– – OHDSI

Observational Health Data Sciences and Informatics (OHDSI)

2025-03-10

Contents

序章 xi
この本の目標 xi
本書の構成 xi
貢献者 xi
ソフトウェアのバージョン xii
ライセンス xii
本書が作成された方法 xiv

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 5
1.4 OHDSIの進展 5
1.5 OHDSIにおける協力 6
1.6 まとめ 7

2 どこから始めようか 9
2.1 旅に参加しよう 9
2.2 どこにフィットするか 13
2.3 まとめ 14

3 オープンサイエンス 15
3.1 オープンサイエンス 15
3.2 オープンサイエンスの実践: the Study-a-Thon 16
3.3 オープンスタンダード 16

3.4 オープンソース 16
3.5 オープンデータ 16
3.6 オープンディスコース 17
3.7 OHDSIとFAIRガイディングプリンシプルズ 17

II 共通データモデル 19

4 共通データモデル 21
4.1 デザインの原則 21
4.2 データモデルの規約 23
4.3 CDM標準化テーブル 27
4.4 追加情報 36
4.5 まとめ 36
4.6 演習 36

5 標準化ボキャブラリ 39
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 . . . 39
5.2 コンセプト 41
5.3 関係 46
5.4 階層 49
5.5 内部参照テーブル 51
5.6 特別な状況 51
5.7 まとめ 52
5.8 演習 52

6 ETL（抽出-変換-読込） 55
6.1 はじめに 55
6.2 ステップ 1: ETLのデザイン 55
6.3 ステップ2: コードマッピングの作成 63
6.4 ステップ 3: ETLの実装 69
6.5 ステップ 4: 品質管理 70
6.6 ETLの規約とTHEMIS 70
6.7 CDMおよびETLのメンテナンス 71
6.8 ETLに関する最終的な考察 71
6.9 まとめ 71
6.10 演習 72

III データ解析 75

7 データ解析の使用例 77
7.1 特性評価 77
7.2 集団レベルの推定 78
7.3 患者レベルの予測 79
7.4 高血圧症におけるユースケース 80
7.5 観察研究の限界 80
7.6 まとめ 81
7.7 演習 81

8 OHDSI 分析ツール 83
8.1 分析の実装 83
8.2 分析戦略 84
8.3 ATLAS 85
8.4 Methods Library 86
8.5 展開戦略 93
8.6 まとめ 94

9 SQLとR 95
9.1 SqlRender 96
9.2 DatabaseConnector 104
9.3 CDMへのクエリ 107
9.4 クエリ実行時にボキャブラリを使用する 109
9.5 QueryLibrary 110
9.6 簡単な研究のデザイン 112
9.7 SQLとRを使用した研究の実施 112
9.8 まとめ 118
9.9 演習 118

10 コホートの定義 121
10.1 コホートとは？ 121
10.2 ルールベースのコホート定義 122
10.3 コンセプトセット 123
10.4 確率的コホート定義 123
10.5 コホート定義の妥当性 124
10.6 高血圧のコホート定義 125
10.7 ATLASを用いたコホートの実装 125
10.8 SQLを使用したコホートの実装 134
10.9 要約 141

10.10 演習 141

11 特性評価 143
11.1 データベースレベルの特性評価 143
11.2 コホート特性評価 144
11.3 治療経路 144
11.4 発生率 145
11.5 高血圧症患者の特性評価 145
11.6 ATLASにおけるデータベースの特性評価 146
11.7 ATLASにおけるコホート特性分析 146
11.8 Rでのコホートの特性評価 156
11.9 ATLASにおけるコホート経路分析 158
11.10 ATLAS における発生率分析 162
11.11 まとめ 166
11.12 演習 166

12 集団レベルの推定 167
12.1 コホートメソッド設計 167
12.2 自己対照コホートデザイン 170
12.3 症例対照デザイン 170
12.4 ケース・クロスオーバーデザイン 171
12.5 自己対照症例シリーズデザイン 172
12.6 高血圧研究のデザイン 173
12.7 ATLASを使用した研究の実施 174
12.8 Rを使用した研究の実施 184
12.9 研究の結果 191
12.10 まとめ 197
12.11 演習 197

13 患者レベル予測 199
13.1 予測課題 199
13.2 データ抽出 201
13.3 モデルの適合 202
13.4 予測モデルの評価 205
13.5 患者レベル予測研究のデザイン 208
13.6 ATLASでの研究の実装 210
13.7 Rでの研究実施 221
13.8 アウトカム普及 226
13.9 患者レベル予測機能の追加 235
13.10 まとめ 235

13.1 演習 236

IV　エビデンスの質 237

14 エビデンスの質 239
14.1 信頼できるエビデンスの属性 239
14.2 エビデンスの質の理解 240
14.3 エビデンスの質の伝達 241
14.4 まとめ 241

15 データ品質 243
15.1 データ品質問題の原因 243
15.2 一般的なデータ品質 244
15.3 研究特有のチェック 247
15.4 実践におけるACHILLES 249
15.5 Data Quality Dashboardの実践 251
15.6 特定の研究チェックの実践 251
15.7 まとめ 254
15.8 演習 254

16 臨床的妥当性 257
16.1 医療データベースの特性 257
16.2 コホートバリデーション 257
16.3 ソースレコード検証 259
16.4 PheValuator 260
16.5 エビデンスの一般化可能性 267
16.6 まとめ 268

17 ソフトウェアの妥当性 269
17.1 研究コードの妥当性 269
17.2 Methods Librarのソフトウェア開発プロセス 270
17.3 Methods Libraryのテスト 272
17.4 まとめ 273

18 方法の妥当性 275
18.1 デザイン特有の診断 275
18.2 推定のための診断 276
18.3 実践におけるメソッド検証 281
18.4 OHDSIメソッド評価ベンチマーク 287
18.5 まとめ 289

V OHDSI研究 291

19 研究の段階 293
19.1 一般的なベストプラクティスガイドライン 293
19.2 詳細な研究手順 294
19.3 まとめ 297

20 OHDSI ネットワーク研究 299
20.1 OHDSI 研究ネットワークとして 299
20.2 OHDSIネットワーク研究 299
20.3 OHDSIネットワーク研究の実行 301
20.4 展望: ネットワーク研究の自動化を利用する 303
20.5 OHDSIネットワーク研究のベストプラクティス 304
20.6 まとめ 305

Appendix 305

A 用語集 307

B コホート定義 311
B.1 ACE阻害薬 311
B.2 ACE阻害薬単剤療法新規ユーザー 312
B.3 急性心筋梗塞（AMI） 315
B.4 血管性浮腫 316
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 . . . 317
B.6 高血圧のための第一選択治療を開始する患者 321
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324
B.8 ACE阻害薬の使用 324
B.9 アンジオテンシン受容体拮抗薬（ARB）の使用 325
B.10 サイアザイドおよびサイアザイド様利尿薬の使用 326
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326
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327
B.13 ベータ遮断薬使用 327
B.14 ループ利尿薬使用 328
B.15 カリウム保持性利尿薬使用 328
B.16 アルファ1遮断薬使用 329

C ネガティブコントロール 331
C.1 ACE阻害薬とサイアザイド・サイアザイド様利尿薬 331

D プロトコルテンプレート 335

E 回答例 337
E.1 共通データモデル 337
E.2 標準化ボキャブラリ 341
E.3 ETL (Extract-Transform-Load) 342
E.4 データ分析のユースケース 343
E.5 SQLとR 343
E.6 コホートの定義 345
E.7 特性評価 350
E.8 集団レベルの推定 356
E.9 患者レベルの予測 361
E.10 データ品質 364
E.11 364

Bibliography 365

Index 379

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コミュニティにより作成され、O
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（訳者注：英語
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、OHDS
データ標準、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の両方を対象

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々です。しか

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 Hripcsak	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。

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

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
https://github.com/OHDSI/TheBookOfOhdsiInJapanese/ から自動的に再構築され、継続的
“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（ユ

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、保険会社や政

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のエビデンスを
(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のデータを分析
(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、私たちはこうした素晴らし

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を導き出すま

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データベースが
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に必要とされ
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、この新しい知識

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、米国食品
(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、真の医薬品安

Figure 1.1: データからエビデンスへの旅

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っ
(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、データ標準
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ために

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(M
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、よ
- 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

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、コミュニテ
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立するた

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、コミュニテ

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、新しい方法
- 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。
- 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、誰もがOHDSIに
- 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優先し、対処する
- 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、コミュニティの
- 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、技術提供者、医療シス
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。

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ワークを
CDMという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アプローチ

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する
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効果のため

[1] https://www.ohdsi.org/who-we-are/collaborators/

Figure 1.2: 2019年8月現在のOHDSI協力者の地図

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
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

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
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
(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
(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
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。

[2]https://github.com/OHDSI

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
- 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であり、こ
- 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、オープンソース分析

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ています。私た

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、誰もが積極的に参加

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話ができるオン

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法についてのデ
- 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開発につい
- 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、統計手法やOH
- 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開発に関するディスカッショ
- 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コミュニティ活動に関連する、母国語での地域のディスカッション

[1] https://forums.ohdsi.org

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

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。フォーラムのアカウン
to OHDSI! - Please introduce yourself（OHDSIへようこそ！-自己紹介をお願いします」というスレッドで
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したいか教えて

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に新しい投稿が

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、定期的に対面イ

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、コラボレーターが全体
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、コミュニティの新規参加者に、テ

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フォーラム
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、適切な観察分析を設計・

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、対面会議を検索し、OHDSIチ

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機会で

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ッションのトピッ

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把握するため

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。各ワークグ
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。ワークグル
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。是非コールに参加して、より

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盤とし	プログラマー、Atlas & WebAPI & JavaScript ソフトウェア開発者

ワークグループ名	目的	対象参加者
CDM & ボキャブラリ	臨床患者データに適用される共通データ標準化されたと標準化された分析を	OHDSIコミュニティの人が参加可能
	グループによって開発された標準化された分析をサポートするため、	
ゲノム解析	OMOP CD-	すべての人が参加可能
	Mを拡張して、患者のゲノムデータを組み込みます。グループは、さ	
	CDM	
	互換スキーマを定義します。	
集団レベルの推定	正確で信頼性が高く、再現性のある集団レベルの効果推定につながる	関心のある人が参加可能
自然言語処理	OHDSI傘下の観察研究で、EHRのテキスト情報の使用を促進します。	すべての人が参加可能
患者レベルの予測	複数の対象とするアウトカムに対して、対象とするあらゆ	すべての人が参加可能
ゴールドスタンダート表現型	現在OHDSIコミュニティのメンバーが、イプのキュレーションを検証する	検証に関心
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加可
	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に	
GIS	OMOP CD-	健康関連の地理属性に興味のあるすべての人
	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付	
臨床試験	OHDSIプラットフォーム	臨床試験に興味があるすべての人が参加可能
	OHDSI	
	ツールの更新の推進を支援します。	
THEMIS	THEMIS	ETL標準化に関心のあるすべての人が参加可
	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-	
	読込）	
	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る	
メタデータ & 注釈	私たちの目標は、人間と機械が作成した注釈データと注釈を	すべての人が参加可能
	共通データモデル	
	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ト	
患者生成医療データ (PGHD)	この	すべての人が参加可能
	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	
	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で	
OHDSI女性グループ	OHDSIコミュニティ内の女性が一堂に会し、共同学習する場を	女性科学者、技術者、学生、教授

ワークグループ名	目的	対象参加者
運営委員会	OHDSIのすべての活動とコミュニティの成長を続けるコミュニティのニーズに合致の使命、ビジョン、価値観を維持します。さらに、このグループは、OHDSIの将来の方向性についてガイダンスを提供することで、コロンビアに拠点を置くOHDSI調整センターの諮問グループとして機能します。	コミュニティ、OHDSI

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。現在、O
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。

2.1.6 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関心を持ってい
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、特定のドメイン

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？」と疑問に思って

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たいと考えてい

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のに役
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にかけており、

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。組織
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、あなたはOHDSI
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。OHDSIの
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研究の種類を理解するのに役立つかもしれません。OHDSIコミュニティは、あ

[2] https://www.ohdsi-europe.org/
[3] https://forums.ohdsi.org/c/For-collaborators-wishing-to-communicate-in-Korean
[4] https://ohdsichina.org/

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 CD-
M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
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。遠慮しないでく

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
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ます）に

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
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ライセ

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。
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。必要に応
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フ
- 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、すべての会議の議事録や資料
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医療データに関し
本章ではこれらの疑問について触れていきます。

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るようになっ
(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、ソフトウェアを含む）
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の多くは今日の

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。情報技術は
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。こうし
[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査読を適用しようとした
(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の誤った結論、
(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、完全な再現性を実現するに
(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、大規模な医療エビデンスの生
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ザインを用いて、

[1] https://ohdsi.github.io/TheBookOfOhdsi/OpenScience.html#fn17
[2] https://www.ehden.eu/webinars/
[3] https://www.nature.com/collections/prbfkwmwvz

共通データモデルにデータを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実現

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
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、st
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
8参照）。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

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
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
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
７章でさらに詳しく説明されており、世界中で使用されている数百の医療コーディングシステ
(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
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
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、
ネットワーク研究（第 20 章参照）の実行コードです。これらのプログラムは、GitHubを介し
第17 章を参照ください。

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
//howoften.orgやhttp://data.ohdsi.orgで公開されている、その他の公開結果セットのような
リサーチネットワーク（ 第20 章参照）を活用して、データをOMOPにマッピングした利用可

[4] https://youtu.be/X5yuoJoL6xs

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

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」ツールを再利用し、マッピングコードをオープンソースとして公開することが

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、それだけでは医療行

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、OHDSI
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を通じて執筆され
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必要があります
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いうビジョンを

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、将来の参照と再現の
(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ます。

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方法として標準
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。しかし、LEGENDやH

[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
[10] https://github.com/OHDSI/TheBookOfOhdsi
[11] https://emif-catalogue.eu

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
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、OH
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との関

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）
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、本書で詳しく取り
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メタ
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管轄

- 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積極的
- 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

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。このデータは世界中でますます多
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究に日常的に使

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ることはできま

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）と組み合わせ

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）やアウトカム（コンテ
- 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、治療や治療の順序などさまざま
- 個々の患者でアウトカムが発生する可能性を予測する一一　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 一一 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：

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

- 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、分析のために最

- 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すべてのデータ

- 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れる、個人中心
- 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ン）で、その
  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、エンティティ-
  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、標準的な医療コンセプトの
- 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、疾病予防管理センターな
- 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場合でも、モテ

- 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分析データセットとして実現できま

- 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臨床観察データ

- 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/OHDSI/C

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の開発者は

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テーブルにリ

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と読み取り
10章を参照ください。これらのテーブルは書き込み可能であり、実行時にCOHORTテーブルにコホートを

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、INTEGER、FLOA

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築され
8章をご覧ください。

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のテ
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、どのテーブルに記録され

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)
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)	51	費用 (Cost)
89645	メジャーメント（測定）(Measurement)	15	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)

コンセプト数	ドメインID	コンセプト数	ドメインID
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現されます

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）が含まれ
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、イベントテーブル間の関係を確立す
[Event]_CONCEPT_ID	CONCEPT参照テーブルの標準コンセプトレコードへの外部キーで = 31967にはS-NOMEDコンセプトの「吐き気」の参照値が含まれています。
[Event]_SOURCE_CONCEPT_ID	CONCEPT参照テーブルのレコードへの外部キーです。このコンセ = 45431665は「吐き気」というコンセプトを、Read用語で示し、同
[Event]_TYPE_CONCEPT_ID	CTコンセプト31967です。標準分析アプリケーションの場合、ソー ソース情報の出所を示す標準化ボキャブラリ内で標準化されたコン
[Event]_SOURCE_VALUE	ソースデータでこのイベントがどのように表現されていたかを反映 9コード787.02に対応する「78702」のレコードが含まれる可能性

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、複数の場所に同等の情報が

- 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。これらは、ICD9CM、
- 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。ほとんどのコン
- 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。ソースコンセプトは、既存
- 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、すべてのデータベースで臨床

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、特定のデー
4.2 で示されているように、コンディション「肺結核」（TB）のICD9CMコードは011です。

Figure 4.2: 肺結核のICD9CMコード

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（メディケアパートA
このような場合に、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を表す
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プ（OM
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関係は
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、サポート

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。

TERM CONNECTIONS (82)

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 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、4つのヘ
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、本章の残りの部分ではある1人の

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。重症になる

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なければなりま

[1] https://github.com/OHDSI/CommonDataModel/wiki

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嚢腫の
//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れ、ビル
GENDER_CONCEPT_ID	8532	女性性別を参照するコンセプトIDは8532で
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ます。
DEATH_DATETIME		
RACE_CONCEPT_ID	8527	白人を示すコンセプトIDは8527です。英国
ETHNICITY_CONCEPT_ID	38003564	これはヒスパニック系の人々を他の人々から に格納されます。米国以外では使用されませ
LOCATION_ID		彼女の住所は不明です。
PROVIDER_ID		彼女のプライマリケア医療従事者は不明です
CARE_SITE		彼女の主な医療施設は不明です。
PERSON_SOURCE_VALUE	1	通常、これはソースデータでの彼女の識別子
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コー F」であり、PCORNet ボキャブラリコンセプトに記載されている場

列名	値	説明
RACE_SOURCE_VALUE	white	ソースに表示されてい人種値がここに格納されます。
RACE_SOURCE_CONCEPT_ID	0	同様にGENDER_SOURCE_CONCEPT_IDの原則が適用
ETHNICITY_SOURCE_VALUE	english	ソースに表示されている民族値がここに格納されます。
ETHNICITY_SOURCE_CONCEPT_ID	0	同様にGENDER_SOURCE_CONCEPT_IDの原則が適用

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、少なく

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しょう。彼女

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かもしれま

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、テーブル内の各レ
PERSON_ID	1	これはPERSONテーブルでLauraのレコードへの外部キ
OBSERVATION_PERIOD_START_DATE	2010-01-06	これは記録上、彼女の最初のエンカウンターの開始日

列名	値	説明
OBSERVATION_PERIOD_END_DATE	2013-01-24	これは記録上、彼女の最後のエンカウンター
PERIOD_TYPE_CONCEPT_ID	44814725	「Obs Period Type」コンセプトクラスを持つボキャブラリ

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

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、各レコー
PERSON_ID	1	これはPERSONテーブルでLaurenのレコー
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。時間が不明なた
VISIT_END_DATE	2013-01-24	ビジットの終了日です。これは1日のビジッ
VISIT_END_DATETIME	2013-01-24 00:00:00	ビジットの終了日と時間です。時間が不明な
VISIT_TYPE_CONCEPT_ID	32034	ビジットレコードの出所を示します。保険請
PROVIDER_ID	NULL	エンカウンターレコードに医療従事者が関連
CARE_SITE_ID	NULL	エンカウンターレコードに関連するケアサイ
VISIT_SOURCE_VALUE	入院	ビジット値はソースでどのように表示される
VISIT_SOURCE_CONCEPT_ID	NULL	ビジット値は、ソースがOHDSIによって認識
ADMITTED_FROM_CONCEPT_ID	NULL	既知の場合、患者が入院した場所を表すコン 8536「自宅」が含まれます。
ADMITTED_FROM_SOURCE_CONCEPT_ID	NULL	患者が入院した元の場所を表すソース値が表

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コンセプ 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。上
PRECEDING_VIS	NULL	現在のビジットの直前のビジットを示します。ADMIT

- 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することが 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、または患者によっ

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。直腸のす

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_OCCURRENCEテーブルでどのように表現されるかを示しています。

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、各レコードに一意
PERSON_ID	1	これは、PERSONテーブルのLauraのレコードへの外部
CONDITION_CONCEPT_ID	194696	SNOMEDコード266599000を表す外部キーは19469
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。時刻は不明な
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が、
CONDITION_END_DATETIME	NULL	既知の場合、コンディションが終了したと見なされる

列名	値	説明
CONDITION_TYPE_CONCEPT_ID	32020	この列は、レコードの由来に関する情報を提 32020「EHR エンカウンター診断」）というコンセプトが
CONDITION_STATUS_CONCEPT_ID	NULL	これが分かると、状況と理由がわかります。 4203942 が使用されました。
STOP_REASON	NULL	既知の場合、ソースデータに示されているコ
PROVIDER_ID	NULL	コンディションレコードに診断を付けた医療 ID がこのフィールドに入ります。これは、その
VISIT_OCCURRENCE_ID	509	コンディションが診断されたビジット（VIS
CONDITION_SOURCE_VALUE	266599000	これはコンディションを表す元のソース値で
CONDITION_SOURCE_CONCEPT_ID	194696	ソースからのコンディションの値が OHDSI で認識されるボキャブラリを使用してコード がここに入ります。月経困難症の例では、ソ コードなので、そのコードを表すコンセプト 194696 です。この場合、CONDITION_CONCEPT_ 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・ステータ OHDSI がサポートするコード化スキームでコード化

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記

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の経口投

列名	値	説明

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を作成するために自動
PERSON_ID	1	PERSONテーブルのLaurenのレコードに対する外部キ
DRUG_CONCEPT_ID	1127433	医薬品のコンセプト。アセトアミノフェンの NDC コードは RxNorm コード 313782 に対応し、コンセプト 1127433を表します。
DRUG_EXPOSURE_ START_DATE	2010-01-06	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。時間が不明なため0時を使用。
DRUG_EXPOSURE_ END_DATE	2010-02-05	薬剤曝露の終了日。様々な情報源によって、既知の日付
DRUG_EXPOSURE_ END_DATETIME	2010-02-05 00:00:00	薬剤曝露の終了日時。DRUG_EXPOSURE_END_DATE
VERBATIM_END_DATE	NULL	情報源が実際の終了日を明確に記録している場合。推定
DRUG_TYPE_ CONCEPT_ID	38000177	この欄は、記録の出所に関する情報（保険請求や処方箋 38000177 (“Prescription written”) が使用されています。
STOP_REASON	NULL	薬剤の投与が中止された理由。理由にはレジメンの完了
REFILLS	NULL	多くの国で処方システムの一部となっている、初回処方 NULL から始まります。Lauren のアセトアミノフェンの場合、リフィルはなかったので NULL です。
QUANTITY	60	最初の処方箋または調剤記録に記録された薬剤の量。
DAYS_SUPPLY	30	処方された薬の処方日数。
SIG	NULL	元の処方箋または調剤記録に記録されている（米国の場 は CDM ではまだ標準化されておらず、逐語的に提供されます。

列名	値	説明
ROUTE_CONCEPT_ID	4132161	このコンセプトは、患者が曝露された薬剤の 4132161 (“Oral（経口）”) が使用されています.
LOT_NUMBER	NULL	製造業者から医薬品の特定の数量またはロッ
PROVIDER_ID	NULL	薬剤レコードに処方プロバイダがリストされ
VISIT_OCCURRENCE_ID	509	薬剤が処方された VISIT_OCCURRENCE テーブルへの外部キー。
VISIT_DETAIL_ID	NULL	A foreign key to the VISIT_DETAIL table during which the Drug was prescribed.
DRUG_SOURCE_VALUE	69842087651	ソース・データに表示される医薬品のソース
DRUG_SOURCE_CONCEPT_ID	750264	薬剤のソースデータでの値を表すコンセプト 750264 ND-Cコードで”Acetaminophen 325 MG Oral Tablet（アセトアミノフェン 325 MG　経口錠）“を表します。
ROUTE_SOURCE_VALUE	NULL	情報源に詳述されている投与経路に関する逐

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命じた

- 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の
- 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の嚢胞があることがわか

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を作成するた
PERSON_ID	1	これはPERSONテーブルのローラのレコードに対する
PROCEDURE_CONCEPT_ID	4127451	骨盤超音波検査のSNOMEDプロシージャーコードは3
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。時刻が不明な場合は
PROCEDURE_TYPE_CONCEPT_ID	38000275	このカラムはプロシージャーの記録の由来に関する情 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 を意味します。例えば、CPT4 のプロシージャーが両側で行われたと記録されている 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、または実施されたプロシージャーの
PROVIDER_ID	NULL	ProcedureレコードにProviderがリストされている場
VISIT_OCCURRENCE_ID	740	情報がある場合には、これはプロシージャーが施行さ
VISIT_DETAIL_ID	NULL	情報がある場合、プロシージャーが実施されたビジッ テーブルから VISIT_detail_id として取得）。
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コ
PROCEDURE_SOURCE_CONCEPT_ID	4127451	プロシージャーのソースデータの値を表すコンセプト
MODIFIER_SOURCE_VALUE	NULL	ソース・データに表示される修飾子のソース・コード。

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
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
- 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、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

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

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：7616800

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、Da

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

[2]https://github.com/OHDSI/CommonDataModel/wiki
[3]http://athena.ohdsi.org/
[4]http://atlas-demo.ohdsi.org

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データセッ

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

CDMデータベーススキーマは「main」です。これはCONDITION_OCCURRENCEテーブルの一行を取得

```
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）のすべてのレコードを取得して

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」のすべてのレコードを取得してく
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、データの内容を定義するこ
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、および典型的な状

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めに作成された
(“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（処置）やサー
“International Classification of Disease (ICD)” を作成しています。各国の政府は、ICD10CM（米国）、

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り、それは使用
Nomenclature of Medicine (SNOMED) や、Logical Observation Identi-
fiers 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あるRxNormを、
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つの目的で利用できま

- 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

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.

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

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
CDM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。

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。これにより、研究者が
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れています。誤り
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、ATHENA[4]から

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むzipファイルを
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れているボキャ

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 ）多くのボキャブラリ
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ト）などの内部管
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。これらはデータレコードの基本的
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（例：コンディ

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。この無意味な整数

[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

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

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。名称はボキャブラ

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。これは数値のC
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、臨床イベントやイベント
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ラ
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
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

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。直感的には、

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

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、HC
垂直	CIEL、HES専門、ICDO3、MeSH、NAACCR、NDFRT、
混在	CPT4、ISBT、LOINC
なし	APC、すべてのタイプコンセプト、民族性、OXMIS、種

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

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D00

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（NU

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が、
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_ANCES
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行わ
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に
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。

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	非標準コンセプトのためのボキャブラリ	分類コンセプトのためのボキャブラリ
コンディション	SNOMED, ICDO3	SNOMED Veterinary	MedDRA

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	階層のためのボキャブラリ
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、現時点では標準化されていない	現時点では標準化されていない
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NDCコード
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、コンセプトク

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ID	コンセプトクラス
35803438	1001	顆粒球コロニー刺激因子	薬剤	HemOnc	コンポーネントクラス
35942070	1001	AJCC TNM Clin T	メジャーメント	NAACCR	NAACCR変数
1036059	1001	アンチピリン	薬剤	RxNorm	成分
38003544	1001	レジデンシャル治療 - 精神科	収益コード	収益コード	収益コード
43228317	1001	アセプロメタゾン	薬剤	ATC	成分
45417187	1001	ブロムフェニラミンマレイン酸塩 10 mg/ml注射用溶液	薬剤	Multum	Multum
45912144	1001	血清	標本	CIEL	標本

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ID	コンセプトクラス

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。そ
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在は使

- 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り込み 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が、現
  - 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り込み 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が欠落
  - 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。通常は
  - 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 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、アップ
  - 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
  - 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
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、
to」関係には反対の関係「Mapped from」があります。

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階層関係。

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_START

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ペアによって
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” “Maps to value”と”Value mapped from”	標準コンセプトはそれ自身にマッピングされ、非標準コンセプトは MEASUREMENTとOBSERVATIONテーブルのVALUE_AS_CONCEP

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、臨床イベントがOMOP
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が同じ意味
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SNOMEDボキャ
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という情報

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、ICD9C

070.43「肝性昏睡を伴うE型肝炎」は、SNOMED 235867002「急性E型肝炎」とSNOMED 72836002「肝性昏睡」の両方にマッピングされます。これは、元のソースコンセプトが肝炎

「Maps to value」関係は、エンティティ-属性-値（EAV）モデルに従って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
to」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。例については図 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。事前に組み合わせら
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され
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り組

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、子コンセプトか
49436004「心房細動」は、SNOMED 17366009「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るか
49436004「心房細動」はSNOMED 40593004「細動」に対しても「Is a」に該当します。

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

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、ボキャブラリの
- 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[6]。

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。例えば、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）。

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のような

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テーブルに

CONCEPT_ANCESTOR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CO
“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の他の関係

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_LEV

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す。プロシ

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

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、CONCEP

5.5

DOMAIN_ID、VOCABULARY_ID、CONCEPT_CLASS_ID（すべてCONCEPTレコード内）およびCONCE
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 *_CON-
CEPT_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
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照するVOCABU

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。代替の性別を
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。このテーブルに

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別であり、人種

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いて、ある程度
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が、部分
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。また、ICDコード

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あると考えられ

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ーディングスキー

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、これらはサービス

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専門

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
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つの性別コ

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、コン
- 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CD
- 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
- 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
- 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
- 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
- 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お

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ATHEN

[7] http://athena.ohdsi.org/
[8] http://atlas-demo.ohdsi.org

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？この標準コンセプト
9CM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。

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に再構築し、標準化

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つの主要なステップが

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、これらのステップ

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とCDMの両方

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：White
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）を理
Rabbitツールです。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、フィールドに
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ス
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
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に
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

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

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
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
all in DB」ボタンをクリックできます。

[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 WhiteRabbit 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に設定されおり
- 「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、White Rabbitはテーブル内の100,000行をランダムに選択してスキャンします。

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
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、各テーブルの各フ
6.3 は、概要タブの例を示しています。

	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

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。各ソーステーブ

レポートはソースデータ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。例えば、図6.
Rabbitは1を男性、2を女性として定義することはなく、データホルダーが通常、ソースシステムに固有の

	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

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in-a-Hatをスクリーンに映し出します。最初のラウンドでは、テーブル間のマッピングを共同で

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 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
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CD
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
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

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

[2]Synthea™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/syntheticheath/synthea/wiki をご覧ください。

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よくあります。これは、イベントに正しいVISIT_C

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。これ
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は患者名、運転

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たロジックを
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。『Source
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テーブル（この
＆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。これは、ソースのid ソースからのidフィールドは、その値を保持し、必要に
GENDER_CONCEPT_ID	gender	性別が「M」の場合、GENDER_CONCEPT_IDは8507
YEAR_OF_BIRTH	birthdate	生年月日から年を取得します。

目的フィールド	ソースフィールド	ロジックとコメント
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、ソース
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、それ以外の場合は
ETHNICITY_CONCEPT_ID	race ethnicity	race = ‘HISPANIC’、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、複数のソース列が 1 つの CDM 列にどのように影響するかを示す良い例です 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	
RACE_SOURCE_CONCEPT_ID		

目的フィールド	ソースフィールド	ロジックとコメント
ETHNICITY_SOURCE_VALUE	ethnicity	この場合、ETHNICITY_SOURCE_VALUEはETHNICIT
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、仕様書全文をご覧ください[3]。

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にデータを変換する際の
10CM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には、relationsh
=「Maps to」を持つCONCEPT_RELATIONSHIPテーブルのレコードを使用できます。例えば、ICD-
10CM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している場合も

- 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
- 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にマッピングされ
- Usagiを使用しましょう。

[3] https://ohdsi.github.io/ETL-Synthea/

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

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。コードの記述のテキスト類

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列が必要

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ピングし
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。コー
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、ソースコードと英語の

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

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、英語にも翻訳

「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使用 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せんが

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。 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。各フィルターについて

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A concept ID列で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（セミコロンで区切りま

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。

[4] https://translate.google.com/
[5] https://github.com/OHDSI/Usagi

Figure 6.7: Usagiコード入力画面

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ー」

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から派生し 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。いずれのテーブル

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 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、オランダ語のコンデ 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名前 が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る @(ref:fig:usagiOverview) では、オランダ語の “Hoesten” は英語の

Figure 6.8: Usagiでのソースコード入力画面

“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ング
(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なし（
= 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

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のような論理演算子はサ

例を続けると、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たとします。検

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にマッピン
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

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
+ Cキーを押すか、上部メニューバーのview （閲覧）-> Concept informa-
tion (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。

Figure 6.9: Usagi コンセプト情報パネル

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

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続行します。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

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、かな
(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取り組む
- 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、一部のコードは適切な
- 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セウポート

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_MAP 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。

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ウポート) を選択します。どのSOURCE_VOCABULARY_I
(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、SOURCE_TO_CO

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の構造はS
(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。マッピ

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

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、それらを

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されるコ

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。ET
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の

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
  - 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

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。 ト

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。典型なパターンは、
>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をテストする方法は

- 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どんな人でも間
  - 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が標準
- 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件ずつ確認すると役立
- 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が生
    性別を持つ人々を削除することを選択しています。なぜなら、そのような人々は分析
    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法で構
    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、予想し
- 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た方
- 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。例え
  - 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、変

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
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べき

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コミュニティが合
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セットがありま

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ことは不可能です
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女神で、秩序、

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多大な労力を要し

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場合もありま

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性を持ってい

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、ボキャブラリは1ヶ月に

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る場合です

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、私たちがすべてユニークなソ

- 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をかけ
  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、ユースケースに基づ
- 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、
- 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うことではあ
- 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、実装者フ

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

[6] https://github.com/OHDSI/CommonDataModel/wiki
[7] https://github.com/OHDSI/Themis
[8] http://forums.ohdsi.org/
[9] https://forums.ohdsi.org/c/implementers

    - 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
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のために書か
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、前の

- 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ジットセットがある場合、CDMで生成されるVISIT_OCCURR

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、実世界のヘルスケ
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はデー
これらの限界については、エビデンスの質に関する本の（第 14 章-
第18章）で詳しく説明しています。

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、経時的な変化に関する問いに答

データが答えを提供できる問い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トプットは以下の通りです：

- 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、ワルファリンはダビガトラ
- 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レベルの予測を

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はどの程度が
- 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を発症する確率

典型的な患者レベルの予測に関する問いは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、予測の重

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、予測モテ

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阻害薬単

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とすることに
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年以内に急性心筋

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、プラセボ治
- 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
- 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

•‡

dX8XR

2+'6,ÙªÑëªËŠ]ût«~µ°ÙªÑŽ•µ°•‡ ‡ ƒv…cªZ gka©•x‘Ág•• k è/„O±<}v…
Í£^ Í†Ž•r«¨• `Šƒc…œâv…i©•ëËÚéûÁØµË†Ž•r•§e‰ `±Z ‰Œ©± v
)XOOHU ±–Æn•tc‘

dX8Xk

2+'6,ÙªÑëªËŠipªS1Ž£Ñ‰s`± v•x‘ÙªÑëªËŠ]ût«ª˜l^×¶ë Ú¸Á•ˆc•k@¡‰ˆ
HWDO Žr•ÚæÕÁŠ*xª-Q‰ì ‘±%¯v…c•x‘

dXe

```
m#…p–†Ž• ƒ•]l‰øªËÃªË•½ØÆüª±ö2v•x‘
m±gÆWŽÄŠ¨ŠUk_r,~j—Å‡ce cŠŠg§e‡v•x‘
mX˜þëý•@ ŽÄÒbÄbŽUj—Å‡ce cŠŠg§e‡v•x‘
m þëý•="ŽÄÈŠŽUk_rª•j—Å‡ce cŠŠg§e‡v•x‘
m="ôÙýŽÒbôÙý†Ža©•z²‘½c="Ò¶Š†ìv…¡´¸Ú½òŠþÏ±?gª‡õgª}3Ža
m#…¿£ÙªÑ†ŽŠg¨«‰c c¡a©•x‘
```

dXd

!L r«¨• Žˆ•øªËÃªË•½ØÆüªŠ&v•xj—

Mï•ËØÿ¶ÛgÛ"Ç ¸16$,'¹±z?t«~ ŠipªÁS ¸*,¹pX•Yr†±CŸxª‘

ëªËû¶ •±gŠB„c…•± • k)¯#û(Š*,pX±<}xªÂ†±CŸxª‘

ÍþÅ¿Éè‡•Çv…ÊÁÿç¹ÜÁŠ§ª*,pX•üËÁ Y±@ xª‘

!L ÊÁÿç¹ÜÁŠ§ª*,pX•üËÁkéûÍî¸f ¹Š•˜…ˆ•*i •ªj±@ v~c‘¿£#…ÙªÑ±–

@§t«ªˆŠŽ•Ýû (±–Æn•tc‘

Ù ª Ñ ˆ ‰ • ¬ Q Á

*? Ti2` 3

P>.aA

Ì 0DUWLMQ 6FKXHPLH)UDQN 'H)DOFR

2+'6,Ž• þëý•#…ÙªÑŠFxªt•u•‰ÙªÑ ‰•øªËÃªË±Çïª Úxª~ • Õc¼ªé ÏªË

Í†Ž•M‰Š ‰± xªt•u•‰ce±œâv• ‰†TQ†Iª©ŠŠƒc…œâv•x‘ãŠ•t•u•‰2+

3XR

: Ž•&'0±¬Qv…ÙªÑëªËŠFv…p–± xª~ ŠÂh†Iªt•u•‰ce±êv…c•x‘

)LJXUH &'0•ÙªÑŠFxª ‰± xªt•u•‰ce

p–± xª~ •)‰´éÿªÓŽ ƒa©•x‘M‰•ceŽ•2+'6,k%¯xªÖªý±Æ”¬Qv‰c½ËÑò

€þ•ceŽ•5† ‰±ŸYv• 2+'6,0HWKRGV/LEUDU\•ÅÊÃÊÊÃÊÊ±xQxªce†x'-‰
Í†Úvnœâv…c•xk•3RVWJUH64/•64/6HUYHU•2UDFOH‰ˆ•t•u•
/LEUDU\•Å ïªß Ú±LxQxªr‡†• ÌŠ½ËÑòÅªÛ±¬QxªÊ §©¡Ä†:ŠtÖ±

€þ•ceŽ•éÿÂûðª†‰n…¡ Õn ‰±Ä†:Š ü†Iª¸¹èëªË•Öªý•
/LEUDULHV±¬Qv•xk• ‰±ÙÈ¶ xª~ •…ƒ‰Âûçµ½ý¶ Ñªç¹¶Ë±%¯v
/LEUDU\†xQZ ‰x˜…•¼éÉù ±ÇïªÚv…cª®p†Ža©•z²']h•p–Ž$7/$

$7/$6‡0HWKRGV/LEUDU\Ž¶‡v~¡•†Ža©•z²'$7/$6†•'pt«ª Œ‰ %
/LEUDU\•ãÕÃªÊ—••'pv±þw… üt«•x'ŸMŠ•0HWKRGV
/LEUDU\†¬Qt«ªÅíªÚŽ•5n•ê •$7/$6†ÙÈ¶ t«…c•x'

3Xk

&'0ŠFxª ‰± xª~ •©ŠŠYg•Ág•½ËÑòÅªÙµ Â£0HWKRGV
/LEUDU\†%¯t«ªÀ€ ‰ÅªÛ•xQ‰ˆ•ºåÙ Ë±rhxª~ • ‰kq±¬Qxª P•©
Ž•2+'6,†TQt«…cª ƒ•©Š±êv…c•x'

)LJXUH µ°• cŠFxªºåÙ Ë±rhxª~ •©Š

M‰•©Š†Ž•º ‰±Ž2•p–‡v…'c•x' ‰ŽéÿÚÅý†ÚÈŠ\ v•ÅªÛ‡v… v•
HW DO ' rr†Ž••yéÿÚÅýktht«•2+'6, 0HWKRGV
/LEUDU\±¬Qv~ ‰ÅªÛkŸYt«•2+'6,ßÕÚ ªÁÌD† üt«•Wbk•‡ ¨«…Ì

^á•©Š†Ž•ü´ýÑ¶ò•~Žš›ü´ýÑ¶ò†± •ÁûË• cŠŠg¨«ª´éüÃªÉù ±ŸYv

^"•©Š†Ž•ŸMŠ cŠÅP±†…•xk•|• c•ÁûËÐ•x˜…•ºåÙ Ë±øe:Šrhv§e
HW DO E 'r•p–†Ž•x˜…•eƒÊå£k• ƒ•]\ñ‰#…p–ÙªÑëªË†* •aª´¸Ú¹
†xQ†I•x'

KWWS GDWD RKGVL RUJ 6\VWHPDWLF(YLGHQFH

$7/$6

3Xj hG a

$7/$6Ž•2+'6,Åñ÷ÝØµkŸYv~•À€St«~ þëý•#…ÙªÑ±
&'0-ò† ‰xª™C‡ ü±½ xª•×Ÿ†ÂŸt«…cª¸¹èëªË•Öªý†x‘$7/$6
Ž•2+'6,:HE$3,‡àž ®z…¸¹è´éüÃªÉù ‡v…HŸt«•þíŽ
$SDFKH7RPFDWã†íËÚt«•x‘ü´ýÑ¶ò ‰±üeŠŽ•&'0
Ð• þëýÙªÑ—•´ÁÍËk´W†aª~ •þíŽà •ç³¶´¸»ªý•âÕÁŠ¶ ËÚªýt«•x‘~•v•ãè

)LJXUH $7/$6øªÈ¶ Ñç¹ªË

: Š$7/$6•ËÁüª ÉùÕÚ±êv•x‘> ŠŽ$7/$6•M§‰Q ±êxÜåÄªÉù âªka©•x
ÙªÑÏªË ÙªÑÏªËŽ•$7/$6éûÕÚç»ªòŠåht«~ºÙªÑÏªË•]r:†À€St«~þïªÚ±þ
î¿õèûük} $7/$6Ž2023À€Sî¿õèûü±k}v•r«¨•î¿õèûüŠˆ•§e‰Å ÍéÚk.e
Å ÍéÚÍÕÚ Å ÍéÚÍÕÚŽ•À€St«~ ‰ÌD†¬QxªÅ ÍéÚ•ÍÕÚ±ó2xª~ Š¬Q†Iª
ÅíªÚ q ÅíªÚ qŽ•Æ N(ÐŠ ƒ•~Ž P•B€±Â~x)•ÍÕÚ±å¯xªQ †a©•r«¨•Åíª
±gÆW ±gÆWŽ• qt«~#ƒ•~Ž P•ÅíªÚ±æ@v•r«¨• X˜•±g±W xª~ • ‰Q †x
ÅíªÚ<å ÅíªÚ<åŽ• ƒ•~Ž P•X˜Ð†Yrxªµ°¶ë Ú•ÉªÃ Ë±#…†Iª ‰Öªý†x‘r
Yr† Yr†Ž•FÛX˜Ð•´¸Ú½ò•Yr†±@ xª~ •Öªý†x‘r•Q Šƒc…Ž•^
éÿç³¶ý éÿç³¶ýŽ•Ž§• •d›:#…ÙªÑ±æ@v•± •Ž)Š_r,…cªóÅ±W xª~ •Öª

X˜þëý@ @ Ž••ÇÅíªÚÙÈ¶ ±¬Qv…X˜þëý•Äb@ p–± qv• ƒ•~Ž

þëý•=" ="xêŽQ—ÌL´ýÆüÌò±<Qv…• þëý•=" ‰±üc•± •ÑªÄ

Êùè Êùèó Ý÷ª þ±Âhv…•:HE$3,±þw… üt«…cªéÿÍË•óL±Âó†l•x'

åh åhóÝ÷ª þ±Âhv…•ÏªËåhÍÁÉù Šåht«~ÙªÑÏªË±Âó†l•x'

çµªÛâÕÁ çµªÛâÕÁü Á±ÁüÕÁxª‡•$7/$6•s`ÿÂŠ´ÁÍËv• vc `•ÿÂ

3XjXR

$7/$6‡:HE$3,Ž•éûÕÚç»ªòÌD†Q £ÙªÑÏªË—•´ÁÍË±c°xª~ •_jcÍ¿÷
6KLURû¶èûü±ÜQv…å¯t«…c•x'Í¿÷üØµÉËØò•Ú_Ž•¼ û¶ •:HE$3
'

3XjXk

$7/$6•Û¿÷ó ÚŽ•$7/$6*LW+XEüïÊÚü•ZLNL†¼ û¶ †Âó†l•x
'r•ZLNLŠŽ•t•u•‰´éüÃªÉù Q Š*xªîY£•¼ û¶ åÙ¼Ó÷ªÚü´ý—•ü

3XjXj

$7/$6•¶ ËÚªýŽ•2+'6,:HE$3,‡àž ®z…üc•x'ºÅ ïªß Ú•¶ ËÚªý¾¶
*LW+XEüïÊÚü•ÍÕÚ´Õé¾¶Û ‡:HE$3,*LW+XEüïÊÚü•¶ ËÚªý¾¶Ú

3X9J2i?Q/b GB#` `v

2+'6,0HWKRGVŽ•LEUDŠÛât«…cª¼ªé ïªË•5ãÕÃªÊ•ÅþÁÉù †x'

r«¨•ãÕÃªÊŽ•&'0Ð•ÙªÑj¨³•©•@ ¡£|«±€Ýpª•C•:Å±rhxª Ì‰#…p–±
¹•~ • i‰À€S ‰±%¯xªr‡¡†l•x'0HWKRGV OL
EUDU\Ž•—âg•L‡g•µ‰ªÅ Ø¿ËÚ†•óÏÕÛ•øt±g•" ¡•|•óÏÕÛŠ§,…rh

KWWSV JLWKXE FRP 2+'6, :HE$3, ZLNL 6HFXULW\ &RQILJXUDWLRQ
KWWSV JLWKXE FRP 2+'6, $7/$6 ZLNL
KWWSV JLWKXE FRP 2+'6, $WODV ZLNL $WODV 6HWXS *XLGH
KWWSV JLWKXE FRP 2+'6, :HE$3, ZLNL :HE$3, ,QVWDOODWLRQ *X

0 H W K R G V / L E U D U \

) L J X U H 2 + ' 6 , 0 H W K R G V / L E U D U \ • ã Õ Ã ª Ê

2+'6, ‰Öªý

0HWKRGV OLEUDU\Žx†Š5n•ÂÅt«~µ%ÞR-ODQG HW DO
'XNH HW DO 5DPFKDUUDQ HW HDLQ VWHLQ HW **DQ**J
HW DO 5\DQ HW DO 9DVKLVKW HW DXOD Q HW DO
-RKQVWRQ HW Ð-OQt«…i©•ceü•p–Š¡xQt«…c•x
6FKXHPLH HW DO 5HSV HW DO 7LDQ HW DO
6FKXHPLH HW DO 5HSV HW DO ‘0HWKRGV OL
EUDU\Ð•óÏÕÛ• •;†gŠƒc…Ž^ Í†œât«…c•x‘

3X9XR

x˜…•ãÕÃªÊ†àž&•«…cªeW‰Q •Æƒ Ž•5n• ‰±Ä†:Š ü†lªr‡†x‘Ág•

r•CŸÄ†Š§©•]\ñ‰ ‰kZ Š‰©•5n• ŠÆiŠ‰Šxªr‡k†l•x‘•~•Å Úÿªy
†œât«…cª§eŠ•<} ŠB„n¿õüèþªÉù ±ürˇ¡ÜZS†x‘

3X9Xk

0HWKRGV OLEUDU\Ž••íŠ]\ñ‰ÙªÑëªËŠFv…¡]ª•ÙªÑ±0ŸCŸ± ü†

]ü •ÙªÑøtŽÙªÑëªËÇªâªã† üt«•x‘ ‰Žþí•ÙªÑëªËÐ•ÌÙªÑ•snA
OLEUDU\Ž6TO5HQGHU£'DWDEDVH&RQQHFWRUãÕÃªÊ±†
]ª•ÿª½ýÙªÑ¼èÊ¹ÁÚŽóôüÄ†•§cce†C.t«•x‘ÿª½ýðÉ ŠÒ¸ ÿªÛt
OLEUDU\Ž6TOÕÃªÊ±¬Qv…]\ñÙªÑ¼èÊ¹ÁÚ±C.•^}v•x‘r«Š§©•óôüŠ?
´WŠ6w… g Å æ÷ªØµ Âk<Qt«•x‘Ág•• &\FOR&ÕÃªÊŽ•
OLEUDU\ÌD†¬Qt«ª•íŠÄ†:‰‰Rº Ê ± v…i©•r«Š§©þíŽ< †l%

3X9Xj

5ŽãÕÃªÊ± ‘Sxª~ •À€:‰ce±%¯v…c•x‘ºãÕÃªÊŠŽ•ãÕÃªÊŠ0•«ªx
/LEUDU\•¸¹èÇ¶ÚãÕÃªÊ•*LW+XEüïÊÚü•&5$1†xQ†l•x‘t¨Š•5•Ðü

ãÕÃªÊðÝ÷´ýŠYg…•5n•ãÕÃªÊŽåßÕÚk%¯t«…c•x‘åßÕÚŽ•± •ÑË
/LEUDU\•¸¹èÇ¶Ú•ãÕÃªÊ•*LW+XEüïÊÚü•&5$1†ì,Z ‰ãÕÃªÊŽ&5

KWWSV RKGVL JLWKXE LR 0HWKRGV/LEUDU\
KWWSV RKGVL JLWKXE LR &RKRUW0HWKRG DUWLFOHV 0XOWLSOH

0HWKRGV /LEUDU\

3X9X9

ÉËØòW[±k’xªcŠŽ•áƒ•Å æ÷ªØµ Â »k*Þv…l•x•ÙªÑëªËÇªâª‡ ‰ ªÁËØªÉ

ÙªÑëªËÇªâªŽ&'0-ò•#…¿£ÙªÑ±Cáxª´Wka©•x‘0HWKRGV
OLEUDU\Ž•^h•ÙªÑëªËÉËØò¸3RVWJUH64/•0LFURVRIW 64/
6HUYHU•2UDFOH¹•ãûþýÙªÑ¸¹´á¸Ë¸0LFURVRIW $36•,%0
1HWH]]D•$PD]RQ5HGVKLIW¹•ŠYgåÕÂÙªÑéûÕÚç»ªò¸,PSDOD<3†•+D
%LJ4XHU\¹‰ˆ• ÕcÙªÑëªË }ÉËØò±ÇïªÚv…c•x‘

‰ ªÁËØªÉù Ž•0HWKRGV OLEUDU\k¶ ËÚªýt« üt«ªêŠ†x‘r«kÿª½ýðÉ ¸
6HUYHUk üt«ªüôªÚÇªâªjŠ*®¨y•5k¶ ËÚªýt«…cª´Wka©•x‘Z †a«•56WX
/LEUDU\†Ž-DYDk¶ ËÚªýt«…cª´Wka©•x‘ ‰ ªÁËØªÉù ŽÙªÑëªËÇªâªŠ–

3X9X8

2+'6,5ãÕÃªÊ± üxª~ Š´W‰ »±¶ ËÚªýxª~ •,‡Žã•þ©†x‘¶ ËÚªýxª´Wkaª

5Ž•C:Å æ÷ªØµ Â »†x‘B :‰øªÈ¶ Ñªç¹ªË‡v…)ŠÅð Ûû¶ ¶ Ñªç¹ªË±%
5WRROVŽ•:LQGRZV†5ãÕÃªÊ±ïªËj¨åýÛxªcŠ´W‰éÿÂûòÆò†x‘
56WXGLRŽ•5±¬c£xnxª,'(¸• ŸY »¹†x‘ÅªÛºÙµÑ•ÙâÕÂ•i§’åÊ÷´ýÖªý
-DYDŽ•2+'6,5ãÕÃªÊ•Æü•Å ïªß Ú•Ág•ÙªÑëªË—•–)Š´W‰Å ïªß Ú±

¤R†Ž•:LQGRZV »†•|«}«•¶ ËÚªýce±œâv•x‘

:LQGRZV†Ž•5‡-DYDŽˆ€¨¡ åÕÚ‡ åÕÚ•´ª¿ØÁÓõka©•x‘5±œc•´ª¿

5•¶ ËÚªý

KWWSV FUDQ U†S•URMHŠFW«ªSUeJŠÄ'RZQORDG
5IRU:LQGRZVÅ•ÄEDVHÅ•‡ŠÁüÕÁv•Ò¸ ÿªÛv…n•tc‘

Ò¸ ÿªÛk ™v~¨•¶ ËÚªû± üv•x‘ ƒ•Á¤±œc…•x˜…Ùç»ýÚ•¼éÉù ±¬Qv.
•y•éÿÂûòç³¶ýŠŽ¶ ËÚªýv‰cck§c†v¦e‘[®©Š•:
•§eŠ•&Ûû¶è•Çèç»ýÒ‡v…=±thv•x‘ãŠ•5‡-DYD•´ª¿ØÁÓõ•½cŠ§ª
•§eŠ åÕÚ´ª¿ØÁÓõ±×ÄŠv•x‘

™xª‡•ËÑªÚóÝ÷ªj¨5±Âh†lª§eŠ‰©•x‘

2 + ' 6 , ‰ Ö ª ý

) L J X U H & 5 $ 1 j ¨ • 5 • Ò ¸ ÿ ª Û

) L J X U H 5 ç » ý Ò • ™

0HWKRGV /LEUDU\

)LJXUH åÕÚâªÊù •5±×ÄS

5WRROV•¶ ËÚªý

KWWSV FUDQ U ŠÚÁÉÄÍÄ WRZQORDG 5 IRU :LQ
GRZVÅ±ÁüÕÁv•ãŠÄ5WRROVÅ±ÁüÕÁv…•M q•5WRROV±Ò¸ ÿªÛv•x

Ò¸ ÿªÛk ™¯•¶ ËÚªû± üv•x‘x˜…Ùç»ýÚ•¼éÉù ±Âhv•x‘

56WXGLR•¶ ËÚªý

KWWSV ZZZ UVWXŠGÁÉÄÍÄ RZQORDG 56WX
GLRÅ•~Ž56WXGLR•R•ÄÒ¸ ÿªÛÅîÑ ±ÁüÕÁv•×Ÿq±Âhv•:
Šêt«ª§eŠ:LQGRZVQ•¶ ËÚªû±Ò¸ ÿªÛv•x‘

Ò¸ ÿªÛk ™¯•¶ ËÚªû± üv•x‘x˜…Ùç»ýÚ•¼éÉù ±Âhv…n•tc‘

-DYD•¶ ËÚªý

KWWSV MDYD FRP HQ GRZŠÚÁÉÄÍÄ :PDQXDO MVS
Šêt«ª§eŠ•:LLQGRZV åÕÚq•¶ ËÚªû±Âhv•x‘ åÕÚq•5¡¶ ËÚªýv…c

Ò¸ ÿªÛ¯•¶ ËÚªû± üv…•x‘

2 + ' 6 , ‰ Ö ª ý

) L J X U H 5 6 W X G L R • Ò ¸ ÿ ª Û

) L J X U H - D Y D • Ò ¸ ÿ ª Û

¶ Ë Ú ª ý • Â ó

r « † € ¡ Ž j , ~ Ž y † x k • ü • ~ Â ó v … i l • v ¦ e ‘ 5 ± _ ž v • R] • § e Š Ñ ¶ é v … n • t c ‘

```
B M b i  H H X T  U F  ]a[;H2b2M/V2`]
H B # `  U a [ H _ 2 M / 2 `  V
i `  M b H U 2  ]a1G1*h hPS Ry  6_PJ- T2QbQ;M2bV[H]
```

O O (R)] a 1 G 1 * h 6 _ P J T 2 ` b Q M G A J A h R y c]
O O i i ` U -] b [H . B H 2 + i] V
O O (R)] T Q b i ; ` 2 b [H]

r • * P Ž - D Y D ± ¬ Q x ª • † • x ˜ … k e • n c p • • 5 ‡ - D Y D • œ c k o v n ¶ Ë Ú ª ý t « … c ª

¡ e Æ ƒ • Ø Ë Ú Ž • Ï ª Ë ã Õ Ã ª Ê k å ý Û † l ª j ˆ e j ± Â ó x ª r ‡ † x ‘ ¤ R • 5 Å ª Û ± ü v … • 2
* L W + X E ü ï Ê Ú ü j ¨ * Q ? Q ` i J 2 ã ? Ã Q Ã ª Ê ± ¶ Ë Ú ª ý v • x •

```
B M b i  H H X T  U F  ]`;2]V
/ `  i , , //_2TUQP>.aV]
B M b i  H H X T  U F  ]*Q2bQ`iJ2iVQ/]
```

H Ÿ © Š

3 X 8

$ 7 / $ 6 £ 0 H W K R G V / L E U D U \ ± 0 Ÿ 2 + ' 6 , Ö ª ý Ë Ñ Õ Á Ì D ± à Ð † H Ÿ x ª r ‡ Ž • • í Š * ß %
: H E 6 H U Y L F H V $: 6 † x ‘

3 X 8 X "R Q / b 2

% U R D G Ž V R Ð N H U Å Ø Ü k q ± ¬ Q v … c • x ‘ 2 + ' 6 , Ö ª ý Ž § . * $ ‡ ‡ ¡ Š • ' R F N H U ¶ ó ª Ê ‡ •
: L Q G R Z V • 0 D F 2 6 • / L Q X [‰ ˆ • š ‡ ² ˆ • ¼ ì þ ª Ø µ Â É Ë Ø ò † x Q Z † x ‘ % U R D G V H D
' R F N H U ¶ ó ª Ê Š Ž • 0 H W K R G V O L E U D U \ £ $ 7 / $ 6 ± 0 Ÿ) ‰ 2 + '
6 , Ö ª ý k 0 • « … c • x ‘

3 X 8 X k K x Q M q a

$ P D] R Q Ž • $: 6 Á û ¸ Û Å æ ÷ ª Ø µ Â » † î Ñ ± Á ü Õ Á x ª • p † ¶ Ë Ñ Ë S † l ª á ƒ • » ± Q ¯
L Q D % R [2 [+ ' 6 , R Q $: 6 † x ‘

2 + ' 6 , L Q D % R [Ž ± Š Ì L » ‡ v … t h t ~ ¡ • † a © • 2 + ' 6 , Å ñ ÷ Ý Ø µ k % ¯ x ª š ‡ ² ˆ • Ó ÷ ª Ú
& ' 0 Ù ª Ñ ð Õ æ Â ‡ (7 / Ö ª ý ¡ • 2 + ' 6 , L Q D % R [Š 0 • « … c • x ‘ 2 + ' 6 ,
L Q D % R [• ´ ª ¿ Ø Á Ó õ Ž • : Š ê t « … c • x ‘

)L J X U H 2 + ' 6 , L Q D % R [• $ P D] R Q : H E 6 H U Y L F H V ´ ª ¿ Ø Á Ó õ

2 + ' 6 , R Q $: 6 Ž • = Ö È p • ð ý Ó ø ª È ª F 6 • ¼ × g £ G á ¦ g Š % « ~ 2 + ' 6 , » • ~ • ü ç ³ þ Ë ´ ª ¿
P • Ç é ý Ù ª Ñ Í Õ Ú k 0 • « … i © • à • c • ê ý Ë Ã ´ Ù ª Ñ ± í ž : Š ÿ ª Û x ª r ‡ ¡ Z † x ‘
Ù ª Ñ Ž $ P D] R Q 5 H G V K L I W Ù ª Ñ ë ª Ë é û Õ Ú ç » ª ò Š ' ª t « • 2 + ' 6 , Ö ª ý Š § , … Ç ï ª Ú t «
$ 7 / $ 6 • º (W b Ž 3 R V W J U H 6 4 / Ù ª Ñ ë ª Ë Š C . t « • x ‘ ø ª È ª Ž ç ÿ Ú º Û † • ¸ ¹ è ¶ Ñ ª ç ¹ ª

K W W S V J L W K X E F R P 2 + ' 6 , % U R D G V H D
K W W S V Z Z Z G R F N H U F R P
K W W S V J L W K X E F R P 2 + ' 6 , 2 + ' 6 , L Q D % R [
K W W S V J L W K X E F R P 2 + ' 6 , 2 + ' 6 , R Q $: 6

2 + ' 6 , ‰ Ö ª ý

6 H U Y H U ± Ü Q ¹ ± þ w … $ 7 / $ 6 £ 5 6 W X G L R Š ´ Á Í Ë † l • x ‘ 5 6 W X G L R Š Ž 2 +
0 H W K R G V / L E U D U \ k x † Š ¶ Ë Ú ª ý t « … i © • Ù ª Ñ ë ª Ë — • –) Š ¬ Q † l • x ‘ 2 +

) L J X U H 2 + ' 6 , R Q $: 6 • $ P D] R Q : H E 6 H U Y L F H V ´ ª ¿ Ø Á Ó õ

3 X e

```
m & ' 0 Ð • Ù ª Ñ Š F v …  ‰ ± ü e Š Ž
      ½ Ë Ñ ò Å ª Û ± t h x ª
      2 + ' 6 , 0 H W K R G V / L E U D U \ • 5 ã Õ Ã ª Ê ± ¬ Q v ~ Å ª Û ± t h x ª
      ¶  Ñ û Á Ø µ è ‰  ‰ é û Õ Ú ç » ª ò $ 7 / $ 6 ± ¬ Q x ª
m 2 + ' 6 , Ö ª ý Ž t • u • ‰  ‰ © Š ± Q c … c • x
      … Æ p –
      ü ´ ý Ñ ¶ ò Á º ü
      ] \ ñ ´ Ü ü Ø µ Á Ë
m 2 + ' 6 , ´ Ü ü Ø µ Á Ë Ö ª ý • š ‡ ² ˆ Ž • ¤ R Š à ž & • « … c • x
      ¶  Ñ û Á Ø µ è ‰  ‰ é û Õ Ú ç » ª ò $ 7 / $ 6
      2 + ' 6 , 0 H W K R G V / L E U D U \ • 5 ã Õ Ã ª Ê
m 2 + ' 6 , Ö ª ý • H Ÿ ± F ± Š x ª c n ƒ j • © Š k a © • x ‘
```

*? Ti2` N

aZG _

Ì 0DUWLMQ 6FKXHPLH 3HWHU 5LMQEHHN

´ þ Ù ª Ñ ô Ù ý ¸ & ' 0 ¹ Ž ü þ ª É ù Ü ý Ù ª Ñ ë ª Ë ô Ù ý † x ¸ x ˜ … • Ù ª Ñ Ž ç µ ª ý Û ± á ƒ Ø ª è ý • þ Å ª
6 4 / 6 H U Y H U ‰ ˆ • Ï ç Ú ¸ ¹ ´ é û Õ Ú ç » ª ò ± ¬ Q v … ü þ ª É ù Ü ý Ù ª Ñ ë ª Ë Š C . t « • x ‘ $ 7 / $ 6
/ L E U D U \ ‰ ˆ • t • u • ‰ 2 + ' 6 , Ö ª ý Ž • â Õ Á Â û ¸ Û † Ù ª Ñ ë ª Ë Š Á º ü ± p x r ‡ † ž t v • x k • < ”
Ö ª ý Ž 5 n • ê • ø ª È ª k Ù ª Ñ ± < ” Š ‰ † I ª § e • ¾ ¶ Û x ª § e Š ™ C t « … c ª ~ • Ù ª Ñ ë ª Ë ± ð –

ü þ ª É ù Ü ý Ù ª Ñ ë ª Ë ± Á º ü x ª À € : ‰ Š ¶ Ž 6 4 / ¸ 6 W U X F W X U H G
4 X H U \ / D Q J X D J H ¹ † • Á º ü £ Ù ª Ñ 7 á Š ¬ Q † I • x ‘ 6 4 / • B Å ð Û Ž Â j Š À € S t « … i © • Ï
Á g • • 6 4 / 6 H U Y H U ã • 3 (5 6 2 1 Ø ª è ý • M ‰ • ü ± * ® x ª Š Ž • ã • § e Š ì - v • x ‘ •

```
a 1 G 1 * hP S R y 6 _ P J 2 ` b Q M c
```

Æ c • 3 R V W J U H 6 4 / † Ž Ÿ w Á º ü Ž ã • § e Š ‰ © • x •

```
a 1 G 1 * h6 _ P J 2 ` b Q G M A J A R y c
```

2 + ' 6 , † Ž • é û Õ Ú ç » ª ò ‘ - • Å ‡ Š § . v ‰ c r ‡ ± , ² † c • x ‘ x ‰ ® € • x ˜ … • 2 + ' 6 , Ù ª Ñ ë ª Ë † Ÿ
2 + ' 6 , 6 4 / Ž) Š 6 4 / 6 H U Y H U 6 4 / Å ‡ • Ç è Í Õ Ú † x ‘ Í † Á ê x ª 6 4 / Ž x ˜ … 2 + ' 6 ,
6 4 / ± ¬ Q v • x ‘

º Ù ª Ñ ë ª Ë é û Õ Ú ç » ª ò Š Ž • 6 4 / ± ¬ Q v ~ Ù ª Ñ ë ª Ë • Á º ü • ~ • ¶ í • Ï ç Ú ¸ ¹ ´ Ö ª ý ¡ Ý & v … c •

64/‡5

|•~ •&'0Š€¢v~ÙªÑëªËŠFv…2+'6,Öªý±¬QzyŠÁºü± ü†l•xk•@§t«

Í†Ž•· k64/•B :‰}ˆ±v…cªr‡±È%‡v…c•x‘•y•6TO5HQGHU‡'DWD
†Ž•&'0ŠÁºü±px~ •64/¸r•ê 2+'6,64/¹±¬Qxªce±œâv•x‘ã•ÍÁÉù †

NXRa[H_2M/2`

6TO5HQGHUÃÕÃªÊŽ &5$1¸&RPSUHKHQVLYH 5 $UFKLYH 1HW
ZRUN¹†ì,Z †a©•¤R•Åð Û†¶ ËÚªý†l•x•

```
BMbi HHXT ŁJFa[;H2b2M/V2`]
```

6TO5HQGHUŽ•^h•ÙªÑëªËÉËØò¸3RVWJUH64/•0LFURVRIW
64/ 6HUYHU•64/LWH•2UDFOH¹£(ÑÙªÑ¸¹´á¸Ë¸0LFURVRIW
$36•,%01HWH]]D•$PD]RQ5HGVKLIW¹ŠYg•åÕÂÙªÑéûÕÚç»ªò
j¨,PSDOD•*RRJOH%LJ4XHU\¹‰ˆ• Õckqéû Õ Úç»ªò±ÇïªÚv…c•x

NXRXaZG

ãÕÃªÊ•Q •'‡ƒŽ•64/•ãûóªÑS±ÇïªÚxªr‡†x'
v•v••cnƒj•ãûóªÑŠB„c…•64/•¬t‰âüºªÉù ±rhxª´Wka©•x'6TO5H

ãûóªÑ¡•ª

! Þ±¬Qv…•þ Òü ÂâŠ c•ãûóªÑ¡‡ª xª´Wkaªãûóª Ñà±êv•x'¤R•Á†
‡ce7Pk64/†ŠŠt«…c•x' `2M/2P••'pv†Ž•r•ãûóªÑ•¡k qt«…c•

```
b[Hl@]a1G1*h 6_PJ +QM+2Ti q>1_1 +QM+2TinB/ 4 ! c]
`2M/2Ub[H- 4 Rkv
```

OO (R)]a1G1*h 6_PJ +QM+2Ti q>1_1 +QM+2TinB/ 4 Rkjc]

š‡²ˆ•ÙªÑëªË }ÉËØòk%¯xªãûóªÑS‡Žµ‰©•Øªèýà£çµªýÛà±¡‡ŸM

```
b[Hl@]a1G1*h 6_PJ !t q>1_1 T2`bQMnB/ 4 ! c]
`2M/2Ub[Ht- 4 ]Q#b2`p iBQM4]Rkv
```

6TO5HQGHU

OO (R)]a1G1*h 6_PJ Q#b2`p iBQM q>1_1 T2`bQMnB/ 4 Rkjc]

ãûóªÑ¡Ž•P¡• ÞÑ•èªü´ 7P•ëÁÚý¸½ ðö”©•üËÚŠ7 t«ª¹‡xªr‡k†l•x‘•

```
b[Hl @]a1G1*h 6_PJ +QM+2Ti q>1_1 +QM+2TinB/ AL U! Vc]
`2M/ÞU`b[H- 4 +URkj kj9- j98VV
```

OO (R)]a1G1*h 6_PJ +QM+2Ti q>1_1 +QM+2TinB/ AL URkj-kj9-j9

,I 7KHQ (OVH

âŠŽ• ƒ•~Ž P•ãûóªÑ•¡ŠB„c…ÅªÛèÿÕÁ±¼ •~Ž¼çŠxª´Wka©•x‘r«Ž•
\ &B7 i`m2' , &B7 ÅHbÞ'v…üc•x‘FRQGLWLRQ k WUXH
•~Ž •ê •LI WUXH èÿÕÁk¬Qt«•|«¤¤•ê Ž LI IDOVH
èÿÕÁk¸.exªê ¹Åêt«•x‘

```
b[Hl @]a1G1*h 6_PJ +Q?Q`i &!t' \ &q>1_1 bm#D2+inB/ 4 R']
`2M/ÞU`b[Ht- 4 6 GaV1
```

OO (R)]a1G1*h 6_PJ +Q?Q`i]

```
`2M/ÞU`b[Ht- 4 h_lV
```

OO (R)]a1G1*h 6_PJ +Q?Q`i q>1_1 bm#D2+inB/ 4 R]

…‰•Ç¡Çïª Út«…c•x•

```
b[Hl @]a1G1*h 6_PJ +Q?Q`i &!t 44 R' \ &q>1_1 bm#D2+inB/ 4 R'c]
`2M/ÞU`b[Ht- 4 RV
```

OO (R)]a1G1*h 6_PJ +Q?Q`i q>1_1 bm#D2+inB/ 4 Rc]

```
`2M/ÞU`b[Ht- 4 kV
```

OO (R)]a1G1*h 6_PJ +Q?Q`i c]

AŸ¶¡Çïª Út«…c•x•

```
b[Hl @]a1G1*h   6_PJ +Q?Q`i &!t AL UR-k-jV' \ &q>1_1 bm#D2+inB
`2M/2U`b[Ht- 4 kV
```

OO (R)]a1G1*h 6_PJ +Q?Q`i q>1_1 bm#D2+inB/ 4 Rc]

NXRXkaZG

6TO5HGÃÕÃÚÊ•¡eÆƒ•Q Ž•2+'6,64/j¨4•64/Å‡—7 xªr‡†x‘Ág••

```
b[Hl @]a1G1*h hPS Ry   6_PJ T2`bQMc]
i` MbH U2b[Hi- `;2i.B H2+i T4Qbi;`2bV[H]
```

OO (R)]a1G1*h 6_PJ T2`bQM GAJAh Ryc]
OO ii`U-]b[H.B H2+i]V
OO (R)]TQbi;`2b[H]

i `;2i.B H2+i ãûóªÑŠŽã•¡k™ Z †x•kRUDFOHp
kSRVWJUHVTOp kSGZp kUHGVKLIWp kLPSDODp kQHWH]]
TXHU\p kVTOLWHp kVTO VHUYHUp‘

> 64/*P£å ±<”Š7 †lªy©ŠŽŒškа©•x‘|•}3Ž•ãÕÃªÊŠŽŒ¨«~P•7
> 64/k¶í• vc64/cŠ‡v…ŸYt«~)‰}3†x‘vjv•Z ‰Œ©• ¶•LYâ±œp‘
> 6HUYHU•å Š^e§eŠv…c•x‘

M]Œ•h-±0nv…¡•Çïª Út«…cªx˜…•éûÕÚç»ªòã†ºûª‰n üt«ª2+'6,
64/±]bxªŠŽ•õ–x˜lPkcnƒja©•x‘¤R†Ž•r«¨•õ–Ú Šƒc…ÚvnœâV•x‘

7UDQVODWHkÇïªÚxª*P‡å

r«¨•64/6HUYHU*PŽØËÚW†a©•ºÅ‡—•oÂ‰7 kÂót«…c•x•

7DEOH kWUDQVODWH ¡ Š§©ÇïªÚt«ª*P‡å

*P	*P	*P
$%6	(;3	5$1'
$&26	)/225	5$1.
$6,1	*(7'$7(	5,*+7
$7$1	+$6+%<7(6	5281'

6 T O 5 H Q G H U

* P	* P	* P
$ 9 *	, 6 1 8 / /	5 2 : B 1 8 0 % (5
& $ 6 7	, 6 1 8 0 (5 , &	5 7 5 , 0
& (, / , 1 *	/ ( ) 7	6 , 1
& + $ 5 , 1 ' (;	/ (1	6 4 5 7
& 2 1 & $ 7	/ 2 *	6 4 8 $ 5 (
& 2 6	/ 2 *	6 7 ' (9
& 2 8 1 7	/ 2 : (5	6 8 0
& 2 8 1 7 B % , *	/ 7 5 , 0	7 $ 1
' $ 7 ($ ' '	0 $;	8 3 3 (5
' $ 7 (' ,))	0 , 1	9 $ 5
' $ 7 ( ) 5 2 0 3 $ 5 7 6	0 2 1 7 +	< ($ 5
' $ 7 (7 , 0 ( ) 5 2 0 3 $ 5 7 6	1 (: , '	
' $ <	3 ,	
(2 0 2 1 7 +	3 2 : (5	

2 U D F O H † Ž ± 2 ‰ l Œ k ´ W † x ‘ 6 4 / L W H † Ž Ÿ ‰ • ¡ • k a © • z ² ‘ Ÿ M Š • 5 n • 6 4 / å å k Ç ï ª Ú t « … c • x ‘ ¤ R Ž • o Â Š ¡ t « ª r ‡ k Â ó t « … c ª • ø e : ‰ ü Ë Ú † x •

```
@@ aBKTH2 b2H2+ib,
a1G1*h6_PïJ #Hc2

@@ a2H2+ib rBi? DQBMb,
a1G1*h6_PïJ #H2nARL L1C_P AiL#H2nFkL 4 #c

@@ L2bi2/ [m2`B2b,
a1G1*h6_PïJa1G1*h6_PïJ #H2nRV ¡k T_14 #c

@@ GBKBiBM; iQ iQT `Qrb,
a1G1*hhP SRy 6_PïJ #Hc2

@@ a2H2+iBM; BMiQ M2r i #H2,
a1G1*hA L hM2rni #6H_2PïJ #Hc2

@@ *`2 iBM; i #H2b,
*_1 hh "G1 #H2U7B2HL Vc

@@ AMb2`iBM; p2`# iBK p Hm2b,
ALa1_AL hQi?2`ni #H2 U7B2HG/h1RV c
```

```
@@ AMb2`iBM; 7`QK a1G1*h,
ALa1_AhLhPQi?2`ni #H2 U7Ba1G1*hPVHm62_PJ #Hc2

@@ aBKTH2 /`QT +QKK M/b,
._PSh "G1 #Hc2

@@ .`QT i #H2 B7 Bi 2tBbib,
A6P"C1*hnA*bUAGG1an M-HvbBbaLPhLlGG
  ._PSh "G1*>AGG1an M HvbBbc

@@ .`QT i2KT i #H2 B7 Bi 2tBbib,
A6P"C1*hnA2bUT/#XXO+Q?QVbAaLPhLlGG
  ._PSh "G1O+Q?Q`ibc

@@ *QKKQM i #H2 2tT`2bbBQMb,
qAh>i2 aUa1G1*h6_PJ #HV2a1G1*h6_PJ i2c

@@ Po1_ +H mb2b,
a1G1*hPqnLlJUVPo1US _hAhA'PLP_.1"u#V
  a]_Qr LmK#62_PJ #Hc2

@@ * a1 q>1L +H mb2b,
a1G1*ha1q>1L4Rh>1L 1Gay1 1L. a p Hm62_PJ #Hc2

@@ lLAPLb,
a1G1*h6_PJ lLAPa1G1*h6_P#c

@@ ALh1_a1*hAPLb,
a1G1*h6_PJ ALh1_a1*hG1*h6_P#c

@@ 1s*1Sh,
a1G1*h6_PJ 1s*1Sah1G1*h6_P#c
```

Þ Ñ • Þ W

Þ Ñ • Þ W Ž • 6 4 / 6 H U Y H U k 4 • Š ¶ § © ¡ ± µ † Ž ‰ c ¬ Â • ƒ † x ‘ 6 4 /
6 H U Y H U k Ž • G 1 * h 7 B `b i n M K 2 Y ^ ^ Y H b i n M K 2 a 7 m H H n M K
6 _ P J i # H ‘ 2 • x k • r « Ž 3 R V W J U H 6 4 / ‡ 2 U D F O H a 1 Ž G 1 * h
7 B `b i n M K 2 % % ^ ^ % % H b i n M K 2 a 7 m H ‰ ‰ p M • K ‰ © 6 z _ P 6 T i O # 5 H
a 1 G 1 * h 7 B `b i n M K 2 Y H b i n M K 2 a 7 m H H a n M ê K • 2 6 T O F 5 J H Q G H 2
| • ~ • a 1 G 1 * h H b i n M K 2 Y * a h U ; 2 a o _ * > _ U j V V a 7 m H

6 T O 5 H Q G H U

6 _ P J i #¡H2n¡ t«•x‘ ªt±œpª~ Š• ƒ¤ã• ÞÑ±ÞWxªê Ž• *P L
*P±¬Qxª•kMÉ•ce†x‘

Øªèýº¶ü´Ë‡$6¿ª ªÛ

5n•64/Å‡†ŽØªèýº¶ü´Ë± qxªcŠ a¿ª ªÛ±¬Q†l•xk•¿ª ªÛ‰v†¡ `‰
6HUYHU•3RVWJUH64/•5HGVKLIW‰ˆ†Ž `‰nžtv•x•

```
@@ lbBM; a F2vrQ`/
a1G1*h
6_PJKvni #H2i #H2nR
ALL1C_PAUL
  a1G1*h6_PQi?2`ni #H2
V ai #H2nk
PLi #H2nRXT2`b4QiM#nHB2nkXT2`bQMnB/c

@@ LQi mbBM; a F2vrQ`/
a1G1*h
6_PJKvni #H2 i #H2nR
ALL1C_PAUL
  a1G1*h6_PQi?2`ni #H2
V i #H2nk
PLi #H2nRXT2`b4QiM#nHB2nkXT2`bQMnB/c
```

vjv•2UDFOH†Ža¿ª ªÛ±¬Qxª‡ºûªkYrv•x‘ã]•Á†Ž•M‰•ÁºüŽü v•x‘|•~ •Øª
a¿ª ªÛ±¬Qv‰crª‡±@§v•x‘¸Á•6TO5HQGHU†Ž2UDFOHk
•¬Q±¦Zv…c‰cØªèýº¶ü´Ë‡2UDFOHk a•¬Q±W‘v…cªçµªýÛº¶ü´Ë±ö2†l

Ø ïØªèý

Ø ïØªèýŽº(Wb±C.xª•Š•íŠ-Q†a©•ovn¬Qxª‡Áºü•ãç»ªð Ë±] ŠÈãtzªr‡k†

µ‰ªøªÈªj¨•Øªèýk³ v‰cšeŠ•Øªèýà Šû Òò‰ ÞÑ±ûYv•x‘
Ø ïûüØªèýktht«ªËª¿ªð±øªÈªk¼ †lª§eŠv•x‘

Ág••

```
b[Hl @]a1G1*h 6_PJ O+?BH/`2Mc]
i` MbHUb[Hi- `;2i.B H2+i]Q` +H2Q` +H2h2KTa+?2K ]?2KTa+?2K ]
```

OO q `MBM;, h?2 ^Q` +H2h2KTa+?2K ^ `;mK2Mi Bb /2T`2+ i2/X lb
OO h?Bb r `MBM; Bb /BbTH v2/ QM+2 2p2`v 3 ?Qm`bX

64/‡5

OO (R)]a1G1*h 6_PJ i2KTnb+?2K XF[?K`xRD+?BH/`2M c]
OO ii`U-]b[H.B H2+i]V
OO (R)]Q` +H2]

øªÈªŽ i2KTnb+?Š2‘K&žlŒ±á,…cª´Wka©•x‘

•~•2UDFOH†ŽØªèýà•ëtk ÞŠcŒt«…cª~ •Ø ïûüØªèýàŽM] Þ¤

t¨Š•2UDFOH†ŽØ ïûüØªèýŽíž:Šuœt«‰c~ •¬Q¯Šâê:Šx˜…•Ø ïû
h_lL* h$1’ ._PŠ…•“‡v~Øªèýk2UDFOH•ÆâË¿ªðŠ³‹v‰c§eŠxª´W

Ÿý•+7

64/6HUYHUk4•Š¶§©¡±µ†aªP-‰cP• ƒŽ•Ÿý•+7 k¦Zt«…cªr‡†x‘Á
6HUYHU†žtv•x•

```
*_1 hh "GDi2KT Uiti_*>V_c

ALa1_AhLhPi2KT
a1G1*hRc

a1G1*h6_PDi2KT>1_iti 4 Rc
```

itik9$5&+$5çµªýÛ†•|«±jP‡•Çv…cª‡v…¡•64/
6HUYHUŽ•Ç±Z Šxª~ Š• ƒ•e€• ƒ±íž:Šovc+Š7 v•x‘r«ŠFv…•3RV
v~k,…•¿õËÚŽíŠâê:Šüe´Wka©•x‘ã]•Á†Ž•M¯•ËØªÚó Ú±¤R•cy«j

```
a1G1*h6_PDi2KT>1_iti 4* abR a o _*>V_c
```

•~Ž

```
a1G1*h6_PDi2KT>1_*1 abiti a ALV 4 Rc
```

ÞÑ•ÇŠipª] Þ”¬ Þ•ö2

64/6HUYHU‰^•Æü•'%06éûÕÚç»ªòŽíŠ] Þ‡¬ Þ±ö2zyŠ ÞÑ•Ç±ü

```
a1G1*h6_PJ+QM+2Ti>1_+1QM+2Tin+H bbnB/ 4^*HB/MB+ H 6BM/BM;^
```

6 T O 5 H Q G H U

[® © Š ¤ R • § e Š] r x ª r ‡ k @ § t « • x •

a 1 G 1 * h 6 _ P J Q M + 2 T i 1 _ G P q U ± Q M + 2 T i n + H b n B W 4 1 + B M B + H 7 B M / B M ; ^

Ë ¿ ª ð ‡ Ù ª Ñ ë ª Ë

6 4 / 6 H U Y H U † Ž • Ø ª è ý Ž Ë ¿ ª ð Ð Š a © • Ë ¿ ª ð Ž Ù ª Ñ ë ª Ë Ð Š a © • x ‘ Á g • • + / K n
Ž + / K n / Ù ª Ñ ë ª Ë Ð • / # Q Ë ¿ ª ð Ð • T 2 ` b Q M è ý ± ¼ v • x ‘ 4 • Š ¶ † ¡ Ÿ M • ð k . e v • x k
6 H U Y H U † Ž • Ù ª Ñ ë ª Ë s ‡ Š þ í ƒ • Ë ¿ ª ð ¸ / # Q • • « ª r ‡ k 5 c ¹ k . e v • ø ª È ª Ž µ ‰ ª Ù ª
6 H U Y H U • Ù ª Ñ ë ª Ë Š † x ª ¡ • k Ë ¿ ª ð † a ª ‡ Š g • x ‘
| • ~ • 6 4 / 6 H U Y H U • Ù ª Ñ ë ª Ë ‡ Ë ¿ ª ð ± ƒ • ã û ó ª Ñ Š W x ª r ‡ ± @ § v • x ‘ þ í • r « ±
! / i # b 2 a + ? 2 K • x ‘ Á g • • ã û ó ª Ñ S t « ~ 6 4 / † Ž ã • § e Š ‰ © • x •

a 1 G 1 * h 6 _ P J / i # b 2 a + ? 2 K X T 2 ` b Q M

6 4 / 6 H U Y H U † Ž • ¡ Š Ù ª Ñ ë ª Ë à ‡ Ë ¿ ª ð à • œ c ± 0 ª r ‡ k † l • x • / i # b 2 a + ? 2
4] + / K n / i X / # Q ê ¡ û Õ Ú ç » ª ò † Ž • Ÿ w Å ª Û ± ¬ Q v • ã û ó ª Ñ ¡ ‡ v … Ë ¿ ª ð • ž ± ¼ v • x •
4] + / K n / ï]

r • c e k ü x ª # Æ • ó Å Ž l a 1 Å ð Û ± ¬ Q v ~ ê † x ‘ l a 1
+ / K n / i X / # Q ª k Y r v • x ‘ v ~ k , … • í Š Ù ª Ñ ë ª Ë Ë ¿ ª ð ± ¼ v … Ø ª è ý • ê Š ± ê x § e Š v •
Å ð Û • ¬ Q ± œ p ª r ‡ ± @ § v • x ‘

ã û ó ª Ñ S t « ~ 6 4 / • Ù â Õ Â

ã û ó ª Ñ S t « ~ 6 4 / • Ù â Õ Â Ž - v Œ Š ‰ ª r ‡ k a © • x ‘ þ Ò ü Â t « ~ 6 4 / • ž k Ù ª Ñ ë ª Ë Ç ª â ª

Ï ª Ë • 6 4 / ± ¶ Ñ û Á Ø µ è Š : X v • þ Ò ü Â i § ’ ¡ t « ~ 6 4 / ± r h x ª ~ • 6 K L Q \ ´ é ü k
6 T O 5 H Q G H U ã Õ Ã ª Ê Š 0 • « … c • x ‘ r • ´ é ü Ž ã • c e † _ ž † l • x •

H m M + ? a [H _ 2 M / 2 ` U 2 V p 2 H Q T 2 `

r « Š § © • : Š ê x § e Š • ´ é ü k Ù ç » ý Ú • è û ¸ È † Ÿ l • x ‘ ´ é ü Ž ¸ ¹ è ã † ¡ Â Ÿ t « … c • x

r • ´ é ü † Ž • 2 + ' 6 , 6 4 / ± ì - v • Ñ ª Ä Õ Ú • c Š ± Â h v • 6 4 / Š Å ê t « ª ã û ó ª Ñ • ¡ ± ì - x ª ‡ • ¡ k í ž

K W W S G D W D R K G V L R U J 6 T O ' H Y H O R S H U

) L J X U H 7 K H 6 T O ' H Y H O R S H U 6 K L Q \ D S S

N X k. i # b 2 * Q M M 2 + i Q `

' D W D E D V H & R Ž • Q H F W R U % & Û û ¶ â ± ¬ Q v … t • u • ‰ Ù ª Ñ ë ª Ë é û Õ Ú ç » ª ò
5 $ U F K L Y H 1 H W Z R U N ¹ † ì , Z † • ã • § e Š ¶ Ë Ú ª ý † l • x •

```
B M b i H H X T U F i 2 # b b 2 * Q M M 2 + i Q ` ]
```

' D W D E D V H & R Q Q H F W R U Ž • ^ h • Ù ª Ñ ë ª Ë É Ë Ø ò ¸ 3 R V W J U H 6 4 / • 0 L F
6 4 / 6 H U Y H U • 6 4 / L W H • i § ' 2 U D F O H ¹ • (Ñ Ù ª Ñ ¸ ¹ ´ á ¸ Ë ¸ 0 L F U R V R I W
$ 3 6 • , % 0 1 H W H]] D • $ P D] R Q ¹ • ‰ ¨ ' Š å Õ Â Ù ª Ñ é û Õ Ú ç » ª ò ¸ + D G R R S ±
% L J 4 X H U \ ¹ ‰ ˆ • Õ y ‰ k q é û Õ Ú ç » ª ò ± Ç ï ª Ú v … c • x ' r • ã Õ Ã ª Ê Š Ž x † Š š ‡

N X k X R

Ù ª Ñ ë ª Ë Š –) x ª Š Ž • Ù ª Ñ ë ª Ë é û Õ Ú ç » ª ò • Ç ª â ª • ¦ ª • ø ª È ª à • ã Ë ª Û ‰ ˆ • 5 n

```
+ Q M M 2 + Q M M 2 # K b ] 4 T Q b i ; ` 2 b - [ H ]
                b 2 ` p 2 ` ] 4 H Q + H ? Q b i f T Q - b i ; ` 2 b ]
                m b 2 ` ] 4 D Q - 2 ]
```

'DWDEDVH&RQQHFWRU

```
T bbrQ`/]4 42+`2-i]
b+?2K ]4 4/KV]
```

OO *QMM2+iBM; mbBM; SQbi;`2aZG /`Bp2`

º é û Õ Ú ç » ª ò Š ´ W ‰ Ú _ î Y Š ƒ c … Ž • \ + Q M M 2 + 4 E n • t c ‘ –) ± * w ~ r ‡ ± ´ y Â ó n • t c

```
/Bb+QMM2+QMMV
```

Ç ª â ª à ± ¼ x ª [® © Š • - ' % & –) Þ Ñ ± % ¯ x ª r ‡ ¡ Z † x ‘ t ¨ Š > x ‰ ê Ž • r € ¨ ± ¬ Q x ª r ‡ ¡ † l

```
+QMMai`BM;/#+,TQbi;`2b[H,ffHQ+ H?Qbi,89jkfTQbi;`2b]
+QMM@+QMMU2#+Kb 4TQbi;`2b-[H]
+QMM2+iBQMai`BM;-
mb2` ]4DQ2]
T bbrQ`/]4 42+`2-i]
b+?2K ]4 4/KV]
```

OO *QMM2+iBM; mbBM; SQbi;`2aZG /`Bp2`

ê Š § , … Ž • –) • Ú _ ± ¢ Š ¼ v • –) ± ¯ Š v ~ c ê ¡ a ª † v ¦ e ‘ Á g • • –) k * P Ð † Â ‡ t « ª ê • Ú _ ±

```
/2i BHb@+`2 i2*QMM2+iBQUM#Kb 4TQbi;`2b-[H]
b2`p2`]4HQ+ H?QbifTQbi;`2b]
mb2` ]4DQ2]
T bbrQ`/]4 42+`2-i]
b+?2K ]4 4/KV]
+QMM@+QMMU2/2ii BHbV
```

OO *QMM2+iBM; mbBM; SQbi;`2aZG /`Bp2`

N X k X k

Ù ª Ñ ë ª Ë Š Á º ü ± ü x ª ~ •) ‰ * P Ž • [m2`va4[2Ht2+mi2+ax[H [m2`važ[ÙHª Ñ k Ù

c n ƒ j • Á ± £ q • x •

```
[m2`vaU[+HQMM]M1G1*h hPS j 6_PJVT2`bQM]
```

O O T 2 ` b Q M n B / ; 2 M / 2 ` n + Q M + 2 T i n B / v 2 ` n Q 7 n # B ` i ?
O O R R 3 8 y d R N d 8
O O k k 3 8 y d R N d e
O O j j 3 8 y d R N d d

```
2 t 2 + m i 2 U + Q M M - l L * h 1 h " G 1 7 Q Q c . _ P S V " G 1 7 Q Q c ]
```

ˆ € ¨ • * P ¡ Õ y ‰ º û ª þ ï ª Ú ± % ¯ v • x • Ç ª â ª Š § , … º û ª k Y r v ~ ê • º û ª ó Õ Í ª Ê ‡ `

N X k X z j / 7

Ù ª Ñ ë ª Ë j ¨ * ® x ª Ù ª Ñ k ó ô ü Š ? • © l ¨ ‰ c š ˆ] l c ê ¡ a © • x ‘ Í Á É ù
† r ˜ ~ § e Š • | • § e ‰ ê Š Ž 7 Ž Ã Õ Ã ª Ê ± ¬ Q v … 5 Ù ª Ñ ¼ è Ê ¹ Á Ú ± ç ³ ¶ ý Š C

```
t l @ [ m 2 ` v a [ H X U + Q M M - 1 G 1 * h 6 _ P J T V 2 ` b Q M ]
```

[Ž ‡ e l l G l ¼ è Ê ¹ Á Ú † x ‘

N X k X 9 a Z G

6 T O 5 H Q G H U ã Õ Ã 2 Ê M / 2 ì ` M b H * i 2 ± M ‰ Š • ’ p x > x ‰ * P k a © • x • ` 2

```
t l @ ` 2 M / 2 ` h ` M b H i 2 Z U + Q M M -
                                b [ H 4 a 1 G 1 * h h P S R y 6 _ P J ! b + ? - 2 K X T 2 ` b Q M
                                b + ? 2 K ] 4 / K n b v M V m 7 ]
```

6 4 / 6 H U Y H U ‘ - • Ä 7 2 3 Å å Ž • 3 R V W J U H 6 4 / † Ž Ä / , 0 , 7 Å ‰ ˆ Š 7 t « •

N X k X 8

Ù ª Ñ ± Ù ª Ñ ë ª Ë Š ö ì x ª Š Ž 2 t 2 + m i 2 * a [± ¬ Q v … 6 4 / Ë Ø ª Ú ó Ú ± x ª r ‡ ¡

```
/ i U K i + ` b V
B M b 2 ` i h U + Q M M - i + ` b ] K i + ` b + ` 2 i 2 h # H 2 _ 4 V
```

r • Á † Ž • P W F D U V Ù ª Ñ ç þ ª ò ± Ç ª â ª ã • Ä P W F D U V Å ‡ c e Ø ª è ý Š ´ Õ é ÿ ª Û v

&'0—•Áºü

NXj*.J

¤R•Á†Ž•2+'6,64/±¬Qv…&'0Š€¢v~ÙªÑëªËŠÁºü± üv•x‘r«¨•Áºü†Ž•&'0•

•y•ÙªÑëªËŠU)•)kcªj±Áºü†*®v…ž•v¦e•

```
a1G1*IPlLUhV  aT2`bQMn+6_PMi+/KXT2`bQMc
```

3(5621B&2817

aªcŽ•#…N(•%ß:‰ë‰tŠËÅkaª•j¡v«•z²•

```
a1G1*b:U. h1.A66U
                Q#b2`p iBQMnT2`BQ/nbi `in/ i2-
                Q#b2`p iBQMnT2`BQ/n2jM8nX/kB2VMmKnv2 `b
6_PJ+/KXQ#b2`p iBQMnT2`BQ/c
```

180B<($56

ØªèÿW v…ûY••C±rhxªr‡k†l•x‘W Žþí•ØªèýÐ•± •çµªýÛŠŸw¡kaªr‡±W‘

```
a1G1*JhsU1 UQ#b2`p iBQMnT2`BQ/@2M/n/ i2V
        v2 `nQ7n#B`a?KVtn ;2
6_PJ+/KXT2`bQM
ALL1C_PA!L+/KXQ#b2`p iBQMnT2`BQ/
  PLT2`bQMXT2`bQM#nB2/p iBQMnT2`BQ/XT2`bQMnB/c
```

0$;B$*(

#…Ÿ³â•ûÎ ã±T xªŠŽ•ŽªjŠ Œ‰Áºük´W†x‘r•Áºü†Ž••y3(5621±2%6(59$7
XXX ±aQ¹•kDJHVp‡•’•x‘r«Š§©•DJHV±M.•Øªèý†aªj•§eŠ–Æxªr‡k†l•x‘l
l X8y •M¬ûÎ±¬Qv•x‘M¬ûÎ‡M]ûÎŽ2§ŠCŸt«•x•

```
q A h >; 2 b
 a U
    a 1 G 1 * h 2 -
        _ P q n L l J U V P o 1 U
            P _ . 1 " u  ; 2
            V  Q ` / 2 ` n M `
    6 _ P U
        a 1 G 1 * u 1  U Q # b 2 ` p  i B Q M n T 2 ` B Q / n b v 2 i n / n Q 2 V n # B à i ? 2
        6 _ P J + / K X T 2 ` b Q M
        A L L 1 C P A U + / K X Q # b 2 ` p  i B Q M n T 2 ` B Q /
            P L T 2 ` b Q M X T 2 ` b Q M # b 2 / p  i B Q M n T 2 ` B Q / X T 2 ` b Q M n B /
        V  ; 2 n + Q K T m i 2 /
    V
a 1 G 1 * J h A U  ; 2 V a K B M n  ; 2 -
    J A U *  a 1
            q > 1 U Q ` / 2 ` n M X k 8  M
                h > 1 N N N N
            1 G a 1 ; 2
            1 L V  a [ k 8 n  ; 2 -
    J A U *  a 1
            q > 1 U Q ` / 2 ` n M X 8 y  M
                h > 1 N N N N
            1 G a 1 ; 2
            1 L V  a K 2 / B  M n  ; 2 -
    J A U *  a 1
            q > 1 U Q ` / 2 ` n M X d 8  M
                h > 1 N N N N
            1 G a 1 ; 2
            1 L V  a [ d 8 n  ; 2 -
    J  s U  ; 2 V a K  t n  ; 2
6 _ P J ; 2 b
* _ P a 6 P A U
    a 1 G 1 * h P l L U V  a M
    6 _ P J ; 2 b
    V  T Q T m H  i B Q M n b B x 2 c
```

0 , 1 B $ * (4 B $ * (0 (' , $ 1 B $ * (4 B $ * (0 $; B $ * (

§ © Œ ‰ C Ÿ Ž • 6 4 / • [® © Š 5 ± ¬ Q v … ü e r ‡ ¡ † l • x ‘ Á g • • Ÿ w W b ± ® ª ~ • ã • 5 Å

Áºü üâŠî¿õèûü±¬Qxª

```
b[HI @]a1G1*h u1 _UQ#b2`p iBQMnT2`BQ/nbi `in/ i2V @
            v2 `nQ7n#B`i?  a  ;2
6_PJ !+/KXT2`bQM
ALL1_ CPAL !+/KXQ#b2`p iBQMnT2`BQ/
  PL T2`bQMXT2`bQMnB/ 4 Q#b2`p iBQMnT2`BQ/XT2`bQMnB/c]
 ;2 I @`2M/2`h` MbH i2Zu+2QwaM7H b[H-4+/KV
[m MiBUH;22(R) - +Uy- yXk-8yX-8yXd-8RVV
```

OO yW k8W 8yW d8W RyyW
OO y e Rd j9 Ny

rr†Ž•Çªâªã†ûÎ±CŸv•x˜…•ûÎ±Ò¸ ÿªÛv•ûÎ ã±CŸv•x‘vjv•r«ŠŽÙªÑëªËÇª

Áºü†Ž•&'0Ð•ÏªË¡±¬Qxªr‡k†l•x‘Ág••M¡Ùi• cÅ ÙµÉù •ÏªËÅªÛ•ÚÕé ±*

```
a1G1*hP SRy+QM/BiBQMnbQm`+2np Hm2-
  *PlLUhV a+Q/2n+QmMi
6_PJ+/KX+QM/BiBQMnQ++m``2M+2
:_PlSu+QM/BiBQMnbQm`+2np Hm2
P_.1"u @PlLUhVc
```

&21',7,21B6285&(B9$/8(&2'(B&2817

,

rr†Ž•&21',7,21B2&&855(1&(ØªèýÐ•&21',7,21B6285&(B9$/8(çµªýÛ•¡

NX9

5n•øt†Ž•î¿õèûük-Q†x‘î¿õèûüØªèýŽ&'0•Æü†a©•64/Áºü±¬Qv…xQ†l•x

```
a 1 G 1 * lP l L U hV  a b m # D 2 + i n + Q m M i -
  + Q M + 2 T i n M  K 2
6 _ P L U + / K X T 2 ` b Q M
A L L 1 C P A L + / K X + Q M + 2 T i
  P L T 2 ` b Q M X ; 2 M / 2 ` n + Q M + 2 T i n B / 4 M + Q M + 2 T i n B / i X + Q M + 2 T i n B /
: _ P l S u + Q M + 2 T i n M  K 2 c
```

6 8 % - (& 7 B & 2 8 1 7 & 2 1 & (3 7 B 1 $ 0 (

) (0 $ / (

0 $ / (

î ¿ õ è û ü • • í Š ½ - ‰ Q • Æ ƒ Ž • | • ð å † x ‘ § n a ª Á º ü Ž • ± • Å Í é Ú ‡
| • x ˜ … • ¶ / ± ‰ x ¡ • † x ‘ Á g • • ¶ é é ÿ ç ¹ ‡ c e h ± 0 Ÿ ˆ c [P ± P g ª ‡ v • x •

```
a 1 G 1 * lP l L U hV  a T ` 2 b + ` B T i B Q M n + Q m M i
6 _ P L U + / K X / ` m ; n 2 t T Q b m ` 2
A L L 1 C P A L + / K X + Q M + 2 T i n  M + 2 b i Q `
  P L / ` m ; n + Q M + 2 T i n B / 4 M + Q M + 2 T i n B / i n B / + 2 M /  M i n + Q M + 2 T i n B /
A L L 1 C P A L + / K X + Q M + 2 T i  B M ; ` 2 / B 2 M i
  P L  M + 2 b i Q ` n + Q M + 2 T i n B / 4 M + Q M + 2 T i n B / i n B / + 2 M ; ` 2 / B 2 M i X + Q M + 2 T i n B /
q > 1 _ G P q U B M ; ` 2 / B 2 M i X + Q M + 2 T i n B / 4 T i n B / M + K 2 V 7 2 M ^
   L . B M ; ` 2 / B 2 M i X + Q M + 2 T i n + A M b b 2 n / B / 2 M i ^
   L . B M ; ` 2 / B 2 M i X b i  M / ` / n + Q M + 2 T i 4 ^ a M
```

3 5 (6 & 5 , 3 7 , 2 1 B & 2 8 1 7

N X 8 Z m 2 ` v G B # ` ` v

4 X H U \ / L E U D U \ Ž • & ' 0 Q • Æ v Š ¬ Q t « ª 6 4 / Á º ü • û ¶ è û ü † x ‘ r « Ž ¼ û ¶ ´ é

r • û ¶ è û ü • þ : Ž • v c ø ª È ª k & ' 0 • Á º ü c e ± Ì L x ª • ± ½ x ª r ‡ † x ‘ û ¶ è û ü Ð • Á º ü

4 X H U \ / L E U D U \ Ž • 6 T O 5 H Q G H U ± x Q v … • Â h v ~ 6 4 / c Š † Á º ü ± ü v • x ‘ ø

K W W S G D W D R K G V L R U J 4 X H U \ / L E U D U \

K W W S V J L W K X E F R P 2 + ' 6 , 4 X H U \ / L E U D U \

4 X H U \ / L E U D U \

) L J X U H Á º ü û ¶ è û ü • & ' 0 Š F x ª 6 4 / Á º ü • û ¶ è û ü ‘

N X e

N X e X R

X g ê 2 Ž • $ & (ã ¦ ¸ $ & (L ¹ • § n ¢ ¨ « ~ t Q † x ‘ 6 O D W H U
H W D O Š § ª ‡ • $ & (L å £ Ÿ ³ ¯ U (• X g ê 2 • Y Ç † Ž) º Á U ‡ @ t « … c •

ü É à é ü ý z ? Ÿ ³ ¯ • M ‰ • U († • X g ê 2 • Y r † Ž • û Î ‡ g 2 † ð 2 S x ª ‡ ˆ • * i † v ¦ e

N X e X k

D æ ± ü É à é ü ý — • ‰ ‰ • D æ ‡ v … q v • x ‘ ‰ ‰ ‡ Ž • ¤ È Š ü É à é ü ý — • D æ k ‰ c

N X e X j

X g ê 2 Ž • ì Ù º • ~ Ž • Ž ¤ h ¸ (5 ¹ 7 # º Š X g ê 2 • # › Å ª Û k] û t « ~ ê ‡ q v • x ‘

N X e X 9

å £ Ÿ ³ ¯ • M ‰ • U (• Y Ç † ± C Ÿ v • x ‘ k U (Š ® ~ , … =) : Š D æ t « ~ j ˆ e j Ž c • z ² ‘

N X da Z G _

2 + ' 6 , Ö ª ý • Á Š G ¨ « ª r ‡ Ž a © • z ² k • Ÿ w € Š ^ e r ‡ Ž - † x ‘ r • ê • 2 + ' 6 , Ö ª ý k ž

```
HB#` Ùv i # b2*QMM2+iQ`V
+QMM@+QMM2#Kb ]4TQbi;`2b-[H]
                b2`p2`]4HQ+ H?QbifTQ-bi;`2b]
                mb2`]4DQ2]
                T bbrQ`/]4 2+`2V]
+/K.#a+?2@K]+/K]
+Q?Q`i.#a+?2@]K]o+` i+?]
+Q?Q`ih #@]2Kvn+Q?Q`ib]

b[HI@]
*_1 h1 h "G1 !+Q?Q`in/#nb+?2K X!+Q?Q`ini #H2 U
  +Q?Q`in/27BMBiBQMnB/ ALh-
  +Q?Q`inbi `in/ i2 . h1-
  +Q?Q`in2M/n/ i2 . h1-
  bm#D2+inB/ "A:ALh
```

6 4 / ‡ 5 ± ¬ Q v ~ p – • À

```
Vc
]
`2M/2`h` MbH i21t2U+Q?Q`i.#a+?2K -
                    +Q?Q`in/#nb+?2K X+Q?Q`ini #H2 U
                    +Q?Q`ini #H2 U+Q?Q`ih #H2V
```

r r † Ž • Ù ª Ñ ë ª Ë Ë ¿ ª ð ‡ Ø ª è ý à ± ã û ó ª Ñ S v … c • x ‘ µ ‰ ª » Š … Š < 6 t z ª r ‡ k † l • x ‘ | • W

N X d X R

ã Š • D æ Å í ª Ú ± t h v • & 2 + 2 5 7 Ø ª è ý Š ö ì v • x •

```
b[Hl @]
ALa1_h ALhP !+Q?Q`in/#nb+?2K X!+Q?Q`ini #H2 U
  +Q?Q`in/27BMBiBQMnB/-
  +Q?Q`inbi `in/ i2-
  +Q?Q`in2M/n/ i2-
  bm#D2+inB/
V
a1G1*h R  a +Q?Q`in/27BMBiBQMnB/-
  +Q?Q`inbi `in/ i2-
  +Q?Q`in2M/n/ i2-
  bm#D2+inB/
6_PJ U
  a1G1*h /`m;n2` nbi `in/ i2  a +Q?Q`inbi `in/ i2-
    /`m;n2` n2M/n/ i2  a +Q?Q`in2M/n/ i2-
    T2`bQMnB/  a bm#D2+inB/
  6_PJ U
    a1G1*h /`m;n2` nbi `in/ i2-
      /`m;n2` n2M/n/ i2-
      T2`bQMnB/-
      _PqnLlJ"1_UV Po1_ U
        S _hAhAPL "u T2`bQMnB/
          P_.1_ "u /`m;n2` nbi `in/ i2
      V Q`/2`nM`
    6_PJ !+/Kn/#nb+?2K X/`m;n2`
    q>1_1 /`m;n+QM+2TinB/ 4 Rjy3kRe @@
  V Q`/2`2/n2tTQbm`2b
  q>1_1 Q`/2`nM` 4 R
V 7B`bin2`
ALL1_ CPAL !+/Kn/#nb+?2K XQ#b2`p iBQMnT2`BQ/
  PL bm#D2+inB/ 4 T2`bQMnB/
```

```
    L. Q#b2`p iBQMnT2`BQ/nbi `in/ i2 I +Q?Q`inbi `in/ i2
    L. Q#b2`p iBQMnT2`BQ/n2M/n/ i2 = +Q?Q`inbi `in/ i2
q>1_1 . h1.A66U. u-
            Q#b2`p iBQMnT2`BQ/nbi `in/ i2-
            +Q?Q`inbi `in/ i2V =4 je8c
]

`2M/2`h` MbH i21t2U+Q?Q`inbi `in/ i2 I +Q?Q`in/#nb+?2K X!+Q?Q`ini #H2 U
                      +Q?Q`in/#nb+?2K X!+Q?Q`ini #H2 U
                      +Q?Q`ini #H2 U +Q?Q`ih #H2-
                      +/Kn/#nb+?2K X+/Kn/#a+?2K V
```

r r † Ž • & ' 0 • À € Ø ª è ý † a ª ' 5 8 * B (5 $ Ø ª è ý ± ¬ Q v • x ‘ r • Ø ª è ý Ž ' 5 8 * B (; 3 2 6 8
Ø ª è ý Š W v •) † ~ © • M ‰ • D æ ± * © p v • x ‘) • k P • # … N (± á ƒ Z g k a ª ~
‡ & 2 + 2 5 7 B 6 7 $ 5 7 B ' $ 7 (• (Š Ž • - ‰ n ‡ ¡ ê • (Ê k ´ W ‡ ‰ © • x ‘

N X d X k

M ¯ Š • ´ ¸ Ú ½ ò Å í ª Ú ± t h x ª ´ W k a © • x •

```
b[HI @]
ALa1_h ALhP !+Q?Q`in/#nb+?2K X!+Q?Q`ini #H2 U
 +Q?Q`in/27BMBiBQMnB/-
 +Q?Q`inbi `in/ i2-
 +Q?Q`in2M/n/ i2-
bm#D2+inB/
V
a1G1*h k a +Q?Q`in/27BMBiBQMnB/-
  +Q?Q`inbi `in/ i2-
  +Q?Q`in2M/n/ i2-
  bm#D2+inB/
6_PJ U
  a1G1*h .AahAL*h T2`bQMnB/ a bm#D2+inB/-
    +QM/BiBQMnbi `in/ i2 a +Q?Q`inbi `in/ i2-
    +QM/BiBQMn2M/n/ i2 a +Q?Q`in2M/n/ i2
  6_PJ !+/Kn/#nb+?2K X+QM/BiBQMnQ++m``2M+2
  ALL1_ CPAL !+/Kn/#nb+?2K X+QM+2Tin M+2biQ`
    PL +QM/BiBQMn+QM+2TinB/ 4 /2b+2M/ Min+QM+2TinB/
  q>1_1 M+2biQ`n+QM+2TinB/ 4 9jkdNR @@
V /BbiBM+inQ++m``2M+2
ALL1_ CPAL !+/Kn/#nb+?2K XpBbBinQ++m``2M+2
  PL bm#D2+inB/ 4 T2`bQMnB/
```

6 4 / ‡ 5 ± ¬ Q v ~ p – • À

```
   L. pBbBinbi `in/ i2 I4 +Q?Q`inbi `in/ i2
   L. pBbBin2M/n/ i2 =4 +Q?Q`inbi `in/ i2
q>1_1 pBbBin+QM+2TinB/ AL Ukek- Nkyj-
    NkyRV @@      1_c
]

`2M/2`h` MbH i21t2 +Q?Q`i.#a+?2K -
                      +Q?Q`in/#nb+?2K 4
                      +Q?Q`ini #H2 +Q?Q`ih #H2-
                      +/Kn/#nb+?2K 4#a+?2K V
```

r r † Ž • & 2 1 ' , 7 , 2 1 B 2 & & 8 5 5 (1 & (Ø ª è ý ± & 2 1 & (3 7 B $ 1 & (6 7 2 5 Ø ª è ý ‡ W v … • X

N X d X j

Å í ª Ú k ™ t « ~ • † • û Î ‡ g 2 † ð 2 S t « ~ Y Ç † ± C Ÿ † l • x •

```
b[Hl @]
qAh> i `  a U
  a1G1*h +QM+2TinM K2  a ;2M/2`-
    6GPP_UUu1 _U+Q?Q`inbi `in/ i2V @
         v2 `nQ7n#B`i?V f RyV  a  ;2-
    bm#D2+inB/-
    +Q?Q`inbi `in/ i2-
    * a1 q>1L . h1 ..U. u- d- +Q?Q`inbi `in/ i2V =
      Q#b2`p iBQMnT2`BQ/n2M/n/ i2
    h>1L Q#b2`p iBQMnT2`BQ/n2M/n/ i2
    1Ga1 . h1 ..U. u- d- +Q?Q`inbi `in/ i2V
    1L.  a +Q?Q`in2M/n/ i2
  6_PJ !+Q?Q`in/#nb+?2K X!+Q?Q`ini #H2
  ALL1_ CPAL !+/Kn/#nb+?2K XQ#b2`p iBQMnT2`BQ/
    PL bm#D2+inB/ 4 Q#b2`p iBQMnT2`BQ/XT2`bQMnB/
      L. Q#b2`p iBQMnT2`BQ/nbi `in/ i2 I +Q?Q`inbi `in/ i2
      L. Q#b2`p iBQMnT2`BQ/n2M/n/ i2 = +Q?Q`inbi `in/ i2
  ALL1_ CPAL !+/Kn/#nb+?2K XT2`bQM
    PL bm#D2+inB/ 4 T2`bQMXT2`bQMnB/
  ALL1_ CPAL !+/Kn/#nb+?2K X+QM+2Ti
    PL ;2M/2`n+QM+2TinB/ 4 +QM+2TinB/
  q>1_1 +Q?Q`in/27BMBiBQMnB/ 4 R @@
V
a1G1*h / vbX;2M/2`-
    / vbX ;2-
```

```
   / vb-
   * a1 q>1L 2p2Mib Aa LlGG h>1L y 1Ga1 2p2Mib 1L.  a 2p2Mib
6_PJ U
  a1G1*h ;2M/2`-
    ;2-
   alJU. h1.A66U. u- +Q?Q`inbi `in/ i2-
     +Q?Q`in2M/n/ i2VV  a / vb
  6_PJ i `
  :_PlS "u ;2M/2`-
    ;2
V / vb
G16h CPAL U
  a1G1*h ;2M/2`-
      ;2-
     *PlLhU V  a 2p2Mib
  6_PJ i `
  ALL1_ CPAL !+Q?Q`in/#nb+?2K X!+Q?Q`ini #H2  M;BQ2/2K
    PL i `Xbm#D2+inB/ 4  M;BQ2/2K Xbm#D2+inB/
      L. i `X+Q?Q`inbi `in/ i2 I4  M;BQ2/2K X+Q?Q`inbi `in/ i2
      L. i `X+Q?Q`in2M/n/ i2 =4  M;BQ2/2K X+Q?Q`inbi `in/ i2
  q>1_1 +Q?Q`in/27BMBiBQMnB/ 4 k @@
  :_PlS "u ;2M/2`-
    ;2
V 2p2Mib
PL / vbX;2M/2` 4 2p2MibX;2M/2`
   L. / vbX ;2 4 2p2MibX ;2c
]

`2bmHi@`2M/2`h` MbH i2Zu+QwaqH- b[H-
                                +Q?Q`in/#nb+?2K  +Q?Q`i.#a+?2K -
                                +Q?Q`ini #H2  +Q?Q`ih #H2-
                                +/Kn/#nb+?2K  +/K.#a+?2K -
                                bM F2* b2hQ* K2H*lV2 4
```

•y•&7(ÄWDUÅ±thv•<”‰üËÁâ(±fex˜…•Dæ±0 •x‘2%6(59$7,21B

VQDNH&DVH7R&DPHO&DVH 758(±Qcª•Ž•64/†ŽçµªýÛàŠVQD
±¬QxªÈkaª¸64/Ž] Þ‡¬ Þ±ö2v‰c~ ¹•ŠFv•5†ŽFDPHO&DVH±¬

JJSORW ãÕÃªÊ±¬Qxª‡•Wb± …ŠéÿÕÚ†l•x•

```
O     A_
`2bmHBbl@Ryyy `2bmHi2bp2Mib`2bmHi/bvbf d
```

6 4 / ‡ 5 ± ¬ Q v ~ p – • À

```
O
`2bmHi;b2 I @`2bmHi;b2   Ry

HB#` Uy;THQikV
;;THQi`2bmHi;b2 bU 4  ;2-v 4B`-;`QmT;2M/2`+QHQ`;2M/2`VY
  ;2QKnHUBVM2
  tH U]  V Y
  vH U]    R-yyyV f  ]
```

N X d X 9

t h v ~ Ø ª è ý ± Á ü ª ´ Õ é v • ~ « y Š –) ± * w • x •

```
b[HI @]
h_lL* h1 h "G1 !+Q?Q`in/#nb+?2K X!+Q?Q`ini #H2c
._PS h "G1 !+Q?Q`in/#nb+?2K X!+Q?Q`ini #H2c
]
`2M/2`h` MbH i21t2U+mQ?Ma[+ b[H-
                        +Q?Q`in/#nb+?2K +Q?Q`i.#a+?2K -
                        +Q?Q`ini #H2 +Q?Q`ih #H2V

/Bb+QMUM+QMMV
```

NXdX8

2+'6,64/‡'DWDEDVH&RQQHFWRU‡6TO5HQGHU±àž ®z…¬Qx'

Ùô ËÚþªÉù QŠ•,tÖ†64/±¬Qv…ÅíªÚ±thxªr‡Šv•v~k•$7/$6†ÅíªÚ
64/±rhxª~ •6TO5HQGHU‡'DWDEDVH&RQQHFWRU‡ …Š!Qxªr‡

NX3

m64/¸6WUXFWXUHG4XHU\/DQJXDJH¹Ž•´þÙªÑôÙý¸&'0¹

mµ‰ªÙªÑëªËéûÕÚç»ªòŽµ‰ª64/Å‡±á,…i©•Æ‡xª~ ŠŽµ‰ª

m6TO5HQGHU‡'DWDEDVH&RQQHFWRU5ãÕÃªÊŽ•&'0Ð•Ú

m5‡64/±!Qxªr‡†•2+'6,Öªý†ŽÇïªÚt«…c‰c½ËÑò ‰± †l•x

m4XHU\/LEUDU\Ž•&'0Q•LxQZ ‰64/Áºü•ÅþÁÉù ±%¯v•x'

NXN

È%ð[

r«¨•!L†Ž•ÍÁÉù Š]bt«…cª§eŠ•5•5
6WXGLR•-DYDk¶ ËÚªýt«…cªr‡±È%‡v•x'•~• 6TO5HQGHU‡'DWDED
(XQRPãÕÃªÊ¡´W†x'¤R•,‡†¶ ËÚªý†l•x'

BMbi HHXT UFU]a2H_2M/2]`.]i # b2*QMM2+ïQ2KQi2VV
`2KQi,2BMbi HHn;Bi??/#Bf1mMQ`2B 4pRXyXy]

(XQRPLDãÕÃªÊŽ•&'0Ð†ÿª½ý5ÍÕÉù Ð†žtxªÉñ÷þªÚt«~ÙªÑÍÕÚ

+QMM2+iBQMI@1BMbQKB 2i1mMQKB *QMM2+iBVQM.2i BHb

&'0ÙªÑëªË•Ë¿ªðŽÄPDLQÅ†x'

!L 64/‡5±¬Qv…•ÙªÑëªËÐŠU)cªj±CŸv•x'

!L 64/‡5±¬Qv…•ÍþÅ¿Éè•ˆc±-‰n‡¡#‰7p~r‡kaª)•)P±CŸv•x

! L

! L 6 4 / ‡ 5 ± ¬ Q v … • Í þ Å ¿ É è • Q º Š Á S p X ‡ # › t « ~) •) P ± C Ÿ v • x ‘ ä Ú Á S p X • Å Í é Ú , ' Ž † x

@ § t « ª ˆ Š Ž Ý û (± – Æ n • t c ‘

6 4 / ‡ 5

*? Ti2` Ry

Ì .ULVWLQ .RVWND

… + ^ × Ùª Ñ¸ü´ý ªýÛÙª Ñ‡¡••«ª¹Ž• •^×óL£¿£•%¯Š*Þxª Ùª Ñ†•t•u•‰îY …

Ág••¿£C{{‘Ùª Ñëª ËŽ•aªÇó¸Á•X gê2¹ŠFv…%¯t«~x˜…•¿£±G v•*Þ›Qk<”

Í†Ž•Åíª Ú q•th‡´-‡ŽUj•Åíª Ú±ŸYxªce•$7/$6•~Ž64/±¬Qv…¶í•Åíª Ú±å¯›

RyXR

2+'6,p–†Ž•Åíª Ú±•aªÆ N(Š ƒ¤ã•<¾B€±Â~x)§•X D‡ qv…c•x‘r•Q¶ŽÅ‡

```
Åíª ÚŽ•aªÆ N(Š ƒ¤ã•<¾B€±Â~x)§•X D†x‘
```

2+'6,†¬Qt«…cªÅíª Ú• qŽ•r• •4•)§k¬Qxª¡•‡Žµ‰ªj¡v«‰crˆ±óóxªrˆkeW
,&' •1'&•+&3&6‰ˆ¹Š¾Ûv…cª‡t«…c•x‘ÅªÛÍÕÚŽÅíª Ú±àž‡…ªc•eW‰
,&' ÅªÛ•‰‰•Yrj—|«‡¡Yrx˜…j—¹‘ âÂŠ qt«~Åíª Ú†Ž• kÅíª ÚŠàžì«¨«

2+'6,•Åíª Ú q±xQxª~ •øÝªÁ‰Ý÷´ ËŠŽ¤Rka©•x‘

- oÆ)•)k P•Åíª ÚŠ&xªZ gka©•x
- oÆ)• k P•µ‰ªN(ŠŸwÅíª ÚŠ&xªZ gka©•x‘
- oÆ)• kŸwN(ÐŠŸwÅíª ÚŠ P‰&xªr‡Ža©•z²‘
- oÅíª ÚŠó âªkÎÿ•~Ž P0•«ªê ka©•x‘

ÅíªÚ• q

ÅíªÚ±å¯xª~ •)‰´éÿªÓŽáƒa©•x•

ýªýëªË•ÅíªÚ qŽ• kÅíªÚŠcªâN±âê:‰ýªý†œâv•x‘r«¨•ýªý• q

Â†ëªË•ÅíªÚ qŽ• kÅíªÚŠ&xª •Â†¸j¨ •(•Â†¹±CŸxªÂ†ôÙ

ã•ÍÁÉù †Ž•r«¨•´éÿªÓŠƒc…ÚvnœâV•x‘

RyXk

ýªýëªË•ÅíªÚ qŽ•± •N(¸Á•Äwž Z¤ÐŠ|•óL±YÇv~)ÅŠÆƒ•~Ž• P•â

r«¨•B€±åhxªcŠ¬QxªÀ€:‰åhWßŽ¤R•þ©†x•

oÛó¶ •ÙªÑk¾ t«…cª&'0Ûó¶ ¸Á•Äˆª¸éÿÉªÊõª¹•YrÅ•Ä dDæÅ¹

oÅ ÍéÚÍÕÚ•FÛ‡xªµ° L±W0xªÆƒ¤ã•À€Å ÍéÚ± qxªÙªÑŠ§.v‰

oÛó¶ ‘-•&g•* •aªµ° LŠ*ÞxªûY•&g¸Á•'58*B(;32685(•'$<6B68

oâ(:‰ÿÊÕÁ•<¾B€‡¶ë Ú•*$kÆWt«ªâ((Ê¸Á•¼ t«~óLŽDæŸ³È•

ÅíªÚ q±å¯xªc•ÅíªÚ&g±ÅxåýÙµ ÂèÿÕÁ•§eŠÛó¶ ±õgª‡>x†x¸:

)LJXUH ÅíªÚ q•åýÙµ ÂèÿÕÁ

ÅíªÚ qthâŠí x˜lcnƒj• ka©•x•

Å íéÚÍÕÚ

- oÅíªÚàì«•â(± qxª‰N¶ë ÚŽUj—
- o‰N¶ë ÚŠ<Qt«ª<¾B€ŽUj—
- oÅíªÚƒz± qxª¡•ŽUj—

ÅíªÚàì«¶ë Ú•ÅíªÚàì«¶ë Ú¸‰N¶ë Ú¹Ž•)§kÅíªÚŠ–Yxªâ P•ƒ•©Åíª Ú¶ ÙÕ

<¾B€•<¾B€Ž•‰N¶ë ÚÅíªÚŠ<Qt«•t¨Š)§•ÍÕÚ±cŒv•x‘º<¾B€Ž•ÙªÑkC.t

ÅíªÚƒzB€•ÅíªÚƒz¶ë ÚŽ•aª)kÅíªÚó âª‡v…•Õ¾±üer‡±êv•x‘ÅíªÚƒzŽ•#

```
2+'6,Öªý†Ž•<¾B€‡œ¤B€•ö2Ža©•z²‘x˜…•B€Ž<¾B€‡v…-òSt«•x‘Ág
```

RyXj

Å íéÚÍÕÚŽ•t•u•‰ ‰†LxQZ ‰Å ïªß Ú‡v…¬Q†lªÅ íéÚ•üËÚ±ÅxÅ‡†x‘r«

- oœ¤•r•Å íéÚ¸i§’Âht«…cªê Ž|•R¦ðŠ0•«ª¡•¹±Å íéÚÍÕÚj¨œ¤v•x‘
- oR¦ðŠ0•«ª•r•Å íéÚ•p†‰n•|•R¦ðŠ0•«ª¡•¡õ–v•x‘
- oðÕæ ÂWž•À€St«…c‰cÅ íéÚ•k}±¦Zv•x‘

Ág••Å íéÚÍÕÚ•Å‡Ž•:Šêt«ª§eŠ ƒ•Å íéÚ±0Ÿr‡k†l•x¸Å
¹‘rr†Ž•Å íéÚ ¸Ä è/„OÅ¹‡|•x˜…•R¦ðŠ0•«ªÅ íéÚ±0ž•xk•Å

7DEOH Å íéÚÍÕÚ•Å‡•Á

Å íéÚ,'	Å íéÚà œ¤ ¶/ ðÕæ ÂFÛ			
	è/„O	ccg	Žc	ccg
	öœg è/„O	Žc	Žc	ccg

:Šêx§eŠ¸: ¹•r«ŽÄ è/„OÅi§’|•R¦ðŠ0•«ª¡•±0Ÿk•Äöœg è/„OÅi§’|•R¦
i§’,&' ÅªÛ¹±iÿv•x‘

RyX9

ýªýëªË•ÅíªÚ qŽ•ÅíªÚ q±àž‡…ª~ •Æv:‰ce†x‘vjv•p–ÅíªÚ±thxª~ Š´W%

Åíª Ú• q

)LJXUH Ä è/„OÅ¸R¦ð±0Ÿ¹±0Ÿk•Äöœg è/„OÅ¸R¦ð±0Ÿ¹±œ¤xªÅ Íé

r•´éÿªÓ±&'0ÙªÑŠ<Qv~Á‡v…•$3+52',7(¸$XWRPDWHG
3+HQRW\SH5RXWLQH IRU2EVHUYDWLRQDO'HILQLWLRQ ,C
DQG(YDOXDWLRQ¹5kãÕÃªÊr•ãÕÃªÊŽ•Ü ÌŠûëýÝpt«~ÙªÑj¨ÌLxª
%DQGD HW DO

RyX8

Åíª Ú±å¯xªc•ã•ˆ€¨keWj±õ–x˜l†x•<¾ ±x˜…uƒpªr‡keWj—
|«‡¡ Â ±á…ª •ž±àžì«ªr‡keWj—

Åíª Ú•å¯©ŠŽ•¦ ^• ¯k ±ˆ•§eŠ qxªj‡ceµ°:‰•¾gŠ§.v•x‘ƒ•©•<”‰

Í•…š†r˜~§eŠ•Åíª Ú qŽ]ût«~ÙªÑj¨#…v~cr‡±@"v§e‡xªÒž†x‘r•Ò
ÃªË•,žÓõª Úþå÷ª¹‡•Çxª ØËÚ‡õgªr‡k†l•x‘Ú_Ž^
¸Äµ°:;†gÅ¹†Úvnœâv…c•x‘

RyX8XPR>.aA

M.•Åíª Ú q‡´ýÆüÌò•¶ ë Úüª‡ÌD:‰ÆW±½ xª~ Š•2+'6,
Æªý ÛËÑ Òª Ûç¹àÑ¶éû¶èûü¸*63/¹ ªÁÂýªék™‡t«•v~‘*63/ ªÁÂýª
r• ªÁÂýªé•p–ŠŽ•È•ÍÁÉù †tüt«~$3+52',7(
%DQGD HW DO £3KH9DOXDWRU6ZÖHªÚGHO HW DO
•4•2+'6,ßÕÚ ªÁÌD†•U¶½ýØ‡ÄàñÁË• H0(5*(

KWWSV JLWKXE FRP 2+'6, $SKURGLWH
KWWSV ZZZ RKGVL RUJ ZHE ZLNL GRNX SKS"LG SURMHFWV ZRUN
ZJ

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

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

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_O
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

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コン

Figure 10.5: 初期イベントの追加

Figure 10.6: 薬剤曝露の定義

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ります 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ことが 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 モジュールを

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

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阻害薬成分を選択し 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するこ 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

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リス inhibitors」と入力し、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に絞 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ット Drug」ボックス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がわか

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。問いでは、新規ユーザーまたはAC attribute」をクリックします。次に「Add first exposure criteria」を選択します。作成する基準 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

Concept Set Expression | Included Concepts 21536 | Included Source Codes | Export | Import

Name:

ACE Inhibitors

Show 25 entries Search:

Showing 1 to 15 of 15 entries Previous 1 Next

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

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

現在のATLASのデザインは一部の人を混乱させるかもしれません。見た目の通り、✖は

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るた

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組み

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。塗り

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。これら 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、これらすべての基 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者数を

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構築方 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につい

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 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。複 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で、 ?(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 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、”+Add 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

Figure 10.11: 基準の適用による患者の適格性の説明

Figure 10.12: 追加の選択基準 1

DRUG_EXPOSURE なので、”Add Drug Exposure (薬剤曝露を追加)“をクリックし、高血圧治
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。

Figure 10.13: 追加の選択基準 2

「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。こ
は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、論理演算子を

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、高血圧治療薬のコ
10.14 と照らし合わせてロジックを確認します。

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。「こ

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ないことが

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。図 10.15 と照らし合わせてロジックを確認しましょう。

Figure 10.15: コホート離脱基準

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
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
(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

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

分かりやすくするために、SQLをいくつかのチャンクに分割し、各チャンクが次のチャンクで

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DatabaseConne

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

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

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必要
`<- "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阻害薬を含むす

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す：

```
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にSE
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することに

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

```
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

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

```
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

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分析ツール全体で使用（
- 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論的なアプロ
- 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
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テーブルに急性心筋

- 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含
- 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リソースとなりま

- 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するための、ト
- 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、これらの方法は「インデックス日」と呼ばれるイベントに対して
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人のインデックス

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。本章では

11.1

関心集団についての特性評価の問いに答える前に、使用するデータベースの特性をまず理解する必要があ
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は？

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ある
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

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
the Reporting of Observation Studies in Epidemiology（STROBE）ガイドラインで詳述されて
(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
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

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）を要

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

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

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

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ための公
11.2 は、単一の人に対する発生率の計算要素を示すことを目的としています：

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の翌日に始まり

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、個人が
発生率を計算するための2つの尺度があります：

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

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。別の言い方をする

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

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。リスク期間中にあ

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レベルのDUSの

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きな利益がもた
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法を提供します。

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リスト
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ます：

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

特定の関心のあるコンデ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
“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、行をクリックします。
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ごと

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、高血
“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索して関心のある

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

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 Id	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

B.7）。

コホート定義

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、図 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

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の事前設定さ
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の使用について詳しく説明

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用する特性

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する説明を
“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情報を計算

- 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、その人の観察期間に
- 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。

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

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、新たにサブグ

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す。この手

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、サブグループは選択

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できます

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、実行リス
“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もできます

11.7.3

CONDITION / Condition Occurrence Long Term / stratified by Female

Export | Export comparison | Show 10 entries | Search: cardi

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
			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

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」の既往を持つ人の割合を 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

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）を見つけることも “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、血管

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が確

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元指

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。カスタム特性に

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、コホート内の人数を特定するカ

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。基準を定義し保 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タム機

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

Export Show 10 entries Search:

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

Export Export comparison Show 10 entries Search: edema

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
			Count	Pct	Female		Count	Pct	Female		
					Count	Pct			Count	Pct	
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: angi

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 1 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 1 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

Figure 11.15: ATLASでのカスタム特性定義

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（共変

- 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
10章で説明されています。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
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

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DatabaseC

```
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
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
`<- "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

```
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（開始日を含む）の

多く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、これら

- 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）と再定義し、短期ウ
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を指定することもできま

```
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に要約統計を提供

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対

```
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
//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て
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。これらの方法は一般化

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要約す

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スタディを

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（付録 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。

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、イベントコホートを定義します。
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:

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るはずです。次に、いくつかの追加の分析設定を決定する必要が

- 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見なされる日数
  + イベントコホート2」として組み合わせます。
- 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、出力から削除され
- 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デザインを実行

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、対象コホートの総人

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す。図の中心は
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、人々にとって2番目のイベ

Figure 11.21: 経路の詳細を表示するパスウェイアウトカム

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。こ

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し
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。

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ザイド
acute myocardial infarction events (急性心筋梗塞イベント) 、 an-

Figure 11.22: 対象およびアウトカム定義の発生率

gioedema 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

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とび終了日を基
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:

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手順に従います。設計が完了したので、一つまたは複数のデータベースに対して設計

11.10.2

[生成] タブをクリックし、 ▶ Generate ボタンをクリックして、分析を実行するデータベースの一覧

Figure 11.25: 発生率分析実行

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、指定された対象コ

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を

それら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。 Reports ボタンをクリックして発生率分析の結果を表示します:

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ます。
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つい

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、 Reports ボタンをクリックして詳細を確認します。

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。

	Persons	Cases	Proportion [+|-] per 1k persons	Time At Risk (years)	Rate [+|-] per 1k years
Summary Statistics:	816,867	1,703	2.08	623,402	2.73

Stratify Rule	N	Cases	Proportion [+|-] per 1k persons	Time At Risk (years)	Rate [+|-] per 1k years
1. Gender = Female	313,518	416	1.33	230,157	1.81

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

	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

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のツ
- 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、複数の時間枠の基準特性
- 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を提

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、アクセス可能なその他のインスタンスを使用でき

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い
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作
(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さ
GI出血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界のエビデンス

- 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する。
- 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、別の曝露（比較曝露

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者に何が起こっ

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まれます。方法

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集団レベルの推定

12.1

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）を開始した対象
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。

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治療

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の
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を

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

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、治療を比較対照と

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

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

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
(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0.4であり、
Jun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：ターゲッ

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、採用された推定方法
“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ツールはそのような実践を
(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ンディショ
(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、どの特性が

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その日に記録され

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駆動型アプロー
(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可能性は低いで
(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なく、異なる研究

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。代わりに、マッチ
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、主要な診断は二つのグループ

$$\ln\left(\frac{F}{1-F}\right) = \ln\left(\frac{S}{1-S}\right) - \ln\left(\frac{P}{1-P}\right)$$

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、少なくとも半数の曝露が選

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始
対照時間	対照時間として使用される期間

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。図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。各患者特性につい(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曝露前

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、曝
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。

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

12.3

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、
12.3 の選択肢が、症例対照の質問を定義します。

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

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、アウトカムを持たない被験者

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から
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）で考慮するか

12.4

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、アウトカムの日付より前の

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率が、アウトカ
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異による交絡に
時間-コントロールデザイン (Suissa, 1995) が開発され、例えば年齢や性別で一致させた対照をケース・ク

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）で考慮
対照時間	対照時間として使用される期間

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と関
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

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ウト

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、時間変動す (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、関

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す (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
(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として現れる
(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
(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
(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ていませ
(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）が考えら

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、次の比較推定質問に対処

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のアンジ

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の急性心

これらは比較効果推定の質問であるため、セクション @ref (Cohort-Method) 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患者

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発生として

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に30日のキ

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、条件、薬剤

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阻害薬の新規使用
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チア
アウトカムコホート	アンジオエデマまたは急性心筋梗塞。
リスク期間	治療開始の翌日から開始し、曝露が終了するまで。
モデル	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。

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、真の効果サ
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の左バー
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付
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。複数の比較と複

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、新しいタ
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」をクリ
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。完了すると、
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

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ない
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。ネガ
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。

Figure 12.6: 比較ダイアログ

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、傾向スコアモデ

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。この
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（その下位層も含まれま

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。

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

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ウ
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、患者ごとの最初の暴露に限定す
minimum required continuous observation time prior to index date for a

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

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、インデックス日の前
というオプションは、コホート定義ですでに設定されていることが多いので、ここでは0のままにしてお

“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、リスク期間の重複
オプションと併せて、対象がターゲットコホートと比較コホートの両方に存在するいる場合にどのように

- “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。このオプショ
- “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ことを示し
- “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します。

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い
12.10 に示します。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ます。こ

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にアウトカ

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トの定義は

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が、アウト
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使用するか

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、共変量のセットを変更で

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仮定した場合の2つのコホートに含まれる

Figure 12.11: 研究対象集団の設定

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、これらの設定は、

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、比較の設定で除外する概念に下位層を追加することを選択していることに注意して下

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定義され

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、治療終了直
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日に設定されます。副作

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、コホート開始日

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。上位

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を

- 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向スコ

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する研

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する

Test each covariate for correlation with the target assignment?
(各共変量とターゲットの割付の相関を検定しますか。) 共変量が異常に高い相関（正または負）

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か。)
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。このよう

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人に設定すること

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する
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。この例で

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。これは傾向モデ

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)を使うこともできます。

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、共変量が多ければ、モデルを適合

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るた

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、操作特性

Figure 12.14: 傾向スコアによる調整の設定

Figure 12.15: アウトカムモデルの設定

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選択し
“ingrown nail of foot (足の陥入爪)” の出現も、上位層の “ingrown nail
(陥入爪)” 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きにど

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ポジティブコントロールも含めることができます。様々な理
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。

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
(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

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

Positive Control Synthesis

Should we perform positive control synthesis? (to calibrate confidence intervals)

Yes

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

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して 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

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。すべて

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 correla-
tion between covariate(s) and treatment detected.”（共変量と治療の間に高い相関が検出さ
です。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る
(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コードを自分で
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、CDM
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につい

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。ACEi、TH
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。 Cohort-
MethodはDatabaseConnectorパッケージを使用しており、`createConnectionDetails`という

```
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バージョンを定義してい
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るため、
<- "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タを抽出するよ

```
#
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。createDefa

(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が

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、cohor
(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、将来

```
saveCohortMethodData(cmData, "AceiVsThzForAngioedema")
```

将来のセッションでデータをロードするには、loadCohortMethodData()関数を使用できます。

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され、

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、firstExposureOnly、
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定義さ

12.8.3

通常、曝露コホートとアウトカムコホートは独立して定義されます。効果サイズの推定値を算出するには

```
studyPop <- createStudyPopulation(cohortMethodData = cmData,
                                  outcomeId = 3,
                                  firstExposureOnly = FALSE,
                                  restrictToCommonPeriod = FALSE,
                                  washoutPeriod = 0,
                                  removeDuplicateSubjects = "remove all",
                                  removeSubjectsWithPriorOutcome = TRUE,
                                  minDaysAtRisk = 1,
                                  riskWindowStart = 1,
                                  startAnchor = "cohort start",
                                  riskWindowEnd = 0,
                                  endAnchor = "cohort end")
```

firstExposureOnlyとremoveDuplicateSubjectsをFALSEに設定し、washoutPeriodを0に設定しているの
= 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た曝露終了時
= 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終了日に切

```
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、各個人に傾向

```
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。プロペ

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

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るこ

```
#      :
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指定します：

```
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し

```
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、AMIの

```
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ます：

```
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。これが満たされているかどう

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要があります。
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けた治療が予測

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が良い考えです。
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。変数があまりにも予測的である場

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

ベータ	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: 蕁麻疹
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。変数が明らかに曝露の一

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を選択
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を超える共変量がな

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します。

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かどうかを
12.20 に示すように、`drawAttritionDiagram`関数を使用して私たちの研究での対象者の脱落を表示できま

レトロスペクティブ研究ではサンプルサイズは固定されており（データはすでに収集されている）、真の

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満

追跡可能なフォローアップの量をよりよく理解するために、フォローアップ時間の分布を検査すること
12.21 に示されるように、両コホートのフォローアップ時間が比較可能であることがわかります。

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

12.9.4

最後に、カプラン–マイヤープロットをレビューし、両コホートの時間経過による生存率を示します。pl
12.22 を作成し、ハザードの比例性の仮定が保持されているかどうかなどを確認できます。カプラン–マ

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。これは、ACEiがTHZと比較して血管浮腫のリスクを増加させることを示してい
- 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。前述の診断ア
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。

Figure 12.22: カプラン–マイヤープロット

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ったかというこ
- 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、適切な反事実を作成するため

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、Databa

```
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データセッ

```
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も使用しま

```
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）出血のリスクは？
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

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からCoh

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、事前にア

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

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

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経路を推測する

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。現在使用されているほとん

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レベルの予測
[1] は、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、透明性に関連するいくつかの

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、共通データモデル（CDM）

この章では、OHDSIの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と評価のため

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、定義された時点（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。予測は、

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、アウトカム

[1] https://www.equator-network.org/reporting-guidelines/tripod-statement/

Figure 13.1: 予測課題

[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]を持つよ
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観察

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含

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
  - 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
- 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
- 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達成するの

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、機械学習の一形

共変量（“予測因子”、“特徴”、または”独立変数”とも呼ばれる）は、患者の特性を説明します。共
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（本書ではこれを基準日

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”とも呼ばれる）

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。コホート定義IDが1のコホ

Table 13.2: 例示的なCOHORTテーブル。簡潔のためにCOHORT_END_DATEは省略しています。

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）から1年間と仮定
1に対して0（非存在）（状態が基準日後に発生）と、個人ID 2に対して1（存在）を持ちます。同様に、
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ID
2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。前の例では、
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、その時点で状

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の関係
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）に対

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としま

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ア

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
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。
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

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習し、最終
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）パッケージを

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、初期分散はアウトカ

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、我々の枠組みでは複数の決定木を
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。バギングの背後にあ

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デー
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独立

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
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	4
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。lear	

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
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は分	10^-7
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、入力を重み付けするた

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
パッケージの別のビネットで、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います。

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、これはこの章の範囲外です。
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致度を測定することによって行うことが

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しなけ
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
- 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、トレイン
  - 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し、
  - 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。データを等しいサイズに
  10）。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訓練し、

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

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の間の値を各
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

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	リスク期間中にアウトカムを持つ	タイプ
1	0.8	1	1	TP
2	0.1	0	0	TN
3	0.7	1	0	FP
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

患者ID	予測リスク	0.5閾値での予測リスク	リスク期間中にアウトカムを持つ	
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）と呼びます。患

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、アウトカムの
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、感度を向上さ

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能力のこ
- 特異度、y軸に感度をプロットすることで作成されます。ROCプロットの例は、この章の後半に図
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を示し、値が

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がどれだけ
13.3 に示されているようにです。

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、閾
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

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人の患者に
＝yの線上に位置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ことを
13.18 で示されています。また、これらの点を使用して線形モデルをフィットし、切片（ゼロに近いはず
Calibration Curvesも実装しています。

Figure 13.3: 識別力に関連する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

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明確に
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を要求することにより、予測問題の適

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る血
(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まで
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の増加
(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、しばしば最初
(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

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に取

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、例えばアウトカ

- 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ソース
- 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では、個人は異なる期間にコホートに複数回適格となる可能性があります
- 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を決める
- 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ターゲ
- 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データベース

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。ある予測問題に
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を許可して
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、性別、年

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、ここでは内
- 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に使用

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365日未満、
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日のリスク期間が必

選択	値
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 ボタ

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つのセク

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。予測モデ

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ます。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は、「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。

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 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てい

我々の例では、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

Figure 13.4: 予測問題設定

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。共変量設定ビューでは、含め

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。これで共変量設定ビュ

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に対して
“What concepts do you want to include in baseline covariates in the
patient-level prediction model? (Leave blank if you want to include ev-
erything) (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めた
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に下位層コン
“Should descendant concepts be added to the list of included con-
cepts? (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を除外する
“What concepts do you want to exclude in baseline covariates in the
patient-level prediction model? (Leave blank if you want to include ev-
erything) (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す
13.6のようになります。

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：

Figure 13.5: 勾配ブースティングマシン設定

Figure 13.6: 共変量の包括と除外設定

- 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意味し、0はそう
- 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意味し、0は
- 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、0はそ
- 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日の月を表
- 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した日数に
- 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した日数に相
- コホート時間： 患者がコホートに属していた日数（コホート終了日ーコホート開始日）に対応する
- 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を表すバイ

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。

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
  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
  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
  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
  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、終了がコホート開始

これが完了すると、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

						Era Start		
Domain	Any Time Prior	Long Term (-365 days)	Medium Term (-180 days)	Short Term (-30 days)	Overlapping	Long Term (-365 days)	Medium Term (-180 days)	Short Term (-30 days)
Condition	☐	☐	☐	☐	☐	☐	☐	☐
Condition Group	☐	☑	☐	☑	☐	☐	☐	☐
Drug	☐	☐	☐	☐	☐	☐	☐	☐
Drug Group	☐	☑	☐	☑	☐	☐	☐	☐

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

- 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_OCCURRENCテーブル
- 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テーブル 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 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 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、MEASUR 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かを示すバイナ
- 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、OBSE
- 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テーブルの
- 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、その時間間隔に記録され 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共変量を構築し、

“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、異なるコンセプト

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。

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	☐	☐	☐	☐	☐	☐	☐
Condition - Primary Inpatient	☐	☐	☐	☐			
Drug	☐	☐	☐	☐	☐	☐	☐
Procedure	☐	☑	☐	☐	☐	☐	☐
Measurement	☐	☑	☐	☐	☐	☐	☐
Measurement - Value	☐	☐	☐	☐			
Measurement - Range Group	☐	☑	☐	☐			
Observation	☐	☑	☐	☐	☐	☐	☐
Device	☐	☑	☐	☐			
Visit - Count		☑	☐	☐			
Visit - Concept Count		☐	☐	☐			

Figure 13.10: 時間制約共変量

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。設
13.11 のようになります。

Set the index score covariates

Index Score Type	
$CHADS_2$	☐
CHA_2DS_2VASc	☑
DCSI	☑
Charlson	☑

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、リスク
研究に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

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、関心の対象で
“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
は、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の
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、最小ベースライ

“Should subjects without time at risk be removed? (リスク時間がない対象は除外すべきですか
が”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、
1）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、最小
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、すべ

“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、リスク時間中にアウ

“Should only the first exposure per subject be included? (対象ごとに最初の曝露のみを含める
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がい

“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みが保持されます

“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するため、
“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。患者が以前にアウト

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。

Figure 13.12: 対象集団設定

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選択します（
  “no (いいえ)”）。“yes (はい)” に設定すると、別のオプションが表示されます：
  “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してテス

- “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、その共
  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）。共変量出現の最小閾値は、

Figure 13.13: 分析の設定

- “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です。

この例では、図 13.14 のように選択します。

Figure 13.14: 実行の設定

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けるかを
トレーニング/テストデータを人別（アウトカムで層別化）または時間別（古いデータをトレーニン
- “Percentage of the data to be used as the test set (0-100%)
(テストセットとして使用するデータの割合(0-100%))” ： テストデータとして使用するデータの割
- “The number of folds used in the cross validation (クロスバリデーションで使用するフォールド数)
最適なハイパーパラメータを選択するために使用するクロスバリデーションのフォールド数を選択
- “The seed used to split the test/train set when using a person type
testSplit (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
(オプション)：)” ： 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、分割に使用す

この例では、図 13.15 のように選択します。

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、選択したモデ

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ドウに
“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て上書きさ

Figure 13.15: トレーニングの設定

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)”」レビューとダウンロード” タブをクリックします。“Downloa

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
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がインスト
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
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

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

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。ここでは
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。コホートのイン
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であ

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、`DatabaseConnect`

```
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バージョンを指定し
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ることに
`<- "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。共

```
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PatientLevelPredictio

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、将来のセッションのた

```
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。例えに
= 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了日に開始
= 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できます。

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

```
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します。データ
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）およびtestFract

```
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デー

runPlp関数には、plpData、plpResults、plpPlots、evaluationなどのオブジェクトを保存す

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モ
= 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
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新しいデー

```
#
plpModel <- loadPlpModel("model")

#   plpData
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ことができます。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('cdms1chema.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。これはアウ

```
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0.74に低下13.17に示されています。

	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

Figure 13.16: Shinyアプリケーションにおける評価統計の要約

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たリスク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モデル

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

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

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、ターゲット人口の大部分が失

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を生成およ

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。

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、ターゲット集団
- 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、関心のあるリ

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ます。行が
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。

Figure 13.20: 予測問題におけるattritionプロット

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

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。 Performance
をクリックして閾値における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（PPV）、負の予
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%がリスク0.00482以

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、精度-
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に対応してい
13.24 はROCおよび精度-再現プロットを示しています。ROCプロットは、モデルが翌年にアウトカムが
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に見えます。

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。

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。こ
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。

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しているよ

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

Figure 13.25: アウトカム有およびアoutingの患者に対する予測リスク分布。重なりが多いほと

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

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、図 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

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。

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は係数、

Model Table

Download Model

Show 10 entries

Search:

	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 Previous 1 2 3 4 5 ... 6790 Next

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ペーパーの
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を作成でき

```
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すること
- 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、具体的な問題とデー
- PatientLevelPredictionパッケージは、OMOP CDMに保存されたデータを使用した予測モデル
- 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外部検証を

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、D

```
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、これをローカル

```
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使用し

```
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、次の年に消化
NSAIDの新規ユーザーコホートのCOHORT_DEFINITION_IDは4です。GI出血コホートのCOH

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、

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`crea`

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の性能

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でしょう。

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であるべきであり、特定の質問に対して同じデータに同じ分析を

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自信を持つ可能
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した際の識別精
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。各データソースに

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過度
(Madigan et al., 2013a) 。 母集団レベルの効果推定では、感度分析には、コホート比較研究や

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。未知の質問に対
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
(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
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
(Botsis et al., 2010; Hersh et al., 2013; Sherman et al., 2016)
。一般的に指摘されるのは、観察研究で使用されるデータはもともと研究目的で収集されたも
(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
(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

> 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

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

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

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
臨床的妥当性	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
ソフトウェアの妥当性	データの変換および分析プロセスが意図した通りに機能
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

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表現する能

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
  - 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る不確実性を表

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。例えば、電子カル
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。懸念される

> 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、正確性。

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ください。

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評価には２つの
(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評価と特定の研究におけ

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は数多く存在
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、各ステップにDQを統

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、欠測値のデフォルト値
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やメタデータの
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などの問題が
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を改善す

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の一部について
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。い
(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

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、このステップにお
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

   - 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
   - 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、

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
   - 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

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、ローカルの知識に焦
- 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

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（デ
Characterization of Health Information at Large-scale Longitudinal Evi-
dence Systems）があります (Huser et al., 2018)。ACHILLESは、CDMに準拠したデータベース
(Huser et al., 2016) 。ACHILLESはスタンドアロンツールとして利用でき、「データソース」機

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あ
DRUG_CONCEPT_IDによるDAYS_SUPPLYの分布」や「506: 性別による死亡時の年齢の分布」などがあ

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。合計 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。各チェック 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 measurement_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。	0.00	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、フィールド、

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れたD

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。デー
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は、一連のユニットテスト
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作成

```
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_C
実行後には、さまざまな期待値を持つ 2 つのエントリが PERSON
テーブルに存在しているはずであるという期待値を指定します。

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
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、テ
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、通常は数百

```
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。

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

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

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す。

15.3

この章では、これまでは一般的な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は、データ

これらの評価の一部は、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。例えば、関心

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、例えば、コホート

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の有病率や経時
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ださい。一部の

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へ
ボ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摘さ
[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完全に

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
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
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
9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

‰ 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

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が、
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
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てい
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、壊疽
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（「肺壊死」

[1] https://github.com/OHDSI/Vocabulary-v5.0/issues

Figure 15.3: 孤立したソースコードのサンプル

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。

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、DatabaseConnectorパッケ

```
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Rに伝えるた
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ることに
`<- 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データと同じ
ACHILLESデータベース特性評価を表示することができます。これは、ATLASをACHILLES結果データベー

```
exportToJson(connectionDetails,
             cdmDatabaseSchema = cdmDatabaseSchema,
             resultsDatabaseSchema = cdmDatabaseSchema,
             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
15.4 はACHILLESデータ密度プロットを示しています。このプロットは、データの大部分が20

Figure 15.4: ACHILLESウエブビューワーでのデータ密度プロット

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ている

Figure 15.5: ACHILLESウエブビューワーでの月次糖尿病の診断コードの有病率

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を説明します。
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、一連のチェックを実行します。
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が正しいバー
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

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。その後、

```
viewDqDashboard(jsonPath)
```

変数jsonPathは、上記のexecuteDqChecks関数を呼び出す際に指定したoutputFolderにある、ダッシュボ

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。このテーブルには、

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結果が表示さ
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少なくとも1つの

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

```
library(MethodEvaluation)
json <- readChar("cohort.json", file.info("cohort.json")$size)
```

OVERVIEW
METADATA
RESULTS
ABOUT

DATA QUALITY ASSESSMENT

SYNTHEA SYNTHETIC HEALTH DATABASE

Results generated at 2019-08-21 18:08:28 in 13 mins

	Verification				Validation				Total			
	Pass	Fail	Total	% Pass	Pass	Fail	Total	% Pass	Pass	Fail	Total	% Pass
Plausibility	159	21	180	88%	283	0	283	100%	442	21	463	95%
Conformance	637	34	671	95%	104	0	104	100%	741	34	775	96%
Completeness	369	17	386	96%	5	10	15	33%	374	27	401	93%
Total	1165	72	1237	94%	392	10	402	98%	1557	82	1639	95%

Figure 15.6: Data Quality Dashboard-におけるデータ品質チェックの概要

RESULTS

SYNTHEA SYNTHETIC HEALTH DATABASE

Results generated at 2019-08-22 14:15:06 in 29 mins

Column visibility CSV

Show 5 entries Search:

STATUS	CONTEXT	CATEGORY	SUBCATEGORY	LEVEL	DESCRIPTION	% RECORDS
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

```
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、血管性浮腫のコホート定義
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見ると、二

‰ 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

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。ここでは

```
orphans <- findOrphanSourceCodes(connectionDetails,
                                 cdmDatabaseSchema = cdmDbSchema,
                                 conceptName = "Angioedema",
                                 conceptSynonyms = c("Angioneurotic edema",
                                                     "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）として見つかったの

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か
- 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
- 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ル
- 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がインストール済みであると想定します。また、SqlRender、DatabaseCo

```
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セットを

```
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ようなものだ。フ

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世界」です。レト
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ります。臨床的妥
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浮腫を引き起こす

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阻害薬の使用と血管浮腫の関係
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は、元のデータ
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にまたがるデー

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、その治療を反映

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とは区別される、

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。1）観察可能な

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に、「コホート

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、研究プロセス

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系利尿
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、他の原因に
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、疾患の発症

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、コホート定義の

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、その定義を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と
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判定です。

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十分なリソー
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する。
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めのガイドラ
5. 臨床専門家がサンプル内の人物について、利用可能なすべてのデータを検証および判定し、各対象者
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、収集したデータから可能な

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）の評価に限定されます。これ
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状態や特徴

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のような単一の指標の

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、米国国立がん研究所（NCI）の実

1. OHDSI がんフェノタイピング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含み 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. ATLAS 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が真陽性か偽陽性かを判断するために、入院患者と外来患者の 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、理想的には合意と評価
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、病理報告書、
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の控えめな推定値か
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、コホート定義に
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推定

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、研究

16.4 PheValuator

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
(Swerdel et al., 2019) 。一般的な考え方は、臨床医がソースレコードの検証で実施するのと同
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時に
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
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、幅広い患者の特徴を予測因子として使用してモデルを適合し、その人
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
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

診断予測モデリングでは、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
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
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
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
コホートは、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 PheValuator

PheValuator 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

PheValuator を使用して心筋梗塞のコホート定義をテストする手順は以下の通りです。

[1] https://github.com/OHDSI/PheValuator

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レコードで、
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。

Figure 16.2: 心筋梗塞の極めて特異的なコホート定義（xSpec）

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、病歴の任意
16.3 は、ATLASにおけるMIのxSensコホート定義を示しています。

Figure 16.3: 心筋梗塞の極度に感度の高いコホート定義（xSens）

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
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

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

モデリングプロセスに含まれる人物の特徴を指定するために使用できるパラメータがいくつか
ID に設定することで、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
次の例を参照してください:

```
setwd("c:/temp")
library(PheValuator)
connectionDetails <- createConnectionDetails(
  dbms = "postgresql",
  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、コホートテーブルの場所（co
- 「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が2010年
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歳から90

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カムの予測確率
`lowerAgeLimit`および`upperAgeLimit`引数として年齢を設定）、性別（`gender`パラメータを男性および/ま

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_myC

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処するためのコホートの使用目的によって異な
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、より特異
PheValuator 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
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）：

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、コホート定義
   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（カウントの期待値に
   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の全人
   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
   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。この手
   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を記
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こ

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、モデ

Figure 16.4: PheValuator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

関数から出力されたRDSファイル「c:/temp/lr_results_5XMI_eval_myCDM_ePPV0.75_2018
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ます
x MI」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

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に

```
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」）も含ま
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。

Phenotype Algorithm	Database	Sens	PPV	Spec	NPV

Table 16.1: pheValuator 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診断条件 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
	OptumGE66	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
	OptumGE66	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
	OptumGE66	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
	OptumGE66	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りますが 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ました。後 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、民間

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
- 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
- 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し
- ソースレコードの検証とPheValuatorによる臨床判定は、コホート定義の検証を推定
- OHDSIネットワーク研究は、データソースの異質性を調査し、実証結果の一般化可能

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、私たちの分析ソフトウェ
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が不可欠であり、
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 Li-
brary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ため、分析全体の

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。次に、ソフ
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。データベースの専
したがって、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。自動化とは、分
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言語が推奨されてい
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、非常に複雑になる可能
(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、多くの書籍が書ける

- 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、コードの行と行の間
- 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、明確に定義する
- 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
  `<- 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

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

- 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

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
end」をexposure end dateを含むと解釈すべきか、それともそうでないか）が必要なため、2

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
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
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
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

17.2 Methods Librar

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
tracker（例えば、CohortMethod issue tracker[1]）とOHDSIフォーラムの2つの場所で議論され

[1] https://github.com/OHDSI/CohortMethod/issues
[2] http://forums.ohdsi.org/

集団レベルの推定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患者レベルの予測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」と呼ばれるシステムを Library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ため、適切 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コミ 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、継続的な評価と改善の対

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バージョン 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、OHDSIコミ 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、こうしたリクエストを 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の変更のあらゆ

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のリー

- 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。新しい機能はな
- 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。後方互換性のみが保
- 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。互換性については保

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じて文書化されて 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。コミュニティからのフィードバ

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。まず、GitHu dratリポジトリに保存されています。

[3] https://ohdsi.github.io/MethodsLibrary/

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、パッチに関して積

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学
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べての

17.2.6

OHDSI Methods LibraryはGitHub[4]システムでホストされています。GitHubのセキュリティ対策
//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Metho
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、マスターブランチに

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

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
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

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

[4] https://github.com/

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、分析全体を実行する
- 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、最良の
- 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することが
- 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プロセ

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します。

> 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、方法の妥当
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、方法の妥当性を良好に
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対処しておく必要

方法の妥当性を確立する上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るかどうかを評

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、研究デザインに特化し
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アップされており、

- 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バランス（ベ
- 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、対象とする初期

- 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
  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
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
18.1 では、2つの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

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奨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

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、関心のあ
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、さらに以

- 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、仮説を否定するもの
  時には、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（例えばハザード比
- 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べきです。この
- 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が互
- 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。例えば、社会保障番

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べきであると
(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。現実に見られる変

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、関係性の
(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情報は自動的に

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、帰無仮説が真
(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、アウトカムを追加でシミュレー

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が示されるか
(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結果を予測

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原因をシミュレー

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、異なる手法では
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じに

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、さ

- 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
- カバー率：真の効果量が95%信頼区間内に収まる頻度。
- 平均精度：精度は $1/(\quad)^2$ として計算され、精度が高いほど信頼区間が狭くなる。精度の
- 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

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
(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

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
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
$\hat{\theta}_i$ で表し、対応する推定標準誤差を $\hat{\tau}_i$ 、 $i = 1, \ldots, n$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
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値の計算と
$\hat{\theta}_i$ は、 $\theta_i + \beta_i$ 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
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
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により求めます:

$$L(\mu,\sigma|\theta,\tau) \propto \prod_{i=1}^{n} \int p(\hat{\theta}_i|\beta_i,\theta_i,\hat{\tau}_i)p(\beta_i|\mu,\sigma)\mathrm{d}\beta_i$$

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て計算しま
アウトカムペアの効果推定 $\hat{\theta}_{n+1}$ 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。真の効果量は、95%信頼
(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の枠

厳密には、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、今回は平均と標準
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、

$$\mu(\theta_i) = a + b \times \theta_i \text{ かつ}$$
$$\sigma(\theta_i)^2 = c + d \times \mid \theta_i \mid$$

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：

$$l(a, b, c, d|\theta, \hat{\theta}, \hat{\tau}) \propto \prod_{i=1}^{n} \int p(\hat{\theta}_i|\beta_i, \theta_i, \hat{\tau}_i) p(\beta_i|a, b, c, d, \theta_i) \mathrm{d}\beta_i,$$

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します
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：

$$\hat{\theta}_{n+1} \sim N(\theta_{n+1} + \hat{a} + \hat{b} \times \theta_{n+1}, \hat{c} + \hat{d} \times \mid \theta_{n+1} \mid + \hat{\tau}_{n+1}^2).$$

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めます

$$\Phi\left(\frac{\theta_{n+1} + \hat{a} + \hat{b} \times \theta_{n+1} - \hat{\theta}_{n+1}}{\sqrt{(\hat{c} + \hat{d} \times \mid \theta_{n+1} \mid) + \hat{\tau}_{n+1}^2}}\right) = 0.025,$$

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

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収
(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
(Schuemie et al., 2018a)。

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
(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ま
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。大
(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

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、その逆は

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、層化傾向スコアマ

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研究の結論

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
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によって
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、次の部で説明されているよ

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の組

ネガティブコントロールの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コンセプトセ
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。

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セット

次に、「Explore Evidence」タブに移動し、 Generate ボタンをクリックします。エビデンス概要の生成
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、結果のリストが表示されます。

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%

Column visibility | Copy | CSV | Show 15 entries | Filter:

Showing 1 to 15 of 13,787 entries | Previous 1 2 3 4 5 ... 920 Next

Suggested Negative Control
No (12777)
Yes (1010)

Found in Publications
No (12398)
Yes (Parent) (1160)
Yes (Exact) (229)

Found on Product Label
No (12667)
Yes (Parent) (878)
Yes (Exact) (242)

Found in Product Label Or Publications
Yes (10576)
No (3211)

Signal in FAERS
No (10951)
Yes (Parent) (1949)

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)
Rift valley fever	Y	13,781	0	0	0	0	0	0
Obstruction due to foreign body accidentally left in operative wound AND/OR body cavity during a procedure	Y	13,780	0	0	0	0	0	0
Infection by Shigella	Y	13,766	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を示す

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、そのようなコホ
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ー
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

```
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いうことで

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とポジティブコ
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易に可能にしま

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ロールについて、

これらの推定値をもとに、MethodEvaluationパッケージの`computeMetrics`関数を使用して、表
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たメソッドの

メトリクス	値
ROC曲線下面積（AUC）	0.96

メトリクス	値
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

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が、表 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が

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Shinyア 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、経験的なキャリブ

```
# 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)
```

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

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
```

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
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、真の効果サイズを含ま

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。ダイアモンド型は関心の

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値は非常に類似

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、信頼区間をキ

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4.3
- 8.08）と1.13（0.59 - 2.18）です。補正されたハザード比はそれぞれ4.75（2.52
- 9.04）および1.15（0.58 - 2.30）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）で分析を実
18.7 は、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と
(DerSimonian and Laird, 1986) を示しています。この図は、EvidenceSynthesisパッケージのp

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利尿薬

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が示
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
を計算すると、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各

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およびサイ

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した。しか
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と血管性浮腫に対する

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（95

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の信頼

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。すべて
ハッキング（望ましい結果を得るためにデザインを調整すること）を避けるため、対象となる効果量

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ですが、
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ウトカム、
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。手法を評価
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、Method
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
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

Figure 18.9: Methods LibraryのMethod-
sに対する95%信頼区間のカバー率。各ドットは分析選択の特定セットの性能を表します。点線
= 自己対照症ケースシリーズ、GI = 消化管、IBD = 炎症性腸疾患。

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
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%信頼区間には含まれていない可能

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よ

[1] http://data.ohdsi.org/MethodEvalViewer/

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回答を
- 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特性を示しませ
- 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、名目上の特性に復元する
- 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検性を維持する限

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を提供するこ

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要約しま

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るため
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、主に自身の施設でテ
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れば、原則と
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は行われない研
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が、多くの場合
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が
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を完全に
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答え
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識的にアプロ
ハッキングと呼ばれることもあります。EHRや保険請求データなどのデータは、時に研究者に無限の可能

や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、詳細

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
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
時には、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

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

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

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
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
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
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
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

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
型糖尿病(T2DM)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

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患者に対する処方方法

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定の患者群

あるいは、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、研究の問いを定義するには、問いに答えるために必要なアプローチがOHDSIツ
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修正して、

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）、どのフィールドにデータが入

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、上記の例に戻りましょう。
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、推奨された

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。観察研究

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の解釈にお

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のに適した
8 および 第 10 章で詳しく説明されています。簡単に説明すると、ATLASは、詳細な包含基準を定義して
UI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・除外の基準、お
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）に基づいて定義することができ
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が標準コード
包含基準で使用する最適なコードセットが明確でない場合は、コホート定義で探索的分析を行うことが妥
あるいは、異なるコードセットを使用してコホートを定義するさまざまな可能性を考慮するために、より

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を生成するため

コホートが作成されると、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や状態の頻度を

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、それらをさまざ
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アウトカム
ツールセットは、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の生成を迅速
18 章 で詳しく説明しています。

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、頑健で検証済みの表

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た

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
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

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、対象コホートと比
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、検出

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）のマッチング手順と関
12 章で詳しく説明されています。さらに、これらの最終的にマッチングされたコホートを使用

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（MDRR

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。このプ
D 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
これらの後者のステップは、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的な研究

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

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

Figure 19.1: 研究プロセスのダイアグラム

19.3

- 研究には明確に定義された問いを検討すべきです。
- 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ることを推奨
- 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、最終的な研究を実行する前
- 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るべきです。

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。このミッションを達
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にわたって透明性

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国際的な協力体制
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とができます。デ

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（臨床コンセプトの
- 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、さまざまな分野のリーダー
- 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臨床特性評価や

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いて説明しました

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比較
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、
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、観察研究の結

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

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
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
このような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
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
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
- 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
- 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
- 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

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
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
コードの選択によって、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・構築するわ
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。データベ

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
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ことです。
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ります。幸

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、共同研究者がコミュニ
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ネットワーク

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パッケージを許可し、インス

**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とは別に定期

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。この担当者は、研

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承認プロセ

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、研究の問いを定義し、研究プロトコル

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プロ

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、提案された研究の
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定
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共
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の研究診断を実行することが含ま
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、内部
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

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

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

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）と結果の提出期限を伝える責任を負います。施設は、送信方法が内部プ

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

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

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
によるデータベースへのアクセス不可など）に関連する課題が生じる可能性があります。参加
CDM 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

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

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

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
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

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/著者名推定ツール）

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っているこ
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シンポジ

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、研究設計、コ

Figure 20.1: ARACHNEネットワーク研究プロセス

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。ARACHN
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研究チーム
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R、Pyth
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、自己完結型パッケー

[1] http://www.icmje.org/recommendations/browse/roles-and-responsibilities/defining-the-role-of-authors-and-contributors.html

Figure 20.2: ARACHNEのネットワークのネットワークス

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ベスト

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デー
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、そ
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った
この測定誤差は、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

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使用

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
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
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
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し、
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を使

20.6

- OHDSI研究は、異なる機関の複数のCDMで実施されると、OHDSIネットワーク研究となります
- OHDSIネットワーク研究は、すべての人に公開されています。ネットワーク研究のリーダーは
- 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、研究の設計と実施
- 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に開発されている

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の分析のデ

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）の単位（

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デザインの

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、ターゲットとする

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コンセプトの 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（同じ分析を変更な
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
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、徴候、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き

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

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ま

薬剤（Drug） 人に投与されたときに特定の生理学的効果を発揮するように調製された生化学物

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が

電子的健康記録（Electronic Health Record, EHR） 医療の過程で生成され、電子システムに記

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

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

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

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記

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）が対

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

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

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

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
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
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

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関わらず

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ダイアレクト（方言）に自動変換できるS
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。詳細は第
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、ソフト
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「脳

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経験する

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だけで

集団レベル推定（Population-level estimation）　因果効果の研究。平均（集団レベル）の効果の大きさを

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と信じられ
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ために
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または実行される活

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を模倣す
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もの。

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ラクティブなソフトウ
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。詳細は第
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から逸脱した

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研

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ために研究の主

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、同義語および定義を提供する

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使用できるかど
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。詳細は第
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めることができる

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OHDSIワ

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か

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単語や
5 章参照。

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日後の間の連続した観察があり

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コードから薬剤

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日後の間の連続した

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間の出現がちょうど1回です。

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：この考え方は、指定されたコンセプトセットで見つかったコードか

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
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	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

コンセプトID	コンセプト名	除外	下位層	マッピング元
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.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日後の間の連続した観察があり、

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デッ
(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ト。

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

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うに限定し

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、インデックス日前0日 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日後の間の連続し

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
  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.11)の明確な薬物曝露期間の出現がちょうど1回です。

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：この考え方は、指定されたコンセプトセットで見つかっ

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
1314577	ネビボロール	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
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日後の間の連続した観察が

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を含む観察

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

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
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
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	エナラプリル	NO	YES	NO
1342439	トランドラプリル	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
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

コンセプトID	コンセプト名	除外	下位層	マッピング元
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	メチルドパ	NO	YES	NO
1307046	メトプロロール	NO	YES	NO
1307863	ベラパミル	NO	YES	NO
1308216	リシノプリル	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	モエキシプリル	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
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	ペンブトロール	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

コンセプトID	コンセプト名	除外	下位層	マッピング元
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日後の

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、初回

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します。インデ

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	キナプリル	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

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

コンセプトID	コンセプト名	除外	下位層	マッピング元
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	ペンブトロール	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（THZ）を比較する場

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	切断足
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	頚部の体性機能障害

コンセプトID	コンセプトの名前
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
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	類表皮嚢胞
4092896	糞便内容物異常
259995	開口部内の異物
40481632	ガングリオン嚢胞
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	白色帯下
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

コンセプトID	コンセプトの名前
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
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）を使用。
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。

カラム名	値	説明
ETHNICITY_CONCEPT_ID	38003564	38003564は「非ヒスパニック」を示す。
LOCATION_ID		住所は不明。
PROVIDER_ID		主治医が不明。
CARE_SITE		主たる医療施設が不明。
PERSON_SOURCE_VALUE	NULL	提供されていない。
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
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とは
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への外部
DRUG_CONCEPT_ID	19078461	提供されたNDCコードは標準コンセプト 19078461 にマッピングされる。
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_CONCEPT_ID	38000177	38000177 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かった。
VISIT_DETAIL_ID	NULL	
DRUG_SOURCE_VALUE	76168009520	提供されたNDCコード。

カラム名	値	説明
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

```
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_SOUR

```
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：

```r
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

A) 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。
B) 医学知識を持つ人がコードマッピングを作成します。
C) エンジニアがETLを実装します。
D) 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を持って
GENDER_CONCEPT_ID	8532	
YEAR_OF_BIRTH	NULL	生年月日の月や日が不明な場合は推測しませ
MONTH_OF_BIRTH	NULL	
DAY_OF_BIRTH	NULL	
RACE_CONCEPT_ID	0	人種はWHITEであり、これは8527にマッピ
ETHNICITY_CONCEPT_ID	8527	民族の情報が提供されていないため、これは
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義すること

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テー

```
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ことに注意してくだ

代わりに、すでに成分レベルにまとめられているDRUG_ERAテーブルを使用することもできま

```
library(DatabaseConnector)
connection <- connect(connectionDetails)

sql <- "SELECT COUNT(DISTINCT(person_id)) AS person_count
FROM @cdm.drug_era
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テーブルに

```
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。なぜな

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日間の連続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層を含む

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。

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

Figure E.2: ジクロフェナクのコンセプトセット

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。

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。

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層を含み、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日間のギャップを

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ンディション

Figure E.6: 広範な悪性腫瘍のコンセプトセット

Figure E.7: コホート・イグジット日の設定

```
library(DatabaseConnector)
connection <- connect(connectionDetails)
sql <- "SELECT person_id AS subject_id,
  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（ここで

```
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ジ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、コンディションとビ

また、ここではコホート終了日は考慮していないことに注意してください。通常、コホートがアウトカム

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コキシ

Figure E.8: データソースの特性評価

演習 11.2

をクリックして「New cohort (新規コホート)」を作成します。コホートに意味のある new users (セレコキシブ新規ユーザー)」）を付け、「Concept Sets (コンセプトセット)」タブに移動します。「New Concept Set (新規コンセプトセット)」をクリックし、コ (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、コンセプトセットにコンセ

Figure E.9: 成分「Celecoxib (セレコキシブ)」の標準コンセプトの選択

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。「+Add 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ンセプトセット attribute… (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日の連続する観察期間 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そのままにします 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 をクリックして「N
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：「Celecox
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
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つのコン
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。特性
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。完了すると、
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お

演習 11.3

をクリックして「New cohort (新規コホート)」を作成します。コホートに意
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。「
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（
bleed (消化管出血)」）を付けます。 Search モジュールを開き、「Gastrointestinal
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
E.13 参照）。

Figure E.11: 特性評価設定

Figure E.12: 特性評価の結果

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「Descemdamts (下位層)」をチェックします（図 E.14 参照）。

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)」をクリックしてから「Add Condition Occurrence (コンディション出現を追加)」をクリックします。先に作成したコンディション出 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はその をクリックして終了します。

Figure E.15: 単純な消化管出血コホート定義

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「New 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。「Add 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を選択し

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を
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。

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。データソ
(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（図
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除

```
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ったことが見てとれます。なぜなら、ここで使用

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、ベースラインの違

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。その後、傾

```
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。これは、非常に小

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違い

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, numberOfStrata = 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(standa
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、標準化
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：

```
adjModel <- fitOutcomeModel(population = strataPop,
                            modelType = "cox",
                            stratified = TRUE)
adjModel
```

```
##      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わかります

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関数を使用します：

```
library(PatientLevelPrediction)
covSettings <- createCovariateSettings(
  useDemographicsGender = TRUE,
  useDemographicsAge = TRUE,
  useConditionGroupEraLongTerm = TRUE,
  useConditionGroupEraAnyTimePrior = TRUE,
  useDrugGroupEraLongTerm = TRUE,
  useDrugGroupEraAnyTimePrior = TRUE,
  useVisitConceptCountLongTerm = TRUE,
  longTermStartDays = -365,
  endDays = -1)

plpData <- getPlpData(connectionDetails = connectionDetails,
                      cdmDatabaseSchema = "main",
                      cohortDatabaseSchema = "main",
                      cohortTable = "cohort",
                      cohortId = 4,
                      covariateSettings = covSettings,
```

```
                     outcomeDatabaseSchema = "main",
                     outcomeTable = "cohort",
                     outcomeIds = 3)

summary(plpData)
```

```
## 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のア

```
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要求することに

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関数を呼び出しま

```
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によるシ

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のAUCが0

	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 Dashbard, 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., Hripcsak, 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 Academy 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., Brannan, 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 Flaig, 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., Hripcsak, 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., Hripcsak, 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., Hripcsak, 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., Scarnecchia, 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., Hripcsak, G., and Ryan, P. B. (2019). PheValuator: 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., Hripcsak, 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 Diabetes 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 Vandenbroucke, 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., Persson, 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 target-

ing the renin-angiotensin-aldosterone system. Nat Rev Drug Discov, 1(8):621–636.

Index

ACE阻害薬, 206
ACHILLES, 242
adaboost, 202
agnostic SQL, see SqlRender
analysis implementation, 81
APHRODITE, 122
arachne, 301
ATHENA, 39
ATLAS, 83, 208
    characterization features, 150
    インストール, 84
    コホートパスウェイ, 83
    コホート定義, 83
    コンセプトセット, 83
    ジョブ, 84
    セキュリティ, 84
    データソース, 83
    ドキュメント, 84
    フィードバック, 84
    プロファイル, 83
    ボキャブラリ検索, 83
    患者レベル予測, 84
    特性の評価, 83
    発生率, 83
    設定, 84
    集団レベル推定, 84
attrition diagram, 191
AUC, 205

back-propagation, 203
balance, see covariate balance
baseline time, 141
best practice for network research, 302
between-database heterogeneity, 278
bigknn, 202
Bill of Mortality, 37

caliper, 167
    scale, 178
case-control design, 168
case-crossover design, 169
case-time-control design, 169
CDM , see Common Data Model
characterization, 75, 141
    cohort, 142
    database level, 142
    treatment pathways, 142
classification concept, 42
Clem McDonald, 238
clinical equipoise, 167
cohort
    entry event, 121
    exit criteria, 121
    inclusion criteria, 121
    probabilistic design, 121
    rule-based design, 120
cohort method, 165
colliders, 167

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, 21
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, 21
community, 5, 238
comparative effect estimation, 165
comparative effectiveness, see comparative effect estimation
comparator cohort, 165
concept, 39
class, 40
code, 43
hierarchy, 46
identifier, 39
mapping, 45
relationship, 44
concept set, 121
conditioned model, 178
confidence interval calibration, 277
confounder, 166
control hypotheses, 86
convolutional neural network, 203
counterfactual, 165
covariate balance, 168
example, 191
Cox proportional hazards model, see Cox regression
Cox regression, 166
cross-validation, 201
Cyclops, 201
data profiling, see White Rabbit
data quality, 241
DatabaseConnector, 93
creating a connection, 102
querying, 103
decision boundary, 200
decision tree, 202
deep learning, 203
descriptive statistics, see characterization
design considerations for network research, 298
direct effect estimation, 165
disease natural history, see characterization
domain
concept, 40
drug utilization, 142

ETL, see 抽出、変換、ロード（ETL）
implementations, 68
quality control, 68
単体テスト, 244
ETL design, see Rabbit-In-A-Hat
evidence quality, 237

FAIR, 15
feature analyses, 147
FeatureExtraction, 154
forum, 9

gradient boosting, 201

high correlation, 182
hyper-parameter, 201

incidence, 143
proportion, 143
rate, 143
index date, 141
instrumental variables, 167

k-nearest neighbors, 202
Kaplan-Meier plot, 193

LASSO, 201
logistic regression, 166, 201
logistical considerations for an OHDSI network study, 299

methods library, 84
minimum detectable relative risk (MDRR), 193
mission, 5
model viewer app, 224

naive bayes, 202
nesting cohort
  case-control design, 168
network study, 297
neural network, 203
no free lunch, 200

objectives, 5
OHDSI Methods Benchmark, 285
OHDSI SQL, see SqlRender
outcome cohort
  case-control design, 168
  case-crossover design, 169
  cohort method, 165
  SCCS design, 170
  self-controlled cohort design, 168

p-value calibration, 276
Pallas system, 39
patient-level prediction, 77
perceptron, 203
person-time, 143
phenotype library, 123
PheValuator, 258
Poisson regression, 166
population-level estimation, 76, 165
post-index time, 141
power, 193
preference score, 167
  example, 189
propensity model, 166
  example, 190
propensity score, 166
  matching, 166
  stratification, 166
  trimming, 178
  weighting, 166
python, 201, 202

quality improvement, see characterization
Query Library, 94
QueryLibrary, 108

R, 94
  installation, 87
Rabbit-In-A-Hat, 58
random forest, 201
randomized trial, 166
recurrent neural networks, 203
regularization, 201
reliable evidence, 238
ROC, 205
running network research, 299

safety surveillance, 165
self-controlled case series (SCCS) design, 170
self-controlled cohort design, 168
sensitivity analysis, 279
sklearn, 201
source code mapping, see Usagi
SQL, 93
SQL Query Library, see Query Library

SqlRender, 93
debugging, 101
parameterization, 94
supported functions, 96
translation, 96
standard concept, 42
Standard SQL Dialect, see SqlRender
standardized vocabularies, 37
download, 39
search, 39
stratified model, conditioned modelを参照178
strongly ignorable, 167
structured query language, see SQL
study diagnostics, 294
study feasibility
single study, 294
study-a-thon, 14
supervised learning(監督学習), 199
survival plot, see Kaplan-Meier plot
system requirements, 86

target cohort
case-control design, 168
case-crossover design, 169
cohort method, 165
SCCS design, 170
self-controlled cohort design, 168
tools deployment, 91
Amazon AWS, 91
Broadsea, 91
treatment utilization, see characterization
TRIPOD, 197

Usagi, 63

variable ratio matching, 166
variance, 201
vignette, 86
vision, 5
vocabulary, 40

White Rabbit, 54
workgroups, 9

xgboost, 201

アウトカムコホート, 197
アウトカムステータス, 199
インデックス日, 206
オープンサイエンス, 13
open standards, 14
オープンソース, 14
オープンディスコース, 15
オープンデータ, 14
キャリブレーション, 205
クラス, 199
コホート, 119
コホート定義, 119
コミュニティ
コミュニティコール, 9
コンセプト
祖先, 48
コードセット, 119
ソースデータ, see 生データ
ソースレコード検証, 257
ターゲットコホート, 197
データ品質
研究特有のチェック, 245
チェック, 242
バリデーション, 242
妥当性, 242
完全性, 242
検証, 242
準拠, 242

ネイティブデータ, see 生データ
ネガティブコントロール, 274
プロトコル, 292
ポジティブコントロール, 275
ラベル, 199
リスク期間, 197
リレーショナルデータモデル, see Common Data Model

予後アウトカム, 197
予測モデル, 197
予測モデルの評価, 203
交差検証, 204

偽陰性, 205
偽陽性, 205
共変量, 199
共通データモデル
　標準化テーブル, 25
　規約, 21
基準日, 199
実証的キャリブレーション, 276
実証的評価, 276

性能指標, 204
患者レベル予測, 197
感度, 204
方法の妥当性, 273
検証
　内部検証, 204
　外部検証, 204
　時間的検証, 204
　空間的検証, 204
機械学習, 197
欠損データ, 79, 200

特異度, 204
生データ, 53
真陰性, 205
真陽性, 205
研究ネットワーク, 297
研究パッケージ, 292
研究診断, 273
精度, 204

臨床意思決定, 197
血管性浮腫, 206
表現型, 119
観察研究の限界, 78
診断アウトカム, 197
識別力, 205

適合率-再現率曲線下の面積, 205
陽性予測値, 204