

Data Analysis on Post-HCT (Hematopoietic Cell Transplantation) Outcomes among Sickle Cell Disease Patients

BIS687: Data Science Capstone

Team Number: 2

Group Member:

Wanjing Chen

Yi Li

Bo Zhang

Scott (Sizhe) Zuo

Date: May.1.2024

Project Summary

This project aims to leverage the comprehensive research database established by the Center for International Blood and Marrow Transplant Research® (CIBMTR) for hematopoietic cell transplantation (HCT) in 2021 to identify features that affect survival performance regarding clinical outcomes of HCT for sickle cell disease (SCD) patients. The project is divided into three specific aims: an exploratory analysis of enrolled patients to understand demographic and clinical characteristics, a survival analysis focusing on the time from HCT to eight different endpoints, and the development of an R Shiny application for dynamic and interactive visualization of study results. This initiative seeks to enhance understanding of post-HCT outcomes, identify factors influencing survival post-transplant, and facilitate data accessibility for clinicians and researchers alike, ultimately contributing to improved patient care and outcomes in the field of cellular therapies.

Project Narrative

Sickle cell disease (SCD) is a group of inherited red blood cell disorders affecting millions of people throughout the world. In someone with Sickle Cell Disease (SCD), their red blood cells contain abnormal hemoglobin, causing them to become misshapen, resembling a sickle. These cells die prematurely, constantly decreasing the body's red blood cell count. They can also block blood flow in small vessels, triggering pain, and severe complications like infections, acute chest syndrome, and stroke. Despite a variety of strategies including supportive care, drug therapies, and red blood cell transfusions are able to potentially alleviate symptoms and extend lifespan of SGD patients, allogeneic hematopoietic cell transplantation (HCT) is the only established potential cure for SCD. Nevertheless, post- HCT SCD patients continue to face severe health challenges because HCT is associated with life-threatening complications most of which occur within the first 2 years after transplantation such as graft-versus-host disease (GVHD) and mortality. However, there is limited study on exploring features of post-HCT SGD patients that affect the outcomes for time to significant endpoints including last contact or death, graft failure, neutrophil engraftment, platelet recovery, acute and chronic graft-vs-host disease (GVHD), post-transplant lymphoproliferative disorder (PTLD), and second malignancy. Moreover, no standardized data analysis platforms have been established for clinical investigators to efficiently interact with and generate insight from the database of post-HCT SCD patients.

Specific Aims

A. Exploratory Data Analysis of Enrolled Patients

We performed exploratory analyses on variables included in the datasets to obtain an overview of the baseline characteristics of patients who received HCT and explore the features that are potentially associated with the survival outcomes of the patients. Specifically, the dataset categorizes the variables into 3 categories: disease-related, patient-related, and transplant-related. By summarizing information on variables in the dataset, we are able to learn the distribution of baseline variables that could facilitate further processing of the data, and provide initial insights on identifying important predictors contributing to predicting the clinical outcomes of patients.

B. Survival Analysis

We performed a survival analysis to assess the time from HCT to 8 endpoints, including last contact or death, graft failure, neutrophil engraftment, platelet recovery, acute and chronic graft-vs-host disease (GVHD), post-transplant lymphoproliferative disorder (PTLD), and second malignancy. These analyses help identify key predictors of survival outcomes which serve as potential factors that could enhance the quality of care for patients undergoing HCT treatment.

C. Shiny App Development for Results Visualization

Finally, an R Shiny application was developed to not only make findings interactive and accessible for clinical investigators but also serve as a standardized platform to generate results accordingly with a more updated database in the future. This interactive tool allows users to explore the

data dynamically, visualize survival curves, compare outcomes across different patient groups, and potentially identify new areas for research or intervention. The R Shiny application is considered as a crucial tool designed to visualize and interpret complex datasets to enhance the understanding and management of Sickle Cell Disease (SCD) post-hematopoietic cell transplantation (HCT).

Research Strategy

A. Significance

Recognizing the significance of HCT as the only established treatment of SGD and the post-HCT potential health obstacles, our team aims to explore crucial patient-related, disease-related, and transplant-related factors contributing to clinical outcomes after HCT among SGD patients by utilizing the database from Center for International Blood and Marrow Transplant Research (CIBMTR) in 2021. The project initiative is to specify the granularity for both evaluating success and subsequent health threats post-HCT and assist clinical professionals with the decision on establishing a more comprehensive treatment plan for SGD patients by taking key patient-related, disease-related, and transplant-related features into consideration.

B. Innovation

1) Analysis Innovation

For the survival analysis of this project, the targeted time-to-events include not only outcomes with implications of HCT failure such as “time to GVHD” and “time to PTLD(Post-transplant lymphoproliferative diseases)” but also implications of HCT success such as “time to neutrophil engraftment” and “time to platelet recovery”. In this way, we are able to have a more comprehensive understanding to assess the HCT performance for SGD patients.

2) Technology Innovation

In this project, we plan to create a standardized and interactive visualization platform to illustrate the analysis results via an R shiny web application. Besides the ease of exploring and engaging with research outcomes, we also recognize the advantage of an R shiny web application to serve as a standardized tool for assisting future researchers to expand upon this project by feeding more recent data to this platform and obtaining more updated results.

C. Data Pre-processing

After comparing the dataset from 2020 to 2021, our team has decided to focus on the dataset from 2021 due to its enhanced data completeness and inclusiveness of the dataset from 2020. To ensure the quality of analysis, columns for variables excluding outcomes with over 20% missing data were removed. In addition, variables with unchanging values were disregarded and rows with any missing data were excluded and resulted in the removal of most patients who received the transplant before 2008. In the end, we have the following numbers of variables in addition to patient ID (*see Appendix - Section A, Table 1-3*):

- a) 11 patient-related variables
- b) 1 disease-related variable
- c) 10 transplant-related variables
- d) 8 time-to-event variables/endpoints
- e) 14 categorical variables related to time-to-event variables

D. Specific Aim 1: Exploratory Analysis of Enrolled Patients

1) Hypothesis

The exploratory data analysis aims to characterize the distribution of baseline variables and to initially assess associations between the distribution of baseline characteristics and the occurrence of clinical endpoint events.

2) Rationale

Conducting exploratory analyses allows us to understand the structure and the pattern in the data, and thus facilitates the process of identification of potentially unbalanced variables and the decision whether to keep these variables in downstream analyses. By testing associations between outcome variables and response variables, we could gain initial insights into characteristics that might be clinically relevant in HCT patients.

3) Experimental Approach

In this aim, we primarily focused on categorical variables which entirely cover the information provided in our dataset. Specifically, for each categorical variable, we calculated the count and proportion of each value level. For continuous variables, we calculated the mean and standard deviation. We then examined the association between occurrence of clinical endpoints and explanatory variables by either Fisher's test or Pearson's Chi-squared test if the counts of all cells in the 2-way table were greater than 5 .

Summary tables were created by processing the data with R packages dplyr and tidyverse and presented with the package gt. When examining associations between explanatory variables and outcome variables, we implemented Pearson's Chi-squared test by the R function chisq.test and Fisher test by the R function fisher.test.

E. Specific Aim 2: Survival Analysis

1) Hypothesis

The time from hematopoietic cell transplantation (HCT) to various clinical endpoints (last contact or death, graft failure, neutrophil engraftment, platelet recovery, acute and chronic graft-vs-host disease (GVHD), post-transplant lymphoproliferative disorder (PTLD), and second malignancy) significantly differ among patient groups defined by specific demographic and clinical characteristics.

2) Rationale

Understanding the time to key post-HCT outcomes is crucial for several reasons. First of all it enables the prediction of patient prognosis by providing estimates on when significant post-transplant events might occur, thus informing both patients and clinicians about possible recovery paths and any potential complications that could arise. Besides, insights gained from analyzing the timing of events such as graft-versus-host disease (GVHD) are instrumental in guiding clinical decisions. Last but not least, by identifying the factors that influence the speed of recovery or the occurrence of delayed complications, adjustments can be made to the transplant process itself.

3) Experimental Approach

Cox Proportional Hazards (CoxPH) Model: To adjust for potential confounders and assess the impact of various predictors on the time to each endpoint, CoxPH models were fitted. A test of the proportional hazard assumption was conducted based on stratified survival plots. Variables included patient demographics, disease characteristics, transplant-related factors, and pre-transplant comorbidities.

Kaplan-Meier (KM) Survival Curves: For each of the eight endpoints, KM survival curves were generated to visualize the unadjusted probability of reaching each endpoint over time post-HCT.

F. Specific Aim 3: R Shiny App Development for Results Visualization

1) Hypothesis

This application serves as an interactive platform for clinicians, researchers, and policy-makers, enabling them to dynamically explore and analyze the wealth of patient-related, disease-related, and transplant-related variables and their impact on post-HCT outcomes.

2) Rationale

The rationale behind this application is twofold. Firstly, given the complexity and multi-dimensionality of the data, traditional static methods of data presentation are insufficient for capturing the nuanced relationships between variables such as age, ethnicity, disease genotype, transplant type, and conditioning regimen. An interactive tool will allow for a more comprehensive and tailored exploration of these variables, fostering a deeper understanding of their interplay and impact on patient outcomes. Secondly, this application aims to democratize data access, allowing users to generate custom analyses and visualizations that can inform clinical decision-making and policy development.

3) Experimental Approach

The app was implemented using R Shiny, leveraging its capabilities for creating interactive, web-based data visualizations. Key features include the ability to filter and stratify data based on specific variables like age group, disease genotype, or transplant type, and visualize these in the form of survival curves and bar charts. For instance, users could compare survival outcomes between different age groups or analyze the impact of donor-recipient HLA matching on post-transplant complications. This approach enables users to interact with the data, providing immediate visual feedback and insights.

Results & Interpretations

A. Specific Aim 1: Exploratory Analysis of Enrolled Patients

1) Categorical Patient-Related Explanatory Variables (*See Appendix - Section B, Table 4*)

Predominantly, over 90% of the patients are identified as non-Hispanic or non-Latino, and a similar proportion applies to African Americans, representing the racial distribution. These skewed demographics may influence disease outcomes and the generalizability of our results. Additionally, more than 80% of the SCD patients exhibited high Karnofsky/Lansky scores, indicating minor or no restrictions with strenuous physical activity at the time of HCT. This suggests that a large segment of our patient population was in relatively good physical condition prior to undergoing transplantation. However, concerning the HCT comorbidity index, approximately 35% of patients had a score higher than 3, which is associated with poorer HCT performance, indicating that a significant portion of the cohort had less favorable prognostic expectations for clinical outcomes post-transplant.

2) Categorical Transplant-Related Explanatory Variables (*See Appendix - Section B, Table 5*)

Approximately 60% of the patients received transplants from HLA identical siblings, suggesting a preference or greater accessibility for this donor type. Regarding the type of graft used, bone marrow was the predominant choice, used in 70% of the cases.

In terms of conditioning intensity, the patients were equally divided with half receiving myeloablative conditioning, which is more intense and aims at completely eradicating the patient's bone marrow before the transplant, and the other half split between reduced-intensity and non-myeloablative conditioning.

Furthermore, donor-recipient HLA matching shows a high level of compatibility, with over 70% of the patients achieving the full score of compatibility (0 to 8 mismatches), which is indicative of careful donor selection and matching processes.

3) Categorical Disease-Related Explanatory Variable (*See Appendix - Section B, Table 6*)

In this dataset, disease genotype is the only variable that fulfills this category. Significant imbalance is identified for this variable with hemoglobin SS taking up 96% and hemoglobin S-beta-thalassemia taking up 4% among the patients.

4) Continuous Patient-Related Explanatory Variable (*See Appendix - Section B, Table 7*)

In this dataset, age is the only variable that fulfills this category. Over 90% of the patients at transplant were less than 30 years old, which were relatively young. The average age at transplant is 14.52 with 10.26 as the standard deviation.

5) Categorical Outcome-related Variables (*See Appendix - Section B, Table 9*)

Among patients in this dataset, the overall death rate is 6.45% which is extremely low. Meanwhile, for subcategories within death which focus on death outcomes with reason exclusions, death

without neutrophil engraftment has the lowest event rate while death without graft-failure has the highest event rate. Furthermore, the results of the subcategories within death imply that there are overlaps for reason exclusions.

For the two positive outcomes including neutrophil engraftment and platelet recovery, both event rates are extremely high with at least 95 percent. Whereas for the 6 negative outcomes, the event rates are relatively low ranging from less than 1% to approximately 20 percent. Considering the high event rates for positive outcomes and low event rate for negative outcomes, the results are fairly consistent.

6) Continuous Outcome Variables (Time to Endpoints) (*See Appendix - Section B, Table 8*)

For the 8 target time to events, on average, the two positive events occur extremely early around 1 month after HCT. On the other hand, negative events occur around 3 years after HCT.

Note:

For specific distribution plots, please refer to (See Appendix - Section B, Plot 1-43)

B. Specific Aim 2: Survival Analysis

Table A: Features Selected by Cox Hazard Models with LASSO for 8 Endpoints (Y: Selected)

	DEAD	GF	PTLD	ANC	PLATELET	AGVHD	CGVHD	SCDMAL_FINAL
RCMVPR	Y		Y			Y		
SEX	Y	Y		Y			Y	
ETHNICIT	Y	Y	Y			Y	Y	Y
DONORF	Y	Y		Y	Y	Y		Y
GRAFTYPE	Y	Y		Y	Y	Y		Y
YEARTX	Y		Y					Y
AGE	Y					Y		Y
AGEGPFF								
KPS								
HCTCIGPF								
SUBDIS1F	Y		Y					Y
ATGF	Y	Y				Y		Y
YEARGPFF								
GVHD_FINAL	Y	Y	Y	Y	Y	Y	Y	Y
CONDGRPF	Y	Y	Y	Y		Y	Y	Y
CONDGRP_FINAL	Y	Y		Y	Y	Y		Y
HLA_FINAL								
FLAG_LANCET			Y	Y		Y		Y
FLAG_BLOOD	Y	Y						

Note: For variable definitions, please refer to (Appendix - Section A, Table 1-2).

Based on the summary table on selected features within each of the endpoints by the Cox hazard model with LASSO feature selection, transplant-related variables have the most significant role in terms of the survival performance regarding the clinical outcomes. Specifically, GVHD prophylaxis is identified as a crucial feature for all of the 8 endpoints; conditional intensity is selected by 7 out of 8 models;

conditioning regimen, donor type, and graft type are present as important features for 6 out of 8 models; if ATG or Alemtuzumab was given as conditioning regimen or GVHD prophylaxis is identified as an important feature in 4 out of 8 models; recipient CMV serostatus is selected by 3 out of 8 models. On the other hand, patient-related features are likely to have a moderate contribution to the Cox hazard models of the 8 endpoints. Specifically speaking, ethnicity is present in 6 out of 8 features and sex is selected as a crucial feature for 4 Cox hazard models. Furthermore, the disease genotype of a disease-related feature is selected by 3 out of the 8 times to outcomes, highlighting the contribution of a disease-related feature to clinical outcomes as well.

Intriguingly, opposite contributions of features on positive and negative outcomes are identified, which further implies consistency regarding the impact of features on clinical outcomes. For example, mismatched unrelated donor and cord blood as a specific donor type has a positive contribution with a coefficient estimation of 0.41 for time to chronic GVHD while it has a negative contribution with a coefficient estimation of -0.29 for time to platelet recovery. In addition, umbilical cord blood as a specific GVHD type has a negative contribution with a coefficient estimation of -0.56 for time to neutrophil engraftment while it has a positive contribution with a coefficient estimation of 0.2 for time to acute GVHD.

Note:

*For tables on coefficient estimation for selected features by 8 target endpoints, please refer to
(See Appendix - Section C, Table 10-17)*

*For Survival plots, please refer to the “Survival Plots” tab of the R shiny application addressed
in aim 3.*

C. Specific Aim 3: R Shiny App Development for Results Visualization

The Shiny app (link: <https://github.com/annieliyi/bis687.2024/tree/master/aim3/app.R>) was developed to serve as a dynamic exploration tool, seamlessly integrating the insights garnered from the first two aims. It enables users to interactively visualize and interrogate the rich dataset, drawing connections between patient demographics, transplant characteristics, and survival outcomes. Through the app, clinicians and researchers can manipulate the variables, filter the cohort, and instantly observe the effects on survival curves and other pertinent visualizations. Beyond visualizing results, the app is a springboard for hypothesis generation. It provides a sandbox for researchers to test conjectures about variable interplay and to discover novel patterns within the data. This empowers users to formulate new research questions and hypotheses directly from the interactive summaries and plots, potentially leading to breakthroughs in patient management and treatment strategies post-transplant.

Below is a succinct summary of the app functionality:

1) Tab - Variable Distribution Plots

This tab serves as a primary gateway for visual data interaction, where users can select a dataset and a variable to view its distribution. The resulting plots provide immediate visual feedback, facilitating a deeper understanding of individual variable characteristics within the patient cohort.

2) Tab - Variable Descriptions

A detailed tabular view lists all variables along with their descriptions, serving as a reference point to assist users in interpreting the variables used in analyses throughout the app.

3) Tab - Outcome and Categorical Variables Summaries

These tabs present comprehensive summaries of patient and transplant-related variables. They showcase the distribution of categorical factors such as race and donor type, as well as outcome variables, providing a snapshot of the cohort's demographic and clinical profile.

4) Tab - Event Coefficients & Event Features Summary

Offering a deeper dive into survival analysis, these tabs display the results of Cox proportional hazards modeling. Users can explore the influence of various features on survival outcomes, identifying which factors bear the most significant impact.

5) Tab - Survival Plots

A pivotal tab for the application, it provides stratified Kaplan-Meier survival plots that adhere to the proportional hazards assumption. As illustrated in the accompanying screenshot, users can select events and features to produce stratified survival curves, such as those comparing different graft types (*See Appendix - Section D, Illustration A*). This visualization aids in understanding how specific treatments and patient conditions influence survival probability over time.

For a hands-on exploration of additional survival plots and to understand how different variables interact to affect patient outcomes post-HCT, users are encouraged to navigate to the last tab of the app. Instructions within the app will guide users on utilizing inputs effectively to generate customized visual analyses tailored to their research questions.

Conclusion

In our study on post-hematopoietic cell transplantation (HCT) outcomes for sickle cell disease (SCD) patients, we conducted a comprehensive three-aim analysis that provided insightful conclusions. Aim 1 involved an exploratory analysis, revealing that while the distribution of positive outcomes like platelet recovery and neutrophil engraftment was high (>95%), negative outcomes such as graft failure occurred at a much lower rate (5%-20%). Notably, positive events typically occurred within one-month post-transplant, whereas negative events surfaced around three years later. The majority of patients at the time of transplant were younger than 30 years, with an average age of 14.52 years (SD: 10.26), indicating a relatively young cohort at high risk.

Aim 2 focused on survival analysis, identifying donor type, graft type, GVHD prophylaxis, conditioning regimen, and conditioning intensity as crucial variables. These factors were consistently highlighted in models for at least six outcomes based on LASSO feature selection, corroborating our findings from Fisher and Chi-squared tests in Aim 1. This alignment emphasizes the significant role of these variables and therefore transplant-related variables in influencing post-HCT outcomes.

Lastly, Aim 3 assessed tool standardization, where we utilized an R Shiny web application. This tool proved to be an effective means for users to interact with and visualize data, thereby enhancing the

understanding and generation of insights regarding post-HCT outcomes for SCD patients. This comprehensive approach not only advances our understanding of the factors that influence post-transplant success but also provides a robust framework for future research and clinical application in this field.

Limitations

In our study, several limitations were identified that may affect the interpretation of the findings. First, data from before 2008 were irretrievably lost, and notably, this period includes over 300 patients who underwent hematopoietic cell transplantation (HCT), creating a gap in the historical data. Meanwhile, data imbalance is identified for all outcome variables and most of the explanatory variables. Additionally, the proportional hazard assumptions underpinning the Cox hazard model may over-rely on stratified survival plots, which were informed by features selected via the LASSO method. Hence, we suggest potential improvements like employing log-negative log survival plots and Schoenfeld residuals to further check the proportional hazard assumption. Furthermore, the format of our data visualization tool was identified as a limitation; it lacks detailed, user-friendly elements such as clear definitions of variables and a comprehensive user guide, which could hinder the accessibility and usability of the data presented.

Future Work

In the future, we aim to address the identified limitations of our current study and expand the scope of our research. A key focus will be the integration of advanced machine learning models to enhance our analysis and predictive capabilities. Specifically, we plan to incorporate survival random forest models, which can relax the stringent assumptions required by the Cox model. Additionally, we intend to explore the use of neural networks for prediction of outcomes, capitalizing on their ability to model complex non-linear relationships.

To facilitate the application of our project by other researchers and ensure its relevance with evolving data landscapes, we will develop a comprehensive guide. This guide will instruct users on how to effectively employ our methodologies and tools on the most updated datasets. By providing clear protocols and support, we hope to make our project a valuable resource for the broader research community, fostering further innovation and discovery in the field.

Appendix

Section A:

Table 1: List of Predictors

Category	Variable.name	Description
Disease-related	subdis1f	Disease genotype
Patient-related	Dummyid	Unique patient identifier
Patient-related	flag_lancet	Cases from 2019 Lancet Haematology publication
Patient-related	flag_blood	Cases from 2016 Blood publication
Patient-related	flag_0601	Cases from BMT CTN 0601
Patient-related	age	Patient age at transplant, years
Patient-related	agegpff	Patient age at transplant, years
Patient-related	sex	Sex
Patient-related	ethnicit	Ethnicity
Patient-related	kps	Karnofsky/Lansky score at HCT
Patient-related	hctcigpf	HCT-comorbidity index
Transplant-related	donorf	Donor type
Transplant-related	grafttype	Graft type
Transplant-related	condgrpf	Conditioning intensity
Transplant-related	condgrp_final	Conditioning regimen
Transplant-related	atgf	ATG/Alemtuzumab given as conditioning regimen/GVHD prophylaxis
Transplant-related	gvhd_final	GVHD prophylaxis
Transplant-related	hla_final	Donor-recipient HLA matching
Transplant-related	rcmvr	Recipient CMV serostatus
Transplant-related	yeargpf	Year of transplant
Transplant-related	yeartx	Year of transplant

Table 2: List of Time-to-Event Variables

Category	Variable.name	Description
Outcomes	intxsurv	Time from HCT to date of last contact or death, months
Outcomes	intxgf	Time from HCT to graft failure, months
Outcomes	intxanc	Time from HCT to neutrophil engraftment, months
Outcomes	intxplatelet	Time from HCT to platelet recovery, months
Outcomes	intxagvh	Time from HCT to acute graft-vs-host disease, months
Outcomes	intxcgvhd	Time from HCT to chronic graft-vs-host disease, months
Outcomes	intxptld	Time from HCT to PTLD, months
Outcomes	intxscdmal	Time from HCT to second malignancy, months

Table 3: List of Categorical Outcome Variables

Category	Variable.name	Description
Outcomes	dead	Survival status at last contact
Outcomes	efs	Event-free survival (Graft failure or death are the events)
Outcomes	gf	Graft failure
Outcomes	dwogf	Death without graft failure
Outcomes	anc	Neutrophil engraftment
Outcomes	dwoanc	Death without neutrophil engraftment
Outcomes	platelet	Platelet recovery
Outcomes	dwoplatelet	Death without platelet recovery
Outcomes	agvhd	Acute graft versus host disease, grades II-IV
Outcomes	dwoagvhd	Death without acute graft versus host disease, grades II-IV
Outcomes	cgvhd	Chronic graft-vs-host disease
Outcomes	dwocgvhd	Death without chronic graft-vs-host disease
Outcomes	ptld	Post-transplant lymphoproliferative disorder (PTLD)
Outcomes	scdmal_final	Secondary malignancy

Section B:

Table 4:Patient-Related Categorical Variables Summary Table 5:Transplant-Related Categorical Variable Summary

Patient-related : Categorical Variable Summary	
	value
Cases from 2019 Lancet Haematology publication	
Yes	747 (65.07%)
No	401 (34.93%)
Cases from 2016 Blood publication	
No	1036 (90.24%)
Yes	112 (9.76%)
Cases from BMT CTN 0601	
No	1114 (97.04%)
Yes	34 (2.96%)
Age Categories (Years)	
<=10	472 (41.11%)
11-17	335 (29.18%)
18-29	236 (20.56%)
30-49	94 (8.19%)
>=50	11 (0.96%)
Sex	
Female	528 (45.99%)
Male	620 (54.01%)
Ethnicity	
Non-Hispanic or non-Latino	1042 (90.77%)
Hispanic or Latino	79 (6.88%)
Non-resident of the U.S.	27 (2.35%)
Race	
African-American	1041 (90.68%)
Asian	6 (0.52%)
Caucasian	89 (7.75%)
Others	12 (1.05%)
Karnofsky/Lansky score at HCT	
>=90	945 (82.32%)
<90	203 (17.68%)
HCT-comorbidity index	
0-2	747 (65.07%)
3+	401 (34.93%)

Transplant-related : Categorical Variable Summary	
	value
Donor type	
HLA identical sibling	678 (59.06%)
Mismatched unrelated donor and cord blood	104 (9.06%)
HLA mismatch relative	234 (20.38%)
Matched unrelated donor	132 (11.5%)
Graft type	
Bone marrow	798 (69.51%)
Umbilical cord blood	111 (9.67%)
Peripheral blood	239 (20.82%)
Conditioning intensity	
Myeloablative	541 (47.13%)
Reduced-intensity conditioning	307 (26.49%)
Non-myeloablative	300 (26.39%)
Conditioning regimen	
Flu/Bu	210 (18.29%)
Bu/Cy	300 (26.3%)
Flu/Mel	219 (19.08%)
Flu/Mel/TT	95 (8.28%)
TBI/Cy/Flu	85 (7.4%)
TBI/Flu	2 (0.17%)
TBI alone (300/400cGy)	106 (9.23%)
TBI/Mel	5 (0.44%)
TBI/Cy	18 (1.57%)
Flu/Bu/TT	19 (1.66%)
TBI/Cy/Flu/TT	85 (7.4%)
Cy/Flu	3 (0.26%)
Cy alone	1 (0.09%)
ATG/Alemtuzumab given as conditioning regimen/GVHD prophylaxis	
Alemtuzumab	570 (49.55%)
ATG	521 (45.38%)
None	57 (4.97%)
GVHD prophylaxis	
CNI + MTX	472 (41.11%)
CNI + MMF	235 (20.47%)
CNI alone	48 (4.18%)
CD 34 selection	46 (4.01%)
Post-CY + MMF + CNI	52 (4.53%)
Ex-vivo T-cell depletion	14 (1.22%)
MMF + MTX	1 (0.09%)
Post-CY + siro +/- MMF	168 (14.63%)
Siro alone	99 (8.62%)
MTX alone	5 (0.44%)
CNI + siro	2 (0.17%)
MMF alone	3 (0.26%)
MTX + siro	2 (0.17%)
MMF + siro	1 (0.09%)
Donor-recipient HLA matching	
8/8	810 (70.56%)
7/8	54 (4.7%)
<=6/8	284 (24.74%)
Recipient CMV serostatus	
Negative	569 (49.56%)
Positive	579 (50.44%)
Year of Transplant Categories (Year)	
2008-2012	273 (23.78%)
< 2008	1 (0.09%)
2013-2017	527 (45.91%)
2018-2020	347 (30.23%)
Year of Transplant (Year)	
2008	37 (3.22%)
2009	57 (4.97%)
2007	1 (0.09%)
2013	74 (6.45%)
2012	74 (6.45%)
2011	63 (5.49%)
2014	108 (9.41%)
2010	42 (3.66%)
2016	126 (10.98%)
2015	100 (8.71%)
2017	119 (10.37%)
2019	126 (10.98%)
2018	153 (13.33%)
2020	68 (5.92%)

Table 6: Disease-Related Categorical Variable Summary

Disease-related : Categorical Variable Summary	
	value
Disease genotype	
Hemoglobin SS	1105 (96.25%)
Hemoglobin S β -thalassemia	43 (3.75%)

Table 7: Patient-Related Continuous Variable Summary

Patient-related : Continuous Variable Summary	
	value
Age (Years)	14.52 (10.26)

Table 8: Outcome-Related Continuous Variables Summary

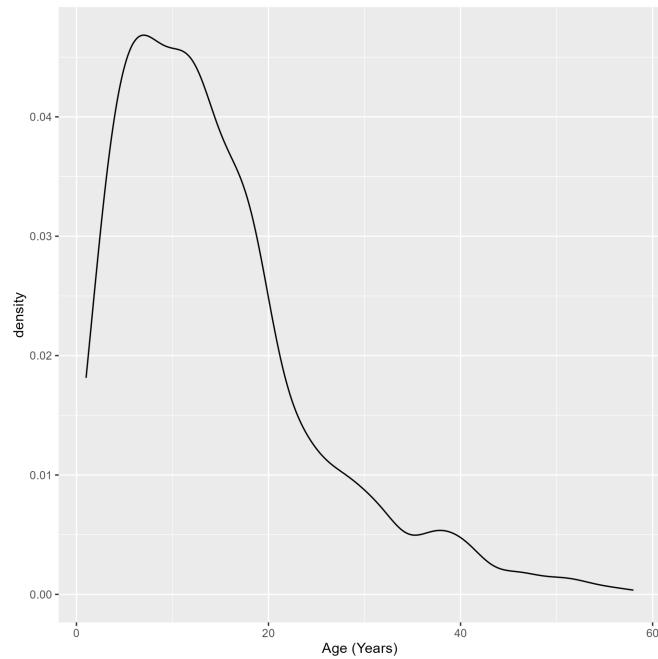
Outcomes : Continuous Variable Summary

	value
Time from HCT to date of last contact or death, months	44.7 (32.88)
Time from HCT to graft failure, months	37.75 (31.88)
Time from HCT to neutrophil engraftment, months	0.62 (0.55)
Time from HCT to platelet recovery, months	1.25 (4.96)
Time from HCT to acute graft-vs-host disease, months	37.25 (33.91)
Time from HCT to chronic graft-vs-host disease, months	36.61 (32.97)
Time from HCT to PTLD, months	44.3 (32.99)
Time from HCT to second malignancy, months	44.48 (32.8)

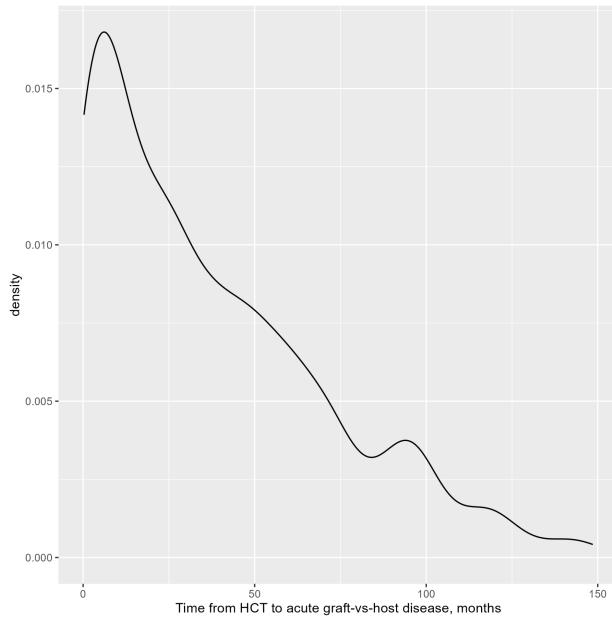
Table 9: Outcome-Related Categorical Variables Summary

Outcomes : Categorical Variable Summary	
	value
Survival status at last contact	
Alive	1074 (93.55%)
Dead	74 (6.45%)
Event-free survival (Graft failure or death are the events)	
No event	889 (77.44%)
Event happened	259 (22.56%)
Graft failure	
No	948 (82.58%)
Yes	200 (17.42%)
Death without graft failure	
No	1089 (94.86%)
Yes	59 (5.14%)
Neutrophil engraftment	
Yes	1134 (98.78%)
No	14 (1.22%)
Death without neutrophil engraftment	
No	1147 (99.91%)
Yes	1 (0.09%)
Platelet recovery	
Yes	1102 (95.99%)
No	46 (4.01%)
Death without platelet recovery	
No	1134 (98.78%)
Yes	14 (1.22%)
Acute graft versus host disease, grades II-IV	
No	985 (85.8%)
Yes	163 (14.2%)
Death without acute graft versus host disease, grades II-IV	
No	1114 (97.04%)
Yes	34 (2.96%)
Chronic graft-vs-host disease	
No	902 (78.57%)
Yes	246 (21.43%)
Death without chronic graft-vs-host disease	
No	1106 (96.34%)
Yes	42 (3.66%)
Post-transplant lymphoproliferative disorder (PTLD)	
No	1136 (98.95%)
Yes	12 (1.05%)
Secondary malignancy	
None	1138 (99.13%)
Myofibroblastic tumor	1 (0.09%)
Clonal cytogenetic abnormality	1 (0.09%)
TP53 mutation	1 (0.09%)
Acute myelogenous leukemia	3 (0.26%)
Acute lymphoblastic leukemia	2 (0.17%)
Sarcoma	1 (0.09%)
Kaposi sarcoma	1 (0.09%)

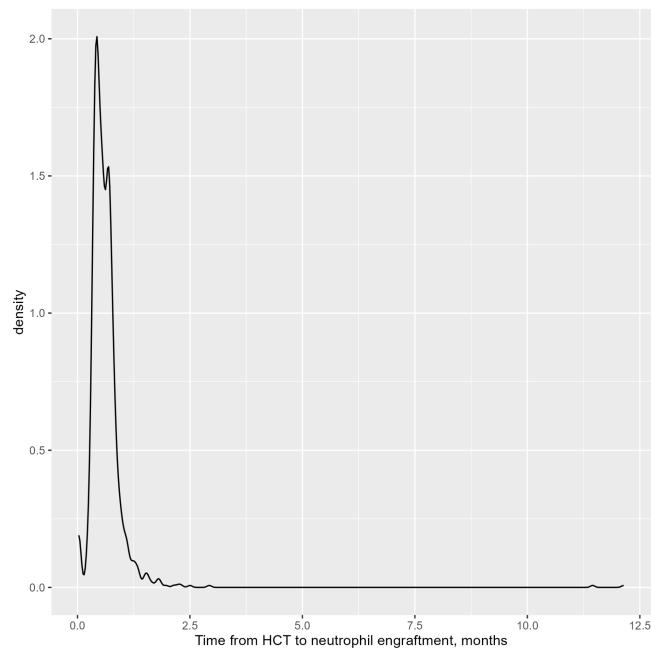
Plot 1: Distribution of Age



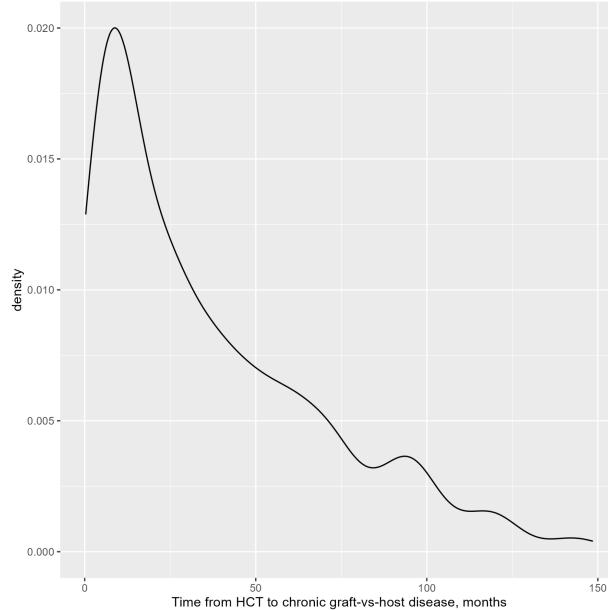
Plot 2: Distribution of Time from HCT to Acute GVHD



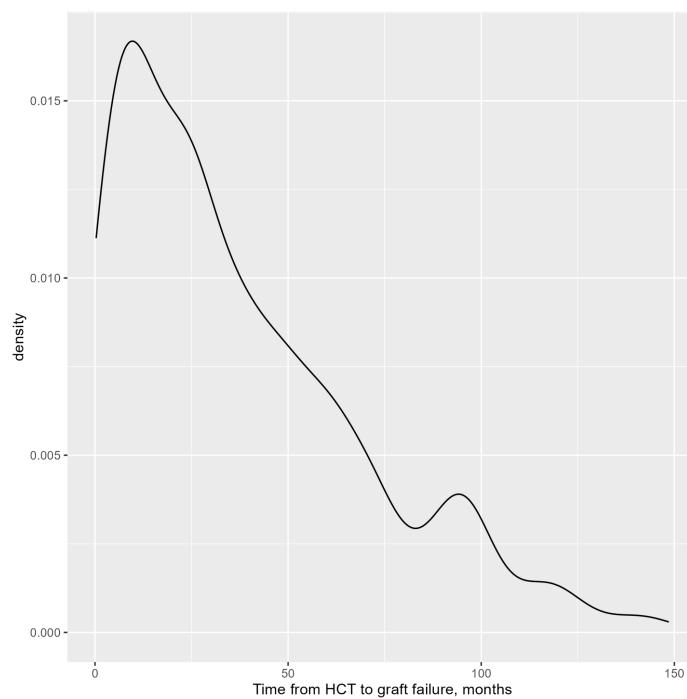
Plot 3: Distribution of Time from HCT to Neutrophil Engraftment



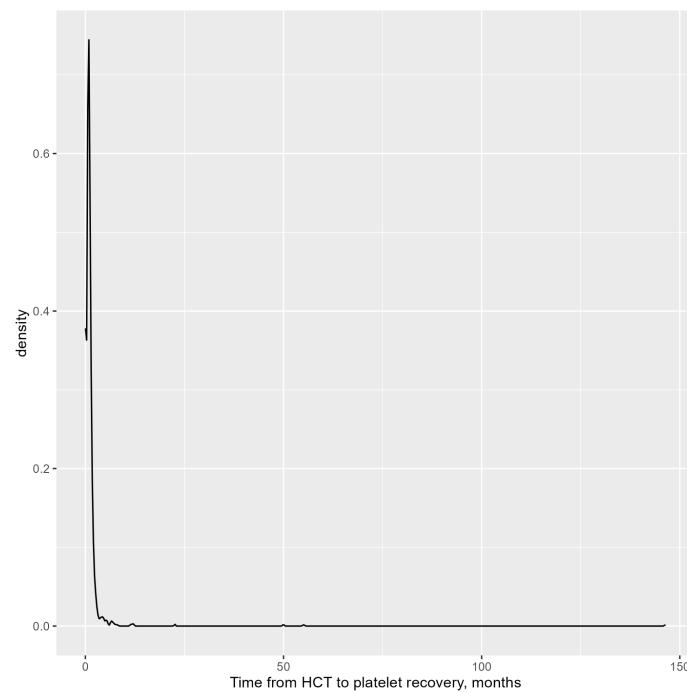
Plot 4: Distribution of Time from HCT to Chronic GVHD



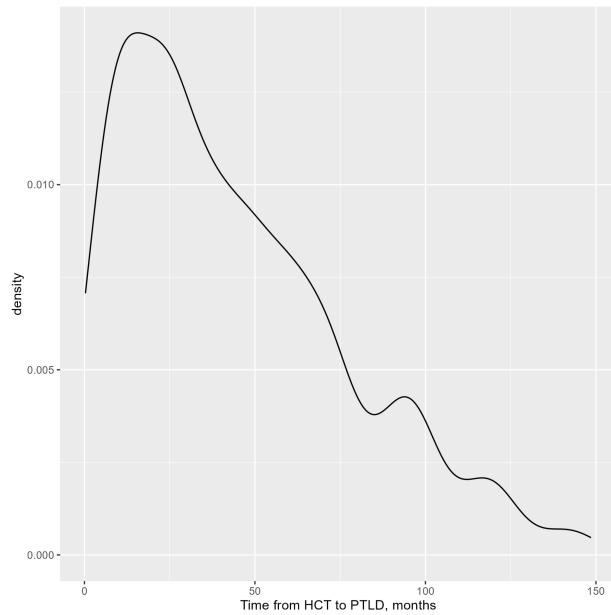
Plot 5: Distribution of Time from HCT to Graft Failure



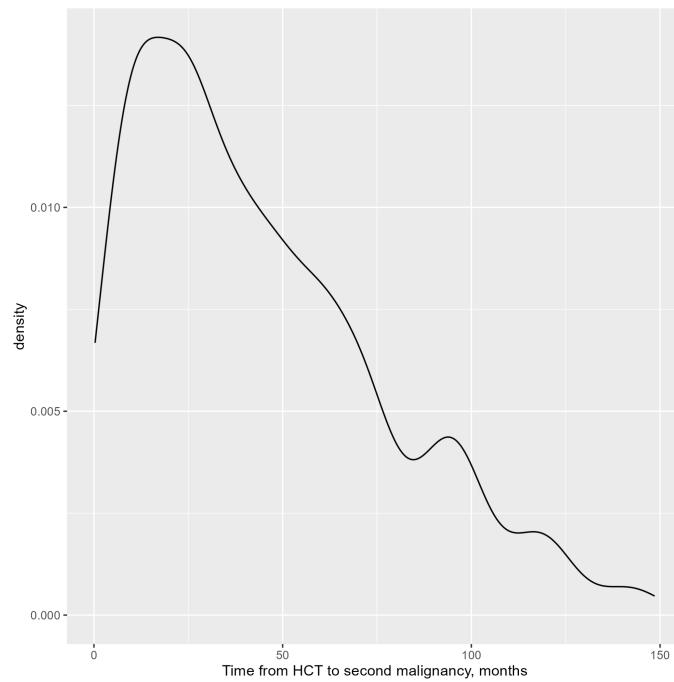
Plot 6: Distribution of time from HCT to Platelet Recovery



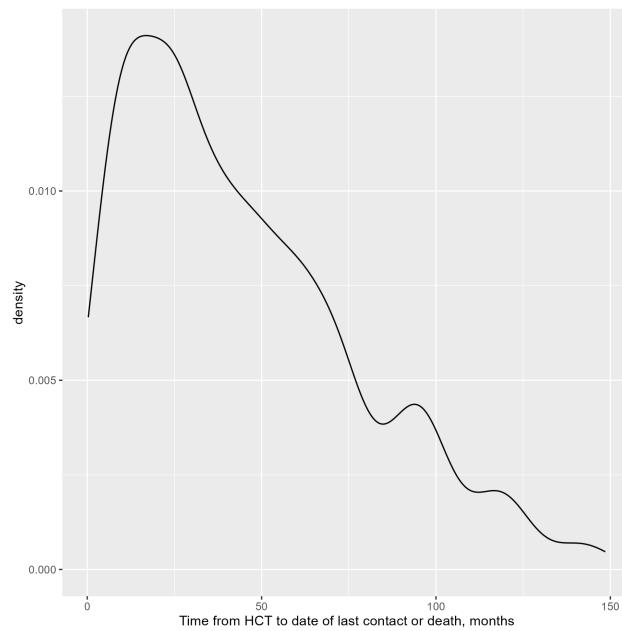
Plot 7: Distribution of Time from HCT to PTLD



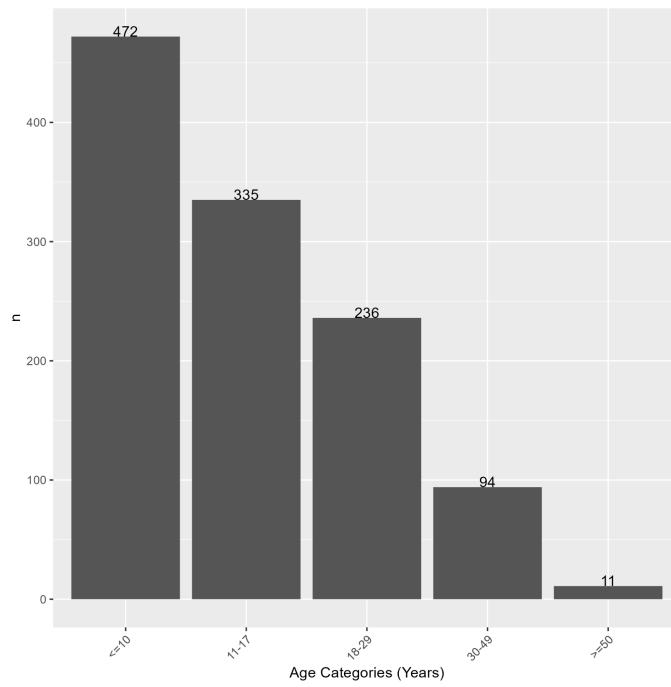
Plot 8: Distribution of Time from HCT to Second Malignancy



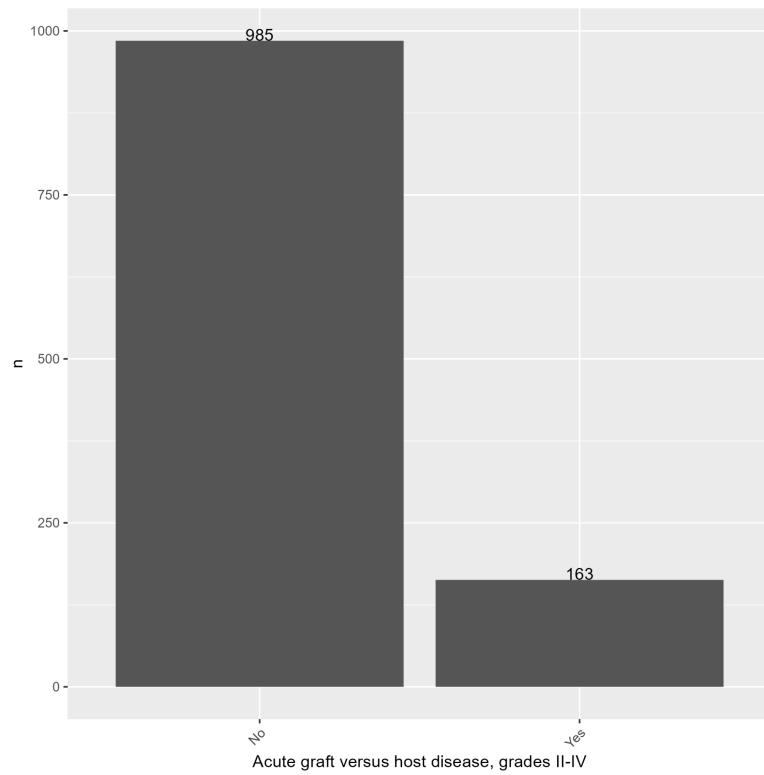
Plot 9: Distribution of Time from HCT to Date of Last Contact or Death



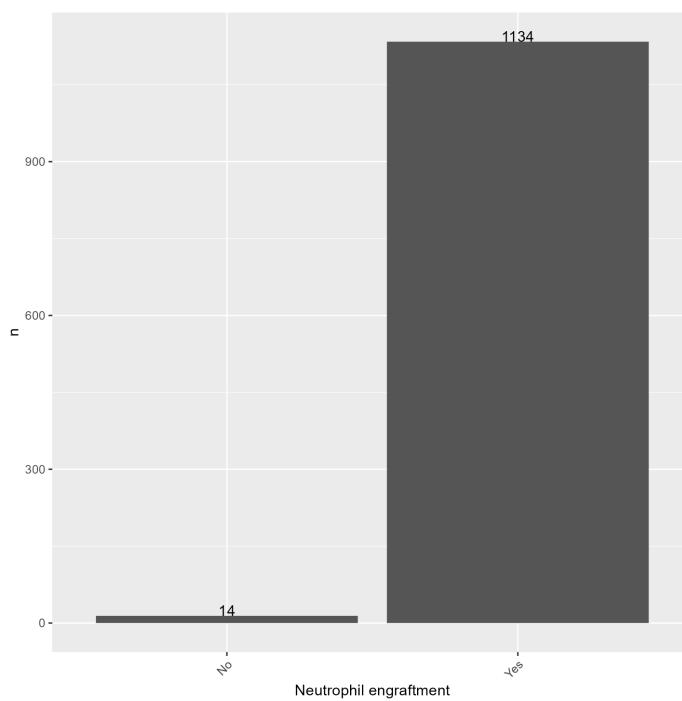
Plot 10: Distribution of Age Categories



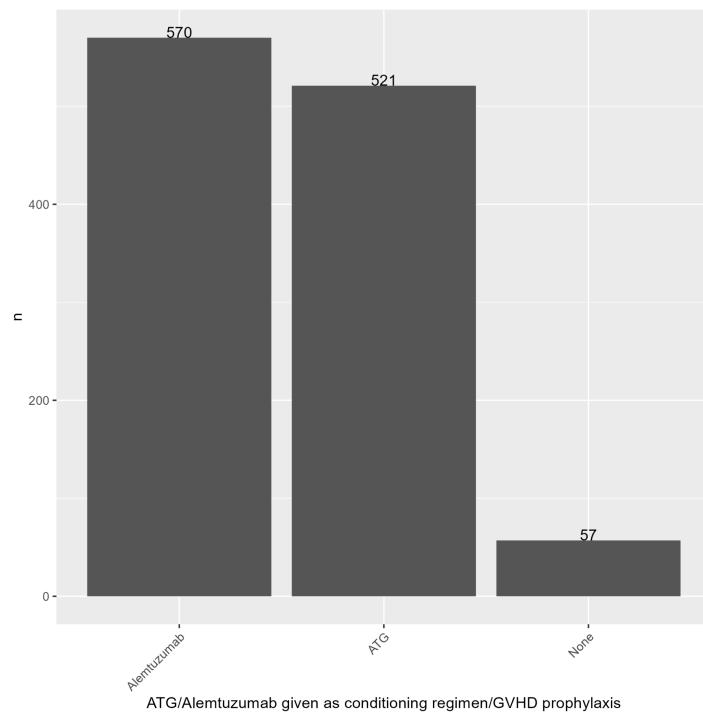
Plot 11: Distribution of Acute GVHD Event



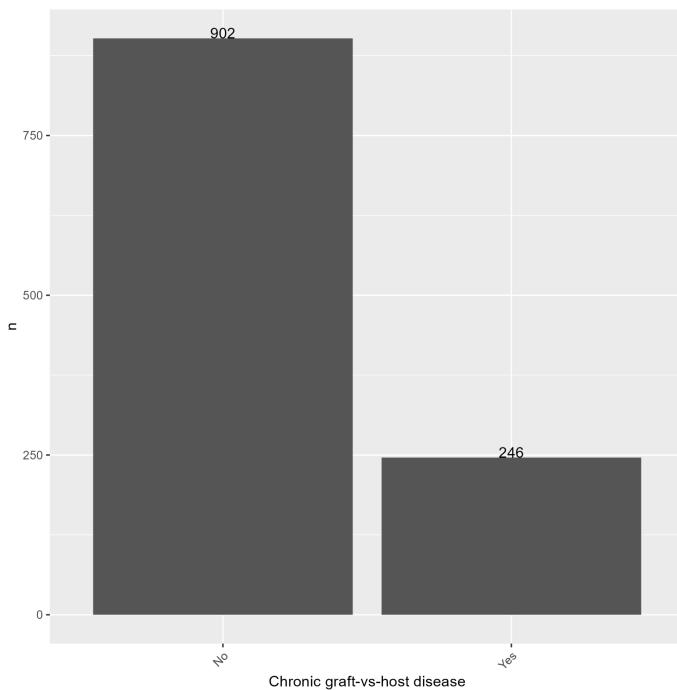
Plot 12: Distribution of Neutrophil Engraftment Event



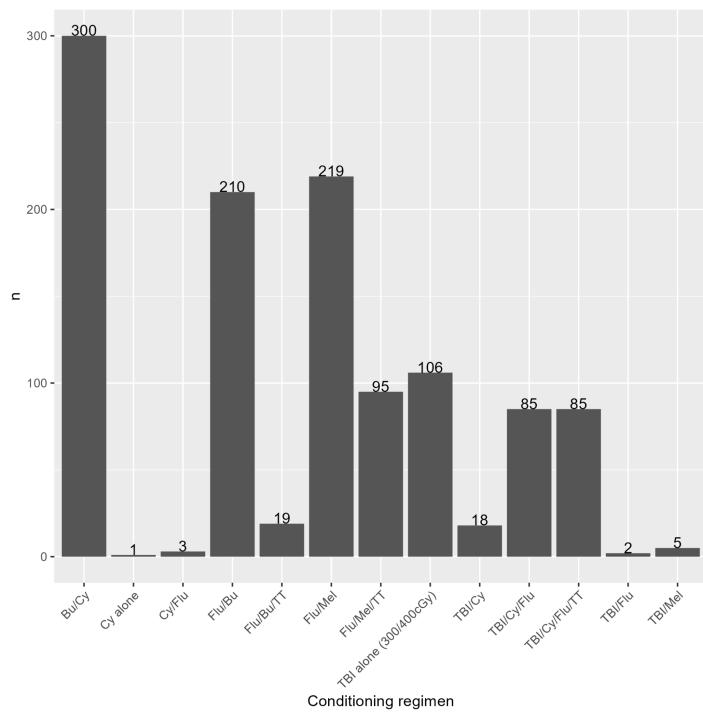
Plot 13: Distribution of ATG/Alemtuzumab Given as Conditioning Regimen/GVHD Prophylaxis



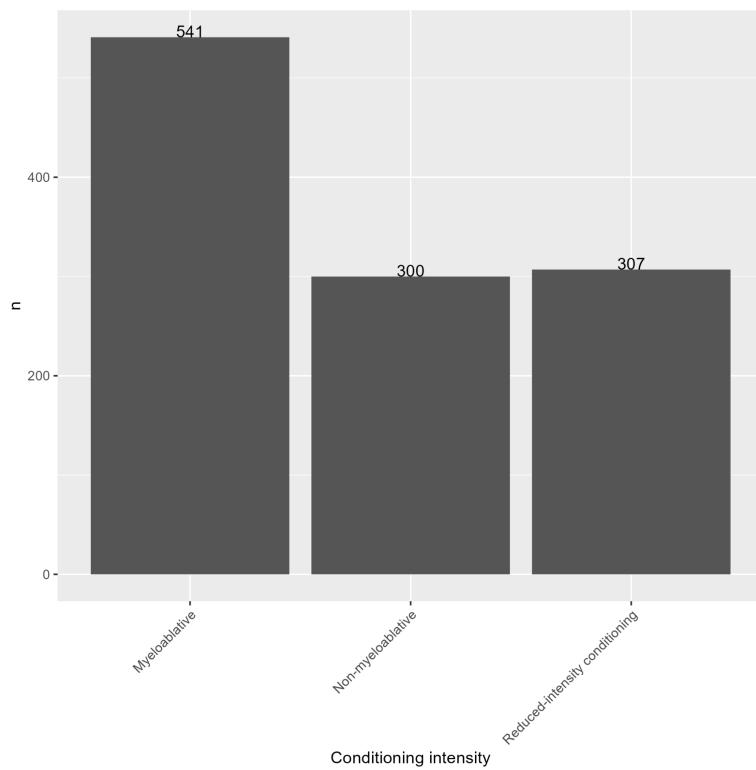
Plot 14: Distribution of Chronic GVHD Event



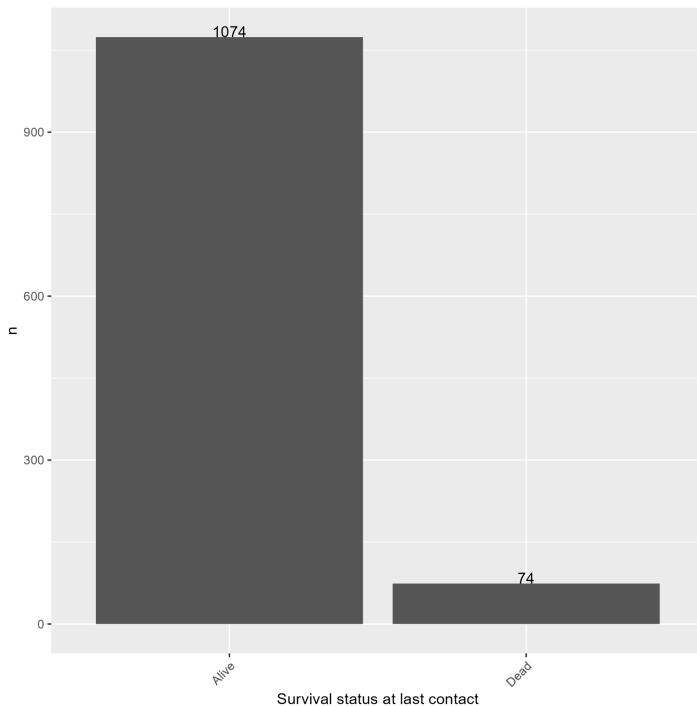
Plot 15: Distribution of Conditioning Regimen



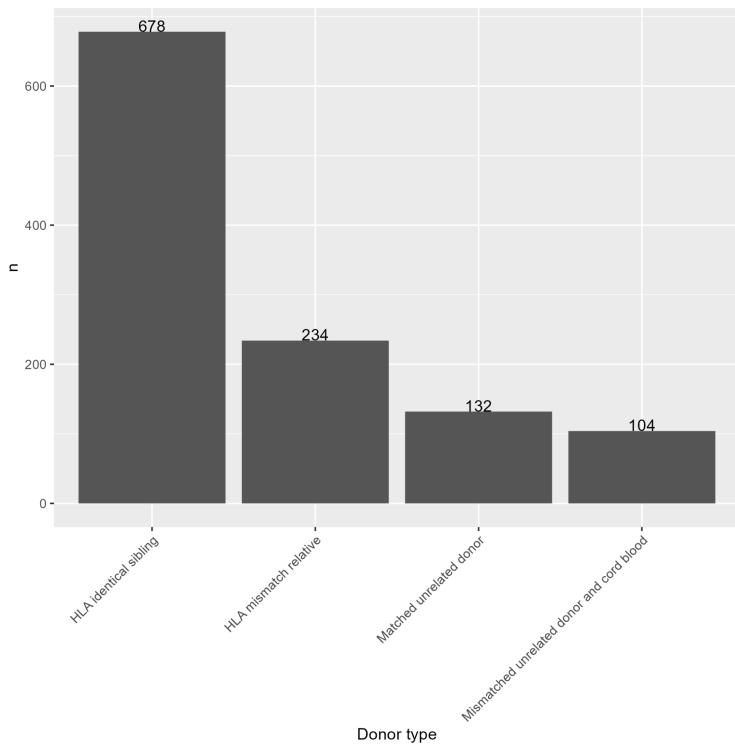
Plot 16: Distribution of Conditioning Intensity



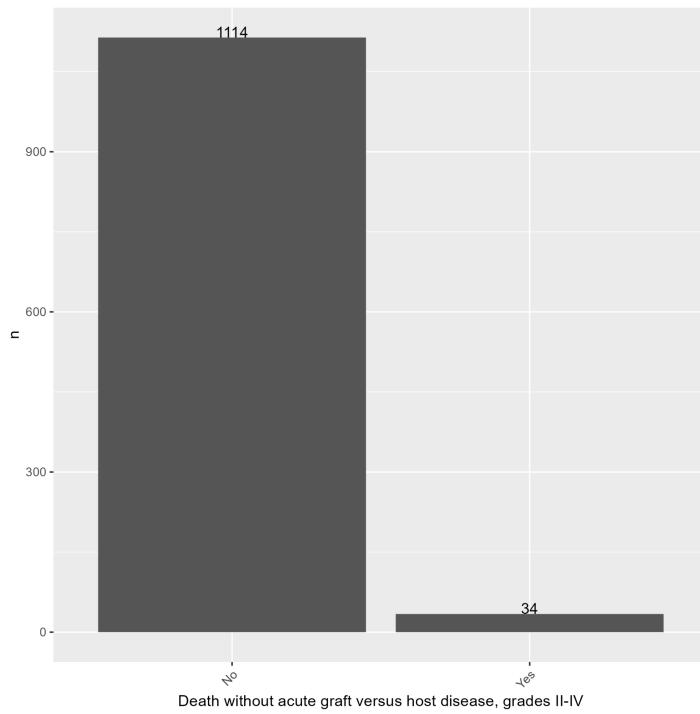
Plot 17: Distribution of Survival Status at Last Contact



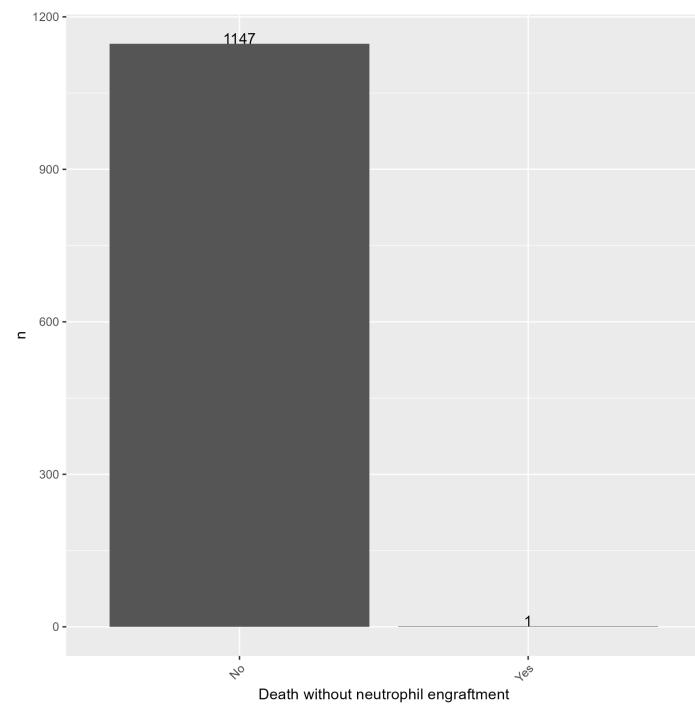
Plot 18: Distribution of Donor Type



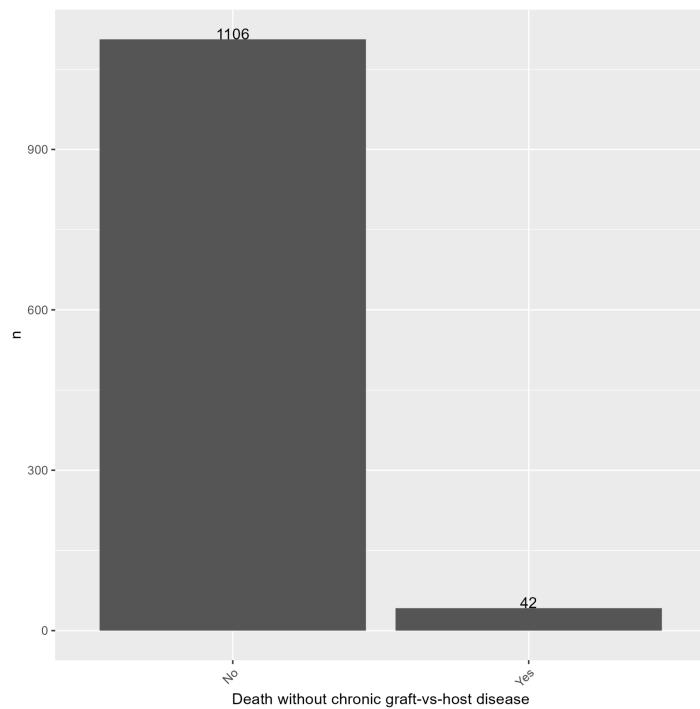
Plot 19: Distribution of Death without Acute GVHD Event



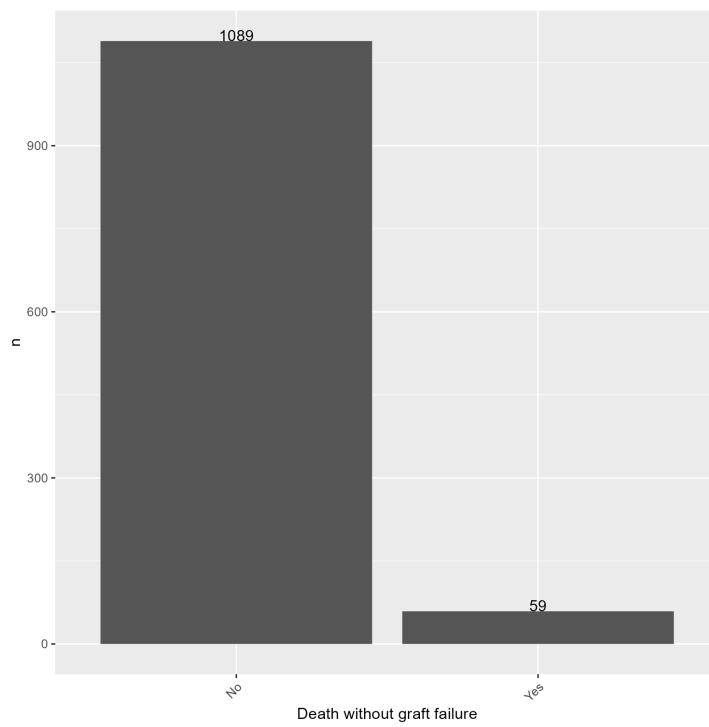
Plot 20: Distribution of Death without Neutrophil Engraftment Event



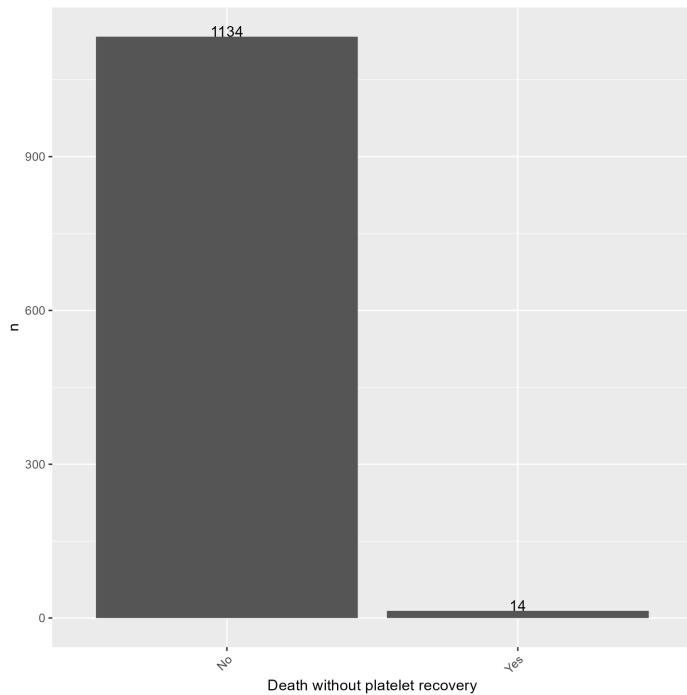
Plot 21: Distribution of Death without Chronic GVHD Event



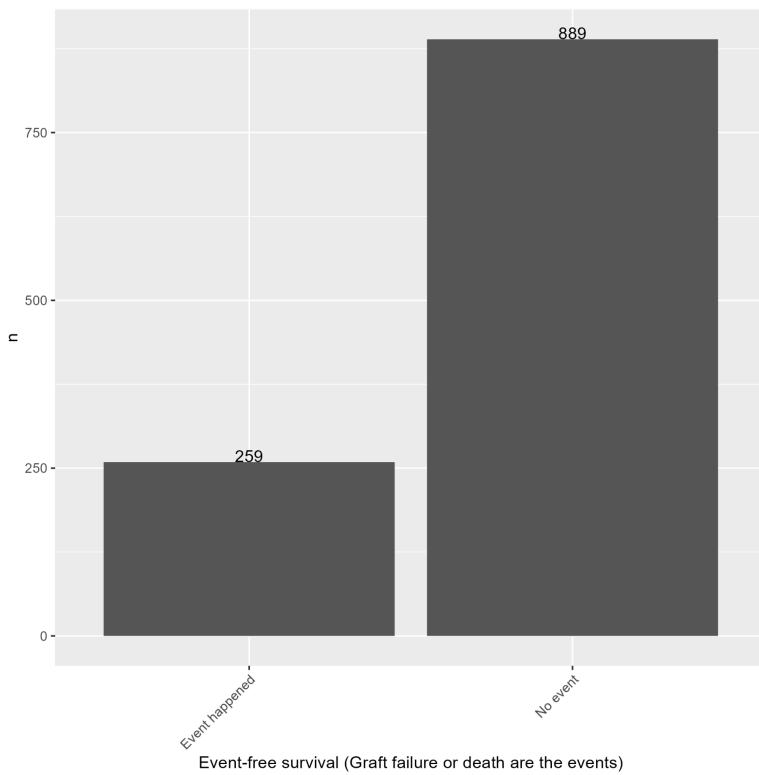
Plot 22: Distribution of Death without Graft Failure Event



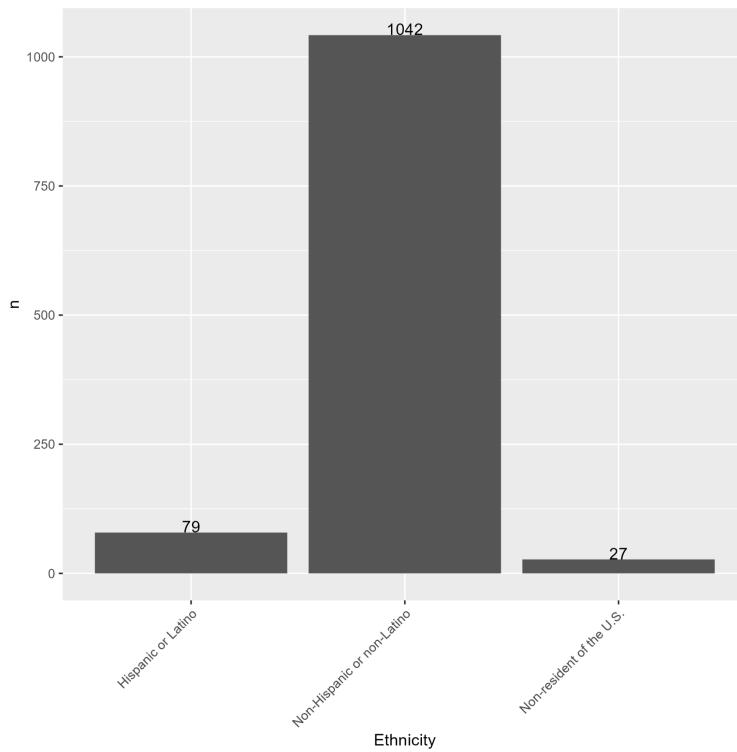
Plot 23: Distribution of Death without Platelet Recovery Event



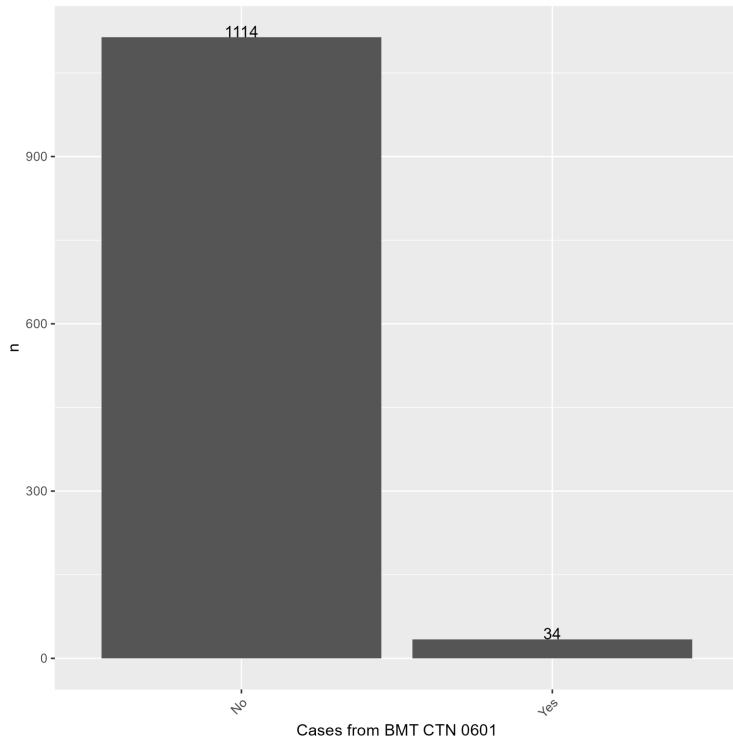
Plot 24: Distribution of Event-Free Survival (Graft Failure or Death are the Events)



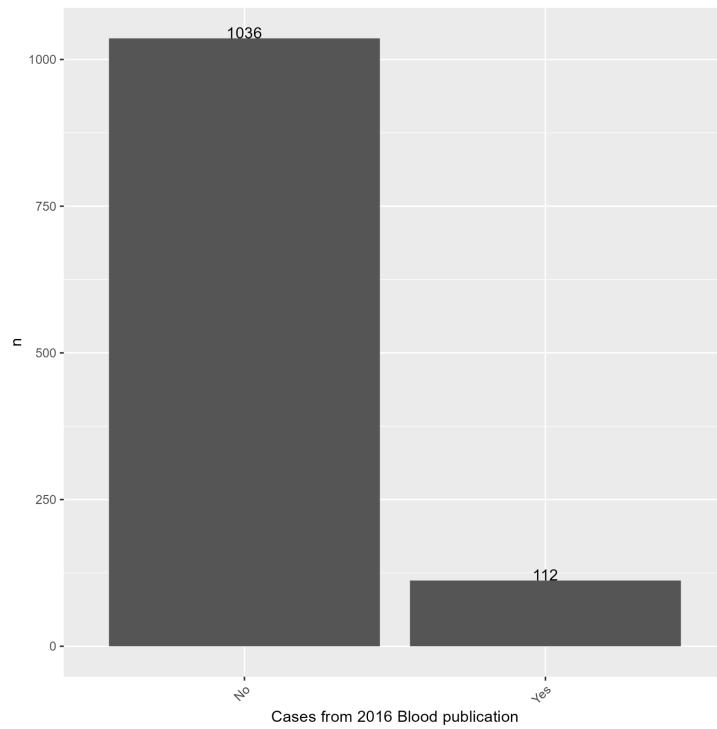
Plot 25: Distribution of Ethnicity



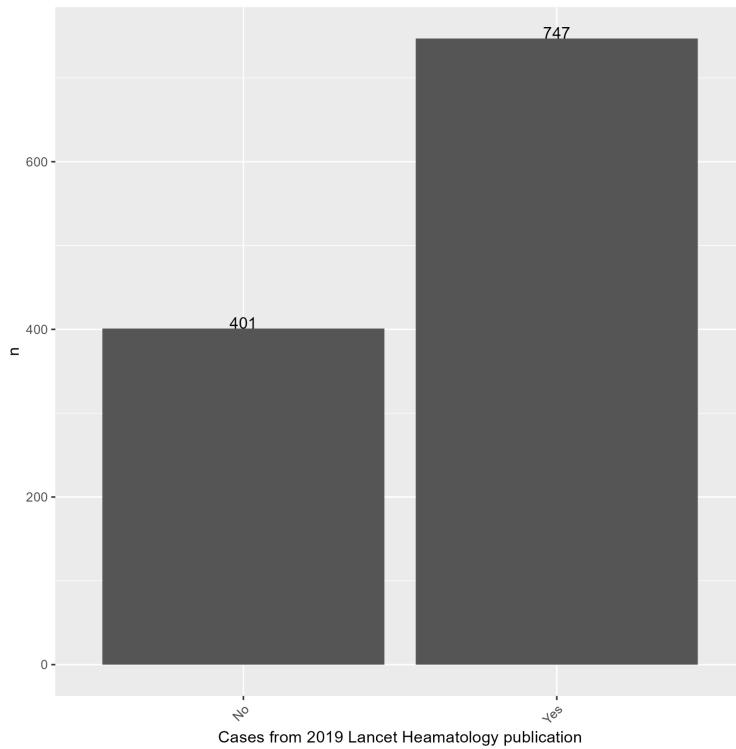
Plot 26: Distribution of Cases from BMT CTN 0601



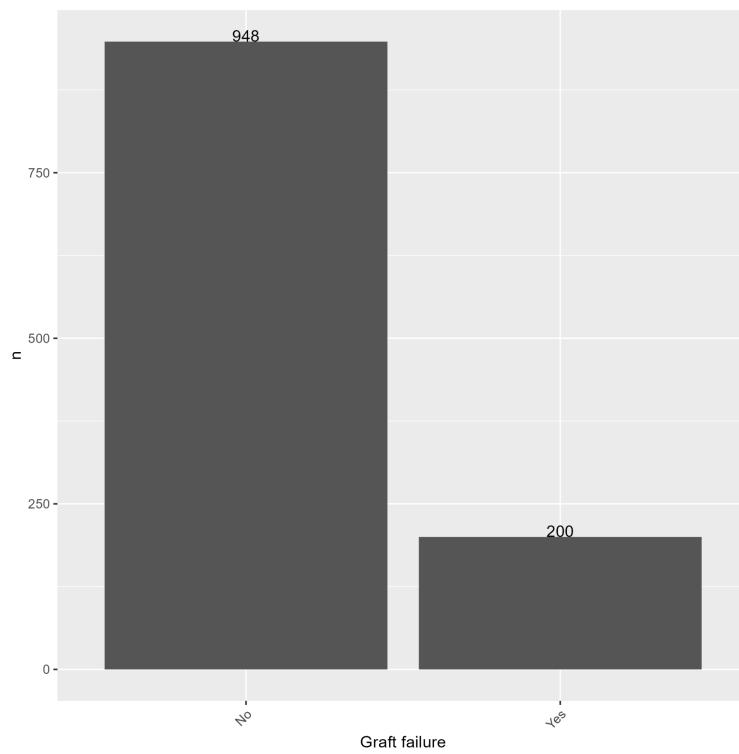
Plot 27: Distribution of Cases from 2016 Blood Publication



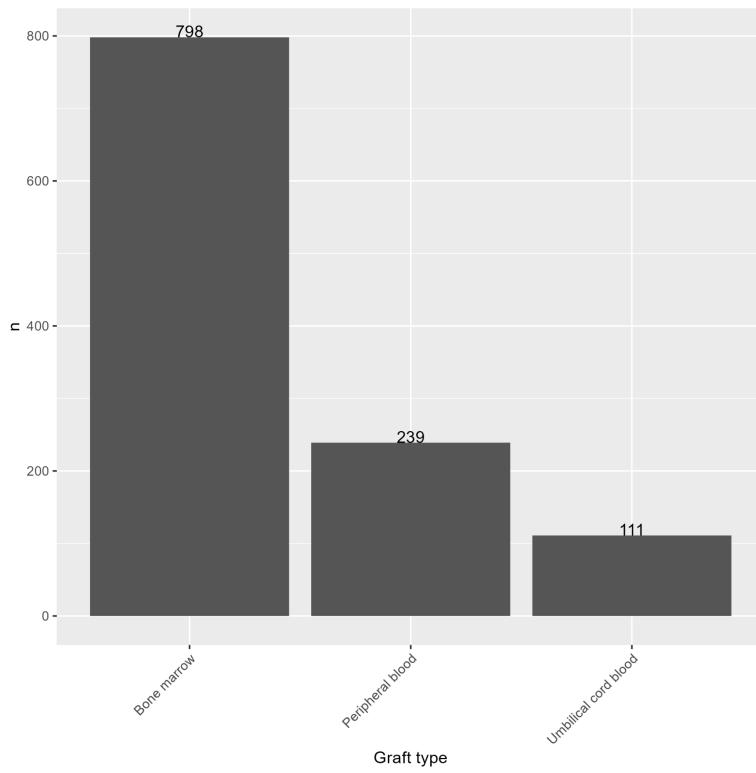
Plot 28: Distribution of Cases from 2019 Lancet Hematology Publication



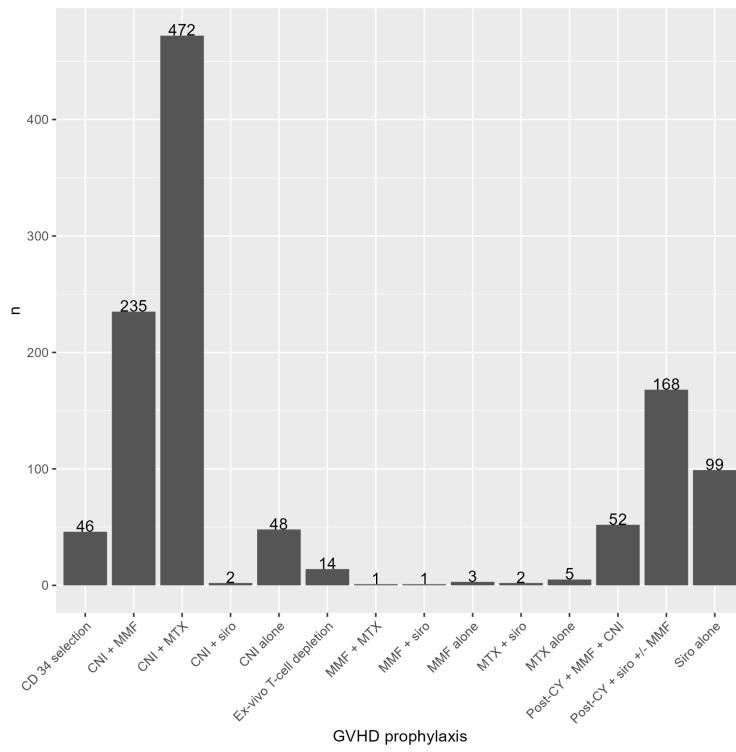
Plot 29: Distribution of Graft Failure Event



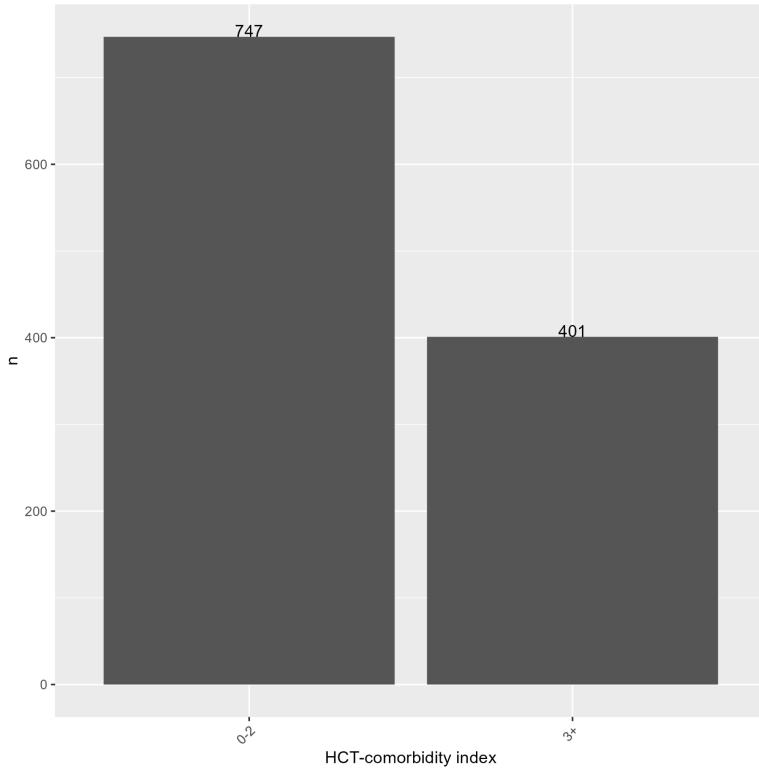
Plot 30: Distribution of Graft Type



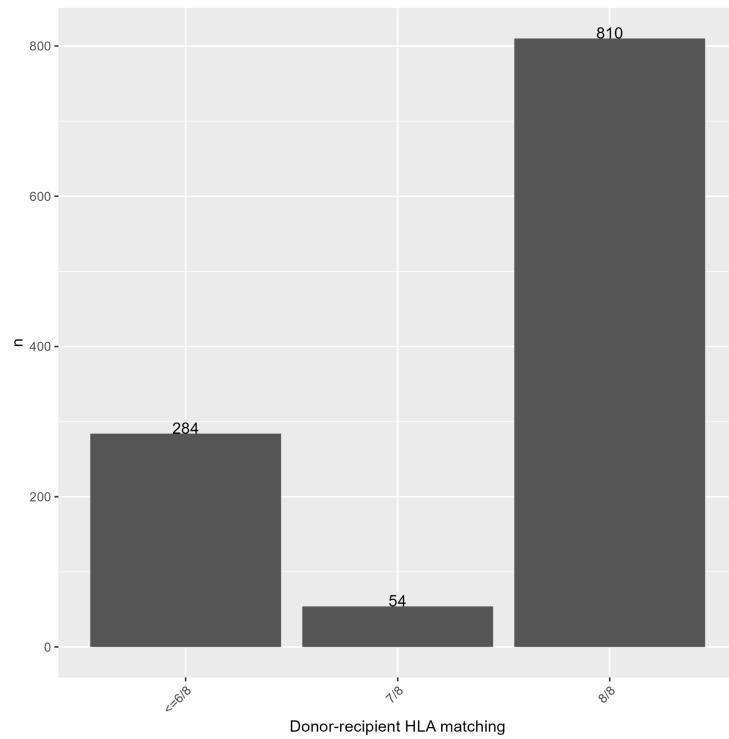
Plot 31: Distribution of GVHD Prophylaxis



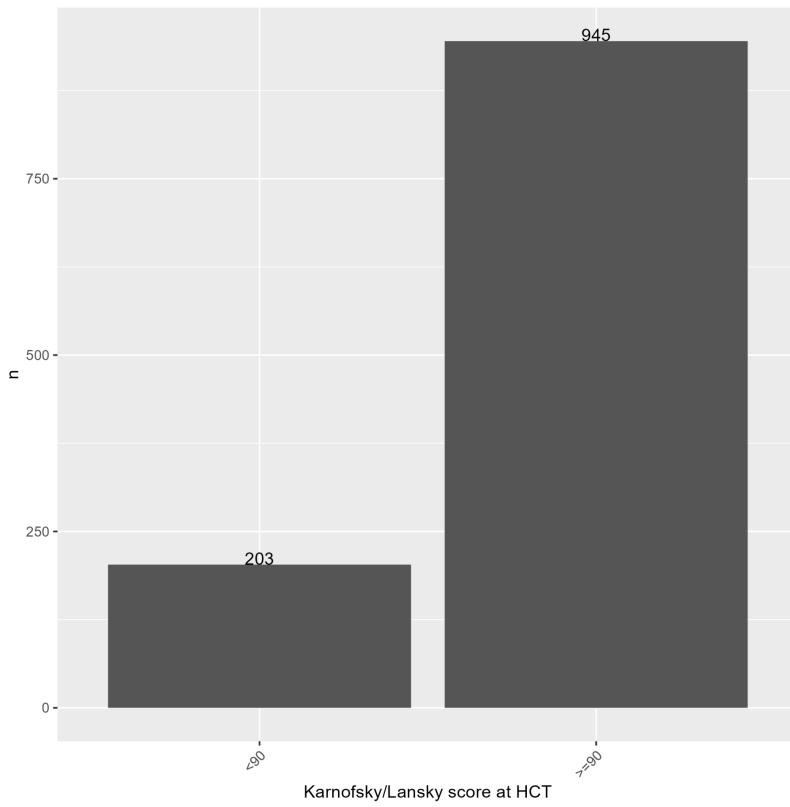
Plot 32: Distribution of HCT-Comorbidity Index



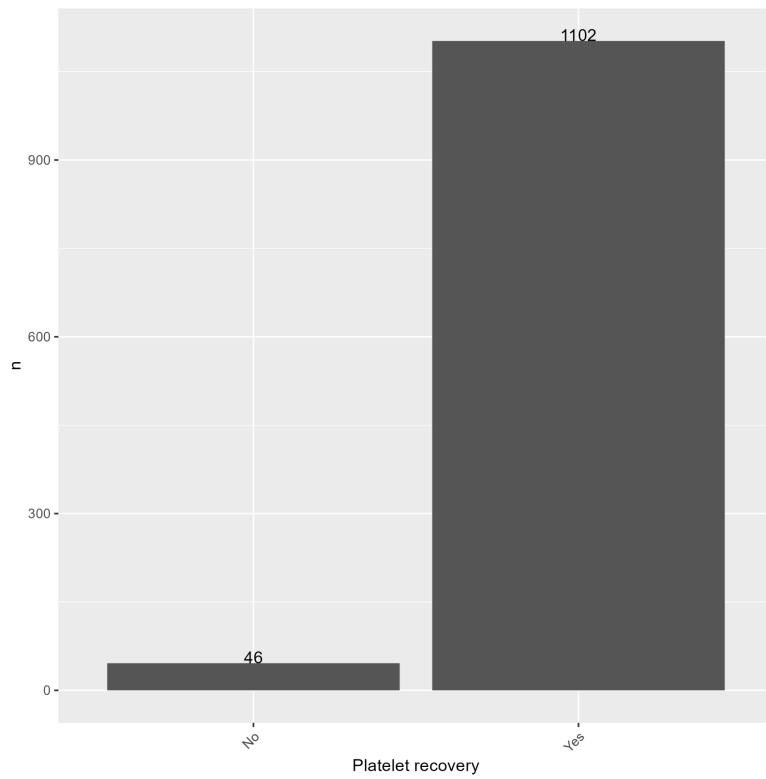
Plot 33: Distribution of Donor-Recipient HLA Matching



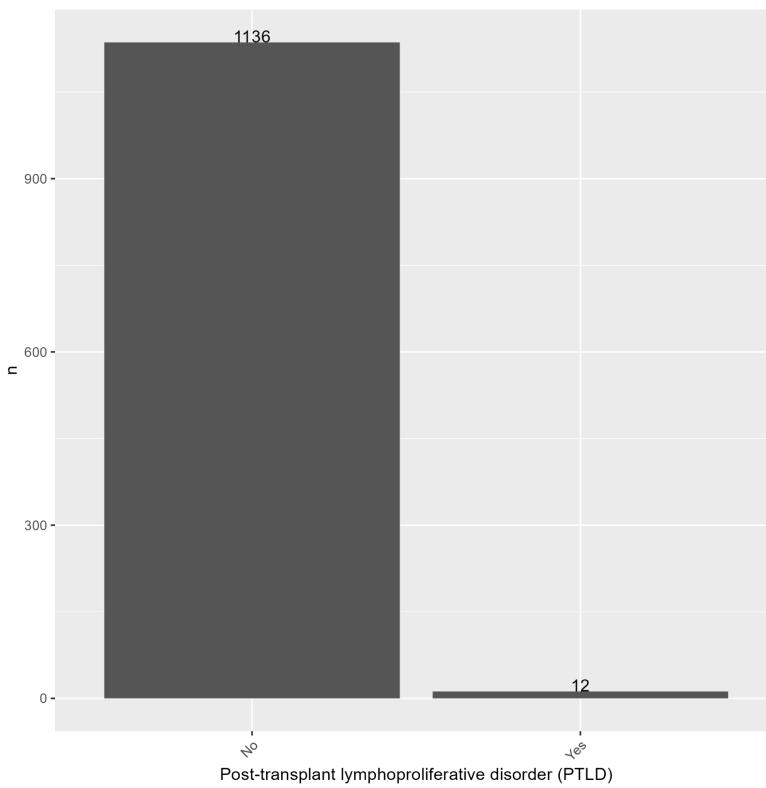
Plot 34: Distribution of Karnofsky/Lansky Score at HCT



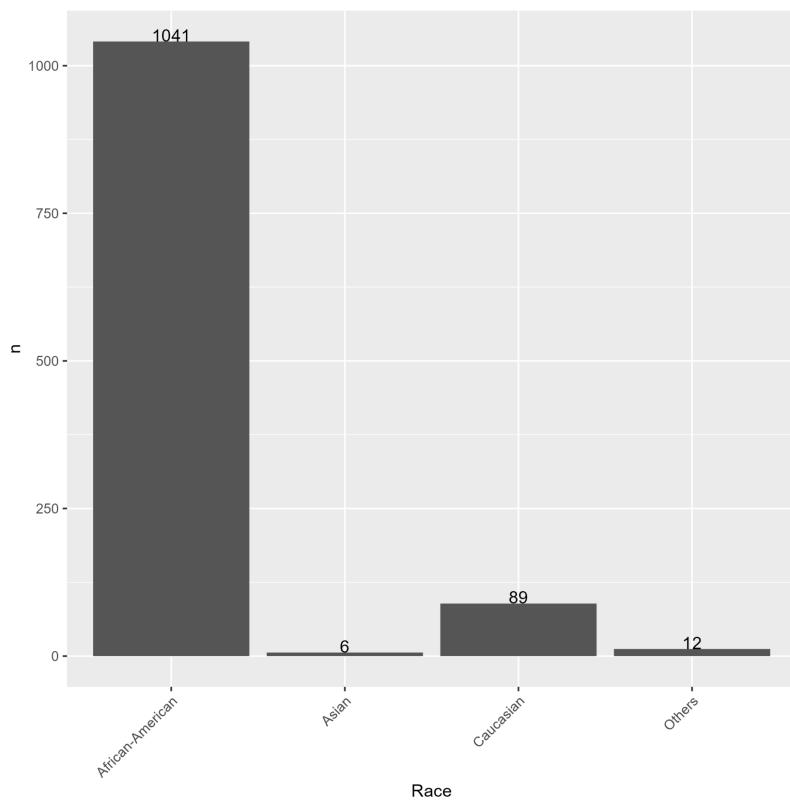
Plot 35: Distribution of Platelet Recovery Event



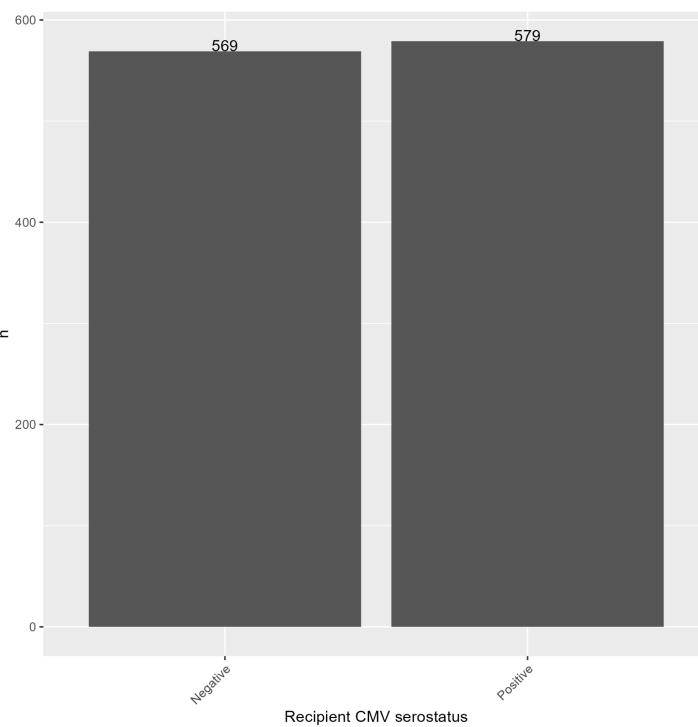
Plot 36: Distribution of PTLD Event



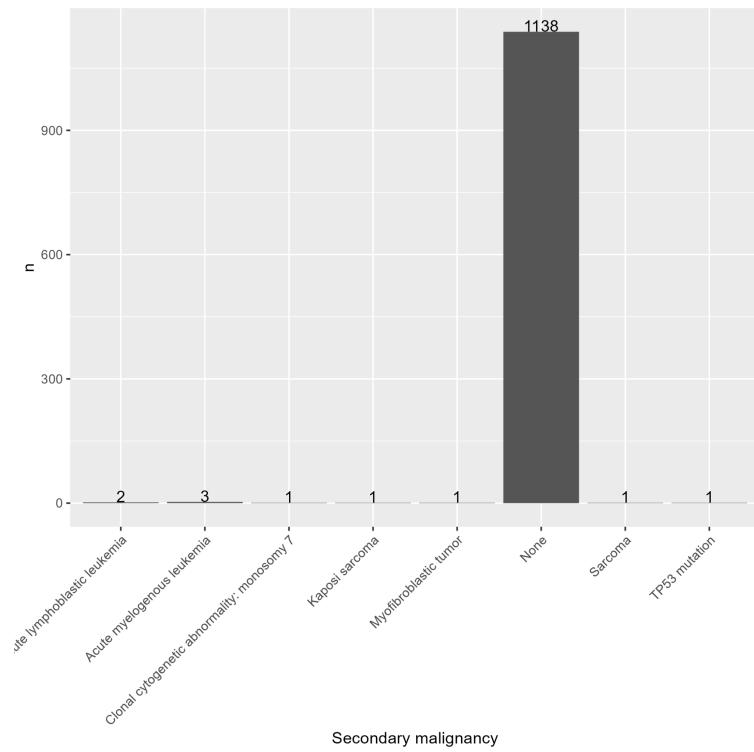
Plot 37: Distribution of Race



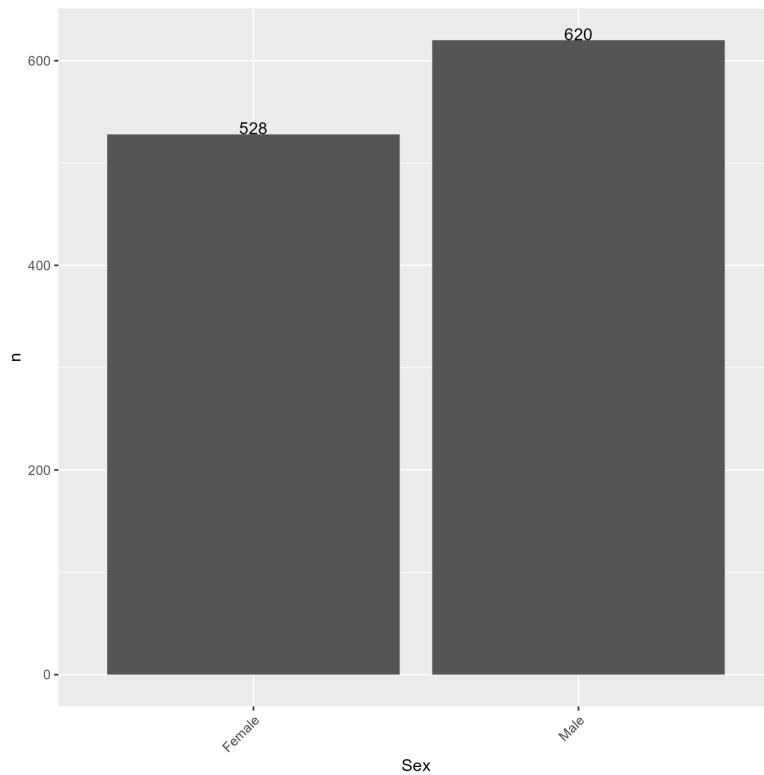
Plot 38: Distribution of Recipient CMV Serostatus



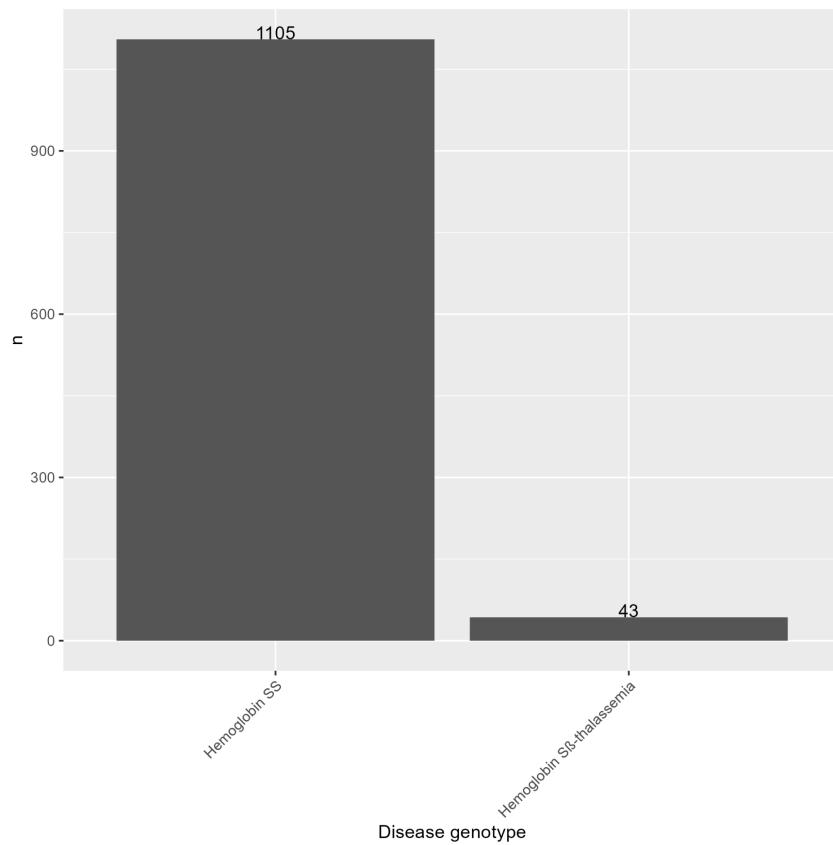
Plot 39: Distribution of Second Malignancy Event



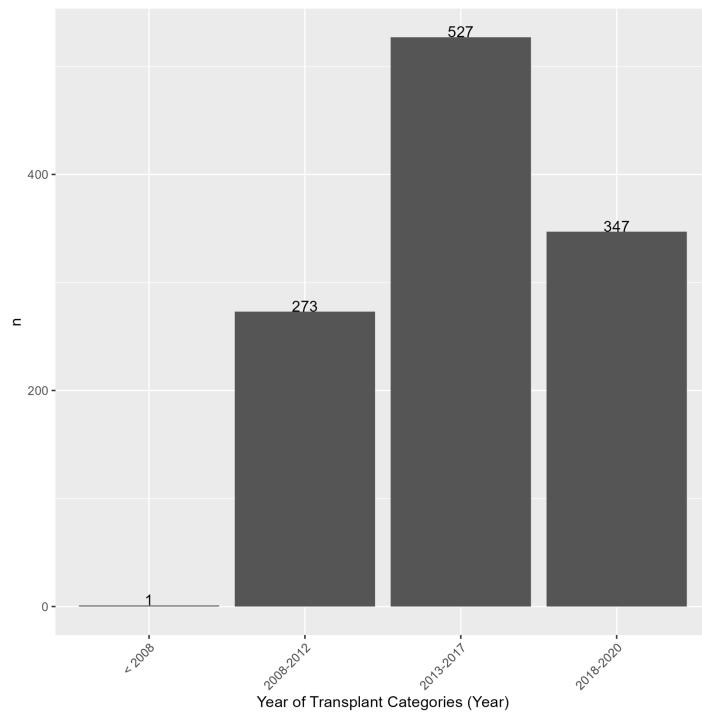
Plot 40: Distribution of Sex



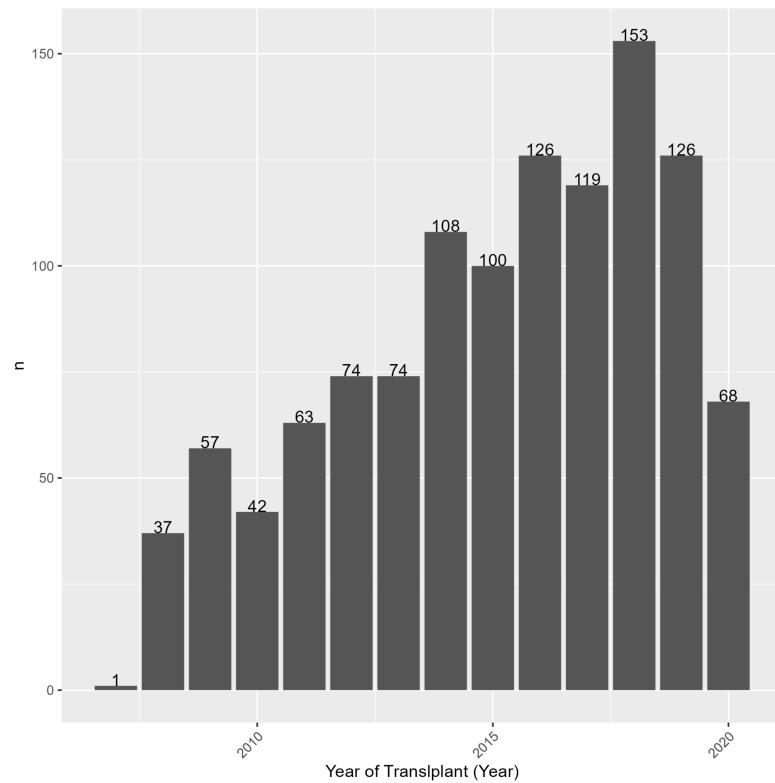
Plot 41: Distribution of Disease Genotype



Plot 42: Distribution of Year of Transplant Categories



Plot 43: Distribution of Year of Transplant



Section C:

Table 10: Coefficient Estimations in the Cox Hazard Model of Second Malignancy

feature	value
ETHNICITNon-Hispanic or non-Latino	0.01207333
ETHNICITNon-resident of the U.S.	0.29206130
YEARTX	0.53649371
AGEGPFF11-17	0.07298373
SUBDIS1FHemoglobin S β -thalassemia	0.16339464
YEARGPF2018-2020	0.70766233
GVHD_FINALPost-CY + MMF + CNI	-0.04723045
FLAG_LANCETYes	-0.06607486

Table 11: Coefficient Estimations in the Cox Hazard Model of Platelet Recovery

feature	value
DONORFMismatched unrelated donor and cord blood	-0.285141383
GRAFTYPEPeripheral blood	0.166442215
GRAFTYPEUmbilical cord blood	-0.371292510
AGEGPFF18-29	0.090657641
KPS>=90	-0.003971782
HCTCIGPF3+	-0.024423507
YEARGPF2008-2012	0.101732833
GVHD_FINALCNI + MMF	-0.003739682
GVHD_FINALCNI alone	-0.010551806
GVHD_FINALMTX + siro	-0.177679755
GVHD_FINALSiro alone	0.954649225
CONDGRP_FINALCy/Flu	-0.391456938
CONDGRP_FINALFlu/Mel	0.394768832
CONDGRP_FINALFlu/Mel/TT	0.284925325
CONDGRP_FINALTBI/Cy/Flu	0.438781550
CONDGRP_FINALTBI/Flu	0.412805923
HLA_FINAL8/8	0.033584846

Table 12: Coefficient Estimations in the Cox Hazard Model of Graft Failure

feature	value
RCMVPRPositive	0.0008738952
SEXMale	0.2711801577
ETHNICITNon-Hispanic or non-Latino	0.1133762353
DONORFMatched unrelated donor	0.7619634990
DONORFMismatched unrelated donor and cord blood	0.3675253648
GRAFTYPEPeripheral blood	0.0848108091
GRAFTYPEUmbilical cord blood	0.1034953265
AGEGPFF>=50	1.0804914571
KPS>=90	0.1441573500
ATGFATG	-0.0136277635
ATGFNone	-0.2000531293
GVHD_FINALCNI + MMF	0.1651548195
GVHD_FINALCNI + MTX	-0.0303100618
GVHD_FINALCNI alone	0.1090796845
GVHD_FINALEx-vivo T-cell depletion	0.4673639135
GVHD_FINALMMF + siro	1.0895918471
GVHD_FINALPost-CY + siro +/- MMF	-0.1299693639
CONDGRP_FINALFlu/Bu	0.0248220851
CONDGRP_FINALFlu/Bu/TT	0.2035258666
CONDGRP_FINALFlu/Mel	0.0319480945
CONDGRP_FINALFlu/Mel/TT	-0.3187810243
CONDGRP_FINALTBI alone (300/400cGy)	0.4698013315
CONDGRP_FINALTBI/Cy/Flu	0.4238854582
CONDGRP_FINALTBI/Mel	0.2908137133
HLA_FINAL8/8	-1.3137791954
FLAG_BLOODYYes	0.7697372857

Table 13: Coefficient Estimations in the Cox Hazard Model of PTLD

feature	value
ETHNICITNon-Hispanic or non-Latino	0.0826684800
ETHNICITNon-resident of the U.S.	0.4989728156
GRAFTYPEPeripheral blood	0.0472841558
GRAFTYPEUmbilical cord blood	0.0030274328
YEARTX	0.5735240797
AGEGPFF11-17	0.1071757341
KPS>=90	0.0002790435
SUBDIS1FHemoglobin S β -thalassemia	0.3453871297
ATGFATG	0.0009297446
YEARGPF2013-2017	-0.1187940900
YEARGPF2018-2020	0.5394242336
GVHD_FINALCNI + siro	0.0166083096
GVHD_FINALPost-CY + MMF + CNI	-0.1752612100
CONDGRPReduced-intensity conditioning	0.0494704261
CONDGRP_FINALFlu/Bu/TT	0.0323243067
CONDGRP_FINALFlu/Mel	0.0020071975
CONDGRP_FINALTBI alone (300/400cGy)	-0.0675968146
CONDGRP_FINALTBI/Cy/Flu/TT	0.0779337189
CONDGRP_FINALTBI/Flu	0.0256583727
FLAG_LANCETYes	-0.0617151364

Table 14: Coefficient Estimations in the Cox Hazard Model of Death or Last Contact

feature	value
SEXMale	-0.37412666
ETHNICITNon-Hispanic or non-Latino	-0.45712417
DONORFMatched unrelated donor	1.74255660
GRAFTYPEPeripheral blood	0.60914410
YEARTX	-0.07652276
AGE	0.05726610
AGEGPFF>=50	0.56461185
AGEGPFF11-17	0.12976430
AGEGPFF18-29	0.31630840
KPS>=90	0.28395577
SUBDIS1FHemoglobin S β -thalassemia	0.09233150
ATGFATG	0.43162455
ATGFNone	-0.44763112
YEARGPF2013-2017	0.25515986
GVHD_FINALCNI + MMF	0.52659224
GVHD_FINALCNI + MTX	0.32606015
GVHD_FINALCNI + siro	-1.97031554
GVHD_FINALCNI alone	0.45079886
GVHD_FINALEx-vivo T-cell depletion	-0.96185657
GVHD_FINALMTX alone	3.21376765
GVHD_FINALPost-CY + MMF + CNI	-0.63879893
GVHD_FINALPost-CY + siro +/- MMF	-0.10220066
CONDGRPNon-myeloablative	-0.60947607
CONDGRP_FINALCy alone	-0.23511771
CONDGRP_FINALFlu/Bu	-0.21324500
CONDGRP_FINALFlu/Bu/TT	1.18908616
CONDGRP_FINALFlu/Mel/TT	-0.70837349
CONDGRP_FINALTBI alone (300/400cGy)	-0.68805567
CONDGRP_FINALTBI/Cy	0.10683612
CONDGRP_FINALTBI/Cy/Flu	-0.50067580
CONDGRP_FINALTBI/Mel	1.30259989
HLA_FINAL8/8	-1.30091743
FLAG_BLOODYes	0.50855267

Table 15: Coefficient Estimations in the Cox Hazard Model of Chronic GVHD

feature	value
SEXMale	-0.06920065
ETHNICITNon-resident of the U.S.	-0.12871798
DONORFMatched unrelated donor	0.79174479
DONORFMismatched unrelated donor and cord blood	0.41252028
GRAFTYPEPeripheral blood	-0.11870919
AGE	0.02652853
AGEGPFF>=50	-0.26556058
AGEGPFF18-29	0.33301499
KPS>=90	-0.05626722
HCTCIGPF3+	0.05603813
ATGFATG	0.04819949
ATGFNone	0.42186069
YEARGPF2008-2012	-0.10091982
YEARGPF2018-2020	0.05533465
GVHD_FINALCNI + MTX	0.03356692
GVHD_FINALEx-vivo T-cell depletion	-0.03811708
GVHD_FINALMTX + siro	0.90372932
GVHD_FINALSiro alone	-0.47051620
CONDGRPFNon-myeloablative	-0.30309555
CONDGRPFReduced-intensity conditioning	0.01252967
CONDGRP_FINALCy/Flu	-0.02798929
CONDGRP_FINALFlu/Bu	-0.30028247
CONDGRP_FINALFlu/Mel	0.19803276
CONDGRP_FINALTBI alone (300/400cGy)	-1.61754534
CONDGRP_FINALTBI/Cy	-0.69513741
CONDGRP_FINALTBI/Cy/Flu	-0.48395209
CONDGRP_FINALTBI/Flu	0.38906330
CONDGRP_FINALTBI/Mel	-0.70783266
HLA_FINAL8/8	-0.04536742

Table 16: Coefficient Estimations in the Cox Hazard Model of Neutrophil Recovery

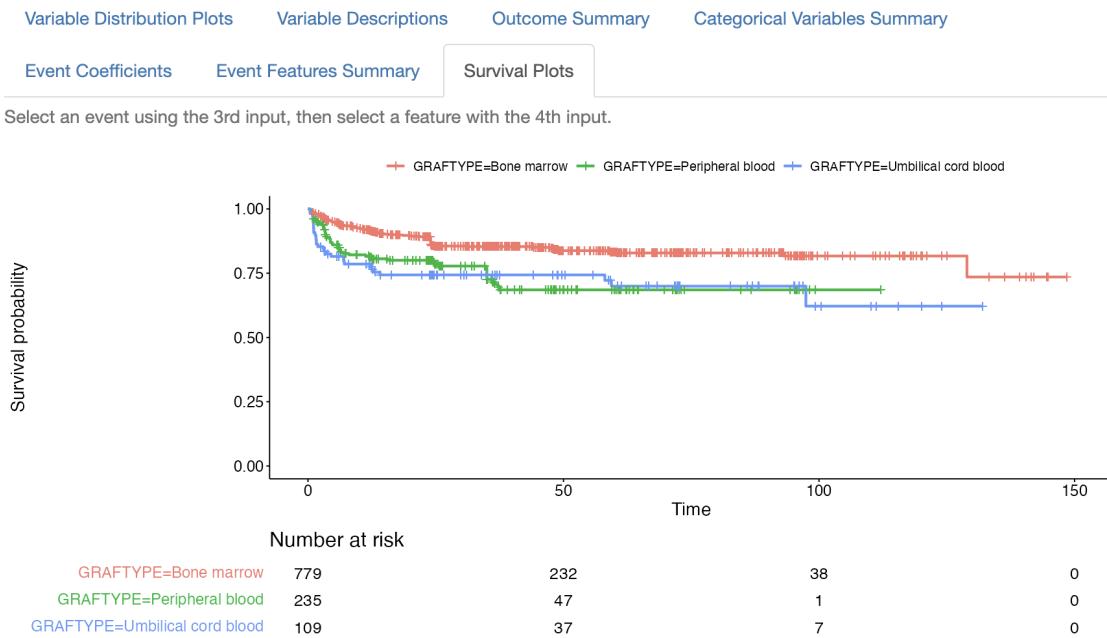
feature	value
DONORFMatched unrelated donor	3.159459e-02
GRAFTYPEPeripheral blood	9.100892e-02
GRAFTYPEUmbilical cord blood	-5.605726e-01
YEARTX	1.209365e-03
AGEGPFF11-17	1.131159e-02
ATGFATG	8.398803e-03
YEARGPF2018-2020	3.747162e-05
GVHD_FINALPost-CY + siro +/- MMF	-2.360926e-01
CONDGRPNon-myeloablative	-4.976128e-01
CONDGRP_FINALFlu/Bu	-1.025819e-01
CONDGRP_FINALFlu/Mel	5.574895e-01
CONDGRP_FINALTBI/Cy/Flu	-4.816363e-02
CONDGRP_FINALTBI/Mel	-1.925069e-01
HLA_FINAL8/8	2.908103e-03
FLAG_LANCETYes	-2.711459e-02

Table 17: Coefficient Estimations in the Cox Hazard Model of Acute GVHD

feature	value
RCMVPRPositive	0.225167503
SEXMale	-0.015290534
ETHNICITNon-resident of the U.S.	-0.294307553
DONORFMatched unrelated donor	1.159015608
DONORFMismatched unrelated donor and cord blood	1.076401125
GRAFTYPEUmbilical cord blood	0.200562075
AGE	0.005666673
AGEGPFF18-29	0.279408272
AGEGPFF30-49	0.438233789
KPS>=90	-0.127463578
ATGFATG	0.449429817
ATGFNone	0.697926590
YEARGPFF2008-2012	-0.361287957
YEARGPFF2013-2017	0.209093980
GVHD_FINALCNI + MMF	0.137773332
GVHD_FINALCNI + MTX	0.051048623
GVHD_FINALCNI alone	-0.029808549
GVHD_FINALEx-vivo T-cell depletion	-0.097356432
GVHD_FINALMMF + siro	-0.253623657
GVHD_FINALMMF alone	-0.656300476
GVHD_FINALMTX alone	0.114083113
GVHD_FINALPost-CY + MMF + CNI	-0.353903665
CONDGRPFNon-myeloablative	-0.411201698
CONDGRP_FINALCy alone	-0.818254598
CONDGRP_FINALCy/Flu	-0.680129520
CONDGRP_FINALFlu/Bu	-0.121370190
CONDGRP_FINALFlu/Mel	0.240456603
CONDGRP_FINALTBI alone (300/400cGy)	-1.285452695
CONDGRP_FINALTBI/Cy/Flu	-0.644616006
CONDGRP_FINALTBI/Cy/Flu/TT	0.019110944
HLA_FINAL7/8	-0.438004933
HLA_FINAL8/8	-0.741684288
FLAG_LANCETYes	-0.328618675

Section D:

Illustration A



References

Centers for Disease Control and Prevention. (2023a, July 6). Data & statistics on Sickle Cell Disease. Centers for Disease Control and Prevention. <https://www.cdc.gov/ncbddd/sicklecell/data.html>

Centers for Disease Control and Prevention. (2023b, July 6). What is sickle cell disease?. Centers for Disease Control and Prevention. <https://www.cdc.gov/ncbddd/sicklecell/facts.html>

Chen, D.-P., Wen, Y.-H., Wang, P.-N., Hour, A.-L., Lin, W.-T., Hsu, F.-P., & Wang, W.-T. (2021). The adverse events of haematopoietic stem cell transplantation are associated with gene polymorphism within human leukocyte antigen region. *Scientific Reports*, 11(1). <https://doi.org/10.1038/s41598-020-79369-w>

Sickle cell disease – outcomes & advances. bethematchclinical.org. (n.d.).
<https://bethematchclinical.org/transplant-indications-and-outcomes/disease-specific-indications-and-outcomes/sickle-cell-disease/>

St. Martin, A., Hebert, K. M., Serret-Larmande, A., Jouhet, V., Hughes, E., Stedman, J., DeSain, T., Pillion, D., Lyons, J. C., Steinert, P., Avillach, P., & Eapen, M. (2022). Long-term survival after hematopoietic cell transplant for sickle cell disease compared to the United States population. *Transplantation and Cellular Therapy*, 28(6). <https://doi.org/10.1016/j.jtct.2022.03.014>

ggplot2: Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN: 978-3-319-24277-4

dplyr: Wickham, H., François, R., Henry, L., & Müller, K. (2021). *dplyr: A Grammar of Data Manipulation*. R package version 1.0.7. <https://CRAN.R-project.org/package=dplyr>

haven: Wickham, H., & Miller, E. (2021). *haven: Import and Export 'SPSS', 'Stata' and 'SAS' Files*. R package version 2.4.3. <https://CRAN.R-project.org/package=haven>

tidy: Wickham, H. (2021). *tidy: Tidy Messy Data*. R package version 1.1.4.
<https://CRAN.R-project.org/package=tidy>

readr: Wickham, H., Hester, J., & Francois, R. (2021). *readr: Read Rectangular Text Data*. R package version 2.1.1. <https://CRAN.R-project.org/package=readr>

stringr: Wickham, H. (2019). *stringr: Simple, Consistent Wrappers for Common String Operations*. R package version 1.4.0. <https://CRAN.R-project.org/package=stringr>

readxl: Wickham, H., & Bryan, J. (2021). *readxl: Read Excel Files*. R package version 1.3.1.
<https://CRAN.R-project.org/package=readxl>

gt: Iannone, R., Cheng, J., & Schloerke, B. (2021). *gt: Easily Create Presentation-Ready Display Tables*. R package version 0.3.0. <https://CRAN.R-project.org/package=gt>

kableExtra: Zhu, H. (2021). kableExtra: Construct Complex Table with 'kable' and Pipe Syntax. R package version 1.3.4. <https://CRAN.R-project.org/package=kableExtra>

survival: Therneau, T. (2021). A Package for Survival Analysis in R. R package version 3.2-11. <https://CRAN.R-project.org/package=survival>

glmnet: Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*, 33(1), 1-22. <https://www.jstatsoft.org/v33/i01/>

boot: Canty, A., & Ripley, B. D. (2021). boot: Bootstrap R (S-Plus) Functions. R package version 1.3-28. <https://CRAN.R-project.org/package=boot>

survminer: Kassambara, A., Kosinski, M., & Biecek, P. (2021). survminer: Drawing Survival Curves using 'ggplot2'. R package version 0.4.9. <https://CRAN.R-project.org/package=survminer>

knitr: Xie, Y. (2021). knitr: A General-Purpose Package for Dynamic Report Generation in R. R package version 1.33. <https://CRAN.R-project.org/package=knitr>

gtExtras: Merlino, D. & Rich, J. (2021). gtExtras: Additional Functions for Table Formatting with the gt Package. GitHub repository. <https://jthomasmock.github.io/gtExtras/>

purrr: Wickham H, Henry L (2023). purrr: Functional Programming Tools. R package version 1.0.2. <https://CRAN.R-project.org/package=purrr>

randomcoloR: Ammar R (2019). randomcoloR: Generate Attractive Random Colors. R package version 1.1.0.1. <https://CRAN.R-project.org/package=randomcoloR>

Shiny: Chang, W., Cheng, J., Allaire, J.J., Xie, Y., & McPherson, J. (2021). shiny: Web Application Framework for R. R package version 1.7.1. <https://CRAN.R-project.org/package=shiny>