

令和4年度 厚生労働行政推進調査事業費補助金
(医薬品・医療機器等レギュラトリーサイエンス政策研究事業)
分担研究報告書

医療情報データベースの活用推進に関する研究

中島 直樹・九州大学病院・教授

研究要旨：

医療情報データベースに基づく医薬品の安全性等を評価する場合には、データベースに含まれる情報を組み合わせて対象とする有害事象（アウトカム）を定義する必要があり、効率的なアウトカム定義の作成手法の検討及び実用化可能なアウトカム定義を確立するため、本研究を実施した。

昨年度に作成したAll Possible Case (APC)定義を基に、「DPC病名の主病名～医療資源1のいずれかにK85を含む」の単独条件またはDPC病名の主病名～医療資源1のいずれかにK86を含む、「経腸成分栄養剤 (ATC_V06DE)」、「血清アミラーゼおよびリパーゼの異常値の有無」、「急性膵炎の重症度分類」のうち3項目に該当する条件を加えた2通りの改良型アウトカム定義を作成した。これらを各施設へ適用した結果、本院ではPPVや感度が若干落ちるものの、他2施設のPPVや感度は80%以上と良好であり、全体的に見ておおむね目標 (PPV・感度80%以上の改良型アウトカム定義) は達成できたと考えられた。

A. 研究目的

製造販売後の医薬品安全性評価は、従来、副作用報告、使用成績調査等の結果に基づくことが主であったが、医療情報データベースの整備等によりリアルワールドでの大規模データに基づく評価が可能となりつつある。

MID-NET[®]は、厚生労働省の医療情報データベース基盤整備事業により構築されたデータベースで、独立行政法人医薬品医療機器総合機構（以下「PMDA」という。）による運用管理の下、平成30年度から運用を開始している。MID-NET[®]は、診療情報明細書（レセプト）、診断群分類（DPC）及び電子カルテの情報が利用可能で、現在では行政による活用の他、臨床研究や製造販売後データベース調査にも活用されており、製造販売後の医薬品安全対策の質の向上が期待されている。

医療情報データベースに基づく医薬品の安全性等の評価では、データベースに含まれる情報から対象とする有害事象（アウトカム）を適切に特定するために、信頼できるアウトカム定義を作成する必要があるが、本邦においてアウトカム定義のバリデーションが実施された例はほとんどない。「MID-NET[®]データの特性解析及びデータ抽出条件・解析手法等に関する研究」（日本医療研究開発機構 医薬品等規制調和・評価研究事業）（以下「先行研究」という。）において、MID-NETを対象にアウトカム定義の作成及びその妥当性の評価を効率的に実施するための

検討を実施し、基礎的な検討手法を確立するとともに、バリデーションされた複数のアウトカム定義が作成された。しかしながら、実用化可能なアウトカム定義を増やすためには、更なる検討手法の効率化や具体的なアウトカム定義の作成を継続的に実施する必要がある。

本研究は、先行研究の成果を踏まえて、研究の流れを見直し All possible cases の定義（以下「APC 定義」という。）を決定した上で以降の検討を行う等の検討手法の改善及び実用化可能なアウトカム定義の確立を目指し、医薬品安全性評価における医療情報データベースの活用促進と、より科学的な根拠に基づく安全対策の実現に繋げることを目的とした。

B. 研究方法

安全対策上の必要性や重要性を考慮し、検討の対象とするアウトカムの中から「急性膵炎」を選定し、昨年度に作成した APC 定義を基に機械学習の手法を取り入れてアウトカム定義を作成した（以下、改良型アウトカム定義）。また、改良型アウトカム定義について、複数医療機関にて妥当性の評価を行った。

改良型アウトカム定義の検討の方針 (B-1) と研究の流れ (B-2) は以下のとおりである。

B-1. 改良型アウトカム定義の検討方針

- ◆ MID-NET 事業では全拠点に対して同じアウトカム定義を用いて製販後 DB 調査を実

施することが想定されることから、複数施設で使用可能な一つの改良型アウトカム定義を開発する

- ◆ 薬剤性の有害事象を検討するための定義として、「疑わしいケース」は含めずに「真のケース」のみで対応できることが望ましいため、「真のケースのみを真症例とした場合」の結果に基づく改良型アウトカム定義を作成する
- ◆ 陽性的中度（以下「PPV」という。）および感度がいずれも 80%程度以上の改良型アウトカム定義の作成を目指す

B-2.1) APC 定義の検討を実施した九州大学病院（以下「本院」という。）、佐賀大学医学部附属病院および香川大学医学部附属病院の MID-NET データを統合してデータセットを作成する

B-2.2) APC 定義を基に機械学習による改良型アウトカム定義を作成する

B-2.3) 2)で作成した改良型アウトカム定義に対して PPV 及び感度を算出する

B-2.4) 3)について複数医療機関の PPV の比較及び医療機関間の差異の要因を検討する

（倫理面への配慮）

医療機関が行う作業は、倫理審査承認（承認番号 2020-330）の下、「人を対象とする生命科学・医学系研究に関する倫理指針」を遵守して行った。

C. 結果

1) 機械学習に用いる解析対象数の検討

本院と他 2 施設は真偽判定を実施した症例数（以下「判定対象 possible case 数」という。）が異なり、単純に 3 施設を合わせたデータセットでは、症例数の多い本院の特性に偏った改良型アウトカム定義ができる可能性が懸念されたため、本院の判定対象 possible case から 100 例ランダムサンプリングして解析対象の症例数を統一した（表 1）。

- ・データ期間： 2011 年 1 月 1 日～
2019 年 12 月 31 日

表 1. APC 定義およびランダムサンプリング後の真偽判定結果

APC 定義の結果

医療機関	Possible Cases数	判定対象 Possible Case数	真のケース数	その他のケース数	PPV[%]	感度[%]
香川	156	100	65	35	65.00	100.00
九州	213	213	109	104	51.17	100.00
佐賀	205	100	87	13	87.00	100.00
合計	574	413	261	152	63.20	100.00

ランダムサンプリング後

医療機関	Possible Cases数	判定対象 Possible Case数	真のケース数	その他のケース数	PPV[%]	感度[%]
香川	156	100	65	35	65.00	100.00
九州	213	100	54	46	54.00	100.00
佐賀	205	100	87	13	87.00	100.00
合計	574	300	206	94	68.67	100.00

2) 機械学習に用いるデータセットの検討

MID-NET 仕様および臨床的観点から機械学習に用いるデータ項目を以下の様に決定した。

データ項目は計 1172 個であった。

- ・データ期間：DPC 病名に該当する入院日～退院日
- ・index date: DPC 病名に該当する入院日
- DPC 傷病情報：
 - ◆ ICD10 コードを先頭 3 桁にまとめ、主傷病名～入院後発症までの 6 種、主傷病名～資源病名 1 までの 3 種のどちらに該当したかが分かるよう、下記のように接頭語を付けた
「DPC 病名 X 種_ICD10 コード(3 桁)」
留意) 主病名～医療資源 1 に該当する病名がある場合、“DPC 病名 6 種”と “DPC 病名 3 種 “の 2 データが存在することとなる
- DPC 患者情報（様式 1）
急性膵炎に特異的な項目である「急性膵炎の重症度分類」を用いた
- レセプト診療行為、管理料、手術情報
 - ◆ レセプト電算コード単位に 2 項（有無）のデータとした
 - ◆ 先行のアウトカム定義の検討※結果に基づいて特定器材は使用しなかった
- 処方オーダ・注射実施情報
 - ◆ 処方・注射を区別せず、ATC 分類の先頭 3 桁のデータを使用した
- 検体検査情報
急性膵炎に特有の血清アミラーゼ血清リパーゼを用いた。
 - ◆ カットオフ値は先行のアウトカム定義の検討で設定した数値を採用し、先行の検討結果と臨床的観点から以下の 2 タイプを設定した

type1: データ期間（入院期間中）の最大値

type2: index date の前後 3 日間での異常値の有無 異常値=フラグ 1

(異常値：リパーゼ ≥ 50 , アミラーゼ ≥ 130)

※先行のアウトカム定義の検討：APC 定義の他に、PMDA が手動的に作成した複数のアウトカム定義について PPV を算出し、定義の妥当性について検討した

2) 機械学習の結果

検体検査値を入院期間中の最大値とした場合の機械学習の結果を図 1 に示す（パターン 1）。その結果、3 施設に共通で、かつ急性膵炎に特有なレセプト診療行為、管理料、手術情報が上位重要変数として挙げられなかった。そのため、診療行為を除いた場合（図 2. パターン 2）と、検体検査値を入院日の前後 3 日間における異常値の有無とし、かつ診療行為を除いた場合についてそれぞれ機械学習を実施した（図 3. パターン 3）。3 パターンのいずれも AUC が 0.8 以上と高かった。先行のアウトカム定義の検討結果と急性膵炎で入院となる場合は入院日または直近の膵酵素値が最高（最悪）であるケースが多いという臨床的観点を考慮して、最終的にパターン 3 の機械学習の結果を基に改良型アウトカム定義に用いる項目を選定した。

図 1. [パターン 1] 検体検査値を入院期間中の最大値とした場合

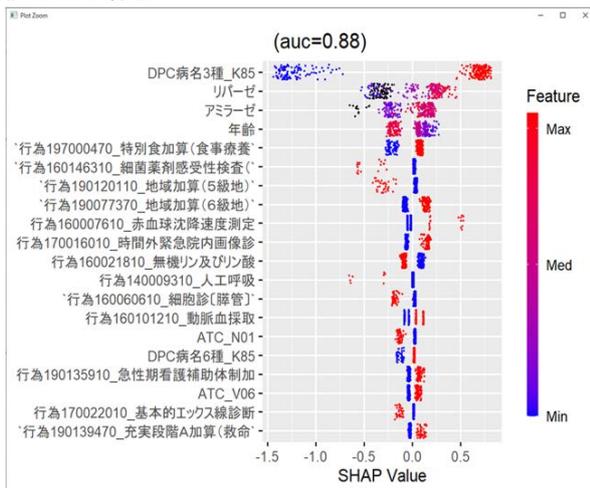


図 2. [パターン 2] 検体検査値が入院期間中の最大値で、かつ診療行為を除いた場合

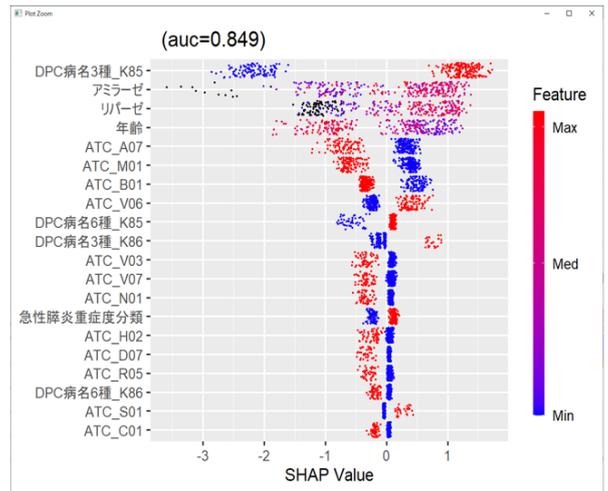
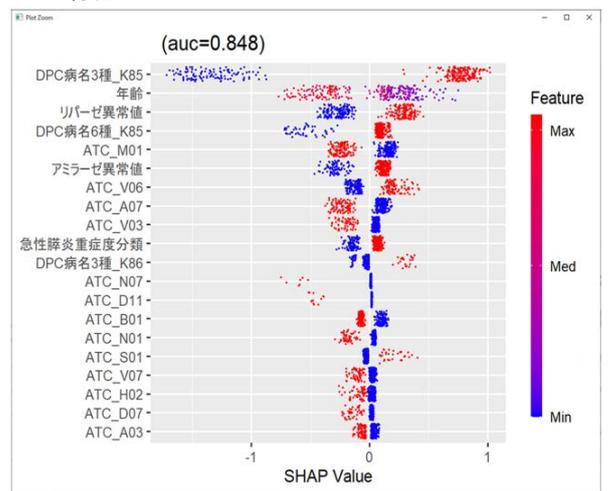


図 3. [パターン 3] 検体検査値を入院日の前後 3 日間における異常値の有無とし、かつ診療行為を除いた場合



パターン 3 の上位重要変数に対する検討結果および採否の関する事項を以下に記す。

➤ DPC 傷病情報：

DPC 病名 3 種_K85 が最上位で、かつ DPC 病名 6 種より上位であったため、ICD10 先頭 3 桁が K85 と K86 の DPC 病名 3 種を採用した

➤ 急性膵炎の重症度分類：

上位重要変数としても抽出され、かつ急性膵炎に特有の項目であることから採用した。

➤ 血清アミラーゼおよびリパーゼ：

パターン 3（検体検査値を入院日の前後 3 日間における異常値の有無）を採用した

➤ 処方オーダ・注射実施情報

図 3 のように、ATC_V06 と ATC_S01 以外は偽症例に多く処方または実施されている傾向があった。ATC_V06（経腸成分栄養剤）については急性膵炎診療ガイドライン 1 で重要急性膵炎（ただし腸管壊死等の合併症がない場合を除く）に対して感染予防のため早期の経腸栄養を

推奨しており、急性膵炎との関連性が大きい変数であったが、ATC_S01（点眼薬）については膵炎との関連性が見受けられなかった。

補足) レセプト診療行為、管理料、手術情報については、施設に共通した項目や急性膵炎に関連する項目がないため不採用とした。

3) 改良型アウトカム定義の検討結果

改良型アウトカム定義は、APC 定義に 2) で決定した重要変数（データ項目）を条件として追加する形で改良した。改良型アウトカム定義に用いたデータ項目を表 2 に示す。

表 2 改良型アウトカム定義に用いたデータ項目および時系列条件

選定項目※	時系列条件
DPC病名3種_K86	入院日～退院日まで
ATC_V06DE	入院日～退院日まで
急性膵炎_重症度分類	入院日～退院日まで
リパーゼ異常値(≥50)	Index dateの前後3日間
アミラーゼ異常値(≥130)	Index dateの前後3日間

また、以下のように改良の方針を決定した。

- ✓ 「DPC 病名 3 種_K85」は膵炎の同定に大きく貢献する変数のため、単独での条件として加える
- ✓ 他の 5 個の変数（以下、選定項目）については組み合わせてスコア化する形で条件に加える

選定項目の組み合わせパターン別(改良 1～4)の PPV と感度の算出結果を図 4 に示す。

改良型アウトカム定義=APC 定義 and 改良 1～4

改良1 DPC病名3種_K85

	真	偽	計	PPV	F値
positive	170	20	190	89.47%	85.86
negative	36	74	110		
計	206	94	300		

感度 82.52%

改良2 (DPC病名3種_K85 or 選定項目2点以上)

	真	偽	計	PPV	F値
positive	190	48	238	79.83%	85.59
negative	16	46	62		
計	206	94	300		

感度 92.23%

改良3 (DPC病名3種_K85 or 選定項目3点以上)

	真	偽	計	PPV	F値
positive	182	26	208	87.50%	87.92
negative	24	68	92		
計	206	94	300		

感度 88.35%

改良4 (DPC病名3種_K85 or 選定項目4点以上)

	真	偽	計	PPV	F値
positive	174	21	195	89.23%	86.78
negative	32	73	105		
計	206	94	300		

感度 84.47%

図 4. 選定項目の組み合わせパターン別の PPV と感度の算出結果

改良 1～4 のうち、DPC 病名 3 種_K85 を加えた条件(改良 1)と DPC 病名 3 種_K85 単独または残り 4 項目のうち 3 項目以上該当する条件(改良 3)が PPV と感度のいずれも 80%以上であり、かつ双方のバランスがよく、改良 1 においては最もシンプルな条件であったため、これら 2 つを改良型アウトカム定義に決定した。

改良型アウトカム定義 1 [改良 1]

PPV=89.5%, 感度=82.5%

[病名 1:DPC 膵炎関連 (6 種)] 時系列 and [医薬品 1: 蛋白分解酵素阻害薬_注射] 時系列 not [処置 1: ERCP、膵手術等] 時系列 and [病名 : DPC 膵炎関連 (3 種) _K85]

↓

上記の条件式を簡潔にすると、

[病名 : DPC 膵炎関連 (3 種) _K85] 時系列 and [医薬品 1: 蛋白分解酵素阻害薬_注射] 時系列 not [処置 1: ERCP、膵手術等]

改良型アウトカム定義 2 [改良 3]

PPV=87.5%, 感度=88.4%

[病名 1:DPC 膵炎関連 (6 種)] 時系列 not [医薬品 1: 蛋白分解酵素阻害薬_注射] 時系列 *not [処置 1: ERCP、膵手術等] 時系列 and ([病名 : DPC 膵炎関連 (3 種) _K85] or [選定項目(表 2) の合計が 3 点以上])

(参考) APC 定義

PPV=68.7%, 感度=100%

[病名 1:DPC 膵炎関連 (6 種)] 時系列 and [医

薬品 1: 蛋白分解酵素阻害薬_注射] 時系列 not
[処置 1: ERCP、膵手術等]

4) 各施設の改良型アウトカム定義の適用結果

改良型アウトカム定義 1 と 2 を、各施設の判定対象 possible case にそれぞれ適用した結果を図 5,6 に示す。香川大学医学部附属病院および佐賀大学医学部附属病院の PPV や感度は、3) で算出した数値と大きく変わりがなかったが、本院のデータに適用した結果、PPV・感度ともに 70% 台まで低下した。

香川		真	偽	計	PPV 87.50%	F値 86.82
	positive	56	8	64		
	negative	9	27	36		
	計	65	35	100		
		感度 86.15%				
九大		真	偽	計	PPV 81.25%	F値 76.47
	positive	39	9	48		
	negative	15	37	52		
	計	54	46	100		
		感度 72.22%				
佐賀		真	偽	計	PPV 96.15%	F値 90.91
	positive	75	3	78		
	negative	12	10	22		
	計	87	13	100		
		感度 86.21%				
<参考>						
九州	213例全数に適用				PPV 75.68%	F値 76.36
	positive	84	27	111		
	negative	25	77	102		
	計	109	104	213		
		感度 77.06%				

図 5. 改良型アウトカム定義 1 [改良 1] の適用結果

香川		真	偽	計	PPV 85.51%	F値 88.06
	positive	59	10	69		
	negative	6	25	31		
	計	65	35	100		
		感度 90.77%				
九州		真	偽	計	PPV 76.79%	F値 78.18
	positive	43	13	56		
	negative	11	33	44		
	計	54	46	100		
		感度 79.63%				
佐賀		真	偽	計	PPV 96.39%	F値 94.12
	positive	80	3	83		
	negative	7	10	17		
	計	87	13	100		
		感度 91.95%				
<参考>						
九州	213例全数に適用				PPV 70.77%	F値 76.99
	positive	92	38	130		
	negative	17	66	83		
	計	109	104	213		
		感度 84.40%				

図 6. 改良型アウトカム定義 2 [改良 3] の適用結果

D. 考察

本研究では、全拠点でを使用すること、薬剤性

の有害事象の検討に使用することを想定して、PPV 及び感度 80% 程度で、3 施設統合した一つの改良型アウトカム定義の作成を検討した。

そして、APC 定義に DPC 病名 3 種_K85 を単独で加えたもの (改良 1) と、APC 定義に「DPC 病名の主病名～医療資源 1 のいずれかに K85 を含む」の単独条件または DPC 病名の主病名～医療資源 1 のいずれかに K86 を含む、「経腸成分栄養剤 (ATC_V06DE)」、「血清アミラーゼおよびリパーゼの異常値の有無」、「急性膵炎の重症度分類」のうち 3 項目に該当する条件を加えた 2 通りの改良型アウトカム定義を作成した。

これら 2 種類の改良型アウトカム定義は、いずれも PPV・感度ともに 80% 程度であり目標が達成された。また、これらの定義を実施施設に適用しても、本院では PPV や感度が若干落ちるものの、他 2 施設の PPV や感度は 80% 以上と良好であり、全体的に見ておおむね目標 (PPV・感度 80% 以上の改良型アウトカム定義) は達成できていたと考えられる。

E. 結論

本検討では、3 施設の MID-NET 個票データを合わせた一つのデータセットを用いて機械学習を実施し、抽出された重要変数を抽出条件として追加することで 2 種類の改良型アウトカム定義を作成した。これらを各施設へ適用した結果、PPV・感度 80% 以上の改良型アウトカム定義の目標をほぼ達成できたと考えられる。

F. 健康危険情報：

(分担研究報告書では記入不要です)

G. 研究発表：

各機関における成果をご記入ください
(発表誌名巻号・頁・発行年等も記入)

1. 論文発表

中島 直樹, 奥井 佑, 野尻 千夏, 伊豆倉 理江子, 診療情報に基づく ePhenotyping と機械学習法, 医薬品医療機器レギュラトリーサイエンス, 53, 6, 460-466, 2022.06.

2. 学会発表

佐藤 直市, 伊豆倉 理江子, 相良 理香子, 神田 橋 忠, 平田 明恵, 安西 慶三, 野村 浩子, 宇山 佳明, 小川 佳宏, 中島 直樹, 大規模臨床データベースにおける糖尿病性ケトアシドーシス抽出アルゴリズムの作成, 第 60 回日本糖尿病学会九州地方会, 2022.10.

H. 知的財産権の出願・登録状況

(予定を含む。)

特記なし

1. 特許取得

特記なし

2. 実用新案登録

特記なし

3. その他