

「統計解析入門⑤： 生存時間解析～その3：応用編～」

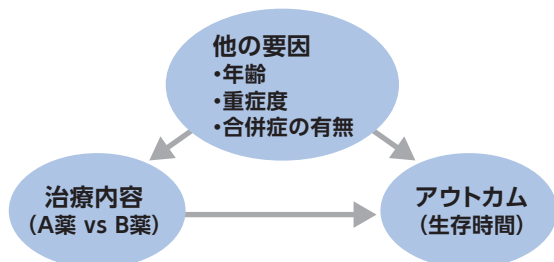
総合医学研究センター 医療統計室 室長・准教授 伊藤 ゆり



生存時間データの多変量解析： Cox比例ハザードモデル

今回は生存時間解析の応用編です。今回は少しだけ数式を使わせてください。第8回で学んだ交絡因子の対応として、層別化、標準化がありますが、それよりも多くの交絡因子を調整したい際に適用する多変量解析の手法について紹介します。今、新薬Aと従来薬Bの効果の違いを生存時間で比較したいとします。この場合、研究デザインとしては無作為化比較試験による検討が最適なのですが、今回はリアルワールドデータによる観察研究で検討することとします。その場合、治療薬の選択およびアウトカム(生存時間)の両方に影響を与える交絡因子が存在します(図1)。

図1. 交絡因子のイメージ

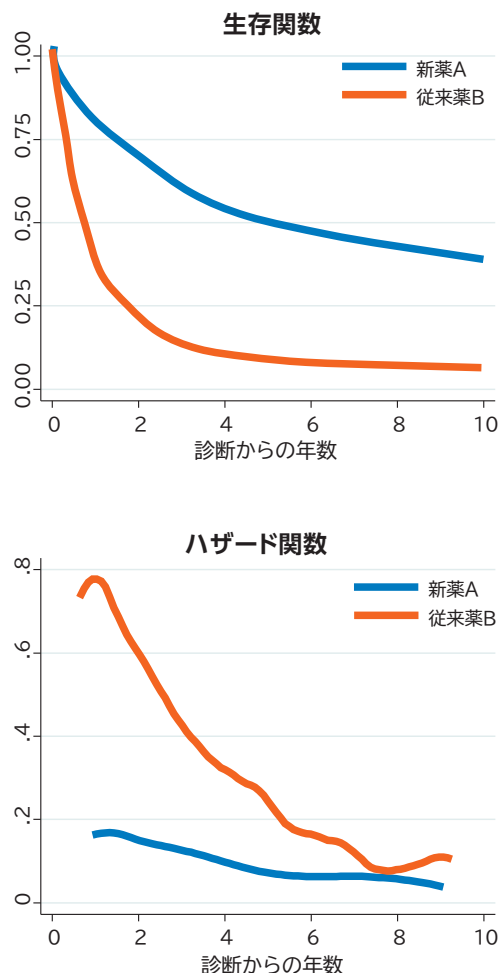


そのため、年齢や重症度、合併症の有無など、治療薬の選択と生存時間に影響を与える交絡因子を調整したうえで、新薬Aと従来薬Bのどちらの生存時間が長くなるのかを分析します。複数の要因を一度に検討する際に用いるのが多変量解析という手法で、生存時間データで用いる場合にはCox比例ハザードモデルを適用します。

このモデルを適用する上で、生存関数とハザード関数の関係を知っておく必要があります。

前回学んだKaplan-Meier法により、生存関数を推定します。この曲線の落ちていく速度、つまり各瞬間の死亡速度を示したものが下のハザード関数になります。Cox比例ハザードモデルではこのハザード関数を比較して、死亡するリスクの大きさを比較することになります(図2)。

図2. 生存関数とハザード関数



Cox比例ハザードモデルは時間 t としたハザード関数 $h(t)$ 、治療薬(従来薬B： $x_1=0$ 、新薬A： $x_1=1$)、合併症の有無(無： $x_2=0$ 、有： $x_2=1$)、ステージ(I： $x_3=0$ 、 $x_4=0$ 、II：

かなり役立つ生涯学習 「統計解析入門⑤：生存時間解析～その3：応用編～」

$x_3=1, x_4=0$, III : $x_3=0, x_4=1$)とします。
そのとき、Cox比例ハザードモデルは以下の式1のようになります。

数式 1

$$h(t|x_1, x_2, x_3, x_4) = h_0(t) \exp(\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4)$$

従来薬使用、合併症無、Stage Iの場合、すべての x に0が入り、この時のハザード関数を、ベースラインハザード $h_0(t)$ と言います。説明を簡単にするために、まず、治療薬のハザード比を推定する単変量解析を想定します。

$$h(t|x_1) = h_0(t) \exp(\beta_1 x_1)$$

この時に、 β_1 が-0.231と推定されたとした場合の従来薬Bに対する新薬Aのハザード比は

$$\begin{aligned} \text{ハザード比} &= \frac{h_0(t) \exp(-0.231)}{h_0(t)} \\ &= 0.794 \end{aligned}$$

となります。他の変数も同時に考慮する場合、式1に当てはめを行い、すべての β を推定し、自然対数 e の β 乗、つまり $\exp(\beta)$ すると、それぞれのハザード比が得られます。ハザード比が0.794、つまり1を下回っているので、従来薬Bに対する新薬Aの死亡のリスクが20.4%低くなっているという結果になります(図3①の結果)。

ただし、これは治療薬の変数しか入っていないため、他の交絡因子の検討をする必要があります。通常、論文などにはそれぞれの単変量解析の結果(図3②は合併症の有無の変数だけを入れた場合の結果、図3③はステージの変数を入れた場合の結果)を提示しておき、その横にすべての変数をモデルに含めた結果(図3④の部分が多変量解析による結果)を合わせて提示し、治療薬の違いにおいて、合併症の有無や、ステージの違いによる影響を補正したハザード比を評価することになります。今回の場合、単変量解析で見た新薬Bの従来薬Aに対するハザード比は0.79でしたが、他の交絡因子を考慮すると、その効果は少し薄まり0.82となります

図3. Cox比例ハザードモデルの結果の示し方

	単変量解析			多変量解析		
	HR	95%CI	p-value	HR	95%CI	p-value
① 従来薬B	(ref)			④ (ref)		
新薬A	0.79	0.68-0.89	0.01	0.82	0.76-0.93	0.02
② 合併症 無	(ref)			(ref)		
有	1.56	1.13-1.78	0.03	1.32	1.21-1.41	0.03
③ ステージ I	(ref)			(ref)		
II	5.01	3.46-7.25	<0.001	2.33	1.54-3.43	<0.001
III	10.60	7.52-14.95	<0.001	4.42	2.95-6.64	<0.001

が、18%の死亡リスクの低下ということになります。

ステージの変数は今回、I, II, IIIとし、順序の関係性がありますが、通常連続値としてモデルに投入することではなく、今回示したような、ステージIと比較したII, IIIというダミー変数という形で投入します。3つ以上のカテゴリカル変数を統計モデルに投入する場合に行う方法ですので、知っておいてください。この場合、ステージIが**reference (ref)** 参照群となります。

なお、比例ハザードモデルを適用する場合に、比例ハザード性の前提という条件があります。これは、比較する二群のハザードの比が診断後の経過時間に応じて変化しないという前提です。ハザード自体は変化しても構いませんが、その関係性としての比が変化しないというものです。通常、生存率を $-\ln(-\ln(S(t)))$ に式変換したものをY軸に、時間tは $\ln(t)$ をX軸にしたプロットで、二つの線が平行になっているかを確認したり、Schoenfeld残差のプロットというもので確認したりすることがあります。今回学んだCox比例ハザードモデルは生存時間を扱う論文で多く使うものですので、解析の方法や結果の解釈について、知っておいていただけるとよいと思います。

文献

※1：佐藤弘樹, 市川度. 生存解析がこれでわかる!
臨床統計まるごと図解. 中山書店

略歴

大阪大学大学院医学系研究科博士前期・後期課程卒業後、大阪府立成人病センター(現大阪国際がんセンター)リサーチ・レジデント、研究員、主任研究員を経て、大阪医科薬科大学 総合医学研究センター医療統計室 准教授(現職)。現在、がん疫学、健康格差、医療統計の研究に主に従事。