Final Project

BIOS 6611: Biostatistical Methods

Joseph Froelicher

12/10/2020

Abstract

Question: Is there a relationship between the length of storage of autologous red blood cells (RBCs), and the number of units of transfused per patient?
Methods: There are 316 Patients divided into three RBC age groups, younger, middle and older. The goal is to examine the median number of units tranfused within each group to make between group comparisons. This is performed using a Dunn's Kruskal-Wallis test for multiple pairwise comparisons. Additionally, the resilience of the medians are examined using a permutation test of the difference of medians between groups.
Results: For patients (n = 316) in three groups, the median units transfused between RBC age groups are not different between groups.
Discussion: The length of storage is not significantly associated with changes in units tranfused in prostatectomy patients that required RBC transfusion.

Introduction

Prostate cancer is very common in men, and the associated surgery frequently requires blood transfusions. Blood transfusions are a complicated medical operation that can result in varying reactions to the transfusion.[1] Additionally, blood used in transfusions is stored for varying amounts of time.[2] In order to test in the future relationships between blood storage and recurrence of prostate cancer, it is necessary to examine the realtionship of the length of storage of blood, with the number of tranfusion units used on a patient. Based on previous a previous study[3], it is hypothesized that the the median number of units between RBC age groups will not be different.

Methods

The data were gathered from from the Cleveland clinic, from patients who recieved laparoscopic radical prostatectomy. The data contain 316 male patients who had a blood transfusion after Prostate surgery. Those patients are subdivided into three strata, younger blood (n = 106), middle blood (n = 103), and older blood (n = 107)(Table 1.). The groups are defined by the number of days the blood spent in storage, where younger is less than thirteen days, middle is thirteen to eighteen days, and older is more than eighteen days. The data contain information on Race and family history of disease, as well as age (Table 1.).

The purpose of this analysis is to examine the median number of units in each of the three groups. This is done in R (4.0.3), with an assumed alpha level of five percent, by performing a Dunn's Kruskal-Wallis tests of pairwise median comparisons. Following the pairwise comparisons, the results are corrected by a Bonferroni correction (n = 3). To examine the resilience of the median as a parameter of interest for number of units transfused, a second pairwise comparison was performed, sampling from the original data without replacement (n = 10,000), and again correcting using a Bonferroni correction (n = 3).[4]

Descriptive Statistics

	Total	< 13 Days	13 - 18 Days	> 18 Days
	n = 316	n = 106	n = 103	n = 107
Race				
Not African American	261 (82.6%)	88 (83%)	85 (82.5%)	88 (82.2%)
African American	55 (17.4%)	18 (17%)	18 (17.5%)	19 (17.8%)
Family history of disease				
No	248 (78.5%)	83 (78.3%)	86 (83.5%)	79 (73.8%)
Yes	68 (21.5%)	23 (21.7%)	17 (16.5%)	28 (26.2%)
Age [Median(SD)]				
	61.8 (7.2)	62 (7)	62.2 (7)	61.7 (7.7)
Number of allogeneic units [Median(SD)]				
	2 (1.9)	2 (1.6)	2 (1.1)	2 (2.6)

Table 1. *Descriptive statistics for the Blood storage data set. Categorical variables are reported with medians and standard deviations.*

Results

The median and standard deviation number of units of allogeneic blood transfused per group are 2(1.6), 2(1.1), and 2(2.6)(Table 1.), for younger blood, middle blood, and older blood respectively. After pairwise comparisons using Dunn's Kruskal-Wallis test for multiple comparisons, and a Bonferroni correction, the estimated mean-rank differences were: -11.7 (p = 0.92) comparing middle blood to younger blood, -31.4 (p = 0.051) comparing older blood to younger blood, and -19.7 (p = 0.51) comparing older blood to middle blood. Blood is not signficiantly different between RBC age groups (alpha = 0.05)(Table 2.).

	Point Estimate	P-value	Adjusted P-value
Middle-Younger	-11.68154	0.3073300	0.9219899
Older-Younger	-31.38256	0.0168752	0.0506255
Older-Middle	-19.70103	0.1688257	0.5064771

Table 2. *Results of Dunn's Kruskal-Wallis between group comparisons. The adjusted P values use a Bonferroni correction (n = 3).*

As can be seen in the box plots, after visual examination the medians do not appear to be different (Figure 1.). Additionally, the interquartile ranges shown in the boxplots in Figure 1. appear to be very narrow, indicating little variablility within RBC age groups. There are a few outliers, however, the median estimate will not be affected significantly by outliers.

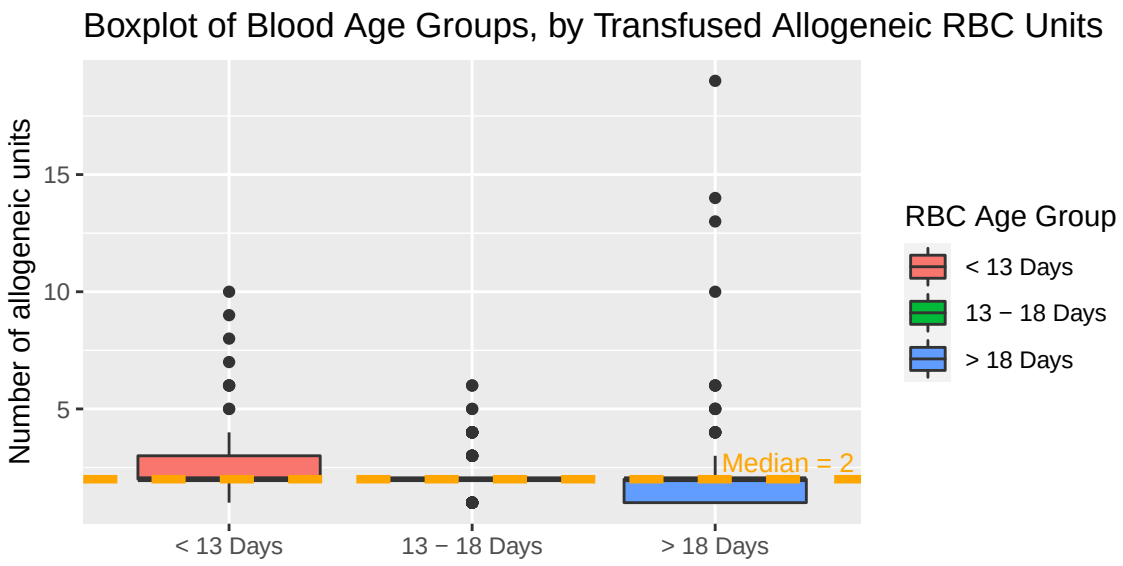

Figure 1. *Boxplots of the number of tranfused allogeneic units by RBC storage duration group. The reported median for each group is two units.*

After sampling the data without replacement (n = 10,000), it cannot be demonstrated that the median is infact a robust estimator of the number of transfused units. The median as an estimator was not demonstrated to have come from a normal distribution. Further analysis could exmine the mean as an estimator. Further more, the Dunn's Kruskal-Wallis test was used to test between group rank differences for the simulated data. This exploration revealed no significant difference between the mean-rank differences of the simulated medians.

Discussion

In response to the initial objective of examining the relationship between the length of storage of autologous red blood cells (RBCs), and the number of units of transfused per patient, the data has shown that their is no significant relationship. The median number of units of RBC blood transfused was not different between RBC age groups. This result is in direct contradiction to the previously published paper by Cata et al. (2011).[3] The previous study was predicated on the idea that the RBC age groups containd pairwise differences between groups. Note that the previous study did not correct for multiple post-hoc comparisons. This analysis did attempt to reproduce the previous, and the result was an identical p-value using the same Kruskal-Wallis test without multiple comparisons or a correction.

The permutaion test performed in this analysis exmaines the robustness of the median as an estimator. Further studies should examine the mean and the mean of the log transformed number of units as an estimator. The median cannot be assumed to come from a normal distribution, however it is likely that a permutation test of the mean would reveal that the mean number of units of RBC blood does in fact come from a normal distribution. Treating the RBC blood age as groups may also be limiting our study. Future studies should examing RBC blood age as a continuous variable and reinvestigate its relationship with number of units tranfused, and in turn exmine the relationship with Prostate cancer recurrence.

This study is limited by the strict definitions of the age groups for RBC blood. With different age groups, there may be a more significant relationship. Additionally, the insignifcant p-value reported for between group differences, between older and younger blood is statistically insignificant, however it should be considered that it may be clinically meaningful. The reported p-value likely carries moderate clinical significance. Meaning that while it cannot be stated statistically that Older blood is associated with a different number of transfusions that younger blood, that should be taken into account when dealing with patients recieving RBC blood transfusions after a laparoscopic radical prostatectomy.

References

1. Han M, Partin AW, Pound CR, Epstein JI, Walsh PC. Long-term biochemical disease-free and cancer-specific survival following anatomic radical retropubic prostatectomy: the 15-year Johns Hopkins experience. Urol Clin North Am. 2001;28(3):555-565.

2. Koch CG, Li L, Sessler DI, et al. Duration of red-cell storage and complications after cardiac surgery. N Engl J Med. 2008;358(12):1229-1239.

3. Cata PC, Klein EA, Hoeltge GA, Dalton JE, Mascha E, O'Hara J, Russell A, Kurz A, Ben-Elihayhu S, Sessler DI. Blood Storage Duration and Biochemical Recurrence of Cancer After Radical Prostatectomy. MDMayo Clin Proc. 2011;86(2):120-127

4. Haynes W in: Dubitzky W, Wolkenhauer O, Cho KH, Yokota H. Bonferroni Correction. Encyclopeda of Systems Biology. 2013;154:154-155

Appendix

```
# import
library(DescTools)
library(dplyr)
library(furniture)
library(ggplot2)
library(kableExtra)

blood = readRDS(file = "C:/Users/froelijo/dev/methods1/final/data/blood.rds")

# prepare data for table 1
blood_tableone = blood
blood_tableone$rbcagegroup <- as.factor(blood_tableone$rbcagegroup)
levels(blood_tableone$rbcagegroup) = c(
  `1` = "< 13 Days",
  `2` = "13 - 18 Days",
  `3` = "> 18 Days"
)

blood_tableone$aa <- as.factor(blood_tableone$aa)
levels(blood_tableone$aa) = c(
  `0` = "Not African American",
  `1` = "African American"
)

blood_tableone$famhx <- as.factor(blood_tableone$famhx)
levels(blood_tableone$famhx) = c(
  `0` = "No",
  `1` = "Yes"
)

table_one = table1(
  blood_tableone,
  factor(aa),
  factor(famhx),
  age,
  units,
  splitby = ~ rbcagegroup,
  var_names = c(
    aa = "Race",
    famhx = "Family history of disease",
    age = "Age [Median(SD)]",
    units = "Number of allogeneic units [Median(SD)]"
  ),
  FUN = function(x) {
    paste0(round(median(x, na.rm=FALSE), 1), " (", round(sd(x, na.rm=FALSE), 1), ")")
  },
  digits = 1,
  total = TRUE,
  output = "latex"
)

# table 1
```

```r
table_one %>%
  row_spec(c(2, 5, 8, 10), bold = T, color = "white", background = "gray")

# Mimmic Paper, p-value is the same
test = kruskal.test(units ~ rbcagegroup, data = blood)

# Post Hoc Comparisons
comparisons = DunnTest(units ~ rbcagegroup, data = blood)

# Adjust for mulitple comparisons
p = p.adjust(comparisons[[1]][,2], method = "bonferroni")

# check medians to compare with paper
young_median = median(blood[blood$rbcagegroup == 1,]$units)
middle_median = median(blood[blood$rbcagegroup == 2,]$units)
old_median = median(blood[blood$rbcagegroup == 3,]$units)

# prepare results for kable
comparisons = data.frame(cbind(comparisons[[1]], p))
rownames(comparisons) = c("Middle-Younger", "Older-Younger", "Older-Middle")

# table 2
kable(
  comparisons,
  col.names = c("Point Estimate", "P-value", "Adjusted P-value")
) %>%
  kable_styling(position = "center")

# figure 2
ggplot(blood_tableone, aes(x = rbcagegroup, y = units, fill = factor(rbcagegroup))) +
  geom_boxplot() +
  ggtitle("Boxplot of Blood Age Groups, by Transfused Allogeneic RBC Units") +
  ylab("Number of allogeneic units") +
  xlab(NULL) +
  guides(fill = guide_legend(title = "RBC Age Group")) +
  geom_hline(
    yintercept = median(blood$units),
    linetype = "dashed",
    color = "orange",
    size = 1.5
  ) +
  annotate("text", x = 3.3, y = 2.7, label = "Median = 2", color = "Orange", size = 3.5)

# Permutation Test
n = dim(blood)[1]
n1 = length(blood[blood$rbcagegroup == 1,])
n2 = length(blood[blood$rbcagegroup == 2,])
n3 = length(blood[blood$rbcagegroup == 3,])

# length
b = 10000

# structure for result storage
```

```r
result_mat = matrix(NA, nrow = b, ncol = 3)

# reproducibility (Jenny don't lose that!)
set.seed(8675309)

for(i in 1:b) {
  # sample same length as groups without replacement
  group1 = sample(n, n1, replace = FALSE)
  group2 = sample(n, n2, replace = FALSE)
  group3 = sample(n, n3, replace = FALSE)

  # store results in previously defined structure
  result_mat[i, 1] = mean(blood[group1,]$units)
  result_mat[i, 2] = mean(blood[group2,]$units)
  result_mat[i, 3] = mean(blood[group3,]$units)
}

# prepare results for Kruskal-Wallis test
blood_new = cbind(
  c(
    rep(1, length( result_mat[,1])),
    rep(2, length( result_mat[,2])),
    rep(3, length( result_mat[,1]))
  ),
  c(result_mat[, 1], result_mat[, 2], result_mat[, 3])
)

colnames(blood_new) <- c("rbcagegroup", "units")

# run Kruskal-Wallis test for comparison
kruskal.test(units ~ rbcagegroup, blood_new)

# between group comparisons, and p-value adjustment
comparisons_new = DunnTest(units ~ rbcagegroup, data = blood_new)
p_new = p.adjust(comparisons_new[[1]][,2], method = "bonferroni")
```