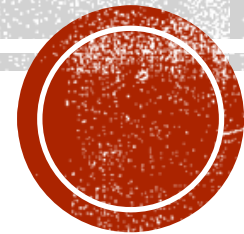


日本におけるアウトカム定義の検証

チャートアブストラクション/病名のみアウトカム定義の実例紹介

および展開するにあたっての留意点



中島直樹、他MICメンバー
九州大学病院 メディカル・インフォメーションセンター

2022年 第1回 医療情報データベース推進コンソーシアム
2022.12.05

COI 開示

日本におけるアウトカム定義の検証

チャートアブストラクション/病名のみアウトカム定義の実例紹介および
展開するにあたっての留意点

発表者名： 中島直樹
九州大学病院 メディカル・インフォメーションセンター

共同研究： 富士通（株）、ファイザー製薬（株）
エクイティ：（株）カルナヘルスサポート



DIA年会2022において、アウトカムバリデーションのセッションが開催された (2022年10月10日、東京ビッグサイト)

LS16

Current Situation and Issues for Conducting Outcome Validation Study Prior to Post-Marketing Database Study



PMDA
丸山穂高氏

ラトガー大学
瀬戸口聡子氏

ファイザーR&D
松居真理氏

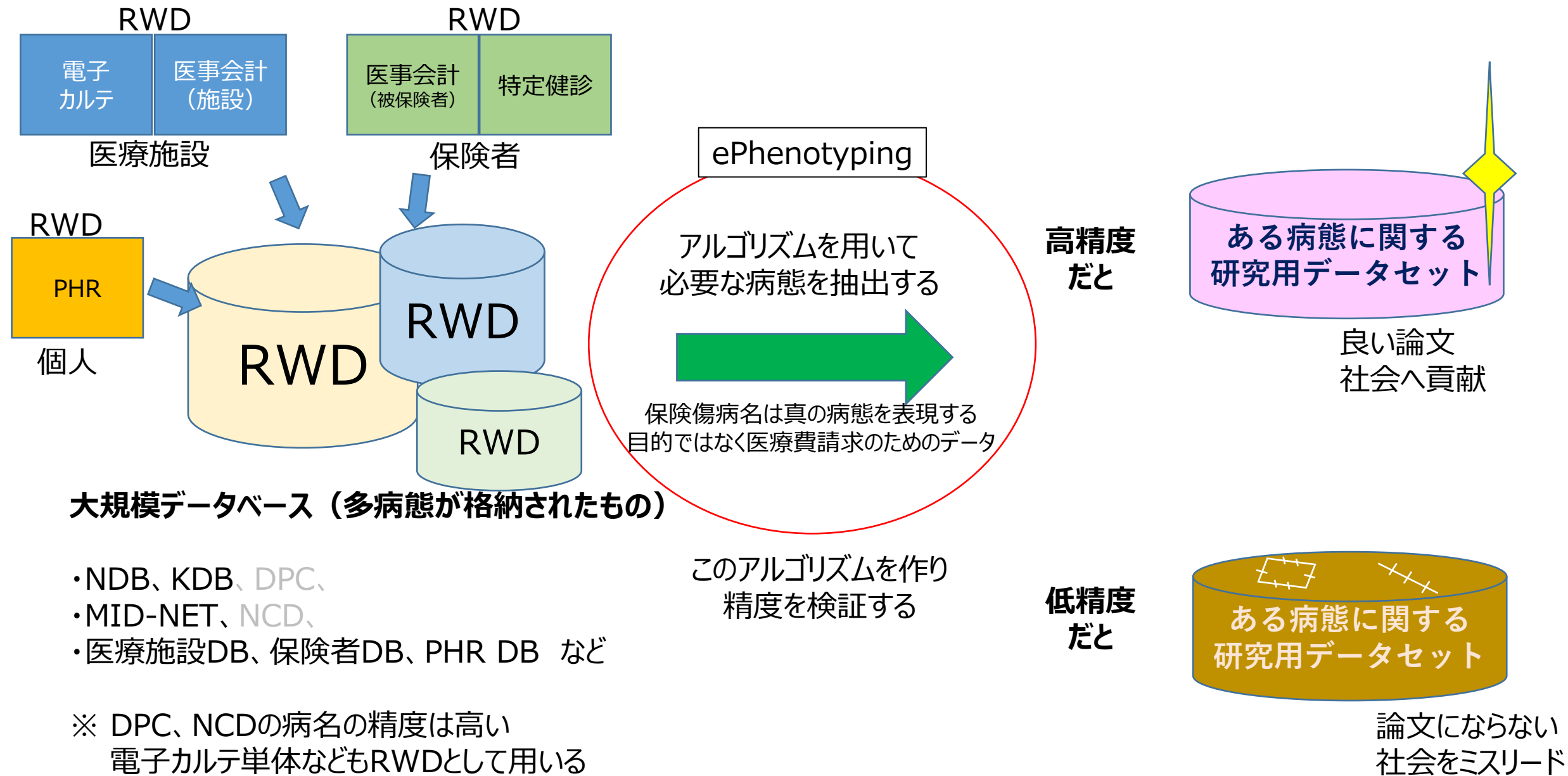


座長 中島

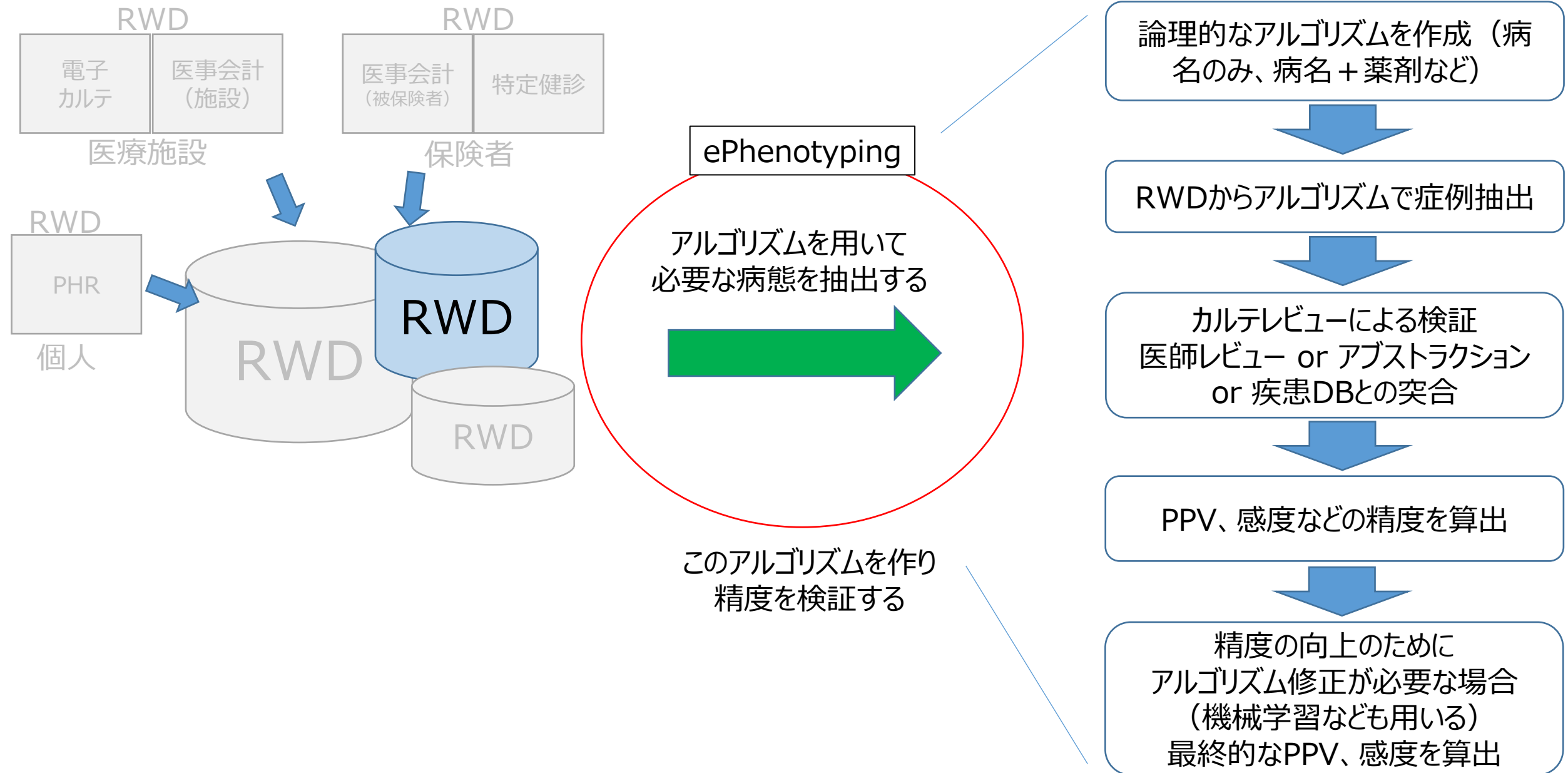
総論として

- ・蓄積したデータから、Phenotype（健康・病態・病名などの表現型）をアルゴリズムで抽出する方法を「ePhenotyping」という
- ・その中で、何らかの暴露に対するアウトカムであるPhenotypeを抽出するアルゴリズムを「アウトカム定義」という。薬剤＝暴露となる薬剤疫学で使われる。
- ・つまり、本コンソーシアムの議論やアウトカム定義の研究は、アウトカム定義のみならず、ePhenotyping全体に貢献する
（例えばgenotypeとphenotypeを突合するPrecision Medicineの開発研究など）
- ・アルゴリズムを作成する環境やアルゴリズムを使う目的別に、「アルゴリズム（アウトカム定義含む）」や「アルゴリズムを作成・検証する手法」は異なる
- ・アルゴリズムは普遍的なものではなく、施設、時期、システムが変われば大きく変わる。
- ・特に今後は、ePhenotyping精度が高くなるための環境（制度を含む）やシステムの向上が求められる

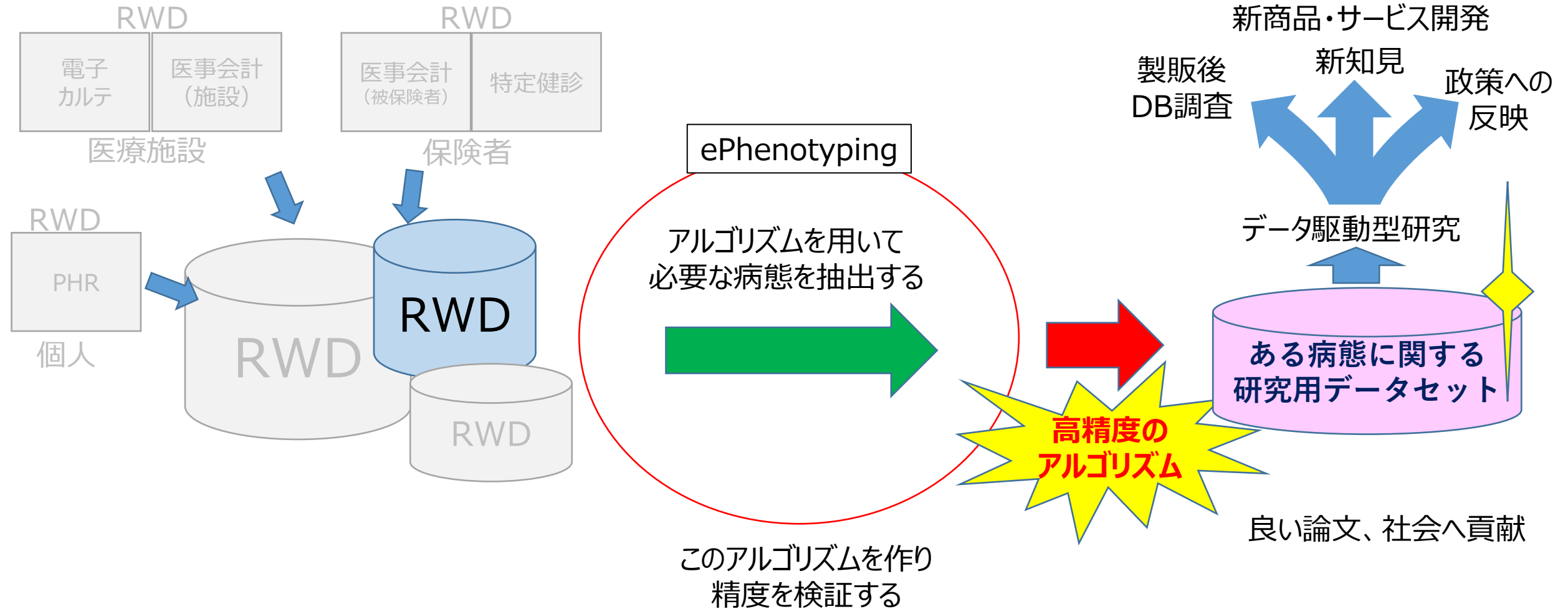
何のためにアルゴリズム（アウトカム定義）を作るのか？



カルテレビューによるアルゴリズム策定のステップ



高い精度のアルゴリズムができると・・・



アルゴリズムの精度の指標は？

- ◆ 利用目的によって異なる
- ◆ 感度・特異度を知りたいが、一般に感度や特異度の測定は難しい。
- ◆ リスク比を知るには特異度が重要だが、PPVでもリスク比は（擬陽性が少ない場合に）近似する
- ◆ それで、現実的にはPPVを測定し、リスク比を算出することが多い

		GSによる判定			
		真	偽		
アルゴリズム	陽性	a (正解)	b (擬陽性)	$\frac{a}{a+b}$	= PPV (陽性的中率)
	陰性	c (擬陰性)	d (正解)	$\frac{d}{c+d}$	= NPV (陰性的中率)
		$\frac{a}{a+c}$ = 感度	$\frac{d}{b+d}$ = 特異度		

アルゴリズムを検証する二つの方法

1. チャートレビュー（カルテレビュー）の結果をGSとする

1 - 1 カルテレビューする人が、直接真偽を判定

専門医、非専門医～学生まで様々（米国）
専門医が主（日本）

1 - 2 カルテからアブストラクトされたデータで間接的に真偽判定

アブストラクションする人 = アブストラクター
真偽判定する人 = アジュディケーター（専門医が主）

2. 悉皆性の高い疾患レジストリDBをGSとして突合する

欧米ではよくやられる。日本は難しいことが多い

1. 法制度の差（データベース突合の可否など）
2. 欧州、台湾では既突合DBで使えるものもある
3. 日本では悉皆性が高くないものが多い。その場合、感度は推定できるがPPVはわからない

	真	GSによる判定	偽		
陽性	a (正解)		b (擬陽性)	$\frac{a}{a+b}$	=PPV (陽性的中率)
陰性	c (擬陰性)		d (正解)	$\frac{d}{c+d}$	=NPV (陰性的中率)
	$\frac{a}{a+c}$	=感度	$\frac{d}{b+d}$	=特異度	

事例 1 : 厚生労働科学研究費補助金 (循環器疾患・糖尿病等生活習慣病対策総合研究事業)

1 型糖尿病の実態調査、客観的診断基準、日常生活・社会生活に着目した重症度評価の作成に関する研究

H27-循環器等-一般-006 研究代表者 田嶋尚子 (東京慈恵会医科大学・内科) 分担研究者 中島直樹 (九州大学)

確定病名のみアウトカム定義の精度

平成27年度研究

DM専門医によるカルテレビュー

		インスリン依存性					総計
		1	2	3	4	5	
1型DM	1	262	11	6	12	35	326
	2	16	14		3	5	38
	3	3	1	45	1	2	52
	4				② 2	2	① 4
	5	30	11	17	10	376	444
総計		311	37	68	28	420	864

216,857人患者からの1型糖尿病の確定診断のみをアウトカム定義とした場合にPPVが52%であったものを、2つのインスリン治療法 (CSII 用のシリンジ ポンプとインスリン針) のクレームコードの両方の症例定義を加えることにより、PPVが90%を超える定義を策定することができた。

Tasuku Okui, Naoki Nakashima, et al.
Performance evaluation of case definitions of type 1 diabetes for health insurance claims data in Japan, *BMC Medical Informatics and Decision Making*, doi:10.1186/s12911-021-01422-z, 2021.02.

1型糖尿病 :

上記①の448名を真とする

1型糖尿病かつインスリン依存 :

上記②の390名を真とする

事例2：疾患登録DBとの突合（2021年度厚労科研宇山班研究）

（未発表データのため九州大学の結果のみ）

◆ 悪性腫瘍

- 肺癌
- 悪性リンパ腫
- 全癌

疾患DBへの登録条件を「対象」で考慮すれば、院内がん登録のような、悉皆性の高いDBをGSとするならば、全て明らかにできる

対象とする事象：

再発や継続治療の症例は含めず、当院で肺がん治療を開始した症例（診断は他院でも可）。

対象期間： 2016～2018年（3年間）

初期アウトカム定義： = APC定義（感度100%を想定）
「肺癌関連の病名が付与されている」
（SS-MIX病名、DPC病名、レセプト病名）

真偽判定には、院内がん登録を用いた

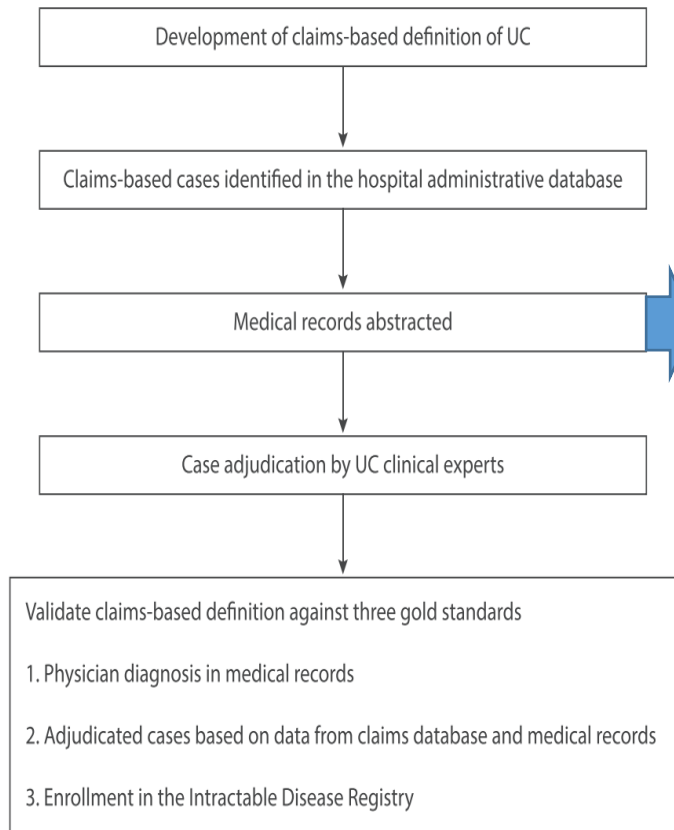
- ✓ 正確な疾患登録DBを用いると、真偽判定労力激減。症例数の制限は緩和し、初期定義が低いPPVでも可能
- ✓ 但し、該当する疾患がすべて登録されている（悉皆である）ことが必要。院内がん登録は上記対象事象を100%反映している

	GSによる判定		
	真	偽	
陽性 アルゴリズム	a (正解)	b (擬陽性)	$\frac{a}{a+b}$ =PPV (陽性的中率)
陰性	c (擬陰性)	d (正解)	$\frac{d}{c+d}$ =NPV (陰性的中率)
	$\frac{a}{a+c}$ =感度	$\frac{d}{b+d}$ =特異度	

事例3 :

医師レビュー、アブストラクション、DB突合の三種のGSの検討を行った研究（ファイザー社との共同研究）

Haruei Ogino, Hiromu Morikubo, Keita Fukaura, Tasuku Okui, Sean Gardiner, Naonobu Sugiyama, Noritoshi Yoshii, Tsutomu Kawaguchi, Haoqian Chen, Edward Nonnenmacher, Soko Setoguchi, Naoki Nakashima, Taku Kobayashi, Validation of a claims-based algorithm to identify cases of ulcerative colitis in Japan, *Journal of Gastroenterology and Hepatology*, doi:10.1111/jgh.15732, 2021.10.



訓練されたアブストラクターがランダムに抽出された20症例で予備のアブストラクションし、専門医と項目などを修正し全部で200症例（各施設）アブストラクション

Table 5 PPV (95% CI) of a claims-based algorithm for identifying cases of UC confirmed by gold standard diagnoses, stratified by gender

Gold standard criteria	Female (N = 187)		Male (N = 206)	
	N	PPV (95% CI)	N	PPV (95% CI)
Physician diagnosis [†]	162	86.6% (81.8–91.5)	194	94.2% (91.0–97.4)
Adjudicated cases [‡]	162	86.6% (81.8–91.5)	194	94.2% (91.0–97.4)
Enrollment in Japanese Intractable Disease Registry	75	40.1% (33.1–47.1)	88	42.7% (36.0–49.5)

[†]From medical records.

[‡]Based on physician diagnosis from medical records and information on core clinical symptoms and colonoscopy from the medical records. CI, confidence interval; N, number of cases meeting gold standard definition; PPV, positive predictive value; UC, ulcerative colitis.

Figure 1 Study flow chart. UC, ulcerative colitis.

医師レビューとアブストラクションのGSでは、PPVに差はなく十分に高かった疾患登録DBをGSとした場合は、PPVが低く出た（登録率が低いため）

事例4：各施設で別々のアブストラクターでアブストラクション行った事例

2021年度 医薬品医療機器総合機構（PMDA）受託調査 アウトカム定義のバリデーションに関連する調査業務
チャート・アブストラクターを活用してアウトカム定義のバリデーションを実施する際の留意事項の検討

（未発表につき概要のみ）

症例：

過去の研究で既に専門医レビューで真偽判定をしている症例群

「入院して治療を行った急性期脳出血」、「入院を要する糖尿病性ケトアシドーシス」

参加施設：

MID-NET協力医療施設のうち7施設

方法：

アブストラクターを各協力医療機関が3名設定したが、アブストラクターのリクルートが難しく、「職種は限定しない。3年以上、電子的な診療情報を扱った実務経験がある」とした。Web説明会に全員が参加した上で、アブストラクションフォームに従ってアブストラクションを行った。その結果により専門医師2名が真偽判定し、過去の専門医レビューの真偽判定結果をGSとして比較した。

結果：

アブストラクターの経験や職種はばらつきが大きかった。

検証の結果、一致率は十分に出ていない。現在解析中であるが、各施設でアブストラクターを養成するのは現実的ではなく、スキルの高いアブストラクターが各参加施設を回ってアブストラクションを行うことが現実的か

事例5： All Possible Caseの設定（2021年度厚労科研宇山班研究）

（未発表データのため九州大学の結果のみ）

◆ 急性膵炎

対象事象：

入院して蛋白分解酵素阻害薬の経静脈的投与による治療を要した急性膵炎または慢性膵炎の急性増悪（ただし外科的操作（ERCP/手術）が起因の膵炎を除く）

対象期間：2011年1月～2019年12月（9年間）

APC=「cが0となる事象」を選ぶ
感度、特異度も算出できるようになる

	GSによる判定		
	真	偽	
陽性 アルゴリズム	a (正解)	b (擬陽性)	$\frac{a}{a+b}$ =PPV (陽性的中率)
陰性	c (擬陰性)	d (正解)	$\frac{d}{c+d}$ =NPV (陰性的中率)
	$\frac{a}{a+c}$ =感度	$\frac{d}{b+d}$ =特異度	

All Possible Cases (APC) 定義

「対象とする事象」を100%拾える初期アウトカム定義（初期感度を100%とする）ことにより、PPV、感度の両方を測定できる（特異度も測定可能）

但し、宇山班の希望としてはAPC定義のPPVを50%以上とする（サンプル数200症例の場合に少なくとも真症例を100症例以上得るため）

結果：

APC定義のPPV 51.2%、感度100%を得た。その後、修正APC定義を検討した結果、PPV 71.6%、感度 80.7%となった

事例6：アウトカムが明確なコホートを使って擬陽性をゼロにする

Wayne A. Ray, et al, Oral Erythromycin and the Risk of Sudden Death from Cardiac Causes. N Engl J Med 2004; 351:1089-1096

テネシーのMedicadeデータで突然死した人のデータを使った研究。
突然死であることが確認されており、つまり擬陽性はない。

PPV、特異度は100%となる。検証の必要がない。
最終的に出てくるリスク比にバイアスが全くかからない

これが0となる

アルゴリズム	GSによる判定		
	真	偽	
陽性	a (正解)	b (擬陽性)	$\frac{a}{a+b}$ =PPV (陽性的中率)
陰性	c (擬陰性)	d (正解)	$\frac{d}{c+d}$ =NPV (陰性的中率)
	$\frac{a}{a+c}$ = 感度	$\frac{d}{b+d}$ = 特異度	

100%となる

本論文ではそのデータを使って、エリスロマイシンの暴露の関与を調べている。

アウトカム定義の注意点

利用目的によって異なる

- ✓ 感度、特異度、PPV何を重視するか？
- ✓ 単施設で用いるのか、複数施設なのか？
- ✓ 同じ施設（群）で用いるのか、外挿するのか？
- ✓ 人が理解できる単純なスクリプトしか使えないのか？それをどう作るのか？
- ✓ 機械学習を使える場合、どの機械学習を使うかによっても結果が変わる

作成した定義の精度はDB間ではばらつきがあり、同じDBでも時期で変わる

- ✓ どのロジックで、どのデータベースを用いたか、どの期間のデータか、の明記
- ✓ 感度、特異度、PPV等の明記

限界は多い

- ✓ 感度の正確な把握は一般に困難、PPVも充分上がらないことも多い
- ✓ いずれ電子カルテ側や保険診療の方法を変えて、より精度の高い構造化データの入力から行うべき

事例から見えてきたこと（再掲）

- 事例 1 より： 保険傷病名だけの定義では、一般にPPVは低い（深刻な課題）
- 事例 2 より： 悉皆性の高い疾患登録DBであれば、その登録条件を抽出対象に設定できれば、良いGSとなり、効率よく正確に感度・特異度・PPV・NPVを把握できる
- 事例 3 より： 訓練されたアブストラクタであれば専門医の直接のカルテレビューと判定結果は変わらない
疾患登録DBも登録の悉皆性が低ければ、良いGSにはならない
- 事例 4 より： アブストラクタを医療施設毎に確保することは、アウトカム検証の精度上でも困難なことが示唆された
- 事例 5 より： 感度も把握したい場合は、擬陰性が出ないAPC定義（感度100%となる）を定める。
多くの場合PPVが低い、APC定義を修正してPPVを上げる（感度を下げすぎない）
- 事例 6 より： アウトカムが明確なコホートを使って擬陽性をゼロにすることにより、PPV、特異度が100%となり最終的なリスク比のバイアスが消える。検証の必要がない。

結論

- ◆ 日本では、現状では悉皆性の高い疾患登録DBは少ない
 - ✓ 疾患や環境によっては存在している
 - 診断が容易であり、生命に直結し、経済的補償を行うためのDB
 - ○院内がん登録（事例2）
 - ○未成年一型糖尿病（小児慢性特定疾病）
 - ×潰瘍性大腸炎（指定難病97）（事例3）
 - 悉皆性の高い疾患登録DBがあれば、GSとすることが勧められる
 - 「DBの登録条件」と「アウトカム定義の利用目的」が合えば、ではある
 - 欧米のようにNDBなど大規模DBとの突合が可能となる制度改革が望ましい
- ◆ 「疾患名」のみ等の簡単なアウトカム定義は精度が低いことが多い。アウトカム定義の検証や修正は重要
- ◆ 専門医師によるカルテレビューは医師の負担が大きい
 - ✓ アブストラクションの手法をさらに洗練化する必要あり。方法も人材も
- ◆ いずれ電子カルテなどの入力機能を改善する必要あり

アウトカム定義のバリデーション手法の発展は、
MID-NETのみならず、日本でのデータ駆動型臨床研究全体に必要です



ご清聴ありがとうございました