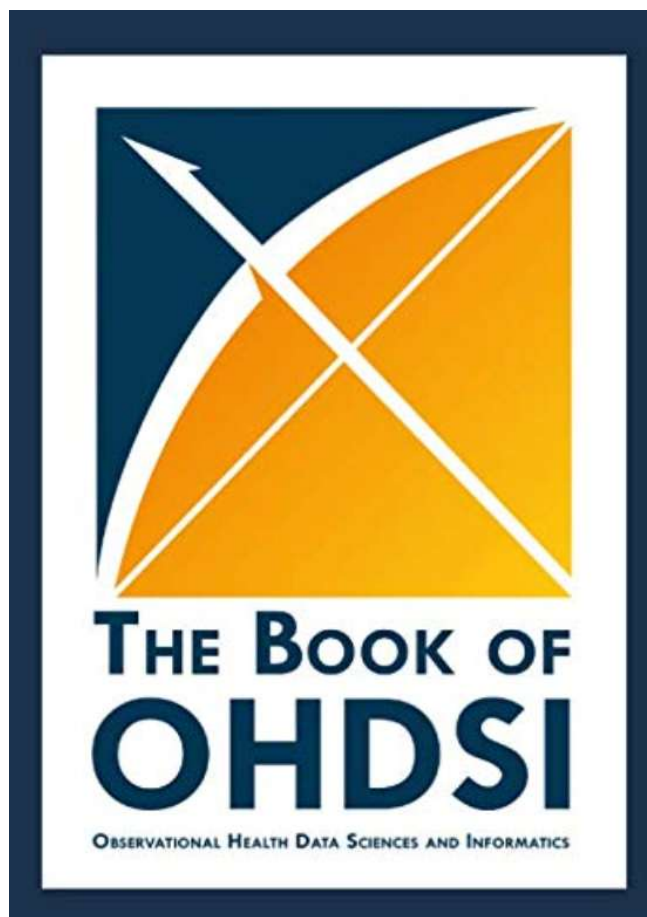




The Book of OHDSI 日本語版
「OHDSIの本」について
(松本先生・浅尾)



The Beginning of the Journey

2025-3-28(金) 17:30 - 19:00

OHDSI Japan イブニングカンファレンス
2025年3月 OHDSI
Japan



RWD R&D

京都大学 大学院 医学研究科
リアルワールドデータ研究開発講座

DEPARTMENT OF REAL WORLD DATA R&D,
GRADUATE SCHOOL OF MEDICINE, KYOTO UNIVERSITY

松本 繁巳



For the Journey



OHDSI

OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

オデッセイ ジャパン

ホーム

概要

OHDSI Japan資料

お問い合わせ

OHDSI (オデッセイ) は、共通データ形式を使った医療ビッグデータ分析を推進するオープンサイエンスコミュニティ (研究会) です。国際的・ボランティアベース・産学官病所属を問わないという特徴をもち、より良い医療を促進するエビデンスを共同して生みだすことを推進し、健康と病気の包括的な理解を観察研究から得られる世界を目指しています。OHDSIは米国で2014年にスタートし今では世界中に参加者がいます。オープンとは言っても、医療データは各参加組織のところで守られ、個人情報各参加組織の外にすることはありません。この方法により開始わずか5年で、OHDSIの国際連携ネットワークには重複を除外して推定6億人以上のデータが含まれるまで成長しました。欧州、中国、韓国に引き続き、2019年秋に日本にも地域コミュニティが活動しています。

※OHDSIは古代ギリシャ叙事詩Odysseyにちなんでおり同じ発音をします。Odysseyの日本語読みは慣用的に「オデッセイ」、英語では人/場面により異なりますが、多くはオデシーと発音されます (オにアクセント)。

■臨中ネット(国がん東病院)「OHDSI Tool Documents」

OHDSI の各種Toolの日本語資料が、国立がん研究センター東病院の青柳先生により臨中ネットの成果として整備されました。
<https://rwd-data-environment-in-hospital.github.io/Documents/Files/>

OHDSI Tool Documents

Usagi
WhiteRabbit & Rabbit-in-a-Hat
Atlas
HADES

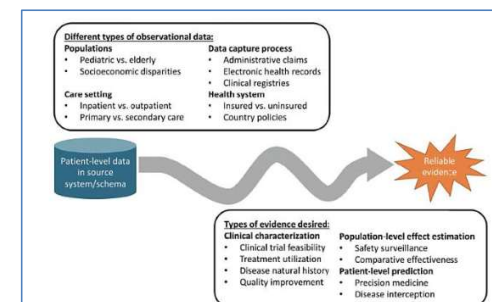
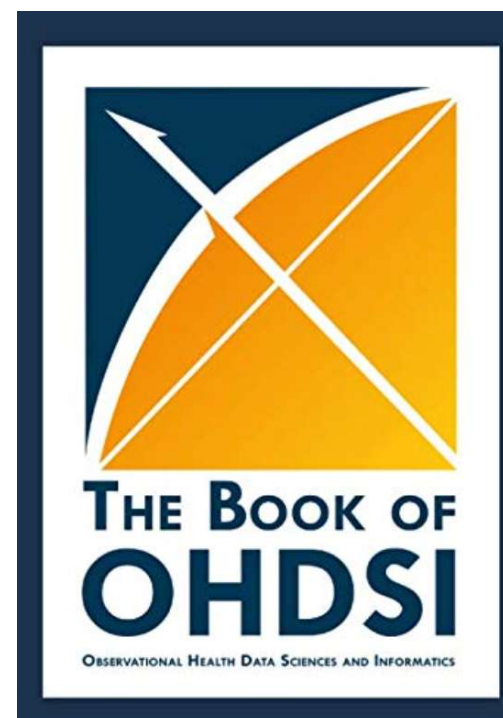


Figure 1.1: The journey from data to evidence

There are different types of observational databases which capture disparate patient-level data in source systems. These databases are as diverse as the healthcare system itself, reflecting different populations, care settings, and data capture processes. There are also different types of evidence that could be useful to inform decision-making, which can be classified by the analytic use cases of clinical characterization, population-level effect estimation, and patient-level prediction. Independent from the origin (source data) and desired destination (evidence), the challenge is further complicated by the breadth of clinical, scientific, and technical competencies that are required to undertake the journey. It requires a thorough understanding of health informatics, including its full provenance of the source data from the point-of-care interaction between a patient and provider through the administrative and clinical systems and into final repository, with an appreciation of the biases that can arise as part of the health policies and behavioral incentives associated with the data capture and curation processes. It requires mastery of epidemiologic principles and statistical methods to translate a clinical question into an observational study design properly suited to produce a relevant answer. It requires the technical ability to implement and execute computationally-efficient data science algorithms to datasets containing millions of patients with billions of clinical observations over years of longitudinal follow-up. It requires the clinical knowledge to synthesize what has been learned across an observational data network with evidence from other information sources, and to determine how this new knowledge should impact health policy and clinical practice. Accordingly, it is quite rare that any one individual would possess the requisite skills and resources to successfully trek from data to evidence alone. Instead, the journey often requires collaboration across multiple individuals and organizations to ensure

ペーパーバック
¥2,962 (30pt)
prime

470p
AMAZON



Marina Bay Sands
Dec 5-6th . 2024

On APAC in Singapore



Challenges in Conducting Federated Analysis in CyberOncology Project in Japan

Shigemitsu Matsumoto¹⁾, Kosuke Tanaka²⁾, Pei Liying³⁾, Masafumi Okada³⁾, Manabu Muto³⁾

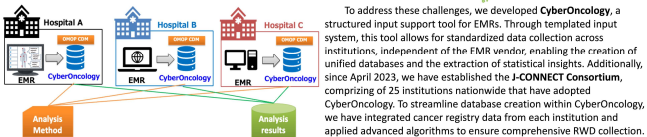
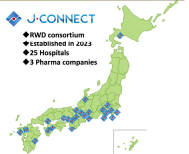
1. Department of Real World Data R & D, Graduate School of Medicine, Kyoto University
2. Department of Medical Oncology, Graduate School of Medicine, Kyoto University
3. Prime Research Institute for Medical RWD, Inc.



Background

There is a growing demand for the development of frameworks to generate Real World Evidence (RWE) using Real World Data (RWD)¹⁾. In recent years, in the field of precision medicine, particularly in oncology. Data on patient backgrounds and treatments, and genomic information, including patient-specific biomarkers, have become critical. Additionally, outcomes related to efficacy, safety, and prognosis are indispensable. However, in real-world database research in oncology, there is a strong need to enhance the quality and transparency of RWD sources²⁾.

In Japan, there are multiple electronic medical record (EMR) vendors, and customization is common in hospitals. This has posed significant challenges for collecting outcome data, particularly concerning efficacy and safety, which are crucial in oncology. Furthermore, the efficient collection and analysis of RWD from multi-institutional EMRs face various regulatory and ethical challenges, including compliance with the Personal Information Protection Act and ethical guidelines.



Method

Mapping strategy

Description	CyberOncology Table	Vocabulary used in CO	OMOP CDM v5.4	CDM vocabulary
Patient demographics	episode	person	person	person
Adverse events (begin - end dates)	reaction	observation_period	observation_period	observation_period
Last visit date	episode	visit_occurrence	visit_occurrence	visit_occurrence
Cancer diagnosis	cancer	condition_occurrence	condition_occurrence	condition_occurrence
Prescriptions / Injection	prescription, injection	YJ Code	drug_exposure	ATC
Laboratory data	laboratory	CTCAE	measurement	LOINC
Adverse events	reaction	CTCAE	observation	MedDRA
Biomarker	biomarker	(Text)	measurement	LOINC
Outcome	outcome		death	death

CyberOncology is a structured oncology database integrated with electronic medical records (EMRs), with the definitions of each table outlined in the Mapping Strategy. Additionally, the data structure is designed to be both compatible with the similar to the OMOP Common Data Model (OMOP CDM). The CDM vocabulary is defined as specified in the corresponding documentation.

The following is an example of a CTCAE transformation recorded in CyberOncology. The variables and vocabulary associated with a case of Grade 3 anemia (Hb 7.0 g/dL) diagnosed on February 15, 2021, and resolved on February 28, 2021, are shown in the table on the right.

Summary

We initiated a project to develop a federated analysis platform for the 25 member institutions of the J-CONNECT consortium by transforming CyberOncology, a structured tool integrated with electronic medical records (EMRs), into the OMOP Common Data Model (OMOP CDM). The platform is expected to become fully operational for analysis by April 2025.

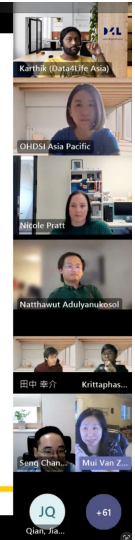
References

1. Concato J, Corrigán-Curay J. Real-World Evidence - Where Are We Now? *Nw Engl J Med* 2022;386:1680-1687
2. Ramsey SD, Omar-Thomas A, Wheeler SB. Real-World Database Studies in Oncology: A Call for Standards. *J Clin Oncol* 2024;42(9).



APAC Community-wide ETL Project Kick-off Meeting

August 1, 2024



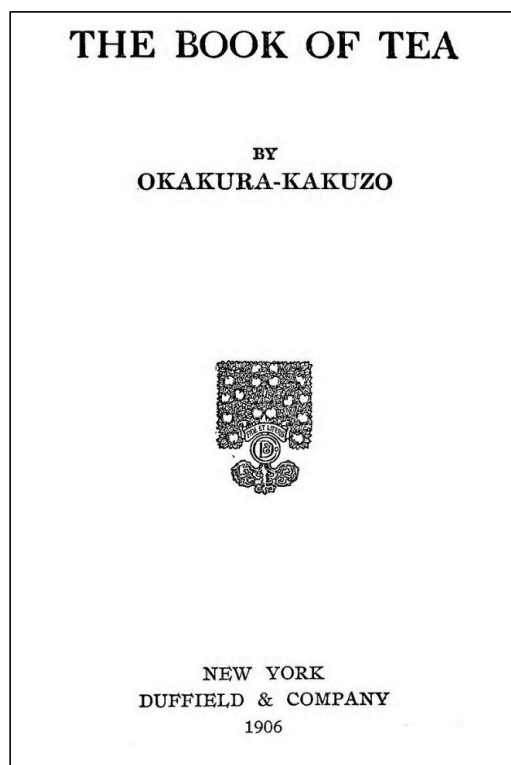
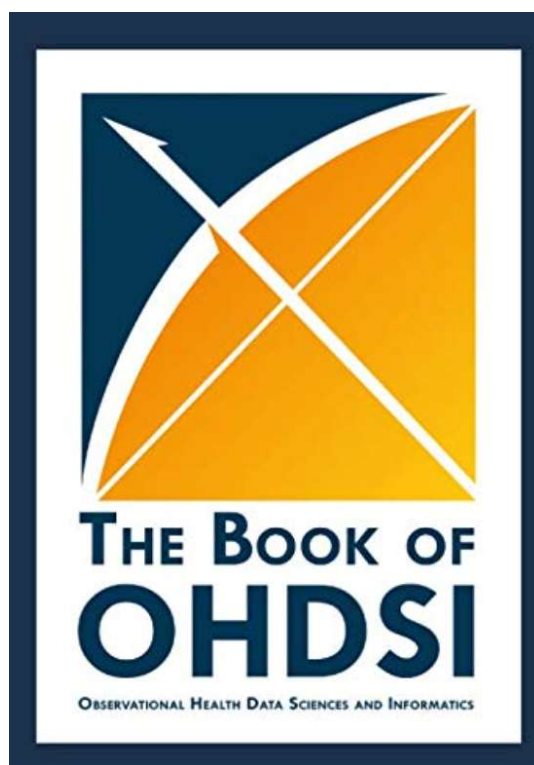
OHDSIの本

Goals of this Book

This book aims to be a central knowledge repository for OHDSI, and it focuses on describing the OHDSI community, OHDSI data standards, and OHDSI tools. It is intended both for both OHDSI newcomers and veterans alike, and aims to be practical, providing the necessary theory and subsequent instructions on how to do things. After reading this book you will understand what OHDSI is, and how you can join the journey. You will learn what the common data model and standard vocabularies are, and how they can be used to standardize

この本の目標

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、OHDSI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としており、必要な理論とそれに続く手順を提供する実用的な内容を目指しています。本書を読んだ後は、OHDSIが何であり、どのようにしてその旅に参加できるかを理解できます。 共通データモデルおよび標準ボキャブラリとは何か、また、それらが観察医療データベースを標準化のためにどのように使用されるかを学びます。これらのデータの主な使用例である特性評価、集団レベルの推定、および患者レベルの予測について学びます。これらの3つすべての活動をサポートするOHDSIのオープンソースツールとその使用方法についても読みます。データ品質、臨床的妥当性、ソフトウェアの妥当性、方法の妥当性に関する章は、生成されたエビデンスの質をどのように確立するかを説明します。最後に、分散された研究ネットワークでこれらの研究を実行するためのOHDSIツールの使用方法を学びます。



Join the Journey



Join the OHDSI forum
OHDSI Tip: Follow topics to receive emails when new posts are added

Introduce yourself!
Let the community know you're here by introducing yourself in the forum or at a community meeting

Join an OHDSI meeting
Sit in on our weekly community meetings



Join the OHDSI research network
By leading a study across the network

OR

By converting data to the OMOP Common Data Model

Join an working group
Or start your own work group!



Join the Journey
Help improve medical decision making today!

Provide feedback
Identify and evaluate ways to use real-world evidence to inform decision making



Atlas

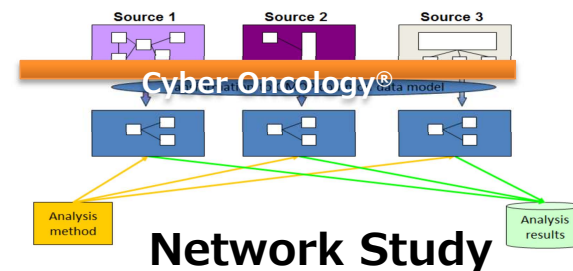


Figure 2.1: 旅に参加しよう - OHDSIのコラボレーターになるには



The Book of OHDSI 日本語版作成プロジェクト

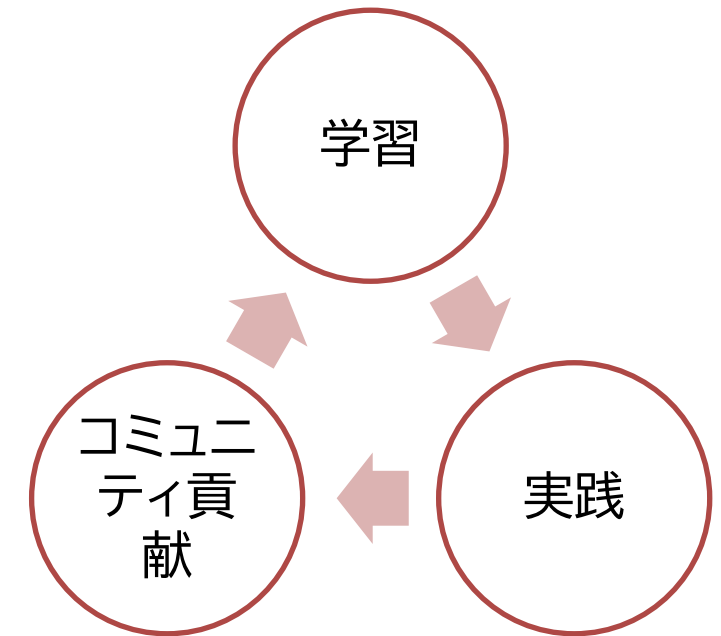
OHDSI Japan イブニング・カンファレンス
03/28/2025

浅尾 啓子(株式会社カッパ・メディカル)



The Book of OHDSI とは

- OHDSIの公式テキストブック
- OMOP CDMとOHDSIツールを用いた研究の実践的ガイド
- 初学者から上級者まで、幅広いレベルの読者に対応
- 50名以上のOHDSIコミュニティメンバーによる共著
- R Markdownで執筆され、GitHubで無償公開 (Creative Commons Zero)





The Book of OHDSI 日本語版作成プロジェクト

- OHDSI/OMOP CDMの理解促進を目的とした翻訳プロジェクト
- 原著(英語):約11.5万語、464頁

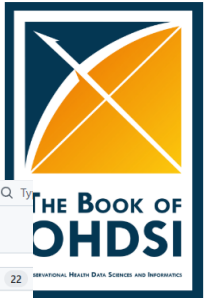
The Book of OHDSI

Observational Health Data Sciences and Informatics

2021-01-11

Preface

This is a book about the Observational Health Data Sciences and Informatics (OHDSI) collaborative. The OHDSI community wrote the book to serve as a central knowledge repository for all things OHDSI. The Book is a living document, community-maintained through open-source development tools, and evolves continuously. The online version, available for free at <http://book.ohdsi.org>, always represents the latest version. A physical copy of the book is available from Amazon at cost price.



The screenshot shows the GitHub repository page for 'TheBookOfOhdsi'. The repository is public and has 22 issues, 10 pull requests, and 652 commits. The commit history is as follows:

Commit	Message	Time
Admin_mschuemmi and Admin_mschuemmi	Fix render errors, including corrupt png header, R...	4ef7abb · 6 months ago
extras	Fix render errors, including corrupt png header. Rendering t...	6 months ago
images	Fix render errors, including corrupt png header. Rendering t...	6 months ago
.gitignore	Adding HTML version of book description. Changing figure L...	6 years ago
.nojekyll	starting Travis CI	7 years ago
.travis.yml	update to remotes::install_github()	5 years ago
Characterization.Rmd	minor spelling edits -part 2	5 years ago
ClinicalValidity.Rmd	Remove trailing whitespace	5 years ago
CohortDefinitions.Rmd	Remove trailing whitespace	5 years ago
Cohorts.Rmd	minor spelling edits -part 3	5 years ago
CommonDataModel.Rmd	minor spelling edits -part 3	5 years ago
Contributors.csv	dedup and correct Chan's name on the contributor list	6 years ago
DESCRIPTION	initial import of bookdown-demo	7 years ago



日本語版作成時に検討した翻訳アプローチ

アプローチ	利点(Pro)	課題(Con)
PDF/HTMLベースの翻訳 (韓国語・中国語版方式)	<ul style="list-style-type: none">• 比較的手軽に開始可能• 表示形式に近い形で訳せる	<ul style="list-style-type: none">• 図・数式・コードの構造が壊れやすい• 原著の構成維持が困難
GitHub + GPT-4o (フランス語版方式)	<ul style="list-style-type: none">• ソースファイル(Rmd)に忠実• 翻訳と構造調整を並行処理	<ul style="list-style-type: none">• ChatGPTの出力精度に依存• 手作業の確認・修正が必須
GitLocalize使用	<ul style="list-style-type: none">• GitHubベースで翻訳管理が容易• アップデートの対応がしやすい	<ul style="list-style-type: none">• R Markdown対応に制限あり• 技術的セットアップが必要

Martijnさん、
ありがとう！



GitLocalizeとは

- GitHubリポジトリのための継続的ローカライゼーションツール
- 自動的に元の文と翻訳をリンクするため、原文の更新により影響を受ける範囲がわかる。
- 頻繁にアップデートする場合には有用。

翻訳文の差分管理

87%

English	locale/en/docs/index.md	Japanese	locale/ja/docs/index.md
title	Docs	title	ドキュメント
layout	docs.hbs	layout	docs.hbs
labels	Its LTS	labels	Its LTS

About Docs

There are several types of documentation available on this website:

- API reference documentation
- ES6 features
- Frequently asked questions
- Guides

API Reference Documentation

The API reference documentation provides detailed information about a function or object in Node.js. This documentation indicates what arguments a method accepts, the return value of that method, and what errors may be related to that method. It also indicates which methods are available for different versions of Node.js.

This documentation describes the built-in modules provided by Node.js. It does not document modules provided by the community.

ドキュメント

ここには3種類のドキュメントがあります:

- API リファレンス
- ES6 の機能
- よくある質問
- Guides

API リファレンス

API リファレンスでは Node.js の関数やオブジェクトの詳細情報を提供しています。このドキュメントでは、メソッドで何の引数を取るのか、そのメソッドの返り値とそのメソッドに関連するエラーについても説明されています。また、異なる Node.js のバージョンでどのメソッドが使えるかについても説明されています。

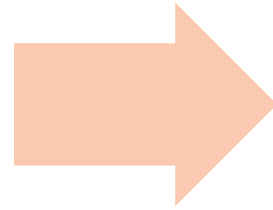
このドキュメントでは Node.js によって提供された組み込みのモジュールについて説明しています。コアに組み込まれていないモジュールは含みませ



翻訳のプロセス

ステップ1: 翻訳・ポストエディット

- GitHub + GPT-4o
- 用語リスト作成で一貫性確保
- Rマークダウン形式に対応したポスト・エディット
- 日本語の読みやすさを考慮したポスト・エディット
- (図については翻訳なし)



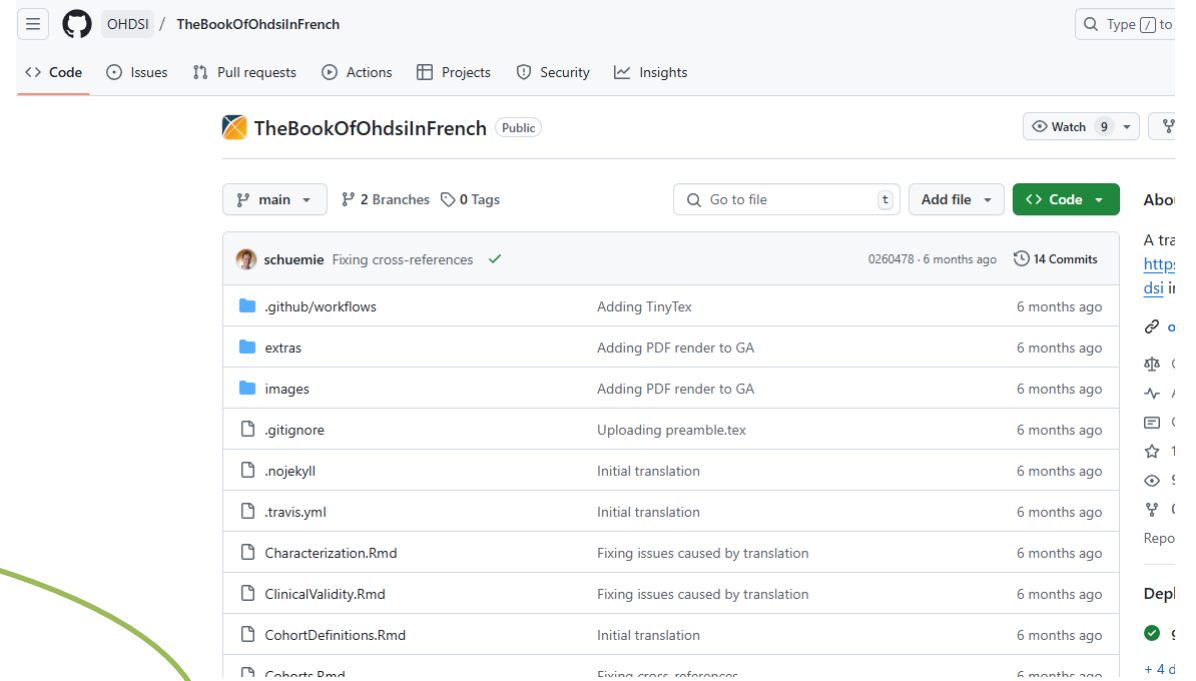
ステップ2: 公開と修正対応

- HTML、PDFで公開
- GitHub他で修正受付体制



GitHub + GPT-4o

- [TheBookOfOhdsiInFrench/extras/TranslateBook.R at main · OHDSI/TheBookOfOhdsiInFrench](https://github.com/OHDSI/TheBookOfOhdsiInFrench/tree/main/extras/TranslateBook.R)
- Martijn Schuemie氏がThe Book of OHDSIをRのプログラミング部分を残しつつ、GPT-4oで作成



M「ちょっとやってみましょうか？」

(およそ2時間後)

M「できましたよ、日本語がどのくらいいいかわからないのですが。。。」



用語リスト作成で一貫性確保

- OHDSI用語の訳語をどうするか？
 - キーワードや頻繁に使われる用語リストアップ(200語程度)
 - 先生方のレビュー・ご意見
- 用語も文脈で訳が変わる
- 原文の中でも用法の揺れがある

- 難しい訳語の例

Observation	観察、オブザベーション(ドメインとして)
Procedure	処置(プロシージャー)
Visit	受診期間(ビジット)
Encounter	受診
The Book of OHDSI	OHDSIの本



Rマークダウン形式に対応したポスト・エディット

HTMLでの見え方

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 (Chapter (ref?)(Cohorts) で説明)。付録には、ターゲットコホート (Appendix (ref?)(AceInhibitorsMono), (ref?)(ThiazidesMono)) およびアウトカムコホート (Appendix (ref?)(Angioedema), (ref?)(Ami), (ref?)(ARBUse)) の完全な定義が提供されています。



直し方は、原文の様式に合わせる。



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 (Chapter 8 で説明)。付録には、ターゲットコホート (Appendix 19.2, 19.5) およびアウトカムコホート (Appendix 19.4, 19.3, 19.9) の完全な定義が提供されています。

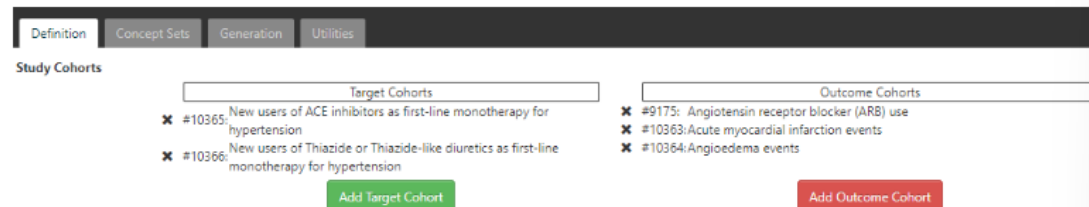


図 9.22: ターゲットおよびアウトカム定義の発生率。

Rマークダウン形式

```
400
401 - ### 設計
402
403 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 (Chapter @ref(Cohorts)
    で説明)。付録には、ターゲットコホート (Appendix
    @ref(AceInhibitorsMono), @ref(ThiazidesMono))
    およびアウトカムコホート (Appendix @ref(Angioedema),
    @ref(Ami), @ref(ARBUse)) の完全な定義が提供されています。
404
405 - ```{r atlasIncidenceCohortSelection,
    fig.cap='ターゲットおよびアウトカム定義の発生率。',echo=FALSE
```



```
: [C:/images/populationlevelEstimation/save.png]
400
401 - ### 設計
402
403 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 (Chapter \@ref(Cohorts)
    で説明)。付録には、ターゲットコホート (Appendix
    \@ref(AceInhibitorsMono), \@ref(ThiazidesMono))
    およびアウトカムコホート (Appendix \@ref(Angioedema),
    \@ref(Ami), \@ref(ARBUse)) の完全な定義が提供されています。
404
405 - ```{r atlasIncidenceCohortSelection.
```

日本語の読みやすさを考慮したポスト・エディット

- 内容がテクニカルに難しい。
 - 特に第III部後半、第IV部(PheValuator、ネガティブコントロールと効果推定値のキャリブレーション、ポジティブコントロール等)。
 - 内容が理解できて初めて、読みやすい訳になる。
- 訳しにくい用語がある。
- ウェブアプリの説明については、英語(日本語)として、アプリの操作としてフォローできるように工夫。

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、その下位層すべてを含め、これらの成分を含むすべての薬剤を含めています。「Included concepts (包含されるコンセプト)」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とができ、「Included Source Codes (包含されるソースコード)」をクリックすると、様々なコーディングシステムに含まれるすべてのソースコードを探索することができます。

Concept Id	Concept Code	Concept Name	Domain	Standard Concept Caption	Exclude	Descendants	Mapped
1335471	18867	benazepril	Drug	Standard	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
1340128	1998	Captopril	Drug	Standard	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
19050216	21102	Cilazapril	Drug	Standard	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
1341927	3827	Enalapril	Drug	Standard	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
1342001	3829	Enalaprilat	Drug	Standard	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
1363749	50166	Fosinopril	Drug	Standard	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
19122327	60245	imidapril	Drug	Standard	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
1308216	29046	Lisinopril	Drug	Standard	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>



HTML、PDFで2025年4月1日公開

- <https://ohdsi.github.io/TheBookOfOhdsiInJapanese/>

- <https://ohdsi.github.io/TheBookOfOhdsiInJapanese/TheBookOfOhdsiInJapanese.pdf>

OHDSIの本

- 序章
 - この本の目標
 - 本書の構成
 - 貢献者
 - ソフトウェアのバージョン
 - ライセンス
 - 本書が作成された方法
- I OHDSI コミュニティ
 - 1 OHDSI コミュニティ
 - 2 どこから始めようか
 - 3 オープンサイエンス
- II 共通データモデル
 - 4 共通データモデル
 - 5 標準化ボキャブラリ
- III データ解析
 - 6 ETL (抽出-変換-読込)
 - 7 データ解析の使用例
- 8 OHDSI 分析ツール

OHDSIの本

Observational Health Data Sciences and Informatics (OHDSI)

2025-03-15

序章

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、OHDSIに関するすべての知識の中心的なリポジトリとして役立つことを目指しています。この本はオープンソース開発ツールを通じてコミュニティにより維持される生きた文書であり、絶えず進化しています。オンライン版は無料で <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：英語版）はAmazonで原価価格で入手可能です。

この本の目標

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、OHDSI データ標準、および

OHDSI の本

Observational Health Data Sciences and Informatics (OHDSI)

2025-03-15



修正・改訂について： ご協力をお願い

- 公開後3か月間、集中的に改訂・修正します。
- 読んで気づいた修正などは、info@kappa-medical.com にご連絡ください。
- 継続的な改訂についてはGitHubのReadMeに記載しています。

The screenshot shows a web browser displaying the GitHub repository page for 'The Book of OHDSI in Japanese'. The address bar shows the URL: <https://github.com/OHDSI/TheBookOfOhdsilnJapanese>. The repository name is 'userDefinedRmd.xml' and it was created 'Initial translation by GPT-4o' '2 months ago'. The page content includes a README section with the following text:

README CC0-1.0 license

The Book of OHDSIの日本語版です。

原文は下記にあります。 <https://github.com/OHDSI/TheBookOfOhdsi>

この翻訳は、The Book OHDSI原文を下記のプログラムにより機械翻訳した後、ポストエディットし作成されました。 <https://github.com/OHDSI/TheBookOfOhdsilnFrench/blob/main/extras/TranslateBook>

日本語版の内容に修正案がある方は、下記の手順に従ってください。

1. このGitHubのレポジトリを自分のアカウントにフォークしてください。
2. RおよびRStudioを使って、Rmd (Rマークダウン) ファイルを編集してください。
3. 変更をコミットしてください。
4. ご自分のフォークから、Contributeしてください。
5. 管理者が確認したのち、Mergeすることによって、修正が反映されます。

Rマークダウンの用法は、原文のレポジトリから、Cheat Sheetや原文のRmd等を参考にしてください。特に、リンク等の書式に気をつけるようお願いいたします。



ポスト・エディットする際に注意が必要だった点

- Gitの使い方
 - コミットする際には、Conflictを解消してからMergeする。
- Rマークダウンの最終形を見ながら編集する。
 - R Studio上でKnitまたはVisualizeを使う。
 - HTMLに適時反映されるものをよく見る。
- 本の構造調整
 - 章やセクション番号、図表番号、参考文献などが適切に反映されるように編集する。
- PDFはさらに調整が必要
 - 日本語フォントを使う際、設定を変更する必要があった。
 - 索引はPDFのみで表出される。
- 機械翻訳の落とし穴
 - 大きな訳抜けが起こる！（段落、表半分など、特に一文が長い場合に注意）



今後の展望:The Book of OHDSIの多言語化

- OHDSIグローバルメンバーで、多言語化に関する検討がある。
- グローバルとの協業・協調
 - フランス語化の経験の応用と協力 (Martijnさん)
 - APACの2025年の目標の1つとしてのThe Bookの多言語化
 - 次のバージョンへの日本からの貢献
- 多言語化を進めるための要件
 - 医療情報・疫学の知識
 - 言語
 - RマークダウンやGitHubの知識
- 原著においても多言語化しやすいものにする工夫
 - わかりやすい、比較的短い原文
 - 訳しやすい構成(アプリの説明についてはアプリ上の指示を図にするなど)