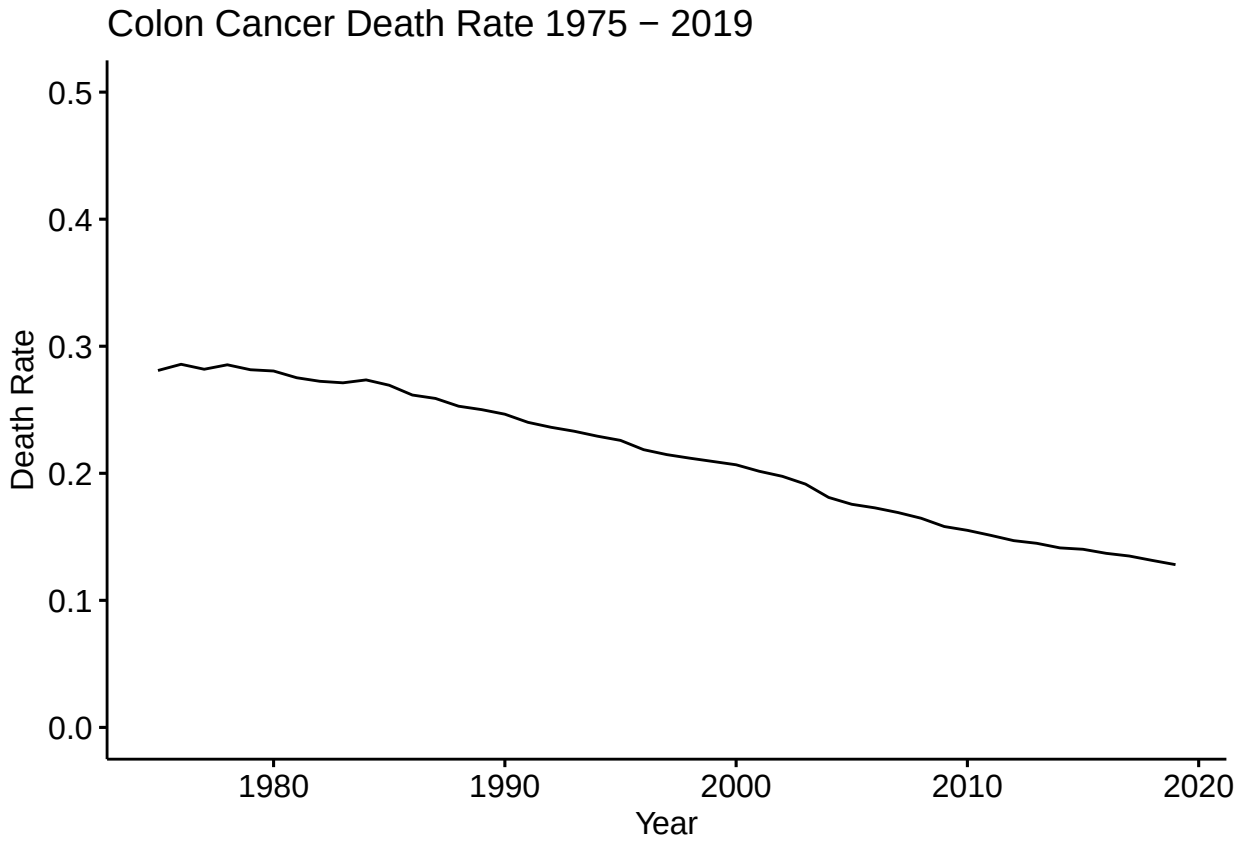
Introdution

Before the development of effective chemotherapy for colon cancer, treatment was limited to surgery. Recurrence was frequent and life-expectancy after initial treatment was not long.

However, from 1985 to 1990, adjuvant chemotherapy began to be developed and implemented in combination with surgery.

After simultaneous advancements in detection and screening, death rates due to colon cancer began to fall: a 54% reduction.

```
library(ggpubr)
# Colon Cancer death rate in the US over time.
cancerData <- read.table("ColonCancerDeathRate.txt", header=TRUE)
cancerData$Prop <- cancerData$Prop / 100
g <- ggline(cancerData, x="Year", y="Prop",
       plot_type="l",
       title="Colon Cancer Death Rate 1975 - 2019",
       xlab="Year",
       ylab="Death Rate")
ggpar(g, ylim=c(0,0.50))
```



Today, excluding skin cancer, colon cancer is the third most common type of cancer in the U.S., but there are now more treatments available, including immunotherapy and radiation therapy. In this study, we will examine the outcomes of a clinical Colon Cancer study that measured the effects of two treatment groups and a control.

Data Source

This data set comes from one of the first successful trials of adjuvant chemotherapy for colon cancer, using a combination of levamisole and fluorouracil.

The trial examined both Stage B2 (locally invasive) and Stage C (regional nodal involvement) colon cancer, but this dataset contains information only on subjects with Stage C.

The data set had a complicated format. For each subject, there were two rows of data entry: one measuring data related to the first recurrence of cancer, such as days until recurrence, and the other measuring data related to death, such as presence of cancer in other regions. The rows were categorized by the "etype" nominal variable, which we used to sort the data.

In addition, each row had a "status" nominal variable, that indicated whether the etype was positive or negative – whether there was or was not recurrence, and whether or not the subject had died. This required additional sorting of the data.

Each subject in the study was assigned a treatment – observation, levamisole (alone), or levamisole and fluorouracil (combined) – and this was recorded by the "rx" nominal variable for each row of data, allowing us to sort, analyze, and compare the effectiveness of the different treatments.

Import Data

The below code imports the data sets. Click show to view it.

```
colonData <- read.table("colon.txt", header = TRUE)

# Grabs Recurrence Data from colonData
studyA <- subset(colonData, etype==1)

# Grabs Death Data from colonData
studyB <- subset(colonData, etype==2)

# Add year column
studyA$years <- studyA$time / 365.25
studyB$years <- studyB$time / 365.25

# Treatment groups
obsA <- subset(studyA, rx=="Obs")
levA <- subset(studyA, rx=="Lev")
combA <- subset(studyA, rx=="Lev+5FU")

obsB <- subset(studyB, rx=="Obs")
levB <- subset(studyB, rx=="Lev")
combB <- subset(studyB, rx=="Lev+5FU")

# Convert treatment to number
studyA$xnum <- studyA$rx
studyB$xnum <- studyB$rx

studyA$xnum[studyA$xnum == "Obs"] <- 0
studyA$xnum[studyA$xnum == "Lev"] <- 1
studyA$xnum[studyA$xnum == "Lev+5FU"] <- 2

studyB$xnum[studyB$xnum == "Obs"] <- 0
studyB$xnum[studyB$xnum == "Lev"] <- 1
studyB$xnum[studyB$xnum == "Lev+5FU"] <- 2
```

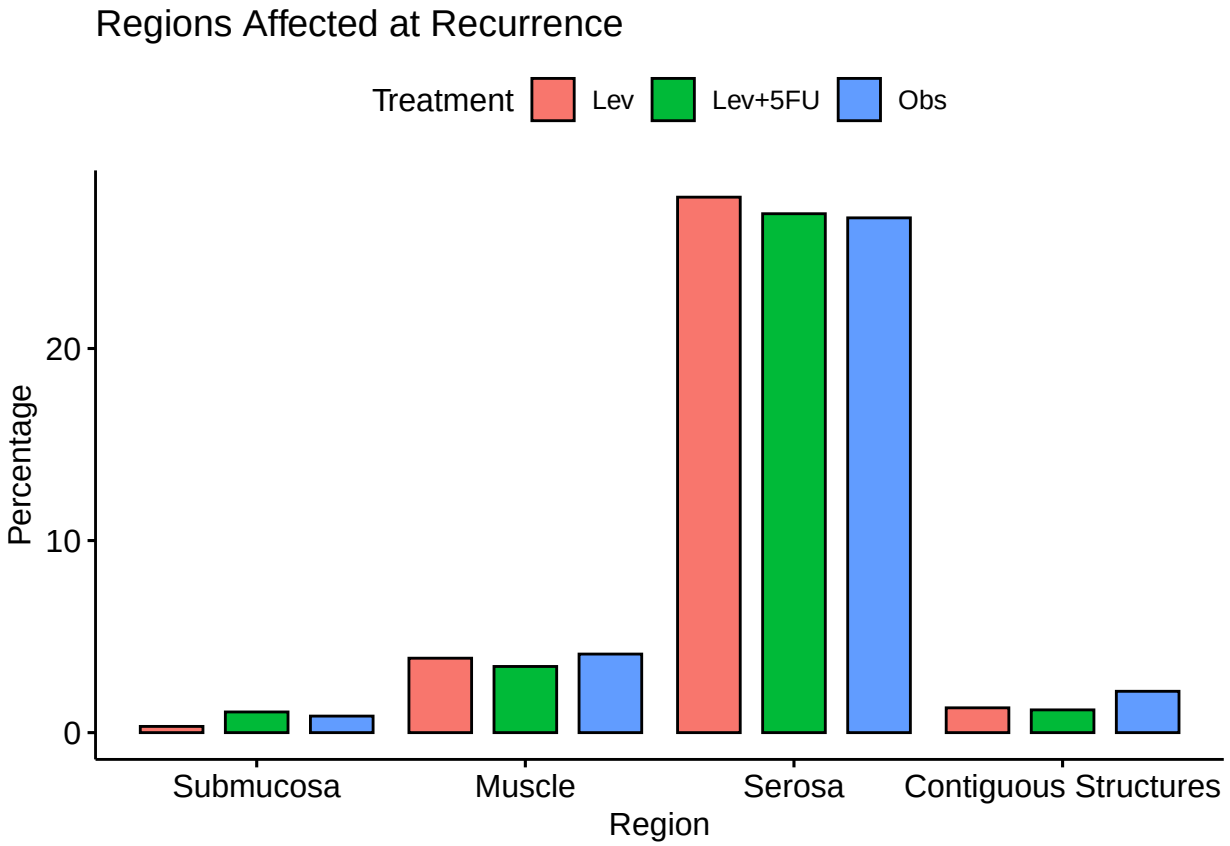
Description of Variables

We first performed exploratory data analysis to see which variables were affected by treatment. Most variables were not affected by treatment - only the "nodes" variable and the "time" variable were affected by treatment.

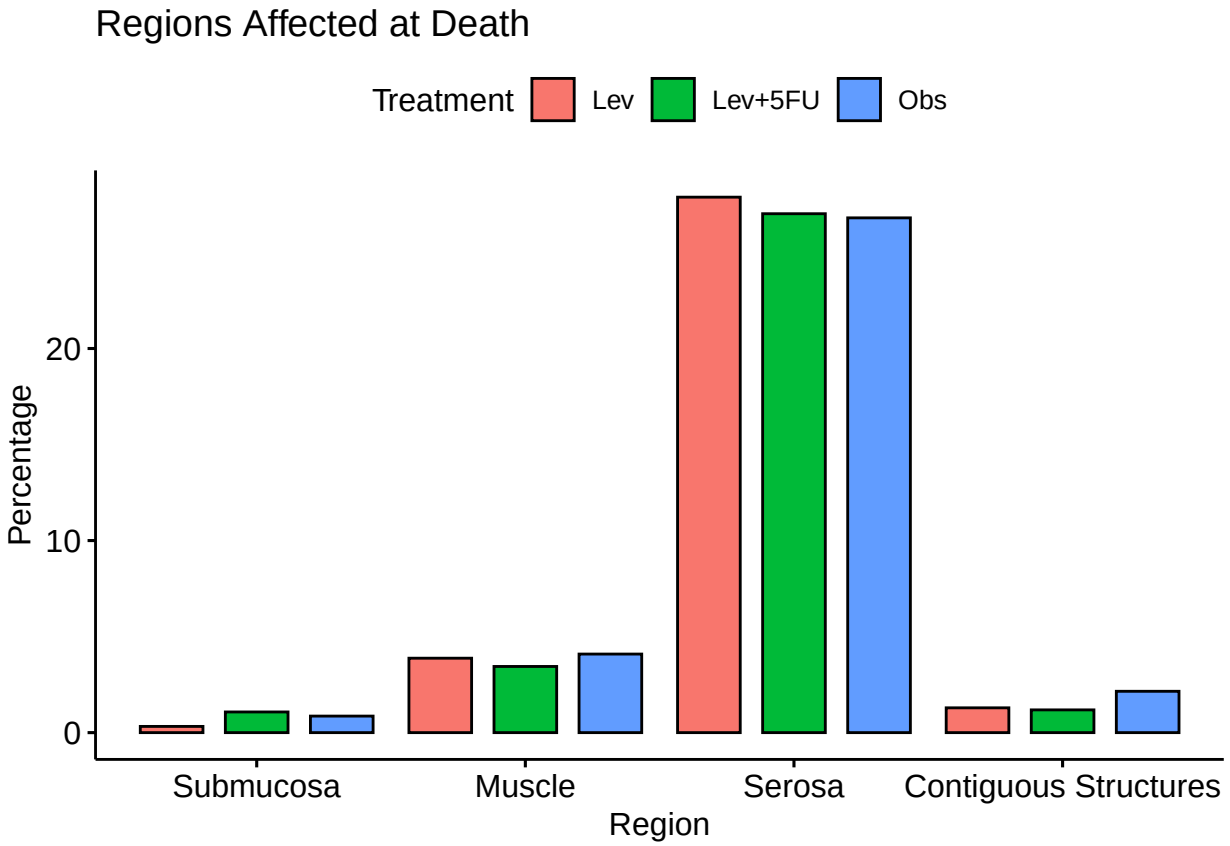
Extent of Local Cancer Spread

The nominal variable "extent", which recorded whether the cancer had spread to specific nearby regions, was not affected by treatment:

```
library(ggpubr)
submucosaAobs <- sum(studyA["rx"] == "Obs" & studyA["extent"] == "1")*100/sum(studyA$study==1)
submucosaAlev <- sum(studyA["rx"] == "Lev" & studyA["extent"] == "1")*100/sum(studyA$study==1)
submucosaAlev5FU <- sum(studyA["rx"] == "Lev+5FU" & studyA["extent"] == "1")*100/sum(studyA$study==1)
muscleAobs <- sum(studyA["rx"] == "Obs" & studyA["extent"] == "2")*100/sum(studyA$study==1)
muscleAlev <- sum(studyA["rx"] == "Lev" & studyA["extent"] == "2")*100/sum(studyA$study==1)
muscleAlev5FU <- sum(studyA["rx"] == "Lev+5FU" & studyA["extent"] == "2")*100/sum(studyA$study==1)
serosaAobs <- sum(studyA["rx"] == "Obs" & studyA["extent"] == "3")*100/sum(studyA$study==1)
serosaAlev <- sum(studyA["rx"] == "Lev" & studyA["extent"] == "3")*100/sum(studyA$study==1)
serosaAlev5FU <- sum(studyA["rx"] == "Lev+5FU" & studyA["extent"] == "3")*100/sum(studyA$study==1)
contiguousAobs <- sum(studyA["rx"] == "Obs" & studyA["extent"] == "4")*100/sum(studyA$study==1)
contiguousAlev <- sum(studyA["rx"] == "Lev" & studyA["extent"] == "4")*100/sum(studyA$study==1)
contiguousAlev5FU <- sum(studyA["rx"] == "Lev+5FU" & studyA["extent"] == "4")*100/sum(studyA$study==1)
extentdf <- data.frame(Region = rep(c("Submucosa", "Muscle", "Serosa", "Contiguous Structures"), each=3
                       Treatment = rep(c("Obs", "Lev", "Lev+5FU"),4),
                       Percentage = c(submucosaAobs, submucosaAlev, submucosaAlev5FU,
                                      muscleAobs, muscleAlev, muscleAlev5FU,
                                      serosaAobs, serosaAlev, serosaAlev5FU,
                                      contiguousAobs, contiguousAlev, contiguousAlev5FU))
ggbarplot(extentdf, "Region", "Percentage", fill="Treatment", position=position_dodge(), title = "Regio
```



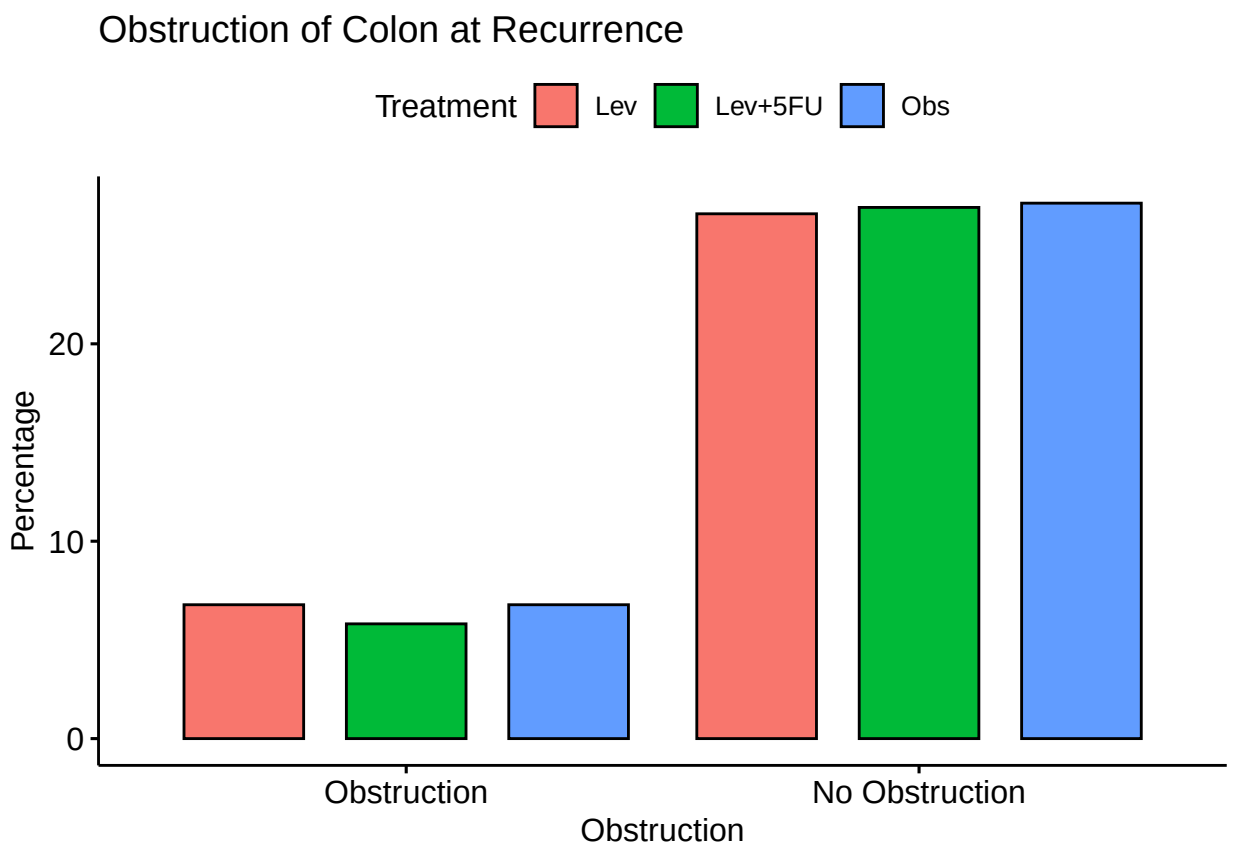

```
submucosaBobs <- sum(studyB["rx"] == "Obs" & studyB["extent"] == "1")*100/sum(studyB$study==1)
submucosaBlev <- sum(studyB["rx"] == "Lev" & studyB["extent"] == "1")*100/sum(studyB$study==1)
submucosaBlev5FU <- sum(studyB["rx"] == "Lev+5FU" & studyB["extent"] == "1")*100/sum(studyB$study==1)
muscleBobs <- sum(studyB["rx"] == "Obs" & studyB["extent"] == "2")*100/sum(studyB$study==1)
muscleBlev <- sum(studyB["rx"] == "Lev" & studyB["extent"] == "2")*100/sum(studyB$study==1)
muscleBlev5FU <- sum(studyB["rx"] == "Lev+5FU" & studyB["extent"] == "2")*100/sum(studyB$study==1)
serosaBobs <- sum(studyB["rx"] == "Obs" & studyB["extent"] == "3")*100/sum(studyB$study==1)
serosaBlev <- sum(studyB["rx"] == "Lev" & studyB["extent"] == "3")*100/sum(studyB$study==1)
serosaBlev5FU <- sum(studyB["rx"] == "Lev+5FU" & studyB["extent"] == "3")*100/sum(studyB$study==1)
contiguousBobs <- sum(studyB["rx"] == "Obs" & studyB["extent"] == "4")*100/sum(studyB$study==1)
contiguousBlev <- sum(studyB["rx"] == "Lev" & studyB["extent"] == "4")*100/sum(studyB$study==1)
contiguousBlev5FU <- sum(studyB["rx"] == "Lev+5FU" & studyB["extent"] == "4")*100/sum(studyB$study==1)
extentBdf <- data.frame(Region = rep(c("Submucosa", "Muscle", "Serosa", "Contiguous Structures"), each=3
                        Treatment = rep(c("Obs", "Lev", "Lev+5FU"),4),
                        Percentage = c(submucosaBobs, submucosaBlev, submucosaBlev5FU, muscleBobs, muscl
                                       serosaBobs, serosaBlev, serosaBlev5FU, contiguousBobs, contiguous
ggbarplot(extentBdf, "Region", "Percentage", fill="Treatment", position=position_dodge(), title = "Regi
```



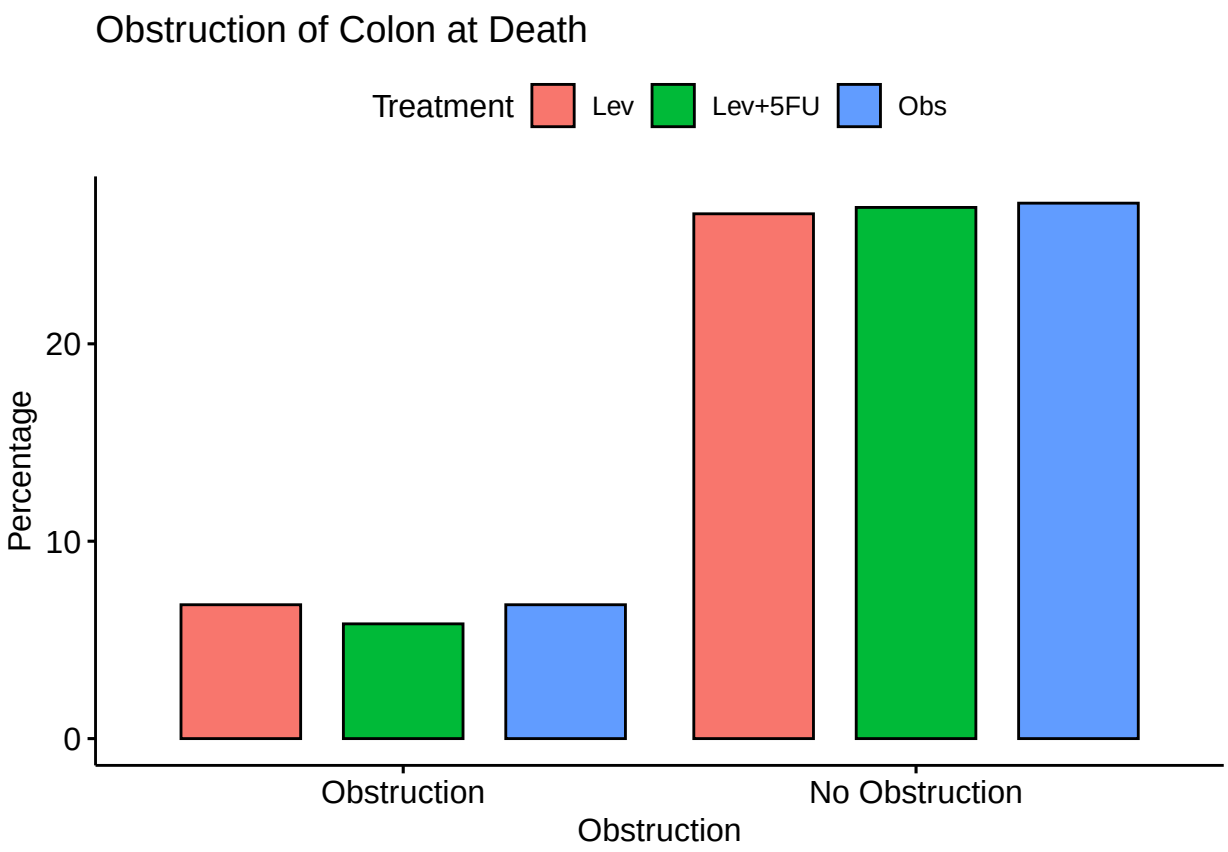
Obstruction of Colon by Tumour

The nominal variable "obstruct", which specified whether the colon was obstructed by the cancerous tumor, was not affected by treatment:

```
obstructAobs <- sum(studyA["rx"] == "Obs" & studyA["obstruct"] == "1")*100/sum(studyA$study==1)
obstructAlev <- sum(studyA["rx"] == "Lev" & studyA["obstruct"] == "1")*100/sum(studyA$study==1)
obstructAlev5FU <- sum(studyA["rx"] == "Lev+5FU" & studyA["obstruct"] == "1")*100/sum(studyA$study==1)
noobstructAobs <- sum(studyA["rx"] == "Obs" & studyA["obstruct"] == "0")*100/sum(studyA$study==1)
noobstructAlev <- sum(studyA["rx"] == "Lev" & studyA["obstruct"] == "0")*100/sum(studyA$study==1)
noobstructAlev5FU <- sum(studyA["rx"] == "Lev+5FU" & studyA["obstruct"] == "0")*100/sum(studyA$study==1
obstructAdf <- data.frame(Obstruction = rep(c("Obstruction", "No Obstruction"), each=3),
                          Treatment = rep(c("Obs", "Lev", "Lev+5FU"),2),
                          Percentage = c(obstructAobs, obstructAlev, obstructAlev5FU,
                                         noobstructAobs, noobstructAlev, noobstructAlev5FU))
ggbarplot(obstructAdf, "Obstruction", "Percentage", fill="Treatment", position=position_dodge(), title =
```



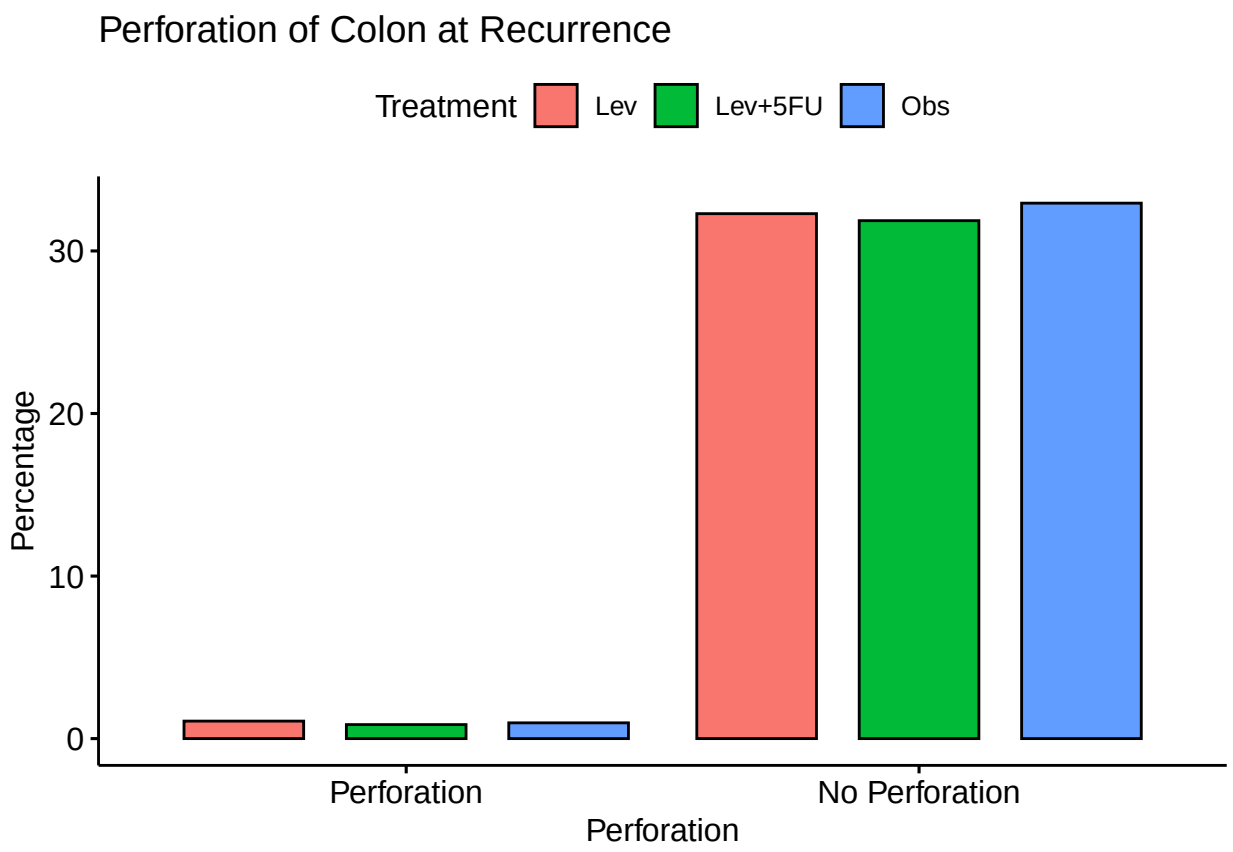

```
obstructBobs <- sum(studyB["rx"] == "Obs" & studyB["obstruct"] == "1")*100/sum(studyB$study==1)
obstructBlev <- sum(studyB["rx"] == "Lev" & studyB["obstruct"] == "1")*100/sum(studyB$study==1)
obstructBlev5FU <- sum(studyB["rx"] == "Lev+5FU" & studyB["obstruct"] == "1")*100/sum(studyB$study==1)
noobstructBobs <- sum(studyB["rx"] == "Obs" & studyB["obstruct"] == "0")*100/sum(studyB$study==1)
noobstructBlev <- sum(studyB["rx"] == "Lev" & studyB["obstruct"] == "0")*100/sum(studyB$study==1)
noobstructBlev5FU <- sum(studyB["rx"] == "Lev+5FU" & studyB["obstruct"] == "0")*100/sum(studyB$study==1
obstructBdf <- data.frame(Obstruction = rep(c("Obstruction", "No Obstruction"), each=3),
                          Treatment = rep(c("Obs", "Lev", "Lev+5FU"),2),
                          Percentage = c(obstructBobs, obstructBlev, obstructBlev5FU,
                                         noobstructBobs, noobstructBlev, noobstructBlev5FU))
ggbarplot(obstructBdf, "Obstruction", "Percentage", fill="Treatment", position=position_dodge(), title =
```



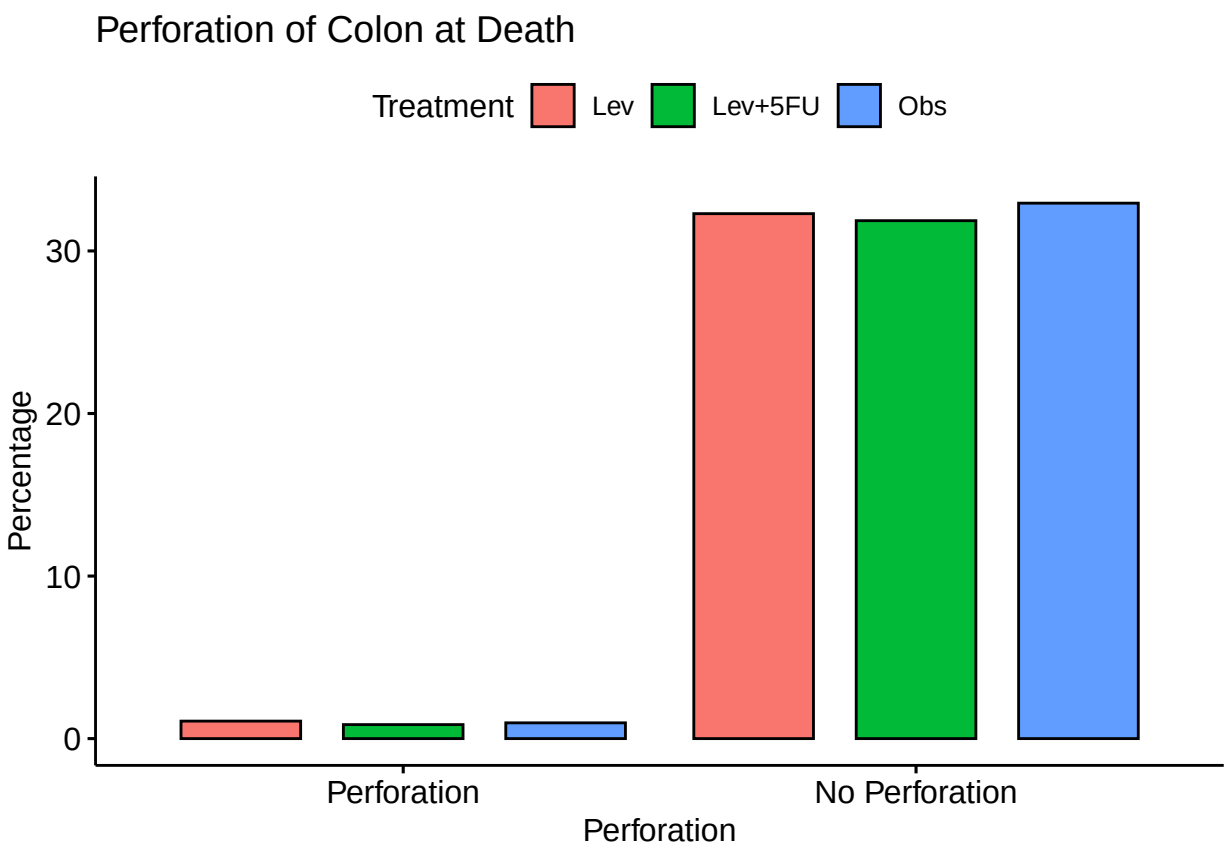
Perforation of Colon by Tumour

The nominal variable "perfor", which specified whether the colon was perforated by the cancerous tumor, was not affected by treatment:

```
perforAobs <- sum(studyA["rx"] == "Obs" & studyA["perfor"] == "1")*100/sum(studyA$study==1)
perforAlev <- sum(studyA["rx"] == "Lev" & studyA["perfor"] == "1")*100/sum(studyA$study==1)
perforAlev5FU <- sum(studyA["rx"] == "Lev+5FU" & studyA["perfor"] == "1")*100/sum(studyA$study==1)
noperforAobs <- sum(studyA["rx"] == "Obs" & studyA["perfor"] == "0")*100/sum(studyA$study==1)
noperforAlev <- sum(studyA["rx"] == "Lev" & studyA["perfor"] == "0")*100/sum(studyA$study==1)
noperforAlev5FU <- sum(studyA["rx"] == "Lev+5FU" & studyA["perfor"] == "0")*100/sum(studyA$study==1)
perforAdf <- data.frame(Perforation= rep(c("Perforation", "No Perforation"), each=3),
                        Treatment = rep(c("Obs", "Lev", "Lev+5FU"),2),
                        Percentage = c(perforAobs, perforAlev, perforAlev5FU,
                                       noperforAobs, noperforAlev, noperforAlev5FU))
ggbarplot(perforAdf, "Perforation", "Percentage", fill="Treatment", position=position_dodge(), title =
```



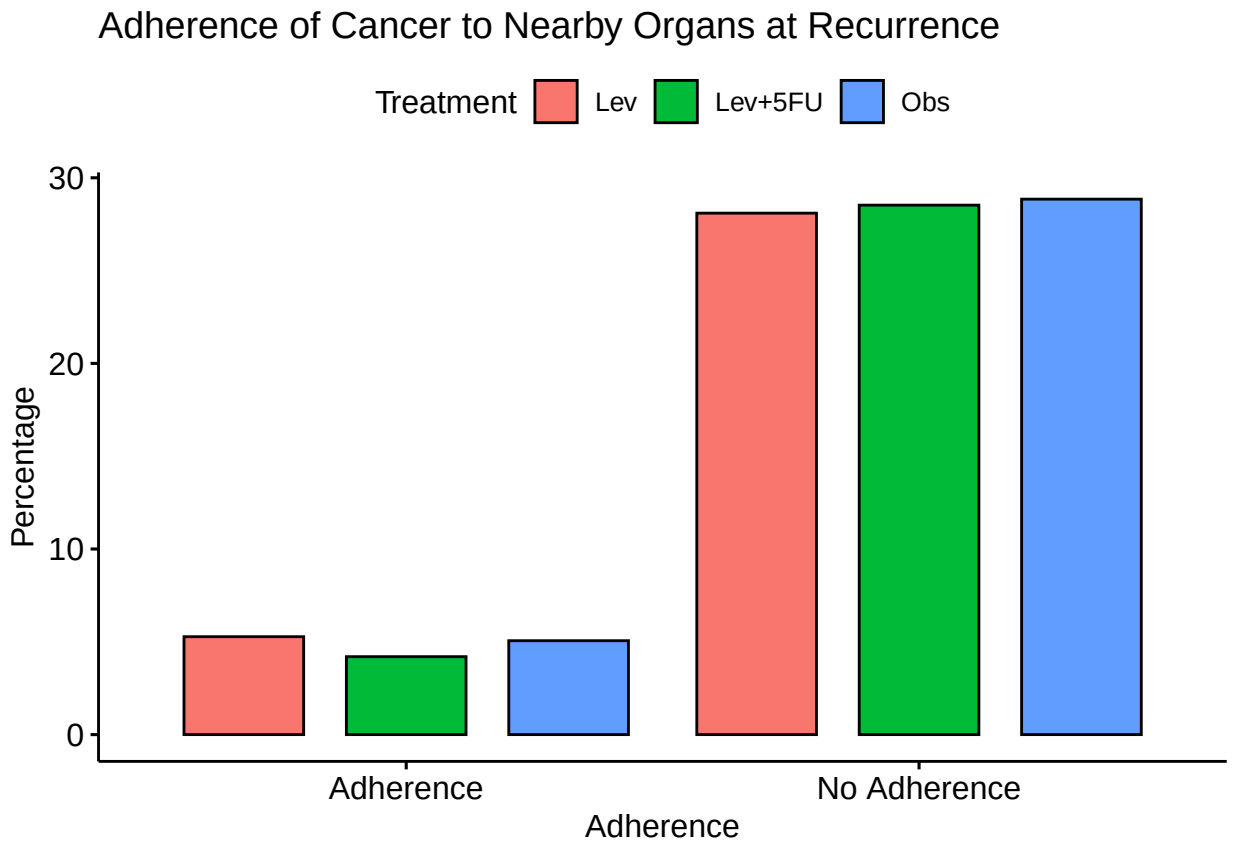

```
perforBobs <- sum(studyB["rx"] == "Obs" & studyB["perfor"] == "1")*100/sum(studyB$study==1)
perforBlev <- sum(studyB["rx"] == "Lev" & studyB["perfor"] == "1")*100/sum(studyB$study==1)
perforBlev5FU <- sum(studyB["rx"] == "Lev+5FU" & studyB["perfor"] == "1")*100/sum(studyB$study==1)
noperforBobs <- sum(studyB["rx"] == "Obs" & studyB["perfor"] == "0")*100/sum(studyB$study==1)
noperforBlev <- sum(studyB["rx"] == "Lev" & studyB["perfor"] == "0")*100/sum(studyB$study==1)
noperforBlev5FU <- sum(studyB["rx"] == "Lev+5FU" & studyB["perfor"] == "0")*100/sum(studyB$study==1)
perforBdf <- data.frame(Perforation = rep(c("Perforation", "No Perforation"), each=3),
                        Treatment = rep(c("Obs", "Lev", "Lev+5FU"),2),
                        Percentage = c(perforBobs, perforBlev, perforBlev5FU,
                                       noperforBobs, noperforBlev, noperforBlev5FU))
ggbarplot(perforBdf, "Perforation", "Percentage", fill="Treatment", position=position_dodge(), title =
```



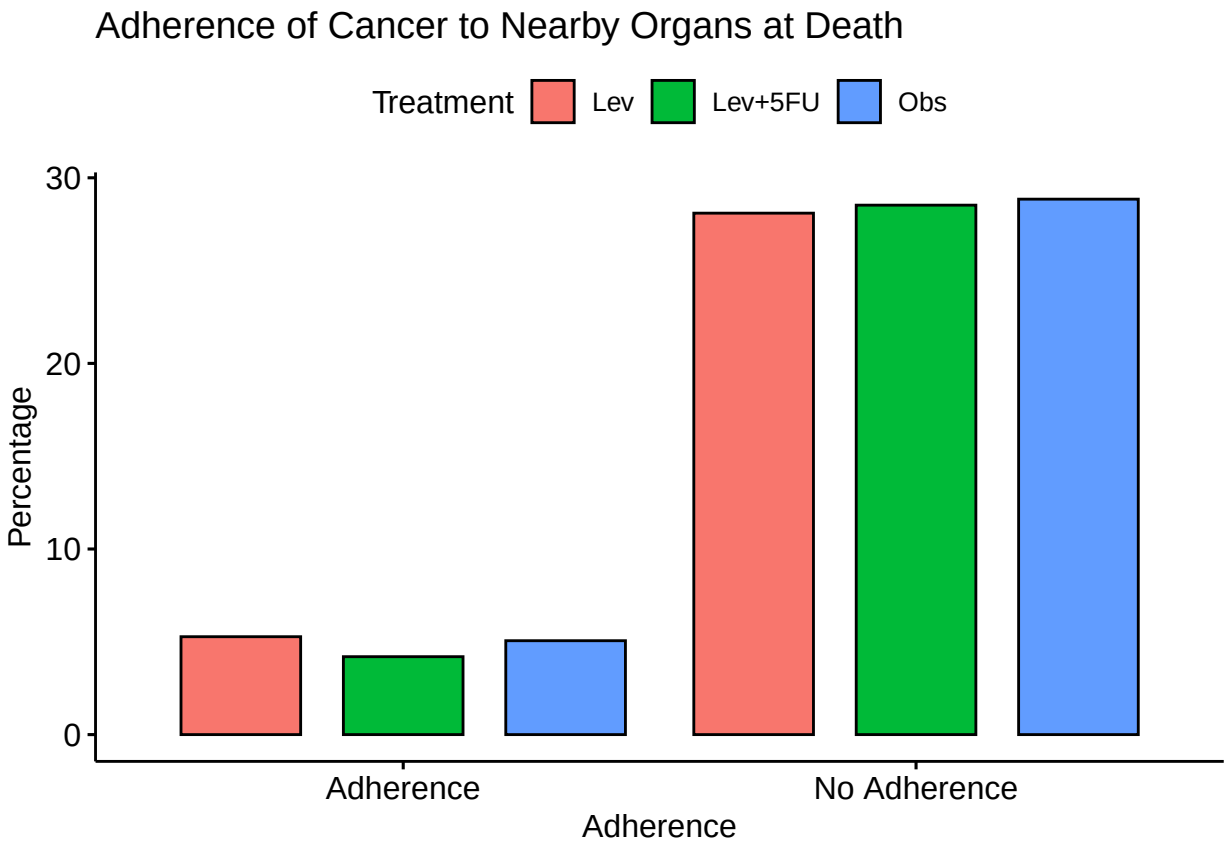
Adherence of Cancer to Nearby Organs

The nominal variable "adhere", which specified whether the cancer adhered to nearby organs, was not affected by treatment:

```
adhereAobs <- sum(studyA["rx"] == "Obs" & studyA["adhere"] == "1")*100/sum(studyA$study==1)
adhereAlev <- sum(studyA["rx"] == "Lev" & studyA["adhere"] == "1")*100/sum(studyA$study==1)
adhereAlev5FU <- sum(studyA["rx"] == "Lev+5FU" & studyA["adhere"] == "1")*100/sum(studyA$study==1)
noadhereAobs <- sum(studyA["rx"] == "Obs" & studyA["adhere"] == "0")*100/sum(studyA$study==1)
noadhereAlev <- sum(studyA["rx"] == "Lev" & studyA["adhere"] == "0")*100/sum(studyA$study==1)
noadhereAlev5FU <- sum(studyA["rx"] == "Lev+5FU" & studyA["adhere"] == "0")*100/sum(studyA$study==1)
adhereAdf <- data.frame(Adherence = rep(c("Adherence", "No Adherence"), each=3),
                        Treatment = rep(c("Obs", "Lev", "Lev+5FU"),2),
                        Percentage = c(adhereAobs, adhereAlev, adhereAlev5FU,
                                       noadhereAobs, noadhereAlev, noadhereAlev5FU))
ggbarplot(adhereAdf, "Adherence", "Percentage", fill="Treatment", position=position_dodge(), title = "A
```



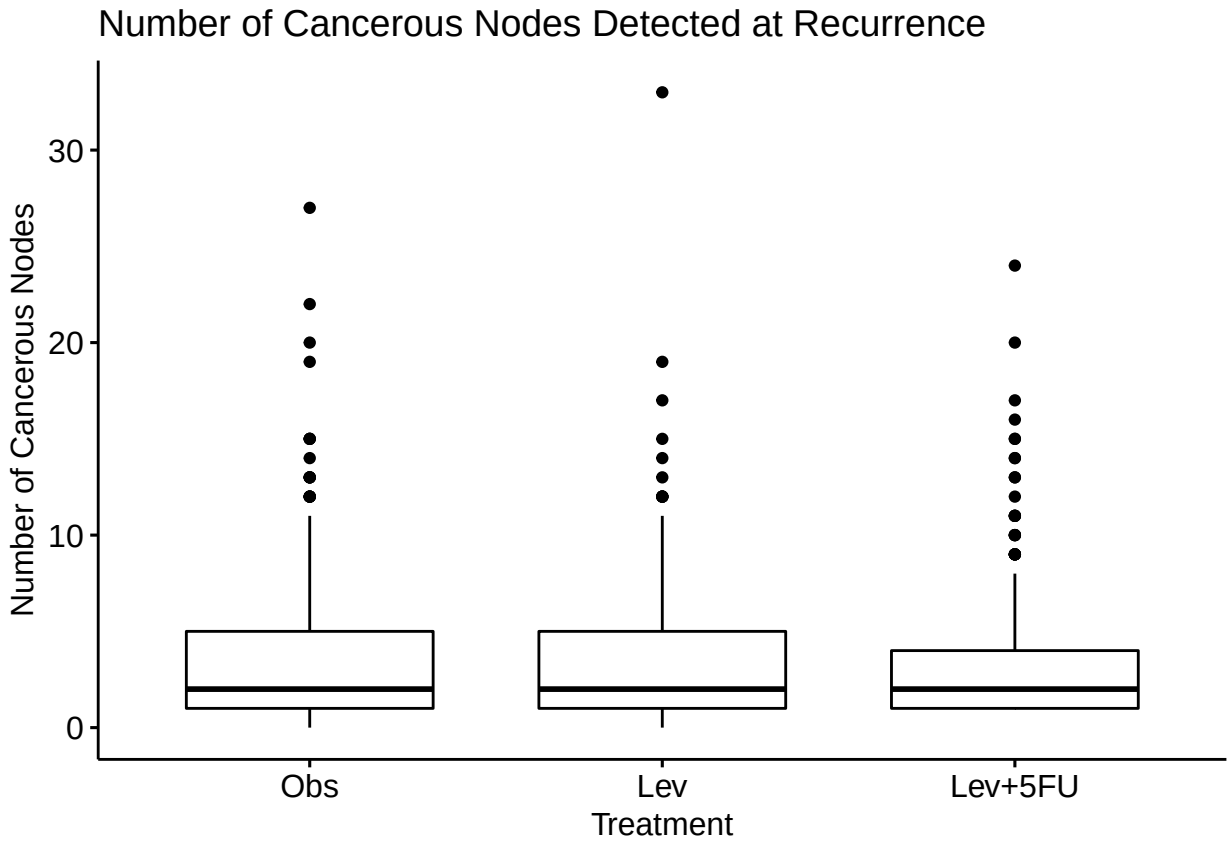

```
adhereBobs <- sum(studyB["rx"] == "Obs" & studyB["adhere"] == "1")*100/sum(studyB$study==1)
adhereBlev <- sum(studyB["rx"] == "Lev" & studyB["adhere"] == "1")*100/sum(studyB$study==1)
adhereBlev5FU <- sum(studyB["rx"] == "Lev+5FU" & studyB["adhere"] == "1")*100/sum(studyB$study==1)
noadhereBobs <- sum(studyB["rx"] == "Obs" & studyB["adhere"] == "0")*100/sum(studyB$study==1)
noadhereBlev <- sum(studyB["rx"] == "Lev" & studyB["adhere"] == "0")*100/sum(studyB$study==1)
noadhereBlev5FU <- sum(studyB["rx"] == "Lev+5FU" & studyB["adhere"] == "0")*100/sum(studyB$study==1)
adhereBdf <- data.frame(Adherence = rep(c("Adherence", "No Adherence"), each=3),
                        Treatment = rep(c("Obs", "Lev", "Lev+5FU"),2),
                        Percentage = c(adhereBobs, adhereBlev, adhereBlev5FU,
                                       noadhereBobs, noadhereBlev, noadhereBlev5FU))
ggbarplot(adhereBdf, "Adherence", "Percentage", fill="Treatment", position=position_dodge(), title = "A
```



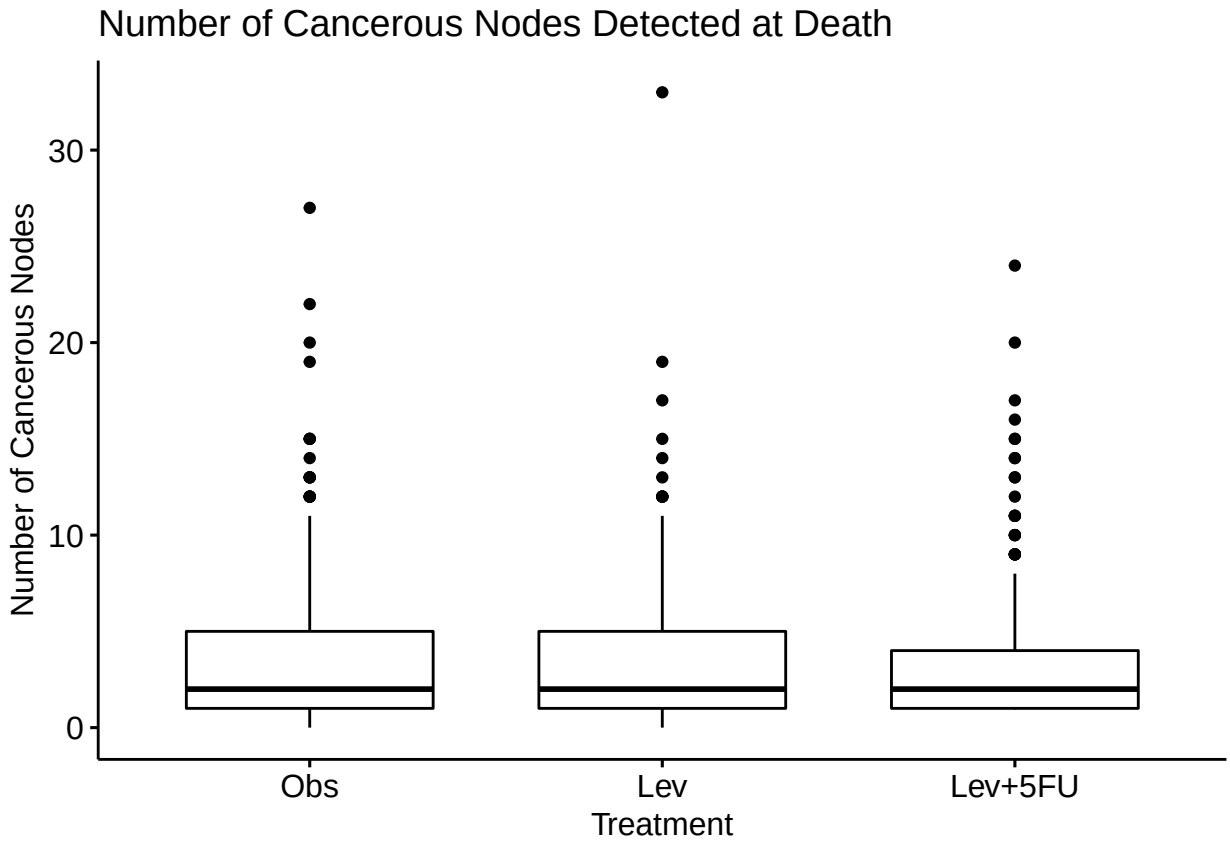
Number of Lymph Nodes with Detectable Cancer

The discrete quantitative variable "nodes", which recorded the number of lymph nodes detected with cancer, was somewhat affected by treatment:

```
ggboxplot(studyA, x = "rx", y = "nodes",
          #color = "rx", palette = c("#00AFBB", "#E7B800", "#FC4E07"),
          order = c("Obs", "Lev", "Lev+5FU"),
          ylab = "Number of Cancerous Nodes", xlab = "Treatment",
          title = "Number of Cancerous Nodes Detected at Recurrence")
```




```
ggboxplot(studyB, x = "rx", y = "nodes",
          #color = "rx", palette = c("#00AFBB", "#E7B800", "#FC4E07"),
          order = c("Obs", "Lev", "Lev+5FU"),
          ylab = "Number of Cancerous Nodes", xlab = "Treatment",
          title = "Number of Cancerous Nodes Detected at Death")
```



Thus, for the analysis, we looked at the discrete quantitative "time" variable (measured in days) on the effect of "status", to determine the effectiveness of the different treatment types and the Hazard Ratios for each treatment.

Research Questions

We sought to answer questions about the effectiveness of the different treatment types:

- Which treatment had a lower recurrence rate?
- Which treatment had a greater amount of time until recurrence
- Which treatment had a higher survival rate?
- Which treatment had a higher life expectancy?

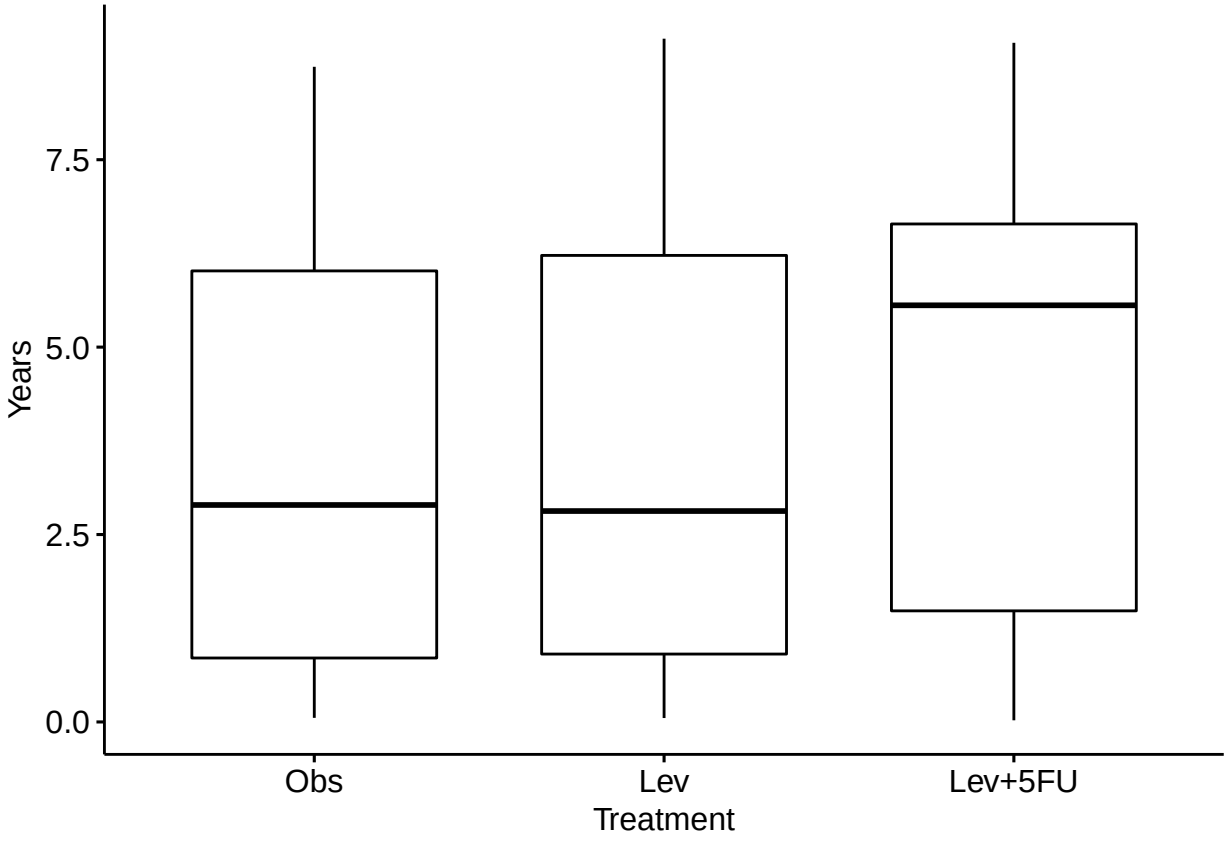
Data Exploration and Munging

Distributions of Treatments

Box plots

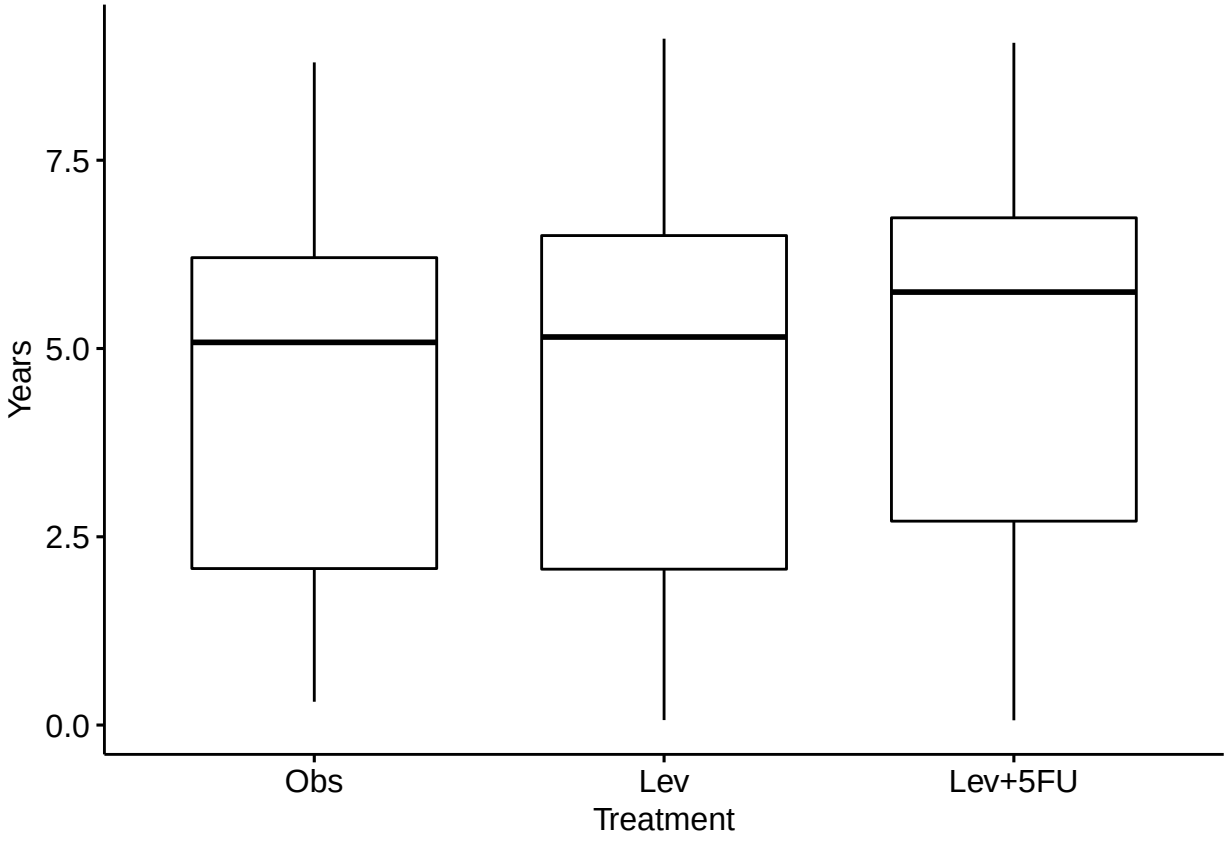
Recurrence Data

```
library(ggpubr)
ggboxplot(studyA, x = "rx", y = "years",
          #color = "rx", palette = c("#00AFBB", "#E7B800", "#FC4E07"),
          order = c("Obs", "Lev", "Lev+5FU"),
          ylab = "Years", xlab = "Treatment",)
```



Death Data

```
ggboxplot(studyB, x = "rx", y = "years",
          #color = "rx", palette = c("#00AFBB", "#E7B800", "#FC4E07"),
          order = c("Obs", "Lev", "Lev+5FU"),
          ylab = "Years", xlab = "Treatment",)
```




```
library(gtsummary)
out <- studyA %>% select(status, rx)
tbl_summary(out, by=rx) %>%
  modify_caption("Table 1. Reoccurance by Treatment")
```

Table 1: Table 1. Reoccurance by Treatment

Characteristic	**Lev**, N = 310	**Lev+5FU**, N = 304	**Obs**, N = 315
status	172 (55%)	119 (39%)	177 (56%)

```
library(gtsummary)
out <- studyB %>% select(status, rx)
tbl_summary(out, by=rx) %>%
  modify_caption("Table 2. Death by Treatment")
```

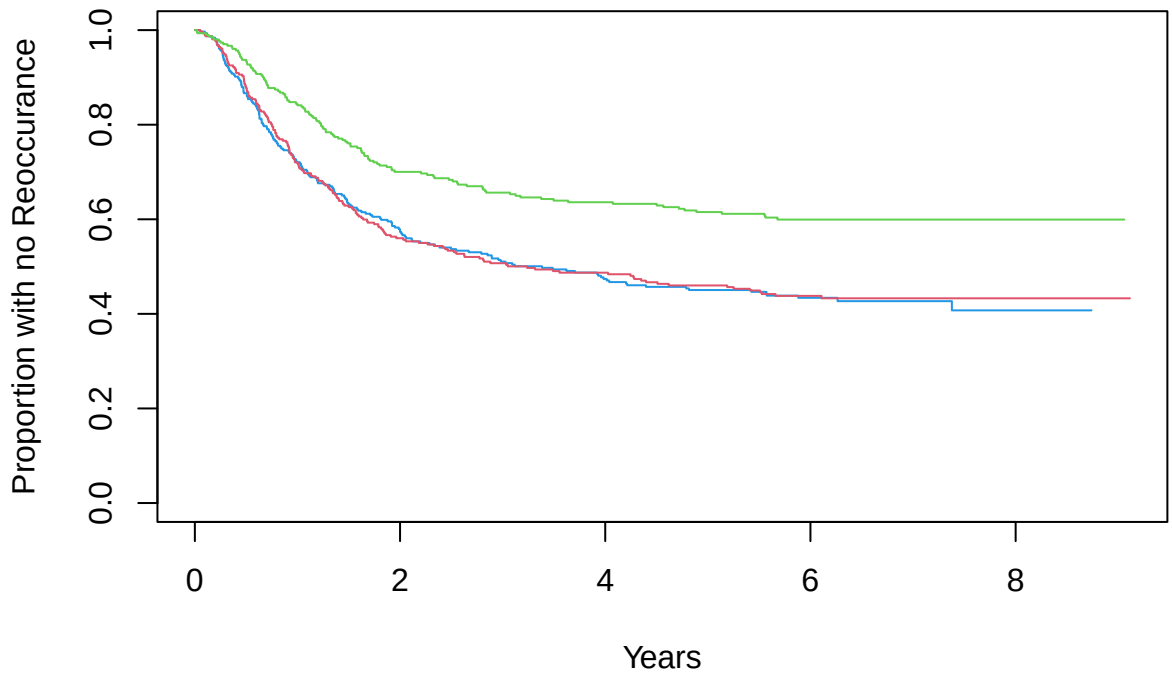
Table 2: Table 2. Death by Treatment

Characteristic	**Lev**, N = 310	**Lev+5FU**, N = 304	**Obs**, N = 315
status	161 (52%)	123 (40%)	168 (53%)

Survival Curve (Kaplan-Meier Survival Plot)

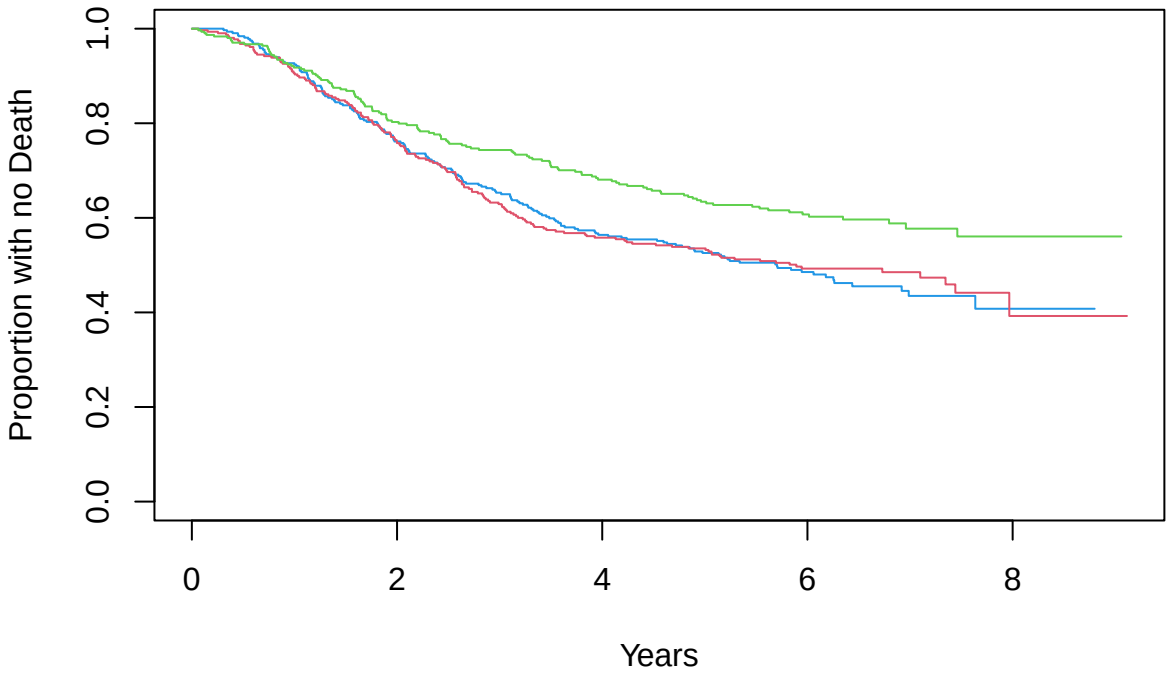
Recurrence data

```r
# Kaplan-Meier Survival Plot (Survival Rate Over Time)
library(survival)
library(survminer)
# Survival curve by treatment
plot(survfit(Surv(years, status) ~ xnum, data=studyA),
     ylab= "Proportion with no Reoccurance",
     xlab = "Years",
     title = "Reoccurance Rate by Treatment",
     col = c(4,2,3),
     pch=c("df","df","df"))
```



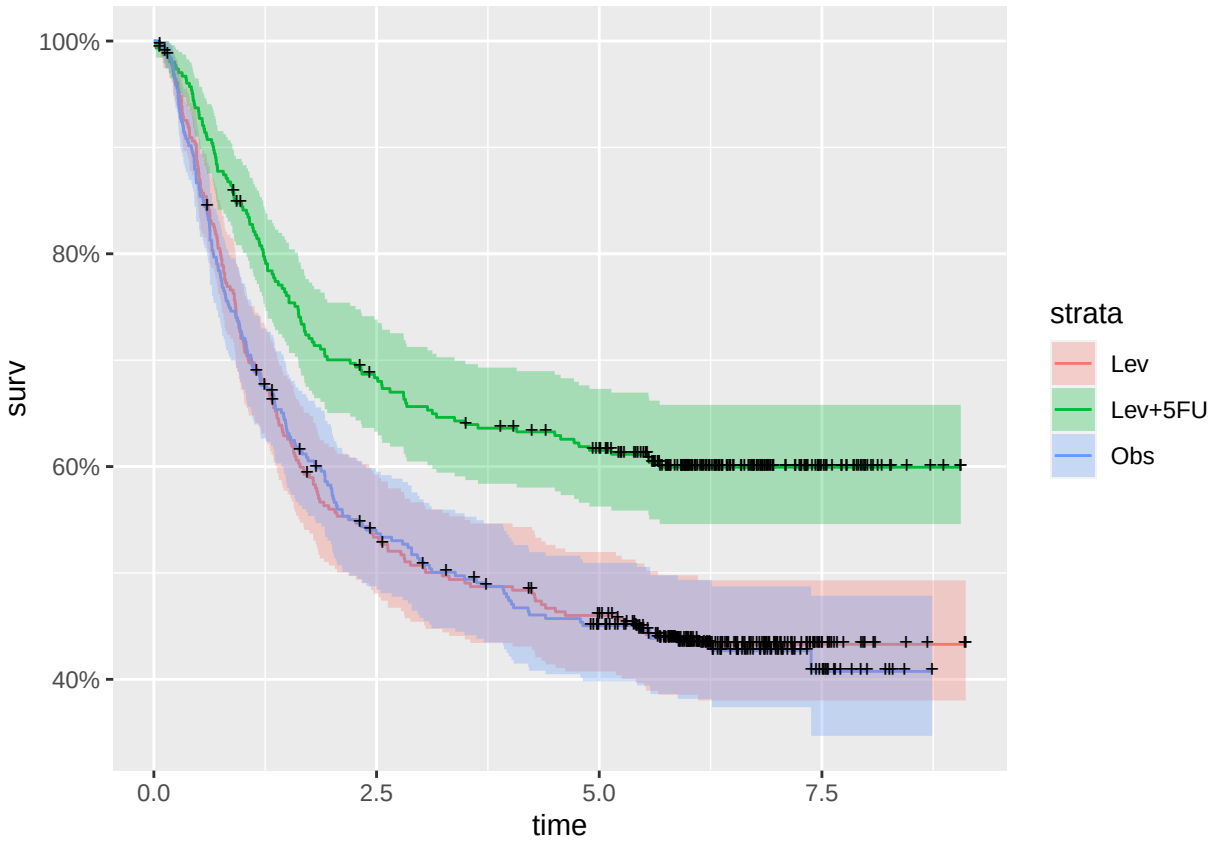
Survival Data

```r
# Kaplan-Meier Survival Plot (Survival Rate Over Time)
library(survival)
library(survminer)
# Survival curve by treatment
plot(survfit(Surv(years, status) ~ xnum, data=studyB),
     ylab= "Proportion with no Death",
     xlab = "Years",
     title = "Death Rate by Treatment",
     col = c(4,2,3)
     )
```

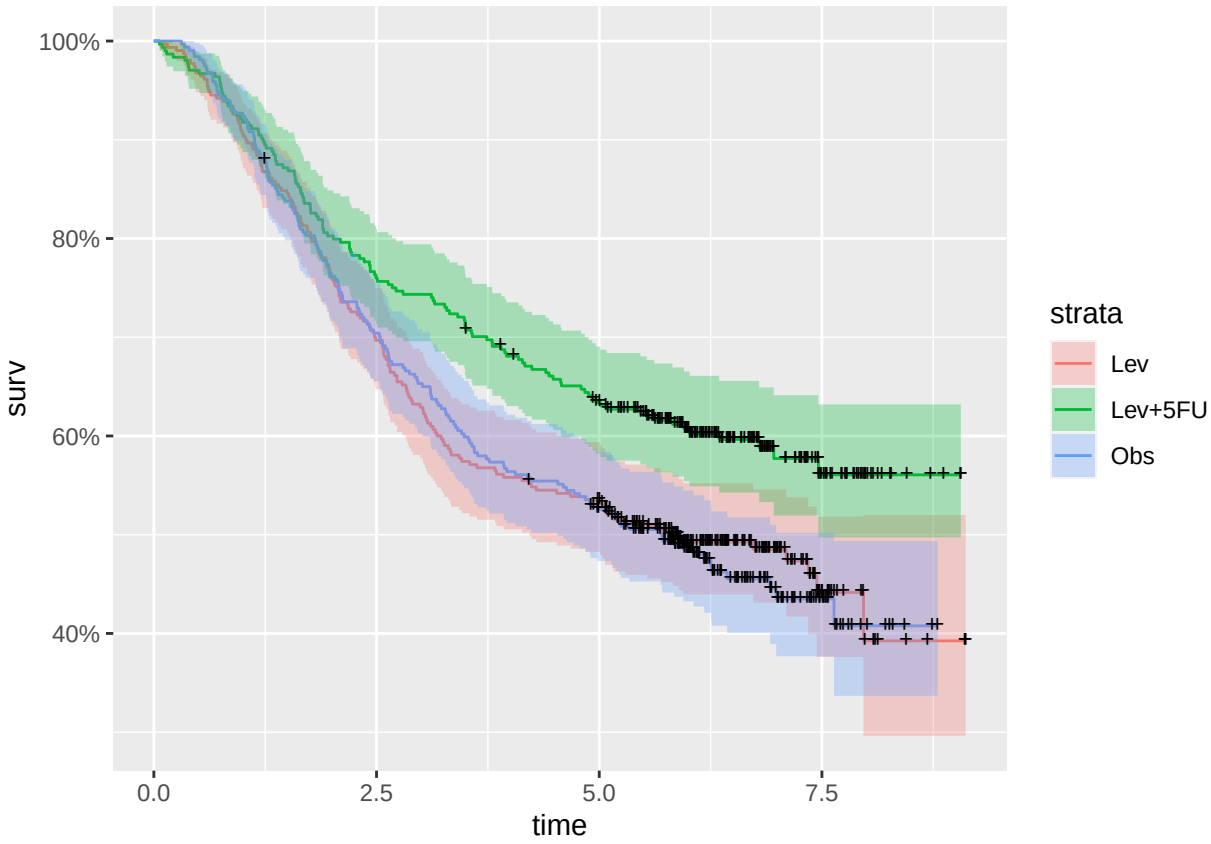


Survival Curves Detail with Confidence Intervals

```
# Clean ouput
library(ggfortify)
library(survival)
fit <- survfit(Surv(years, status) ~ rx, data=studyA)
autoplot(fit)
```




```
fit <- survfit(Surv(years, status) ~ rx, data=studyB)
autoplot(fit)
```



T-Test

A t-test was conducted to quickly see if there is a significance between the means between the control and combined therapy groups.

Assumptions * Continuous variables * Simple Random Sample * Data is approximately normal * Large sample sizes * Homogeneity of variance

```
# T test proves sig. diff. in reoccurrence rates between obs and comb
# treatment groups
t.test(obsA$status, combA$status)
```

```
    Welch Two Sample t-test

data:  obsA$status and combA$status
t = 4.3017, df = 616.77, p-value = 1.97e-05
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.09264048 0.24827431
sample estimates:
mean of x mean of y
0.5619048 0.3914474
```

```
# T test also proves sig. diff. in death rates between obs and comb
# treatment groups
t.test(obsB$status, combB$status)
```

```
    Welch Two Sample t-test
```

```
data:  obsB$status and combB$status
t = 3.2307, df = 616.77, p-value = 0.001301
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.05047816 0.20697798
sample estimates:
mean of x mean of y
0.5333333 0.4046053
```

T tests for both recurrence and death means are significant with a p-value of 1.97e-05 and 0.001 respectively.

Cox Regression

The Proportional hazards model is a class of survival model. Survival models relate the time that passes, before some event occurs, to one or more covariates that may be associated with that quantity of time. In a proportional hazards model, the unique effect of a unit increase in a covariate is multiplicative in regard to the hazard rate.

The Cox Regression or Proportional-Hazards Model can be estimated as follows:

$$\lambda(t|X) = \lambda_0(t) * \exp(\beta_1 X_1 + \dots \beta_p X_p)$$

where:

$$\lambda : \text{hazard} \beta : \text{measure of impact of the covariates} X : \text{covariates}$$

Hypothesis

Based on an exploration of the data, we will examine the following hypotheses:

$H_0 : \beta = 1$ There is no difference between the treatment Hazard Ratio and the control.$H_A : \beta < 1$ The treatnent Harard Ra

Model

The following is the Cox Hazard ratio model. The control group is 0, the single treatment group is 1, and the combined treatment is 2.

```
library(survival)
coxModelA <- coxph(Surv(time, status) ~ xnum, data=studyA)
summary(coxModelA)
```

```
Call:
coxph(formula = Surv(time, status) ~ xnum, data = studyA)

  n= 929, number of events= 468

          coef exp(coef) se(coef)      z Pr(>|z|)
xnum1 -0.01512   0.98499  0.10708 -0.141    0.888
xnum2 -0.51209   0.59924  0.11863 -4.317 1.58e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef) exp(-coef) lower .95 upper .95
xnum1    0.9850      1.015    0.7985    1.2150
```

```
xnum2    0.5992      1.669    0.4749    0.7561

Concordance= 0.554  (se = 0.013 )
Likelihood ratio test= 24.34  on 2 df,   p=5e-06
Wald test            = 22.58  on 2 df,   p=1e-05
Score (logrank) test = 23.07  on 2 df,   p=1e-05
```

```
coxModelA %>%
  gtsummary::tbl_regression(exp = TRUE, label = xnum ~ "Treatment") %>%
  modify_caption("Table 3. Reoccurance Cox Regression")
```

Table 3: Table 3. Reoccurance Cox Regression

Characteristic	HR	95% CI	p-value
Treatment			
0	—	—	
1	0.98	0.80, 1.21	0.9
2	0.60	0.47, 0.76	<0.001

Given the Hazard Ratio of 0.98 for the levamisole treatment, it can be stated that this treatment only resulted in a 2% reduction in the cancer recurrence rate. A p-value of 0.89 for this metric also indicates that the result is not statistically significant.

However, given the Hazard Ratio of 0.60 for the levamisole and fluorouracil treatment, it can be stated that this treatment resulted in a 40% reduction in the cancer recurrence rate. A p-value of 1.58e-05 for this metric indicates that the result is statistically significant.

```
library(survival)
coxModelB <- coxph(Surv(time, status) ~ xnum, data=studyB)
summary(coxModelA)
```

```
Call:
coxph(formula = Surv(time, status) ~ xnum, data = studyA)

  n= 929, number of events= 468

          coef exp(coef) se(coef)      z Pr(>|z|)
xnum1 -0.01512   0.98499  0.10708 -0.141    0.888
xnum2 -0.51209   0.59924  0.11863 -4.317 1.58e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef) exp(-coef) lower .95 upper .95
xnum1    0.9850      1.015    0.7985    1.2150
xnum2    0.5992      1.669    0.4749    0.7561

Concordance= 0.554  (se = 0.013 )
Likelihood ratio test= 24.34  on 2 df,   p=5e-06
Wald test            = 22.58  on 2 df,   p=1e-05
Score (logrank) test = 23.07  on 2 df,   p=1e-05
```

```
coxModelB %>%
  gtsummary::tbl_regression(exp = TRUE, label = list(xnum ~ "Treatment")) %>%
  modify_caption("Table 4. Death Rate Cox Regression")
```

Table 4: Table 4. Death Rate Cox Regression

Characteristic	HR	95% CI	p-value
Treatment			
0	—	—	
1	0.97	0.78, 1.21	0.8
2	0.69	0.55, 0.87	0.002

Given the Hazard Ratio of 0.97 for the levamisole treatment, it can be stated that this treatment only resulted in a 3% reduction in the cancer death rate. A p-value of 0.80 for this metric also indicates that the result is not statistically significant.

However, given the Hazard Ratio of 0.69 for the levamisole and fluorouracil treatment, it can be stated that this treatment resulted in a 31% reduction in the cancer recurrence rate. A p-value of 0.002 for this metric indicates that the result is statistically significant.

Conclusions

Overall, we have determined that there is a significant difference in the survival rate between the different groups. Given the recurrence and death rate p-values for the levamisole and fluorouracil group, we found that they were statistically significant and thus we are able to reject the null hypothesis for the combined treatments. This indicates that the hazard ratios between the two groups are statistically significant and could suggest that the combined treatment has a lower recurrence and death rate. For the levamisole group, we did not find any statistical significance in the recurrence or death rate, indicating that the null hypothesis must be rejected for both outcomes for that group. This indicates that there is no difference between the hazard ratio in the observed and levamisole groups.

For future studies, additional combinations of drug therapies could be examined. Since this study was published over 20 years ago, advances in medicine could have led to the creation of new therapies that could be tested in addition to these medications. Additionally, the effects of the combined treatment on the spread of the cancer could be examined as well. The spread of cancer is an important consideration that should be examnied in addition to recurrence and death rates.