

2019/10/29 大学院特別講義@東京理科大学

予後予測モデルの構築

北海道大学 医学統計学
横田 勲

今日の内容

- ▶ Time-to-eventアウトカム・生存時間解析
- ▶ 予後予測モデルの構築
- ▶ 予後予測モデルの性能評価
- ▶ 動的予測

Time-to-event アウトカム

- ▶ 連続量、カテゴリカルのほか、医学研究でよく登場するアウトカム
- ▶ あらかじめ定義した「イベント」が起こるまでの時間
 - ▶ 死亡、再発、入院、ある基準の達成、など
 - ▶ at risk : まだイベントを起こしていない状態

打ち切りのあるデータ

- ▶ ある時点までイベントを起こしていない
- ▶ その先で起こるはずのイベントの正確な時点が分からない
 - ▶ 脱落や研究終了等による
- ▶ 適切に考慮する解析方法が生存時間解析
 - ▶ 単に除外すると有病率を過大評価しがち
 - ▶ イベントなしとすると有病率を過小評価
 - ▶ 無情報な打ち切りの仮定

Kaplan-Meier法①

- ▶ 直前までat riskである人について、イベントを起こさなかった確率を乗じる
 - ▶ 生存例は、それまでの間、常に生存してきた
- ▶ 以下のデータセットを想定

イベント発生時点 (年)	内容
1	死亡 (イベント)
2	死亡 (イベント)
3	脱落 (打ち切り)
4	死亡 (イベント)
5	研究終了 (打ち切り)

Kaplan-Meier法②

5人中1人死亡したため、生存率は $1 - \frac{1}{5} = 0.8$

5人中2人死亡したため、生存率は $1 - \frac{2}{5} = 0.6$

直前での生存率は0.8であり、4人中1人死亡したため、 $0.8 \times (1 - \frac{1}{4}) = 0.6$

直前での生存率は0.6であり、2人中1人死亡したため、 $0.6 \times (1 - \frac{1}{2}) = 0.3$

打ち切り例の扱い 7

- ▶ 3年で打ち切りとなった対象者
 - ▶ 1年、2年での生存率を計算する際には、at riskであった人として解析に寄与
 - ▶ 4年、5年での生存率計算では分母に入らず
 - ▶ 生存率の計算自体には反映されている

追跡開始時は5人でスタート 8

- ▶ 1人あたり、20%の確率をもつ
- ▶ 3年で打ち切りとなった人の予後は分からない
- ▶ 3年でat riskな人の予後で置き換えよう
- ▶ 4年では、
 2人 + 1人 = 30%
 だけ生存率が低下

無情報な打ち切り noninformative censoring 9

- ▶ Kaplan-Meier法、ログランク検定、Cox回帰で置かれる仮定
 - ▶ ランダムな打ち切り、とも
- ▶ 打ち切りとイベント発生が無関係
 - ▶ 研究終了時の生存
 - ▶ 偶然の事故による追跡不能
- ▶ 打ち切り例の予後を、at risk例で置き換えるため

解析の目的 10

- ▶ ランダム化臨床試験
 - ▶ ログランク検定；生存曲線が群間で異なるか
 - ▶ Cox回帰モデル；治療効果の大きさを評価
- ▶ 予測モデルの構築
 - ▶ イベントの発生しやすさをモデル化
 - ▶ ロジスティックモデル；特定の時点における発生有無に注目
 - ▶ Cox回帰モデル；期間全体における発生しやすさに注目
 - ▶ 作ったモデルの予測性能を評価

予測モデル構築の流れ 11

12

Kobayashi T, et al. Blood Cancer J. 2016. e383

DLBCLの新規予後予測モデル 13

- ▶びまん性大細胞型B細胞リンパ腫
- ▶全生存予後を予測したい
- ▶臨床で簡単に利用できるスコアを作りたい
 - ▶年齢、血清LDH、Ann Arborステージ、ECOG-Performance Status、血清CRP、低アルブミン血症、節外（骨髄、骨、皮膚、肺/胸膜）病変
 - ▶変数選択により、予測に用いる因子を決定

変数選択 14

- ▶回帰分析において、複数の因子候補から、関連の強そうなものだけに絞る方法
 - ▶予測モデルをシンプルにするためには便利
- ▶目的やデータの特徴に応じた使い分け
 - ▶変数増加法 forward
 - ▶変数減少法 backward
 - ▶ステップワイズ法 stepwise
 - ▶LASSO法、elastic net法

モデル構築のアプローチ 15

Collins GS, et al. *Ann Intern Med.* 2015. TRIPOD guideline

ランダム分割を採用 16

- ▶465例のデータ
 - ▶323例(70%)をトレーニングコホート
 - ▶142例(30%)をバリデーションコホート
- ▶トレーニングコホートで予測モデルを構築
- ▶バリデーションコホートで他の予測モデルとの性能を比較
 - ▶モデル構築に用いていないデータであるため、公平な性能比較を行えるだろう

最終モデル 17

- ▶変数減少ステップワイズ法を利用

因子	ハザード比	95%信頼区間	回帰係数	スコア
LDH ≤ 1×ULN	1	-	0	
LDH > 1×ULN, ≤3×ULN	2.47	1.20-5.08	0.91	1点
LDH > 3×ULN	3.68	1.57-8.66	1.31	2点
ECOG-PS ≥ 2	2.50	1.40-4.45	0.91	1点
ALB < 3.5mg/dL	2.52	1.36-4.69	0.93	1点
特定部位への節外病変	1.71	1.03-2.84	0.54	1点

- ▶合計点を基にさらにリスク分類

合計点	0点	1-2点	3点	4-5点
リスク分類	低	低中間	高中間	高

各コホートでのリスク分類 18

トレーニング バリデーション

予測性能指標による評価 19

- ▶ 予後の悪い対象者を特定するための予測モデルがどれだけ有用かを知りたい
- ▶ 他の予測モデルと比較したい
- ▶ 予測モデルを構築する上で、overfittingを避けたい
 - ▶ ノイズまでモデルをあてはめてしまい、将来の対象者への予測性能が悪くなること

ところで「予測性能がよい」とはどういうこと？

生存時間解析における予測指標 20

- ▶ すべて二値の予測指標を基に拡張
 - ▶ ある程度、研究が進んだと感じるもの
 - ▶ Brierスコア (予測誤差)
 - ▶ C-index (判別能力)
 - ▶ 予測確率をw年生存率で置き換え、二値の議論をそのまま適用するもの
 - ▶ Net Benefit (臨床的有用性)
 - ▶ Calibration plot (較正)
 - ▶ Net Reclassification Improvement (再分類)

Notation 21

- ▶ 二値アウトカム $D (= 0, 1)$
 - ▶ イベント発生を $D = 1$ 、未発生を $D = 0$
- ▶ 生存時間アウトカム T
 - ▶ イベント発生までの時間、短いほど予後が悪い
 - ▶ 観察イベント/打ち切り時点 \tilde{T}
 - ▶ イベント有無の指示変数 $\delta = I(\tilde{T} = T)$
- ▶ 予測変数 (マーカー) M
- ▶ イベント発生の予測確率 $P = g(M)$
 - ▶ M と P の間は単調増加関係
- ▶ M や P が大きいほど、予後が悪いという関係

Brierスコア 22

- ▶ 2値アウトカムの偏差平方
 - ▶ 連続量アウトカムでは、Mean Squared Errorとしておなじみ
- ▶ 期待Brierスコアの上限
 - ▶ 無情報 (切片だけ) モデル

$$\underbrace{\{1 - S(t)\}}_{\text{イベント発生確率}} \underbrace{\{0 - S(t)\}^2}_{\text{イベント発生例の Brierスコア}} + \underbrace{S(t)}_{\text{イベント未発生確率}} \underbrace{\{1 - S(t)\}^2}_{\text{イベント未発生例の Brierスコア}}$$

Graf E, et al. Stat Med. 1999. 2529-2545.

期待Brierスコアの範囲 23

- ▶ 上限は $S(t)\{1 - S(t)\}$
 - ▶ 最大でも0.25 ($S(t) = 0.5$ のとき)
- ▶ 下限は予測したい対象者の真のモデルを特定できた時
 - ▶ 特定できたか実際のデータからは不明
 - ▶ 予測したい対象者が真のモデル $S(t|M)$ に従う

$$\underbrace{S(t|M)\{1 - S(t|M)\}}_{\text{真のバラツキ (期待値の下限)}} + \underbrace{\{S(t|M) - S(t|M)\}^2}_{\text{モデル誤差 (今はゼロ)}}$$

平均Brierスコアの数値例 24

- ▶ 2人死亡、2人生存という仮想例
- ▶ 無情報モデル

ID	生存 / 死亡	予測確率	Brierスコア
1	死亡	0.5	$(1 - 0.5)^2 = 0.25$
2	死亡	0.5	$(1 - 0.5)^2 = 0.25$
3	生存	0.5	$(0 - 0.5)^2 = 0.25$
4	生存	0.5	$(0 - 0.5)^2 = 0.25$

平均Brierスコア 0.25

- ▶ 予測モデル

ID	生存 / 死亡	予測確率	Brierスコア
1	死亡	0.9	$(1 - 0.9)^2 = 0.01$
2	死亡	0.6	$(1 - 0.6)^2 = 0.16$
3	生存	0.3	$(0 - 0.3)^2 = 0.09$
4	生存	0.2	$(0 - 0.2)^2 = 0.04$

平均Brierスコア 0.075

相対Brierスコア減少 25

- ▶ 期待Brierスコアのとりうる範囲は0から0.25
 - ▶ しかも0に近いほど「予測性能がよい」
 - ▶ 集団全体の生存確率によって、上限が変化
- ▶ 無情報モデルに対する、予測モデルでの期待Brierスコアを小さくした割合
 - ▶ 0から1をとり、1に近いほど「予測性能がよい」

$$\frac{\text{Brier}_{\text{無情報モデル}} - \text{Brier}_{\text{予測モデル}}}{\text{Brier}_{\text{無情報モデル}}}$$

打ち切りを含むデータでの推定 26

- ▶ 対象者はいずれか3パターン
 1. $I(\tilde{t}_i > t)$ (t でイベント未発生)
 2. $I(\tilde{t}_i \leq t)$ かつ $\delta_i = 1$ (t でイベント発生)
 3. $I(\tilde{t}_i \leq t)$ かつ $\delta_i = 0$ (t での状態不明)

IPCW法の利用 27

- ▶ 時点 t にて打ち切りがない確率 $G(t)$
 - ▶ 例えばKaplan-Meier法で推定
- ▶ パターン3のBrierスコアが計算不能
 - ▶ パターン1,2のBrierスコアを打ち切られない確率の逆数で膨らませる

$$\frac{1}{n} \left[\sum_i \frac{1}{\hat{G}(t)} \{0 - \hat{S}(t|M_i)\}^2 + \sum_i \frac{1}{\hat{G}(\tilde{t}_i)} \{1 - \hat{S}(t|M_i)\}^2 \right]$$

全員 パターン1は評価時点 t の打ち切りなし確率 パターン2は各イベント時点 \tilde{t}_i の打ち切りなし確率

判別 discrimination 28

- ▶ アウトカムの異なる対象者を分ける予測
- ▶ 連続変数 M を用いて、二値アウトカム $D(=0,1)$ を診断
 - ▶ 感度: $\Pr(M > c | D = 1)$
 - ▶ 特異度: $\Pr(M < c | D = 0)$
 - ▶ 偽陽性率: $1 - \text{特異度}, \Pr(M > c | D = 0)$

カットオフ値をもって評価 29

イベントなし ($D=0$) イベントあり ($D=1$)

予測変数 M

カットオフ値

陰性と判断 陽性と判断

特異度 感度

感度と特異度はトレードオフ 30

イベントなし ($D=0$) イベントあり ($D=1$)

評価値

カットオフ値

- ▶ 感度を上げれば、特異度は下がる
- ▶ 特異度を上げれば、感度は下がる

ROC曲線 31

- ▶ 縦軸に感度、横軸に偽陽性率（1-特異度）
- ▶ カットオフ値を全範囲で動かした場合の感度と偽陽性率をプロット

AUC (Area Under the Curve)

(ROC-)AUC 32

- ▶ ROC曲線の要約指標
 - ▶ 判別能力を表す指標として解釈
- ▶ AUC自体はモデルに依らずに計算される
 - ▶ AUC=0.5であれば、no discriminative ability
 - ▶ AUCが1に近づくほど、判別能力がよい
 - ▶ 絶対的な解釈は困難

AUCを代数的に表現 33

- ▶ M_1 は $M|D = 1$ から、 M_0 は $M|D = 0$ から独立にサンプリング
 - ▶ n_0, n_1 はそれぞれ $D = 0, D = 1$ の人数

$$(n_0 n_1)^{-1} \sum_{i=1}^{n_0} \sum_{j=1}^{n_1} \{I(m_{1i} > m_{0j}) + 0.5I(m_{1i} = m_{0j})\}$$

- ▶ イベント有無と予測変数の
 - ▶ 大小関係が一致していれば1点
 - ▶ 大小関係が一致していなければ0点
 - ▶ 値が等しければ0.5点（引き分け）

C (concordance) index 34

- ▶ アウトカムと予測変数の大小関係が一致した (concordantであった) 割合
 - ▶ Kendall's τ 、Goodman-Kruskal's γ 、Somers's Dの拡張ともみなせる
 - ▶ 二値アウトカムの場合、ROC-AUCと一致

Harrell FE, et al. JAMA. 1982. 2543-2546.
Harrell FE, et al. Stat Med. 1996. 361-387.

c-indexの数値例① 35

- ▶ 死亡例の予測確率：A:0.9, B:0.7, C:0.4
- ▶ 生存例の予測確率：D:0.6, E:0.3, F:0.1
- ▶ 総当たり表

		生存例			
		0.6	0.3	0.1	
死亡例	0.4	×	○	○	c-indexは 8/9=0.89
	0.7	○	○	○	
	0.9	○	○	○	

○：一致 ×：不一致

c-indexの数値例② 36

- ▶ 死亡例の予測確率：A:0.9, B:0.7, C:0.4
- ▶ 生存例の予測確率：D:0.6, E:0.3, F:0.1

c-indexの数値例③

37

- ▶ 死亡例の予測確率：A:0.9, B:0.7, C:0.4
- ▶ 生存例の予測確率：D:0.6, E:0.3, F:0.1

感度

1-特異度

死亡したAの0.9が生存したDの0.6より大きい（一致）

死亡したCの0.4が生存したDの0.6より小さい（不一致）

生存時間解析のc-index

38

- ▶ 生存時間の短長関係と死亡確率の高低がそろえば一致
- ▶ 生存時間が長いのに死亡確率が高い場合は不一致
- ▶ 打ち切りを考慮した、Uno's c-indexを利用

Uno H, et al. *Stat Med.* 2011. 1105-1117.

DLBCL予測モデル研究

39

	PFS		OS	
	c-index	RBSR	c-index	RBSR
R-IPI	0.668	0.122	0.642	0.135
NCCN-IPI	0.749	0.172	0.736	0.251
提案スコア(4段階)	0.703	0.183	0.740	0.305
元の0-5点スコア	0.711	0.215	0.754	0.356

RBSR：相対Brierスコア減少

- ▶ 提案スコアが従来スコアより概ね性能がよいことを示した

キャリブレーション（校正）

40

- ▶ 予測リスクと実際のリスクの近さ
- ▶ Hosmer-Lemeshow統計量
 - ▶ 予測確率の順に集団を均等に分割し、各グループの粗イベント割合と予測確率の平均値を比較
- ▶ キャリブレーションプロット
 - ▶ 各グループの粗イベント割合と予測確率の平均値をプロット

生存時間解析でのcalibration plot

41

Teramukai S, et al. *J Clin Oncol.* 2007. 3302-3306.

- ▶ 粗生存率と予測生存率の平均をプロット

観察イベント確率を平滑化

42

Austin PC, Steyerberg EW. *Stat Med.* 2014. 517-535.

Overfittingの検討

43

- ▶ ブートストラップサンプルでの粗生存率と予測生存率の平均をプロット
- ▶ これを繰り返し、重ねてプロット

Colling GS, et al. *Stat Med*. 2016. 4124-4135.

Calibration (plot) の指標

44

- ▶ \hat{p} : 予測確率、 \hat{p}_c : 平滑化した観察確率
- ▶ $E_{\max}(a, b) = \max_{a \leq \hat{p} \leq b} |\hat{p} - \hat{p}_c|$
 - ▶ Harrell FE. *Regression Modeling Strategies*. Springer. 1996.
- ▶ $ICI = \int_0^1 |\hat{p} - \hat{p}_c| \phi(\hat{p}) d\hat{p}$
 - ▶ $\phi(\hat{p})$: 予測確率の分布の密度関数
 - ▶ Integrated Calibration Index
 - ▶ Austin PC, Steyerberg EW. *Stat Med*. 2019.

治療途中における予測

45

- ▶ ある時点 s まで生存した下での条件付き推測
 - ▶ 例) 診断後2年生存した対象者の4年後の死亡確率
- ▶ 動的予測 (dynamic prediction)
 - ▶ $s = 0$ にて通常の w 年生存割合と同様

$$\Pr\{X \in (s, s + w] \mid X > s\}$$

- ▶ X : イベント時点を表す確率変数

経時データ $Z(t)$ の利用

46

- ▶ 予測開始時点 s までに観察された患者の経過をせっかくなので取り入れたい
- ▶ 時間依存性共変量Coxモデル
 - ▶ 対象者自身の健康状態を表す変数は使えない

- ▶ $Z(t)$ が観察されたということは、対象者が生存していることが確定
- ▶ ハザード λ との対応が不成立

$$\Pr\{T > t \mid Z(t)\} \neq \exp\left[-\int_0^t \lambda\{u \mid Z(u)\} du\right]$$

ランドマークモデル

47

- ▶ ランドマーク時点 s における at risk 例で条件付け
- ▶ 時点 s までに得られた情報を利用
- ▶ 単一の生存時間である場合
(vanHouwelingen 2007, *Scand Stat Theory Appl*)

$$\lambda(t|s, Z(t)) = \lambda_0(t|s) \exp\{Z(s)^T \beta(s)\}$$

- ▶ $Z(s)$ はこのモデルにおいてベースライン共変量と同等

多変量生存時間に拡張

48

Yokota I, Matsuyama Y. *BMC Med Res Methodol*. 2019.

- ▶ 大腸がん肝転移追跡例データ
 - ▶ 東京大学医学部附属病院肝胆膵外科にて大腸がん肝転移の先行肝切除を受けた336名
 - ▶ 1996年1月～2010年12月
 - ▶ 補助化学療法の行われなかった263名を解析対象

切除不能再発が起こっていない対象者 49

- ▶ 3年後に以下のいずれにあるかを予測
 - ▶ 前回切除時の腫瘍情報も利用可能
 - ▶ 初回肝切除後の再発回数 (0回 / 1回以上)
 - ▶ 腫瘍個数 (単発 / 多発)

予測開始時点

- ▶ 無再発
 - ▶ 初回肝切除時の腫瘍情報
- ▶ 再発あり, 切除可能
 - ▶ 前回再発時の腫瘍情報

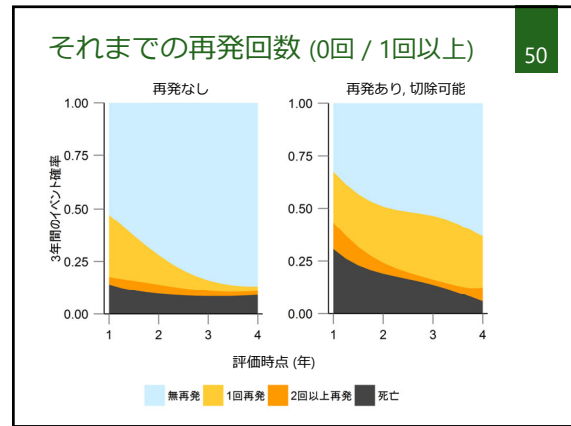
予測開始時点から3年後

無再発

1回再発

2回以上再発

死亡



おわりに 51

- ▶ 予後予測モデルの構築と予測性能のいくつかの側面を表す指標
 - ▶ Brier score
 - ▶ ROC-AUC, c-index
 - ▶ Calibration
- ▶ 動的予測による経時データの組入