

# 白血病予後因子解析

白血病再発の予後因子について仮想データを用いて検討したいと思います。エンドポイントを病気の再発とし、諸々の予後因子についても検討したいと思います。つまり何が再発と関係するかを知りたいのです。治療効果のみをみるのであれば、Kaplan Meier 生存曲線でもよいのですが、予後因子はお互いに影響し合っていることもよくあるので、1つ1つ解析せずに同時に解析する必要がでてきます。そのような場合には Kaplan Meier は不適切であり、Cox を用います。Proportional hazard とは曲線が似た形をしていることを意味します。

まずこのデータを生存曲線として扱うことをコンピュータに宣言します。ttf は time to failure すなわち再発するまでの日数を示しています。データはこの章の最後に付してあります。cens は censor を示しており、cens = 0 は 42 日の経過観察で再発した、あるいは追跡不能例です。15 人に関して最後まで追跡調査することができませんでした。

```
. stset ttf cens

      failure event:  cens ~= 0 & cens ~= .
obs. time interval:  (0, ttf]
exit on or before:  failure

-----

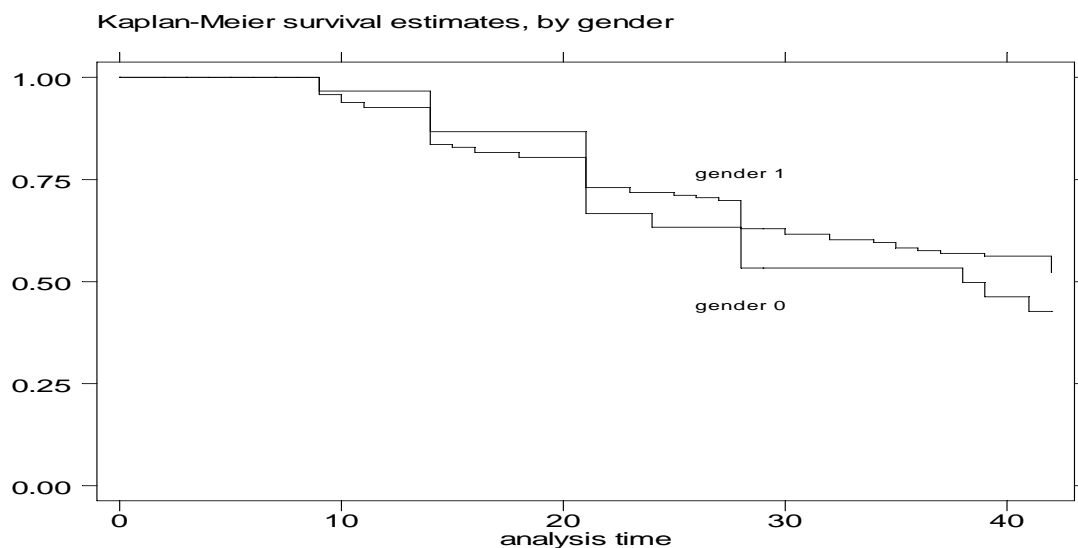
      224 total obs.
       0 exclusions

-----

      224 obs. remaining, representing
      120 failures in single record/single failure data
      6305 total analysis time at risk, at risk from t =          0
              earliest observed entry t =          0
              last observed exit t =          42
```

224 人について観察し、各々 0 42 日観察し、のべ日数は 6305 日です。次に性別(gender)によって再発する時期に差がないか Kaplan-Meier グラフでみてみます。少なくとも、年齢と性別はどのような臨床研究でもまず考慮に入れます。

```
. sts graph, by(gender)
```

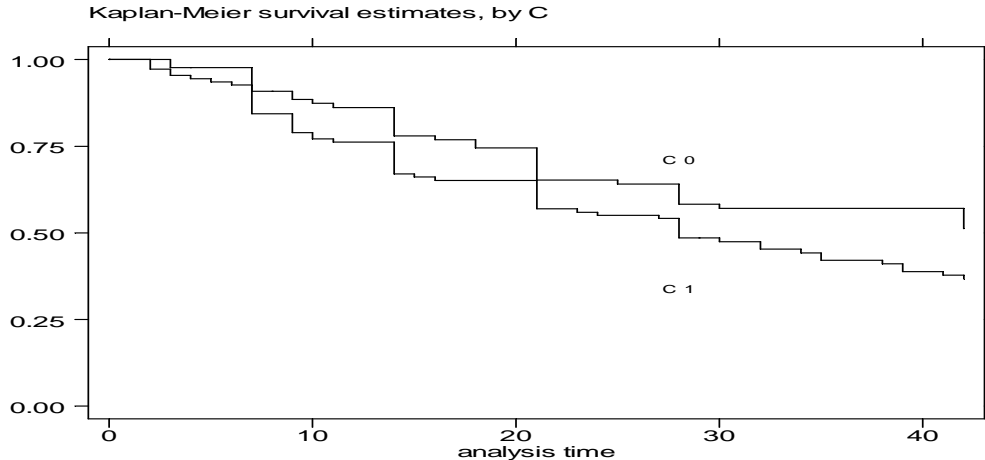


gender 0 は女性で gender 1 は男性です。途中でクロスしています。みたところ大きな差はなさそうです。

同様に治療開始日、検査値 C が正常範囲であれば=0, 異常値であれば=1 としてあります。

```
. sts graph, by(C)
```

```
failure _d: cens  
analysis time _t: ttf
```



見た目、差がはっきりしません。検査値 C は他の検査値と平行に動くかもしれませんが、Kaplan-Meier 生存曲線は1つのパラメータについて検討することができますが、多くの因子を一度に検討することができません。多因子を同時に解析したいので Cox proportional hazard model を用います。

1. 最初に年齢、性別、C、D、E と再発についての関係を個別に検討してください。次に全部同時に検討してください。どれか省略しても問題の無い変数がありますか？

### 年齢と再発の関係

```
. stcox age
```

```
failure _d: cens  
analysis time _t: ttf
```

```
Iteration 0: log likelihood = -608.20335 ← Null model  
Iteration 1: log likelihood = -604.99821  
Iteration 2: log likelihood = -604.98229  
Iteration 3: log likelihood = -604.98229  
Refining estimates:  
Iteration 0: log likelihood = -604.98229
```

```
Cox regression -- Breslow method for ties
```

No. of subjects =	224	Number of obs =	224
No. of failures =	120		
Time at risk =	6305		
		LR chi2(1) =	6.44

```
Log likelihood = -604.98229 ← Fitted model Prob > chi2 = 0.0111
```

_t   _d	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
age	.9785644	.0085478	-2.481	0.013	.9619537 .995462

pが0.05未満であり、95%CIは1.0を含んでいないので、年齢と再発の間には有意な関係があると結論できます。Hazard ratioが0.98ということは、年齢が1歳上がると再発のリスクは0.98になる、年齢が10歳上がるとリスクは0.82になる、つまりリスクは18%低下することになります。

下は $\chi^2$ を示しています。上のFitted modeにおけるLog likelihoodからNull modelにおけるLog Likelihoodを引いて2倍すると、 $\chi^2$ を算出することができます。この場合自由度は1なので、3.8より大きければ有意にモデルが改善されたことを示します。つまり、年齢を考慮に入れることにより、再発をより正確に予測できることを意味します。

```
. di 2*((-604.98) - (-608.20))  
6.44
```

下はpseudo  $\chi^2$ を示しています。

```
. di 2*((-604.98) - (-608.20)) / (0 - (-608.20))  
.01058862
```

下はHazard ratioを示しています。上の表と照らし合わせてください。

```
. di exp(-.0216687)  
.97856438
```

下はHazard ratio lower limitを示しています。

```
. di exp(-.004455)  
.99555491
```

下はHazard ratio upper limitを示しています。

```
. di 2*((-462.04)-(-463.28))  
2.48
```

## 性別と再発の関係

```
. stcox gender

      failure _d:  cens
      analysis time _t:  ttf

Iteration 0:  log likelihood = -608.20335
Iteration 1:  log likelihood = -607.6032
Iteration 2:  log likelihood = -607.59767
Iteration 3:  log likelihood = -607.59767
Refining estimates:
Iteration 0:  log likelihood = -607.59767

Cox regression -- Breslow method for ties

No. of subjects =          224          Number of obs =          224
No. of failures =          120
Time at risk   =          6305
Log likelihood  = -607.59767          LR chi2(1)      =          1.21
                                          Prob > chi2    =          0.2711

-----+-----
      _t |
      _d | Haz. Ratio   Std. Err.      z    P>|z|      [95% Conf. Interval]
-----+-----
gender |   .7691464   .1784707    -1.131  0.258    .4880884   1.212047
-----+-----
```

性別は再発に有意な影響を与えていません。

## 検査値 C、D、E の正常・異常と再発の関係

治療開始時、検査値 C、D、E それぞれが正常か異常かと再発の関係について検討してみましよう。

```
. stcox C

      failure _d:  cens
      analysis time _t:  ttf

Iteration 0:  log likelihood = -538.6235
Iteration 1:  log likelihood = -536.50858
Iteration 2:  log likelihood = -536.50778
Refining estimates:
Iteration 0:  log likelihood = -536.50778

Cox regression -- Breslow method for ties

No. of subjects =          197          Number of obs =          197
No. of failures =          109
Time at risk   =          5495
Log likelihood  = -536.50778          LR chi2(1)      =          4.23
                                          Prob > chi2    =          0.0397

-----+-----
      _t |
      _d | Haz. Ratio   Std. Err.      z    P>|z|      [95% Conf. Interval]
-----+-----
C      |   1.493287   .2946279     2.032  0.042    1.014378   2.198299
-----+-----
```

```
. stcox D

      failure _d:  cens
      analysis time _t:  ttf

Iteration 0:  log likelihood = -608.20335
```

```
Iteration 1: log likelihood = -605.8117
Iteration 2: log likelihood = -605.74487
Iteration 3: log likelihood = -605.74485
Refining estimates:
Iteration 0: log likelihood = -605.74485
```

Cox regression -- Breslow method for ties

```
No. of subjects = 224          Number of obs = 224
No. of failures = 120
Time at risk = 6305
Log likelihood = -605.74485    LR chi2(1) = 4.92
                               Prob > chi2 = 0.0266
```

_t   _d	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
D	1.620264	.3383853	2.311	0.021	1.076009	2.439808

```
. stcox E
```

```
failure _d: cens
analysis time _t: ttf
```

```
Iteration 0: log likelihood = -596.60753
Iteration 1: log likelihood = -594.54622
Iteration 2: log likelihood = -594.50875
Iteration 3: log likelihood = -594.50874
Refining estimates:
Iteration 0: log likelihood = -594.50874
```

Cox regression -- Breslow method for ties

```
No. of subjects = 220          Number of obs = 220
No. of failures = 118
Time at risk = 6212
Log likelihood = -594.50874    LR chi2(1) = 4.20
                               Prob > chi2 = 0.0405
```

_t   _d	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
E	1.545657	.3174746	2.120	0.034	1.033423	2.311788

検査値 C、D、Eそれぞれが異常値であると再発するリスクが上がります。

**年齢、性別、検査値 C, D, E 全ての変数を合わせた場合**

今まで個々に検討してきた変数について、全部一緒に検討してみましょう。これができる点が Cox の利点です。

```
. stcox age gender C D E
```

```
failure _d: cens
analysis time _t: ttf
```

```
Iteration 0: log likelihood = -528.04681
Iteration 1: log likelihood = -517.7733
Iteration 2: log likelihood = -517.59148
Iteration 3: log likelihood = -517.5914
Refining estimates:
```

```
Iteration 0: log likelihood = -517.5914
```

```
Cox regression -- Breslow method for ties
```

```
No. of subjects = 195 Number of obs = 195
No. of failures = 107
Time at risk = 5446
LR chi2(5) = 20.91
Log likelihood = -517.5914 Prob > chi2 = 0.0008
```

_t   _d	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
age	.975415	.0088091	-2.756	0.006	.9583015	.9928342
gender	.9106448	.2310515	-0.369	0.712	.553834	1.497333
C	1.479876	.3042192	1.907	0.057	.9891065	2.214152
D	1.627401	.3650534	2.171	0.030	1.048469	2.526
E	1.583393	.3417421	2.129	0.033	1.037229	2.417145

```
. lrtest, saving(null)
```

個々の因子の有意性に関してはp値や95%CIで解決します。しかし、全体として予後をよく表しているかどうかは、Log likelihoodで検討します。

年齢に関してはほとんど同じHazard ratioに対して95%CIは狭くなり、P値も小さくなっています。データを合わせることによってより信頼性の高い値となったのがわかります。一方性別に関してみると、hazard ratioは大きく変動し、P値は大きく、95%CIも幅広となっており、他の変数と合わせることにより信頼性は落ちています。D、Eは相変わらず再発と相関関係にあります。変数を合わせても大きな変化はありません。一方Cに関してhazard ratioはほぼ同じであるにもかかわらず、変数を同時に検討することによって有意な相関関係を失ってしまいました。つまり、1つ1つで検討した結果と全体で検討した結果は異なることがあります。何故ならお互いの因子が連動するためです。それでは変数を合わせるによって大きく数値の変化した性別を全体の中から除いたらどうなるのでしょうか？

**性別を変数から除いたらどうなる？**

```
. stcox age C D E
```

```
failure _d: cens
analysis time _t: ttf
```

```
Iteration 0: log likelihood = -528.04681
Iteration 1: log likelihood = -517.81737
Iteration 2: log likelihood = -517.65838
Iteration 3: log likelihood = -517.65833
Refining estimates:
Iteration 0: log likelihood = -517.65833
```

```
Cox regression -- Breslow method for ties
```

```
No. of subjects = 195 Number of obs = 195
No. of failures = 107
Time at risk = 5446
LR chi2(4) = 20.78
Log likelihood = -517.65833 Prob > chi2 = 0.0004
```

_t   _d	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
------------	------------	-----------	---	------	----------------------	--

```

age | .9753554 .0088605 -2.747 0.006 .9581428 .9928772
C | 1.505854 .3009727 2.048 0.041 1.017782 2.22798
D | 1.647741 .3653798 2.252 0.024 1.066936 2.544716
E | 1.584473 .3419288 2.133 0.033 1.037996 2.418654
-----
. di ((-517.591) - (-517.658))*2
.134

. di 20.91 - 20.78
.13

. lrtest, saving(alt)

. lrtest, model(null) using(alt)
Cox: likelihood-ratio test                    chi2(1) = 0.17
                                              Prob > chi2 = 0.6776

```

年齢、D、Eに関する数値はあまり変動しませんが、Cに関してはP値が 0.05 未満となり、95%CIも 1.0 を含まず、Cは再発と関係するという結果になりました。0.05 に近いところにあるので微妙です。おそらくはCの異常値にわずかながら性差があるためでしょう。Log likelihood はあまり変化していません。 $\chi^2$ は 0.134 で有意ではありません。また検定している内容は同じなのですが、先の性別を含めたモデルをnullとして保存し[lrtest, saving(null)], 性別を除いたモデルも保存し[lrtest, saving(alt)], そして2つのモデルに差がないかどうか検討します。このlikelihood ratio test でも性別を検討すべき変数から除外しても、モデルの精度には影響がないとでています。

2. 先の問題の最後でいくつかの変数を同時に検討しましたが、これに別の予後因子 A, B をそれぞれ、あるいは同時に加えたらどうなるでしょうか？ log likelihoods, estimates, standard errors の変化も同時に検討してみてください。

### Bを加える

```

. stcox age gender B C D E

      failure _d:  cens
      analysis time _t:  ttf

Iteration 0:  log likelihood = -528.04681
Iteration 1:  log likelihood = -515.75797
Iteration 2:  log likelihood = -515.60895
Iteration 3:  log likelihood = -515.6089
Refining estimates:
Iteration 0:  log likelihood = -515.6089

Cox regression -- Breslow method for ties

No. of subjects =          195                Number of obs =          195
No. of failures =           107                LR chi2(6) =          24.88
Time at risk =          5446                Prob > chi2 =          0.0004

Log likelihood =    -515.6089

```

```

-----
      _t |
      _d | Haz. Ratio  Std. Err.      z    P>|z|    [95% Conf. Interval]
-----+-----
age | .9733815   .0089222   -2.943  0.003   .9560505   .9910266
gender | .8498389   .2171111   -0.637  0.524   .5150832   1.402154
B | 1.731153   .4523839    2.100  0.036   1.037289   2.889156
C | 1.519354   .3118546    2.038  0.042   1.016122   2.27181
D | 1.595395   .3575971    2.084  0.037   1.028199   2.47548
E | 1.627915   .3522842    2.252  0.024   1.065197   2.487904

```

```
-----
. di 24.88 - 20.91
3.97
```

年齢、B、C、D、Eは有意です。Log likelihood も - 517.59 から - 515.61、Chi2 も 20.91 から 24.88 まで若干増加しており、B を加えることによってモデルは有意に改善したといえます。

### Aを加える

```
. stcox age gender A C D E
```

```
      failure _d:  cens
      analysis time _t:  ttf
```

```
Iteration 0:  log likelihood = -528.04681
Iteration 1:  log likelihood = -517.60872
Iteration 2:  log likelihood = -517.43562
Iteration 3:  log likelihood = -517.43555
Refining estimates:
Iteration 0:  log likelihood = -517.43555
```

```
Cox regression -- Breslow method for ties
```

```
No. of subjects =          195          Number of obs =          195
No. of failures =          107
Time at risk   =          5446

LR chi2(6)      =          21.22
Prob > chi2    =          0.0017

Log likelihood = -517.43555
```

```
-----
      _t |
      _d | Haz. Ratio  Std. Err.      z    P>|z|    [95% Conf. Interval]
-----+-----
      age | .9748824   .0088335   -2.807  0.005   .9577218   .9923505
gender |
      A | 1.147307   .2788128    0.565  0.572   .7125655   1.847287
      C | 1.481096   .3046864    1.909  0.056   .9896383   2.216612
      D | 1.623458   .3640473    2.161  0.031   1.046083   2.51951
      E | 1.589285   .3431317    2.146  0.032   1.040937   2.426492
-----
```

Aを加えたこと自体、上の解析では有意でなく、Log likelihood も chi2 も大きな変化を示していません。

### 全て一緒に

```
. stcox age gender A B C D E
```

```
      failure _d:  cens
      analysis time _t:  ttf
```

```
Iteration 0:  log likelihood = -528.04681
Iteration 1:  log likelihood = -515.76644
Iteration 2:  log likelihood = -515.60786
Iteration 3:  log likelihood = -515.60781
Refining estimates:
Iteration 0:  log likelihood = -515.60781
```

```
Cox regression -- Breslow method for ties
```

```
No. of subjects =          195          Number of obs =          195
No. of failures =          107
Time at risk   =          5446

LR chi2(7)      =          24.88
```



Log likelihood = -515.60781

Prob > chi2 = 0.0008

_t _d	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
age	.9734074	.0089423	-2.934	0.003	.9560377	.9910927
gender	.8507174	.2181297	-0.631	0.528	.5146731	1.406174
A	.9879947	.2553591	-0.047	0.963	.5953212	1.639676
B	1.738428	.4801664	2.002	0.045	1.011693	2.987203
C	1.519041	.3118107	2.037	0.042	1.015886	2.271402
D	1.59593	.3578965	2.084	0.037	1.028317	2.476856
E	1.626965	.3526514	2.245	0.025	1.063841	2.488168

B のみを加えたきとあまり代わり映えしません。A は重要な要素であるとはいえません。今までみてきてわかる通り、 $p$  が 0.05 に近い場合には条件によって有意になったり有意にならなったりします。そこで、上の例では性別や A のようなモデルに加えても加えなくてもモデルの正確性が変らないものは除去してもよいと判断できます。このようにどの予後因子を加えるかどうかは、モデル全体の正確性を示す log likelihood を 0 になるべく近付けるかどうかで判断します。つまり、ある因子を加えても log likelihood が有意に改善しなければ加える必要はなく、改善すればその因子の  $p$  値が有意でなくても加えるべきであるといえます。

3. 今度は変数を組み合わせて新しい変数を作成し、再発予測因子にならないか検討してみましょう。連続変数と、カテゴリ変数でどちらが有利ですか？

まずは C と A に関する変数を掛け合わせてみます。病態生理から考えると A は C を誘導するとします。新しい変数を AC と設定します。

```
. gen AC = A*C  
(27 missing values generated)
```

\* はかける(X)のコマンドです。上は A および C 両方が異常(=1) のとき  $AC = 1$  となり、それ以外は全て 0 になります。

AC のどちらかが欠けていると新しい変数は創れませんが、27 のデータ欠損があります。もしも再発例と非再発例の間でデータ欠損に偏りがなければ、この程度は許容範囲かと思います

次に新しい変数 (AC) を創ります。A の正常・異常、C の正常・異常を組み合わせると 4 種類の組み合わせができます。下の表のように  $C \times 2 + A$  にすると 4 つの組み合わせを、1, 2, 3 のどれかで表現することができます。

C	C*2	A	=ACp	意味
0	0	0	0	A も C も正常
0	0	1	1	A は異常だが C は正常
1	2	0	2	A は正常だが C は異常
1	2	1	3	A も C も異常

```
. gen ACp = C * 2 + A  
(27 missing values generated)
```

```
. tabulate ACp, generate (ACp)
```

ACp	Freq.	Percent	Cum.
0	73	37.06	37.06
1	15	7.61	44.67
2	85	43.15	87.82
3	24	12.18	100.00
Total	197	100.00	

それぞれのカテゴリに当てはまる人数は 27 の missing data を除いて上記のような分布になります。やや A は異常で C が正常の人の割合が低いようです。

#### Model 1: 従来ある変数のみの組み合わせ

これは先に行なった解析と同じです。

```
. stcox age gender A B C D E  
  
      failure _d:  cens  
      analysis time _t:  ttf
```

```

Iteration 0: log likelihood = -528.04681
Iteration 1: log likelihood = -515.76644
Iteration 2: log likelihood = -515.60786
Iteration 3: log likelihood = -515.60781
Refining estimates:
Iteration 0: log likelihood = -515.60781

```

Cox regression -- Breslow method for ties

```

No. of subjects =          195          Number of obs =          195
No. of failures =          107
Time at risk   =          5446
Log likelihood  = -515.60781
LR chi2(7)     =          24.88
Prob > chi2    =          0.0008

```

_t   _d	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
age	.9734074	.0089423	-2.934	0.003	.9560377	.9910927
gender	.8507174	.2181297	-0.631	0.528	.5146731	1.406174
A	.9879947	.2553591	-0.047	0.963	.5953212	1.639676
B	1.738428	.4801664	2.002	0.045	1.011693	2.987203
C	1.519041	.3118107	2.037	0.042	1.015886	2.271402
D	1.59593	.3578965	2.084	0.037	1.028317	2.476856
E	1.626965	.3526514	2.245	0.025	1.063841	2.488168

**Model2: 先のデータに AC を加えてみます。**

```
. stcox age gender A B C D E AC
```

```

failure _d: cens
analysis time _t: ttf

```

```

Iteration 0: log likelihood = -528.04681
Iteration 1: log likelihood = -512.7434
Iteration 2: log likelihood = -512.31027
Iteration 3: log likelihood = -512.30936
Iteration 4: log likelihood = -512.30936
Refining estimates:
Iteration 0: log likelihood = -512.30936

```

Cox regression -- Breslow method for ties

```

No. of subjects =          195          Number of obs =          195
No. of failures =          107
Time at risk   =          5446
Log likelihood  = -512.30936
LR chi2(8)     =          31.47
Prob > chi2    =          0.0001

```

_t   _d	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
age	.97148	.0089522	-3.140	0.002	.9540914	.9891854
gender	.8372675	.2123444	-0.700	0.484	.5093144	1.376393
A	.3879747	.1954911	-1.879	0.060	.1445118	1.041606
B	2.024257	.5595879	2.551	0.011	1.177494	3.479949
C	1.178192	.2640789	0.732	0.464	.7593267	1.828116
D	1.674378	.3763596	2.293	0.022	1.077765	2.601255
E	1.618362	.3519225	2.214	0.027	1.056761	2.478418
AC	<b>3.937333</b>	<b>2.237464</b>	<b>2.412</b>	<b>0.016</b>	<b>1.292679</b>	<b>11.99261</b>

Cのみ、あるいはA単独では再発に対して有意な影響をもたらしません、AとCを掛け合わせたものを含めると有意であり、Log likelihood は 515 から 512 になり chi2(8)

は 24 から 31 まで上昇しモデルの改善をみています。有意な影響を持ちます。つまり、検査値 A と C の両方が異常であったときに再発に影響すると言えます。

### Model 3: C と A の組み合わせを詳細に検討する

C があるか A に治療を受けている場合、あるいはその両方の場合を加えています。その際 C にウエイトを置いてみます。

```
. stcox age gender B D E ACp2 ACp3 ACp4
      failure _d:  cens
      analysis time _t:  ttf

Iteration 0:  log likelihood = -528.04681
Iteration 1:  log likelihood = -512.7434
Iteration 2:  log likelihood = -512.31027
Iteration 3:  log likelihood = -512.30936
Iteration 4:  log likelihood = -512.30936
Refining estimates:
Iteration 0:  log likelihood = -512.30936

Cox regression -- Breslow method for ties

No. of subjects =          195          Number of obs   =          195
No. of failures =          107
Time at risk    =          5446
Log likelihood  = -512.30936          LR chi2(8)       =          31.47
                                          Prob > chi2    =          0.0001
```

_____	_____	_____	_____	_____	_____	_____
_t						
_d	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
age	.97148	.0089522	-3.140	0.002	.9540914	.9891854
gender	.8372675	.2123444	-0.700	0.484	.5093144	1.376393
B	2.024257	.5595879	2.551	0.011	1.177494	3.479949
D	1.674378	.3763596	2.293	0.022	1.077765	2.601255
E	1.618362	.3519225	2.214	0.027	1.056761	2.478418
ACp2	.3879747	.1954911	-1.879	0.060	.1445118	1.041606
ACp3	1.178192	.2640789	0.732	0.464	.7593267	1.828116
ACp4	1.79979	.5445898	1.942	0.052	.9946282	3.256738

ACp2,3,4 の中で有意なものはありませんでした。しかし Log likelihood および chi2 は AC (A x C)と同程度改善しています。

### Continuous variables

今まで D が異常値(=1)か正常(=0)かの indicator variable として扱ってきましたが、観察された実数、すなわち連続変数として扱ってみましょう。CC は検査 D の実際の値であるとします。

### Model 4: 0/1 でなく連続変数で検討する

```
. stcox age gender CC E B C A AC
      failure _d:  cens
      analysis time _t:  ttf

Iteration 0:  log likelihood = -528.04681
Iteration 1:  log likelihood = -516.08042
Iteration 2:  log likelihood = -509.80432
```

```
Iteration 3: log likelihood = -509.74945
Iteration 4: log likelihood = -509.74943
Refining estimates:
Iteration 0: log likelihood = -509.74943
```

Cox regression -- Breslow method for ties

```
No. of subjects = 195          Number of obs = 195
No. of failures = 107
Time at risk = 5446

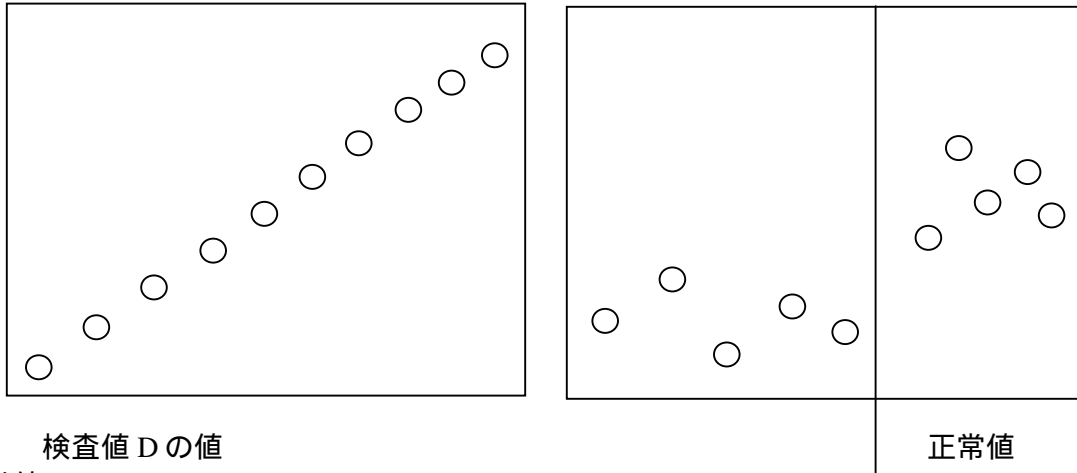
LR chi2(8) = 36.59
Prob > chi2 = 0.0000

Log likelihood = -509.74943
```

Variable	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
age	.972334	.0089407	-3.051	0.002	.9549676 .9900163
gender	.8768137	.2228907	-0.517	0.605	.5327548 1.44307
CC	1.000049	.0000119	4.103	0.000	1.000025 1.000072
E	1.534339	.3419877	1.921	0.055	.9912834 2.374899
B	2.118163	.585022	2.717	0.007	1.232717 3.639614
C	1.149935	.2586459	0.621	0.535	.7399787 1.787012
A	.4022491	.2015523	-1.818	0.069	.1506569 1.073992
AC	3.939228	2.233186	2.418	0.016	1.296752 11.96645

D を CC 連続変数として扱ったところ強い相関関係を見出せました。しかも、全体として Log likelihood および chi2 が少しずつ改善しています。(512.3 - 509.7) x 2 > 3.8 ですから、統計学的にも有意です。全か無の変数を連続変数に変えることによってモデルの質をあげることができました。

### 再発率



もしも検査値 D の値と再発率が左のような関係であれば、検査値 D を indicator variable としてとらえるよりは continuous variable としてとらえる方がモデルの改善を望めます。一方右のような関係であれば、indicator variable としてとらえた方が continuous variable としてとらえるよりもモデルとして改善します。このように最も再発を予測しえるモデルを作成することが、Cox の目的なのです。この例では、連続変数としてとらえた方が正常・異常でとらえるより再発を予測しやすいという結果となりました。変数はしばしば 0 として判定もできれば、数値として読むこともできます。変数をどのように扱ったときに log likelihood の値を 0 に近くすることができるかいろいろ検討しなくて

はなりません。また、相乗効果についても組み合わせは沢山あります。ですから、同じデータを扱っても人によって結果が微妙に異なるかもしれません。しかし、その施行錯誤の過程は論文からはうかがい知ることができません。

4. 治療開始後 7 ヶ月以下に再発するもの、7 ヶ月以降に再発するもの間に何か違いはあるでしょうか？

```
. tabulate cens
```

cens	Freq.	Percent	Cum.
0	104	46.43	46.43
1	120	53.57	100.00
Total	224	100.00	

再発したものは 224 例中 120 例、センサーとなったものは 104 例です。追跡期間が 8 ヶ月未満 (7 ヶ月以下) に限ってみるとどうでしょうか？

```
. tabulate cens if ttf<8
```

cens	Freq.	Percent	Cum.
0	2	6.90	6.90
1	27	93.10	100.00
Total	29	100.00	

センサーとなった例は 29 例中 2 例しかありませんでした。27 例は 7 ヶ月の間に再発した患者さんです。この患者さん達を cens1 として区別します。

```
. generate cens1 =cens
```

cens1 でないものは cens0 となりますが、これは 7 ヶ月以降に再発になるかセンサーとなった人々です。

```
. replace cens1 = 0 if ttf>7  
(93 real changes made)
```

```
. tabulate cens if ttf >7
```

cens	Freq.	Percent	Cum.
0	102	52.31	52.31
1	93	47.69	100.00
Total	195	100.00	

224 から 29 人引きますから 195 人です。今度はセンサーとなった人の方が若干多くなっています。

この人達を cens2 としましょう。

```
. generate cens2 = cens
```

cens2 でないものは 8 ヶ月未満で観察を打ち切っているものです。

```
. replace cens2 = 0 if ttf <8  
(27 real changes made)
```

**B は最初の 7 ヶ月間に再発する症例と関連がありますか？**

```
. stset ttf cens1

      failure event:  cens1 ~= 0 & cens1 ~= .
obs. time interval:  (0, ttf]
exit on or before:  failure

-----
      224 total obs.
       0 exclusions

-----
      224 obs. remaining, representing
      27 failures in single record/single failure data
6305 total analysis time at risk, at risk from t =           0
      earliest observed entry t =           0
      last observed exit t =           42

. stcox age gender D E B A AC

      failure _d:  cens1
analysis time _t:  ttf

Iteration 0:  log likelihood = -125.54063
Iteration 1:  log likelihood = -118.97083
Iteration 2:  log likelihood = -118.59129
Iteration 3:  log likelihood = -118.59092
Iteration 4:  log likelihood = -118.59092
Refining estimates:
Iteration 0:  log likelihood = -118.59092

Cox regression -- Breslow method for ties

No. of subjects =           195                Number of obs   =           195
No. of failures =            24
Time at risk    =           5446

LR chi2(7)      =           13.90
Prob > chi2     =           0.0530
Log likelihood  = -118.59092

-----
      _t |
      _d | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
      age | .9495189   .0199973    -2.460  0.014   .9111228   .9895331
gender | .9842169   .5197798    -0.030  0.976   .3495895   2.770915
      D  | 1.821698   .8351541     1.308  0.191   .7417282   4.474126
      E  | 2.439475   1.026158     2.120  0.034   1.069648   5.563545
      B  | 1.333872   .8982392     0.428  0.669   .3563765   4.992516
      A  | .4884225   .5301526    -0.660  0.509   .0581928   4.099421
      AC | 1.563443   1.961872     0.356  0.722   .1336472  18.28961
-----
```

7 ヶ月以内に再発となる場合、年齢と E が重要な予測因子となります。一方 B は最初の 7 ヶ月間に再発とは関係ありませんでした。

**B は 7 ヶ月間よりあとに再発になる症例と関連がありますか？**

```
. stset ttf cens2

      failure event:  cens2 ~= 0 & cens2 ~= .
obs. time interval:  (0, ttf]
exit on or before:  failure

-----
      224 total obs.
       0 exclusions

-----
```



```

224 obs. remaining, representing
93 failures in single record/single failure data
6305 total analysis time at risk, at risk from t = 0
      earliest observed entry t = 0
      last observed exit t = 42

. stcox age gender D E B A AC

      failure _d:  cens2
      analysis time _t:  ttf

Iteration 0:  log likelihood = -402.50618
Iteration 1:  log likelihood = -391.55474
Iteration 2:  log likelihood = -390.83308
Iteration 3:  log likelihood = -390.83081
Iteration 4:  log likelihood = -390.83081
Refining estimates:
Iteration 0:  log likelihood = -390.83081

Cox regression -- Breslow method for ties

No. of subjects =          195          Number of obs =          195
No. of failures =           83
Time at risk =          5446

LR chi2(7) =          23.35
Log likelihood = -390.83081      Prob > chi2 =          0.0015

-----
      _t |
      _d | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
      age | .9758802   .0099795    -2.388  0.017    .9565155    .995637
      gender | .7854184   .2222799    -0.853  0.393    .4510285    1.367723
      D | 1.568872   .4112266     1.718  0.086    .9385856    2.622413
      E | 1.406593   .3643699     1.317  0.188    .8465829    2.337047
      B | 2.265086   .6933456     2.671  0.008    1.243171    4.127039
      A | .3348033   .1831784    -2.000  0.046    .1145719    .9783658
      AC | 5.85768    3.398843     3.047  0.002    1.878568    18.2652
-----

```

7ヶ月より後では、Bが関係していました。他にも年齢とACが関係していました。しかしEは関係ないようです。つまり再発になる時期によって予測因子が異なることとなります。これは effect modification にあたります。

#### まとめ

年齢、B、D、(C)の異常値が再発を決定付ける重要な予後因子でした。AとCの異常値を同時に持つ人も再発に関係していました。更に、7ヶ月以内に再発の頻度は、Eの異常値をもつもので高い傾向にありました。一方、7ヶ月以降では、Bの異常値があると再発となりやすい傾向にありました。