` に書き込み権限を持っている必要があります。

また、Oracleではテーブル名の長さが30文字に制限されているため、テンポラリテーブル。

さらに、Oracleではテンポラリテーブルは自動的に削除されないため、使用後に明示的に `TRUNCATE` および `DROP` して、孤立したテーブルがOracleの一時スキーマに蓄積しないよう

暗黙の型変換

SQL Serverが他の言語よりも特異である数少ない点の1つは、暗黙の型変換が許可されてServerで動作します：

```
CREATE TABLE #temp (txt VARCHAR);

INSERT INTO #temp
SELECT '1';

SELECT * FROM #temp WHERE txt = 1;
```

`txt` がVARCHARフィールドで、それを整数と比較しているとしても、SQL Serverは比較を可能にするために、2つのうちの1つを自動的に正しい型に変換します。これがやって、キャストは常に明示的に行う必要があります。上記の例では、最後のステー

```
SELECT * FROM #temp WHERE txt = CAST(1 AS VARCHAR);
```

または

```
SELECT * FROM #temp WHERE CAST(txt AS INT) = 1;
```

文字列比較における大文字・小文字の区別

SQL Serverなどの一部のDBMSプラットフォームは常に大文字と小文字を区別せずに文字

```
SELECT * FROM concept WHERE concept_class_id = 'Clinical Finding'
```

代わりに以下のように記述することが推奨されます：

```
SELECT * FROM concept WHERE LOWER(concept_class_id) = 'clinical finding'
```

スキーマとデータベース

SQL Serverでは、テーブルはスキーマ内にあり、スキーマはデータベース内にあります。例えは cdm_data データベース内の dbo スキーマ内の person テーブルを指します。他の言語でも同 Serverでは、データベースごとに通常1つのスキーマ (dbo と呼ばれることが多い) が存在し、Serverのデータベースに相当するものがスキーマであると言えます。そのため、SQL Serverのデータベースとスキーマを1つのパラメータに結合することを推奨します。通常、これ @databaseSchema と呼びます。例えば、パラメータ化されたSQLでは次のようにになります：

```
SELECT * FROM @databaseSchema.person
```

SQL Serverでは、値にデータベース名とスキーマ名の両方を含めることができます：database
= "cdm_data.dbo"。他のプラットフォームでは、同じコードを使用し、パラメータ値として
= "cdm_data"。

この方法が失敗する唯一の状況は USE コマンドを使用した場合です。USE
cdm_data.dbo; エラーが発生します。したがって、常にデータベース/スキーマを指定してテー
コマンドの使用を避けることを推奨します。

パラメータ化されたSQLのデバッグ

パラメータ化されたSQLのデバッグは少し複雑になることがあります。レンダリングされたSQL

ソースのSQLをインタラクティブに編集し、レンダリングおよび翻訳されたSQLを生成するた
SqlRender パッケージに含まれています。このアプリは次の方法で起動できます：

```
launchSqlRenderDeveloper()
```

これにより、図 9.1 に示すように、アプリがデフォルトのブラウザで開きます。アプリはウェ
このアプリでは、OHDSI SQLを入力し、ターゲットの方言を選択し、SQLに表示されるパラメ

9.2 DatabaseConnector

DatabaseConnectorは、JavaのJDBC ドライバを使用してさまざまなデータベースプラットフ
R Archive Network) で入手可能で、次のようにインストールできます：

¹<http://data.ohdsi.org/SqlDeveloper/>

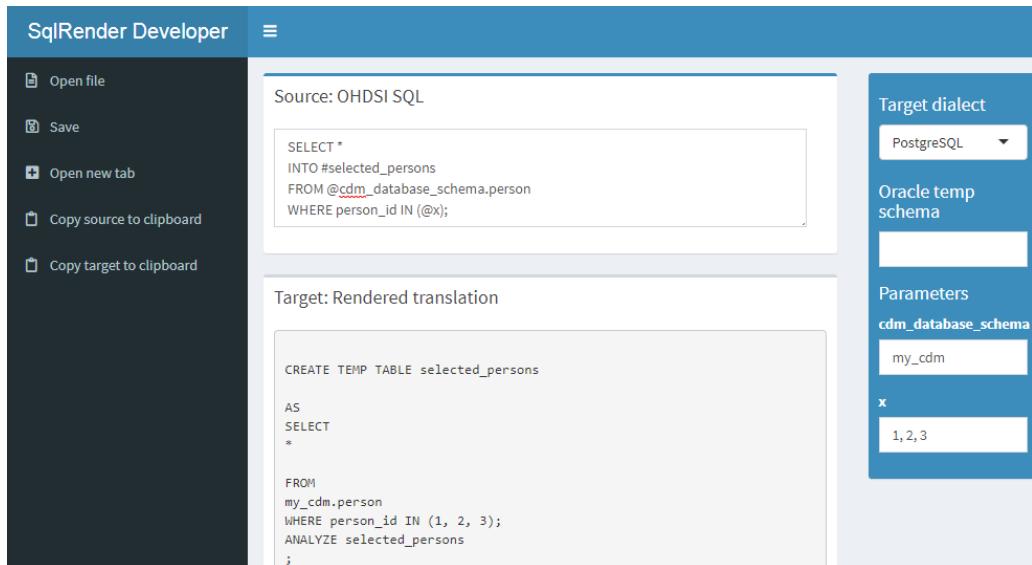


Figure 9.1: The SqlDeveloper Shiny app.

```
install.packages("DatabaseConnector")
```

DatabaseConnectorは、従来のデータベースシステム（PostgreSQL、Microsoft SQL Server、SQLite、およびOracle）、並列データウェアハウス（Microsoft APS、IBM Netezza、Amazon）、ならびにビッグデータプラットフォーム（Hadoopを介したBigQuery）など、広範な技術プラットフォームをサポートしています。このパッケージは

9.2.1

データベースに接続するには、データベースプラットフォーム、サーバーの位置、ユーザ名

```
conn <- connect(dbms = "postgresql",
                 server = "localhost/postgres",
                 user = "joe",
                 password = "secret",
                 schema = "cdm")
```

```
## Connecting using PostgreSQL driver
```

各プラットフォームに必要な詳細情報については、?connectを参照ください。接続を閉じるには

```
disconnect(conn)
```

サーバー名を指定する代わりに、JDBC接続文字列を提供することも可能です。さらに便

```

connString <- "jdbc:postgresql://localhost:5432/postgres"
conn <- connect(dbms = "postgresql",
                connectionString = connString,
                user = "joe",
                password = "secret",
                schema = "cdm")

```

Connecting using PostgreSQL driver

場合によっては、接続の詳細を先に指定し、接続を後にしたい場合もあるでしょう。例えば、

```

details <- createConnectionDetails(dbms = "postgresql",
                                      server = "localhost/postgres",
                                      user = "joe",
                                      password = "secret",
                                      schema = "cdm")
conn <- connect(details)

```

Connecting using PostgreSQL driver

9.2.2

データベースにクエリを実行するための主な関数は、`querySql`と`executeSql`です。`querySql`

いくつかの例を挙げます：

```
querySql(conn, "SELECT TOP 3 * FROM person")
```

```

##   person_id gender_concept_id year_of_birth
## 1          1              8507      1975
## 2          2              8507      1976
## 3          3              8507      1977

```

```
executeSql(conn, "TRUNCATE TABLE foo; DROP TABLE foo;")
```

どちらの関数も広範なエラーレポートを提供します：サーバーによってエラーが発生した場合

9.2.3 ffdf

データベースから取得するデータがメモリに収まりきらないほど大きい場合もあります。セクターで述べたように、そのような場合にはffパッケージを使用してRデータオブジェクトをファイル

```
x <- querySql.ffdf(conn, "SELECT * FROM person")
```

xは現在ffdfオブジェクトです。

9.2.4 SQL

SqlRenderパッケージのrenderとtranslate関数を最初に呼び出す便利な関数があります

```
x <- renderTranslateQuerySql(conn,
                               sql = "SELECT TOP 10 * FROM @schema.person",
                               schema = "cdm_synpuf")
```

SQL Server固有の「TOP 10」構文は、PostgreSQLでは「LIMIT 10」などに変換され、SQL Server固有の「TOP 10」構文は、PostgreSQLでは「LIMIT 10」などに変換され、SQL

9.2.5

データをデータベースに挿入するにはexecuteSql関数を使用してSQLステートメントを送ります

```
data(mtcars)
insertTable(conn, "mtcars", mtcars, createTable = TRUE)
```

この例では、mtcarsデータフレームをサーバー上の「mtcars」というテーブルにアップロードします

9.3 CDM

以下の例では、OHDSI SQLを使用してCDMに準拠したデータベースにクエリを実行します。データベースに何人の人がいるかをクエリで取得してみましょう：

```
SELECT COUNT(*) AS person_count FROM @cdm.person;
```

| PERSON_COUNT |
|--------------|
| 26299001 |

あるいは、観察期間の平均的な長さに興味があるのかもしれません：

```
SELECT AVG(DATEDIFF(DAY,
                     observation_period_start_date,
                     observation_period_end_date) / 365.25) AS num_years
FROM @cdm.observation_period;
```

| NUM_YEARS |
|-----------|
| 1.980803 |

テーブルを結合して追加の統計を生成することができます。結合は通常、テーブル内の特定の

```
SELECT MAX(YEAR(observation_period_end_date) - year_of_birth) AS max_age
FROM @cdm.person
INNER JOIN @cdm.observation_period
    ON person.person_id = observation_period.person_id;
```

| MAX_AGE |
|---------|
| 90 |

観察開始時の年齢分布を決定するには、はるかに複雑なクエリが必要です。このクエリでは、... ASを使用)、“ages”と呼びます。これにより、agesを既存のテーブルであるかのように<.50*nの最小年齢を使用します。最小年齢と最大年齢は別々に計算されます：

```
WITH ages
AS (
    SELECT age,
        ROW_NUMBER() OVER (
            ORDER BY age
        ) order_nr
    FROM (
        SELECT YEAR(observation_period_start_date) - year_of_birth AS age
        FROM @cdm.person
        INNER JOIN @cdm.observation_period
            ON person.person_id = observation_period.person_id
    ) age_computed
)
SELECT MIN(age) AS min_age,
    MIN(CASE
        WHEN order_nr < .25 * n
            THEN 9999
        ELSE age
        END) AS q25_age,
    MIN(CASE
        WHEN order_nr < .50 * n
            THEN 9999
        ELSE age
        END) AS median_age,
    MIN(CASE
        WHEN order_nr < .75 * n
```

```

        THEN 9999
    ELSE age
END) AS q75_age,
MAX(age) AS max_age
FROM ages
CROSS JOIN (
    SELECT COUNT(*) AS n
    FROM ages
) population_size;

```

| MIN_AGE | Q25_AGE | MEDIAN_AGE | Q75_AGE | MAX_AGE |
|---------|---------|------------|---------|---------|
| 0 | 6 | 17 | 34 | 90 |

より複雑な計算は、SQLの代わりにRを使用して行うこともできます。例えば、同じ結果

```

sql <- "SELECT YEAR(observation_period_start_date) -
        year_of_birth AS age
FROM @cdm.person
INNER JOIN @cdm.observation_period
    ON person.person_id = observation_period.person_id;"
age <- renderTranslateQuerySql(conn, sql, cdm = "cdm")
quantile(age[, 1], c(0, 0.25, 0.5, 0.75, 1))

```

```

##   0%   25%   50%   75% 100%
##   0     6    17    34    90

```

ここでは、サーバー上で年齢を計算し、すべての年齢をダウンロードし、年齢分布を計算

クエリでは、CDM内のソース値を使用することができます。例えば、最も頻度の高いコ

```

SELECT TOP 10 condition_source_value,
    COUNT(*) AS code_count
FROM @cdm.condition_occurrence
GROUP BY condition_source_value
ORDER BY -COUNT(*);

```

| CONDITION_SOURCE_VALUE | CODE_COUNT |
|------------------------|------------|
| 4019 | 49094668 |
| 25000 | 36149139 |
| 78099 | 28908399 |
| 319 | 25798284 |
| 31401 | 22547122 |
| 317 | 22453999 |

| CONDITION_SOURCE_VALUE | CODE_COUNT |
|------------------------|------------|
| 311 | 19626574 |
| 496 | 19570098 |
| I10 | 19453451 |
| 3180 | 18973883 |

ここでは、CONDITION_OCCURRENCEテーブル内のCONDITION_SOURCE_VALUEフィールド

9.4

多くの操作では、ボキャブラリが有用です。ボキャブラリテーブルはCDMの一部であり、SQ

```
SELECT COUNT(*) AS subject_count,
       concept_name
  FROM @cdm.person
 INNER JOIN @cdm.concept
    ON person.gender_concept_id = concept.concept_id
 GROUP BY concept_name;
```

| SUBJECT_COUNT | CONCEPT_NAME |
|---------------|--------------|
| 14927548 | FEMALE |
| 11371453 | MALE |

ボキャブラリの非常に強力な機能の一つは、その階層構造です。よくあるクエリは、特定の概念のすべての子孫を探すものです。例えば、イブプロフェンという成分を含む処方件数を数え

```
SELECT COUNT(*) AS prescription_count
  FROM @cdm.drug_exposure
 INNER JOIN @cdm.concept_ancestor
    ON drug_concept_id = descendant_concept_id
 INNER JOIN @cdm.concept ingredient
    ON ancestor_concept_id = ingredient.concept_id
 WHERE LOWER(ingredient.concept_name) = 'ibuprofen'
   AND ingredient.concept_class_id = 'Ingredient'
   AND ingredient.standard_concept = 'S';
```

| PRESCRIPTION_COUNT |
|--------------------|
| 26871214 |

9.5 QueryLibrary

QueryLibraryは、CDM用の一般に使用されるSQLクエリのライブラリです。これはオンラインで利用可能で、CDM用のSQLクエリを簡単に実行できます。

The screenshot shows the QueryLibrary interface. On the left, there is a search bar and a table listing various SQL queries. The table has two columns: 'Group' and 'Name'. The 'Group' column contains categories like 'drug exposure', and the 'Name' column contains the names of the queries, such as 'DEX01 Counts of persons with any number of exposures to a certain drug'. On the right, there is a detailed view of one query, 'DEX01', with sections for 'Query Description', 'Description', and 'Query'.

| Group | Name |
|-------------------|---|
| ["drug exposure"] | All |
| drug exposure | DEX01 Counts of persons with any number of exposures to a certain drug |
| drug exposure | DEX02 Counts of persons taking a drug, by age, gender, and year of exposure |
| drug exposure | DEX03 Distribution of age, stratified by drug |
| drug exposure | DEX04 Distribution of gender in persons taking a drug |
| drug exposure | DEX05 Counts of drug records for a particular drug |
| drug exposure | DEX06 Counts of distinct drugs in the database |
| drug exposure | DEX07 Maximum number of drug exposure events per person over some time period |

Query Description

DEX01: Counts of persons with any number of exposures to a certain drug

Description

This query is used to count the persons with at least one exposures to a certain drug (drug_concept_id). See vocabulary queries for obtaining valid drug_concept_id values. The input to the query is a value (or a comma-separated list of values) of a drug_concept_id. If the input is omitted, all drugs in the data table are summarized.

Query

The following is a sample run of the query. The input parameters are highlighted in blue.

```

SELECT
    c.concept_name,
    drug_concept_id,
    COUNT(person_id) AS num_persons
FROM cdm.drug_exposure
INNER JOIN cdm.concept c
ON drug_concept_id = c.concept_id
WHERE domain_id = 'Drug'
  
```

Figure 9.2: クエリライブラリ：CDMに対するSQLクエリのライブラリ。

このライブラリの目的は、新しいユーザーがCDMのクエリ方法を学習するのを支援する。
QueryLibraryは、SqlRenderを利用して、選択したSQL方言でクエリを実行します。ユーザーはクエリを入力し、実行ボタンをクリックすれば、そのSQL文が自動的に選択された方言に変換され、実行結果が表示されます。

9.6

9.6.1

血管性浮腫は、ACE阻害薬（ACEi）のよく知られた副作用です。Slater et al. (1988)によると、ACEi治療開始後1週間の血管性浮腫の発症率は3,000人中1例/週と推定されています。²

リシノプリル投与開始後の最初の1週間での血管性浮腫の発生率は、年齢と性別で以下のようになります。

9.6.2

曝露をリシノプリルへの初回の曝露として定義します。初回とは、以前にリシノプリルへ

²<http://data.ohdsi.org/QueryLibrary>

³<https://github.com/OHDSI/QueryLibrary>

9.6.3

血管性浮腫は、入院中または救急外来（ER）受診中に血管性浮腫の診断コードが記録された場

9.6.4

治療開始後の最初の1週間の発症率を計算します。患者が1週間にわたって継続的に曝露された

9.7 SQL R

OHDSIツールの慣例に縛られることはありませんが、同じ原則に従うことは有益です。この場

```
library(DatabaseConnector)
conn <- connect(dbms = "postgresql",
                 server = "localhost/postgres",
                 user = "joe",
                 password = "secret")
cdmDbSchema <- "cdm"
cohortDbSchema <- "scratch"
cohortTable <- "my_cohorts"

sql <- "
CREATE TABLE @cohort_db_schema.@cohort_table (
  cohort_definition_id INT,
  cohort_start_date DATE,
  cohort_end_date DATE,
  subject_id BIGINT
);
"
renderTranslateExecuteSql(conn, sql,
                          cohort_db_schema = cohortDbSchema,
                          cohort_table = cohortTable)
```

ここでは、データベーススキーマとテーブル名をパラメータ化しています。異なる環境に簡単

9.7.1

次に、曝露コホートを作成し、COHORTテーブルに挿入します：

```
sql <- "
INSERT INTO @cohort_db_schema.@cohort_table (
  cohort_definition_id,
  cohort_start_date,
  cohort_end_date,
  subject_id
)"
```

```

SELECT 1 AS cohort_definition_id,
       cohort_start_date,
       cohort_end_date,
       subject_id
FROM (
  SELECT drug_era_start_date AS cohort_start_date,
         drug_era_end_date AS cohort_end_date,
         person_id AS subject_id
  FROM (
    SELECT drug_era_start_date,
           drug_era_end_date,
           person_id,
           ROW_NUMBER() OVER (
             PARTITION BY person_id
             ORDER BY drug_era_start_date
           ) order_nr
    FROM @cdm_db_schema.drug_era
    WHERE drug_concept_id = 1308216 --
  ) ordered_exposures
  WHERE order_nr = 1
) first_era
INNER JOIN @cdm_db_schema.observation_period
  ON subject_id = person_id
  AND observation_period_start_date < cohort_start_date
  AND observation_period_end_date > cohort_start_date
WHERE DATEDIFF(DAY,
                observation_period_start_date,
                cohort_start_date) >= 365;
"
renderTranslateExecuteSql(conn, sql,
                        cohort_db_schema = cohortDbSchema,
                        cohort_table = cohortTable,
                        cdm_db_schema = cdmDbSchema)

```

ここでは、CDMの標準テーブルであるDRUG ERAテーブルを使用します。このテーブルに結合し、1人当たりの最初の薬物曝露を取り出します。1人の患者が複数の観察と COHORT_START_DATE の間には、少なくとも 365 日の間隔が必要となります。

9.7.2

最後に、アウトカムコホートを作成する必要があります：

```

sql <- "
INSERT INTO @cohort_db_schema.@cohort_table (
  cohort_definition_id,
  cohort_start_date,

```

```

cohort_end_date,
subject_id
)
SELECT 2 AS cohort_definition_id,
cohort_start_date,
cohort_end_date,
subject_id
FROM (
  SELECT DISTINCT person_id AS subject_id,
    condition_start_date AS cohort_start_date,
    condition_end_date AS cohort_end_date
  FROM @cdm_db_schema.condition_occurrence
  INNER JOIN @cdm_db_schema.concept_ancestor
    ON condition_concept_id = descendant_concept_id
  WHERE ancestor_concept_id = 432791 --
) distinct_occurrence
INNER JOIN @cdm_db_schema.visit_occurrence
  ON subject_id = person_id
  AND visit_start_date <= cohort_start_date
  AND visit_end_date >= cohort_start_date
WHERE visit_concept_id IN (262, 9203,
  9201) -- ER;
"

```

```

renderTranslateExecuteSql(conn, sql,
  cohort_db_schema = cohortDbSchema,
  cohort_table = cohortTable,
  cdm_db_schema = cdmDbSchema)

```

ここでは、CONDITION_OCCURRENCEテーブルをCONCEPT_ANCESTORテーブルと結合し

9.7.3

コホートが設定されたので、年齢と性別で層別化された発症率を計算できます：

```

sql <- "
WITH tar AS (
  SELECT concept_name AS gender,
    FLOOR((YEAR(cohort_start_date) -
      year_of_birth) / 10) AS age,
    subject_id,
    cohort_start_date,
    CASE WHEN DATEADD(DAY, 7, cohort_start_date) >
      observation_period_end_date
    THEN observation_period_end_date
    ELSE DATEADD(DAY, 7, cohort_start_date)
    END AS cohort_end_date

```

```

        FROM @cohort_db_schema.@cohort_table
        INNER JOIN @cdm_db_schema.observation_period
            ON subject_id = observation_period.person_id
            AND observation_period_start_date < cohort_start_date
            AND observation_period_end_date > cohort_start_date
        INNER JOIN @cdm_db_schema.person
            ON subject_id = person.person_id
        INNER JOIN @cdm_db_schema.concept
            ON gender_concept_id = concept_id
        WHERE cohort_definition_id = 1 --
    )
SELECT days.gender,
       days.age,
       days,
       CASE WHEN events IS NULL THEN 0 ELSE events END AS events
FROM (
    SELECT gender,
           age,
           SUM(DATEDIFF(DAY, cohort_start_date,
                         cohort_end_date)) AS days
    FROM tar
    GROUP BY gender,
             age
) days
LEFT JOIN (
    SELECT gender,
           age,
           COUNT(*) AS events
    FROM tar
    INNER JOIN @cohort_db_schema.@cohort_table angioedema
        ON tar.subject_id = angioedema.subject_id
        AND tar.cohort_start_date <= angioedema.cohort_start_date
        AND tar.cohort_end_date >= angioedema.cohort_start_date
    WHERE cohort_definition_id = 2 --
    GROUP BY gender,
             age
) events
ON days.gender = events.gender
   AND days.age = events.age;
"
results <- renderTranslateQuerySql(conn, sql,
                                    cohort_db_schema = cohortDbSchema,
                                    cohort_table = cohortTable,
                                    cdm_db_schema = cdmDbSchema,
                                    snakeCaseToCamelCase = TRUE)

```

まず、CTE「tar」を作成し、適切なリスク時間を伴うすべての曝露を含めます。OBSERV

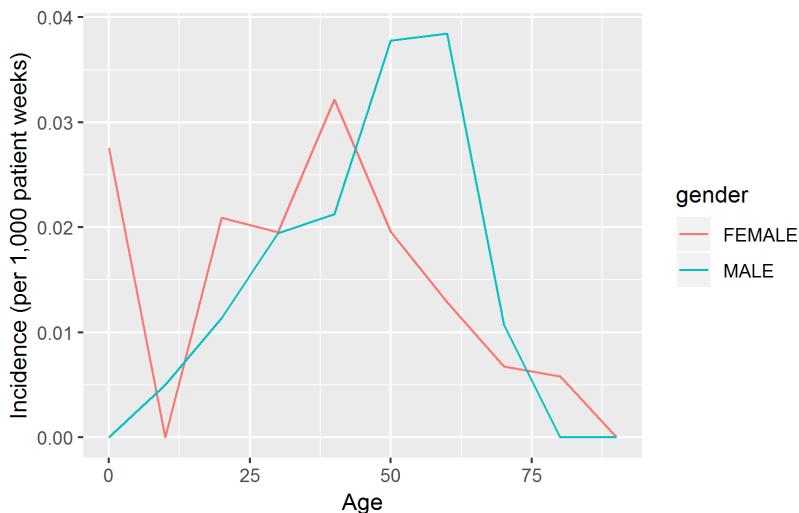
`snakeCaseToCamelCase = TRUE` を用いるのは、SQLではフィールド名に`snake_case`を使用する傾向がある（SQLは大文字と小文字を区別しないため）のに対し、Rでは`camelCase`

ggplot2パッケージを使用すると、結果を簡単にプロットできます：

```
#   IR
results$ir <- 1000 * results$events / results$days / 7

#
results$age <- results$age * 10

library(ggplot2)
ggplot(results, aes(x = age, y = ir, group = gender, color = gender)) +
  geom_line() +
  xlab(" ") +
  ylab(" 1,000 / ")
```



9.7.4

作成したテーブルをクリーンアップし、忘れずに接続を閉じます：

```
sql <- "
TRUNCATE TABLE @cohort_db_schema.@cohort_table;
DROP TABLE @cohort_db_schema.@cohort_table;
"
renderTranslateExecuteSql(conn, sql,
                          cohort_db_schema = cohortDbSchema,
                          cohort_table = cohortTable)

disconnect(conn)
```

9.7.5

OHDSI SQLとDatabaseConnectorとSqlRenderを組み合わせて使用するため、ここで紹介します。デモンストレーション用に、手作業でSQLを使用してコホートを作成することにしました。SQLを生成するため、SqlRenderとDatabaseConnectorと簡単に併用することができます。

9.8



- SQL (Structured Query Language) は、共通データモデル (CDM) に準拠します。
- 異なるデータベースプラットフォームは異なるSQL表現を持っており、照会する方法が異なります。
- SqlRenderとDatabaseConnectorRパッケージは、CDM内のデータを照会する方法を定義しています。
- RとSQLを併用することで、OHDSIツールではサポートされていないカスタムクエリを実現できます。
- QueryLibraryは、CDM用の再利用可能なSQLクエリのコレクションを提供します。

9.9

前提条件

これらの演習では、セクション 8.4.5 に記載されているように、R、R-Studio、Java がインストールされていることを前提とします。また、SqlRender、DatabaseConnector、Eunomia パッケージも必要です。以下の手順でインストールできます。

```
install.packages(c("SqlRender", "DatabaseConnector", "remotes"))
remotes::install_github("ohdsi/Eunomia", ref = "v1.0.0")
```

Eunomia パッケージは、CDM 内でローカル R セッション内で動作するシミュレートされたデータベースです。

```
connectionDetails <- Eunomia::getEunomiaConnectionDetails()
```

CDM データベースのスキーマは「main」です。

演習 9.1. SQL と R を使用して、データベース内に何人いるかを計算します。

演習 9.2. SQL と R を使用して、セレコキシブの処方を少なくとも 1 回受けたことがある人の数を計算します。

演習 9.3. SQL と R を使用して、セレコキシブの服用中に消化管出血と診断された人の人数を計算します。消化管出血のコンセプト ID は 192671 です)

推奨される解答は付録 E.5 を参照ください。

Chapter 10

著者: Kristin Kostka

観察型健康データ（リアルワールドデータとも呼ばれる）は、患者の健康状態や医療の提供を記録するためのデータです。例えば、医療保険請求データベースは、ある症状（例：血管性浮腫）に対して提供された治療情報を記録します。本章では、コホート定義の作成と共有とは何か、コホートを開発する方法、ATLASまたは他のツールを使用してデータを分析する方法について学びます。

10.1

OHDSI研究では、コホートを、ある一定期間に1つ以上の適格基準を満たす人々の集合体と定義します。



コホートは、ある一定期間に1つ以上の適格基準を満たす人々の集合体です。

OHDSIで使用されているコホートの定義は、この分野の他の人々が使用するものとは異なります。たとえば、ICD-10コード（ICD-9/CM、ICD-10、NDC、HCPCSなど）に類似しているとされています。コードセットはコホートを構築するための基準となります（たとえば、ICD-10コードの初回の発生か？それとも発生すべてか？）。明確に定義されたコホートは、複数の基準を満たす人々の集合体です。

OHDSIのコホート定義を利用するためのユニークなニュアンスには以下があります。

- 一人の人が複数のコホートに属する可能性があります。
- 一人の患者が複数の異なる期間に同じコホートに属する可能性があります。
- 一人の患者が同じ期間内に同じコホートに複数回属することはありません。
- コホートにメンバーがゼロまたは複数含まれる場合があります。

コホートを構築するための主なアプローチは二つあります：

1. ルールベースのコホート定義は、患者がコホートにいる時期を明示的なルールで説明する方法です。
2. 確率ベースのコホート定義は、患者がコホートに属する患者の確率（0から100%の範囲）を示す方法です。

次のセクションでは、これらのアプローチについて詳しく説明します。

10.2

ルールベースのコホート定義は、特定の期間（例：「過去6ヶ月以内にその状態を発症した人」）これらを構成する際に使用する標準的な構成要素は以下の通りです：

- ・ ドメイン：データが格納されているCDMドメイン（例：「処置（プロシージャ）」の発生）
- ・ コンセプトセット：対象とする臨床実態を包含する一つ以上の標準コンセプトを定義する
- ・ ドメイン固有の属性：関心のある臨床実態に関連する追加の属性（例：DRUG_EXPOSURE）
- ・ 時間的なロジック：適格基準とイベントの関係が評価される時間間隔（例：指定された状況）

コホート定義を構築する際、コホート属性を表すビルディングブロックのようにドメインを考慮



Figure 10.1: コホート定義のビルディングブロック

コホート定義作成時に自問すべきいくつかの質問があります：

- ・ コホート組入れの時間を定義する初期イベントは何か？
- ・ 初期イベントに適用される適格基準は何か？
- ・ コホート離脱を定義するものは何か？

コホート組入れイベント：コホート組入れイベント（初期イベント）は、人々がコホートに参加する開始点を定義する

適格基準：適格基準は、初期イベントコホートに適用され、さらに人々のセットを制限します

コホート離脱基準：コホート離脱イベントは、ある人がコホートメンバーとしての資格を失う



OHDSIツールでは、適格基準と除外基準の区別はありません。すべての基準は適格

10.3

コンセプトセットは、さまざまな分析で再利用可能なコンポーネントとして使用できる二

- ・除外：このコンセプト（および選択されている場合はその下位層に含まれるもの）
- ・下位層に含まれる：このコンセプトだけでなく、その下位層に含まれるものも考慮
- ・マッピング済み：標準化されていないコンセプトの検索を許可します。

例えば、コンセプトセットの表現は、図に示されるように2つのコンセプトを含むことが10.1）。ここでは、コンセプト4329847（「心筋梗塞」）とそのすべての下位層に含ま

Table 10.1: コンセプトセットの表現の例

| コンセプトID | コンセプト名 | 除外 | 子孫 | マッピング対象 |
|---------|---------|-----|----|---------|
| 4329847 | 心筋梗塞 | いいえ | はい | いいえ |
| 314666 | 陳旧性心筋梗塞 | はい | はい | いいえ |

図に示すように（図10.2）、これは「心筋梗塞」およびその下位層に含まれるもの（ICD-9およびICD-10コード）を反映します。

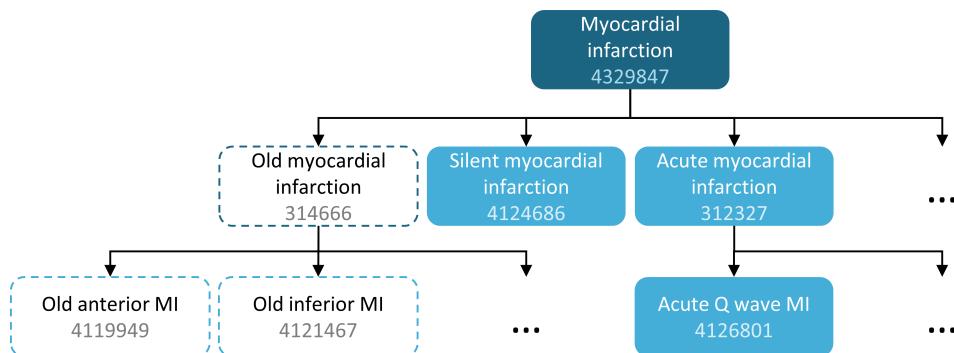


Figure 10.2: 「心筋梗塞」（下位層を含む）を含むが、「陳旧性心筋梗塞」（下位層を含

10.4

ルールベースのコホート定義は、コホート定義を組み立てるための一般的な方法です。し

このアプローチをCDMデータに適用した例として、APHRODITE（Automated PHenotype Routine for Observational Definition, Identification, Training and Evaluation）Rパッケージ¹があります。このパッケージは、不完全にラベル付けされたデータ（Banda et al., 2017）。

10.5

コホートを構築する際、次のどちらが重要かを考慮すべきです：適格患者をすべて見つけることそれとも 確信を持てる患者のみを組み入れることが重要か？

コホートの構築戦略は、専門家の合意が疾患をどのように定義するかという臨床的な厳格性に本章の冒頭で述べたように、コホート定義は記録されたデータから観察したいことを推測しよ（ケースの手動チャートレビュー）と比較するテストを考えることができます。詳細は第16（「臨床的妥当性」）で詳しく説明しています。

10.5.1 OHDSI

既存のコホート定義とアルゴリズムのインベントリーと全体的な評価を支援するために、OHDSI ゴールドスタンダードフェノタイプライブラリ（GSPL）ワークグループが設立されました。このワークグループの研究には、前のセクションで議論されたAPHRODITE（Banda et al., 2017）やPheValuatorツール（Swerdel et al., 2019）の他、OHDSIネットワーク全eMERGE Phenotype Library を共有するための取り組みなどが含まれます（Hripcsak et al., 2019）。表現型のキュレーションに関心がある場合は、このワークグループへ

10.6

コホート定義をルールベースのアプローチでまとめることによって、コホートスキルの練習をこのコンテキストを念頭に、コホートを構築します。この演習を通して、標準的な脱落チャート10.3は、このコホートをどのように構築するかの論理的なフレームワークを示しています。コホートはATLASのユーザーインターフェースで作成することも、CDMに対して直接クエリを

10.7 ATLAS

まずATLASで始めるには、 Cohort Definitions モジュールをクリックします。モジュールを読み込む「New cohort」をクリックします。次の画面では空のコホート定義が表示されます。図10.4に示す内

まず最初に、「New Cohort Definition」からコホートの名前を固有の名前に変更することをお

¹<https://github.com/OHDSI/Aphrodite>

²<https://www.ohdsi.org/web/wiki/doku.php?id=projects:workgroups:gold-library-wg>

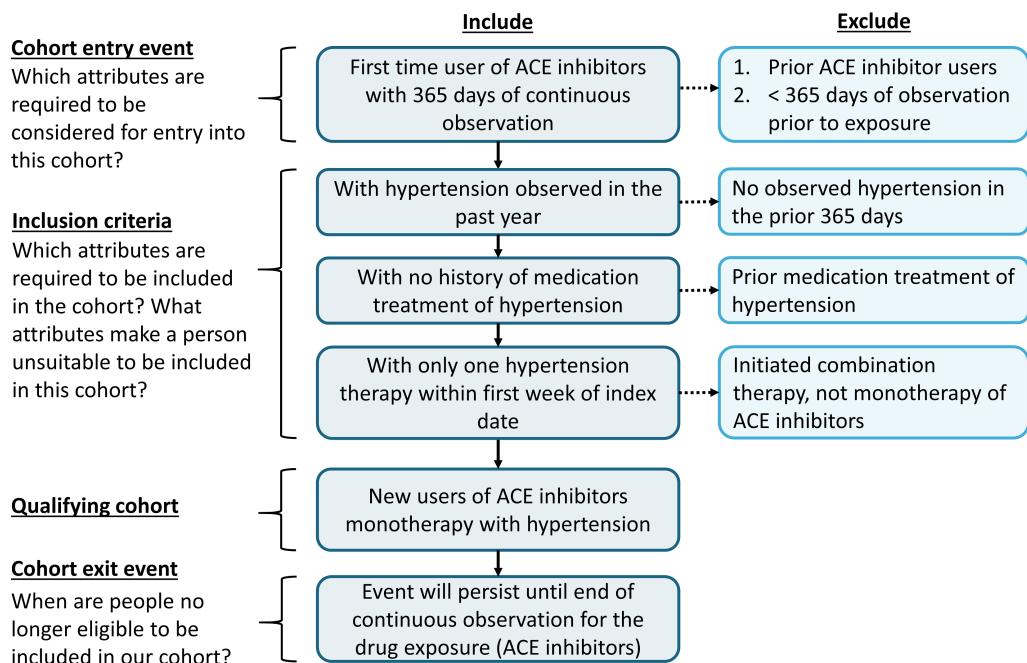


Figure 10.3: 目標とするコホートの論理図

New Cohort Definition

New Cohort Definition

Definition Concept Sets Generation Reporting Export

enter a cohort definition description here

Cohort Entry Events

Events having any of the following criteria:

with continuous observation of at least days before and days after event index date

Limit initial events to: earliest event per person.

+ Add Initial Event

Restrict initial events

Inclusion Criteria

New inclusion criteria

Limit qualifying events to: earliest event per person.

Figure 10.4: 新しいコホート定義



ATLASは二つのコホートに全く同じ名前を付けることはできません。他のATLASコホートには同じ名前が付いています。

名前を入力後、をクリックしてコホートを保存します。

10.7.1

では、初期コホートイベントの定義に進みます。「Add initial event」をクリックします。どの

The screenshot shows the ATLAS Cohort Entry Events interface. At the top, there is a header bar with tabs: Definition (selected), Concept Sets, Generation, Reporting, and Export. Below the header is a search bar with placeholder text 'enter a cohort definition description here'. The main area is titled 'Cohort Entry Events' and contains sections for 'Events having any of the following criteria:' and 'Inclusion Criteria'. Under 'Events having any of the following criteria:', there is a dropdown for 'with continuous observation of at least [0] days before and [0] days after event index date' and a dropdown for 'Limit initial events to: earliest event per person'. A green button labeled 'Restrict initial events' is visible. On the right side, a sidebar lists various criteria with descriptions: Add Condition Era (Find patients with specific diagnosis era), Add Condition Occurrence (Find patients with specific diagnoses), Add Death (Find patients based on death), Add Device Exposure (Find patients based on device exposure), Add Dose Era (Find patients with dose eras), Add Drug Era (Find patients with exposure to drugs over time), Add Drug Exposure (Find patients with exposure to specific drugs or drug classes). A blue button labeled '+ Add Initial Event' is located in the top right corner of the main section.

Figure 10.5: 初期イベントの追加

図10.5に示されているように、ATLASは各基準の説明を提供しています。もしCONDITION_C「Drug Exposure」をクリックします。

画面は選択した基準を表示するように更新されますが、まだ終了ではありません。図10.6を参考

The screenshot shows the ATLAS Cohort Entry Events interface with the 'Drug Exposure' criteria selected. The main section now displays the details for 'Drug Exposure': 'a drug exposure of Any Drug' and 'with continuous observation of at least [0] days before and [0] days after event index date'. A red 'Delete Criteria' button is visible next to the 'Add attribute...' button. The sidebar on the right still lists other criteria with their descriptions.

Figure 10.6: 薬剤曝露の定義

10.7.2

コンセプトセットを定義するためには、▼をクリックして、ACE阻害薬を定義するための

シナリオ1: コンセプトセットを構築していない場合

基準に適用するコンセプトセットをまだ作成していない場合は、先にそれを行う必要がある。 「Concept set」タブに移動し、「New Concept Set」をクリックしてコンセプトセットを作成する。 「Concept Set」から任意の名前に変更する必要があります。そこから、**Search** モジュール

The screenshot shows the ATLAS search interface. At the top, there's a navigation bar with a back arrow, the text 'EXAMPLE: new users of ACE inhibitors as first-line mono-therapy for hypertension > ACE Inhibitors', and a search icon. Below that is a search bar with the placeholder 'Search' and a dropdown menu with 'Import'. The main area has a search bar containing 'ace inhibitors' with a magnifying glass icon. Below it are buttons for 'Column visibility', 'Copy', 'CSV', and 'Show 15 entries'. A 'Filter:' input field is also present. The results table has columns: Vocabulary, Id, Code, Name, Class, RC, DRC, Domain, and Vocabulary. It shows 9 entries, with the first four being from the 'Vocabulary' section and the last five from the 'Class' section. The 'Name' column lists terms like 'ACE inhibitors, plain', 'ACE INHIBITORS, PLAIN', 'ACE inhibitors and diuretics', and 'ACE INHIBITORS, COMBINATIONS'. The 'Class' column indicates they belong to ATC 4th.

| Vocabulary | Id | Code | Name | Class | RC | DRC | Domain | Vocabulary |
|--------------|----------|-------|------------------------------|---------|----|---------|--------|------------|
| ATC (6) | 21601784 | C09AA | ACE inhibitors, plain | ATC 4th | 0 | 507,772 | Drug | ATC |
| Multilex (1) | 21601783 | C09A | ACE INHIBITORS, PLAIN | ATC 3rd | 0 | 507,772 | Drug | ATC |
| VA Class (1) | 21601802 | C09BA | ACE inhibitors and diuretics | ATC 4th | 0 | 10,982 | Drug | ATC |
| LOINC (1) | | | | | | | | |
| Class | 21601801 | C09B | ACE INHIBITORS, COMBINATIONS | ATC 3rd | 0 | 10,982 | Drug | ATC |
| ATC 4th (4) | | | | | | | | |

Figure 10.7: 語彙の検索 - ACE阻害薬

使用したいボキャブラリを見つけたら、をクリックし、そのコンセプトを選択します。

図10.8はコンセプトセット表現を示しています。対象とするすべてのACE阻害薬成分を選択するため、「Concepts」をクリックして、この表現に含まれている21,536のコンセプトすべてを確認できます。また、「Source Codes」をクリックすると、様々なコーディングシステムに含まれるすべてのソースコードが表示されます。

シナリオ2: すでにコンセプトセットを構築している場合

すでにコンセプトセットを作成してATLASに保存している場合、「Import Concept Set」をクリックします。ダイアログボックスが開き、ATLASのコンセプトセット「ACE inhibitors」を入力し、コンセプトセットのリストがマッチングする名前のコンセプトのみが表示されます。コンセプトセットを選択するとダイアログボックスは消えます。) 選択したコンセプトセツトが選択したコンセプトセットに更新されると、この操作が成功したことになります。

10.7.3

コンセプトセットを添付したら、作業終了ではありません。問い合わせでは、新規ユーザーまたは「Add attribute」をクリックします。次に「Add first exposure criteria」を選択します。作成する属性を選択したら、ウィンドウは自動的に閉じます。この追加属性は最初の基準と同じボックスに表示されます。

Concept Set Expression Included Concepts 21536 Included Source Codes Export Import

Name: ACE Inhibitors

Show 25 entries Search: Previous 1 Next

Showing 1 to 15 of 15 entries

| | Concept Id | Concept Code | Concept Name | Domain | Standard Concept Caption | Exclude | Descendants | Mapped |
|----------|------------|--------------|--------------|----------|--------------------------|-------------------------------------|-------------------------------------|--------|
| 1335471 | 18867 | benazepril | Drug | Standard | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | |
| 1340128 | 1998 | Captopril | Drug | Standard | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | |
| 19050216 | 21102 | Cilazapril | Drug | Standard | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | |
| 1341927 | 3827 | Enalapril | Drug | Standard | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | |
| 1342001 | 3829 | Enalaprilat | Drug | Standard | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | |
| 1363749 | 50166 | Fosinopril | Drug | Standard | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | |
| 19122327 | 60245 | imidapril | Drug | Standard | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | |
| 1308216 | 29046 | Lisinopril | Drug | Standard | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | |
| 1310756 | 30131 | moexipril | Drug | Standard | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | |
| 1373225 | 54552 | Perindopril | Drug | Standard | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | |
| 1331235 | 35208 | quinapril | Drug | Standard | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | |
| 1334456 | 35296 | Ramipril | Drug | Standard | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | |
| 19040051 | 36908 | spirapril | Drug | Standard | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | |
| 1342439 | 38454 | trandolapril | Drug | Standard | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | |
| 19102107 | 39990 | zofenopril | Drug | Standard | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | |

Classification Non-Standard Standard

Figure 10.8: ACE阻害薬を含むコンセプトセット

Import Concept Set From Repository...

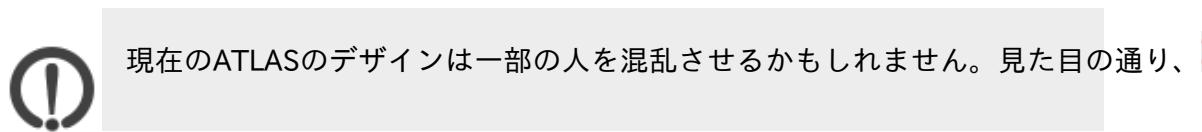
New Concept Set

Show 10 entries Filter Repository Concept Sets: ace inhibitors

| ID | Title | Created | Modified | Author |
|---------|---|---------------------|---------------------|-----------|
| 1794480 | [OHDSI EU 2019] Excluded concepts of ACE inhibitors or Thiazide diuretics | 03/28/2019 11:04 AM | 03/28/2019 11:04 AM | anonymous |
| 963 | ACE Inhibitors | | | anonymous |
| 3268 | COPY OF: ACE Inhibitors | | | anonymous |
| 99283 | Ace Inhibitors | | | anonymous |
| 142965 | PheKB ACE-I ACE inhibitors | | | anonymous |

Showing 1 to 5 of 5 entries (filtered from 11,667 total entries) Previous 1 Next

Figure 10.9: ATLASリポジトリからのコンセプトセットのインポート



これで最初の適格イベントが構築できました。最初に観察された薬剤曝露を確実に捕捉す

The screenshot shows the 'Cohort Entry Events' interface. It displays the following criteria:

- Events having any of the following criteria:
- a drug exposure of **ACE inhibitors**
- X for the first time in the person's history**
- with continuous observation of at least **365** days before and **0** days after event index date
- Limit initial events to: **earliest event** per person.

Buttons include: + Add Initial Event, + Add attribute..., Delete Criteria, and Restrict initial events.

Figure 10.10: インデックス日付前に必要な継続的観察を設定

このロジックがどのように組み合わさるかをさらに説明するため、患者のタイムラインを図10.11では、各線はコホートに参加資格がある可能性のある単一の患者を表しています。

10.7.4

コホートエントリイベントを指定すると、追加の適格イベントを「Restrict initial events」または「New inclusion criteria」のいずれかに追加することができます。この「initial events」に追加基準を加えると、ATLASでカウントを生成するときに、これらすべての「inclusion criteria」に基準を追加すると、追加の包括基準を適用することによって失う患

コホートのメンバーシップに関するロジックをさらに追加するために「New inclusion criteria」をクリックします。このセクションの機能は、前述のコホート基準の「inclusion criteria」をクリックします。基準に名前を付け、必要に応じて探している内容

新しい条件に注釈を付けたら、「+Add criteria to group」ボタンをクリックして、このルールを「Initial Event」と同様に機能します。ただし、初期イベントを指定するわけではありません。「criteria to group」と指定されています。たとえば、疾患を見つけるための方法が複数ある場合、「condition occurrence」を追加します。このレコードにコンセプトセットを添付すること?(fig:ATLASIC1)と照らし合わせてロジックを確認ください。

次に、患者を検索するための別の条件を追加します：インデックス開始日の前日から当日までの「inclusion criteria(新規の選択基準)」「ボタンをクリックし、この条件に注釈を追加し、」「criteria to group(グループに条件を追加)」「をクリックして開始します。これはDRUG_EXPOSUREなので、」Add Drug Exposure(薬剤曝露を追加)「をクリックし、高さ正確に0であることを確認ください。ここで、図10.13で、ロジックを確認します。

「発生なし」が「正確に0回出現」としてコード化される理由がわからないかもしれません。適格条件のみを処理します。特定の属性が存在しないことを指定する場合は、論理演算

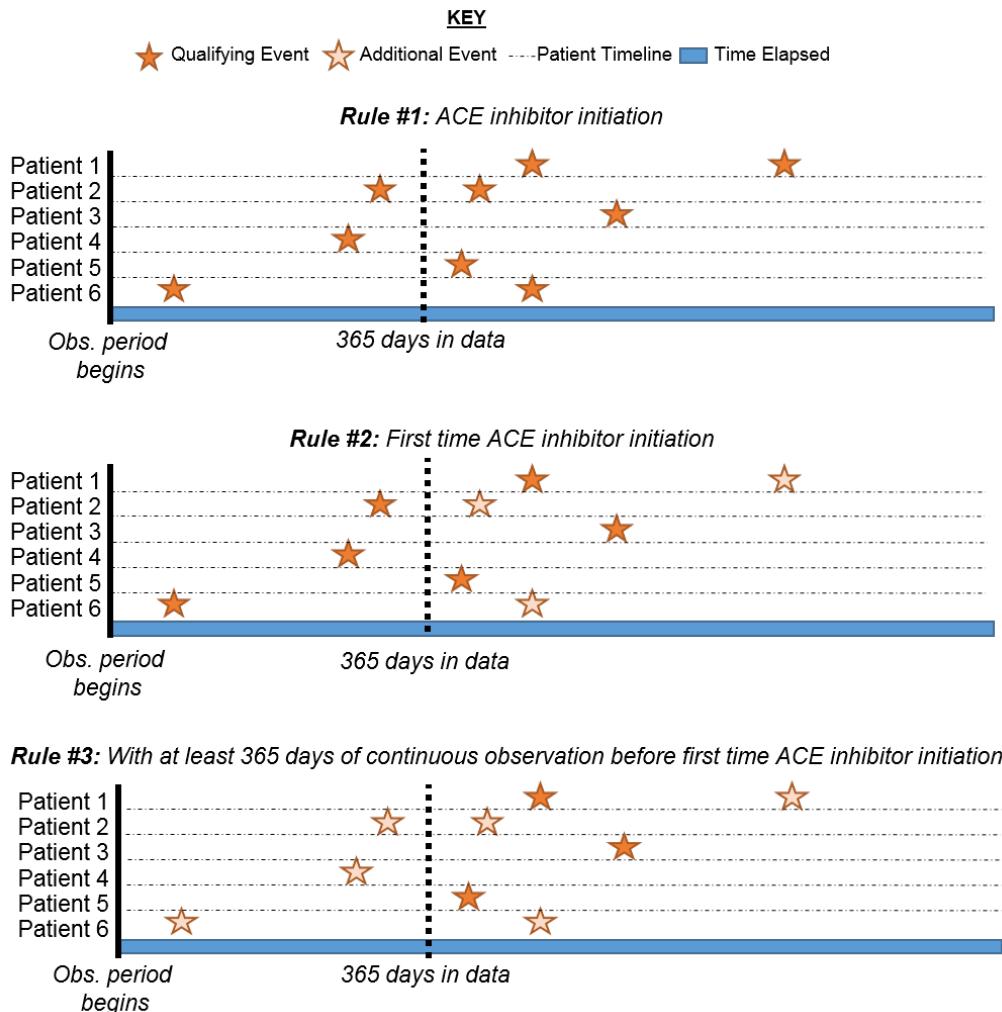


Figure 10.11: 基準の適用による患者の適格性の説明

Inclusion Criteria

New inclusion criteria

has hypertension diagnosis in 1 yr prior to treatment

Copy Delete

1. has hypertension diagnosis in 1 yr prior to treatment

enter an inclusion rule description

having all of the following criteria:

+ Add criteria to group...

with at least 1 using all occurrences of:
a condition occurrence of Hypertensive disorder + Add attribute...

where event starts between 365 days Before and 0 days After index start date add additional constraint

restrict to the same visit occurrence
 allow events from outside observation period

Delete Criteria

Limit qualifying events to: earliest event per person.

Figure 10.12: 追加の選択基準 1

Inclusion Criteria

New inclusion criteria

Has no prior antihypertensive drug exposures in medical history

Copy Delete

1. has hypertension diagnosis in 1 yr prior to treatment

2. Has no prior antihypertensive drug exposures in medical history

enter an inclusion rule description

having all of the following criteria:

+ Add criteria to group...

with exactly 0 using all occurrences of:
a drug exposure of Hypertension drugs + Add attribute...

where event starts between All days Before and 1 days Before index start date add additional constraint

restrict to the same visit occurrence
 allow events from outside observation period

Delete Criteria

Limit qualifying events to: earliest event per person.

Figure 10.13: 追加の選択基準 2

最後に、患者を絞り込むために、もう一つ別の条件を追加します：インデックス開始日の0日前 inclusion criteria(新しい選択基準) “ボタンをクリックし、この条件に注釈を追加してから” criteria to group(グループに条件を追加) “をクリックして開始します。これは DRUG ERAなので、” Add Drug Era(薬剤曝露期間を追加) “をクリックし、高血圧治療薬の2010.14と照らし合わせてロジックを確認します。

The screenshot shows the 'Inclusion Criteria' section of a software application. A green button labeled 'New inclusion criteria' is selected. The main area displays the following rule:

Is only taking ACE as monotherapy, with no concomitant combination treatments

Below this, there is a description field and a dropdown menu for specifying the number of occurrences (with exactly 1 using distinct occurrences of). A dropdown for 'a drug era of' is set to 'Hypertension drugs'. Below that, a 'where' clause specifies 'event starts' between 0 and 7 days 'Before' the index start date. There is also an option to allow events from outside the observation period.

Figure 10.14: 追加の選択基準 3

10.7.5

これで、すべての適格基準が追加されました。次に、コホート離脱基準を指定する必要があります。なぜギャップが許容されるのでしょうか？データセットによっては、受療の一部しか観察できません。これを設定するには、イベントは”end of a continuous drug exposure (連続した薬剤曝露の終了)”で継続するを選択します。次に、持続期間を”allow for a maximum of 30 days (最大30日間)”に設定し、「ACE阻害剤」のコンセプトセットを追加します。

このコホートの場合、他に打ち切りイベントはありません。しかし、打ち切りを指定する必要がある場合は、ボタンをクリックしてください。おめでとうございます！コホートの作成は、OHDSIツールで完了しました。“出力”タブを使用して、SQLコードまたはATLASに読み込むためのJSONファイルの形式によって出力されます。

10.8 SQL

ここでは、SQLとRを使用して同じコホートを作成する方法について説明します。第9章で説明したように、SQLを分かりやすくするために、SQLをいくつかのチャunkに分割し、各チャunkが次のチャunkでどのように接続されるかを示します。

10.8.1

最初にRに対してサーバーへの接続方法を指示する必要があります。ここではDatabaseConnectionオブジェクトを作成します。

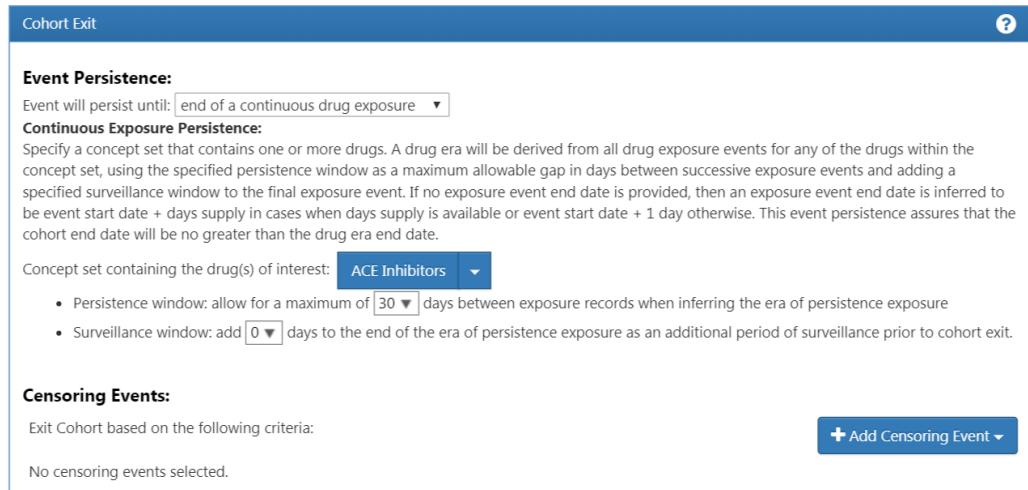


Figure 10.15: コホート離脱基準

```
library(CohortMethod)
connDetails <- createConnectionDetails(dbms = "postgresql",
                                         server = "localhost/ohdsi",
                                         user = "joe",
                                         password = "supersecret")

cdmDbSchema <- "my_cdm_data"
cohortDbSchema <- "scratch"
cohortTable <- "mycohorts"
```

最後の3行で、`cdmDbSchema`、`cohortDbSchema`、および`cohortTable`の変数を定義しています。SQL Serverの場合、データベースのスキーマはデータベースとスキーマの両方を指定する
`<- "my_cdm_data.dbo"`となります。

10.8.2

可読性を高めるために、必要なコンセプトIDをRで定義し、それらをSQLに渡します：

```
aceI <- c(1308216, 1310756, 1331235, 1334456, 1335471, 1340128, 1341927,
        1342439, 1363749, 1373225)

hypertension <- 316866

allHtDrugs <- c(904542, 907013, 932745, 942350, 956874, 970250, 974166,
                 978555, 991382, 1305447, 1307046, 1307863, 1308216,
                 1308842, 1309068, 1309799, 1310756, 1313200, 1314002,
                 1314577, 1317640, 1317967, 1318137, 1318853, 1319880,
                 1319998, 1322081, 1326012, 1327978, 1328165, 1331235,
```

```
1332418, 1334456, 1335471, 1338005, 1340128, 1341238,
1341927, 1342439, 1344965, 1345858, 1346686, 1346823,
1347384, 1350489, 1351557, 1353766, 1353776, 1363053,
1363749, 1367500, 1373225, 1373928, 1386957, 1395058,
1398937, 40226742, 40235485)
```

10.8.3

まず、各患者のACE阻害薬の初回使用を見つけます：

```
conn <- connect(connDetails)

sql <- "SELECT person_id AS subject_id,
  MIN(drug_exposure_start_date) AS cohort_start_date
INTO #first_use
FROM @cdm_db_schema.drug_exposure
INNER JOIN @cdm_db_schema.concept_ancestor
  ON descendant_concept_id = drug_concept_id
WHERE ancestor_concept_id IN (@ace_i)
GROUP BY person_id;"

renderTranslateExecuteSql(conn,
  sql,
  cdm_db_schema = cdmDbSchema,
  ace_i = aceI)
```

DRUG_EXPOSUREテーブルをCONCEPT_ANCESTORテーブルと結合することで、ACE阻害薬

10.8.4 365

次に、OBSERVATION_PERIODテーブルと結合して365日間の連続した事前の観察を要求しま

```
sql <- "SELECT subject_id,
  cohort_start_date
INTO #has_prior_obs
FROM #first_use
INNER JOIN @cdm_db_schema.observation_period
  ON subject_id = person_id
  AND observation_period_start_date <= cohort_start_date
  AND observation_period_end_date >= cohort_start_date
WHERE DATEADD(DAY, 365, observation_period_start_date) < cohort_start_date;"

renderTranslateExecuteSql(conn, sql, cdm_db_schema = cdmDbSchema)
```

10.8.5

365日以内の高血圧の診断が必要です：

```
sql <- "SELECT DISTINCT subject_id,
  cohort_start_date
INTO #has_ht
FROM #has_prior_obs
INNER JOIN @cdm_db_schema.condition_occurrence
  ON subject_id = person_id
    AND condition_start_date <= cohort_start_date
    AND condition_start_date >= DATEADD(DAY, -365, cohort_start_date)
INNER JOIN @cdm_db_schema.concept_ancestor
  ON descendant_concept_id = condition_concept_id
WHERE ancestor_concept_id = @hypertension;"

renderTranslateExecuteSql(conn,
  sql,
  cdm_db_schema = cdmDbSchema,
  hypertension = hypertension)
```

過去に複数の高血圧診断がある場合でも、重複するコホートエントリーを作成しないよう DISTINCTを使用していることに注意ください。

10.8.6

高血圧症の治療歴がないことを求めます：

```
sql <- "SELECT subject_id,
  cohort_start_date
INTO #no_prior_ht_drugs
FROM #has_ht
LEFT JOIN (
  SELECT *
  FROM @cdm_db_schema.drug_exposure
  INNER JOIN @cdm_db_schema.concept_ancestor
    ON descendant_concept_id = drug_concept_id
    WHERE ancestor_concept_id IN (@all_ht_drugs)
) ht_drugs
  ON subject_id = person_id
    AND drug_exposure_start_date < cohort_start_date
WHERE person_id IS NULL;"

renderTranslateExecuteSql(conn,
  sql,
  cdm_db_schema = cdmDbSchema,
  all_ht_drugs = allHtDrugs)
```

LEFT JOINを使用し、DRUG_EXPOSUREテーブルからのperson_idがNULLの場合のみ行を許可

10.8.7

コホート組入れの最初の7日間に高血圧症治療への曝露が一回のみである必要があります：

```
sql <- "SELECT subject_id,
  cohort_start_date
INTO #monotherapy
FROM #no_prior_ht_drugs
INNER JOIN @cdm_db_schema.drug_exposure
  ON subject_id = person_id
    AND drug_exposure_start_date >= cohort_start_date
    AND drug_exposure_start_date <= DATEADD(DAY, 7, cohort_start_date)
INNER JOIN @cdm_db_schema.concept_ancestor
  ON descendant_concept_id = drug_concept_id
WHERE ancestor_concept_id IN (@all_ht_drugs)
GROUP BY subject_id,
  cohort_start_date
HAVING COUNT(*) = 1;"
```

```
renderTranslateExecuteSql(conn,
  sql,
  cdm_db_schema = cdmDbSchema,
  all_ht_drugs = allHtDrugs)
```

10.8.8

コホートの終了日を除いて、これでコホートは完全に指定されました。コホートは曝露が停止（「magic」）と呼ばれることがよくあります）。まず、統合したいすべての曝露を含む一時テーブルを作成します：

```
sql <- "
SELECT person_id,
  CAST(1 AS INT) AS concept_id,
  drug_exposure_start_date AS exposure_start_date,
  drug_exposure_end_date AS exposure_end_date
INTO #exposure
FROM @cdm_db_schema.drug_exposure
INNER JOIN @cdm_db_schema.concept_ancestor
  ON descendant_concept_id = drug_concept_id
WHERE ancestor_concept_id IN (@ace_i);"
renderTranslateExecuteSql(conn,
  sql,
  cdm_db_schema = cdmDbSchema,
  ace_i = aceI)
```

次に、連続する曝露を統合するための標準コードを実行します：

```
sql <- "
SELECT ends.person_id AS subject_id,
       ends.concept_id AS cohort_definition_id,
       MIN(exposure_start_date) AS cohort_start_date,
       ends.era_end_date AS cohort_end_date
INTO #exposure_era
FROM (
    SELECT exposure.person_id,
           exposure.concept_id,
           exposure.exposure_start_date,
           MIN(events.end_date) AS era_end_date
    FROM #exposure exposure
    JOIN (
--cteEndDates
        SELECT person_id,
               concept_id,
               DATEADD(DAY, - 1 * @max_gap, event_date) AS end_date
        FROM (
            SELECT person_id,
                   concept_id,
                   event_date,
                   event_type,
                   MAX(start_ordinal) OVER (
                       PARTITION BY person_id ,concept_id ORDER BY event_date,
                           event_type ROWS UNBOUNDED PRECEDING
                   ) AS start_ordinal,
                   ROW_NUMBER() OVER (
                       PARTITION BY person_id, concept_id ORDER BY event_date,
                           event_type
                   ) AS overall_ord
            FROM (
-- select the start dates, assigning a row number to each
                SELECT person_id,
                       concept_id,
                       exposure_start_date AS event_date,
                       0 AS event_type,
                       ROW_NUMBER() OVER (
                           PARTITION BY person_id, concept_id ORDER BY exposure_start_date
                       ) AS start_ordinal
                FROM #exposure exposure
                    UNION ALL
-- add the end dates with NULL as the row number, padding the end dates by
-- @max_gap to allow a grace period for overlapping ranges.

                SELECT person_id,
                       concept_id,
                       DATEADD(day, @max_gap, exposure_end_date),
                       1 AS event_type,
```

```

        NULL
    FROM #exposure exposure
    ) rawdata
) events
WHERE 2 * events.start_ordinal - events.overall_ord = 0
) events
ON exposure.person_id = events.person_id
    AND exposure.concept_id = events.concept_id
    AND events.end_date >= exposure.exposure_end_date
GROUP BY exposure.person_id,
    exposure.concept_id,
    exposure.exposure_start_date
) ends
GROUP BY ends.person_id,
    concept_id,
    ends.era_end_date;"

renderTranslateExecuteSql(conn,
    sql,
    cdm_db_schema = cdmDbSchema,
    max_gap = 30)

```

このコードは、その後のすべての曝露をマージし、max_gap引数で定義された曝露間のギャップを埋めます。次に、ACE阻害薬の曝露期間を元のコホートに結合し、期間終了日をコホートの終了日として定義します。

```

sql <- "SELECT ee.subject_id,
    CAST(1 AS INT) AS cohort_definition_id,
    ee.cohort_start_date,
    ee.cohort_end_date
INTO @cohort_db_schema.@cohort_table
FROM #monotherapy mt
INNER JOIN #exposure_era ee
    ON mt.subject_id = ee.subject_id
    AND mt.cohort_start_date = ee.cohort_start_date;"

renderTranslateExecuteSql(conn,
    sql,
    cohort_db_schema = cohortDbSchema,
    cohort_table = cohortTable)

```

ここで、先に定義したスキーマとテーブルに最終的なコホートを格納します。同じテーブルに複数のコホートが格納されることがあります。

10.8.9

最後に、作成した一時テーブルをすべてクリーンアップし、データベースサーバーから切断します。

```
sql <- "TRUNCATE TABLE #first_use;  
DROP TABLE #first_use;  
  
TRUNCATE TABLE #has_prior_obs;  
DROP TABLE #has_prior_obs;  
  
TRUNCATE TABLE #has_ht;  
DROP TABLE #has_ht;  
  
TRUNCATE TABLE #no_prior_ht_drugs;  
DROP TABLE #no_prior_ht_drugs;  
  
TRUNCATE TABLE #monotherapy;  
DROP TABLE #monotherapy;  
  
TRUNCATE TABLE #exposure;  
DROP TABLE #exposure;  
  
TRUNCATE TABLE #exposure_era;  
DROP TABLE #exposure_era;"  
  
renderTranslateExecuteSql(conn, sql)  
  
disconnect(conn)
```

10.9



- コホートとは、一定期間に1つ以上の適格基準を満たす人の集合体を指します。
- コホート定義とは、特定のコホートを識別するために使用されるロジックの説明です。
- コホートは、対象とする曝露やアウトカムを定義するために、OHDSI分析ツールを使用します。
- コホートを構築するには、2つの主要なアプローチがあり、ルールベースと確率的アプローチがあります。
- ルールベースのコホート定義は、ATLASまたはSQLを使用して作成できます。

10.10

前提条件

最初の演習には、ATLASインスタンスへのアクセスが必要です。以下のインスタンス <http://atlas-demo.ohdsi.org> またはアクセス可能な他のインスタンスを使用できます。

演習 10.1. 以下の条件に従ってATLASでコホート定義を作成してください。：

- ・ジクロフェナクの新規ユーザー
- ・16歳以上
- ・曝露前に少なくとも365日の継続的な観察期間があること
- ・以前に（非ステロイド性抗炎症薬（NSAID）への曝露がないこと
- ・以前に癌の診断がないこと
- ・コホートからの離脱は、曝露の中止（30日間のギャップを許容）と定義すること

前提条件

2番目の演習では、R、R-Studio、Javaがインストールされていることを前提とします。セクション8.4.5で説明されている。また、SqlRender、DatabaseConnector、Eunomiaパッケージが必要です。これらは、以下の方法でインストールできます：

```
install.packages(c("SqlRender", "DatabaseConnector", "remotes"))
remotes::install_github("ohdsi/Eunomia", ref = "v1.0.0")
```

Eunomiaパッケージは、ローカルのRセッション内で実行されるCDM内のシミュレートされた

```
connectionDetails <- Eunomia::getEunomiaConnectionDetails()
```

CDMデータベーススキーマは「main」です。

演習 10.2. 以下の基準に従って、SQLおよびRを使用して、既存のCOHORTテーブルに急性心

- ・心筋梗塞の診断の発生（コンセプト4329847「心筋梗塞」およびそのすべての下位層に
- ・入院または救急外来受診期間（コンセプト9201、9203、262；それぞれ「入院ビジット」

提案された解答は、付録 E.6 を参照ください。

Chapter 11

著者: Anthony Sena & Daniel Prieto-Alhambra

観察医療データベースは、さまざまな特性に基づく集団の差異を理解するための貴重なり

- ・データベースの特性評価：データベース全体のデータプロファイルを全体的に理解
- ・コホート特性評価：集団をその累積的な医療履歴に基づいて記述します。
- ・治療経路：特定の期間に受けた一連の介入を説明します。
- ・発生率：リスク期間における集団のアウトカムの発生率を測定する。

データベースレベルの特性評価を除き、これらのことばは「インデックス日」と呼ばれる10章で説明されているようにコホートとして定義されます。コホートは対象集団内の各

特性評価のユースケースには、疾患の自然経過、治療の利用状況、品質向上などが含まれます。

11.1

関心集団についての特性評価の問い合わせに答える前に、使用するデータベースの特性をまず理解します。データベースの定量的評価には、通常、以下のような質問が含まれます。：

- ・このデータベースには全体で何人が含まれていますか？
- ・年齢分布は？
- ・このデータベースで観察されている期間は？
- ・時間の経過とともに{治療、コンディション、処置など}が記録・処方された人の割合

これらのデータベースレベルの記述統計は、研究者がデータベースに欠けている可能性の15章では、データ品質についてさらに詳しく説明します。

11.2

コホート特性評価は、コホート内の人々のベースラインとポストインデックスの特徴を記述します。

コホートの特性評価の方法は、特定の治療を受けている患者の適応症や禁忌の有病率を推定する方法です。the Reporting of Observation Studies in Epidemiology (STROBE) ガイドラインで詳述されています (von Elm et al., 2008)。

11.3

集団の特性を評価するもう一つの方法は、インデックス後の期間における治療シーケンスを記述する方法です。et al. (2016) は、OHDSIの共通データ標準を利用して、2型糖尿病、高血圧症、抑うつ症に対する治療経路分析を行っています。

経路分析は、特定の疾患を診断された人が最初の薬剤処方/供給を受けた治療（イベント）を起点として、その後の治療経路を示す分析です。

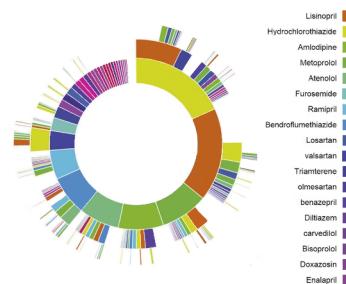


Figure 11.1: 高血圧症のOHDSI治療経路「サンバースト」グラフ

例として、図 11.1 は高血圧症治療を開始する患者集団を表しています。中央にある最初の円は、治療経路分析の起点です。

経路分析は、集団における治療利用に関する重要なエビデンスを提供します。この分析から、et al. (2016) はメトホルミンが糖尿病治療に対して最も一般的に処方されている薬剤であることを示しています。

従来のDUS（薬剤使用実態研究）用語では、治療経路分析は、指定された集団における一つまたは複数の治療経路を示す分析です。

11.4

発生率および発生割合は、時間の経過とともに集団における新たなアウトカムの発生を評価する指標です。図 11.2 は、単一の人に対する発生率の計算要素を示す目的としています：

図 11.2 では、人がデータで観察される期間が観察開始と終了時間によって示されています。次に、発生率を計算するための2つの尺度があります：

$$= \frac{\#}{\#}$$

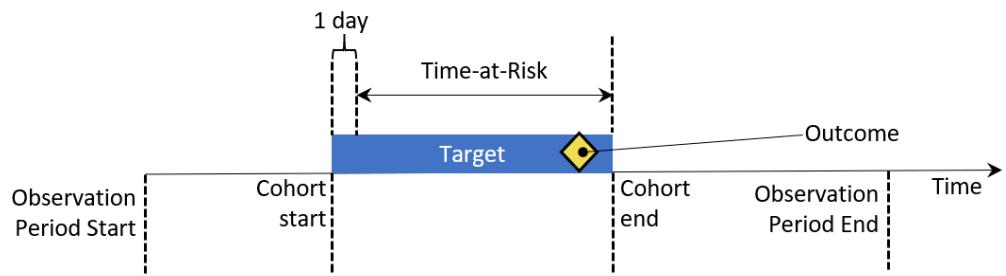


Figure 11.2: 発生率計算要素の人単位のビュー。この例では、リスク時間はコホート開始

発生割合は、リスク期間中に集団内で発生した新規のアウトカムの割合を提供します。別

$$= \frac{\#}{\text{_____}}$$

発生率は、集団の累積的なリスク期間内に新規のアウトカムの数を測定する指標です。リ

治療に対して計算される場合、発生割合および発生率は、特定の治療の使用における集団

11.5

世界保健機関（WHO）の高血圧症に関するグローバル概要（Who, 2013）によると、高血圧症の早期発見、適切な治療、良好な管理には、健康と経済上の両面で大

観察研究のデータソースは、WHOが行ったように高血圧症患者集団の特性を評価する方

11.6 ATLAS

ここでは、ACHILLES で作成されたデータベースの特性評価統計を調査するために、ATLAS の Data Sources をクリックして開始します。ATLAS に表示される最初のドロップダウン 「Occurrence」 を選択し、データベースに存在するすべての症状のツリーマップを表示し

特定の関心のあるコンディションを検索するには、テーブルタブをクリックして、データベース内の「hypertension (高血圧)」を含む項目に基づいてリストをフィルタリングできます：

特定のコンディションの詳細なドリルダウンレポートを表示するには、行をクリックして、「hypertension (本態性高血圧)」を選択し、選択されたコンディションの経時的および性別

高血圧症のコンセプトの有無と経時的な傾向についてデータベースの特性を確認した後、「Drug Era (薬剤曝露期間)」 レポートを使用します。データベースの特性を探して関心

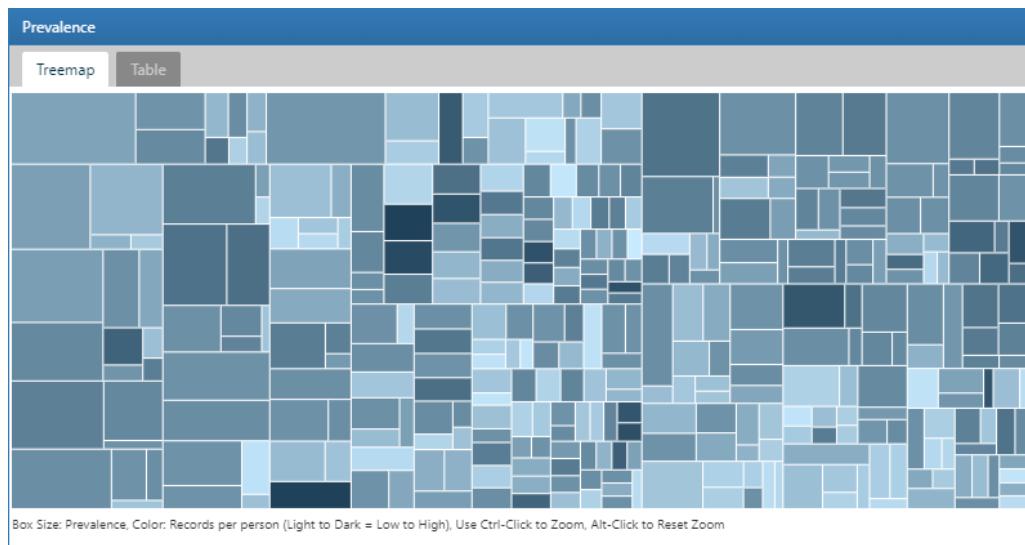


Figure 11.3: ATLASデータソース: コンディション出現のツリーマップ

| Prevalence | | | | |
|---|---|-----------------------|------------|--------------------|
| Treemap | | Table | | |
| | | Column visibility | Copy | CSV |
| Show 15 ▾ entries | | | | |
| Filter: hypertension | | | | |
| Showing 1 to 15 of 47 entries (filtered from 15,907 total entries) | | Previous 1 2 3 4 Next | | |
| Concept | Name | Person Count | Prevalence | Records per person |
| 320128 | Essential hypertension | 17,814,076 | 12.30% | 5.80 |
| 312648 | Benign essential hypertension | 11,014,877 | 7.61% | 4.35 |
| 317898 | Malignant essential hypertension | 1,021,441 | 0.70% | 2.22 |
| 381290 | Ocular hypertension | 521,264 | 0.36% | 2.40 |
| 441922 | Transient hypertension of pregnancy | 209,317 | 0.14% | 2.45 |
| 44782429 | Chronic kidney disease due to hypertension | 170,534 | 0.12% | 3.60 |
| 137940 | Transient hypertension of pregnancy - delivered | 153,806 | 0.11% | 1.07 |
| 321080 | Hypertension complicating pregnancy, childbirth and the puerperium | 148,728 | 0.10% | 2.15 |
| 314423 | Benign essential hypertension complicating pregnancy, childbirth and the puerperium - not delivered | 132,245 | 0.09% | 3.94 |
| 44782690 | Chronic kidney disease stage 5 due to hypertension | 119,375 | 0.08% | 5.20 |
| 44783618 | Heritable pulmonary arterial hypertension | 104,737 | 0.07% | 3.61 |
| 319826 | Secondary hypertension | 96,356 | 0.07% | 2.14 |
| 4167493 | Pregnancy-induced hypertension | 91,675 | 0.06% | 2.60 |
| 321074 | Pre-existing hypertension complicating pregnancy, childbirth and puerperium | 74,311 | 0.05% | 2.99 |
| 192680 | Portal hypertension | 71,240 | 0.05% | 3.11 |
| Showing 1 to 15 of 47 entries (filtered from 15,907 total entries) | | Previous 1 2 3 4 Next | | |

Figure 11.4: ATLASデータソース: コンセプト名に ” hypertension (高血圧)” が含まれるコンディション

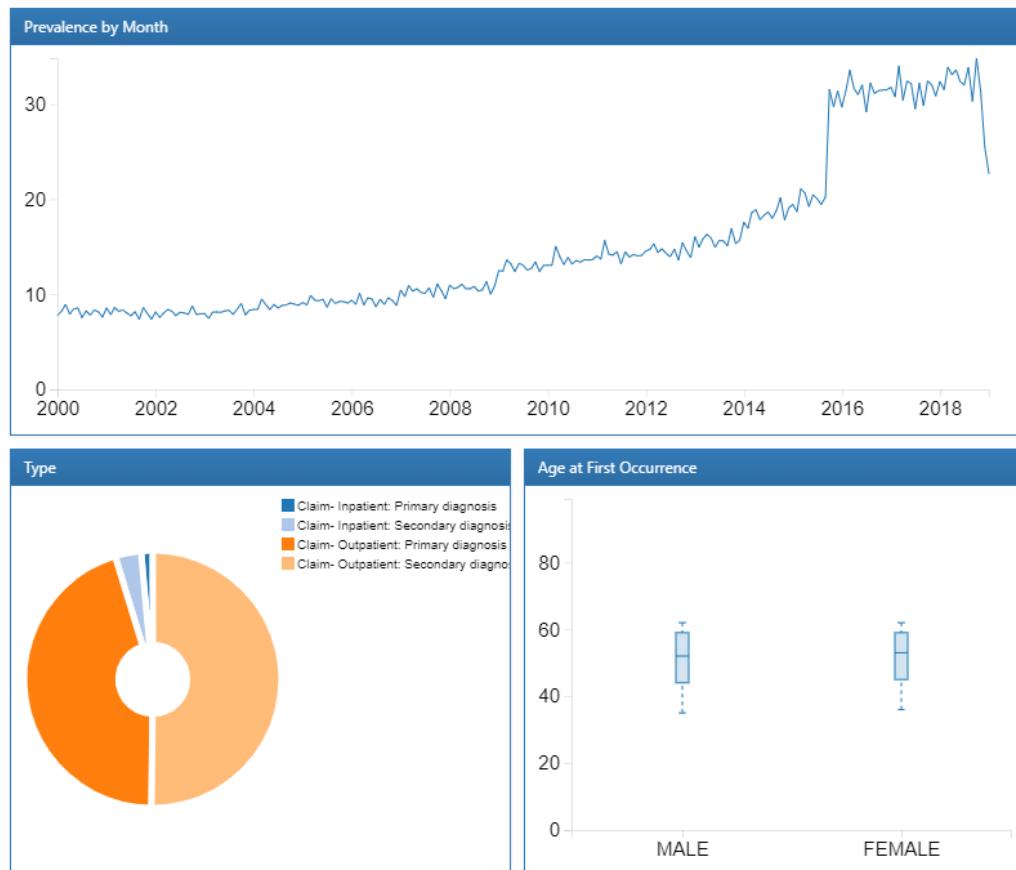


Figure 11.5: ATLASデータソース: 本態性高血圧ドリルダウンレポート

11.7 ATLAS

ここでは、ATLASを使用して複数のコホートの大規模な特性評価を行う方法を示します。左側

11.7.1

特性評価には、少なくとも1つのコホートと少なくとも1つの特性が必要です。この例では、2(B.6)。2つ目のコホートは、最初のコホートと同様ですが、1年間の代わりに少なくとも3年間(B.7)。

コホート定義

The screenshot shows the 'Cohort characterization' section of the ATLAS interface. At the top, there are three tabs: 'Design' (selected), 'Executions', and 'Utilities'. Below the tabs, a detailed description of cohort characterization is provided:

Cohort characterization is defined as the process of generating cohort level descriptive summary statistics from person level covariate data. Summary statistics of these person level covariates may be count, mean, sd, var, min, max, median, range, and quantiles. In addition, covariates during a period may be stratified into temporal units of time for time-series analysis such as fixed intervals of time relative to cohort_start_date (e.g. every 7 days, every 30 days etc.), or in absolute calendar intervals such as calendar-week, calendar-month, calendar-quarter, calendar-year.

Below this, the 'Cohort definitions' section is shown. It includes an 'Import' button and a table of cohorts:

| ID | Name | Edit cohort | Remove |
|-------|--|-------------|--------|
| 10447 | Patients initiating first-line therapy for hypertension with >1 yr follow-up | Edit cohort | Remove |
| 10448 | Patients initiating first-line therapy for hypertension with >3 yr follow-up | Edit cohort | Remove |

At the bottom, it says 'Showing 1 to 2 of 2 entries' and has 'Previous' and 'Next' buttons.

Figure 11.6: 特性設計タブ - コホート定義の選択

コホートは既にATLASで作成されていると仮定しています（第10章を参照）。Import をクリックし、Figure 11.6 に示すようにコホートを選択します。次に、これらのコホートを特性評価するために使用

特徴量の選択

ATLASにはOMOP CDMでモデル化された臨床ドメイン全体で特性評価を行うため、約100のRパッケージを利用しています。次のセクションでは、FeatureExtractionとRの使用について説明

Import をクリックして、特性を選択します。以下は、これらのコホートを特性評価するために使

上の図は、選択された機能のリストと、各機能が各コホートについて何を特徴付けるかを説明。“Demographics(人口動態的特性)”で始まる特性は、コホート開始日における各人の人口統計

- Any time prior (任意の期間) : コホート開始前のすべての利用可能な期間で、その人の毎

Feature analyses

Import

Show 25 ▾ entries Search:

| ID | Name | Description | Actions |
|-----|-------------------------------------|---|------------------------|
| 43 | Drug Era Short Term | One covariate per drug in the drug_era table overlapping with any part of the short window. | Remove |
| 49 | Charlson Index | The Charlson comorbidity index (Romano adaptation) using all conditions prior to the window end. | Remove |
| 67 | Condition Occurrence Long Term | One covariate per condition in the condition_occurrence table starting in the long term window. | Remove |
| 71 | Demographics Age Group | Age of the subject on the index date (in 5 year age groups) | Remove |
| 72 | Demographics Race | Race of the subject. | Remove |
| 73 | Demographics Prior Observation Time | Number of continuous days of observation time preceding the index date. | Remove |
| 74 | Demographics Gender | Gender of the subject. | Remove |
| 76 | Condition Occurrence Medium Term | One covariate per condition in the condition_occurrence table starting in the medium term window. | Remove |
| 77 | Demographics Age | Age of the subject on the index date (in years). | Remove |
| 79 | Demographics Time In Cohort | Number of days of observation time during cohort period. | Remove |
| 80 | Demographics Index Year | Year of the index date. | Remove |
| 81 | Demographics Post Observation Time | Number of continuous days of observation time following the index date. | Remove |
| 87 | Procedure Occurrence Any Time Prior | One covariate per procedure in the procedure_occurrence table any time prior to index. | Remove |
| 103 | Visit Count Long Term | The number of visits observed in the long term window. | Remove |

Figure 11.7: 特性設計タブ - 特性選択

- Long term (長期) : コホート開始日を含む最大365日前まで。
- Medium term (中期) : コホート開始日を含む最大180日前まで。
- Short term (短期) : コホート開始日を含む最大30日前まで。

サブグループ分析

性別に基づいて異なる特性を作成したい場合、「サブグループ分析」セクションを用いて、新サブグループを作成するには、サブグループのメンバーシップの条件をクリックして追加しま

Subgroup analyses

New subgroup

Female

Calculate subgroup analyses only

having all of the following criteria:

+ Add criteria to group... □

with the following event criteria:

+ Add attribute... □

✖ with a gender of: FEMALE Add Import Delete Criteria

Figure 11.8: 特性評価の設計 - 女性サブグループ分析



ATLASのサブグループ分析は階層とは異なります。階層は相互に排他的ですが、サブグローブ

11.7.2

特性評価のデザインが完了したら、環境内の1つ以上のデータベースに対してこのデザインを実行します。

Design Executions Utilities

Executions

| | | | |
|-----------|------------|----------------------|--------------------|
| SYNPUF 1K | ▶ Generate | 🕒 View latest result | All executions (3) |
| SYNPUF 5% | ▶ Generate | 🕒 View latest result | All executions (3) |

Figure 11.9: 特性評価設計の実行 - CDMソース選択

分析が完了したら、「All Executions (すべてを実行)」ボタンをクリックしてレポートを表示します。

“View Reports (レポートを見る)” を選択します。あるいは、“View latest result(最新の結果を見る)” をクリックして、最後に実行されたアウトカムを表示すること

11.7.3

| CONDITION / Condition Occurrence Long Term / stratified by Female | | | | | | | | | | | | |
|---|-----------|------------|--|--|-------|--------|-------|--|-------|--------|------------|---------|
| | | | | Patients initiating first-line therapy for hypertension with >1 yr follow-up | | | | Patients initiating first-line therapy for hypertension with >3 yr follow-up | | | Std diff ▼ | |
| Covariate | Explore | Concept ID | | Count | Pct | Female | | Count | Pct | Female | | |
| | | | | | | Count | Pct | | | Count | Pct | |
| Tachycardia | Explore ▾ | 444070 | | 17,322 | 1.04% | 9,042 | 1.18% | 6,547 | 0.78% | 3,530 | 0.90% | -0.0193 |
| Cardiomegaly | Explore ▾ | 314658 | | 20,958 | 1.26% | 8,007 | 1.04% | 9,016 | 1.08% | 3,465 | 0.89% | -0.0121 |
| Cardiac arrhythmia | Explore ▾ | 44784217 | | 30,474 | 1.83% | 13,221 | 1.72% | 14,540 | 1.74% | 6,318 | 1.62% | -0.0052 |

Showing 1 to 3 of 3 entries (filtered from 206 total entries)

Previous 1 Next

Figure 11.10: 特性アウトカム - 過去1年間の疾患発生

結果は、デザインで選択した各コホートについて、さまざまな特徴を一覧表示します。図 11.10 では、コホート開始日の前の365日間に存在するすべての条件の概要が提供されています。

検索ボックスを使用してアウトカムをフィルタリングし、「不整脈」の既往を持つ人の書類（図 11.11 参照）。

コホートのすべての条件コンセプトを特性評価したため、“explore(探索する)” オプションを使用して、選択されたコンセプト（この場合は不整脈）のすべての上位層と

この特性結果を用いて、高血圧症治療に禁忌のある条件（例：血管性浮腫）を見つけることを目的とした検索（例：“edema(浮腫)” を検索します（図 11.12 を参照）。

再度、“explore(探索する)” 機能を使用して、高血圧症集団における浮腫の特性を調べます。

ここでは、降圧薬を開始する前の1年間に血管性浮腫の既往歴がこの集団の一部にあることを示す

ドメイン共変量は、コホート開始前の時間枠にコードの記録が存在したかどうかを示す二値

11.7.4

プリセットの機能に加えて、ATLASはユーザー定義のカスタム機能を用いることもできます。Analysisタブをクリックして、New Feature Analysisボタンをクリックします。カスタム

この例では、ACE阻害剤の服用歴が各コホート開始後にある、コホート内的人数を特定す

上で定義した基準は、コホート開始日に適用されることを前提としています。基準を定義Analysisセクションに移動します。Importボタンをクリックし、メニューから新しいカス

| Exploring condition_occurrence during day -365 through 0 days relative to index: Cardiac arrhythmia | | | | | | |
|---|-----|-----------------------------------|--------------|-------------|--------|--------|
| Cohort: Patients initiating first-line therapy for hypertension with >1 yr follow-up | | | | | | |
| Relationship type | | Distance | Concept name | All stratas | | Female |
| Count | Pct | Count | Pct | | | |
| Explore Ancestor | 4 | Disorder by body site | 32 | 0.00% | 17 | 0.00% |
| Explore Ancestor | 4 | Finding of trunk structure | 991 | 0.06% | 605 | 0.08% |
| Explore Ancestor | 3 | Disorder of trunk | 23 | 0.00% | 14 | 0.00% |
| Explore Ancestor | 3 | Disorder of thorax | 241 | 0.01% | 104 | 0.01% |
| Explore Ancestor | 3 | Disorder of body system | 4,135 | 0.25% | 1,992 | 0.26% |
| Explore Ancestor | 2 | Disorder of cardiovascular system | 12,979 | 0.78% | 6,073 | 0.79% |
| Explore Ancestor | 2 | Disorder of mediastinum | 138 | 0.01% | 62 | 0.01% |
| Explore Ancestor | 2 | Disorder of body cavity | 24 | 0.00% | 10 | 0.00% |
| Explore Ancestor | 1 | Heart disease | 4,691 | 0.28% | 1,869 | 0.24% |
| Explore Selected | 0 | Cardiac arrhythmia | 30,474 | 1.83% | 13,221 | 1.72% |

Showing 1 to 10 of 62 entries Previous [1] 2 3 4 5 6 7 Next

Figure 11.11: 特性アウトカム - 単一コンセプトの探索

CONDITION / Condition Occurrence Long Term / stratified by Female

| Covariate | | Explore | Concept ID | Patients initiating first-line therapy for hypertension with >1 yr follow-up | | | | Patients initiating first-line therapy for hypertension with >3 yr follow-up | | | | Std diff | |
|-----------|-----------|------------|------------|--|--------|-------|--------|--|--------|-------|---------|----------|--|
| Covariate | Explore | Concept ID | | | Female | | | | Female | | | | |
| | | | Count | Pct | Count | Pct | Count | Pct | Count | Pct | Count | | |
| Edema | Explore ▾ | 433595 | 32,243 | 1.94% | 20,200 | 2.63% | 15,173 | 1.81% | 9,684 | 2.48% | -0.0066 | | |

Showing 1 to 1 of 1 entries (filtered from 206 total entries) Previous [1] Next

Figure 11.12: 特性評価の結果 - 禁忌条件の探索

Exploring condition_occurrence during day -365 through 0 days relative to index: Edema

Cohort: Patients initiating first-line therapy for hypertension with >1 yr follow-up

Export Show 10 entries Search:

| Relationship type | Distance | Concept name | All stratas | | Female | |
|--------------------|----------|--------------|-------------|-------|--------|-------|
| | | | Count | Pct | Count | Pct |
| Explore Descendant | -2 | Angioedema | 2,605 | 0.16% | 1,506 | 0.20% |

Showing 1 to 1 of 1 entries (filtered from 56 total entries) Previous Next

Figure 11.13: 特性アウトカム - 禁忌条件の詳細を探索

DEMOGRAPHICS / Demographics Age

Export Export comparison Show 10 entries Search:

| Strata | Patients initiating first-line therapy for hypertension with >1 yr follow-up | | | | Patients initiating first-line therapy for hypertension with >3 yr follow-up | | | | Std diff |
|-------------|--|-------|---------|--------|--|-------|---------|--------|----------|
| | Count | Avg | Std Dev | Median | Count | Avg | Std Dev | Median | |
| Female | 768,180 | 49.39 | 9.78 | 51.00 | 390,693 | 49.01 | 9.03 | 51.00 | -0.0291 |
| All stratas | 1,661,604 | 48.96 | 10.00 | 50.00 | 837,459 | 48.64 | 9.26 | 50.00 | -0.0232 |

Showing 1 to 2 of 2 entries Previous Next

Figure 11.14: 各コホートとサブグループの年齢特性アウトカム

Design

Criteria Custom

Analysis type:

Prevalence

Add Criteria feature

Ace inhibitor exposure after index Delete

having of the following criteria: + Add criteria to group...

with using occurrences of:
a drug era of + Add attribute...

where between days and days [add additional constraint](#)

allow events from outside observation period Delete Criteria

Figure 11.15: ATLASでのカスタム特性定義

| DRUG / Ace inhibitor exposure after index / stratified by Female | | | | | | | | | | | | |
|--|-----------|------------|--|--------|---------|--------|--|--------|---------|--------|----------|--|
| | | | Patients initiating first-line therapy for hypertension with >1 yr follow-up | | | | Patients initiating first-line therapy for hypertension with >3 yr follow-up | | | | Std diff | |
| Covariate | Explore | Concept ID | Count | Pct | Female | | Count | Pct | Female | | | |
| | | | | | Count | Pct | | | Count | Pct | | |
| Ace inhibitor exposure after index | Explore ▾ | 0 | 686,034 | 41.29% | 289,215 | 17.41% | 426,280 | 50.90% | 182,219 | 21.76% | 0.1001 | |

Showing 1 to 1 of 1 entries Previous [1] Next

Figure 11.16: カスタム機能の結果表示

11.8 R

Rを使用してコホートをの特性を評価することもできます。このセクションでは、OHDSI RパッケージであるFeatureExtractionを使用して、高血圧症コホートのベースライン特性（共

- ・デフォルトの共変量セットを選択する
- ・事前に指定された分析セットから選択する
- ・カスタム分析セットを作成する

FeatureExtractionは、個人レベルの特徴と集約された特徴の2つの異なる方法で共変量を作成

11.8.1

最初に、特性を評価するためにコホートをインスタンス化する必要があります。コホートのインスタンス化は、OHDSI RパッケージのcreateCohort()関数で実現されています。この例では、高血圧症に対して一次治療を開始し、1年間のフォローアップを行った患者（付録B）を対象としています（付録B.6）。付録Bの他のコホートの特性評価は、読者への練習問題として残しておきます。ここでは、高血圧症コホートのインスタンス化を行います。

11.8.2

まず、Rにサーバーへの接続方法を指示する必要があります。FeatureExtractionはDatabaseConnectionオブジェクトを用いてデータベースへの接続情報を定義します。

```
library(FeatureExtraction)
connDetails <- createConnectionDetails(dbms = "postgresql",
                                         server = "localhost/ohdsi",
                                         user = "joe",
                                         password = "supersecret")

cdmDbSchema <- "my_cdm_data"
cohortsDbSchema <- "scratch"
```

```
cohortsDbTable <- "my_cohorts"
cdmVersion <- "5"
```

最後の4行は、`cdmDbSchema`、`cohortsDbSchema`、`cohortsDbTable`変数、およびCDMバージョンを定義します。SQL Serverの場合、データベーススキーマはデータベースとスキーマの両方を指定する必要があります。`<- "my_cdm_data.dbo"`となります。

11.8.3

`createCovariateSettings`関数は、ユーザーが定義済みの多くの共変量から選択できるようにします。

```
settings <- createCovariateSettings(
  useDemographicsGender = TRUE,
  useDemographicsAgeGroup = TRUE,
  useConditionOccurrenceAnyTimePrior = TRUE)
```

これにより、性別、年齢（5歳との年齢グループ）、およびコホート開始日までの期間（期間）などの事前に指定された分析は、短期、中期、長期の時間枠を参照しています。デフォルトでは、

- 長期: コホート開始日を含む365日前まで
- 中期: コホート開始日を含む180日前まで
- 短期: コホート開始日を含む30日前まで

ただし、ユーザーはこれらの値を変更できます。例を以下に示します：

```
settings <- createCovariateSettings(useConditionEraLongTerm = TRUE,
                                       useConditionEraShortTerm = TRUE,
                                       useDrugEraLongTerm = TRUE,
                                       useDrugEraShortTerm = TRUE,
                                       longTermStartDays = -180,
                                       shortTermStartDays = -14,
                                       endDays = -1)
```

これは、長期ウィンドウをコホート開始日の180日前から当日まで（当日を含まず）と再構築します。また、共変量を構築する際に使用すべき、または使用すべきでないコンセプトIDを指定することができます。

```
settings <- createCovariateSettings(useConditionEraLongTerm = TRUE,
                                       useConditionEraShortTerm = TRUE,
                                       useDrugEraLongTerm = TRUE,
                                       useDrugEraShortTerm = TRUE,
                                       longTermStartDays = -180,
                                       shortTermStartDays = -14,
                                       endDays = -1,
                                       excludedCovariateConceptIds = 1124300,
```

```
addDescendantsToExclude = TRUE,
aggregated = TRUE)
```



上記すべての例について、「aggregated = TRUE」の使用は、FeatureExtractionに要約統計を生成する。

11.8.4

次のコードブロックは、コホートの集計統計を生成します：

```
covariateSettings <- createDefaultCovariateSettings()

covariateData2 <-getDbCovariateData(
  connectionDetails = connectionDetails,
  cdmDatabaseSchema = cdmDatabaseSchema,
  cohortDatabaseSchema = resultsDatabaseSchema,
  cohortTable = "cohorts_of_interest",
  cohortId = 1,
  covariateSettings = covariateSettings,
  aggregated = TRUE)

summary(covariateData2)
```

出力は次のようにになります：

```
## CovariateData Object Summary
##
## Number of Covariates: 41330
## Number of Non-Zero Covariate Values: 41330
```

11.8.5

集計されたcovariateDataオブジェクトの主なコンポーネントは、二値および連続の共変量に

```
covariateData2$covariates
covariateData2$covariatesContinuous
```

11.8.6

FeatureExtractionは、カスタム共変量を定義および利用する機能も提供します。これらの詳細は[//ohdsi.github.io/FeatureExtraction/](https://ohdsi.github.io/FeatureExtraction/)

11.9 ATLAS

経路分析の目標は、1つまたは複数の対象とするコホート内で治療がどのように順序づけられるかをHripcsak et al. (2016)によって報告されたデザインに基づいています。これらの方法はPathwaysという機能に組み込まれました。

コホート経路の目的は、1つまたは複数の対象とするコホートのコホート開始日以降のイベントを示すことです。



Figure 11.17: 単一の人物におけるパスウェイ分析の文脈

図 11.17 では、その人物が開始日と終了日が定義された対象コホートに属していることを示しています。まず、ATLASの左側のバーで Cohort Pathways をクリックして、新しいコホートパスウェイを作成します。

11.9.1

まず、高血圧症の第一選択療法を開始するコホートと、1年および3年間のフォローアップ（B.6、B.7）を継続して使用します。ボタンを使用して、2つのコホートをインポートします。

| ID | Name | Edit cohort | Remove |
|-------|---|-----------------------------|------------------------|
| 10447 | Patients initiating first-line therapy for hypertension with >1 yr follow-up | Edit cohort | Remove |
| 10448 | Patients initiating first-line therapy for hypertension with >3 yr follow-up | Edit cohort | Remove |

Figure 11.18: 対象コホートを選択したパスウェイ分析

次に、対象となる各第一選択の降圧薬のイベントコホートを作成して、イベントコホート（B.8 – B.16）に記載されていることを確認ください。完了したら、Import ボタンをクリックして、これらの定義を経路デザインのイベントコホートセクションにインポートします。

Event Cohorts

Each Event Cohort defines the step in a pathway that may occur for a person in the Target Cohort.

| Import | | | |
|-------------------|---|------------------------------|------------------------|
| Show 10 ▾ entries | | Search: <input type="text"/> | |
| ID | Name | | |
| 9174 | ACE inhibitor use | Edit cohort | Remove |
| 9175 | Angiotensin receptor blocker (ARB) use | Edit cohort | Remove |
| 9176 | Thiazide or thiazide-like diuretic use | Edit cohort | Remove |
| 9177 | dihydropyridine Calcium Channel Blocker (dCCB) use | Edit cohort | Remove |
| 9178 | non-dihydropyridine Calcium Channel Blocker (ndCCB) use | Edit cohort | Remove |
| 9179 | beta blocker use | Edit cohort | Remove |
| 9180 | Diuretic-loop use | Edit cohort | Remove |
| 9181 | Diuretic-potassium sparing use | Edit cohort | Remove |
| 9182 | alpha-1 blocker use | Edit cohort | Remove |

Showing 1 to 9 of 9 entries

Previous [1](#) Next

Figure 11.19: 初回第一選択降圧治療を開始するためのイベントコホート

完了すると、デザインは上記のようになります。次に、いくつかの追加の分析設定を示します。

- 組み合わせウィンドウ: この設定では、イベント間の重複がイベントの組み合わせと「+ イベントコホート2」として組み合われます。
- 最小セル数: この人数に満たないイベントコホートは、プライバシー保護のため、出力されません。
- 最大経路長: 分析の対象となる一連のイベントの最大数を指します。

11.9.2

パスウェイ分析のデザインが完了すると、環境内の1つ以上のデータベースに対してこの

11.9.3

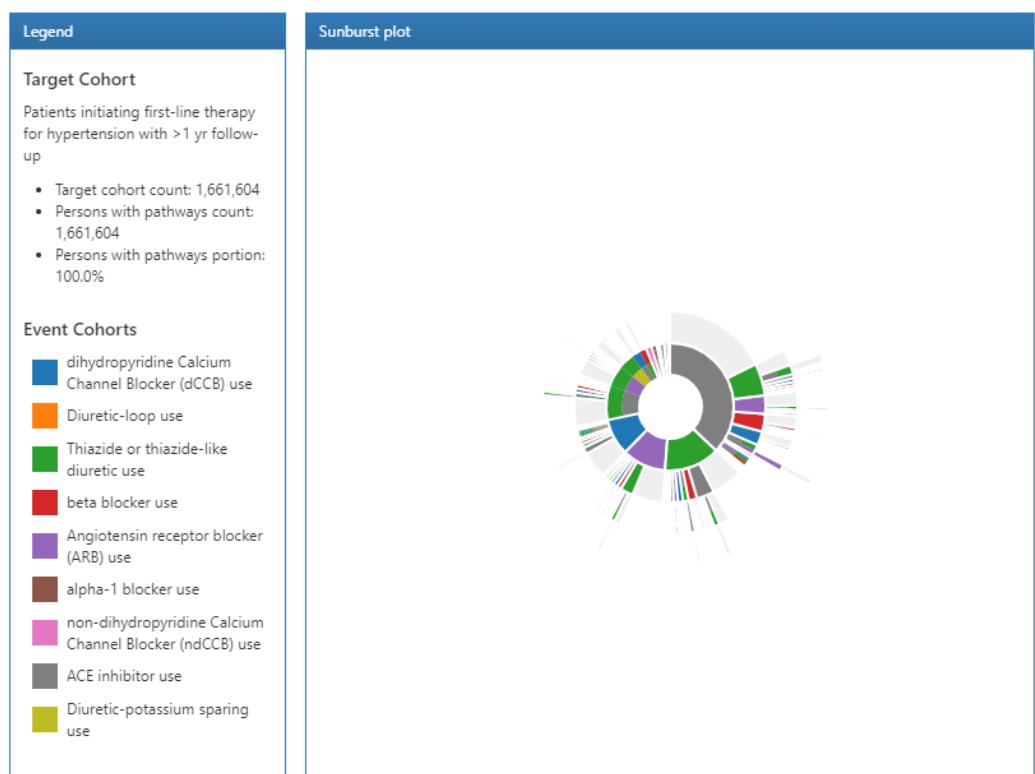


Figure 11.20: 経路結果の凡例とサンバースト図

パスウェイ分析の結果は3つのセクションに分かれています：凡例セクションでは、対象

サンバースト図は、時間の経過に伴うさまざまなイベント経路を視覚的に表現したもので（ACE阻害薬、アンジオテンシン受容体拮抗薬など）。2番目のリングセットは、人々にと

サンバーストプロットのセクションをクリックすると、右側に経路の詳細が表示されます

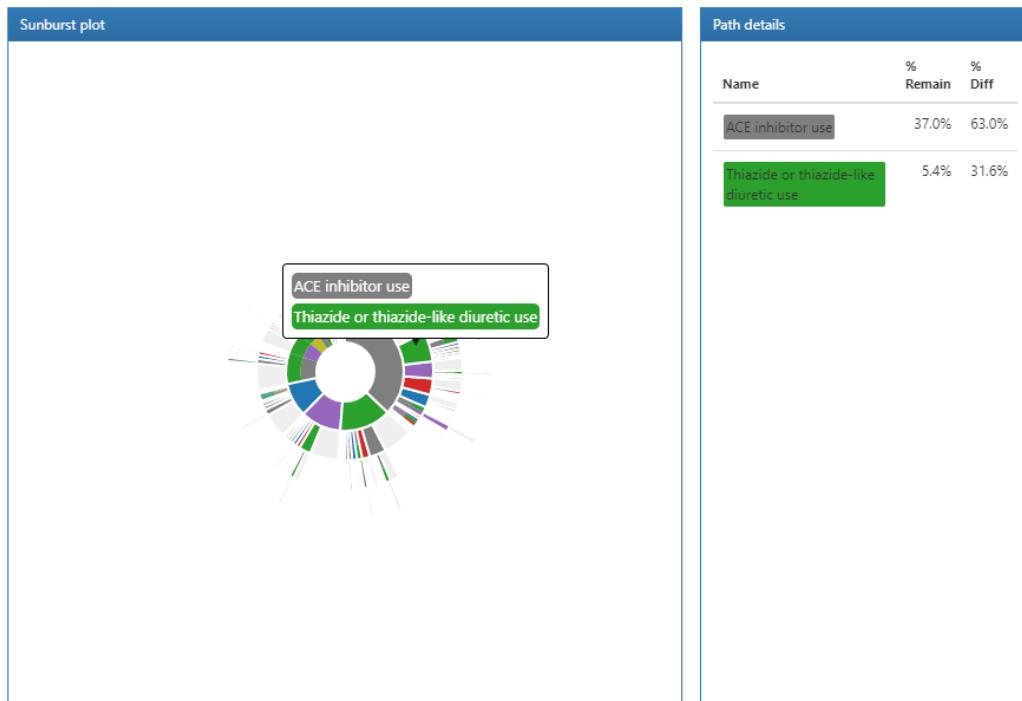


Figure 11.21: 経路の詳細を表示するパスウェイアウトカム

11.10 ATLAS

発生率の算出では、以下の内容を記述します
ク期間中に、対象コホートに属する人の中で、ア
まず、ATLAS の左側のバーにある **Incidence Rates** をクリックし、新規の発生率分析を作成
Define をクリックします。

11.10.1

本例で使用されるコホートは、既に ATLAS に作成されていると仮定します(第
10章で説明)。付録には、対象コホート(付録 B.2、B.5)およびアウトカムコホート
(付録 B.4、B.3、B.9)の完全な定義が記載されています。

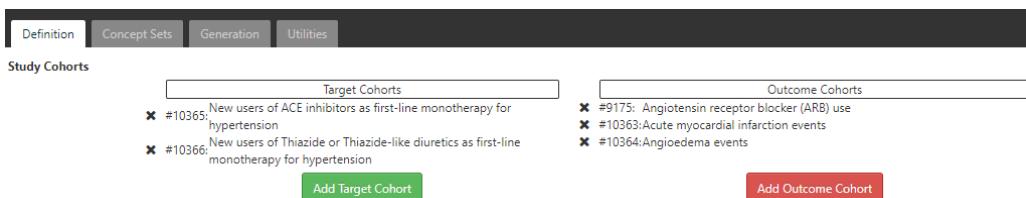


Figure 11.22: 対象およびアウトカム定義の発生率

定義タブで、 New users of ACE inhibitors (ACE阻害薬の新規ユーザー)

コホートと New users of Thiazide or Thiazide-like diuretics (サイアザイドまたはサイア+acute myocardial infarction events (急性心筋梗塞イベント)、 angioedema events、 および Angiotensin receptor blocker (ARB) use (アンジオテンシン受容体拮抗薬 (ARB) の新規ユーザー) のアウトカムコホートを選択します。再びウィンドウを閉じて、

Time At Risk

Time at risk defines the time window relative to the cohort start or end date with an offset to consider the person 'at risk' of the outcome.

- Time at risk starts with plus days.
- Time at risk ends with plus days.

No study window defined.

Figure 11.23: 対象およびアウトカム定義の発生率

次に、分析のリスク期間を定義します。上に示すように、リスク期間はコホートの開始日およびTHZ コホートの定義には、薬剤曝露が終了する時点をコホート終了日としています。

ATLAS では、分析仕様の一部として対象コホートを層別化する方法も提供しています:

Stratify Criteria: You can provide optional stratification criteria to the analysis that will divide the population into unique groups based on their satisfied criteria.

New stratify criteria

1. Gender = Female

Gender = Female

Copy Delete

enter an inclusion rule description

having all of the following criteria:

+ Add criteria to group... ▾

with the following event criteria:

+ Add attribute... ▾

✖ with a gender of: FEMALE Add Import

Delete Criteria

Figure 11.24: 女性における発生率の層別定義

これを行うには、[New Stratify Criteria] ボタンをクリックし、第 11 章で説明されている

11.10.2

[生成] タブをクリックし、 ボタンをクリックして、分析を実行するデータベース一つ以上のデータベースを選択し、「Generation」ボタンをクリックして、指定された

11.10.3

「Generation」タブでは、画面の上部でターゲットおよびアウトカムを選択してアウトカムのドロップダウンリストから ACEi 使用者のターゲットコホートと急性心筋梗塞 (AMI) を選択します。 ボタンをクリックして発生率分析の結果を表示します:

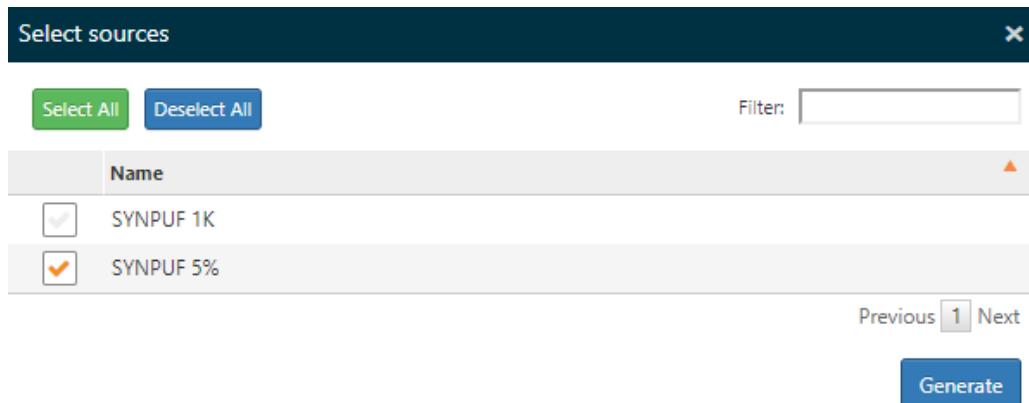


Figure 11.25: 発生率分析実行

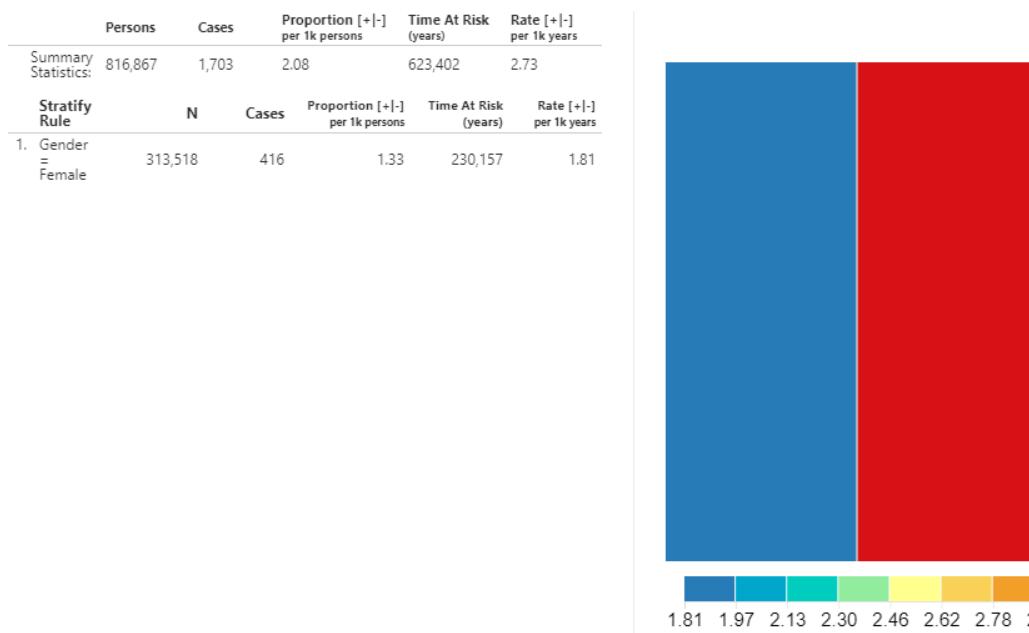


Figure 11.26: 発生率分析の出力 - AMI のアウトカムを持つ新規 ACEi 使用者

データベースの要約は、TAR 期間中に観察されたコホート内の総人数と総症例数を示します。1000人当たりの症例数を示しています。対象コホートのリスク期間は年単位で計算され、1000人年当たりの症例数として表されます。

設計で定義した層の発生率メトリクスも見ることができます。上記のメトリクスは各層に

ACEi 集団の中で ARB 新規使用の発生率を確認するために、同じ情報を収集することができます。ARB 使用に変更し、 ボタンをクリックして詳細を確認します。

| | Persons | Cases | Proportion [+/-] per 1k persons | Time At Risk (years) | Rate [+/-] per 1k years |
|------------------------|---------|--------|------------------------------------|-------------------------|----------------------------|
| Summary Statistics: | 818,520 | 43,868 | 53.59 | 621,094 | 70.63 |
| Stratify Rule | N | Cases | Proportion [+/-] per 1k persons | Time At Risk (years) | Rate [+/-] per 1k years |
| 1. Gender = Female | 313,832 | 21,049 | 67.07 | 228,369 | 92.17 |

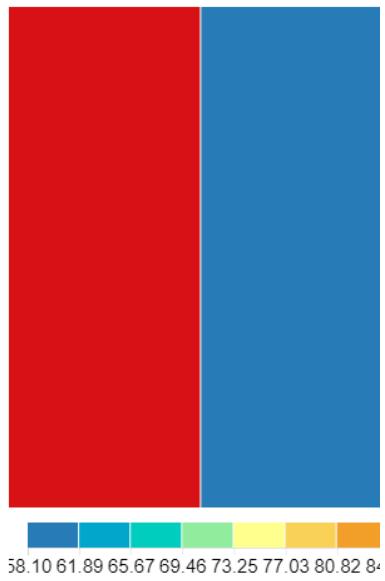


Figure 11.27: 発生率 - ACEi 曝露中に ARB 処理を受けている新規 ACEi 使用者

示されているように、算出されたメトリクスは同じですが、解釈は異なります。なぜなら、入力 (ARB 使用) が健康アウトカムではなく薬剤使用量の推定値を参照しているためです。

11.11



- OHDSI は、データベース全体または対象とするコホートの特性を評価するため
- コホートの特徴付けは、インデックス日（ベースライン）前およびインデックス日後
- ATLAS の特徴付けモジュールと OHDSI Methods Library は、複数の時間枠の基
- ATLAS の経路および発生率モジュールは、ポストインデックス期間中の記述統計

11.12

前提条件

これらの演習には、ATLAS インスタンスへのアクセスが必要です。<http://atlas-demo.ohdsi.org> のインスタンスや、アクセス可能な他のインスタンスを使用できます。

演習 11.1. セレコキシブが実世界でどのように使用されているかを理解したいと思います。まず、このデータベースがこの薬についてどのようなデータを持っているかを理解したいと思 ATLASデータソースモジュールを使用して、セレコキシブに関する情報を検索します。

演習 11.2. セレコキシブの使用者の疾患の自然経過について、より深く理解したいと思います。365日間のウォッシュアウト期間を使用して、セレコキシブの新規使用者の単純なコホートを(第 10 章を参照してください)、ATLAS を使用して、併存疾患と薬剤曝露を示すこのコホート

演習 11.3. セレコキシブ処方開始後に消化管出血(GI 出血)がどのくらいの頻度で発生するのか (“消化管出血”) またはその下位層に含まれるいずれかのコンセプトの発生として単純に定義 GI 出血イベントのコホートを作成します。前の演習で定義した曝露コホートを使用して、セレコ GI 出血イベントの発生率を計算してください。

推奨される解答は付録E.7 を参照ください。

Chapter 12

著者: Martijn Schuemie, David Madigan, Marc Suchard & Patrick Ryan

保険請求データや電子健康記録などの観察的な医療データは、治療の効果に関する現実世

- 直接効果推定: アウトカムのリスクに対する曝露の効果を、曝露なしと比較して推定
- 比較効果推定: アウトカムのリスクに対する曝露（ターゲット曝露）の効果を、別の

いずれの場合でも、患者レベルの因果効果は事実のアウトカム、すなわち曝露を受けた患

集団レベルの効果推定のユースケースには、治療選択、安全性監視、および比較効果が含

本章ではまず、OHDSI Methods LibraryとしてRパッケージで実装されているさまざまな

12.1

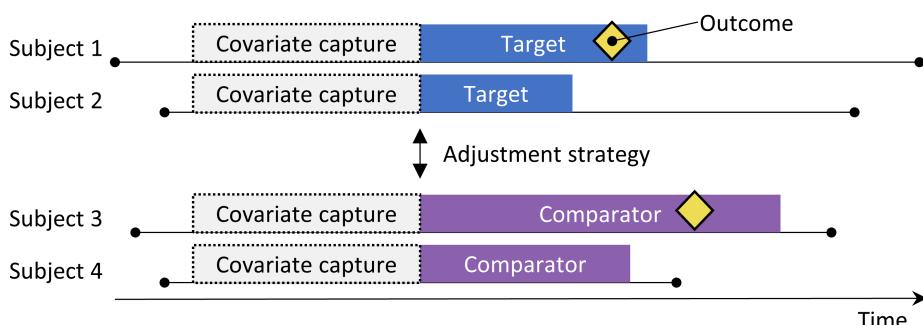


Figure 12.1: 新規ユーザーコホートデザイン。ターゲット治療を開始した対象は比較対照

コホートメソッドはランダム化臨床試験を模倣することを試みます (Hernan and Robins, 2016)。ある治療（ターゲット）を開始した対象は別の治療（比較対照）を開始します。12.1にハイライトされた5つの選択を行うことで具体化されます。

Table 12.1: 比較コホートデザインの主要な設計選択

| 選択 | 説明 |
|-----------|----------------------------|
| ターゲットコホート | ターゲット治療を代表するコホート |
| 比較対照コホート | 比較対照治療を代表するコホート |
| アウトカムコホート | 関心のあるアウトカムを代表するコホート |
| リスク期間 | どの時点で（通常はターゲットおよび比較対照コホート） |
| モデル | ターゲットと比較対照の間の違いを調整しながら効果 |

モデルの選択には、他の要素の中でも、アウトカムモデルの種類が含まれます。例えば、ロジスティック回帰モデルは、各治療群のリスクを直接比較する方法です。



新規ユーザーコホートメソッドは本質的に比較効果推定の方法であり、治療を比較対照する方法を用いています。

重要な懸念は、ターゲット治療を受ける患者が比較対照治療を受ける患者と系統的に異なる可能性があります。

12.1.1

ランダム化試験では、（仮想的な）コイントスが患者を各グループに割り当てます。したがって、各治療群の構成は、ランダム化によって決定されます。

ある患者に対する傾向スコア (PS) は、その患者がターゲット治療を受ける確率です (Rosenbaum and Rubin, 1983)。バランスの取れた二群ランダム化試験では、傾向スコアはすべての患者で等しくなります (Rassen et al., 2012)。

例えば、一対一のPSマッチングを用いるとします。Janのターゲット治療を受ける事前確率が0.6で、Junのターゲット治療を受ける事前確率が0.4です。この場合、JanとJunの因果コントラストの推定をもたらします。次に、推定のための手順は次のようになります。

傾向スコアは測定された交絡因子を制御します。実際、測定された特性がない場合、採用されません。しかし、この問題は「強い無視可能性」は実際にはテストできない前提です。この問題についての詳細は第18章で説明します。

12.1.2

以前は、PSは手動で選択された特性に基づいて計算されていましたが、OHDSIツールはそのように機能しません (Tian et al., 2018)。これらの特性には、人口統計情報、治療開始前および当日に観察されたコホート (Suchard et al., 2013) を使用して適合させ、Cyclopsパッケージで実装します。本質的には、



典型的には、治療開始日の特性は治療の原因となる診断などの多くの関連データが

一部の人々は、因果構造を特定するために臨床専門知識に依存しない共変量選択のデータ (Hernan et al., 2002)。しかし、これらの懸念は現実的なシナリオでは大きな影響を与える (Schneeweiss, 2018)。さらに、医学においては真の因果構造が判明することはほとんど

12.1.3

傾向スコアは0から1の連続体上にあるため、厳密なマッチングはほとんど不可能です。Austin (2011) に従い、ロジットスケールで標準偏差の0.2を使用します。

12.1.4

傾向スコア方法は一致する患者が存在することを必要とします。このため、主要な診断は

$$\ln \left(\frac{F}{1-F} \right) = \ln \left(\frac{S}{1-S} \right) - \ln \left(\frac{P}{1-P} \right)$$

ここで F は選好スコア、 S は傾向スコア、 P はターゲット治療を受ける患者の割合です。

Walker et al. (2013) は「経験的均衡」のコンセプトを述べています。彼らは、少なくとも

12.1.5

良い実践は常にPS調整がバランスの取れた患者群を生成するかどうかをチェックします。12.19はバランスをチェックするための標準的なOHDSI出力を示しています。各患者特性 (Rubin, 2001)。

12.2

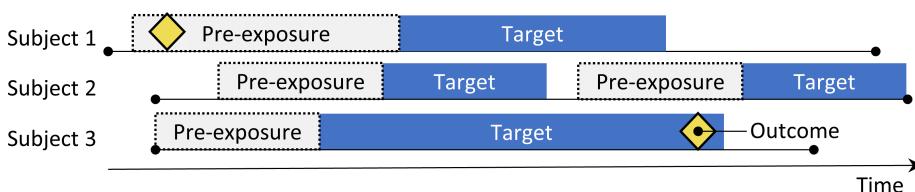


Figure 12.2: 自己対照コホートデザイン。ターゲットへの曝露中のアウトカムの発生率を

自己対照コホート (SCC) デザイン (Ryan et al., 2013a) は曝露中のアウトカムの発生率を 12.2 に示す4つの選択肢が、自己対照コホートの質問を定義します。

Table 12.2: 自己対照コホートデザインの主要な設計選択肢。

| 選択 | 説明 |
|-----------|---------------------------------|
| ターゲットコホート | 治療を表すコホート |
| アウトカムコホート | 関心のあるアウトカムを表すコホート |
| リスク時間 | アウトカムのリスクをどのタイミング（通常ターゲットコホートの開 |
| 対照時間 | 対照時間として使用される期間 |

Table 12.3: 症例対照デザインの主要な設計選択肢

| 選択 | 説明 |
|------------|--|
| アウトカムコホート | 症例（興味のあるアウトカム）を表すコホート |
| 対照コホート | 対照を表すコホート。通常、選択ロジックを使用してアウトカムコホートと対照コホートを割り当てる |
| ターゲットコホート | 治療を表すコホート |
| ネスティングコホート | 任意で症例および対照が抽出されるサブポピュレーションを定義する |
| リスク時間 | 曝露ステータスをどのタイミング（通常インデックス日が基準）で測定するか |

曝露群を構成する同じ被験者が対照群としても使用されるため、被験者間の差異について調整する必要がある。

12.3

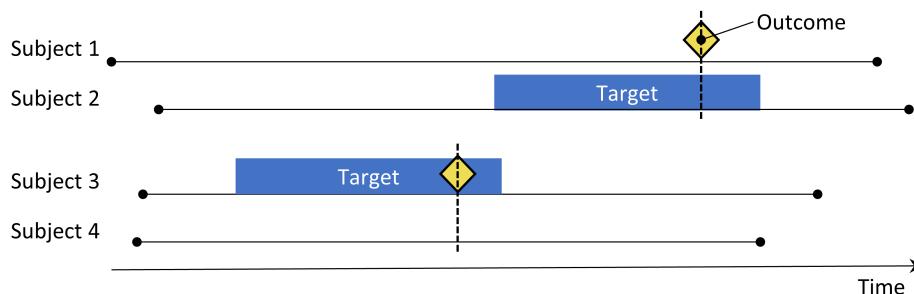


Figure 12.3: 症例対照デザイン。アウトカムを持つ被験者（「症例」）は、アウトカムを持つ被験者（「対照」）と比較される。

症例対照研究 (Vandenbroucke and Pearce, 2012) は、「特定の疾患のアウトカムを持つ人が、他の疾患のアウトカムを持つ人よりも、特定の暴露に対するリスクが高い」という問題を定義します。

通常、症例を年齢や性別などの特性で一致させて対照を選択し、症例と対照を比較しやすくします。

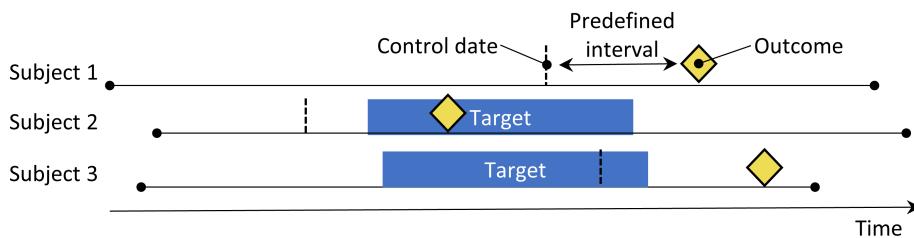


Figure 12.4: ケース・クロスオーバーデザイン。アウトカムの周りの時間を、アウトカム

Table 12.4: ケース・クロスオーバーデザインの主要な設計選択肢

| 選択 | 説明 |
|-----------|-------------------------------|
| アウトカムコホート | 症例（興味のあるアウトカム）を表すコホート |
| ターゲットコホート | 治療を表すコホート |
| リスク時間 | 曝露ステータスをどのタイミング（通常インデックス日が基準） |
| 対照時間 | 対照時間として使用される期間 |

12.4

ケース・クロスオーバー (Maclure, 1991) デザインは、アウトカムのタイミングでの曝露 12.4 は、ケース・クロスオーバーの質問を定義する選択肢を示します。

症例は自分自身の対照として機能します。自己対照デザインとして、これらは人間間の差 時間-コントロールデザイン (Suiswa, 1995) が開発され、例えば年齢や性別で一致させた

12.5

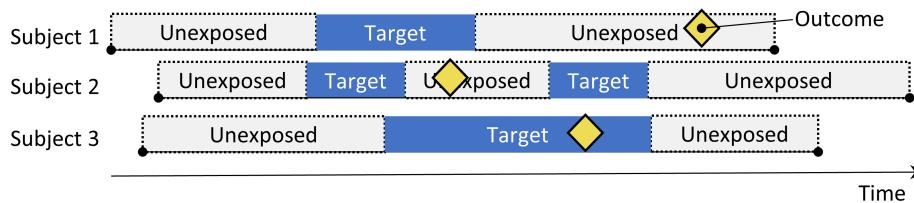


Figure 12.5: 自己対照症例シリーズデザイン。曝露中のアウトカム発生率と非曝露中のア

自己対照症例シリーズ (SCCS) デザイン (Farrington, 1995; Whitaker et al., 2006) は、曝露中のアウトカム発生率を、すべての非曝露期間中の発生率、これには曝露 12.5 の選択肢が SCCS の質問を定義します。

他の自己対照デザイン同様、SCCS は個人間の違いによる交絡に対して頑健ですが、時間 (Simpson et al., 2013)。これにより、モデルに数千の追加変数が追加されます。この場合

Table 12.5: 自己対照症例シリーズデザインの主な設計選択肢

| 選択 | 説明 |
|-----------|---------------------------------|
| ターゲットコホート | 治療を代表するコホート |
| アウトカムコホート | 関心のあるアウトカムを代表するコホート |
| リスク時間 | どの時点（多くの場合、ターゲットコホートの開始日または終了日） |
| モデル | 時間変動する交絡因子の調整を含む効果の推定モデル |

SCCSの重要な前提条件の一つは、観察期間の終了がアウトカムの日付と独立していることである (Farrington et al., 2011)。

12.6

12.6.1

ACE阻害薬（ACEi）は、高血圧や虚血性心疾患を持つ患者、特にうっ血性心不全、糖尿病、慢性腎臓病などに有効である (Zaman et al., 2002)。アンジオエデマは、唇、舌、口、喉頭、咽頭、または眼窩周囲の腫れである (Sabroe and Black, 1997)。しかし、これらの薬剤使用に関連するアンジオエデマの絶対および相対リスクはまだ不明である (Powers et al., 2012)。いくつかの観察研究は、ACEiをβ遮断薬と比較してアンジオエデマのリスクを評価している (Magid et al., 2010; Toh et al., 2012) が、β遮断薬はもはや高血圧の一線級治療として推奨されない (Whelton et al., 2018)。有力な代替治療法として、チアジド類およびチアジド様利尿薬（THZ）がある。

以下では、観察医療データに我々の集団レベル推定フレームワークを適用して、次の比較推定を行います。

ACE阻害薬の新規使用者とチアジドおよびチアジド様利尿薬の新規使用者を比較した場合

ACE阻害薬の新規使用者とチアジドおよびチアジド様利尿薬の新規使用者を比較した場合

これらは比較効果推定の質問であるため、セクション @ref (CohortMethod) で述べたコホート方法を適用します。

12.6.2

高血圧の治療を初めて観察された患者をACEiまたはTHZクラスのどちらかの単剤療法として利用する

12.6.3

アンジオエデマは、入院または救急部（ER）訪問中の血管浮腫のコンディションコンセプトの間

12.6.4

リスク期間を治療開始の翌日から開始し、曝露が終了するまでと定義し、後続の薬剤曝露の間

12.6.5

デフォルトの共変量セットを使用してPSモデルを適合させます。このセットには、人口統計学的特徴

12.6.6

Table 12.6: 私たちの比較コホート研究の主な設計選択肢

| 選択肢 | 値 |
|-----------|--|
| ターゲットコホート | 高血圧の第一選択単剤療法としてのACE阻害薬 |
| 比較コホート | 高血圧の第一選択単剤療法としてのチアジドまたはアンジオエデマまたは急性心筋梗塞。 |
| アウトカムコホート | 治療開始の翌日から開始し、曝露が終了するまで |
| リスク期間 | 可変比マッチングを用いたコックス比例ハザード |
| モデル | モデル |

12.6.7

我々の研究デザインが真実と一致する推定を生成するかどうかを評価するために、真の効果を18章で説明しています。

12.7 ATLAS

ここでは、ATLASの推定機能を使用してこの研究を実施する方法を示します。ATLASの左側のツールバーにある  **Estimation** をクリックし、新しい推定研究を作成します。研究に簡単に認識できる名前を入力して、 ボタンをクリックして保存できます。

推定設計機能には、比較、分析設定、評価設定の3つのセクションがあります。複数の比較

12.7.1

研究には1つ以上の比較を含めることができます。「比較を追加」をクリックすると、新しい比較  をクリックしてターゲットおよび比較コホートを選択します。「アウトカムを追加」(10章を参照)。ターゲット(付録B.2)、比較(付録B.5)、およびアウトカム(付録@ref(Angioedema) および付録B.3)コホートの完全な定義は付録にあります。完了する12.6のようになります。

ターゲットと比較コホートのペアに対して複数のアウトカムを選択できることに注意してください。

ネガティブコントロールアウトカム

ネガティブコントロールアウトカムは、ターゲットまたは比較対照によって引き起こされる効果を評価するための对照です。18章で説明されているとおりすでに作成されていると仮定し、それを選択するだけです。12.7は、この研究に使用されたネガティブコントロールコンセプトセットを示しています。

Comparison
Add or update the target, comparator, outcome(s) cohorts and negative control outcomes

Choose your target cohort:
New users of ACE inhibitors as first-line monotherapy for hypertension

Choose your comparator cohort:
New users of Thiazide-like diuretics as first-line monotherapy for hypertension

Choose your outcome cohorts:

Add Outcome

Show 10 entries Search:

| ID | Name | Edit cohort | Remove |
|---------|-------------------------------------|-----------------------------|------------------------|
| 1770712 | Angioedema outcome | Edit cohort | Remove |
| 1770713 | Acute myocardial infarction outcome | Edit cohort | Remove |

Showing 1 to 2 of 2 entries Previous 1 Next

Figure 12.6: 比較ダイアログ

Negative controls for ACEi and THZ Optimize

| Concept Set Expression | Included Concepts (75) | Included Source Codes | Explore Evidence | Export | Compare |
|------------------------------|--|-----------------------|------------------|--------|---------|
| Search: <input type="text"/> | Previous 1 2 3 Next | | | | |

Show 25 entries

Showing 1 to 25 of 75 entries

| Concept Id | Concept Code | Concept Name | Domain | Standard Concept Caption | Exclude | Descendants | Mapped |
|------------|--------------|--------------------------------|-----------|--------------------------|-------------------------------------|-------------------------------------|-------------------------------------|
| 72748 | 74779009 | Strain of rotator cuff capsule | Condition | Standard | <input checked="" type="checkbox"/> | <input type="checkbox"/> | <input checked="" type="checkbox"/> |
| 73241 | 197210001 | Anal and rectal polyp | Condition | Standard | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |
| 73560 | 55260003 | Calcaneal spur | Condition | Standard | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |
| 75911 | 65358001 | Acquired hallux valgus | Condition | Standard | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |
| 76786 | 63643000 | Derangement of knee | Condition | Standard | <input checked="" type="checkbox"/> | <input type="checkbox"/> | <input checked="" type="checkbox"/> |

Figure 12.7: ネガティブコントロールコンセプトセット

含めるコンセプト

コンセプトを選択するときに、生成したい共変量を指定できます。たとえば、傾向スコア

除外するコンセプト

含めるコンセプトを指定する代わりに、除外するコンセプトを指定することもできます。12.8は、これらのコンセプトを含むコンセプトセットを示しています（その下位層も含みます）。

The screenshot shows the 'Concept Set #1798551' interface. At the top, there is a header bar with buttons for 'Concepts to exclude for ACEi and THZ' (with a count of 38225), 'Optimize', and 'Compare'. Below this is a table titled 'Concept Set Expression' with columns: 'Included Concepts' (38225), 'Included Source Codes', 'Explore Evidence', 'Export', and 'Compare'. The table has a search bar at the top right and a dropdown for 'Show 25 entries'. It displays 14 entries from a total of 14, with columns for 'Concept Id', 'Concept Code', 'Concept Name', 'Domain', 'Standard Concept Caption', 'Exclude' (checkbox), 'Descendants' (checkbox), and 'Mapped' (checkbox). The data includes:

| Concept Id | Concept Code | Concept Name | Domain | Standard Concept Caption | Exclude | Descendants | Mapped |
|------------|--------------|--------------|--------|--------------------------|--------------------------|-------------------------------------|--------------------------|
| 1342439 | 38454 | trandolapril | Drug | Standard | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/> |
| 1334456 | 35296 | Ramipril | Drug | Standard | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/> |
| 1331235 | 35208 | quinapril | Drug | Standard | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/> |
| 1373225 | 54552 | Perindopril | Drug | Standard | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/> |
| 1310756 | 30131 | moexipril | Drug | Standard | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/> |

Figure 12.8: 除外するコンセプトを定義するコンセプトセット

ネガティブコントロールと除外する共変量を選択した後、比較ダイアログの下半分は図12.9のようになります。

The screenshot shows the 'Comparison' dialog. At the top, it says 'Negative control concept set:' followed by a list box labeled 'Negative controls for ACEi and THZ' with a 'File' and 'X' button. Below this is a section titled 'Covariate selection' with a note: 'Concepts to include when constructing the covariates to be used in this study. (Leave blank if you want to include every concept).*' A text input field is shown with a 'File' and 'X' button. A note below it states: '* Concepts defined here are combined with those defined in the Analysis settings section.' Below this is another section titled 'Concepts to exclude when constructing the covariates to be used in this study.*' A list box labeled 'Concepts to exclude for ACEi and THZ' is shown with a 'File' and 'X' button. A note below it states: '* Concepts defined here are combined with those defined in the Analysis settings section.'

Figure 12.9: ネガティブコントロールおよび除外するコンセプトセットを示す比較ウィンドウ

12.7.2

比較ダイアログを閉じた後、「分析設定を追加」をクリックできます。「分析名」とラベル付

研究集団

分析に入る被験者のセットである研究集団を指定するさまざまなオプションがあります。多くの研究開始および終了日を使用して、特定の期間に分析を制限できます。研究終了日はリスク警告のため)であり、特定の方法で実践された時間にのみ興味がある場合、研究開始および終

オプション “Should only the first exposure per subject be included? (各対象の初回の曝露のみが含まれるべきか)” を使用すると、患者ごとの最初の曝露に限定する minimum required continuous observation time prior to index date for a person to be included in the cohort (その個人をコホートに含めるための要件として、インデックスというオプションは、コホート定義すでに設定されていることが多いので、ここでは0のま

“Remove subjects that are in both target and comparator cohort? (ターゲットコホートと比較群コホートの両方に含まれる対象を除外しますか)” は、“If a subject is in multiple cohorts, should be censored time-at-risk when the new time-at-risk starts to prevent overlap? (対象が複数のコホートに含まれる場合、リスク期

- “Keep All(すべてを保持します)” は、両方のコホートの対象を保持することを示す。
- “Keep First(最初のコホートに保持します)” は、最初に発生したコホートに対象を残す
- “Remove All(すべてから除外します)” は、すべてのコホートから対象を除外することを示す。

もし “Keep All” または “Keep First” のオプションが選択された場合、ある対象が両方のコホートに含まれる場合、その対象は両方のコホートに含まれます。デフォルトでは、リスク時間はコホートの開始日と終了日を基準に定義され

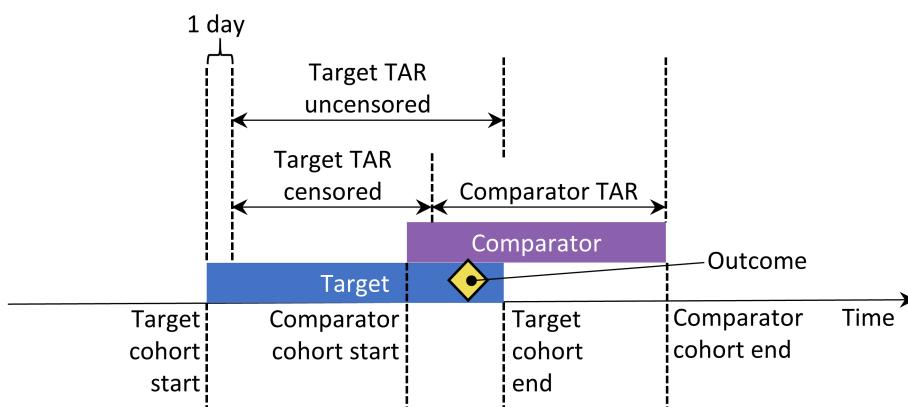


Figure 12.10: リスク時間 (Time-at-risk (TAR)) が薬剤曝露開始日から薬剤曝露終了時までと

アウトカムの2回目の出現は1回目の継続であることが多いため、リスクウインドウが始まる前私たちの研究の事例での選択は、図 12.11 に示されています。対象コホートと比較対象コホー

 Study Population

Study start date - a calendar date specifying the minimum date that a cohort index can appear (leave blank to use all time):

Study end date - a calendar date specifying the maximum date that a cohort index can appear (leave blank to use all time). **Important:** the study end date is also used to truncate risk windows, meaning no outcomes beyond the study end date will be considered.

Restrict the study to the period when both exposures are present in the data? (E.g. when both drugs are on the market)

Should only the first exposure per subject be included?

The minimum required continuous observation time (in days) prior to index date for a person to be included in the cohort.

Remove subjects that are in both the target and comparator cohort?

If a subject is in multiple cohorts, should time-at-risk be censored when the new time-at-risk start to prevent overlap?

Remove subjects that have the outcome prior to the risk window start?

How many days should we look back when identifying prior outcomes?

If either the target or the comparator cohort is larger than this number it will be sampled to this size. (0 for this value indicates no maximum size)

Figure 12.11: 研究対象集団の設定

共変量の設定

ここでは共変量の構成を指定します。これらの共変量は通常傾向スコアモデルで使用されます。to view details (詳細を見るにはここをクリック)をクリックすると、どの共変量の組み合わせをinclude (組入れ)および/またはexclude (除外)するコンセプトを指定することで、共変量のセクション 12.7.1 にある比較のための設定と同じです。これらの設定が2つの場所にある理由は、これら

図 12.12 は、この研究で選択した内容を示しています。図 12.9 に定義するように、比較の設定

The screenshot shows the 'Covariate Settings' interface. It has two main sections:

- Concepts to include:** A text input field with a blue 'Save' button and a red 'Delete' button. Below it is a note: "* Concepts defined here are combined with those defined in the Comparisons section."
- Concepts to exclude:** A text input field with a blue 'Save' button and a red 'Delete' button. Below it is a note: "* Concepts defined here are combined with those defined in the Comparisons section."

Figure 12.12: 共変量の設定

リスク時間

リスク時間 (Time-at-risk) は、対象コホートや比較コホートにおける開始日と終了日を基準に

リスク時間の終了日は、コホートの終了日、つまり曝露が停止した時点としました。例えば、to-treat デザインと呼ばれることもあります。

リスク時間がゼロの患者は何の情報も提供しないので、最小リスク日数は通常1日に設定され



コホート研究を計画する際の鉄則は、バイアスが含まれる可能性を除外するため、コホー

The screenshot shows the 'Time At Risk' settings interface. It includes the following fields:

- Start Date: "1 days from cohort start date"
- End Date: "0 days from cohort end date"
- Minimum Days at Risk: "1"

Figure 12.13: リスク時間設定

傾向スコアによる調整

傾向スコア値が極端な人を除外して、研究対象集団を切り取って整えることができます。

トリミングに加えて、またはトリミングの代わりに、傾向スコアで層別化またはマッチング

- ・傾向スコア尺度：傾向スコアそのもの
- ・標準化尺度：傾向スコア分布の標準偏差による
- ・標準化ロジット尺度：傾向スコアをより正規分布にするためのロジット変換後の傾

疑問がある場合は、デフォルト値を使用するか、Austin (2011) によるこのトピックに関

大規模な傾向モデルの適合には計算コストがかかることがあるので、モデルの適合に使用

Test each covariate for correlation with the target assignment? (各共変量とターゲットの共変量が異常に高い相関（正または負）を持つ場合、これはエラーを出力します。これに

Use regularization when fitting the model? (モデルをフィットするときに正則化をします。標準的な手順は、傾向モデルに多くの共変量（通常10,000以上）を含めることです。これに

図 12.14 は、この研究での選択を示しています。最大マッチング人数を100人に設定する

アウトカムモデルの設定

最初に、対象コホートと比較コホート間のアウトカムの相対リスクを推定するために使用 12.1 で簡単に述べたように、Cox、Poisson、ロジスティック回帰から選択できます。こ

また、共変量をアウトカムモデルに追加して分析を調整することもできます。これは傾向

傾向スコアで層別化またはマッチングする代わりに、逆確率重み付け (IPTW) を使うこと

結果モデルにすべての共変量を含めることを選択した場合、共変量が多くれば、モデルを

図 12.15 は、この研究での私たちの選択を示しています。可変比率マッチングを用いてい

12.7.3

第 18 章にあるように、ネガティブコントロールとポジティブコントロールを検討し、操

ネガティブコントロールアウトカムコホートの定義

セクション 12.7.1 では、ネガティブコントロールアウトカムを表すコンセプトセットを “*ingrown nail of foot* (足の陷入爪)” の出現も、上位層の “*ingrown nail* (陷入爪)” の出現として数えることができます。3番目の選択肢は、コンセプトを探すと

Propensity Score Adjustment

How do you want to trim your cohorts based on the propensity score distribution?

None ▼

Do you want to perform matching or stratification?

Match on propensity score ▼

What is the maximum number of persons in the comparator arm to be matched to each person in the target arm within the defined caliper? (0 = means no maximum - all comparators will be assigned to a target person):

100 ▼

What is the caliper for matching:

0.2

What is the caliper scale:

Standardized Logit ▼

What is the maximum number of people to include in the propensity score model when fitting? Setting this number to 0 means no down-sampling will be applied:

250000 ▼

Test each covariate for correlation with the target assignment? If any covariate has an unusually high correlation (either positive or negative), this will throw an error.

Yes ▼

Use regularization when fitting the propensity model?

Yes ▼

Control Settings ▼ Prior ▼

Figure 12.14: 傾向スコアによる調整の設定

Outcome Model Settings

Specify the statistical model used to estimate the risk of outcome between target and comparator cohorts:

Cox proportional hazards ▼

Should the regression be conditioned on the strata defined in the population object (e.g. by matching or stratifying on propensity scores)?

Yes ▼

Should the covariates also be included in the outcome model?

No ▼

Use inverse probability of treatment weighting?

No ▼

Use regularization when fitting the outcome model? *

Yes ▼

* Regularization will not be applied to the exposure of interest, just to the covariates.

Control Settings ▾ **Prior** ▾

Figure 12.15: アウトカムモデルの設定

Negative Control Outcome Cohort Definition

This expression will define the criteria for inclusion and duration of time for cohorts intended for use as negative control outcomes. The type of occurrence of the event when selecting from the domain.

First occurrence ▼

When true, descendant concepts for the negative control outcome concept IDs will be used to detect the outcome and roll up the occurrence to the concept ID.

Yes ▼

What domains should be considered to detect negative control outcomes? (Hold control to select multiple domains)

Condition
Drug
Device
Measurement
Observation
Procedure
Visit

Figure 12.16: ネガティブコントロールアウトカムコホートの定義の設定

ポジティブコントロールの合成

ネガティブコントロールに加えて、因果関係があると思われる曝露-結果ペアで、効果量が既知である。18章で説明したように、ネガティブコントロールから得られる合成ポジティブコントロールを「はい」であれば、モデル・タイプを選択しなければなりませんが、現在の選択は「Poisson」「survival」です。私たちの集団レベルの推定の研究では生存（Cox）モデルを使用するので、12.15にポジティブコントロール合成の設定を示します。

Positive Control Synthesis

Should we perform positive control synthesis? (to calibrate confidence intervals)

Model Type:

Survival

Using OHDSI covariates for model. ([Click to view details](#))

Define the time-at-risk window start, relative to target/comparator cohort entry:

1 days from cohort start date

Define the time-at-risk window end:

0 days from cohort end date

The minimum required continuous observation time (in days) prior to exposure:

365

Should only the first exposure per subject be included?

Yes

Should only the first outcome per person be considered when modeling the outcome?

Yes

Remove people with prior outcomes?

Yes

Advanced Settings start here

[Additional Settings](#)

Figure 12.17: ポジティブコントロールアウトカムの定義の設定

12.7.4

これで研究の定義が完了したので、実行可能なRパッケージとしてエクスポートできます。この（ユーティリティ）タブをクリックしてください。ここで、実行される一連の分析をレビューで

パッケージの名前を指定し、“Download(ダウンロード)”をクリックしてzipファイルをダウンロード(Wickham, 2015)。このパッケージを使用するには、R Studioの使用をお勧めします。R Studioをローカルで実行している場合は、ファイルを解凍し、.RprojファイルをダブルクリックR Studioで開きます。RスタジオをRスタジオサーバーで実行している場合は、 UploadをクリックR Studioでプロジェクトを開いたら、READMEファイルを開き、指示に従ってください。すべ

研究の実行時に表示される一般的なエラーメッセージは、“High correlation between covariate(s) and treatment detected.”（共変量と治療の間に高い相関が検出されました。これは傾向モデルのフィッティングの際に、いくつかの共変量が曝露と高い相関があります（セクション 12.1.2 参照）。

12.8 R

研究を実行するためのRコードを記述するためにATLASを使用する代わりに、RコードをRパッケージが提供する機能と組み合わせる必要があります。

例として、CohortMethodパッケージを使用して研究を実行します。CohortMethodは、12.8.6において、これを拡張してAMIとネガティブコントロールアウトカムを含める方法

12.8.1

最初にターゲットコホートおよびアウトカムコホートをインスタンス化する必要があります（10）で説明しています。付録にはターゲット（付録 B.2）、比較（付録 B.5）、およびアウトカム（付録 B.4）コホートの完全な定義が示されています。ACE 1、2、3である scratch.my_cohorts という表にインスタンス化されていると仮定します

12.8.2

最初に、Rにサーバーへの接続方法を教える必要があります。CohortMethodはDatabaseConnectorパッケージを使用しており、createConnectionDetailsと

```
library(CohortMethod)
connDetails <- createConnectionDetails(dbms = "postgresql",
                                         server = "localhost/ohdsi",
                                         user = "joe",
                                         password = "supersecret")

cdmDbSchema <- "my_cdm_data"
cohortDbSchema <- "scratch"
cohortTable <- "my_cohorts"
cdmVersion <- "5"
```

最後の4行はcdmDbSchema、cohortDbSchema、およびcohortTable変数と、CDMバージョンSQL Serverの場合、データベーススキーマはデータベースとスキーマの両方を指定する必要があります。“<- "my_cdm_data.dbo"”のようになります。

次に、CohortMethodにコホートを抽出し、共変量を構築し、分析に必要なすべてのデー

```
# aceI <- c(1335471, 1340128, 1341927, 1363749, 1308216, 1310756, 1373225,
       1331235, 1334456, 1342439)
```

```

thz <- c(1395058,974166,978555,907013)

#
cs <- createDefaultCovariateSettings(excludedCovariateConceptIds = c(aceI,
                                                                     thz),
                                       addDescendantsToExclude = TRUE)

#
cmData <- getDbCohortMethodData(connectionDetails = connectionDetails,
                                   cdmDatabaseSchema = cdmDatabaseSchema,
                                   oracleTempSchema = NULL,
                                   targetId = 1,
                                   comparatorId = 2,
                                   outcomeIds = 3,
                                   studyStartDate = "",
                                   studyEndDate = "",
                                   exposureDatabaseSchema = cohortDbSchema,
                                   exposureTable = cohortTable,
                                   outcomeDatabaseSchema = cohortDbSchema,
                                   outcomeTable = cohortTable,
                                   cdmVersion = cdmVersion,
                                   firstExposureOnly = FALSE,
                                   removeDuplicateSubjects = FALSE,
                                   restrictToCommonPeriod = FALSE,
                                   washoutPeriod = 0,
                                   covariateSettings = cs)

cmData

```

```

## CohortMethodData
##
##      ID 1
##      ID 2
##      ID(s) 3

```

多くのパラメーターがありますが、すべてCohortMethodマニュアルに文書化されています。
(12.1)で述べたように、共変量のセットからターゲットと比較対照となる治療を除外する必要

コホート、アウトカム、および共変量に関するすべてのデータはサーバーから抽出され、cohort
(8.4.2)で述べたように）。

抽出したデータの詳細を確認するために、汎用summary()関数を使用できます：

```
summary(cmData)
```

```

## CohortMethodData
##
##      ID 1
##      ID 2

```

```
## ID(s) 3
## 67166
## 35333
##
## 3 980 891
##
## 58349
## 24484665
```

`cohortMethodData` ファイルの作成にはかなりの計算時間がかかる可能性がありますので、

```
saveCohortMethodData(cmData, "AceiVsThzForAngioedema")
```

将来のヤッショーンでデータをロードするには、`loadCohortMethodData()` 関数を使用できま

新しいユーザーの定義

通常、新しいユーザーは薬剤（ターゲットか比較対照のいずれか）の初回使用として定義

1. コホートの定義時。
 2. コホートをgetDbCohortMethodData関数を使用してロードする際、firstExposureOnDateを指定する。
 3. createStudyPopulation関数を使用して研究集団を定義する際（下記参照）。

オプション1の利点は、入力コホートがすでにCohortMethodパッケージの外部で完全に

12.8.3

通常、曝露ヨホートとアウトカムヨホートは独立して定義されます。効果サイズの推定値

`firstExposureOnly`と`removeDuplicateSubjects`を`FALSE`に設定し、`washoutPeriod`を0に設定 = 1および`startAnchor = "cohort start"`）、リスクウィンドウはコホート定義で定義され = 0および`endAnchor = "cohort end"`）。リスクウィンドウは自動的に観察終了時または研究

```
getAttritionTable(studyPop)
```

```
##                                     ...
## 1                         67212          35379 ...
## 2                         67166          35333 ...
## 3                         67061          35238 ...
## 4     1                     66780          35086 ...
```

12.8.4

`getDbcohortMethodData()`で構築された共変量を使用してプロペンシティモデルを適合し、各

```
ps <- createPs(cohortMethodData = cmData, population = studyPop)
```

`createPs`関数はCyclopsパッケージを使用して大規模な正則化ロジスティック回帰を適合します。ここでは、変数比のマッチングを使用してPSを使用します：

```
matchedPop <- matchOnPs(population = ps, caliper = 0.2,
                           caliperScale = "standardized logit", maxRatio = 100)
```

あるいは、PSを`trimByPs`、`trimByPsToEquipoise`、または`stratifyByPs`関数で使用することもできます。

12.8.5

アウトカムモデルは、どの変数がアウトカムと関連しているかを説明するモデルです。厳密な

```
outcomeModel <- fitOutcomeModel(population = matchedPop,
                                    modelType = "cox",
                                    stratified = TRUE)
```

```
outcomeModel
##      COX
##      TRUE
##      FALSE
##      FALSE
##      OK
##
##              95%    95%   logRr   seLogRr
##      4.3203  2.4531  8.0771  1.4633  0.304
```

12.8.6

一般的に、ネガティブコントロールを含む多くのアウトカムに対して複数の分析を実行す

```
#       :
ois <- c(3, 4) # Angioedema, AMI

#
ncs <- c(434165,436409,199192,4088290,4092879,44783954,75911,137951,77965,
       376707,4103640,73241,133655,73560,434327,4213540,140842,81378,
       432303,4201390,46269889,134438,78619,201606,76786,4115402,
       45757370,433111,433527,4170770,4092896,259995,40481632,4166231,
       433577,4231770,440329,4012570,4012934,441788,4201717,374375,
       4344500,139099,444132,196168,432593,434203,438329,195873,4083487,
       4103703,4209423,377572,40480893,136368,140648,438130,4091513,
       4202045,373478,46286594,439790,81634,380706,141932,36713918,
       443172,81151,72748,378427,437264,194083,140641,440193,4115367)

tcos <- createTargetComparatorOutcomes(targetId = 1,
                                         comparatorId = 2,
                                         outcomeIds = c(ois, ncs))

tcosList <- list(tcos)
```

次に、先ほどの例で説明した様々な関数を呼び出す際に、どのような引数を使うべきかを

```
aceI <- c(1335471,1340128,1341927,1363749,1308216,1310756,1373225,
         1331235,1334456,1342439)
thz <- c(1395058,974166,978555,907013)

cs <- createDefaultCovariateSettings(excludedCovariateConceptIds = c(aceI,
                                                                     thz),
                                         addDescendantsToExclude = TRUE)

cmdArgs <- createGetDbCohortMethodDataArgs(
  studyStartDate = "",
  studyEndDate = "",
  firstExposureOnly = FALSE,
  removeDuplicateSubjects = FALSE,
  restrictToCommonPeriod = FALSE,
  washoutPeriod = 0,
  covariateSettings = cs)

spArgs <- createCreateStudyPopulationArgs(
  firstExposureOnly = FALSE,
  restrictToCommonPeriod = FALSE,
  washoutPeriod = 0,
  removeDuplicateSubjects = "remove all",
```

```

removeSubjectsWithPriorOutcome = TRUE,
minDaysAtRisk = 1,
startAnchor = "cohort start",
addExposureDaysToStart = FALSE,
endAnchor = "cohort end",
addExposureDaysToEnd = TRUE)

psArgs <- createCreatePsArgs()

matchArgs <- createMatchOnPsArgs(
  caliper = 0.2,
  caliperScale = "standardized logit",
  maxRatio = 100)

fomArgs <- createFitOutcomeModelArgs(
  modelType = "cox",
  stratified = TRUE)

```

次に、これらを1つの分析設定オブジェクトに結合し、一意の分析IDといくつかの説明を提供

```

cmAnalysis <- createCmAnalysis(
  analysisId = 1,
  description = "Propensity score matching",
  getDbCohortMethodDataArgs = cmdArgs,
  createStudyPopArgs = spArgs,
  createPs = TRUE,
  createPsArgs = psArgs,
  matchOnPs = TRUE,
  matchOnPsArgs = matchArgs
  fitOutcomeModel = TRUE,
  fitOutcomeModelArgs = fomArgs)

cmAnalysisList <- list(cmAnalysis)

```

これで、すべての比較と分析設定を含む研究を実行することができます：

```

result <- runCmAnalyses(connectionDetails = connectionDetails,
                        cdmDatabaseSchema = cdmDatabaseSchema,
                        exposureDatabaseSchema = cohortDbSchema,
                        exposureTable = cohortTable,
                        outcomeDatabaseSchema = cohortDbSchema,
                        outcomeTable = cohortTable,
                        cdmVersion = cdmVersion,
                        outputFolder = outputFolder,
                        cmAnalysisList = cmAnalysisList,
                        targetComparatorOutcomesList = tcosList)

```

`result`オブジェクトには、作成されたすべての成果物への参照が含まれます。例えば、`A`

```
omFile <- result$outcomeModelFile[result$targetId == 1 &
                                    result$comparatorId == 2 &
                                    result$outcomeId == 4 &
                                    result$analysisId == 1]
outcomeModel <- readRDS(file.path(outputFolder, omFile))
outcomeModel
```

```
## Model type: cox
## Stratified: TRUE
## Use covariates: FALSE
## Use inverse probability of treatment weighting: FALSE
## Status: OK
##
##           95%    95%   logRr   seLogRr
##     1.1338    0.5921    2.1765  0.1256    0.332
```

また、1つのコマンドですべてのアウトカムに対する効果量推定値を取得することもでき

```
summ <- summarizeAnalyses(result, outputFolder = outputFolder)
head(summ)
```

| ## | ID | ID | ID | ID | ... | | | | |
|------|----|----|----|----|-----|---|-------|-----------|-----|
| ## 1 | | 1 | | 1 | | 2 | 72748 | 0.9734698 | ... |
| ## 2 | | 1 | | 1 | | 2 | 73241 | 0.7067981 | ... |
| ## 3 | | 1 | | 1 | | 2 | 73560 | 1.0623951 | ... |
| ## 4 | | 1 | | 1 | | 2 | 75911 | 0.9952184 | ... |
| ## 5 | | 1 | | 1 | | 2 | 76786 | 1.0861746 | ... |
| ## 6 | | 1 | | 1 | | 2 | 77965 | 1.1439772 | ... |

12.9

私たちの推定値は、いくつかの仮定が満たされている場合にのみ有効です。これが満たさ

12.9.1

まず、ターゲットコホートと比較対象コホートがある程度比較可能かどうかを評価する必
12.18に示すような選好スコア分布も生成できます。この図から、多くの人々にとって受

一般的に、特にモデルが非常に予測的である場合には、傾向モデル自体も検査することか
12.7は、私たちの傾向モデルにおける主要な予測因子を示しています。変数があまりに

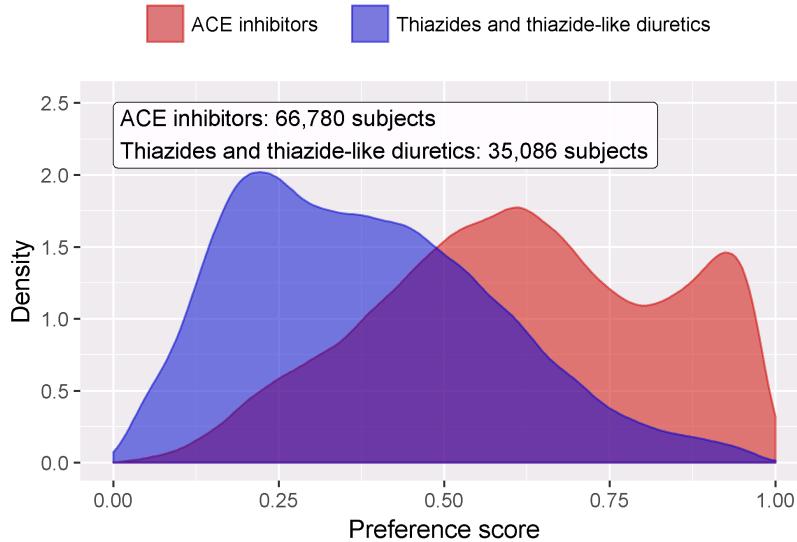


Figure 12.18: 選好スコアの分布

Table 12.7: ACEiおよびTHZの傾向モデルにおける上位10予測因子。正の値は、共変量

| ペーテ | 共変量 |
|-------|---|
| -1.42 | 基準日から-30日から0日までの期間の疾患エラ: 浮腫 |
| -1.11 | 基準日から0日から0日までの期間の薬剤エラ: 塩化カリウム |
| 0.68 | 年齢グループ: 05-09 |
| 0.64 | 基準日から-365日から0日までの期間のメジャーメント: レニン |
| 0.63 | 基準日から-30日から0日までの期間の疾患エラ: 莎麻疹 |
| 0.57 | 基準日から-30日から0日までの期間の疾患エラ: タンパク尿 |
| 0.55 | 基準日から-365日から0日までの期間の薬剤エラ: インスリン及び類似体 |
| -0.54 | 人種: 黒人またはアフリカ系アメリカ人 |
| 0.52 | (切片) |
| 0.50 | 性別: 男性 |



変数が非常に予測的であると判明した場合、2つの可能な結論があります。変数が明らかに

12.9.2

PSを使用する目的は、2つのグループを比較可能にすることです（少なくとも比較可能なグループ12.19を生成できます。指標の1つの目安は、傾向スコア調整後の絶対標準化差が0.1を超える

12.9.3

アウトカムモデルを適合させる前に、特定の効果サイズを検出するための十分なパワーがある

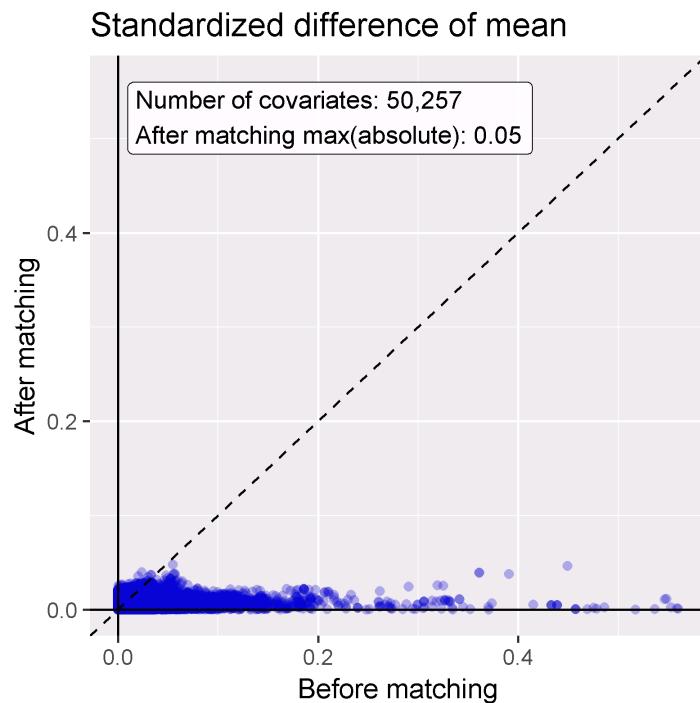


Figure 12.19: 共変量バランスの図。傾向スコアマッチング前およびマッチング後の平均の絶対標準化差を示す。各ドットは共変量を表す。

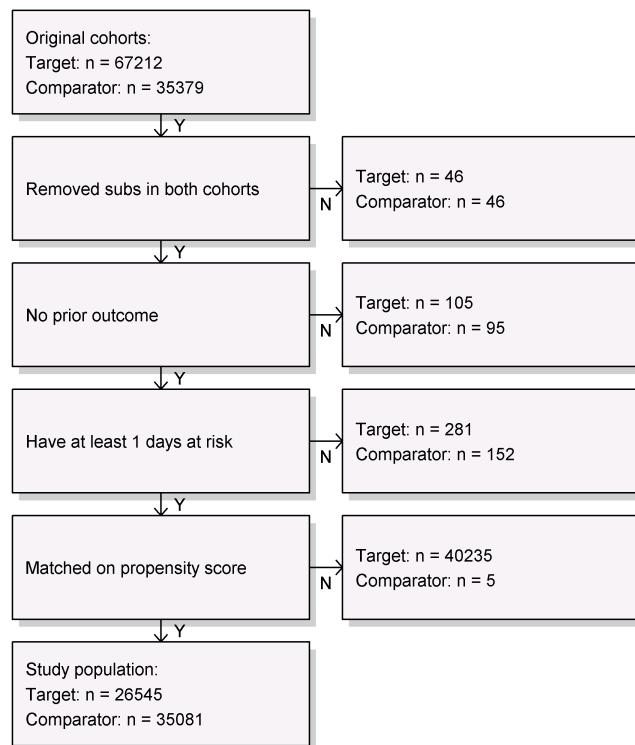


Figure 12.20: 脱落図。上部に示されているカウントは目標および比較対象コホートの定義を示す。

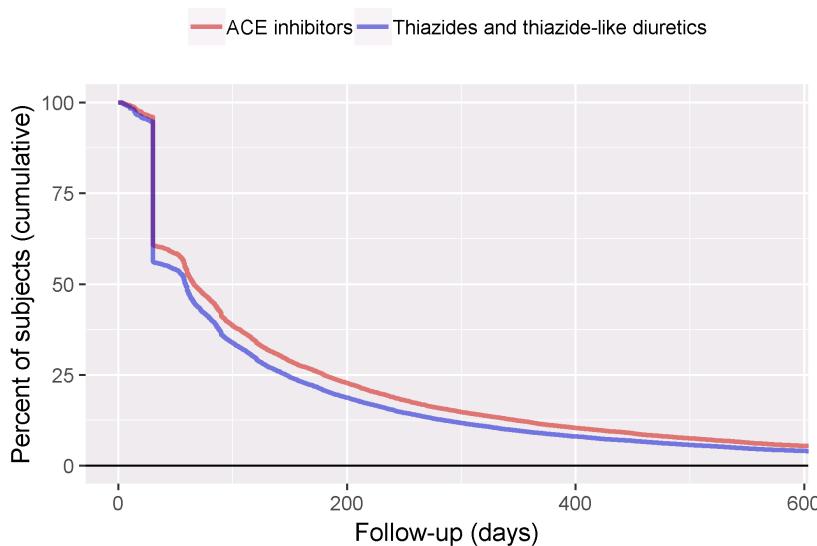


Figure 12.21: ターゲットおよび比較対象コホートのフォローアップ時間の分布

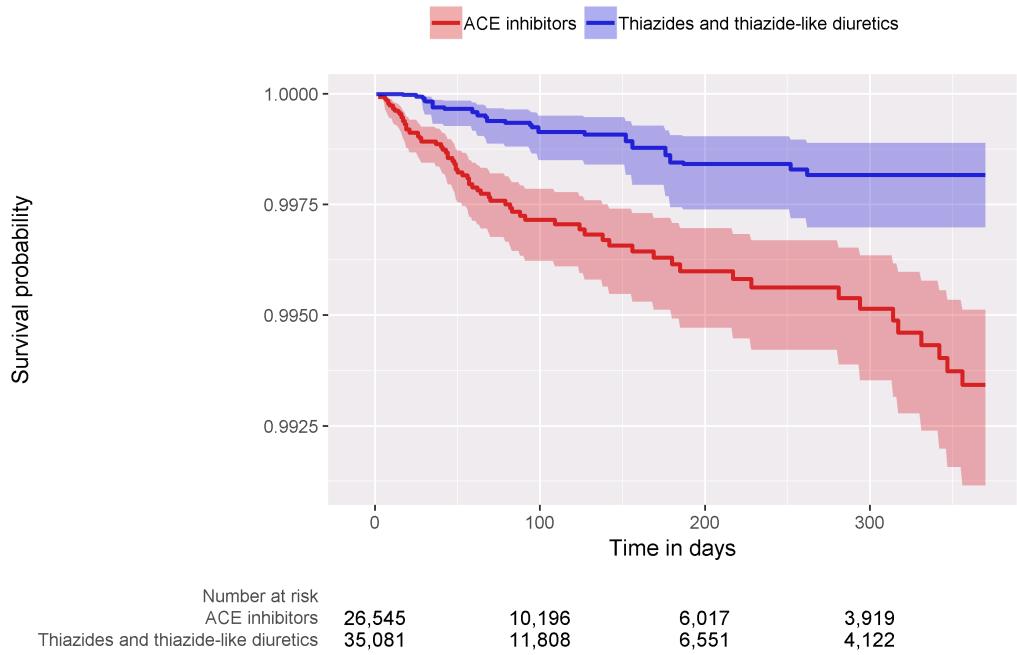


Figure 12.22: カプラン-マイヤー プロット

12.9.5

私たちは血管浮腫に対するハザード比は4.32（95%信頼区間：2.45 - 8.08）を観察しました（2.18）を観察し、AMIに対してはほとんどまたは全く効果がないことを示唆しています。14章で説明されている研究診断ではカバーされていない多くの要因に依存します。

12.10



- 集団レベルの推定は、観察データから因果効果を推測することを目的としています。
- 反事実とは、被験者が別の曝露または何も曝露を受けなかった場合に何が起こるかを想定する方法です。
- 異なる設計は、異なる方法で反事実を構築しようとします。
- OHDSIメソッドライブラリに実装されているさまざまな設計は、適切な反事実分析を可能にします。

12.11

前提条件

これらの演習を行うためには、R、R-Studio、およびJavaがセクション 8.4.5 で説明されているようにインストールされていることを前提とします。また、SqlRender、Da

```
install.packages(c("SqlRender", "DatabaseConnector", "remotes"))
remotes::install_github("ohdsi/Eunomia", ref = "v1.0.0")
remotes::install_github("ohdsi/CohortMethod")
```

Eunomiaパッケージは、ローカルのRセッション内で実行されるCDM内のシミュレートされた

```
connectionDetails <- Eunomia::getEunomiaConnectionDetails()
```

CDMデータベースのスキーマは「main」です。また、これらの演習ではいくつかのコホート：

```
Eunomia::createCohorts(connectionDetails)
```

問題定義

セレコキシブの新規使用者とジクロフェナクの新規使用者における消化管（GI）出血のリスクを評価する。

セレコキシブ新規使用者コホートのCOHORT_DEFINITION_IDは1です。ジクロフェナク新規使用者コホートのCOHORT_DEFINITION_IDは2です。

演習 12.1. CohortMethod Rパッケージを使用して、デフォルトの共変量セットを使用し、CDMデータベース内の新規使用者コホートを評価します。

演習 12.2. createStudyPopulation関数を使用して、180日のウォッシュアウト期間を要求します。

演習 12.3. 調整を行わずにコックス比例ハザードモデルを適合させます。これを行う際に何が考慮されるべきか？

演習 12.4. 傾向スコアモデルを適合させます。2つの群は比較可能ですか？

演習 12.5. 5つの層を使用してPS階層化を行います。共変量バランスは達成されましたか？

演習 12.6. PS階層を使用してコックス比例ハザードモデルを適合させます。そのアウトカムが何であるか？

付録 E.8 に回答例があります。

Chapter 13

著者: Peter Rijnbeek & Jenna Reps

臨床意思決定は、利用可能な患者の病歴と現在の臨床ガイドラインに基づいて診断や治療方針を決定するプロセスです。

過去10年間で、臨床予測モデルを説明する出版物の数が大幅に増加しました。現在使用される多くの臨床予測モデルは、OHDSIなどの大規模データセットを用いて開発されています。

大規模データセットの解析に対する機械学習の進歩により、この種類のデータに対する患者個別の予測が可能になりました。¹ 予測モデルの開発と検証を報告するための明確な推奨事項を提供し、透明性に関する標準化されたフレームワークを確立するための取り組みが進められています。

OHDSIのおかげで、大規模、患者特異的予測モデリングが現実のものとなり、共通データベースによる標準化された患者レベル予測のフレームワーク (Reps et al., 2018) を説明し、開発と検証のための確立されたベストプラクティスを実装する PatientLevelPrediction R パッケージについて説明します。まず、患者レベルの予測の開発手順について概要を述べます。

13.1

図 13.1 は私たちが取り組む予測課題を示しています。リスク集団の中で、定義された時間 (t=0) でどの患者がリスク期間中にあるアウトカムを経験するかを予測することを目指します。

表 13.1 に示すように、予測課題を定義するには、ターゲットコホートによって t=0 を定義します。

[ターゲットコホートの定義]の中で、リスク期間内に[アウトカムコホートの定義]を定義します。さらに、開発したいモデルのデザイン選択肢を検討し、内部および外部検証を行うための評価指標を定義します。

¹<https://www.equator-network.org/reporting-guidelines/tripod-statement/>

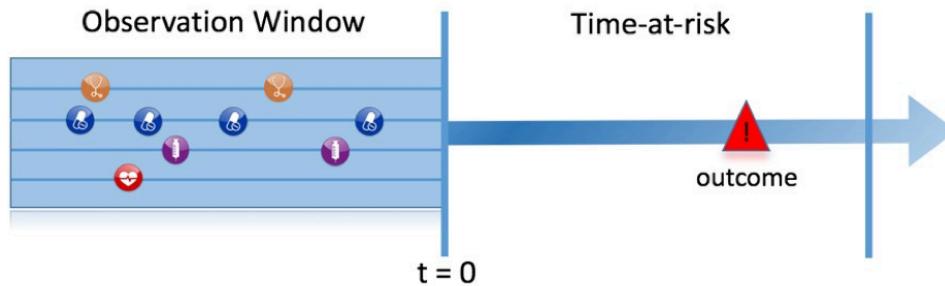


Figure 13.1: 予測課題

Table 13.1: 予測デザインにおける主要なデザイン選択肢

| 選択肢 | 説明 |
|-----------|-------------------------------|
| ターゲットコホート | 予測したい人物のコホートをどのように定義しますか？ |
| アウトカムコホート | 予測したいアウトカムをどのように定義しますか？ |
| リスク時間 | $t=0$ に対してどの時間ウィンドウで予測を行いますか？ |
| モデル | どのアルゴリズムを使用し、どの潜在的な予測変数を |

この概念的フレームワークは、すべての種類の予測課題に適用されます。例えば：

- 疾病の発症と進行
 - 構造: [病気]と新たに診断された患者の中で、[診断からの時間枠内]に[別の病気または]
 - 例: 心房細動と新たに診断された患者の中で、次の3年の間に虚血性脳卒中を発症する
- 治療選択
 - 構造: [適応された疾患]を持ち、[治療1]または[治療2]で治療された患者の中で、[治療]
 - 例: ワルファリンまたはリバロキサバンを服用した心房細動患者の中で、ワルファリ
- 治療反応
 - 構造: [治療]の新規使用者の中で、[時間枠内]に[ある効果]を経験するのは誰ですか？
 - 例: メトホルミンを開始した糖尿病患者のうち、3年間メトホルミンを継続するのは
- 治療安全性
 - 構造: [治療]の新規使用者の中で、[時間枠内]に[副作用]を経験するのは誰ですか？
 - 例: ワルファリンの新規使用者の中で、1年以内に消化管出血を経験するのは誰ですか？
- 治療遵守
 - 構造: [治療]の新規使用者の中で、[時間枠]で[遵守指標]を達成するのは誰ですか？
 - 例: メトホルミンを開始した糖尿病患者のうち、1年後に80%以上の日数カバー率を

13.2

予測モデルを作成する際には、監督学習として知られるプロセスを使用します。これは、機械共変量（“予測因子”、“特徴”、または“独立変数”とも呼ばれる）は、患者の特性を説明

11章で説明しています。予測のためには、個人がターゲットコホートに入る日付（本書また、リスク期間中の全ての患者のアウトカムステータス（“ラベル”または”クラス”）

13.2.1

表13.2は、2つのコホートが含まれたCOHORTテーブルの例を示しています。コホート

Table 13.2: 例示的なCOHORTテーブル。簡潔のためにCOHORT_END_DATE

| COHORT_DEFINITION_ID | SUBJECT_ID | COHORT_START_DATE |
|----------------------|------------|-------------------|
| 1 | 1 | 2000-06-01 |
| 1 | 2 | 2001-06-01 |
| 2 | 2 | 2001-07-01 |

表13.3は、例示的なCONDITION_OCCURRENCEテーブルを示しています。Concept ID 320128は「本態性高血圧」に該当します。

Table 13.3: 例示的なCONDITION_OCCURRENCEテーブル。簡潔のため、3つの列のみ表示しています。

| PERSON_ID | CONDITION_CONCEPT_ID | CONDITION_START_DATE |
|-----------|----------------------|----------------------|
| 1 | 320128 | 2000-10-01 |
| 2 | 320128 | 2001-05-01 |

この例示的なデータに基づき、時間のリスクが基準日（ターゲットコホートの開始日）か1に対して0（非存在）（状態が基準日後に発生）と、個人ID2に対して1（存在）を持ち1に対して0（この人はアウトカムコホートにエントリがない）、個人ID2に対して1（基準日から1年以内にアウトカムが発生）となります。

13.2.2 vs

観察医療データは、値が否定的か欠損しているかを示すことは滅多にありません。前の例1の人が基準日前に本態性高血圧の発生がなかったことを観察しました。これは、その時

13.3

予測モデルの適合を行う際には、ラベル付きの例から共変量と観測されたアウトカム状態13.2を参照）。この図では、データポイントの形が患者のアウトカム状態（例：脳卒中）

指導付き学習モデルは、2つのアウトカムクラスを最適に分離する決定境界を見つけよう

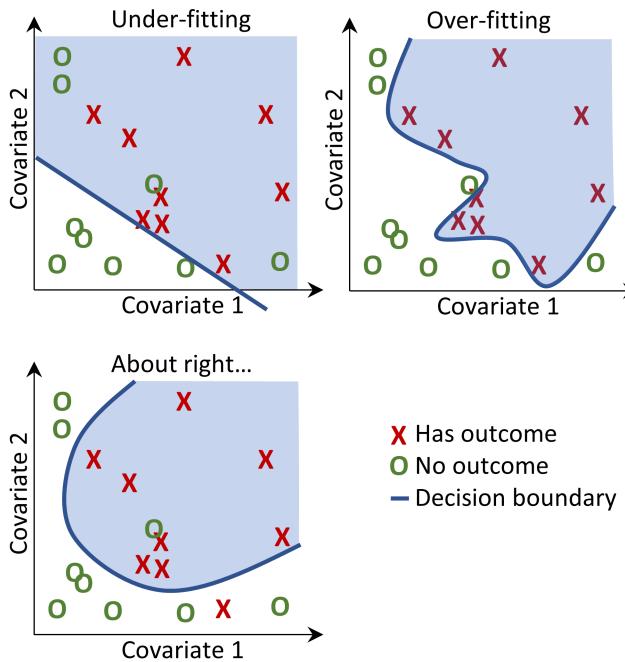


Figure 13.2: 決定境界

図 13.2 では、三つの異なる決定境界を見ることができます。境界は新しいデータポイントの予測を可能にするための規則です。各指導付き学習アルゴリズムは決定境界を学習する異なる方法を持っており、どのアルゴリズムが最も効率的かは未だ明確ではありません。Free Lunch 定理が述べているように、1つのアルゴリズムがすべての予測問題に対して常に他のアルゴリズムよりも優れているわけではありません。したがって、患者レベルの予測モデルを開発する際には、複数の指導付き学習アルゴリズムを試して、最も適切なものを選択する必要があります。

この後に示すアルゴリズムは以下から取得可能です： PatientLevelPrediction パッケージ

13.3.1

LASSO（最小絶対収縮および選択オペレーター）ロジスティック回帰は、変数の線形結合を学ぶためのアルゴリズムです。Cyclops（ロジスティック、ポアソン、サバイバル分析のためのサイクリック座標降下法）パッケージ

Table 13.4: 正則化ロジスティック回帰のハイパーパラメータ

| パラメータ | 説明 | 典型的な値 |
|-------|-----------|-------|
| 初期分散 | 事前分布の初期分散 | 0.1 |

分散はクロスバリデーションでのサンプル外の尤度を最大化して最適化されるため、初期分散

13.3.2

勾配ブースティングマシンはブースティングアンサンブル技術であり、我々の枠組みではRパッケージを使用しています。

Table 13.5: 勾配ブースティングマシンのハイパーパラメータ

| パラメータ | 説明 | 典型的な値 |
|----------------|-----------------|------------------|
| earlyStopRound | 改善がない場合の停止ラウンド数 | 25 |
| learningRate | ブースティングの学習率 | 0.005, 0.01, 0.1 |
| maxDepth | 木の最大深さ | 4, 6, 17 |
| minRows | ノード内の最小データポイント数 | 2 |
| ntrees | 木の数 | 100, 1000 |

13.3.3

ランダムフォレストは複数の決定木を組み合わせるバギングアンサンブル技術です。バギングアンサンブル技術

Table 13.6: ランダムフォレストのハイパーパラメータ

| パラメータ | 説明 | 典型的な値 |
|----------|--------|----------------------|
| maxDepth | 木の最大深さ | 4, 10, 17 |
| mtries | 各木の変数数 | -1 = 総変数数の平方根, 5, 20 |
| ntrees | 木の数 | 500 |

13.3.4 K-

K-近傍法 (K-nearest neighbors; KNN) は、ある距離メトリクスを使用して新しい未ラベル BigKnn パッケージが含まれています。

Table 13.7: K-近傍法のハイパーパラメータ

| パラメータ | 説明 | 典型的な値 |
|-------|-----|-------|
| k | 近傍数 | 1000 |

13.3.5

ナイーブベイズアルゴリズムは、クラス変数の値が与えられた全ての特徴ペアの条件付き

13.3.6 AdaBoost

AdaBoostはブースティングアンサンブル技術です。ブースティングは、繰り返し分類器を追加するAdaboostClassifier実装を使用しています。

Table 13.8: AdaBoostのハイパーパラメータ

| パラメータ | 説明 | 典型的な値 |
|--------------|---|-------|
| nEstimators | ブースティングが停止される最大推定器数 | |
| learningRate | 学習率が各分類器の貢献をlearning_rateによって抑え込みます。learning_rate | |

13.3.7

決定木は、貪欲法を用いた個々のテストを使って変数空間を分割する分類器です。クラスを分けるDecisionTreeClassifier実装を使用しています。

Table 13.9: 決定木のハイパーパラメータ

| パラメータ | 説明 | 典型的な値 |
|------------------|---------------------------------------|-------|
| classWeight | “Balance” または “None” | None |
| maxDepth | 木の最大深さ | 10 |
| minImpuritySplit | 木の成長中に早期停止するための閾値。ノードの不純物が閾値を上回る場合に停止 | |
| minSamplesLeaf | 各リーフの最小サンプル数 | 10 |
| minSamplesSplit | 各分割の最小サンプル数 | 2 |

13.3.8

多層パーセプトロンは、複数の層のノードを含むニューラルネットワークであり、入力を重み

Table 13.10: 多層パーセプトロンのハイパーパラメータ

| パラメータ | 説明 | 典型的な値 |
|-------|---------|---------|
| alpha | L2正則化 | 0.00001 |
| size | 隠れノードの数 | 4 |

13.3.9

深層学習は、ディープネット、畳み込みニューラルネットワーク、またはリカレントニューラルネットワークなどの別のビネットで、これらのモデルとハイパーパラメータの詳細について説明して

13.3.10

患者レベルの予測フレームワークには他のアルゴリズムを追加できますが、これはこの章のパッケージの“Adding Custom Patient-Level Prediction Algorithms” ビネットをご覧ください。

13.4

13.4.1

予測モデルの評価は、モデルの予測と観測されたアウトカムの一一致度を測定することによ



評価には、モデルの開発に使用されたデータセットとは異なるデータセットを使用（例：13.3 を参照）、新しい患者には適切に機能しない可能性があります。

評価の種類には、以下のものがあります：

- 内部検証：同じデータベースから抽出された異なるデータセットを使用してモデルを開発する。
- 外部検証：一つのデータベースでモデルを開発し、別のデータベースで評価します。

内部検証の方法には、次の2つがあります：

- ホールドアウトセットアプローチ：ラベル付きデータを独立した2つのセット、トレーニングセットと検証セットに分割する。
 - 時間に基づいた分割（時間的検証）：例えば、特定の日付より前のデータで訓練する、それ以降のデータで検証する。
 - 地理的位置に基づいた分割（空間的検証）。
- クロスバリデーション：データが限られている場合に有用です。データを等しいサブセット（例：10）に分割し、各セットに対して、そのセットのデータを除いた全てのデータでモデルを開発する。

外部検証は、モデルが開発された設定外の別のデータベースに対するモデルの性能を評価する方法です。

13.4.2

閾値測定

予測モデルは、リスク期間中に患者がアウトカムを持つリスクに対応する0から1の間の閾値を設定することができます。たとえば、13.11 にあるように閾値を0.5と設定すると、患者1、3、7、および10は閾値0.5以上の予測リスクを持っています。

Table 13.11: 予測確率に対する閾値の利用例

| 患者ID | 予測リスク | 0.5閾値での予測リスク | アウトカムを持つ期間中にアウトカムを持つ | |
|------|-------|--------------|----------------------|----|
| 1 | 0.8 | 1 | 1 | TP |
| 2 | 0.1 | 0 | 0 | TN |
| 3 | 0.7 | 1 | 0 | FP |

| 患者ID | 予測リスク | 0.5閾値での予測 | 約1ヶ月間中にアウトカムを持つ | |
|------|-------|-----------|-----------------|----|
| 4 | 0 | 0 | 0 | TN |
| 5 | 0.05 | 0 | 0 | TN |
| 6 | 0.1 | 0 | 0 | TN |
| 7 | 0.9 | 1 | 1 | TP |
| 8 | 0.2 | 0 | 1 | FN |
| 9 | 0.3 | 0 | 0 | TN |
| 10 | 0.5 | 1 | 0 | FP |

患者が予測されたアウトカムを持ち、実際にアウトカムを持つ場合、それを真陽性 (TP) と呼ぶ

以下の閾値ベースの指標を計算できます：

- 精度: $(TP + TN)/(TP + TN + FP + FN)$
- 感度: $TP/(TP + FN)$
- 特異度: $TN/(TN + FP)$
- 陽性予測値: $TP/(TP + FP)$

これらの値は、閾値が下げられると増減する可能性があります。分類器の閾値を下げると、 TP が増加する一方で FN は一定です）。このため、分類器の閾値を下げることで真陽性アウトカム数を増やし、感度を高めることができます。

識別力

識別力とは、リスク期間中にアウトカムを経験する患者に対して、より高いリスクを割り当てることで、正しくリスクを割り当てる能力を指します。ROCプロットの例は、この章の後半の13.17で示されています。受信者動作特性曲線下の面積 (AUC) は、識別力の全体的な指標となります。

AUCは、リスク期間中にアウトカムを経験する患者と経験しない患者との間で予測リスク分布を比較する指標です。AUCの値は、0.5から1.0まで変化します。AUCが1.0の場合、モデルは完璧にリスクを割り当てる能力があります。

非常に稀なアウトカムに対しては、AUCが高くても実際には実用的でない場合があります。なぜなら、稀なアウトカムに対する感度が低いためです。その代わりに、再現率曲線下の面積 (AUPRC) と呼ばれる別の値が推奨されます。AUPRCは、感度をx軸（再現率）として示すことで、稀なアウトカムに対するモデルの性能をより正確に評価することができます。

キャリブレーション

キャリブレーションは、モデルが正しいリスクを割り当てる能力です。例えば、モデルが100%の確率でアウトカムを予測するかどうかを確認します。これがモデルが適切にキャリブレーションされているかを示す指標です。AUCもまた、モデルが適切にキャリブレーションされているかを示す指標ですが、AUCはモデルがリスクを割り当てる能力を総合的に評価する指標です。

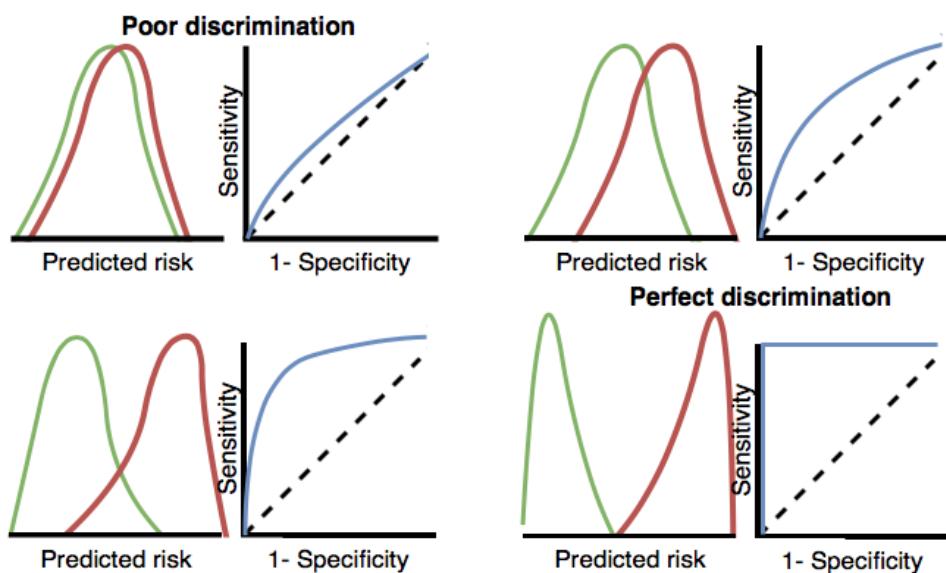


Figure 13.3: 識別力に関するROCプロット。2つのクラスの予測リスクの分布が類似している場合。

13.5

このセクションでは予測研究のデザイン方法を実演します。最初のステップは予測問題を13.1に定義された主要な選択肢を明示的に定義することにより、予測問題を明確化します。

13.5.1

血管性浮腫はACE阻害薬のよく知られた副作用であり、ACE阻害薬のラベルに記載されています(Byrd et al., 2006)。この副作用を監視することは重要です。なぜなら、血管性浮腫は稀である(Norman et al., 2013)。さらに、血管性浮腫が最初に認識されないと、その原因を特定するのに時間がかかる場合があります(Brown et al., 2013; Thompson and Frable, 1993)。アフリカ系アメリカ人患者におけるリスクは高い(Byrd et al., 2006)。ほとんどの反応は初めての治療の最初の週または月以内に、しばしば2週間以内(Cicardi et al., 2004)。しかし、一部の症例は治療開始から数年後に発生することもあります(O’ Mara and O’ Mara, 1996)。リスクのある人を特定する特定の診断テストは利用できません。

患者レベル予測フレームワークを観察医療データに適用して、次の患者レベルの予測問題を明確化します。

初めてACE阻害薬を開始した患者の中で、翌年に血管性浮腫を経験するのは誰か？

13.5.2

最終的な研究集団はターゲットコホートのサブセットであることが多いです。なぜなら、ターゲットコホートは複数の要件を満たす必要があります。

- ターゲットコホートの開始前にどの程度の観察期間が必要ですか？
この選択肢は、トレーニングデータで利用可能な患者時間や、将来モデルを適用したいデータの範囲を定義するのに役立ちます。
- 患者がターゲットコホートに複数回入ることができますか？ ターゲットコホートの定義によっては、複数回の入院や訪問が許可される場合があります。
- 以前にアウトカムを経験した人をコホートに含めることができますか？ ターゲットコホートに適格となる前にアウトカムを経験した人をコホートに含めるかどうかは、研究の目的によって異なります。
- ターゲットコホート開始日に対してアウトカムを予測する期間をどう定義しますか？ この質問に答えるために、2つの決定を下す必要があります。最初に、リスク期間の開始日を定義します。
- 最小リスク期間を要求しますか？ アウトカムが発生しなかったが、リスク期間終了前に発生する可能性がある場合は、最小リスク期間を設定する必要があります。

13.5.3

予測モデルを開発するために、どのアルゴリズムを訓練するかを決める必要があります。ある13.3に記載されているように多くのアルゴリズムを実装し、他のアルゴリズムを追加すること Boosting Machines (GBM) を一つのアルゴリズムとして選択します。

さらに、モデルを訓練するために使用する共変量を決定する必要があります。私たちの例では、モデルを訓練するためには年齢、性別、疾患歴、投薬歴などの情報を必要とします。

13.5.4

最後に、どのようにモデルを評価するかを定義する必要があります。シンプルさを追求して、約25%の分割を使用します。非常に大規模なデータセットでは、より多くのデータをトレーニング用に使用する方が良い場合があります。

13.5.5

これで、表 13.12 に示されるように、研究を完全に定義しました。

Table 13.12: 私たちの研究の主なデザイン選択

| 選択 | 値 |
|-----------|--|
| ターゲットコホート | 初めてACE阻害薬を開始した患者。以前の観察期間が3ヶ月未満である。 |
| アウトカムコホート | 血管性浮腫。 |
| リスク期間 | コホート開始後1日から365日。少なくとも364日のデータを含む。 |
| モデル | Gradient Boosting Machine with hyper-parameters ntree: 5000, max depth: 4 or 7 or 10 and learning rate: 0.001 or 0.01 or 0.1 or 0.9. Covariates will include gender, age, conditions, drugs, drug groups, and visit count. データ分割: 75% トレーニング - 25% テスト、個人ごとにランダムに割り当てられます。 |

13.6 ATLAS

予測研究をデザインするインターフェースは、ATLASメニューの左側にある  Prediction

予測デザイン機能には、予測の問題設定、分析設定、実行設定、トレーニング設定の4つ

13.6.1

ここでは、分析の対象となる母集団コホート群とアウトカムコホートを選択します。予測対象となる母集団集団コホートを選択するには、事前にATLASで定義しておく必要があり10章で説明しています。この例で使用する対象（付録B.1）およびアウトカム（付録B.4）コホートの完全な定義は付録に掲載しています。対象集団をコホートに追加するに「Target Cohort」ボタンをクリックします。アウトカムコホートの追加も同様に、「Add Outcome Cohort」ボタンをクリックすることで行います。完了すると、ダイアログが図13.4 のようになります。

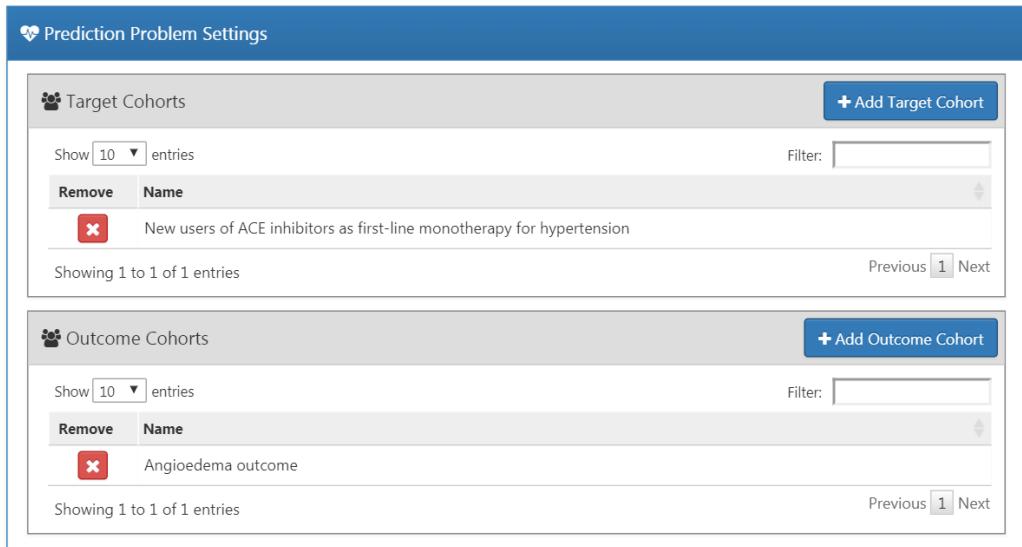


Figure 13.4: 予測問題設定

13.6.2

分析設定では、教師あり学習アルゴリズム、共変量と集団設定を選択できます。

モデル設定

モデル開発のために1つ以上の教師あり学習アルゴリズムを選ぶことができます。教師あり学習アルゴリズムを選択するには、「Model Settings」ボタンをクリックします。現在ATLASインターフェースでサポートされ

るモデルには、勾配ブースティングマシンを選択し、図13.5に示すようにハイパーパラメータを設定します。

Gradient Boosting Machine Model Settings
Use the options below to edit the model settings

The boosting learn rate (default = 0.01,0.1):

| Boosting learn rate | Action |
|---------------------|------------------------|
| 0.001 | Remove |
| 0.01 | Remove |
| 0.1 | Remove |
| 0.9 | Remove |

[Add](#) [Reset to default](#)

Maximum number of interactions - a large value will lead to slow model training (default = 4,6,17):

| Maximum number of interactions | Action |
|--------------------------------|------------------------|
| 4 | Remove |
| 7 | Remove |
| 10 | Remove |

[Add](#) [Reset to default](#)

The minimum number of rows required at each end node of the tree (default = 20):

| Minimum number of rows | Action |
|------------------------|------------------------|
| 20 | Remove |

[Add](#) Using default

The number of trees to build (default = 10,100):

| Trees to build | Action |
|----------------|------------------------|
| 5000 | Remove |

[Add](#) [Reset to default](#)

The number of computer threads to use (how many cores do you have?) (default = 20):

| | |
|----|---------------|
| 20 | Using default |
|----|---------------|

Figure 13.5: 勾配ブースティングマシン設定

共変量設定

CDM形式の観察データから抽出できる標準共変量のセットを定義しました。共変量設定研究に共変量設定を追加するには、「Add Covariate Settings」をクリックします。これ共変量設定ビューの最初の部分は除外/包括オプションです。共変量は一般に任意のコンセプトを含むことができます。“What concepts do you want to include in baseline covariates in the patient-level prediction model? (Leave blank if you want to include everything) (患者レベルの予測モデルにおけるベースライン共変量として、どのようなコンセプトを含むか?)”の下でをクリックしてコンセプトセットを選択します。コンセプトセット内のコンセプトを含むかどうかを確認するための確認用の質問 “Should descendant concepts be added to the list of included concepts? (含まれるコンセプトのリストに下位層コンセプトを追加すべきでしょうか?)” の質問に「yes」と答えます。同じプロセスを、共変量に対応する選択されたコンセプト “What concepts do you want to exclude in baseline covariates in the patient-level prediction model? (Leave blank if you want to include everything) (患者レベルの予測モデルにおけるベースライン共変量から除外したいコンセプトは何ですか?)” の質問にも繰り返します。最後のオプション “A comma delimited list of covariate IDs that should be restricted to (制限すべき共変数IDのコンマ区切りリスト)” では、共変量ID（コンセプトIDではなく）をカンマ区切りで追加し、これらがモデルに含まれるようになります。

The screenshot shows the 'Add Covariate Settings' interface with three main sections:

- Inclusion:** A text input field asking "What concepts do you want to include in baseline covariates in the propensity score model? (Leave blank if you want to include everything)". Below it is a blue save button and a red cancel button.
- Exclusion:** A dropdown menu set to "No" asking "Should descendant concepts be added to the list of included concepts?". Below it is a text input field asking "What concepts do you want to exclude in baseline covariates in the propensity score model? (Leave blank if you want to include everything)". Below it is a blue save button and a red cancel button.
- Restriction:** A dropdown menu set to "No" asking "Should descendant concepts be added to the list of included concepts?". Below it is a text input field asking "A comma delimited list of covariate IDs that should be restricted to:". Below it is a blue save button and a red cancel button.

Figure 13.6: 共変量の包括と除外設定

次のセクションでは、時間に依存しない変数の選択ができます：

- ・性別：男性または女性の性別を示す二値変数
- ・年齢：年単位の連続変数
- ・年齢グループ：5年ごとのバイナリ変数（0-4、5-9、…、95+）
- ・人種：各人種に関するバイナリ変数で、1は患者がその人種を記録していることを示す
- ・民族：各民族性に関するバイナリ変数で、1は患者がその民族性を記録していることを示す
- ・インデックス年：各コホート開始日の年を表すバイナリ変数で、1は患者のコホート開始年を示す

- インデックス月：各コホート開始日の月を表すバイナリ変数で、1は患者のコホート開始月
 - 前観察期間：[予測には推奨されません]コホート開始日以前に患者がデータベースに存在する期間
 - 後観察期間：[予測には推奨されません]患者がコホート開始日以降データベースに存在する期間
 - コホート時間：患者がコホートに属していた日数（コホート終了日－コホート開始日）
 - インデックス年と月：[予測には推奨されません]各コホート開始日の年と月の組み合わせ
- これが完了すると、このセクションは図 13.7 のようになるはずです。

Select Covariates

| | Gender | Age | Age Groups | Race | Ethnicity | Index Year | Index Month | Prior Observation Time | Post Observation Time | Time In Cohort | Index Year & Month |
|--------------|-------------------------------------|--------------------------|-------------------------------------|-------------------------------------|-------------------------------------|--------------------------|-------------------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| Demographics | <input checked="" type="checkbox"/> | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |

Figure 13.7: 共変量の選択

標準共変量は共変量の柔軟な3つの時間間隔を可能にします：

- 終了日までの日数：コホート開始日からの終了日まで[デフォルトは0]
- 長期 [デフォルトはコホート開始前365日から終了日まで]
- 中期 [デフォルトはコホート開始前180日から終了日まで]
- 短期 [デフォルトはコホート開始前30日から終了日まで]

これが完了すると、このセクションは図 13.8 のようになるはずです。

Time bound covariates

Set the time windows for the time bound covariates in days relative to the cohort index

| | Any Time Prior | Long Term | Medium Term | Short Term | End Days |
|--------------|----------------|-----------|-------------|------------|----------|
| Time Windows | All Time | -365 | -180 | -30 | 0 |

Figure 13.8: 時間に依存する共変量

次のオプションは、期間テーブルから抽出される共変量です：

- コンディション：選択された各コンディションコンセプトIDと時間間隔ごとに共変量を構築し、1、そうでない場合は 0。
- コンディショングループ：選択されたコンディションコンセプトIDと時間間隔ごとに共変量を構築し、1、そうでない場合は 0。
- 薬剤：選択された各薬剤コンセプトIDと時間間隔ごとに共変量を構築し、DRUG ERAテーブルを用いて1、そうでない場合は 0。
- 薬剤グループ：選択された各薬剤コンセプトIDと時間間隔ごとに共変量を構築し、DRUG GROUPテーブルを用いて1、そうでない場合は 0。

オーバーラップ設定には、薬剤または症状がコホート開始日以前に開始し、終了がコホート終了すると、このセクションは図 13.9 のようになるはずです。

Set the time bound era covariates

| Domain | Any Time Prior | Long Term (-365 days) | Medium Term (-180 days) | Short Term (-30 days) | Overlapping | Era Start | | |
|-----------------|--------------------------|-------------------------------------|----------------------------|-------------------------------------|--------------------------|--------------------------|----------------------------|--------------------------|
| | | | | | | Long Term (-365 days) | Medium Term (-180 days) | Short Term (-30 days) |
| Condition | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Condition Group | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Drug | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Drug Group | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |

Figure 13.9: 期間時間共変量.

次のオプションは、各ドメインでのコンセプトIDに基づく共変量に基づきます：

- コンディション：選択されたコンディションコンセプトIDと時間間隔ごとに共変量を構築し、CONDITION_OCCURRENCE 1、そうでない場合は 0。
- 主たる入院コンディション (Condition Primary Inpatient) : condition_occurrence テーブルで入院患者の主たる診断として、CONDITION_OCCURRENCE 1、そうでない場合は 0。
- 薬剤：選択された薬剤コンセプトIDと時間間隔ごとに共変量を構築し、DRUG_EXPOSURE 1、そうでない場合は 0。
- 処置（プロシージャー）：選択されたプロシージャーコンセプトIDと時間間隔ごとに共変量を構築し、PROCEDURE 1、そうでない場合は 0。
- 測定（メジャーメント）：選択されたメジャーメントコンセプトIDと時間間隔ごとに共変量を構築し、MEASUREMENT 1、そうでない場合は 0。
- 測定値：測定値が伴う選択された測定値コンセプトIDと時間間隔ごとに共変量を構築し、MEASURE_VALUE 1、そうでない場合は 0。
- 測定値範囲グループ：測定値が正常範囲以下、範囲内、または正常範囲以上である場合を示す共変量。
- 観察（オブザベーション）：選択された観察コンセプトIDと時間間隔ごとに共変量を構築し、OBSERVATION 1、そうでない場合は 0。
- デバイス：選択されたデバイスコンセプトIDと時間間隔ごとに共変量を構築し、DEVICE 1、そうでない場合は 0。
- ビジット回数：選択されたビジット回数と時間間隔ごとに共変量を構築し、その時変量値としてカウントします。
- ビジットコンセプト数：選択された各ビジット、ドメイン、および時間間隔ごとに “distinct count (重複を除いたカウント)” オプションは、ドメインと時間間隔ごとに、異なるビジットコンセプト数をカウントします。

これらが終了すると、このセクションは下図 13.10 の様になっているはずです。

最後のオプションは、一般的に使われるリスクスコアを共変量として含めるかどうかです 13.11 のようになります。

Set the time bound covariates

| Domain | Any Time Prior | Long Term (-365 days) | Medium Term (-180 days) | Short Term (-30 days) | Distinct Count | | |
|-------------------------------|--------------------------|-------------------------------------|----------------------------|--------------------------|--------------------------|----------------------------|--------------------------|
| | | | | | Long Term (-365 days) | Medium Term (-180 days) | Short Term (-30 days) |
| Condition | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Condition - Primary Inpatient | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | | | |
| Drug | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Procedure | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Measurement | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Measurement - Value | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | | | |
| Measurement - Range Group | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | | | |
| Observation | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Device | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | | | |
| Visit - Count | | <input checked="" type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | | | |
| Visit - Concept Count | | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | | | |

Figure 13.10: 時間制約共変量

Set the index score covariates

| Index Score Type | |
|---------------------------------------|-------------------------------------|
| CHADS ₂ | <input type="checkbox"/> |
| CHA ₂ DS ₂ VASc | <input checked="" type="checkbox"/> |
| DCSI | <input checked="" type="checkbox"/> |
| Charlson | <input checked="" type="checkbox"/> |

Figure 13.11: リスクスコア共変量設定

研究対象集団設定

対象集団の設定は、追加の適格基準をターゲット集団に適用できる場所であり、また、リサーチマネージャーが対象集団の設定を追加するには、“Add Population Settings(対象集団の追加)”ボタンをクリックします。これにより、対象集団の設定ビューが表示されます。

最初のオプションセットでは、リスク時間を指定することができます。これは、関心の対象が “Has outcome (アウトカムあり)” に分類し、そうでない場合は “Has outcome (アウトカムなし)” に分類します。“Define the time-at-risk window start, relative to target cohort entry: (ターゲットコホートの開始または終了日を基準とした、リスク時間ウインドウの開始を定義します。同様に、” Define the time-at-risk window end: (ターゲットコホートの開始または終了日を基準とした、リスク時間は、リスク時間ウインドウの終了を定義します。

“Minimum lookback period applied to target cohort (ターゲットコホートに適用される最小ベース期間)” は、患者がコホート開始日より前の継続的に観察された日数の最低値である、最小ベース期間です。

“Should subjects without time at risk be removed? (リスク時間がない対象は除外すべきか)” Yes (はい) に設定されている場合、“Minimum time at risk: (最低リスク時間:)” の値も必要となります。これにより、追跡不能となった人（すなわち、リスク時間中にデータが記録されない）となります。もし、その全期間にわたって観察された患者のみを含めたいのであれば subjects without time at risk be removed? (リスク時間がない対象は削除すべきですか。) を “No (いいえ)” に設定すると、リスク期間中にデータベースから脱落した患者も含めません。

“Include people with outcomes who are not observed for the whole at risk period? (全リスク時間で観察されていないアウトカムを持つ人々を含めますか。)” というオプションは、前のオプションに関連しています。Yes (はい) に設定すると、指定された最低リスク時間で観察されていない場合でも、リスク時間中にデータが記録された場合は該当する対象として扱われます。

“Should only the first exposure per subject be included? (対象ごとに最初の曝露のみを含めますか。)” というオプションは、ターゲットコホートに異なるコホート開始日で複数回含まれる患者を “Yes (はい)” を選択すると、分析では患者ごとに最も早いターゲットコホートの日付の曝露のみが選択されます。

“Remove patients who have observed the outcome prior to cohort entry? (コホート組入れの前にアウトカムが観察された患者を除外しますか。)” を “Yes (はい)” に設定すると、リスク時間開始日より前にアウトカムを経験した患者を除外します。 “No (いいえ)” が選択されると、患者は以前にアウトカムを持つ可能性があります。患者が過去にアウトカムを経験する可能性がある場合は、このオプションを選択する必要があります。

完了すると、対象集団設定のダイアログは図 13.12 のようになります。

これで分析の設定が終わり、ダイアログ全体が図 13.13 になります。

13.6.3

オプションは3つあります：

- Perform sampling(サンプリングの実行)：ここでサンプリングを実行するかどうかを “no (いいえ)” または “yes (はい)” で選択します。“yes (はい)” に設定すると、別のオプションが表示されます：

Population Settings
Add or update the population settings

Define the time-at-risk window start, relative to target cohort entry:
 days from

Define the time-at-risk window end:
 days from

Minimum lookback period applied to target cohort:

Should subjects without time at risk be removed?
 Yes Minimum time at risk: days

Include people with outcomes who are not observed for the whole at risk period?
 Yes

Should only the first exposure per subject be included?
 Yes

Remove patients who have observed the outcome prior to cohort entry?
 No

Figure 13.12: 対象集団設定

“How many patients to use for a subset? (サブセットに何人の患者を含めますか。)”
で、サンプルの大きさを指定できます。サンプリングは、患者のサンプルでモデルを作成

- “Minimum covariate occurrence: If a covariate occurs in a fraction of the target population less than this value, it will be removed: (最小共変量出現率：もし共変量がこの値より小さい割合でターゲット集団に出現する場合、ここでは共変量の出現率の最小値を選択します（デフォルトは0.001）。共変量出現の最小値）”
- “Normalize covariate (共変量を正規化する)”: ここで共変量を正規化するかどうかを選択します（“yes (はい)”）。共変量の正規化は、通常、LASSOモデルをうまく実行するために必要な手順です。

この例では、図 13.14 のように選択します。

13.6.4

4つのオプションがあります：

- “Specify how to split the test/train set (テストセットとトレーニングセットをどのように分割する)”: テストデータとトレーニングデータを人別（アウトカムで層別化）または時間別（古いデータをテストセットとして使用する）で分割します。
- “Percentage of the data to be used as the test set (0-100%)”: テストデータとして使用する割合（0-100%）を指定します。
- “The number of folds used in the cross validation (クロスバリデーションで使用する分割数)”: クロスバリデーションで使用する分割数を指定します。
- “The seed used to split the test/train set when using a person type test-Split (optional): (人単位で testSplit を使う場合の、テストセットとトレーニングセットを分割するための種子番号)”: 人単位でテストセットとトレーニングセットを分割する場合に、分割用の種子番号を指定します。

Analysis Settings

Model Settings

Show 10 entries Filter:

Remove **Model** Options

X GradientBoostingMachineSettings {"nrtrees": [5000], "nthread": 20, "maxDepth": [4, 7, 10], "minRows": [20], "learnRate": [0.001, 0.01, 0.1, 0.9], "seed": null}

Showing 1 to 1 of 1 entries Previous **1** Next

Covariate Settings

Column visibility Copy CSV Show 10 entries Filter:

Remove Options

X DemographicsGender, DemographicsAgeGroup, DemographicsRace, DemographicsEthnicity, DemographicsIndexMonth, ConditionGroupEraLongTerm (+12 more covariate settings)

Showing 1 to 1 of 1 entries Previous **1** Next

Population Settings

Column visibility Copy CSV Show 10 entries Filter:

| Remove | Risk Window Start | Risk Window End | Washout Period | Include All Outcomes | Remove Subjects With Prior Outcome | Minimum Time At Risk |
|----------|---------------------------|-----------------------------|----------------|----------------------|------------------------------------|----------------------|
| X | 1d from cohort start date | 365d from cohort start date | 365d | true | false | 364d |

Showing 1 to 1 of 1 entries Previous **1** Next

Figure 13.13: 分析の設定

Execution Settings

Perform sampling: **Yes**

How many patients to use for a subset? **500000** patients

Minimum covariate occurrence: If a covariate occurs in a fraction of the target population less than this value, it will be removed: **0.001**

Normalize covariates: **Yes**

Figure 13.14: 実行の設定

この例では、図 13.15 のように選択します。

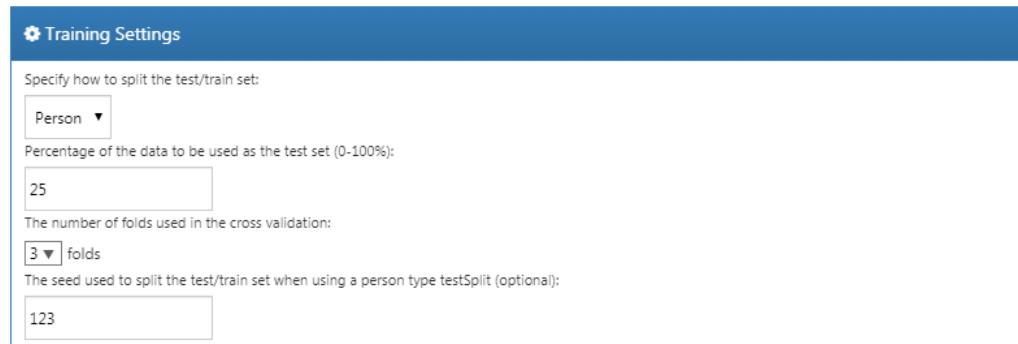


Figure 13.15: トレーニングの設定

13.6.5

研究をエクスポートするには、“Utilities (ユーティリティ)” の下にある “Export (エクスポート)” タブをクリックします。ATLASは、研究の名称、コホート定義、選択し

研究をインポートするには、“Utilities (ユーティリティ)” の下にある “Import (インポート)” タブをクリックします。患者レベルの予測研究のJSONの内容をこのウィン “Import (インポート)” ボタンをクリックします。これにより、その研究の以前の設定がすべ

13.6.6

“Utilities (ユーティリティ)” の下にある “Review & Download (レビューとダウンロード)” 」タブをクリックします。“Download Study Package (研究パッケージをダウンロード)”

セクションで、Rパッケージのわかりやすい名前を入力します。Rで使えない文字は、ATLASにをクリックして、Rパッケージをローカルフォルダにダウンロードします。

13.6.7

セクション 8.4.5 の説明のように、Rパッケージを実行するには、R、RStudio、Javaがインス PatientLevelPrediction パッケージも必要です：

```
install.packages("drat")
drat::addRepo("OHDSI")
install.packages("PatientLevelPrediction")
```

機械学習アルゴリズムの中には、追加ソフトウェアのインストールが必要なものがあります。パッケージのインストール方法の詳細については、“Patient-Level Prediction Installation Guide” vignetteを参照してください。

study Rパッケージを使用するには、R Studioの使用をお勧めします。R Studioをローカルで実行している場合は、ATLASで生成されたファイルを解凍し、.RprojファイルをR Studioで開きます。RスタジオをRスタジオサーバーで実行している場合は、 Upload をR Studioでプロジェクトを開いたら、READMEファイルを開き、指示に従ってください。

13.7 R

研究デザインをATLASで実装する代わりに、Rで直接コードを記述して実施することもできます。CDMに変換されたデータベースからデータを抽出し、モデルを構築し、評価することができます。

13.7.1

まず、ターゲットコホートとアウトカムコホートをインスタンス化する必要があります。10章で説明しています。付録にはターゲットコホート(付録B.1)とアウトカムコホート(付録B.4)の完全な定義があります。この例では、ACE阻害薬コホートのIDが1、血管浮腫コホートのIDが2です。

13.7.2

まず、Rにサーバへの接続方法を伝える必要があります。PatientLevelPredictionは、データベース接続情報を定義するための関数を提供します。

```
library(PatientLevelPrediction)
connDetails <- createConnectionDetails(dbms = "postgresql",
                                         server = "localhost/ohdsi",
                                         user = "joe",
                                         password = "supersecret")

cdmDbSchema <- "my_cdm_data"
cohortsDbSchema <- "scratch"
cohortsDbTable <- "my_cohorts"
cdmVersion <- "5"
```

最後の4行はcdmDbSchema、cohortsDbSchema、cohortsDbTable変数の定義と、CDMバージョンを定義するための変数です。SQL Serverの場合、データベーススキーマはデータベースとスキーマの両方を指定する必要があります。たとえば、`<- "my_cdm_data.dbo"`のように指定します。

まず、コホート作成が成功したかを確認するために、コホートエントリの数をカウントします。

```
sql <- paste("SELECT cohort_definition_id, COUNT(*) AS count",
             "FROM @cohortsDbSchema.cohortsDbTable",
             "GROUP BY cohort_definition_id")
conn <- connect(connDetails)
renderTranslateQuerySql(connection = conn,
                        sql = sql,
```

```
cohortsDbSchema = cohortsDbSchema,
cohortsDbTable = cohortsDbTable)
```

```
## cohort_definition_id count
## 1 1 527616
## 2 2 3201
```

PatientLevelPredictionに我々の分析に必要なすべてのデータを抽出するように指示します。

```
covariateSettings <- createCovariateSettings(
  useDemographicsGender = TRUE,
  useDemographicsAge = TRUE,
  useConditionGroupEraLongTerm = TRUE,
  useConditionGroupEraAnyTimePrior = TRUE,
  useDrugGroupEraLongTerm = TRUE,
  useDrugGroupEraAnyTimePrior = TRUE,
  useVisitConceptCountLongTerm = TRUE,
  longTermStartDays = -365,
  endDays = -1)
```

データ抽出の最終ステップは、getPlpData関数を実行し、接続詳細、コホートが保存されてい

```
plpData <- getPlpData(connectionDetails = connDetails,
                      cdmDatabaseSchema = cdmDbSchema,
                      cohortDatabaseSchema = cohortsDbSchema,
                      cohortTable = cohortsDbSchema,
                      cohortId = 1,
                      covariateSettings = covariateSettings,
                      outcomeDatabaseSchema = cohortsDbSchema,
                      outcomeTable = cohortsDbSchema,
                      outcomeIds = 2,
                      sampleSize = 10000
)
```

getPlpData関数には多くの追加パラメータがあります。これらはすべてPatientLevelPrediction plpDataオブジェクトの生成にはかなりの計算時間がかかることがあります。将来のセッションの

```
savePlpData(plpData, "angio_in_ace_data")
```

将来的なセッションでデータをロードするにはloadPlpData()関数を使用します。

13.7.3

最終的な研究対象集団は、前述の2つのコホートに追加の制約を適用することによって得られます。riskWindowStart = 30、riskWindowEnd = 365と設定します。場合によっては、リスクウィンドウをコホート終

= TRUEを設定し、コホート（曝露）期間を開始日に加算することで実現できます。

以下の例では、我々の研究のために定義したすべての設定を適用します：

```
population <- createStudyPopulation(plpData = plpData,
                                      outcomeId = 2,
                                      washoutPeriod = 364,
                                      firstExposureOnly = FALSE,
                                      removeSubjectsWithPriorOutcome = TRUE,
                                      priorOutcomeLookback = 9999,
                                      riskWindowStart = 1,
                                      riskWindowEnd = 365,
                                      addExposureDaysToStart = FALSE,
                                      addExposureDaysToEnd = FALSE,
                                      minTimeAtRisk = 364,
                                      requireTimeAtRisk = TRUE,
                                      includeAllOutcomes = TRUE,
                                      verbosity = "DEBUG"
)
```

13.7.4

アルゴリズムの設定関数において、ユーザーは各ハイパーパラメータの候補値のリストを指定します。例えば、次の設定をグラデーションブースティングマシン（Gradient Boosting Machine）で使用するとします：ntrees = c(100,200), maxDepth = 4。このグリッドサーチは、ntrees = 100およびmaxDepth = 4、またはntrees = 200およびmaxDepth = 4の設定でデフォルトの他のハイパーパラメータ設定を含めてケ

```
gbmModel <- setGradientBoostingMachine(ntrees = 5000,
                                         maxDepth = c(4,7,10),
                                         learnRate = c(0.001,0.01,0.1,0.9))
```

runPlp関数は集団、plpData、およびモデル設定を使用してモデルをトレーニングし評価を25%に分割して患者レベルの予測パイプラインを実行するためにtestSplit（人/時間）を

```
gbmResults <- runPlp(population = population,
                      plpData = plpData,
                      modelSettings = gbmModel,
                      testSplit = 'person',
                      testFraction = 0.25,
                      nfold = 2,
                      splitSeed = 1234)
```

このパッケージは内部的にRのxgboostパッケージを使用して、75%のデータを用いてグランジモードでモデルを訓練します。runPlp関数には、plpData、plpResults、plpPlots、evaluationなどのオブジェクトを

モデルを保存するには：

```
savePlpModel(gbmResults$model, dirPath = "model")
```

モデルをロードするには：

```
plpModel <- loadPlpModel("model")
```

完全なアウトカム構造を保存することもできます：

```
savePlpResult(gbmResults, location = "gbmResults")
```

完全なアウトカム構造をロードするには：

```
gbmResults <- loadPlpResult("gbmResults")
```

13.7.5

研究を実行すると、runPlp関数はトレーニング済みモデルとトレイン/テストセットに対する予測を生成します（gbmResults = gbmResults）を実行します。これによりShinyアプリケーションが開き、フレームワークによる実行が開始されます（セクション 13.16 参照）。

評価プロットをフォルダーに生成して保存するには、次のコードを実行します：

```
plotPlp(gbmResults, "plots")
```

プロットの詳細については、セクション 13.4.2 を参照してください。

13.7.6

いつでも外部バリデーションを行うことをお勧めします。すなわち、最終モデルを可能な限り複数回評価する必要があります。

```
# Load the saved model
plpModel <- loadPlpModel("model")

# Load new data
plpData <- loadPlpData("newData")

population <- createStudyPopulation(plpData = plpData,
                                      outcomeId = 2,
                                      washoutPeriod = 364,
                                      firstExposureOnly = FALSE,
                                      removeSubjectsWithPriorOutcome = TRUE,
```

```

        priorOutcomeLookback = 9999,
        riskWindowStart = 1,
        riskWindowEnd = 365,
        addExposureDaysToStart = FALSE,
        addExposureDaysToEnd = FALSE,
        minTimeAtRisk = 364,
        requireTimeAtRisk = TRUE,
        includeAllOutcomes = TRUE
    )

# apply the trained model on the new data
validationResults <- applyModel(population, plpData, plpModel)

```

さらに簡単にできるように、必要なデータの抽出も行う外部検証を行うための externalValidatePlp 関数も提供しています。 result <- runPlp(...) を実行したと仮定すると、モデルに必要なデータを抽出して、新しいデータで評価するご ID 1 と 2 のテーブル mainschema.dob.cohort にあり、CDM データがスキーマ cdmschema.dob にあると仮定すると：

```

valResult <- externalValidatePlp(
    plpResult = result,
    connectionDetails = connectionDetails,
    validationSchemaTarget = 'mainschema.dob',
    validationSchemaOutcome = 'mainschema.dob',
    validationSchemaCdm = 'cdmschema.dbo',
    databaseNames = 'new database',
    validationTableTarget = 'cohort',
    validationTableOutcome = 'cohort',
    validationIdTarget = 1,
    validationIdOutcome = 2
)

```

モデルを検証する複数のデータベースがある場合、以下を実行できます：

```

valResults <- externalValidatePlp(
    plpResult = result,
    connectionDetails = connectionDetails,
    validationSchemaTarget = list('mainschema.dob',
                                  'difschema.dob',
                                  'anotherschema.dob'),
    validationSchemaOutcome = list('mainschema.dob',
                                   'difschema.dob',
                                   'anotherschema.dob'),
    validationSchemaCdm = list('cdms1schema dbo',
                              'cdm2schema dbo',
                              'cdm3schema dbo'),
)

```

```

databaseNames = list('new database 1',
                     'new database 2',
                     'new database 3'),
validationTableTarget = list('cohort1',
                             'cohort2',
                             'cohort3'),
validationTableOutcome = list('cohort1',
                             'cohort2',
                             'cohort3'),
validationIdTarget = list(1,3,5),
validationIdOutcome = list(2,4,6)
)

```

13.8

13.8.1

予測モデルの性能を探索する最も簡単な方法は、viewPlp関数を使用することです。これはア

```

plpResult <- loadPlpResult(file.path(outputFolder,
                                       'Analysis_1',
                                       'plpResult'))

```

ここで「Analysis_1」は以前に特定した分析に対応しています。

次に、以下を実行してShinyアプリケーションを起動できます。

```
viewPlp(plpResult)
```

Shinyアプリケーションはテストセットとトレインセットの性能指標の要約から始まります（図13.16参照）。アウトカムは、トレインセットのAUCは0.78であり、これがテストセットでは13.17に示されています。

図13.18に示されているキャリブレーションプロットは、一般的に観察されたリスクが予測されるリスクよりも高くなることを示すプロットです。図13.19に示されている人口統計学的キャリブレーションプロットは、40歳未満の若年患者についてのリスクを示すプロットです。

最後に、選択基準に基づくラベル付きデータからの患者の脱落を示すattritionプロットがあります（図13.20参照）。プロットは、リスク期間全体で観察されていなかったため、ターゲット人口の

13.8.2

ATLASで生成されたスタディパッケージは、異なる予測問題に対して多くの異なる予測モデル

The screenshot shows a Shiny application window titled "c/rwd - Shiny" with the URL "http://127.0.0.1:4888". The application has tabs for "PatientLevelPrediction Explorer", "Internal Validation", and "External Validation". The "PatientLevelPrediction Explorer" tab is active, and its sub-tabs include "Evaluation Summary" (which is selected and highlighted with a blue border), "Characterization", "ROC", "Calibration", "Demographics", "Preference", "Box Plot", and "Settings".

The "Evaluation Summary" section contains a table with 11 rows of data. The columns are "Metric", "test", and "train". The data is as follows:

| Metric | test | train |
|----------------------------------|-------------|-------------|
| 1 AUC | 0.72130 | 0.75348 |
| 2 AUC_lb95ci | 0.70057 | 0.74215 |
| 3 AUC_ub95ci | 0.74203 | 0.76482 |
| 4 AUPRC | 0.10971 | 0.13571 |
| 5 BrierScaled | 0.03755 | 0.04902 |
| 6 BrierScore | 0.03355 | 0.03304 |
| 7 CalibrationIntercept.Intercept | -0.00089 | -0.00813 |
| 8 CalibrationSlope.Gradient | 1.02041 | 1.22457 |
| 9 outcomeCount | 601.00000 | 1802.00000 |
| 10 populationSize | 16685.00000 | 50054.00000 |
| 11 Incidence | 3.60204 | 3.60011 |

At the bottom left, it says "Showing 1 to 11 of 11 entries". At the bottom right, there are buttons for "Previous", a page number "1", and "Next".

Figure 13.16: Shinyアプリケーションにおける評価統計の要約

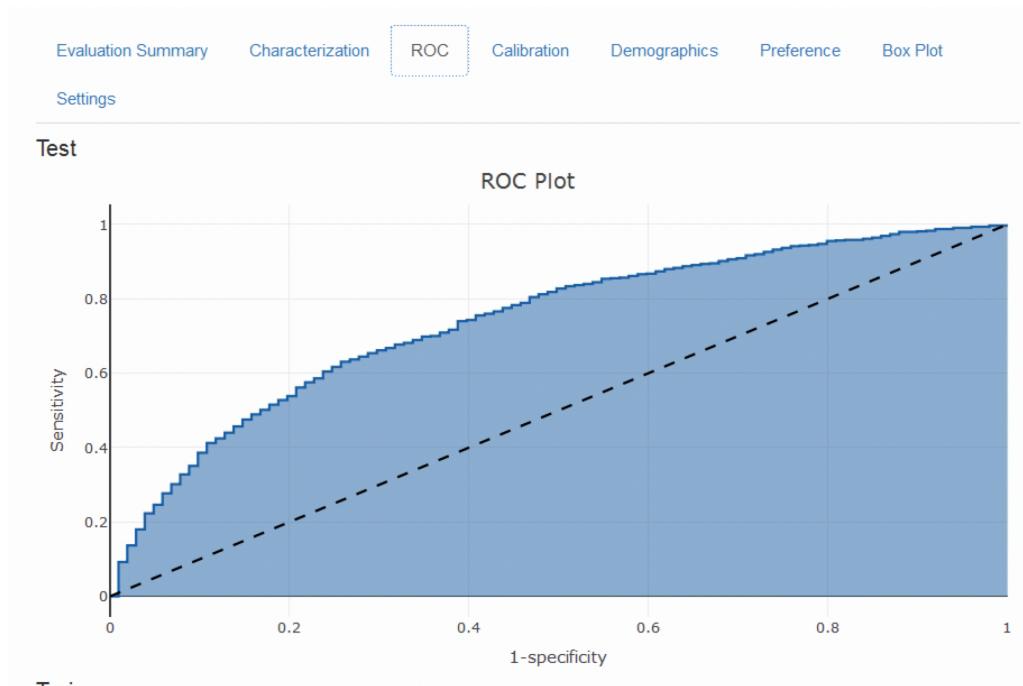


Figure 13.17: ROCプロット

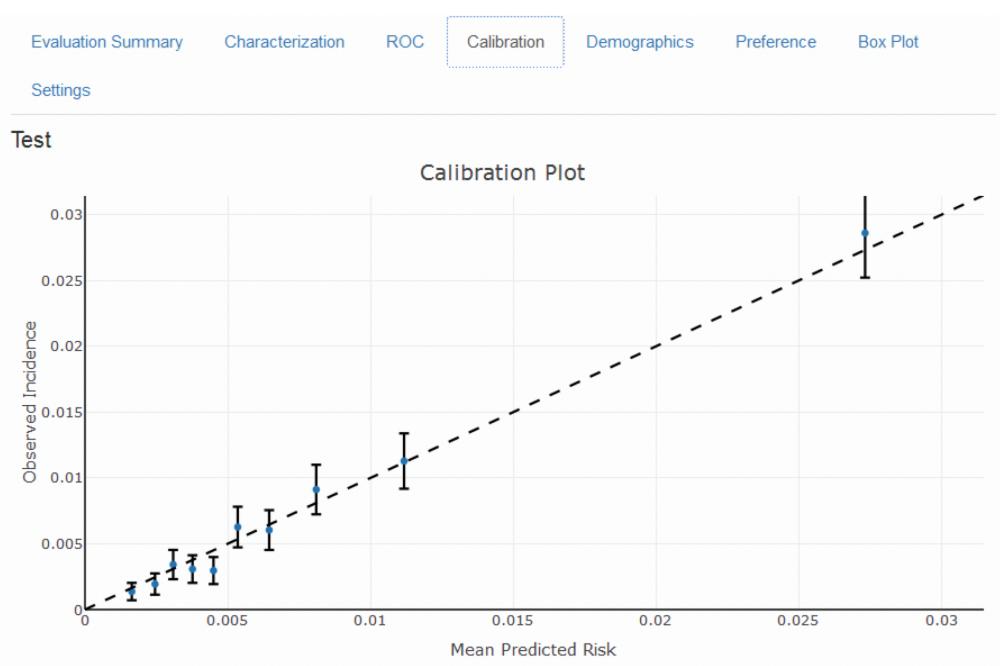


Figure 13.18: モデルのキャリブレーション

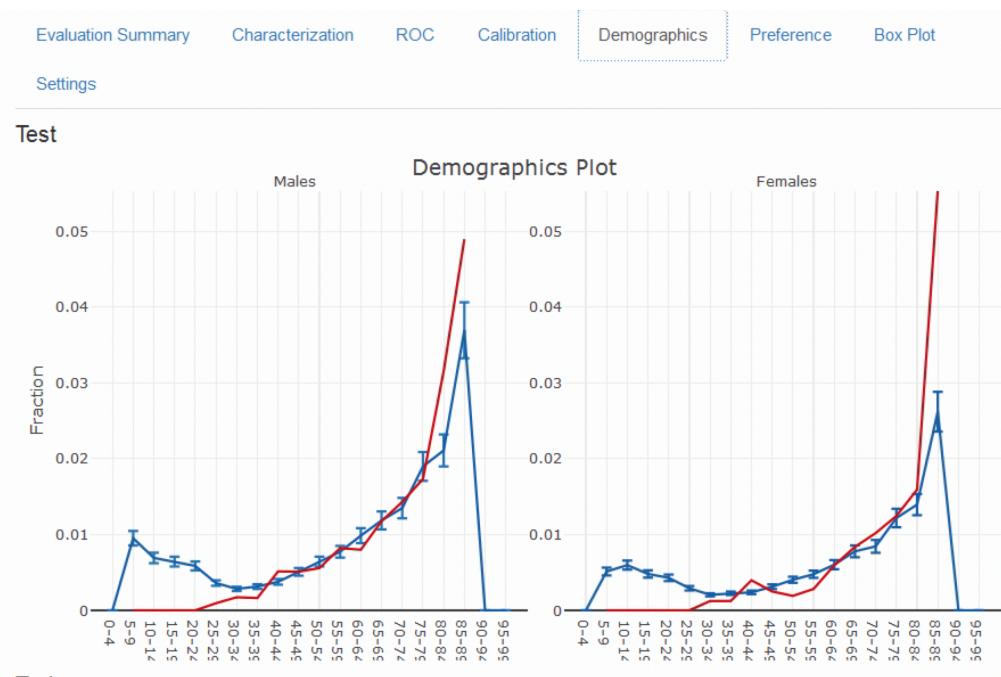


Figure 13.19: モデルの人口統計学的キャリブレーション

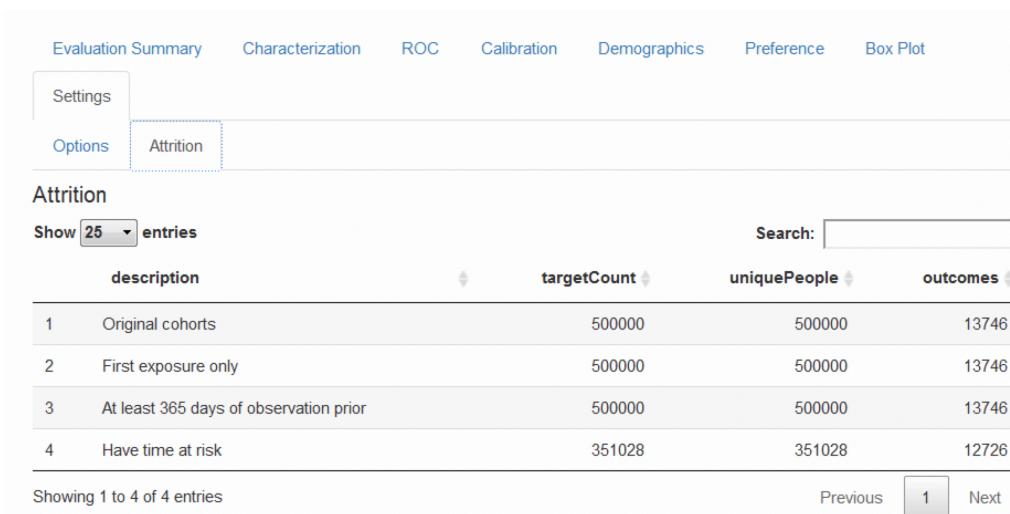


Figure 13.20: 予測問題におけるattritionプロット

モデルの要約と設定の表示

インタラクティブなShinyアプリは、図13.21に示す要約ページから始まります。

The screenshot shows a Shiny application interface for a medical study. On the left, there are four filter sections: 'Development Database' (All), 'Validation Database' (All), 'Target Cohort' (New users of ACE inhibitors as first-line monotherapy for hypertension), and 'Outcome Cohort' (All). At the top right, there are tabs for 'Results', 'Model Settings', 'Population Settings', and 'Covariate Settings'. Below these is a search bar and a dropdown for 'Show 10 entries'. The main area contains a table with columns: Analysis, Dev, Val, T, Model, TAR start, TAR end, AUC, AUPRC, T size, O Count, and Incidence (%). The table lists four rows: Analysis_1 (Lasso Logistic Regression), Analysis_3 (Lasso Logistic Regression), Analysis_5 (Random forest), and Analysis_7 (Random forest). Each row includes a detailed description of the analysis (e.g., 'New users of ACE inhibitors as first-line monotherapy for hypertension'). The table footer indicates 'Showing 1 to 4 of 4 entries'.

Figure 13.21: 各モデルのホールドアウトセットの主要な性能指標を含むShinyの要約ページ

この要約ページの表には以下が含まれています：

- モデルに関する基本情報（例：データベース情報、分類器タイプ、リスク期間設定、ターゲットコホート）
- ホールドアウトターゲット人口数およびアウトカム発生率
- 判別指標：AUC、AUPRC

テーブルの左側にはフィルターオプションがあり、開発/検証データベース、モデルの種類、関連するターゲットコホートを指定できます。

モデルを詳細に探るには、対応する行をクリックします。選択された行はハイライト表示され、各モデルの開発設定を確認できます。「Model Settings」タブをクリックしてモデルを開発する際に使用した設定を調べることができます。

The screenshot shows the 'Model Settings' tab of the Shiny application. It has tabs for 'Results', 'Model Settings' (which is active), 'Population Settings', and 'Covariate Settings'. Below these is a search bar and a dropdown for 'Show 10 entries'. The main area is a table with columns 'Setting' and 'Value'. It lists three settings: 'Model' (lr_lasso), 'variance' (0.01), and 'seed' (50975614). The table footer indicates 'Showing 1 to 3 of 3 entries'.

Figure 13.22: モデルを開発する際に使用した設定を表示する

同様に、他のタブでモデルを生成するために使用された人口および共変量の設定を調べることもできます。

モデル性能の表示

モデル行が選択されると、モデル性能も表示できます。性能評価の要約を表示します（図13.23参照）。

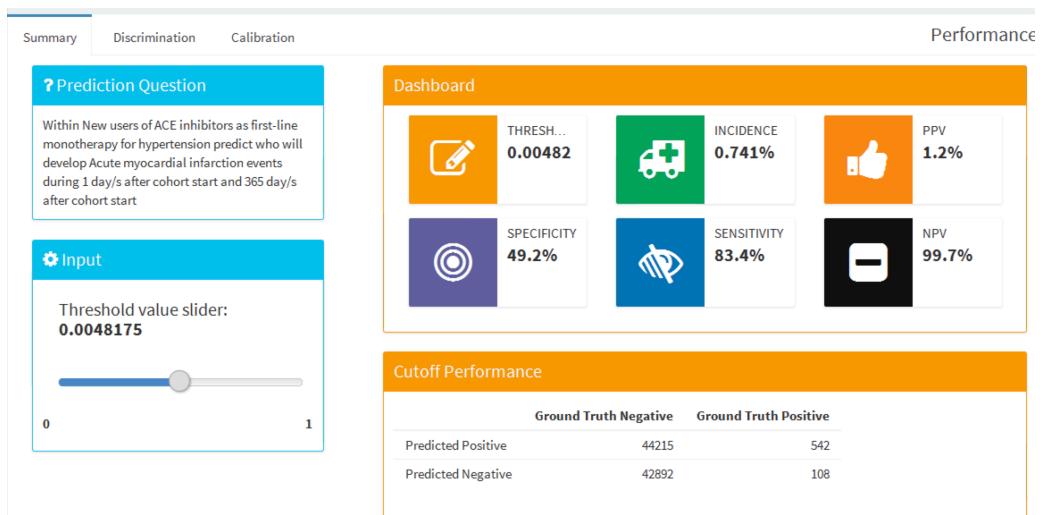


Figure 13.23: 特定の閾値における性能評価の要約

この要約ビューは標準形式で選択された予測問題を表示し、閾値セレクタと正の予測値（13.4.2 参照）などの重要な閾値ベースの指標を含むダッシュボードを表示します。図 13.23 では、閾値が0.00482で感度は83.4%（翌年のアウトカムを持つ患者の83.4%が

モデル全体の判別を確認するには、「Discrimination」タブをクリックしてROCプロット、再現プロット、および分布プロットを表示します。プロットの線は選択された閾値ポイントを示しています。ROCプロットは、モデルが翌年再現プロットを見ると性能はそれほど印象的でないよう見えます。

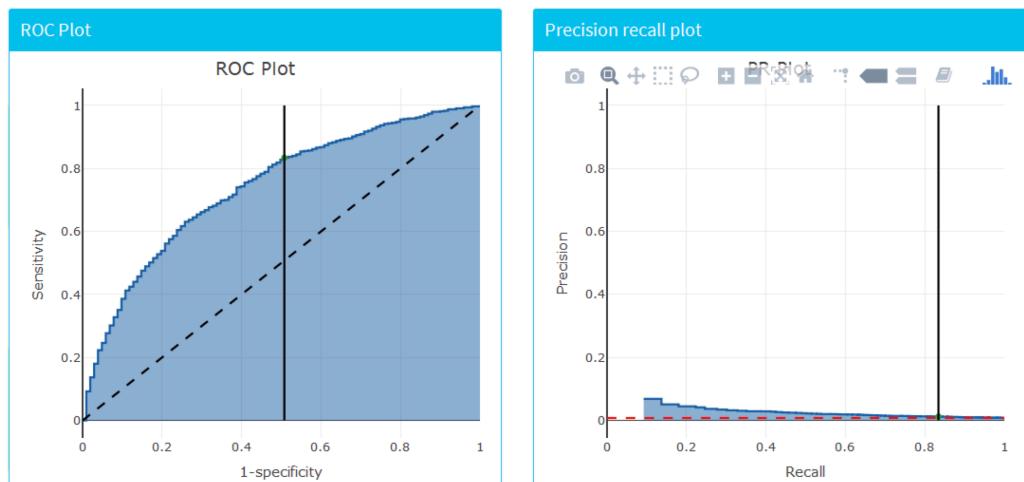


Figure 13.24: ROCおよび精度-再現プロットを使用してモデルの判別能力全体を評価する

図 13.25 は予測リスクおよび選好スコア分布を示しています。

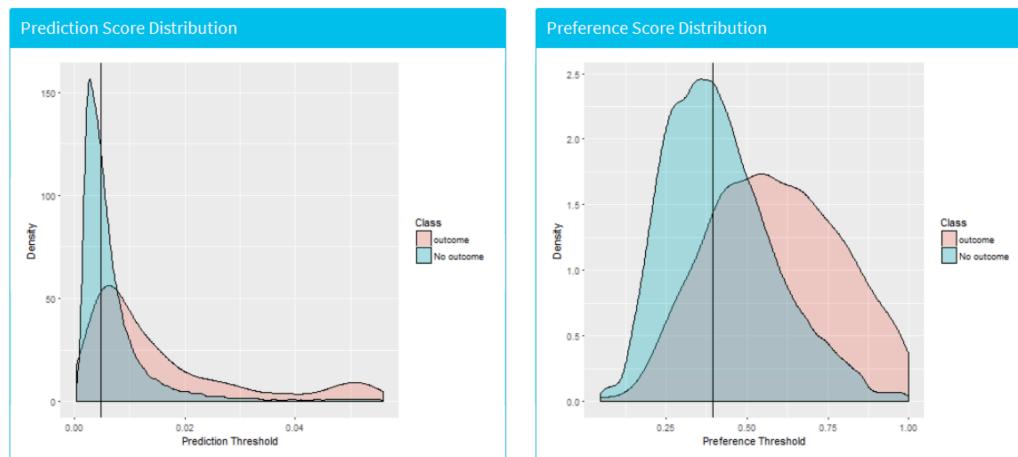


Figure 13.25: アウトカム有およびアウトカム無の患者に対する予測リスク分布。重なりが多いほど、モデルが不確実であることを示す。

最後に、「Calibration」タブをクリックしてモデルのキャリブレーションを確認することもできます。Figure 13.26 に示されるキャリブレーションプロットおよび人口統計学的キャリブレーションが表示されます。

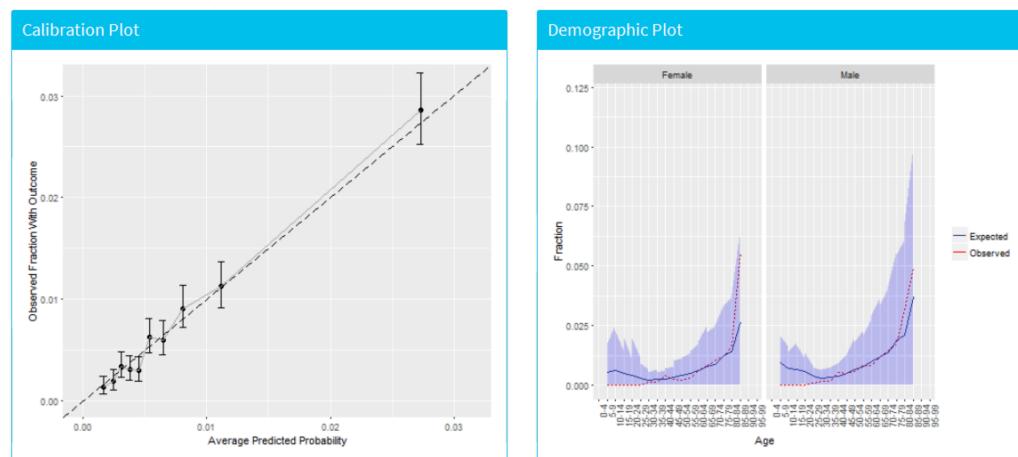


Figure 13.26: リスク層別キャリブレーションおよび人口統計学的キャリブレーション

1年以内のアウトカムを経験したグループの予測リスクと観察されたアウトカムの割合が一致する。

モデルの表示

最終モデルを検査するには、左側のメニューから Model オプションを選択します。これにより、Figure 13.27 に示すモデル内の各変数のプロットと図 13.28 に示すすべての候補共変量を要約するテーブルが表示されます。

図 13.28 に示すテーブルでは、すべての候補共変量の名前、値（一般線形モデルを使用する場合）

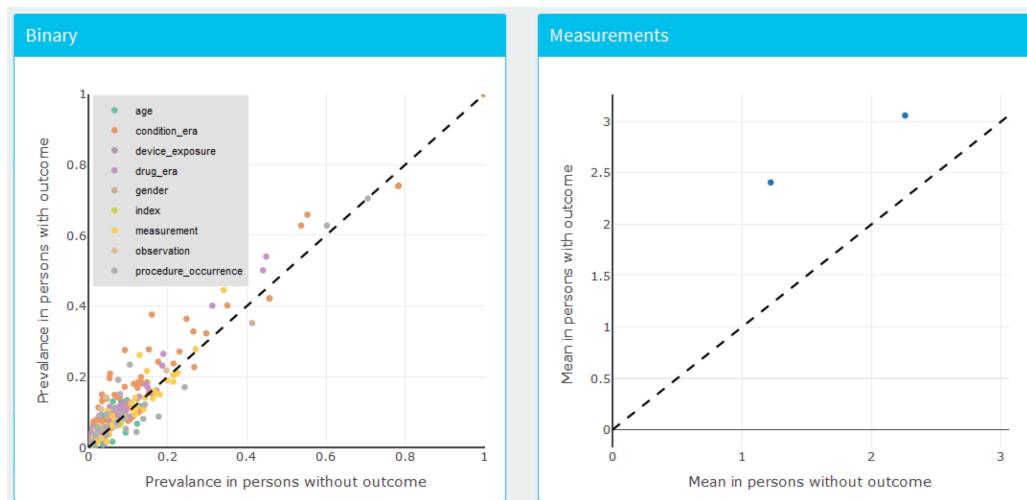


Figure 13.27: モデル要約プロット。各点はモデルに含まれる変数に対応します。

Model Table

| Covariate Name | Value | Outcome Mean | Non-outcome Mean |
|--|--------|--------------|------------------|
| 1 age group: 00-04 | 0 | 0.0004 | 0.0001 |
| 2 age group: 05-09 | 0 | 0 | 0.0003 |
| 3 index month: 1 | 0 | 0.1307 | 0.1096 |
| 4 observation during day -365 through 0 days relative to index: Domain | 0 | 0.1188 | 0.0514 |
| 5 Charlson index - Romano adaptation | 0 | 2.4783 | 1.3817 |
| 6 Diabetes Comorbidity Severity Index (DCSI) | 0.1478 | 2.4056 | 1.2207 |
| 7 CHADS2VASc | 0.9279 | 3.0573 | 2.2576 |
| 8 visit_occurrence concept count during day -365 through 0 concept_count relative to index | 0 | 19.5263 | 13.8837 |
| 9 age group: 10-14 | 0 | 0 | 0.001 |
| 10 index month: 2 | 0 | 0.0934 | 0.0909 |

Showing 1 to 10 of 67,897 entries

Figure 13.28: モデル詳細テーブル



予測モデルは因果モデルではなく、予測変数を原因と誤解しないようにしてください。図 13.28 のいずれかの変数を変更することでアウトカムのリスクが影響を受ける保証はありません。

13.9

13.9.1

自動的にワードドキュメントを生成する機能を追加しました。このドキュメントはジャーナル「1」を任意で追加することもできます。この機能を実行することでジャーナルペーパーの草稿

```
createPlpJournalDocument(plpResult = <your plp results>,
    plpValidation = <your validation results>,
    plpData = <your plp data>,
    targetName = "<target population>",
    outcomeName = "<outcome>",
    table1 = F,
    connectionDetails = NULL,
    includeTrain = FALSE,
    includeTest = TRUE,
    includePredictionPicture = TRUE,
    includeAttritionPlot = TRUE,
    outputLocation = "<your location>")
```

詳細は関数のヘルプページを参照してください。

13.10



- 患者レベルの予測は、過去のデータを使用して将来の出来事を予測するモデルを開発します。
- モデル開発に最適な機械学習アルゴリズムの選択は経験的な問題であり、具体的な問題によって異なる場合があります。
- PatientLevelPredictionパッケージは、OMOP CDMに保存されたデータを使用してモデルを開発します。
- モデルとその性能指標の発信はインタラクティブなダッシュボードを通じて行います。
- OHDSIの予測フレームワークは、臨床採用の前提条件である予測モデルの大規模な評価と検証を支援します。

13.11

前提条件

これらの演習では、セクション 8.4.5 で説明されているように、R、R-Studio、およびJavaがインストールされていることを前提としています。また、SqlRender

```
install.packages(c("SqlRender", "DatabaseConnector", "remotes"))
remotes::install_github("ohdsi/Eunomia", ref = "v1.0.0")
remotes::install_github("ohdsi/PatientLevelPrediction")
```

Eunomiaパッケージは、CDM内のシミュレートされたデータセットを提供し、これをロードするには、

```
connectionDetails <- Eunomia::getEunomiaConnectionDetails()
```

CDMデータベースのスキーマは「main」です。これらの演習ではいくつかのコホートも定義されます。

```
Eunomia::createCohorts(connectionDetails)
```

問題定義

初めてNSAIDs（非ステロイド性抗炎症剤）を使用し始めた患者において、次の年にGI出血コホートを定義するためのターゲット集団を定義するためのデザインを定義します。

演習 13.1. PatientLevelPrediction Rパッケージを使用して、予測に使用する共変量を定義します。

演習 13.2. 最終的なターゲット集団を定義するためのデザインの選択肢を再考し、これを実装します。

演習 13.3. LASSOを使用して予測モデルを構築し、Shinyアプリケーションを使用してそれを可視化します。

提案された回答は、付録 E.9で見つけることができます。

Part IV

Chapter 14

著者: Patrick Ryan & Jon Duke

14.1

どんな旅も出発する前に、理想の目的地がどのようなものかを思い描いておくことが役に立つ

| 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: 信頼できる証拠の望ましい属性

信頼性の高いエビデンスは再現性があるべきであり、特定の質問に対して同じデータ

私たちは、再現可能であることが示されれば、そのエビデンスが信頼できるものであると (1965)。患者レベルの予測では、再現性は外部検証の価値と、異なるデータベースに適用可能 (Madigan et al., 2013b) は、効果推定値がデータの選択に敏感であることを示しました。

信頼性の高い証拠は頑健であるべきであり、つまり、分析の中でなされる主観的な選択には (Madigan et al., 2013a)。母集団レベルの効果推定では、感度分析には、コホート比較研究

最後に、しかし最も重要なこととして、エビデンスは校正されるべきである。未知の質問には、ネガティブコントロールは、観察研究における系統的エラーを特定し、軽減するための強力なツール (Schuemie et al., 2016, 2018a,b)。

14.2

しかし、研究結果が十分に信頼できるものであるかどうかを、どうすれば判断できるのでしょうか？臨床現場での使用に耐えうるのでしょうか？規制当局の意思決定に利用できるのでしょうか？将来の研究の基礎として役立つのでしょうか？ランダム化比較試験、観察研究、その他の分析

観察研究や「リアルワールドデータ」の利用に関してよく挙げられる懸念事項のひとつに、データの正確性と信頼性があります (Botsis et al., 2010; Hersh et al., 2013; Sherman et al., 2016)。一般的に指摘されるのは、観察研究 (Kahn et al., 2012; Liaw et al., 2013; Weiskopf and Weng, 2013)。OHD-SIコミュニティは、こうした研究の強力な推進者であり、コミュニティのメンバーは、OMOP-CDMとOHDSIネットワークにおけるデータ品質を調査する多くの研究を主導し、または参加しています (Huser et al., 2016; Kahn et al., 2015; Callahan et al., 2017; Yoon et al., 2016)。

この分野における過去10年間の調査結果を踏まえると、データ品質は完璧ではなく、今後も完璧にはならないことが示されています。

データの正確性が損なわれるのは、医師の頭脳からカルテにデータが移動する時点から始まります。

したがって、コミュニティとして私たちは次のような問いかけをしなければなりません。不完璧なデータはなぜ生まれるのか？その答えは、「エビデンスの質」を全体的に見ることにあります。すなわち、データからエビデンスへと至る過程で何が起こっているのか？

次の章では、表 14.1 にリストされているエビデンスの質の4つのコンポーネントを探ります。

Table 14.1: エビデンスの質の4つのコンポーネント

| エビデンスの質のコンポーネント | 測定するもの |
|-----------------|---|
| データの質 | データが合意された構造と規約に準拠した形で、完全に実施された分析が臨床的な意図とどの程度一致しているか |
| 臨床的妥当性 | データの変換および分析プロセスが意図した通りに機械化されているか |
| ソフトウェアの妥当性 | データの強みと弱点を考慮した上で、その方法論が研究目的に適切であるか |
| 方法の妥当性 | データの強みと弱点を考慮した上で、その方法論が研究目的に適切であるか |

14.3

エビデンスの質に関する重要な側面は、データからエビデンスに至る過程で生じる不確実性を捉えることです。

14.4



- 我々が生成するエビデンスは、再現可能、再現実験が可能、複製可能、一般化可能、頑健性、そして較正済みでなければなりません。
- エビデンスが信頼できるかどうかを判断する際には、データの質だけでなく、エビデンスの質の4つのコンポーネントを考慮する必要があります。

- * データの品質
- * 臨床的妥当性
- * ソフトウェアの妥当性
- * 方法の妥当性

- エビデンスを伝える際には、エビデンスの質に対するさまざまな課題から生じ

Chapter 15

著者: Martijn Schuemie, Vojtech Huser & Clair Blacketer

医療観察研究に用いられるデータのほとんどは、研究目的で収集されたものではない。例 Lei(1991)は「データは収集された目的のみに使用されるべきである」とさえ述べています。

研究目的に対して、データの品質は十分でしょうか？

データ品質 (DQ) を次のように定義できます (Roebuck, 2012):

特定の使用目的に適したデータとなるような、完全性、妥当性、一貫性、適時性、データが完璧であることはまずありませんが、目的には十分である可能性があることに留意する必要があります。DQは直接観察することができませんが、それを評価する方法論が開発されています。DQ (Weiskopf and Weng, 2013): DQを全般的に評価する評価と、特定の研究におけるDQを評価する評価です。

本章では、まずDQの問題の原因となり得るものを検討し、その後、一般的なDQ評価と特徴的なDQ問題について述べます。

15.1

第14章で述べたように、医師が自身の考えを記録する段階からデータの品質に対する脅威がある。Hansen and Johnson (2003) は、データのライフサイクルにおいて次のステップを区別し、各ステップで問題となる点を示しています。

1. データ収集と統合。考えられる問題としては、手入力の誤り、バイアス（例: 保険請求におけるコーディングの誤り）、EHRでのテーブルの誤った結合、欠測値
2. データの保存と知識の共有。考えられる問題としては、データモデルの文書化不足
3. データ分析。不正確なデータ変換、データの誤った解釈、不適切な方法論の使用など
4. データの公開。下流での使用のためにデータを公開する際の問題。

私たちが使用するデータはすでに収集され統合されていることが多いため、ステップ1を考慮する必要はありません。

同様に、私たちは特定の形式でデータを頻繁に受け取っているため、ステップ2の一部に該当する場合があります。15.2.2で説明するように、DQを維持するための厳格な保護策を構築することができます。

(Defalco et al., 2013; Makadia and Ryan, 2014; Matcho et al., 2014; Voss et al., 2015a,b; Hripcsak et al., 2018)によれば、正しく実行されれば、CDMへの変換時にほとんどのステップ3（データ分析）もまた、私たちの管理下にあります。OHDSIでは、このステップに16、17、18章で詳しく議論しています。

15.2

観察研究の一般的な目的に対してデータが適しているかどうかを問うことができます。Kahn et al. (2016)は、このような一般的なデータ品質 (DQ) を次の3つの要素から構成されるもの

1. 適合性: データ値が指定された基準や形式に従っているでしょうか。3つのサブタイプに分類されます。
 - ・ 値: 記録されたデータ要素が指定された形式に合致しているでしょうか。例えば、すべての値が整数である。
 - ・ 関係: 記録されたデータが指定された関係制約に合致しているでしょうか。例えば、データの PROVIDER_ID が PROVIDER テーブルの対応するレコードと一致している。
 - ・ 計算: データに対する計算結果が意図したとおりになっているでしょうか。例えば、年齢が誕生日から計算されている。
2. 完全性: 特定の変数が存在するかどうか（例：診察室で測定された体重が記録されている）。
3. 妥当性: データ値は信頼できるでしょうか。3つのサブタイプが定義されています：
 - ・ 一意性: 例えば、PERSON テーブルで各 PERSON_ID は一度しか出現しないでしょうか。
 - ・ 非一時的: 値、分布、密度が期待される値と一致しているでしょうか？例えば、データの分布が正規分布に一致している。
 - ・ 一時的: 値の変化は期待と一致しているでしょうか？例えば、予防接種の順序は推奨通りである。

各コンポーネントは2つの方法で評価できます：

- ・ 検証では、モデルとメタデータのデータ制約、システムの前提条件、ローカルの知識に基づいています。
- ・ 妥当性（バリデーション）では、関連する外部ベンチマークとのデータ値の整合性に焦点を当てます。

15.2.1

カーンは、データが所定の要件に適合しているかどうかをテストするデータ品質チェック（Data Quality Check）（Data Quality Characterization of Health Information at Large-scale Longitudinal Evidence Systems）があります (Huser et al., 2018)。ACHILLESは、CDMに準拠したデータベースの特徴を評価するツールです (Huser et al., 2016)。ACHILLESはスタンダードアロンツールとして利用でき、「データソース」欄

ACHILLESは、170以上のデータ特性評価を事前に計算します。各分析には分析IDと分析の簡単な説明が含まれます。たとえば、「506: 性別による死亡時の年齢の分布」や「DRUG_CONCEPT_IDによるDAYS_SUPPLYの分布」。

コミュニティが作成したもう一つのツールで、DQ を評価するものに、Data Quality Dashboard (DQD) があります。ACHILLES が特性評価を実行して CDM インスタンスの全体像を視覚的に把握できるようにするのに対し、DQD は表ごとに、またフィールドごとに、CDM 内の所定の仕様を満たさないレコード数を数値化します。たとえば、1,500 を超えるチェックが実行され、それぞれが Kahn のフレームワークで整理されています。15.1 は、いくつかのチェックの例を示しています。

Table 15.1: データ品質ダッシュボードのデータ品質ルールの例

| 違反行の割合 | チェックの説明 | 閾値 | 状態 |
|--------|--|------|------|
| 0.34 | VISIT_OCCURRENCE の provider_id が仕様に基づく期待されるデータ型であるかどうかを示す yes、 no の値。 | 0.05 | FAIL |
| 0.99 | MEASUREMENT テーブルの measure-ment_source_value フィールドにある異なるソース値の数やパーセントが 0 にマッピングされている。 | 0.30 | FAIL |
| 0.09 | DRUG_ERA テーブルの drug_concept_id フィールドにある値が成分クラスに適合しない記録の数、パーセント。 | 0.10 | PASS |
| 0.02 | DRUG_EXPOSURE テーブルの DRUG_EXPOSURE_END_DATE フィールドの値が DRUG_EXPOSURE_START_DATE フィールドの日付より前に発生する記録の数、パーセント。 | 0.05 | PASS |
| 0.00 | PROCEDURE_OCCURRENCE テーブルの procedure_occurrence_id フィールドに重複する値がある記録の数、パーセント。 | | PASS |

このツールでは、チェックは複数の方法で整理されており、その一つはテーブル、フィー



ACHILLESとDQDはCDM内のデータに対して実行されます。このようにして特定さ

15.2.2 ETL - -

高度なデータ品質のチェックに加え、個別レベルでのデータチェックも実施すべきです。のように複雑なプロセスに変更を加えると予期せぬ結果を招く可能性があり、ETL のすべての側面を再考し、再評価する必要が生じます。

ETLが期待通りに動作し、その状態を維持できるようにするために、一連のユニットテ 6章で説明したRabbit-in-a-Hatツールを使用すると、このようなユニットテストを簡単に

```
source("Framework.R")
declareTest(101, "Person gender mappings")
add_enrollment(member_id = "M000000102", gender_of_member = "male")
add_enrollment(member_id = "M000000103", gender_of_member = "female")
expect_person(PERSON_ID = 102, GENDER_CONCEPT_ID = 8507)
expect_person(PERSON_ID = 103, GENDER_CONCEPT_ID = 8532)
```

この例では、Rabbit-in-a-Hatによって生成されたフレームワークがソースとして読み込まれ、in-a-Hatによって作成されたadd_enrollment関数を使用して、MEMBER_IDおよびGENDER_CONCEPT_IDの期待値を持つ2つのエントリがPERSONテーブルに存在しているはず。

ENROLLMENTテーブルには他にも多くのフィールドがありますが、このテストのコンテキストがレコードを破棄したりエラーを発生させたりする可能性があります。この問題を克服し、データ（Rabbitスキャンレポートで観測された最も一般的な値）を割り当てます。

同様のユニットテストをETLの他のすべてのロジックに対して作成することもでき、通常は複数のSQL文を生成します。

```
insertSql <- generateInsertSql(databaseSchema = "source_schema")
testSql <- generateTestSql(databaseSchema = "cdm_test_schema")
```

全体のプロセスは図 15.1 に示されています。

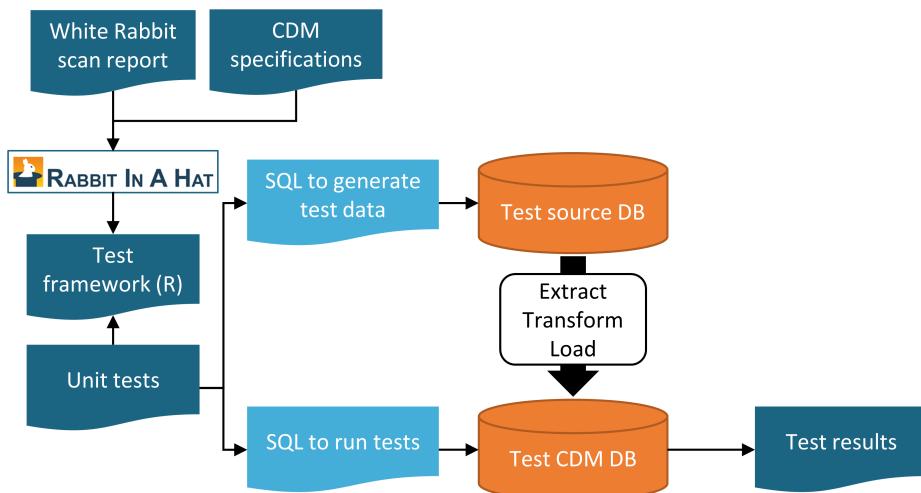


Figure 15.1: Rabbit-in-a-Hat テストフレームワークを使用した ETL (Extract-Transform-Load) プロセスの単体テスト

テスト用のSQLは、表 15.2 のようなテーブルを返します。このテーブルでは、先に定義した2

Table 15.2: ETL 単体テスト結果の例

| ID | 説明 | 状態 |
|-----|------------------------|------|
| 101 | Person gender mappings | PASS |
| 101 | Person gender mappings | PASS |

これらの単体テストの力は、ETL プロセスが変更されたときに簡単に再実行できることで

15.3

この章では、これまで一般的なDQチェックに焦点を当ててきました。このようなチェックの評価の一部は、調査に特に関連するDQルールという形式を取ることができます。標準的な評価は、ACHILLESで研究に最も関連するコンセプトを検討することであり、別の評価方法としては、研究用に開発されたコホート定義を使用して生成されたコホートを16章で説明されているように、臨床的妥当性のコンセプトと重複していることに留意く。

15.3.1

私たちの管理下で明確に該当するエラーの可能性として、ソースコードから標準コンセプトキャブラリのマッピングは入念に作成されており、コミュニティのメンバーによって指

¹に報告され、今後のリリースで修正されます。しかし、すべてのマッピングを手作業で

対応づけられたソースコードをレビューする方法のひとつに、RパッケージMethodEvaluationの15.2の出力例は、「うつ病性障害」と呼ばれるコンセプトセットの一部内訳を示しています。対象のデータベースにおけるこのコンセプトセットで最も頻度の高いコンセプトは、コンセプトID「F32.8」です。このデータベースでは、ICD-9コード3.11、ICD-10コードF32.8、F32.89の3つのソースコードが使用されなくなったことによるものであることが分かります。これはICD-10コードが使用され始めた時期と一致していますが、ICD-10コードの合計の有病率はICD-9コードの有病率よりもはるかに低くなっています。この特定の例は、ICD-10コードのF32.9（「大うつ病性障害、単一エピソード、特定不能」）もまた、このコンセプト

前述の例では、マッピングされていないソースコードが偶然発見されたことを示しています。RパッケージのfindOrphanSourceCodes関数を使用する方法があります。この関数を使用すると、得られた結果として得られたソースコードのセットは、次に、手元のCDMデータベースに表示されることがあります。例えば、研究では「壊疽性疾患」（439928）という概念と、その子孫すべてを使用して15.3に示されています。ICD-10のJ85.0（「肺壊疽および壊死」）は、4324261（「肺

15.4 ACHILLES

ここでは、CDM形式のデータベースに対してACHILLESを実行する方法を説明します。

¹<https://github.com/OHDSI/Vocabulary-v5.0/issues>

| % per month | Max monthly % | Person count | Description |
|-------------|---------------|--------------|---|
| | 26.81 | 92,019,885 | Depressive Disorder |
| | 6.64 | 15,969,198 | Depressive disorder 440383 |
| | 6.64 | 15,686,275 | 311 (ICD9CM) Depressive disorder, not elsewhere classified |
| | 0.46 | 188,230 | F328 (ICD10CM) Other depressive episodes |
| | 0.38 | 94,693 | F3289 (ICD10CM) Other specified depressive episodes |
| | 3.10 | 12,010,783 | Adjustment disorder with mixed emotional features 433454 |
| | 3.07 | 9,839,712 | 30928 (ICD9CM) Adjustment disorder with mixed anxiety and depressed mood |
| | 3.03 | 2,049,618 | F4323 (ICD10CM) Adjustment disorder with mixed anxiety and depressed mood |
| | 0.04 | 121,453 | 3091 (ICD9CM) Prolonged depressive reaction |
| | 3.17 | 9,237,192 | Dysthymia 433440 |

Figure 15.2: checkCohortSourceCodes関数のサンプル出力

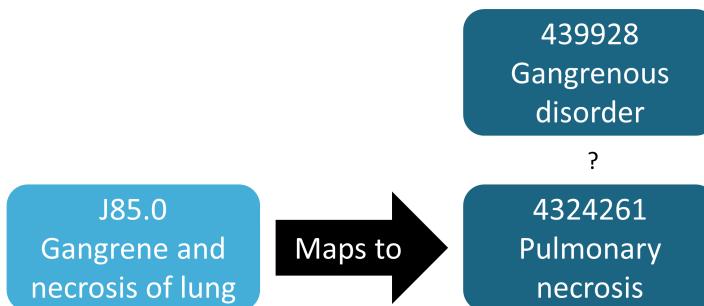


Figure 15.3: 孤立したソースコードのサンプル

まず、Rにサーバーへの接続方法を指示する必要があります。ACHILLESは、DatabaseConnectionDetailsオブジェクトを用いて接続情報を定義します。

```
library(Achilles)
connDetails <- createConnectionDetails(dbms = "postgresql",
                                         server = "localhost/ohdsi",
                                         user = "joe",
                                         password = "supersecret")

cdmDbSchema <- "my_cdm_data"
cdmVersion <- "5.3.0"
```

最後の2行では、cdmDbSchema変数とCDMのバージョンを定義しています。これらは、CDM形式のデータがどこに存在し、どのバージョンのCDMが使用されているかを示します。SQL Serverでは、データベーススキーマではデータベースとスキーマの両方を指定する必要があります。
`<- my_cdm_data.dbo`となります。

次に、ACHILLESを実行します：

```
result <- achilles(connectionDetails,
                     cdmDatabaseSchema = cdmDbSchema,
                     resultsDatabaseSchema = cdmDbSchema,
                     sourceName = "My database",
                     cdmVersion = cdmVersion)
```

この関数は、resultsDatabaseSchema内に複数のテーブルを作成します。ここでは、CDMの評価結果を表示します。これは、ATLASをACHILLESで実行する場合と同様です。

```
exportToJson(connectionDetails,
             cdmDatabaseSchema = cdmDbSchema,
             resultsDatabaseSchema = cdmDbSchema,
             outputPath = "achillesOut")
```

JSONファイルはachillesOutサブフォルダに書き込まれ、AchillesWebウェブアプリケーションで表示されます。このプロットは、データの大部分が正常な範囲内であることを示しています。

別の例を図15.5に示します。これは、糖尿病の診断コードの有病率に急激な変化が生じたことを示すプロットです。

15.5 Data Quality Dashboard

ここでは、CDM形式のデータベースに対してデータ品質ダッシュボードを実行する方法について説明します。これを行うには、セクション15.4で説明されているCDM接続に対して、一連のチェックを行います。現時点では、DQDはCDM v5.3.1のみをサポートしているため、接続する前にデータベース名を指定する必要があります。ACHILLESの場合と同様に、Rにデータの検索先を指示するために、変数

cdmDbSchema

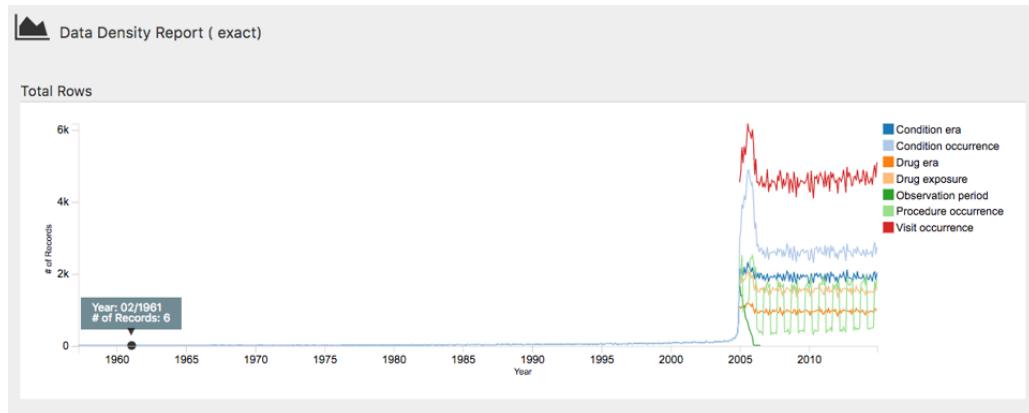


Figure 15.4: ACHILLES ウエブビューウーでのデータ密度プロット

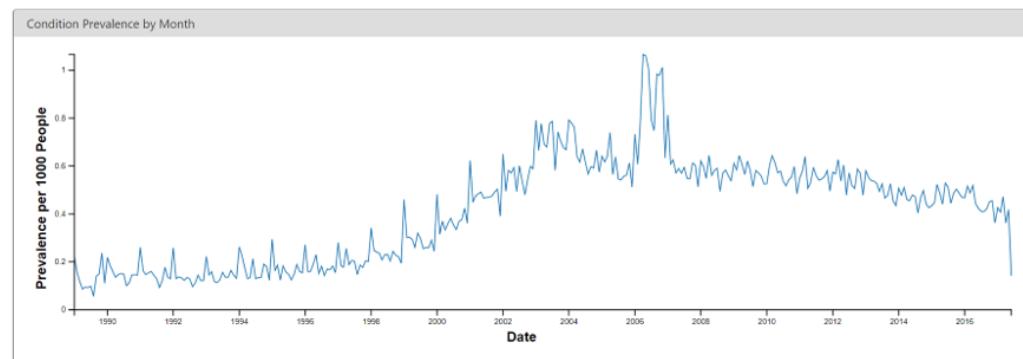


Figure 15.5: ACHILLES ウエブビューウーでの月次糖尿病の診断コードの有病率

を作成する必要があります。

```
cdmDbSchema <- "my_cdm_data.dbo"
```

次に、Dashboardを実行します…

```
DataQualityDashboard::executeDqChecks(connectionDetails = connectionDetails,  
                                         cdmDatabaseSchema = cdmDbSchema,  
                                         resultsDatabaseSchema = cdmDbSchema,  
                                         cdmSourceName = "My database",  
                                         outputFolder = "My output")
```

上記の関数は、指定されたスキーマ上で利用可能なすべてのデータ品質チェックを実行します。

```
viewDqDashboard(jsonPath)
```

変数(jsonPath)は、上記のexecuteDqChecks関数を呼び出す際に指定したoutputFolderにあります。

最初にダッシュボードを開くと、図15.6に示すような概要テーブルが表示されます。この

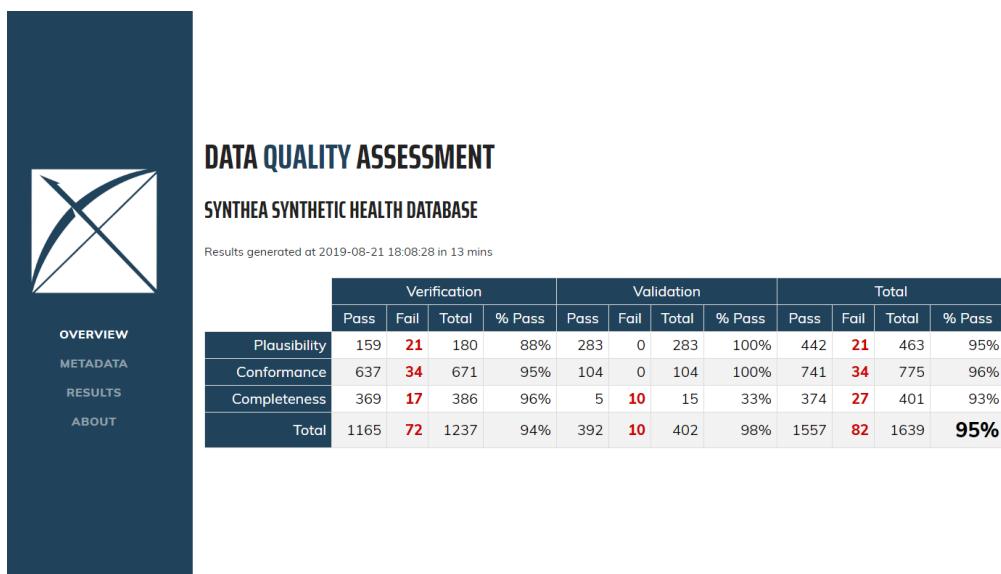


Figure 15.6: Data Quality Dashboardにおけるデータ品質チェックの概要

左側のメニューでResults（結果）をクリックすると、実行された各チェックのドリルダウン（図15.7）。この例では、個々のCDMテーブルの完全性、すなわち、指定されたテーブルに

RESULTS

SYNTHEA SYNTHETIC HEALTH DATABASE

Results generated at 2019-08-22 14:15:06 in 29 mins

| | STATUS | CONTEXT | CATEGORY | SUBCATEGORY | LEVEL | DESCRIPTION | % RECORDS |
|-----|--------|--------------|--------------|-------------|-------|---|-----------|
| | Show | 5 entries | | | FIELD | Column visibility | CSV |
| [+] | FAIL | Verification | Plausibility | Atemporal | FIELD | The number and percent of records with a value in the gap_days field of the DRUG_ERA table less than 0. (Threshold=0%). | 24.07% |
| [+] | FAIL | Verification | Completeness | None | FIELD | The number and percent of records with a value of 0 in the standard concept field race_concept_id in the PERSON table. (Threshold=0%). | 16.74% |
| [+] | FAIL | Verification | Conformance | Relational | FIELD | The number and percent of records that have a value in the ethnicity_concept_id field in the PERSON table that does not exist in the CONCEPT table. (Threshold=0%). | 16.15% |
| [+] | PASS | Verification | Completeness | None | FIELD | The number and percent of records with a NULL value in the condition_end_date of the CONDITION_OCCURRENCE. (Threshold=100%). | 13.24% |
| [+] | PASS | Verification | Completeness | None | FIELD | The number and percent of records with a NULL value in the condition_end_datetime of the CONDITION_OCCURRENCE. (Threshold=100%). | 13.24% |

Showing 71 to 75 of 1,327 entries (filtered from 1,639 total entries)

Previous 1 ... 14 15 16 ... 266 Next

Figure 15.7: Data Quality Dashboardにおけるデータ品質チェックの詳細

15.6

次に、付録 B.4 で提供されている血管性浮腫コホート定義に特化したいくつかのチェックを実行します。15.4 で説明されているように設定済みであり、コホート定義のJSONとSQLはそれぞれ「cohort.json」と「cohort.sql」で定義されています。

```
library(MethodEvaluation)
json <- readChar("cohort.json", file.info("cohort.json")$size)
sql <- readChar("cohort.sql", file.info("cohort.sql")$size)
checkCohortSourceCodes(connectionDetails,
                        cdmDatabaseSchema = cdmDbSchema,
                        cohortJson = json,
                        cohortSql = sql,
                        outputFile = "output.html")
```

出力ファイルをウェブブラウザで開くことができます（図 15.8）。ここでは、血管性浮腫の二つのICD-9コードと二つのICD-10コードによって見つかります。それぞれのコードのスパークラインを確認できます。

次に、標準コンセプトコードにマッピングされていない孤立したソースコードを検索できます。

```
orphans <- findOrphanSourceCodes(connectionDetails,
                                    cdmDatabaseSchema = cdmDbSchema,
                                    conceptName = "Angioedema",
                                    conceptSynonyms = c("Angioneurotic edema",
```

| % per month | Max monthly % | Person count | Description |
|-------------|---------------|--------------|---|
| | 60.60 | 24,189,656 | Inpatient or ER visit |
| | 39.50 | 15,003,249 | Emergency Room Visit 9203 |
| | 39.50 | 15,003,249 | ER (None) No matching concept |
| | 23.90 | 9,186,407 | Inpatient Visit 9201 |
| | 23.90 | 9,186,407 | IP (None) No matching concept |
| | 0.27 | 76,711 | Angioedema |
| | 0.27 | 76,711 | Angioedema 432791 |
| | 0.26 | 64,726 | 9951 (ICD9CM) Angioneurotic edema, not elsewhere classified |
| | 0.20 | 8,822 | T783XXA (ICD10CM) Angioneurotic edema, initial encounter |
| | 0.09 | 3,163 | T783XXD (ICD10CM) Angioneurotic edema, subsequent encounter |

Figure 15.8: 血管性浮腫のコホート定義で使用されるソースコード

```

"Giant hives",
"Giant urticaria",
"Periodic edema"))
View(orphans)

```

| コード | 説明 | 語彙ID | 全体のカウント |
|-----------|---|---------|---------|
| T78.3XXS | Angioneurotic edema, sequela | ICD10CM | 508 |
| 10002425 | Angioedemas | MedDRA | 0 |
| 148774 | Angioneurotic Edema of Larynx | CIEL | 0 |
| 402383003 | Idiopathic urticaria and/or angioedema | SNOMED | 0 |
| 232437009 | Angioneurotic edema of larynx | SNOMED | 0 |
| 10002472 | Angioneurotic edema, not elsewhere classified | MedDRA | 0 |

データで実際に使用されている潜在的な孤児コード（祖父も子孫もない）として見つかっ

15.7



- ほとんどの観察医療データは研究のために収集されたものではありません。
- データの品質チェックは研究に不可欠な要素です。データが研究目的に十分な品質であることを確認する必要があります。
- 一般に研究目的でデータの品質を評価すべきであり、特定の研究においては特に慎重に評価する必要があります。
- データ品質の一部は、Data Quality Dashboardのような大規模な事前に定義された評価基準に基づいています。
- 特定の研究に関連するコードのマッピングを評価するための他のツールも存在します。

15.8

前提条件

これらの演習では、セクション 8.4.5で説明されているように、R、R-Studio、およびJavaがインストールされている必要があります。

```
install.packages(c("SqlRender", "DatabaseConnector", "remotes"))
remotes::install_github("ohdsi/Achilles")
remotes::install_github("ohdsi/DataQualityDashboard")
remotes::install_github("ohdsi/Eunomia", ref = "v1.0.0")
```

Eunomiaパッケージは、ローカルのRセッション内で実行されるCDMのシミュレーションデータベースです。

```
connectionDetails <- Eunomia::getEunomiaConnectionDetails()
```

CDMデータベーススキーマは「main」です。

演習 15.1. Eunomiaデータベースに対してACHILLESを実行してください。

演習 15.2. Eunomiaデータベースに対してData Quality Dashboardを実行してください。

演習 15.3. DQDのチェックリストを抽出してください。

回答例は付録 E.10にあります。

Chapter 16

著者: Joel Swerdel, Seng Chan You, Ray Chen & Patrick Ryan

物質をエネルギーに変える可能性は、鳥がほとんどいない国で暗闇の中で鳥を撃つ
OHDSIのビジョンは、「観察研究によって健康と疾病に関する包括的な理解が得られる
14章で説明したように、妥当性のさまざまな側面に対する脅威に満ちている可能性があ
12章)。「ACE阻害薬は、サイアザイド系またはサイアザイド系利尿薬と比較して血管

16.1

私たちが発見したのは、ACE阻害薬の処方と血管浮腫の関係であり、ACE阻害薬の使用と
15章)で議論しました。Common Data Model (CDM)に変換されたデータベースの質
それらのデータは患者の完全な病歴を表しているとは限らず、また、複数の医療システム
観察データから信頼性の高いエビデンスを生成するには、患者が治療を求めた瞬間から、

16.2

Hripcsak and Albers (2017)は、「表現型とは、生物の遺伝的構成から導かれる遺伝子型
この説明は、臨床的妥当性を検討する際に強化すべきいくつかの属性を強調しています。
OHDSIでは、一定期間にわたって1つ以上の包含基準を満たす人々の集合を定義するため
臨床的特性、集団レベルの影響の推定、患者レベルの予測など、ほとんどの観察分析では
集団レベルの推定の例(第12章)「ACE阻害薬は、サイアザイド系またはサイアザイド
薬剤誘発の可能性がある血管性浮腫事象を、食物アレルギーやウイルス感染など、他の原
曝露状況と結果の発生率との間に時間的な関連性を導くことに自信が持てるほど、疾患の
本章では、コホート定義の妥当性を検証する方法について説明します。まず、コホート定

16.2.1

対象コホートの定義が確定すると、その妥当性を評価することができます。妥当性を評価する16.1は、混同行列の要素を示しています。

| | | Gold Standard | |
|-------------------|-------|----------------|----------------|
| | | True | False |
| Cohort Definition | True | True Positive | False Positive |
| | False | False Negative | True Negative |

Figure 16.1: コンフュージョンマトリックス

コホート定義による真の結果と偽の結果は、その定義がある集団に適用することで決定されま表現型指定の二値表示におけるエラーに加えて、健康状態のタイミングも不正確である可能性次のステップは、ゴールドスタンダードとコホート定義の一一致度を評価することです。ゴール

1. コホート定義の感度 - 集団内の表現型に真に属する人のうち、コホート定義に基づいて健
 感度 = 真陽性 / (真陽性 + 偽陰性)
2. コホート定義の特異度 - 集団内の表現型に属さない人のうち、コホート定義に基づいて健
 特異度 = 真の陰性数 / (真の陰性数 + 偽陽性数)
3. コホート定義の陽性的中率 (PPV) - コホート定義によって健康状態にあると特定された
 PPV = 真の陽性数 / (真の陽性数 + 偽陽性数)
4. コホート定義の陰性的中率 (NPV) - コホート定義によって特定された健康状態ではない
 NPV = 真陰性 / (真陰性 + 偽陰性)

これらの指標の満点は100%です。観察データの性質上、満点は通常、標準からかけ離れた値 Rubbo et al. (2015) は、心筋梗塞のコホート定義を検証した研究をレビューしました。彼らがコホート定義のパフォーマンス指標が確立された後は、これらの定義を使用する研究の結果を

16.3

コホートの定義を検証するために一般に用いられる方法は、ソースレコードの検証による臨床

1. カルテレビューを含む調査を実施するため、必要に応じて現地のIRB (Institutional Review Board) および／または関係者から許諾を得る。
2. 評価対象のコホートの定義を用いてコホートを生成する。コホート全体を審査するのに一
3. 対象者の記録を審査するのに十分な臨床的専門知識を有する1人または複数の人物を特定
4. 対象者が対象とする臨床状態または特性について陽性または陰性であるかを判定するた
5. 臨床専門家がサンプル内的人物について、利用可能なすべてのデータを検証および判定し

6. コホートの定義分類や臨床判定分類に従って対象者を混同行列に分類し、収集したチャートレビューの結果は、通常、1つの性能特性である陽性的中率（PPV）の評価に限る。チャートレビューのステップ3から6を繰り返して、これらの患者が本当に臨床的に関心のあるソース記録の検証による臨床判定には、多くの限界があります。前述の通り、PPVのよう

16.3.1

コロンビア大学アーヴィング医療センター（CUIMC）による研究では、米国国立がん研究

1. OHDSI がんフェノタイプ研究のための提案を提出し、IRB の承認を取得しました。
2. 前立腺がんの集団定義を開発：ボキャブラリを調査するために ATHENA と ATLAS を使用し、前立腺悪性腫瘍（概念 ID 4163261）の発生状態の患者をすべて（概念 ID 4314337）または前立腺非ホジキンリンパ腫（概念 ID 4048666）を除く集団定義を作成しました。
3. ATLAS¹ を使用して生成されたコホートから、手動レビュー用に 100 人の患者を無作為に抽出し、マッピングテーブルを使用して各 PERSON_ID を患者 MRN にマッピングしました。100 人の患者は、PPV のパフォーマンス指標について、望ましいレベルの統計的精度を達成するよう抽出されました。
4. 無作為に抽出されたサブセット内の人々が真陽性か偽陽性かを判断するために、入院 EHR の記録を手動で確認しました。
5. 手動レビューと臨床判定は1人の医師によって実施されました(ただし、将来、理想では複数の医師によって実施される予定です)。
6. 参照基準の決定は、入手可能な電子的な患者記録のすべてに記録されている臨床記録を用いて行われました。
7. 患者は、1) 前立腺がん、2) 前立腺がんではない、3) 判断不能、のいずれかに分類されました。
8. 前立腺がん/（前立腺がんではない+判断不能）という計算式で、PPVの控えめな推定を行いました。
9. 次に、腫瘍登録を追加のゴールドスタンダードとして使用し、CUIMC 全体の集団における基準標準を特定しました。腫瘍レジストリにおいて、コホートの特異度を評価しました。
10. 推定された感度、陽性適中率、および有病率を用いて、このコホート定義の特異度を評価しました。

Rubbo et al. (2015) らによる心筋梗塞（MI）コホート定義の妥当性評価のレビューでは、

16.4 PheEvaluator

OHDSIコミュニティは、診断予測モデルを用いてゴールドスタンダードを構築する別のアプローチを提供しています（Swerdel et al., 2019）。一般的な考え方では、臨床医がソースレコードの検証で実施するのではなく、Level Predictionパッケージの機能を使用しています。

プロセスは以下の通り：

1. 極めて特異的な（「xSpec」）コホートを作成する：診断予測モデルのトレーニング

¹<https://github.com/OHDSI/PheEvaluator>

2. 極めて感度の高い（「xSens」）コホートを作成する：結果が得られる可能性のある人を対象にします。
3. xSpecとxSensコホートを使用して予測モデルを適合：第13で説明したように、幅広い年齢層と性別の人々を対象にします。
4. コホート定義の性能を評価するために使用される、除外された人々のセットに対して、結果を評価します。
5. コホート定義の性能特性を評価します：予測確率をコホート定義の二値分類と比較します。

このアプローチを使用する際の主な限界は、健康アウトカムのある人の確率の推定がデータベースによって異なることです。診断予測モデルでは、疾患を持つ人と持たない人を識別するモデルを作成します。患者レコード（第13章）で説明されているように、予測モデルは対象コホートと結果コホートを使用して開発されます。プロセスでは、予測モデルのアウトカムコホートを決定するために、非常に特異的なコホートを定義します。コホートは、定義を使用して、対象疾患の罹患確率が極めて高い人を見つけ出します。xSpecでは、対象の健康アウトカムについて複数の条件発生記録を持つ人々として定義される（Suchard et al., 2013）。このアルゴリズムは簡潔なモデルを生成し、通常、データセット全体

16.4.1 PheEvaluator

PheEvaluatorを使用して、急性心筋梗塞を患ったことがある人を特定する必要がある研究で使われる手順は以下の通りです。

ステップ 1: xSpec コホートの定義

MIの可能性が高いものを特定します。心筋梗塞またはその子孫のコンセプトを持つ条件発生記録（第16.2）は、ATLASにおけるMIのこのコホート定義を示しています。

ステップ 2: xSens コホートの定義

次に、極めて感度の高いコホート（xSens）を開発します。このコホートは、MIについては、ATLASにおけるMIのxSensコホート定義を示しています。

ステップ 3: 予測モデルの適合

関数createPhenoModelは、評価コホートにおいて対象の健康アウトカムとなる確率を評価するためのモデルを構築します。コホートは、モデリングプロセスで使用されるターゲットコホートから除外すべきであることを示します。

xSpec コホートを定義するために使用されたすべてのコンセプトは、モデリングプロセスから除外されます。モデルに含める性別を指定することもできます。デフォルトでは、パラメータの値は次の例を参照してください：

```
setwd("c:/temp")
library(PheEvaluator)
connectionDetails <- createConnectionDetails(
  dbms = "postgresql",
```

Cohort #10934

MI xSpec Cohort

Definition Concept Sets Generation Reporting Export

[460] MI xSpec Model

Cohort Entry Events

Events having any of the following criteria:

+ Add Initial Event

a condition occurrence of [460] Myocardial Infarction

+ Add attribute...

with continuous observation of at least 365 days before and 0 days after event index date

Limit initial events to: earliest event per person.

Restrict initial events to:

having all of the following criteria:

+ Add criteria to group...

with at least 1 using all occurrences of:

a condition occurrence of [460] Myocardial Infarction

+ Add attribute...

with a Visit occurrence of: Inpatient Visit

where event starts between 0 days Before and 5 days After

restrict to the same visit occurrence
 allow events from outside observation period

and with at least 4 using all occurrences of:

a condition occurrence of [460] Myocardial Infarction

+ Add attribute...

where event starts between 1 days After and 365 days After

restrict to the same visit occurrence
 allow events from outside observation period

Limit initial events to: earliest event per person.

Remove initial event restriction

Figure 16.2: 心筋梗塞の極めて特異的なコホート定義 (xSpec)

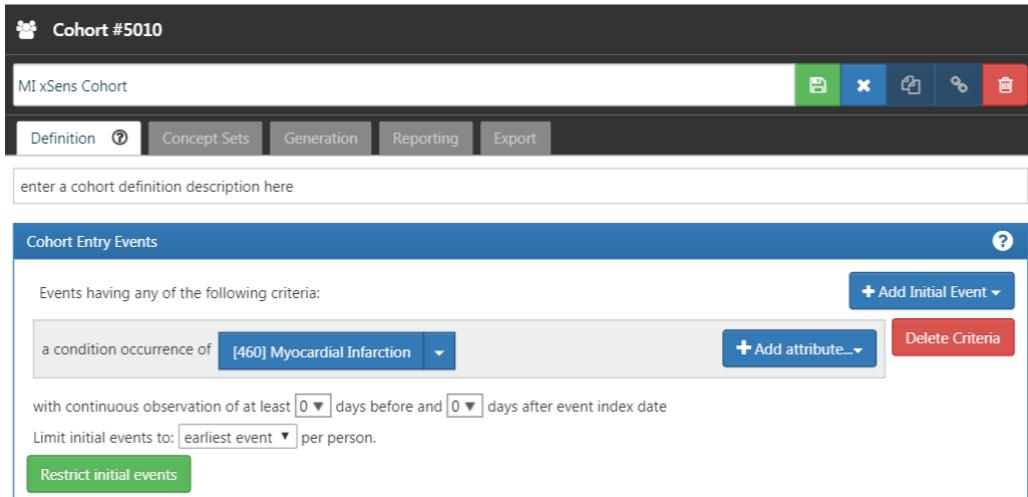


Figure 16.3: 心筋梗塞の極度に感度の高いコホート定義 (xSens)

```
server = "localhost/ohdsi",
user = "joe",
password = "supersecret")

phenoTest <- createPhenoModel(
  connectionDetails = connectionDetails,
  xSpecCohort = 10934,
  cdmDatabaseSchema = "my_cdm_data",
  cohortDatabaseSchema = "my_results",
  cohortDatabaseTable = "cohort",
  outDatabaseSchema = "scratch.dbo", #
  trainOutFile = "5XMI_train",
  exclCohort = 1770120, #xSens
  prevCohort = 1770119, #
  modelAnalysisId = "20181206V1",
  excludedConcepts = c(312327, 314666),
  addDescendantsToExclude = TRUE,
  cdmShortName = "myCDM",
  mainPopnCohort = 0, #
  lowerAgeLimit = 18,
  upperAgeLimit = 90,
  gender = c(8507, 8532),
  startDate = "20100101",
  endDate = "20171231")
```

この例では、「my_results」データベースで定義されたコホートを使用し、コホートテーブル - 「my_results.cohort」）と、モデルに条件、薬物曝露などを知らせる場所（cdmDatabaseSchema - 「my_cdm_data」）を指定しました。モデルに含まれる対象者は、CDMにおける初回来院日 312327、314666、およびそれらの子孫は、除外しています。これらの初回来院時の年齢は18歳未満とされています。

ステップ 4: 評価コホートの作成

関数createEvalCohortは、パッケージ関数applyModelを使用して、対象とする健康アウトリリース（lowerAgeLimitおよびupperAgeLimit引数として年齢を設定）、性別（genderパラメータ）を指定します。例えば：

```
setwd("c:/temp")
connectionDetails <- createConnectionDetails(
  dbms = "postgresql",
  server = "localhost/ohdsi",
  user = "joe",
  password = "supersecret")

evalCohort <- createEvalCohort(
  connectionDetails = connectionDetails,
  xSpecCohort = 10934,
  cdmDatabaseSchema = "my_cdm_data",
  cohortDatabaseSchema = "my_results",
  cohortDatabaseTable = "cohort",
  outDatabaseSchema = "scratch.dbo",
  testOutFile = "5XMI_eval",
  trainOutFile = "5XMI_train",
  modelAnalysisId = "20181206V1",
  evalAnalysisId = "20181206V1",
  cdmShortName = "myCDM",
  mainPopnCohort = 0,
  lowerAgeLimit = 18,
  upperAgeLimit = 90,
  gender = c(8507, 8532),
  startDate = "20100101",
  endDate = "20171231")
```

この例では、パラメータにより、関数がモデルファイル「c:/temp/lr_results_5XMI_train」を作成します。

ステップ 5: コホート定義の作成とテスト

次のステップは、評価対象のコホート定義を作成し、テストすることです。望ましい性能特性は、対象とする研究課題に対応するためのコホートの使用目的によって異なります。特定の研究課題には非常に感度の高いアルゴリズムが必要となる場合もありますが、より効率的なアルゴリズムを使用する場合は、PheEvaluatorを使用してコホート定義の性能特性を決定するプロセスを図16.4に示します。

図16.4のパートAでは、私たちは、テスト対象となるコホート定義の人物を調査し、コホート（人物ID 016、019、022、023、025）と、含まれない評価コホートの人物（人物ID 017、018、020、021、024）を見つけました。これらの対象者/非対象者それぞれについて、

真陽性、真陰性、偽陽性、偽陰性の値は、以下のように推定しました（図16.4のパートB）：

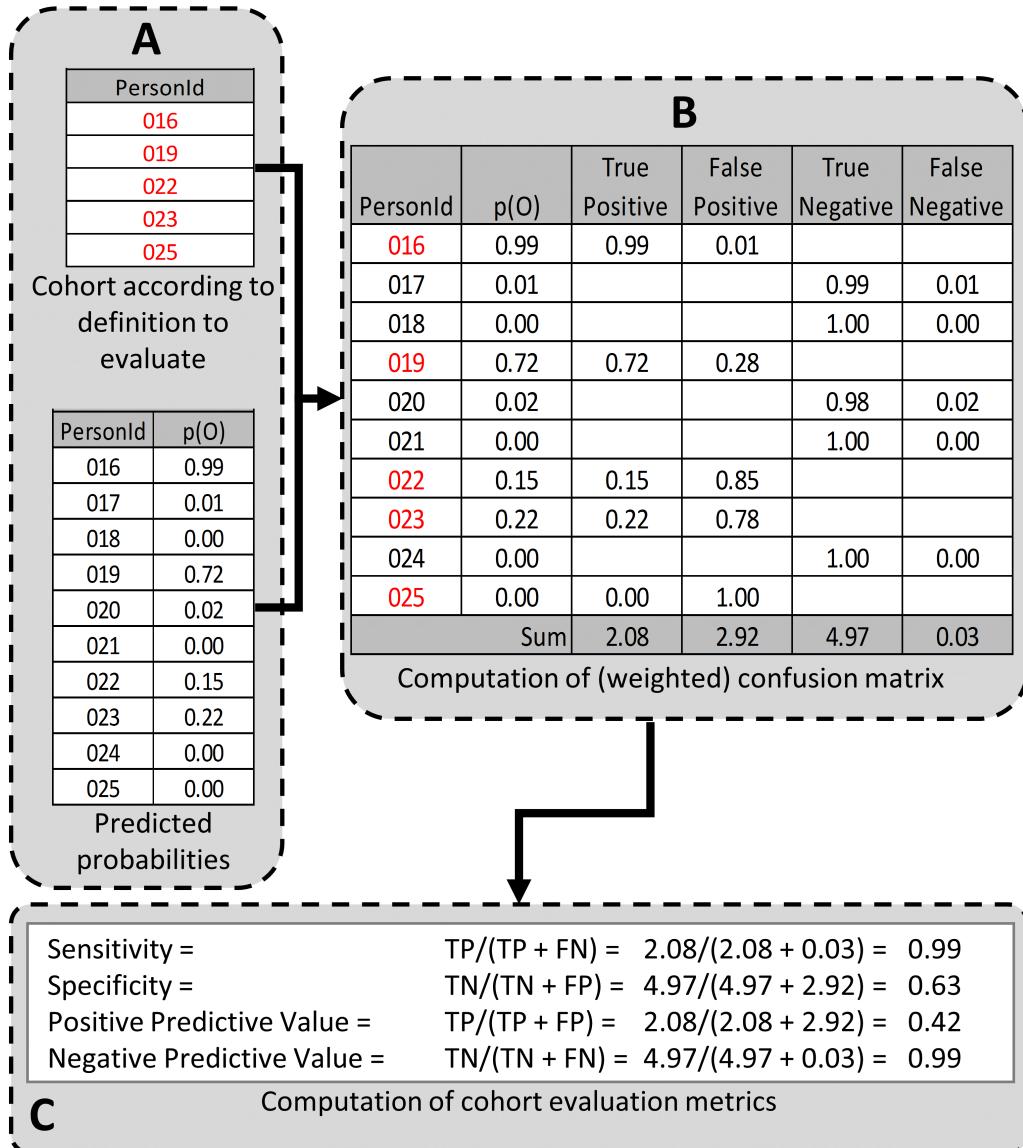


Figure 16.4: PheEvaluatorを使用したコホート定義の性能特性の決定 $p(O)$ = 結果の確率; TP = 真陽性; FN = 偽陰性; TN = 真陰性; FP = 偽陽性

1. コホート定義に評価コホートに属する人物が含まれていた場合、すなわち、コホート 016 の健康結果の存在の予測確率は 99% であり、0.99 が真陽性（カウントの期待値 0.99 を追加）に追加され、 $1.00 - 0.99 = 0.01$ が偽陽性（0.01 の期待値）に追加されました。この処理は、コホート定義に含まれる評価コホート 019、022、023、および 025）に対して繰り返されました。
2. 同様に、コホート定義が評価コホートに属する人物を含んでいなかった場合、すなわち、コホート 017 の健康アウトカムの存在に対する予測確率は 1%（および、対応する健康アウトカム 99%）であり、 $1.00 - 0.01 = 0.99$ が真陰性に、0.01 が偽陰性に追加されました。018、020、021、024）に対して繰り返しました。

これらの値を評価コホート内の全対象者について加算した後、4つのセルに各セルの期待値（「EV」）を割り当てます（図 16.4 のパート C）。これらの期待セルカウントは、推定値の分散を評価するために使用する

コホート定義の性能特性を決定するには、関数 `testPhenotype` を使用します。この関数は、関数から出力された RDS ファイル 「c:/temp/lr_results_5XMI_eval_myCDM_ePPV0.75.rds」 の ID に設定します。 `phenText` パラメータを、コホート定義として人が読める説明文で設定し、「All MI by Phenotype 1 X In-patient, 1st Position」 などです。このステップの出力は、テストされたコホート定義の性能特性を含むデータフレームです。

`cutPoints`

パラメータの設定は、性能特性の結果を導き出すために使用される値のリストです。性能特性

`cutPoints`

パラメータのリストに「EV」を含めます。また、特定の予測確率、すなわちカットポイント

```
setwd("c:/temp")
connectionDetails <- createConnectionDetails(
  dbms = "postgresql",
  server = "localhost/ohdsi",
  user = "joe",
  password = "supersecret")

phenoResult <- testPhenotype(
  connectionDetails = connectionDetails,
  cutPoints = c(0.1, 0.2, 0.3, 0.4, 0.5, "EV", 0.6, 0.7, 0.8, 0.9),
  resultsFileName =
    "c:/temp/lr_results_5XMI_eval_myCDM_ePPV0.75_20181206V1.rds",
  modelFileName =
    "c:/temp/lr_results_5XMI_train_myCDM_ePPV0.75_20181206V1.rds",
  cohortPheno = 1769702,
  phenText = "All MI by Phenotype 1 X In-patient, 1st Position",
  order = 1,
  testText = "MI xSpec Model - 5 X MI",
  cohortDatabaseSchema = "my_results",
  cohortTable = "cohort",
  cdmShortName = "myCDM")
```

この例では、予測閾値の幅広い範囲（`cutPoints`）が提供されており、期待値（「EV」）

このプロセスを使用すると、表 16.1 は、5つのデータセットにおけるMIの4つのコホート定義「X HOI, In-Patient」では、平均PPVは67%（範囲：59%～74%）であることが分かりました。

Table 16.1: pheEvaluator を使用して複数のデータセット上で心筋梗塞を診断するためのパラメータ。Sens – 感度； PPV – 陽性適中率； Spec – 特異度； NPV – 陰性適中率； Dx Code – コホートの診断コード。

| Phenotype Algorithm | Database | Sens | PPV | Spec | NPV |
|-----------------------------------|-----------|-------|-------|-------|-------|
| >=1 X HOI | CCAE | 0.761 | 0.598 | 0.997 | 0.999 |
| | Optum1862 | 0.723 | 0.530 | 0.995 | 0.998 |
| | OptumGE60 | 0.643 | 0.534 | 0.973 | 0.982 |
| | MDCD | 0.676 | 0.468 | 0.990 | 0.996 |
| | MDCR | 0.665 | 0.553 | 0.977 | 0.985 |
| >= 2 X HOI | CCAE | 0.585 | 0.769 | 0.999 | 0.998 |
| | Optum1862 | 0.495 | 0.693 | 0.998 | 0.996 |
| | OptumGE60 | 0.382 | 0.644 | 0.990 | 0.971 |
| | MDCD | 0.454 | 0.628 | 0.996 | 0.993 |
| | MDCR | 0.418 | 0.674 | 0.991 | 0.975 |
| >=1 X HOI, In-Patient | CCAE | 0.674 | 0.737 | 0.999 | 0.998 |
| | Optum1862 | 0.623 | 0.693 | 0.998 | 0.997 |
| | OptumGE60 | 0.521 | 0.655 | 0.987 | 0.977 |
| | MDCD | 0.573 | 0.593 | 0.995 | 0.994 |
| | MDCR | 0.544 | 0.649 | 0.987 | 0.980 |
| 1 X HOI, In-Patient, 1st Position | CCAE | 0.633 | 0.788 | 0.999 | 0.998 |
| | Optum1862 | 0.581 | 0.754 | 0.999 | 0.997 |
| | OptumGE60 | 0.445 | 0.711 | 0.991 | 0.974 |
| | MDCD | 0.499 | 0.666 | 0.997 | 0.993 |
| | MDCR | 0.445 | 0.711 | 0.991 | 0.974 |

16.5

コホートは、特定の観察データベースの文脈内で明確に定義され、十分に評価される可能性がある。Madigan et al. (2013b) は、データベースの選択が観察研究の結果に影響を与えることを実証した。OHDSIネットワーク全体を見ると、観察データベースは、対象とする集団（例えば、小児対高齢者）

16.6



- 臨床的妥当性は、基礎となるデータソースの特性を理解し、分析内のコホートの性能を評価する。
- コホートの定義は、コホートの定義と利用可能な観察データに基づいてコホート内での実現可能性を検証する。
- コホート定義の検証には、感度、特異度、陽性適中率など、複数の性能特性を推定する。

- ソースレコードの検証とPheEvaluatorによる臨床判定は、コホート定義の検証
- OHDSIネットワーク研究は、データソースの異質性を調査し、実証結果の一般

Chapter 17

著者: Martijn Schuemie

ソフトウェアの妥当性に関する中心的な問題は

ソフトウェアが期待通りに動作しているか？

ソフトウェアの妥当性は、エビデンスの質にとって不可欠な要素です。つまり、私たちの17.1.1で説明されているように、すべての研究をソフトウェア開発の演習と見なすことか8.1で説明されているように、分析全体をカスタムコードとして記述することも、OHDSI Methods Libraryで利用可能な機能を用いることもできます。Methods Libraryを使用する利点は、その妥当性を確保するためにすでに多大な注意が払われていること

この章では、まず有効な分析コードの記述に関するベストプラクティスについて説明します。Libraryがどのように検証されているかについて説明していきます。

17.1

17.1.1

従来、観察研究はプロセスというよりも旅路として捉えられることがよくありました。ただし、エビデンスを生み出す分析はすべて完全に自動化されなければなりません。分析スクリプトは、どのようなコンピュータ言語でも実装できますが、OHDSIではR言語 Methods Libraryの他のRパッケージを通じて、多くの高度な分析機能を利用できます。

17.1.2

観察分析は、最終結果を得るまでに多くのステップを必要とするため、非常に複雑になる(Martin, 2008)。これらのベストプラクティスについて詳しく説明すると、多くの書籍が

- ・抽象化：すべてを実行する巨大なスクリプトを1つ書くのではなく、コードの行と行
- ・カプセル化：抽象化を機能させるには、関数の依存関係を最小限に抑え、明確に定義

- わかりやすい命名：変数や関数はわかりやすい名前を付けるべきであり、コードはほとんどの場合、
`<- spl(y, 100)` の代わりに、`sampledPatients <- takeSample(patients, sampleSize = 100)` と記述できます。省略したくなる衝動を抑えるようにしてください。
- 再利用：明確で、うまくカプセル化された関数を書くことの利点のひとつは、それらを再利用する能力です。

17.1.3

ソフトウェアコードの妥当性を検証する方法はいくつかありますが、観察研究を実施するコードの検証には、以下の方法が一般的です。

- コードレビュー：1人がコードを書き、別の1人がそのコードをレビューする。
- ダブルコーディング：2人がそれぞれ独立して分析コードを書き、その後、2つのスクリプトを比較する。

コードレビューには通常、作業量が少ないという利点がありますが、欠点としては、レビュアが「exposure end date」を「exposure end」に誤解する（「exposure end date」を含むと解釈すべきか、それともそうでないか）必要なため、2人のレビュアが必ずしも同じ結論を得るとは限らないことがあります。

ユニットテストなどの他のソフトウェア検証手法は、研究が通常、入力（CDMのデータ）と出力（結果）を自動的に比較するため、Methods Libraryで適用されていることに注意してください。

17.1.4 Methods Library

OHDSI Methods Library は、多数の関数を提供しており、ほとんどの観察研究は数行のコードで実現可能になります。OHDSI Methods Library を使用することで、研究コードの妥当性を立証する負担のほとんどが Methods Library に移行されます。Methods Library の妥当性は、そのソフトウェア開発プロセスと広範な検証によって確保されています。

17.2 Methods Library

OHDSI Methods Library は OHDSI コミュニティによって開発されています。Libraryへの変更提案は GitHub の issue tracker¹ と OHDSI フォーラム² の 2 つの場所で議論され、OHDSI フォーラムでは、OHDSI 患者レベルの予測ワークグループのリーダー（Marc Suchard 博士と Martijn Schuemie 博士）および OHDSI 集団レベルの推定ワークグループのリーダー（Peter Rijnbeek 博士と Jenna Reps 博士）のみが行います。

ユーザーは GitHub のリポジトリにあるマスター ブランチから直接、または「drat」と呼ばれる CRAN パッケージをインストールすることができます。Methods Library のパッケージの多くは R の Comprehensive R Archive Network (CRAN) を通じて入手でき、この数は今後さらに増える見込みです。

OHDSI では、Methods Library のパフォーマンスの正確性、信頼性、一貫性を最大限に高めるよう努めています。Methods Library が Apache License V2 の条件に基づいてリリースされているため、Methods Library の基盤となるすべてのソースコード（R、C++、SQL、Java のいずれであっても）は、OHDSI の GitHub リポジトリに公開されています。Methods Library に具現化されたすべての機能は、その正確性、信頼性、一貫性に関して、継続的な評価と監視が行われています。

¹<https://github.com/OHDSI/CohortMethod/issues>

²<http://forums.ohdsi.org/>

17.2.1

Methods Libraryのソースコードはすべて、GitHubを通じて一般公開されているソースコードです。Libraryのリポジトリはアクセス制御されています。世界中の誰もがソースコードを閲覧できます。人口レベル推定作業部会および患者レベル予測作業部会のリーダーシップのみが、こうしたリポジトリ内では、コード変更の継続的なログが管理されており、コードとドキュメント

新しいバージョンは、OHDSI集団レベルの推定ワークグループや患者レベルの予測ワークグループによって生成されます。

- ・新しいマイクロバージョン（例：4.3.2から4.3.3）は、バグ修正のみを示します。
- ・新しいマイナーバージョン（例：4.3.3から4.4.0）は、機能追加を示します。後方互換性を維持するため、既存の機能に影響を与える可能性があります。
- ・新しいメジャーバージョン（例：4.4.0から5.0.0）は、大幅な改訂を示します。互換性が失われる可能性があります。

17.2.2

Methods Library内のすべてのパッケージは、Rの内部ドキュメントフレームワークを通して、Libraryウェブサイトで閲覧できます³。

Methods Libraryのソースコードはすべてエンドユーザーが利用できます。コミュニティ

17.2.3

Methods Libraryパッケージの現在および過去のバージョンは、2つの場所で入手できます。dratリポジトリに保存されています。

17.2.4

Methods Libraryの各最新バージョンは、OHDSIによりバグレポート、修正、パッチに関する情報が含まれます。

17.2.5

OHDSIコミュニティのメンバーは、複数の統計分野を代表しており、複数の地域にまたがるOHDSI集団レベルの推定ワークグループやOHDSI患者レベルの予測ワークグループのす

17.2.6

OHDSI Methods LibraryはGitHub⁴システムでホストされています。GitHubのセキュリティ://github.com/securityを参照してください。OHDSIコミュニティのすべてのメンバーがLibraryに変更を加えるには、ユーザー名とパスワードが必要です。また、マスター

³<https://ohdsi.github.io/MethodsLibrary/>

⁴<https://github.com/>

17.2.7

OHDSI Methods Libraryは GitHub システム上でホストされています。GitHub の災害復旧施設については、<https://github.com/security> に説明があります。

17.3 Methods Library

Methods Libraryで実行されるテストには、パッケージ内の個々の関数に対するテスト（いわゆる単体テスト）が含まれます。

17.3.1

OHDSIにより、ソースコードを既知のデータおよび既知の結果に対してテストできるように、これらのテストは、パッケージに変更が加えられた場合（具体的には、変更がパッケージリポジトリや Methods Library のインストールに関する正確性、信頼性、一貫性に関する追加の文書および客観的な証明）を確認するためです。

17.3.2

より複雑な機能については、入力に対して期待される出力がどのようなものであるべきかが常に明確に定義されています。

17.4



- 再現性と透明性を確保するため、観察研究は CDM のデータから結果まで、分析全体を再現可能であることを確認します。
- カスタムスタディコードは、抽象化、カプセル化、明確な命名、コードの再利用などを実現します。
- カスタムスタディコードは、コードレビューまたはダブルコーディングにより検証されます。
- Methods Library は、観察研究で使用できる検証済みの機能を提供しています。
- Methods Library は、有効なソフトウェアを作成することを目的としたソフトウェア開発ツールです。

Chapter 18

著者: Martijn Schuemie

方法の妥当性を検討する際、次の質問に答えようとなります。

この方法は、この質問に答えるために妥当ですか？

「方法」には研究デザインだけでなく、データやデザインの実施も含まれます。したがって多くの場合、データ品質、臨床的妥当性、ソフトウェアの妥当性が良くなければ、方法の妥当性を検討する前に、エビデンスの質に関して、これらの側面はすでに個別に検討されます。

方法の妥当性を確立するまでの中心的な活動は、分析における重要な仮定が満たされています。

本章では、集団レベルの推定で使用される手法の妥当性に焦点を当てます。まず、研究データの品質とOHDSI Methods BenchmarkとOHDSI Methods Libraryへの応用について、高度なトピックも紹介します。

18.1

各研究デザインには、そのデザインに特有の診断法があります。これらの診断法の多くは、OHDSI Methods LibraryのRパッケージで実装されており、すぐに利用できます。例えば、セクション 12.9 では、CohortMethodパッケージで生成される幅広い診断法がリストされています。

- コホートの初期の比較可能性を評価するための傾向スコア分布。
- モデルから除外すべき潜在変数を特定するための傾向モデル。
- 傾向スコア調整によりコホートが比較可能になったかどうかを評価するための共変量調整。
- さまざまな分析ステップで除外された対象者の数を観察するための脱落。これは、脱落率と脱落原因を示す統計量です。
- 質問に答えるのに十分なデータが利用可能かどうかを評価するための検出力。
- 典型的な発症までの時間を評価し、Cox モデルの基礎となる比例性の仮定が満たされていることを確認するための検定。
- Kaplan-Meier 曲線。

他の研究デザインでは、それらのデザインの異なる仮説を検証するために、異なる診断が必要です。Figure 18.1 に示すプロットを生成することで評価できます。このプロットは、打ち切りとなったものと打ち切りなしで見られる分布の間にわずかな違いしか見られず、私たちの仮説が正しいことが示唆されます。

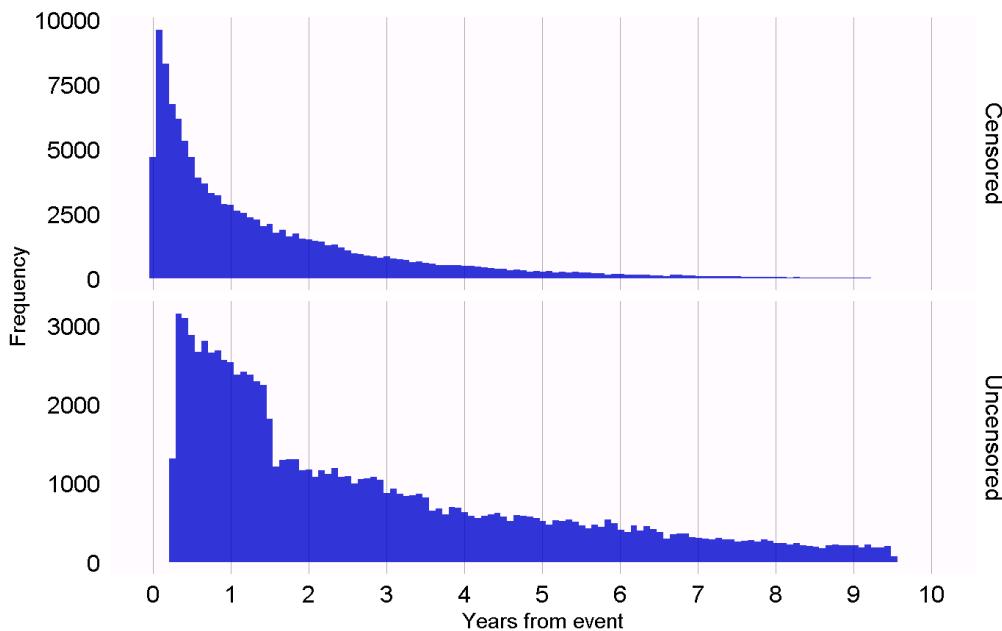


Figure 18.1: 打ち切りありと打ち切りなしとされた対象者の観察終了までの時間。

18.2

デザイン固有の診断に加え、因果効果の推定法全般に適用できる診断もいくつかあります。こ

18.2.1

ネガティブコントロールとは、因果関係が存在しないと考えられる曝露と結果の組み合わせで (Lipsitch et al., 2010)、選択バイアス、測定誤差 (Arnold et al., 2016) を検出する方法として推薦されているネガティブコントロールまたは「偽陰性エンドポイント」 (and Jena, 2013) が含まれます。たとえば、小児期の疾患と後の多発性硬化症 (MS) との関係 (Zaadstra et al., 2008) では、著者は MS の原因とは考えられていない 3 つのネガティブコントロール

私たちは、関心のある仮説と比較可能なネガティブコントロールを選択すべきであり、通常はアウトカムの組み合わせ（いわゆる「アウトカムコントロール」）または同じ結果を持つ曝露アウトカムの組み合わせ（「曝露コントロール」）を選択します。ネガティブコントロールは

- 曝露がアウトカムを引き起こすべきではない。因果関係を考える一つの方法は、仮説を否定する場合には、これは明らかです。例えば、ACE阻害薬は血管性浮腫を引き起こすことが知られています。
- 曝露はアウトカムを予防または治療すべきではありません。これは、真の効果量（例えは

- ・ネガティブコントロールはデータ内に存在すべきであり、理想的には十分な数である。
- ・ネガティブコントロールは理想的には独立しているべきです。例えば、ネガティブコントロールは、ある程度の偏りの可能性があることが理想的です。例

また、ネガティブコントロールは、注目する曝露とアウトカムのペアと同じ交絡構造を持つ (Lipsitch et al., 2010)。しかし、この交絡の構造は不明であると私たちは考えています。

曝露とアウトカムの間に因果関係がないことは、ほとんど文書化されていません。その代わり (Voss et al., 2016)。簡単に説明すると、文献、製品ラベル、および自発報告から得られた

18.2.2

真の相対リスクが1より小さい場合、または1より大きい場合の方法の動作を理解するには (Noren et al., 2014)。

そのため、OHDSIでは合成したポジティブコントロール (Schuemie et al., 2018a) を使用しています。これは、曝露のリスクにさらされる期間中に、アウトカムを

重要な問題として、交絡因子の保存が挙げられます。ネガティブコントロールでは強い交絡 (Suchard et al., 2013) により、予測モデルを適合させます。次に、曝露中のシミュレーション

図 18.2 は、このプロセスを示しています。この手順では、いくつかの重要なバイアスの

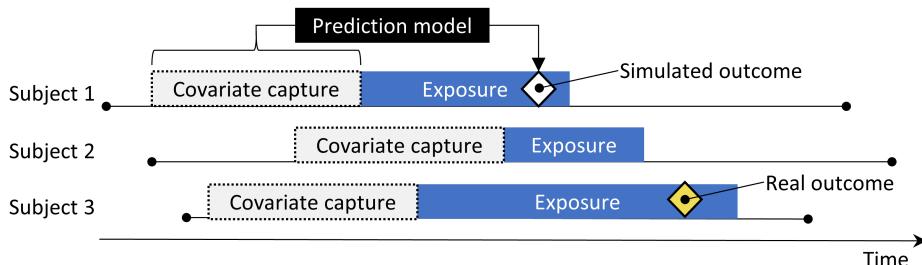


Figure 18.2: ネガティブコントロールからのポジティブコントロールの合成

各コントロールについて、単一の真の「効果の大きさ」を参照していますが、異なる手法 (ATE) にも当てはまります。結果はすべてまれであるため、オッズ比は相対リスクとほぼ同

18.2.3

ネガティブコントロールとポジティブコントロールに対する特定の手法の推定値に基づき

- ・ROC曲線下面積 (AUC) : ポジティブコントロールとネガティブコントロールを識別する能力。
- ・カバー率 : 真の効果量が95%信頼区間に収まる頻度。
- ・平均精度 : 精度は $1/()^2$ として計算され、精度が高いほど信頼区間が狭くなる。
- ・平均二乗誤差 (MSE) : 効果量の推定値の対数と真の効果量の対数との間の平均二乗誤差。

- 第1種の過誤：ネガティブコントロールの場合、帰無仮説が棄却された頻度 ($\alpha = 0.05$)。これは偽陽性率、もしくは「1 - 特異度」と同等です。
- 第2種の過誤：ポジティブコントロールの場合、どのくらいの頻度で帰無仮説が棄却され 0.05 。これは偽陰性率、もしくは「1 - 感度」と同等である。
- 推定なし：推定値を算出できなかったコントロールはいくつあったか？推定値が算出でき ユースケースに応じて、これらの操作特性が目的に適しているかどうかを評価することができます。

18.2.4 P

しばしば、第1種の過誤 ($\alpha = 0.05$) は5%よりも大きくなります。言い換えれば、実際には帰無仮説が棄却される頻度は > 0.05 (Schuemie et al., 2014)。ネガティブコントロールの実際の効果推定値から経験的帰無分布を導く

具体的には、各推定値のサンプリングエラーを考慮して、推定値にガウス確率分布を当てはめます。
 i 番目のネガティブコントロールとアウトカムの組から推定された対数効果推定値（相対リスク）を $\hat{\theta}_i$ で表し、対応する推定標準誤差を $\hat{\tau}_i$ 、 $i = 1, \dots, n$ で表します。 θ_i を真の対数効果量とし（ネガティブコントロールでは0と仮定）、 β_i を対 i に関連する真の（ただし未知の）バイアス、すなわち、コントロールが非常に大きかった場合に i に対して返すであろう推定値の対数と真の効果量の対数の差とします。標準的なp値の計算では、 $\hat{\theta}_i + \beta_i$ を平均とし、 $\hat{\tau}_i^2$ を標準偏差とする正規分布に従うと仮定します。従来のp値の計算では常にゼロと仮定されていましたが、我々は、 μ を平均とし σ^2 を分散とする正規分布から生じる β_i を仮定します。これは、帰無（バイアス）分布を表します。最尤法により μ と σ^2 を推定します。つまり、以下の仮定を置きます。

$$\beta_i \sim N(\mu, \sigma^2) \text{かつ } \hat{\theta}_i \sim N(\theta_i + \beta_i, \tau_i^2)$$

ここで $N(a, b)$ は平均値 a 、分散 b のガウス分布を示し、 μ と σ^2 を次の尤度を最大にすること

$$L(\mu, \sigma | \theta, \tau) \propto \prod_{i=1}^n \int p(\hat{\theta}_i | \beta_i, \theta_i, \hat{\tau}_i) p(\beta_i | \mu, \sigma) d\beta_i$$

ここから最大尤度推定 $\hat{\mu}$ と $\hat{\sigma}$ を得ます。キャリブレートされたp値を実証的な帰無分布を用いたアダプティブペアの効果推定 $\hat{\theta}_{n+1}$ を取り、対応する推定標準誤差 $\hat{\tau}_{n+1}$ を用います。前述の仮定の下で β_{n+1} が同じ帰無分布から発生したとして、次が得られます。

$$\hat{\theta}_{n+1} \sim N(\hat{\mu}, \hat{\sigma} + \hat{\tau}_{n+1})$$

ここでは、 $\hat{\theta}_{n+1}$ は $\hat{\mu}$ より小さく、新しいペアのキャリブレーションされた片側P値は、

$$\phi \left(\frac{\theta_{n+1} - \hat{\mu}}{\sqrt{\hat{\sigma}^2 + \hat{\tau}_{n+1}^2}} \right)$$

ここでは $\phi(\cdot)$ は、標準正規分布の累積分布関数を表します。また、 $\hat{\theta}_{n+1}$ が $\hat{\mu}$ より大きいとき、キャリブレーションされた片側P値は、

$$1 - \phi \left(\frac{\theta_{n+1} - \hat{\mu}}{\sqrt{\hat{\sigma}^2 + \hat{\tau}_{n+1}^2}} \right)$$

18.2.5

同様に、通常、95%信頼区間のカバー率は95%未満であることが観察されます。真の効果 (Schuemie et al., 2018a) では、ポジティブコントロールも活用することで、p値キャリブレーションには、 β_i (i に関連するバイアス) は再びガウス分布から得られると仮定しますが、 θ_i と線形関係にあるものを使用します：

$$\beta_i \sim N(\mu(\theta_i), \sigma^2(\theta_i))$$

ここでは、

$$\begin{aligned}\mu(\theta_i) &= a + b \times \theta_i \text{ かつ} \\ \sigma(\theta_i)^2 &= c + d \times |\theta_i|\end{aligned}$$

a 、 b 、 c 、 d は、未観測の β_i を積分した以下の周辺尤度を最大化することで推定します：

$$l(a, b, c, d | \theta, \hat{\theta}, \hat{\tau}) \propto \prod_{i=1}^n \int p(\hat{\theta}_i | \beta_i, \theta_i, \hat{\tau}_i) p(\beta_i | a, b, c, d, \theta_i) d\beta_i,$$

そしてこれにより最尤推定値 $(\hat{a}, \hat{b}, \hat{c}, \hat{d})$ を求めます。

系統誤差モデルを用いてキャリブレーションされた信頼区間を計算します。再び $\hat{\theta}_{n+1}$ を新しい対象結果に対する効果推定値の対数とし、 $\hat{\tau}_{n+1}$ を対応する推定標準誤差と β_{n+1} が同じ系統誤差モデルから生じると仮定すると、次のようにになります：

$$\hat{\theta}_{n+1} \sim N(\theta_{n+1} + \hat{a} + \hat{b} \times \theta_{n+1}, \hat{c} + \hat{d} \times |\theta_{n+1}| + \hat{\tau}_{n+1}^2).$$

この式を θ_{n+1} について解くことで、キャリブレーションされた95%信頼区間の下限を求

$$\Phi \left(\frac{\theta_{n+1} + \hat{a} + \hat{b} \times \theta_{n+1} - \hat{\theta}_{n+1}}{\sqrt{(\hat{c} + \hat{d} \times |\theta_{n+1}|) + \hat{\tau}_{n+1}^2}} \right) = 0.025,$$

ここで、 $\Phi(\cdot)$ は標準正規分布の累積分布関数を表します。確率0.975についても同様に上限をp値キャリブレーションと信頼区間のキャリブレーションの両方がEmpiricalCalibrationパッケージで実装されています。

18.2.6

別の方法検証の形として、異なる母集団、異なる医療システム、および/または異なるデータベース間の異質性を表現する一つの方法として、 I^2 スコアがあります。これは、偶然でなく有意な異質性を示す統計量です (Madigan et al., 2013b)。これは、異なる母集団では効果が大幅に異なるか、または異なるデータベース間で効果が異なるかを検証するための統計量です (Schuemie et al., 2018a)。

I^2 スコアは、データベース間の異質性を表現する一つの方法として、 I^2 スコアがあります。これは、偶然でなく有意な異質性を示す統計量です (Higgins et al., 2003)。 I^2 の値を単純に分類することは、すべての状況に適切であるとはいえない場合があります。たとえば、 I^2 の値にそれぞれ「低」、「中程度」、「高」という形容詞を仮に割り当てることはできます。ただし、(Schuemie et al., 2018b) では、推定値の58%のみが I^2 が25%未満であることが観察され、実際には多くの研究で I^2 が25%以上であることが報告されています。



データベース間の異質性を観察すると、推定値の妥当性に疑問が生じます。残念ながら、

18.2.7

研究を計画する際には、不確実なデザイン上の選択肢がしばしば存在します。例えば、層化傾向がある場合は、層化因子を考慮する必要があります。この感度分析の定義は、例えば Rosenbaum (2005) が「さまざまな規模の隠れた偏りによって影響される」と述べています。

18.3

ここでは、第12章の例を基に、ACE阻害薬（ACEi）が血管性浮腫および急性心筋梗塞（AMI）のリスクに対する効果を検討します。12.7で説明されているようにATLASを使用して研究が実施された場合、これらの診断にはATLASが使用されます。12.8で説明されているように、代わりにRを使用して研究が実施された場合、次の部で説明さ

18.3.1

私たちは、因果効果は存在しないと考えられるネガティブコントロール、すなわち曝露とアウトカムとの関連性がない候補リストを作成するには、まず、関心のある曝露をすべて含むことを確認します。18.3に示すように、ACEiおよびTHZクラスのすべての成分を選択します。

次に、「Explore Evidence」タブに移動し、 Generate ボタンをクリックします。エビデンスリストから View Evidence ボタンをクリックできます。これにより、図18.4に示されるように、結果の

| ACEi and THZ combined | | | | | Optimize | | | |
|---|--|-----------------------|--------------|------------------|--|---|--|--|
| Concept Set Expression | Included Concepts 14 | Included Source Codes | | Explore Evidence | Export | Compare | | |
| Show 25 entries | | | | | Search: <input type="text"/> | | | |
| Showing 1 to 14 of 14 entries | | | | | Previous 1 | Next | | |
| | Concept Id | Concept Code | Concept Name | Domain | Standard Concept Caption | Exclude <input type="checkbox"/> <input checked="" type="checkbox"/> Descendants <input type="checkbox"/> Mapped | | |
| | 1342439 | 38454 | trandolapril | Drug | Standard | <input type="checkbox"/> <input checked="" type="checkbox"/> <input type="checkbox"/> | | |
| | 1334456 | 35296 | Ramipril | Drug | Standard | <input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> | | |
| | 1331235 | 35208 | quinapril | Drug | Standard | <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> | | |
| | 1373225 | 54552 | Perindopril | Drug | Standard | <input checked="" type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> | | |
| | 1310756 | 30131 | moexipril | Drug | Standard | <input type="checkbox"/> <input checked="" type="checkbox"/> <input type="checkbox"/> | | |

Figure 18.3: 対象とする曝露および比較対照の曝露を定義するコンセプトを含むコンセプト

| Evidence for all conditions for ACEi and THZ combined | | | | | | | | | | | |
|---|--|----------------------------|--|--|--|---|---|---|---|---|------|
| Save New Concept Set From Selection Below | | | View database record counts (RC) and descendant record counts (DRC) for: SYNPUF 5% | | | | | | | | |
| | | | | | | Show 15 entries | Filter: <input type="text"/> | | | | |
| Showing 1 to 15 of 13,787 entries | | | | | Previous 1 | 2 | 3 | 4 | 5 | ... 920 | Next |
| | Name | Suggested Negative Control | Sort Order | Publication Count (Descendant Concept Match) | Publication Count (Exact Concept Match) | Publication Count (Parent Concept Match) | Product Label Count (Descendant Concept Match) | Product Label (Exact Concept Match) | Product Label (Parent Concept Match) | Product Label (Parent Concept Match) | |
| | Rift valley fever | Y | 13,781 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| | Obstruction due to foreign body accidentally left in operative wound | Y | 13,780 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| | AND/OR body cavity during a procedure | Y | 13,766 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| | Infection by Shigella | Y | 13,766 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |

Figure 18.4: 文献、製品ラベル、および自発的な報告から見つかったエビデンスの概要

このリストには、条件のコンセプトと、その条件を私たちが定義した曝露のいずれかと関連付

次のステップは、候補リストを手動で確認することです。通常はリストの上から始め、最も頻
18.2.1 で述べた基準を考慮しながら臨床医に確認してもらいます。

今回の研究例では、付録 C.1 にリストされた 76 のネガティブコントロールを選択します。

18.3.2

ネガティブコントロールのセットを定義したら、それらを調査に含める必要があります。まず
12.7.3 では、ATLAS がユーザーが選択するいくつかのオプションに基づいて、そのようなコオ
R で実施される場合、SQL（構造化問い合わせ言語）を使用してネガティブコントロールコホ
9 章では、SQL および R を使用してコホートを作成する方法について説明しています。適切な
SQL および R を記述する方法については、読者の練習問題とします。

OHDSI ツールは、ネガティブコントロールから派生したポジティブコントロールを自動的に生
12.7.3 で説明されている ATLAS の「評価設定」セクションにあり、MethodEvaluation パッケージ

```
library(MethodEvaluation)
#      ACEi = 1
#
eoPairs <- data.frame(exposureId = 1,
                      outcomeId = ncs)

pcs <- synthesizePositiveControls(
  connectionDetails = connectionDetails,
  cdmDatabaseSchema = cdmDbSchema,
  exposureDatabaseSchema = cohortDbSchema,
  exposureTable = cohortTable,
  outcomeDatabaseSchema = cohortDbSchema,
  outcomeTable = cohortTable,
  outputDatabaseSchema = cohortDbSchema,
  outputTable = cohortTable,
  createOutputTable = FALSE,
  modelType = "survival",
  firstExposureOnly = TRUE,
  firstOutcomeOnly = TRUE,
  removePeopleWithPriorOutcomes = TRUE,
  washoutPeriod = 365,
  riskWindowStart = 1,
  riskWindowEnd = 0,
  endAnchor = "cohort end",
  exposureOutcomePairs = eoPairs,
  effectSizes = c(1.5, 2, 4),
  cdmVersion = cdmVersion,
  workFolder = file.path(outputFolder, "pcSynthesis"))
```

注意すべきは、推定研究のデザインで使用されたリスク時間設定を模倣しなければならないと

次に、効果を推定するために使用したのと同じ研究を実行して、ネガティブコントロール Methods Library のすべての推定パッケージは、多くの効果を効率的に推定することを容

18.3.3

図 18.5 は、私たちの研究例に含まれているネガティブコントロールとポジティブコント

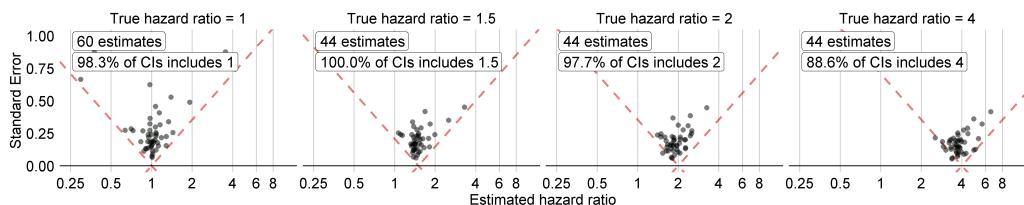


Figure 18.5: ネガティブコントロール（真のハザード比=1）およびポジティブコントロール（真のハザード比=1）に関する推定値。各点はコントロールを表します。点線の下にある推定値は、真の効果を示すものです。

これらの推定値とともに、MethodEvaluationパッケージのcomputeMetrics関数を使用してTable 18.1に示すメトリクスを計算することができます。

Table 18.1: ネガティブコントロールとポジティブコントロールの推定値から得られるメトリクス

| メトリクス | 値 |
|----------------|-------|
| ROC曲線下面積 (AUC) | 0.96 |
| カバー率 | 0.97 |
| 平均精度 | 19.33 |
| 平均二乗誤差 (MSE) | 2.08 |
| 第1種の過誤 | 0.00 |
| 第2種の過誤 | 0.18 |
| 推定なし | 0.08 |

カバー率と第1種の過誤は、それぞれ95%と5%という公称値に非常に近く、AUCも非常に高いです。

図 18.5 では、真のハザード比が1である場合の信頼区間すべてに1が含まれていませんか？Table 18.1 の第1種の過誤は0%です。これは例外的な状況であり、Cyclopsパッケージの信頼区間推定が正確であることを示しています。

18.3.4 P

ネガティブコントロールの推定値を使用して、p値を調整することができます。これはSection 12.8.6で説明したように、要約オブジェクトsummを作成したと仮定すると、経験的なキーワード

```
# Estimates for negative controls (ncs) and outcomes of interest (ois):
ncEstimates <- summ[summ$outcomeId %in% ncs, ]
oiEstimates <- summ[summ$outcomeId %in% ois, ]
```

```
library(EmpiricalCalibration)
plotCalibrationEffect(logRrNegatives = ncEstimates$logRr,
                      seLogRrNegatives = ncEstimates$seLogRr,
                      logRrPositives = oiEstimates$logRr,
                      seLogRrPositives = oiEstimates$seLogRr,
                      showCis = TRUE)
```

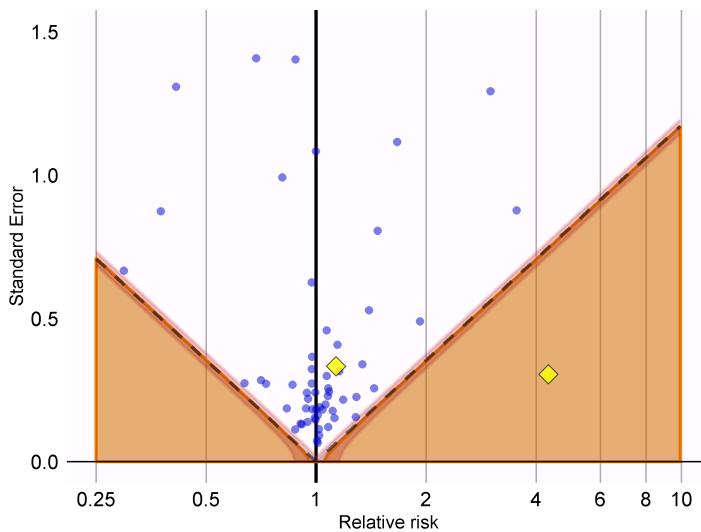


Figure 18.6: P値のキャリブレーション：破線以下の推定値は従来の $p < 0.05$ 。網掛け部分の推定値はキャリブレーションされた $p < 0.05$ 。網掛け部分の端の狭い帯は95%信用区間。点はネガティブコントロール。ダイアモンドはポジティブコントロール。

図 18.6 では、陰影部分が破線で示された領域とほぼ完全に重なっていることがわかります。これで補正済みのp値を計算することができます。

```
null <- fitNull(logRr = ncEstimates$logRr,
                 seLogRr = ncEstimates$seLogRr)
calibrateP(null,
            logRr= oiEstimates$logRr,
            seLogRr = oiEstimates$seLogRr)
```

```
## [1] 1.604351e-06 7.159506e-01
```

そして、キャリブレーションされていないp値と比較してみましょう。

```
oiEstimates$p
```

```
## [1] [1] 1.483652e-06 7.052822e-01
```

予想通り、バイアスはほとんど観察されなかったため、未補正および補正後のp値は非常によく似ています。

18.3.5

同様に、ネガティブコントロールとポジティブコントロールの推定値を用いて、信頼区間

キャリブレーション前の推定ハザード比（95%信頼区間）は、血管性浮腫とAMIでそれぞれ2.22 (1.72-2.88) と4.32 (2.45-8.08) です。補正されたハザード比はそれぞれ2.69 (2.11-3.43) と3.51 (2.34-5.40) です。

18.3.6

1つのデータベース（この場合はIBM MarketScan Medicaid (MDCD) データベース）で分析した結果、血管性浮腫の結果について、合計5つのデータベースにわたるフォレストプロット (DerSimonian and Laird, 1986) を示しています。この図は、EvidenceSynthesisパッケージ

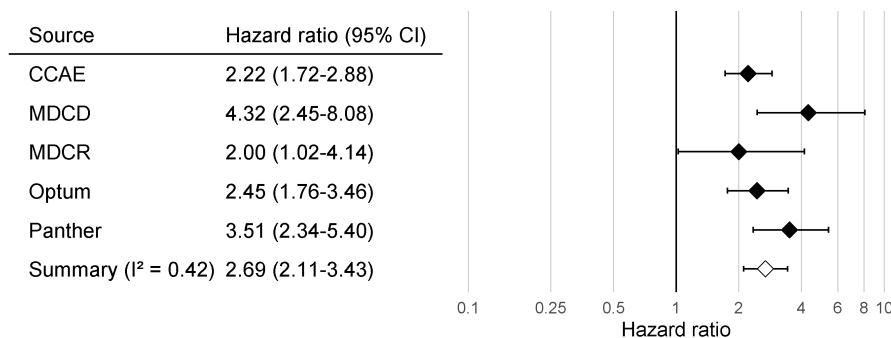


Figure 18.7: 血管浮腫のリスクについてACE阻害薬とサイアザイドおよびサイアザイド様薬の効果。

すべての信頼区間が1より大きいため、何らかの影響があるという点では一致していることはデータベース間の異質性を示唆しています。しかし、図18.8で示したようにキャリブレーションすると、この異質性は、ネガティブコントロールとポジティブコントロールを通じて

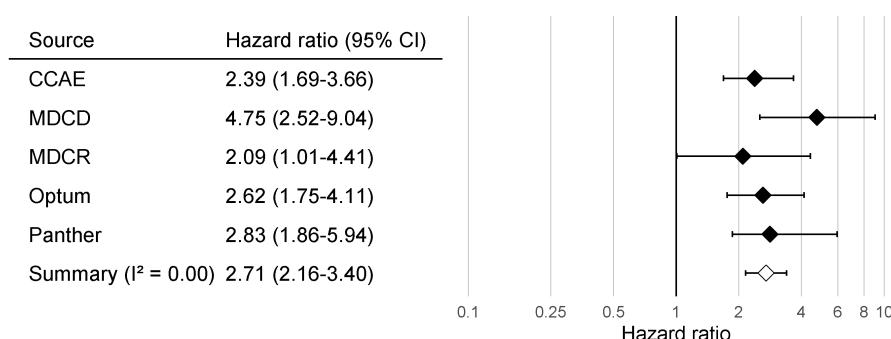


Figure 18.8: キャリブレーションされた血管浮腫のリスクについてACE阻害薬とサイアザイド様薬の効果。

18.3.7

分析におけるデザインの選択肢のひとつは、傾向スコアで変数比マッチングを使用することです。18.2は、変数比マッチングと層別化（10の均等サイズの層）を使用した場合のAMIと血管性浮腫

Table 18.2: 2つの分析におけるキャリブレーション前とキャリブレーション後のハザード比

| アウトカム | 調整方法 | キャリブレーションなし | キャリブレーションあり |
|--------|-------|--------------------|--------------------|
| 血管性浮腫 | マッチング | 4.32 (2.45 - 8.08) | 4.75 (2.52 - 9.04) |
| 血管性浮腫 | 層別化 | 4.57 (3.00 - 7.19) | 4.52 (2.85 - 7.19) |
| 急性心筋梗塞 | マッチング | 1.13 (0.59 - 2.18) | 1.15 (0.58 - 2.30) |
| 急性心筋梗塞 | 層別化 | 1.43 (1.02 - 2.06) | 1.45 (1.03 - 2.06) |

マッチングと層化分析による推定値は、強い一致を示しており、層化分析の信頼区間はマッチング



研究診断により、研究を完全に実施する前でもデザインの選択肢を評価することができます。18.2.2で説明されているように、600件の合成されたポジティブコントロールが導かれます。18.2.3で説明されている評価基準を計算することができます。ベンチマークは公開されており、パッケージの「OHDSI Methods Benchmark vignette」で説明されているように展開することができます。

18.4 OHDSI

推奨される方法は、適用される文脈の中で、その手法のパフォーマンスを経験的に評価することができます。メソッド評価ベンチマークが開発された理由です。このベンチマークは、慢性または急性のアレルギー18.2.2で説明されているように、600件の合成されたポジティブコントロールが導かれます。18.2.3で説明されている評価基準を計算することができます。ベンチマークは公開されており、パッケージの「OHDSI Methods Benchmark vignette」で説明されているように展開することができます。

私たちは、OHDSI Methods Library に収められているすべての手法をこのベンチマークで実行します。18.9に示されているように、第1種の過誤が高く、95%信頼区間のカバー率が低いことが示されています。

このことは、経験的評価とキャリブレーションの必要性を強調しています。経験的評価が実施され18.9の結果から事前に情報を得て、真の効果の大きさが95%信頼区間に含まれていない可能性があります。

また、Methods Libraryにおけるデザインの評価では、経験則に基づくキャリブレーションに依存する可能性があります。

18.5



- 手法の妥当性は、その手法の前提条件が満たされているかどうかによって決まります。
- 可能な場合、これらの前提条件は研究診断を用いて経験的に検証されるべきです。
- コントロール仮説、すなわち答えが既知の質問は、特定の研究デザインが真と一致する必要があります。

¹<http://data.ohdsi.org/MethodEvalViewer/>

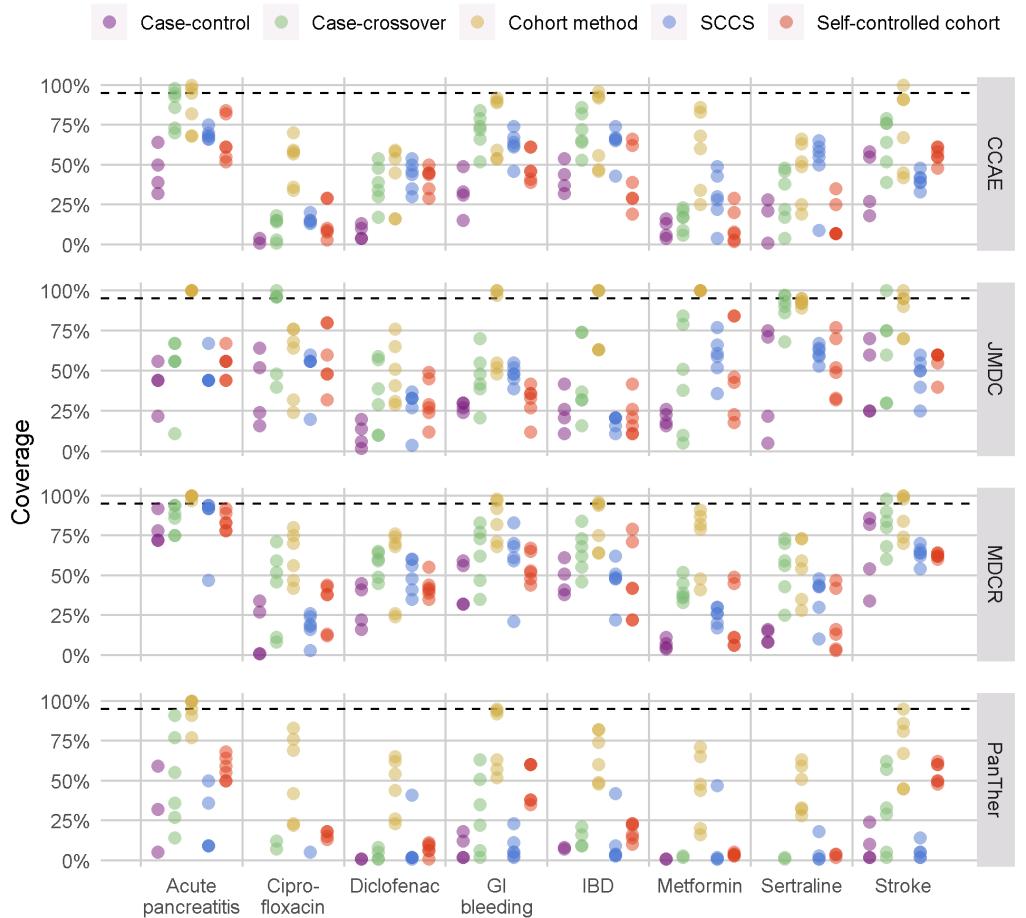


Figure 18.9: Methods LibraryのMethodsに対する95%信頼区間のカバー率。各ドットは自己対照症ケースシリーズ、GI = 消化管、IBD = 炎症性腸疾患。

- 多くの場合、p値や信頼区間は、コントロール仮説を用いて測定された名目上の
- これらの特性は、経験的なキャリブレーションによって多くの場合、名目上の
- 研究診断は、研究者がp-hackingを回避するために関心のある効果に対して盲視

Part V

OHDSI

Chapter 19

著者: Sara Dempster & Martijn Schuemie

ここでは、OHDSIツールを用いた観察研究のデザインと実施に関する一般的な段階的ガイドラインを示します。

さらに、OHDSIコミュニティが推奨する観察研究のためのガイドラインとベストプラクティスを示します。

本章では、OHDSIツール、R、SQLのインフラが読者にとって利用可能であることを前提としています（第 8 章）および第 9 章を参照）。また、読者は OMOP CDM のデータベースを使用して、主に ETL については第 6 章を参照）。ただし、以下で説明するように研究パッケージが準備されています（第 20 章で詳しく説明します）。

19.1

19.1.1

観察研究とは、定義上、患者は単に観察されるだけで、特定の患者の治療に介入する試みではありません。時には、レジストリ研究のように特定の目的のために観察データが収集されることもあることがあります。後者のタイプのデータとしてよく見られる例としては、電子的健康記録（EHR）や保険請求データなどです。観察研究を実施する際の基本的な指針は、研究の疑問を明確に説明し、研究実施前にアプローチする必要があります。この点において、観察研究は臨床試験と変わりません。ただし、臨床試験では、特定の疑問に対する回答を求める場合があります（第 12 章および第 18 章を参照ください）。

19.1.2

観察研究のデザインとパラメータの事前規定は、望ましい結果を得るために無意識または意図的に行われる「ハッキング」と呼ばれることがあります。EHR や保険請求データなどのデータは、時に研究の対象となることがあります。PLP において特に重要です。探索的な理由のみで実施される特性評価研究の場合でも、

19.1.3

観察研究計画は、研究実施前に作成されるプロトコルという形式で文書化されるべきです。少くとも感度分析は、研究デザインの選択が研究結果全体に及ぼす潜在的な影響を評価するために設計する場合には、プロトコルが完了した後で、予期せぬ問題が発生し、プロトコルの修正が必要になるこのような事態が生じた場合、プロトコル自体に変更内容と変更理由を記録することが極めて（sandboxなど）に記録することが理想的です。そうすれば、そのバージョンや修正をタイムスケーリングで追跡することができます。

19.1.4

OHDSIのユニークな利点は、観察研究で繰り返し尋ねられる質問（第2、7、11、12、13章）は、実際にはいくつかの主要なカテゴリーに分類できることを認識し、繰り返される側面を考慮してアドバイスを行います。

OHDSIのアプローチは、これらのステップを共通の枠組みとツール内で比較的簡単に実行でき（19.2.4 を参照）。

19.1.5

標準化されたテンプレートやデザインのもう一つの利点は、研究者がプロトコルの形で研究が実行される方法を理解するための手引書を入手できる点です。

19.1.6 CDM

OHDSIの研究は、観察データベースがOMOP共通データモデル（CDM）に変換されることを前提としています（第4章を参照）。したがって、この前提を満たすためのETL（抽出-変換-読込）プロセス（第6章を参照）が、特定のデータソースについて十分に文書化されていることも重要です。このCDMの目的は、サイト固有のデータ表現を減らす方向に向かうことですが、これは完璧なプロセスによって達成されるべきです。CDMに変換されたソースデータに精通している人々と協力することが重要です。

CDMに加えて、OMOP標準化ボキャブラリシステム（第5章）も、OHDSIフレームワークを構成する重要な要素です。繰り返しになりますが、OMOP CDMへのデータベースの電子タグ付け（ETLing）とOMOPボキャブラリの統合は、OHDSIフレームワークの強みの一つです。

19.2

19.2.1

最初のステップは、研究の関心を、観察研究で対処できる正確な質問に変換することです。たとえば、T2DM（2型糖尿病）の患者に提供されるケアの質を調査したいとします。この大きな目的を、第7章で最初に説明した3つのタイプの質問のいずれかに該当する、はるかに具体的な質問に分解します。

特性評価研究では、「特定の医療環境において、軽度のT2DM患者と重度のT2DM患者に対する治療効果を比較する」という目的があります。また、T2DM治療の処方ガイドラインが、T2DMと心臓病の両方を患っている患者のような特徴的な症例を対象としている場合、軽度のT2DMから重度のT2DMへと進行する患者を予測するモデルを開発することもできます。

純粋に実用的な観点から、研究の問い合わせを定義するには、問い合わせに答えるために必要なアプローチ（第7章を参照ください）。もちろん、独自の分析ツールを設計したり、現在利用可能なツールを活用したりすることもできます。

19.2.2

特定の研究課題に着手する前に、データの質を確認し（第15章を参照）、どのフィールドが何を示すかを理解する。また、軽度のT2DMから重度のT2DMへの進行を予測するモデルを開発するという、上記の例には7章を参照ください。

もう一つの一般的な問題は、特定のケア環境に関する情報の不足です。上述のPLEの例では、

19.2.3

研究対象集団または対象集団を定義することは、あらゆる研究における基本的なステップです。洗練されたコホート定義を作成するには、適切な科学文献のレビューと、特定のデータベースや調査対象集団の定義が説明されたら、OHDSIツールのATLASは、関連するコホートを作成する方法が詳しく説明されています。簡単に説明すると、ATLASは、詳細な包含基準でコホート定義を実装できず、手動でカスタムSQLコードを必要とする場合もあります。

ATLAS UIでは、多数の選択基準に基づくコホートの定義が可能です。コホートへの登録には、CDMのあらゆるドメイン（コンディション、薬剤、プロシージャーなど）に基づいて定義できます（第5章を参照）。この機能を使用するには、ETL プロセス（第6章参照）中にすべてのコードを含む必要があります。ATLAS UIでは、複数の選択基準に基づくコホートの定義が可能です。コホートへの登録には、CDMのあらゆるドメイン（コンディション、薬剤、プロシージャーなど）に基づいて定義できます（第5章を参照）。この機能を使用するには、ETL プロセス（第6章参照）中にすべてのコードを含む必要があります。ATLAS UIでは、複数の選択基準に基づくコホートの定義が可能です。コホートへの登録には、CDMのあらゆるドメイン（コンディション、薬剤、プロシージャーなど）に基づいて定義できます（第5章を参照）。この機能を使用するには、ETL プロセス（第6章参照）中にすべてのコードを含む必要があります。

ATLASがデータベースに接続するように適切に設定されている場合、定義されたコホートが自動的に作成されると、患者の人口統計学的特性の要約と、最も頻繁に観察された薬剤が表示されます。実際には、ほとんどの研究では、複数のコホートまたは複数のコホートセットを指定します（第12章を参照）。さらに、完全な PLE の比較効果研究を実施するには、ネガティブコントロールセットは、これらのネガティブコントロールおよびポジティブコントロールコホートを定義する必要があります（第18章を参照）。

最後に、研究のためのコホートを定義する際には、OHDSIコミュニティで進行中の、頑健な研究パッケージを確認してください。

19.2.4

コホートが定義され生成されたら、利用可能なデータソースで研究の実行可能性を検討する。

この段階における主な活動は、生成したコホートが希望する臨床的特性と一致していることを確認することです。PLE 研究や PLP 研究では、これらのステップを特徴抽出ステップとともに研究パッケージに組み込まれます。

PLE や PLP の実行可能性を評価するもう一つの一般的な重要なステップは、対象コホートの発症率機能を使用してこれらのカウントを特定し、他の箇所で説明されているように、

PLE研究に強く推奨されるもう一つのオプションは、傾向スコア（PS）のマッチング手順を用いたコホートの構成です（第12章を参照）。さらに、これらの最終的にマッチングされたコホートが、研究の目的に適切であることを確認する。

OHDSIコミュニティでは、利用可能なサンプルサイズを考慮した最小検出相対リスク（N

19.2.5

これまでのすべてのステップの準備が完了したら、最終的なプロトコルをまとめます。このPDFでは、PLE研究のフルプロトコルのサンプル目次を提供しています。この目次は、OHDSI githubでも入手できます。このサンプルは、包括的なガイドおよびチェックリストとして提供

図19.1に示されているように、人間が読める形式での最終的な研究プロトコルの作成は、最終的な後者のステップは、以下の図では研究実施と呼ばれています。

これには、ATLASからの最終的な研究パッケージのエクスポートや、必要に応じてカスタムコ

その後、完成したスタディパッケージを使用して、プロトコルに記載されている予備診断ステ

重要なのは、この段階で最終的なプロトコルを臨床協力者や利害関係者に確認してもらうこと

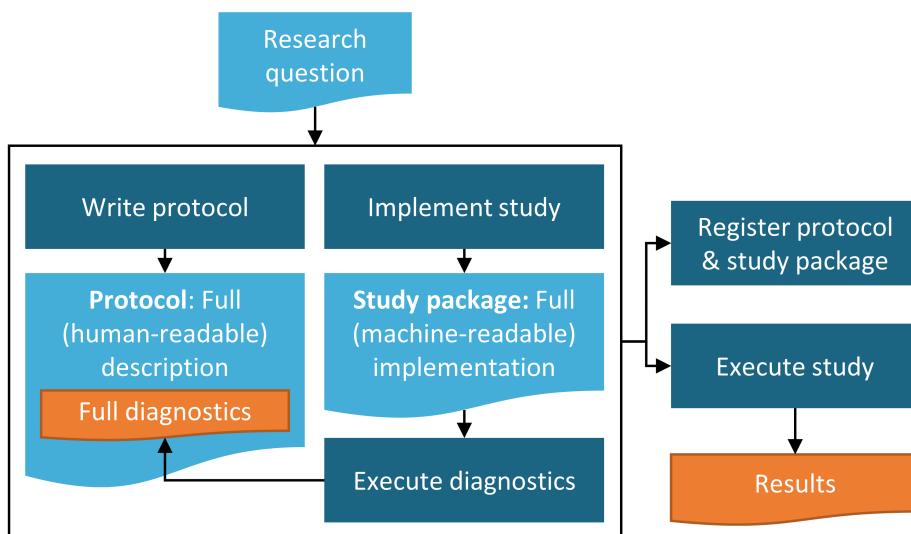


Figure 19.1: 研究プロセスのダイアグラム

19.2.6

これまでのすべてのステップが完了すれば、試験の実施は理想的には単純明快なはずです。も

19.2.7

サンプルサイズが十分で、データの質も妥当な、よく定義された研究では、結果の解釈は多くしかし、解釈がより困難になり、慎重なアプローチが必要となる一般的な状況もいくつかあり

1. サンプルサイズが有意性の境界線上にある場合、信頼区間が大きくなる
2. PLEに特異的な場合：ネガティブコントロールを用いたp値のキャリブレーションにより
3. 研究実施中に予期せぬデータ品質の問題が明らかになる

どのような研究においても、上記の懸念事項について報告し、それに応じて研究結果の解釈を

19.3



- 研究には明確に定義された問い合わせを検討すべきです。
- データの品質、完全性、関連性の事前チェックを適切に行いましょう。
- 可能であれば、プロトコルの開発プロセスにソースデータベースの専門家を含めます。
- 事前にプロトコルに研究計画を記載します。
- 書面によるプロトコルと並行して研究パッケージコードを生成し、最終的な研究報告書を作成します。
- 研究は実施に先立ち登録し、必要に応じて承認を得るべきです。
- 最終報告書または論文の原稿は、臨床専門家やその他の関係者による査読を受けます。

Chapter 20

OHDSI

著者: Kristin Kostka, Greg Klebanov & Sara Dempster

OHDSIのミッションは、観察研究を通じて質の高いエビデンスを生み出すことです。この19章) などが含まれます。OHDSIネットワーク研究は、地理的に分散した多数のデータ

20.1 OHDSI

OHDSI研究ネットワークは、医療における観測データ研究の進展を目指す研究者たちのCDMに変換し、ネットワークの研究に参加することで、このネットワークに参加するこ



データ保持者がOHDSIネットワークに参加するメリット

- 無料ツールへのアクセス: OHDSIは、データの特性評価や標準化された分析(
- 一流の研究コミュニティへの参加: ネットワーク研究の作成と公開、さまざま
- 医療のベンチマークの機会: ネットワーク研究により、データパートナー全体で

20.2 OHDSI

前章(第19章)では、CDMを使用した研究実施に関する一般的な設計上の考慮事項につ

20.2.1 OHDSI

観察研究の典型的な使用例としては、「リアルワールド」の環境における治療の有効性の観察研究の結果は、服薬アドヒアランス、遺伝的多様性、環境要因、全体的な健康状態などをしたがって、ネットワーク研究は、幅広い設定とデータソースを調べることで、観察研究

20.2.2 OHDSI



どのような研究がネットワーク研究と見なされるのでしょうか？ OHDSI研究は、異なる機関の複数のCDMで実施された場合に、OHDSIネットワーク研究とな

OHDSIのアプローチによるネットワーク研究では、OMOP CDMと標準化されたツールや研究

ネットワーク研究は、OHDSI研究コミュニティの重要な一部です。しかし、OHDSI研究をパッケージ化するためのCDMおよびOHDSI Methods Libraryを使用して研究を実施したり、研究対象を一部の機関に限定したりするなどの研究貢献もコミュニティにとって同様に重要です。研究を単一のデータベースで実施する場合、OHDSIコミュニティが実施するオープンなネットワーク研究について説明します。

オープンなOHDSIネットワーク研究の要素：OHDSIネットワーク研究をオープンに実施するためには、研究を特徴づける要素はいくつかあります。これには以下が含まれます。

- すべての文書、研究コード、その後の結果は、OHDSI GitHub で一般公開されます。
- 研究者は、実施する分析の範囲と意図を詳細に記した公開研究プロトコルを作成し、公開します。
- 研究者は、CDMに準拠したコードを含む研究パッケージ（通常はRまたはSQL）を作成します。
- 研究者は、OHDSIネットワーク研究の共同研究者を募るために、OHDSIコミュニティコラボレーションプラットフォームを使用します。
- 分析終了後、集計された研究結果はOHDSI GitHubで公開されます。
- 可能な場合は、研究者は研究 R Shiny アプリケーションを data.ohdsi.org に公開することが推奨されます。

次のセクションでは、独自のネットワーク研究を作成する方法と、ネットワーク研究を実施するための手順について説明します。

20.2.3 OHDSI

OHDSIネットワーク全体で実行する研究を設計するには、研究コードの設計や作成方法のパラメータを選択する必要があります。ただし、データが特定のケア環境（例：プライマリケア、外来診療）や特定の地域（例：米国）によって、コホートの定義が偏ってしまう可能性があるからです。

OHDSIネットワーク研究では、もはや自分のデータのみを対象とした研究パッケージを設計するのではなく、CDMに移植可能な包括的なコホート定義を作成することが推奨されます。OHDSI研究パッケージでは、すべての機関で同じパラメータ化されたコードセットを使用しています。

臨床コーディングのばらつきに加えて、各地域の技術インフラのばらつきも想定して設計する必要があります。研究コードはもはや単一の技術環境で実行されるものではありません。

各OHDSIネットワークサイトは、独自のデータベース層を選択します。

つまり、特定のデータベース方言に研究パッケージをハードコードすることはできないという研究コードは、その方言の演算子に簡単に修正できるSQLの種類にパラメータ化する必要がある

20.2.4 OHDSI

OHDSIはオープンサイエンスのコミュニティであり、OHDSI中央調整センターは、共同研究用CDMとOHDSIツールスタックの当該サイトでの導入の成熟度によって決まります。OHDSI

各研究において、機関での初期活動には以下が含まれます。

- 必要に応じて、研究を機関審査委員会（または同等の委員会）に登録する
- 必要に応じて、研究実施の承認を機関審査委員会から受ける
- 承認済みのCDMにスキーマを読み書きするためのデータベースレベルの権限を取得する
- 研究パッケージを実行するための機能的なRStudio環境の構成を確認する
- 研究コードに技術的な異常がないか確認する
- 技術的な制約内でパッケージを実行するために必要な依存関係のあるRパッケージをインストールする



**データ品質とネットワーク調査：第6章で説明したように、品質管理はETL（抽出-変換-読込）プロセスの基本かつ反復的な要素です。これはネットワーク調査プロセスにおいても同様です。

各機関には、研究パッケージを実行するローカルのデータアナリストが配置されます。このデータアナリストは、研究結果を共有する際には、結果の送信方法や結果の外部公開に関する責任者です。

20.3 OHDSI

OHDSIネットワーク研究を実行するには、以下の三つの一般的な段階があります：

- 研究デザインと実行可能性
- 研究実行
- 結果の公表と発表

20.3.1

研究の実行可能性の段階（または事前研究段階）では、研究の問い合わせを定義し、研究プロトコルを作成します。実行可能性の評価段階の結果として、ネットワークでの実施に備えて公表された最終的な実行可能性の段階は、明確に定義されたプロセスではありません。これは、提案された研究用CDMにアクセスできる必要はありません。研究責任者は、合成データ（例えば、CMS Synthetic Public Use Files、Mitre社のSyntheticMass、Synthea）を使用して対象コホートからのコホート定義のJSONファイルを使用してコホートを作成し特性を明らかにするようになります。Rパッケージを検証すること、第19章で説明されている初期診断を実行することが含まれます。OHDSI研究を承認するための組織固有のプロセスを開始することもあります。すなわち、機関審査委員会の承認などです。実行可能性の段階で、これらの組織固有の活動を完了させることで、研究が実行可能であると判断されます。

20.3.2

実行可能性の検討を完了した後、研究は実行段階に進みます。この期間は、OHDSIネットワーク研究責任者がOHDSIコミュニティに働きかけて、OHDSIネットワーク研究の新規の実施を正式GitHubに研究プロトコルを公開します。研究責任者は、OHDSIコミュニティの週例電話会議や各施設では、研究チームが研究参加の承認を得るための施設内手続きに従い、研究パッケージを提出します。研究責任者は、結果の受け取り方法（SFTPまたは安全なAmazon S3バケット経由など）と結果の分析方法を定義します。実行段階において、妥当な調整が必要な場合は、統合研究チーム（研究責任者および参加施設）で調整を行います。最終的には、研究責任者およびサポートするデータサイエンティスト/統計学者が、必要に応じてデータを分析します。研究責任医師は、参加施設の状況を監視し、参加施設と定期的に連絡を取り合うことで、パッチ（データベースへのアクセス不可など）による問題を早期に発見し、関連する課題が生じる可能性があります。参加施設で発生した問題の解決に役立つ適切なリソースを確保するかどうかは、最終的には参加施設の判断となります。OHDSI研究は迅速に実施できますが、すべての参加施設が研究を実施し、結果を公表するためには時間がかかる場合があります。研究責任者はプロトコルに研究マイルストーンを設定し、事前に終了予定日を伝えて、研究全般のスケジュールを管理します。

20.3.3

結果の公表と公開段階では、研究責任者は他の参加者と協力して、原稿の作成やデータ可視化ツール（例如：JANE Application）の作成と公表を行います。研究責任者がOHDSI研究骨格（Atlasによって生成される）



OHDSIネットワーク研究をどこで発表するかお悩みですか？JANE（ジャーナル/著者名挙げ方）

原稿が作成されると、参加している各共同研究者は、その成果物が外部での出版プロセスに従事する機会を満たすことが期待されています。結果の発表は、研究者が選択する任意のフォーラム（OHDSI研究会議や学術雑誌）で行われます。

20.4 :

現在のネットワーク研究プロセスはマニュアルで行われており、研究チームのメンバーは、研究プロセスを理解する必要があります。ARACHNEは、ネットワーク研究の実施プロセスを合理化し自動化するプラットフォームです。ARACHNEは、データ提供者、研究員、スポンサー、データサイエンティストといった参加組織を単一の共同データベースで管理するツールです。このツールは、データ管理者によって管理される承認ワークフローを含む、完全な標準ベースの研究プロセスを実現します。ARACHNEは、ACHILLESレポートやATLAS設計アーティファクトのインポート機能、自己完結型の論文作成機能などを備えています。

¹<http://www.icmje.org/recommendations/browse/roles-and-responsibilities/defining-the-role-of-authors-and-contributors.html>

Network Study Workflow

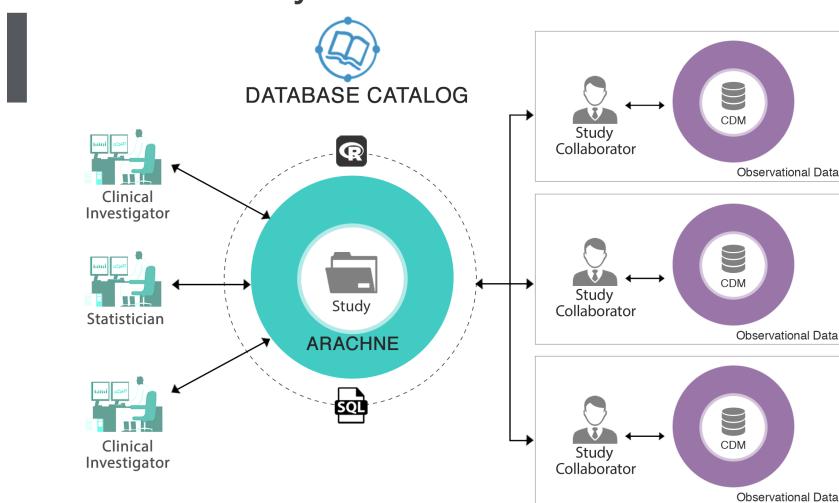


Figure 20.1: ARACHNEネットワーク研究プロセス

20.5 OHDSI

ネットワーク研究を実施する際には、OHDSIコミュニティがOHDSIネットワーク研究の

研究デザインと実現可能性：ネットワーク研究を実施する際には、研究デザインが単一のCDMへのデータ変換に関する標準化された規定をどの程度厳密に遵守しているかによって医師が処方箋を書いた記録と、薬剤師が処方箋を調剤した時間、患者が薬局で薬を受け取るまでの測定誤差は、あらゆる分析ユースケースの結果に潜在的にバイアスを生じさせる可能性がある。

研究の実行：可能な場合、研究責任者は、ATLAS、OHDSI Methods Library、OHDSI Study Skeletonsを活用し、標準化された分析パッケージをできるだけ多く利用する。

結果と普及：研究責任者は、結果を共有する前に、各サイトがローカルのガバナンスルールに基づいており、OMOP CDM の原則と標準化されたボキャブラリを再確認し、OHDSIネットワーク研究は、すべての文書とその後の結果を OHDSI GitHub リポジトリまたは data.ohdsi.org R Shiny サーバーに公開し、完全に透明性のあるものとなっています。論文やOMOP CDM の各サイト間でデータがどのように異なりうるかをジャーナルが理解できるように、4章で説明されているOMOPの観察期間が、適格性ファイル（保険請求データベースに存在するOMOP CDMの観察期間がどのように作成されるかを参照し、ソースシステム内のエンカウンターフィールド）

20.6



- OHDSI研究は、異なる機関の複数のCDMで実施されると、OHDSIネットワーク研究は、すべての人に公開されています。ネットワーク研

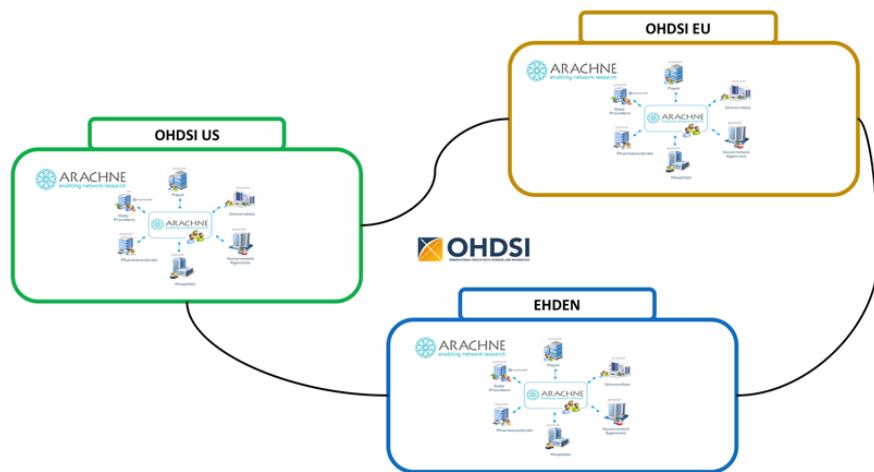


Figure 20.2: ARACHNEのネットワークのネットワークス

- ネットワーク研究の実施でお困りですか？OHDSI研究育成委員会に相談し、研究の共有は思いやりです。すべての研究文書、コード、結果は、OHDSI GitHub または R Shiny アプリケーションで公開されています。研究責任者は、OHDSIイベントで研究発表を行うよう推奨されています。

Chapter A

ACHILLES データベースレベルの特性評価レポート。

ARACHNE 連携ネットワーク研究のオーケストレーションおよび実行を可能にするための基盤。

ATLAS 患者レベルの臨床データからリアルワールドエビデンスを創生するための観察データベース。

バイアス (Bias) 誤差（真の値と推定値の差）の期待値。

ブーリアン変数 (Boolean) 2つの値（真または偽）のみを持つ変数。

医療施設 (Care site) 医療提供が実施される一意に識別された制度上（物理的または組織的）。

症例対照 (研究) (Case control) 集団レベルの効果推定のためのレトロスペクティブ研究。

因果効果 (Causal effect) 集団レベルの推定が関心を寄せるもの。「因果効果」を、タスク。

特性評価 (Characterization) コホートまたは全データベースの記述的研究。詳細は第 11 章参照。

保険請求データ (Claims data) 医療保険会社への請求目的で作成されたデータ。

臨床試験 (Clinical trial) 介入臨床研究。

コホート (Cohort) ある期間内に、1つ以上の選択基準を満たす個人の集団。詳細は第 10 章参照。

コンセプト (Concept) 医学用語で定義された表現（コードが付随する）（例：SNOMED CT）。詳細は第 5 章参照。

コンセプトセット (Concept set) 様々な分析で再利用可能な構成要素として使用できる。詳細は第 10 章参照。

共通データモデル (Common Data Model, CDM) 分析のポータビリティ（同じ分析を異なるシステムで実行可能）。詳細は第 7 章参照。

比較効果 (Comparative Effectiveness) 関心のあるアウトカムに対する2つの異なる曝露の効果を比較する研究。詳細は第 12 章参照。

- コンディション (Condition) 医療従事者が観察したまたは患者が報告した診断、徵候、また
交絡 (Confounding) 主たる関心の曝露が、アウトカムと関連する他の要因と混同されるとされる要因。
共変量 (Covariate) 独立変数として統計モデルで使用されるデータ要素（例：体重）。
データ品質 (Data quality) そのデータが特定の用途に適していると判断するためのデータの属性。
デバイス (Device) 化学作用を超えた機序によって診断または治療の目的で使用される異物。
薬剤 (Drug) 人に投与されたときに特定の生理学的效果を発揮するように調製された生化学的活性物質。
ドメイン (Domain) 共通データモデルのテーブルにおける標準化フィールドに対して使用される分類。
電子的健康記録 (Electronic Health Record, EHR) 医療の過程で生成され、電子システムによって管理される個人の健康記録。
疫学 (Epidemiology) 一定の集団における健康および疾患の分布、パターン、および決定要因。
エビデンスに基づく医療 (Evidence-based medicine) 個々の患者の医療に関する意思決定に用いられる証拠。
ETL (抽出-変換-読込) (Extract-Transform-Load) データをある形式から別の形式に変換するプロセス。
マッチング (Matching) 多くの集団レベルの効果推定のためのアプローチは、曝露された患者群と非曝露群との間の特徴を比較する。
メジャーメント (測定) (Measurement) 患者または患者の検体の体系的かつ標準化された評価。
測定誤差 (Measurement error) 記録された測定値（例：血圧、患者の年齢、治療期間）が実際の値と異なる誤差。
メタデータ (Metadata) 他のデータについて説明し、情報を提供するデータの一式。メタデータはデータの構造や意味を記述する。
Methods Library 観察研究を実行するためにOHDSIコミュニティによって開発された一連のRパッケージ。
モデルの誤特定 (Model misspecification) 多くのOHDSIで用いる方法は、比例ハザード回帰分析。
ネガティブコントロール (Negative control) 曝露がアウトカムを引き起こさないまたは予防する効果を確認する。
アウトカムの組み合わせ。効果推定が真実に沿った結果を生成するかどうかを評価するための検証。
18 章参照。
- オブザベーション (観察) (Observation) 診察、問診または処置のコンテキストで得られた情報。
観察期間 (Observation period) 患者がソースシステム内で臨床イベントの有無にかかわらず観察される期間。
観察研究 (Observational study) 研究者が介入を制御しない研究。
- OHDSI SQL RパッケージSqlRenderを使用して様々な他のSQL方言（方言）に自動変換される。SQLは主にSQL Server SQLのサブセットであるが、追加のパラメータ化が可能である。詳しい説明は9章参照。
- オープンサイエンス (Open science) 科学研究（パブリケーション、データ、物理的サンプル等）を公開する。3章参照。
- アウトカム (Outcome) 分析の焦点となる観察。例えば、患者レベルの予測モデルは、アウトカムを予測する。
患者レベルの予測 (Patient-level prediction) ベースライン特性に基づいて将来のアウトカムを予測する。
フェノタイプ (Phenotype) 身体的特性の説明。これには、体重や髪の色のような可視的な特徴から、遺伝子や脳活動などの生物学的特徴まで幅広い範囲があります。

- 集団レベル推定 (Population-level estimation) 因果効果の研究。平均（集団レベル）の効果を測定する。
ポジティブコントロール (Positive control) 曝露がアウトカムを引き起こすまたは予防する効果を確認する。
アウトカムの組み合わせ。効果推定方法が真実に沿った結果を生成するかどうかを評価する。
18 章参照。
- 処置 (Procedure) 患者に対して診断または治療目的で医療従事者によって命じられる操作。
傾向スコア (Propensity score, PS) 観察研究において、2つの治療群間の均衡をとり無作為化する。
12 章参照。
- プロトコル (Protocol) 研究のデザインを完全に指定するドキュメントで、人が読んで理解できる。
Rabbit-in-a-Hat ソース形式から共通データモデルへのETLを定義を支援するインターフェース。
Rabbitによって生成されたデータベースプロファイルを入力として使用する。詳細は 7 章参照。
- 選択バイアス (Selection bias) データ内の患者集団が統計分析を歪める形で母集団の患者を除外する。
自己対照デザイン (Self-controlled designs) 同一患者内で異なる曝露期間中のアウトカムを比較する。
感度分析 (Sensitivity analysis) 不確実性が存在する分析に関する選択の影響を評価する。
SNOMED 臨床ドキュメントおよび報告書で使用するため、コード、用語、同義語および定義。
研究診断 (Study diagnostics) 特定の分析アプローチが特定の研究質問に対する回答に影響を与える。
18 章参照。
- 研究パッケージ (Study package) 研究を完全に実行するコンピュータ実行プログラム。
17 章参照。
- ソースコード (Source code) ソースデータベースで使用されるコード。例えば、ICD-10 コード。
- 標準コンセプト (Standard Concept) 妥当であると指定され、共通データモデルに含まれる。
- THEMIS 共通データモデル仕様に関して高い粒度と詳細を持つデータ形式に取り組むオープンソース。
- ビジット (Visit) 医療システム内で特定の設定の医療施設において、1人以上の医療従事者が訪問する。
- ボキャブラリ (Vocabulary) 通常アルファベット順に並べられ、定義または翻訳された用語。
- White Rabbit 共通データモデルへのETLを定義する前にデータベースをプロファイリングする。
6 章参照。

Chapter B

この付録には、本書全体で使用されるコホート定義が含まれています。

B.1 ACE

初回イベントコホート

以下のいずれかを持つ人：

- その人の履歴において初回のACE阻害薬（表 B.1）への曝露

かつ、インデックス日から遡って少なくとも365日前からインデックス日0日後の間の連続格コホートとしてイベントを次のように限定します：その個人における全てのイベント

終了日の考え方

カスタム薬剤曝露期間の終了基準：ここでは、指定されたコンセプトセットで見つかったACE阻害薬（表 B.1）の曝露期間終了日を使用

- 曝露期間の間隔は30日を許容
- 曝露期間終了後に0日を追加

コホート圧縮の考え方

30日間のギャップによりコホートを圧縮します。

コンセプトセット定義

Table B.1: ACE阻害薬

| コンセプトID | コンセプト名 | 除外 | 下位層 | マッピング元 |
|---------|----------|-----|-----|--------|
| 1308216 | リシノプリル | いいえ | はい | いいえ |
| 1310756 | モエキシプリル | いいえ | はい | いいえ |
| 1331235 | キナプリル | いいえ | はい | いいえ |
| 1334456 | ラミプリル | いいえ | はい | いいえ |
| 1335471 | ベナゼプリル | いいえ | はい | いいえ |
| 1340128 | カプトプリル | いいえ | はい | いいえ |
| 1341927 | エナラプリル | いいえ | はい | いいえ |
| 1342439 | トランドラプリル | いいえ | はい | いいえ |
| 1363749 | フォシノプリル | いいえ | はい | いいえ |
| 1373225 | ペリンドプリル | いいえ | はい | いいえ |

B.2 ACE

初回イベントコホート

以下のいずれかを持つ人：

- ・その人の履歴において初回のACE阻害薬（表 B.2）への曝露

かつ、インデックス日から遡って少なくとも365日前からインデックス日0日後の間に連続して

選択ルール

選択基準#1：治療開始前1年間に高血圧の診断を受けています。

以下の全ての基準を満たします：

- ・インデックス開始日から遡って365日前からインデックス開始日0日後の間に少なくともB.3)のコンディションが出現します。

選択基準#2：病歴に高血圧治療薬の使用がありません。

以下の全ての基準を満たします：

- ・インデックス開始日1日前までのすべての日に始まる高血圧薬（表 B.4）の薬物使用が完全に0回です。

選択基準#3：ACE単剤療法のみを受けており、併用治療を行っていません。

以下のすべての基準を満たします：

- ・インデックス開始日0日前から7日後の間に始まる高血圧薬（表 B.4）の明確な薬物曝露期

適格コホートとしてイベントを次のように限定します：その個人における最も早いイベント。

終了日の考え方

カスタム薬剤曝露期間の終了の基準：この考え方は、指定されたコンセプトセットで見つけるACE阻害薬の曝露期間の終了日（表B.2）

- ・薬剤曝露間隔が30日を許容します。
- ・薬剤曝露終了後0日追加します。

コホート圧縮の考え方

0日間のギャップサイズで期間によりコホート圧縮します。

コンセプトセット定義

Table B.2: ACE阻害薬

| コンセプトID | コンセプト名 | 除外 | 下位層 | マッピング元 |
|---------|----------|-----|-----|--------|
| 1308216 | リシノプリル | いいえ | はい | いいえ |
| 1310756 | モエキシプリル | いいえ | はい | いいえ |
| 1331235 | キナプリル | いいえ | はい | いいえ |
| 1334456 | ラミプリル | いいえ | はい | いいえ |
| 1335471 | ベナゼプリル | いいえ | はい | いいえ |
| 1340128 | カプトプリル | いいえ | はい | いいえ |
| 1341927 | エナラプリル | いいえ | はい | いいえ |
| 1342439 | トランドラプリル | いいえ | はい | いいえ |
| 1363749 | フォシノプリル | いいえ | はい | いいえ |
| 1373225 | ペリンドプリル | いいえ | はい | いいえ |

Table B.3: 高血圧性障害

| コンセプトID | コンセプト名 | 除外 | 下位層 | マッピング元 |
|---------|--------|-----|-----|--------|
| 316866 | 高血圧性障害 | いいえ | はい | いいえ |

Table B.4: 高血圧治療薬

| コンセプトID | コンセプト名 | 除外 | 下位層 | マッピング元 |
|---------|---------|-----|-----|--------|
| 904542 | トリアムテレン | いいえ | はい | いいえ |
| 907013 | メトラゾン | いいえ | はい | いいえ |
| 932745 | ブメタニド | いいえ | はい | いいえ |
| 942350 | トルセミド | いいえ | はい | いいえ |
| 956874 | フロセミド | いいえ | はい | いいえ |

| コンセプトID | コンセプト名 | 除外 | 下位層 | マッピング元 |
|---------|------------|-----|-----|--------|
| 970250 | スピロノラクトン | いいえ | はい | いいえ |
| 974166 | ヒドロクロロチアジド | いいえ | はい | いいえ |
| 978555 | インダパミド | いいえ | はい | いいえ |
| 991382 | アミロリド | いいえ | はい | いいえ |
| 1305447 | メチルドパ | いいえ | はい | いいえ |
| 1307046 | メトプロロール | いいえ | はい | いいえ |
| 1307863 | ベラパミル | いいえ | はい | いいえ |
| 1308216 | リシノプリル | いいえ | はい | いいえ |
| 1308842 | バルサルタン | いいえ | はい | いいえ |
| 1309068 | ミノキシジル | いいえ | はい | いいえ |
| 1309799 | エプレレノン | いいえ | はい | いいえ |
| 1310756 | モエキシプリル | いいえ | はい | いいえ |
| 1313200 | ナドロール | いいえ | はい | いいえ |
| 1314002 | アテノロール | いいえ | はい | いいえ |
| 1314577 | ネビボロール | いいえ | はい | いいえ |
| 1317640 | テルミサルタン | いいえ | はい | いいえ |
| 1317967 | アリストキレン | いいえ | はい | いいえ |
| 1318137 | ニカルジピン | いいえ | はい | いいえ |
| 1318853 | ニフェジピン | いいえ | はい | いいえ |
| 1319880 | ニソルジピン | いいえ | はい | いいえ |
| 1319998 | アセブトロール | いいえ | はい | いいえ |
| 1322081 | ベタキソロール | いいえ | はい | いいえ |
| 1326012 | イスラジピン | いいえ | はい | いいえ |
| 1327978 | ベンブトロール | いいえ | はい | いいえ |
| 1328165 | ジルチアゼム | いいえ | はい | いいえ |
| 1331235 | キナプリル | いいえ | はい | いいえ |
| 1332418 | アムロジピン | いいえ | はい | いいえ |
| 1334456 | ラミプリル | いいえ | はい | いいえ |
| 1335471 | ベナゼプリル | いいえ | はい | いいえ |
| 1338005 | ビソプロロール | いいえ | はい | いいえ |
| 1340128 | カプトプリル | いいえ | はい | いいえ |
| 1341238 | テラゾシン | いいえ | はい | いいえ |
| 1341927 | エナラプリル | いいえ | はい | いいえ |
| 1342439 | トランドラプリル | いいえ | はい | いいえ |
| 1344965 | グアンファシン | いいえ | はい | いいえ |
| 1345858 | ピンドロール | いいえ | はい | いいえ |
| 1346686 | エプロサルタン | いいえ | はい | いいえ |
| 1346823 | カルベジロール | いいえ | はい | いいえ |
| 1347384 | イルベサルタン | いいえ | はい | いいえ |
| 1350489 | プラゾシン | いいえ | はい | いいえ |
| 1351557 | カンデサルタン | いいえ | はい | いいえ |
| 1353766 | プロプラノロール | いいえ | はい | いいえ |
| 1353776 | フェロジピン | いいえ | はい | いいえ |

| コンセプトID | コンセプト名 | 除外 | 下位層 | マッピング元 |
|----------|---------|-----|-----|--------|
| 1363053 | ドキサゾシン | いいえ | はい | いいえ |
| 1363749 | フォシノプリル | いいえ | はい | いいえ |
| 1367500 | ロサルタン | いいえ | はい | いいえ |
| 1373225 | ペリンドプリル | いいえ | はい | いいえ |
| 1373928 | ヒドララジン | いいえ | はい | いいえ |
| 1386957 | ラベタロール | いいえ | はい | いいえ |
| 1395058 | クロルタリドン | いいえ | はい | いいえ |
| 1398937 | クロニジン | いいえ | はい | いいえ |
| 40226742 | オルメサルタン | いいえ | はい | いいえ |
| 40235485 | アジルサルタン | いいえ | はい | いいえ |

B.3 AMI

初回イベントコホート

以下のいずれかを持つ人：

- 急性心筋梗塞（表 B.5）のコンディション出現

かつ、インデックス日から遡って少なくとも0日前からインデックス日0日後の間に連続

主要イベントがありとなるのは、以下のいずれかの基準を満たす人：

- インデックス日から遡るすべての日から0日後の間に始まり、ビジット終了日がイン（表 B.6）のビジット出現が少なくとも1件。

インデックスイベントのコホート次のように限定します：その個人における全てのイベン

終了日の考え方

日付オフセット終了基準：

このコホート定義の終了日は、インデックスイベントの開始日から7日後とします。

コホート圧縮の考え方

180日間のギャップサイズで期間によりコホートを圧縮します。

コンセプトセット定義

Table B.5: 急性心筋梗塞

| コンセプトID | コンセプト名 | 除外 | 下位層 | マッピング元 |
|---------|---------|-----|-----|--------|
| 314666 | 陳旧性心筋梗塞 | はい | はい | いいえ |
| 4329847 | 心筋梗塞 | いいえ | はい | いいえ |

| コンセプトID | コンセプト名 | 除外 | 下位層 | マッピング元 |
|---------|--------|----|-----|--------|
|---------|--------|----|-----|--------|

Table B.6: 入院または救急室ビジット

| コンセプトID | コンセプト名 | 除外 | 下位層 | マッピング元 |
|---------|--------------|-----|-----|--------|
| 262 | 救急室および入院ビジット | いいえ | はい | いいえ |
| 9201 | 入院ビジット | いいえ | はい | いいえ |
| 9203 | 救急室ビジット | いいえ | はい | いいえ |

B.4

初回イベントコホート

以下のいずれかを持つ人：

- 血管性浮腫のコンディション出現（表 B.7）

イベント発生日の前と後少なくとも0日間の連続した観察期間を持ち、初回イベントを次のよ

主要イベントがありとなるのは、以下のいずれかの基準を満たす人：

- 入院または救急室ビジットインデックス日前全ての日と後0日の間に開始し、インデックス（B.8）で特定されるビジットが少なくとも1回発生します。

初回イベントのコホートを次のように限定します：各個人の全てのイベント。

適格なコホートを次のように限定します：各個人の全てのイベント。

終了日の考え方

このコホート定義の終了日はインデックスイベントの開始日から7日後とします。

コホート圧縮の考え方

30日間のギャップサイズで期間によりコホートを圧縮します。

コンセプトセット定義

Table B.7: 血管性浮腫

| コンセプトID | コンセプト名 | 除外 | 下位層 | マッピング元 |
|---------|--------|-----|-----|--------|
| 432791 | 血管性浮腫 | いいえ | はい | いいえ |

Table B.8: 入院または救急室ビジット

| コンセプトID | コンセプト名 | 除外 | 下位層 | マッピング元 |
|---------|--------------|-----|-----|--------|
| 262 | 救急室および入院ビジット | いいえ | はい | いいえ |
| 9201 | 入院ビジット | いいえ | はい | いいえ |
| 9203 | 救急室ビジット | いいえ | はい | いいえ |

B.5

初回イベントコホート

以下のいずれかを持つ人：

- ・その人の履歴において初回のサイアザイドまたはサイアザイド様利尿薬（表B.9）への曝露

かつ、インデックス日からの遡って少なくとも365日間からインデックス日0日後の間の

選択ルール

選択基準1：治療前の1年間に高血圧の診断をうけています。

以下の全ての基準を満たします：

- ・インデックス開始日から遡って365日前からインデックス開始日0日後の間に少なくとも1回の高血圧性障害（表B.10）のコンディションが出現します。

選択基準#2：病歴に高血圧治療薬の使用がありません。

以下の全ての基準を満たします：

- ・インデックス開始日1日前までのすべての日に始まる高血圧薬（表B.11）の薬物使用が確認できます。

選択基準#3：ACE単剤療法のみを受けており、併用治療を行っていません。

以下のすべての基準を満たします：

- ・インデックス開始日0日前から7日後の間に始まる高血圧薬（表B.11）の明確な薬物曝露が確認できます。

適格コホートとしてイベントを次のように限定します：その個人における最も早いイベント

終了日の考え方

カスタム薬剤曝露期間の終了の基準：この考え方では、指定されたコンセプトセットで見つかる

サイアザイドまたはサイアザイド様利尿薬（表B.9）の曝露期間の終了日

- ・薬剤曝露間隔が30日を許容します。
- ・薬剤曝露終了後0日追加します。

コホート圧縮の考え方

0日間のギャップによりコホートを圧縮します。

コンセプトセット定義

Table B.9: サイアザイドまたはサイアザイド様利尿薬

| コンセプトID | コンセプト名 | 除外 | 下位層 | マッピング元 |
|---------|------------|-----|-----|--------|
| 907013 | メトラゾン | いいえ | はい | いいえ |
| 974166 | ヒドロクロロチアジド | いいえ | はい | いいえ |
| 978555 | インダパミド | いいえ | はい | いいえ |
| 1395058 | クロルタリドン | いいえ | はい | いいえ |

Table B.10: 高血圧性障害

| コンセプトID | コンセプト名 | 除外 | 下位層 | マッピング元 |
|---------|--------|-----|-----|--------|
| 316866 | 高血圧性障害 | いいえ | はい | いいえ |

Table B.11: 高血圧治療薬

| コンセプトID | コンセプト名 | 除外 | 下位層 | マッピング元 |
|---------|------------|-----|-----|--------|
| 904542 | トリアムテレン | いいえ | はい | いいえ |
| 907013 | メトラゾン | いいえ | はい | いいえ |
| 932745 | ブメタニド | いいえ | はい | いいえ |
| 942350 | トルセミド | いいえ | はい | いいえ |
| 956874 | フロセミド | いいえ | はい | いいえ |
| 970250 | スピロノラクトン | いいえ | はい | いいえ |
| 974166 | ヒドロクロロチアジド | いいえ | はい | いいえ |
| 978555 | インダパミド | いいえ | はい | いいえ |
| 991382 | アミロライド | いいえ | はい | いいえ |
| 1305447 | メチルドパ | いいえ | はい | いいえ |
| 1307046 | メトプロロール | いいえ | はい | いいえ |
| 1307863 | ベラパミル | いいえ | はい | いいえ |
| 1308216 | リシノプリル | いいえ | はい | いいえ |
| 1308842 | バルサルタン | いいえ | はい | いいえ |
| 1309068 | ミノキシジル | いいえ | はい | いいえ |
| 1309799 | エプレレノン | いいえ | はい | いいえ |
| 1310756 | モエキシプリル | いいえ | はい | いいえ |
| 1313200 | ナドロール | いいえ | はい | いいえ |
| 1314002 | アテノロール | いいえ | はい | いいえ |

| コンセプトID | コンセプト名 | 除外 | 下位層 | マッピング元 |
|----------|----------|-----|-----|--------|
| 1314577 | ネビボロール | いいえ | はい | いいえ |
| 1317640 | テルミサルタン | いいえ | はい | いいえ |
| 1317967 | アリスキレン | いいえ | はい | いいえ |
| 1318137 | ニカルディピン | いいえ | はい | いいえ |
| 1318853 | ニフェジピン | いいえ | はい | いいえ |
| 1319880 | ニソルジピン | いいえ | はい | いいえ |
| 1319998 | アセブトロール | いいえ | はい | いいえ |
| 1322081 | ベタキソール | いいえ | はい | いいえ |
| 1326012 | イスラジピン | いいえ | はい | いいえ |
| 1327978 | ペンブトロール | いいえ | はい | いいえ |
| 1328165 | ジルチアゼム | いいえ | はい | いいえ |
| 1331235 | キナプリル | いいえ | はい | いいえ |
| 1332418 | アムロジピン | いいえ | はい | いいえ |
| 1334456 | ラミプリル | いいえ | はい | いいえ |
| 1335471 | ベナゼプリル | いいえ | はい | いいえ |
| 1338005 | ビソプロロール | いいえ | はい | いいえ |
| 1340128 | カプトプリル | いいえ | はい | いいえ |
| 1341238 | テラゾシン | いいえ | はい | いいえ |
| 1341927 | エナラプリル | いいえ | はい | いいえ |
| 1342439 | トランドラプリル | いいえ | はい | いいえ |
| 1344965 | グアンファシン | いいえ | はい | いいえ |
| 1345858 | ピンドロール | いいえ | はい | いいえ |
| 1346686 | エプロサルタン | いいえ | はい | いいえ |
| 1346823 | カルベジロール | いいえ | はい | いいえ |
| 1347384 | イルベサルタン | いいえ | はい | いいえ |
| 1350489 | プラゾシン | いいえ | はい | いいえ |
| 1351557 | カンデサルタン | いいえ | はい | いいえ |
| 1353766 | プロプラノロール | いいえ | はい | いいえ |
| 1353776 | フェロジピン | いいえ | はい | いいえ |
| 1363053 | ドキサゾシン | いいえ | はい | いいえ |
| 1363749 | フォシノプリル | いいえ | はい | いいえ |
| 1367500 | ロサルタン | いいえ | はい | いいえ |
| 1373225 | ペリンドプリル | いいえ | はい | いいえ |
| 1373928 | ヒドララジン | いいえ | はい | いいえ |
| 1386957 | ラベタロール | いいえ | はい | いいえ |
| 1395058 | クロルタリドン | いいえ | はい | いいえ |
| 1398937 | クロニジン | いいえ | はい | いいえ |
| 40226742 | オルメサルタン | いいえ | はい | いいえ |
| 40235485 | アジルサルタン | いいえ | はい | いいえ |

B.6

初回イベントコホート

以下のいずれかを持つ人：

- ・その人の履歴において初回の第一選択高血圧治療薬（表 B.12）への曝露かつ、インデックス日から遡って少なくとも365日前からインデックス日365日後の間の連続

選択ルール

以下の全ての基準を満たすこと：

- ・インデックス開始日1日前までのすべての日に少なくとも1回の高血圧治療薬（表 B.13）の薬剤への曝露がちょうど0回出現します。
- ・かつ、インデックス開始日からさかのぼって365日と後0日までの間に高血圧性障害（表 B.14）のコンディションが少なくとも1回出現します。

初回イベントのコホートを次のように限定します：その個人における最も早いイベント。
適格コホートとしてイベントを次のように限定します：その個人における最も早いイベント。

終了日の考え方

終了日の考え方は選択されません。デフォルトでは、コホート終了日はインデックスイベント

コホート圧縮の考え方

0日間のギャップによりコホートを圧縮します。

コンセプトセット定義

Table B.12: 第一選択高血圧治療薬

| コンセプトID | コンセプト名 | 除外 | 下位層 | マッピング元 |
|---------|------------|----|-----|--------|
| 907013 | メトラゾン | NO | YES | NO |
| 974166 | ヒドロクロロチアジド | NO | YES | NO |
| 978555 | インダパミド | NO | YES | NO |
| 1307863 | ベラパミル | NO | YES | NO |
| 1308216 | リシノプリル | NO | YES | NO |
| 1308842 | バルサルタン | NO | YES | NO |
| 1310756 | モエキシプリル | NO | YES | NO |
| 1317640 | テルミサルタン | NO | YES | NO |
| 1318137 | ニカルジピン | NO | YES | NO |
| 1318853 | ニフェジピン | NO | YES | NO |
| 1319880 | ニソルジピン | NO | YES | NO |

| コンセプトID | コンセプト名 | 除外 | 下位層 | マッピング元 |
|----------|----------|----|-----|--------|
| 1326012 | イスラジピン | NO | YES | NO |
| 1328165 | ジルチアゼム | NO | YES | NO |
| 1331235 | キナプリル | NO | YES | NO |
| 1332418 | アムロジピン | NO | YES | NO |
| 1334456 | ラミプリル | NO | YES | NO |
| 1335471 | ベナゼプリル | NO | YES | NO |
| 1340128 | カプトプリル | NO | YES | NO |
| 1341927 | エナラブリル | NO | YES | NO |
| 1342439 | トランドラブリル | NO | YES | NO |
| 1346686 | エプロサルタン | NO | YES | NO |
| 1347384 | イルベサルタン | NO | YES | NO |
| 1351557 | カンデサルタン | NO | YES | NO |
| 1353776 | フェロジピン | NO | YES | NO |
| 1363749 | ホシノブリル | NO | YES | NO |
| 1367500 | ロサルタン | NO | YES | NO |
| 1373225 | ペリンドブリル | NO | YES | NO |
| 1395058 | クロルタリドン | NO | YES | NO |
| 40226742 | オルメサルタン | NO | YES | NO |
| 40235485 | アジルサルタン | NO | YES | NO |

Table B.13: 高血圧治療薬

| コンセプトID | コンセプト名 | 除外 | 下位層 | マッピング元 |
|---------|------------|----|-----|--------|
| 904542 | トリアムテレン | NO | YES | NO |
| 907013 | メトラゾン | NO | YES | NO |
| 932745 | ブメタニド | NO | YES | NO |
| 942350 | トルセミド | NO | YES | NO |
| 956874 | フロセミド | NO | YES | NO |
| 970250 | スピロノラクトン | NO | YES | NO |
| 974166 | ヒドロクロロチアジド | NO | YES | NO |
| 978555 | インダパミド | NO | YES | NO |
| 991382 | アミロイド | NO | YES | NO |
| 1305447 | メチルドバ | NO | YES | NO |
| 1307046 | メトプロロール | NO | YES | NO |
| 1307863 | ベラパミル | NO | YES | NO |
| 1308216 | リシノブリル | NO | YES | NO |
| 1308842 | バルサルタン | NO | YES | NO |
| 1309068 | ミノキシジル | NO | YES | NO |
| 1309799 | エプレレノン | NO | YES | NO |
| 1310756 | モエキシブリル | NO | YES | NO |
| 1313200 | ナドロール | NO | YES | NO |
| 1314002 | アテノロール | NO | YES | NO |

| コンセプトID | コンセプト名 | 除外 | 下位層 | マッピング元 |
|----------|----------|----|-----|--------|
| 1314577 | ネビボロール | NO | YES | NO |
| 1317640 | テルミサルタン | NO | YES | NO |
| 1317967 | アリスキレン | NO | YES | NO |
| 1318137 | ニカルジピン | NO | YES | NO |
| 1318853 | ニフェジピン | NO | YES | NO |
| 1319880 | ニソルジピン | NO | YES | NO |
| 1319998 | アセブトロール | NO | YES | NO |
| 1322081 | ベタキソロール | NO | YES | NO |
| 1326012 | イスラジピン | NO | YES | NO |
| 1327978 | ベンブトロール | NO | YES | NO |
| 1328165 | ジルチアゼム | NO | YES | NO |
| 1331235 | キナプリル | NO | YES | NO |
| 1332418 | アムロジピン | NO | YES | NO |
| 1334456 | ラミプリル | NO | YES | NO |
| 1335471 | ベナゼプリル | NO | YES | NO |
| 1338005 | ビソプロロール | NO | YES | NO |
| 1340128 | カプトプリル | NO | YES | NO |
| 1341238 | テラゾシン | NO | YES | NO |
| 1341927 | エナラプリル | NO | YES | NO |
| 1342439 | トランドラプリル | NO | YES | NO |
| 1344965 | グアンファシン | NO | YES | NO |
| 1345858 | ピンドロール | NO | YES | NO |
| 1346686 | エプロサルタン | NO | YES | NO |
| 1346823 | カルベジロール | NO | YES | NO |
| 1347384 | イルベサルタン | NO | YES | NO |
| 1350489 | プラゾシン | NO | YES | NO |
| 1351557 | カンデサルタン | NO | YES | NO |
| 1353766 | プロプラノロール | NO | YES | NO |
| 1353776 | フェロジピン | NO | YES | NO |
| 1363053 | ドキサゾシン | NO | YES | NO |
| 1363749 | ホシノプリル | NO | YES | NO |
| 1367500 | ロサルタン | NO | YES | NO |
| 1373225 | ペリンドプリル | NO | YES | NO |
| 1373928 | ヒドララジン | NO | YES | NO |
| 1386957 | ラベタロール | NO | YES | NO |
| 1395058 | クロルタリドン | NO | YES | NO |
| 1398937 | クロニジン | NO | YES | NO |
| 40226742 | オルメサルタン | NO | YES | NO |
| 40235485 | アジルサルタン | NO | YES | NO |

Table B.14: 高血圧性障害

| コンセプトID | コンセプト名 | 除外 | 下位層 | マッピング元 |
|---------|--------|----|-----|--------|
| 316866 | 高血圧性障害 | NO | YES | NO |

B.7 3

コホート定義 B.6 と同じだが、インデックス日から遡って少なくとも365日前から1095日

B.8 ACE

初回イベントコホート

以下のいずれかを持つ人：

- ACE阻害薬（表 B.15）の薬物への曝露

かつ、インデックス日から遡って少なくとも0日前から0日後の間の連続した観察があり適格なコホートを個人ごとのすべてのイベントに限定します。

終了日の考え方

この考え方では、指定されたコンセプトセットで見つかったコードから薬剤曝露期間を作成
ACE阻害薬の曝露期間の終了日（表 B.15）

- 薬剤曝露間隔が30日を許容します。
- 薬剤曝露終了後0日追加します。

コホート圧縮の考え方

30日間のギャップサイズで期間によりコホート圧縮します。

コンセプトセット定義

Table B.15: ACE阻害薬

| コンセプトID | コンセプト名 | 除外 | 下位層 | マッピング元 |
|---------|---------|----|-----|--------|
| 1308216 | リシノプリル | NO | YES | NO |
| 1310756 | モエキシプリル | NO | YES | NO |
| 1331235 | キナブリル | NO | YES | NO |
| 1334456 | ラミブリル | NO | YES | NO |
| 1335471 | ベナゼブリル | NO | YES | NO |

| コンセプトID | コンセプト名 | 除外 | 下位層 | マッピング元 |
|---------|----------|----|-----|--------|
| 1340128 | カプトプリル | NO | YES | NO |
| 1341927 | エナラプリル | NO | YES | NO |
| 1342439 | トランドラプリル | NO | YES | NO |
| 1363749 | ホシノプリル | NO | YES | NO |
| 1373225 | ペリンドプリル | NO | YES | NO |

B.9 ARB

コホート定義 B.8 と同じですが、アンジオテンシン受容体拮抗薬（ARB）（表 B.16）がACE阻害薬（表 B.15）の代わりに使用されます。

コンセプトセット定義

Table B.16: アンジオテンシン受容体拮抗薬（ARB）

| コンセプトID | コンセプト名 | 除外 | 下位層 | マッピング元 |
|----------|---------|----|-----|--------|
| 1308842 | バルサルタン | NO | YES | NO |
| 1317640 | テルミサルタン | NO | YES | NO |
| 1346686 | エプロサルタン | NO | YES | NO |
| 1347384 | イルベサルタン | NO | YES | NO |
| 1351557 | カンデサルタン | NO | YES | NO |
| 1367500 | ロサルタン | NO | YES | NO |
| 40226742 | オルメサルタン | NO | YES | NO |
| 40235485 | アジルサルタン | NO | YES | NO |

B.10

コホート定義 B.8 と同じですが、サイアザイドおよびサイアザイド様利尿薬（表 B.17）がACE阻害薬（表 B.15）の代わりに使用されます。

コンセプトセット定義

Table B.17: サイアザイドおよびサイアザイド様利尿薬

| コンセプトID | コンセプト名 | 除外 | 下位層 | マッピング元 |
|---------|------------|----|-----|--------|
| 907013 | メトラゾン | NO | YES | NO |
| 974166 | ヒドロクロロチアジド | NO | YES | NO |
| 978555 | インダパミド | NO | YES | NO |
| 1395058 | クロルタリドン | NO | YES | NO |

| コンセプトID | コンセプト名 | 除外 | 下位層 | マッピング元 |
|---------|--------|----|-----|--------|
|---------|--------|----|-----|--------|

B.11 DCCB

コホート定義 B.8 と同じですが、ジヒドロピリジン系カルシウムチャネル遮断薬（DCCB）（表 B.18）がACE阻害薬（表 B.15）の代わりに使用されます。

コンセプトセット定義

Table B.18: ジヒドロピリジン系カルシウムチャネル遮断薬（DCCB）

| コンセプトID | コンセプト名 | 除外 | 下位層 | マッピング元 |
|---------|--------|----|-----|--------|
| 1318137 | ニカルジピン | NO | YES | NO |
| 1318853 | ニフェジピン | NO | YES | NO |
| 1319880 | ニソルジピン | NO | YES | NO |
| 1326012 | イスラジピン | NO | YES | NO |
| 1332418 | アムロジピン | NO | YES | NO |
| 1353776 | フェロジピン | NO | YES | NO |

B.12 NDCCB

コホート定義 B.8 と同じですが、非ジヒドロピリジン系カルシウムチャネル遮断薬（NDCCB）（表 B.19）がACE阻害薬（表 B.15）の代わりに使用されます。

コンセプトセット定義

Table B.19: 非ジヒドロピリジン系カルシウムチャネル遮断薬（NDCCB）

| コンセプトID | コンセプト名 | 除外 | 下位層 | マッピング元 |
|---------|--------|----|-----|--------|
| 1307863 | ベラパミル | NO | YES | NO |
| 1328165 | ジルチアゼム | NO | YES | NO |

B.13

コホート定義 B.8 と同じですが、ベータ遮断薬（表 B.20）が ACE阻害剤（表 B.15）の代わりに使用されます。

コンセプトセット定義

Table B.20: ベータ遮断薬

| コンセプトID | コンセプト名 | 除外 | 下位層 | マッピング元 |
|---------|----------|----|-----|--------|
| 1307046 | メトプロロール | NO | YES | NO |
| 1313200 | ナドロール | NO | YES | NO |
| 1314002 | アテノロール | NO | YES | NO |
| 1314577 | ネビボロール | NO | YES | NO |
| 1319998 | アセブトロール | NO | YES | NO |
| 1322081 | ベタキソロール | NO | YES | NO |
| 1327978 | ベンブトロール | NO | YES | NO |
| 1338005 | ビソプロロール | NO | YES | NO |
| 1345858 | ピンドロール | NO | YES | NO |
| 1346823 | カルベジロール | NO | YES | NO |
| 1353766 | プロプラノロール | NO | YES | NO |
| 1386957 | ラベタロール | NO | YES | NO |

B.14

ACE阻害剤使用 B.8 と同じですが、ループ利尿薬（表 B.21）がACE阻害剤（表 B.15）の代わりに使用されます。

コンセプトセット定義

Table B.21: ループ利尿薬

| コンセプトID | コンセプト名 | 除外 | 下位層 | マッピング元 |
|---------|--------|----|-----|--------|
| 932745 | ブメタニド | NO | YES | NO |
| 942350 | トルセミド | NO | YES | NO |
| 956874 | フロセミド | NO | YES | NO |

B.15

ACE阻害剤使用 B.8 と同じですが、カリウム保持性利尿薬（表 B.22）がACE阻害剤（表 B.15）の代わりに使用されます。

コンセプトセットの定義

Table B.22: カリウム保持性利尿薬

| コンセプトID | コンセプト名 | 除外 | 下位層 | マッピング元 |
|---------|---------|----|-----|--------|
| 904542 | トリアムテレン | NO | YES | NO |
| 991382 | アミロライド | NO | YES | NO |

B.16 1

ACE阻害剤使用 B.8 と同じですが、アルファ1遮断薬（表 B.23）が ACE阻害剤（表 B.15）の代わりに使用されます。

コンセプトセットの定義

Table B.23: アルファ1遮断薬

| コンセプトID | コンセプト名 | 除外 | 下位層 | マッピング元 |
|---------|--------|----|-----|--------|
| 1341238 | テラゾシン | NO | YES | NO |
| 1350489 | プラゾシン | NO | YES | NO |
| 1363053 | ドキサゾシン | NO | YES | NO |

Chapter C

この付録には、本書のさまざまな章で使用されるネガティブコントロールが含まれています。

C.1 ACE

Table C.1: ACE阻害剤（ACEi）とサイアザイドおよびサイアザイド様利尿薬（

| コンセプトID | コンセプトの名前 |
|----------|--------------------------|
| 434165 | 子宮頸部スメア異常 |
| 436409 | 瞳孔異常 |
| 199192 | 感染を伴わない体幹の擦過傷および/または摩擦熱傷 |
| 4088290 | 乳房欠損 |
| 4092879 | 腎臓欠損 |
| 44783954 | 胃酸逆流 |
| 75911 | 後天性外反母趾 |
| 137951 | 後天性角質変性症 |
| 77965 | 後天性ばね指 |
| 376707 | 急性結膜炎 |
| 4103640 | 切斷足 |
| 73241 | 肛門および直腸ポリープ |
| 133655 | 前腕の熱傷 |
| 73560 | 踵骨の骨棘 |
| 434327 | 大麻乱用 |
| 4213540 | 頸部の体性機能障害 |
| 140842 | 皮膚の質感の変化 |
| 81378 | 膝蓋骨軟骨軟化症 |
| 432303 | コカイン乱用 |
| 4201390 | 人工肛門あり |
| 46269889 | クローン病による合併症 |

| コンセプトID | コンセプトの名前 |
|----------|-----------------|
| 134438 | 接触性皮膚炎 |
| 78619 | 膝の打撲傷 |
| 201606 | クローン病 |
| 76786 | 膝関節運動障害 |
| 4115402 | 睡眠困難 |
| 45757370 | 再建乳房の不均衡 |
| 433111 | 空腹の影響 |
| 433527 | 子宮内膜症 |
| 4170770 | 類表皮囊胞 |
| 4092896 | 糞便内容物異常 |
| 259995 | 開口部内の異物 |
| 40481632 | ガングリオン囊胞 |
| 4166231 | 遺伝的素因 |
| 433577 | 槌状趾 |
| 4231770 | 遺伝性血栓症 |
| 440329 | 合併症を伴わない帯状疱疹 |
| 4012570 | ハイリスクの性行動 |
| 4012934 | ホモシスチン尿症 |
| 441788 | ヒトパピローマウイルス感染 |
| 4201717 | 人工回腸あり |
| 374375 | 耳垢塞栓 |
| 4344500 | 肩関節インピンジメント症候群 |
| 139099 | 嵌入爪 |
| 444132 | 膝損傷 |
| 196168 | 月経不順 |
| 432593 | クワシオルコル |
| 434203 | 打撲の後遺症 |
| 438329 | 自動車事故の後遺症 |
| 195873 | 白色帶下 |
| 4083487 | 網膜ドルーゼン |
| 4103703 | メレナ（黒色便） |
| 4209423 | ニコチン依存症 |
| 377572 | 内耳に対する騒音の影響 |
| 40480893 | 非特異的ツベルクリンテスト反応 |
| 136368 | 非毒性多結節性甲状腺腫 |
| 140648 | 皮膚糸状菌による爪白癬 |
| 438130 | オピオイド乱用 |
| 4091513 | 放屁 |
| 4202045 | ウイルス感染後疲労症候群 |
| 373478 | 老視 |
| 46286594 | ライフスタイルに関連する問題 |
| 439790 | 精神疼痛 |
| 81634 | 下垂乳房 |

| コンセプトID | コンセプトの名前 |
|----------|--------------|
| 380706 | 正乱視 |
| 141932 | 老人性角化症 |
| 36713918 | 腰部の体性機能障害 |
| 443172 | 大きな開放創のない顔の刺 |
| 81151 | 足首の捻挫 |
| 72748 | 肩腱板のストレイン損傷 |
| 378427 | 涙液不足 |
| 437264 | タバコ依存症候群 |
| 194083 | 膣炎および外陰膣炎 |
| 140641 | 尋常性疣贅 |
| 440193 | 手首下垂 |
| 4115367 | 手関節痛 |

Chapter D

1. 目次
2. 略語一覧
3. 要約
4. 変更と更新
5. マイルストーン
6. 研究の根拠と背景
7. 研究目的
 - 主要仮説
 - 二次仮説
 - 主要目的
 - 二次目的
8. 研究方法
 - 研究デザイン
 - データソース
 - 研究対象集団
 - 曝露
 - 結果（アウトカム）
 - 共変量
9. データ解析計画
 - リスク期間の計算
 - モデル仕様
 - データベース間の効果推定の統合
 - 実施する解析
 - 出力
 - エビデンス評価
10. 研究診断
 - サンプルサイズと検出力
 - コホートの比較可能性
 - 系統的誤差の評価

11. 研究方法の強みと限界
12. 研究対象者の保護
13. 有害事象と有害反応の管理および報告
14. 研究結果の普及およびコミュニケーション計画
15. 付録：ネガティブコントロール
16. 参考文献

Chapter E

この付録には、本書の演習に対する回答例が含まれています。

E.1

演習 4.1

演習で説明されている内容に基づくと、ジョンのレコードは表 E.1 のようになるはずです

Table E.1: PERSONテーブル

| カラム名 | 値 | 説明 |
|----------------------|---------------------|-----------------------------|
| PERSON_ID | 2 | 一意の整数。 |
| GENDER_CONCEPT_ID | 8507 | 男性のコンセプト ID は 8507。 |
| YEAR_OF_BIRTH | 1974 | |
| MONTH_OF_BIRTH | 8 | |
| DAY_OF_BIRTH | 4 | |
| BIRTH_DATETIME | 1974-08-04 00:00:00 | 時間が不明な場合は0時（00:00:00） |
| DEATH_DATETIME | NULL | |
| RACE_CONCEPT_ID | 8516 | アフリカ系アメリカ人のコンセプト ID は 8516。 |
| ETHNICITY_CONCEPT_ID | 38003564 | 38003564 |
| LOCATION_ID | | 「非ヒスパニック」を示す。 |
| PROVIDER_ID | | 住所は不明。 |
| CARE_SITE | | 主治医が不明。 |
| PERSON_SOURCE_VALUE | NULL | 主たる医療施設が不明。 提供されていない。 |

| カラム名 | 値 | 説明 |
|-----------------------------|------------------|---------------|
| GENDER_SOURCE_VALUE | Man | 説明で使用されたテキスト。 |
| GENDER_SOURCE_CONCEPT_ID | 0 | |
| RACE_SOURCE_VALUE | African American | 説明で使用されたテキスト。 |
| RACE_SOURCE_CONCEPT_ID | 0 | |
| ETHNICITY_SOURCE_VALUE | NULL | |
| ETHNICITY_SOURCE_CONCEPT_ID | 0 | |

演習 4.2

演習で説明されている内容に基づくと、ジョンのレコードは表 E.2 のようになるはずです。

Table E.2: OBSERVATION_PERIODテーブル

| カラム名 | 値 | 説明 |
|-------------------------------|------------|----------------------------|
| OBSERVATION_PERIOD_ID | 2 | 一意の整数。 |
| PERSON_ID | 2 | これはPERSONテーブルのジョンのレコードのID。 |
| OBSERVATION_PERIOD_START_DATE | 2015-01-01 | 加入日の日付。 |
| OBSERVATION_PERIOD_END_DATE | 2019-07-01 | データ抽出日以降のデータが存在するこの期間。 |
| PERIOD_TYPE_CONCEPT_ID | 44814722 | 44814724 は「保険加入期間」を示す。 |

演習 4.3

演習で説明されている内容に基づくと、ジョンのレコードは表 E.3 のようになるはずです。

Table E.3: DRUG_EXPOSUREテーブル

| カラム名 | 値 | 説明 |
|------------------|----------|--|
| DRUG_EXPOSURE_ID | 1001 | 一意の整数。 |
| PERSON_ID | 2 | これはPERSONテーブルのジョンのレコードのID。 |
| DRUG_CONCEPT_ID | 19078461 | 提供されたNDCコードは標準コンセプトID 19078461 にマッピングされる。 |

| カラム名 | 値 | 説明 |
|------------------------------|---------------------|--|
| DRUG_EXPOSURE_START_DATE | 2019-05-01 | 薬剤への曝露開始日。 |
| DRUG_EXPOSURE_START_DATETIME | 2019-05-01 00:00:00 | 時間が不明なため0時を使用。 |
| DRUG_EXPOSURE_END_DATE | 2019-05-31 | 開始日 + 処方日数に基づく。 |
| DRUG_EXPOSURE_END_DATETIME | 2019-05-31 00:00:00 | 時間が不明なため0時を使用。 |
| VERBATIM_END_DATE | NULL | 提供されていない。 |
| DRUG_TYPE | 38000177 | 38000177 |
| CONCEPT_ID | | は「書かれた処方箋」を示す。 |
| STOP_REASON | NULL | |
| REFILLS | NULL | |
| QUANTITY | NULL | 提供されていない。 |
| DAYS_SUPPLY | 30 | 演習に記述されている通り。 |
| SIG | NULL | 提供されていない。 |
| ROUTE_CONCEPT_ID | 4132161 | 4132161 は「経口」を示す。 |
| LOT_NUMBER | NULL | 提供されていない。 |
| PROVIDER_ID | NULL | 提供されていない。 |
| VISIT_OCCURRENCE_ID | NULL | ビジットに関する情報は提供されない。 |
| VISIT_DETAIL_ID | NULL | |
| DRUG_SOURCE_VALUE | 76168009520 | 提供されたNDCコード。 |
| DRUG_SOURCE_CONCEPT_ID | 583945 | 583945 は薬剤のソースコードの値を表す (ND-Cコード「76168009520」)。 |
| ROUTE_SOURCE_VALUE | NULL | |

演習 4.4

一連のレコードを見つけるには、CONDITION_OCCURRENCEテーブルをクエリする必要がある。

```
library(DatabaseConnector)
connection <- connect(connectionDetails)
sql <- "SELECT *
FROM @cdm.condition_occurrence
WHERE condition_concept_id = 192671;"
```

```
result <- renderTranslateQuerySql(connection, sql, cdm = "main")
head(result)
```

```
##   CONDITION_OCCURRENCE_ID PERSON_ID CONDITION_CONCEPT_ID ...
## 1                   4657      273           192671 ...
## 2                   1021      61           192671 ...
## 3                   5978      351           192671 ...
## 4                   9798      579           192671 ...
## 5                   9301      549           192671 ...
## 6                   1997     116           192671 ...
```

演習 4.5

一連のレコードを見つけるには、CONDITION_OCCURRENCEテーブルのCONDITION_SOURCE_VALUE列を確認します。

```
sql <- "SELECT *
FROM @cdm.condition_occurrence
WHERE condition_source_value = 'K92.2';"

result <- renderTranslateQuerySql(connection, sql, cdm = "main")
head(result)
```

```
##   CONDITION_OCCURRENCE_ID PERSON_ID CONDITION_CONCEPT_ID ...
## 1                   4657      273           192671 ...
## 2                   1021      61           192671 ...
## 3                   5978      351           192671 ...
## 4                   9798      579           192671 ...
## 5                   9301      549           192671 ...
## 6                   1997     116           192671 ...
```

演習 4.6

この情報はOBSERVATION_PERIODテーブルに保存されています：

```
library(DatabaseConnector)
connection <- connect(connectionDetails)
sql <- "SELECT *
FROM @cdm.observation_period
WHERE person_id = 61;"

renderTranslateQuerySql(connection, sql, cdm = "main")

##   OBSERVATION_PERIOD_ID PERSON_ID OBSERVATION_PERIOD_START_DATE ...
## 1                   61          61           1968-01-21 ...
```

E.2

演習 5.1

コンセプト ID 192671 (“消化管出血”)

演習 5.2

ICD-10CMコード：

- K29.91 “出血を伴う胃十二指腸炎、詳細不明”
- K92.2 “消化管出血、詳細不明”

ICD-9CMコード：

- 578 “消化管出血”
- 578.9 “消化管出血、詳細不明”

演習 5.3

MedDRA基本語 (Preferred Terms, PT) :

- “消化管出血” (コンセプトID 35707864)
- “腸出血” (コンセプトID 35707858)

E.3 ETL (Extract-Transform-Load)

演習 6.1

- データの専門家とCDMの専門家が協力してETLの設計を行います。
- 医学知識を持つ人がコードマッピングを作成します。
- エンジニアがETLを実装します。
- 全員が品質管理に関与します。

演習 6.2

| カラム | 値 | 回答 |
|----------------------|----------|---------------------|
| PERSON_ID | A123B456 | このカラムは整数型のデータタイプ |
| GENDER_CONCEPT_ID | 8532 | |
| YEAR_OF_BIRTH | NULL | 生年月日の月や日が不明な場合は |
| MONTH_OF_BIRTH | NULL | |
| DAY_OF_BIRTH | NULL | |
| RACE_CONCEPT_ID | 0 | 人種はWHITEであり、これは8527 |
| ETHNICITY_CONCEPT_ID | 8527 | 民族の情報が提供されていないため |

| カラム | 値 | 回答 |
|------------------------|---------------|----|
| PERSON_SOURCE_VALUE | A123B456 | |
| GENDER_SOURCE_VALUE | F | |
| RACE_SOURCE_VALUE | WHITE | |
| ETHNICITY_SOURCE_VALUE | NONE PROVIDED | |

演習 6.3

| カラム | 値 |
|---------------------|------------|
| VISIT_OCCURRENCE_ID | 1 |
| PERSON_ID | 11 |
| VISIT_START_DATE | 2004-09-26 |
| VISIT_END_DATE | 2004-09-30 |
| VISIT_CONCEPT_ID | 9201 |
| VISIT_SOURCE_VALUE | inpatient |

E.4

演習 7.1

1. 特性評価
2. 患者レベルの予測
3. 集団レベルの推定

演習 7.2

そうではないかもしれません。ジクロフェナク曝露コホートと比較可能な非曝露コホートを定

E.5 SQL R

演習 9.1

人数を計算するには、単純にPERSONテーブルをクエリすればよいです：

```
library(DatabaseConnector)
connection <- connect(connectionDetails)
sql <- "SELECT COUNT(*) AS person_count
FROM @cdm.person;"

renderTranslateQuerySql(connection, sql, cdm = "main")

##    PERSON_COUNT
## 1      2694
```

演習 9.2

セレコキシブの処方を少なくとも1回受けた人の人数を計算するには、DRUG_EXPOSURE

```
library(DatabaseConnector)
connection <- connect(connectionDetails)
sql <- "SELECT COUNT(DISTINCT(person_id)) AS person_count
FROM @cdm.drug_exposure
INNER JOIN @cdm.concept_ancestor
  ON drug_concept_id = descendant_concept_id
INNER JOIN @cdm.concept ingredient
  ON ancestor_concept_id = ingredient.concept_id
WHERE LOWER(ingredient.concept_name) = 'celecoxib'
  AND ingredient.concept_class_id = 'Ingredient'
  AND ingredient.standard_concept = 'S';"

renderTranslateQuerySql(connection, sql, cdm = "main")

##    PERSON_COUNT
## 1      1844
```

COUNT(DISTINCT(person_id))を使用して、重複することない人数を求めるに注意し
代わりに、すでに成分レベルにまとめられているDRUG_EXPOSUREテーブルを使用することも

```
library(DatabaseConnector)
connection <- connect(connectionDetails)

sql <- "SELECT COUNT(DISTINCT(person_id)) AS person_count
FROM @cdm.drug_exposure
INNER JOIN @cdm.concept ingredient
  ON drug_concept_id = ingredient.concept_id
WHERE LOWER(ingredient.concept_name) = 'celecoxib'
  AND ingredient.concept_class_id = 'Ingredient'
  AND ingredient.standard_concept = 'S';"

renderTranslateQuerySql(connection, sql, cdm = "main")
```

```
## PERSON_COUNT
## 1 1844
```

演習 9.3

曝露期間中の診断の数を計算するには、以前のクエリを拡張してCONDITION_OCCURRENCEテーブルを追加します。

```
library(DatabaseConnector)
connection <- connect(connectionDetails)
sql <- "SELECT COUNT(*) AS diagnose_count
FROM @cdm.drug_era
INNER JOIN @cdm.concept ingredient
  ON drug_concept_id = ingredient.concept_id
INNER JOIN @cdm.condition_occurrence
  ON condition_start_date >= drug_era_start_date
    AND condition_start_date <= drug_era_end_date
INNER JOIN @cdm.concept_ancestor
  ON condition_concept_id = descendant_concept_id
WHERE LOWER(ingredient.concept_name) = 'celecoxib'
  AND ingredient.concept_class_id = 'Ingredient'
  AND ingredient.standard_concept = 'S'
  AND ancestor_concept_id = 192671;"

renderTranslateQuerySql(connection, sql, cdm = "main")
```

```
## DIAGNOSE_COUNT
## 1 41
```

この場合、DRUG_EXPOSUREテーブルではなくDRUG ERAテーブルを使用することが重要です。

E.6

演習 10.1

以下の要件をコード化する初期イベント基準を作成します:

- ・ジクロフェナクの新規ユーザー
- ・年齢は16歳以上
- ・曝露前に少なくとも365日間の連續した観察期間があること

完了したときには、コホート・エントリー・イベントのセクションは図E.1のようになります。

ジクロフェナクのコンセプトセットは図E.2のように、成分「Diclofenac」とそのすべての下位概念が含まれます。次に、図E.3に示されるように、NSAIDの曝露の履歴がないことを求めます。

Cohort Entry Events

Events having any of the following criteria:

a drug era of diclofenac ▾

for the first time in the person's history

with age in years at era start Greater or Equal To 16

+ Add attribute... ▾

+ Add Initial Event ▾

Delete Criteria

with continuous observation of at least 365 days before and 0 days after event index date

Limit initial events to: earliest event per person.

Restrict initial events

Figure E.1: ジクロフェナクの新規ユーザーのコホート・エントリー・イベント設定

| Concept Set Expression | | Included Concepts 11473 | Included Source Codes | Export | Import |
|------------------------|--------------|-------------------------|-----------------------|--------------------------|---|
| Name: | | | | | |
| diclofenac | | | | | Search: <input type="text"/> |
| Show | 25 ▾ entries | | | | Previous 1 Next |
| | | | | | |
| Concept Id | Concept Code | Concept Name | Domain | Standard Concept Caption | Exclude Descendants Mapped |
| 1124300 | 3355 | Diclofenac | Drug | Standard | <input type="checkbox"/> <input checked="" type="checkbox"/> <input type="checkbox"/> |

Classification Non-Standard Standard

Figure E.2: ジクロフェナクのコンセプトセット

Inclusion Criteria

New inclusion criteria

Without prior exposure to any NSAID

Copy Delete

1. Without prior exposure to any NSAID

Excluding subjects with prior exposure to any NSAID

having all of the following criteria:

+ Add criteria to group... ▾

with exactly 0 using all occurrences of:

a drug exposure of NSAIDs ▾

+ Add attribute... ▾

where event starts between All days Before and 1 days Before

index start date add additional constraint

restrict to the same visit occurrence

allow events from outside observation period

Delete Criteria

Limit qualifying events to: earliest event per person.

Figure E.3: NSAID曝露の履歴がないことの要求

NSAIDsのコンセプトセットは、NSAIDsクラスとそのすべての下位層を含めるように、図E.4のようになるはずで、よって、NSAIDを含むすべての薬剤を含むことになります。

The screenshot shows a user interface for creating a concept set. At the top, there are tabs: 'Concept Set Expression' (highlighted in blue), 'Included Concepts (23112)', 'Included Source Codes', 'Export', and 'Import'. Below the tabs, the 'Name:' field contains 'NSAIDs'. Underneath, there's a search bar with 'Show 25 entries' and a 'Search:' input field. A table displays one entry: 'ANTIINFLAMMATORY AND ANTRHEUMATIC PRODUCTS, NON-STEROIDS' (Concept ID: 21603933, Concept Code: M01A). The table has columns: Concept Id, Concept Code, Concept Name, Domain, Standard Concept Caption, Exclude, Descendants, and Mapped. The 'Classification' checkbox is checked. A legend at the bottom indicates: Classification (purple square), Non-Standard (red square), and Standard (blue square).

Figure E.4: NSAIDsのコンセプトセット

さらに、図E.5のよう、以前にがんの診断がないことも要求します。

The screenshot shows an 'Inclusion Criteria' editor. It includes a 'New inclusion criteria' button and a 'Without prior diagnose of cancer' section with 'Copy' and 'Delete' buttons. Below it, a section for excluding subjects with prior cancer diagnosis is shown. The main area contains a list of criteria: 'Excluding subjects with prior exposure to any NSAID' and 'Without prior diagnose of cancer'. The 'Without prior diagnose of cancer' section is expanded, showing a detailed configuration. It specifies 'having all' of the following criteria: 'with exactly 0 using all occurrences of: a condition occurrence of Broad malignancies'. There are options to 'Add attribute...' and 'Delete Criteria'. Below this, there are fields for specifying time periods: 'event starts between All days Before and 0 days Before', 'index start date', and 'add additional constraint'. There are also checkboxes for 'restrict to the same visit occurrence' and 'allow events from outside observation period'. At the bottom, there's a note: 'Limit qualifying events to: earliest event per person.'

Figure E.5: 以前にがんの診断がないことの要求

「広範な悪性腫瘍」のコンセプトセットは、上位コンセプト「悪性腫瘍」とそのすべての下位E.6のようになるはずです。

最後に、図E.7のように。曝露中断をコホート・エグジット基準として定義します（30日間）

演習 10.2

読みやすくするため、ここではSQLを2つのステップに分けます。まず、すべての心筋梗塞コ

```
library(DatabaseConnector)
connection <- connect(connectionDetails)
sql <- "SELECT person_id AS subject_id,
```

Concept Set Expression Included Concepts 4401 Included Source Codes Export Import

Name: Broad malignancies

Show 25 entries Search:

Showing 1 to 1 of 1 entries Previous 1 Next

| | Concept Id | Concept Code | Concept Name | Domain | Standard Concept Caption | Exclude | Descendants | Mapped |
|--|------------|--------------|------------------------------|-----------|--------------------------|--------------------------|-------------------------------------|--------------------------|
| | 443392 | 363346000 | Malignant neoplastic disease | Condition | Standard | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/> |

Classification Non-Standard Standard

Figure E.6: 広範な悪性腫瘍のコンセプトセット

Cohort Exit

Event Persistence:
Event will persist until: end of a continuous drug exposure ▾

Continuous Exposure Persistence:
Specify a concept set that contains one or more drugs. A drug era will be derived from all drug exposure events for any of the drugs within the concept set, using the specified persistence window as a maximum allowable gap in days between successive exposure events and adding a specified surveillance window to the final exposure event. If no exposure event end date is provided, then an exposure event end date is inferred to be event start date + days supply in cases when days supply is available or event start date + 1 day otherwise. This event persistence assures that the cohort end date will be no greater than the drug era end date.

Concept set containing the drug(s) of interest: diclofenac ▾

- Persistence window: allow for a maximum of 30 days between exposure records when inferring the era of persistence exposure
- Surveillance window: add 0 days to the end of the era of persistence exposure as an additional period of surveillance prior to cohort exit.

Censoring Events:
Exit Cohort based on the following criteria: + Add Censoring Event ▾

No censoring events selected.

Figure E.7: コホート・イグジット日の設定

```

    condition_start_date AS cohort_start_date
INTO #diagnoses
FROM @cdm.condition_occurrence
WHERE condition_concept_id IN (
    SELECT descendant_concept_id
    FROM @cdm.concept_ancestor
    WHERE ancestor_concept_id = 4329847 --
)
AND condition_concept_id NOT IN (
    SELECT descendant_concept_id
    FROM @cdm.concept_ancestor
    WHERE ancestor_concept_id = 314666 --
);
"

renderTranslateExecuteSql(connection, sql, cdm = "main")

```

次に、入院または救急室受診時に出現したもののみを選択し、特定のCOHORT_DEFINITION_ID

```

sql <- "INSERT INTO @cdm.cohort (
    subject_id,
    cohort_start_date,
    cohort_definition_id
)
SELECT subject_id,
    cohort_start_date,
    CAST (1 AS INT) AS cohort_definition_id
FROM #diagnoses
INNER JOIN @cdm.visit_occurrence
    ON subject_id = person_id
        AND cohort_start_date >= visit_start_date
        AND cohort_start_date <= visit_end_date
WHERE visit_concept_id IN (9201, 9203, 262);      ;"
"

renderTranslateExecuteSql(connection, sql, cdm = "main")

```

コンディションの日がビギット開始日と終了日の間になることを要求する代わりに、コンディションの終了日を考慮しないように注意してください。通常、コホートが一時テーブルが不要になったらクリーンアップすることをお勧めします：

```

sql <- "TRUNCATE TABLE #diagnoses;
DROP TABLE #diagnoses;"

renderTranslateExecuteSql(connection, sql)

```

E.7

演習 11.1

ATLASで  **Data Sources** をクリックし、興味のあるデータソースを選択します。図 E.8 のように、薬剤曝露レポートを選択し、「Table(表)」タブを選択して「celecoxib(セレコキシブ)」を検索することができます。ここでは、この特定のデータベースがセレ

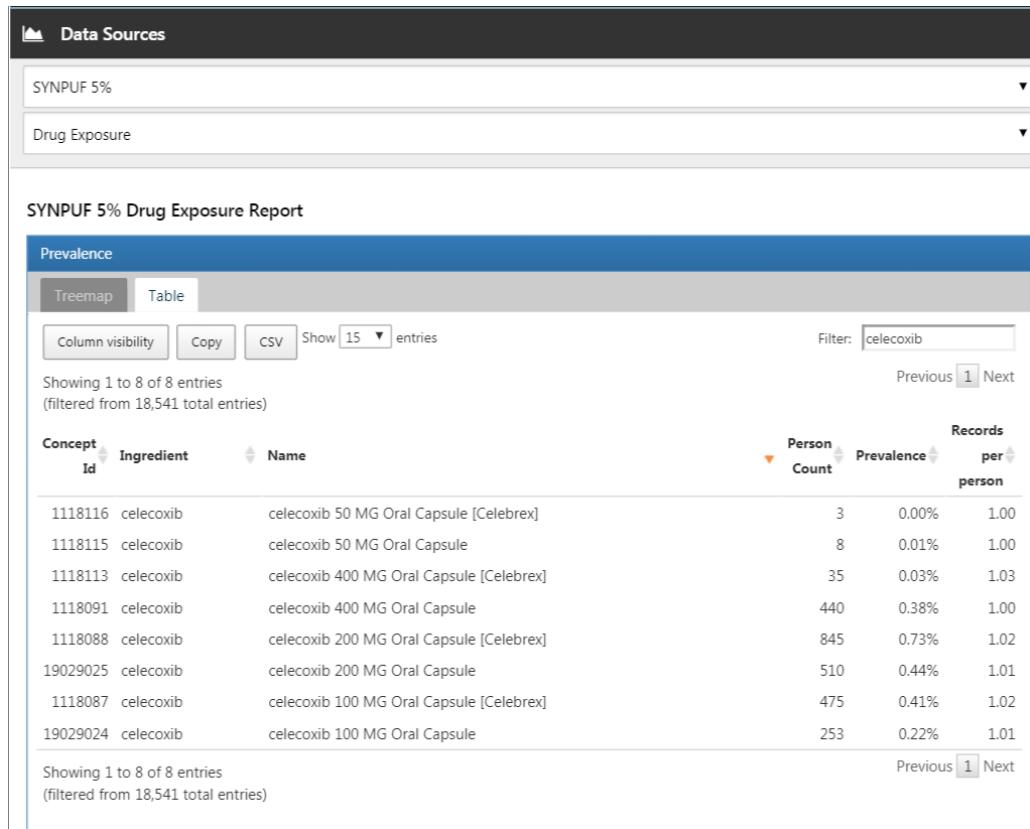


Figure E.8: データソースの特性評価

演習 11.2

 **Cohort Definitions** をクリックして「New cohort(新規コホート)」を作成します。コホート名を「new users(セレコキシブ新規ユーザー)」) と付け、「Concept Sets(コンセプトセット)」を「Concept Set(新規コンセプトセット)」をクリックし、コンセプトセットに意味のある名前(セレコキシブ)」) を付けます。  **Search** モジュールを開き、「Celecoxib(セレコキシブ)」を検索し、クラスを「Ingredient(成分)」、標準コンセプトを「Standard(標準)」とするように限定し、 をクリックして、図 E.9 に示されるように、コンセプトセットを登録します。

図 E.9 の左最上部に表示されている左矢印をクリックしてコホート定義に戻ります。「+」

◀ Celecoxib new users ▶ Celecoxib

Search

Search Import

celecoxib 

Advanced Options

Column visibility Copy CSV Show 15 entries Filter: Previous 1 Next

| | Id | Code | Name | Class | RC | DRC | Domain | Vocabulary |
|---|---------|--------|-----------|------------|-------|-------|--------|------------|
|  | 1118084 | 140587 | celecoxib | Ingredient | 2,587 | 5,184 | Drug | RxNorm |

Showing 1 to 1 of 1 entries Previous 1 Next

Vocabulary

- RxNorm Extension (1376)
- NDC (1337)
- SPL (449)
- DPD (167)
- SNOMED (75)

Class

- Ingredient (7)**
 - Clinical Drug Form (5)
 - Clinical Drug Comp (5)
 - Lab Test (5)
- Medicine (1)

Domain

- Drug (3570)
- Measurement (18)
- Observation (1)
- Meas Value (1)

Standard Concept

- Non-Standard (1831)
- Standard (1292)**
- Classification (467)

Figure E.9: 成分「Celecoxib (セレコキシブ)」の標準コンセプトの選択

Initial Event (初回イベントを追加)」をクリックしてから「Add Drug Era (薬剤曝露期間を追加)」をクリックします。薬剤曝露期間基準のために以前に作成したコホート属性… (属性を追加…)」をクリックして「Add First Exposure Criteria (最初の曝露基準を追加)」を選択します。インデックス日の前に少なくとも365日の連続E.10のようになるはずです。選択基準、コホート・イグジット、コホート期間の選択は…をクリックしてコホート定義を保存することを忘れないでください。  をクリックして終了します。

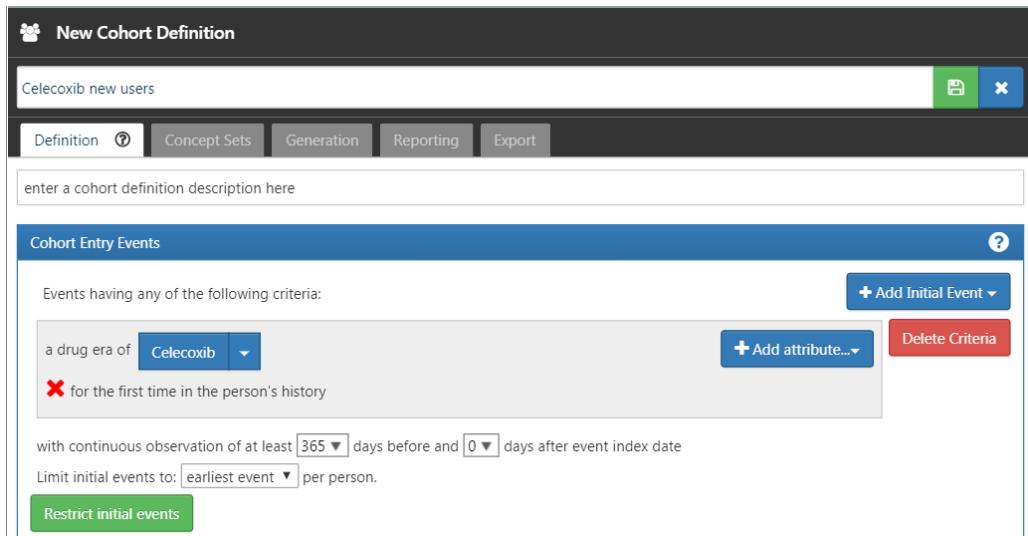


Figure E.10: セレコキシブ新規ユーザーの単純なコホート定義

コホートを定義したので、その特性評価ができます。 Characterizations をクリックし Characterization (新規特性評価)」を選択します。特性評価に意味のある名前 (例：「Celecoxib new users characterization (セレコキシブ新規ユーザーの特性評価)」) を付けます。コホート定義の右側にある「Analyses (特徴量分析)」の下で、「Import (インポート)」をクリックし、少なくとも1つの「Group Era Any Time Prior (任意の期間前の薬剤グループ曝露期間)」と「Condition Group Era Any Time Prior (任意の期間前のコンディショングループ期間)」を選択します。E.11のようになっているはずです。  をクリックして特性評価の設定を保存してください。

「Executions (実行)」タブをクリックし、1つのデータソースについて「Generate (作成)」をクリックします。作成が完了するまで時間を要する場合があります。完了する「latest results (最新の結果を表示)」をクリックできます。結果画面は、図E.12のように見えるはずで、図には例えば痛みや関節症が一般的に観察されることを示しています。

演習 11.3

 Cohort Definitions をクリックして「New cohort (新規コホート)」を作成します。コホート名 (例：「Gastrointestinal bleed (消化管出血)」) を付け、「Concept Sets (コンセプトセット)」タブに移動します。新規の「Concept Set (新しいコンセプトセット)」をクリックし、コンセプトセットに意味のある名前 (例：「Gastrointestinal bleed (消化管出血)」) を付けます。  Search モジュールを開き、「Gastrointestinal

New Characterization

Celecoxib new users characterization

Design Executions Utilities

Cohort characterization is defined as the process of generating cohort level descriptive summary statistics from person level covariate data. Summary statistics of these person level covariates may be count, mean, sd, var, min, max, median, range, and quantiles. In addition, covariates during a period may be stratified into temporal units of time for time-series analysis such as fixed intervals of time relative to cohort_start_date (e.g. every 7 days, every 30 days etc.), or in absolute calendar intervals such as calendar-week, calendar-month, calendar-quarter, calendar-year.

Cohort definitions

Import

Show 10 entries Search:

| ID | Name | Actions |
|---------|---------------------|--------------------|
| 1771701 | Celecoxib new users | Edit cohort Remove |

Showing 1 to 1 of 1 entries Previous 1 Next

Feature analyses

Import

Show 10 entries Search:

| ID | Name | Description | Actions |
|----|------------------------------------|--|---------|
| 15 | Drug Group Era Any Time Prior | One covariate per drug rolled up to ATC groups in the drug_era table overlapping with any time prior to index. | Remove |
| 27 | Condition Group Era Any Time Prior | One covariate per condition era rolled up to groups in the condition_era table overlapping with any time prior to index. | Remove |

Showing 1 to 2 of 2 entries Previous 1 Next

Figure E.11: 特性評価設定

Characterization #69

Celecoxib new users characterization

Design Executions Utilities

Executions > Reports for SYNPUF 5%

Date: 08/23/2019 12:53 PM Design: -1840810470 Results: 2 reports

Filter panel

Cohorts Analyses Domains

Celecoxib new users Condition Group Era Any Time P Condition, Drug

CONDITION / Condition Group Era Any Time Prior

Export Show 10 entries Search:

| Covariate | Explore | Concept ID | Count | Pct |
|---------------------------------------|---------|------------|-------|--------|
| Pain | Explore | 4329041 | 1,140 | 78.62% |
| Pain finding at anatomical site | Explore | 4132926 | 1,135 | 78.28% |
| Inflammation of specific body systems | Explore | 4178818 | 1,135 | 78.28% |
| Arthropathy | Explore | 73553 | 1,122 | 77.38% |

Figure E.12: 特性評価の結果

hemorrhage(消化管出血)」を検索し、一番上のコンセプトの横にある をクリックしてコンセプトを選択（図 E.13 参照）。

| Vocabulary | Id | Code | Name | Class | RC | DRC | Domain | Vocabulary |
|-----------------------|---------|----------|-----------------------------------|------------------|-----|--------|-----------|------------|
| SNOMED (17) | 192671 | 74474003 | Gastrointestinal hemorrhage | Clinical Finding | 919 | 37,144 | Condition | SNOMED |
| ICD10CM (2) | | | | | | | | |
| ICD9CM (2) | | | | | | | | |
| DRG (2) | | | | | | | | |
| NIDERT (1) | | | | | | | | |
| Class | | | | | | | | |
| Clinical Finding (17) | 4100660 | 27719009 | Lower gastrointestinal hemorrhage | Clinical Finding | 0 | 15,617 | Condition | SNOMED |
| | | | Acute gastrointestinal hemorrhage | Clinical Finding | 0 | 9,852 | Condition | SNOMED |

Figure E.13: 「Gastrointestinal hemorrhage(消化管出血)」の標準コンセプトの選択

図 E.13 の左最上部に表示されている左矢印をクリックしてコホート定義に戻ります。「Concept Sets(コンセプトセット)」タブを再度開き、消化管出血のコンセプトの横にある「Descendants(下位層)」をチェックします（図 E.14 参照）。

| Concept Id | Concept Code | Concept Name | Domain | Standard Concept Caption | Exclude | Descendants | Mapped |
|------------|--------------|-----------------------------|-----------|--------------------------|--------------------------|-------------------------------------|--------------------------|
| 192671 | 74474003 | Gastrointestinal hemorrhage | Condition | Standard | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/> |

Figure E.14: 「Gastrointestinal hemorrhage(消化管出血)」のすべての下位層を追加

「Definition(定義)」タブに戻り、「+Add Initial Event(初回イベントを追加)」をクリックして「Condition Occurrence(コンディション出現を追加)」をクリックします。先に作成したコンセプト E.15 のようになっているはずです。選択基準、コホート・イグジット、コホート期間のセクションをクリックして終了します。

これでコホートが定義されたので、発生率が計算できます。 Incidence Rates をクリックして「Analysis(新規分析)」を選択します。分析に意味の通じる名前（例：「Incidence of GI bleed after celecoxib initiation(セレコキシブ開始後の消化管出血発生率)」）を付けます。Target Cohort(ターゲットコホートを追加)」をクリックし、セレコキシブ新規ユーザーコホート

GI bleed

Definition Concept Sets Generation Reporting Export

enter a cohort definition description here

Cohort Entry Events

Events having any of the following criteria:

a condition occurrence of GI bleed ▾

+ Add Initial Event ▾

+ Add attribute... ▾ Delete Criteria

with continuous observation of at least [0] days before and [0] days after event index date

Limit initial events to: earliest event ▾ per person.

Restrict initial events

Figure E.15: 単純な消化管出血コホート定義

Outcome Cohort(アウトカムコホートを追加)」をクリックし、作成した消化管出血コホート定義がFigure E.16のようになっているはずです。をクリックして分析設定を保存してください。

Incidence of GI bleed after celecoxib initiation

Definition Concept Sets Generation Utilities

Study Cohorts

Target Cohorts

#1771701:Celecoxib new users

Add Target Cohort

Outcome Cohorts

#1771702:GI bleed

Add Outcome Cohort

Time At Risk

Time at risk defines the time window relative to the cohort start or end date with an offset to consider the person 'at risk' of the outcome.

- Time at risk starts with start date ▾ plus [0] days.
- Time at risk ends with start date ▾ plus [1095] days.

No study window defined. Add Study Window

Stratify Criteria: You can provide optional stratification criteria to the analysis that will divide the population into unique groups based on their satisfied criteria.

New stratify criteria

Please select a qualifying inclusion criteria to edit.

Figure E.16: 発生率の分析

「Generation(作成)」タブをクリックし、「Generation(作成)」をクリックします。データ(作成)」をクリックします。完了すると、計算された発生率と発生割合が表示されます (Figure E.17 参照)。

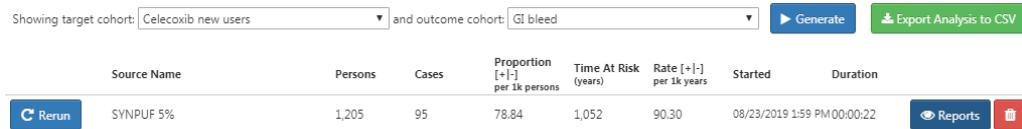


Figure E.17: 発生率の結果

E.8

演習 12.1

デフォルトの共変量セットを指定しますが、比較している2つの薬剤を含むすべての下位層を

```
library(CohortMethod)
nsaids <- c(1118084, 1124300) #
covSettings <- createDefaultCovariateSettings(
  excludedCovariateConceptIds = nsaids,
  addDescendantsToExclude = TRUE)

#
cmData <- getDbCohortMethodData(
  connectionDetails = connectionDetails,
  cdmDatabaseSchema = "main",
  targetId = 1,
  comparatorId = 2,
  outcomeIds = 3,
  exposureDatabaseSchema = "main",
  exposureTable = "cohort",
  outcomeDatabaseSchema = "main",
  outcomeTable = "cohort",
  covariateSettings = covSettings)
summary(cmData)
```

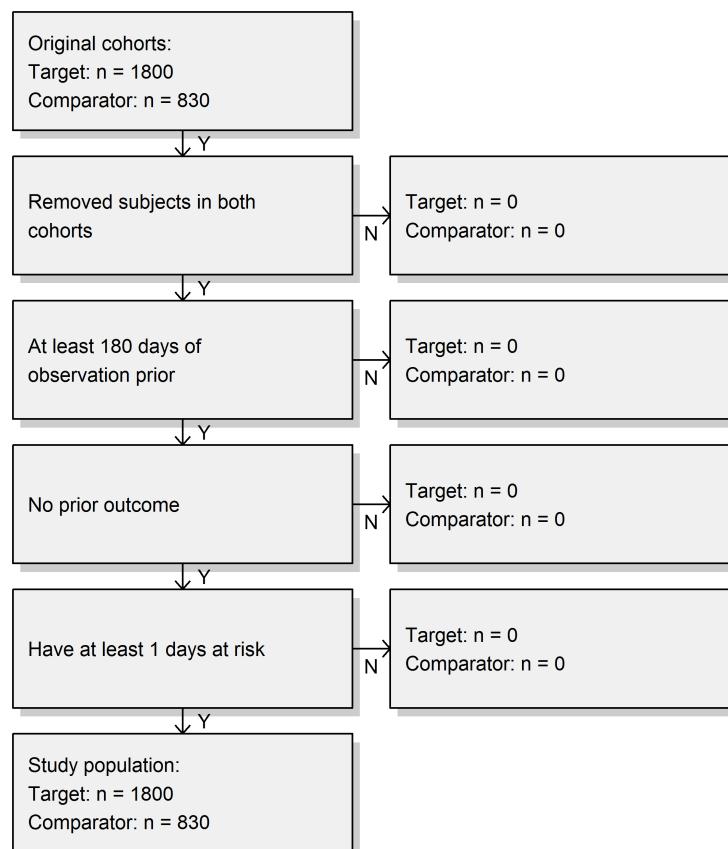
```
## CohortMethodData
##
##      ID 1
##      ID 2
##      ID 3
##
##      1800
##      830
##
##      3       479      479
##
```

```
##  
##      389  
##     26923
```

演習 12.2

仕様に従って研究対象集団を作成し、脱落を示す図を出力します：

```
studyPop <- createStudyPopulation(  
  cohortMethodData = cmData,  
  outcomeId = 3,  
  washoutPeriod = 180,  
  removeDuplicateSubjects = "remove all",  
  removeSubjectsWithPriorOutcome = TRUE,  
  riskWindowStart = 0,  
  startAnchor = "cohort start",  
  riskWindowEnd = 99999)  
drawAttritionDiagram(studyPop)
```



元のコホートと比較して研究対象が失われなかつたことが見てとれます。なぜなら、ここで使

演習 12.3

Cox回帰モデルを使用して単純なアウトカムをフィットさせます：

```
model <- fitOutcomeModel(population = studyPop,
                           modelType = "cox")
model

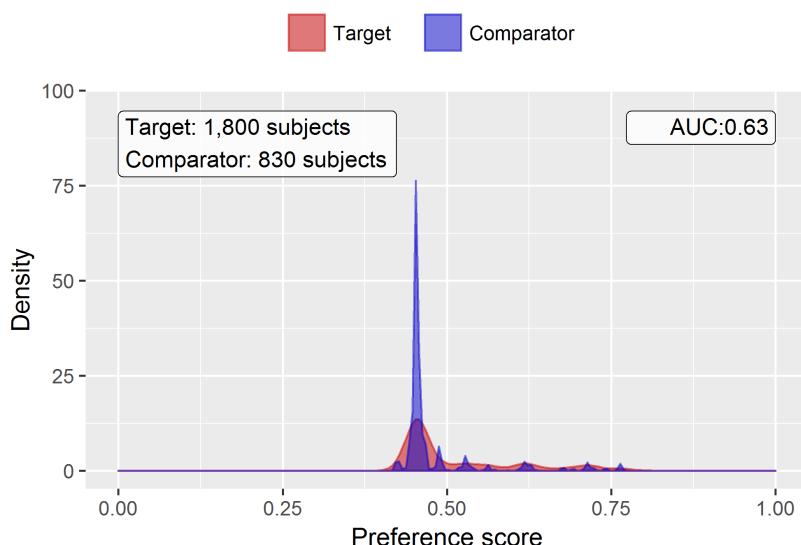
##      COX
##    FALSE
##    FALSE
##      FALSE
##    OK
##
##      .95     .95
##    1.34612   1.10065   1.65741       0.29723           0.1044
```

セレコキシブのユーザーとジクロフェナクのユーザーが交換可能でない可能性が高く、ベース

演習 12.4

抽出したすべての共変量を使用し、研究集団に対して傾向スコアモデルをフィットさせます。

```
ps <- createPs(cohortMethodData = cmData,
                population = studyPop)
plotPs(ps, showCountsLabel = TRUE, showAucLabel = TRUE)
```

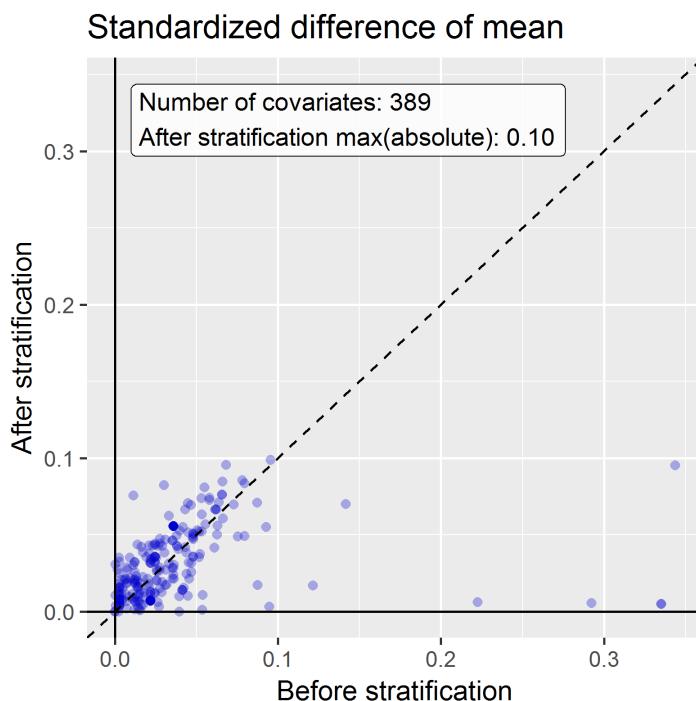


この分布には、いくつかのスパイクがあり少し奇妙に見えることに注意してください。この傾向スコアモデルは0.63のAUCを達成し、ターゲットのコホートと比較群コホートの間には

演習 12.5

傾向スコアに基づいて集団を層別化し、層別化前後の共変量バランスを計算します：

```
strataPop <- stratifyByPs(ps, number0fStrata = 5)
bal <- computeCovariateBalance(strataPop, cmData)
plotCovariateBalanceScatterPlot(bal,
                                showCovariateCountLabel = TRUE,
                                showMaxLabel = TRUE,
                                beforeLabel = " ", 
                                afterLabel = " ")
```



さまざまなベースライン共変量が、層別化前（x軸）には大きな (>0.3) 標準化平均差(standardized mean difference, SMD)を示していることがわかります。層別化後のバランスが向上し、最大値は0.1です。

演習 12.6

傾向スコアによる層別化し、Cox回帰モデルをフィットさせます：

```

##      COX
##      TRUE
##      FALSE
##      (IPTW)  FALSE
##      OK
##
##          .95     .95
##      1.13211   0.92132   1.40008           0.12409      0.1068

```

調整後の推定値が未調整の推定値より低くなり、95%信頼区間が1を含むようになったことが

E.9

演習 13.1

共変量設定のセットを指定し、データベースからデータを抽出するためにgetPlpData関数を使

```
## plpData
##
## At risk      ID -1
##           ID 3
##
## 2630
##
##
## 3          479    479
##
## 245
##      54079
```

演習 13.2

関心の対象であるアウトカムの研究対象集団を（この場合は抽出したデータに対して唯一

```
population <- createStudyPopulation(plpData = plpData,
                                      outcomeId = 3,
                                      washoutPeriod = 364,
                                      firstExposureOnly = FALSE,
                                      removeSubjectsWithPriorOutcome = TRUE,
                                      priorOutcomeLookback = 9999,
                                      riskWindowStart = 1,
                                      riskWindowEnd = 365,
                                      addExposureDaysToStart = FALSE,
                                      addExposureDaysToEnd = FALSE,
                                      minTimeAtRisk = 364,
                                      requireTimeAtRisk = TRUE,
                                      includeAllOutcomes = TRUE)

nrow(population)
```

```
## [1] 2578
```

この場合、アウトカムがすでに出現した対象を除外することと364日以上のリスク期間を

演習 13.3

LASSOモデルを実行するために、まずモデル設定オブジェクトを作成し、その後runPlp

```
lassoModel <- setLassoLogisticRegression(seed = 0)

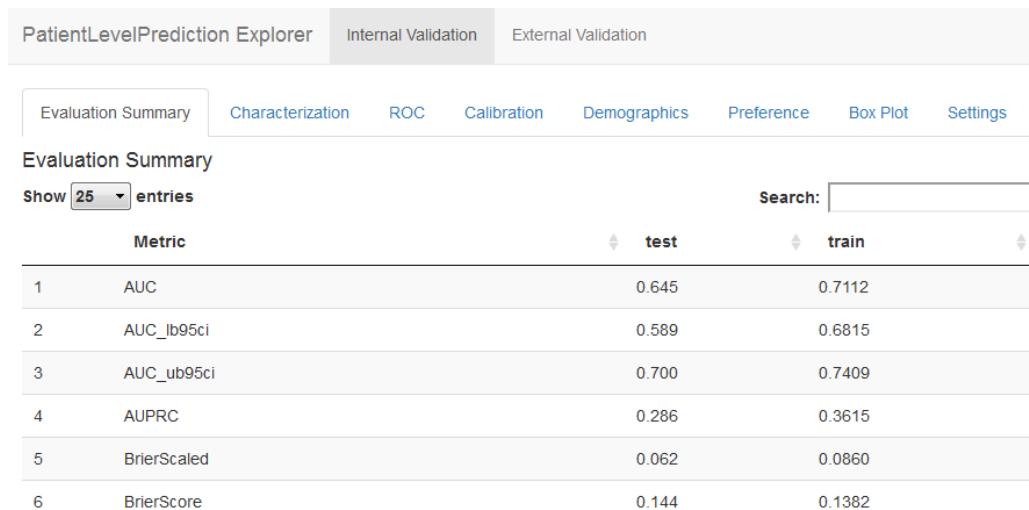
lassoResults <- runPlp(population = population,
                        plpData = plpData,
```

```
modelSettings = lassoModel,
testSplit = 'person',
testFraction = 0.25,
nfold = 2,
splitSeed = 0)
```

この例では、LASSOのクロスバリデーションと訓練用・テスト用データ分割の両方に対して乱数を固定してShinyアプリを使用して結果を表示することができます：

```
viewPlp(lassoResults)
```

これにより、図 E.18に示されるようにアプリが起動されます。ここで、テスト用データセット



The screenshot shows the 'PatientLevelPrediction Explorer' Shiny application interface. At the top, there are three tabs: 'PatientLevelPrediction Explorer' (selected), 'Internal Validation', and 'External Validation'. Below the tabs is a navigation bar with links: 'Evaluation Summary' (selected), 'Characterization', 'ROC', 'Calibration', 'Demographics', 'Preference', 'Box Plot', and 'Settings'. The main content area is titled 'Evaluation Summary' and contains a table. The table has two columns: 'Metric' and 'test' (with a 'train' header). The table rows are numbered 1 to 6 and show the following data:

| Metric | test | train |
|---------------|-------|--------|
| 1 AUC | 0.645 | 0.7112 |
| 2 AUC_lb95ci | 0.589 | 0.6815 |
| 3 AUC_ub95ci | 0.700 | 0.7409 |
| 4 AUPRC | 0.286 | 0.3615 |
| 5 BrierScaled | 0.062 | 0.0860 |
| 6 BrierScore | 0.144 | 0.1382 |

Figure E.18: Shinyアプリによる患者レベルの予測の表示

E.10

演習 15.1

ACHILLESを実行するには：

```
library(ACHILLES)
result <- achilles(connectionDetails,
                      cdmDatabaseSchema = "main",
                      resultsDatabaseSchema = "main",
                      sourceName = "Eunomia",
                      cdmVersion = "5.3.0")
```

演習 15.2

データ品質ダッシュボード(Data Quality Dashboard, DQD)を実行するには:

```
DataQualityDashboard::executeDqChecks(  
    connectionDetails,  
    cdmDatabaseSchema = "main",  
    resultsDatabaseSchema = "main",  
    cdmSourceName = "Eunomia",  
    outputFolder = "C:/dataQualityExample")
```

演習 15.3

データ品質チェックのリストを見るには:

```
DataQualityDashboard::viewDqDashboard(  
    "C:/dataQualityExample/Eunomia/results_Eunomia.json")
```

E.11

Bibliography

- Allison, D. B., Brown, A. W., George, B. J., and Kaiser, K. A. (2016). Reproducibility: A tragedy of errors. *Nature*, 530(7588):27–29.
- Arnold, B. F., Ercumen, A., Benjamin-Chung, J., and Colford, J. M. (2016). Brief Report: Negative Controls to Detect Selection Bias and Measurement Bias in Epidemiologic Studies. *Epidemiology*, 27(5):637–641.
- Austin, P. C. (2011). Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. *Pharmaceutical statistics*, 10(2):150–161.
- Banda, J. M., Halpern, Y., Sontag, D., and Shah, N. H. (2017). Electronic phenotyping with APHRODITE and the Observational Health Sciences and Informatics (OHDSI) data network. *AMIA Jt Summits Transl Sci Proc*, 2017:48–57.
- Boland, M. R., Parhi, P., Li, L., Miotto, R., Carroll, R., Iqbal, U., Nguyen, P. A., Schuemie, M., You, S. C., Smith, D., Mooney, S., Ryan, P., Li, Y. J., Park, R. W., Denny, J., Dudley, J. T., Hripcsak, G., Gentine, P., and Tatonetti, N. P. (2017). Uncovering exposures responsible for birth season - disease effects: a global study. *J Am Med Inform Assoc*.
- Botsis, T., Hartvigsen, G., Chen, F., and Weng, C. (2010). Secondary use of ehr: data quality issues and informatics opportunities. *Summit on Translational Bioinformatics*, 2010:1.
- Byrd, J. B., Adam, A., and Brown, N. J. (2006). Angiotensin-converting enzyme inhibitor-associated angioedema. *Immunol Allergy Clin North Am*, 26(4):725–737.
- Callahan, T. J., Bauck, A. E., Bertoch, D., Brown, J., Khare, R., Ryan, P. B., Staab, J., Zozus, M. N., and Kahn, M. G. (2017). A comparison of data quality assessment checks in six data sharing networks. *eGEMS*, 5(1).
- Cepeda, M. S., Reps, J., Fife, D., Blacketer, C., Stang, P., and Ryan, P. (2018). Finding treatment-resistant depression in real-world data: How a data-

- driven approach compares with expert-based heuristics. *Depress Anxiety*, 35(3):220–228.
- Chen, X., Dallmeier-Tiessen, S., Dasler, R., Feger, S., Fokianos, P., Gonzalez, J. B., Hirvonsalo, H., Kousidis, D., Lavasa, A., Mele, S., Rodriguez, D. R., Šimko, T., Smith, T., Trisovic, A., Trzcinska, A., Tsanaktsidis, I., Zimmermann, M., Cranmer, K., Heinrich, L., Watts, G., Hildreth, M., Iglesias, L. L., Lassila-Perini, K., and Neubert, S. (2018). Open is not enough. *Nature Physics*, 15(2):113–119.
- Cicardi, M., Zingale, L. C., Bergamaschini, L., and Agostoni, A. (2004). Angioedema associated with angiotensin-converting enzyme inhibitor use: outcome after switching to a different treatment. *Arch. Intern. Med.*, 164(8):910–913.
- Dasu, T. and Johnson, T. (2003). Exploratory data mining and data cleaning, volume 479. John Wiley & Sons.
- Defalco, F. J., Ryan, P. B., and Soledad Cepeda, M. (2013). Applying standardized drug terminologies to observational healthcare databases: a case study on opioid exposure. *Health Serv Outcomes Res Methodol*, 13(1):58–67.
- DerSimonian, R. and Laird, N. (1986). Meta-analysis in clinical trials. *Control Clin Trials*, 7(3):177–188.
- Duke, J. D., Ryan, P. B., Suchard, M. A., Hripcak, G., Jin, P., Reich, C., Schwalm, M. S., Khoma, Y., Wu, Y., Xu, H., Shah, N. H., Banda, J. M., and Schuemie, M. J. (2017). Risk of angioedema associated with levetiracetam compared with phenytoin: Findings of the observational health data sciences and informatics research network. *Epilepsia*, 58(8):e101–e106.
- Farrington, C. P. (1995). Relative incidence estimation from case series for vaccine safety evaluation. *Biometrics*, 51(1):228–235.
- Farrington, C. P., Anaya-Izquierdo, K., Whitaker, H. J., Hocine, M. N., Douglas, I., and Smeeth, L. (2011). Self-controlled case series analysis with event-dependent observation periods. *Journal of the American Statistical Association*, 106(494):417–426.
- Fuller, W. A. (2009). Measurement error models, volume 305. John Wiley & Sons.
- Garza, M., Del Fiol, G., Tenenbaum, J., Walden, A., and Zozus, M. N. (2016). Evaluating common data models for use with a longitudinal community registry. *J Biomed Inform*, 64:333–341.
- Hernan, M. A., Hernandez-Diaz, S., Werler, M. M., and Mitchell, A. A. (2002). Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology. *Am. J. Epidemiol.*, 155(2):176–184.

- Hernan, M. A. and Robins, J. M. (2016). Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. *Am. J. Epidemiol.*, 183(8):758–764.
- Hersh, W. R., Weiner, M. G., Embi, P. J., Logan, J. R., Payne, P. R., Bernstam, E. V., Lehmann, H. P., Hripcsak, G., Hartzog, T. H., Cimino, J. J., et al. (2013). Caveats for the use of operational electronic health record data in comparative effectiveness research. *Medical care*, 51(8 0 3):S30.
- Higgins, J. P., Thompson, S. G., Deeks, J. J., and Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ*, 327(7414):557–560.
- Hill, A. B. (1965). THE ENVIRONMENT AND DISEASE: ASSOCIATION OR CAUSATION? *Proc. R. Soc. Med.*, 58:295–300.
- Hripcsak, G. and Albers, D. J. (2017). High-fidelity phenotyping: richness and freedom from bias. *J Am Med Inform Assoc*.
- Hripcsak, G., Duke, J. D., Shah, N. H., Reich, C. G., Huser, V., Schuemie, M. J., Suchard, M. A., Park, R. W., Wong, I. C. K., Rijnbeek, P. R., van der Lei, J., Pratt, N., Norén, G. N., Li, Y.-C., Stang, P. E., Madigan, D., and Ryan, P. B. (2015). Observational Health Data Sciences and Informatics (OHDSI): Opportunities for Observational Researchers. *Studies in health technology and informatics*, 216:574–578.
- Hripcsak, G., Levine, M. E., Shang, N., and Ryan, P. B. (2018). Effect of vocabulary mapping for conditions on phenotype cohorts. *J Am Med Inform Assoc*, 25(12):1618–1625.
- Hripcsak, G., Ryan, P. B., Duke, J. D., Shah, N. H., Park, R. W., Huser, V., Suchard, M. A., Schuemie, M. J., DeFalco, F. J., Perotte, A., Banda, J. M., Reich, C. G., Schilling, L. M., Matheny, M. E., Meeker, D., Pratt, N., and Madigan, D. (2016). Characterizing treatment pathways at scale using the OHDSI network. *Proceedings of the National Academy of Sciences*, 113(27):7329–7336.
- Hripcsak, G., Shang, N., Peissig, P. L., Rasmussen, L. V., Liu, C., Benoit, B., Carroll, R. J., Carrell, D. S., Denny, J. C., Dikilitas, O., Gainer, V. S., Marie Howell, K., Klann, J. G., Kullo, I. J., Lingren, T., Mentch, F. D., Murphy, S. N., Natarajan, K., Pacheco, J. A., Wei, W. Q., Wiley, K., and Weng, C. (2019). Facilitating phenotype transfer using a common data model. *J Biomed Inform*, page 103253.
- Huser, V., DeFalco, F. J., Schuemie, M., Ryan, P. B., Shang, N., Velez, M., Park, R. W., Boyce, R. D., Duke, J., Khare, R., Utidjian, L., and Bailey, C. (2016). Multisite Evaluation of a Data Quality Tool for Patient-Level Clinical Data Sets. *EGEMS (Washington, DC)*, 4(1):1239.
- Huser, V., Kahn, M. G., Brown, J. S., and Gouripeddi, R. (2018). Methods

- for examining data quality in healthcare integrated data repositories. Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing, 23:628–633.
- Johnston, S. S., Morton, J. M., Kalsekar, I., Ammann, E. M., Hsiao, C. W., and Reps, J. (2019). Using Machine Learning Applied to Real-World Healthcare Data for Predictive Analytics: An Applied Example in Bariatric Surgery. *Value Health*, 22(5):580–586.
- Kahn, M. G., Brown, J. S., Chun, A. T., Davidson, B. N., Meeker, D., Ryan, P. B., Schilling, L. M., Weiskopf, N. G., Williams, A. E., and Zozus, M. N. (2015). Transparent reporting of data quality in distributed data networks. EGEMS (Washington, DC), 3(1):1052.
- Kahn, M. G., Callahan, T. J., Barnard, J., Bauck, A. E., Brown, J., Davidson, B. N., Estiri, H., Goerg, C., Holve, E., Johnson, S. G., Liaw, S.-T., Hamilton-Lopez, M., Meeker, D., Ong, T. C., Ryan, P. B., Shang, N., Weiskopf, N. G., Weng, C., Zozus, M. N., and Schilling, L. (2016). A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data. EGEMS (Washington, DC), 4(1):1244.
- Kahn, M. G., Raebel, M. A., Glanz, J. M., Riedlinger, K., and Steiner, J. F. (2012). A pragmatic framework for single-site and multisite data quality assessment in electronic health record-based clinical research. *Medical care*, 50.
- Liaw, S.-T., Rahimi, A., Ray, P., Taggart, J., Dennis, S., de Lusignan, S., Jalaludin, B., Yeo, A., and Talaei-Khoei, A. (2013). Towards an ontology for data quality in integrated chronic disease management: a realist review of the literature. *International journal of medical informatics*, 82(1):10–24.
- Lipsitch, M., Tchetgen Tchetgen, E., and Cohen, T. (2010). Negative controls: a tool for detecting confounding and bias in observational studies. *Epidemiology*, 21(3):383–388.
- MacLure, M. (1991). The case-crossover design: a method for studying transient effects on the risk of acute events. *Am. J. Epidemiol.*, 133(2):144–153.
- Madigan, D., Ryan, P. B., and Schuemie, M. (2013a). Does design matter? Systematic evaluation of the impact of analytical choices on effect estimates in observational studies. *Ther Adv Drug Saf*, 4(2):53–62.
- Madigan, D., Ryan, P. B., Schuemie, M., Stang, P. E., Overhage, J. M., Hartzema, A. G., Suchard, M. A., DuMouchel, W., and Berlin, J. A. (2013b). Evaluating the impact of database heterogeneity on observational study results. *Am. J. Epidemiol.*, 178(4):645–651.
- Magid, D. J., Shetterly, S. M., Margolis, K. L., Tavel, H. M., O’ Connor, P. J., Selby, J. V., and Ho, P. M. (2010). Comparative effectiveness of angiotensin-

- converting enzyme inhibitors versus beta-blockers as second-line therapy for hypertension. *Circ Cardiovasc Qual Outcomes*, 3(5):453–458.
- Makadia, R. and Ryan, P. B. (2014). Transforming the Premier Perspective Hospital Database into the Observational Medical Outcomes Partnership (OMOP) Common Data Model. *EGEMS (Wash DC)*, 2(1):1110.
- Martin, R. C. (2008). *Clean Code: A Handbook of Agile Software Craftsmanship*. Prentice Hall PTR, Upper Saddle River, NJ, USA, 1 edition.
- Matcho, A., Ryan, P., Fife, D., and Reich, C. (2014). Fidelity assessment of a clinical practice research datalink conversion to the OMOP common data model. *Drug Saf*, 37(11):945–959.
- Noren, G. N., Caster, O., Juhlin, K., and Lindquist, M. (2014). Zoo or savannah? Choice of training ground for evidence-based pharmacovigilance. *Drug Saf*, 37(9):655–659.
- Norman, J. L., Holmes, W. L., Bell, W. A., and Finks, S. W. (2013). Life-threatening ACE inhibitor-induced angioedema after eleven years on lisinopril. *J Pharm Pract*, 26(4):382–388.
- Oliveira, J. L., Trifan, A., and Silva, L. A. B. (2019). EMIF catalogue: A collaborative platform for sharing and reusing biomedical data. *International Journal of Medical Informatics*, 126:35–45.
- Olsen, L., Aisner, D., McGinnis, J. M., et al. (2007). The learning healthcare system: workshop summary. *Natl Academ Pr*.
- O’ Mara, N. B. and O’ Mara, E. M. (1996). Delayed onset of angioedema with angiotensin-converting enzyme inhibitors: case report and review of the literature. *Pharmacotherapy*, 16(4):675–679.
- Overhage, J. M., Ryan, P. B., Reich, C. G., Hartzema, A. G., and Stang, P. E. (2012). Validation of a common data model for active safety surveillance research. *J Am Med Inform Assoc*, 19(1):54–60.
- Perkins, N. J., Cole, S. R., Harel, O., Tchetgen Tchetgen, E. J., Sun, B., Mitchell, E. M., and Schisterman, E. F. (2017). Principled approaches to missing data in epidemiologic studies. *American journal of epidemiology*, 187(3):568–575.
- Powers, B. J., Coeytaux, R. R., Dolor, R. J., Hasselblad, V., Patel, U. D., Yancy, W. S., Gray, R. N., Irvine, R. J., Kendrick, A. S., and Sanders, G. D. (2012). Updated report on comparative effectiveness of ACE inhibitors, ARBs, and direct renin inhibitors for patients with essential hypertension: much more data, little new information. *J Gen Intern Med*, 27(6):716–729.
- Prasad, V. and Jena, A. B. (2013). Prespecified falsification end points: can they validate true observational associations? *JAMA*, 309(3):241–242.

- Ramcharran, D., Qiu, H., Schuemie, M. J., and Ryan, P. B. (2017). Atypical Antipsychotics and the Risk of Falls and Fractures Among Older Adults: An Emulation Analysis and an Evaluation of Additional Confounding Control Strategies. *J Clin Psychopharmacol*, 37(2):162–168.
- Rassen, J. A., Shelat, A. A., Myers, J., Glynn, R. J., Rothman, K. J., and Schneeweiss, S. (2012). One-to-many propensity score matching in cohort studies. *Pharmacoepidemiol Drug Saf*, 21 Suppl 2:69–80.
- Reps, J. M., Rijnbeek, P. R., and Ryan, P. B. (2019). Identifying the DEAD: Development and Validation of a Patient-Level Model to Predict Death Status in Population-Level Claims Data. *Drug Saf*.
- Reps, J. M., Schuemie, M. J., Suchard, M. A., Ryan, P. B., and Rijnbeek, P. R. (2018). Design and implementation of a standardized framework to generate and evaluate patient-level prediction models using observational healthcare data. *Journal of the American Medical Informatics Association*, 25(8):969–975.
- Roebuck, K. (2012). Data quality: high-impact strategies-what you need to know: definitions, adoptions, impact, benefits, maturity, vendors. Emereo Publishing.
- Rosenbaum, P. (2005). Sensitivity Analysis in Observational Studies. American Cancer Society.
- Rosenbaum, P. and Rubin, D. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70:41–55.
- Rubbo, B., Fitzpatrick, N. K., Denaxas, S., Daskalopoulou, M., Yu, N., Patel, R. S., Hemingway, H., Danesh, J., Allen, N., Atkinson, M., Blaveri, E., Branen, R., Brayne, C., Brophy, S., Chaturvedi, N., Collins, R., deLusignan, S., Denaxas, S., Desai, P., Eastwood, S., Gallacher, J., Hemingway, H., Hotopf, M., Landray, M., Lyons, R., O’ Neil, T., Pringle, M., Sprosen, T., Strachan, D., Sudlow, C., Sullivan, F., Zhang, Q., and Flraig, R. (2015). Use of electronic health records to ascertain, validate and phenotype acute myocardial infarction: A systematic review and recommendations. *Int. J. Cardiol*, 187:705–711.
- Rubin, D. B. (2001). Using propensity scores to help design observational studies: application to the tobacco litigation. *Health Services and Outcomes Research Methodology*, 2(3-4):169–188.
- Ryan, P. B., Buse, J. B., Schuemie, M. J., DeFalco, F., Yuan, Z., Stang, P. E., Berlin, J. A., and Rosenthal, N. (2018). Comparative effectiveness of canagliflozin, SGLT2 inhibitors and non-SGLT2 inhibitors on the risk of hospitalization for heart failure and amputation in patients with type 2 diabetes mellitus: A real-world meta-analysis of 4 observational databases (OBSERVE-4D). *Diabetes Obes Metab*, 20(11):2585–2597.

- Ryan, P. B., Madigan, D., Stang, P. E., Overhage, J. M., Racoosin, J. A., and Hartzema, A. G. (2012). Empirical assessment of methods for risk identification in healthcare data: results from the experiments of the Observational Medical Outcomes Partnership. *Stat Med*, 31(30):4401–4415.
- Ryan, P. B., Schuemie, M. J., and Madigan, D. (2013a). Empirical performance of a self-controlled cohort method: lessons for developing a risk identification and analysis system. *Drug Saf*, 36 Suppl 1:95–106.
- Ryan, P. B., Schuemie, M. J., Ramcharran, D., and Stang, P. E. (2017). Atypical Antipsychotics and the Risks of Acute Kidney Injury and Related Outcomes Among Older Adults: A Replication Analysis and an Evaluation of Adapted Confounding Control Strategies. *Drugs Aging*, 34(3):211–219.
- Ryan, P. B., Stang, P. E., Overhage, J. M., Suchard, M. A., Hartzema, A. G., DuMouchel, W., Reich, C. G., Schuemie, M. J., and Madigan, D. (2013b). A comparison of the empirical performance of methods for a risk identification system. *Drug Saf*, 36 Suppl 1:S143–158.
- Sabroe, R. A. and Black, A. K. (1997). Angiotensin-converting enzyme (ACE) inhibitors and angio-oedema. *Br. J. Dermatol.*, 136(2):153–158.
- Schneeweiss, S. (2018). Automated data-adaptive analytics for electronic healthcare data to study causal treatment effects. *Clin Epidemiol*, 10:771–788.
- Schuemie, M. J., Hripcak, G., Ryan, P. B., Madigan, D., and Suchard, M. A. (2016). Robust empirical calibration of p-values using observational data. *Stat Med*, 35(22):3883–3888.
- Schuemie, M. J., Hripcak, G., Ryan, P. B., Madigan, D., and Suchard, M. A. (2018a). Empirical confidence interval calibration for population-level effect estimation studies in observational healthcare data. *Proc. Natl. Acad. Sci. U.S.A.*, 115(11):2571–2577.
- Schuemie, M. J., Ryan, P. B., DuMouchel, W., Suchard, M. A., and Madigan, D. (2014). Interpreting observational studies: why empirical calibration is needed to correct p-values. *Stat Med*, 33(2):209–218.
- Schuemie, M. J., Ryan, P. B., Hripcak, G., Madigan, D., and Suchard, M. A. (2018b). Improving reproducibility by using high-throughput observational studies with empirical calibration. *Philos Trans A Math Phys Eng Sci*, 376(2128).
- Sherman, R. E., Anderson, S. A., Dal Pan, G. J., Gray, G. W., Gross, T., Hunter, N. L., LaVange, L., Marinac-Dabic, D., Marks, P. W., Robb, M. A., et al. (2016). Real-world evidence—what is it and what can it tell us. *N Engl J Med*, 375(23):2293–2297.

- Simpson, S. E., Madigan, D., Zorych, I., Schuemie, M. J., Ryan, P. B., and Suchard, M. A. (2013). Multiple self-controlled case series for large-scale longitudinal observational databases. *Biometrics*, 69(4):893–902.
- Slater, E. E., Merrill, D. D., Guess, H. A., Roylance, P. J., Cooper, W. D., Inman, W. H. W., and Ewan, P. W. (1988). Clinical Profile of Angioedema Associated With Angiotensin Converting-Enzyme Inhibition. *JAMA*, 260(7):967–970.
- Stang, P. E., Ryan, P. B., Racoosin, J. A., Overhage, J. M., Hartzema, A. G., Reich, C., Welebob, E., Scarneccchia, T., and Woodcock, J. (2010). Advancing the science for active surveillance: rationale and design for the Observational Medical Outcomes Partnership. *Ann. Intern. Med.*, 153(9):600–606.
- Suchard, M. A., Simpson, S. E., Zorych, I., Ryan, P. B., and Madigan, D. (2013). Massive parallelization of serial inference algorithms for a complex generalized linear model. *ACM Trans. Model. Comput. Simul.*, 23(1):10:1–10:17.
- Suissa, S. (1995). The case-time-control design. *Epidemiology*, 6(3):248–253.
- Swerdel, J. N., Hripcak, G., and Ryan, P. B. (2019). PheEvaluator: Development and Evaluation of a Phenotype Algorithm Evaluator. *J Biomed Inform*, page 103258.
- Thompson, T. and Frable, M. A. (1993). Drug-induced, life-threatening angioedema revisited. *Laryngoscope*, 103(1 Pt 1):10–12.
- Tian, Y., Schuemie, M. J., and Suchard, M. A. (2018). Evaluating large-scale propensity score performance through real-world and synthetic data experiments. *Int J Epidemiol*, 47(6):2005–2014.
- Toh, S., Reichman, M. E., Houstoun, M., Ross Southworth, M., Ding, X., Hernandez, A. F., Levenson, M., Li, L., McCloskey, C., Shoaibi, A., Wu, E., Zornberg, G., and Hennessy, S. (2012). Comparative risk for angioedema associated with the use of drugs that target the renin-angiotensin-aldosterone system. *Arch. Intern. Med.*, 172(20):1582–1589.
- van der Lei, J. (1991). Use and abuse of computer-stored medical records. *Methods of information in medicine*, 30(02):79–80.
- Vandenbroucke, J. P. and Pearce, N. (2012). Case-control studies: basic concepts. *Int J Epidemiol*, 41(5):1480–1489.
- Vashisht, R., Jung, K., Schuler, A., Banda, J. M., Park, R. W., Jin, S., Li, L., Dudley, J. T., Johnson, K. W., Shervey, M. M., Xu, H., Wu, Y., Natrajan, K., Hripcak, G., Jin, P., Van Zandt, M., Reckard, A., Reich, C. G., Weaver, J., Schuemie, M. J., Ryan, P. B., Callahan, A., and Shah, N. H. (2018). Association of Hemoglobin A1c Levels With Use of Sulfonylureas, Dipeptidyl Peptidase 4 Inhibitors, and Thiazolidinediones in Patients With Type 2 Di-

- abetes Treated With Metformin: Analysis From the Observational Health Data Sciences and Informatics Initiative. *JAMA Netw Open*, 1(4):e181755.
- von Elm, E., Altman, D. G., Egger, M., Pocock, S. J., Gøtzsche, P. C., and Vandebroucke, J. P. (2008). The strengthening the reporting of observational studies in epidemiology (strobe) statement: guidelines for reporting observational studies. *Journal of Clinical Epidemiology*, 61(4):344 – 349.
- Voss, E. A., Boyce, R. D., Ryan, P. B., van der Lei, J., Rijnbeek, P. R., and Schuemie, M. J. (2016). Accuracy of an Automated Knowledge Base for Identifying Drug Adverse Reactions. *J Biomed Inform*.
- Voss, E. A., Ma, Q., and Ryan, P. B. (2015a). The impact of standardizing the definition of visits on the consistency of multi-database observational health research. *BMC Med Res Methodol*, 15:13.
- Voss, E. A., Makadia, R., Matcho, A., Ma, Q., Knoll, C., Schuemie, M., DeFalco, F. J., Londhe, A., Zhu, V., and Ryan, P. B. (2015b). Feasibility and utility of applications of the common data model to multiple, disparate observational health databases. *J Am Med Inform Assoc*, 22(3):553–564.
- Walker, A. M., Patrick, A. R., Lauer, M. S., Hornbrook, M. C., Marin, M. G., Platt, R., Roger, V. L., Stang, P., and Schneeweiss, S. (2013). A tool for assessing the feasibility of comparative effectiveness research. *Comp Eff Res*, 3:11–20.
- Wang, Y., Desai, M., Ryan, P. B., DeFalco, F. J., Schuemie, M. J., Stang, P. E., Berlin, J. A., and Yuan, Z. (2017). Incidence of diabetic ketoacidosis among patients with type 2 diabetes mellitus treated with SGLT2 inhibitors and other antihyperglycemic agents. *Diabetes Res. Clin. Pract.*, 128:83–90.
- Weinstein, R. B., Ryan, P., Berlin, J. A., Matcho, A., Schuemie, M., Swerdel, J., Patel, K., and Fife, D. (2017). Channeling in the Use of Nonprescription Paracetamol and Ibuprofen in an Electronic Medical Records Database: Evidence and Implications. *Drug Saf*, 40(12):1279–1292.
- Weiskopf, N. G. and Weng, C. (2013). Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. *Journal of the American Medical Informatics Association: JAMIA*, 20(1):144–151.
- Whelton, P. K., Carey, R. M., Aronow, W. S., Casey, D. E., Collins, K. J., Dennison Himmelfarb, C., DePalma, S. M., Gidding, S., Jamerson, K. A., Jones, D. W., MacLaughlin, E. J., Muntner, P., Ovbiagele, B., Smith, S. C., Spencer, C. C., Stafford, R. S., Taler, S. J., Thomas, R. J., Williams, K. A., Williamson, J. D., and Wright, J. T. (2018). 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA Guideline for the Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults: Executive Summary: A Report of the American

- College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *Circulation*, 138(17):e426–e483.
- Whitaker, H. J., Farrington, C. P., Spiessens, B., and Musonda, P. (2006). Tutorial in biostatistics: the self-controlled case series method. *Stat Med*, 25(10):1768–1797.
- Who, A. (2013). Global brief on hypertension. World Health Organization.
- Wickham, H. (2015). R Packages. O’ Reilly Media, Inc., 1st edition.
- Wikipedia (2019a). Open science — Wikipedia, the free encyclopedia. <http://en.wikipedia.org/w/index.php?title=Open%20science&oldid=900178688>. [Online; accessed 24-June-2019].
- Wikipedia (2019b). Science 2.0 — Wikipedia, the free encyclopedia. <http://en.wikipedia.org/w/index.php?title=Science%202.0&oldid=887565958>. [Online; accessed 09-July-2019].
- Wikiquote (2019). Ronald fisher — wikiquote,. [Online; accessed 2-August-2019].
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J. W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., Gray, A. J., Groth, P., Goble, C., Grethe, J. S., Heringa, J., ’t Hoen, P. A., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S. J., Martone, M. E., Mons, A., Packer, A. L., Person, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S. A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., and Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data*, 3:160018.
- Yoon, D., Ahn, E. K., Park, M. Y., Cho, S. Y., Ryan, P., Schuemie, M. J., Shin, D., Park, H., and Park, R. W. (2016). Conversion and Data Quality Assessment of Electronic Health Record Data at a Korean Tertiary Teaching Hospital to a Common Data Model for Distributed Network Research. *Healthc Inform Res*, 22(1):54–58.
- Yuan, Z., DeFalco, F. J., Ryan, P. B., Schuemie, M. J., Stang, P. E., Berlin, J. A., Desai, M., and Rosenthal, N. (2018). Risk of lower extremity amputations in people with type 2 diabetes mellitus treated with sodium-glucose co-transporter-2 inhibitors in the USA: A retrospective cohort study. *Diabetes Obes Metab*, 20(3):582–589.
- Zaadstra, B. M., Chorus, A. M., van Buuren, S., Kalsbeek, H., and van Noort, J. M. (2008). Selective association of multiple sclerosis with infectious mononucleosis. *Mult. Scler.*, 14(3):307–313.

- Zaman, M. A., Oparil, S., and Calhoun, D. A. (2002). Drugs targeting the renin-angiotensin-aldosterone system. *Nat Rev Drug Discov*, 1(8):621–636.

Index

- ACE阻害薬, 185
ACHILLES, 220
adaboost, 182
agnostic SQL, see SqlRender
analysis implementation, 73
APHRODITE, 110
arachne, 272
ATHENA, 37
ATLAS, 74, 187
 characterization features, 135
 インストール, 76
 コホートパスウェイ, 75
 コホート定義, 75
 コンセプトセット, 75
 ジョブ, 75
 セキュリティ, 76
 データソース, 75
 ドキュメント, 76
 フィードバック, 75
 プロファイル, 75
 ボキャブラリ検索, 75
 患者レベル予測, 75
 特性の評価, 75
 発生率, 75
 設定, 75
 集団レベル推定, 75
attrition diagram, 172
AUC, 184
- back-propagation, 182
balance, see covariate balance
baseline time, 127
best practice for network research,
 273
- between-database heterogeneity, 252
bigknn, 181
Bill of Mortality, 35
- caliper, 151
 scale, 161
- case-control design, 152
- case-crossover design, 153
- case-time-control design, 153
- CDM , see Common Data Model
- characterization, 67, 127
 cohort, 128
 database level, 127
 treatment pathways, 128
- classification concept, 39
- Clem McDonald, 216
- clinical equipoise, 151
- cohort
 entry event, 108
 exit criteria, 108
 inclusion criteria, 108
 probabilistic design, 109
 rule-based design, 108
- cohort method, 149
- colliders, 151
- Common Data Model, 19
 スケーラビリティ, 21
 ソースコード, 21
 デザインの原則, 19
 データモデル図, 19
 データ保護, 19
 データ損失防止, 21
 ドメイン, 21

- 後方互換性, 21
- 技術の中立性, 21
- 目的適合性, 19
- community, 5, 216
- comparative effect estimation, 149
- comparative effectiveness, see comparative effect estimation
- comparator cohort, 150
- concept, 37
 - class, 38
 - code, 41
 - hierarchy, 43
 - identifier, 37
 - mapping, 42
 - relationship, 42
- concept set, 109
- conditioned model, 161
- confidence interval calibration, 251
- confounder, 150
- control hypotheses, 76
- convolutional neural network, 182
- counterfactual, 149
- covariate balance, 151
 - example, 172
- Cox proportional hazards model, see Cox regression
- Cox regression, 150
- cross-validation, 180
- Cyclops, 180
- data profiling, see White Rabbit
- data quality, 219
- DatabaseConnector, 85
 - creating a connection, 93
 - querying, 94
- decision boundary, 179
- decision tree, 182
- deep learning, 182
- descriptive statistics, see characterization
- design considerations for network research, 270
- direct effect estimation, 149
- disease natural history, see characterization
- domain
 - concept, 38
- drug utilization, 128
- ETL, see 抽出、変換、ロード (ETL)
- implementations, 60
- quality control, 61
- 単体テスト, 221
- ETL design, see Rabbit-In-A-Hat
- evidence quality, 215
- FAIR, 15
- feature analyses, 132
- FeatureExtraction, 138
- forum, 7
- gradient boosting, 181
- high correlation, 165
- hyper-parameter, 180
- incidence, 128
 - proportion, 129
 - rate, 129
- index date, 127
- instrumental variables, 151
- k-nearest neighbors, 181
- Kaplan-Meier plot, 172
- LASSO, 180
- logistic regression, 150, 180
- logistical considerations for an OHDSI network study, 271
- methods library, 76
- minimum detectable relative risk (MDRR), 172
- mission, 4
- model viewer app, 202
- naive bayes, 181
- nesting cohort
 - case-control design, 152
- network study, 269
- neural network, 182
- no free lunch, 180

- objectives, 5
- OHDSI Methods Benchmark, 258
- OHDSI SQL, see SqlRender
- outcome cohort
 - case-control design, 152
 - case-crossover design, 153
 - cohort method, 150
 - SCCS design, 153
 - self-controlled cohort design, 151
- p-value calibration, 250
- Pallas system, 37
- patient-level prediction, 69
- perceptron, 182
- person-time, 129
- phenotype library, 110
- PheEvaluator, 233
- Poisson regression, 150
- population-level estimation, 68, 149
- post-index time, 127
- power, 172
- preference score, 151
 - example, 171
- propensity model, 150
 - example, 171
- propensity score, 150
 - matching, 150
 - stratification, 150
 - trimming, 161
 - weighting, 150
- python, 181, 182
- quality improvement, see characterization
- Query Library, 85
- QueryLibrary, 99
- R, 85
 - installation, 78
- Rabbit-In-A-Hat, 50
- random forest, 181
- randomized trial, 150
- recurrent neural networks, 182
- regularization, 180
- reliable evidence, 216
- ROC, 184
- running network research, 271
- safety surveillance, 149
- self-controlled case series (SCCS) design, 153
- self-controlled cohort design, 151
- sensitivity analysis, 252
- sklearn, 181
- source code mapping, see Usagi
- SQL, 85
- SQL Query Library, see Query Library
- SqlRender, 85
 - debugging, 92
 - parameterization, 86
 - supported functions, 88
 - translation, 87
- standard concept, 39
- Standard SQL Dialect, see SqlRender
- standardized vocabularies, 35
 - download, 37
 - search, 37
- stratified model, conditioned modelを参照161
- strongly ignorable, 150
- structured query language, see SQL
- study diagnostics, 265
- study feasibility
 - single study, 265
- study-a-thon, 14
- supervised learning(監督学習), 178
- survival plot, see Kaplan-Meier plot
- system requirements, 78
- target cohort
 - case-control design, 152
 - case-crossover design, 153
 - cohort method, 150
 - SCCS design, 153
 - self-controlled cohort design, 151
- tools deployment, 81
 - Amazon AWS, 82
 - Broadsea, 82

- treatment utilization, see characterization
 TRIPOD, 177
- Usagi, 55
- variable ratio matching, 150
 variance, 180
 vignette, 78
 vision, 5
 vocabulary, 38
- White Rabbit, 48
 workgroups, 9
- xgboost, 181
- アウトカムコホート, 177
 アウトカムステータス, 179
 インデックス日, 185
 オープンサイエンス, 13
 open standards, 14
 オープンソース, 14
 オープンディスクース, 14
 オープンデータ, 14
 キャリブレーション, 184
 クラス, 179
 コホート, 107
 コホート定義, 107
 コミュニティ
 コミュニティコール, 9
 コンセプト
 祖先, 44
 コードセット, 107
 ソースデータ, see 生データ
 ソースレコード検証, 232
 ターゲットコホート, 177
 データ品質
 研究特有のチェック, 223
 チェック, 220
 バリデーション, 220
 妥当性, 220
 完全性, 220
 検証, 220
 準拠, 220
 ネイティブデータ, see 生データ
- ネガティブコントロール, 248
 プロトコル, 264
 ポジティブコントロール, 249
 ラベル, 179
 リスク期間, 177
 リレーションナルデータモデル, see Common Data Model
- 予後アウトカム, 177
 予測モデル, 177
 予測モデルの評価, 183
 交差検証, 183
- 偽陰性, 184
 偽陽性, 184
 共変量, 178
 共通データモデル
 標準化テーブル, 23
 規約, 21
 基準日, 179
 実証的キャリブレーション, 250
 実証的評価, 249
- 性能指標, 183
 患者レベル予測, 177
 感度, 183
 方法の妥当性, 247
 検証
 内部検証, 183
 外部検証, 183
 時間的検証, 183
 空間的検証, 183
 機械学習, 177
 欠損データ, 70, 179
- 特異度, 183
 生データ, 47
 真陰性, 184
 真陽性, 184
 研究ネットワーク, 269
 研究パッケージ, 264
 研究診断, 247
 精度, 183
- 臨床意思決定, 177
 血管性浮腫, 185

- 表現型, 107
 - 観察研究の限界, 70
 - 診断アウトカム, 177
 - 識別力, 184
- 適合率-再現率曲線下の面積, 184
- 陽性予測値, 183