

20250527 第11回協会けんぽ調査研究フォーラム
@一橋大学 一橋講堂

機械学習を用いた生活習慣病の 治療行動予測モデルの構築

研究代表者 丹野高三 岩手医科大学 衛生学公衆衛生学講座

研究分担者 西谷直之 岩手医科大学 薬学部 臨床薬学講座 情報薬科学分野
米倉佑貴 聖路加国際大学大学院 看護学研究科 看護情報学
林 邦好 京都女子大学 データサイエンス学部

背景

糖尿病や高血圧などの生活習慣病は将来の脳心血管疾患や認知症、フレイル・要介護の原因であり、長期的に継続的な服薬管理が必要である。

糖尿病 vs. 非糖尿病

脳心血管疾患

1.6 倍

(Hirakawa Y, et al. J Epidemiol 2017;27:123–129)

認知症

1.7 倍

(Ohara T, et al. Neurology 2011;77:1126–1134)

日常生活動作低下

1.7 倍

(Hoang PTN, et al. J Diabetes Investig 2022; 13: 1897–1904)

目的

本研究の目的は、機械学習を用いて糖尿病及び高血圧の

①治療中断予測モデル

②治療コントロール不良予測モデル

を構築することである。

レセプトデータを用いた 治療中断予測モデル

対象者

糖尿病治療者

糖尿病治療者 **677,572** 人

2015～2020年度までの糖尿病関連受診レセプト ^{*1}
23,787,174 件



2015年度定期受診者 ^{*2}

337,549 人

2015～2020年度までの糖尿病関連受診レセプト
17,398,075 件



解析対象者 ^{*3} **315,111** 人

高血圧治療者

高血圧治療者 **2,125,864** 人

2015～2020年度までの高血圧関連受診レセプト ^{*1}
72,303,928 件



2015年度定期受診者 ^{*2}

1,065,039 人

2015～2020年度までの高血圧関連受診レセプト
53,623,767 件



解析対象者 ^{*3} **1,019,439** 人

*1：傷病名に糖尿病（E10-E14）かつ糖尿病薬処方かつ外来受診あり

*2：通院回数4回以上/年度

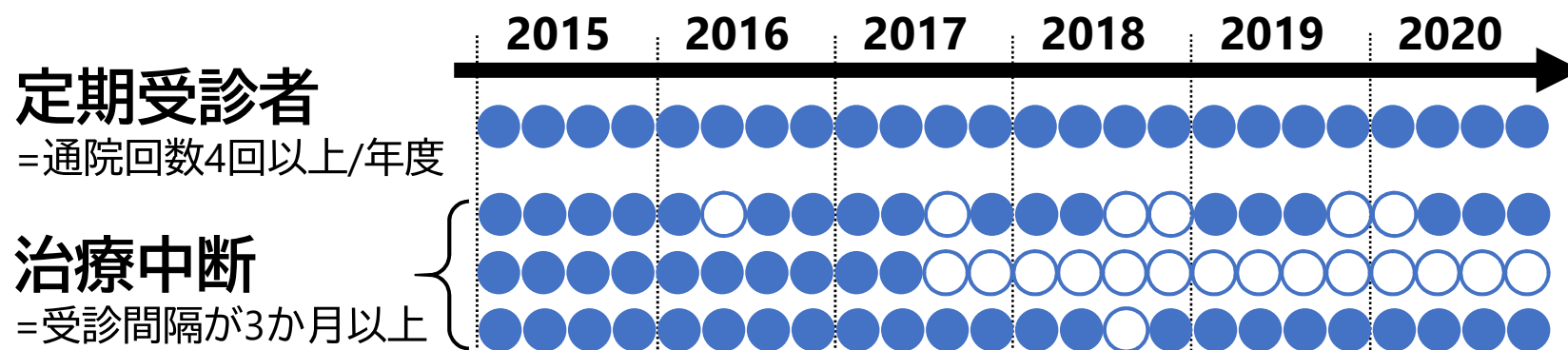
*3：2015年度に年齢が40歳から74歳かつ使用する変数に欠損なし

*1：傷病名に高血圧（I10）かつ降圧薬処方かつ外来受診あり

*2：通院回数4回以上/年度

*3：2015年度に年齢が40歳から74歳かつ使用する変数に欠損なし

治療中断アウトカム



糖尿病治療定期受診者

315,111人



糖尿病治療中断者

48,752人(15.5%)

高血圧治療定期受診者

1,019,439人



高血圧治療中断者

181,088人(17.8%)

統計解析

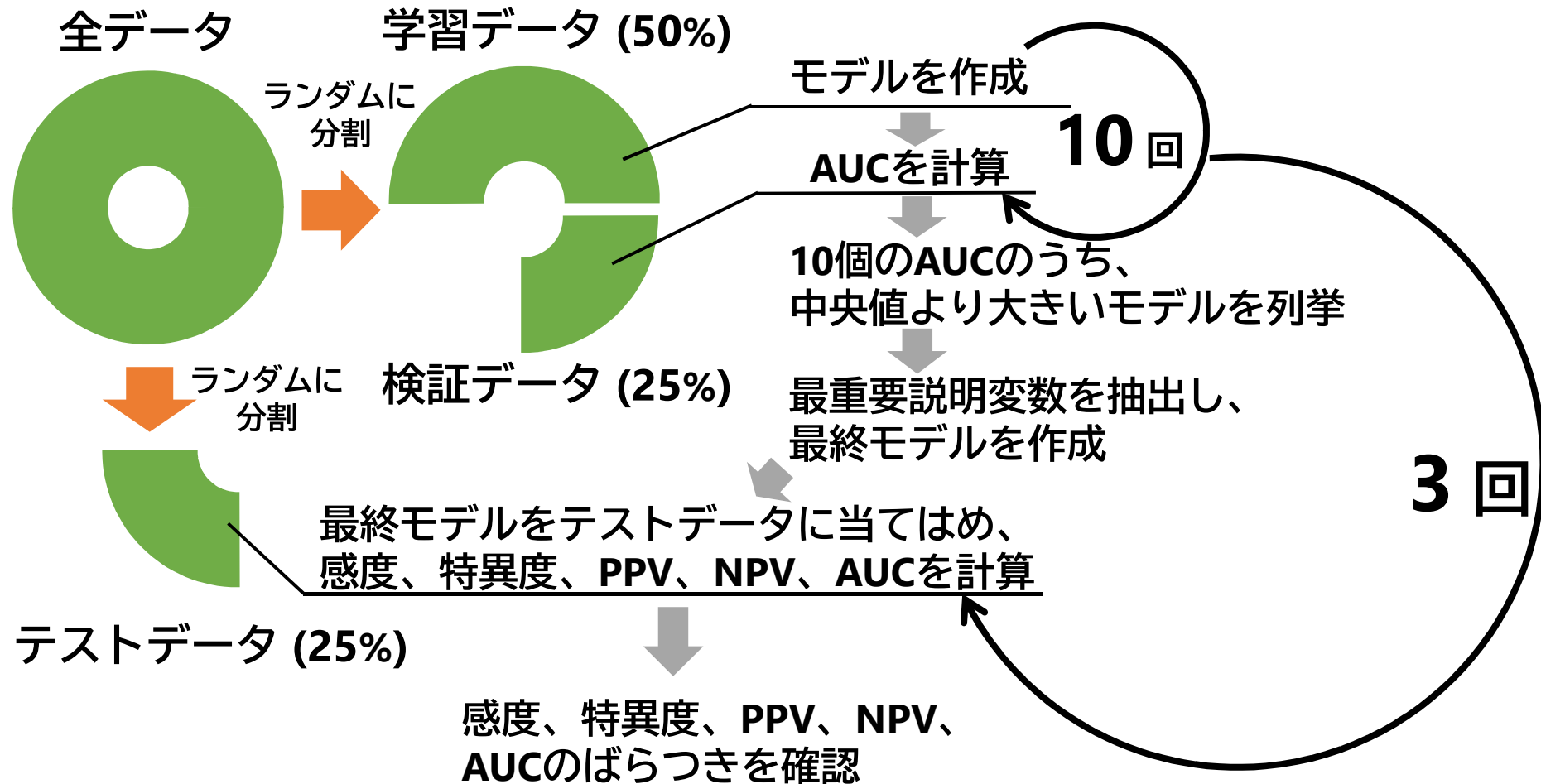
Classification and Regression Trees (CART) モデルによって説明変数の変数重要度を決定

1. 複雑さパラメータ (complexity parameter, cp) を0.00001に設定してCARTを実行した。
2. 1で最低1つの分岐が発生するcpから、0.00001までcpを変化させ、cpに対応した決定木のサイズ、決定木の交差妥当化誤差 (cross validation error, xerror)、交差妥当化誤差の標準誤差 (xstd) を推定した。
3. 最小のxerror $\pm 1 \times \text{xstd}$ になるようなcpのうち、最小の決定木サイズになるようなcpを最良モデルとして、採用した (1-SEルール)。
4. 1-SEルールで採択されたCARTの決定木で選択された変数（変数重要度が算出された変数）をロジスティック回帰分析の説明変数として、ロジスティック回帰モデルを推定した。

CARTモデル (CART)、及び、ロジスティック回帰モデル (logistic) で採択・推定したモデルによる予測の当てはまりをROC分析で評価

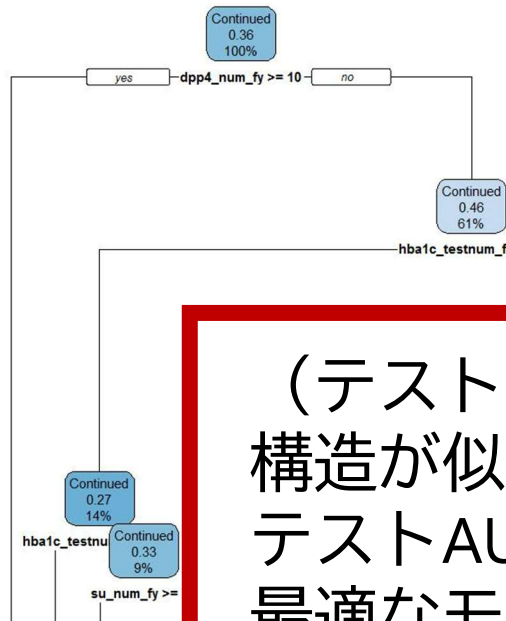
モデルの評価方法

(学習、検証、テストデータの比率はHastie et al.(2001, p.196)を参照)

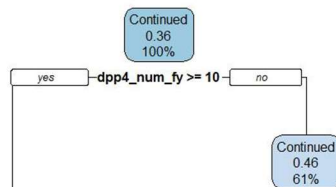


糖尿病治療中断：CART

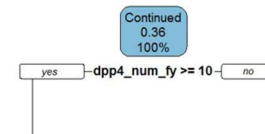
テスト1



テスト2



テスト3



(テストデータを用いた) 3つのモデルのうち、構造が似ている2つのモデルに絞り、最終的にテストAUC値の高いモデルを糖尿病治療中断の最適なモデルとした。



糖尿病治療中断：Logistic

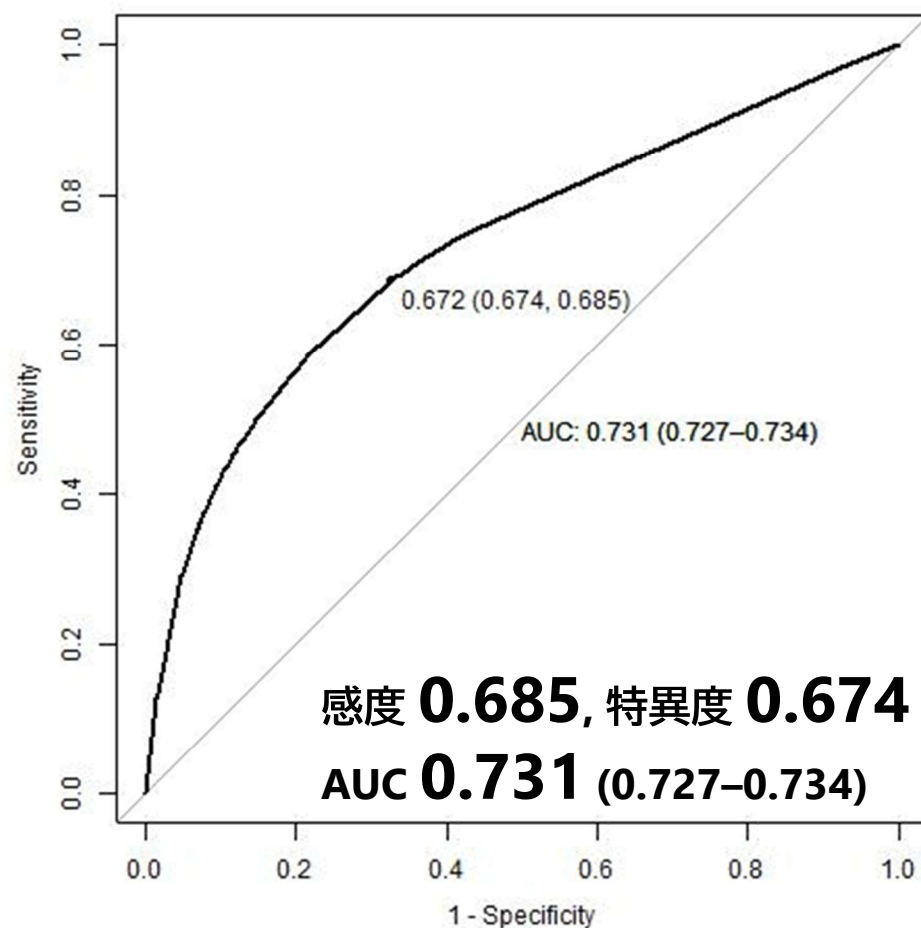
テスト3

最重要変数	OR	95% CI 下限	95% CI 上限	最重要変数	OR	95% CI 下限	95% CI 上限
年齢	0.9816	0.9804	0.9828	年度内糖尿病薬処方種類数	1.4644	1.4456	1.4834
女性	0.8484	0.8265	0.8709	年度内薬処方種類数	1.0230	1.0211	1.0248
DPP-4阻害薬処方量/年度	1.0008	1.0007	1.0009	月内最大処方薬種類数	0.9749	0.9716	0.9783
DPP-4阻害薬処方回数/年度							0.9447
αグルコシダーゼ阻害薬処方量/年度							1.3122
αグルコシダーゼ阻害薬処方回数/年度							1.0113
SU剤処方量/年度							1.0001
SU剤処方回数/年度							1.3637
ビッグアナログ処方量/年度							1.1123
ビッグアナログ処方回数/年度							1.1148
チアゾリジン処方量/年度							1.1737
チアゾリジン処方回数/年度							1.4931
グリニド処方量/年度							1.1896
グリニド処方回数/年度							1.0916
GLP-1受容体拮抗薬処方量/年度							0.9514
GLP-1受容体拮抗薬処方回数/年度							1.0866
インスリン製剤処方量/年度	0.9990	0.9982	0.9998	情報通信	1.1924	1.0985	1.2944
インスリン製剤処方回数/年度	0.9639	0.9611	0.9666	生活関連サービス・娯楽	1.1540	1.0758	1.2378
SGLT2阻害薬処方量/回	1.0002	1.0000	1.0005	電気・ガス・熱供給・水道	0.9703	0.8611	1.0933
SGLT2阻害薬処方回数/年度	0.9473	0.9388	0.9560	農林水産業	1.0248	0.9310	1.1279
糖尿病配合剤処方量/年度	1.0003	1.0001	1.0004	不動産・物品賃貸	1.2829	1.2072	1.3633
糖尿病配合剤処方回数/年度	0.8978	0.8911	0.9046	複合サービス	0.9616	0.8758	1.0559
歯科レセプト回数/年度	0.9936	0.9926	0.9947	その他サービス	1.0864	1.0420	1.1328

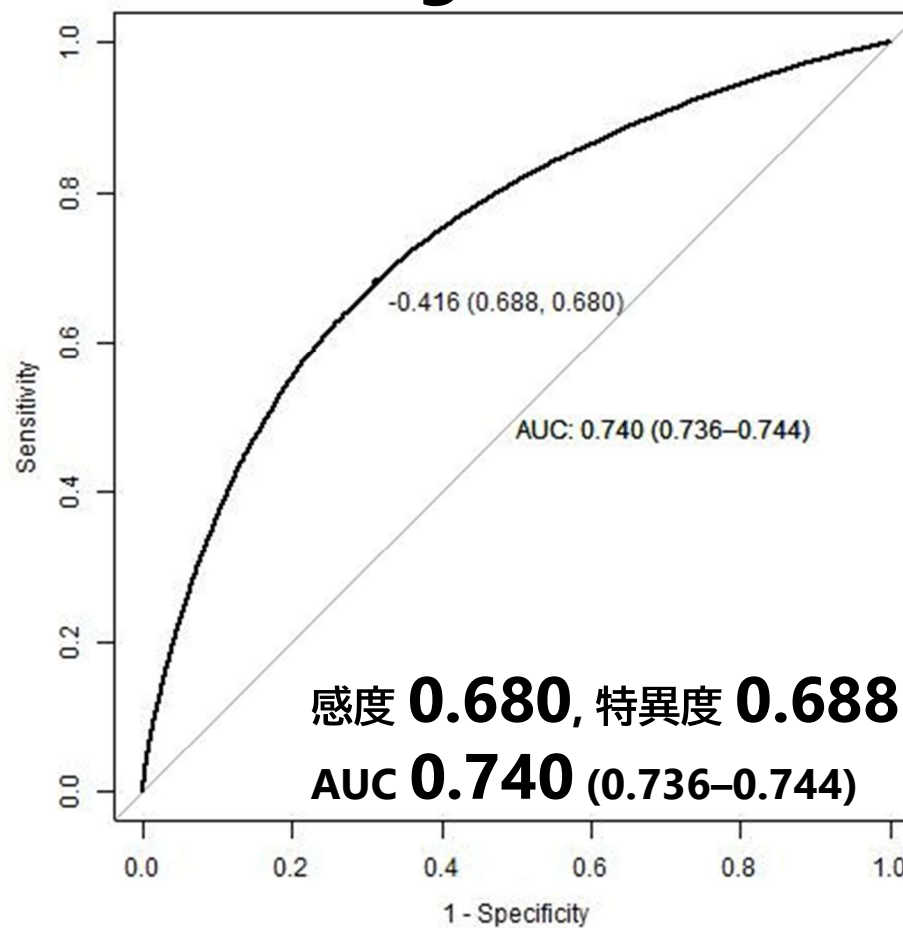
（テストデータを用いた）3つのモデルのうち、構造が似ている2つのモデルに絞り、最終的にテストAUC値の高いモデルを糖尿病治療中断の最適なモデルとした。

糖尿病治療中断予測の最適モデル

CART

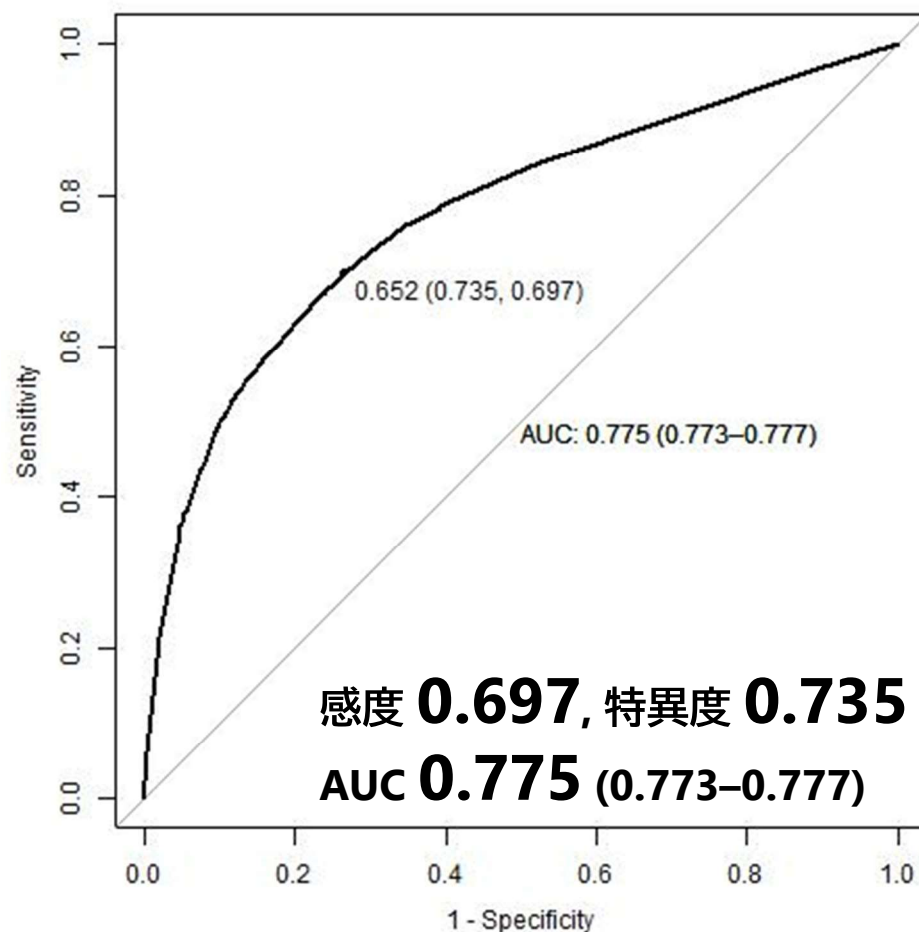


Logistic

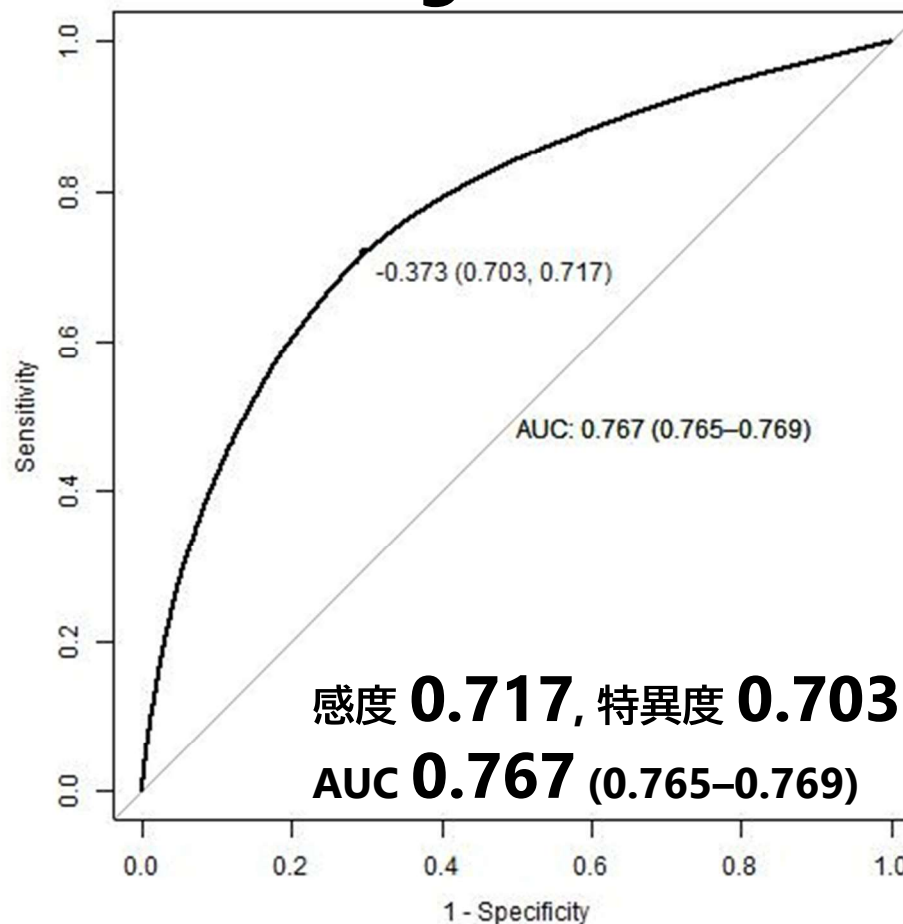


高血圧治療中断予測の最適モデル

CART



Logistic



モデル・データ別の治療中断の予測性能

アウトカム	モデル	レセプトデータ				
		感度	特異度	PPV	NPV	AUC
糖尿病 5年間 の治療中断	CART	0.685	0.674	0.544	0.790	0.731
	Logistic	0.680	0.688	0.554	0.790	0.740
糖尿病 翌年 の治療中断	CART	0.554	0.810	0.349	0.909	0.739
	Logistic	0.703	0.707	0.305	0.929	0.769
高血圧 5年間 の治療中断	CART	0.697	0.735	0.625	0.793	0.775
	Logistic	0.717	0.703	0.606	0.797	0.767
高血圧 翌年 の治療中断	CART	0.766	0.690	0.348	0.932	0.794
	Logistic	0.746	0.718	0.364	0.929	0.802

モデル・データ別の治療中断の予測性能

		レセプトデータのみ					レセプトデータ+健診データ				
アウトカム	モデル	感度	特異度	PPV	NPV	AUC	感度	特異度	PPV	NPV	AUC
糖尿病 5年間の 受診中断	CART	0.685	0.674	0.544	0.790	0.731	0.641	0.758	0.543	0.825	0.735
	Logistic	0.680	0.688	0.554	0.790	0.740	0.686	0.718	0.525	0.835	0.756
糖尿病 翌年の 受診中断	CART	0.554	0.810	0.349	0.909	0.739	0.567	0.852	0.355	0.932	0.769
	Logistic	0.703	0.707	0.305	0.929	0.769	0.720	0.741	0.286	0.948	0.798
高血圧 5年間の 受診中断	CART	0.697	0.735	0.625	0.793	0.775	0.691	0.754	0.605	0.817	0.774
	Logistic	0.717	0.703	0.606	0.797	0.767	0.715	0.732	0.593	0.825	0.784
高血圧 翌年の 受診中断	CART	0.766	0.690	0.348	0.932	0.794	0.645	0.819	0.406	0.924	0.800
	Logistic	0.746	0.718	0.364	0.929	0.802	0.746	0.748	0.356	0.940	0.821

健診データを用いた コントロール不良予測モデル

対象者

糖尿病治療者

2015年度糖尿病治療者 *1 **349,463** 人
2015～2020年度までの健診データ
1,371,633 件



2015-2020年健診受診者 **124,217** 人
2015～2020年度までの健診データ
745,302 件



解析対象者 *2 **74,324** 人

高血圧治療者

2015年度糖尿病治療者 *1 **1,187,510** 人
2015～2020年度までの高血圧関連受診レセプト *1
4,696,586 件



2015-2020年健診受診者 **440,590** 人
2015～2020年度までの健診データ
2,643,540 件



解析対象者 *2 **294,442** 人

*1：2015年度に「インスリン注射または血糖を下げる薬」を服用中と回答

*2：2015年度に年齢が40歳から74歳かつ使用する変数に欠損なし

*1：2015年度に「血圧を下げる薬」を服用中と回答

*2：2015年度に年齢が40歳から74歳かつ使用する変数に欠損なし

コントロール不良アウトカム

コントロール不良者：2016–2020年度の健診結果のうち以下のいずれかの場合

■糖尿病

空腹時血糖値 130 mg/dL or 随時血糖値 180 mg/dL or HbA1c 7.0% を上回った場合
(糖尿病診療ガイドライン2019)

■高血圧

収縮期血圧 140 mmHg or 拡張期血圧 90 mmHg を上回った場合
(高血圧治療ガイドライン2019)

糖尿病治療者

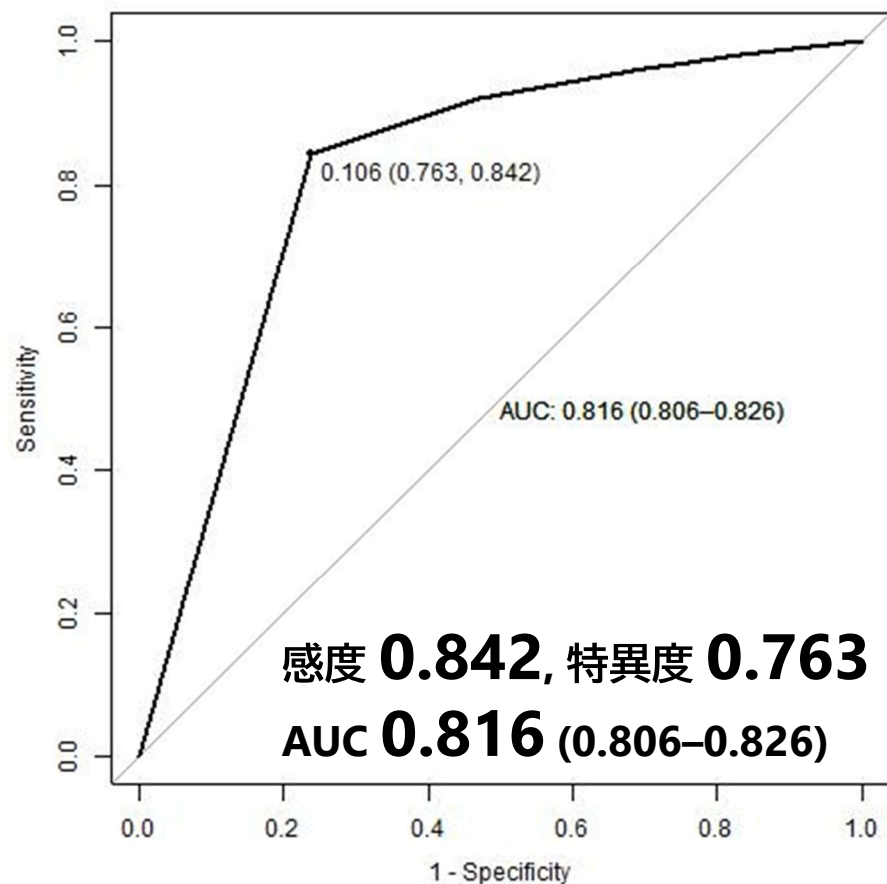
74,324人  糖尿病コントロール不良者 **42,617**人(57.3%)

高血圧治療定期受診者

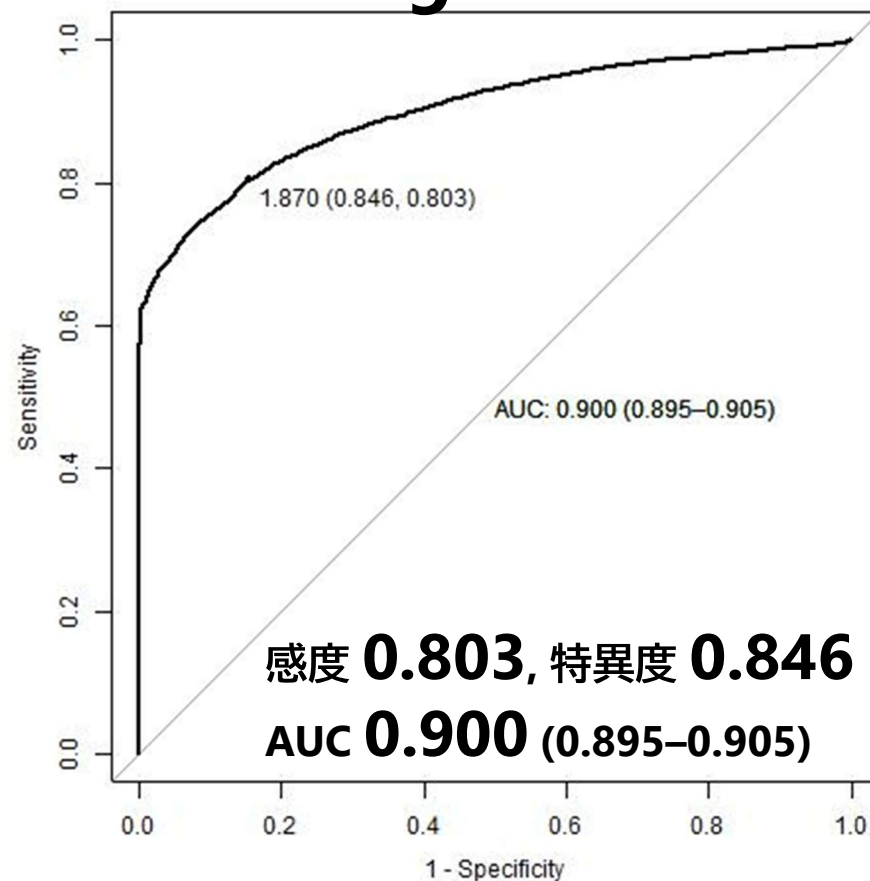
294,442人  高血圧コントロール不良者 **106,647**人(36.2%)

糖尿病コントロール不良予測の最適モデル

CART

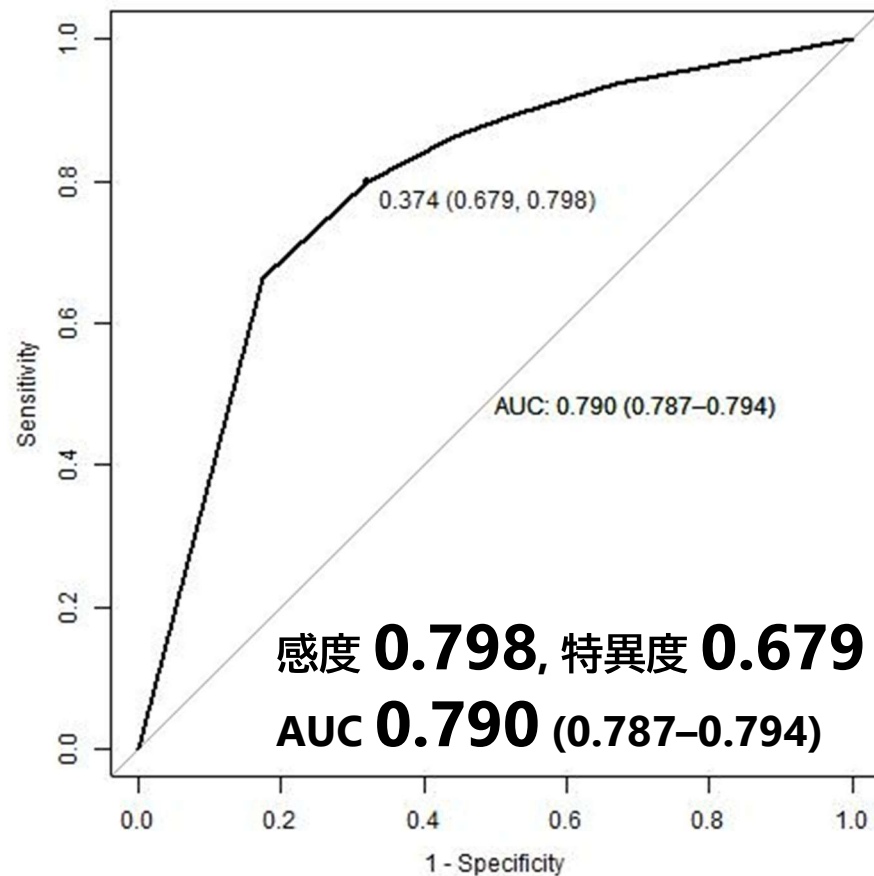


Logistic

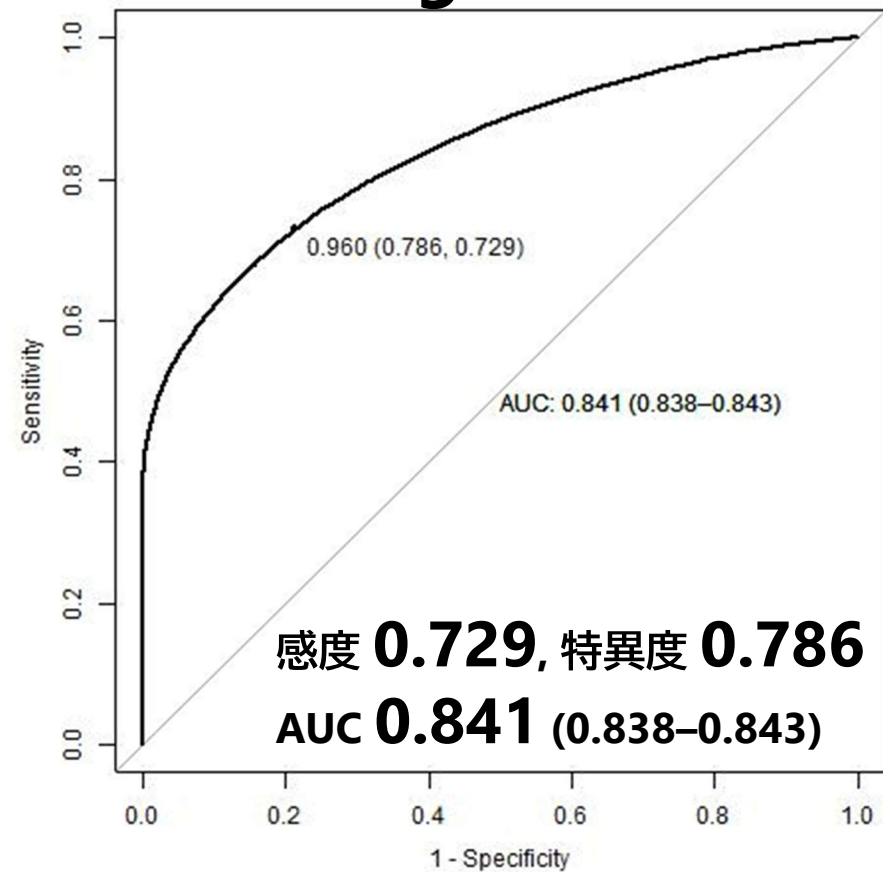


高血圧コントロール不良予測の最適モデル

CART



Logistic



モデル・データ別のコントロール不良の予測性能

		健診データ				
アウトカム	モデル	感度	特異度	PPV	NPV	AUC
糖尿病 5年間の コントロール不良	CART	0.842	0.763	0.965	0.383	0.816
	Logistic	0.803	0.846	0.976	0.357	0.900
糖尿病 翌年の コントロール不良	CART	0.814	0.651	0.785	0.691	0.733
	Logistic	0.751	0.726	0.811	0.650	0.803
高血圧 5年間の コントロール不良	CART	0.798	0.679	0.858	0.581	0.790
	Logistic	0.729	0.786	0.892	0.545	0.841
高血圧 翌年の コントロール不良	CART	0.749	0.688	0.575	0.829	0.761
	Logistic	0.714	0.726	0.596	0.818	0.792

モデル・データ別のコントロール不良の予測性能

アウトカム	モデル	健診データのみ					健診データ+レセプトデータ				
		感度	特異度	PPV	NPV	AUC	感度	特異度	PPV	NPV	AUC
糖尿病 5年間の コントロール不良	CART	0.842	0.763	0.965	0.383	0.816	0.768	0.883	0.983	0.305	0.859
	Logistic	0.803	0.846	0.976	0.357	0.900	0.802	0.842	0.978	0.323	0.900
糖尿病 翌年の コントロール不良	CART	0.814	0.651	0.785	0.691	0.733	0.805	0.654	0.789	0.677	0.748
	Logistic	0.751	0.726	0.811	0.650	0.803	0.742	0.729	0.817	0.634	0.801
高血圧 5年間の コントロール不良	CART	0.798	0.679	0.858	0.581	0.790	0.774	0.710	0.864	0.570	0.787
	Logistic	0.729	0.786	0.892	0.545	0.841	0.716	0.791	0.890	0.540	0.836
高血圧 翌年の コントロール不良	CART	0.749	0.688	0.575	0.829	0.761	0.744	0.689	0.563	0.833	0.759
	Logistic	0.714	0.726	0.596	0.818	0.792	0.720	0.711	0.574	0.825	0.788

まとめ

- CARTモデルとロジスティック回帰モデルによって
治療中断及びコントロール不良の予測モデルを構築した。
- 予測能は中等度
 - ・ 治療中断モデル AUC 0.6～0.7
 - ・ コントロール不良モデル AUC 0.7～0.9
- 既存のレセプトデータと健診データを用いることで、
治療中断及びコントロール不良を一定程度予測できる
ことが示唆された。