

## 4. 治療効果判定のための実用的方法

— 予後因子を補正した生存率曲線の検定 —

青木国雄\*, 水野正一\*, 浜島信之\*,  
佐々木隆一郎\*, 浅野明彦\*, 勝田信行\*

### 〔はじめに〕

治療効果を評価する際には、通常生存率曲線が計算され、さらに2群間での生存率曲線の優劣を統計学的に検討するために検定が行なわれる。これらの具体的方法については、日本癌治療学会においても規約が作られ<sup>1)</sup>、医学研究者の間でも一定の理解に達したと言えよう。

しかしながら、2群の治療法を比較する際に常に問題となるのは、各群での予後因子の偏りであって、この点を明確にしなければ治療効果を正確に判定することは不可能となる。いずれの研究においても予後因子が偏らないよう最善の努力が払われているのであるが、実際には偏りをなくすことはかなり難しいのが現状であろう。多要因を同時に補正して治療効果をみる手法として Cox の Proportional Hazard Model<sup>2)</sup>が注目されているが、わが国では汎用のプログラムが未だ入手し難い。

そこで、一般医学研究者が実際に用いることができるような手法で、予後因子を補正する実用的な一方法を考案したので紹介する。

### 〔方法〕

比較したい治療法をA、Bとする。治療法以外で予後を左右する重要な因子を $(\alpha_1, \alpha_2, \dots, \alpha_n)$ とする。治療の対象者全体をSとし、これを予後因子 $(\alpha_1, \alpha_2, \dots, \alpha_n)$ によって層別化し、それらの群 $S_1, S_2, \dots, S_m$ では、予後因子をそろえておく。先ず各群 $S_k$  ( $k=1, 2, \dots,$

$m$ )において、治療法A、Bの差を Generalized Wilcoxon Test<sup>3)</sup>によって検討する。日本癌治療学会規約<sup>1)</sup>にしたがうと、以下のように計算できる。

以下の計算は先ず群 $S_k$ での計算である。治療法A群の生存期間を $x$ 、治療法B群の生存期間を $y$ で表す。群 $S_k$ での治療法Aの対象数を $n_1$ 、治療法Bの対象数を $n_2$ とする。

治療法A群の観察打ち切り時生存例を

$$x_1', \dots, x_{r_1}' : r_1 \text{ 例生存}$$

死亡例を

$$x_{r_1+1}, \dots, x_{n_1} : (n_1 - r_1) \text{ 例死亡}$$

とする。

治療法B群の観察打ち切り時生存例を

$$y_1', \dots, y_{r_2}' : r_2 \text{ 例生存}$$

死亡例を

$$y_{r_2+1}, \dots, y_{n_2} : (n_2 - r_2) \text{ 例死亡}$$

とする。

次にA群とB群の症例を各ペアに対してスコア $U_{ij}$ を次のように計算する。

$$U_{ij} = \begin{cases} -1 : x_i < y_j \text{ 又は } x_i \leq y_j' \\ 0 : x_i = y_j \text{ 又は } x_i' = y_j' \\ \text{又は } x_i' < y_j \text{ 又は } x_i > y_j' \\ +1 : x_i > y_j \text{ 又は } x_i' \geq y_j \end{cases}$$

群 $S_k$ において各ペアに対するスコアの総和

$$W_k = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} U_{ij}$$

を計算する。次にこの $W_k$ の分散統計量 $V_k$

$$V_k = \frac{n_1 n_2}{(n_1 + n_2)(n_1 + n_2 - 1)} \sum_{i=1}^{n_1 + n_2} U_i^2$$

\* 名古屋大学医学部予防医学教室

を計算する。ただし、 $U_i$  は各症例  $i$  を残りの全症例の観察期間と比較して、 $U_{ij}$  の場合と同じ規則で計算したスコアである。

以上のように、各群  $S_k$  において治療効果判定のもととなる統計量  $W_k$  と、その分散統計量  $V_k$  を計算する。そしてまとめの統計量となる

$$W = \sum W_k$$

を計算する。この統計量  $W$  の分散推定量  $V$  は、各群  $S_k$  が独立されているという条件のもと

$$V = \text{var}(W) = \text{var}(\sum W_k) = \sum \text{var}(W_k) = \sum V_k$$

と計算できる。それゆえ、このまとめの統計量  $W$  に対して標準化 ( $Z$  変換) を行い、

$$Z = \frac{W}{\sqrt{V}}$$

この統計量  $Z$  を、平均 0 分散 1 の標準正規分布により有意差検定 ( $Z$  検定) する。

### 〔資料〕

応用例として、当班の東京小児白血病グループの 8 次案による治療成績結果<sup>9)</sup>を用いた。8 次案には頭蓋照射群 (A 群) と MTX 髄注群 (B 群) があるが、その詳細<sup>10)</sup>は省略する。

ここでは、我々の方法を適用し、一つの子後因子を補正して A 群、B 群を比較してみた。年齢の偏りを除くため、対象者を 2~5 歳の 113 名に限定し、一つの子後因子として初診時の白血球数を用いた。この外にも、子後因子は考えられるであろうが、重要な子後因子の一つであったので例としてこれを選んだ。

層別化のために、判別値を仮に 20,000/cmm とした。

### 〔結果〕

Kaplan-Meier<sup>9)</sup>法により計算した治療法別の生存率曲線の単純比較を、図 1 に示した。治療法 A のほうが予後がよいのが分かる。有意水準を Generalized Wilcoxon Test の結果でみると、 $Z = 2.723$  で有意差 ( $p < 0.01$ ) が得られる。図 2 に、初診時白血球数が 2 万以上の群と未満の群の単純比較を示した。初診時白血球が少ない群のほ

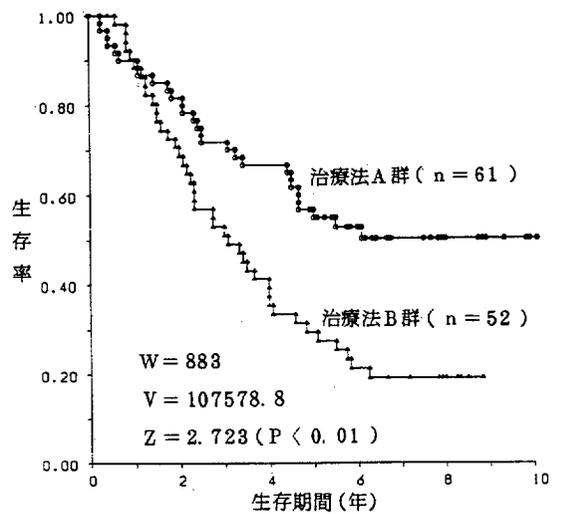


図 1 治療法群別の生存率曲線

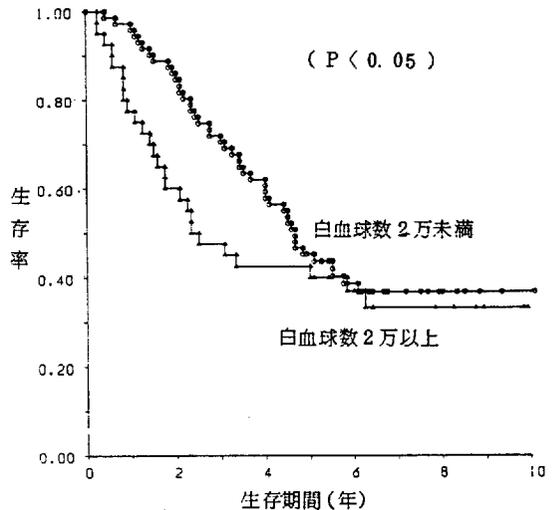


図 2 白血球数別生存率曲線

うが予後がよい ( $p < 0.05$ )。そこで、白血球が 2 万未満の場合と 2 万以上の場合と別々に、A 群と B 群の生存率曲線を計算し、図 3、図 4 に比較を示した。白血球数が 2 万未満の場合には、検定統計量  $Z = 1.732$  万以上の場合には  $Z = 1.82$  となり、A 群の生存率のほうが良好である傾向にあるが、どちらも 5% の有意水準に達していない。しかし、前述のまとめの統計量を計算すると、

$$W = W_1 + W_2 = 289 + 131 = 420$$

$$V = V_1 + V_2 = 28073.6 + 5164.0 = 33237.6$$

$$Z = 420 / \sqrt{33237.6} = 2.30 \quad (p < 0.025)$$

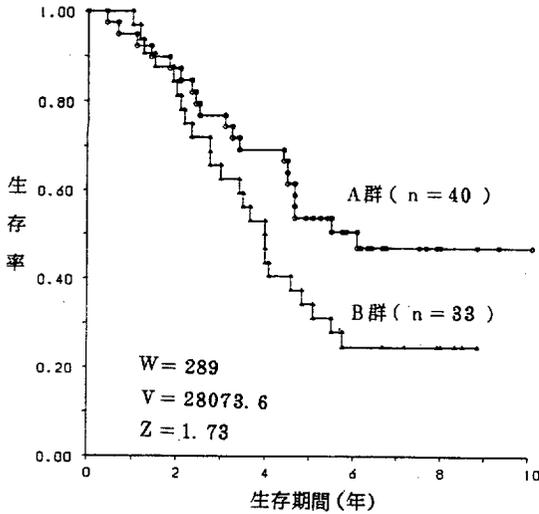


図3 白血球数2万未満における治療法別生存率曲線

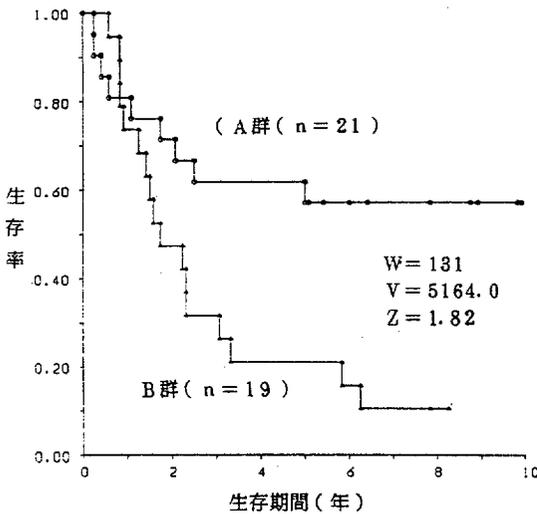


図4 白血球数2万以上における治療法別生存率曲線

となり有意水準に達し、白血球数を補正した結果においても、A群のほうがB群より統計学的によいとの結論が得られる。

〔考 察〕

Gehan<sup>3)</sup>が治療効果判定のために Generalized Wilcoxon Test を導入した時は、各治療群への症例の割当ては無作為化 (randomized) を想定して、かかる状況下では、治療群間の予後因子の

偏りはないものとしている。しかしながら、実際には randomized trial であっても、偶然片方の治療法に予後の比較的良好な患者が集まることもあり、予後因子をそろえて全体を層別化し、再検討することが重要となる。

また、historical control との比較においては、一般に、両者の予後因子を揃えるのは難しいので、予後因子の補正は不可欠となる。

前述の方法により、かかる状況のもとでも予後因子を補正した治療法の検定が可能となる。統計学的原理としては、一般によく用いられている Mantel-Haenszel procedure<sup>7)</sup>と同じように、予後因子を揃えたところで、それぞれの比較統計量を算出し、まとめの統計量によって全体の傾向の有意性を検出するものである。このようにして、他の因子を補正した検定が可能となる。

いずれの群においても、同じような傾向があれば、一つ一つの群においては有意水準に達していなくても、例題のごとく予後因子を補正した全体の有意性が検出できる。

ただし、本方法を用いるにあたり注意すべきは、群によっては治療に対する反応が異なるような場合である。統計学的には層別化された各群に特异性 (heterogeneity) が存在する場合であり、この場合には本方法の結果の解釈には注意を要する。

なお、例題における結果をみると、補正した結果は白血球数で層別化しなかった場合より有意性がやや弱められている。このことは、治療法A群に予後の良い症例がやや集まっているかもしれないと思わせるが、実際表1のごとく、A群には初診時白血球数が少ない症例の割合がやや多く、最終結果には補正の効果が表れたと考えられる。

表1 白血球数別治療法別症例数

白血球数	治療法		計
	A	B	
2万未満	40	33	73
2万以上	21	19	40
計	61	52	113

## 〔要 約〕

二つの治療法 A, B の生存率曲線を, 予後因子の影響を補正し, 検定する統計学的一方法を開発し, 当班の資料に一つの応用を示した。

治療法以外の重要な予後因子によって治療群全体 S を層別化する。層別化された各群を ( $S_1, S_2, \dots, S_m$ ) とする。各群  $S_k$  においては重要な予後因子はそろっているものとする。そして, 各群  $S_k$  において治療法 A と B を Generalized Wilcoxon Test を用い検定する。それより得られた統計量  $W_k$  と, その分散統計量  $V_k$  から全体のまとめの統計量  $W$  とその分散統計量  $V$  を,

$$W = \sum W_k, \quad V = \sum V_k$$

と計算し, そしてこの統計量  $W$  の標準化 ( $Z$  変換)

$$Z = \frac{W}{\sqrt{V}}$$

を計算し, これを正規分布により有意差検定する。この方法により治療期間における予後因子の偏りを補正して, 治療法の検定を容易に行うことができる。

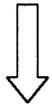
## § 文 献

- 1) 日本癌治療学会: 日本癌治療学会. 生存率算出規約 (案), 日本癌治療学会誌, **19**: 137-162, 1984.
- 2) Cox, D. R.: Regression models and life tables. J. Royal Stat. Soc. Series B, **34**: 187~220, 1972.
- 3) Gehan, E.: A generalized Wilcoxon test for comparing arbitrarily singly-censored samples. Biometrika, **52**: 203-224, 1956.
- 4) 富永祐民: 治療効果判定のための実用統計学. 蟹書房, 1982.
- 5) 青木国雄, 浜島信之, 月本一郎, 山本正生, 植田 穰, 伊勢 泰, 西村昂三: 東京小児白血病研究班の 8 次・9 次治療成績 (第 1 報). 厚生省心身障害研究報告書小児慢性疾患の診断・治療・管理に関する研究班, 昭和 58 年度研究業績. 139-148, 1984.
- 6) Kaplan, E. L., Meier, P.: Nonparametric estimation for incomplete observations. J. Am. Stat. Assoc., **53**: 457-481, 1958.
- 7) Mantel, N.: Chi-square tests with one degree of freedom: Extensions of the Mantel-Haenszel procedure. J. Stat. Assoc., **58**: 690-700, 1963.



## 検索用テキスト OCR(光学的文字認識)ソフト使用

論文の一部ですが、認識率の関係で誤字が含まれる場合があります



〔はじめに〕

治療効果を評価する際には、通常生存率曲線が計算され、さらに2群間での生存率曲線の優劣を統計学的に検討するために検定が行なわれる。これらの具体的方法については、日本癌治療学会においても規約が作られ<sup>1)</sup>、医学研究者の間でも一定の理解に達したと言えよう。しかしながら、2群の治療法を比較する際に常に問題となるのは、各群での予後因子の偏りであって、この点を明確にしなければ治療効果を正確に判定することは不可能となる。いずれの研究においても予後因子が偏らないよう最善の努力が払われているのであるが、実際には偏りをなくすことはかなり難しいのが現状であろう。多要因を同時に補正して治療効果をみる手法としてCoxのProportional Hazard Model<sup>2)</sup>が注目されているが、わが国では汎用のプログラムが未だ入手し難い。

そこで、一般医学研究者が実際に用いることができるような手法で、予後因子を補正する実用的な一方法を考案したので紹介する。