

令和3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。

・**糖尿病性ケトアシドーシス（令和2年度より継続）：**

前年度 APC 定義の検討を実施した 3 施設の MID-NET 個票データを用いて改良型アウトカム定義の作成を試みた。機械学習法より抽出された DKA の特定に重要な変数のうち、pH と HCO₃⁻ を追加条件に加えた 2 通りを改良型アウトカム定義案とし、これらを各実施施設の個票データに適用した。その結果、施設間で PPV や感度にバラつきが見受けられ、症例数や施設背景が影響している可能性が高かった。今後、全拠点で利用するなアウトカム定義の作成において考慮しておく必要がある。

・**悪性腫瘍（令和2年度より継続）：**

肺癌、悪性リンパ腫、全癌において改良型アウトカム定義の作成を試みた。令和2年度実施した初期ルールの検討結果をもとに病名抽出源別でPPVや感度を算出した結果、SS-MIX疑い病名でPPVが1%未満、レセプト病名で25%未満と非常に低く、改良型アウトカム定義ではこれらを除外した。機械学習法により抽出された重要変数のうち、DPC様式1の「がん患者/初発・再発（DPCコード：CAN0010）」が、3種に共通して真症例の特定に最も有効な変数（寄与率50%以上）であった。この変数と傷病名を組み合わせた条件において、PPVや感度が比較的高くバランスが良かったため、改良型アウトカム定義案として決定した。今後は、複数拠点において定義の妥当性を検討していく必要がある。

・**急性膵炎（令和2年度より継続）：**

前年度に作成したAPC定義候補から抽出された213例に対して専門医による真偽判定を実施し、PPVが50%以上で、かつ真症例を100例以上抽出可能なAPC定義かどうかを検討した。その結果、外科的操作に伴う膵酵素上昇又は膵炎を除外する条件の見直しや、膵酵素上昇のない偽症例の除外条件の検討等の課題はあるものの、当初の目標であるPPV 50%以上で真症例100例以上を確認できるAPC定義である可能性が高いAPC定義候補を見出した。今後は複数拠点においてこれらの課題を含め、定義の妥当性について検討していく必要がある。

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少ない。「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、感度を100%とする All possible cases (以下 APC) の定義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。なお、本年度検討を実施したテーマは令和2年度から継続して実施しているものである。

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を複数選定した上で、各アウトカムについて、従来法又は機械学習の手法を取り入れて作成した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。

研究の流れは以下のとおりである。

B-1. 糖尿病性ケトアシドーシス (以下、DKA) (令和2年度より継続)

B-1.1) 複数拠点で検討対象とする APC 定義及びアウトカム定義の検討

B-1.2) 評価基準の作成

B-1.3) APC 定義より抽出された対象アウトカムについて、カルテレビューによる真症例の特定と陽性的中度 (以下「PPV」という) の算出および評価

B-1.4) 機械学習手法を用いた DKA 特定に重要な変数の選定

B-1.5) 改良型アウトカム定義に対する PPV 及び感度の算出・評価

B-1.6) 改良型アウトカム定義案の決定

B-1.7) 複数医療機関の PPV の比較及び医療機

関間の差異の要因検討

B-2. 悪性腫瘍 (令和2年度より継続)

B-2.1) 初期ルールの検討

B-2.2) 評価基準の作成

B-2.3) 初期ルールより抽出された対象アウトカムについて院内がん登録を Gold Standard とした真症例の特定と PPV 及び感度の算出・評価

B-2.4) 機械学習手法を用いた悪性腫瘍の特定に重要な変数の選定

B-2.5) 改良型アウトカム定義に対する PPV 及び感度の算出・評価

B-2.6) 改良型アウトカム定義案の決定

B-2.7) 複数医療機関の PPV の比較及び医療機関間の差異の要因検討

B-3. 急性膵炎 (令和2年度より継続)

<事前検討>

B-3.1) APC 定義候補の作成

B-3.2) 評価基準の作成

B-3.3) 対象アウトカムについて、カルテレビューによる真症例の特定と PPV の算出および評価：症例数 10 例

<本検討>

B-3.4) APC 定義候補から抽出された対象アウトカムについて、カルテレビューによる真症例の特定と PPV の算出および評価：症例数 213 例

B-3.5) APC 定義案の決定

B-3.6) 複数医療機関の PPV の比較及び医療機関間の差異の要因検討

(倫理面への配慮)

医療機関が行う作業は、文部科学省・厚生労働省、人を対象とする医学系研究に関する倫理指針 (平成26年文部科学省・厚生労働省告示第3号) を遵守して行った。

C. 研究結果

C-1. 糖尿病性ケトアシドーシス

・対象の病態：入院を要する DKA

・期間：2011年4月1日～2019年3月31日

2021年度は **B-1.の4)～7)** について実施した。

<APC 定義>

前年度作成した APC 定義を表 C-1-1 に示す。

表 C-1-1. DKA の APC 定義

([病名 1 (DPC)] or [病名 2 (DPC)]) and [病名 3 (SS-MIX)]

病名 1 or 病名 3:DKA 病名

病名 2 : DKA に関連した病名 (例 : アシドーシス)

<改良型アウトカム定義の作成について>

これまで、改良型アウトカム定義は自施設データを用いて作成していたため、自施設に特異的な定義となり、一般化が難しいことが懸念されてきた。そこで今回、昨年度 APC 定義の検討を実施した九州大学病院、佐賀大学医学部付属病院および徳洲会 (グループ) の MID-NET 個票データを用いて、MID-NET 協力医療施設内としてはより汎化された改良型アウトカム定義の作成を目指した。

4) 機械学習手法を用いた DKA 特定に重要な変数の選定

今後、アウトカム定義は薬剤性有害事象の検討などで用いることが想定されるため、できるだけ真のケースのみを最終的な「真症例」と定義した改良型アウトカム定義の検討が望ましいとの見解から、真偽判定において「真のケース」と判定された症例のみを「最終的な真症例」とした場合の個票データを用いて作成したものを、DKA の改良型アウトカム定義 (案) として提示することに決定した。

4)-1 APC 定義の検討結果 (前年度実施)

3 施設中 PPV が最も高い施設は佐賀大学の 75.58% で、最も PPV が低い施設は徳洲会の 39.0% であった。3 施設のデータを合わせた場合の PPV は 54.27% であった (表 C-1-2)。

次に、徳洲会において全体の推定される真症例数を算出すると、真症例数は 384.9 ケースで 3 施設全体の PPV は 42.19% となった (表 C-1-3)。

表 C-1-2. 3 施設における PPV の算出結果

医療機関	Possible Cases 数	判定対象 Possible Case 数	真症例数	偽症例数	PPV (%)	95%信頼区間 [下限値-上限値]
3 施設全て	1121	234	127	107	54.27	[47.66-60.78]
九州大学	48	48	23	25	47.92	[33.29-62.81]
佐賀大学	86	86	65	21	75.58	[65.13-84.20]
徳洲会	987	100	39	61	39.00	[29.40-49.27]

九州大学 : 九州大学病院、佐賀大学 : 佐賀大学医学部付属病院、徳洲会 : 徳洲会 (グループ)

表 C-1-3. 症例数を調整した場合の PPV 算出結果 (施設全体の推定)

医療機関	Possible Cases 数	真症例数	偽症例数	PPV (%)	感度 (%)
3 施設全て	1121	472.93	648.07	42.19	100.00
九州大学	48	23	25	47.92	100.00
佐賀大学	86	65	21	75.58	100.00
徳洲会	987	384.93	602.07	39.00	100.0

九州大学 : 九州大学病院、佐賀大学 : 佐賀大学医学部付属病院、徳洲会 : 徳洲会 (グループ)

4)-2 データセットに用いたデータ種目

MID-NET データベースに保有されているデータ種目のうち、DKA の治療と関連性の低いレセプト/DPC 手術情報、レセプト/DPC 特定器材情報、細菌検査情報を除外した。また、医薬品情報は持参薬情報などを含む SS-MIX のみとした。

検体検査情報については、DKA 診断において重要な指標となる pH, HCO₃⁻ を選定し、材料 (動脈血 or 静脈血) や測定法については限定しなかった。なお、尿中/血清ケトン体については MID-NET 仕様上利用できないため、本検討では使用しなかった。本検討で用いたデータ種目を表 C-1-4 に示す。

- ・データ期間 : 入院日の前 30 日から退院日まで
- ・データ項目の総計 : 1971 変数

表 C-1-4. 改良型アウトカム定義に用いたデータ種目

データ種目	用いる標準コード	特記事項
傷病情報 (SS-MIX/DPC)	SS-MIX : 病名交換用コード DPC : ICD10 コード	2 項分類
診療行為情報 (DPC/レセプト)	レセプト電算コード	2 項分類
管理料情報 (DPC/レセプト)	レセプト電算コード	2 項分類
医薬品情報【処方・実施】 (SS-MIX)	YJ コード先頭 7 桁 (一般名レベル)	2 項分類
検体検査情報	JLAC10 コード	pH, HCO ₃ ⁻ : 期間内における最小値 ケトン体 : 尿中・血清とも MID-NET 仕様上、利用できなかった

計 1971

4)-3 機械学習の結果

4)-1 に示すように、3 施設全ての真偽判定を実施した症例 234 例中、真症例は 127 例、偽症例は 107 例であった。この結果を目的変数、4)-2 のデータ項目 1971 個を説明変数として機械学習を実施した。上位 20 個の重要変数に対する SHAP プロットを図 C-1-1 に示す。

結果、AUC は 0.883 と精度が高かった。上位 5 つの重要変数は、上から順に pH, HCO₃⁻, E872 (DPC 病名の ICD10 コード)、インスリンアスパルト, E101 (DPC 病名の ICD10 コード) であった。うち、E872 においては、青のプロットがプラスの方に集中しており、偽症例に多く病名付与されている傾向がみられた。

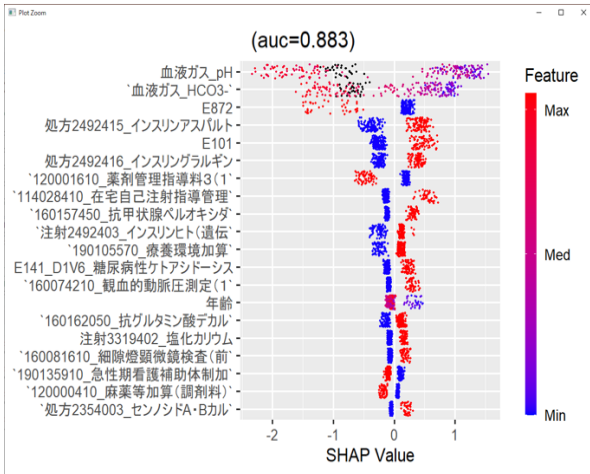


図 C-1-1. SHAP プロット
真：n=127, 偽：n=107 説明変数：1971

・ 上位重要変数：傷病名 E872 および E101 について

傷病名 E872 (DPC 病名) および E101 (DPC 病名) に対して APC 定義の判定結果と照合したものを表 4 に示す。真症例 127 例のうち 5 例において E872 があり、うち 3 例は DKA の病名も併せて持っていた。図 1 の SHAP プロットと表 C-1-5 の結果を総合して、E872 は改良型アウトカム定義において除外条件とすることとした。

表 C-1-5. 傷病名と判定結果との照合結果

	判定結果			計		判定結果			計
	真	偽	計			真	偽	計	
E872	あり	5	35	40	E101	あり	65	18	83
	なし	122	72	194		なし	62	89	151
計	127	107	234	計	127	107	234		

・ 上位重要変数：医薬品_インスリンアスパルトについて

超即効型インスリン製剤であるインスリンアスパルトを含むインスリン製剤に対して APC 定義の判定結果と照合したものを表 C-1-6 に示す。結果、インスリンアスパルトに限らず、インスリン全般において真症例に多く使用している傾向が見受けられたため、インスリンに関しては改良型アウトカム定義の条件として全種を追加することとした。

表 C-1-6. 医薬品と判定結果との照合結果

YJ7	医薬品一般名	実在した医薬品商品名	真	偽
2492415	インスリンアスパルト(遺伝子組換え)	ノボラビッド30ミックス注フレックスペン	4	1
2492415	インスリンアスパルト(遺伝子組換え)	ノボラビッド注300フレックスペン		3
2492415	インスリンアスパルト(遺伝子組換え)	ノボラビッド注フレックスタッチ	44	21
2492415	インスリンアスパルト(遺伝子組換え)	ノボラビッド注フレックスペン	39	26
2492415	インスリンアスパルト(遺伝子組換え)	ノボラビッド注ベンフィル	3	2
2492416	インスリングラルギン(遺伝子組換え)	ランタスXR注ソロスター	14	6
2492416	インスリングラルギン(遺伝子組換え)	ランタス注ソロスター	59	37
2492403	【控】インスリンヒト(遺伝子組換え)	ノボリンR注100	11	5
2492403	インスリンヒト(遺伝子組換え)	ノボリンR注100単位/mL	38	33
2492403	インスリンヒト(遺伝子組換え)	ヒューマリンR注100単位/mL	69	22

・ 検体検査情報_pH および HCO3-のカットオフ値の調整

APC 定義の検討にて作成した判定票で設定した基準値に準拠することを前提としたうえで、九州大学病院で真偽判定を実施した 48 例の検査データを用いてカットオフ値を微調整した。

具体的には、データ期間における最小値を昇順に配列し (図 C-1-2)、真のケースの分布を目視で確認した。結果、pH を 7.31 未満、HCO3- は判定票の基準値とした。

pH	判定結果	HCO3-	判定結果
7.293	その他	16.4	真
7.296	真	16.7	その他
7.297	真	16.8	真
7.3	その他	17	疑わしいA
7.3	真	17.1	真
7.304	疑わしいB	17.1	真
7.305	真	17.1	真
7.305	真	17.2	疑わしいA
7.306	真	17.3	疑わしいA
7.306	真	17.6	疑わしいA
7.308	真	17.8	疑わしいA
7.309	真	17.8	真
7.31	疑わしいA	18	疑わしいA
7.316	疑わしいA	18	その他
7.317	疑わしいA	18.1	その他
7.32	その他	18.5	真
7.322	疑わしいA	18.5	疑わしいB
7.325	その他	18.6	疑わしいA
7.325	その他	18.8	真
7.327	その他	19	真
7.329	疑わしいA	19.2	真
7.329	真	19.2	その他
7.329	疑わしいA	19.2	疑わしいB
7.329	疑わしいA	19.2	その他

図 C-1-2. 48 例の pH、HCO3-の分布：抜粋

4) 改良型アウトカム定義に用いる重要変数と条件

改良型アウトカム定義の追加条件として用いる重要変数と条件を表 C-1-7 に示す。

表 C-1-7. 改良型アウトカム定義に用いる重要変数と条件

重要変数	条件
pH	< 7.310
HCO3-	< 18.0
E872	除外条件として設定：以下の傷病名コードを除外 DPC: E872 SS-MIX : E872 に紐づく病名交換コード (「EU8R : ケトアシドーシス」、「PEJS : 乳児ケトアシドーシス」)
E101	-
インスリンアスパルト	インスリン全般を採用

5) 改良型アウトカム定義に対する PPV 及び感度の算出・評価

APC 定義から 4) 4 で設定した E872 および E872 に紐づく病名交換用コードを除外したものの (以下、除外後病名) : (1) の PPV と感度はそれぞれ 61.8% と 96.9% であり、APC 定義に比べ PPV がわずかに上昇した。(1) に pH と HCO3- を組み合わせると、2 つの検査値を OR 条件とし

た場合(2)の PPV と感度はそれぞれ 88.1%・75.6%となり、AND 条件の場合 (3)では PPV が 80.0%、感度は 81.9%となった。図 C-1-3 に示す 5通りの定義案のうち、(3)の定義案において PPV vs 感度のバランスが良かった。一方で、(1)にインスリンを組み合わせた場合 (1)' の PPV と感度はそれぞれ 64.0%・95.3%となり、(1)の結果とほぼ変わらなかった。(1)' に pH と HCO₃⁻を AND 条件で組み合わせた場合 (2)' の PPV と感度は 89.5%・74.0%で、(2)の結果とほぼ変わらなかった。

(1) 除外後病名がある				(1)' 除外後病名 and インスリン				
	真陽判定	真陰判定	計		真陽判定	真陰判定	計	
追加条件	positive	123	76	199	positive	121	68	189
	negative	4	31	35	negative	6	39	45
	計	127	107	234	計	127	107	234
	感度	96.9%		PPV	61.8%		F値	0.75
				感度	95.3%			
(2) 除外後病名 and pH and HCO₃⁻				(2)' 除外後病名 and pH and HCO₃⁻ and インスリン				
	真陽判定	真陰判定	計		真陽判定	真陰判定	計	
追加条件	positive	96	13	109	positive	94	71	165
	negative	31	94	125	negative	33	96	129
	計	127	107	234	計	127	107	234
	感度	75.6%		PPV	89.5%		F値	0.81
				感度	74.0%			
(3) 除外後病名 and (pH or HCO₃⁻)				(3)' 除外後病名 and (pH or HCO₃⁻) and インスリン				
	真陽判定	真陰判定	計		真陽判定	真陰判定	計	
追加条件	positive	104	26	130	positive	104	26	130
	negative	23	81	104	negative	23	81	104
	計	127	107	234	計	127	107	234
	感度	81.9%		PPV	80.0%		F値	0.81
				感度	81.9%			

図 C-1-3. 条件の組み合わせにおける PPV・感度および F 値の算出結果

6) 改良型アウトカム定義案の決定

5)の検討結果より、図 3 の(2) と (3) の 2 通りを改良型アウトカム定義 (案) に決定した (図 C-1-4)。※案 1: 図 C-1-3 の (2)、案 2: 図 C-1-3 の (3)

案 1) PPV 重視型: PPV の上昇に着目した定義 (PPV 88.1%、感度 75.6%) ((DPC 病名 1]時系列 and ([SS-MIX 病名 3]not 除外後病名))and pH and HCO ₃ ⁻ 除外後病名。
案 2) バランス型: PPV が高く、かつ感度の低下が小さい定義 (PPV 80.0%、感度 81.9%) ((DPC 病名 1]時系列 and ([SS-MIX 病名 3]not 除外後病名))and (pH or HCO ₃ ⁻) 除外後病名。
条件項目>> DPC 病名 1: DKA に関連した DPC 病名 (APC 定義の条件項目と同じ)。 SS-MIX 病名 3: DKA に関連した SSMIX 病名 (APC 定義の条件項目と同じ) 除外後病名: DPC_E872 (病名 2)。 SSMIX_E872 に紐づく「EUSR: ケトアシドーシス」。 「PEJ5: 乳児ケトアシドーシス」。 血液ガス: pH <7.310, HCO ₃ ⁻ <18。

図 C-1-4. 改良型アウトカム定義案

7) 複数医療機関の PPV の比較及び医療機関間の差異の要因検討

7)-1 改良型アウトカム定義案 1 の適用結果

定義案 1 を適用した場合の各施設の PPV と感度を表 C-1-8 に、3 施設全体の推定真症例数、PPV 及び感度を表 C-1-9 に示す。

PPV は九州大学で 85.72%、佐賀大学で 93.34%、徳洲会では 80.0%と、3 施設ともに 80%以上の高さであった。一方、感度においては、3 施設のうち九州大学が 52.18%と最も低かった。

3 施設中ランダムサンプリングを実施した徳洲会に対して施設全体の推定真症例数を算出すると 276.4 例と推定され、3 施設全体において推定される PPV は 82.10%、感度は 72.81%となった。

表 C-1-8. 定義案 1 の PPV と感度の算出結果

施設	APC 定義			改良型アウトカム定義案 1						
	判定対象 possible cases 数	真症例 数	PPV (%)	Positive 数	真症例 数	偽症例 数	PPV (%)	95%信頼区間 [下限値-上限値]	感度 (%)	95%信頼区間 [下限値-上限値]
3 施設全て	234	127	54.27	106	96	13	88.07	[80.47-93.49]	75.59	[67.18-82.77]
九州大学	48	23	47.92	14	12	2	85.71	[57.19-98.22]	52.17	[30.59-73.18]
佐賀大学	86	85	75.58	60	56	4	93.33	[83.80-98.15]	88.15	[75.34-93.47]
徳洲会	100	39	39.00	35	28	7	80.00	[63.06-91.56]	71.79	[55.13-85.02]

九州大学: 九州大学病院、佐賀大学: 佐賀大学医学部付属病院、徳洲会: 徳洲会 (グループ)

表 C-1-9. 症例数を調整した場合の PPV と感度の算出結果 (施設全体の推定)

施設	APC 定義			改良型アウトカム定義案 1					
	Possible Cases 数	真症例 数	偽症例数	PPV (%)	Positive 数	真症例 数	偽症例 数	PPV (%)	感度 (%)
3 施設全て	1121	472.5	107	42.19	419.45	344.36	75.09	82.10	72.81
九州大学	48	23	25	47.92	14	12	2	85.71	52.17
佐賀大学	86	65	21	75.58	60	56	4	93.33	86.15
徳洲会	987	384.5	602.5	39.00	345.45	276.36	69.09	80.00	71.79

九州大学: 九州大学病院、佐賀大学: 佐賀大学医学部付属病院、徳洲会: 徳洲会 (グループ)

7)-2 改良型アウトカム定義案 2 の適用結果

定義案 2 を適用した場合の各施設の PPV と感度を表 C-1-10 に、3 施設全体の推定真症例数、PPV 及び感度を表 C-1-11 に示す。

3 施設のうち、佐賀大学において PPV が 92.19%、感度が 90.77 と、PPV vs 感度のバランスがよかった。九州大学では PPV が 72.22%に対して感度は 56.52%まで低下した。一方で、徳洲会では PPV が 66.67%に対して感度は 82.05%と高かった。徳洲会に対して施設全体の推定真症例数を算出すると 315.8 ケースと推定され、3 施設全体において推定される PPV は 69.79%、感度は 82.01%となった。

表 C-1-10. 定義案 1 の PPV と感度の算出結果

施設	APC 定義			改良型アウトカム定義案 2						
	判定対象 possible cases 数	真症例 数	PPV (%)	Positive 数	真症例 数	偽症例 数	PPV (%)	95%信頼区間 [下限値-上限値]	感度 (%)	95%信頼区間 [下限値-上限値]
3 施設全て	234	127	54.27	130	104	26	80.00	[72.08-88.50]	81.89	[74.09-88.16]
九州大学	48	23	47.92	18	13	5	72.22	[46.52-90.31]	56.52	[34.49-76.81]
佐賀大学	86	85	75.58	64	59	5	92.19	[82.70-97.41]	90.77	[80.99-96.54]
徳洲会	100	39	39.00	48	32	16	66.67	[51.59-79.62]	82.05	[66.47-92.46]

表 C-1-11. 症例数を調整した場合の PPV と感度の算出結果（施設全体の推定）

施設	APC 定義				改良型アウトカム定義案 2				感度 (%)
	Possible Cases 数	真症例数	偽症例数	PPV (%)	Positive 数	真症例数	偽症例数	PPV (%)	
3 施設全て	1121	472.93	648.07	42.19	555.76	387.84	167.92	69.79	82.01
九州大学	48	23	25	47.92	13	5	5	72.22	56.52
佐賀大学	86	65	21	75.58	64	59	5	92.19	90.77
徳洲会	987	384.93	602.7	39.00	473.76	315.84	157.92	66.67	82.05

C-2. 悪性腫瘍

- 対象の病態：初回治療を実施した原発性悪性腫瘍（2 次性や転移、再発を除く）
- 期間：2016 年 1 月 1 日～2018 年 12 月 31 日

対象となる癌種は以下の通りである。

- 肺癌
- 悪性リンパ腫
- 全癌

2021 年度は B-3. の 4)～6) について実施した。

<初期ルール>

前年度作成した初期ルールを C-2-1 に示す。

表 C-2-1. 悪性腫瘍の初期ルールの定義

[病名 (SS-MIX2)] OR [病名 (レセプト)] OR [病名 (DPC)]

※病名：悪性腫瘍に関連する病名

4) 機械学習手法を用いた悪性腫瘍の特定に重要な変数の選定

DKA のケースと同様に、アウトカム定義の実用性を考慮し、真のケースのみを最終的な「真症例」とした場合の個票データを用いて作成した改良型アウトカム定義案の検討結果を記載する。

4)-1 初期ルールの検討結果（前年度実施）

初期ルールの検討において、真偽判定を実施した症例の PPV 算出結果を表 C-2-2 に示す。

PPV は肺癌 13.1%、悪性リンパ腫 7.7% と非常に低く、全癌では 21.6% であった。

表 C-2-2. 初期ルールの PPV 算出結果

癌種	初期ルール	真症例数	偽症例数	PPV	がん登録ケース数	真症例数	がん登録のみに存在	感度
肺癌	6345	831	5514	13.1%	833	831	2	99.8%
悪性リンパ腫	4589	352	4237	7.7%	352	352	0	100.0%
全癌	39847	8615	31232	21.6%	8782	8615	167	98.1%

4)-2 機械学習に用いるデータ種別/項目の選定

病名抽出源別に PPV と感度を算出した結果を表図 C-2-1 に示す。病名抽出源のうち、SS-MIX 疑い病名のみでは、肺癌、悪性リンパ

腫、全癌のいずれも PPV 及び感度が 1% 未満と非常に低かった。

肺癌									
病名抽出源	MID-NET 抽出症例数	真症例数	偽症例数	PPV	がん登録症例数	真症例数	がん登録のみに存在	感度	
初期ルール(SS-MIX or DPC or レセプト)	6345	831	5514	13.1%	833	831	2	99.8%	
SS-MIX 確定	1903	816	1087	42.9%	833	816	17	98.0%	
DPC	1578	811	767	51.4%	833	811	22	97.4%	
SS-MIX 疑いのみ	4099	0	4099	0.0%	833	0	833	0.0%	
レセプト	5804	806	4998	13.9%	833	806	27	96.8%	
SS-MIX 確定 or DPC	2197	830	1367	37.8%	833	830	3	99.8%	
※SS-MIX 疑いのみ：SS-MIX 確定病名も DPC 病名も無い									
悪性リンパ腫									
病名抽出源	MID-NET 抽出症例数	真症例数	偽症例数	PPV	がん登録症例数	真症例数	がん登録のみに存在	感度	
初期ルール(SS-MIX or DPC or レセプト)	4589	352	4237	7.7%	352	352	0	100.0%	
SS-MIX 確定	1553	348	1205	22.4%	352	348	4	98.9%	
DPC	762	281	481	36.9%	352	281	71	79.8%	
SS-MIX 疑いのみ	2737	3	2734	0.1%	352	3	349	0.9%	
レセプト	4171	341	3830	8.2%	352	341	11	96.9%	
SS-MIX 確定 or DPC	1748	349	1397	20.0%	352	349	3	99.1%	
全癌									
病名抽出源	MID-NET 抽出症例数	真症例数	偽症例数	PPV	がん登録症例数	真症例数	がん登録のみに存在	感度	
初期ルール(SS-MIX or DPC or レセプト)	39847	8615	31232	21.6%	8782	8615	167	98.1%	
SS-MIX 確定	17160	8452	8708	49.3%	8782	8452	330	96.2%	
DPC	13025	7439	5586	57.1%	8782	7439	1343	84.7%	
SS-MIX 疑いのみ	19876	74	19802	0.4%	8782	74	8708	0.8%	
レセプト	35479	8390	27089	23.6%	8782	8390	392	95.5%	
SS-MIX 確定 or DPC	18212	8526	10686	44.4%	8782	8526	256	97.1%	

図 C-2-1. 病名抽出源別の PPV・感度算出結果

4)-3 データセットに用いたデータ種目

4)-1 の結果から、病名については、SS-MIX の確定病名と DPC 病名を採用した。また、医師との協議にて MID-NET データベースに保有されているデータ種目のうち、検体検査情報を除外した。データ期間については、Index date から半年以内には対象の癌種に対する治療を何かしら実施する可能性が高いことから、末日を index date+180 日に設定した。

- データ期間：index date（入院日/病名開始日）～index date+180 日
- データ項目の総計：肺癌：3150 変数
悪性リンパ腫：3451 変数
全癌：5312 変数

機械学習に用いたデータ種目を表 C-2-3 に示す。

表 C-2-3. 改良型アウトカム定義に用いたデータ種目

データ種別	標準コード	備考
病名 (SS-MIX2/DPC)	SS-MIX2：ICD10コード DPC：ICD10コード	
診療行為情報 (DPC/レセプト)	診療行為コード	
管理料 (DPC/レセプト)	診療行為コード	
処方依頼・注射実施 (SS-MIX2)	YJコード先頭7桁 (一般名レベル)	
DPC患者情報 (DPC様式1：悪性腫瘍の関連項目)		がんの初発・再発、 UTIC病期分類、喫煙情

また、本検討では、対象の癌種別に真偽判定を実施した症例の 7 割を学習用、3 割を改良型アウトカム定義の検証用に分けて機械学習および改良型アウトカム定義の検討を実施した。

4)-4 機械学習の結果

4)-4-1 機械学習 1 回目

・肺癌

AUC は 0.966 で「がんの初発」が最も貢献度の高い変数であった。

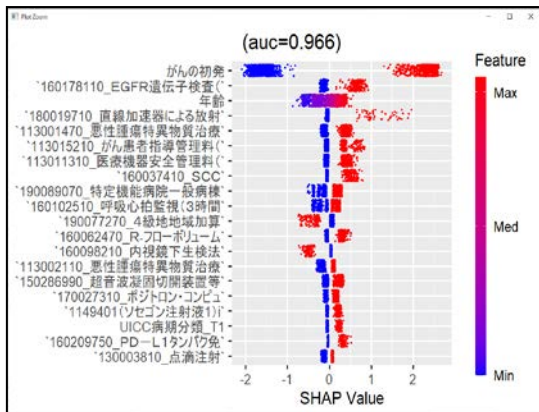


図 C-2-2. 肺癌 SHAP プロット

・悪性リンパ腫

AUC は 0.942 で、肺癌と同様「がんの初発」が最も貢献度の高い変数であった。

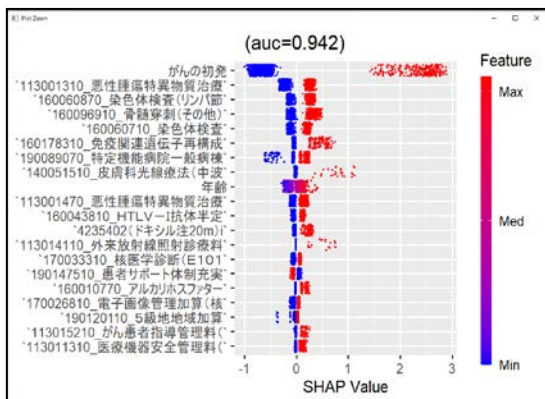


図 C-2-3. 悪性リンパ腫 SHAP プロット

・全癌

AUC は 0.948 で「がんの初発」が最も貢献度の高い変数であった。

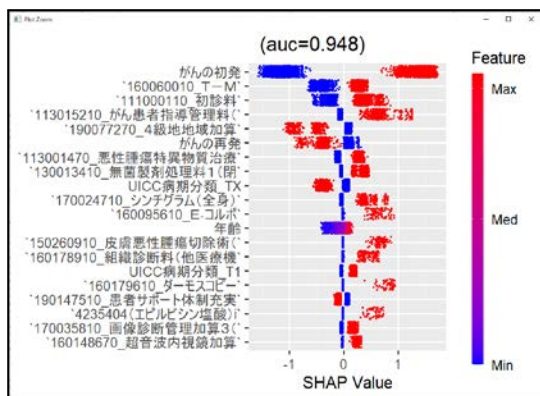


図 C-2-4. 全癌 SHAP プロット

4)-4-2 機械学習 2 回目

他に改良型アウトカム定義に用いる重要変数を探索すべく、4)-4-1 にて最も貢献度の高い重要変数である「がんの初発」と、本院の病院形態を示す算定項目（特定機能病院やがん拠点病院等）を除いたうえで再度機械学習を実施した。

・肺癌

AUC は 0.947 で、重要変数の中に加算や施設基準に関するレセプト情報が複数みられた。

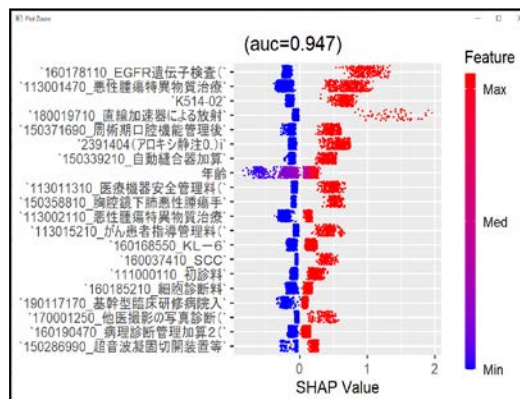


図 C-2-5. 肺癌 SHAP プロット

・悪性リンパ腫

AUC は 0.921 で、重要変数の中に加算に関するレセプト情報が複数みられた。

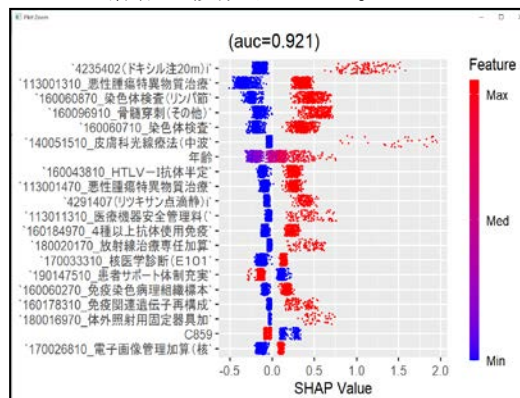


図 C-2-6. 悪性リンパ腫 SHAP プロット

・全癌

AUC は 0.921 と高く、重要変数の中に加算や施設基準に関するレセプト情報が複数みられた。また、抽出された診療行為については本院において症例数の多い癌種に影響されている傾向があった。

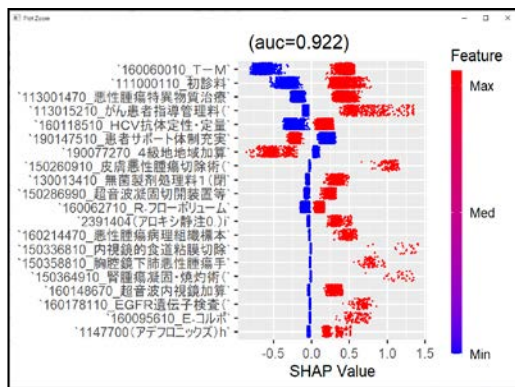


図 C-2-7. 全癌 SHAP プロット

4)-5 改良型アウトカム定義に用いる抽出条件の設定

4)-5-1 定義に用いる重要変数の選定

改良型アウトカム定義に用いる重要変数の選定は以下の手順で実施した。

- ① 4)-4-2 の各癌種の SHAP プロット図に表示された上位 20 個の重要変数を医師と協議の下で目的別（確定診断のための検査、スクリーニング目的、治療行為、等）に分類した
- ② 加算の要件は変更されることが多く、施設基準においては施設基準の厳格化から施設によって算定にばらつきが生じることが予想されたため、加算をなるべく定義には使用せず、代替として加算の元になる診療行為を用いた
- ③ 単一の診療行為が上位変数であった場合、関連する診療行為コードを全て含めて一つの条件とした

なお、全癌においては、件数の多い癌種の影響が強く出ているため、上位 5 位までの重要変数のうち、癌に関連のある「T-M (組織切片)」、「悪性腫瘍特異物質治療管理料」および「がん患者指導管理料」を採用した。

肺癌と悪性リンパ腫の上位重要変数の分類と採否の結果をそれぞれ表 C-2-4 と C-2-5 に示す。

表 C-2-4. 肺癌の上位重要変数と採否結果

順位	コード	名称	分類	採否
1.	160178110.	E G F R 遺伝子検査 (リアルタイム PCR 法以外)	確定診断	○
2.	113001470.	悪性腫瘍特異物質治療管理料 (その他・1 項目初回)	加算 (非特異的)	△
3.	K514-02.	胸腔鏡下肺悪性腫瘍手術	治療 (種類・進行度特異的)	○
4.	180019710.	直線加速器による放射線治療 (定位放射線治療)	治療 (種類・進行度特異的)	○
5.	150371690.	周術期口腔機能管理後手術加算	加算 (非特異的)	△
6.	2391404.	アロキシ筋注 0.75 mg 5 mL	治療 (種類・進行度特異的)	○
7.	150339210.	自動吻合器加算	治療 (種類・進行度特異的)	△
8.		年齢	患者属性 (非特異的)	△
9.	113011310.	医療機器安全管理料 (放射線治療計画確定)	治療 (種類・進行度特異的)	△
10.	150358810.	胸腔鏡下肺悪性腫瘍手術 (肺葉切除又は 1 肺葉を超えるもの)	治療 (種類・進行度特異的)	○Kコードで指定
11.	113002110.	悪性腫瘍特異物質治療管理料 (その他・2 項目以上)	加算 (非特異的)	△
12.	113015210.	がん患者指導管理料 (医師と看護師の共同診療方針等を文書等で提供)	加算 (非特異的)	△
13.	160168550.	KL-6	スクリーニング	△
14.	160037410.	SCC	確定診断	△
15.	111000110.	初診料	患者属性 (非特異的)	△
16.	160185210.	細胞診料	確定診断	△
17.	190117170.	基幹型臨床研修病院内院加算	加算 (非特異的)	△
18.	170001250.	他院撮影の写真診断 (単純撮影・イ)	確定診断	△
19.	160190470.	病理診断管理加算 2 (細胞診)	確定診断	△
20.	150286990.	超音波凝固切開装置等加算	治療 (種類・進行度特異的)	△

表 C-2-5. 悪性リンパ腫の上位重要変数と採否結果

順位	コード	名称	分類	採否
1.	4235402.	ドキシル注 20mg10mL	治療 (特異的)	○
2.	113001310.	悪性腫瘍特異物質治療管理料 (その他・1 項目)	加算 (非特異的)	△
3.	160060870.	分染法加算 (染色体検査)	確定診断	○
4.	160096910.	骨髄穿刺 (その他) (検査)	確定診断	○
5.	160060710.	染色体検査	確定診断	△
6.	140051510.	皮膚科光線療法 (中波紫外線、308nm以上 313nm以下)	治療 (特異的)	○
7.		年齢	患者属性 (非特異的)	△
8.	160043810.	HTLV-1 抗体半定量	スクリーニング	△
9.	113001470.	腫瘍マーカー検査初回月加算	確定診断	△
10.	4291407.	リツキサン点滴静注	治療 (特異的)	○
11.	113011310.	医療機器安全管理料 (放射線治療計画確定)	治療 (特異的)	△
12.	160184970.	4 種以上抗体使用免疫染色加算	確定診断	△
13.	180020170.	放射線治療専任加算 (放射線治療管理料)	治療 (特異的)	△
14.	170033310.	核医学診断 (E101-2~E101-5)		
15.	190147510.	患者サポート体制充実加算	加算 (非特異的)	△
16.	160060270.	免疫染色病理組織標本作製 (その他)	確定診断	○
17.	160178310.	免疫関連遺伝子再構成	確定診断	○
18.	180016970.	体外照射用固定器具加算	治療 (特異的)	△
19.	C859.	非ホジキン<non-Hodgkin>リンパ腫、詳細不明	病名 (部分的)	△
20.	170026810.	電子画像管理加算 (核医学診断科)	進行度診断	△

4)-5-2 改良型アウトカム定義に用いる抽出条件

改良型アウトカム定義に用いる抽出条件を表 C-2-6 に示す。4)-4-1 の結果より「がんの初発」は 3 種共通して最も貢献度の高い変数のため追加条件として採用した。肺癌と悪性リンパ腫においては「検査」「治療」に大別し、そのうち各項目において該当があれば 1 項目 1 点とし、その組み合わせや点数に応じて Positive、Negative と判定する方法を用いた。

表 C-2-6. 改良型アウトカム定義に用いる抽出条件

癌種	項目分類	項目	説明
共通	病名	SS-MIX2 確定病名と DPC 病名	。
	がんの初発	DPC 様式 1「がんの初発、再発」	腫瘍・悪性リンパ腫 「医療資源を最も投入した傷病名」が肺癌、悪性リンパ腫それぞれに対応する初発ルール ICD10 コードである場合、該当あり(初発)とする。 全癌 制限なし。
肺癌	検査	① E G F R 遺伝子検査	。
	治療	① 肺悪性腫瘍手術 (Kコード: K514-00, K514-02)	肺悪性腫瘍の術式全てのコードを含める。
		② 直線加速器による放射線治療	放射線治療に関するコードを含める。
	③ アロキサン静注	。	
悪性リンパ腫	検査	① 骨髄穿刺(その他)	。
		② 染色体検査	。
		③ 免疫染色病理組織標本作製(その他)	。
		④ 免疫関連遺伝子再構成	。
治療	① 皮膚科光線療法(中波紫外線、308nm以上313nm以下)	。	
	② ドキシル注	。	
	③ リツキサン点滴静注	。	
	④ 体外照射(高エネルギー放射線治療)	体外照射に関するコードを全て含める。	
全癌	病理検査	① T-M(組織切片)	。
	管理科	① 悪性腫瘍特異物質治療管理科 ② がん患者指導管理科	。

5) 改良型アウトカム定義に対する PPV 及び感度の算出・評価

肺癌、悪性リンパ腫及び全癌における、追加条件の組み合わせパターン別の PPV と感度の算出結果を図 C-2-8~C-2-10 に示す。

・肺癌

肺癌病名がある and						
初発=1						
	真	偽	計	PPV	F値	
positive	536	108	644	83.2%	0.87	
negative	56	882	938			
計	592	990	1582			
ケース内感度	90.5%					
初発=1 and (検査1点以上 or 治療1点以上)						
	真	偽	計	PPV	F値	
positive	494	55	549	90.0%	0.87	
negative	98	935	1033			
計	592	990	1582			
ケース内感度	83.4%					
初発=1 and (検査1点以上 and 治療1点以上)						
	真	偽	計	PPV	F値	
positive	159	7	166	95.8%	0.42	
negative	433	983	1416			
計	592	990	1582			
ケース内感度	26.9%					
(初発=1 or (検査1点以上 and 治療1点以上))						
	真	偽	計	PPV	F値	
positive	540	117	657	82.2%	0.86	
negative	52	873	925			
計	592	990	1582			
ケース内感度	81.2%					
{(初発=1 and (検査1点以上 or 治療1点以上)) or (初発=0 and 検査1点以上 and 治療1点以上)}						
	真	偽	計	PPV	F値	
positive	498	64	562	88.6%	0.86	
negative	94	926	1020			
計	592	990	1582			
ケース内感度	84.1%					
not (初発=0 and 検査0点 and 治療0点)						
	真	偽	計	PPV	F値	
positive	565	222	787	71.8%	0.82	
negative	27	768	795			
計	592	990	1582			
ケース内感度	95.4%					

図 C-2-8. 肺癌の組み合わせパターン別 PPV 及び感度の算出結果

・悪性リンパ腫

悪性リンパ腫病名がある and						
基本						
初発=1						
	真	偽	計	PPV	F値	
positive	196	78	274	71.5%	0.75	
negative	54	938	992			
計	250	1016	1266			
ケース内感度	78.4%					
定型1						
初発=1 and (検査1点以上 or 治療1点以上)						
	真	偽	計	PPV	F値	
positive	196	65	261	75.1%	0.77	
negative	54	951	1005			
計	250	1016	1266			
ケース内感度	78.4%					
定型2						
初発=1 and (検査1点以上 and 治療1点以上)						
	真	偽	計	PPV	F値	
positive	141	11	152	92.8%	0.70	
negative	109	1005	1114			
計	250	1016	1266			
ケース内感度	56.4%					
定型3						
(初発=1 or (検査1点以上 and 治療1点以上))						
	真	偽	計	PPV	F値	
positive	213	134	347	61.4%	0.71	
negative	37	882	919			
計	250	1016	1266			
ケース内感度	85.2%					
定型1+3						
{(初発=1 and (検査1点以上 or 治療1点以上)) or (初発=0 and 検査1点以上 and 治療1点以上)}						
	真	偽	計	PPV	F値	
positive	213	121	334	63.8%	0.73	
negative	37	895	932			
計	250	1016	1266			
ケース内感度	85.2%					
案1						
(初発=1 not (検査1点以下 and 治療0点))						
	真	偽	計	PPV	F値	
positive	194	42	236	82.2%	0.80	
negative	56	974	1030			
計	250	1016	1266			
ケース内感度	77.6%					

図 C-2-9. 悪性リンパ腫の組み合わせパターン別 PPV 及び感度の算出結果

・全癌

悪性腫瘍病名がある and						
基本						
初発=1						
	真	偽	計	PPV	F値	
positive	4827	1084	5911	81.7%	0.81	
negative	1142	6500	7642			
計	5969	7584	13553			
ケース内感度	80.9%					
案1						
初発=1 and 病理検査						
	真	偽	計	PPV	F値	
positive	4382	736	5118	85.6%	0.79	
negative	1587	6848	8435			
計	5969	7584	13553			
ケース内感度	73.4%					
案2						
初発=1 and 病理検査 and 管理科1点以上						
	真	偽	計	PPV	F値	
positive	3289	432	3721	88.4%	0.68	
negative	2680	7152	9832			
計	5969	7584	13553			
ケース内感度	55.1%					
案3						
初発=1 and (病理検査 or 管理科1点以上)						
	真	偽	計	PPV	F値	
positive	4682	931	5613	83.4%	0.81	
negative	1287	6653	7940			
計	5969	7584	13553			
ケース内感度	78.4%					
案4						
初発=1 or 病理検査						
	真	偽	計	PPV	F値	
positive	5459	3154	8613	63.4%	0.75	
negative	510	4430	4940			
計	5969	7584	13553			
ケース内感度	91.5%					

図 C-2-10. 全癌の組み合わせパターン別 PPV 及び感度の算出結果

6) 改良型アウトカム定義案の決定

最終的な改良型アウトカム定義案はPMDAとの協議の下で決定した。

条件式は以下の通りである

[病名 (SS-MIX2) : 確定] OR [病名 (DPC)] AND [追加ロジック (表 C-2-7)]

表 C-2-7. 改良型アウトカム定義案

癌種	パターン名	追加ロジック	機械学習用データ (全体の7割)			評価用データ (全体の3割)		
			PPV (%)	感度 (%)	F値	PPV (%)	感度 (%)	F値
肺癌	バランス型	初発=1	83.23	90.54	0.87	85.29	85.29	0.85
肺癌	PPV 重複型	初発=1 and (検査1点以上 or 治療1点以上)	89.98	83.45	0.87	90.15	76.89	0.83
悪性リンパ腫	バランス型	初発=1 and (検査1点以上 or 治療1点以上)	75.10	78.40	0.77	75.28	67.68	0.71
悪性リンパ腫	PPV 重複型	初発=1 and (検査1点以上 and 治療1点以上)	92.76	56.40	0.70	92.31	48.48	0.64
全癌	バランス型	初発=1	81.66	80.87	0.81	82.83	81.31	0.82
全癌	PPV 重複型	初発=1 and 病理検査	85.62	73.41	0.79	86.20	73.02	0.79

C-3. 急性膵炎

- 対象の病態：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する、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）を起因とした膵炎を除く）
- 期間：2011年1月1日～2019年12月31日
2021年度はB-3の4)、5)について実施した。

<APC 定義>

前年度作成した APC 定義候補を表 C-3-1 に示す。

表 C-3-1. 急性膵炎の APC 定義候補

[病名(DPC)] 時系列 AND*1 [注射オーダー・実施(SS-MIX/レセプト/DPC)]時系列 NOT*2[診療行為(DPC)]

<DPC 病名と各条件間における時間的關係>

*1：DPC 病名がある入院日から退院日までの間に DPC 注射がある

*2：DPC 病名がある入院日から退院日までの間に DPC 診療行為がある

※注射オーダー・実施情報：蛋白分解酵素阻害剤

※診療行為：ERCP/手術

4) APC 定義候補の対象アウトカムに対するし真偽判定と PPV の算出

4)-1 判定結果

APC 定義より抽出された全症例 213 例に対して、専門医 2 名により真偽判定を実施した。

カッパ係数は 0.68 (95%信頼区間: 0.58 - 0.78) であった。判定結果を表 C-3-2 に示す。

表 C-3-2. 判定結果

Possible cases数	判定対象 possible cases数	真のケース	疑わしいケース	その他のケース
213	213	109	10	94

4)-2 PPV の算出結果

真のケースを最終的な真症例とした場合と真のケース及び疑わしいケースを最終的な真症例とした場合の 2 パターンにおける PPV の算出結果を表 C-3-3、C-3-4 に示す。

真のケースのみを最終的な真症例とした場合の PPV は 51.17% であった。

表 C-3-3. PPV の算出結果

定義 No.	Possible case	判定対象 possible cases	真症例 数	偽症例 数	PPV (%)	95%信頼区間 [下限値-上限値]	感度 (%)	95%信頼区間 [下限値-上限値]
APC	213	213	109	104	51.17	[44.25-58.06]	100.00	--

真のケース及び疑わしいケースを最終的な真症例とした場合の PPV は 55.87% であった。

表 C-3-4. PPV の算出結果

定義 No.	Possible case	判定対象 possible cases	真症例 数	偽症例 数	PPV (%)	95%信頼区間 [下限値-上限値]	感度 (%)	95%信頼区間 [下限値-上限値]
APC	213	213	119	94	55.87	[48.92-62.65]	100.00	--

4)-3 「その他のケース」の判定の根拠

「その他のケース」と判定した理由を表 C-3-5 に示す。外科的侵襲に伴う膵酵素上昇または膵炎の症例が 28 例存在した。

表 C-3-5. その他のケースの判定理由

判定理由	件数
処置 1 以外の外科的処置に関連した膵酵素上昇又は膵炎	16
ERCP による膵酵素上昇又は膵炎	12
膵腫瘍 (膵酵素上昇なし)	7
膵膵炎もしくは圧迫による膵酵素上昇	5
膵嚢胞性疾患 (膵酵素上昇なし)	5
腫瘍による閉塞性膵炎 (膵酵素上昇あり)	4
慢性膵炎 (膵酵素上昇なし)	4
ショックによる膵酵素上昇	3
肝・胆・膵疾患以外 (膵酵素上昇あり) 例) クローン、Lammel 症候群	3
肝・胆・膵疾患以外 (膵酵素上昇なし)	3
胆道・胆管系疾患 (膵機能上昇なし)	3
胆道・胆管系疾患 (膵機能上昇なし)、ERCP あり	3
MOF に伴う膵酵素上昇	2
外傷	2
高アミラーゼ血症	2
骨髄移植後の膵酵素上昇	2
腫瘍による閉塞性膵炎 (膵酵素上昇なし)	2
胆道・胆管系疾患 (膵機能上昇あり)	2
膵炎の進行に伴う膵腫瘍	2
慢性膵炎の急性増悪 (蛋白分解酵素阻害剤投与なし)	2
膵炎後合併症 (膵酵素上昇あり)	2
その他	8

5) APC 定義案の決定

4)の結果から、APC 定義候補(表 C-3-1)は PPV50%で真症例 100 例以上の抽出が可能であった。したがって、APC 定義案として提示することに決定した。

D. 考察

D-1 糖尿病性ケトアシドーシス

本検討では、APC 定義から E872 と E872 に紐づく病名交換用コードを持つ病名を除き、かつ pH と HCO₃⁻の数値を AND 条件または OR 条件で加えた 2 通り(案 1, 案 2)の改良型アウトカム定義を作成した。これらの定義を各実施施設で真偽判定をした症例の個票データに適用すると、施設間で PPV vs 感度にバラつきがみられた。まず、九州大学病院では他施設に比べて感度の低下が大きく、これは病院形態による運用の関係上、初期治療がなされた後に同院に入院するケースが多いことが一因と考えられた。つまり、初期治療の実施施設での検査値が最も悪く、紹介状を通じて得られるこの情報は MID-NET データベースに保有されないため、多くの真症例が改良型アウトカム定義に該当しないものとして除外された可能性が高い。

一方、佐賀大学医学部付属病院では、2 通りのアウトカム定義の PPV vs 感度がいずれも 80%後半から 90%台と非常に高く、前年度実施した APC 定義の検討においても 3 施設中最も真症例の比率が高かった。改良型アウトカム定義は真偽判定を実施した症例の範囲内で作成したため、この施設の真症例の特徴を反映した定義となった可能性は否めない。

なお、今回は 3 施設データを合わせて機械学習などに供した際には、施設別の症例数の重みを配慮しなかったが、症例数に大きく差があるため、症例数を調整して行った方がより適切な PPV、感度が得られた可能性はあり、今後の課題とする。

本検討では、全拠点で用いる一つの改良型アウトカム定義の作成を試みたが、例えば、自宅から直搬送される症例が多い市中病院と近医で治療後に転院する症例が多い大学病院の個票データを分けてアウトカム定義を作成する等、施設背景を考慮したアウトカム定義が必要なかもしれない。また、APC 定義の検討結果から 1 施設の真症例数が他 2 施設の 10 倍以上であると推定され、検討初期の段階からこの施設における真症例の特定に着目した改良型アウトカム定義の作成も一案だったのかもしれない。

他にも、MID-NET 利用時には施設選択が可能であるが、また今回の結果から、その際に疾患特性や施設特性を考慮することが重要だろう。

D-2 悪性腫瘍

肺癌、悪性リンパ腫および全癌のいずれも初期ルールにおいて PPV が 50%を大きく下回っていた。病名抽出源別に PPV を算出した結果、SS-MIX においては【「疑い」病名のみ】の条件で PPV が 1%未満と極めて低くレセプトにおいては、レセプト病名単独の条件において真症例の割合が 3 割にも満たないことから、SS-MIX と同様に検査目的で「疑い」病名が登録された症例が多かった可能性がある。これらの病名抽出源は、改良型アウトカム定義の検討において除外したが、初期ルールの段階で含めないことも有用であったと考える。

機械学習から抽出された重要変数のうち DPC 様式 1 の「がん患者/初発・再発 (DPC コード: CAN0010)」が、肺癌、悪性リンパ腫、全癌に共通して真症例の特定に最も有効な変数(寄与率 50%以上)であった。DPC 様式 1 には疾患に特化した項目があることから、テーマによっては DPC 様式 1 との組み合わせが真症例の抽出に有効なアウトカム定義となる可能性がある。この変数と本院の病院形態を示す算定項目を除いて機械学習を再度実施した結果、3 種のいずれも上位重要変数の中に施設基準に関連した加算や管理料などが点在していた。これらのレセプト情報は、診療報酬改定に伴い算定要件が少しずつ変わり、かつ施設基準の厳格化から結果的に施設間で算定にバラつきが生じる=レセプト情報がバラつく可能性がある。本検討では採用しなかったが、今後のアウトカム定義の作成において十分留意していく必要がある。

改良型アウトカム定義において、肺癌では DPC 様式 1 の「がん患者/初発・再発 (DPC コード: CAN0010)」を条件に追加するだけで PPV・感度ともに 80%を超え、かつ PPV vs 感度のバランスも良かった。一方で、悪性リンパ腫では DPC 様式 1 の追加条件で PPV・感度のどちらも 70%台であり、検査・治療を条件に追加してもほぼ変わらなかった。この理由としては、入院していない(DPC 病名が無い)、検査・治療を実施していない等が挙げられるが、悪性リンパ腫の種類、部位、進行度等によって検査や治療の方針が異なることが影響していると考えられた。

全癌において機械学習から得られた重要変数は、本院において症例数が多い癌種の診療内容

を反映している傾向があった。そのため、上位重要変数のうち癌種に影響しないレセプト診療行為とDPC様式1情報を用いて複数パターンアウトカム定義を作成し、うちDPC様式1「がん患者/初発・再発 (DPCコード: CAN0010)」を追加したアウトカム定義がPPV・感度ともに80%台と高くバランスがよかった。

全癌においては複数拠点で適用することを予定しており、施設背景も含めて様々な観点から検討していく必要がある。

D-3 急性膵炎

本検討にて、前年度作成したAPC定義候補は目標であるPPV 50%以上、かつ真症例100例以上を確保できる定義の可能性が高いことが分かった。他MID-NET協力医療機関でも同様の結果が得られるか、複数施設における妥当性を検討する必要がある。

真偽判定において「その他のケース」と判定された症例の中に、外科的侵襲に伴う膵酵素上昇または膵炎が含まれており、[処置1]で設定した診療行為コードでは不十分であった。実際にEUS-FNAやHGSなど超音波内視鏡下での穿刺術を起因とした膵酵素上昇または膵炎の症例が16名存在しており、抽出条件の見直しが必要であると思われる。

また、APC定義では膵酵素の検査情報を条件に含めなかったため、肝・胆・膵疾患で膵炎所見がなく、何らかの理由でDPC病名がついたため抽出されたケースが10名以上存在していた。膵炎の疑いで精査のため疑い病名として登録された可能性もあるが、MID-NETの仕様上、DPC病名において疑いを除くことができない。これらは検査情報を条件に加えることで回避できるケースであり、改良型アウトカム定義を作成する場合には考慮すべき事項であろう。

E. 結論

E-1 糖尿病性ケトアシドーシス

3施設のMID-NET個票データを用いた一つの改良型アウトカム定義を作成し、各実施施設へ適用したところ、施設間でPPVや感度にバラつきがみとめられた。MID-NET協力医療機関全体において、可能な限り真症例を正確に抽出するためには、症例数や施設背景を考慮したアウトカム定義の検討が重要である。

E-2 悪性腫瘍

PPVや感度が比較的高く、かつバランスの良

い改良型アウトカム定義は、PPV低下の要因である疑い病名を含む病名抽出源を除き、癌に特化したDPC様式1項目を条件に追加した定義であった。今後は、複数施設にて定義の妥当性を検討する必要がある。

E-3 急性膵炎

本検討にて、外科的操作に伴う膵酵素上昇又は膵炎を除外する条件の見直しや、膵酵素上昇のない偽症例の除外条件の検討等の課題はあるものの、当初の目標であるPPV 50%以上で真症例100例以上を確認できるAPC定義である可能性が高いことがわかった。

今後は複数拠点においてこれらの課題を含め、定義の妥当性について検討していく必要がある。

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：

1. 論文発表

- 1) Toyoshi Inoguchi, Tasuku Okui, Chinatsu Nojiri, Erina Eto, Nao Hasuzawa, Yukihiro Inoguchi, Kentaro Ochi, Yuichi Takashi, Fujiyo Hiyama, Daisuke Nishida, Fumio Umeda, Teruaki Yamauchi, Daiji Kawanami, Kunihisa Kobayashi, Masatoshi Nomura, Naoki Nakashima, A Simplified Prediction Model for End-stage Kidney Disease in Patients With Diabetes, Scientific Reports, <https://doi.org/10.21203/rs.3.rs-1440453/v1>, doi:10.21203/rs.3.rs-1440453/v1, 2022.03.
- 2) Hajime Nagasu, Yuichiro Yano, Hiroshi Kanegae, Hideo J.L. Heerspink, Masaomi Nangaku, Yosuke Hirakawa, Yuka Sugawara, Naoki Nakagawa, Yuji Tani, Jun Wada, Hitoshi Sugiyama, Kazuhiko Tsuruya, Toshiaki Nakano, Shoichi Maruyama, Takashi Wada, Kunihiko Yamagata, Ichiei Narita, Kouichi Tamura, Motoko Yanagita, Yoshio Terada, Takashi Shigematsu, Tadashi Sofue, Takafumi Ito, Hirokazu Okada, Naoki Nakashima, Hiromi Kataoka, Kazuhiko Ohe, Mihoko Okada, Seiji Itano, Akira Nishiyama, Eiichiro Kanda, Kohjiro Ueki, Naoki Kashihara, Kidney Outcomes Associated With SGLT2 Inhibitors Versus Other Glucose-Lowering Drugs in Real-world Clinical Practice: The Japan Chronic Kidney Disease Database, Diabetes Care, 44(11): 2542-2551, doi:10.2337/dc21-1081, 2021.11.
- 3) 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.
- 4) Tasuku Okui, Masayuki Ochiai, Naoki Nakashima, An Association between Maternal Occupations and Low Birth Weight Infants in Japan from 1995 to 2015, *International Journal of Environmental Research and Public Health*, 18(15), 8040; doi:10.3390/ijerph18158040, 2021.07.
- 5) Toyoshi Inoguchi, Yasunobu Nohara, Chinatsu Nojiri, Naoki Nakashima, Association of serum bilirubin levels with risk of cancer development and total death, *Scientific Reports*, doi:10.1038/s41598-021-92442-2, 2021.06.
- 6) Takanori Yamashita, Rieko Izukura, Naoki Nakashima, Performance Verification of a Text Analyzer Using Machine Learning for Radiology Reports Toward Phenotyping, *Innovation in Medicine and Healthcare*, 242:171-182, doi:10.1007/978-981-16-3013-2_14, 2021.06.

2. 学会発表

- 1) 山下 貴範, 陰山 卓哉, 永島 里美, 内山 田 健次, 高田 敦史, 野尻 千夏, 岸場 真理, 大田 有佳里, 堀田 多恵子, 宇山 佳明, 大江 和彦, 中島 直樹, 康 東天, ここまで来た、病院情報システムに蓄積するデータの標準化と 2 次利活用手法～MID-NET をモデルとして～, 第 41 回医療情報学連合大会 (第 22 回日本医療情報学会学術大会), 2021.11, 国内, 口頭.
- 2) 中島 直樹, 山本 隆一, 寺本 振透, 井上 悠輔, 阿部 浩幸, 高林 克日己, 研究から社会実装に向けた医療データ活用への展開—個人情報保護法や次世代医療基盤法などの枠組みでの活用—, 第 41 回医療情報学連合大会 (第 22 回日本医療情報学会学術大会), 2021.11, 国内, 口頭.
- 3) 白鳥 義宗, 岡田 美保子, 大津 敦, 中島 直樹, 松田 晋哉, 長島 公之, 野村 由美子, Real World Evidence 実現のために, 第 41 回医療情報学連合大会 (第 22 回日本医療情報学会学術大会), 2021.11, 国内, 口頭.

H. 知的財産権の出願・登録状況 (予定を含む。)

1. 特許取得
特になし
2. 実用新案登録
特になし
3. その他