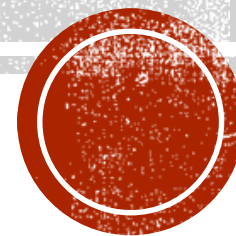


アウトカム指標のバリデーション簡易化 の工夫・機械学習の活用

中島直樹
奥井 佑
野尻千夏

九州大学病院 メディカル・インフォメーションセンター



「アウトカム定義」について

利用目的や利用するシステムに応じて、アウトカム定義は変わる

- ✓ 感度重視か、特異度重視か、PPV重視か
- ✓ 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
- ✓ 予測であって、医学解析のような相関/因果関係の推論まではしないことが多い（当たれば良い）
- ✓ 感度・PPVが低くても、その両方が正しいならば、症例数を正確にベイズ推定できる

厚労科研 宇山班 (アウトカム定義の作成)

- MID-NETシステムで利用できるデータ項目の中から選択された変数の組み合わせ
 - ✓ SS-MIX2、レセプトデータ、DPC
- MID-NETシステムでは、症例抽出に直接機械学習は使えず、単純なスクリプトしか使えない
 - ✓ 機械学習を使うとしてもその結果をヒトが解釈して、ロジックを模索するしかない
- アウトカム定義は、10の協力医療施設グループのうちの数施設で作成する
- 作成されたアウトカム定義は、グループの一部あるいは全体で用いる
- 感度を正確に測定するためにAll Possible Cases (APC) の概念を用いることとなった
 - ✓ 最初のロジックは感度を100%と想定できる
 - ✓ できるだけ高い感度と高い陽性的中率の組み合わせを模索する

初期アウトカム定義：APC定義。感度100%、PPV50%以上(医師が真偽判定する場合、200症例のランダム抽出で100例以上の真症例を得るため)



改良アウトカム定義 (感度はできれば80%以上、PPVできれば80%以上、こだわり過ぎない)

事例紹介：MID-NETで用いている手法の紹介と精度の検証

- 複数の手法でロジックを作成し性能を評価
 - MICスタッフ作成のロジック（本手法）や機械学習モデルなど
- 訓練データと検証データを分ける（機械学習モデル）
 - 訓練データでモデルをあてはめてロジックやモデルを作成
 - 検証データで性能評価
 - 機械学習モデルの評価では一般的な方法
 - 同じデータを分けても所詮は同じデータではあるが、やらないよりはベター

2021年度に九州大学が検討したテーマ

◆ 急性膵炎

◆ ケトアシドーシス

◆ 悪性腫瘍

• 肺癌

• 悪性リンパ腫

• 全癌

All Possible Cases (APC) 定義の具体
「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%とするため）

急性膵炎では、対象事象を「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）」とした（APC定義のPPVを50%以上とするため（感度100%は前提））

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

- ✓ 医師の真偽判定だと、労力が大変だし、200症例サンプリングが限界（真症例を100例以上確保するには、初期定義からPPV50%以上必要）
- ✓ 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

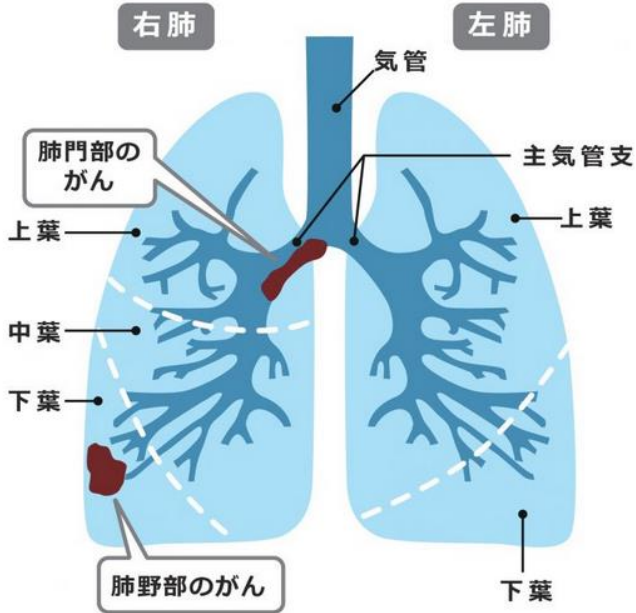
MID-NET対象データ項目からの
「肺癌」抽出

肺癌の初期ルール⇒病名(下記コードリスト)

DPC病名を抽出	SS-MIX2病名を抽出	レセプト病名を抽出
----------	--------------	-----------

ICD10コード: **C34** 気管支及び肺の悪性新生物<腫瘍>

	病名	ICD10コード	病名交換用コード	傷病名コード
1	主気管支の悪性腫瘍	C340	KGDF	8834569
2	肺門部小細胞癌	C340	BB18	8847675
3	肺門部腺癌	C340	R63U	8847676
4	肺門部大細胞癌	C340	GLQ7	8847677
5	肺門部肺癌	C340	HUDK	8842835
6	肺門部非小細胞癌	C340	LERJ	8847678
7	肺門部扁平上皮癌	C340	KLVH	8847679
8	パネコースト症候群	C341	JE7P	8845183
9	上葉小細胞肺癌	C341	N9VR	8847633
10	上葉肺癌	C341	KF4E	8835493
11	上葉肺腺癌	C341	UL6D	8847634
12	上葉肺大細胞癌	C341	M8T4	8847635
13	上葉肺扁平上皮癌	C341	Q5FE	8847636
14	上葉非小細胞肺癌	C341	F682	8847637
15	中葉小細胞肺癌	C342	FSPU	8847660
16	中葉肺癌	C342	FD1R	8837666
17	中葉肺腺癌	C342	DJ77	8847661
18	中葉肺大細胞癌	C342	VVL0	8847662
19	中葉肺扁平上皮癌	C342	PKB6	8847663
20	中葉非小細胞肺癌	C342	K4VD	8847664
21	下葉小細胞肺癌	C343	VS6Q	8847594
22	下葉肺癌	C343	T5K3	8831458
23	下葉肺腺癌	C343	M6MC	8847595
24	下葉肺大細胞癌	C343	GMLD	8847596
25	下葉肺扁平上皮癌	C343	SHQ1	8847597
26	下葉非小細胞肺癌	C343	PE1M	8847598
27	細気管支肺胞上皮癌	C348	FF48	8833932



28	ALK融合遺伝子陽性非小細胞肺癌	C349		D84E	8847272
29	EGFR遺伝子変異陽性非小細胞肺癌	C349		KU19	8847732
30	ROS1融合遺伝子陽性非小細胞肺癌	C349		VMN0	8849238
31	気管支カルチノイド	C349		HFLL	8846553
32	気管支癌	C349		REQN	8832157
33	原発性肺癌	C349		EALN	1629003
34	小細胞肺癌	C349		JJJD	8842185
35	小児胸膜肺芽腫	C349		VJEU	8849759
36	成人型肺芽腫	C349		F4TC	8838805
37	肺カルチノイド	C349		EU91	8838804
39	肺癌	C349		LFJS	1629006
40	肺癌による閉塞性肺炎	C349	J178	PMM5	8847242
41	肺癌肉腫	C349		RG78	8842831
42	肺腺癌	C349		BNCP	8838844
43	肺腺様のう胞癌	C349		C38E	8842833
44	肺腺扁平上皮癌	C349		ASQL	8842832
45	肺大細胞癌	C349		NR1N	8838852
46	肺大細胞神経内分泌癌	C349		L392	8842057
47	肺肉腫	C349		MQQP	1629009
48	肺粘表皮癌	C349		KUPF	8842834
49	肺胞上皮癌	C349		UPJO	8838901
50	肺未分化癌	C349		QHE2	8838904
51	肺扁平上皮癌	C349		EBLV	8838898
52	肺絨毛癌	C349		UT40	8848775
53	非小細胞肺癌	C349		D813	8842053

ICD10コード: **D02** 中耳及び呼吸器系の上皮内癌

	病名	ICD10コード	病名交換用コード	傷病名コード
2	気管上皮内癌	D021	RV1D	8843042
3	気管支上皮内癌	D022	A63J	8832166
5	肺上皮内癌	D022	F7Q6	2312002

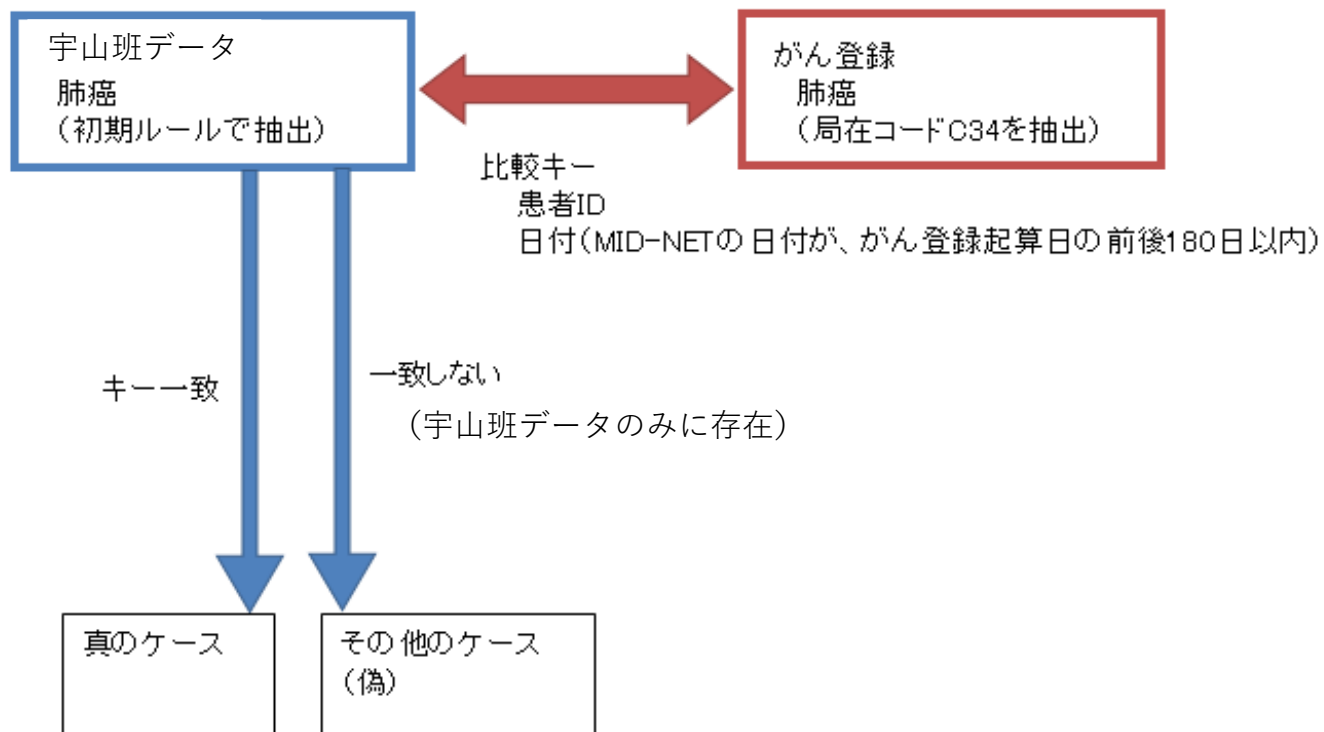
院内がん登録からの 「肺癌」抽出

ICD-O-3

国際疾病分類腫瘍学第3版(3.1版)の局在コード及び和名

局在コード	和名	採否	疾患区分
C34.0	主気管支	○	肺癌
C34.1	上葉, 肺	○	肺癌
C34.2	中葉, 肺	○	肺癌
C34.3	下葉, 肺	○	肺癌
C34.8	肺の境界部病巣	○	肺癌
C34.9	肺, NOS	○	肺癌

真偽判定方法



肺癌 真偽判定結果

病名抽出源を変えて集計を実施

病名抽出源	宇山班データ ケース数	真	その他	PPV	がん登録 ケース数	真	その他	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	99.6%

※SS-MIX疑いのみ: SS-MIX確定病名もDPC病名も無い、という意味。

機械学習を用いたアウトカム定義の修正はこれをスタートとした

まずは臨床的検討！いきなり機械学習をするわけではない

少なくとも数例の治療パターンを確認し、使用するデータ種別と期間を検討した。

症例 1

5月25日	初診	
6月1日	気管支鏡検査	入院
7月5日	手術	入院

症例 2

1月4日	初診	
1月15日	気管支鏡検査	入院
2月9日	放射線治療	入院
4月13日	抗がん剤治療開始	

症例 3

4月4日	初診	
4月9日	PET 検査結果	
4月27日	気管支鏡検査	入院
6月8日	手術	入院

症例 4

9月25日	初診	
11月2日	放射線治療開始	

初診日が病名開始日である場合が多く、手術は初診日から2か月後、3か月後などに実施されている例を確認した。

共通

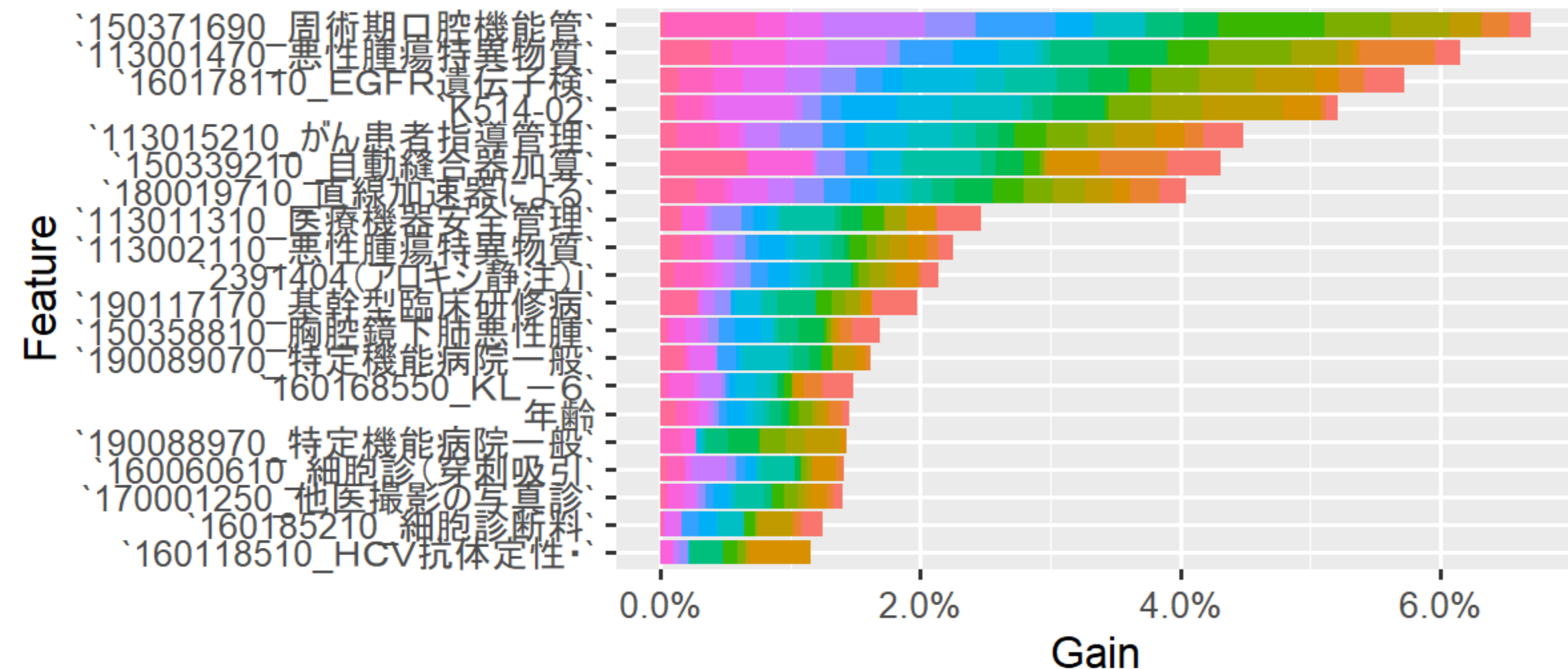
月1などの外来で	
検査(シフラ、SOC)に対し「悪性腫瘍特異性物質治療管理料」	

決めたこと

- まずはシンプルに、レセプト項目（診療行為、管理料、手術）と医薬品（処方オーダー、注射実施）を使用してみよう。手術はKコードも使用する。
- 使用するデータはindex date（病名開始日or入院日）以降3か月ぐらいのデータを見てみよう。
⇒機械学習の結果で+150日までのデータを使用することに仮決めた。

訓練データへの適用結果（GBDT）（上位20位までを図示）

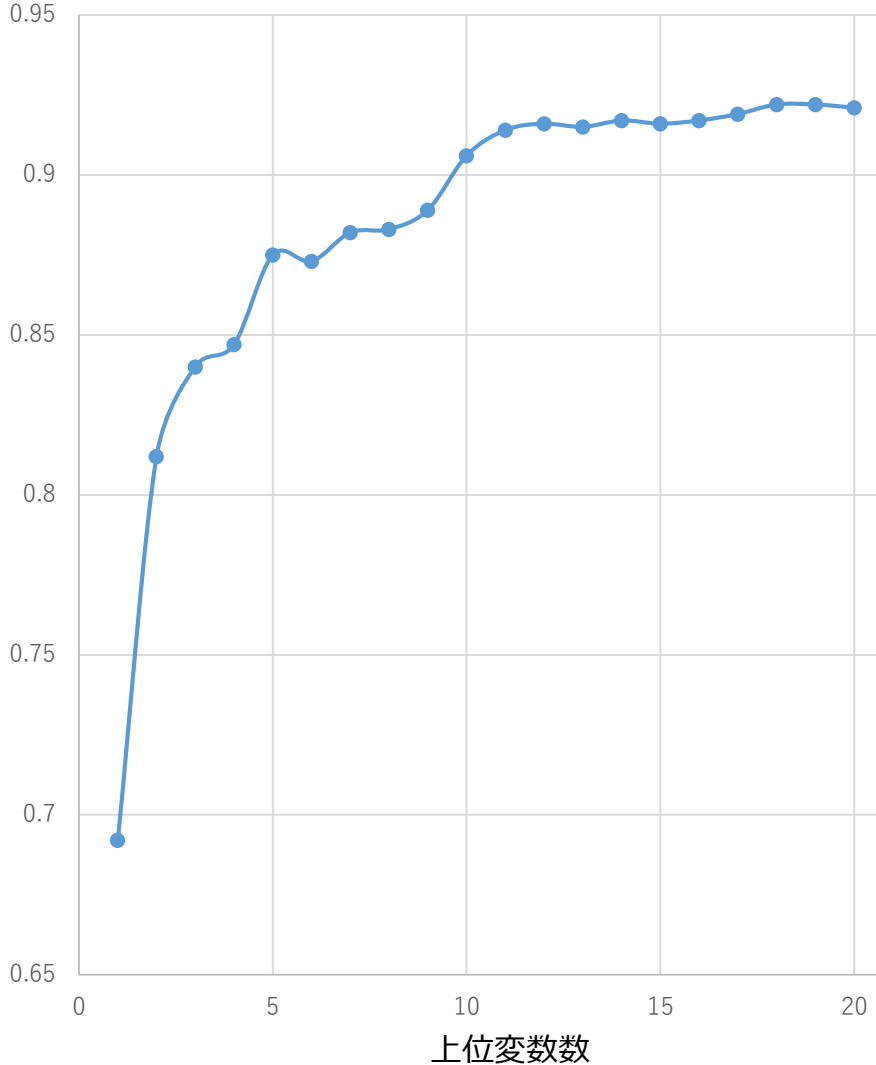
(auc=0.929)



上位20変数について上から順に変数を用いた場合のAUC

	変数名	累積でのAUC値
上位1変数	`150371690_周術期口腔機能管`	0.692
上位2変数	`113001470_悪性腫瘍特異物質`	0.812
上位3変数	`160178110_EGFR遺伝子検`	0.84
上位4変数	`K514-02`	0.847
上位5変数	`113015210_がん患者指導管理`	0.875
上位6変数	`150339210_自動縫合器加算`	0.873
上位7変数	`180019710_直線加速器による`	0.882
上位8変数	`113011310_医療機器安全管理`	0.883
上位9変数	`113002110_悪性腫瘍特異物質`	0.889
上位10変数	`2391404 (アロキシ静注) i`	0.906
上位11変数	`190117170_基幹型臨床研修病`	0.914
上位12変数	`150358810_胸腔鏡下肺悪性腫`	0.916
上位13変数	`190089070_特定機能病院一般`	0.915
上位14変数	`160168550_K L - 6`	0.917
上位15変数	年齢	0.916
上位16変数	`190088970_特定機能病院一般`	0.917
上位17変数	`160060610_細胞診(穿刺吸引)`	0.919
上位18変数	`170001250_他医撮影の写真診`	0.922
上位19変数	`160185210_細胞診断料`	0.922
上位20変数	`160118510_H C V 抗体定性・`	0.921

AUC



GBDT結果から、MICスタッフが経験的に作成したロジック

入院	入院	入院	管理料	管理料	管理料	管理料	手術	手術	手術	検査	検査	PET検査	放射線治療	放射線治療	管理料
190088970_特定機能病院一般病棟入院期間加算(14日以内)	190089070_特定機能病院一般病棟入院期間加算(15日以上30日以内)	入院期間加算	113001310_悪性腫瘍特異物質治療管理料(その他・1項目)	113001470_悪性腫瘍特異物質治療管理料(その他・1項目初回)	113002110_悪性腫瘍特異物質治療管理料(その他・2項目以上)	悪性腫瘍特異物質治療管理料	K514-00、K514-02	150339210_自動縫合器加算	150371690_周術期口腔機能管理後手術加算	160178110_EGFR遺伝子検査(リアルタイムPCR法以外)	160185210_細胞診断料	170027310_ポジトロン・コンピューター断層複合撮影(18FDG使用)	180019710_直線加速器による放射線治療(定位放射線治療)	113011310_医療機器安全管理料(放射線治療計画策定)	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)
1	1	1		1	1	1	1	1	1			1	1		

赤線太枠部分の合計点で症例を抽出した場合

	PPV	感度	F値	
2点以上	49.0%	98.9%	0.66	
3点以上	60.9%	93.5%	0.74	△
4点以上	75.3%	83.7%	0.79	○
5点以上	83.5%	68.1%	0.75	△
6点以上	90.8%	51.0%	0.65	
7点	95.4%	30.1%	0.46	

コードリスト	
190088970	特定機能病院一般病棟入院期間加算(14日以内)
190089070	特定機能病院一般病棟入院期間加算(15日以上30日以内)
113001310	悪性腫瘍特異物質治療管理料(その他・1項目)
113001470	悪性腫瘍特異物質治療管理料(その他・1項目初回)
113002110	悪性腫瘍特異物質治療管理料(その他・2項目以上)
150339210	自動縫合器加算
150371690	周術期口腔機能管理後手術加算
160178110	EGFR遺伝子検査(リアルタイムPCR法以外)
160185210	細胞診断料
170027310	ポジトロン・コンピューター断層複合撮影(18FDG使用)
180019710	直線加速器による放射線治療(定位放射線治療)
113011310	医療機器安全管理料(放射線治療計画策定)
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)

最終アウトカム定義はさらに修正したものである
がサンプルとして主な経過を提示した

検証データでモデルの性能を比較

- ルールベース
 1. MICスタンプ作成のロジック（GBDT結果の上位変数の経験的・臨床的選択）
- ロジスティックモデル
 1. GBDTでの上位20の変数について変数選択してモデル作成(Lasso)
 2. GBDTでの上からのAUCが0.9までの変数を使ってモデル作成
 3. 全変数から変数選択してモデル作成
- GBDTモデル
 1. 上位20の変数を用いてGBDTで判別モデルを作成
 2. 全変数を用いてモデル作成

GBDTでの上位20の変数について変数選択してモデル作成（Lasso）

変数名	標準化偏回帰係数
(Intercept)	-1.01463229
150339210_自動縫合器加算	0.49268359
150371690_周術期口腔機能管	0.526012649
113001470_悪性腫瘍特異物質	0.490427375
160178110_E G F R 遺伝子検	0.474882285
190117170_基幹型臨床研修病	0.290564599
160190470_病理診断管理加算	0.102016314
2391404（アロキシ静注） i	0.417465475
113015210_がん患者指導管理	0.359804133
113011310_医療機器安全管理	0.20866592
113002110_悪性腫瘍特異物質	0.246381772
180019710_直線加速器による	0.305010163
190088970_特定機能病院一般	0.313826678
K514-02	0.192726466
2242001（コデインリン） h	0.445839854
年齢	0.261153482
150332810_閉鎖循環式全身麻	0.079009771
2391008（イメンドカプ） h	0
4291403（カルボプラチ） i	0.325802608
190089070_特定機能病院一般	0.049232885
150358810_胸腔鏡下肺悪性腫	0.211142499

検証データでの評価結果

		感度 †	PPV
・ ルールベース (GBDT+ルール)			
MICスタッフの作成したロジック(4点以上)※ (15変数)		0.833	0.751
・ GBDT+ロジスティックモデル(正則化による変数選択)			
1) GBDTでの上位20の変数について変数選択したモデル (19変数)		0.848	0.838
2) GBDTでのAUCが0.9になる変数までを使ったモデル (10変数)		0.828	0.810
3) 全変数から変数選択したモデル (151変数)		0.836	0.844
・ GBDTモデル			
1) GBDTでの上位20の変数でのGBDT判別モデル		0.789	0.796
2) 全変数 を用いたGBDTモデル		0.843	0.852

※検証データも含めてロジックを作成しているのでその分性能がよくなっている面がある

† 本当の意味での感度ではない

MICスタッフ（ロジック作成担当）の宇山班での機械学習に関するコメント

<注意・課題>

1) 機械学習に投入するデータに関する工夫

✓データ種別や期間の選定

✓ 基本的に使えるデータは全て使用してよいが、各データ定義のためにあらかじめ臨床的検討が必要

✓ AUC高値となるよう、事前に複数パターン（粒度、期間設定など）でGBDTを試すこともある

✓検体検査結果データの取り扱い

✓ 初期値？最大値？最小値？平均？複数（経時データなど）？欠損値は？

✓同種の医薬品、診療行為、手術などのグルーピング

✓ より良い結果を出すために粒度調整を必要とする場合がある

2) 対象アウトカムが複数アウトカムの集合体の場合、件数が多いものの内容が強く機械学習結果に反映

3) 結果からアウトカム定義が簡単にできるわけではない

✓ 予測重視であれば重みづけをするべきだが、スクリプトが複雑になる上、その詳細な重みづけを他医療施設に適用できるのか？

✓複数施設での検討

✓ 施設特有であろう変数は定義に使用しにくい（標準的でない治療方法、施設基準による加算など）

✓ 運用の違い、病院背景の違いで全く違う結果が出ることもある（感度、PPVなど）

検討事項

- ルールベースか機械学習ベースか
 - MICスタッフ作成のロジック（スコア制で何点以上） = GBDT + ルールベース
 - これは係数をすべて1にした線形の判別式モデルとも解釈できる
 - GBDT + ロジスティック判別モデルを使って重みをつけた方が予測精度が良いが、運用面で使いにくい
 - ロジックに用いる変数の数は多くても運用上問題ないかどうか（例えば、どんな長いスクリプトでも投入できるか？）
 - ヒトがスクリプトを作る場合はミスも起きるので困難
- 相関が強い変数は除くべきか、グループ化すべきか
 - 試行錯誤が必要で、検討に相当な時間がかかる
 - 正則化項付きのロジスティックモデルや他の機械学習モデルを使う場合、ある程度こういう問題を考えなくて済むというメリットがある