OMOP CDMを活用した医療情報データベースにおける疾患を特定する定義の妥当性向上

OHDSI Japan Evening Conference 2026-09-29

粕谷 奉幸 Tomoyuki Kasuya

0. 今回の発表の経緯

- 発表者：粕谷奉幸
- 所属：日本セルヴィエ株式会社 Medical & Patient Affairs
- 発表内容は2024年4月から2026年3月まで在学していた国際医療福祉大学大学院・医学研究科・公衆衛生学専攻での課題研究です
- 本研究は個人として在学し実施した研究で、当時及び現所属企業と本研究に資金面等の関連はありません

1. 背景と目的

- 既存の医療情報データベースを研究に二次利用する場合、目的とする解析が可能かどうかは一次利用時のデータ記録状況に依存するため、解析の妥当性を担保する工夫が必要である。

- 二次利用においては正しく定義された疾患を集計できるように、複数の条件設定で定義することとなる。具体的には、データベース内に存在する複数の情報（特に傷病名コード、薬剤コード、検査コード）を組み合わせて目的とする疾患を特定するためのアルゴリズムを作成する。

1. 背景と目的

- 作成した疾患定義の妥当性を評価するためには、多くの場合医療従事者が研究対象の病院に訪問しカルテを目視で確認（カルテレビュー）し、Gold Standardを定義する。この方法では、レビュー担当者の判断が信頼ものであることを保証する要件や時間などの面で制約が大きい。

- OMOP CDMというデータ標準規格で様々なデータソースを一元的に取り扱い、PheValuatorというツールで機械学習によりGold Standardの定義とアルゴリズムの該当度の算出を行う方法が提案されている[1]。

1.J.N. Swerdel, G. Hripcsak, P.B. Ryan, Phevaluator: Development and evaluation of a phenotype algorithm evaluator, J. Biomed. Inform. 97 (2019) 103258

1. 背景と目的

- このツールを用いた先行研究では、循環器領域の疾患に対し海外の保険請求データベース等での妥当性指標を、別のカルテレビューによる結果と比較しており、性能面での整合性が示された[1]。

  例：Acute Myocardial Infarctionにおいて傷病名のみで陽性的中度 0.431 - 0.598

  傷病名＋coronary revascularizationで陽性的中度 0.667 - 0.967

- **本研究では、OMOP CDMとPheValuatorを用いることで、本邦の病院データについてカルテレビューを行わず効率的に疾患定義の評価が可能か、その妥当性を検証する。**

1.J.N. Swerdel, G. Hripcsak, P.B. Ryan, Phevaluator: Development and evaluation of a phenotype algorithm evaluator, J. Biomed. Inform. 97 (2019) 103258

2. 方法

研究デザイン

記述研究（バリデーション研究）

対象集団

2019年1月 – 2024年12月に国際医療福祉大学病院（栃木県那須塩原市）に外来受診・入院した20歳以上の患者：約9万3千名

選択基準

年齢が20歳以上の患者

除外基準

オプトアウトにより診療記録等の研究への二次利用を許可しない申し出があった患者

国際医療福祉大学研究倫理審査委員会より承認

2. 方法

対象とするアウトカム

Acute Myocardial Infarction
Heart Failure
Stroke
Atrial Fibrillation
Angina Pectoris

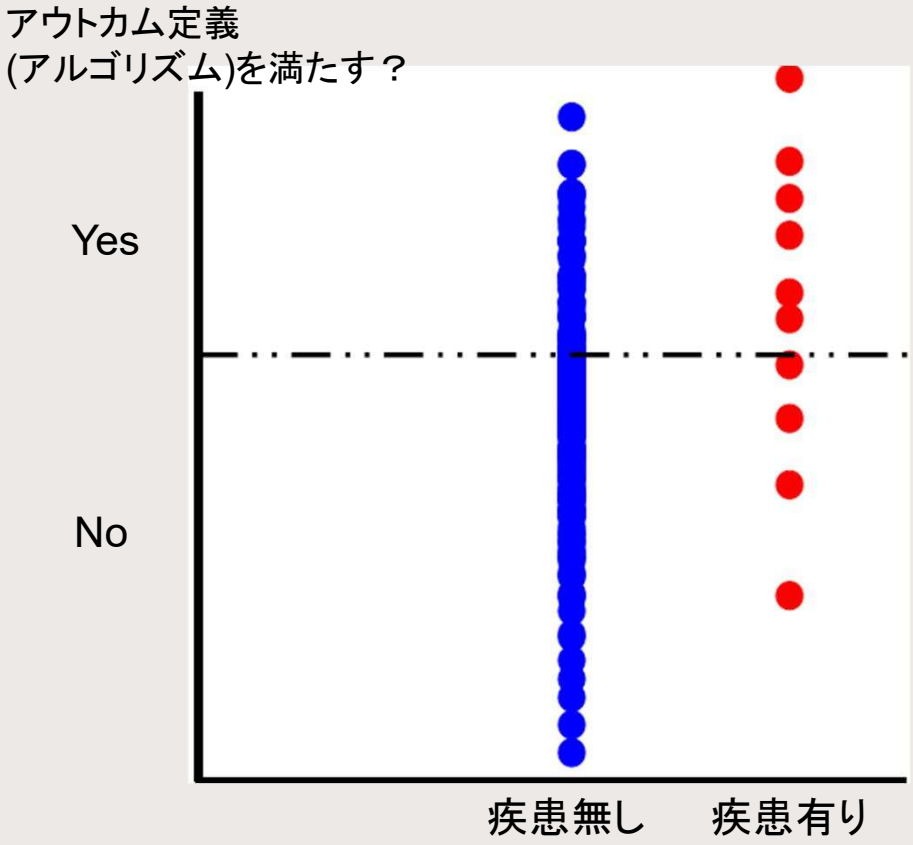
研究結果で示す指標

感度・特異度
陽性的中度・陰性的中度

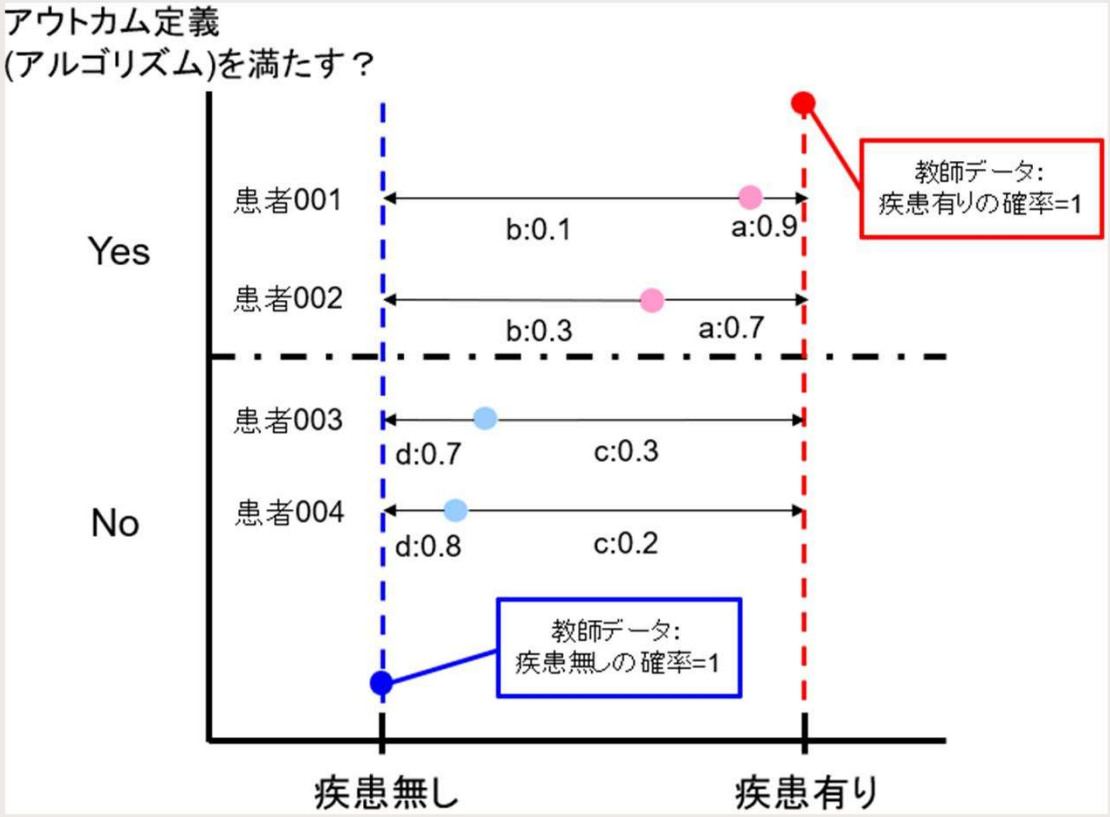
		Gold Standardの定義を満たすか			
		Yes	No	合計	
アウトカム定義（アルゴリズム）を満たすか	Yes	a（真の陽性）	b（偽陽性）	a+b	陽性的中度（PPV）=a/a+b
	No	c（偽陰性）	d（真の陰性）	c+d	陰性的中度（NPV）=d/c+d
	合計	a+c	b+d	a+b+c+d	
		感度=a/a+c	特異度=d/b+d		

2. 方法

感度・特異度・陽性的中度・陰性的中度を知りたいので
全患者で疾患有り・無しを確認する必要がある



全患者をカルテレビューすることは難しいため
機械学習によりGold Standardの定義とアルゴリズムの該当度の算出を行う→
各患者の医療情報がどれだけ教師データに似ているかを機械学習で数値化

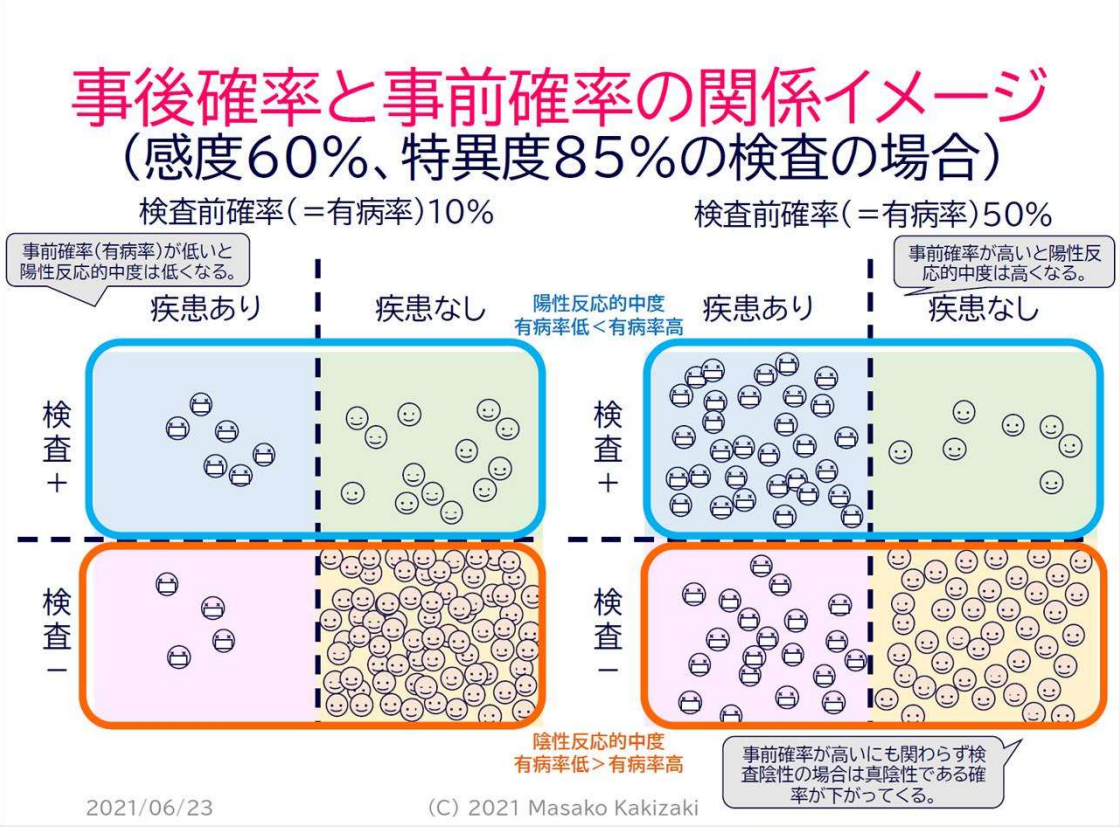


2. 方法

▪一般的な臨床検査の目的（除外診断・確定診断など）とは異なり、疾患定義のバリデーション研究では薬剤疫学研究におけるアウトカムを正しく定義できるか（ひいては相対リスクを正しく評価できるか）を目的とする。

		Gold Standardの定義を満たすか			
		Yes	No	合計	
アウトカム定義（アルゴリズム）を満たすか	Yes	a（真の陽性）	b（偽陽性）	a+b	陽性的中度（PPV）=a/a+b
	No	c（偽陰性）	d（真の陰性）	c+d	陰性的中度（NPV）=d/c+d
	合計	a+c	b+d	a+b+c+d	
		感度=a/a+c	特異度=d/b+d		

疾患がある人を、疾患ありと診断できるか・除外診断→感度
疾患がない人を、きちんと健康と診断できるか・確定診断→特異度



https://www.jsph.jp/covid/files/5BAA6E3.pdf?utm_source=chatgpt.com

2. 方法

▪バリデーション研究において、有病率は変動する値として取り扱わない。

▪比較される2 群（例：糖尿病薬A 群vs. 糖尿病薬B 群）間のアウトカムに対する相対リスクは、アウトカム定義の陽性的中度が高く、感度が（その大小にかかわらず）2群間で同じ場合に正しく求まる[2]。

		Gold Standardの定義を満たすか			
		Yes	No	合計	
アウトカム定義（アルゴリズム）を満たすか	Yes	a（真の陽性）	b（偽陽性）	a+b	陽性的中度（PPV）=a/a+b
	No	c（偽陰性）	d（真の陰性）	c+d	陰性的中度（NPV）=d/c+d
	合計	a+c	b+d	a+b+c+d	
		感度=a/a+c	特異度=d/b+d		

バリデーション研究後に行う薬剤疫学研究のイメージ

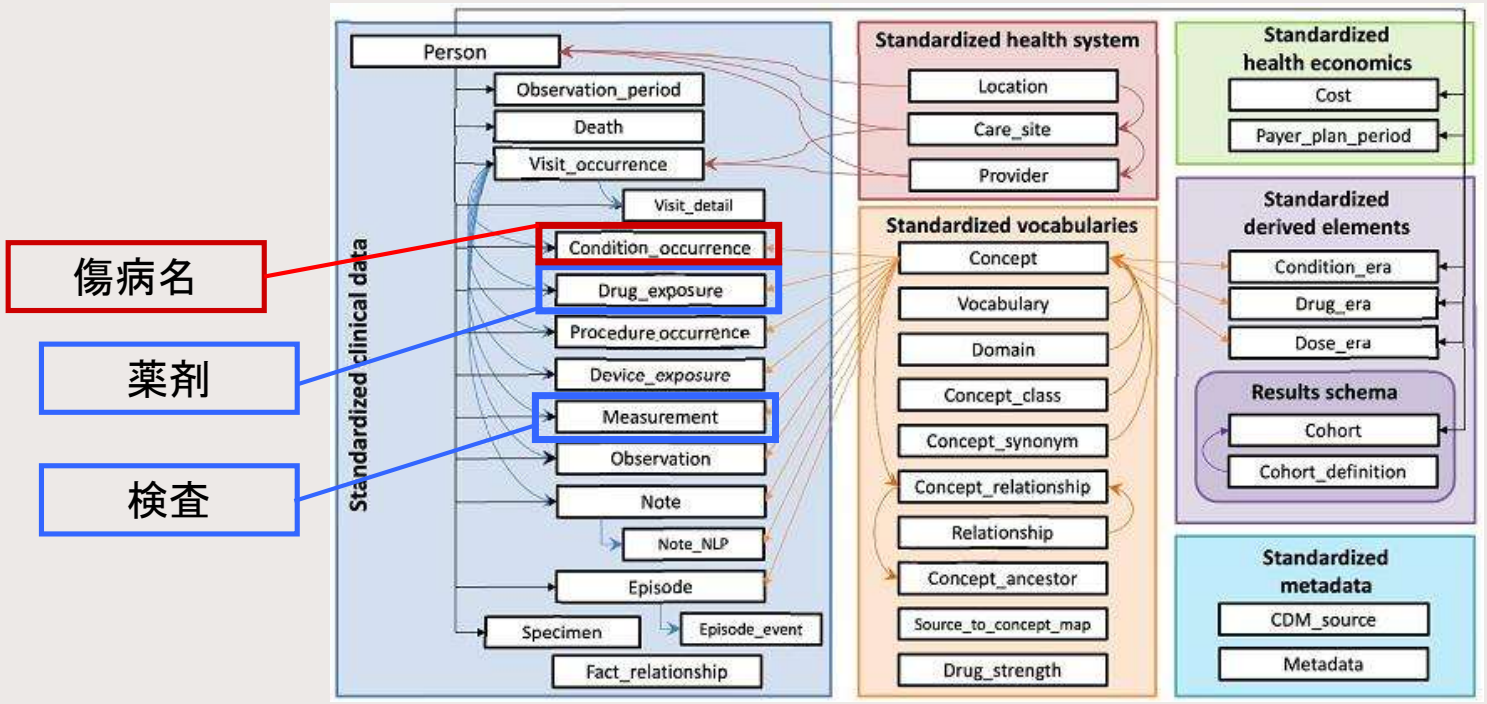
要因	罹患		計
	あり	なし	
薬剤A	A	B	A+B
薬剤B	C	D	C+D

$$\text{相対リスク} = \frac{\text{薬剤Aの罹患リスク}}{\text{薬剤Bの罹患リスク}} = \frac{\frac{A}{A+B}}{\frac{C}{C+D}}$$

2.岩上 将夫, 青木 事成, 赤沢 学, et al.「日本における傷病名を中心とするレセプト情報から得られる指標のバリデーションに関するタスクフォース」報告書. 薬剤疫学. 2018;23(2):95-123

2. 方法

- OMOP CDMは、Real World Data の分析用標準化を実現するグローバルなオープン規格である。以下の図表のようなデータテーブルと、その中の項目、格納形態が形式化されている[3] 。
- 本研究では、国際医療福祉大学病院の電子カルテからSS-MIX2形式で抽出されたデータが、OMOP CDMに変換された。アルゴリズム設定においては主に傷病名、薬剤、検査のデータを用いた。手技については電子カルテ・SS-MIX2の特性上、変換されなかった。



3.Observational Health Data Sciences and Informatics (OHDSI). Our Journey. Accessed Jul 08, 2024. https://www.ohdsi.org/wp-content/uploads/2021/09/OHDSI-OurJourney2021-Final.pdf

3. 結果

TABLE1. 各アルゴリズムの評価指標

Acute Myocardial Infraction

アルゴリズム区分	含まれる定義
A	傷病名：Acute Myocardial Infractionを定義する傷病名コード
B	抗血小板薬：クロピドグレル・プラスグレル・チカグレロルのいずれかを使用
C	心関連臨床検査：Creatine Kinase (CK)・CK-MB・Troponin・BNP・NTproBNPのいずれかを実施

アルゴリズム	感度	陽性的中度	特異度	陰性的中度
A	0.148 (0.126 - 0.173)	0.168 (0.143 - 0.195)	0.987 (0.986 - 0.988)	0.985 (0.984 - 0.986)
A + B	0.039 (0.026 - 0.055)	0.074 (0.051 - 0.104)	0.993 (0.992 - 0.993)	0.985 (0.984 - 0.986)
A + C	0.148 (0.126 - 0.173)	0.170 (0.145 - 0.197)	0.987 (0.986 - 0.988)	0.985 (0.984 - 0.986)
A + B + C	0.025 (0.016 - 0.038)	0.282 (0.190 - 0.390)	0.999 (0.999 - 0.999)	0.983 (0.982 - 0.984)

()内は95%CI

3. 結果

TABLE1. 各アルゴリズムの評価指標

Heart Failure

アルゴリズム区分	含まれる定義
A	傷病名：Heart Failureを定義する傷病名コード
B	心関連臨床検査：Creatine Kinase (CK)・CK-MB・Troponin・BNP・NTproBNPのいずれかを実施

アルゴリズム	感度	陽性的中度	特異度	陰性的中度
A	0.371 (0.357 - 0.386)	0.466 (0.450 - 0.483)	0.962 (0.960 - 0.963)	0.944 (0.942 - 0.946)
A + B	0.371 (0.357 - 0.386)	0.471 (0.455 - 0.488)	0.962 (0.961 - 0.964)	0.944 (0.942 - 0.946)

3. 結果

TABLE1. 各アルゴリズムの評価指標

Stroke

アルゴリズム区分	含まれる定義
A	傷病名：Strokeを定義する傷病名コード

アルゴリズム	感度	陽性的中度	特異度	陰性的中度
A	0.064 (0.058 - 0.071)	0.151 (0.137 - 0.166)	0.955 (0.954 - 0.957)	0.892 (0.889 - 0.895)

*Atrial FibrillationとAngina Pectorisについてはプログラムでデータが少なすぎる(too few)というエラーとなり評価指標が算出できなかった

4. 考察

先行研究では数千万以上の患者数のデータベースが使用されたと推察され、本研究と比べ患者数およびマッピングテーブル数に大きなギャップがあったと考えられる。本研究の患者数、マッピングテーブル数ではGold Standardを期待通りに設定できず、アルゴリズムが適切に評価されなかった可能性がある。

*** 先行論文からの抜粋**

先行研究におけるデータベース名		年齢 (歳)
CCAE	IBM® MarketScan® Commercial Claims and Encounters Database	18–62
MDCR	IBM® MarketScan® Medicare Supplemental and Coordination of Benefits Database	>=66
MDCD	IBM® MarketScan® Multi-State Medicaid	18–62
Optum1862	Optum© De-Identified Clinformatics® Data Mart Database – Date of Death	18–62
OptumGE66	Optum© De-Identified Clinformatics® Data Mart Database – Date of Death	>=66
PanTher1862	Optum© de-identified Electronic Health Record Dataset	18–62
PanTherGE66	Optum© de-identified Electronic Health Record Dataset	>=66

J.N. Swerdel, G. Hripcsak, P.B. Ryan, Phevaluator: Development and evaluation of a phenotype algorithm evaluator, J. Biomed. Inform. 97 (2019) 103258

4. 考察

＊ 先行論文からの抜粋

AMIを特定する定義	データベース	感度	陽性的中度	特異度	陰性的中度
傷病名が1回以上	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
	PanThar1862	0.630	0.479	0.994	0.977
	PanTharGE66	0.574	0.431	0.971	0.984
傷病名＋coronary revascularization	CCAE	0.062	0.967	0.999	0.995
	Optum1862	0.073	0.944	0.999	0.993
	OptumGE66	0.037	0.916	0.999	0.955
	MDCD	0.037	0.905	0.999	0.987
	MDCR	0.026	0.915	0.999	0.959
	PanThar1862	0.017	0.771	0.999	0.991
	PanTharGE66	0.010	0.667	0.999	0.964

J.N. Swerdel, G. Hripcsak, P.B. Ryan, Phevaluator: Development and evaluation of a phenotype algorithm evaluator, J. Biomed. Inform. 97 (2019) 103258

4. 考察

教師データを設定する際、PheValuator開発者が推奨する傷病名レコードの複数回発生を基準にしたが、本研究のデータでは1患者あたりの傷病名の発生回数が少なかった。一方先行研究では外来および入院診療や手技を広範にカバーする保険請求データが主に使用されていたので、1患者あたりの傷病名の回数が多いと推察された。このギャップが、Gold Standardを期待通りに設定できない原因となった可能性がある。

傷病名	症例数	1患者当たりの傷病名の回数
Acute Myocardial Infarction	1,146	1.37
Heart Failure	4,533	1.48
Stroke	4,559	1.53
Atrial Fibrillation	2,238	<1.3
Angina Pectoris	1,450	<1.3

4. 考察

本研究の計画が実現できる要件

- 複数病院からのデータ統合
  - データ標準規格の特性を生かし、一度作成した変換設定を利用して複数病院のデータをOMOP CDM形式として格納し症例数を確保する
- OMOP CDM変換のデータソースの拡大
  - 電子カルテに加えレセプト情報なども変換対象として加え、参照できる項目を増加させる
- 教師データ定義の変更
  - 「疾患有りの確率が1である患者」を定義する条件を回数（レコード数）以外で設定する

4. 考察

疑い病名が含まれていることの影響

- 本研究で対象とした疾患で疑い病名のレコード数は正式にはカウントしていなかったが*AMI*については疑い病名フラグのレコードが多く含まれている可能性があった
- PheValuator上で疑い病名を除く処理は困難であったため実行できなかった
- もし仮に疑い病名が多い場合は、教師データ作成の設定時にノイズとなった可能性が高く、算出される指標に影響すると考えられる
- 本研究の計画が実現できる要件として、疑い病名フラグを教師データ作成時に除くなど検討する必要がある

5. 結語

本研究では、OMOP CDMおよびPheValuatorを用いて効率的な疾患定義の妥当性検証を試みたが、方法論がデータベースの特性に適合していない可能性があった。データベース研究では、適用する研究手法がそのデータベースの規模と特性に合致しているかをfeasibility調査などで十分に吟味することが重要である。