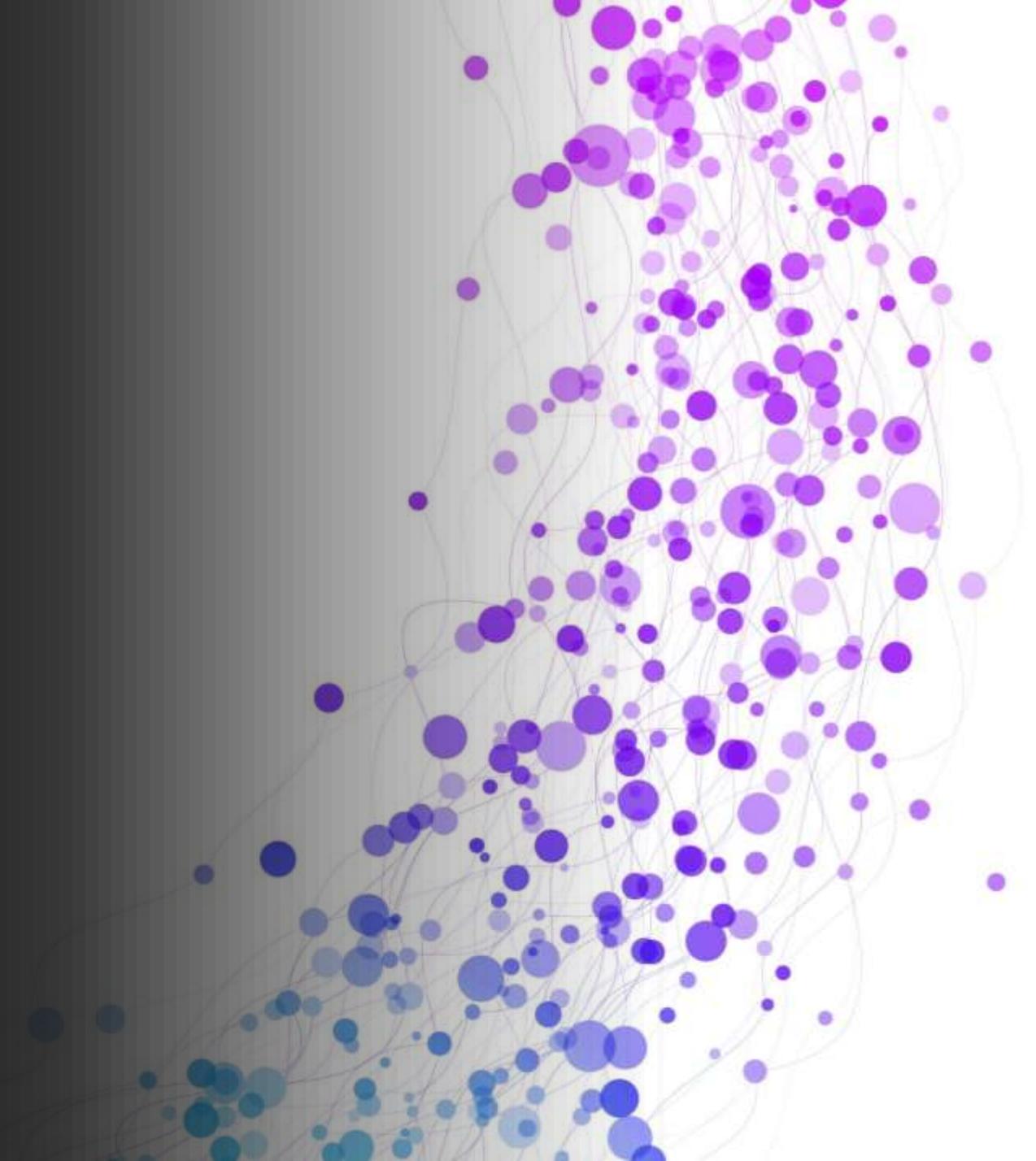




## 電子カルテデータから OMOP CDMへの変換と 分析

---

- @OHDSI one-day event in Tokyo  
2024-04-17
  - 国立がん研究センター東病院
  - 医療情報部 兼)臨床研究支援  
部門 臨床研究推進部
  - 青柳吉博
- 

# アジェンダ

1. はじめに

2. 方法

3. 結果

5. 考察

# アジェンダ

1. はじめに

2. 方法

3. 結果

5. 考察

# はじめに

- OMOP CDMは観察研究のための共通データモデル
- 電子カルテデータをOMOP CDMに変換することで、OHDSIが提供する分析ツールや手法を用いることができる
  - 世界中の医療データを統合し、観察研究を促進する有力なプラットフォーム
- 本邦において電子カルテデータをOMOP CDMに変換した事例はまだ報告が無く、その実現可能性については十分検証されていない
- 技術的な検証を目的としてOMOP利用環境を構築し、OMOP CDMに変換した電子カルテデータがOHDSIの分析ツール（ATLAS）で利用可能かを評価した

# OMOP CDM関連の概要

## OMOP CDMはOHDSIによって管理される標準データモデル

- 診療情報データを統合的に扱うことができる
- 国際的な観察研究にて世界規模でのエビデンスを利用できる可能性がある
- 電子カルテデータだけではなく様々なデータを変換することも想定されている

## OHDSIは様々なツールを公開

- 用語の整理、データ変換、データの表示および分析などに利用できるツールが公開されている。

## 分析に利用できる用語が決められている

- OHDSI的には標準用語でのみ分析を行うことができる。
- 非標準用語から標準用語へのマッピングはボランティアや研究として行われている場合がある。
- 日本のコード群は現状これらの管理群には含まれていない。

# 電子カルテの利活用と国際化

国内は電子カルテを利用した研究が行われている。

一方で治験、臨床研究はグローバル化がすすでおり、今後の研究促進の一端として国際化は必要かなと考えている。

国内で行われている観察研究も国際化されることでよりインパクトのある研究が実施できるかもしれない。

一方で、国内で利用されている電子カルテには標準コードが付与されていない、どのような情報がOMOPに変換できるかがわからない、などの課題が想定されるものの、実データを通じた考察は行われていないため、今回のプロジェクトを通じて、課題を考察する。

# アジェンダ

1. はじめに

2. 方法

3. 結果

5. 考察

# 利用した環境・ソフトウェア

- OMOP CDM ver. 5.3
  - Windows server 2019
  - Apache Maven ver. 3.8.3
  - Apache Tomcat ver. 8.5.72
  - Oracle Java JDK 8.0
  - Postgres SQL ver.10
  - R ver.4.1.1
  - R studio 2021.09.0 Build 351
  - Rtools 4.0.1
  - ATLAS ver.2.10.0
  - Data Quality Dashboard ver.2.6.0
- 環境構築はOHDSI Japanの「OHDSI Tool Documents」を参考に行った。また、Postgres SQLやJavaは別途インストールした。その他、不足する手順がある場合にはOHDSIウェブサイトを参考にして設定等を行った。

# 電子カルテからATLASにデータが登録されるまでの流れ

## Step1 (変換プログラムの作成)

- 電子カルテからデータを抽出
- ドメインごとに電子カルテで利用される用語を整理
- 標準用語とのマッピング表の作成
- Source to concept mapの作成
- 変換プログラムの作成

## Step2 (臨床データ変換)

- 電子カルテより変換対象の症例を抽出 (テキストファイルが生成)
- 変換プログラムの実行 (テキストファイル→Source Valueテーブルへの登録、OMOPへの変換)

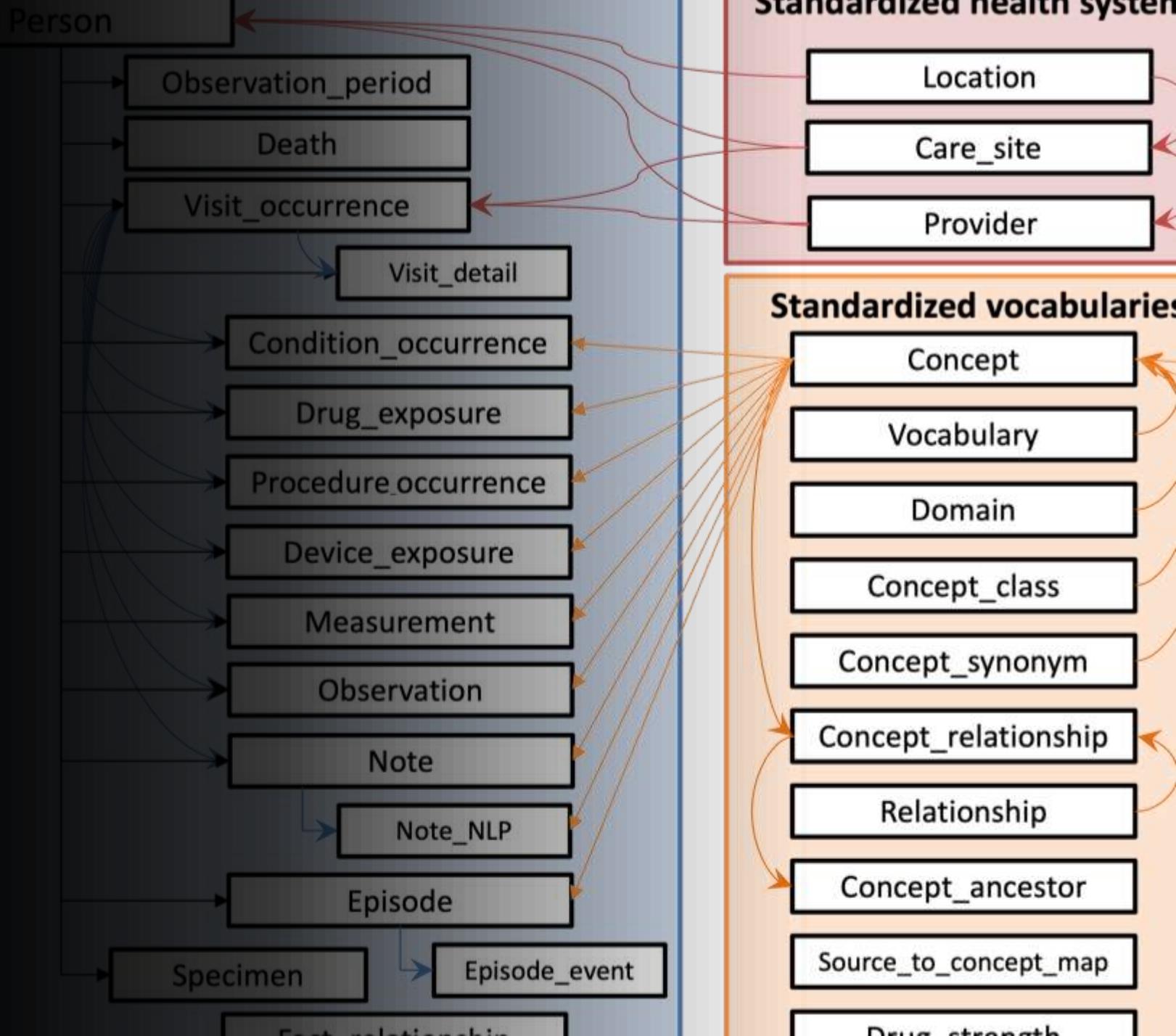
# 用語の対応表作成

- OHDSI標準用語と院内のローカルコードとの対応表を作成
  - OHDSI標準用語はATHENAより入手した。
- source to concept mapを作成
  - 利用頻度の少ない用語は作成の対象外とした。
  - ICD10コードはATHENAに対応表が含まれているため、source to concept mapには含めなかった。

source_code	source_vocabulary_id	target_concept_id	concept_name	concept_code	domain_id	standard_concept	vocabulary_id	valid_start_date	valid_end_date
0000001000 T P	JAPAN_NCCE_Measurement	4119650	Total protein measurement	304383000	Measurement	S	SNOMED	20020131	20991231
0000001530 沈渣目視	JAPAN_NCCE_Measurement	3022113	Urinalysis microscopic panel - Urine sediment	24365-9	Measurement	S	LOINC	20150628	20991231
0000001531 赤血球	JAPAN_NCCE_Measurement	4030871	Red blood cell count	14089001	Measurement	S	SNOMED	20020131	20991231
0000001532 白血球	JAPAN_NCCE_Measurement	4298431	White blood cell count	767002	Measurement	S	SNOMED	20020131	20991231
0000001533 扁平上皮	JAPAN_NCCE_Measurement	3008325	Epithelial cells.squamous [Presence] in Urine sediment by Light microscopy	12258-0	Measurement	S	LOINC	19700101	20991231
0000001534 尿路上皮	JAPAN_NCCE_Measurement	36303790	Epithelial cells [Presence] in Urine	87830-6	Measurement	S	LOINC	20180615	20991231
0000001535 尿管管上	JAPAN_NCCE_Measurement	1176420	Epithelial cells.renal [Presence] in Urine sediment	90044-9	Measurement	S	LOINC	20181214	20991231
0000001536 細菌	JAPAN_NCCE_Measurement	36304419	Bacteria [Presence] in Urine	87829-8	Measurement	S	LOINC	20180615	20991231
0000001541 硝子円柱	JAPAN_NCCE_Measurement	3033308	Hyaline casts [Presence] in Urine sediment by Light microscopy	25162-9	Measurement	S	LOINC	19700101	20991231
0000001542 上皮円柱	JAPAN_NCCE_Measurement	3038106	Epithelial casts [Presence] in Urine sediment by Light microscopy	25157-9	Measurement	S	LOINC	19700101	20991231
0000001543 顆粒円柱	JAPAN_NCCE_Measurement	3001873	Granular casts [Presence] in Urine sediment by Light microscopy	25160-3	Measurement	S	LOINC	19700101	20991231
0000001544 蠟様円柱	JAPAN_NCCE_Measurement	3030603	Waxy casts [Presence] in Urine sediment by Light microscopy	33862-4	Measurement	S	LOINC	19700101	20991231
0000001545 脂肪円柱	JAPAN_NCCE_Measurement	3022090	Fatty casts [Presence] in Urine sediment by Light microscopy	25159-5	Measurement	S	LOINC	19700101	20991231
0000001546 赤血球円	JAPAN_NCCE_Measurement	3046326	RBC casts [Presence] in Urine sediment by Light microscopy	33804-6	Measurement	S	LOINC	19700101	20991231

# データ変換

- 対象患者
  - 2015年1月1日～2022年12月31日の期間に乳がんを診断を受けた患者
  - 研究責任者が適切ではないと判断した患者を除外
  - 8388症例
- データのダウンロード
  - 患者プロフィール、病名、処方・注射オーダ、検体検査結果、入院中バイタルサイン、既往歴、手術歴、手術記録等
  - DWHよりダウンロード
  - カルテIDや氏名等を削除
- マッピングプログラムによりデータ変換



# アジェンダ

1. はじめに

2. 方法

3. 結果

5. 考察

# ソフトウェアインストール

OMOP CDMに変換するためのツール類およびATLASインストール可能であった。

- JavaおよびPostgres SQLは、動作のための設定や検証が必要であった。

Data Quality Dashboardはオンラインインストールが可能であったが、オフライン領域ではインストールできなかった。

# 用語の対応表作成

ドメイン	カウント
Condition	3811
Condition Status	3
Device	1
Drug	1156
Gender	2
Meas Value	64
Meas Value Operator	10
Measurement	1024
Observation	81
Procedure	239
Provider	9
Route	37
Spec Anatomic Site	1
Unit	100
Visit	3
その他	40
総計	6581

ドメインごとのマッピング数

OHDSI標準用語	カウント
ABMS	3
Cancer Modifier	1
Condition Status	3
dm+d	1
Gender	2
HES Specialty	2
LOINC	207
NUCC	3
OMOP Extension	1
OMOP Genomic	30
Provider	1
RxNorm	479
RxNorm Extension	676
SNOMED	5029
UCUM	100
Visit	3
その他	40
総計	6581

OHDSI標準用語の利用数

# 用語の対応表作成

---

今回の研究により、6581の用語に対して、OHDSI標準用語へのマッピングが行われた。

---

病名や薬剤、検体検査等が多くマッピングされた。

---

SNOMEDやLOINCがよく利用された。

---

薬剤はRxNormだけではなく、RxNorm Extensionも多く利用された。

---

単位についてはUCUMが利用された。

---

病名はSNOMEDが利用された。

---

検体検査は、SNOMEDが多くLOINCは少なかった。

---

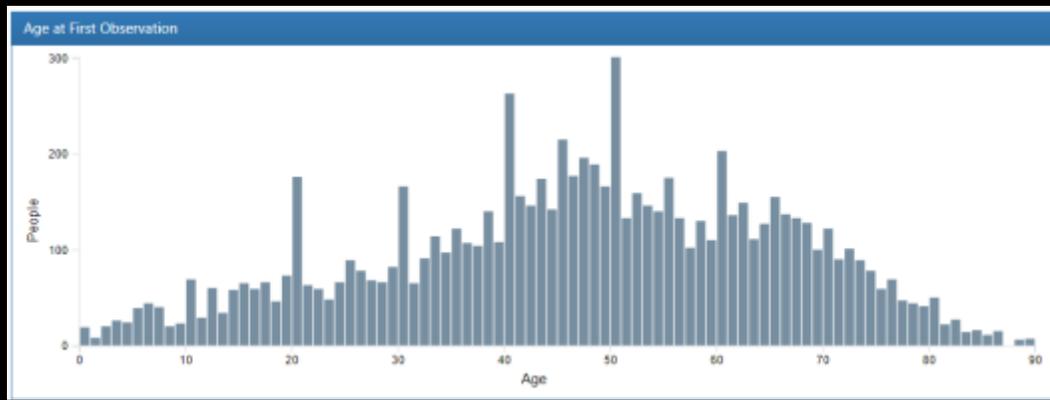
ゲノム情報はOMOP Genomicが利用された。

---

マッピング不可、マッピングの誤り、などはまだ散見された。

---

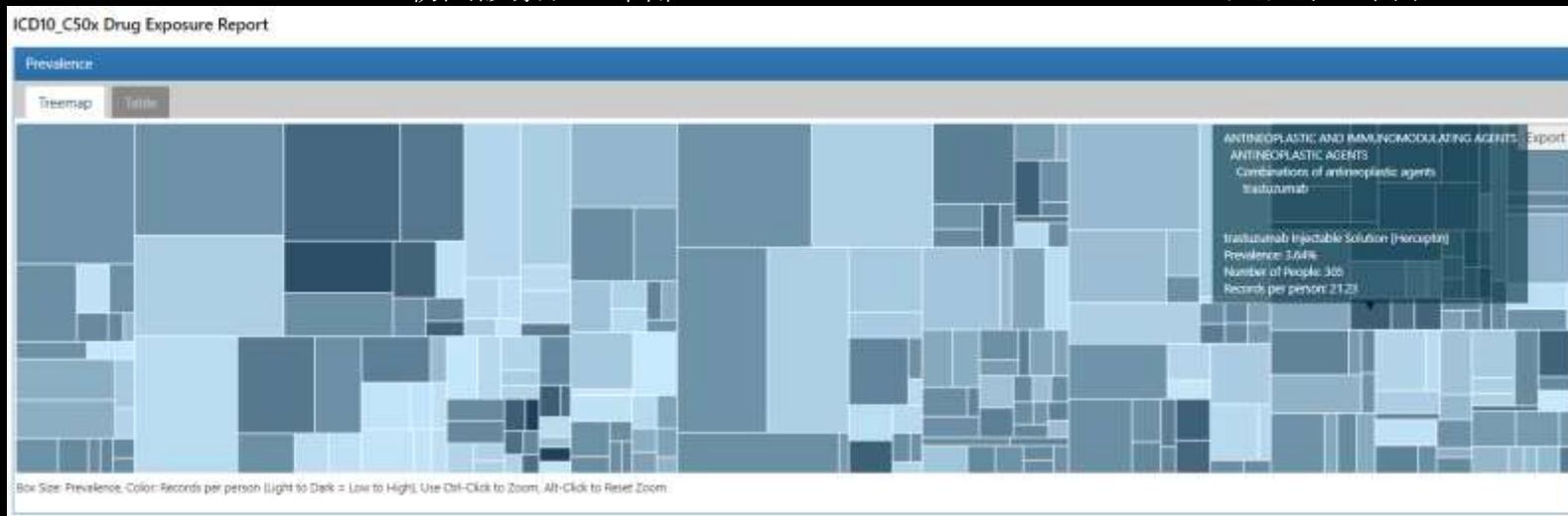
ATLASによる  
症例データの  
の表示



初回診察日の年齢

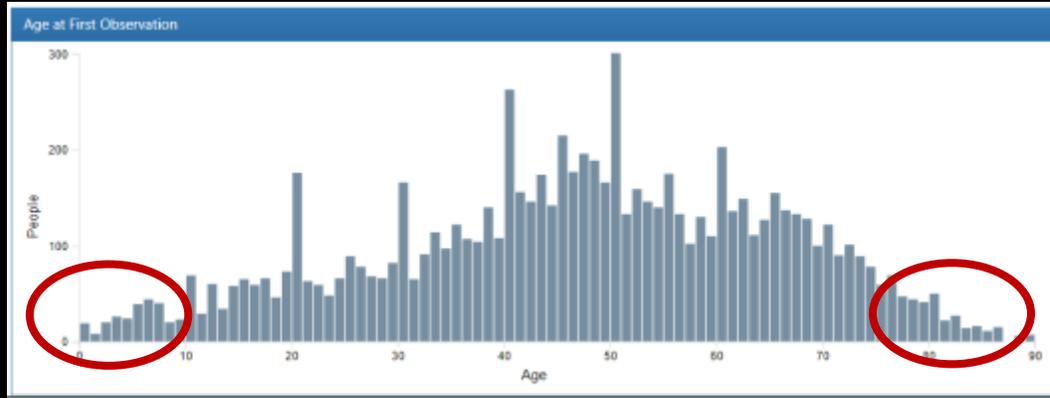


性別毎の年齢

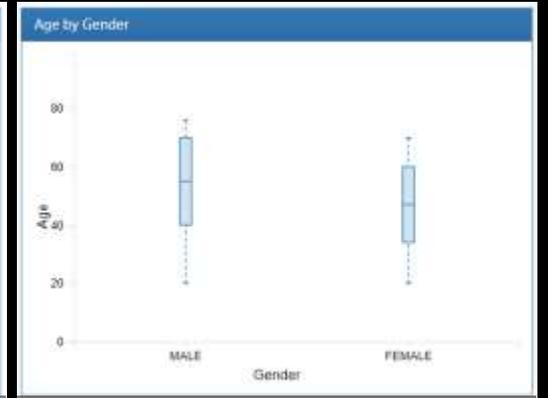


薬剤オーダ

ATLASによる  
症例データ  
の表示



初回診察日の年齢



性別毎の年齢



薬剤オーダ

# データ変換

---

研究対象8388例すべての症例に変換プログラムを適応できた。また、生成されたデータはATLASへ登録・表示することができた。

---

人種データ等OMOP CDMでは必須だが、電子カルテからは取得できないデータが存在した。

---

薬剤オーダは処方情報はオーダ情報であり、実施情報ではなかった。一方、注射情報は実施であった。

---

抗がん剤の規格選択が複雑な薬剤は変換工夫が必要だった。

---

マッピングの誤りなどの理由で、適切に変換できない例もあった。

# アジェンダ

1. はじめに

2. 方法

3. 結果

5. 考察

# 考察

OHDSIで利用されているソフトウェアのインストールが完了

- 用語の対応表作成、データ変換、ATLASによる症例データの表示までの技術的な検証が完了した。

電子カルテデータをOMOP CDMへ変換し、ATLASにおいて利用可能であることが明らかになった。

- 院内の用語もOHDSIで利用される用語に変換可能であることが明らかとなった。
- 既にATHENAに登録されており、標準用語にマッピングが行われていると、症例データの変換作業も効率的になると考えられた。  
(例：病名コード)

技術的な観点での検証が完了し、今後対象症例を増やすことも容易であるため、様々な観察研究に活用できることが示唆された。

今後は用語マッピング結果や変換結果の検証を進めてデータのクオリティを向上させていく。

ネットワーク研究の実施についても検討していく。