

	Variables	Value	Value label	コメント
		4		4座ミスマッチ
		5		5座ミスマッチ
		6		6座ミスマッチ
血清型 A,B,C, DRミスマッチ 座数	seromis8abcdr	0		0座ミスマッチ
		1		1座ミスマッチ
		2		2座ミスマッチ
		3		3座ミスマッチ
		4		4座ミスマッチ
		5		5座ミスマッチ
		6		6座ミスマッチ
		7		7座ミスマッチ
		8		8座ミスマッチ
血清型 A,B,C, DRミスマッチ 座数	seromis8abcdr2	0		0座ミスマッチ
		1		1座ミスマッチ
		2		2座ミスマッチ
		3		3座ミスマッチ
		4		4座ミスマッチ
		5		5座ミスマッチ
		6		6座ミスマッチ
		7		7座ミスマッチ
		8		8座ミスマッチ
遺伝子型 A GVH方向ミスマッチ 座数	genogvhmisa	0		0座ミスマッチ
		1		1座ミスマッチ
		2		2座ミスマッチ
遺伝子型 A HVG方向ミスマッチ 座数	genohvgmisa	0		0座ミスマッチ
		1		1座ミスマッチ
		2		2座ミスマッチ
遺伝子型 B GVH方向ミスマッチ 座数	genogvhmisb	0		0座ミスマッチ
		1		1座ミスマッチ
		2		2座ミスマッチ
遺伝子型 B HVG方向ミスマッチ 座数	genohvgmisb	0		0座ミスマッチ
		1		1座ミスマッチ
		2		2座ミスマッチ
遺伝子型 C GVH方向ミスマッチ 座数	genogvhmisc	0		0座ミスマッチ
		1		1座ミスマッチ
		2		2座ミスマッチ
遺伝子型 C HVG方向ミスマッチ 座数	genohvgmisc	0		0座ミスマッチ
		1		1座ミスマッチ
		2		2座ミスマッチ
遺伝子型 DR GVH方向ミスマッチ 座数	genogvhmisdr	0		0座ミスマッチ
		1		1座ミスマッチ
		2		2座ミスマッチ
遺伝子型 DR HVG方向ミスマッチ 座数	genohvgmisdr	0		0座ミスマッチ
		1		1座ミスマッチ
		2		2座ミスマッチ
遺伝子型 A,B,DRB1 GVH方向ミスマッチ 座数	genogvhmis6abdr	0		0座ミスマッチ
		1		1座ミスマッチ
		2		2座ミスマッチ
		3		3座ミスマッチ
		4		4座ミスマッチ
		5		5座ミスマッチ
		6		6座ミスマッチ
遺伝子型 A,B,DRB1 HVG方向ミスマッチ 座数	genohvgmis6abdr	0		0座ミスマッチ
		1		1座ミスマッチ
		2		2座ミスマッチ
		3		3座ミスマッチ
		4		4座ミスマッチ
		5		5座ミスマッチ
		6		6座ミスマッチ

	Variables	Value	Value label	コメント
遺伝子型 Aミスマッチ座数	genomisa	0		genogvhmisaとgenohvgmisaのうち大きい方
		1		
		2		
遺伝子型 Bミスマッチ座数	genomisb	0		genogvhmisaとgenohvgmisaのうち大きい方
		1		
		2		
遺伝子型 Cミスマッチ座数	genomisc	0		genogvhmiscとgenohvgmiscのうち大きい方
		1		
		2		
遺伝子型 DRミスマッチ座数	genomisd	0		genogvhmisdとgenohvgmisdのうち大きい方
		1		
		2		
遺伝子型 A,B,DRB1 ミスマッチ 座数	genomis6abdr	0		0座ミスマッチ
		1		1座ミスマッチ
		2		2座ミスマッチ
		3		3座ミスマッチ
		4		4座ミスマッチ
		5		5座ミスマッチ
		6		6座ミスマッチ
遺伝子型 A,B,DRB1 ミスマッチ 座数	genomis6abdr2	0		0座ミスマッチ
		1		1座ミスマッチ
		2		2座ミスマッチ
		3		3座ミスマッチ
		4		4座ミスマッチ
		5		5座ミスマッチ
		6		6座ミスマッチ
遺伝子型 A,B,C,DRB1 GVH方向ミスマッチ 座数	genogvhmis8abcdr	0		0座ミスマッチ
		1		1座ミスマッチ
		2		2座ミスマッチ
		3		3座ミスマッチ
		4		4座ミスマッチ
		5		5座ミスマッチ
		6		6座ミスマッチ
		7		7座ミスマッチ
		8		8座ミスマッチ
遺伝子型 A,B,C,DRB1 HVG方向ミスマッチ 座数	genohvgmis8abcdr	0		0座ミスマッチ
		1		1座ミスマッチ
		2		2座ミスマッチ
		3		3座ミスマッチ
		4		4座ミスマッチ
		5		5座ミスマッチ
		6		6座ミスマッチ
		7		7座ミスマッチ
		8		8座ミスマッチ
遺伝子型 A,B,C,DRB1 ミスマッチ 座数	genomis8abcdr	0		0座ミスマッチ
		1		1座ミスマッチ
		2		2座ミスマッチ
		3		3座ミスマッチ
		4		4座ミスマッチ
		5		5座ミスマッチ
		6		6座ミスマッチ
		7		7座ミスマッチ
		8		8座ミスマッチ
遺伝子型 A,B,C,DRB1 ミスマッチ 座数	genomis8abcdr2	0		0座ミスマッチ
		1		1座ミスマッチ
		2		2座ミスマッチ
		3		3座ミスマッチ
		4		4座ミスマッチ
		5		5座ミスマッチ

	Variables	Value	Value label	コメント
		6		6座ミスマッチ
		7		7座ミスマッチ
		8		8座ミスマッチ
HLA適合度別ドナーグループ				

	Variables	Value	Value label	コメント
生存状況	event_os	0		生存
		1		死亡
最終確認日	ldate			
生存期間 day	lday			ldate - tdate
生存期間 year	lyear			lday / 365.25
生存解析可能症例	sa	0		生存状況データなし、最終確認日データなし、最終確認日データ < 0
		1		
Relapse-free Survival	event_rfs	0		
		1		死亡または再発あり
Relapse-free Survivalイベントまでの期間	rfs_day			移植後の疾患の状況=="移植後も寛解ならず"の場合は day0.1で再発とする
	rfs_year			移植後の疾患の状況=="移植後も寛解ならず"の場合は day0.1で再発とする

	Variables	Value	Value label	コメント
再発				
clinical/hematological relapse	event_relapse	1	再発	移植後の疾患の状況 =="移植後も寛解ならず" "の場合はday0.1で再発とする。
		2	寛解時死亡	
		3	Censor	
再発日	rel_date			
再発までの期間 day	rel_day			rel_date - tdate 数値、移植後の疾患の状況 =="移植後も寛解ならず" "の場合はday0で再発とする。
time variable for event_relapse	ci_rel_day			再発あり=再発までの期間、寛解ならず、再発なし=生存期間
好中球/血小板回復				
好中球500以上到達イベント	event_neut500	1	回復	500以下にならなかった場合はday0.1で回復とする
		2	回復前死亡	
		3	Censor	
好中球500以上到達日付	neut500_d			
好中球500以上到達までの期間 day	neut500_day			neut500_d - tdate
time variable for event_neut	ci_neut500			
血小板2万以上到達イベント	event_plt2	1	回復	
		2	回復前死亡	
		3	Censor	plt2_d - tdate
血小板2万以上到達日付	plt2_d			
血小板2万以上到達までの期間 day	plt2_day			
time variable for event_plt2	ci_plt2			
血小板5万以上到達イベント	event_plt5	1	回復	
		2	回復前死亡	
		3	Censor	
血小板5万以上到達日付	plt5_d			
血小板5万以上到達までの期間 day	plt5_day			plt5_date - tdate
time variable for event_plt5	ci_plt5			
急性GVHD				
急性GVHD Grade	agvhd_grade_manip	0		No aGVHD
		1		Grade1
		2		Grade2
		3		Grade3
		4		Grade4
		8		評価不能
		9		未記入
Grade2 to Grade4 acuteGVHD	event_agvhd24	1	Grade2-4 aGVHD	
		2	Death before grade2-4 agvhd	
		3	Relapse before grade2-4 agvhd	
		4	Censor	
	agvhd_day2			皮膚出現日,肝出現日, 腸管出現日のうちの最も 早い日付(aGVHD発症日、 Ⅱ度以上の発症日、 Ⅲ度以上の発症日いずれ のデータもない例に、 代入するため)

	Variables	Value	Value label	コメント
Grade2 to Grade4 発症の有無				
Grade2 to Grade4 発症までの期間 day	agvhd24_day			
time variable for event_agvhd24	ci_agvhd24			上限を200日
Grade3 to Grade4 acuteGVHD	event_agvhd34	1	Grade3-4 aGVHD	
		2	Death before grade3-4 agvhd	
		3	Relapse before grade3-4 agvhd	
		4	Censor	
Grade3 to Grade4 発症の有無				
Grade3 to Grade4 発症までの期間 day	agvhd34_day			
time variable for event_agvhd34	ci_agvhd34			上限を200日
慢性GVHD	event_cgvhd	1	cGVHD	
		2	Death before cGVHD	
		3	Relapse before cGVHD	
		4	Censor	
慢性GVHD発症の有無				
限局型慢性GVHD				
広範型慢性GVHD				
chronic GVHD発症日付	cgvhd_d			
chronic GVHD発症までの期間 day	cgvhd_day			
time variable for event_cgvhd	ci_cgvhd			
	ci_cgvhd_year			
extensive-type chronic GVHD	event_excgvhd	1	Ex cGVHD	
		2	Death before ex cGVHD	
		3	Relapse before ex cGVHD	
		4	Censor	
time variable for event_cgvhd_grade	ci_excgvhd			
	ci_excgvhd_year			

《スクリプト集》

```
1 *StataSeminar用に一部改変
2
3 ///***データの取り込み***///
4 *****Stata12からはエクセルファイルのまま取り込みが可能*****
5 /*
6 横断的WGでフルデータを取り込む場合など、Excelファイルの大きさが
7 40メガバイトを超える場合はExcelファイルからの取り込みができません。
8 その場合は、CSVファイルから取り込んでください。
9 */
10
11 *****事前準備*****
12 **項目が英語名のデータセットを使用する
13 **2010年度データセットに関しては、JSHCT HPに英語項目名変換用のエクセルマクロあるのでそれを使用
14 **ファイルの名前も英語にしておく (TRUMP2010とか)
15 **最終行に入力率が入っている場合はこれを削除する
16
17 *****読み込みコマンド*****
18 **読み込みコマンド、以下のusing以降はファイルの場所 (directory) を示します。
19 **directoryですが、以下は、Cの中に"data"というフォルダを作成し、元データを入れてあります
20 **オプションコマンドのnameは、第一行のデータが項目名であることを示す。
21 **ちなみに、オプションコマンドとは、カンマ以降のコマンドのことです。
22
23 *****ミニ知識*****
24 *米印が先頭に来る場合、あるいは/*と*/で囲んだところは、メモ書きとしてStataは認識します。
25
26 *****以下はcsvファイルからの取り込み*****
27 insheet using C:\%data%\StataSeminar2012.csv, name
28 *読み込みに多少時間はかかります、あせらないでくださいね!!
29
30 *****以下はエクセルファイル (xlsファイル) からの取り込み*****
31 import excel using C:\%data%\StataSeminar2012.xls, firstrow
32
33
34 *****以下は最近のエクセルファイル (xlsxファイル) からの取り込み*****
35 import excel using C:\%data%\StataSeminar2012.xlsx, firstrow
36
37
38
39
```



```

1  //****変数作成の基礎 ****//
2
3
4  ****患者・ドナー性別****
5  *患者、ドナーそれぞれに0 "Female" 1 "Male"という変数を作りたい。
6  *TRUMP構造は?
7
8  tab sex
9  gen pt_sex = 0 if sex=="女"
10 replace pt_sex = 1 if sex=="男"
11
12 *if以下の"等しい"は、"=="
13 *文字列の値は、""で囲む。数値は囲まない
14
15 *コードの意味を忘れてしまうこともあるが、以下のラベルが便利
16 label define sex_label 0 "Female" 1 "Male"
17 label values pt_sex sex_label
18
19 *後で書かれるもので上書きされます。
20 tab do_sex
21 gen donor_sex = 0
22 replace donor_sex = 1 if do_sex=="男"
23
24 *これだとどうなるか
25 tab donor_sex do_sex, missing
26 drop donor_sex
27
28 gen donor_sex = 0 if do_sex=="女"
29 replace donor_sex = 1 if do_sex=="男"
30
31 label values donor_sex sex_label
32
33 **help operator
34 **& (and かつ), | (or または), = (equal 等しい), != (not equal 等しくない)
35
36 ****Sex mismatch1****
37 *患者・ドナー間の性別一致度の変数、 0 "Match" 1 "Mismatch"を作りたい。
38 gen sex_mismatch1 = 0 if sex=="男" & do_sex=="男"
39 replace sex_mismatch1 = 0 if sex=="女" & do_sex=="女"
40 replace sex_mismatch1 = 1 if sex=="男" & do_sex=="女"
41 replace sex_mismatch1 = 1 if sex=="女" & do_sex=="男"
42
43 label define sex_mismatch1_label 0 "Match" 1 "Mismatch"
44 label values sex_mismatch1 sex_mismatch1_label
45
46 ///***ラベルに関して***///
47 *ラベルは、変数名のラベルと、値のラベルがあります。
48 *つけたラベルがわからなくなってしまうたら；
49 tab sex_mismatch1
50 tab sex_mismatch1, nolabel
51 describe sex_mismatch1
52 label list sex_mismatch1_label
53
54 ///***やってみよう！***///
55 ****Sex mismatch2****
56 *患者・ドナー間の性別一致度の変数、 0 "Match" 1 "M to F" 2 "F to M"を作りたい。
57 gen sex_mismatch2 =
58
59
60
61 ****ABO mismatch****
62 *患者・ドナー間のABO血液型一致度の変数、0 "Matched" 1 "Minor mismatch" 2 "Major mismatch" 3
  "Major-Minor mismatch"を作りたい。
63 *TRUMP dataは
64 tab pt_abo, missing
65 tab do_abo, missing
66 tab abo_match, missing
67 *abo_matchの入力のほうが少ないが、患者、ドナーABO型が未入力で、abo_match情報があるものもある？
68 *data sheetを開かないでdatasetの内容を把握する

```

```

69 tab abo_match if pt_abo=="
70 tab abo_match if do_abo=="
71 *=="", 文字列の場合の空欄を示す。数値の場合は=.
72
73 gen abo_mismatch = 0 if pt_abo==do_abo & pt_abo!="" & do_abo!="
74 replace abo_mismatch = 1 if pt_abo=="AB" & do_abo=="A" | pt_abo=="AB" & do_abo=="B"
75 replace abo_mismatch = 1 if pt_abo=="AB" & do_abo=="0" | pt_abo=="A" & do_abo=="0" | pt_abo=="
76 replace abo_mismatch = 2 if pt_abo=="0" & do_abo=="AB" | pt_abo=="0" & do_abo=="A" | pt_abo=="
77 replace abo_mismatch = 2 if pt_abo=="A" & do_abo=="AB" | pt_abo=="B" & do_abo=="AB"
78 replace abo_mismatch = 3 if pt_abo=="A" & do_abo=="B" | pt_abo=="B" & do_abo=="A"
79
80 label define abo_mismatch_label 0 "Matched" 1 "Minor mismatch" 2 "Major mismatch" 3
81 label values abo_mismatch abo_mismatch_label
82
83 tab abo_mismatch, missing
84 tab abo_mismatch abo_match
85
86 ///***やってみよう!***///
87 *****ABO major mismatch*****
88 *Major mismatchは1, そうでない場合 (match or minor mismatch) 0
89 gen abo_majormis =
90
91 *****ABO minor mismatch*****
92 *Minor mismatchは1, そうでない場合 (match or major mismatch) 0
93 gen abo_minormis =
94
95
96
97
98 *****移植時患者年齢*****
99 *9999は年齢未記入の場合の入力のため、未記入"."とする
100 sum age, detail
101 gen pt_age = age
102 replace pt_age = . if pt_age==9999
103
104 sum pt_age, detail
105 count if pt_age==.
106
107 ///***やってみよう!***///
108 *****移植時患者年齢グループ1*****
109 *移植時患者年齢ごとに1 "0-15" 2 "16-39" 3 "40-"とする患者年齢グループ変数を作成したい。
110 gen pt_age_group1 =
111
112
113
114 *****移植日と移植年*****
115 gen tdate = date(tx_date, "YMD")
116
117 *Stataは1960年1月1日を基準に整数にて日付を計算している。
118 *""31/Jan/1960"などの記載の場合は、これに合わせて"DMY"とする。
119 **Excelファイルからデータを取り込んだ場合、Excel日付データはStata日付データに自動変換されてい
120 るため、上記の日付コマンド実行は不要です。
121 /*以下のコマンドを実行してください。
122 (*と*/間のコマンドを実行する場合は、「/*」 「*/」を取ってください)
123 gen tdate = tx_date
124 */
125
126 gen tyear = year(tdate)
127 tab tyear
128
129 *****診断から移植までの日数*****
130 gen ddate = date(dx_date, "YMD")
131 gen dx_to_sct = tdate - ddate
132
133 ///***やってみよう!***///

```

```

134 *****最終確認日をもとに、観察期間 (days and years) *****
135 *最終確認日 : last_date
136 gen ldate = date(
137 gen lday =
138 gen lyear =
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165 ///***やってみよう、solution***///
166 *****Sex mismatch2*****
167 gen sex_mismatch2 = 0 if sex=="男" & do_sex=="男"
168 replace sex_mismatch2 = 0 if sex=="女" & do_sex=="女"
169 replace sex_mismatch2 = 1 if sex=="女" & do_sex=="男"
170 replace sex_mismatch2 = 2 if sex=="男" & do_sex=="女"
171
172 label define sex_mismatch2_label 0 "Match" 1 "M to F" 2 "F to M"
173 label values sex_mismatch2 sex_mismatch2_label
174
175 *****ABO major mismatch*****
176 gen abo_majormis = 0 if abo_mismatch<=1
177 replace abo_majormis = 1 if abo_mismatch>=2 & abo_mismatch<.
178
179 *****ABO minor mismatch*****
180 gen abo_minormis = 0 if abo_mismatch==0 | abo_mismatch==2
181 replace abo_minormis = 1 if abo_mismatch==1 | abo_mismatch==3
182
183 *****移植時患者年齢グループ1*****
184 count if pt_age==.
185 gen pt_age_group1 = 1 if pt_age<16
186 replace pt_age_group1 = 2 if pt_age>=16 & pt_age<40
187 replace pt_age_group1 = 3 if pt_age>=40 & pt_age<.
188 *未記入"."は、連続変数の場合最も大きい数字と認識されているため、上記のようにしないとグループ3
    にふくまれてしまう。
189
190 label var pt_age_group1 "年齢グループ1"
191 label define pt_age_group1_label 1 "0-15" 2 "16-39" 3 "40-"
192 label values pt_age_group pt_age_group1_label
193 tab pt_age_group
194
195 *****最終確認日をもとに、観察期間 (days and years) *****
196 gen ldate = date(last_date, "YMD")
197 gen lday = ldate - tdate
198 *Day0死亡に対応するために以下を追加する
199 replace lday = 0.1 if lday==0
200 *生死不明の場合は欠損値とする
201 replace lday = . if survival=="

```

```
202   gen lyear = lday / 365.25
203
204
205   *****変数およびラベルのdrop*****
206   drop pt_sex donor_sex sex_mismatch1 sex_mismatch2 abo_mismatch abo_majormis abo_minormis /*
207   */ pt_age pt_age_group1 tdate tyear ddate dx_to_sct ldate lday lyear
208
209   label drop sex_label sex_mismatch1_label sex_mismatch2_label abo_mismatch_label
   pt_age_group1_label
210
211
212
213
214
```

```

1  /////*****変数作成の復習*****/////
2
3  *****やってみよう！その1
4  *****ps_cat
5  *****PS (変数名 ps_tx)を1=0, 1と2=2, 3, 4の2グループに分類してみましょう。
6  *****ラベルも0-1, 2-4とつけましょう。
7
8  gen      ps_cat=
9
10
11 *****やってみよう！その2
12 *****ps_cat2
13 *****PS
14 (変数名 ps_tx)を1=0, 1と2=2, 3, 4の2グループに加えて9=不明のグループに分類してみましょう。
15 *****ラベルも0-1, 2-4, Unknownとつけましょう。
16
17 gen      ps_cat2=
18
19 *****やってみよう！その3
20 *****レシピエント・ドナーのCMV抗体の組み合わせで分類しましょう。
21 *****レシピエントCMV抗体(変数名 cmvab_tx)、ドナーCMV抗体(変数名 do_cmvab)
22 *****0=R-D-, 1=R+D-, 2=R-D+, 3=R+D+に分類しましょう。
23 *****臍帯血(source->4)のドナーCMV抗体は陰性としましょう。
24
25 gen      cmvab_cat=
26
27
28 *****やってみよう！その4
29 *****レシピエント・ドナーのCMV抗体の組み合わせで分類しましょう。
30 *****レシピエントCMV抗体(変数名 cmvab_tx)、ドナーCMV抗体(変数名 do_cmvab)
31 *****0=R-D-, 1=R+D-, 2=R-D+, 3=R+D+, 9=Unknownに分類しましょう。
32 *****臍帯血(source->4)のドナーCMV抗体は陰性としましょう。
33
34 gen      cmvab_cat2=
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60 *****やってみよう！ Answer
61 *****ps_cat
62 tab ps_tx, missing
63 gen      ps_cat=1 if ps_tx=="0" | ps_tx=="1"
64 replace ps_cat=2 if ps_tx=="2" | ps_tx=="3" | ps_tx=="4"
65 label define ps_cat_label 1 "0-1" 2 "2-4"
66 label values ps_cat ps_cat_label
67 tab ps_tx ps_cat, missing
68

```

```

69 *****ps_cat2
70 gen ps_cat2=9
71 replace ps_cat2=1 if ps_tx=="0" | ps_tx=="1"
72 replace ps_cat2=2 if ps_tx=="2" | ps_tx=="3" | ps_tx=="4"
73 label define ps_cat2_label 1 "0-1" 2 "2-4" 9 "Unknown"
74 label values ps_cat2 ps_cat2_label
75 tab ps_tx ps_cat2, missing
76
77 *****cmvab_cat
78 tab cmvab_tx, missing
79 tab do_cmvab, missing
80 gen cmvab_cat=0 if cmvab_tx=="無" & do_cmvab=="無"
81 replace cmvab_cat=1 if cmvab_tx=="有" & do_cmvab=="無"
82 replace cmvab_cat=2 if cmvab_tx=="無" & do_cmvab=="有"
83 replace cmvab_cat=3 if cmvab_tx=="有" & do_cmvab=="有"
84 replace cmvab_cat=0 if cmvab_tx=="無" & (tx_type=="臍帯血(CBSCT)" | tx_type==
"複数臍帯血(MultiCBSCT)")
85 replace cmvab_cat=1 if cmvab_tx=="有" & (tx_type=="臍帯血(CBSCT)" | tx_type==
"複数臍帯血(MultiCBSCT)")
86 label define cmvab_cat_label 0 "R-D-" 1 "R+D-" 2 "R-D+" 3 "R+D+"
87 label values cmvab_cat cmvab_cat_label
88 tab cmvab_cat, missing
89
90 *****cmvab_cat2
91 gen cmvab_cat2=9
92 replace cmvab_cat2=0 if cmvab_tx=="無" & do_cmvab=="無"
93 replace cmvab_cat2=1 if cmvab_tx=="有" & do_cmvab=="無"
94 replace cmvab_cat2=2 if cmvab_tx=="無" & do_cmvab=="有"
95 replace cmvab_cat2=3 if cmvab_tx=="有" & do_cmvab=="有"
96 replace cmvab_cat2=0 if cmvab_tx=="無" & (tx_type=="臍帯血(CBSCT)" | tx_type==
"複数臍帯血(MultiCBSCT)")
97 replace cmvab_cat2=1 if cmvab_tx=="有" & (tx_type=="臍帯血(CBSCT)" | tx_type==
"複数臍帯血(MultiCBSCT)")
98 label define cmvab_cat2_label 0 "R-D-" 1 "R+D-" 2 "R-D+" 3 "R+D+" 9 "Unknown"
99 label values cmvab_cat2 cmvab_cat2_label
100 tab cmvab_cat2, missing
101
102
103
104
105
106
107 ///***string variable***///
108 *自由記載情報の整理
109 tab cod_main_o
110 /*敗血症でも、“他(敗血症)”, “敗血症”, “敗血症ショック”など様々である
111 敗血症をこの中から抽出する場合には*/
112 gen cod_sepsis = 0 if cod_main_o=="他(敗血症)"
113 replace cod_sepsis = 0 if cod_main_o=="敗血症"
114 replace cod_sepsis = 0 if cod_main_o=="敗血症ショック"
115 replace cod_sepsis = 0 if cod_main_o=="敗血症性ショック"
116 tab cod_sepsis, missing
117 tab cod_main_o cod_sepsis, missing
118
119 *あるいは以下なども利用できる
120 gen cod_sepsis2 =indexnot("敗血症", cod_main_o)
121 *cod_main_oに敗血症が含まれれば0、含まれなければ1やその他の数字
122 tab cod_sepsis2, missing
123 tab cod_main_o cod_sepsis2, missing
124
125 *String変数の整理
126 tab cod_main
127 *":以降で詳細な分類がなされている
128 split cod_main, p(":")
129 tab cod_main1
130 tab cod_main2 if cod_main1=="臓器不全"
131
132
133 *****変数およびラベルのdrop*****

```

```

134 drop ps_cat ps_cat2 cmvab_cat cmvab_cat2
135
136 label drop ps_cat_label ps_cat2_label cmvab_cat_label cmvab_cat2_label
137
138 drop cod_sepsis cod_sepsis2 cod_main1 cod_main2
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153 //****共有スクリプトを利用した、解析可能なデータシートを作成****//
154 /*
155 11_gv_characteristics_120522
156 12_gv_outcome_120627
157 のスクリプトをそれぞれすべて実行する
158 */
159
160 /*もしエラー等のためデータを再度読み込む必要がでた場合は、
161 clear
162 insheet using C:\data\StataSeminar2012.csv, name
163 を行い、
164 (01_label_variable_wg17-21_120326)
165 11_gv_characteristics_120522
166 12_gv_outcome_120627
167 のscriptをまわせば大丈夫なはず
168 */
169
170
171
172
173
174
175 //****解析データセットの作成****//
176 /*Sample protocolより
177 選択規準：
178 1) ・対象疾患：急性白血病（AML, ALL）
179 2) ・患者年齢：年齢制限なし（不明は除く）
180 3) ・移植時期：2000年1月から2009年12月までに移植を受けた症例
181 4) ・移植回数：第1回目の同種移植症例のみ
182 5) ・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
183
184 除外規準：
185 6) ・複数臍帯血移植
186 7) ・UR-PBSCT, BMT+PBSCTを除く（極めて少数のため）
187 8) ・幹細胞種類が不明、あるいはその他
188 9) ・ドナーが不明のもの
189 10) ・GVHD予防なしのもの
190 11) ・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）
191 */
192
193
194
195
196 tab diagnosis2, missing
197 keep if diagnosis2==1 | diagnosis2==2 /* (1) */
198
199 count if pt_age==.
200 drop if pt_age==. /* (2) */
201
202 tab tyear, missing

```

```

203 keep if tyear>=2000 & tyear<=2010 /* (3) */
204
205 tab no, missing
206 keep if no==1 /* (4) */
207
208 tab abo_mismatch, missing
209 drop if abo_mismatch==. /* (5) 13% of patients were excluded*/
210
211 tab tx_type, missing
212 drop if tx_type=="複数臍帯血(MultiCBSCT)" /* (6) */
213
214 tab source, missing
215 tab source, missing nolabel
216
217 drop if source==3 /* (7) */
218 drop if source==5 /* (8) */
219 drop if source==. /* (8) */
220
221 tab sct_type, missing
222 tab sct_type, missing nolabel
223 drop if sct_type==5 /* (7) */
224 drop if sct_type==. /* (9) */
225
226 tab gvhd_pro, missing
227 tab gvhd_pro, missing nolabel
228 drop if gvhd_pro==0 /* (10) */
229
230 tab sa, missing
231 keep if sa==1 /* (11) */
232 count
233
234
235 *一度dropしたデータは戻ってきません。
236 *もし解析対象外のデータも残しておきたいなら、、、
237 *解析対象をexclude = 0として解析対象外をexclude = 1としてみましよう
238 gen exclude = 0
239
240 tab diagnosis2, missing
241 replace exclude = 1 if diagnosis2!=1 & diagnosis2!=2 /* (1) */
242
243 count if pt_age==.
244 replace exclude = 1 if pt_age==. /* (2) */
245
246 /*以下省略、、、*/
247
248
249
250
251
252
253 //****背景表の作成と群間の検定****//
254 ****解析結果のログを取る****
255 ****ログファイルを作成****
256 *結果はStata
257 Resultsウィンドウに表示されますが、結果が長くなると遡って見ることができなくなります。
258 *ログを取ると結果を遡って見るのに便利です。テキストファイルで保存されるため、メモ帳などで見る
259 *ことができます。
260 *例
261 /*log using ログファイル名
262 xxx
263 log off
264 yyy
265 log on
266 zzz
267 log close
268 translate ログファイル名.smcl ログファイル名.txt*/
269 *ログファイルに「xxx」と「zzz」の部分のログは記録されますが、yyyの部分は記録されません。
270 *textファイルをExcelで開くことができるため、table作成の際に便利です。
271 *Excelで開く際には、開く→「カンマやタブなどの区切り文字によってフィールドごとに区切られたデー

```



```
270 タ」を選択
271 *→「次へ」→区切り文字として「タブ」に加え「スペース」をチェック→「次へ」→「完了」
272
273 log using characteristics
274
275 tab abo_mismatch, missing
276
277 tab pt_age_group1 abo_majormis, missing
278 sort abo_majormis
279 by abo_majormis: sum pt_age, detail
280 tab pt_sex abo_majormis, missing
281 tab sex_mismatch2 abo_majormis, missing
282 tab diagnosis2 abo_majormis, missing
283 tab stage1 abo_majormis, missing
284 tab sct_type abo_majormis, missing
285 tab ric_mac abo_majormis, missing
286 tab cond abo_majormis, missing
287 tab gvhd_pro abo_majormis, missing
288
289 *検定
290 ranksum pt_age, by(abo_majormis)
291 tab pt_sex abo_majormis, chi2
292 tab sex_mismatch2 abo_majormis, chi2
293 tab diagnosis2 abo_majormis, chi2
294 tab stage1 abo_majormis, chi2
295 tab sct_type abo_majormis, chi2
296 tab ric_mac abo_majormis, chi2
297 tab cond abo_majormis, chi2
298 tab gvhd_pro abo_majormis, chi2
299
300 log close
301 translate characteristics.smcl characteristics.txt
302
303 /*
304 少数のテーブルのみエクセルファイルに取り込みたいときは、
305 テーブルを最小限かつスペースが含まれないようにしっかりと選択して、
306 右クリック、copy tableを選択（ここがポイント）
307 エクセル上で張り付けると、綺麗にテーブルのまま
308 貼りつけることができます
309 */
310
311
312
313
314
315
316
```

```

1  ////*****11_gv_characteristics*****////
2  ////*****HLA scriptの概説*****////
3  /*HLA script*/
4  /*hla_low_pt_a1自体を変更したくないので、あらたに変数を作成してから作業を行う*/
5  gen serora1=hla_low_pt_a1
6  gen serora2=hla_low_pt_a2
7  gen serorb1=hla_low_pt_b1
8  gen serorb2=hla_low_pt_b2
9
10 /*2006年までの臍帯血移植の処理、片方のみ空欄なら“-”のマークを入れる*/
11 replace serora1="-" if serora1="" & serora2!="" & (source1==3 | source1==5) & (txdt_y>=1950
& txdt_y<=2006)
12 replace serora2="-" if serora2="" & serora1!="" & (source1==3 | source1==5) & (txdt_y>=1950
& txdt_y<=2006)
13
14 /*個別対応 一つの項目に二つの抗原を書き入れてしまったもの←手作業での確認が必要であり個別対応
とした*/
15 ***T20582-1
16 replace serorb1="44" if hla_low_pt_b1="B44,46" & hla_low_pt_b2="B4,6"
17 replace serorb2="46" if hla_low_pt_b1="B44,46" & hla_low_pt_b2="B4,6"
18
19 ***HLA-A
20 /*foreachというcommandを用いて、繰り返しの作業のスク립トをシンプルにする*/
21 /*`var'がそれぞれserora1 serora2 seroda1 seroda2の変数であった場合の
22 {}内の演算を行う*/
23 foreach var in serora1 serora2 seroda1 seroda2 {
24 replace `var'="NA" if `var'="未検査"
25 }
26
27 foreach var in serora1 serora2 seroda1 seroda2 {
28
29 /*文字列の前後にある空白を取る: trim(" this ") = "this" */
30 replace `var'=trim(`var')
31
32 /*全角を半角に←結構、全角文字での記入例があるためすべて半角に変換する*/
33 /*その際にsubstrというcommandを用いる*/
34 /*
35 substr("this is this","is","X",1) = "thX is this"
36 substr("this is this","is","X",2) = "thX X this"
37 substr("this is this","is","X",.) = "thX X thX" */
38
39 /*`var'の文字列で左から大文字のゼロがあれば半角の0に変換する、それをx回繰り返す、ただし"."の場
合はすべて置換されるまで繰り返す。*/
40 replace `var'=substr(`var',"0","0",.) /*←左側の0は全角、右側の0は半角*/
41 replace `var'=substr(`var',"1","1",.)
42 replace `var'=substr(`var',"2","2",.)
43 replace `var'=substr(`var',"3","2",.)
44 replace `var'=substr(`var',"4","4",.)
45 replace `var'=substr(`var',"5","5",.)
46 replace `var'=substr(`var',"6","6",.)
47 replace `var'=substr(`var',"7","7",.)
48 replace `var'=substr(`var',"8","8",.)
49 replace `var'=substr(`var',"9","9",.)
50
51 /*~A-2402, B-54などのAやBを取り除きたい~
52 不要な記号の削除 */
53 replace `var'=substr(`var',"A","",.)
54
55
56 /*~-2402, -54などの-を取り除きたい~
57 lengthは文字列の長さを示す length("ab") = 2
58 substrは、文字列の何番目から何個の文字を抜き出すことを
59 指令する
60 substr("abcdef",2,3) = "bcd"
61 substr("abcdef",3,1) = "c"
62 */
63 /*
64 文字の最初にある“-”の文字を消したい、ただし、ホモを示す“-”は残したい

```



```
129 gen seromisdr=serogvhmisdr if serogvhmisdr>=serohvgmisdr
130 replace seromisdr=serohvgmisdr if serohvgmisdr>=serogvhmisdr
131
132 gen seromis6abdr2=seromisa + seromisb + seromisdr
133 gen seromis8abcdr2=seromisa + seromisb + seromisc + seromisdr
134
135 tab seromis6abdr2
136 tab seromis8abcdr2
137
138 /*HLA matching end*/
139
140
141 ***HLA矛盾症例を除外
142 count if seromis6abdr!=. & genomis6abdr!=. & (seromis6abdr>genomis6abdr)
143 count if seromis8abcdr!=. & genomis8abcdr!=. & (seromis8abcdr>genomis8abcdr)
144
145
146
```