

データセットをData Editorで確認

Figure 1: Screenshot of Stata Data Editor showing a dataset with variables survival, sex, last_date, tx_date, and Variables. The Variables list is circled, and the data rows are also circled.

survival	sex	last_date	tx_date	Variables
22 生存	男	2011/8/29	2006/4/	Variable Label
23 死亡	男	2007/2/15	2005/11	survival 生存状況
24 生存	男	2006/5/12	2001/4/	sex 性
25	男	2007/3/7		last_date 生乳量検閲日
26 生存	女	2010/11/24	2006/6/	tx_date 検検日
27 生存	女	2011/7/29	2009/6/	age 検検年齢
28 死亡	女	2011/1/1	2010/1/	donor 主治医名
29 死亡	男	2004/5/27	2003/9/	Properties
30 死亡	女	2011/7/13	2009/8/	Name sex
31 死亡	男	2009/7/5	2006/4/	Label 性
32 死亡	男	2004/7/4	2004/4/	Type str2

データセットをData Editorで確認

変数に対するLabelと各項目に対する Value Label

Figure 2: Screenshot of Stata Data Editor showing a dataset with variables diagnosis, dr, and Variables. The Variables list and Value Labels are circled.

diagnosis	dr	Variables
22 急性リンパ性白血病(ALL)	ALL	Variable Label
23 慢性骨髄性白血病(CML)	CML	sex 性
24 急性リンパ性白血病(ALL)	ALL	last_date 生乳量検閲日
25 慢性骨髄性白血病(CML)	CML	tx_date 検検日
26 急性リンパ性白血病(ALL)	ALL	age 検検年齢
27 慢性骨髄性白血病(CML)	CML	donor 主治医名
28 急性リンパ性白血病(ALL)	ALL	Properties
29 慢性骨髄性白血病(CML)	CML	Name sex
30 急性リンパ性白血病(ALL)	ALL	Label 性
31 慢性骨髄性白血病(CML)	CML	Type str2
32 急性リンパ性白血病(ALL)	ALL	Format %w95

ログファイルの作成 ①

配布資料: S1¥Minidata_2014: 46行目

log using "C:¥data¥mini20140913.smcl"

Command
log using "C:¥data¥mini20140913.smcl"

```
do "C:¥Users¥FYACHI¥1¥AppData¥Local¥Temp¥STD0g000000.tmp"
```

```
log using "C:¥data¥mini20140913.smcl", append
```

```
name: <unnamed>
log: C:¥data¥mini20140913.smcl
log type: smcl
opened on: 12 Sep 2014, 11:03:47
```

Logファイルの記録が開始される

appendで、以前に
作ったログへの追
記も可能

Command

ログファイルの作成 ②

資料47行目

*** あるいは、新たなログファイルを
"Log Begin"のタブから作成

Figure 3: Screenshot of Stata Results window showing the Log Begin tab. The 'Log Begin' tab is circled.

STATA 12.1 Copyright 198 StataCorp
4905 Lakeway College Stati
800-STATA-PC 979-696-4600
979-696-4601

Single-user Stata perpetual license:

Figure 4: Screenshot of Stata Review window showing the Log Begin tab. The 'Log Begin' tab and the file name 'mini20140913' are circled.

保存

```

do "C:\Users\YACHIYO\AppData\Local\Temp\STD0g000000.tmp"
log using "C:\data\mini20140913.smcl", append
name: <unnamed>
log: C:\data\mini20140913.smcl
log type: smcl
opened on: 12 Sep 2014, 11:03:47

```

Command

S1¥Minidata_2014 49行目~

```

/*症例数確認*/
count

```

Command
count

Command windowに入力

S1¥Minidata_2014 49行目~

```

/*症例数確認*/
count

```

```

. use "C:\Users\Yachiyo Kuwatsuka\AppData\Local\Temp

```

```

. count
57

```

57例含まれる

Command

S1¥Minidata_2013 52行目~

```

/****不要データの削除****/
*****
* STEP
* 1) AML, ALL, CMLの中からCMLを削除
* 2) survival dataが欠損の症例を削除
* last_date (最終観察日)が欠損の症例を削除
*****

```

新たな変数作成 NOTES
S1¥Minidata_2014: 2~18行目

drop, keep
変数・症例の削除

generate, replace
新たな変数の作成・データの置換

新たな変数作成の方法
S1¥Minidata_2014: 11~22行目

```

1 ///////////////****新たな変数作成 NOTES****/////////////////
2 /**generate, replaceについて**/
3 *generate new_disease="" 文字のデータ
4   *新たな変数new_diseaseが作成されます
5   *replace new_disease=dx
6   *replace new_disease="AML" if diagnosis==1
7   *replace new_disease="ALL" if diagnosis==2
8
9   *generate new_disease2=. 数字のデータ
10  *replace new_disease2=1 if new_disease=="AML"
11  *replace new_disease2=2 if new_disease=="ALL"

```

順番に疾患データの中身を確認
60. tab dx, missing (欠損値も表に含める)

tab dx, missing

主疾患名	Freq.	Percent	Cum.
急性リンパ性白血病(ALL)	30	52.63	52.63
急性骨髄性白血病(AML)	24	42.11	94.74
慢性骨髄性白血病(CML)	3	5.26	100.00
Total	57	100.00	

順番に疾患データの中身を確認
61. tab diagnosis, missing (欠損値も表に含める)

tab diagnosis, missing

疾患名1	Freq.	Percent	Cum.
AML	24	42.11	42.11
ALL	30	52.63	94.74
CML	3	5.26	100.00
Total	57	100.00	

Labelが付いた状態

順番に疾患データの中身を確認
62. tab diagnosis, missing nolabel (ラベルを付けない)

tab diagnosis, missing

疾患名1	Freq.	Percent	Cum.
AML	24	42.11	42.11
ALL	30	52.63	94.74
CML	3	5.26	100.00
Total	57	100.00	

tab diagnosis, missing nolabel

疾患名1	Freq.	Percent	Cum.
1	24	42.11	42.11
2	30	52.63	94.74
4	3	5.26	100.00
Total	57	100.00	

2x2 tableを作る
64. tab dx diagnosis, missing

2x2 tableを作る
64. tab dx diagnosis, missing

tab dx diagnosis, missing

主疾患名	疾患名1			Total
	AML	ALL	CML	
急性リンパ性白血病(ALL)	0	30	0	30
急性骨髄性白血病(AML)	24	0	0	24
慢性骨髄性白血病(CML)	0	0	3	3
Total	24	30	3	57

S1¥Minidata_2014 65行目~

/*AMLとALLのみをkeep*/

* if以下の"等しい"は、"=="

* 文字列の値は、""で囲む。数値は囲まない

* & (and かつ),

| (or または)

== (equal 等しい)

!= (not equal 等しくない)

69. keep if diagnosis==1 | diagnosis==2

70. drop if diagnosis==4

どちらかを使用して、CML症例をデータから削除

71. count

66. keep if diagnosis==1 | diagnosis==2
68. count

diagnosis

diagnosis	disease_1st	seronisfab72	
急性血球(ALL)	ALL	(F)	0
慢性血球(AML)	AML	(M)	2
慢性血球(ALL)	ALL	(M)	1
慢性血球(ALL)	ALL	(F)	0
慢性血球(AML)	AML	(F)	0
慢性血球(AML)	AML	(M)	0
慢性血球(AML)	AML	(F)	1

```

keep if diagnosis==1 | diagnosis==2
(3 observations deleted)

count
54
  
```

S1¥Minidata_2014 73行目~

survival (生存状況)のデータと、last_date(最終確認日)が欠損の症例を削除

まずは、データの確認
74. tab survival, missing
75. tab last_date, missing

次に、76か77 どちらかで欠損症例を削除
76. drop if survival=="" | last_date==""
77. keep if survival!="" & last_date!=""
78. count

74. tab survival, missing

```

(3 observations deleted)

count
54

tab survival, missing
  
```

生存状況	Freq.	Percent	Cum.
死亡	2	3.70	3.70
生存	25	46.30	50.00
Total	54	100.00	

75. tab last_date, missing

```

tab last_date, missing
  
```

生死最終確認日	Freq.	Percent	Cum.
2001/2/13	4	7.41	7.41
2001/8/18	1	1.85	9.26
2001/9/24	1	1.85	12.96
2002/11/3	1	1.85	14.81
2002/2/21	1	1.85	16.67
2002/8/2	1	1.85	18.52
2003/3/9	1	1.85	20.37
2004/10/17	1	1.85	22.22
2004/11/1	1	1.85	24.07
2004/3/18	1	1.85	25.93

スペースキーなどで下を見られます

76. drop if survival=="" | last_date==""
78. count

```

drop if survival=="" | last_date==""
(4 observations deleted)

count
50
  
```

survival	sex	last_date	lx_date	age	donor
46	死亡	男	2004/3/18	2003/12/9	14 同種_allogeneic 急性リンパ
46	生存	女	2006/12/3	2006/3/15	26 同種_allogeneic 急性骨髄
47	生存	男	2011/7/7	2010/11/2	9 同種_allogeneic 急性リンパ
48	生存	女	2011/6/25	2009/7/14	7 同種_allogeneic 急性骨髄
48	死亡	女	2009/3/9	2009/7/24	17 同種_allogeneic 急性骨髄
50	生存	女	2010/6/23	2006/9/16	17 同種_allogeneic 急性リンパ

S1¥Minidata_2014 80行目~

/* ** dataの加工 ** */

* 性別dataの作成

* 「男」「女」のdataでは後に多変量解析ができないため、数値(0, 1, 2,)の変数を作成する必要があります

STEP

- 1) 「男」「女」というデータから、女性が0 男性が1の新たな変数を作り
- 2) それに"Female" "Male"のラベルを付けます
- 3) 作成変数の確認

性別データの確認
92. tab sex, missing

```

JU
-----
tab sex, missing
-----

```

性	Freq.	Percent	Cum.
女	22	44.00	44.00
男	28	56.00	100.00
Total	50	100.00	

新たな変数 "pt_sex" を作成、
sex=="女" の症例は "0"
93. gen pt_sex = 0 if sex=="女"

```

Edit View Data Tools
-----
gen pt_sex = 0 if sex=="女"
-----

```

diagnosis	disease_st %	seronostab %	pt_sex
1	All	(R)	0
2	AME	NP	0
3	All	HR	0
4	All	(R)	0
5	AME	FC	0
6	AME	HR	0
-			

gen pt_sex = 0 if sex=="女"
(28 missing values generated)

変数 "pt_sex" のデータを、
sex=="男" の症例は "1" に変更する

95. replace pt_sex = 1 if sex=="男"

```

di
-----
replace pt_sex = 1 if sex=="男"
-----

```

diagnosis	disease_st %	seronostab %	pt_sex
1	All	(R)	1
2	AME	NP	1
3	All	HR	1
4	All	(R)	1
5	AME	FC	1
6	AME	HR	1
7	AME	(R)	1

replace pt_sex = 1 if sex=="男"
(28 real changes made)

「sex_label」という名前のラベルの定義をする
102. label define sex_label 0 "Female" 1 "Male"

```

di
-----
label define sex_label 0 "Female" 1 "Male"
-----

```

Stataの中で名札が作られた
だけの状態

変数 "pt_sex" にラベル "sex_label" を
関連付ける (名札をつける作業)

103. label values pt_sex sex_label

```

di
-----
label define sex_label 0 "Female" 1 "Male"
label values pt_sex sex_label
-----

```

diagnosis	disease_st %	seronostab %	pt_sex
1	All	(R)	Male
2	AME	NP	Female
3	All	HR	Female
4	All	(R)	Female
5	AME	FC	Male
6	AME	HR	Male
7	AME	(R)	Male
8	AME	(R)	Male

2x2 table を作って、間違えていないか確認
110. tab sex pt_sex, missing

	男	0	28	28
Total	22	28	50	

```

tab sex pt_sex, missing
-----

```

性	pt_sex		Total
	Female	Male	
女	22	0	22
男	0	28	28
Total	22	28	50

S1¥Minidata_2014 112行目~ ミスマッチ変数の作成

Characteristic	HLA matched N (%)	HLA mismatched N (%)	Total
Total	29 (58.00)	21 (42.00)	50
Year of transplant			
2006-2007	17 (34.00)	17 (34.00)	34
2008-2010	12 (24.00)	4 (8.00)	16

S1¥Minidata_2014 112行目~

tab seromis6abdr2, missing 120行目

```
seromis6abdr2 |
-----+-----
```

r2	Freq.	Percent	Cum.
0	29	58.00	58.00
1	17	34.00	92.00
2	4	8.00	100.00
Total	50	100.00	

HLA ミスマッチあり1 なし0 の変数を作成

- 1) seromis6abdr2 でミスマッチ数0個を "mismatch"=0
seromis6abdr2 でミスマッチ数1個以上を "mismatch"=1
- 2) 0 "match" 1 "mismatch" のラベルを付けます
- 3) 作成した変数の確認

"mismatch" の作成: ミスマッチなし0, ミスマッチあり1

121. generate mismatch=0 if seromis6abdr2==0
122. replace mismatch=1 if seromis6abdr2==1 | seromis6abdr2==2

```
generate mismatch=0 if seromis6abdr2==0
(21 missing values generated)

replace mismatch=1 if seromis6abdr2==1 | seromis6abdr2==2
(21 real changes made)
```

変数 "mismatch" にラベル「mismatch_label」をつける
0, match; 1, mismatch

- 127 label define mismatch_label 0 "match" 1 "mismatch"
- 128 label values mismatch mismatch_label

```
label define mismatch_label 0 "match" 1 "mismatch"
label values mismatch mismatch_label
```

2x2 tableで間違えないか確認

129. tab mismatch seromis6abdr2, missing nolabel

tab mismatch seromis6abdr2, missing nolabel

mismatch	seromis6abdr2			Total
	0	1	2	
0	29	0	0	29
1	0	17	4	21
Total	29	17	4	50

S1¥Minidata_2014 148行目~

- ```

```
- | last_date | tx_date    |
|-----------|------------|
| 2001/6/18 | 2001/1/28  |
| 2011/7/11 | 2002/8/9   |
| 2007/7/21 | 2007/3/26  |
| 2011/8/2  | 2009/8/12  |
| 2011/8/1  | 2010/10/13 |
| 2007/5/7  | 2006/5/29  |
- ```
*****
```
- * 日付と日数の計算
 - * STEP
 - * 1) stata日付データの作成
 - * 2) 移植日から移植年の作成
 - * 3) 移植年カテゴリーの作成

- * Excel 1900年1月1日がシリアル値「1」
2013年10月20日はシリアル値「41567」
- * Stata 1960年1月1日がシリアル値「1」
2013年10月20日はシリアル値「19651」
- * 1960/1/1を「1」とした5桁の変数「tdata」を作成
- * 10/20/2013と入力されたデータの場合は"MDY"と指定
- * Excelデータを取り込んだ場合は、この作業は不要です。

移植日付“tx_date”から、stata日付データ“tdate”を作成

161 generate tdate= date(tx_date, "YMD")

st	seromisbab^2	pt_sex	mismatch	tdate
CR1	0	Male	match	15001
NR	2	Female	mismatch	15561
NR	1	Female	mismatch	17245
CR1	0	Female	match	18128
CR2	0	Male	match	18548
NR	0	Male	match	16950
CR1	0	Male	match	16212
CR1	1	Male	mismatch	16357

generate tdate= date(tx_date, "YMD")

今作った“tdate”から、移植年変数“tyear”を作成

162 gen tyear=year(tdate)

seromisbab^2	pt_sex	mismatch	tdate	tyear
0	Male	match	15001	2001
2	Female	mismatch	15561	2002
1	Female	mismatch	17245	2007
0	Female	match	18128	2009
0	Male	match	18548	2010
0	Male	match	16950	2006
0	Male	match	16212	2004
1	Male	mismatch	16357	2004

gen tyear=year(tdate)

移植年 categorical dataの作成【宿題】

- ・ 1: 2000-2005年、2: 2006-2010年 変数“tyear_cat”
- ・ ラベルをつける 1 "2000-2005" 2 "2006-2010"

- ・ 変数“tyear_cat”にも変数label(表札)をつける

164 tab tyear

167 gen tyear_cat = 1 if tyear>=2000 & tyear<=2005
あるいは

168 gen tyear_cat = 1 if tyear==2000 | tyear==2001 |
tyear==2002 | tyear==2003 | tyear==2004 | tyear==2005

169 replace tyear_cat = 2 if tyear>=2006 & tyear<=2010

167 gen tyear_cat = 1 if tyear>=2000 & tyear<=2005

169 replace tyear_cat = 2 if tyear>=2006 & tyear<=2010

mismatch	tdate	tyear	tyear_cat
match	15001	2001	1
mismatch	15561	2002	1
mismatch	17245	2007	2
match	18128	2009	2
match	18548	2010	2
match	16950	2006	2
match	16212	2004	1
mismatch	16357	2004	1

gen tyear_cat = 1 if tyear>=2000 & tyear<=2005
(25 missing values generated)

replace tyear_cat = 2 if tyear>=2006 & tyear<=2010
(25 real changes made)

171 label var tyear_cat "移植年category"

172 label define tyear_cat_label 1 "2000-2005" 2 "2006-2010"

173 label values tyear_cat tyear_cat_label

175 tab tyear_cat tyear, missing

tab tyear_cat tyear, missing

移植年category	2000	2001	2002	2003	Total
2000-2005	3	5	5	5	25
2006-2010	0	0	0	0	25
Total	3	5	5	5	50

移植年category	2004	2005	2006	2007	Total
2000-2005	4	3	0	0	25
2006-2010	0	0	6	5	25
Total	4	3	6	5	50

移植年category	2008	2009	2010	Total
2000-2005	0	0	0	25
2006-2010	3	6	5	25

Table 1. の作成

ミニデータセット ワークシート

Characteristic ^b	HLA-matched ^c		HLA-mismatched ^c		P-value ^d
	N ^e	(%) ^e	N ^e	(%) ^e	
Total ^f					
Age at transplant, years, median (range) ^g					
Gender ^h					
Male ^h	(36) ^h	(36) ^h	(36) ^h	(36) ^h	
Female ^h	(36) ^h	(36) ^h	(36) ^h	(36) ^h	
Disease ⁱ					
AML ⁱ	(36) ⁱ	(36) ⁱ	(36) ⁱ	(36) ⁱ	
ALL ⁱ	(36) ⁱ	(36) ⁱ	(36) ⁱ	(36) ⁱ	
Year of transplant ^j					
2000-2005 ^j	(36) ^j	(36) ^j	(36) ^j	(36) ^j	
2006-2010 ^j	(36) ^j	(36) ^j	(36) ^j	(36) ^j	

179 tab mismatch, missing

. tab mismatch, missing

mismatch	Freq.	Percent	Cum.
match	29	58.00	58.00
mismatch	21	42.00	100.00
Total	50	100.00	

55

180 sum age if mismatch==0, detail

. sum age if mismatch==0, detail

Percentiles		Smallest		
1%	1	1		
5%	1			
10%	2	2	Obs	29
25%	4	3	Sum of Wgt.	29
50%	10		Mean	11
		Largest	Std. Dev.	7.559289
75%	16	21	Variance	57.14286
90%	24	24	Skewness	.4397324
95%	25	25	Kurtosis	2.105173
99%	26	26		

56

181 sum age if mismatch==1, detail

. sum age if mismatch==1, detail

Percentiles		Smallest		
1%	1	1		
5%	3	3		
10%	3	3	Obs	21
25%	6	3	Sum of Wgt.	21
50%	17		Mean	14.14286
		Largest	Std. Dev.	9.172163
75%	21	27	Variance	84.12857
90%	27	27	Skewness	.0578875
95%	27	27	Kurtosis	1.573266
99%	27	27		

57

Wilcoxon Rank Sum Test

187 ranksum age, by (mismatch)

. ranksum age, by (mismatch)

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

mismatch	obs	rank sum	expected
match	29	675	739.5
mismatch	21	600	535.5
combined	50	1275	1275

unadjusted variance 2588.25
adjustment for ties -11.81

adjusted variance 2576.44

Ho: age(mismatch==match) = age(mismatch==mismatch)

$z = -1.271$
Prob > |z| = 0.2038

58

189 tab pt_sex mismatch, chi2 colum

. tab pt_sex mismatch, chi2 colum

Key
frequency
column percentage

pt_sex	mismatch		Total
	match	mismatch	
Female	13 44.83	9 42.86	22 44.00
Male	16 55.17	12 57.14	28 56.00
Total	29 100.00	21 100.00	50 100.00

Pearson chi2(1) = 0.0192 Pr = 0.890

59

190 tab diagnosis mismatch, chi2 colum

192 tab tyear_cat mismatch, chi2 colum

. tab diagnosis mismatch, chi2 colum

. tab tyear_cat mismatch, chi2 colum

Key
frequency
column percentage

疾患名	mismatch		Total
	match	mismatch	
AML	12 41.38	10 47.62	22 44.00
ALL	17 58.62	11 52.38	28 56.00
Total	29 100.00	21 100.00	50 100.00

Pearson chi2(1) = 0.1925 Pr = 0.661

Key
frequency
column percentage

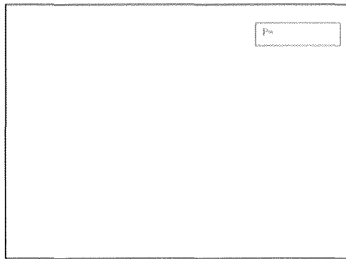
移植年	mismatch		Total
	match	mismatch	
2000-2005	14 48.28	11 52.38	25 50.00
2006-2010	15 51.72	10 47.62	25 50.00
Total	29 100.00	21 100.00	50 100.00

Pearson chi2(1) = 0.0821 Pr = 0.774

60

Figure 1. の作成

Figure 1. Kaplan-Meier Survival Curve



41

S1¥Minidata_2014 187行目~

**** Survival dataの作成 **** /

last_date	tdate	Survat
2001/3/18	2001/1/20	0
2011/7/11	2002/8/9	0
2007/7/21	2007/3/20	0
2011/3/22	2009/3/19	0

日付データの取り扱い

- 1) 生死最終確認日 "last_date" から stata日付変数 "ldate" の作成
- 2) "ldate" と "tdate" から 移植日から最終観察日までの日数 "lday" を作成

195 generate ldate=date(last_date,"YMD")
Stata日付データ、5ケタの数字

196 generate lday=ldate-tdate

197 sum lday, detail

生存日が移植日より前のデータなどないか？

62

```
. generate ldate=date(last_date, "YMD")
. generate lday=ldate-tdate
. sum lday, detail
```

Percentiles		Smallest		
1%	10	10	Obs	50
5%	33	19	Sum of Wgt.	50
10%	89.5	33	Mean	961.46
25%	247	47	Std. Dev.	1003.367
50%	554.5		Variance	1006746
75%	1579	Largest	Skewness	1.321552
90%	2621	3217	Kurtosis	3.867608
95%	3258	3258		
99%	3847	3847		

生死データの作成

200 tab survival, missing nolabel

. tab survival, missing nolabel

生存状況	Freq.	Percent	Cum.
死亡	24	48.00	48.00
生存	26	52.00	100.00
Total	50	100.00	

survival	
1	死亡
2	生存
3	死亡

変数作成 "event_os" 生存: 0 死亡: 1

201 gen event_os = 1 if survival == "死亡"

202 replace event_os = 0 if survival == "生存"

203 *tab survival event_os, missing

205 generate lyear=lday/365.25

event_os, missing	
event_os	Freq.
0	26
1	24
Total	50

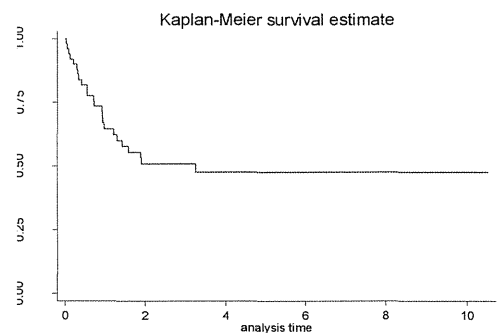
最終観察日, failure の宣言

207 stset lyear, failure (event_os==1)

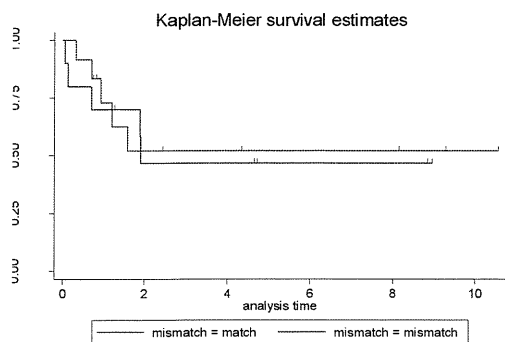
```
. stset lyear, failure (event_os==1)
failure event: event_os = 1
obs. time interval: (0, lyear]
exit on or before: failure
```

50	total observations
0	exclusions
50	observations remaining, representing
24	failures in single-record/single-failure data
131.6167	total analysis time at risk and under observation
	at risk from t = 0
	earliest observed entry t = 0
	last observed exit t = 10.53251

208 sts graph



209 sts graph if diagnosis==1, by (mismatch) censored(multiple)



生存解析の結果の表示(点推定値)
216 sts list

```
. sts list
      failure_d: event_os == 1
      analysis time _t: lyear
```

Time	Req. Total	Fail	Net Lost	Survivor Function	Std. Error	[95% Conf. Int.]	
-.0274	50	1	0	0.9800	0.0190	0.8664	0.9972
-.052	49	1	0	0.9600	0.0277	0.8484	0.9659
-.0903	48	1	0	0.9400	0.0336	0.8254	0.9802
-.1287	47	1	0	0.9200	0.0384	0.8007	0.9692
-.2026	46	1	0	0.9000	0.0424	0.7763	0.9571
-.2875	45	0	1	0.9000	0.0424	0.7763	0.9571
-.2902	44	1	0	0.8795	0.0461	0.7513	0.9440
-.3176	43	1	0	0.8591	0.0494	0.7280	0.9302
-.3388	42	1	0	0.8386	0.0523	0.7030	0.9159
-.4189	41	1	0	0.8182	0.0549	0.6796	0.9010
-.5476	40	1	0	0.7977	0.0572	0.6566	0.8857
-.5585	39	1	0	0.7773	0.0593	0.6339	0.8700
-.6782	38	0	1	0.7773	0.0593	0.6339	0.8700
-.7064	37	1	0	0.7563	0.0613	0.6107	0.8536
-.7146	36	1	0	0.7353	0.0631	0.5879	0.8369
-.7255	35	0	1	0.7353	0.0631	0.5879	0.8369
-.7395	34	0	1	0.7353	0.0631	0.5879	0.8369
-.8254	33	1	0	0.7130	0.0650	0.5656	0.8191
-.8281	32	1	0	0.6907	0.0666	0.5396	0.8009
-.8991	31	1	0	0.6684	0.0681	0.5161	0.7824
-.9892	30	1	0	0.6461	0.0694	0.4930	0.7656
1.213	29	1	0	0.6239	0.0705	0.4703	0.7445

Log rank 検定

221 sts test mismatch

```
. sts test mismatch
```

```
      failure_d: event_os == 1
      analysis time _t: lyear
```

```
***
```

Log-rank test for equality of survivor functions

mismatch	Events observed	Events expected
match	14	14.43
mismatch	10	9.57
Total	24	24.00

```
chi2(1) = 0.03
Pr>chi2 = 0.8582
```

簡単な演習

模擬データセットにて別変数の作成、生存曲線の描出（群分け）、検定

H26年度 TRUMP統計セミナー
基礎編

簡単な演習

模擬データセットにて別変数の作成、生存曲線の
描出(群わけ)、検定

担当: 熱田由子

変数の決定と変数リストの作成

- 変数の種類
 - 患者背景変数
 - 患者年齢・性別
 - 疾患情報
 - 移植関連情報
 - アウトカム変数
 - イベント変数
 - 時間変数
- 研究計画が定まったら、解析に必要な変数リストを作成する

例1: ABO mismatch

データを覗いてみると・・・

	pt_abo	do_abo	abo_mach
1519.	O	O	適合
1520.			
1521.			
1522.	B	B	適合
1523.			
1524.	A	B	適合
1525.			
1526.	O	O	
1527.			
1528.			
1529.	O	O	
1530.			
1531.			
1532.	A	A	適合
1533.	A	A	

変数の作成
エクセル上で作業をしない！！

- 150-200例未満程度の場合は目視・手入力でも出来るかもしれないが・・・
- ソートを繰り返していたらデータが無茶苦茶・・・
- 気付いたらどこかで一行ずれてしまっていた・・・
- 変数作成のアルゴリズムを忘れてしまうことも・・・

“・・・”だらけ

日本造血細胞移植学会
JRMCT The Japanese Society for Hematopoietic Cell Transplantation

変数一覧(基本変数リスト)
(Stata/EZR共有スクリプト対応)

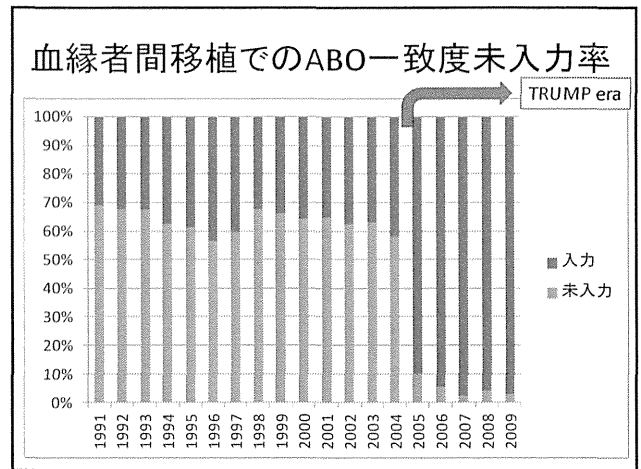
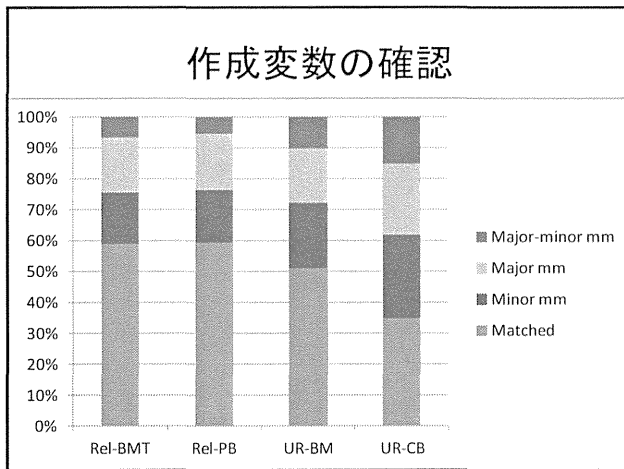
ダウンロード可能!!

変数一覧(基本変数リスト)
(Stata/EZR共有スクリプト対応)

ダウンロード可能!!

変数作成

- ABO matching
 - Matched
 - Minor mismatch
 - Major mismatch
 - Major-minor mismatch
- ABO major mismatch
 - (0) Matched or minor mismatch
 - (1) Major (major-minor) mismatch



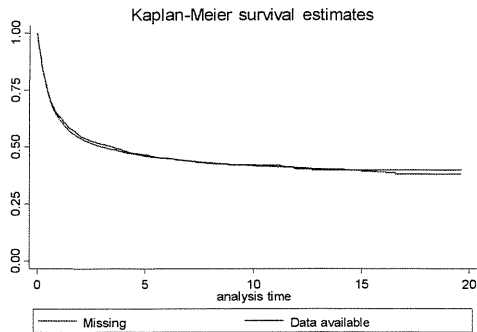
どうしよう?

- 研究課題
- ①. 「血縁者間移植においてABO一致度が生存に与える影響」
- ②. 「血縁者間移植におけるBMとPBの比較」

研究①「ABO一致度の影響」の場合

- 2005年以降の移植に限定した解析を行う
 - 観察期間が短い
- 入力があるものを採用する。その場合、2004年以前に関しては対象に入るのは3-4割程度。
 - 施設でのTRUMP以降の自主的入力
 - 未入力の場合と大きな差が(なんらかのバイアスがないか確認しておく必要がある。
- 2004年以前移植に関しては、入力率が7割以上の施設のデータのみを対象

1991-2004移植での確認



研究②「RBM vs. RPB」の場合

- “ABO matching”変数の使用目的
 - RBM vs. RPBの比較の際の補正因子
 - (主アウトカムは全生存率)
- 補正に用いる因子の候補
 - 検討するアウトカムに有意に影響することが過去の研究で既に明らかになっているもの
 - 比較群間に偏りがあるもの
- 選択
 - 補正因子として用いない
 - 補正因子として用いる

Missing dataの取り扱い -補正因子として用いる場合-

- 3つの方法
 - 1. 欠損値を解析対象からはずす
 - 2. 欠損値をひとつのカテゴリとする
 - 3. imputation: 似た患者の情報の平均値を入れる
- Klein JP says (Klein et al. BMT 2001):*
- 欠損値が少ないとき $\leq 10\%$ (EBMT guideline, 5%)
 - **1番** ただし、はずした場合とはずさなかった場合で結果が変わっていないかを確認する
- 欠損値が比較的多いとき $> 10\%$
 - **2番・3番** データ欠損患者をその他と比較することも可能

欠損値をひとつのカテゴリとする?

	HR	n value	
Stem Cell Source			
BM	1.00	67%	29%
PB	1.14	0.007	
ABO matching			
ABO matched or minor	1.00		
ABO major	0.99	0.923	
ABO missing	1.08	0.145	
Disease stage			
Standard	1.00		
Advanced	2.91	<0.0001	
Patient age			
age<40	1.00	100%	100%
Age>40	1.48	<0.0001	
		3,327例	2,255例

研究課題：移植アウトカムに対するドナーとレシピエントの血液型メジャーミスマッチの影響

1.0 目的：

ドナーとレシピエントの血液型メジャー不一致が、全生存、移植関連死亡を含む移植アウトカムに与える影響について検討する。

2.0 利用データ

- ・日本造血細胞移植学会：成人血縁者間移植
- ・日本小児血液学会：小児血縁者間移植
- ・骨髄移植推進財団：非血縁者間骨髄移植
- ・日本さい帯血バンクネットワーク：非血縁者間臍帯血移植

3.0 データセット

TRUMP Seminar 2014用サンプルデータ (TRUMP 2011固定データセットからの無作為抽出データ)

- ・ TRUMPSeminar2014

4.0 解析対象：

選択規準：

- ・対象疾患：急性白血病
(AML,ALL)
- ・患者年齢：年齢制限なし
- ・移植時期：2000年1月から2010年12月までに移植を受けた症例
- ・移植回数：第1回目の同種移植症例のみ
- ・ドナー・レシピエントの血液型一致度が判明しているもの

除外規準：

- ・複数臍帯血移植
- ・UR-PBSCT, BMT+PBSCTを除く (極めて少数のため)
- ・幹細胞種類が不明、あるいはその他
- ・ドナーが不明のもの
- ・GVHD予防なしのもの
- ・生存状況あるいは生存時間が不明のもの (主要評価項目の解析が不可)

5.0 アウトカム :

Primary endpoint

- ・ 全生存

Secondary endpoints

- ・ 非再発死亡(non-relapse mortality, NRM)
- ・ 再発
- ・ 好中球回復
- ・ grade 2 to 4 acute GVHD

解析に用いる変数 :

Main effects :

- ・ matched or minor mismatched vs. major mismatched or major-minor mismatched

Recipient-related :

- ・ 年齢 : 0-15, 16-39, 40-
- ・ 原疾患 : AML, ALL
- ・ 移植時病期 :
白血病のみ
Standard(AML : CR1, CR2, ALL : 1CR, CML : 1CP, MDS : RA/RARS)
Advanced(all others)
- ・ 移植前PS : 0-1 vs. ≥ 2

Donor-related :

- ・ ドナーとの関係 : 血縁 vs. 非血縁
- ・ 幹細胞種類 : BM vs. PBSC vs. CB
- ・ HLA適合 : HLA well matched(including identical) vs. HLA partially matched vs. mismatched
- ・ (こちらを採用) ドナー-細胞種類 combination :
Rel-BM
Rel-PB
UR-BM
UR-CB

Transplant-related :

- ・ 移植前処置 : TBI照射なし vs. TBI照射あり
- ・ 移植前処置 : 骨髄破壊的(MAC) vs. 骨髄非破壊的(RIC)
- ・ GVHD予防 : CsA based vs. Tac based

6.0 研究デザイン

本邦において1回目の血縁者間末梢血幹細胞・骨髄移植、非血縁者間骨髄移植、非血縁者間臍帯血移植を施行された成人の造血器悪性疾患症例を対象として、ドナーとレシピエントの血液型の組み合わせが、全生存、非再発死亡、好中球回復、急性GVHD、あるいは再発に及ぼす影響を検討する。

血液型の組み合わせでは、ドナー・レシピエント間マッチに対しての他の組み合わせのもの全生存に対するリスクを単変量、および多変量解析で実施し、最終解析における群分け方法を決定する。

ドナー・移植種類によるサブグループ解析を追加し、血液型の影響がどのような移植において明らかであるか検討する。

7.0 統計解析

7.0.1 主評価項目の解析

(1) 全生存

Main effect 群分け変数において、移植日 (day0) を起算日とし、原因を問わない死亡をイベントとして Kaplan-Meier 法を用いて全生存曲線の描出および全生存率の算出を行う。群間比較は logrank test で検定する。同アウトカムに関して 8.0.4 で示す方法での多変量解析による補正解析を実施する。

7.0.2 副次的評価項目の解析

(1) 非再発死亡 (Non-relapse mortality)

Main effect 群分け変数において、移植日 (day0) を起算日とし、寛解時死亡(非再発死亡)をイベントとし、再発を競合リスクイベントとして Cumulative incidence 法を用いて再発率を算出し、Cumulative incidence curve を描出する。群間比較は logrank test および Gray's test で検定する。同アウトカムに関して 8.0.4 で示す方法での多変量解析による補正解析を実施する。

(2) 再発

Main effect 群分け変数において、移植日 (day0) を起算日とし、再発をイベントとし、寛解時死亡を競合リスクイベントとして Cumulative incidence 法を用いて再発率を算出し、Cumulative incidence curve を描出する。群間比較は logrank test および Gray's test で検定する。同アウトカムに関して 8.0.4 で示す方法での多変量解析による補正解析を実施する。

(3) 好中球回復

Main effect 群分け変数において、移植日 (day0) を起算日とし、好中球回復 (3ポイント以上好中球 $500/\mu\text{l}$ 以上を測定できた場合に好中球回復とし、最初の検査日を回復日とする) をイベントとし、回復前死亡を競合リスクイベントとして Cumulative incidence 法を用いて好中球回復率を算出し、Cumulative incidence

curve を描出する。群間比較は logrank test および Gray's test で検定する。同アウトカムに関して 8.0.4 で示す方法での多変量解析による補正解析を実施する。

(4) Grade 2 to 4 急性 GVHD

Main effect 群分け変数において、移植日 (day0) を起算日とし、II 度以上の急性 GVHD をイベントとし、II 度以上の急性 GVHD 発症前死亡と II 度以上の急性 GVHD 発症前再発を競合リスクイベントとして Cumulative incidence 法を用いて好中球回復率を算出し、Cumulative incidence curve を描出する。群間比較は logrank test および Gray's test で検定する。同アウトカムに関して 8.0.4 で示す方法での多変量解析による補正解析を実施する。

7.0.3 サブグループ解析

以下の移植種別における主評価項目および副次的評価項目のサブグループ解析を行う。

- ・血縁者間移植
- ・非血縁者間骨髄移植
- ・非血縁者間臍帯血移植

7.0.4 多変量解析

以下の因子で補正し、再発率、全生存率(OS)および移植関連死亡(TRM)の多変量解析を行う。

- ・年齢 (0-15, 16-39, 40-)
- ・患者性別 (F/M)
- ・患者・ドナー性別一致度 (match, M to F, F to M)
- ・原疾患 (AML, ALL)
- ・移植時病期 (standard, advanced)
- ・ドナー---細胞種類 combination :
 - Rel-BM
 - Rel-PB
 - UR-BM
 - UR-CB
- ・移植前処置 (MAC / RIC)
- ・GVHD 予防 (CyA-based, Tac-based)

全生存は、Cox 比例ハザードモデルを用い、競合リスクアウトカム (非再発死亡、再発、好中球回復、急性 GVHD) は、競合リスク回帰モデルを用いる。

多変量解析のモデルに加える補正因子の選択方法は、いくつかあるが、セミナーでは以下の方法①および③-1 を実践する。

- ① 変数をすべて用いるフルモデルで補正する。

：文献や医学的見地を参照し、主要評価項目全生存に対するリスク因子を用いる。

- ② 群間において比較(検定)を行い、単変量解析において、有意にあるいは marginal に ($P < 0.1$) 検討するエンドポイントに影響を与えた因子をモデルに加える。
- ③ 選択した因子をモデルに加える。
 - ③-1 : 全ての検討する因子を含めたフルモデルからスタートし、血液型の変数は常に残した上で、backward stepwise にて P 値が高い変数から順にフルモデルから除き血液型の変数以外は有意な因子のみを残した最終モデルとする。
 - ③-2 : 血液型を除いた変数のみを用いて、forward or backward stepwise にて有意な因子のみを残したモデルを最初に作成する。絞り込まれた因子と目的の因子をモデルに加えて最終モデルとする。

8.0 背景表

TRUMP統計セミナー資料集
(応用編)