Int J Hematol (2016) 103:11–19
DOI 10.1007/s12185-015-1907-9

Scripts for TRUMP data analyses. Part II (HLA-related data): statistical analyses specific for hematopoietic stem cell transplantation

Junya Kanda[1]

Received: 1 October 2015 / Revised: 5 November 2015 / Accepted: 10 November 2015 / Published online: 20 November 2015

Abstract The Transplant Registry Unified Management Program (TRUMP) made it possible for members of the Japan Society for Hematopoietic Cell Transplantation (JSHCT) to analyze large sets of national registry data on autologous and allogeneic hematopoietic stem cell transplantation. However, as the processes used to collect transplantation information are complex and differed over time, the background of these processes should be understood when using TRUMP data. Previously, information on the HLA locus of patients and donors had been collected using a questionnaire-based free-description method, resulting in some input errors. To correct minor but significant errors and provide accurate HLA matching data, the use of a Stata or EZR/R script offered by the JSHCT is strongly recommended when analyzing HLA data in the TRUMP dataset. The HLA mismatch direction, mismatch counting method, and different impacts of HLA mismatches by stem cell source are other important factors in the analysis of HLA data. Additionally, researchers should understand the statistical analyses specific for hematopoietic stem cell transplantation, such as competing risk, landmark analysis, and time-dependent analysis, to correctly analyze transplant data. The data center of the JSHCT can be contacted if statistical assistance is required.

Keywords HLA mismatch · Mismatch direction · Competing risk · Landmark analysis · Time-dependent analysis

✉ Junya Kanda
jkandajp@gmail.com

[1] Division of Hematology, Saitama Medical Center, Jichi Medical University, 1-847 Amanuma-cho, Omiya-ku, Saitama City, Saitama 330-8503, Japan

Introduction

In this article, I have focused on treating and analyzing human leukocyte antigen (HLA) data collected using Transplant Registry Unified Management Program (TRUMP) and explained several factors regarding the statistical analyses specific for hematopoietic stem cell transplantation.

There are numerous important issues related to the analysis of TRUMP data. HLA information for each locus was collected through free description by physicians or data managers, resulting in errors in HLA matching counts, which needs correction. Mismatch direction should be considered when biological significance of HLA mismatch is analyzed. The counting method and impact of HLA mismatches differ according to the stem cell source. These differences should be considered in the analysis of HLA data obtained using TRUMP.

HLA information in TRUMP data

Information for the HLA locus of patients and donors was collected by free description using the pre-TRUMP and TRUMP version 1 questionnaire form. Automatic calculation of the number of HLA matching is based on whether the digits of HLA locus for the patients are the same as that in the donors. Therefore, the data is considered a mismatch in the absence of proper input of HLA information for either a patient or a donor. For example, donor “2402” and recipient “A2402” at the HLA-A locus are considered a mismatch. Further, the Japanese font of “2402” is considered different from the English font of “2402”. To correct these minor but significant errors and provide accurate HLA matching data, we developed an HLA script for the analysis of TRUMP dataset on the webpage of the Japan Society

Table 1 Examples of the number of mismatches in the GVH and HVG directions at the HLA-A locus

	Donor		Recipient		GVH MM	HVG MM	MM
(1) Examples of 4-digit mismatches at the HLA-A locus							
Case 1	A*24:02	A*24:02	A*24:02	A*24:02	0	0	0
Case 2	A*24:02	A*24:02	A*24:02	**A*24:20**	1	0	1
Case 3	A*24:02	**A*24:20**	A*24:02	A*24:02	0	1	1
Case 4	A*24:02	**A*24:07**	A*24:02	**A*24:20**	1	1	1
(2) Examples of 2-digit mismatches at the HLA-A locus							
Case 5	A*24:02	A*24:07	A*24:02	A*24:20	0	0	0
Case 6	A*24:02	A*24:07	A*24:02	**A*11:01**	1	0	1
Case 7	A*24:02	**A*33:03**	A*24:07	A*24:20	0	1	1
Case 8	A*24:02	**A*33:03**	A*24:07	**A*11:01**	1	1	1

HLA mismatched at 4- or 2-digit level is highlighted in boldface type

GVH MM mismatch in the graft-versus-host direction, *HVG MM* mismatch in the host-versus-graft direction

for Hematopoietic Cell Transplantation (JSHCT), which allows limited access to only JSHCT members (http://www.jshct.com/memdir/download/wg.shtml). Further, the HLA script replaces the original with the retyped HLA data for the HLA-A, -B, -C, and -DRB1 loci in unrelated bone marrow transplantation provided by the research group of Dr. Morishima [1]. HLA information provided from JCBBN is also considered in the HLA script. This makes the HLA data more accurate, particularly for the HLA data collected previously. If HLA 2-digit data (i.e., antigen data) for a locus are both blank but 4-digit information (i.e., allele data) are available, 2-digit data are replaced with 4-digit information. When one of the two 2-digit or 4-digit data at a locus are missing, there are two possibilities: one indicates a homologous locus and the other indicates missing data. In the HLA script, we consider these as missing data and excluded them from the analysis because we are unable to determine their status. By using the HLA script, accuracy of the number of HLA mismatches will be substantially improved. For example, the number of HLA-A 2-digit mismatches in the GVH direction before and after the use of HLA script is 0 in 38,203 and 39,660 patients, 1 in 5516 and 4919 patients, and 2 in 491 and 71 patients, respectively. It should be noted that there are several cases showing a contradiction between the 2-digit and 4-digit data. These should be managed in each study if necessary. In TRUMP version 2, which was started in 2015, HLA information must be selected from the pull-down menu, creating a risk of an improper input minimum.

HLA counting method

The number of HLA mismatches between patients and donors is typically counted as a total without considering the HLA mismatch direction. However, the effect of the immune reaction caused by HLA mismatch differs according to whether the mismatches are in the GVH or HVG direction. A mismatched locus in the GVH direction may be a major target for donor T cells and can cause GVHD, whereas a mismatched locus in the HVG direction may be a major target for the remaining recipient T cells and can lead to graft rejection. Therefore, from a biological perspective, the impact of HLA mismatch should be discussed separately according to the mismatch direction. The risk of GVHD should be evaluated with HLA mismatches in the GVH direction, whereas the risk of engraftment should be evaluated with those in the HVG direction. The risk of overall mortality may be evaluated with those in GVH and/or HVG direction, depending on the study objective.

Examples of HLA 2-digit (antigen) or 4-digit (allele) mismatches at the HLA-A locus are shown in Table 1. An HLA mismatch is considered in the GVH direction when the recipient's antigens or alleles (Case 2 & 4, A*24:20) are not shared with the donor (Case 2, A*24:02; Case 4, A*24:02, A*24:07). An HLA mismatch is considered in the HVG direction when the donor's antigens or alleles (Case 3, A*24:20; Case 4, A*24:07) are not shared with the recipient (Case 3, A*24:02; Case 4, A*24:02, A*24:20). The total number of HLA mismatches can be counted in two ways. If we focus on the number of mismatches in the GVH or HVG direction, the total number of mismatches in the GVH direction or the HVG direction should be counted and the larger number of mismatches for either the GVH or HVG direction should be selected (Table 2). However, if we focus on the number of mismatched loci, the number of mismatched loci should be counted regardless of the mismatch direction. In most cases, there is no discrepancy between these 2 counting methods (cases 9 and 10 in Table 2). However, as shown in case 11 in Table 2, a locus is mismatched only in the GVH direction and another locus is mismatched only in the HVG direction, and thus these numbers differ. The method should be chosen according to the study objective

Table 2 Examples of the number of mismatches at the HLA-A, -B, -C, and -DRB1 loci

Donor		Recipient		GVH MM	HVG MM	MM at each locus
Case 9						
A*24:02	A*24:02	A*24:02	**A*24:20**	1	0	1
B*52:01	B*52:01	B*52:01	B*52:01	0	0	0
C*12:01	C*12:01	C*12:01	C*12:01	0	0	0
DRB1*15:02	DRB1*15:02	DRB1*15:02	DRB1*15:02	0	0	0
Total				1	0	1
Case 10						
A*24:02	**A*24:20**	A*24:02	A*24:02	0	1	1
B*52:01	B*52:01	B*52:01	B*52:01	0	0	0
C*12:01	C*12:01	C*12:01	C*12:01	0	0	0
DRB1*15:01	DRB1*15:02	DRB1*15:02	DRB1*15:02	0	1	1
Total				0	2	2
Case 11						
A*24:02	A*24:02	A*24:02	**A*24:20**	1	0	1
B*52:01	B*52:01	B*52:01	B*52:01	0	0	0
C*12:01	C*12:01	C*12:01	C*12:01	0	0	0
DRB1*15:01	DRB1*15:02	DRB1*15:02	DRB1*15:02	0	1	1
Total				1	1	1 or 2

HLA mismatched at 4-digit level is highlighted in boldface type

GVH MM mismatch in the graft-versus-host direction, *HVG MM* mismatch in the host-versus-graft direction

and design. In the HLA script, we offer variables for each method such as, .HLA.Geno.mis8/genomis8abcdr versus . HLA.Geno.mis8.2/genomis8abcdr2 (Table 3). In case 11, the value of .HLA.Geno.mis8/genomis8abcdr is 1 and that of.HLA.Geno.mis8.2/genomis8abcdr2 is 2.

Impact of HLA mismatch according to stem cell sources

In related transplantation, the presence of an HLA antigen mismatch in the GVH direction was associated with a higher incidence of GVHD compared to an HLA mismatch in the HVG direction [2, 3]. In contrast, the presence of an HLA antigen mismatch in the HVG direction was associated with a higher incidence of graft failure than the HLA match [3, 4]. In a recent analyses of unrelated transplantation, one-allele mismatch (HLA-A, -B, -C, -DRB1) only in the GVH direction, but not one-allele mismatch only in the HVG direction, was associated with a higher incidence of grades III–IV acute GVHD compared with the HLA match. In contrast, allele mismatch in either the GVH or HVG direction was not associated with neutrophil engraftment [5, 6]. This difference between related and unrelated transplantation may be partly explained by more frequent 2-digit (antigen) mismatches in related transplantation and by the improvement in the conditioning regimen and GVHD prophylaxis. Recent studies revealed that a high titer of donor-specific HLA antibody was associated with graft failure, suggesting that multiple 2-digit mismatches in HLA mismatched related transplantation may increase the risk of graft failure, unless donor-specific HLA antibodies are examined [7–10]. Avoidance of a donor to whom the patients have a donor-specific HLA antibody would improve the rate of engraftment. These findings in both related and unrelated transplantation indicate the importance of HLA mismatch direction for interpreting clinical outcomes.

Difference in HLA matching between Western countries and Japan

In Japan, HLA matching is counted as 2-digit level in unrelated cord blood transplantation (UCBT), and up to 2 mismatches in this counting method are allowed for UCB unit selection [11]. In Europe and the U.S., HLA matching is generally counted as 2-digit level for HLA-A and HLA-B loci and as 4-digit level for the HLA-DRB1 locus. However, there is no robust evidence to support counting of the HLA-DRB1 locus on the allele level. We previously analyzed the difference between the impacts of 2-digit or 4-digit level mismatches in the HLA-DRB1 locus [11, 12]. However, we found no significant difference in impact between these mismatches. More importantly, the impact of HLA mismatch was very small or negligible in adult patients who received UCBT in Japan [12]. Although

Table 3 Variables related to HLA data

Variables	Variable name for EZR/R	Variable name for Stata	Value	Description
2-digit HLA MM				
GVH MM at HLA-A	.HLA.Sero.GVHmis.A	serogvhmisa	0–2	
HVG MM at HLA-A	.HLA.Sero.HVGmis.A	serohvgmisa	0–2	
GVH MM at HLA-B	.HLA.Sero.GVHmis.B	serogvhmisb	0–2	
HVG MM at HLA-B	.HLA.Sero.HVGmis.B	serohvgmisb	0–2	
GVH MM at HLA-C	.HLA.Sero.GVHmis.C	serogvhmisc	0–2	
HVG MM at HLA-C	.HLA.Sero.HVGmis.C	serohvgmisc	0–2	
GVH MM at HLA-DRB1	.HLA.Sero.GVHmis.DR	serogvhmisdr	0–2	
HVG MM at HLA-DRB1	.HLA.Sero.HVGmis.DR	serohvgmisdr	0–2	
GVH MM at HLA-ABDRB1	.HLA.Sero.GVHmis	serogvhmis6abdr	0–6	serogvhmisa + serogvhmisb + serogvhmisdr
HVG MM at HLA-ABDRB1	.HLA.Sero.HVGmis	serohvgmis6abdr	0–6	serohvgmisa + serohvgmisb + serohvgmisdr
GVH MM at HLA-ABCDRB1	.HLA.Sero.GVHmis8	serogvhmis8abcdr	0–8	serogvhmisa + serogvhmisb + serogvh-misc + serogvhmisdr
HVG MM at HLA-ABCDRB1	.HLA.Sero.HVGmis8	serohvgmis8abcdr	0–8	serohvgmisa + serohvgmisb + serohvg-misc + serohvgmisdr
MM at HLA-ABDRB1	.HLA.Sero.mis6	seromis6abdr	0–6	Larger number of serogvhmis6abdr or serohvg-mis6abdr
MM at HLA-ABCDRB1	.HLA.Sero.mis8	seromis8abcdr	0–8	Larger number of serogvhmis8abcdr or serohvg-mis8abcdr
MM at HLA-A	.HLA.Sero.mis.A	seromisa	0–2	Larger number of serogvhmisa or serohvgmisa
MM at HLA-B	.HLA.Sero.mis.B	seromisb	0–2	Larger number of serogvhmisb or serohvgmisb
MM at HLA-C	.HLA.Sero.mis.C	seromisc	0–2	Larger number of serogvhmisc or serohvgmisc
MM at HLA-DRB1	.HLA.Sero.mis.DR	seromisdr	0–2	Larger number of serogvhmisdr or serohvgmisdr
MM at HLA-ABDRB1	.HLA.Sero.mis6.2	seromis6abdr2	0–6	seromisa + seromisb + seromisdr
MM at HLA-ABCDRB1	.HLA.Sero.mis8.2	seromis8abcdr2	0–8	seromisa + seromisb + seromisc + seromisdr
4-digit HLA MM				
GVH MM at HLA-A	.HLA.Geno.GVHmis.A	genogvhmisa	0–2	
HVG MM at HLA-A	.HLA.Geno.HVGmis.A	genohvgmisa	0–2	
GVH MM at HLA-B	.HLA.Geno.GVHmis.B	genogvhmisb	0–2	
HVG MM at HLA-B	.HLA.Geno.HVGmis.B	genohvgmisb	0–2	
GVH MM at HLA-C	.HLA.Geno.GVHmis.C	genogvhmisc	0–2	
HVG MM at HLA-C	.HLA.Geno.HVGmis.C	genohvgmisc	0–2	
GVH MM at HLA-DRB1	.HLA.Geno.GVHmis.DRB1	genogvhmisdr	0–2	
HVG MM at HLA-DRB1	.HLA.Geno.HVGmis.DRB1	genohvgmisdr	0–2	
GVH MM at HLA-ABDRB1	.HLA.Geno.GVHmis	genogvhmis6abdr	0–6	genogvhmisa + genogvhmisb + genogvhmisdr
HVG MM at HLA-ABDRB1	.HLA.Geno.HVGmis	genohvgmis6abdr	0–6	genohvgmisa + genohvgmisb + genohvgmisdr
GVH MM at HLA-ABCDRB1	.HLA.Geno.GVHmis8	genogvhmis8abcdr	0–8	genogvhmisa + genogvhmisb + genogvh-misc + genogvhmisdr
HVG MM at HLA-ABCDRB1	.HLA.Geno.HVGmis8	genohvgmis8abcdr	0–8	genohvgmisa + genohvgmisb + genohvg-misc + genohvgmisdr
MM at HLA-ABDRB1	.HLA.Geno.mis6	genomis6abdr	0–6	Larger number of genogvhmis6abdr or genohvg-mis6abdr
MM at HLA-ABCDRB1	.HLA.Geno.mis8	genomis8abcdr	0–8	Larger number of genogvhmis8abcdr or genohvg-mis8abcdr
MM at HLA-A	.HLA.Geno.mis.A	genomisa	0–2	Larger number of genogvhmisa or genohvgmisa
MM at HLA-B	.HLA.Geno.mis.B	genomisb	0–2	Larger number of genogvhmisb or genohvgmisb
MM at HLA-C	.HLA.Geno.mis.C	genomisc	0–2	Larger number of genogvhmisc or genohvgmisc
MM at HLA-DRB1	.HLA.Geno.mis.DR	genomisdr	0–2	Larger number of genogvhmisdr or genohvgmisdr
MM at HLA-ABDRB1	.HLA.Geno.mis6.2	genomis6abdr2	0–6	genomisa + genomisb + genomisdr
MM at HLA-ABCDRB1	.HLA.Geno.mis8.2	genomis8abcdr2	0–8	genomisa + genomisb + genomisc + genomisdr

GVH MM mismatch in the graft-versus-host direction, *HVG MM* mismatch in the host-versus-graft direction

there was no significant difference between the impacts of HLA-DRB1 antigen and allele mismatch, it is important to determine which HLA matching methods researchers will use before they begin to analyze transplant outcomes in UCBT. To directly compare outcomes of UCBT between studies conducted in Europe, the U.S., and Japan, researchers may follow the counting method employed in Europe and the U.S. However, for clinical practice in Japan, the results are easily interpreted if Japanese counting method is used.

In most CIBMTR studies analyzing unrelated bone marrow or peripheral blood stem cell transplantation, the impact of HLA mismatches at the HLA-A, -B, -C, and -DRB1 loci was evaluated regardless of mismatch levels, including the 2-digit or 4-digit levels [13, 14]. However, the impact of HLA allele mismatches was evaluated among 2-digit level matched pairs for the HLA-A, -B, and -DR loci in Japan, following the standard donor selection process of the Japan Marrow Donor Program, as such a donor can be found for more than 90 % of the patients in Japan [11, 15]. Therefore, for the HLA-C mismatch, 80–90 % of HLA-C allele mismatches were at the antigen level in the study. To directly compare the impact of HLA allele mismatch between Japanese studies, it may be better to include only 2-digit level matched pairs for the HLA-A, -B, and -DR loci.

Statistical analyses specific for hematopoietic stem cell transplantation

Survival analysis is the most-frequently used analysis method in the field of hematopoietic stem cell transplantation as well as in other hematological and solid organ malignancies. However, since the incidence of transplant-related mortality is not negligible, specific consideration is needed to calculate the cumulative incidence of post-transplant events, such as a relapse incidence. Further, analysis of the effect of post-transplant events, such as GVHD, on subsequent transplant outcomes requires specialized statistical techniques and consideration. These statistical analyses have been also reviewed in other articles [16–18].

Time-to-event analysis

Time-to-event analysis or survival analysis treats the time from a certain time point until a target event is analyzed. In the time-to-event analysis in hematopoietic stem cell transplantation, the Kaplan–Meier method is used to estimate overall survival, disease-free survival, or progression-free survival rates. In this analysis, an event is defined as death for overall survival, death or relapse for disease-free survival, and death, relapse, or progression for progression-free survival (Tables 4, 5). Since the follow-up time for patients without an event is variable, these patients are treated as censored at the last follow-up. The log-rank test is used to evaluate the overall differences among different groups, and the Cox proportional hazards model is used for univariate and multivariate analyses.

Competing event

A competing event is defined as an event that does not concurrently occur with a target event (i.e., a mutually exclusive event). If relapse is defined as an event, death without relapse is defined as a competing event. In this situation, there are three possible conditions for a patient, including relapse, death without relapse, and alive without relapse. The sum of the incidence of relapse and non-relapse mortality (death without relapse) and probability of disease-free survival (alive without relapse) should be 100 %. If the Kaplan–Meier method is used to calculate the incidence of relapse, that is, a 1-Kaplan–Meier probability, the incidence of relapse is overestimated, as patients who die early after transplantation before relapse are censored and excluded from the patients at risk. This means that the sum of the incidence of relapse and non-relapse mortality and probability of disease-free survival will be greater than 100 %, which is incorrect. Therefore, cumulative incidences should be calculated using the cumulative incidence function to account for competing risks [19]. As shown in Tables 4 and 5, we defined variables for an event and a corresponding competing risk event for neutrophil and platelet engraftment, acute GVHD, chronic GVHD, relapse, and non-relapse mortality. The definition of competing risk and eligibility criteria for the analysis should be determined according to the study design. Gray's test is used to evaluate overall differences among cumulative incidence functions [20]. The Fine and Gray proportional hazards model is used for univariate and multivariate analyses [21]. Log-rank test and Cox proportional hazards model is also acceptable [22].

Landmark analysis and time-dependent covariate

When we analyze the effect of post-transplant events, such as GVHD, on transplant outcomes, we cannot predict whether GVHD will occur at the start of observation, such as at the time of transplantation. If the occurrence of GVHD is treated as a time-fixed variable, patients with GVHD are supposed to live at least for the day of GVHD occurrence, which is biased towards showing a survival advantage for

Table 4 Event and competing event used in the analysis of hematopoietic stem cell transplantation

Outcomes	Event	Competing event	Variable name for EZR/R	Variable name for Stata
Overall survival	Death	–	.OS	event_os
Disease-free survival	Death or relapse	–	.DFS	event_rfs
Progression-free survival	Death, relapse or progression	–	not provided	not provided
Neutrophil engraftment	Neutrophil counts > 500 for 3 days	Death without an engraftment	.CompRisk.Engraftment	event_neut500
Platelet engraftment (>20,000)	Platelet counts > 20,000 for 7 days without transfusion	Death without an engraftment	.CompRisk.EngraftmentPlt2	event_plt2
Platelet engraftment (>50,000)	Platelet counts > 50,000 for 7 days without transfusion	Death without an engraftment	.CompRisk.EngraftmentPlt5	event_plt5
Grades 2–4 aGVHD	Grades 2–4 aGVHD	Death or relapse without aGVHD	.CompRisk.AGVHD24	event_agvhd24
		Death without aGVHD	.CompRisk.AGVHD24_2	event_agvhd24_2
Grades 3–4 aGVHD	Grades 3–4 aGVHD	Death or relapse without aGVHD	.CompRisk.AGVHD34	event_agvhd34
		Death without aGVHD	.CompRisk.AGVHD34_2	event_agvhd34_2
cGVHD	cGVHD	Death or relapse without cGVHD	.CompRisk.CGVHD	event_cgvhd
		Death without cGVHD	.CompRisk.CGVHD_2	event_cgvhd_2
Extensive cGVHD	Extensive cGVHD	Death or relapse without cGVHD	.CompRisk.CGVHD. Extensive	event_excgvhd
		Death without cGVHD	.CompRisk.CGVHD.Extensive_2	event_excgvhd_2
Relapse	Relapse	Death without relapse	.CompRisk.Relapse	event_relapse
Non-relapse mortality	Death without relapse	Relapse	.CompRisk.Relapse	event_relapse

aGVHD acute graft-versus-host direction, *cGVHD* chronic graft-versus-host direction

patients with GVHD. In this situation, the starting time of observation should be changed to a specific post-transplant period (i.e., landmark analysis) or the occurrence of GVHD should be treated as a time-dependent covariate. In landmark analysis of acute GVHD, the occurrence of acute GVHD at a specific post-transplant point is regarded as a time-fixed variable. Acute GVHD occurs until day 60 after transplantation for 90 % of patients. If landmark day is set at day 60 after transplantation, patients who have or have not experienced acute GVHD at day 60 are categorized into the GVHD group or no GVHD group, respectively. Even if patients have acute GVHD more than 60 days after transplantation, they are considered under the no GVHD group and are not included in the GVHD group. Patients who have had a target event by day 60 should be excluded from analysis. The results may change according to the landmark day. If the landmark day is set to an earlier day, the number of patients with GVHD will decrease, while if it is set to a later day, the number of total patients analyzed will decrease. Landmark analysis may be performed by setting the landmark day to various days in order to test the robustness of the results.

In regression analysis, the variable that changes over time can be incorporated in the model treating this variable as a time-dependent covariate. In the case of GVHD, the variable is 0 from the time of transplantation until GVHD occurs, and becomes 1 after GVHD occurrence. An example of Stata script analyzing a time-dependent covariate is shown in Fig. 1. How to analyze a time-dependent covariate using EZR/R is shown in another paper [23].

Conclusions

Since the processes used to collect transplant information differ according to time and are complicated, the background of the process should be clearly understood when these data are used. Particularly, use of a script offered by JSHCT is strongly recommended for analyzing HLA data in the JSHCT dataset. Researchers should also understand the statistical analyses specific for hematopoietic stem cell transplantation to correctly analyze transplant data. Researchers can contact the data center of the JSHCT if statistical help is required.

Table 5 Variables frequently used in the analysis of hematopoietic stem cell transplantation

Variables	Explanation	Variable name for EZR/R	R-Value	Variable name for Stata	Stata-Value
Overall survival	Alive (censor)	.OS	0	event_os	0
	Death		1		1
Time variable for.OS/event_os		.Days.OS		lday	
				lyear	
Disease-free survival	Alive without relapse (censor)	.DFS	0	event_rfs	0
	Death or relapse		1		1
Time variable for.DFS/event_rfs		.Days.DFS		rfs_day	
Clinical/hematological relapse	Relapse	.CompRisk.Relapse	1	event_relapse	1
Non-relapse mortality	Death without relapse		2		2
	Alive without relapse (censor)		0		3
Time variable for .CompRisk. Relapse/event_relapse		.DaysDFS		ci_rel_day	
Neutrophil counts > 500 for 3 days	Neutrophil recovery	.CompRisk.Engraftment	1	event_neut500	1
	Death without recovery		2		2
	Alive without recovery (censor)		0		3
Time variable for .CompRisk. Engraftment/event_neut500		.DaysCompRisk.Engraftment		ci_neut500	
Platelet counts >20,000 for 7 days without platelet transfusion	Platelet recovery	.CompRisk.EngraftmentPlt2	1	event_plt2	1
	Death without recovery		2		2
	Alive without recovery (censor)		0		3
Time variable for.CompRisk. EngraftmentPlt2/eventevent_ plt2		.DaysEngraftmentPlt2		ci_plt2	
Platelet counts >50,000 for 7 days without platelet transfusion	Platelet recovery	.CompRisk.EngraftmentPlt5	1	event_plt5	1
	Death without recovery		2		2
	Alive without recovery (censor)		0		3
Time variable for.CompRisk. EngraftmentPlt5/eventevent_ plt5		.DaysEngraftmentPlt5		ci_plt5	
Grades 2–4 aGVHD	Grades 2-4 aGVHD	.CompRisk.AGVHD24	1	event_agvhd24	1
	Death without aGVHD		2		2
	Relapse without aGVHD		3		3
	Alive without aGVHD or relapse (censor)		0		4
	Grades 2-4 aGVHD	.CompRisk.AGVHD24_2	1	event_agvhd24_2	1
	Death without aGVHD		2		2
	Alive without aGVHD (censor)		0		3
Time variable for.CompRisk. AGVHD24/event_agvhd24		.DaysCompRisk.AGVHD24		ci_agvhd24	
Time variable for.CompRisk. AGVHD24_2/event_ agvhd24_2		.DaysCompRisk.AGVHD24_2		ci_agvhd24_2	
Grades 3-4 aGVHD	Grades 3-4 aGVHD	.CompRisk.AGVHD34	1	event_agvhd34	1
	Death without aGVHD		2		2
	Relapse without aGVHD		3		3
	Alive without aGVHD or relapse (censor)		0		4
	Grade3-4 aGVHD	.CompRisk.AGVHD34_2	1	event_agvhd34_2	1
	Death without aGVHD		2		2

Table 5 continued

Variables	Explanation	Variable name for EZR/R	R-Value	Variable name for Stata	Stata-Value
	Alive without aGVHD (censor)		0		3
Time variable for.CompRisk. AGVHD34/event_agvhd34		.DaysCompRisk.AGVHD34		ci_agvhd34	
Time variable for.CompRisk. AGVHD34_2/event_ agvhd34_2		.DaysCompRisk.AGVHD34_2		ci_agvhd34_2	
cGVHD	cGVHD	.CompRisk.CGVHD	1	event_cgvhd	1
	Death without cGVHD		2		2
	Relapse without cGVHD		3		3
	Alive without cGVHD (censor)		0		4
	cGVHD	.CompRisk.CGVHD_2	1	event_cgvhd_2	1
	Death without cGVHD		2		2
	Alive without cGVHD (censor)		0		3
Time variable for.CompRisk. CGVHD/event_cgvhd		.DaysCompRisk.CGVHD		ci_cgvhd	
				ci_cgvhd_year	
Time variable for.CompRisk. CGVHD_2/event_cgvhd_2		.DaysCompRisk.CGVHD_2		ci_cgvhd_2	
				ci_cgvhd_2_year	
Extensive cGVHD	Extensive cGVHD	.CompRisk.CGVHD.Extensive	1	event_excgvhd	1
	Death without cGVHD or relapse		2		2
	Relapse without cGVHD		3		3
	Alive without cGVHD or relapse (censor)		0		4
	Extensive cGVHD	.CompRisk.CGVHD.Exten sive_2	1	event_excgvhd_2	1
	Death without cGVHD		2		2
	Alive without cGVHD (censor)		0		3
Time variable for.CompRisk. CGVHD.Extensive/event_ excgvhd		.DaysCompRisk.CGVHD. Extensive		ci_excgvhd	
				ci_excgvhd_year	
Time variable for.CompRisk. CGVHD.Extensive_2/event_ excgvhd_2		.DaysCompRisk.CGVHD. Extensive_2		ci_excgvhd_2	
				ci_excgvhd_2_year	

Abbreviations: aGVHD, acute graft-versus-host direction; cGVHD, chronic graft-versus-host direction

```
sort id
stset lday, failure(event_os==1) id(id)

stsplit agvhd24_split1 if event_agvhd24==1, at(0) after(time=ci_agvhd24)
gen agvhd24_split2 = 0 if event_agvhd24>=2 & event_agvhd24<=4
replace agvhd24_split2 = 0 if agvhd24_split1==-1
replace agvhd24_split2 = 1 if agvhd24_split1==0

sort id
list id event_agvhd24 agvhd24_split2 _t0 _t in 1/20

sts graph, by(agvhd24_split2) risktable

stcox agvhd24_split2
stcox agvhd24_split2 pt_age pt_sex i.genogvhmis8abcdr
```

Fig. 1 Example of Stata script analyzing an impact of grades 2–4 acute GVHD on overall survival

Acknowledgments I thank Dr Yoshiko Atsuta (Japanese Data Center for Hematopoietic Cell Transplantation) and Dr Yoshinobu Kanda (Jichi Medical University) for making a great effort to develop the Stata and EZR/R scripts for TRUMP data analysis.

Compliance with ethical standards

Conflict of interest The author declares no competing financial interests.

References

1. Kawase T, Morishima Y, Matsuo K, Kashiwase K, Inoko H, Saji H, et al. High-risk HLA allele mismatch combinations

responsible for severe acute graft-versus-host disease and implication for its molecular mechanism. Blood. 2007;110:2235–41.
2. Anasetti C, Beatty PG, Storb R, Martin PJ, Mori M, Sanders JE, et al. Effect of HLA incompatibility on graft-versus-host disease, relapse, and survival after marrow transplantation for patients with leukemia or lymphoma. Hum Immunol. 1990;29:79–91.
3. Kanda Y, Chiba S, Hirai H, Sakamaki H, Iseki T, Kodera Y, et al. Allogeneic hematopoietic stem cell transplantation from family members other than HLA-identical siblings over the last decade (1991–2000). Blood. 2003;102:1541–7.
4. Anasetti C, Amos D, Beatty PG, Appelbaum FR, Bensinger W, Buckner CD, et al. Effect of HLA compatibility on engraftment of bone marrow transplants in patients with leukemia or lymphoma. N Engl J Med. 1989;320:197–204.
5. Kanda J, Ichinohe T, Fuji S, Maeda Y, Ohashi K, Fukuda T, et al. Impact of HLA mismatch direction on the outcome of unrelated bone marrow transplantation: a retrospective analysis from the Japan Society for Hematopoietic Cell Transplantation. Biol Blood Marrow Transplant. 2015;21:305–11.
6. Hurley CK, Woolfrey A, Wang T, Haagenson M, Umejiego J, Aljurf M, et al. The impact of HLA unidirectional mismatches on the outcome of myeloablative hematopoietic stem cell transplantation with unrelated donors. Blood. 2013;121:4800–6.
7. Yoshihara S, Maruya E, Taniguchi K, Kaida K, Kato R, Inoue T, et al. Risk and prevention of graft failure in patients with preexisting donor-specific HLA antibodies undergoing unmanipulated haploidentical SCT. Bone Marrow Transplant. 2012;47:508–15.
8. Ciurea SO, de Lima M, Cano P, Korbling M, Giralt S, Shpall EJ, et al. High risk of graft failure in patients with anti-HLA antibodies undergoing haploidentical stem-cell transplantation. Transplantation. 2009;88:1019–24.
9. Yoshihara S, Taniguchi K, Ogawa H, Saji H. The role of HLA antibodies in allogeneic SCT: is the 'type-and-screen' strategy necessary not only for blood type but also for HLA? Bone Marrow Transplant. 2012;47:1499–506.
10. Takanashi M, Atsuta Y, Fujiwara K, Kodo H, Kai S, Sato H, et al. The impact of anti-HLA antibodies on unrelated cord blood transplantations. Blood. 2010;116:2839–46.
11. Kanda J, Atsuta Y, Wake A, Ichinohe T, Takanashi M, Morishima Y, et al. Impact of the direction of HLA mismatch on transplantation outcomes in single unrelated cord blood transplantation. Biol Blood Marrow Transplant. 2013;19:247–54.
12. Atsuta Y, Kanda J, Takanashi M, Morishima Y, Taniguchi S, Takahashi S, et al. Different effects of HLA disparity on transplant outcomes after single-unit cord blood transplantation between pediatric and adult patients with leukemia. Haematologica. 2013;98:814–22.
13. Flomenberg N, Baxter-Lowe LA, Confer D, Fernandez-Vina M, Filipovich A, Horowitz M, et al. Impact of HLA class I and class II high-resolution matching on outcomes of unrelated donor bone marrow transplantation: HLA-C mismatching is associated with a strong adverse effect on transplantation outcome. Blood. 2004;104:1923–30.
14. Lee SJ, Klein J, Haagenson M, Baxter-Lowe LA, Confer DL, Eapen M, et al. High-resolution donor-recipient HLA matching contributes to the success of unrelated donor marrow transplantation. Blood. 2007;110:4576–83.
15. Morishima Y, Yabe T, Matsuo K, Kashiwase K, Inoko H, Saji H, et al. Effects of HLA allele and killer immunoglobulin-like receptor ligand matching on clinical outcome in leukemia patients undergoing transplantation with T-cell-replete marrow from an unrelated donor. Biol Blood Marrow Transplant. 2007;13:315–28.
16. Klein JP, Rizzo JD, Zhang MJ, Keiding N. Statistical methods for the analysis and presentation of the results of bone marrow transplants. Part I: unadjusted analysis. Bone Marrow Transplant. 2001;28:909–15.
17. Klein JP, Rizzo JD, Zhang MJ, Keiding N. Statistical methods for the analysis and presentation of the results of bone marrow transplants. Part 2: regression modeling. Bone Marrow Transplant. 2001;28:1001–11.
18. Iacobelli S. Suggestions on the use of statistical methodologies in studies of the European Group for Blood and Marrow Transplantation. Bone Marrow Transplant. 2013;48(Suppl 1):S1–37.
19. Gooley TA, Leisenring W, Crowley J, Storer BE. Estimation of failure probabilities in the presence of competing risks: new representations of old estimators. Stat Med. 1999;18:695–706.
20. Gray RJ. A class of k-sample tests for comparing the cumulative incidence of a competing risk. Ann Stat. 1988;16:1141–54.
21. Fine JP, Gray RJ. A proportional hazards model for subdistribution of a competing risk. J Am Stat Assoc. 1999;94:456–509.
22. Dignam JJ, Kocherginsky MN. Choice and interpretation of statistical tests used when competing risks are present. J Clin Oncol. 2008;26:4027–34.
23. Kanda Y. Investigation of the freely available easy-to-use software 'EZR' for medical statistics. Bone Marrow Transplant. 2013;48:452–8.