BIOST 536 Homework 1

Benjamin Stan

October 9, 2021

Question 1

The table showing summary statistics by treatment group in the study is shown below.

Table 1: Descriptive statistics for both treatment groups and study population overall.

	Daunorubicin	Idarubicin	Overall
Number of subjects	65	65	130
Number of male subjects	35	30	65
Number of female subjects	35	30	65
Minimum age	19	17	17
Maximum age	60	61	61
Mean age	39.8	38.0	38.9
Mean Karnofsky score	79.5	79.5	79.5
SD Karnofsky score	12.6	11.6	12.1
Mean baseline white blood cells ($10^3/mm^3$)	43.3	29.0	36.1
SD baseline white blood cells ($10^3/mm^3$)	55.0	36.3	46.9
Mean baseline platelets ($10^3/mm^3$)	93.6	66.6	80.0
SD baseline platelets ($10^3/mm^3$)	92.4	57.8	77.8
Mean baseline hemoglobin (g/dl)	9.64	9.22	9.4
SD baseline hemoglobin (g/dl)	1.49	1.82	1.7

Question 2

The 2x2 table for males:

	Daunorubicin ($\bar{E}$)	Idarubicin (E)	Total
Complete Remission (D)	18	9	27
Not Complete Remission ($\bar{D}$)	17	21	38
Total	35	30	

The 2x2 table for females:

	Daunorubicin ($\bar{E}$)	Idarubicin (E)	Total
Complete Remission (D)	9	5	14
Not Complete Remission ($\bar{D}$)	21	30	51
Total	30	35	

Question 3

Based on the results of a logistic regression, the odds of complete remission are 2.59 times greater (95% robust CI: 1.20-5.59) for a population taking Idarubicin compared to a population taking Daunorubicin. The summary measure that most AML patients would be interested in would be relative risk (RR) because this measure is the most intuitive to understand. It presents the data in the form of a multiplier on a baseline probability.

Question 4

Both sexes had an equal number of subjects in this trial, 65. For males, 58% of patients achieved complete remission, compared 78% of females. However, as the primary relationship of interest in this analysis is that between the treatment and remission status, the sex of the patient would also have to affect the treatment assignment for it to be a confounder. As the treatment assignment was random, and males and females have equal numbers of subjects in each treatment arm, it can be determined that the treatment effect is likely not confounded by sex.

Question 5

Based on the results of a logistic regression model, the odds of complete remission are 2.51 times greater (95% robust CI: 1.14-5.51) for a population taking Idarubicin compared to a population of the same sex taking Daunorubicin.

Question 6

Based on the results of a logistic regression, the odds of complete remission are 2.47 times greater (95% robust CI: 0.89-6.88) for a population of males taking Idarubicin compared to a population of males taking Daunorubicin.

Question 7

Based on the results of a logistic regression, the odds of complete remission are 2.57 times greater (95% robust CI: 0.75-8.77) for a population of females taking Idarubicin compared to a population of females taking Daunorubicin.

Question 8

(a) The population attributable risk (as given in Lecture 1) as a function of the rate of exposure, P[E], and the relative risk of disease, RR, is shown below:

$$PAR = \frac{P[E](RR-1)}{P[E]*RR+1-P[E]}$$

(b) Using the equation above, the estimate PAR can be computed as follows:

$$PAR = \frac{P[E](RR-1)}{P[E]*RR+1-P[E]} = \frac{0.35*(22-1)}{0.35*22+1-0.35} = 0.88$$

It can be determined that the 88% of the overall risk of fatal lung cancer is due to smoking in this population.

(c) Using the same equation, the estimated PAR can be computed as follows:

$$PAR = \frac{P[E](RR-1)}{P[E]*RR+1-P[E]} = \frac{0.05*(22-1)}{0.05*22+1-0.35} = 0.60$$

It can be determined that the 60% of the overall risk of fatal lung cancer is due to smoking in this population.

Appendix:

```
## Load libraries and set working directory
setwd("/Users/bstan/Documents/UW/Courses/BIOST 536")
rm(list = ls())
library(tidyverse)
library(tidyr)
library(tinytex)
library("sandwich")
leukemia <- read_csv("data/leukemia_data.csv")

## Question 1
### Create summary statistics by treatment group
table1 <- leukemia %>%
  group_by(tx) %>% # idarubicin or daunorubicin
  summarise(n = n(),
            n_male = sum(ifelse(sex=="M",1,0)),
            n_female = sum(ifelse(sex=="M",1,0)),
            min_age = min(age, na.rm=T),
            max_age = max(age, na.rm=T),
            #p50_age = quantile(age, 0.5, na.rm=T),
            mean_age = mean(age, na.rm=T),
            mean_Karnofsky_score = mean(karn, na.rm=T),
            sd_Karnofsky_score = sd(karn, na.rm=T),
            mean_baseline_white_blood_cells = mean(wbc, na.rm=T),
            sd_baseline_white_blood_cells = sd(wbc, na.rm=T),
            mean_baseline_platelets = mean(plt, na.rm=T),
            sd_baseline_platelets = sd(plt, na.rm=T),
            mean_baseline_hemoglobin = mean(hgb, na.rm=T),
            sd_baseline_hemoglobin = sd(hgb, na.rm=T),
            )
### Create summary statistics for all
table2 <- leukemia %>%
```

```
  summarise(n = n(),
            n_male = sum(ifelse(sex=="M",1,0)),
            n_female = sum(ifelse(sex=="M",1,0)),
            min_age = min(age, na.rm=T),
            max_age = max(age, na.rm=T),
            #p50_age = quantile(age, 0.5, na.rm=T),
            mean_age = mean(age, na.rm=T),
            mean_Karnofsky_score = mean(karn, na.rm=T),
            sd_Karnofsky_score = sd(karn, na.rm=T),
            mean_baseline_white_blood_cells = mean(wbc, na.rm=T),
            sd_baseline_white_blood_cells = sd(wbc, na.rm=T),
            mean_baseline_platelets = mean(plt, na.rm=T),
            sd_baseline_platelets = sd(plt, na.rm=T),
            mean_baseline_hemoglobin = mean(hgb, na.rm=T),
            sd_baseline_hemoglobin = sd(hgb, na.rm=T),
            )
t(table2)

## Question 2
### Males
male_table <- table(leukemia %>% filter(sex =="M") %>% select(cr,tx))
rownames(male_table) <- c("Complete_Remission"," Not Complete Remission")
knitr::kable(male_table,
             col.names = c("Daunorubicin","Idarubicin"))

### Females
female_table <- table(leukemia %>% filter(sex =="F") %>% select(cr,tx))
rownames(female_table) <- c("Complete_Remission"," Not Complete Remission")
knitr::kable(female_table,
             col.names = c("Daunorubicin","Idarubicin"))

## Question 3
leukemia <- leukemia %>% mutate(cr_bin = ifelse(cr=="Y",1,0))
glm1 <- glm(cr_bin ~ tx, data = leukemia, family = "binomial")
glm1$coef %>% exp

coef1 <- glm1$coef
normal_se1 <- summary(glm1)$coefficients[, 2]
rob_se1 <- sqrt(diag(vcovHC(glm1, type = "HC0")))
conf_int_tx1 <- coef1[2] + c(0, qnorm(c(0.025, 0.975))) * rob_se1[2]
conf_int_tx1 %>% exp

## Question 4
knitr::kable(table(leukemia$sex,leukemia$cr))

## Question 5
glm2 <- glm(cr_bin ~ tx + sex, data = leukemia, family = "binomial")
glm2$coef %>% exp

coef2 <- glm2$coef
normal_se2 <- summary(glm2)$coefficients[, 2]
rob_se2 <- sqrt(diag(vcovHC(glm2, type = "HC0")))
conf_int_tx2 <- coef2[2] + c(0, qnorm(c(0.025, 0.975))) * rob_se2[2]
```

```r
conf_int_tx2 %>% exp

## Question 6
leukemia_m <- leukemia %>% filter(sex=="M")
glm3 <- glm(cr_bin ~ tx, data = leukemia_m, family = "binomial")
glm3$coef %>% exp

coef3 <- glm3$coef
normal_se3 <- summary(glm3)$coefficients[, 2]
rob_se3 <- sqrt(diag(vcovHC(glm3, type = "HC0")))
conf_int_tx3 <- coef3[2] + c(0, qnorm(c(0.025, 0.975))) * rob_se3[2]
conf_int_tx3 %>% exp

## Question 7
leukemia_f <- leukemia %>% filter(sex=="F")
glm4 <- glm(cr_bin ~ tx, data = leukemia_f, family = "binomial")
glm4$coef %>% exp

coef4 <- glm4$coef
normal_se4 <- summary(glm4)$coefficients[, 2]
rob_se4 <- sqrt(diag(vcovHC(glm4, type = "HC0")))
conf_int_tx4 <- coef4[2] + c(0, qnorm(c(0.025, 0.975))) * rob_se4[2]
conf_int_tx4 %>% exp
```