

Code ▾

Cohen et al_CLIA CRC_Clinical analysis

11022024

```
library(swimplot) library(pheatmap) library(reshape2) library(coxphf) library(grid) library(gtable) library(readr) library(mosaic) library(dplyr)
library(survival) library(broom) library(survminer) library(ggplot2) library(scales) library(coxphf) library(ggthemes) library(tidyverse)
library(gtsummary) library(flextable) library(parameters) library(car) library(ComplexHeatmap) library(tidyverse) library(readxl) library(survival)
library(janitor) library(openxlsx) library(writexl) library(rms) library(DT)
```

#ctDNA Detection rate by Stage and Window

Hide

```
#MRD Window
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA_CRC_Clinical Data 122023.csv")
circ_data <- circ_data[circ_data$CLIA.CRC==TRUE,]
circ_data$ctDNA.MRD <- factor(circ_data$ctDNA.MRD, levels=c("NEGATIVE","POSITIVE"))
circ_data$Stage <- factor(circ_data$Stage, levels=c("I","II","III"))
circ_data <- subset(circ_data, ctDNA.MRD %in% c("NEGATIVE", "POSITIVE"))
positive_counts_by_stage <- aggregate(circ_data$ctDNA.MRD == "POSITIVE", by=list(circ_data$Stage), FUN=sum)
total_counts_by_stage <- aggregate(circ_data$ctDNA.MRD, by=list(circ_data$Stage), FUN=length)
combined_data <- data.frame(
  Stage = total_counts_by_stage$Group.1,
  Total_Count = total_counts_by_stage$x,
  Positive_Count = positive_counts_by_stage$x,
  Rate = (positive_counts_by_stage$x / total_counts_by_stage$x) * 100 # Convert to percentage
)
combined_data$Rate <- sprintf("%.2f%%", combined_data$Rate)
overall_total_count <- nrow(circ_data)
overall_positive_count <- nrow(circ_data[circ_data$ctDNA.MRD == "POSITIVE",])
overall_positivity_rate <- (overall_positive_count / overall_total_count) * 100 # Convert to percentage
overall_row <- data.frame(
  Stage = "Overall",
  Total_Count = overall_total_count,
  Positive_Count = overall_positive_count,
  Rate = sprintf("%.2f%%", overall_positivity_rate)
)
combined_data <- rbind(combined_data, overall_row)
print(combined_data)
```

Stage	Total_Count	Positive_Count	Rate
<fctr>	<int>	<int>	<chr>
I	19	2	10.53%
II	158	13	8.23%
III	228	56	24.56%
Overall	405	71	17.53%
4 rows			

Hide

```

#Surveillance Window
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA CRC_Clinical Data 122023.csv")
circ_data <- circ_data[circ_data$CLIA.CRC==TRUE,]
circ_data$ctDNA.Surveillance <- factor(circ_data$ctDNA.Surveillance, levels=c("NEGATIVE","POSITIVE"))
circ_data$Stage <- factor(circ_data$Stage, levels=c("I","II","III"))
circ_data <- subset(circ_data, ctDNA.Surveillance %in% c("NEGATIVE", "POSITIVE"))
positive_counts_by_stage <- aggregate(circ_data$ctDNA.Surveillance == "POSITIVE", by=list(circ_data$Stage), FUN=sum)
total_counts_by_stage <- aggregate(circ_data$ctDNA.Surveillance, by=list(circ_data$Stage), FUN=length)
combined_data <- data.frame(
  Stage = total_counts_by_stage$Group.1,
  Total_Count = total_counts_by_stage$x,
  Positive_Count = positive_counts_by_stage$x,
  Rate = (positive_counts_by_stage$x / total_counts_by_stage$x) * 100 # Convert to percentage
)
combined_data$Rate <- sprintf("%.2f%%", combined_data$Rate)
overall_total_count <- nrow(circ_data)
overall_positive_count <- nrow(circ_data[circ_data$ctDNA.Surveillance == "POSITIVE",])
overall_positivity_rate <- (overall_positive_count / overall_total_count) * 100 # Convert to percentage
overall_row <- data.frame(
  Stage = "Overall",
  Total_Count = overall_total_count,
  Positive_Count = overall_positive_count,
  Rate = sprintf("%.2f%%", overall_positivity_rate)
)
combined_data <- rbind(combined_data, overall_row)
print(combined_data)

```

Stage	Total_Count	Positive_Count	Rate
I	39	3	7.69%
II	214	16	7.48%
III	370	64	17.30%
Overall	623	83	13.32%
4 rows			

Hide

```

#Anytime post-surgery
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA CRC_Clinical Data 122023.csv")
circ_data <- circ_data[circ_data$CLIA.CRC==TRUE,]
circ_data$ctDNA.anytime <- factor(circ_data$ctDNA.anytime, levels=c("NEGATIVE","POSITIVE"))
circ_data$Stage <- factor(circ_data$Stage, levels=c("I","II","III"))
circ_data <- subset(circ_data, ctDNA.anytime %in% c("NEGATIVE", "POSITIVE"))
positive_counts_by_stage <- aggregate(circ_data$ctDNA.anytime == "POSITIVE", by=list(circ_data$Stage), FUN=sum)
total_counts_by_stage <- aggregate(circ_data$ctDNA.anytime, by=list(circ_data$Stage), FUN=length)
combined_data <- data.frame(
  Stage = total_counts_by_stage$Group.1,
  Total_Count = total_counts_by_stage$x,
  Positive_Count = positive_counts_by_stage$x,
  Rate = (positive_counts_by_stage$x / total_counts_by_stage$x) * 100 # Convert to percentage
)
combined_data$Rate <- sprintf("%.2f%%", combined_data$Rate)
overall_total_count <- nrow(circ_data)
overall_positive_count <- nrow(circ_data[circ_data$ctDNA.anytime == "POSITIVE",])
overall_positivity_rate <- (overall_positive_count / overall_total_count) * 100 # Convert to percentage
overall_row <- data.frame(
  Stage = "Overall",
  Total_Count = overall_total_count,
  Positive_Count = overall_positive_count,
  Rate = sprintf("%.2f%%", overall_positivity_rate)
)
combined_data <- rbind(combined_data, overall_row)
print(combined_data)

```

Stage	Total_Count	Positive_Count	Rate
	<int>	<int>	<chr>
I	46	9	19.57%
II	244	36	14.75%
III	452	134	29.65%
Overall	742	179	24.12%

4 rows

#ctDNA MRD Detection rate Stage I/II vs III

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA CRC_Clinical Data 122023.csv")
circ_data <- circ_data[circ_data$CLIA.CRC==TRUE,]
circ_data$ctDNA.MRD <- factor(circ_data$ctDNA.MRD, levels = c("NEGATIVE", "POSITIVE"))
circ_data$Stage_Grouped <- factor(ifelse(circ_data$Stage %in% c("I", "II"), "I/II", "III"))
contingency_table <- table(circ_data$Stage_Grouped, circ_data$ctDNA.MRD)
chi_square_test <- chisq.test(contingency_table)
print(contingency_table)
```

	NEGATIVE	POSITIVE
I/II	162	15
III	172	56

```
print(chi_square_test)
```

```
Pearson's Chi-squared test with Yates' continuity correction
```

```
data: contingency_table
X-squared = 16.741, df = 1, p-value = 4.285e-05
```

#Demographics Table

```

rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA CRC_Clinical Data 122023.csv")
circ_data <- circ_data[circ_data$CLIA.CRC==TRUE,]

circ_data_subset <- circ_data %>%
  select(
    Age,
    Gender,
    PrimSite,
    pT,
    pN,
    Stage,
    Grade,
    NAC,
    ACT,
    MSI,
    BRAF.V600E,
    RAS,
    DFS.Event,
    OS.months) %>%
  mutate(
    Age = as.numeric(Age),
    Gender = factor(Gender, levels = c("Male", "Female")),
    PrimSite = factor(PrimSite, levels = c("Right-sided colon", "Left-sided colon")),
    pT = factor(pT, levels = c("T0", "T1-T2", "T3-T4")),
    pN = factor(pN, levels = c("N0", "N1-N2")),
    Stage = factor(Stage, levels = c("I", "II", "III")),
    Grade = factor(Grade, levels = c("G1", "G2", "G3", "GX")),
    NAC = factor(NAC, levels = c("TRUE", "FALSE"), labels = c("Neoadjuvant Chemotherapy", "Upfront Surgery")),
    ACT = factor(ACT, levels = c("TRUE", "FALSE"), labels = c("Adjuvant Chemotherapy", "Observation")),
    MSI = factor(MSI, levels = c("MSS", "MSI-High")),
    BRAF.V600E = factor(BRAF.V600E, levels = c("WT", "MUT"), labels = c("BRAF WT", "BRAF V600E")),
    RAS = factor(RAS, levels = c("WT", "MUT"), labels = c("RAS WT", "RAS Mut")),
    DFS.Event = factor(DFS.Event, levels = c("TRUE", "FALSE"), labels = c("Recurrence", "No Recurrence")),
    OS.months = as.numeric(OS.months))
table1 <- circ_data_subset %>%
 tbl_summary(
  statistic = list(
    all_continuous() ~ "{median} ({min} - {max})",
    all_categorical() ~ "{n} ({p}%)") %>%
  bold_labels()
table1

```

Characteristic	N = 795 ¹
Age	61 (13 - 91)
Gender	
Male	407 (51%)
Female	388 (49%)
PrimSite	
Right-sided colon	411 (52%)
Left-sided colon	384 (48%)
pT	
T0	3 (0.4%)
T1-T2	133 (17%)
T3-T4	654 (83%)
Unknown	5
pN	
N0	310 (39%)
N1-N2	482 (61%)
Unknown	3

¹ Median (Range); n (%)

Characteristic	N = 795 ¹
Stage	
I	47 (5.9%)
II	262 (33%)
III	486 (61%)
Grade	
G1	84 (11%)
G2	559 (72%)
G3	127 (16%)
GX	4 (0.5%)
Unknown	21
NAC	
Neoadjuvant Chemotherapy	0 (0%)
Upfront Surgery	795 (100%)
ACT	
Adjuvant Chemotherapy	522 (66%)
Observation	273 (34%)
MSI	
MSS	664 (84%)
MSI-High	131 (16%)
BRAF.V600E	
BRAF WT	699 (88%)
BRAF V600E	96 (12%)
RAS	
RAS WT	459 (58%)
RAS Mut	336 (42%)
DFS.Event	
Recurrence	141 (18%)
No Recurrence	654 (82%)
OS.months	
27 (0 - 103)	
¹ Median (Range); n (%)	

Hide

```
fit1 <- as_flex_table(
  table1,
  include = everything(),
  return_calls = FALSE,
  strip_md_bold = TRUE)
```

Warning: The `strip_md_bold` argument of `as_flex_table()` is deprecated as of gtsummary 1.6.0.

Hide

```
fit1
```

Characteristic	N = 795 ¹
Age	61 (13 - 91)

¹Median (Range); n (%)

Characteristic	N = 795 ¹
Gender	
Male	407 (51%)
Female	388 (49%)
PrimSite	
Right-sided colon	411 (52%)
Left-sided colon	384 (48%)
pT	
T0	3 (0.4%)
T1-T2	133 (17%)
T3-T4	654 (83%)
Unknown	5
pN	
N0	310 (39%)
N1-N2	482 (61%)
Unknown	3
Stage	
I	47 (5.9%)
II	262 (33%)
III	486 (61%)
Grade	
G1	84 (11%)
G2	559 (72%)
G3	127 (16%)
GX	4 (0.5%)
Unknown	21
NAC	
Neoadjuvant Chemotherapy	0 (0%)
Upfront Surgery	795 (100%)
ACT	
Adjuvant Chemotherapy	522 (66%)
Observation	273 (34%)
MSI	
MSS	664 (84%)
MSI-High	131 (16%)
BRAF.V600E	
BRAF WT	699 (88%)
BRAF V600E	96 (12%)
RAS	
RAS WT	459 (58%)
RAS Mut	336 (42%)
DFS.Event	
Recurrence	141 (18%)
No Recurrence	654 (82%)
OS.months	
	27 (0 - 103)

¹Median (Range); n (%)

Hide

```
save_as_docx(fit1, path= "~/Downloads/table1.docx")
```

#Heatmap with Clinical & Genomics Factors

Hide

```

rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA CRC_Clinical Data 122023.csv")
circ_data <- circ_data[circ_data$CLIA.CRC==TRUE,]
circ_data <- circ_data %>% arrange(Stage)
circ_datadf <- as.data.frame(circ_data)

ha <- HeatmapAnnotation(
  Stage = circ_data$Stage,
  Gender = circ_data$Gender,
  PrimSite = circ_data$PrimSite,
  pT = circ_data$pT,
  pN = circ_data$pN,
  Grade = circ_data$Grade,
  ACT = circ_data$ACT,
  MSI = circ_data$MSI,
  BRAF.V600E = circ_data$BRAF.V600E,
  RAS = circ_data$RAS,
  ctDNA.MRD = circ_data$ctDNA.MRD,
  ctDNA.Surveillance = circ_data$ctDNA.Surveillance,
  ctDNA.anytime = circ_data$ctDNA.anytime,
  DFS.Event = circ_data$DFS.Event,

  col = list(Stage = c("I" = "seagreen1", "II" = "orange", "III" = "purple"),
             Gender = c("Female" = "goldenrod", "Male" = "blue4"),
             PrimSite = c("Right-sided colon" = "brown", "Left-sided colon" = "darkgreen"),
             pT = c("T0" = "khaki", "T1-T2" = "khaki", "T3-T4" = "brown2"),
             pN = c("N0" = "cornflowerblue", "N1-N2" = "orange2"),
             Grade = c("GX" = "grey", "G1" = "coral", "G2" = "darkgreen", "G3" = "yellow3"),
             ACT = c("TRUE" = "#C1211A", "FALSE" = "#008BCE"),
             MSI = c("MSS" = "grey", "MSI-High" = "black"),
             BRAF.V600E = c("WT" = "grey", "MUT" = "black"),
             RAS = c("WT" = "grey", "MUT" = "black"),
             ctDNA.MRD = c("POSITIVE" = "red3", "NEGATIVE" = "blue"),
             ctDNA.Surveillance = c("POSITIVE" = "red3", "NEGATIVE" = "blue"),
             ctDNA.anytime = c("POSITIVE" = "red3", "NEGATIVE" = "blue"),
             DFS.Event = c("TRUE" = "red3", "FALSE" = "blue")
  )
)
)

ht <- Heatmap(matrix(nrow = 0, ncol = length(circ_data$Stage)), show_row_names = FALSE, cluster_rows = F, cluster_columns = FALSE, top_annotation = ha)
pdf("heatmap.pdf", width = 7, height = 7)
draw(ht, annotation_legend_side = "bottom")
dev.off()

```

#DFS by ctDNA at the MRD Window - all stages

Hide

```

rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA CRC_Clinical Data 122023.csv")
circ_data <- circ_data[circ_data$CLIA.CRC==TRUE,]
circ_data <- circ_data[circ_data$ctDNA.MRD!=""]
circ_data$DFS.months=circ_data$DFS.months-2.5
circ_data <- circ_data[circ_data$DFS.months>=0,]
circ_datadf <- as.data.frame(circ_data)

circ_data$DFS.Event <- as.logical(circ_data$DFS.Event)
circ_data$DFS.months <- as.numeric(circ_data$DFS.months)
survfit(Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event) ~
  ctDNA.MRD, data = circ_data)

```

Call: survfit(formula = Surv(time = circ_data\$DFS.months, event = circ_data\$DFS.Event) ~
 ctDNA.MRD, data = circ_data)

	n	events	median	0.95LCL	0.95UCL
ctDNA.MRD=NEGATIVE	328	29	NA	NA	NA
ctDNA.MRD=POSITIVE	62	36	11.3	8.67	NA

Hide

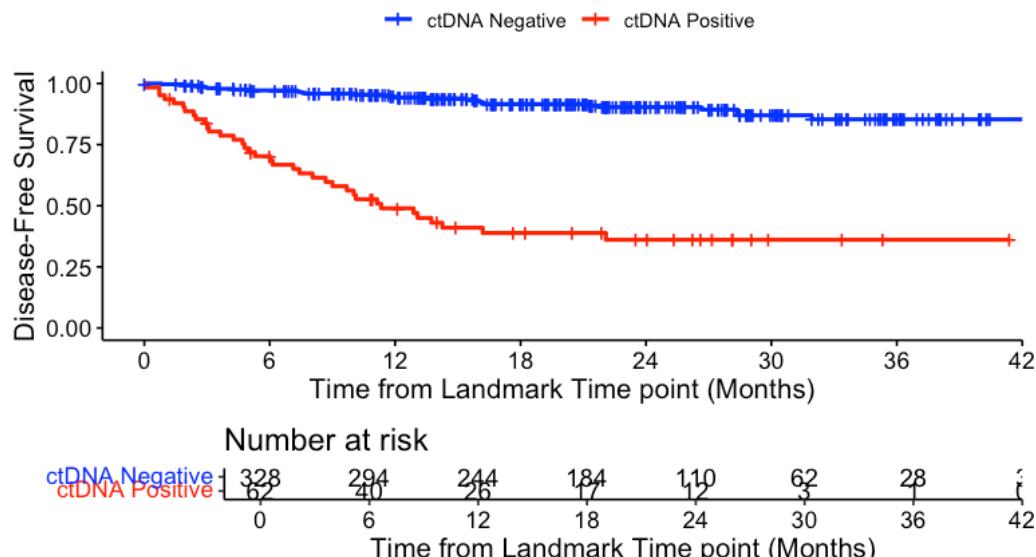
```
event_summary <- circ_data %>%
  group_by(ctDNA.MRD) %>%
  summarise(
    Total = n(),
    Events = sum(DFS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)
```

ctDNA.MRD	Total	Events	Fraction	Percentage
<chr>	<int>	<int>	<dbl>	<dbl>
NEGATIVE	328	29	0.08841463	8.841463
POSITIVE	62	36	0.58064516	58.064516
2 rows				

Hide

```
surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
KM_curve <- survfit(surv_object ~ ctDNA.MRD, data = circ_data, conf.int=0.95, conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette=c("blue","red"), title="DFS - ctDNA MRD window | All pts", ylab= "Disease-Free Survival", xlab="Time from Landmark Time point (Months)", legend.labs=c("ctDNA Negative", "ctDNA Positive"), legend.title="")
```

DFS - ctDNA MRD window | All pts



Hide

```
summary(KM_curve, times= c(24))
```

```
Call: survfit(formula = surv_object ~ ctDNA.MRD, data = circ_data,
  conf.int = 0.95, conf.type = "log-log")
```

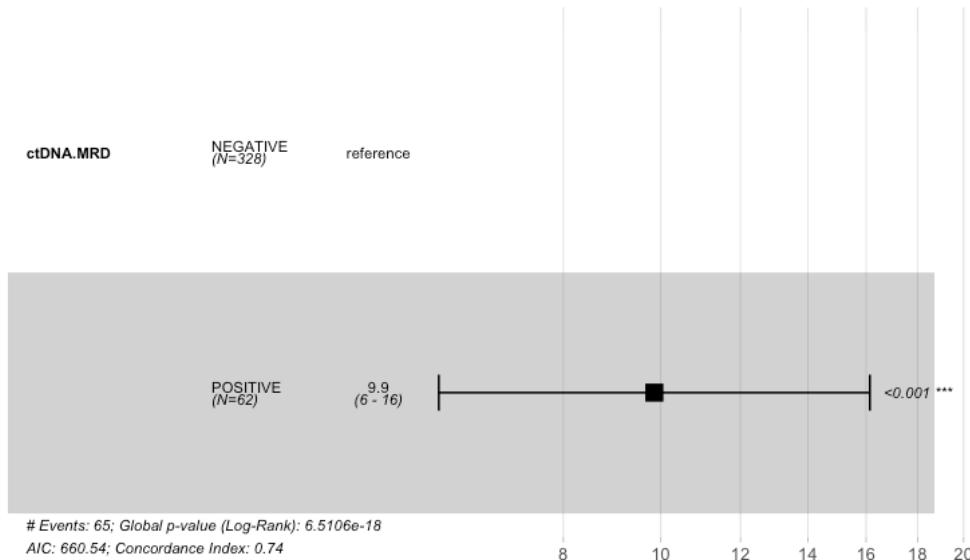
ctDNA.MRD=NEGATIVE						
time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
24.000	110.000	25.000	0.903	0.019	0.859	0.935

ctDNA.MRD=POSITIVE						
time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
24.000	12.0000	36.0000	0.3608	0.0669	0.2332	0.4898

Hide

```
circ_data$ctDNA.MRD <- factor(circ_data$ctDNA.MRD, levels=c("NEGATIVE","POSITIVE"))
cox_fit <- coxph(surv_object ~ ctDNA.MRD, data=circ_data)
ggforest(cox_fit, data = circ_data)
```

Hazard ratio



Hide

```
summary(cox_fit)
```

```
Call:
coxph(formula = surv_object ~ ctDNA.MRD, data = circ_data)

n= 390, number of events= 65

      coef  exp(coef)  se(coef)      z  Pr(>|z|)
ctDNA.MRDPOSITIVE 2.2879    9.8539  0.2517 9.088  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef)  exp(-coef) lower .95 upper .95
ctDNA.MRDPOSITIVE    9.854      0.1015    6.016    16.14

Concordance= 0.742  (se = 0.029 )
Likelihood ratio test= 74.36 on 1 df,  p=<2e-16
Wald test            = 82.6 on 1 df,  p=<2e-16
Score (logrank) test = 124 on 1 df,  p=<2e-16
```

Hide

```
cox_fit_summary <- summary(cox_fit)

# Extract values for HR, 95% CI, and p-value
HR <- cox_fit_summary$coefficients[2]
lower_CI <- cox_fit_summary$conf.int[3]
upper_CI <- cox_fit_summary$conf.int[4]
p_value <- cox_fit_summary$coefficients[5]
label_text <- paste0("HR = ", round(HR, 2), " (", round(lower_CI, 2), "-", round(upper_CI, 2), "); p = ", round(p_value, 3))
print(label_text)
```

```
[1] "HR = 9.85 (6.02-16.14); p = 0"
```

#ctDNA sample positive in the MRD Window - 2-10 week intervals

[Hide](#)

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA CRC cfDNA data_ctDNA timepoints.csv")
circ_data <- circ_data[circ_data$MRD.Window == TRUE,]
circ_data <- circ_data[!is.na(circ_data$cfDNAconc), ]
filtered_data <- circ_data %>%
  filter(ctDNA.MRD >= 2 & ctDNA.MRD <= 10) #intervals for 2-10 weeks
filtered_data$ctDNA_bucket <- cut(filtered_data$ctDNA.MRD,
  breaks = c(2, 4, 6, 8, 10),
  right = FALSE,
  labels = c("2-4", "4-6", "6-8", "8-10"))
filtered_data <- filtered_data %>%
  filter(!is.na(ctDNA_bucket))
rate_by_bucket <- filtered_data %>%
  group_by(ctDNA_bucket) %>%
  summarise(
    n_total = n(), # Total number of patients in the bucket
    n_positive = sum(biomarker_status == "POSITIVE"), # Number of positive cases
    n_negative = sum(biomarker_status == "NEGATIVE"), # Number of negative cases
    percentage_positive = mean(biomarker_status == "POSITIVE") * 100, # Positivity rate
    percentage_negative = mean(biomarker_status == "NEGATIVE") * 100 # Negativity rate
  )
overall_stats <- filtered_data %>%
  summarise(
    total_samples = n(),
    total_positive = sum(biomarker_status == "POSITIVE"),
    total_negative = sum(biomarker_status == "NEGATIVE"),
    overall_percentage_positive = mean(biomarker_status == "POSITIVE") * 100
  )

combined_results <- bind_rows(rate_by_bucket, overall_stats)
print(combined_results)
```

ctDNA_bucket	n_total	n_positive	n_negative	percentage_positive	percentage_negative	total_samples
2-4	140	17	123	12.14286	87.85714	NA
4-6	173	32	141	18.49711	81.50289	NA
6-8	84	22	62	26.19048	73.80952	NA
8-10	56	11	45	19.64286	80.35714	NA
NA	NA	NA	NA	NA	NA	453

5 rows | 1-7 of 10 columns

[Hide](#)

```
# Create the stacked bar plot for positivity and negativity rates by bucket
bar_midpoints <- barplot(
  t(as.matrix(rate_by_bucket[, c("percentage_positive", "percentage_negative")])), # Transpose to get the correct format
  names.arg = rate_by_bucket$ctDNA_bucket,
  col = c("red", "blue"), # Colors: red for positive, blue for negative
  main = '% ctDNA Positive and Negative Samples at the MRD Window',
  xlab = 'Weeks from Surgery',
  ylab = '% ctDNA Samples',
  ylim = c(0, 100),
  legend = c("% Positive", "% Negative"), # Adding a legend for clarification
  args.legend = list(x = "topright")
)
par(new = TRUE)
```

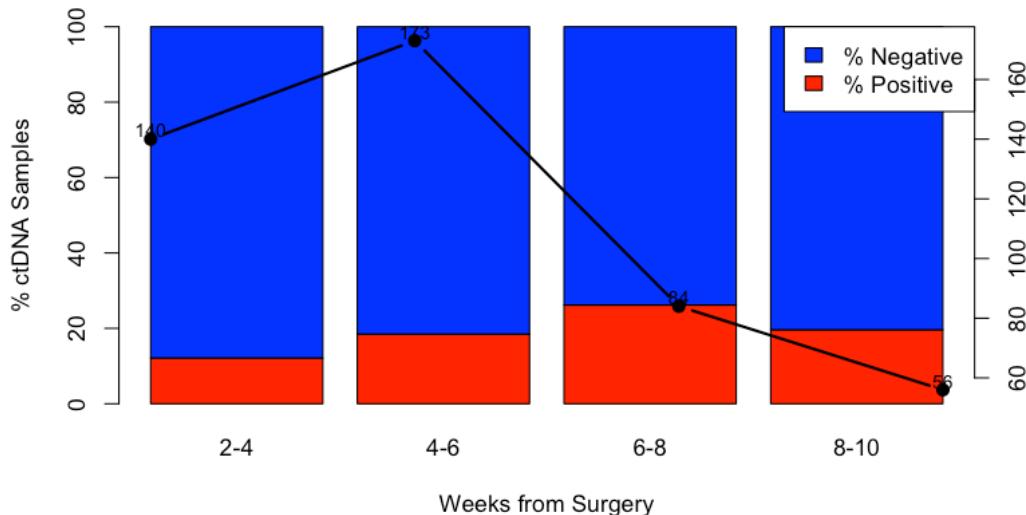
Hide

```
plot(bar_midpoints, rate_by_bucket$n_total, type = "b", col = "black", pch = 19, axes = FALSE, xlab = "", ylab = "", lwd = 2)
axis(side = 4) # Add the secondary y-axis on the right
```

Hide

```
mtext("Total Number of Samples", side = 4, line = 3) # Label for the secondary y-axis
text(bar_midpoints, rate_by_bucket$n_total + 3, labels = rate_by_bucket$n_total, col = "black", cex = 0.8)
```

% ctDNA Positive and Negative Samples at the MRD Window



Hide

```

rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA CRC cfDNA data_ctDNA timepoints.csv")
circ_data <- circ_data[circ_data$MRD.Window == TRUE,]
circ_data <- circ_data[!is.na(circ_data$cfDNAconc), ]
filtered_data <- circ_data %>%
  filter(ctDNA.MRD >= 0 & ctDNA.MRD <= 10) #intervals for 0-10 weeks
filtered_data$ctDNA_bucket <- cut(filtered_data$ctDNA.MRD,
  breaks = c(0, 2, 4, 6, 8, 10),
  right = FALSE,
  labels = c("0-2", "2-4", "4-6", "6-8", "8-10"))
filtered_data <- filtered_data %>%
  filter(!is.na(ctDNA_bucket))
rate_by_bucket <- filtered_data %>%
  group_by(ctDNA_bucket) %>%
  summarise(
    n_total = n(), # Total number of patients in the bucket
    n_positive = sum(biomarker_status == "POSITIVE"), # Number of positive cases
    n_negative = sum(biomarker_status == "NEGATIVE"), # Number of negative cases
    percentage_positive = mean(biomarker_status == "POSITIVE") * 100, # Positivity rate
    percentage_negative = mean(biomarker_status == "NEGATIVE") * 100 # Negativity rate
  )
overall_stats <- filtered_data %>%
  summarise(
    total_samples = n(),
    total_positive = sum(biomarker_status == "POSITIVE"),
    total_negative = sum(biomarker_status == "NEGATIVE"),
    overall_percentage_positive = mean(biomarker_status == "POSITIVE") * 100
  )
)

combined_results <- bind_rows(rate_by_bucket, overall_stats)
print(combined_results)

```

ctDNA_bucket	n_total	n_positive	n_negative	percentage_positive	percentage_negative	total_samples
<fctr>	<int>	<int>	<int>	<dbl>	<dbl>	<int>
0-2	11	1	10	9.090909	90.90909	NA
2-4	140	17	123	12.142857	87.85714	NA
4-6	173	32	141	18.497110	81.50289	NA
6-8	84	22	62	26.190476	73.80952	NA
8-10	56	11	45	19.642857	80.35714	NA
NA	NA	NA	NA	NA	NA	464

6 rows | 1-7 of 10 columns

[Hide](#)

```

# Create the stacked bar plot for positivity and negativity rates by bucket
bar_midpoints <- barplot(
  t(as.matrix(rate_by_bucket[, c("percentage_positive", "percentage_negative")])), # Transpose to get the correct format
  names.arg = rate_by_bucket$ctDNA_bucket,
  col = c("red", "blue"), # Colors: red for positive, blue for negative
  main = '% ctDNA Positive and Negative Samples at the MRD Window',
  xlab = 'Weeks from Surgery',
  ylab = '% ctDNA Samples',
  ylim = c(0, 100),
  legend = c("% Positive", "% Negative"), # Adding a legend for clarification
  args.legend = list(x = "topright")
)
par(new = TRUE)

```

[Hide](#)

```

plot(bar_midpoints, rate_by_bucket$n_total, type = "b", col = "black", pch = 19, axes = FALSE, xlab = "", ylab = "",
  lwd = 2)
axis(side = 4) # Add the secondary y-axis on the right

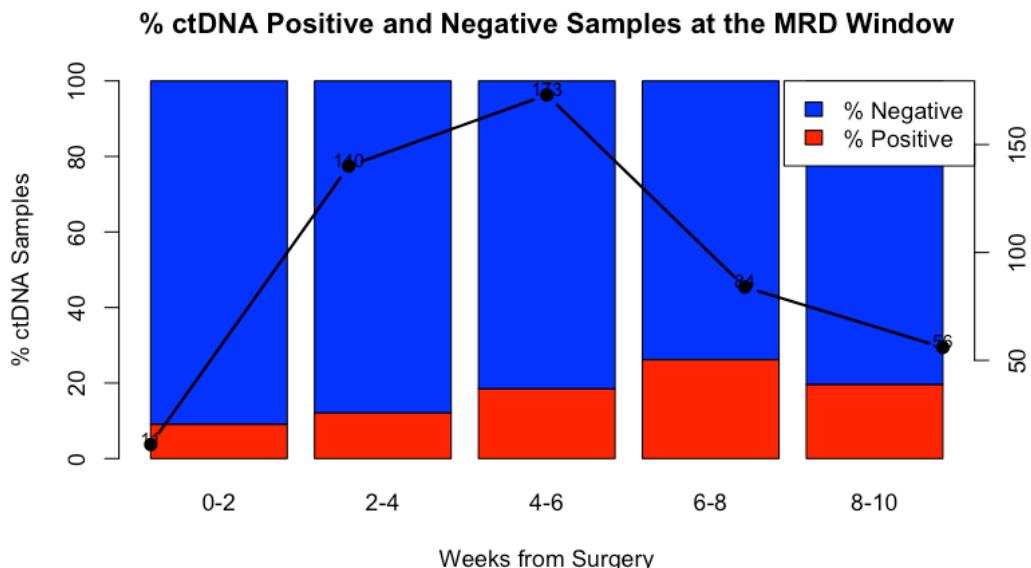
```

[Hide](#)

```

mtext("Total Number of Samples", side = 4, line = 3) # Label for the secondary y-axis
text(bar_midpoints, rate_by_bucket$n_total + 3, labels = rate_by_bucket$n_total, col = "black", cex = 0.8)

```



```
#Median number of timepoints in the MRD Window
```

```
rm(list=ls())
setwd("~/Downloads")
filtered_data <- read.csv("CLIA CRC cfDNA data_ctDNA timepoints.csv")
filtered_data <- filtered_data[filtered_data$MRD.Window==TRUE,]
filtered_data <- filtered_data[!is.na(filtered_data$cfDNAconc), ]
filtered_data <- filtered_data %>%
  filter(ctDNA.MRD >= 2 & ctDNA.MRD <= 10) #intervals for 2-10 weeks

# Calculate the median number of ctDNA tests per patient
median_ctDNA_tests <- filtered_data %>%
  group_by(pts_id) %>%
  summarise(num_tests = n()) %>%
  summarise(median_tests = median(num_tests))
ctDNA_stats <- filtered_data %>%
  group_by(pts_id) %>%
  tally() %>%
  summarise(
    median_tests = median(n),
    min_tests = min(n),
    max_tests = max(n)
  )

print(median_ctDNA_tests)
```

median_tests
<dbl>
1

1 row

```
print(ctDNA_stats)
```

median_tests	min_tests	max_tests
<dbl>	<int>	<int>
1	1	3

1 row

```
#Multivariate cox regression for DFS at the MRD Window & Age threshold as 50 years - all stages
```

Hide

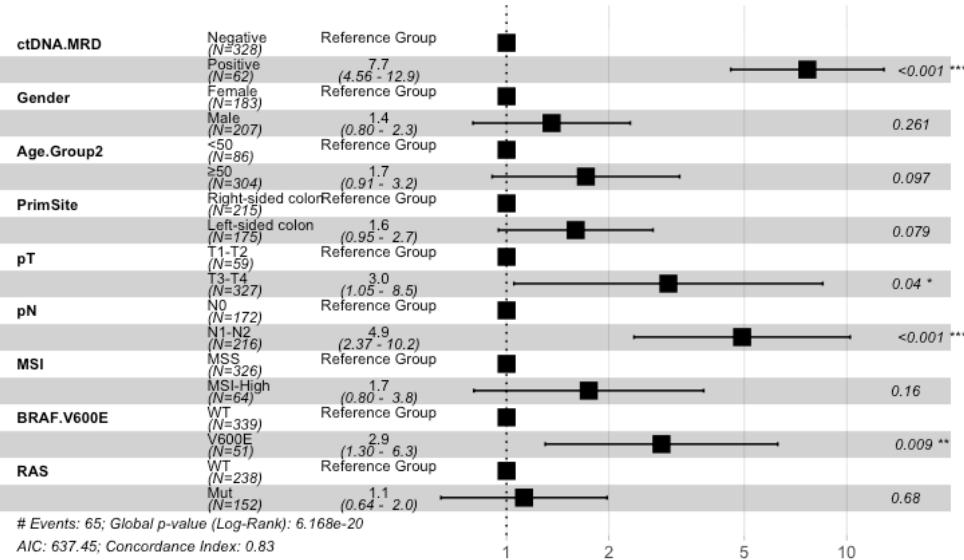
```

rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA CRC_Clinical Data 122023.csv")
circ_data <- circ_data[circ_data$CLIA.CRC==TRUE,]
circ_data <- circ_data[circ_data$ctDNA.MRD!="",]
circ_data$DFS.months=circ_data$DFS.months-2.5
circ_data <- circ_data[circ_data$DFS.months>=0,]
circ_dataadf <- as.data.frame(circ_data)

circ_data$DFS.Event <- as.logical(circ_data$DFS.Event)
circ_data$DFS.months <- as.numeric(circ_data$DFS.months)
circ_data$ctDNA.MRD <- factor(circ_data$ctDNA.MRD, levels=c("NEGATIVE","POSITIVE"), labels = c("Negative", "Positive"))
circ_data$Gender <- factor(circ_data$Gender, levels = c("Female", "Male"))
circ_data$Age.Group2 <- factor(circ_data$Age.Group2, levels = c("1", "2"), labels = c("<50", "≥50"))
circ_data$PrimSite <- factor(circ_data$PrimSite, levels = c("Right-sided colon", "Left-sided colon"))
circ_data$pT <- factor(circ_data$pT, levels = c("T1-T2", "T3-T4"))
circ_data$pN <- factor(circ_data$pN, levels = c("N0", "N1-N2"))
circ_data$MSI <- factor(circ_data$MSI, levels = c("MSS", "MSI-High"))
circ_data$BRAF.V600E <- factor(circ_data$BRAF.V600E, levels = c("WT", "MUT"), labels = c("WT", "V600E"))
circ_data$RAS <- factor(circ_data$RAS, levels = c("WT", "MUT"), labels = c("WT", "Mut"))
surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
cox_fit <- coxph(surv_object ~ ctDNA.MRD + Gender + Age.Group2 + PrimSite + pT + pN + MSI + BRAF.V600E + RAS, data=circ_data)
ggforest(cox_fit, data = circ_data, main = "Multivariate Regression Model for DFS", refLabel = "Reference Group")

```

Multivariate Regression Model for DFS



Hide

```
test.ph <- cox.zph(cox_fit)

# Adjust p-values using False Discovery Rate (FDR) adjustment (Benjamini-Hochberg method)
p_values <- summary(cox_fit)$coefficients[, 5]
adjusted_p_values <- p.adjust(p_values, method = "fdr")
results <- data.frame(
  Variable = rownames(summary(cox_fit)$coefficients),
  Original_P_Value = p_values,
  FDR_Adjusted_P_Value = adjusted_p_values
)
print(results)
```

	Variable <chr>	Original_P_Value <dbl>	FDR_Adjusted_P_Value <dbl>
ctDNA.MRDPPositive	ctDNA.MRDPPositive	1.331163e-14	1.198047e-13
GenderMale	GenderMale	2.610542e-01	2.936859e-01
Age.Group2≥50	Age.Group2≥50	9.700013e-02	1.455002e-01
PrimSiteLeft-sided colon	PrimSiteLeft-sided colon	7.893121e-02	1.420762e-01
pTT3-T4	pTT3-T4	3.993210e-02	8.984723e-02
pNN1-N2	pNN1-N2	1.931204e-05	8.690419e-05
MSIMSI-High	MSIMSI-High	1.603395e-01	2.061508e-01
BRAF.V600EV600E	BRAF.V600EV600E	8.963156e-03	2.688947e-02
RASMut	RASMut	6.798914e-01	6.798914e-01

9 rows

#Univariate cox regression for factors used in MVA - all stages

Hide

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA_CRC_Clinical Data 122023.csv")
circ_data <- circ_data[circ_data$CLIA.CRC==TRUE,]
circ_data <- circ_data[circ_data$ctDNA.MRD!="",]
circ_data$DFS.months=circ_data$DFS.months-2.5
circ_data <- circ_data[circ_data$DFS.months≥0,]
circ_datadf <- as.data.frame(circ_data)
circ_data$DFS.Event <- as.logical(circ_data$DFS.Event)
circ_data$DFS.months <- as.numeric(circ_data$DFS.months)
surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
circ_data$Gender <- factor(circ_data$Gender, levels = c("Female", "Male")) #univariate for gender
cox_fit <- coxph(surv_object ~ Gender, data=circ_data)
summary(cox_fit)
```

```
Call:
coxph(formula = surv_object ~ Gender, data = circ_data)

n= 390, number of events= 65

      coef exp(coef)  se(coef)    z Pr(>|z|)    
GenderMale 0.3259    1.3853   0.2550 1.278    0.201  
                    exp(coef) exp(-coef) lower .95 upper .95  
GenderMale    1.385      0.7219    0.8404    2.284  

Concordance= 0.535  (se = 0.032 )
Likelihood ratio test= 1.67  on 1 df,   p=0.2
Wald test            = 1.63  on 1 df,   p=0.2
Score (logrank) test = 1.65  on 1 df,   p=0.2
```

Hide

```
cox_fit_summary <- summary(cox_fit)
HR <- cox_fit_summary$coefficients[2]
lower_CI <- cox_fit_summary$conf.int[3]
upper_CI <- cox_fit_summary$conf.int[4]
p_value <- cox_fit_summary$coefficients[5]
label_text <- paste0("HR = ", round(HR, 2), " (", round(lower_CI, 2), "-", round(upper_CI, 2), "); p = ", round(p_value, 3))
print(label_text)
```

```
[1] "HR = 1.39 (0.84-2.28); p = 0.201"
```

[Hide](#)

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA_CRC_Clinical Data 122023.csv")
circ_data <- circ_data[circ_data$CLIA.CRC==TRUE,]
circ_data <- circ_data[circ_data$ctDNA.MRD!="",]
circ_data$DFS.months=circ_data$DFS.months-2.5
circ_data <- circ_data[circ_data$DFS.months>=0,]
circ_dataadf <- as.data.frame(circ_data)
circ_data$DFS.Event <- as.logical(circ_data$DFS.Event)
circ_data$DFS.months <- as.numeric(circ_data$DFS.months)
surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
circ_data$Age.Group2 <- factor(circ_data$Age.Group2, levels = c("1", "2"), labels = c("<50", "≥50")) #univariate for Age
cox_fit <- coxph(surv_object ~ Age.Group2, data=circ_data)
summary(cox_fit)
```

```
Call:
coxph(formula = surv_object ~ Age.Group2, data = circ_data)

n= 390, number of events= 65

            coef exp(coef)  se(coef)      z Pr(>|z|)
Age.Group2≥50 0.1934    1.2133   0.3102  0.623    0.533

            exp(coef) exp(-coef) lower .95 upper .95
Age.Group2≥50    1.213     0.8242   0.6606   2.229

Concordance= 0.526  (se = 0.024 )
Likelihood ratio test= 0.4  on 1 df,  p=0.5
Wald test        = 0.39  on 1 df,  p=0.5
Score (logrank) test = 0.39  on 1 df,  p=0.5
```

[Hide](#)

```
cox_fit_summary <- summary(cox_fit)
HR <- cox_fit_summary$coefficients[2]
lower_CI <- cox_fit_summary$conf.int[3]
upper_CI <- cox_fit_summary$conf.int[4]
p_value <- cox_fit_summary$coefficients[5]
label_text <- paste0("HR = ", round(HR, 2), " (", round(lower_CI, 2), "-", round(upper_CI, 2), "); p = ", round(p_value, 3))
print(label_text)
```

```
[1] "HR = 1.21 (0.66-2.23); p = 0.533"
```

[Hide](#)

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA_CRC_Clinical Data 122023.csv")
circ_data <- circ_data[circ_data$CLIA.CRC==TRUE,]
circ_data <- circ_data[circ_data$ctDNA.MRD!="",]
circ_data$DFS.months=circ_data$DFS.months-2.5
circ_data <- circ_data[circ_data$DFS.months>=0,]
circ_dataadf <- as.data.frame(circ_data)
circ_data$DFS.Event <- as.logical(circ_data$DFS.Event)
circ_data$DFS.months <- as.numeric(circ_data$DFS.months)
surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
circ_data$PrimSite <- factor(circ_data$PrimSite, levels = c("Right-sided colon", "Left-sided colon")) #univariate for Tumor Location
cox_fit <- coxph(surv_object ~ PrimSite, data=circ_data)
summary(cox_fit)
```

```

Call:
coxph(formula = surv_object ~ PrimSite, data = circ_data)

n= 390, number of events= 65

            coef exp(coef) se(coef)      z Pr(>|z|)    
PrimSiteLeft-sided colon 0.2250     1.2523   0.2482 0.907    0.365
                           exp(coef) exp(-coef) lower .95 upper .95    
PrimSiteLeft-sided colon     1.252      0.7985     0.77    2.037

Concordance= 0.518 (se = 0.032 )
Likelihood ratio test= 0.82 on 1 df,  p=0.4
Wald test             = 0.82 on 1 df,  p=0.4
Score (logrank) test = 0.83 on 1 df,  p=0.4

```

[Hide](#)

```

cox_fit_summary <- summary(cox_fit)
HR <- cox_fit_summary$coefficients[2]
lower_CI <- cox_fit_summary$conf.int[3]
upper_CI <- cox_fit_summary$conf.int[4]
p_value <- cox_fit_summary$coefficients[5]
label_text <- paste0("HR = ", round(HR, 2), " (", round(lower_CI, 2), "-", round(upper_CI, 2), "); p = ", round(p_value, 3))
print(label_text)

```

```
[1] "HR = 1.25 (0.77-2.04); p = 0.365"
```

[Hide](#)

```

rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA_CRC_Clinical Data 122023.csv")
circ_data <- circ_data[circ_data$CLIA.CRC==TRUE,]
circ_data <- circ_data[circ_data$ctDNA.MRD!=""]
circ_data$DFS.months=circ_data$DFS.months-2.5
circ_data <- circ_data[circ_data$DFS.months>=0,]
circ_datadf <- as.data.frame(circ_data)
circ_data$DFS.Event <- as.logical(circ_data$DFS.Event)
circ_data$DFS.months <- as.numeric(circ_data$DFS.months)
surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
circ_data$pT <- factor(circ_data$pT, levels = c("T1-T2", "T3-T4")) #univariate for Overall T Stage
cox_fit <- coxph(surv_object ~ pT, data=circ_data)
summary(cox_fit)

```

```

Call:
coxph(formula = surv_object ~ pT, data = circ_data)

n= 386, number of events= 65
(4 observations deleted due to missingness)

            coef exp(coef) se(coef)      z Pr(>|z|)    
pTT3-T4 1.0960     2.9923   0.5163 2.123    0.0338 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

            exp(coef) exp(-coef) lower .95 upper .95    
pTT3-T4     2.992      0.3342     1.088     8.231

Concordance= 0.549 (se = 0.017 )
Likelihood ratio test= 6.26 on 1 df,  p=0.01
Wald test             = 4.51 on 1 df,  p=0.03
Score (logrank) test = 4.98 on 1 df,  p=0.03

```

[Hide](#)

```

cox_fit_summary <- summary(cox_fit)
HR <- cox_fit_summary$coefficients[2]
lower_CI <- cox_fit_summary$conf.int[3]
upper_CI <- cox_fit_summary$conf.int[4]
p_value <- cox_fit_summary$coefficients[5]
label_text <- paste0("HR = ", round(HR, 2), " (", round(lower_CI, 2), "-", round(upper_CI, 2), "); p = ", round(p_value, 3))
print(label_text)

```

```
[1] "HR = 2.99 (1.09-8.23); p = 0.034"
```

[Hide](#)

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA CRC_Clinical Data 122023.csv")
circ_data <- circ_data[circ_data$CLIA.CRC==TRUE,]
circ_data <- circ_data[circ_data$ctDNA.MRD!="",]
circ_data$DFS.months=circ_data$DFS.months-2.5
circ_data <- circ_data[circ_data$DFS.months>=0,]
circ_dataadf <- as.data.frame(circ_data)
circ_data$DFS.Event <- as.logical(circ_data$DFS.Event)
circ_data$DFS.months <- as.numeric(circ_data$DFS.months)
surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
circ_data$pN <- factor(circ_data$pN, levels = c("N0", "N1-N2")) #univariate for Overall N Stage
cox_fit <- coxph(surv_object ~ pN, data=circ_data)
summary(cox_fit)
```

Call:

```
coxph(formula = surv_object ~ pN, data = circ_data)
```

n= 388, number of events= 65
(2 observations deleted due to missingness)

	coef	exp(coef)	se(coef)	z	Pr(> z)
pNN1-N2	1.6966	5.4556	0.3592	4.723	2.32e-06 ***

Signif. codes:	0	***	0.001	**	0.01
	*	0.05	.	0.1	' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95
pNN1-N2	5.456	0.1833	2.698	11.03

Concordance= 0.667 (se = 0.023)

Likelihood ratio test= 31.89 on 1 df, p=2e-08

Wald test = 22.31 on 1 df, p=2e-06

Score (logrank) test = 28.18 on 1 df, p=1e-07

[Hide](#)

```
cox_fit_summary <- summary(cox_fit)
HR <- cox_fit_summary$coefficients[2]
lower_CI <- cox_fit_summary$conf.int[3]
upper_CI <- cox_fit_summary$conf.int[4]
p_value <- cox_fit_summary$coefficients[5]
label_text <- paste0("HR = ", round(HR, 2), " (", round(lower_CI, 2), " - ", round(upper_CI, 2), "); p = ", round(p_value, 3))
print(label_text)
```

```
[1] "HR = 5.46 (2.7-11.03); p = 0"
```

[Hide](#)

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA CRC_Clinical Data 122023.csv")
circ_data <- circ_data[circ_data$CLIA.CRC==TRUE,]
circ_data <- circ_data[circ_data$ctDNA.MRD!="",]
circ_data$DFS.months=circ_data$DFS.months-2.5
circ_data <- circ_data[circ_data$DFS.months>=0,]
circ_dataadf <- as.data.frame(circ_data)
circ_data$DFS.Event <- as.logical(circ_data$DFS.Event)
circ_data$DFS.months <- as.numeric(circ_data$DFS.months)
surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
circ_data$MSI <- factor(circ_data$MSI, levels = c("MSS", "MSI-High")) #univariate for MSI
cox_fit <- coxph(surv_object ~ MSI, data=circ_data)
summary(cox_fit)
```

```

Call:
coxph(formula = surv_object ~ MSI, data = circ_data)

n= 390, number of events= 65

      coef exp(coef)  se(coef)    z Pr(>|z|)
MSIMSI-High 0.2018    1.2236   0.3312 0.609    0.542
                           exp(coef) exp(-coef) lower .95 upper .95
MSIMSI-High      1.224     0.8173   0.6393  2.342

Concordance= 0.531 (se = 0.028 )
Likelihood ratio test= 0.35 on 1 df,  p=0.6
Wald test            = 0.37 on 1 df,  p=0.5
Score (logrank) test = 0.37 on 1 df,  p=0.5

```

```

cox_fit_summary <- summary(cox_fit)
HR <- cox_fit_summary$coefficients[2]
lower_CI <- cox_fit_summary$conf.int[3]
upper_CI <- cox_fit_summary$conf.int[4]
p_value <- cox_fit_summary$coefficients[5]
label_text <- paste0("HR = ", round(HR, 2), " (", round(lower_CI, 2), "-", round(upper_CI, 2), "); p = ", round(p_value, 3))
print(label_text)

```

```
[1] "HR = 1.22 (0.64-2.34); p = 0.542"
```

```

rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA CRC_Clinical Data 122023.csv")
circ_data <- circ_data[circ_data$CLIA.CRC==TRUE,]
circ_data <- circ_data[circ_data$ctDNA.MRD!="",]
circ_data$DFS.months=circ_data$DFS.months-2.5
circ_data <- circ_data[circ_data$DFS.months>=0,]
circ_datadf <- as.data.frame(circ_data)
circ_data$DFS.Event <- as.logical(circ_data$DFS.Event)
circ_data$DFS.months <- as.numeric(circ_data$DFS.months)
surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
circ_data$BRAF.V600E <- factor(circ_data$BRAF.V600E, levels = c("WT", "MUT"), labels = c("WT", "V600E")) #univariate for BRAF
cox_fit <- coxph(surv_object ~ BRAF.V600E, data=circ_data)
summary(cox_fit)

```

```

Call:
coxph(formula = surv_object ~ BRAF.V600E, data = circ_data)

n= 390, number of events= 65

      coef exp(coef)  se(coef)    z Pr(>|z|)
BRAF.V600EV600E 0.5464    1.7271   0.3199 1.708    0.0876 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef) exp(-coef) lower .95 upper .95
BRAF.V600EV600E     1.727     0.579    0.9226    3.233

Concordance= 0.552 (se = 0.028 )
Likelihood ratio test= 2.59 on 1 df,  p=0.1
Wald test            = 2.92 on 1 df,  p=0.09
Score (logrank) test = 2.99 on 1 df,  p=0.08

```

```

cox_fit_summary <- summary(cox_fit)
HR <- cox_fit_summary$coefficients[2]
lower_CI <- cox_fit_summary$conf.int[3]
upper_CI <- cox_fit_summary$conf.int[4]
p_value <- cox_fit_summary$coefficients[5]
label_text <- paste0("HR = ", round(HR, 2), " (", round(lower_CI, 2), "-", round(upper_CI, 2), "); p = ", round(p_value, 3))
print(label_text)

```

```
[1] "HR = 1.73 (0.92-3.23); p = 0.088"
```

[Hide](#)

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA CRC_Clinical Data 122023.csv")
circ_data <- circ_data[circ_data$CLIA.CRC==TRUE,]
circ_data <- circ_data[circ_data$ctDNA.MRD!="",]
circ_data$DFS.months=circ_data$DFS.months-2.5
circ_data <- circ_data[circ_data$DFS.months>=0,]
circ_datadf <- as.data.frame(circ_data)
circ_data$DFS.Event <- as.logical(circ_data$DFS.Event)
circ_data$DFS.months <- as.numeric(circ_data$DFS.months)
surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
circ_data$RAS <- factor(circ_data$RAS, levels = c("WT", "MUT"), labels = c("WT", "Mut")) #univariate for RAS
cox_fit <- coxph(surv_object ~ RAS, data=circ_data)
summary(cox_fit)
```

Call:
`coxph(formula = surv_object ~ RAS, data = circ_data)`

n= 390, number of events= 65

	coef	exp(coef)	se(coef)	z	Pr(> z)
RASMut	0.1691	1.1843	0.2507	0.675	0.5
	exp(coef)	exp(-coef)	lower .95	upper .95	
RASMut	1.184	0.8444	0.7246	1.936	

Concordance= 0.511 (se = 0.031)
Likelihood ratio test= 0.45 on 1 df, p=0.5
Wald test = 0.46 on 1 df, p=0.5
Score (logrank) test = 0.46 on 1 df, p=0.5

[Hide](#)

```
cox_fit_summary <- summary(cox_fit)
HR <- cox_fit_summary$coefficients[2]
lower_CI <- cox_fit_summary$conf.int[3]
upper_CI <- cox_fit_summary$conf.int[4]
p_value <- cox_fit_summary$coefficients[5]
label_text <- paste0("HR = ", round(HR, 2), " (", round(lower_CI, 2), " - ", round(upper_CI, 2), "); p = ", round(p_value, 3))
print(label_text)
```

```
[1] "HR = 1.18 (0.72-1.94); p = 0.5"
```

[Hide](#)

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA CRC_Clinical Data 122023.csv")
circ_data <- circ_data[circ_data$CLIA.CRC==TRUE,]
circ_data <- circ_data[circ_data$ctDNA.MRD!="",]
circ_data <- circ_data[!(circ_data$Stage %in% c("I", "III"))]
circ_data$DFS.months=circ_data$DFS.months-2.5
circ_data <- circ_data[circ_data$DFS.months>=0,]
circ_datadf <- as.data.frame(circ_data)

circ_data$DFS.Event <- as.logical(circ_data$DFS.Event)
circ_data$DFS.months <- as.numeric(circ_data$DFS.months)
survfit(Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event) ~
ctDNA.MRD, data = circ_data)
```

Call: `survfit(formula = Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event) ~ ctDNA.MRD, data = circ_data)`

	n	events	median	0.95LCL	0.95UCL
ctDNA.MRD=NEGATIVE	141	5	NA	NA	NA
ctDNA.MRD=POSITIVE	10	3	NA	11.1	NA

[Hide](#)

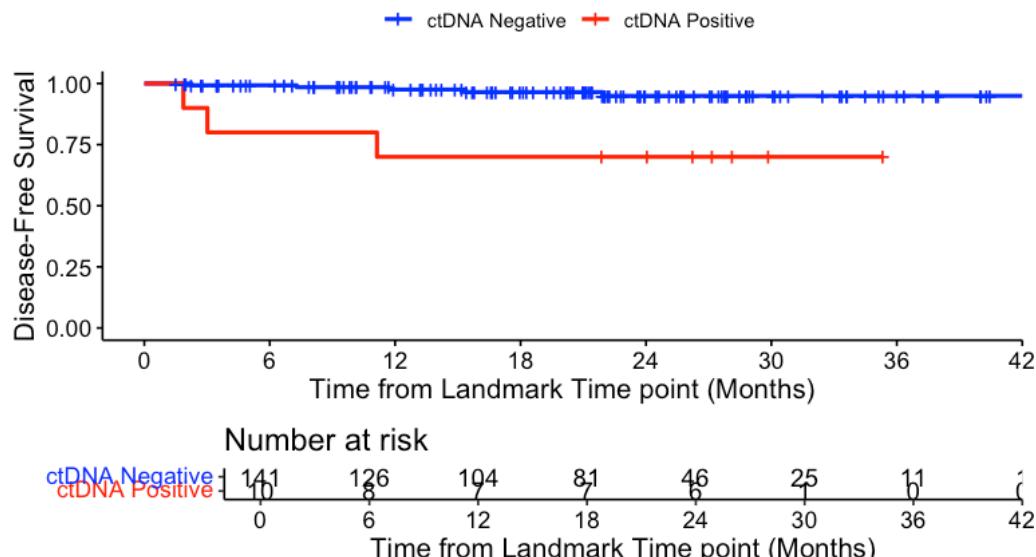
```
event_summary <- circ_data %>%
  group_by(ctDNA.MRD) %>%
  summarise(
    Total = n(),
    Events = sum(DFS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)
```

ctDNA.MRD	Total	Events	Fraction	Percentage
<chr>	<int>	<int>	<dbl>	<dbl>
NEGATIVE	141	5	0.03546099	3.546099
POSITIVE	10	3	0.30000000	30.000000
2 rows				

Hide

```
surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
KM_curve <- survfit(surv_object ~ ctDNA.MRD, data = circ_data, conf.int=0.95, conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette
e=c("blue","red"), title="DFS - ctDNA MRD window | Stage II", ylab= "Disease-Free Survival", xlab="Time from Land
mark Time point (Months)", legend.labs=c("ctDNA Negative", "ctDNA Positive"), legend.title="")
```

DFS - ctDNA MRD window | Stage II



Hide

```
summary(KM_curve, times= c(24))
```

```
Call: survfit(formula = surv_object ~ ctDNA.MRD, data = circ_data,
  conf.int = 0.95, conf.type = "log-log")
```

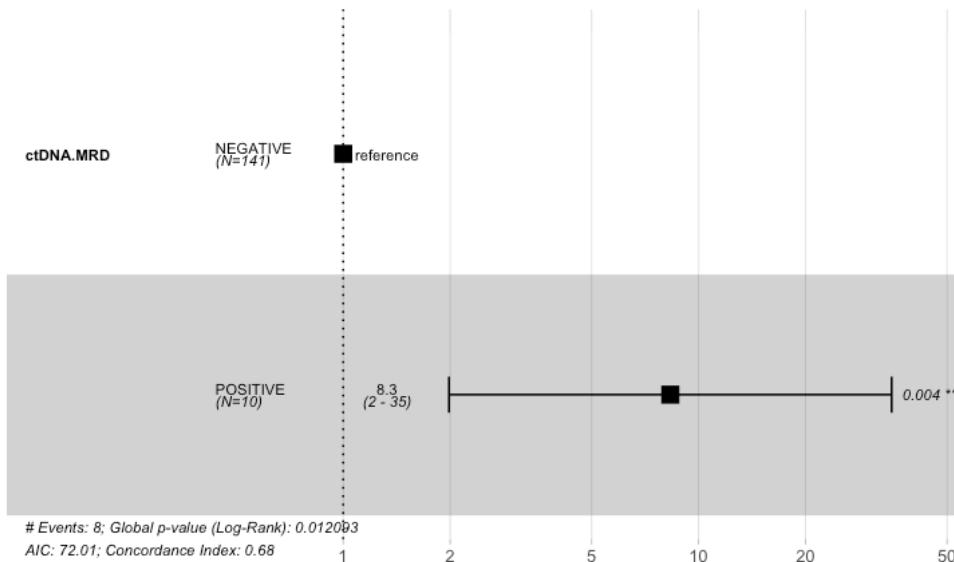
ctDNA.MRD=NEGATIVE						
time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
24.000	46.000	5.000	0.949	0.023	0.878	0.979

ctDNA.MRD=POSITIVE						
time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
24.000	6.000	3.000	0.700	0.145	0.329	0.892

Hide

```
circ_data$ctDNA.MRD <- factor(circ_data$ctDNA.MRD, levels=c("NEGATIVE","POSITIVE"))
cox_fit <- coxph(surv_object ~ ctDNA.MRD, data=circ_data)
ggforest(cox_fit,data = circ_data)
```

Hazard ratio



Hide

```
summary(cox_fit)
```

```
Call:
coxph(formula = surv_object ~ ctDNA.MRD, data = circ_data)

n= 151, number of events= 8

      coef  exp(coef)  se(coef)      z Pr(>|z|)    
ctDNA.MRDPOSITIVE 2.1194    8.3262  0.7317 2.897  0.00377 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef)  exp(-coef) lower .95 upper .95    
ctDNA.MRDPOSITIVE    8.326      0.1201    1.985    34.93    

Concordance= 0.68  (se = 0.09 )
Likelihood ratio test= 6.3  on 1 df,  p=0.01
Wald test            = 8.39  on 1 df,  p=0.004
Score (logrank) test = 12.02 on 1 df,  p=5e-04
```

Hide

```
cox_fit_summary <- summary(cox_fit)

# Extract values for HR, 95% CI, and p-value
HR <- cox_fit_summary$coefficients[2]
lower_CI <- cox_fit_summary$conf.int[3]
upper_CI <- cox_fit_summary$conf.int[4]
p_value <- cox_fit_summary$coefficients[5]
label_text <- paste0("HR = ", round(HR, 2), " (", round(lower_CI, 2), "-", round(upper_CI, 2), "); p = ", round(p_value, 3))
print(label_text)
```

```
[1] "HR = 8.33 (1.98-34.93); p = 0.004"
```

#DFS by ctDNA at the MRD Window - Stage III

Hide

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA_CRC_Clinical Data 122023.csv")
circ_data <- circ_data[circ_data$CLIA.CRC==TRUE,]
circ_data <- circ_data[circ_data$ctDNA.MRD!="",]
circ_data <- circ_data[!(circ_data$Stage %in% c("I", "II"))]
circ_data$DFS.months=circ_data$DFS.months-2.5
circ_data <- circ_data[circ_data$DFS.months>=0,]
circ_datadf <- as.data.frame(circ_data)

circ_data$DFS.Event <- as.logical(circ_data$DFS.Event)
circ_data$DFS.months <- as.numeric(circ_data$DFS.months)
survfit(Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event) ~
  ctDNA.MRD, data = circ_data)
```

Call: survfit(formula = Surv(time = circ_data\$DFS.months, event = circ_data\$DFS.Event) ~
 ctDNA.MRD, data = circ_data)

	n	events	median	0.95LCL	0.95UCL
ctDNA.MRD=NEGATIVE	170	24	NA	NA	NA
ctDNA.MRD=POSITIVE	51	33	10	7.13	16.2

Hide

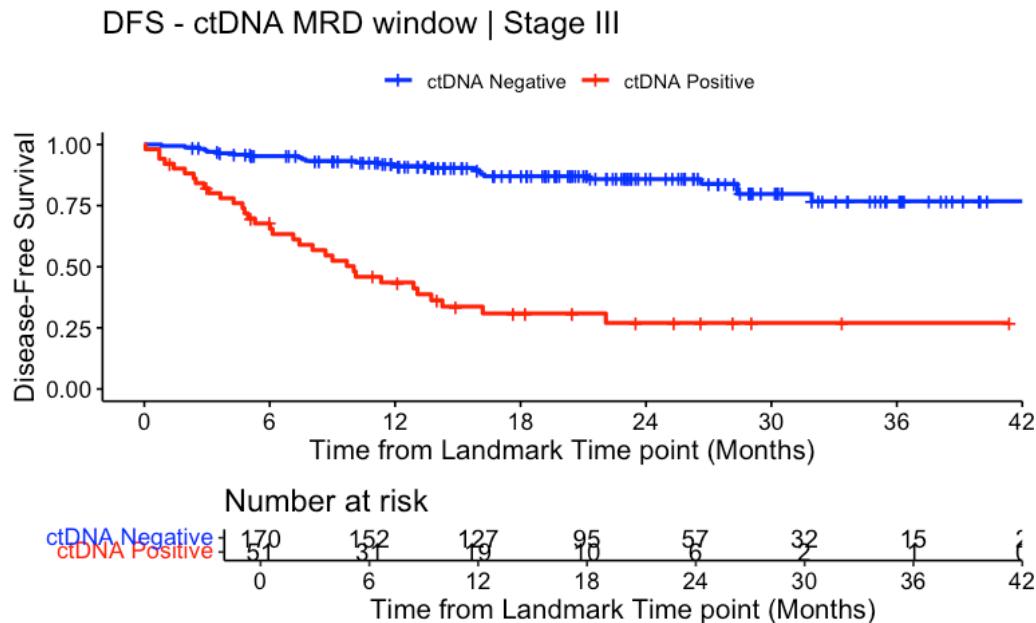
```
event_summary <- circ_data %>%
  group_by(ctDNA.MRD) %>%
  summarise(
    Total = n(),
    Events = sum(DFS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)
```

ctDNA.MRD	Total	Events	Fraction	Percentage
<chr>	<int>	<int>	<dbl>	<dbl>
NEGATIVE	170	24	0.1411765	14.11765
POSITIVE	51	33	0.6470588	64.70588

2 rows

Hide

```
surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
KM_curve <- survfit(surv_object ~ ctDNA.MRD, data = circ_data, conf.int=0.95, conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette=c("blue","red"), title="DFS - ctDNA MRD window | Stage III", ylab= "Disease-Free Survival", xlab="Time from Landmark Time point (Months)", legend.labs=c("ctDNA Negative", "ctDNA Positive"), legend.title="")
```



Hide

```
summary(KM_curve, times= c(24))
```

```
Call: survfit(formula = surv_object ~ ctDNA.MRD, data = circ_data,
  conf.int = 0.95, conf.type = "log-log")
```

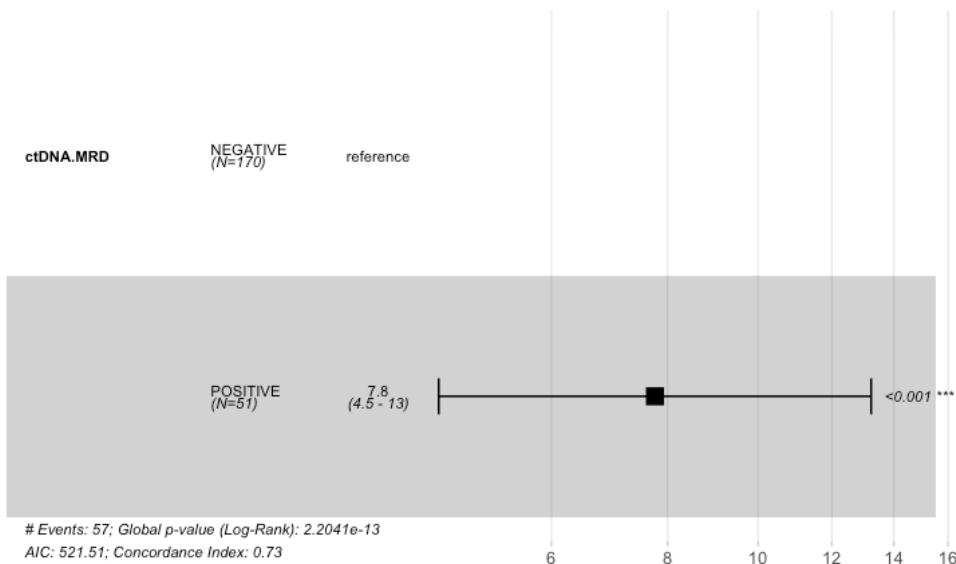
ctDNA.MRD=NEGATIVE						
time	n.risk	n.event	survival	std.err	lower	95% CI upper
24.0000	57.0000	20.0000	0.8582	0.0303	0.7863	0.9073

ctDNA.MRD=POSITIVE						
time	n.risk	n.event	survival	std.err	lower	95% CI upper
24.0000	6.0000	33.0000	0.2704	0.0714	0.1433	0.4146

Hide

```
circ_data$ctDNA.MRD <- factor(circ_data$ctDNA.MRD, levels=c("NEGATIVE","POSITIVE"))
cox_fit <- coxph(surv_object ~ ctDNA.MRD, data=circ_data)
ggforest(cox_fit,data = circ_data)
```

Hazard ratio



Hide

```
summary(cox_fit)
```

Call:
`coxph(formula = surv_object ~ ctDNA.MRD, data = circ_data)`

n= 221, number of events= 57

	coef	exp(coef)	se(coef)	z	Pr(> z)						
ctDNA.MRDPOSITIVE	2.0479	7.7515	0.2728	7.506	6.08e-14 ***						

Signif. codes:	0	'***'	0.001	'**'	0.01	'*'	0.05	.'	0.1	' '	1

	exp(coef)	exp(-coef)	lower .95	upper .95
ctDNA.MRDPOSITIVE	7.752	0.129	4.541	13.23

Concordance= 0.731 (se = 0.031)
Likelihood ratio test= 53.81 on 1 df, p=2e-13
Wald test = 56.35 on 1 df, p=6e-14
Score (logrank) test = 77.35 on 1 df, p=<2e-16

Hide

```
cox_fit_summary <- summary(cox_fit)

# Extract values for HR, 95% CI, and p-value
HR <- cox_fit_summary$coefficients[2]
lower_CI <- cox_fit_summary$conf.int[3]
upper_CI <- cox_fit_summary$conf.int[4]
p_value <- cox_fit_summary$coefficients[5]
label_text <- paste0("HR = ", round(HR, 2), " (", round(lower_CI, 2), " - ", round(upper_CI, 2), "); p = ", round(p_value, 3))
print(label_text)
```

```
[1] "HR = 7.75 (4.54-13.23); p = 0"
```

#DFS by ctDNA at the MRD Window - Stage II & Risk Groups

Hide

```

rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA CRC_Clinical Data 122023.csv")
circ_data <- circ_data[circ_data$CLIA.CRC==TRUE,]
circ_data <- circ_data[circ_data$ctDNA.MRD!="",]
circ_data <- circ_data[!(circ_data$Stage %in% c("I", "III")),]
circ_data <- circ_data[circ_data$Risk.Group!="",]
circ_data$DFS.months=circ_data$DFS.months-2.5
circ_data <- circ_data[circ_data$DFS.months>=0,]

circ_data$ctDNA.Stage.II.Risk <- NA #first we create the variable for the ctDNA & NAC combination, and we assign
values
circ_data <- circ_data %>%
  mutate(ctDNA.Stage.II.Risk = case_when(
    ctDNA.MRD == "NEGATIVE" & Risk.Group == "Low" ~ 1,
    ctDNA.MRD == "POSITIVE" & Risk.Group == "Low" ~ 2,
    ctDNA.MRD == "NEGATIVE" & Risk.Group == "High" ~ 3,
    ctDNA.MRD == "POSITIVE" & Risk.Group == "High" ~ 4
  ))
  
circ_data$DFS.Event <- as.logical(circ_data$DFS.Event)
circ_data$DFS.months <- as.numeric(circ_data$DFS.months)
circ_data <- circ_data[!is.na(circ_data$ctDNA.Stage.II.Risk), ]
survfit(Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)~ctDNA.Stage.II.Risk, data = circ_data)

```

Call: survfit(formula = Surv(time = circ_data\$DFS.months, event = circ_data\$DFS.Event) ~
ctDNA.Stage.II.Risk, data = circ_data)

	n	events	median	0.95LCL	0.95UCL
ctDNA.Stage.II.Risk=1	30	1	NA	NA	NA
ctDNA.Stage.II.Risk=2	2	0	NA	NA	NA
ctDNA.Stage.II.Risk=3	110	4	NA	NA	NA
ctDNA.Stage.II.Risk=4	8	3	NA	11.1	NA

Hide

```

event_summary <- circ_data %>%
  group_by(ctDNA.Stage.II.Risk) %>%
  summarise(
    Total = n(),
    Events = sum(DFS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)

```

ctDNA.Stage.II.Risk	Total	Events	Fraction	Percentage
<dbl>	<int>	<int>	<dbl>	<dbl>
1	30	1	0.03333333	3.333333
2	2	0	0.00000000	0.000000
3	110	4	0.03636364	3.636364
4	8	3	0.37500000	37.500000

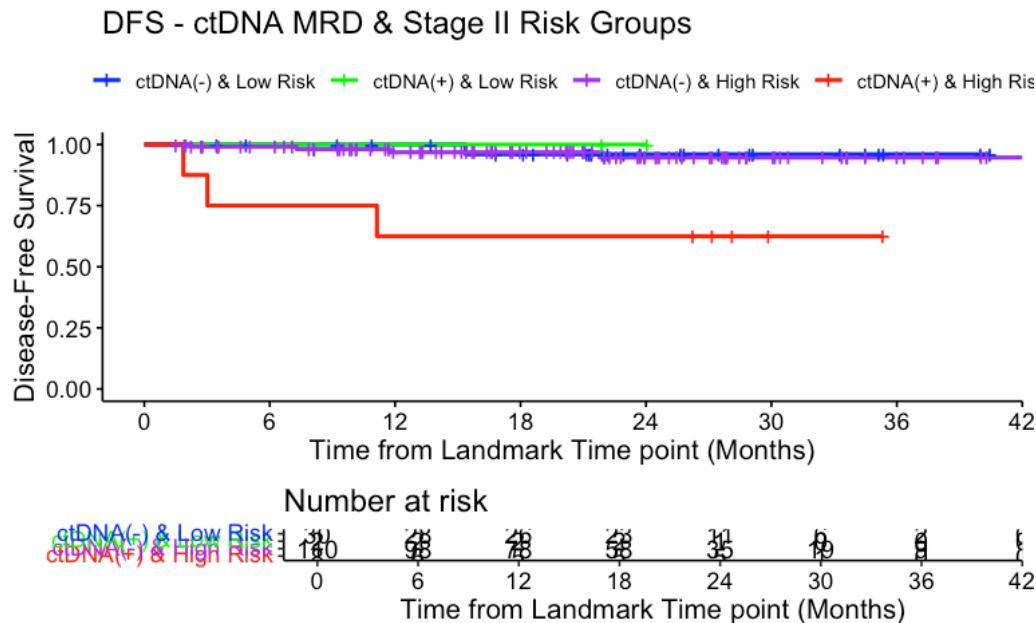
4 rows

Hide

```

surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
KM_curve <- survfit(surv_object ~ ctDNA.Stage.II.Risk, data = circ_data, conf.int=0.95, conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette = c("blue", "green", "purple", "red"), title="DFS - ctDNA MRD & Stage II Risk Groups", ylab= "Disease-Free Survival", xlab="Time from Landmark Time point (Months)", legend.labs=c("ctDNA(-) & Low Risk", "ctDNA(+) & Low Risk", "ctDNA(-) & High Risk", "ctDNA(+) & High Risk"), legend.title=""

```



Hide

```
summary(KM_curve, times= c(24))
```

```
Call: survfit(formula = surv_object ~ ctDNA.Stage.II.Risk, data = circ_data,
  conf.int = 0.95, conf.type = "log-log")

  ctDNA.Stage.II.Risk=1
  time    n.risk    n.event    survival    std.err lower 95% CI upper 95% CI
  24.0000  11.0000    1.0000    0.9600    0.0392    0.7484    0.9943

  ctDNA.Stage.II.Risk=2
  time    n.risk    n.event    survival    std.err lower 95% CI upper 95% CI
  24.0000    1.0000    0.0000    1.0000    0.0000      NA      NA

  ctDNA.Stage.II.Risk=3
  time    n.risk    n.event    survival    std.err lower 95% CI upper 95% CI
  24.0000  35.0000    4.0000    0.9465    0.0278    0.8549    0.9809

  ctDNA.Stage.II.Risk=4
  time    n.risk    n.event    survival    std.err lower 95% CI upper 95% CI
  24.0000    5.0000    3.0000    0.6250    0.1710    0.2290    0.8610
```

Hide

```

circ_data$ctDNA.Stage.II.Risk <- factor(circ_data$ctDNA.Stage.II.Risk, levels=c("1","2","3","4"), labels = c("ctDNA(-) & Low Risk", "ctDNA(+) & Low Risk", "ctDNA(-) & High Risk", "ctDNA(+) & High Risk"))
cox_fit <- coxphf(surv_object ~ ctDNA.Stage.II.Risk, data=circ_data)
summary(cox_fit)

```

```

coxphf(formula = surv_object ~ ctDNA.Stage.II.Risk, data = circ_data)

Model fitted by Penalized ML
Confidence intervals and p-values by Profile Likelihood

            coef se(coef)  exp(coef) lower 0.95 upper 0.95      Chisq
p
ctDNA.Stage.II.RiskctDNA(+) & Low Risk  1.54746514 1.827452  4.6995424 0.03219659  88.15970 0.670679229 0.41281
490
ctDNA.Stage.II.RiskctDNA(-) & High Risk -0.06180044 1.054866  0.9400705 0.17388788  9.38239 0.004245486 0.94804
868
ctDNA.Stage.II.RiskctDNA(+) & High Risk  2.30998495 1.092244 10.0742730 1.65326183 104.09858 6.196660240 0.01279
916

Likelihood ratio test=9.240477 on 3 df, p=0.02625872, n=150
Wald test = 9.971067 on 3 df, p = 0.01881368

Covariance-Matrix:
            ctDNA.Stage.II.RiskctDNA(+) & Low Risk ctDNA.Stage.II.RiskctDNA(-) & High Risk
Risk ctDNA.Stage.II.RiskctDNA(+) & High Risk
ctDNA.Stage.II.RiskctDNA(+) & Low Risk           3.3395808 0.83
33478 0.8321149
ctDNA.Stage.II.RiskctDNA(-) & High Risk           0.8333478 1.11
27416 0.8335422
ctDNA.Stage.II.RiskctDNA(+) & High Risk           0.8321149 0.83
35422 1.1929966

```

#DFS by ctDNA at the MRD Window - Stage III & Risk Groups

Hide

```

rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA_CRC_Clinical Data 122023.csv")
circ_data <- circ_data[circ_data$CLIA.CRC==TRUE,]
circ_data <- circ_data[circ_data$ctDNA.MRD!="",]
circ_data <- circ_data[!(circ_data$Stage %in% c("I", "II")),]
circ_data$DFS.months=circ_data$DFS.months-2.5
circ_data <- circ_data[circ_data$DFS.months>=0,]

circ_data$ctDNA.Stage.III.Risk <- NA #first we create the variable for the ctDNA & NAC combination, and we assign
values
circ_data <- circ_data %>%
  mutate(ctDNA.Stage.III.Risk = case_when(
    ctDNA.MRD == "NEGATIVE" & Risk.Group == "Low" ~ 1,
    ctDNA.MRD == "POSITIVE" & Risk.Group == "Low" ~ 2,
    ctDNA.MRD == "NEGATIVE" & Risk.Group == "High" ~ 3,
    ctDNA.MRD == "POSITIVE" & Risk.Group == "High" ~ 4
  ))
  
circ_data$DFS.Event <- as.logical(circ_data$DFS.Event)
circ_data$DFS.months <- as.numeric(circ_data$DFS.months)
circ_data <- circ_data[!is.na(circ_data$ctDNA.Stage.III.Risk), ]
survfit(Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)~ctDNA.Stage.III.Risk, data = circ_data)

```

```

Call: survfit(formula = Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event) ~
  ctDNA.Stage.III.Risk, data = circ_data)

```

	n	events	median	0.95LCL	0.95UCL
ctDNA.Stage.III.Risk=1	103	10	NA	NA	NA
ctDNA.Stage.III.Risk=2	13	9	12.9	6.14	NA
ctDNA.Stage.III.Risk=3	67	14	NA	NA	NA
ctDNA.Stage.III.Risk=4	38	24	9.0	6.01	NA

Hide

```
event_summary <- circ_data %>%
  group_by(ctDNA.Stage.III.Risk) %>%
  summarise(
    Total = n(),
    Events = sum(DFS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)
```

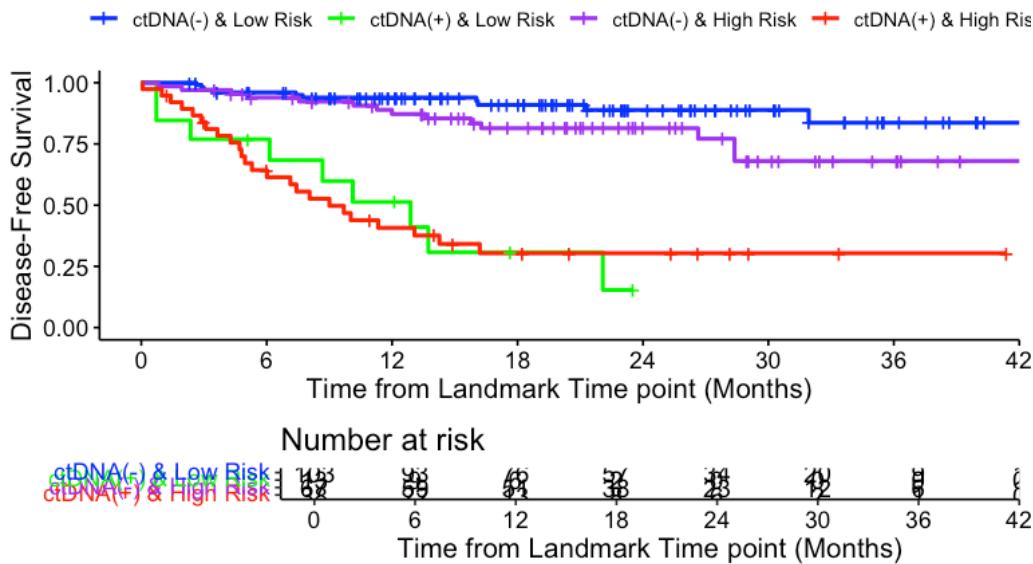
ctDNA.Stage.III.Risk <dbl>	Total <int>	Events <int>	Fraction <dbl>	Percentage <dbl>
1	103	10	0.09708738	9.708738
2	13	9	0.69230769	69.230769
3	67	14	0.20895522	20.895522
4	38	24	0.63157895	63.157895

4 rows

[Hide](#)

```
surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
KM_curve <- survfit(surv_object ~ ctDNA.Stage.III.Risk, data = circ_data, conf.int=0.95, conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette =c("blue","green","purple", "red"), title="DFS - ctDNA MRD & Stage III Risk Groups", ylab= "Disease-Free Survival", xlab="Time from Landmark Time point (Months)", legend.labs=c("ctDNA(-) & Low Risk", "ctDNA(+) & Low Risk", "ctDNA(-) & High Risk", "ctDNA(+) & High Risk"), legend.title="")
```

DFS - ctDNA MRD & Stage III Risk Groups



Hide

```
summary(KM_curve, times= c(18, 24))
```

```
Call: survfit(formula = surv_object ~ ctDNA.Stage.III.Risk, data = circ_data,
  conf.int = 0.95, conf.type = "log-log")
```

```
ctDNA.Stage.III.Risk=1
time n.risk n.event survival std.err lower 95% CI upper 95% CI
 18     57      8    0.909  0.0312    0.825    0.954
 24     34      1    0.889  0.0367    0.791    0.942
```

```
ctDNA.Stage.III.Risk=2
time      n.risk      n.event      survival      std.err      lower 95% CI      upper 95% CI
18.0000  2.0000  8.0000  0.3077  0.1417  0.0793  0.5780
```

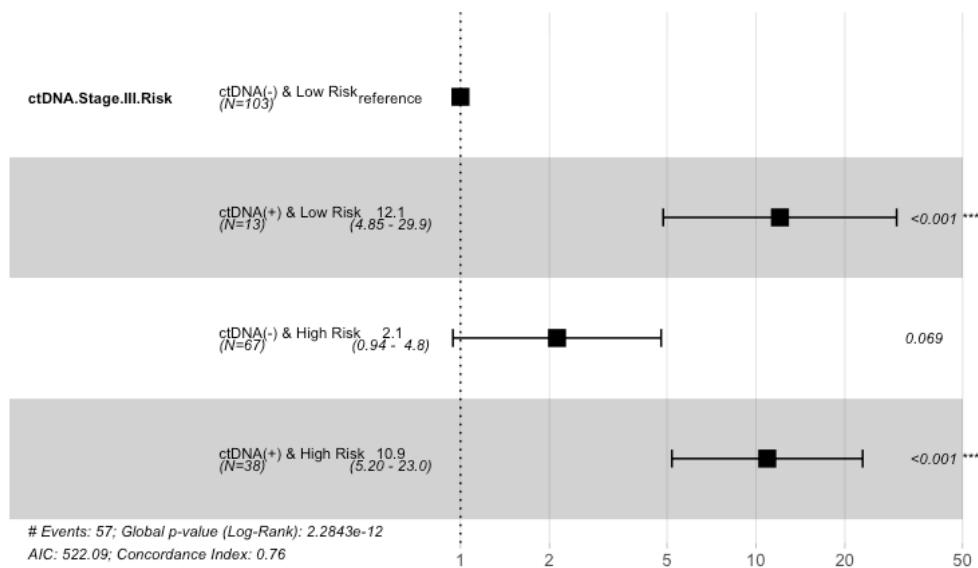
```
ctDNA.Stage.III.Risk=3
time n.risk n.event survival std.err lower 95% CI upper 95% CI
 18     38      11    0.814  0.0512    0.688    0.893
 24     23      0    0.814  0.0512    0.688    0.893
```

```
ctDNA.Stage.III.Risk=4
time n.risk n.event survival std.err lower 95% CI upper 95% CI
 18     8      24    0.304  0.0812    0.157    0.464
 24     6      0    0.304  0.0812    0.157    0.464
```

Hide

```
circ_data$ctDNA.Stage.III.Risk <- factor(circ_data$ctDNA.Stage.III.Risk, levels=c("1","2","3","4"), labels = c("ctDNA(-) & Low Risk", "ctDNA(+) & Low Risk", "ctDNA(-) & High Risk", "ctDNA(+) & High Risk"))
cox_fit <- coxph(surv_object ~ ctDNA.Stage.III.Risk, data=circ_data)
ggforest(cox_fit, data = circ_data)
```

Hazard ratio



Hide

summary(cox_fit)

Call:
coxph(formula = surv_object ~ ctDNA.Stage.III.Risk, data = circ_data)

n= 221, number of events= 57

	coef	exp(coef)	se(coef)	z	Pr(> z)
ctDNA.Stage.III.Risk	ctDNA(+) & Low Risk	2.4892	12.0518	0.4642	5.362 8.23e-08 ***
ctDNA.Stage.III.Risk	ctDNA(-) & High Risk	0.7524	2.1221	0.4141	1.817 0.0692 .
ctDNA.Stage.III.Risk	ctDNA(+) & High Risk	2.3908	10.9222	0.3791	6.307 2.85e-10 ***

Signif. codes:	0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				
		exp(coef)	exp(-coef)	lower .95	upper .95
	ctDNA.Stage.III.Risk	ctDNA(+) & Low Risk	12.052	0.08297	4.8517 29.937
	ctDNA.Stage.III.Risk	ctDNA(-) & High Risk	2.122	0.47123	0.9425 4.778
	ctDNA.Stage.III.Risk	ctDNA(+) & High Risk	10.922	0.09156	5.1955 22.961

Concordance= 0.757 (se = 0.033)

Likelihood ratio test= 57.24 on 3 df, p=2e-12

Wald test = 55.9 on 3 df, p=4e-12

Score (logrank) test = 79.25 on 3 df, p=<2e-16

Hide

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA_CRC_Clinical Data 122023.csv")
circ_data <- circ_data[circ_data$CLIA.CRC==TRUE,]
circ_data <- circ_data[circ_data$ctDNA.MRD!="",]
circ_data <- circ_data[!(circ_data$Stage %in% c("I", "II")),]
circ_data$DFS.months=circ_data$DFS.months-2.5
circ_data <- circ_data[circ_data$DFS.months>=0,]

circ_data$ctDNA.Stage.III.Risk <- NA #first we create the variable for the ctDNA & NAC combination, and we assign values
circ_data <- circ_data %>%
  mutate(ctDNA.Stage.III.Risk = case_when(
    ctDNA.MRD == "NEGATIVE" & Risk.Group == "Low" ~ 1,
    ctDNA.MRD == "POSITIVE" & Risk.Group == "Low" ~ 2,
    ctDNA.MRD == "NEGATIVE" & Risk.Group == "High" ~ 3,
    ctDNA.MRD == "POSITIVE" & Risk.Group == "High" ~ 4
  ))

circ_data$DFS.Event <- as.logical(circ_data$DFS.Event)
circ_data$DFS.months <- as.numeric(circ_data$DFS.months)
survfit(Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)~ctDNA.Stage.III.Risk, data = circ_data)
```

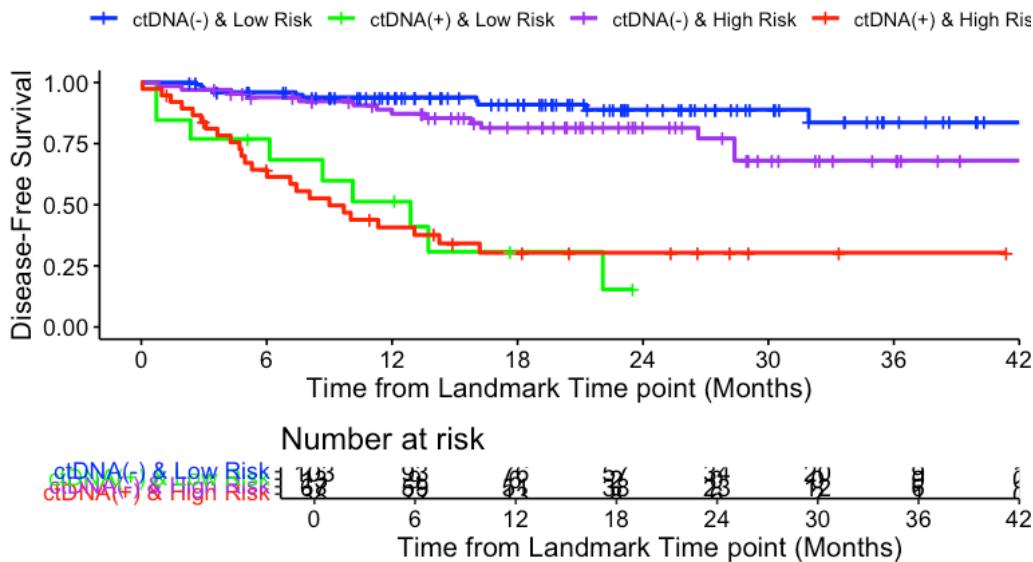
```
Call: survfit(formula = Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event) ~
  ctDNA.Stage.III.Risk, data = circ_data)
```

	n	events	median	0.95LCL	0.95UCL
ctDNA.Stage.III.Risk=1	103	10	NA	NA	NA
ctDNA.Stage.III.Risk=2	13	9	12.9	6.14	NA
ctDNA.Stage.III.Risk=3	67	14	NA	NA	NA
ctDNA.Stage.III.Risk=4	38	24	9.0	6.01	NA

Hide

```
surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
KM_curve <- survfit(surv_object ~ ctDNA.Stage.III.Risk, data = circ_data, conf.int=0.95, conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette
e=c("blue","green","purple","red"), title="DFS - ctDNA MRD & Stage III Risk Groups", ylab= "Disease-Free Survival", xlab= "Time from Landmark Time point (Months)", legend.labs=c("ctDNA(-) & Low Risk", "ctDNA(+) & Low Risk", "ctDNA(-) & High Risk", "ctDNA(+) & High Risk"), legend.title="")
```

DFS - ctDNA MRD & Stage III Risk Groups



Hide

```
summary(KM_curve, times= c(18))
```

```
Call: survfit(formula = surv_object ~ ctDNA.Stage.III.Risk, data = circ_data,
  conf.int = 0.95, conf.type = "log-log")
```

ctDNA.Stage.III.Risk=1		survival		std.err		95% CI		upper 95% CI	
time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI	95% CI	upper 95% CI	
18.0000	57.0000	8.0000	0.9095	0.0312	0.8248	0.9543	0.9543	0.9543	

ctDNA.Stage.III.Risk=2		survival		std.err		95% CI		upper 95% CI	
time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI	95% CI	upper 95% CI	
18.0000	2.0000	8.0000	0.3077	0.1417	0.0793	0.5780	0.5780	0.5780	

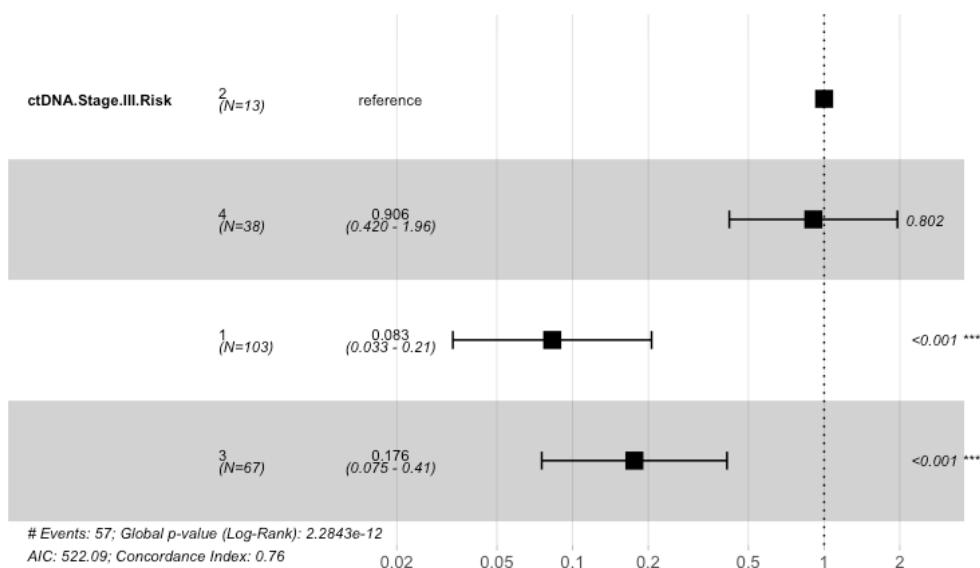
ctDNA.Stage.III.Risk=3		survival		std.err		95% CI		upper 95% CI	
time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI	95% CI	upper 95% CI	
18.0000	38.0000	11.0000	0.8141	0.0512	0.6878	0.8932	0.8932	0.8932	

ctDNA.Stage.III.Risk=4		survival		std.err		95% CI		upper 95% CI	
time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI	95% CI	upper 95% CI	
18.0000	8.0000	24.0000	0.3038	0.0812	0.1574	0.4641	0.4641	0.4641	

Hide

```
circ_data$ctDNA.Stage.III.Risk <- factor(circ_data$ctDNA.Stage.III.Risk, levels=c("2","4","1","3"))
cox_fit <- coxph(surv_object ~ ctDNA.Stage.III.Risk, data=circ_data)
ggforest(cox_fit, data = circ_data)
```

Hazard ratio



Hide

```
summary(cox_fit)
```

Call:
`coxph(formula = surv_object ~ ctDNA.Stage.III.Risk, data = circ_data)`

n= 221, number of events= 57

	coef	exp(coef)	se(coef)	z	Pr(> z)
ctDNA.Stage.III.Risk4	-0.09842	0.90627	0.39258	-0.251	0.802
ctDNA.Stage.III.Risk1	-2.48922	0.08297	0.46424	-5.362	8.23e-08 ***
ctDNA.Stage.III.Risk3	-1.73680	0.17608	0.43279	-4.013	5.99e-05 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95
ctDNA.Stage.III.Risk4	0.90627	1.103	0.41985	1.9562
ctDNA.Stage.III.Risk1	0.08297	12.052	0.03340	0.2061
ctDNA.Stage.III.Risk3	0.17608	5.679	0.07539	0.4113

Concordance= 0.757 (se = 0.033)
 Likelihood ratio test= 57.24 on 3 df, p=2e-12
 Wald test = 55.9 on 3 df, p=4e-12
 Score (logrank) test = 79.25 on 3 df, p=<2e-16

#OS by ctDNA at the MRD Window - all stages

Hide

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA_CRC_Clinical Data 122023.csv")
circ_data <- circ_data[circ_data$CLIA.CRC==TRUE,]
circ_data <- circ_data[circ_data$ctDNA.MRD!=""]
circ_data$OS.Event <- as.logical(circ_data$OS.Event)
circ_data$OS.months <- as.numeric(circ_data$OS.months)
circ_data$DFS.months=circ_data$OS.months-2.5
circ_data <- circ_data[circ_data$OS.months>=0,]
circ_dataadf <- as.data.frame(circ_data)

survfit(Surv(time = circ_data$OS.months, event = circ_data$OS.Event)~ctDNA.MRD, data = circ_data)
```

Call: `survfit(formula = Surv(time = circ_data$OS.months, event = circ_data$OS.Event) ~ ctDNA.MRD, data = circ_data)`

	n	events	median	0.95LCL	0.95UCL
ctDNA.MRD=NEGATIVE	334	8	NA	NA	NA
ctDNA.MRD=POSITIVE	71	7	40.4	40.4	NA

Hide

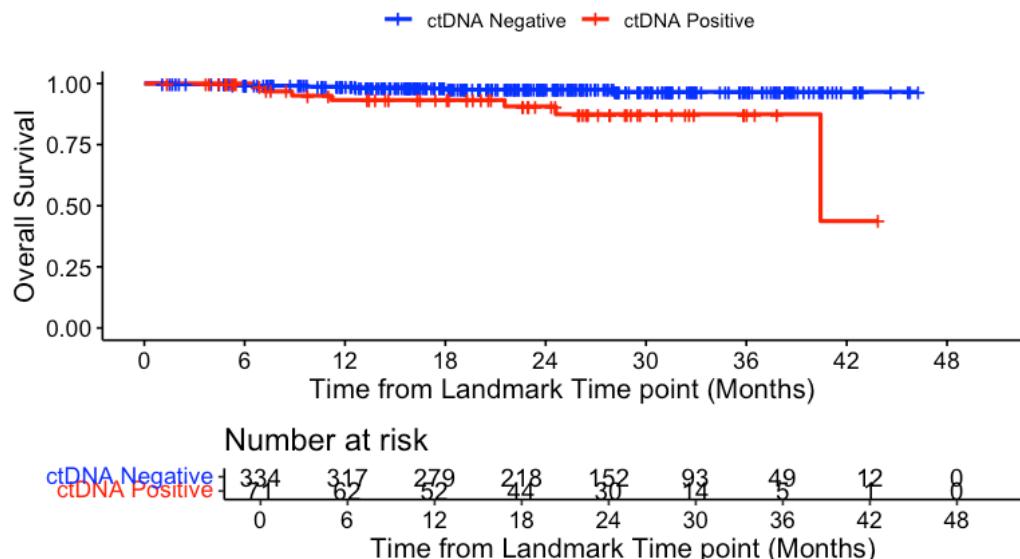
```
event_summary <- circ_data %>%
  group_by(ctDNA.MRD) %>%
  summarise(
    Total = n(),
    Events = sum(OS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)
```

ctDNA.MRD	Total	Events	Fraction	Percentage
<chr>	<int>	<int>	<dbl>	<dbl>
NEGATIVE	334	8	0.02395210	2.395210
POSITIVE	71	7	0.09859155	9.859155
2 rows				

Hide

```
surv_object <- Surv(time = circ_data$OS.months, event = circ_data$OS.Event)
KM_curve <- survfit(surv_object ~ ctDNA.MRD, data = circ_data, conf.int=0.95, conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette=c("blue","red"), title="OS - ctDNA MRD window | All pts", ylab= "Overall Survival", xlab="Time from Landmark Time point (Months)", legend.labs=c("ctDNA Negative", "ctDNA Positive"), legend.title="")
```

OS - ctDNA MRD window | All pts



Hide

```
summary(KM_curve, times= c(24))
```

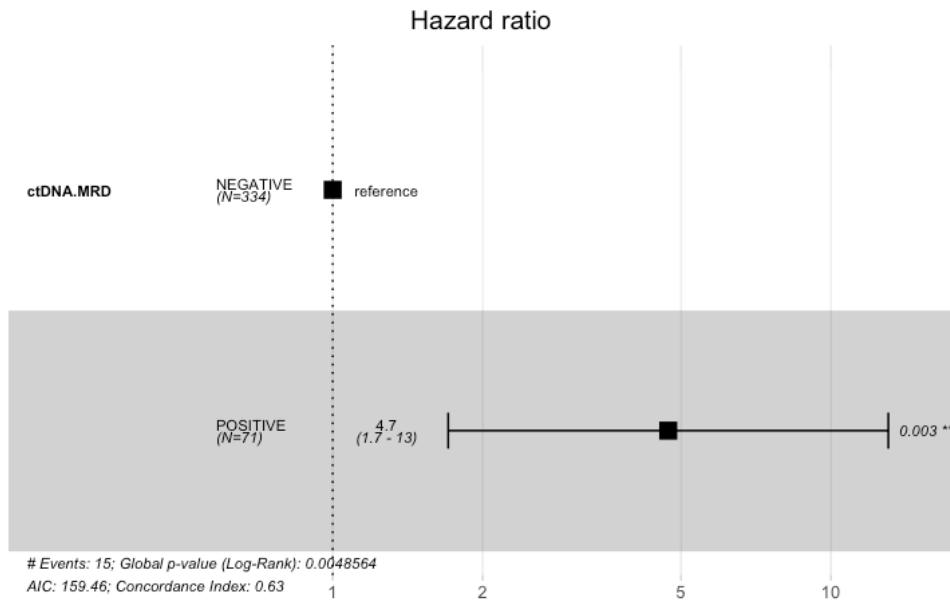
```
Call: survfit(formula = surv_object ~ ctDNA.MRD, data = circ_data,
  conf.int = 0.95, conf.type = "log-log")

  ctDNA.MRD=NEGATIVE
  time    n.risk    n.event    survival    std.err lower 95% CI upper 95% CI
  2.40e+01    1.52e+02    7.00e+00    9.75e-01    9.49e-03    9.48e-01    9.88e-01

  ctDNA.MRD=POSITIVE
  time    n.risk    n.event    survival    std.err lower 95% CI upper 95% CI
  24.0000    30.0000    5.0000    0.9064    0.0408    0.7856    0.9607
```

Hide

```
circ_data$ctDNA.MRD <- factor(circ_data$ctDNA.MRD, levels=c("NEGATIVE","POSITIVE"))
cox_fit <- coxph(surv_object ~ ctDNA.MRD, data=circ_data)
ggforest(cox_fit,data = circ_data)
```



Hide

```
summary(cox_fit)
```

```
Call:
coxph(formula = surv_object ~ ctDNA.MRD, data = circ_data)

n= 405, number of events= 15

      coef exp(coef) se(coef)   z Pr(>|z|)
ctDNA.MRDPOSITIVE 1.551     4.717    0.519 2.988   0.0028 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef) exp(-coef) lower .95 upper .95
ctDNA.MRDPOSITIVE     4.717      0.212     1.705    13.04

Concordance= 0.633  (se = 0.068 )
Likelihood ratio test= 7.93  on 1 df,  p=0.005
Wald test             = 8.93  on 1 df,  p=0.003
Score (logrank) test = 10.85 on 1 df,  p=0.001
```

Hide

```
cox_fit_summary <- summary(cox_fit)

# Extract values for HR, 95% CI, and p-value
HR <- cox_fit_summary$coefficients[2]
lower_CI <- cox_fit_summary$conf.int[3]
upper_CI <- cox_fit_summary$conf.int[4]
p_value <- cox_fit_summary$coefficients[5]
label_text <- paste0("HR = ", round(HR, 2), " (", round(lower_CI, 2), "-", round(upper_CI, 2), "); p = ", round(p_value, 3))
print(label_text)
```

```
[1] "HR = 4.72 (1.71-13.04); p = 0.003"
```

#DFS by ctDNA at the Surveillance Window - all stages

Hide

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA_CRC_Clinical Data 122023.csv")
circ_data <- circ_data[circ_data$CLIA.CRC==TRUE,]
circ_data <- circ_data[circ_data$ctDNA.Surveillance!="",]
circ_data$DFS.months=circ_data$DFS.months-2.5
circ_data <- circ_data[circ_data$DFS.months>=0,]
circ_datadf <- as.data.frame(circ_data)

circ_data$DFS.Event <- as.logical(circ_data$DFS.Event)
circ_data$DFS.months <- as.numeric(circ_data$DFS.months)
survfit(Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event) ~
  ctDNA.Surveillance, data = circ_data)
```

Call: survfit(formula = Surv(time = circ_data\$DFS.months, event = circ_data\$DFS.Event) ~
 ctDNA.Surveillance, data = circ_data)

	n	events	median	0.95LCL	0.95UCL
ctDNA.Surveillance=NEGATIVE	540	20	NA	NA	NA
ctDNA.Surveillance=POSITIVE	83	54	16.2	13.7	28.4

Hide

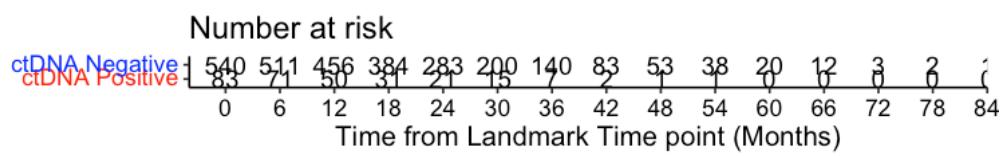
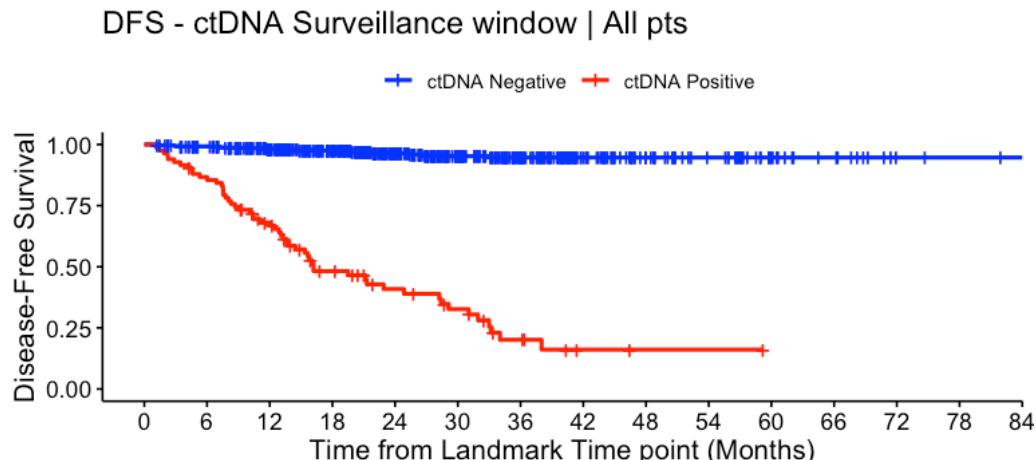
```
event_summary <- circ_data %>%
  group_by(ctDNA.Surveillance) %>%
  summarise(
    Total = n(),
    Events = sum(DFS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)
```

ctDNA.Surveillance	Total	Events	Fraction	Percentage
<chr>	<int>	<int>	<dbl>	<dbl>
NEGATIVE	540	20	0.03703704	3.703704
POSITIVE	83	54	0.65060241	65.060241

2 rows

Hide

```
surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
KM_curve <- survfit(surv_object ~ ctDNA.Surveillance, data = circ_data, conf.int=0.95, conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette=c("blue","red"), title="DFS - ctDNA Surveillance window | All pts", ylab= "Disease-Free Survival", xlab="Time from Landmark Time point (Months)", legend.labs=c("ctDNA Negative", "ctDNA Positive"), legend.title="")
```



Hide

```
summary(KM_curve, times= c(24))
```

```
Call: survfit(formula = surv_object ~ ctDNA.Surveillance, data = circ_data,
  conf.int = 0.95, conf.type = "log-log")
```

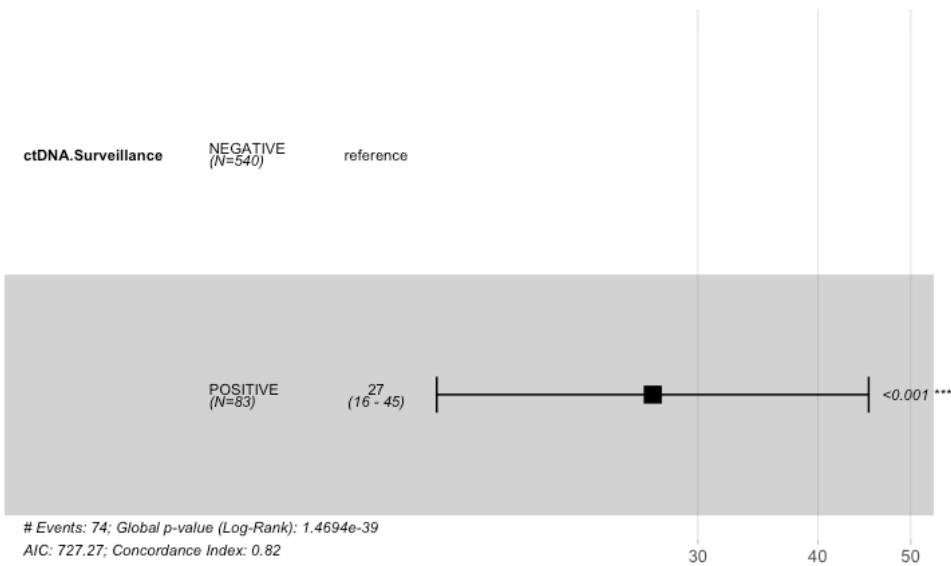
ctDNA.Surveillance=NEGATIVE						
time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
2.40e+01	2.83e+02	1.60e+01	9.64e-01	9.06e-03	9.41e-01	9.78e-01

ctDNA.Surveillance=POSITIVE						
time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
24.0000	21.0000	44.0000	0.4087	0.0597	0.2916	0.5222

Hide

```
circ_data$ctDNA.Surveillance <- factor(circ_data$ctDNA.Surveillance, levels=c("NEGATIVE","POSITIVE"))
cox_fit <- coxph(surv_object ~ ctDNA.Surveillance, data=circ_data)
ggforest(cox_fit, data = circ_data)
```

Hazard ratio



Hide

```
summary(cox_fit)
```

Call:
`coxph(formula = surv_object ~ ctDNA.Surveillance, data = circ_data)`

n= 623, number of events= 74

	coef	exp(coef)	se(coef)	z	Pr(> z)
ctDNA.SurveillancePOSITIVE	3.2926	26.9135	0.2645	12.45	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95
ctDNA.SurveillancePOSITIVE	26.91	0.03716	16.03	45.19

Concordance= 0.817 (se = 0.026)
Likelihood ratio test= 173.2 on 1 df, p=<2e-16
Wald test = 155 on 1 df, p=<2e-16
Score (logrank) test = 344 on 1 df, p=<2e-16

Hide

```
cox_fit_summary <- summary(cox_fit)

# Extract values for HR, 95% CI, and p-value
HR <- cox_fit_summary$coefficients[2]
lower_CI <- cox_fit_summary$conf.int[3]
upper_CI <- cox_fit_summary$conf.int[4]
p_value <- cox_fit_summary$coefficients[5]
label_text <- paste0("HR = ", round(HR, 2), " (", round(lower_CI, 2), " - ", round(upper_CI, 2), "); p = ", round(p_value, 3))
print(label_text)
```

```
[1] "HR = 26.91 (16.03-45.19); p = 0"
```

#DFS by ctDNA at the Surveillance Window - Stages II

Hide

```

rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA CRC_Clinical Data 122023.csv")
circ_data <- circ_data[circ_data$CLIA.CRC==TRUE,]
circ_data <- circ_data[!(circ_data$Stage %in% c("I", "III")),]
circ_data <- circ_data[circ_data$ctDNA.Surveillance!="",]
circ_data$DFS.months=circ_data$DFS.months-2.5
circ_data <- circ_data[circ_data$DFS.months>=0,]
circ_dataadf <- as.data.frame(circ_data)

circ_data$DFS.Event <- as.logical(circ_data$DFS.Event)
circ_data$DFS.months <- as.numeric(circ_data$DFS.months)
survfit(Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)~ctDNA.Surveillance, data = circ_data)

```

```

Call: survfit(formula = Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event) ~
  ctDNA.Surveillance, data = circ_data)

n  events median 0.95LCL 0.95UCL
ctDNA.Surveillance=NEGATIVE 198      6      NA      NA      NA
ctDNA.Surveillance=POSITIVE  16      9    21.2    11.1      NA

```

Hide

```

event_summary <- circ_data %>%
  group_by(ctDNA.Surveillance) %>%
  summarise(
    Total = n(),
    Events = sum(DFS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)

```

ctDNA.Surveillance	Total	Events	Fraction	Percentage
<chr>	<int>	<int>	<dbl>	<dbl>
NEGATIVE	198	6	0.03030303	3.030303
POSITIVE	16	9	0.56250000	56.250000
2 rows				

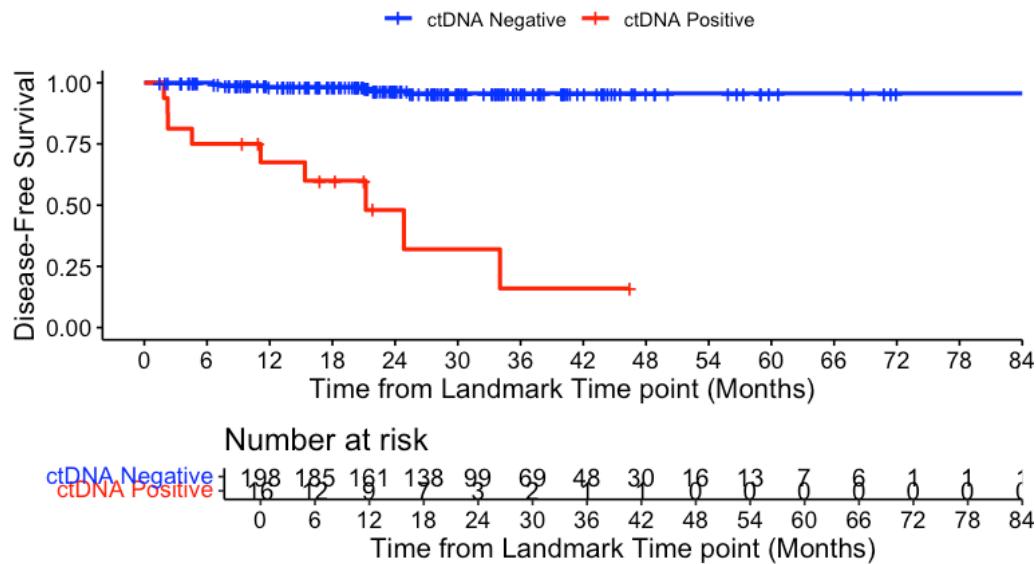
Hide

```

surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
KM_curve <- survfit(surv_object ~ ctDNA.Surveillance, data = circ_data, conf.int=0.95, conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette
e=c("blue","red"), title="DFS - ctDNA Surveillance window | Stage II", ylab= "Disease-Free Survival", xlab="Time
from Landmark Time point (Months)", legend.labs=c("ctDNA Negative", "ctDNA Positive"), legend.title="")

```

DFS - ctDNA Surveillance window | Stage II



Hide

```
summary(KM_curve, times= c(24))
```

Call: survfit(formula = surv_object ~ ctDNA.Surveillance, data = circ_data, conf.int = 0.95, conf.type = "log-log")

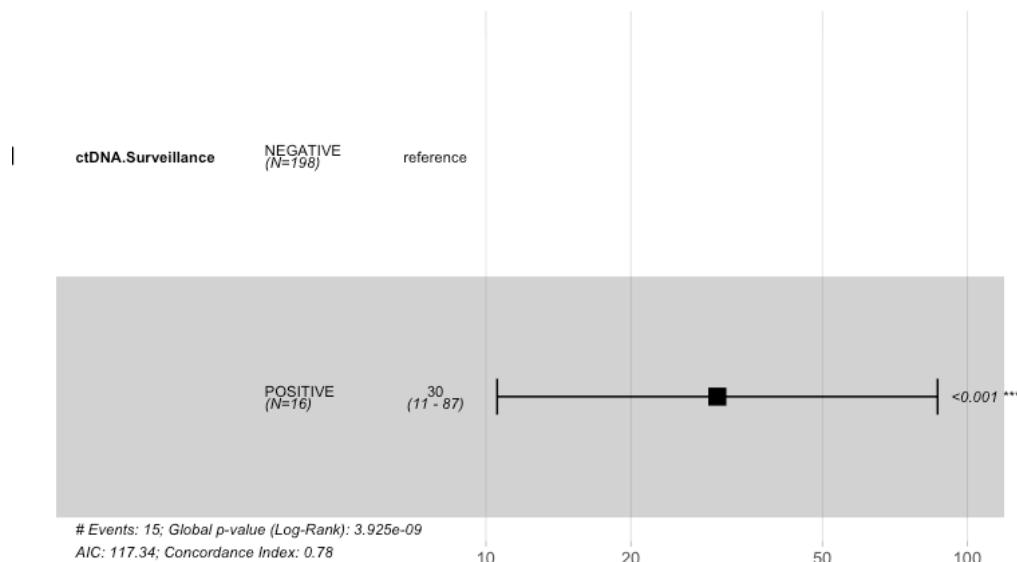
ctDNA.Surveillance=NEGATIVE						
time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
24.000	99.000	5.000	0.967	0.015	0.920	0.986

ctDNA.Surveillance=POSITIVE						
time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
24.000	3.000	7.000	0.480	0.149	0.187	0.725

Hide

```
circ_data$ctDNA.Surveillance <- factor(circ_data$ctDNA.Surveillance, levels=c("NEGATIVE","POSITIVE"))
cox_fit <- coxph(surv_object ~ ctDNA.Surveillance, data=circ_data)
ggforest(cox_fit, data = circ_data)
```

Hazard ratio



Hide

```
summary(cox_fit)
```

```

Call:
coxph(formula = surv_object ~ ctDNA.Surveillance, data = circ_data)

n= 214, number of events= 15

            coef exp(coef)  se(coef)      z Pr(>|z|)
ctDNA.SurveillancePOSITIVE 3.4093   30.2449   0.5373 6.345 2.23e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

            exp(coef) exp(-coef) lower .95 upper .95
ctDNA.SurveillancePOSITIVE   30.24    0.03306    10.55    86.7

Concordance= 0.782 (se = 0.063 )
Likelihood ratio test= 34.66 on 1 df,  p=4e-09
Wald test             = 40.26 on 1 df,  p=2e-10
Score (logrank) test = 94.31 on 1 df,  p=<2e-16

```

Hide

```

cox_fit_summary <- summary(cox_fit)

# Extract values for HR, 95% CI, and p-value
HR <- cox_fit_summary$coefficients[2]
lower_CI <- cox_fit_summary$conf.int[3]
upper_CI <- cox_fit_summary$conf.int[4]
p_value <- cox_fit_summary$coefficients[5]
label_text <- paste0("HR = ", round(HR, 2), " (", round(lower_CI, 2), "-", round(upper_CI, 2), "); p = ", round(p_value, 3))
print(label_text)

```

```
[1] "HR = 30.24 (10.55-86.7); p = 0"
```

#DFS by ctDNA at the Surveillance Window - Stages III

Hide

```

rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA CRC_Clinical Data 122023.csv")
circ_data <- circ_data[circ_data$CLIA.CRC==TRUE,]
circ_data <- circ_data[!(circ_data$Stage %in% c("I", "II")),]
circ_data <- circ_data[circ_data$ctDNA.Surveillance!=""]
circ_data$DFS.months=circ_data$DFS.months-2.5
circ_data <- circ_data[circ_data$DFS.months>=0,]
circ_datadf <- as.data.frame(circ_data)

circ_data$DFS.Event <- as.logical(circ_data$DFS.Event)
circ_data$DFS.months <- as.numeric(circ_data$DFS.months)
survfit(Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)~ctDNA.Surveillance, data = circ_data)

```

```

Call: survfit(formula = Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event) ~
ctDNA.Surveillance, data = circ_data)

```

	n	events	median	0.95LCL	0.95UCL
ctDNA.Surveillance=NEGATIVE	306	13	NA	NA	NA
ctDNA.Surveillance=POSITIVE	64	45	15.8	13.1	28.4

Hide

```

event_summary <- circ_data %>%
  group_by(ctDNA.Surveillance) %>%
  summarise(
    Total = n(),
    Events = sum(DFS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)

```

ctDNA.Surveillance	Total	Events	Fraction	Percentage
<chr>	<int>	<int>	<dbl>	<dbl>
NEGATIVE	306	13	0.04248366	4.248366
POSITIVE	64	45	0.70312500	70.312500

2 rows

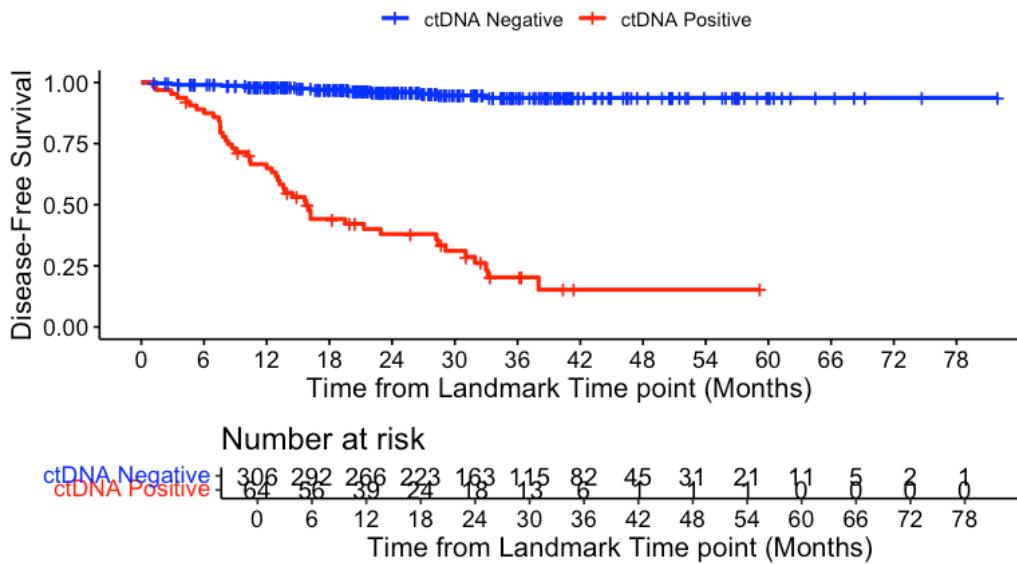
Hide

```

surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
KM_curve <- survfit(surv_object ~ ctDNA.Surveillance, data = circ_data, conf.int=0.95, conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette
e=c("blue","red"), title="DFS - ctDNA Surveillance window | Stage III", ylab= "Disease-Free Survival", xlab="Time
from Landmark Time point (Months)", legend.labs=c("ctDNA Negative", "ctDNA Positive"), legend.title="")

```

DFS - ctDNA Surveillance window | Stage III



Hide

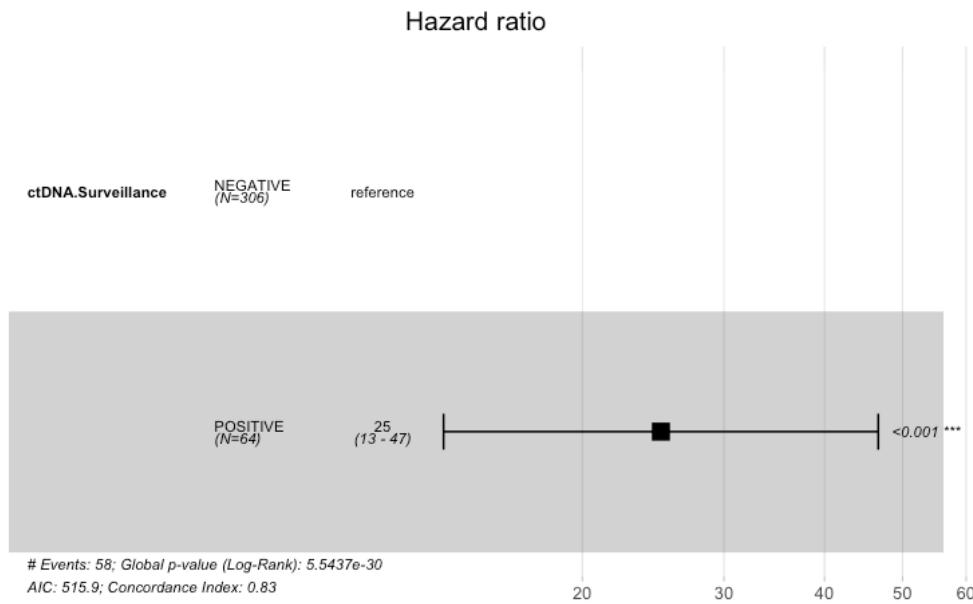
```
summary(KM_curve, times= c(24))
```

```
Call: survfit(formula = surv_object ~ ctDNA.Surveillance, data = circ_data,
conf.int = 0.95, conf.type = "log-log")
```

ctDNA.Surveillance=NEGATIVE						
time	n.risk	n.event	survival	std.err	lower	95% CI upper
24.0000	163.0000	10.0000	0.9610	0.0123	0.9281	0.9790
ctDNA.Surveillance=POSITIVE						
time	n.risk	n.event	survival	std.err	lower	95% CI upper
24.0000	18.0000	37.0000	0.3800	0.0646	0.2553	0.5037

Hide

```
circ_data$ctDNA.Surveillance <- factor(circ_data$ctDNA.Surveillance, levels=c("NEGATIVE","POSITIVE"))
cox_fit <- coxph(surv_object ~ ctDNA.Surveillance, data=circ_data)
ggforest(cox_fit,data = circ_data)
```



Hide

```
summary(cox_fit)
```

```
Call:
coxph(formula = surv_object ~ ctDNA.Surveillance, data = circ_data)

n= 370, number of events= 58

            coef exp(coef)  se(coef)      z Pr(>|z|)
ctDNA.SurveillancePOSITIVE 3.2208   25.0476   0.3173 10.15 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

            exp(coef) exp(-coef) lower .95 upper .95
ctDNA.SurveillancePOSITIVE   25.05     0.03992    13.45    46.65

Concordance= 0.827 (se = 0.027 )
Likelihood ratio test= 129.4 on 1 df,  p=<2e-16
Wald test                 = 103 on 1 df,  p=<2e-16
Score (logrank) test = 222.1 on 1 df,  p=<2e-16
```

Hide

```
cox_fit_summary <- summary(cox_fit)

# Extract values for HR, 95% CI, and p-value
HR <- cox_fit_summary$coefficients[2]
lower_CI <- cox_fit_summary$conf.int[3]
upper_CI <- cox_fit_summary$conf.int[4]
p_value <- cox_fit_summary$coefficients[5]
label_text <- paste0("HR = ", round(HR, 2), " (", round(lower_CI, 2), " - ", round(upper_CI, 2), "); p = ", round(p_value, 3))
print(label_text)
```

```
[1] "HR = 25.05 (13.45-46.65); p = 0"
```

#OS by ctDNA at the Surveillance Window - all stages

Hide

```

rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA_CRC_Clinical Data 122023.csv")
circ_data <- circ_data[circ_data$CLIA.CRC==TRUE,]
circ_data <- circ_data[circ_data$ctDNA.Surveillance!="",]
circ_data$OS.Event <- as.logical(circ_data$OS.Event)
circ_data$OS.months <- as.numeric(circ_data$OS.months)
circ_data$DFS.months=circ_data$OS.months-2.5
circ_data <- circ_data[circ_data$OS.months>=0,]
circ_dataadf <- as.data.frame(circ_data)

survfit(Surv(time = circ_data$OS.months, event = circ_data$OS.Event)~ctDNA.Surveillance, data = circ_data)

```

```

Call: survfit(formula = Surv(time = circ_data$OS.months, event = circ_data$OS.Event) ~
  ctDNA.Surveillance, data = circ_data)

```

	n	events	median	0.95LCL	0.95UCL
ctDNA.Surveillance=NEGATIVE	540	4	NA	NA	NA
ctDNA.Surveillance=POSITIVE	83	4	NA	NA	NA

Hide

```

event_summary <- circ_data %>%
  group_by(ctDNA.Surveillance) %>%
  summarise(
    Total = n(),
    Events = sum(OS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)

```

ctDNA.Surveillance	Total	Events	Fraction	Percentage
<chr>	<int>	<int>	<dbl>	<dbl>
NEGATIVE	540	4	0.007407407	0.7407407
POSITIVE	83	4	0.048192771	4.8192771

2 rows

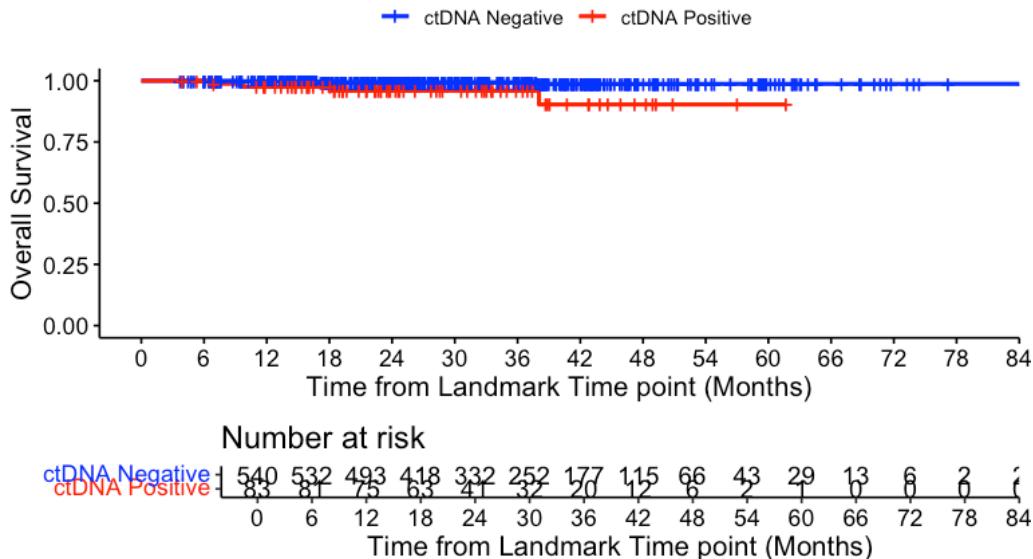
Hide

```

surv_object <- Surv(time = circ_data$OS.months, event = circ_data$OS.Event)
KM_curve <- survfit(surv_object ~ ctDNA.Surveillance, data = circ_data, conf.int=0.95, conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette=c("blue","red"), title="OS - ctDNA Surveillance window | All pts", ylab= "Overall Survival", xlab="Time from Landmark Time point (Months)", legend.labs=c("ctDNA Negative", "ctDNA Positive"), legend.title="")

```

OS - ctDNA Surveillance window | All pts



Hide

```
summary(KM_curve, times= c(24))
```

```
Call: survfit(formula = surv_object ~ ctDNA.Surveillance, data = circ_data,
  conf.int = 0.95, conf.type = "log-log")
```

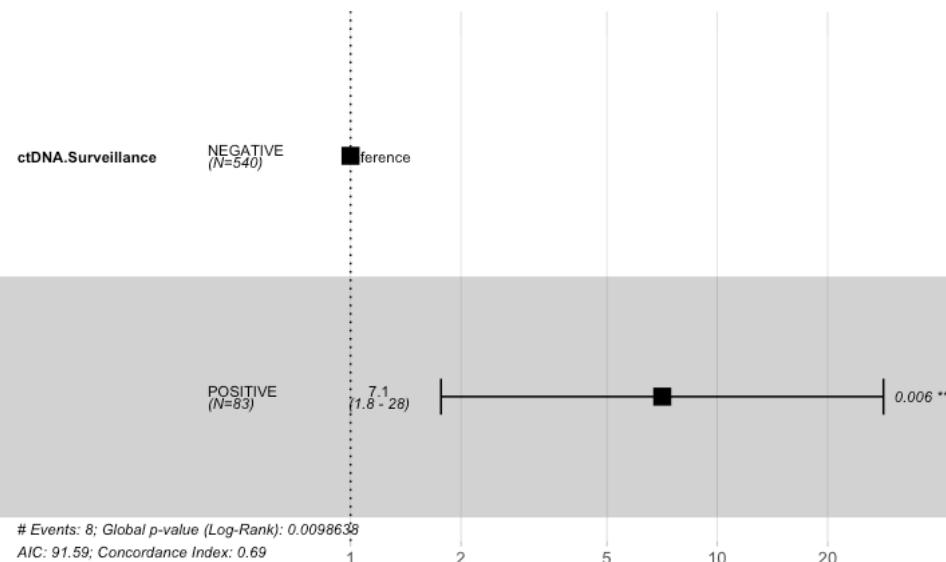
ctDNA.Surveillance=NEGATIVE						
time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
2.40e+01	3.32e+02	3.00e+00	9.94e-01	3.65e-03	9.80e-01	9.98e-01

ctDNA.Surveillance=POSITIVE						
time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
24.0000	41.0000	3.0000	0.9594	0.0231	0.8785	0.9868

Hide

```
circ_data$ctDNA.Surveillance <- factor(circ_data$ctDNA.Surveillance, levels=c("NEGATIVE","POSITIVE"))
cox_fit <- coxph(surv_object ~ ctDNA.Surveillance, data=circ_data)
ggforest(cox_fit, data = circ_data)
```

Hazard ratio



Hide

```
summary(cox_fit)
```

```

Call:
coxph(formula = surv_object ~ ctDNA.Surveillance, data = circ_data)

n= 623, number of events= 8

            coef exp(coef) se(coef)      z Pr(>|z|)
ctDNA.SurveillancePOSITIVE 1.9570    7.0779  0.7088 2.761  0.00576 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

            exp(coef) exp(-coef) lower .95 upper .95
ctDNA.SurveillancePOSITIVE    7.078     0.1413    1.764    28.39

Concordance= 0.687 (se = 0.094 )
Likelihood ratio test= 6.66 on 1 df,  p=0.01
Wald test            = 7.62 on 1 df,  p=0.006
Score (logrank) test = 10.36 on 1 df,  p=0.001

```

Hide

```

cox_fit_summary <- summary(cox_fit)

# Extract values for HR, 95% CI, and p-value
HR <- cox_fit_summary$coefficients[2]
lower_CI <- cox_fit_summary$conf.int[3]
upper_CI <- cox_fit_summary$conf.int[4]
p_value <- cox_fit_summary$coefficients[5]
label_text <- paste0("HR = ", round(HR, 2), " (", round(lower_CI, 2), "-", round(upper_CI, 2), "); p = ", round(p_value, 3))
print(label_text)

```

```
[1] "HR = 7.08 (1.76-28.39); p = 0.006"
```

#ctDNA sample positive in the Surveillance Window - 6mo intervals

Hide

```

rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA CRC cfDNA data_ctDNA timepoints.csv")
circ_data <- circ_data[circ_data$Surveillance.Window == TRUE,]
circ_data <- circ_data[!is.na(circ_data$cfDNAconc), ]
circ_data$ctDNA_bucket <- cut(circ_data$ctDNA.Surv,
                                breaks = seq(0, max(circ_data$ctDNA.Surv, na.rm = TRUE), by = 6),
                                right = FALSE,
                                labels = paste0(seq(0, max(circ_data$ctDNA.Surv, na.rm = TRUE) - 6, by = 6),
                                               "-",
                                               seq(6, max(circ_data$ctDNA.Surv, na.rm = TRUE), by = 6)))
rate_by_bucket <- circ_data %>%
  group_by(ctDNA_bucket) %>%
  summarise(
    n_total = n(), # Total number of patients in the bucket
    n_positive = sum(biomarker_status == "POSITIVE"), # Number of positive cases
    n_negative = sum(biomarker_status == "NEGATIVE"), # Number of negative cases
    percentage_positive = mean(biomarker_status == "POSITIVE") * 100, # Positivity rate
    percentage_negative = mean(biomarker_status == "NEGATIVE") * 100 # Negativity rate
  )
overall_totals <- circ_data %>%
  summarise(
    ctDNA_bucket = "Total", # Label for the total row
    n_total = n(), # Total number of patients across all buckets
    n_positive = sum(biomarker_status == "POSITIVE"), # Total number of positive cases across all buckets
    percentage_positive = mean(biomarker_status == "POSITIVE") * 100, # Overall positivity rate
    n_negative = sum(biomarker_status == "NEGATIVE"), # Total number of negative cases across all buckets
    percentage_negative = mean(biomarker_status == "NEGATIVE") * 100 # Overall negativity rate
  )
positivity_rate_with_total <- bind_rows(rate_by_bucket, overall_totals)
print(positivity_rate_with_total)

```

ctDNA_bucket	n_total	n_positive	n_negative	percentage_positive	percentage_negative
<chr>	<int>	<int>	<int>	<dbl>	<dbl>
0-6	856	72	784	8.411215	91.58879
6-12	669	43	626	6.427504	93.57250
12-18	516	15	501	2.906977	97.09302

ctDNA_bucket	n_total	n_positive	n_negative	percentage_positive	percentage_negative
<chr>	<int>	<int>	<int>	<dbl>	<dbl>
18-24	362	12	350	3.314917	96.68508
24-30	201	8	193	3.980100	96.01990
30-36	142	5	137	3.521127	96.47887
36-42	73	3	70	4.109589	95.89041
42-48	46	2	44	4.347826	95.65217
48-54	23	0	23	0.000000	100.00000
54-60	12	0	12	0.000000	100.00000

1-10 of 14 rows

Previous 1 2 Next

Hide

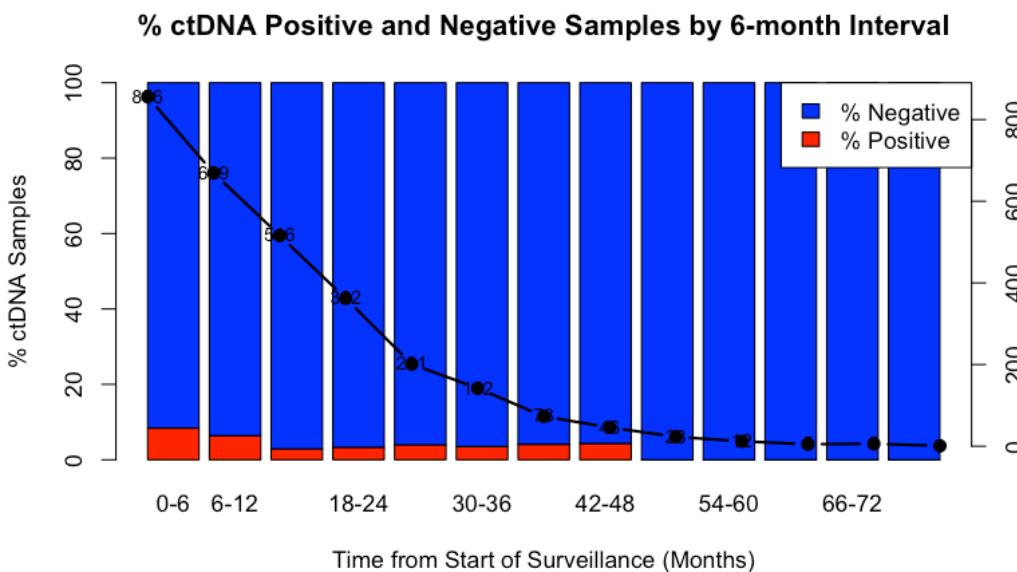
```
bar_midpoints <- barplot(
  t(as.matrix(rate_by_bucket[, c("percentage_positive", "percentage_negative")))), # Transpose to get correct format
  names.arg = rate_by_bucket$ctDNA_bucket,
  col = c("red", "blue"), # Red for positivity and blue for negativity
  main = '% ctDNA Positive and Negative Samples by 6-month Interval',
  xlab = 'Time from Start of Surveillance (Months)',
  ylab = '% ctDNA Samples',
  ylim = c(0, 100),
  legend = c("% Positive", "% Negative"), # Adding a legend for clarification
  args.legend = list(x = "topright")
)
par(new = TRUE)
```

Hide

```
plot(bar_midpoints, rate_by_bucket$n_total, type = "b", col = "black", pch = 19, axes = FALSE, xlab = "", ylab = "", lwd = 2)
axis(side = 4) # Add the secondary y-axis on the right
```

Hide

```
mtext("Total Number of Samples", side = 4, line = 3) # Label for the secondary y-axis
text(bar_midpoints, rate_by_bucket$n_total + 3, labels = rate_by_bucket$n_total, col = "black", cex = 0.8)
```



#Median number of timepoints in the Surveillance Window

Hide

```

rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA CRC cfDNA data_ctDNA timepoints.csv")
circ_data <- circ_data[circ_data$Surveillance.Window == TRUE,]
circ_data <- circ_data[!is.na(circ_data$cfDNAconc),]

# Calculate the median number of ctDNA tests per patient
ctDNA_stats <- circ_data %>%
  group_by(pts_id) %>%
  tally() %>%
  summarise(
    median_tests = median(n),
    min_tests = min(n),
    max_tests = max(n)
  )

# Print the result
print(ctDNA_stats)

```

	median_tests <dbl>	min_tests <int>	max_tests <int>
	4	1	18
1 row			

#Time between imaging at the Surveillance window by ctDNA status

[Hide](#)

```

rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA CRC Imaging Frequency.csv")
tally(~ctDNA_Surveillance, data=circ_data, margins = TRUE)

```

ctDNA_Surveillance	NEGATIVE	POSITIVE	Total
	151	22	173

[Hide](#)

```

median_AverageDateDifference <- aggregate(AverageDateDifference ~ ctDNA_Surveillance, data = circ_data, FUN = median)
print(median_AverageDateDifference)

```

ctDNA_Surveillance <chr>	AverageDateDifference <dbl>
NEGATIVE	369.0000
POSITIVE	157.7411
2 rows	

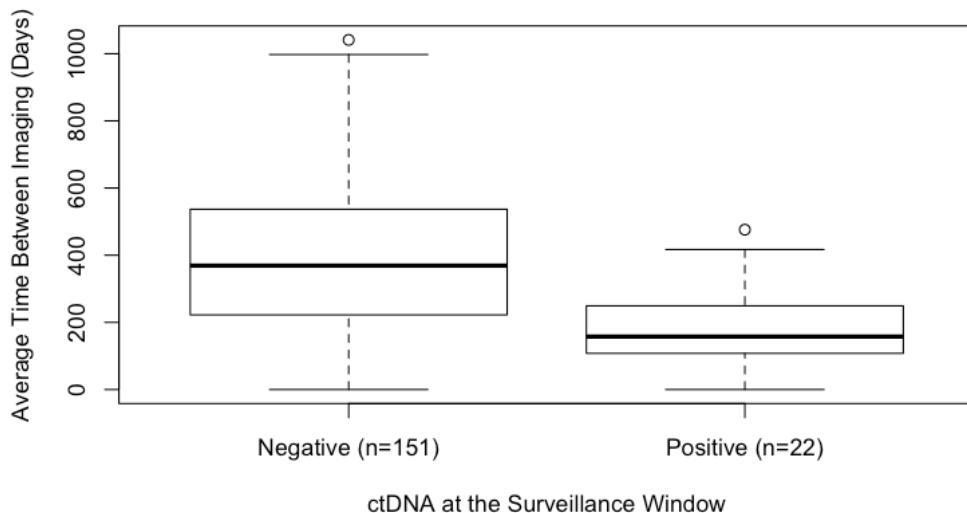
[Hide](#)

```

circ_data$ctDNA_Surveillance <- factor(circ_data$ctDNA_Surveillance, levels = c("NEGATIVE", "POSITIVE"), labels = c("Negative (n=151)", "Positive (n=22)"))
boxplot(AverageDateDifference~ctDNA_Surveillance, data=circ_data, main="Average Time between Imaging at Surveillance Window", xlab="ctDNA at the Surveillance Window", ylab="Average Time Between Imaging (Days)", col="white", border="black")

```

Average Time between Imaging at Surveillance Window



Hide

```
m1<-wilcox.test(AverageDateDifference ~ ctDNA_Surveillance, data=circ_data, na.rm=TRUE, exact=FALSE, conf.int=TRUE)
print(m1)
```

Wilcoxon rank sum test with continuity correction

```
data: AverageDateDifference by ctDNA_Surveillance
W = 2327, p-value = 6.322e-06
alternative hypothesis: true location shift is not equal to 0
95 percent confidence interval:
 106.25 269.00
sample estimates:
difference in location
 181
```

Hide

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA CRC Imaging Frequency.csv")
tally(~ctDNA_Surveillance, data=circ_data, margins = TRUE)
```

ctDNA_Surveillance	NEGATIVE	POSITIVE	Total
	151	22	173

Hide

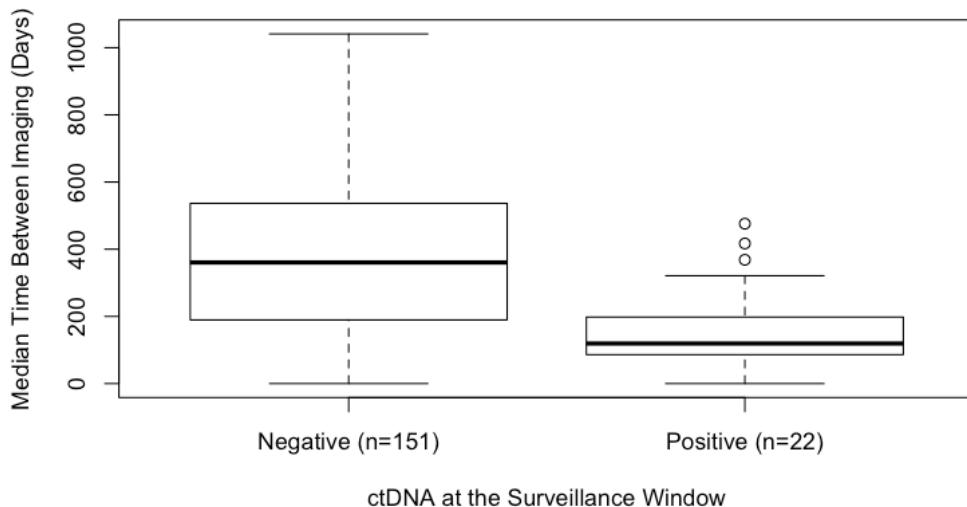
```
median_MedianDateDifference <- aggregate(MedianDateDifference ~ ctDNA_Surveillance, data = circ_data, FUN = median)
print(median_MedianDateDifference)
```

ctDNA_Surveillance	MedianDateDifference
<chr>	<dbl>
NEGATIVE	360.50
POSITIVE	119.25
2 rows	

Hide

```
circ_data$ctDNA_Surveillance <- factor(circ_data$ctDNA_Surveillance, levels = c("NEGATIVE", "POSITIVE"), labels = c("Negative (n=151)", "Positive (n=22)"))
boxplot(MedianDateDifference~ctDNA_Surveillance, data=circ_data, main="Median Time between Imaging at Surveillance Window", xlab="ctDNA at the Surveillance Window", ylab="Median Time Between Imaging (Days)", col="white", border = "black")
```

Median Time between Imaging at Surveillance Window



Hide

```
m2<-wilcox.test(MedianDateDifference ~ ctDNA_Surveillance, data=circ_data, na.rm=TRUE, exact=FALSE, conf.int=TRUE)
print(m2)
```

Wilcoxon rank sum test with continuity correction

```
data: MedianDateDifference by ctDNA_Surveillance
W = 2336.5, p-value = 5.008e-06
alternative hypothesis: true location shift is not equal to 0
95 percent confidence interval:
 90.00009 269.00006
sample estimates:
difference in location
 158.7506
```

#Time-dependent analysis in surveillance window - all stages

Hide

```
rm(list=ls())
setwd("~/Downloads")
dt_final <- read.csv("CLIA CRC Time_Dependent_analysis.csv")
dt_final <- dt_final[dt_final$CRC.Cohort=="TRUE",]
dt_final <- dt_final[dt_final$tstart!="",]

datatable(dt_final, filter = "top")
```

Show 10 entries

Search:

	pts_id	tstart1	tstop1	dfs_time	dfs_event	biomarker_status	ACT	Surveillance	tstart	tstop	CRC.
				AI	All	All		All			All
1	1004874	0	215	459	0	NEGATIVE	true	211	0	4	true
2	1004874	525	572	459	0	POSITIVE	true	211	314	361	true
3	1004874	215	299	459	0	NEGATIVE	true	211	4	88	true
4	1004874	390	456	459	0	POSITIVE	true	211	179	245	true
5	1004874	456	525	459	0	POSITIVE	true	211	245	314	true
6	1004874	299	390	459	0	POSITIVE	true	211	88	179	true
7	1004874	572	670	459	1	POSITIVE	true	211	361	459	true

pts_id	tstart1	tstop1	dfs_time	dfs_event	biomarker_status	ACT	Surveillance	tstart	tstop	CRC.
NA										
9	1007149	504	544	474	1	POSITIVE	false	70	434	474 true
10	1007149	411	504	474	0	POSITIVE	false	70	341	434 true

Showing 1 to 10 of 4,667 entries

Previous 1 2 3 4 5 ... 467 NextHide

```
# Syntax if there is not time-dependent covariate
# fit <- coxph(Surv(dfs_time, dfs_event) ~ biomarker_status,
#                 data = dt_final)
# summary(fit)

fit <- coxph(Surv(tstart, tstop, dfs_event) ~ biomarker_status,
              data = dt_final)
```

Warning: Stop time must be > start time, NA created

Hide

```
summary(fit)
```

```
Call:
coxph(formula = Surv(tstart, tstop, dfs_event) ~ biomarker_status,
       data = dt_final)

n= 3456, number of events= 95
(1211 observations deleted due to missingness)

            coef exp(coef)  se(coef)      z Pr(>|z|)
biomarker_statusPOSITIVE  3.6369   37.9742   0.2235 16.27  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

            exp(coef) exp(-coef) lower .95 upper .95
biomarker_statusPOSITIVE   37.97    0.02633    24.51    58.84

Concordance= 0.816 (se = 0.024 )
Likelihood ratio test= 256.4 on 1 df,  p=<2e-16
Wald test      = 264.9 on 1 df,  p=<2e-16
Score (logrank) test = 679.7 on 1 df,  p=<2e-16
```

#Time-dependent analysis in surveillance window - all stages ACT-treated

Hide

```
rm(list=ls())
setwd("~/Downloads")
dt_final <- read.csv("CLIA CRC Time_Dependent_analysis.csv")
dt_final <- dt_final[dt_final$CRC.Cohort=="TRUE",]
dt_final <- dt_final[dt_final$ACT=="TRUE",]
dt_final <- dt_final[dt_final$tstart!=""]

datatable(dt_final, filter = "top")
```

Show 10 entriesSearch:

pts_id	tstart1	tstop1	dfs_time	dfs_event	biomarker_status	ACT	Surveillance	tstart	tstop	CRC.
			AI	All	All		All			All
1	1004874	0	215	459	0	NEGATIVE	true	211	0	4 true
2	1004874	525	572	459	0	POSITIVE	true	211	314	361 true
3	1004874	215	299	459	0	NEGATIVE	true	211	4	88 true

pts_id	tstart1	tstop1	dfs_time	dfs_event	biomarker_status	ACT	Surveillance	tstart	tstop	CRC.	
4	1004874	390	456	459	0	POSITIVE	true	211	179	245	true
5	1004874	456	525	459	0	POSITIVE	true	211	245	314	true
6	1004874	299	390	459	0	POSITIVE	true	211	88	179	true
7	1004874	572	670	459	1	POSITIVE	true	211	361	459	true
27	1016085	281	305	56	1	POSITIVE	true	249	32	56	true
28	1016085	197	281	56	0	POSITIVE	true	249	0	32	true
NA											

Showing 1 to 10 of 3,192 entries

Previous 1 2 3 4 5 ... 320 NextHide

```
# Syntax if there is not time-dependent covariate
# fit <- coxph(Surv(dfs_time, dfs_event) ~ biomarker_status,
#                 data = dt_final)
# summary(fit)

fit <- coxph(Surv(tstart, tstop, dfs_event) ~ biomarker_status,
              data = dt_final)
```

Warning: Stop time must be > start time, NA created

Hide

```
summary(fit)
```

```
Call:
coxph(formula = Surv(tstart, tstop, dfs_event) ~ biomarker_status,
       data = dt_final)

n= 2178, number of events= 69
(1014 observations deleted due to missingness)

            coef exp(coef)  se(coef)      z Pr(>|z|)
biomarker_statusPOSITIVE  3.5174   33.6971   0.2601 13.52  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

            exp(coef) exp(-coef) lower .95 upper .95
biomarker_statusPOSITIVE      33.7     0.02968     20.24     56.11

Concordance= 0.807  (se = 0.029 )
Likelihood ratio test= 175.9  on 1 df,  p=<2e-16
Wald test      = 182.8  on 1 df,  p=<2e-16
Score (logrank) test = 443.9  on 1 df,  p=<2e-16
```

#Time-dependent analysis in surveillance window - all stages Non-ACT-treated

Hide

```
rm(list=ls())
setwd("~/Downloads")
dt_final <- read.csv("CLIA CRC Time_Dependent_analysis.csv")
dt_final <- dt_final[dt_final$CRC.Cohort=="TRUE",]
dt_final <- dt_final[dt_final$ACT=="FALSE",]
dt_final <- dt_final[dt_final$tstart!="",]

datatable(dt_final, filter = "top")
```

Show 10 entriesSearch:

pts_id	tstart1	tstop1	dfs_time	dfs_event	biomarker_status	ACT	Surveillance	tstart	tstop	CR
			All	All	All		All			A
NA										
9	1007149	504	544	474	1	POSITIVE	false	70	434	474
10	1007149	411	504	474	0	POSITIVE	false	70	341	434
11	1007149	332	411	474	0	NEGATIVE	false	70	262	341
NA.1										
13	1007149	231	332	474	0	NEGATIVE	false	70	161	262
14	1007149	154	231	474	0	NEGATIVE	false	70	84	161
15	1007149	98	154	474	0	NEGATIVE	false	70	28	84
16	1007149	54	98	474	0	NEGATIVE	false	70	0	28
22	1016066	302	440	542	0	NEGATIVE	false	70	232	370

Showing 1 to 10 of 1,475 entries

Previous 1 2 3 4 5 ... 148 NextHide

```
# Syntax if there is not time-dependent covariate
# fit <- coxph(Surv(dfs_time, dfs_event) ~ biomarker_status,
#                 data = dt_final)
# summary(fit)

fit <- coxph(Surv(tstart, tstop, dfs_event) ~ biomarker_status,
              data = dt_final)
```

Warning: Stop time must be > start time, NA created

Hide

```
summary(fit)
```

```
Call:
coxph(formula = Surv(tstart, tstop, dfs_event) ~ biomarker_status,
       data = dt_final)

n= 1278, number of events= 26
(197 observations deleted due to missingness)

            coef exp(coef)  se(coef)    z Pr(>|z|)
biomarker_statusPOSITIVE 3.8078   45.0513   0.4362 8.73  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

            exp(coef) exp(-coef) lower .95 upper .95
biomarker_statusPOSITIVE   45.05      0.0222    19.16   105.9

Concordance= 0.835  (se = 0.043 )
Likelihood ratio test= 76.31  on 1 df,  p=<2e-16
Wald test      = 76.21  on 1 df,  p=<2e-16
Score (logrank) test = 214.3 on 1 df,  p=<2e-16
```

#DFS by ctDNA subgroups at the Surveillance window

Hide

```

rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA_CRC_Clinical Data 122023.csv")
circ_data <- circ_data[circ_data$CLIA.CRC==TRUE,]
circ_data <- circ_data[circ_data$Surveillance.Groups!="",]
circ_data$DFS.months=circ_data$DFS.months-2.5
circ_data <- circ_data[circ_data$DFS.months>=0,]

circ_data$DFS.Event <- as.logical(circ_data$DFS.Event)
circ_data$DFS.months <- as.numeric(circ_data$DFS.months)
survfit(Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)~Surveillance.Groups, data = circ_data)

```

Call: survfit(formula = Surv(time = circ_data\$DFS.months, event = circ_data\$DFS.Event) ~ Surveillance.Groups, data = circ_data)

	n	events	median	0.95LCL	0.95UCL
Surveillance.Groups=All Negative	540	20	NA	NA	NA
Surveillance.Groups=All Positive	35	30	8.05	6.9	13.6
Surveillance.Groups=Negative-Positive	48	24	31.04	21.3	NA

[Hide](#)

```

event_summary <- circ_data %>%
  group_by(Surveillance.Groups) %>%
  summarise(
    Total = n(),
    Events = sum(DFS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)

```

Surveillance.Groups	Total	Events	Fraction	Percentage
<chr>	<int>	<int>	<dbl>	<dbl>
All Negative	540	20	0.03703704	3.703704
All Positive	35	30	0.85714286	85.714286
Negative-Positive	48	24	0.50000000	50.000000

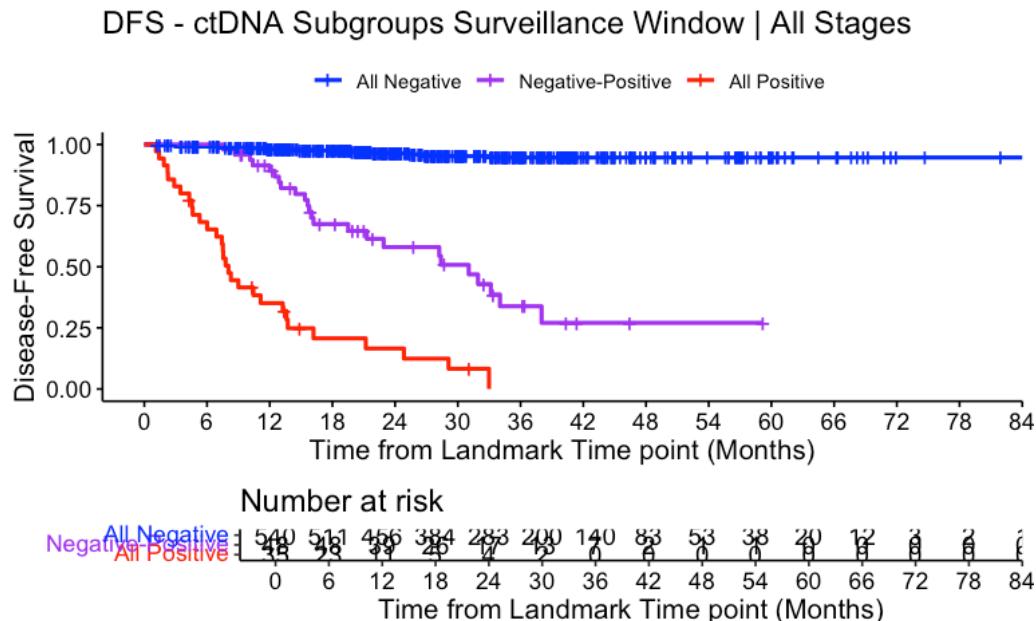
3 rows

[Hide](#)

```

surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
circ_data$Surveillance.Groups <- factor(circ_data$Surveillance.Groups, levels=c("All Negative","Negative-Positive","All Positive"))
KM_curve <- survfit(surv_object ~ Surveillance.Groups, data = circ_data, conf.int=0.95, conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette=c("blue","purple", "red"), title="DFS - ctDNA Subgroups Surveillance Window | All Stages", ylab= "Disease-Free Survival", xlab="Time from Landmark Time point (Months)", legend.labs=c("All Negative","Negative-Positive","All Positive"), legend.title="")

```



Hide

```
summary(KM_curve, times= c(24))
```

```
Call: survfit(formula = surv_object ~ Surveillance.Groups, data = circ_data,
  conf.int = 0.95, conf.type = "log-log")

  Surveillance.Groups=All Negative
  time      n.risk      n.event      survival      std.err lower 95% CI upper 95% CI
  2.40e+01      2.83e+02      1.60e+01      9.64e-01      9.06e-03      9.41e-01      9.78e-01

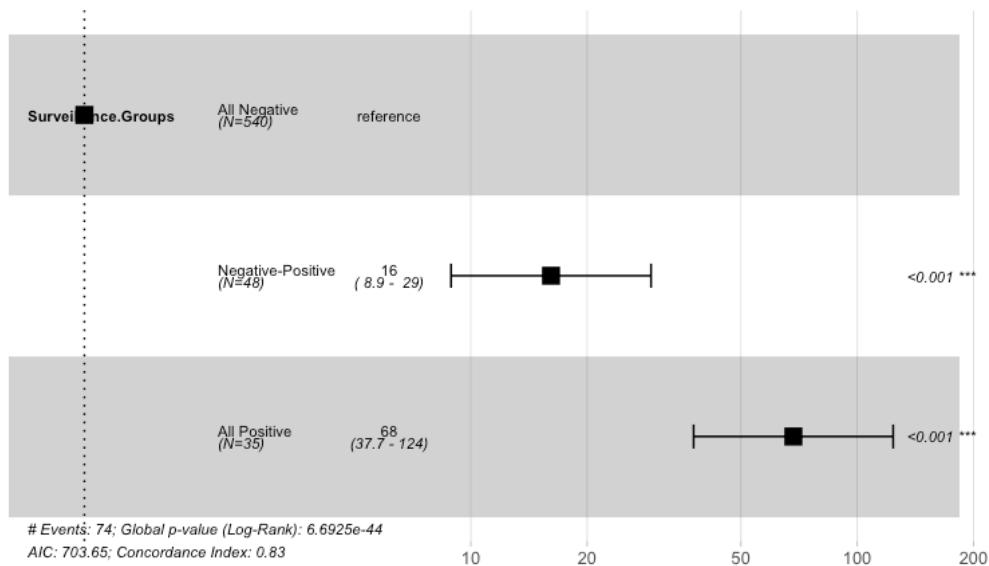
  Surveillance.Groups=Negative-Positive
  time      n.risk      n.event      survival      std.err lower 95% CI upper 95% CI
  24.0000      17.0000      17.0000      0.5803      0.0801      0.4086      0.7182

  Surveillance.Groups=All Positive
  time      n.risk      n.event      survival      std.err lower 95% CI upper 95% CI
  24.0000      4.0000      27.0000      0.1657      0.0702      0.0576      0.3222
```

Hide

```
cox_fit <- coxph(surv_object ~ Surveillance.Groups, data=circ_data)
ggforest(cox_fit,data = circ_data)
```

Hazard ratio



Hide

```
summary(cox_fit)
```

```
Call:
coxph(formula = surv_object ~ Surveillance.Groups, data = circ_data)

n= 623, number of events= 74

            coef exp(coef)  se(coef)      z Pr(>|z|)
Surveillance.GroupsNegative-Positive  2.7806  16.1288  0.3041  9.144 <2e-16 ***
Surveillance.GroupsAll Positive       4.2247  68.3543  0.3033 13.929 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

            exp(coef) exp(-coef) lower .95 upper .95
Surveillance.GroupsNegative-Positive   16.13    0.06200    8.887    29.27
Surveillance.GroupsAll Positive        68.35    0.01463   37.721   123.86

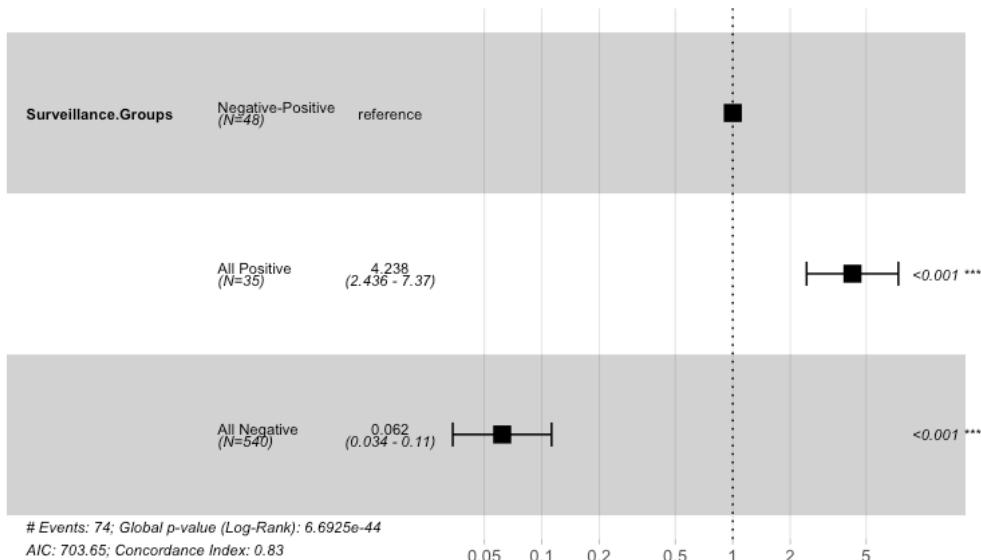
Concordance= 0.833 (se = 0.027 )
Likelihood ratio test= 198.8 on 2 df,  p=<2e-16
Wald test            = 196.3 on 2 df,  p=<2e-16
Score (logrank) test = 536.6 on 2 df,  p=<2e-16
```

Hide

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA_CRC_Clinical Data 122023.csv")
circ_data <- circ_data[circ_data$CLIA.CRC==TRUE,]
circ_data <- circ_data[circ_data$Surveillance.Groups!="",]
circ_data$DFS.months=circ_data$DFS.months-2.5
circ_data <- circ_data[circ_data$DFS.months>=0,]

surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
circ_data$Surveillance.Groups <- factor(circ_data$Surveillance.Groups, levels=c("Negative-Positive","All Positive", "All Negative"))
cox_fit <- coxph(surv_object ~ Surveillance.Groups, data=circ_data)
ggforest(cox_fit,data = circ_data)
```

Hazard ratio



Hide

```
summary(cox_fit)
```

Call:
`coxph(formula = surv_object ~ Surveillance.Groups, data = circ_data)`

`n= 623, number of events= 74`

	coef	exp(coef)	se(coef)	z	Pr(> z)
Surveillance.GroupsAll Positive	1.4441	4.2380	0.2826	5.111	3.2e-07 ***
Surveillance.GroupsAll Negative	-2.7806	0.0620	0.3041	-9.144	< 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95
Surveillance.GroupsAll Positive	4.238	0.236	2.43591	7.3734
Surveillance.GroupsAll Negative	0.062	16.129	0.03416	0.1125

Concordance= 0.833 (se = 0.027)

Likelihood ratio test= 198.8 on 2 df, p=<2e-16

Wald test = 196.3 on 2 df, p=<2e-16

Score (logrank) test = 536.6 on 2 df, p=<2e-16

#Demographics table for ctDNA positive at the Surveillance window with no radiological recurrence

Hide

```

rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA CRC_Clinical Data 122023.csv")
circ_data <- circ_data[circ_data$CLIA.CRC==TRUE,]
circ_data <- circ_data[circ_data$ctDNA.Surveillance=="POSITIVE",]
circ_data <- circ_data[circ_data$DFS.Event==FALSE,]

circ_data_subset <- circ_data %>%
  select(
    Age,
    Gender,
    PrimSite,
    pT,
    pN,
    Stage,
    Grade,
    NAC,
    ACT,
    MSI,
    BRAF.V600E,
    RAS,
    OS.months) %>%
  mutate(
    Age = as.numeric(Age),
    Gender = factor(Gender, levels = c("Male", "Female")),
    PrimSite = factor(PrimSite, levels = c("Right-sided colon", "Left-sided colon")),
    pT = factor(pT, levels = c("T0", "T1-T2", "T3-T4")),
    pN = factor(pN, levels = c("N0", "N1-N2")),
    Stage = factor(Stage, levels = c("I", "II", "III")),
    Grade = factor(Grade, levels = c("G1", "G2", "G3", "GX")),
    NAC = factor(NAC, levels = c("TRUE", "FALSE"), labels = c("Neoadjuvant Chemotherapy", "Upfront Surgery")),
    ACT = factor(ACT, levels = c("TRUE", "FALSE"), labels = c("Adjuvant Chemotherapy", "Observation")),
    MSI = factor(MSI, levels = c("MSS", "MSI-High")),
    BRAF.V600E = factor(BRAF.V600E, levels = c("WT", "MUT"), labels = c("BRAF WT", "BRAF V600E")),
    RAS = factor(RAS, levels = c("WT", "MUT"), labels = c("RAS WT", "RAS Mut")),
    OS.months = as.numeric(OS.months))
table1 <- circ_data_subset %>%
 tbl_summary(
  statistic = list(
    all_continuous() ~ "{median} ({min} - {max})",
    all_categorical() ~ "{n} ({p}%)") %>%
  bold_labels()
table1

```

Characteristic	N = 29 ¹
Age	57 (28 - 86)
Gender	
Male	12 (41%)
Female	17 (59%)
PrimSite	
Right-sided colon	12 (41%)
Left-sided colon	17 (59%)
pT	
T0	0 (0%)
T1-T2	4 (14%)
T3-T4	25 (86%)
pN	
N0	9 (31%)
N1-N2	20 (69%)
Stage	
I	3 (10%)

¹ Median (Range); n (%)

Characteristic	N = 29 [†]
II	7 (24%)
III	19 (66%)
Grade	
G1	4 (14%)
G2	19 (66%)
G3	6 (21%)
GX	0 (0%)
NAC	
Neoadjuvant Chemotherapy	0 (0%)
Upfront Surgery	29 (100%)
ACT	
Adjuvant Chemotherapy	19 (66%)
Observation	10 (34%)
MSI	
MSS	26 (90%)
MSI-High	3 (10%)
BRAF.V600E	
BRAF WT	26 (90%)
BRAF V600E	3 (10%)
RAS	
RAS WT	17 (59%)
RAS Mut	12 (41%)
OS.months	22 (7 - 62)

[†] Median (Range); n (%)

Hide

```
fit1 <- as_flex_table(
  table1,
  include = everything(),
  return_calls = FALSE,
  strip_md_bold = TRUE)
fit1
```

Characteristic	N = 29 [†]
Age	57 (28 - 86)
Gender	
Male	12 (41%)
Female	17 (59%)
PrimSite	
Right-sided colon	12 (41%)
Left-sided colon	17 (59%)
pT	
T0	0 (0%)
T1-T2	4 (14%)
T3-T4	25 (86%)
pN	

[†]Median (Range); n (%)

Characteristic	N = 29 ¹
N0	9 (31%)
N1-N2	20 (69%)
Stage	
I	3 (10%)
II	7 (24%)
III	19 (66%)
Grade	
G1	4 (14%)
G2	19 (66%)
G3	6 (21%)
GX	0 (0%)
NAC	
Neoadjuvant Chemotherapy	0 (0%)
Upfront Surgery	29 (100%)
ACT	
Adjuvant Chemotherapy	19 (66%)
Observation	10 (34%)
MSI	
MSS	26 (90%)
MSI-High	3 (10%)
BRAF.V600E	
BRAF WT	26 (90%)
BRAF V600E	3 (10%)
RAS	
RAS WT	17 (59%)
RAS Mut	12 (41%)
OS.months	22 (7 - 62)

¹Median (Range); n (%)

Hide

```
save_as_docx(fit1, path= "~/Downloads/table1.docx")
```

#DFS by ctDNA at 12 months post-surgery

Hide

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA CRC_Clinical Data 122023.csv")
circ_data <- circ_data[circ_data$CLIA.CRC==TRUE,]
circ_data <- circ_data[circ_data$ctDNA.12months!=""]
circ_data$DFS.months=circ_data$DFS.months-12
circ_data <- circ_data[circ_data$DFS.months>=0,]
circ_dataadf <- as.data.frame(circ_data)

circ_data$DFS.Event <- as.logical(circ_data$DFS.Event)
circ_data$DFS.months <- as.numeric(circ_data$DFS.months)
survfit(Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)~ctDNA.12months, data = circ_data)
```

```
Call: survfit(formula = Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event) ~
ctDNA.12months, data = circ_data)
```

	n	events	median	0.95LCL	0.95UCL
ctDNA.12months=NEGATIVE	136	6	NA	NA	NA
ctDNA.12months=POSITIVE	14	7	6.69	3.74	NA

Hide

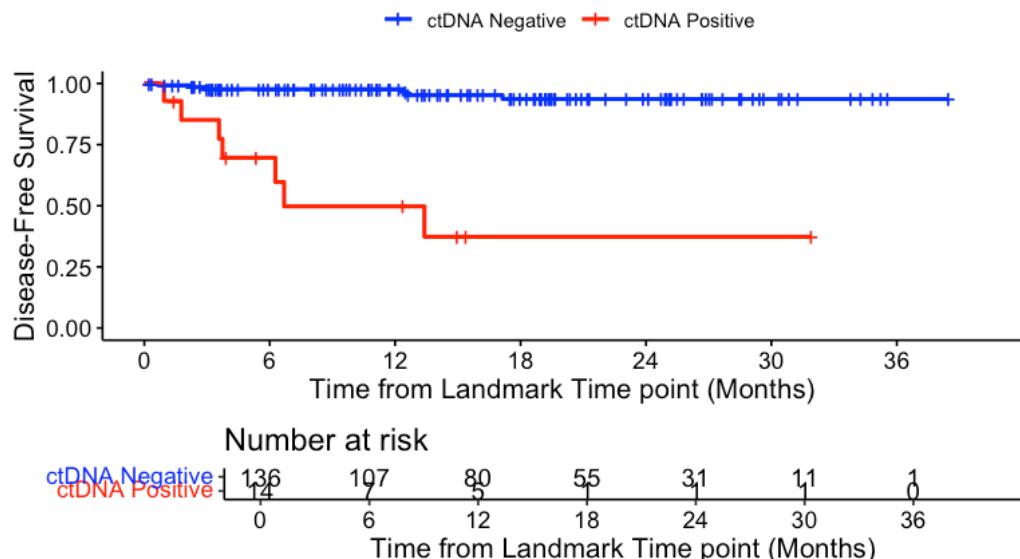
```
event_summary <- circ_data %>%
  group_by(ctDNA.12months) %>%
  summarise(
    Total = n(),
    Events = sum(DFS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)
```

ctDNA.12months	Total	Events	Fraction	Percentage
	<int>	<int>	<dbl>	<dbl>
NEGATIVE	136	6	0.04411765	4.411765
POSITIVE	14	7	0.50000000	50.000000
2 rows				

Hide

```
surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
KM_curve <- survfit(surv_object ~ ctDNA.12months, data = circ_data, conf.int=0.95, conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette=c("blue","red"), title="DFS - ctDNA at 12 months | All pts", ylab= "Disease-Free Survival", xlab="Time from Landmark Time point (Months)", legend.labs=c("ctDNA Negative", "ctDNA Positive"), legend.title="")
```

DFS - ctDNA at 12 months | All pts



Hide

```
summary(KM_curve, times= c(24))
```

```
Call: survfit(formula = surv_object ~ ctDNA.12months, data = circ_data,
  conf.int = 0.95, conf.type = "log-log")
```

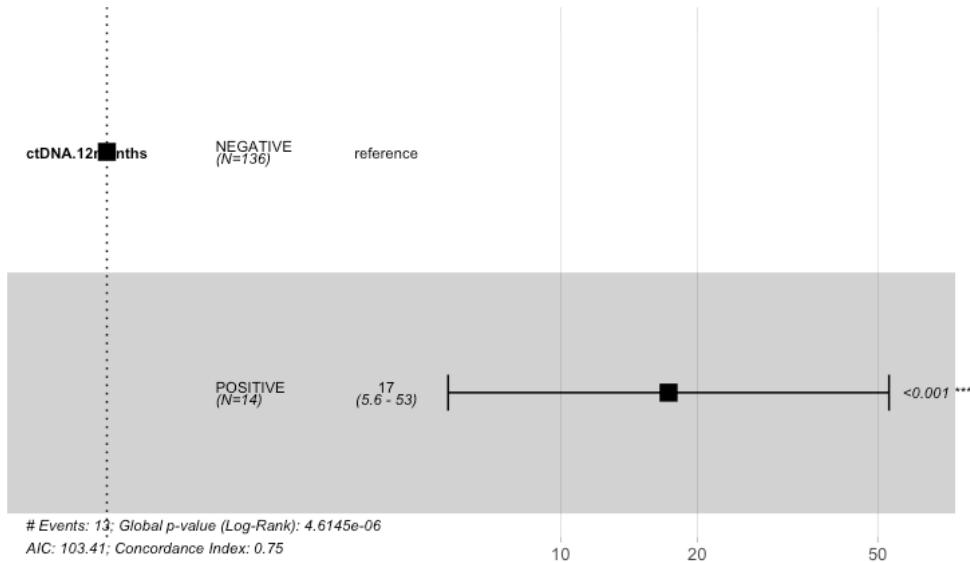
ctDNA.12months=NEGATIVE						
time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
24.000	31.000	6.0000	0.9365	0.0263	0.8593	0.9720

ctDNA.12months=POSITIVE						
time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
24.00	1.000	7.000	0.373	0.156	0.105	0.650

Hide

```
circ_data$ctDNA.12months <- factor(circ_data$ctDNA.12months, levels=c("NEGATIVE","POSITIVE"))
cox_fit <- coxph(surv_object ~ ctDNA.12months, data=circ_data)
ggforest(cox_fit, data = circ_data)
```

Hazard ratio



Hide

```
summary(cox_fit)
```

```
Call:
coxph(formula = surv_object ~ ctDNA.12months, data = circ_data)
```

```
n= 150, number of events= 13
```

	coef	exp(coef)	se(coef)	z	Pr(> z)
ctDNA.12monthsPOSITIVE	2.8501	17.2889	0.5708	4.993	5.94e-07 ***

```
---
```

```
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

	exp(coef)	exp(-coef)	lower .95	upper .95
ctDNA.12monthsPOSITIVE	17.29	0.05784	5.648	52.92

```
Concordance= 0.748 (se = 0.068 )
```

```
Likelihood ratio test= 20.99 on 1 df, p=5e-06
```

```
Wald test = 24.93 on 1 df, p=6e-07
```

```
Score (logrank) test = 45.91 on 1 df, p=1e-11
```

Hide

```
cox_fit_summary <- summary(cox_fit)

# Extract values for HR, 95% CI, and p-value
HR <- cox_fit_summary$coefficients[2]
lower_CI <- cox_fit_summary$conf.int[3]
upper_CI <- cox_fit_summary$conf.int[4]
p_value <- cox_fit_summary$coefficients[5]
label_text <- paste0("HR = ", round(HR, 2), " (", round(lower_CI, 2), "-", round(upper_CI, 2), "); p = ", round(p_value, 3))
print(label_text)
```

```
[1] "HR = 17.29 (5.65-52.92); p = 0"
```

#DFS by ctDNA 4-weeks post-adjuvant chemotherapy

[Hide](#)

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA_CRC_Clinical Data 122023.csv")
circ_data <- circ_data[circ_data$CLIA.CRC==TRUE,]
circ_data <- circ_data[circ_data$ctDNA.postACT!=""]
circ_data$DFS.months=circ_data$DFS.months-2.5
circ_data <- circ_data[circ_data$DFS.months>=0,]
circ_datadf <- as.data.frame(circ_data)

circ_data$DFS.Event <- as.logical(circ_data$DFS.Event)
circ_data$DFS.months <- as.numeric(circ_data$DFS.months)
survfit(Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event) ~
  ctDNA.postACT, data = circ_data)
```

Call: survfit(formula = Surv(time = circ_data\$DFS.months, event = circ_data\$DFS.Event) ~
 ctDNA.postACT, data = circ_data)

	n	events	median	0.95LCL	0.95UCL
ctDNA.postACT=NEGATIVE	112	17	NA	NA	NA
ctDNA.postACT=POSITIVE	6	4	7.54	3.45	NA

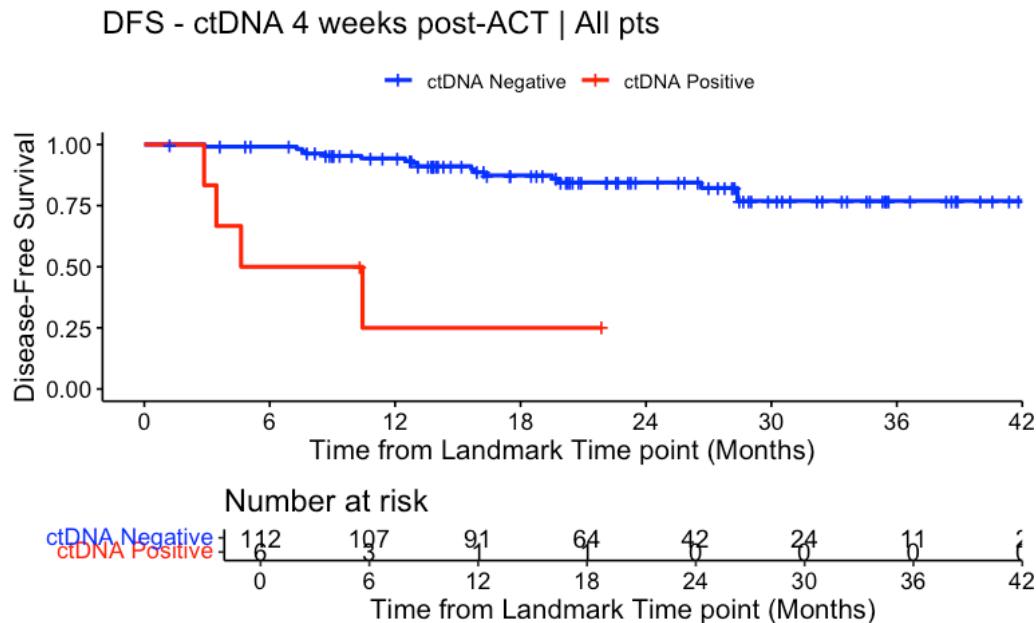
[Hide](#)

```
event_summary <- circ_data %>%
  group_by(ctDNA.postACT) %>%
  summarise(
    Total = n(),
    Events = sum(DFS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)
```

ctDNA.postACT	Total	Events	Fraction	Percentage
<chr>	<int>	<int>	<dbl>	<dbl>
NEGATIVE	112	17	0.1517857	15.17857
POSITIVE	6	4	0.6666667	66.66667
2 rows				

[Hide](#)

```
surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
KM_curve <- survfit(surv_object ~ ctDNA.postACT, data = circ_data, conf.int=0.95, conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette=c("blue","red"), title="DFS - ctDNA 4 weeks post-ACT | All pts", ylab= "Disease-Free Survival", xlab="Time from Landmark Time point (Months)", legend.labs=c("ctDNA Negative", "ctDNA Positive"), legend.title="")
```



Hide

```
summary(KM_curve, times= c(24))
```

```
Call: survfit(formula = surv_object ~ ctDNA.postACT, data = circ_data,
  conf.int = 0.95, conf.type = "log-log")

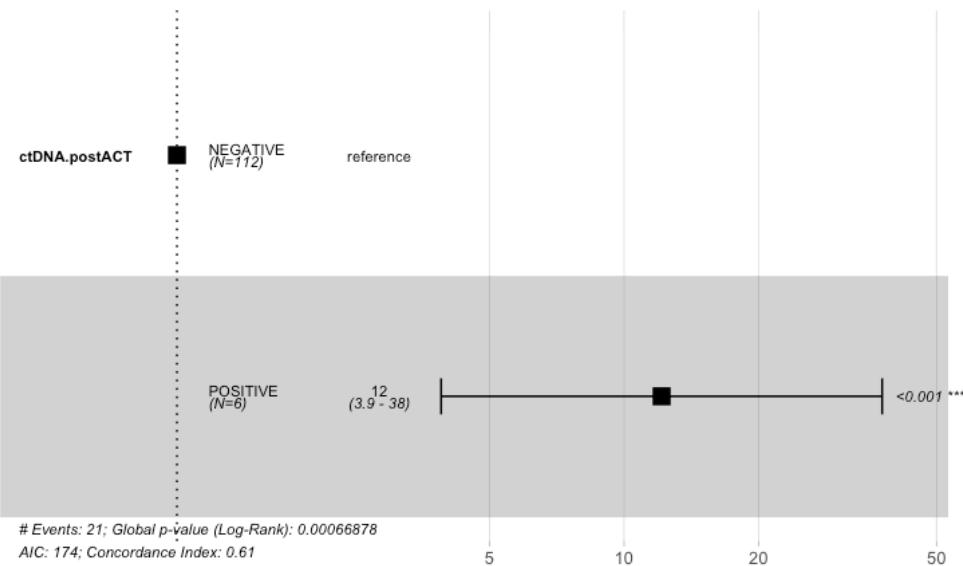
  ctDNA.postACT=NEGATIVE
  time    n.risk    n.event    survival    std.err lower 95% CI upper 95% CI
 24.000    42.000    14.000      0.845      0.039    0.749      0.906

  ctDNA.postACT=POSITIVE
  time n.risk n.event survival std.err lower 95% CI upper 95% CI
```

Hide

```
circ_data$ctDNA.postACT <- factor(circ_data$ctDNA.postACT, levels=c("NEGATIVE","POSITIVE"))
cox_fit <- coxph(surv_object ~ ctDNA.postACT, data=circ_data)
ggforest(cox_fit, data = circ_data)
```

Hazard ratio



Hide

```
summary(cox_fit)
```

Call:
`coxph(formula = surv_object ~ ctDNA.postACT, data = circ_data)`

n= 118, number of events= 21

	coef	exp(coef)	se(coef)	z	Pr(> z)
ctDNA.postACTPOSITIVE	2.4960	12.1334	0.5794	4.308	1.65e-05 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95
ctDNA.postACTPOSITIVE	12.13	0.08242	3.897	37.77

Concordance= 0.612 (se = 0.051)
Likelihood ratio test= 11.57 on 1 df, p=7e-04
Wald test = 18.56 on 1 df, p=2e-05
Score (logrank) test = 30.09 on 1 df, p=4e-08

Hide

```
cox_fit_summary <- summary(cox_fit)
```

```
# Extract values for HR, 95% CI, and p-value
HR <- cox_fit_summary$coefficients[2]
lower_CI <- cox_fit_summary$conf.int[3]
upper_CI <- cox_fit_summary$conf.int[4]
p_value <- cox_fit_summary$coefficients[5]
label_text <- paste0("HR = ", round(HR, 2), " (", round(lower_CI, 2), "-", round(upper_CI, 2), "); p = ", round(p_value, 3))
print(label_text)
```

```
[1] "HR = 12.13 (3.9-37.77); p = 0"
```

#DFS by ACT treatment in MRD negative - High Risk Stage II or Stage III

Hide

```

rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA CRC_Clinical Data 122023.csv")
circ_data <- circ_data[circ_data$CLIA.CRC==TRUE,]
circ_data <- circ_data[circ_data$Risk.Stage==TRUE,]
circ_data <- circ_data[circ_data$ctDNA.MRD!=""]
circ_data <- circ_data[circ_data$ctDNA.MRD=="NEGATIVE",]
circ_data$DFS.months=circ_data$DFS.months-
circ_data <- circ_data[circ_data$DFS.months>=0,]
circ_datadf <- as.data.frame(circ_data)

circ_data$DFS.Event <- as.logical(circ_data$DFS.Event)
circ_data$DFS.months <- as.numeric(circ_data$DFS.months)
survfit(Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)~ACT, data = circ_data)

```

```

Call: survfit(formula = Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event) ~
  ACT, data = circ_data)

  n events median 0.95LCL 0.95UCL
ACT=FALSE  93      5     NA     NA     NA
ACT=TRUE   187     23     NA     NA     NA

```

Hide

```

event_summary <- circ_data %>%
  group_by(ACT) %>%
  summarise(
    Total = n(),
    Events = sum(DFS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)

```

ACT	Total	Events	Fraction	Percentage
FALSE	93	5	0.05376344	5.376344
TRUE	187	23	0.12299465	12.299465

2 rows

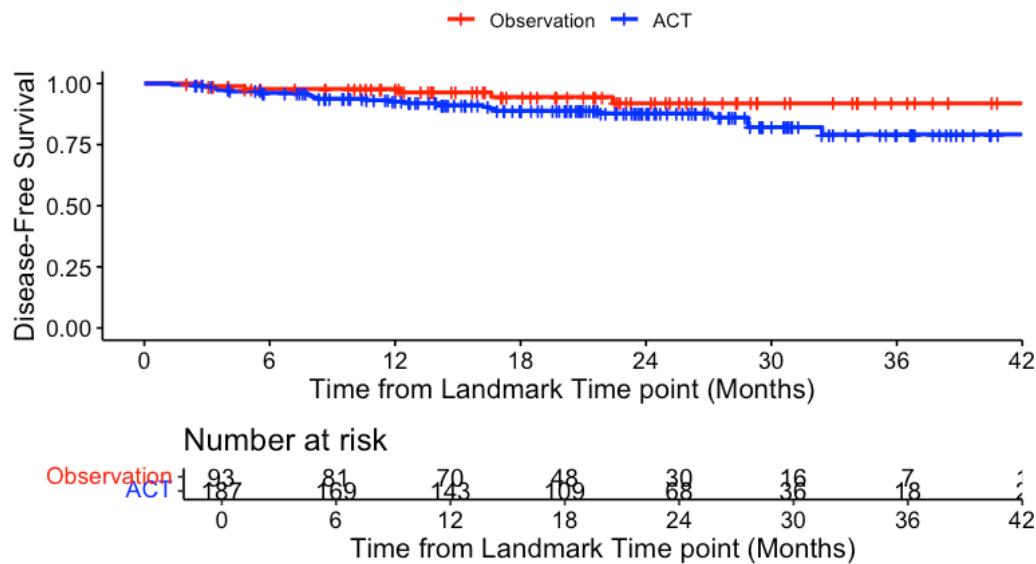
Hide

```

surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
KM_curve <- survfit(surv_object ~ ACT, data = circ_data, conf.int=0.95, conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette=c("red","blue"), title="DFS - ctDNA MRD Negative ACT vs Observation | High Risk Stage II or Stage III", ylab="Disease-Free Survival", xlab="Time from Landmark Time point (Months)", legend.labs=c("Observation", "ACT"), legend.title="")

```

DFS - ctDNA MRD Negative ACT vs Observation | High Risk Stage

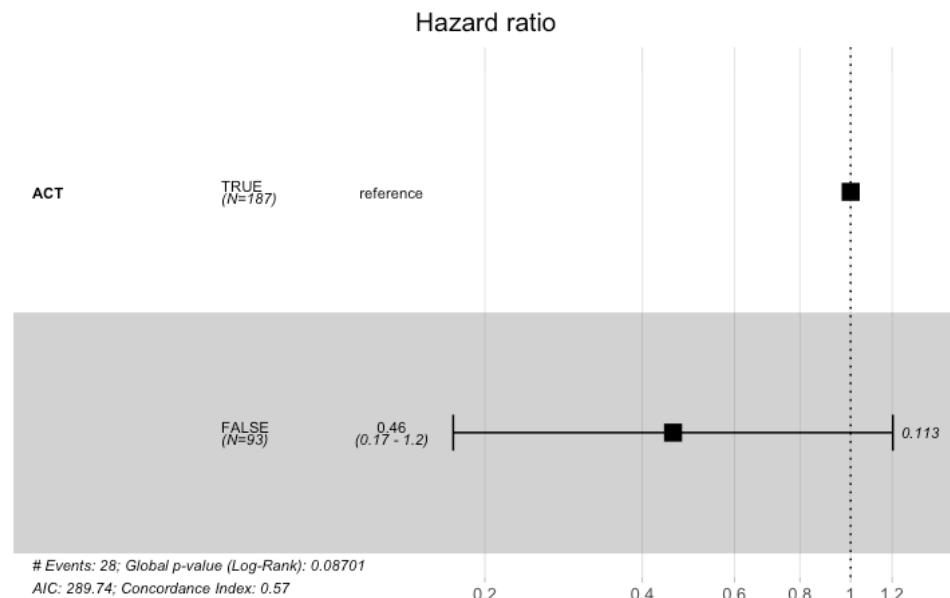


```
summary(KM_curve, times= c(24))
```

```
Call: survfit(formula = surv_object ~ ACT, data = circ_data, conf.int = 0.95,
  conf.type = "log-log")
```

time	ACT=FALSE		ACT=TRUE	
	n.risk	n.event	survival	std.err
24.0000	30.0000	5.0000	0.9186	0.0372
				0.8054 0.9673
24.00	68.00	19.000	0.878	0.027
				0.813 0.921

```
circ_data$ACT <- factor(circ_data$ACT, levels=c("TRUE","FALSE"))
cox_fit <- coxph(surv_object ~ ACT, data=circ_data)
ggforest(cox_fit,data = circ_data)
```



```
summary(cox_fit)
```

```

Call:
coxph(formula = surv_object ~ ACT, data = circ_data)

n= 280, number of events= 28

      coef exp(coef) se(coef)      z Pr(>|z|)    
ACTFALSE -0.7813    0.4578   0.4935 -1.583    0.113  
                    exp(coef) exp(-coef) lower .95 upper .95  
ACTFALSE    0.4578      2.184    0.174    1.204  

Concordance= 0.571  (se = 0.04 )
Likelihood ratio test= 2.93  on 1 df,  p=0.09
Wald test      = 2.51  on 1 df,  p=0.1
Score (logrank) test = 2.64  on 1 df,  p=0.1

```

Hide

```

cox_fit_summary <- summary(cox_fit)

# Extract values for HR, 95% CI, and p-value
HR <- cox_fit_summary$coefficients[2]
lower_CI <- cox_fit_summary$conf.int[3]
upper_CI <- cox_fit_summary$conf.int[4]
p_value <- cox_fit_summary$coefficients[5]
label_text <- paste0("HR = ", round(HR, 2), " (", round(lower_CI, 2), "-", round(upper_CI, 2), "); p = ", round(p_value, 3))
print(label_text)

```

```
[1] "HR = 0.46 (0.17-1.2); p = 0.113"
```

Hide

```

#Adjusted HR "ACT vs no ACT" - age, gender, MSI and pathological stage
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA CRC_Clinical Data 122023.csv")
circ_data <- circ_data[circ_data$CLIA.CRC==TRUE,]
circ_data <- circ_data[circ_data$Risk.Stage==TRUE,]
circ_data <- circ_data[circ_data$ctDNA.MRD!=""]
circ_data <- circ_data[circ_data$ctDNA.MRD=="NEGATIVE",]
circ_data$DFS.months=circ_data$DFS.months-2
circ_data <- circ_data[circ_data$DFS.months>=0,]
circ_dataadf <- as.data.frame(circ_data)

circ_data$DFS.Event <- as.logical(circ_data$DFS.Event)
circ_data$DFS.months <- as.numeric(circ_data$DFS.months)
circ_data$ACT <- factor(circ_data$ACT, levels=c("TRUE","FALSE"))
circ_data$Age.Group <- factor(circ_data$Age.Group, levels = c("1", "2"), labels = c("<70", "≥70"))
circ_data$Gender <- factor(circ_data$Gender, levels = c("Female", "Male"))
circ_data$MSI <- factor(circ_data$MSI, levels = c("MSS", "MSI-High"))
circ_data$Stage <- factor(circ_data$Stage, levels = c("II", "III"))
surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
cox_fit <- coxph(surv_object ~ ACT + Age.Group + Gender + MSI + Stage, data=circ_data)
summary(cox_fit)

```

```

Call:
coxph(formula = surv_object ~ ACT + Age.Group + Gender + MSI +
  Stage, data = circ_data)

n= 280, number of events= 28

      coef exp(coef) se(coef)      z Pr(>|z|)
ACTFALSE    -0.1808    0.8346   0.6123 -0.295   0.7678
Age.Group≥70  0.3185    1.3750   0.4494  0.709   0.4785
GenderMale    0.5723    1.7724   0.4009  1.428   0.1534
MSIMSI-High   0.2271    1.2550   0.5780  0.393   0.6944
StageIII     1.3172    3.7330   0.6395  2.060   0.0394 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef) exp(-coef) lower .95 upper .95
ACTFALSE      0.8346    1.1982   0.2513    2.772
Age.Group≥70   1.3750    0.7273   0.5699    3.318
GenderMale     1.7724    0.5642   0.8078    3.889
MSIMSI-High   1.2550    0.7968   0.4042    3.896
StageIII      3.7330    0.2679   1.0658   13.075

Concordance= 0.668 (se = 0.055 )
Likelihood ratio test= 10.58 on 5 df,  p=0.06
Wald test        = 8.62 on 5 df,  p=0.1
Score (logrank) test = 9.59 on 5 df,  p=0.09

```

Hide

```

# Adjust p-values using False Discovery Rate (FDR) adjustment (Benjamini-Hochberg method)
p_values <- summary(cox_fit)$coefficients[, 5]
adjusted_p_values <- p.adjust(p_values, method = "fdr")
results <- data.frame(
  Variable = rownames(summary(cox_fit)$coefficients),
  Original_P_Value = p_values,
  FDR_Adjusted_P_Value = adjusted_p_values
)
print(results)

```

	Variable	Original_P_Value	FDR_Adjusted_P_Value
	<chr>	<dbl>	<dbl>
ACTFALSE	ACTFALSE	0.76781459	0.7678146
Age.Group≥70	Age.Group≥70	0.47850003	0.7678146
GenderMale	GenderMale	0.15342732	0.3835683
MSIMSI-High	MSIMSI-High	0.69437583	0.7678146
StageIII	StageIII	0.03943392	0.1971696

5 rows

#DFS by ACT treatment in MRD positive - High Risk Stage II or Stage III

Hide

```

rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA CRC_Clinical Data 122023.csv")
circ_data <- circ_data[circ_data$CLIA.CRC==TRUE,]
circ_data <- circ_data[circ_data$Risk.Stage==TRUE,]
circ_data <- circ_data[circ_data$ctDNA.MRD!="",]
circ_data <- circ_data[circ_data$ctDNA.MRD=="POSITIVE",]
circ_data$DFS.months=circ_data$DFS.months-
circ_data <- circ_data[circ_data$DFS.months>=0,]
circ_dataadf <- as.data.frame(circ_data)

circ_data$DFS.Event <- as.logical(circ_data$DFS.Event)
circ_data$DFS.months <- as.numeric(circ_data$DFS.months)
survfit(Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)~ACT, data = circ_data)

```

```

Call: survfit(formula = Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event) ~
  ACT, data = circ_data)

```

	n	events	median	0.95LCL	0.95UCL
ACT=FALSE	10	9	2.41	1.22	NA
ACT=TRUE	51	29	13.38	10.19	NA

Hide

```
event_summary <- circ_data %>%
  group_by(ACT) %>%
  summarise(
    Total = n(),
    Events = sum(DFS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)
```

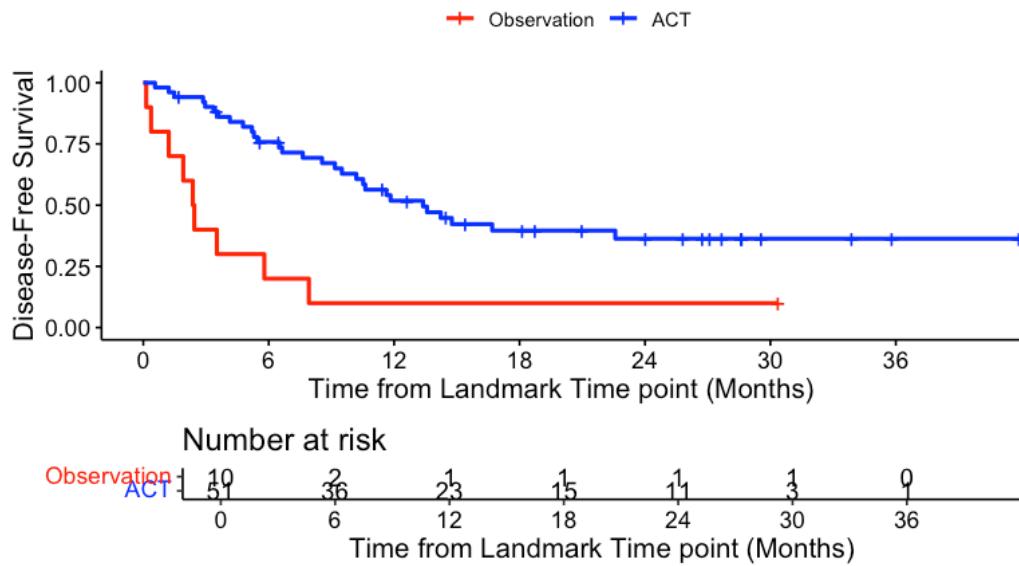
ACT <lgI>	Total <int>	Events <int>	Fraction <dbl>	Percentage <dbl>
FALSE	10	9	0.9000000	90.00000
TRUE	51	29	0.5686275	56.86275

2 rows

Hide

```
surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
KM_curve <- survfit(surv_object ~ ACT, data = circ_data, conf.int=0.95, conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette=c("red","blue"), title="DFS - ctDNA MRD Positive ACT vs Observation | High Risk Stage II or Stage III", ylab="Disease-Free Survival", xlab="Time from Landmark Time point (Months)", legend.labs=c("Observation", "ACT"), legend.title="")
```

DFS - ctDNA MRD Positive ACT vs Observation | High Risk Stage



Hide

```
summary(KM_curve, times= c(24))
```

```
Call: survfit(formula = surv_object ~ ACT, data = circ_data, conf.int = 0.95,
  conf.type = "log-log")
```

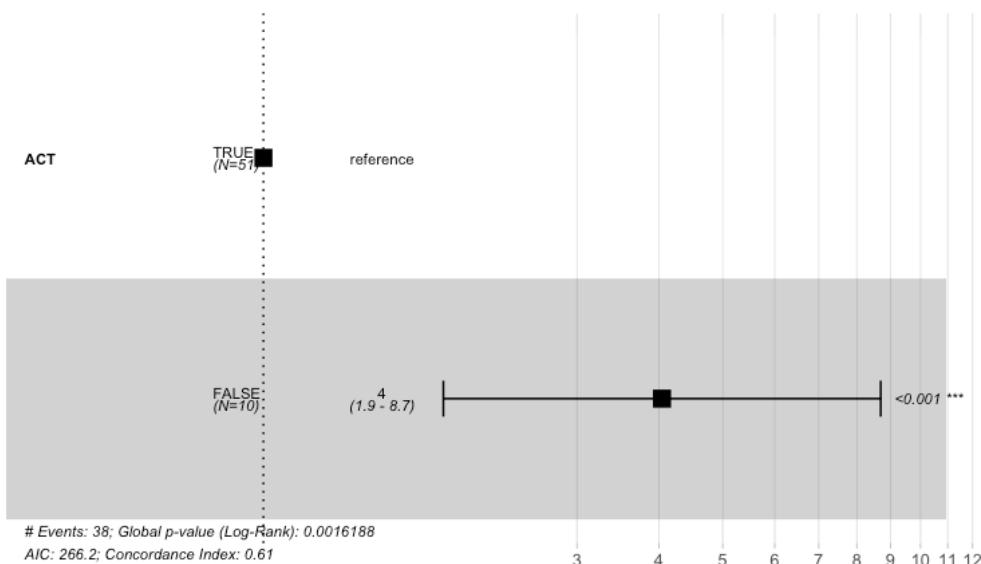
ACT=FALSE						
time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
24.00000	1.00000	9.00000	0.10000	0.09487	0.00572	0.35813

ACT=TRUE						
time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
24.000	11.0000	29.0000	0.3632	0.0744	0.2218	0.5060

Hide

```
circ_data$ACT <- factor(circ_data$ACT, levels=c("TRUE","FALSE"))
cox_fit <- coxph(surv_object ~ ACT, data=circ_data)
ggforest(cox_fit,data = circ_data)
```

Hazard ratio



Hide

```
summary(cox_fit)
```

```

Call:
coxph(formula = surv_object ~ ACT, data = circ_data)

n= 61, number of events= 38

      coef exp(coef)  se(coef)      z Pr(>|z|)
ACTFALSE 1.3968     4.0422   0.3909 3.573 0.000353 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef) exp(-coef) lower .95 upper .95
ACTFALSE     4.042      0.2474     1.879     8.697

Concordance= 0.613  (se = 0.036 )
Likelihood ratio test= 9.94 on 1 df,  p=0.002
Wald test            = 12.77 on 1 df,  p=4e-04
Score (logrank) test = 14.86 on 1 df,  p=1e-04

```

Hide

```

cox_fit_summary <- summary(cox_fit)

# Extract values for HR, 95% CI, and p-value
HR <- cox_fit_summary$coefficients[2]
lower_CI <- cox_fit_summary$conf.int[3]
upper_CI <- cox_fit_summary$conf.int[4]
p_value <- cox_fit_summary$coefficients[5]
label_text <- paste0("HR = ", round(HR, 2), " (", round(lower_CI, 2), "-", round(upper_CI, 2), "); p = ", round(p_value, 3))
print(label_text)

```

```
[1] "HR = 4.04 (1.88-8.7); p = 0"
```

Hide

```

#Adjusted HR "ACT vs no ACT" - age, gender, MSI and pathological stage
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA CRC_Clinical Data 122023.csv")
circ_data <- circ_data[circ_data$CLIA.CRC==TRUE,]
circ_data <- circ_data[circ_data$Risk.Stage==TRUE,]
circ_data <- circ_data[circ_data$ctDNA.MRD!=""]
circ_data <- circ_data[circ_data$ctDNA.MRD=="POSITIVE",]
circ_data$DFS.months=circ_data$DFS.months-2
circ_data <- circ_data[circ_data$DFS.months>=0,]
circ_dataadf <- as.data.frame(circ_data)

circ_data$DFS.Event <- as.logical(circ_data$DFS.Event)
circ_data$DFS.months <- as.numeric(circ_data$DFS.months)
circ_data$ACT <- factor(circ_data$ACT, levels=c("TRUE","FALSE"))
circ_data$Age.Group <- factor(circ_data$Age.Group, levels = c("1", "2"), labels = c("<70", "≥70"))
circ_data$Gender <- factor(circ_data$Gender, levels = c("Female", "Male"))
circ_data$MSI <- factor(circ_data$MSI, levels = c("MSS", "MSI-High"))
circ_data$Stage <- factor(circ_data$Stage, levels = c("II", "III"))
surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
cox_fit <- coxph(surv_object ~ ACT + Age.Group + Gender + Stage, data=circ_data)
summary(cox_fit)

```

```

Call:
coxph(formula = surv_object ~ ACT + Age.Group + Gender + Stage,
      data = circ_data)

n= 61, number of events= 38

      coef exp(coef) se(coef)     z Pr(>|z|)
ACTFALSE    1.9310   6.8967  0.4580  4.216 2.48e-05 ***
Age.Group≥70 0.5599   1.7505  0.3603  1.554   0.1202
GenderMale   0.2086   1.2320  0.3471  0.601   0.5478
StageIII     1.3374   3.8093  0.5942  2.251   0.0244 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef) exp(-coef) lower .95 upper .95
ACTFALSE      6.897    0.1450   2.8106   16.923
Age.Group≥70   1.750    0.5713   0.8640    3.547
GenderMale     1.232    0.8117   0.6239    2.432
StageIII       3.809    0.2625   1.1887   12.208

Concordance= 0.669 (se = 0.044 )
Likelihood ratio test= 18.74 on 4 df,  p=9e-04
Wald test       = 20.61 on 4 df,  p=4e-04
Score (logrank) test = 22.94 on 4 df,  p=1e-04

```

Hide

```

# Adjust p-values using False Discovery Rate (FDR) adjustment (Benjamini-Hochberg method)
p_values <- summary(cox_fit)$coefficients[, 5]
adjusted_p_values <- p.adjust(p_values, method = "fdr")
results <- data.frame(
  Variable = rownames(summary(cox_fit)$coefficients),
  Original_P_Value = p_values,
  FDR_Adjusted_P_Value = adjusted_p_values
)
print(results)

```

	Variable	Original_P_Value	FDR_Adjusted_P_Value
	<chr>	<dbl>	<dbl>
ACTFALSE	ACTFALSE	2.482166e-05	9.928663e-05
Age.Group≥70	Age.Group≥70	1.201553e-01	1.602071e-01
GenderMale	GenderMale	5.478245e-01	5.478245e-01
StageIII	StageIII	2.439518e-02	4.879036e-02
4 rows			

#DFS by ctDNA at the MRD Window & ACT - MSS Stable all stages

Hide

```

rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA_CRC_Clinical Data 122023.csv")
circ_data <- circ_data[circ_data$CLIA.CRC==TRUE,]
circ_data <- circ_data[circ_data$ctDNA.MRD!="",]
circ_data <- circ_data[circ_data$MSI=="MSS",]
circ_data$DFS.months=circ_data$DFS.months-2.5
circ_data <- circ_data[circ_data$DFS.months>=0,]

circ_data$ctDNA.ACT <- NA #first we create the variable for the ctDNA & NAC combination, and we assign values
circ_data <- circ_data %>%
  mutate(ctDNA.ACT = case_when(
    ctDNA.MRD == "NEGATIVE" & ACT == "TRUE" ~ 1,
    ctDNA.MRD == "POSITIVE" & ACT == "TRUE" ~ 2,
    ctDNA.MRD == "NEGATIVE" & ACT == "FALSE" ~ 3,
    ctDNA.MRD == "POSITIVE" & ACT == "FALSE" ~ 4
  ))

circ_data$DFS.Event <- as.logical(circ_data$DFS.Event)
circ_data$DFS.months <- as.numeric(circ_data$DFS.months)
survfit(Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)~ctDNA.ACT, data = circ_data)

```

```
Call: survfit(formula = Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event) ~
  ctDNA.ACT, data = circ_data)
```

	n	events	median	0.95LCL	0.95UCL
ctDNA.ACT=1	169	19	NA	NA	NA
ctDNA.ACT=2	48	27	13.07	9.69	NA
ctDNA.ACT=3	106	6	NA	NA	NA
ctDNA.ACT=4	3	2	7.42	5.29	NA

Hide

```
event_summary <- circ_data %>%
  group_by(ctDNA.ACT) %>%
  summarise(
    Total = n(),
    Events = sum(DFS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)
```

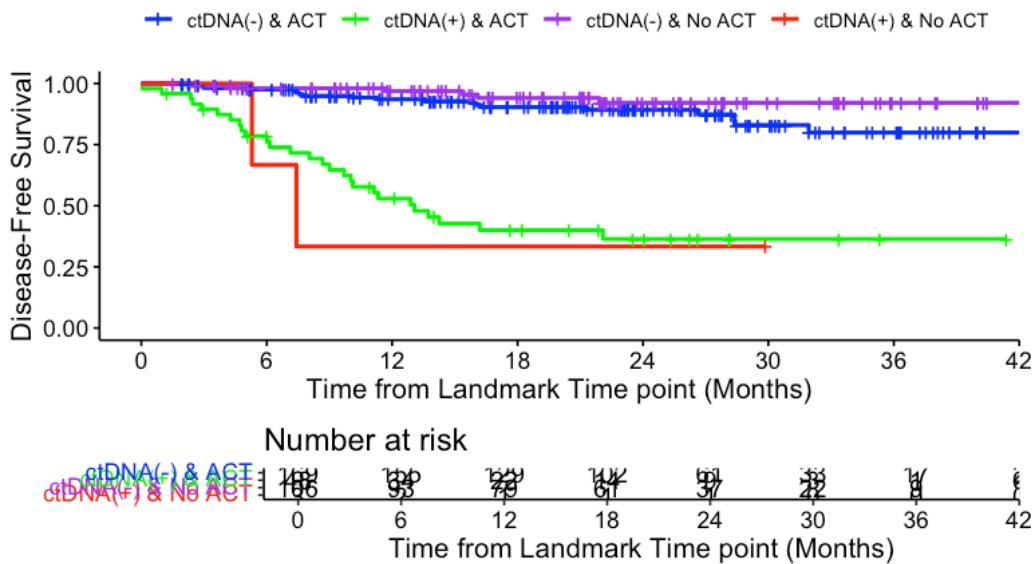
ctDNA.ACT	Total	Events	Fraction	Percentage
1	169	19	0.11242604	11.242604
2	48	27	0.56250000	56.250000
3	106	6	0.05660377	5.660377
4	3	2	0.66666667	66.666667

4 rows

Hide

```
surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
KM_curve <- survfit(surv_object ~ ctDNA.ACT, data = circ_data, conf.int=0.95, conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette = c("blue", "green", "purple", "red"), title="DFS - ctDNA MRD & ACT | MSS pts", ylab= "Disease-Free Survival", xlab = "Time from Landmark Time point (Months)", legend.labs=c("ctDNA(-) & ACT", "ctDNA(+) & ACT", "ctDNA(-) & No ACT", "ctDNA(+) & No ACT"), legend.title="")
```

DFS - ctDNA MRD & ACT | MSS pts



Hide

```
summary(KM_curve, times= c(24))
```

```
Call: survfit(formula = surv_object ~ ctDNA.ACT, data = circ_data,
  conf.int = 0.95, conf.type = "log-log")
```

ctDNA.ACT=1						
time	n.risk	n.event	survival	std.err	lower	95% CI upper
24.000	61.000	15.000	0.892	0.027	0.825	0.934
ctDNA.ACT=2						
24.0000	9.0000	27.0000	0.3640	0.0773	0.2174	0.5120
ctDNA.ACT=3						
24.000	37.000	6.000	0.921	0.032	0.829	0.965
ctDNA.ACT=4						
24.00000	1.00000	2.00000	0.33333	0.27217	0.00896	0.77415

Hide

```
circ_data$ctDNA.ACT <- factor(circ_data$ctDNA.ACT, levels=c("1","2","3","4"), labels = c("ctDNA(-) & ACT", "ctDNA(+) & ACT", "ctDNA(-) & No ACT", "ctDNA(+) & No ACT"))
cox_fit <- coxph(surv_object ~ ctDNA.ACT, data=circ_data) #modify maxexit to reveal NA values in cox_fit
summary(cox_fit)
```

```
Call:
coxph(formula = surv_object ~ ctDNA.ACT, data = circ_data)
```

```
n= 326, number of events= 54
```

	coef	exp(coef)	se(coef)	z	Pr(> z)
ctDNA.ACTctDNA(+) & ACT	2.0237	7.5661	0.3021	6.700	2.09e-11 ***
ctDNA.ACTctDNA(-) & No ACT	-0.6820	0.5056	0.4683	-1.456	0.14536
ctDNA.ACTctDNA(+) & No ACT	2.1276	8.3950	0.7463	2.851	0.00436 **

```
---
```

```
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '
```

	exp(coef)	exp(-coef)	lower .95	upper .95
ctDNA.ACTctDNA(+) & ACT	7.5661	0.1322	4.1856	13.677
ctDNA.ACTctDNA(-) & No ACT	0.5056	1.9778	0.2019	1.266
ctDNA.ACTctDNA(+) & No ACT	8.3950	0.1191	1.9443	36.247

```
Concordance= 0.759 (se = 0.034 )
```

```
Likelihood ratio test= 61 on 3 df, p=4e-13
```

```
Wald test = 65.55 on 3 df, p=4e-14
```

```
Score (logrank) test = 98.87 on 3 df, p=<2e-16
```

Hide

```

rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA CRC_Clinical Data 122023.csv")
circ_data <- circ_data[circ_data$CLIA.CRC==TRUE,]
circ_data <- circ_data[circ_data$ctDNA.MRD!="",]
circ_data <- circ_data[circ_data$MSI=="MSS",]
circ_data$DFS.months=circ_data$DFS.months-2.5
circ_data <- circ_data[circ_data$DFS.months>=0,]

circ_data$ctDNA.ACT <- NA #first we create the variable for the ctDNA & NAC combination, and we assign values
circ_data <- circ_data %>%
  mutate(ctDNA.ACT = case_when(
    ctDNA.MRD == "NEGATIVE" & ACT == "TRUE" ~ 1,
    ctDNA.MRD == "POSITIVE" & ACT == "TRUE" ~ 2,
    ctDNA.MRD == "NEGATIVE" & ACT == "FALSE" ~ 3,
    ctDNA.MRD == "POSITIVE" & ACT == "FALSE" ~ 4
  ))
  
circ_data$DFS.Event <- as.logical(circ_data$DFS.Event)
circ_data$DFS.months <- as.numeric(circ_data$DFS.months)
survfit(Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)~ctDNA.ACT, data = circ_data)

```

Call: survfit(formula = Surv(time = circ_data\$DFS.months, event = circ_data\$DFS.Event) ~
ctDNA.ACT, data = circ_data)

	n	events	median	0.95LCL	0.95UCL
ctDNA.ACT=1	169	19	NA	NA	NA
ctDNA.ACT=2	48	27	13.07	9.69	NA
ctDNA.ACT=3	106	6	NA	NA	NA
ctDNA.ACT=4	3	2	7.42	5.29	NA

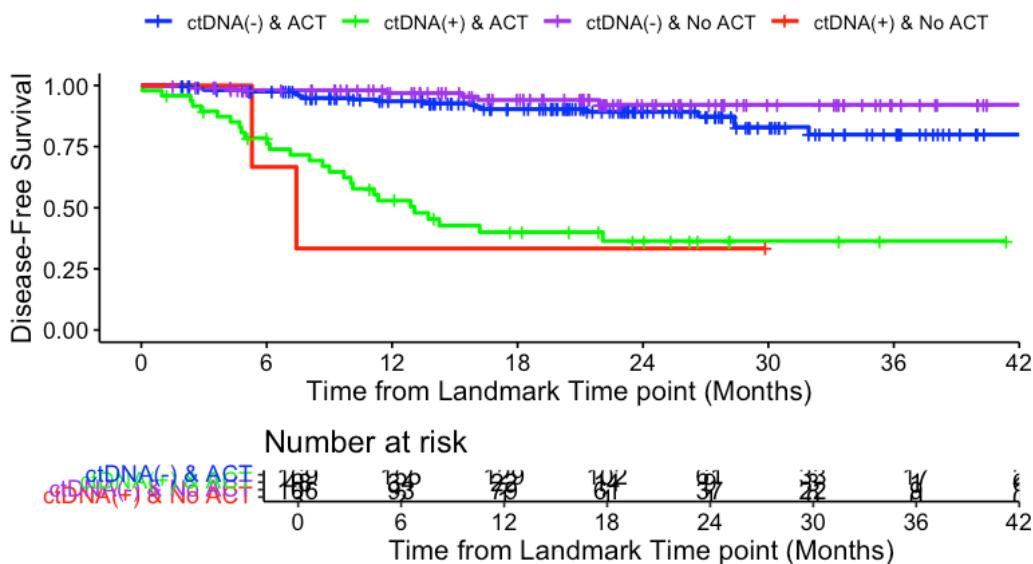
Hide

```

surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
KM_curve <- survfit(surv_object ~ ctDNA.ACT, data = circ_data, conf.int=0.95, conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette
e=c("blue","green","purple","red"), title="DFS - ctDNA MRD & ACT | MSS pts", ylab= "Disease-Free Survival", xlab
="Time from Landmark Time point (Months)", legend.labs=c("ctDNA(-) & ACT", "ctDNA(+) & ACT", "ctDNA(-) & No ACT",
"ctDNA(+) & No ACT"), legend.title="")

```

DFS - ctDNA MRD & ACT | MSS pts



Hide

```
summary(KM_curve, times= c(24))
```

```
Call: survfit(formula = surv_object ~ ctDNA.ACT, data = circ_data,
conf.int = 0.95, conf.type = "log-log")

  ctDNA.ACT=1
  time    n.risk    n.event    survival    std.err lower 95% CI upper 95% CI
  24.000  61.000    15.000    0.892      0.027    0.825    0.934

  ctDNA.ACT=2
  time    n.risk    n.event    survival    std.err lower 95% CI upper 95% CI
  24.000  9.0000    27.0000    0.3640     0.0773    0.2174    0.5120

  ctDNA.ACT=3
  time    n.risk    n.event    survival    std.err lower 95% CI upper 95% CI
  24.000  37.000     6.000     0.921      0.032    0.829    0.965

  ctDNA.ACT=4
  time    n.risk    n.event    survival    std.err lower 95% CI upper 95% CI
  24.0000 1.00000   2.00000    0.33333    0.27217   0.00896   0.77415
```

Hide

```
circ_data$ctDNA.ACT <- factor(circ_data$ctDNA.ACT, levels=c("3","1","2","4"), labels = c("ctDNA(-) & No ACT","ctDNA(-) & ACT", "ctDNA(+) & ACT", "ctDNA(+) & No ACT"))
cox_fit <- coxph(surv_object ~ ctDNA.ACT, data=circ_data)
summary(cox_fit)
```

Call:
 coxph(formula = surv_object ~ ctDNA.ACT, data = circ_data)

n= 326, number of events= 54

	coef	exp(coef)	se(coef)	z	Pr(> z)
ctDNA.ACTctDNA(-) & ACT	0.6820	1.9778	0.4683	1.456	0.145361
ctDNA.ACTctDNA(+) & ACT	2.7056	14.9638	0.4533	5.969	2.39e-09 ***
ctDNA.ACTctDNA(+) & No ACT	2.8096	16.6032	0.8192	3.430	0.000605 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95
ctDNA.ACTctDNA(-) & ACT	1.978	0.50562	0.7898	4.953
ctDNA.ACTctDNA(+) & ACT	14.964	0.06683	6.1544	36.383
ctDNA.ACTctDNA(+) & No ACT	16.603	0.06023	3.3331	82.704

Concordance= 0.759 (se = 0.034)

Likelihood ratio test= 61 on 3 df, p=4e-13

Wald test = 65.55 on 3 df, p=4e-14

Score (logrank) test = 98.87 on 3 df, p=<2e-16

#DFS by ctDNA at the MRD Window & ACT - MSI High all stages

Hide

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA_CRC_Clinical Data 122023.csv")
circ_data <- circ_data[circ_data$CLIA.CRC==TRUE,]
circ_data <- circ_data[circ_data$ctDNA.MRD!="",]
circ_data <- circ_data[circ_data$MSI=="MSI-High",]
circ_data$DFS.months=circ_data$DFS.months-2.5
circ_data <- circ_data[circ_data$DFS.months>=0,]

circ_data$ctDNA.ACT <- NA #first we create the variable for the ctDNA & NAC combination, and we assign values
circ_data <- circ_data %>%
  mutate(ctDNA.ACT = case_when(
    ctDNA.MRD == "NEGATIVE" & ACT == "TRUE" ~ 1,
    ctDNA.MRD == "POSITIVE" & ACT == "TRUE" ~ 2,
    ctDNA.MRD == "NEGATIVE" & ACT == "FALSE" ~ 3,
    ctDNA.MRD == "POSITIVE" & ACT == "FALSE" ~ 4
  ))

circ_data$DFS.Event <- as.logical(circ_data$DFS.Event)
circ_data$DFS.months <- as.numeric(circ_data$DFS.months)
survfit(Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)~ctDNA.ACT, data = circ_data)
```

```
Call: survfit(formula = Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event) ~
  ctDNA.ACT, data = circ_data)
```

	n	events	median	0.95LCL	0.95UCL
ctDNA.ACT=1	19	4	NA	NA	NA
ctDNA.ACT=2	5	2	NA	3.09	NA
ctDNA.ACT=3	34	0	NA	NA	NA
ctDNA.ACT=4	6	5	1.91	1.41	NA

Hide

```
event_summary <- circ_data %>%
  group_by(ctDNA.ACT) %>%
  summarise(
    Total = n(),
    Events = sum(DFS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)
```

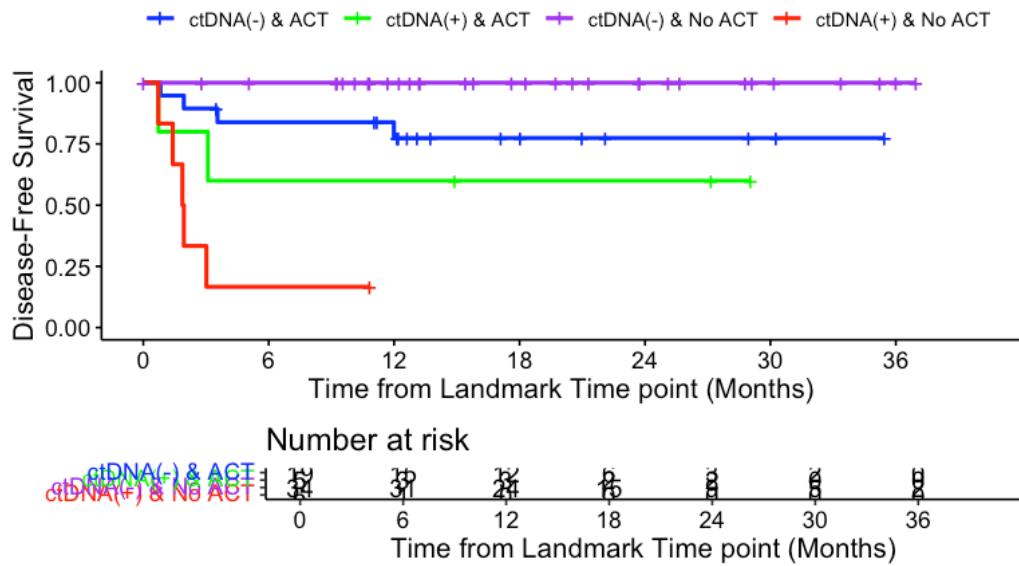
ctDNA.ACT	Total	Events	Fraction	Percentage
<dbl>	<int>	<int>	<dbl>	<dbl>
1	19	4	0.2105263	21.05263
2	5	2	0.4000000	40.00000
3	34	0	0.0000000	0.00000
4	6	5	0.8333333	83.33333

4 rows

Hide

```
surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
KM_curve <- survfit(surv_object ~ ctDNA.ACT, data = circ_data, conf.int=0.95, conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette
e=c("blue","green","purple","red"), title="DFS - ctDNA MRD & ACT | MSI-High pts", ylab= "Disease-Free Survival",
xlab="Time from Landmark Time point (Months)", legend.labs=c("ctDNA(-) & ACT", "ctDNA(+) & ACT", "ctDNA(-) & No A
CT", "ctDNA(+) & No ACT"), legend.title="")
```

DFS - ctDNA MRD & ACT | MSI-High pts



```
summary(KM_curve, times= c(6, 24))
```

```
Call: survfit(formula = surv_object ~ ctDNA.ACT, data = circ_data,
  conf.int = 0.95, conf.type = "log-log")
```

```
ctDNA.ACT=1
time n.risk n.event survival std.err lower 95% CI upper 95% CI
  6     15      3     0.839  0.0854      0.579     0.945
 24      3      1     0.774  0.1003      0.502     0.910
```

```
ctDNA.ACT=2
time n.risk n.event survival std.err lower 95% CI upper 95% CI
  6     3      2     0.6    0.219      0.126     0.882
 24     2      0     0.6    0.219      0.126     0.882
```

```
ctDNA.ACT=3
time n.risk n.event survival std.err lower 95% CI upper 95% CI
  6     31      0     1      0        NA        NA
 24     9      0     1      0        NA        NA
```

```
ctDNA.ACT=4
time n.risk n.event survival std.err lower 95% CI upper 95% CI
 6.00000  1.00000  5.00000  0.16667  0.15215  0.00772  0.51680
```

Hide

```
circ_data$ctDNA.ACT <- factor(circ_data$ctDNA.ACT, levels=c("3","1","2","4"), labels = c("ctDNA(-) & ACT", "ctDNA(+) & ACT", "ctDNA(-) & No ACT", "ctDNA(+) & No ACT"))
cox_fit <- coxphf(surv_object ~ ctDNA.ACT, data=circ_data, maxit = 500, maxstep = 1) #modify maxexit to reveal NA values in cox_fit
summary(cox_fit)
```

Hide

```

coxphf(formula = surv_object ~ ctDNA.ACT, data = circ_data, maxit = 500,
       maxstep = 1)

Model fitted by Penalized ML
Confidence intervals and p-values by Profile Likelihood

      coef se(coef) exp(coef) lower 0.95 upper 0.95      Chisq      p
ctDNA.ACTctDNA(+) & ACT  2.813027 1.622557 16.66027  1.778749 2207.805 6.609234 1.014513e-02
ctDNA.ACTctDNA(-) & No ACT 3.712830 1.686348 40.96957  3.329543 5652.756 8.497597 3.556157e-03
ctDNA.ACTctDNA(+) & No ACT 4.942760 1.627985 140.15657 14.930527 18714.894 24.151928 8.902699e-07

Likelihood ratio test=26.64366 on 3 df, p=6.992091e-06, n=64
Wald test = 14.30441 on 3 df, p = 0.002518759

Covariance-Matrix:
      ctDNA.ACTctDNA(+) & ACT ctDNA.ACTctDNA(-) & No ACT ctDNA.ACTctDNA(+) & No ACT
ctDNA.ACTctDNA(+) & ACT          2.632691          2.370004          2.371709
ctDNA.ACTctDNA(-) & No ACT      2.370004          2.843769          2.373722
ctDNA.ACTctDNA(+) & No ACT      2.371709          2.373722          2.650336

```

Hide

```

rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA CRC_Clinical Data 122023.csv")
circ_data <- circ_data[circ_data$CLIA.CRC==TRUE,]
circ_data <- circ_data[circ_data$ctDNA.MRD!="",]
circ_data <- circ_data[circ_data$MSI=="MSI-High",]
circ_data$DFS.months=circ_data$DFS.months-2.5
circ_data <- circ_data[circ_data$DFS.months>=0,]

circ_data$ctDNA.ACT <- NA #first we create the variable for the ctDNA & NAC combination, and we assign values
circ_data <- circ_data %>%
  mutate(ctDNA.ACT = case_when(
    ctDNA.MRD == "NEGATIVE" & ACT == "TRUE" ~ 1,
    ctDNA.MRD == "POSITIVE" & ACT == "TRUE" ~ 2,
    ctDNA.MRD == "NEGATIVE" & ACT == "FALSE" ~ 3,
    ctDNA.MRD == "POSITIVE" & ACT == "FALSE" ~ 4
  ))

circ_data$DFS.Event <- as.logical(circ_data$DFS.Event)
circ_data$DFS.months <- as.numeric(circ_data$DFS.months)
survfit(Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)~ctDNA.ACT, data = circ_data)

```

Call: survfit(formula = Surv(time = circ_data\$DFS.months, event = circ_data\$DFS.Event) ~ ctDNA.ACT, data = circ_data)

	n	events	median	0.95LCL	0.95UCL
ctDNA.ACT=1	19	4	NA	NA	NA
ctDNA.ACT=2	5	2	NA	3.09	NA
ctDNA.ACT=3	34	0	NA	NA	NA
ctDNA.ACT=4	6	5	1.91	1.41	NA

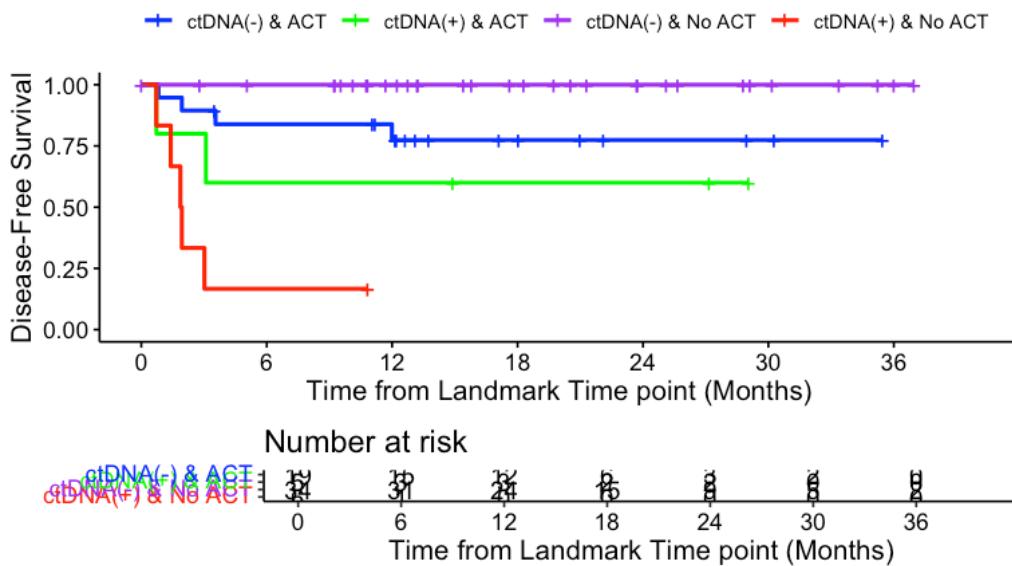
Hide

```

surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
KM_curve <- survfit(surv_object ~ ctDNA.ACT, data = circ_data, conf.int=0.95, conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette
e=c("blue","green","purple","red"), title="DFS - ctDNA MRD & ACT | MSI-High pts", ylab= "Disease-Free Survival",
xlab="Time from Landmark Time point (Months)", legend.labs=c("ctDNA(-) & ACT", "ctDNA(+) & ACT", "ctDNA(-) & No A
CT", "ctDNA(+) & No ACT"), legend.title="")

```

DFS - ctDNA MRD & ACT | MSI-High pts



Hide

```
summary(KM_curve, times= c(6))
```

```
Call: survfit(formula = surv_object ~ ctDNA.ACT, data = circ_data,
  conf.int = 0.95, conf.type = "log-log")
```

ctDNA.ACT=1						
time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
6.0000	15.0000	3.0000	0.8388	0.0854	0.5788	0.9451

ctDNA.ACT=2						
time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
6.000	3.000	2.000	0.600	0.219	0.126	0.882

ctDNA.ACT=3						
time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
6	31	0	1	0	NA	NA

ctDNA.ACT=4						
time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
6.0000	1.0000	5.00000	0.16667	0.15215	0.00772	0.51680

Hide

```
circ_data$ctDNA.ACT <- factor(circ_data$ctDNA.ACT, levels=c("2","4","1","3"))
cox_fit <- coxphf(surv_object ~ ctDNA.ACT, data=circ_data, maxit = 500) #modify maxexit to reveal NA values in co
x_fit
summary(cox_fit)
```

```
coxphf(formula = surv_object ~ ctDNA.ACT, data = circ_data, maxit = 500)
```

Model fitted by Penalized ML
Confidence intervals and p-values by Profile Likelihood

coef	se(coef)	exp(coef)	lower 0.95	upper 0.95	Chisq	p
ctDNA.ACT4	1.2299304	0.8640952	3.42099155	0.771944824	20.3070261	2.600353 0.106839899
ctDNA.ACT1	-0.8998028	0.8581680	0.40664983	0.090118999	2.3336235	1.163885 0.280661276
ctDNA.ACT3	-3.7128298	1.6863479	0.02440835	0.000176905	0.3003415	8.497597 0.003556157

Likelihood ratio test=26.64366 on 3 df, p=6.992091e-06, n=64
Wald test = 14.30441 on 3 df, p = 0.002518759

Covariance-Matrix:

ctDNA.ACT4	ctDNA.ACT1	ctDNA.ACT3	
ctDNA.ACT4	0.7466605	0.4717519	0.4700468
ctDNA.ACT1	0.4717519	0.7364524	0.4737650
ctDNA.ACT3	0.4700468	0.4737650	2.8437691

#DFS by ctDNA Dynamics from MRD to Surveillance Window - all stages

Hide

```

rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA CRC_Clinical Data 122023.csv")
circ_data <- circ_data[circ_data$CLIA.CRC==TRUE,]
circ_data <- circ_data[circ_data$ctDNA.MRD!=""]
circ_data$DFS.months=circ_data$DFS.months-2.5
circ_data <- circ_data[circ_data$DFS.months>=0,]

circ_data <- circ_data %>%
  mutate(ctDNA.Dynamics = case_when(
    ctDNA.MRD == "NEGATIVE" & ctDNA.Surveillance == "NEGATIVE" ~ 1,
    ctDNA.MRD == "POSITIVE" & ctDNA.Surveillance == "NEGATIVE" ~ 2,
    ctDNA.MRD == "NEGATIVE" & ctDNA.Surveillance == "POSITIVE" ~ 3,
    ctDNA.MRD == "POSITIVE" & ctDNA.Surveillance == "POSITIVE" ~ 4
  )) %>%
  filter(!is.na(ctDNA.Dynamics))

circ_data$DFS.Event <- as.logical(circ_data$DFS.Event)
circ_data$DFS.months <- as.numeric(circ_data$DFS.months)
circ_data <- circ_data[circ_data$ctDNA.Dynamics!=""]
survfit(Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)~ctDNA.Dynamics, data = circ_data)

```

Call: survfit(formula = Surv(time = circ_data\$DFS.months, event = circ_data\$DFS.Event) ~ ctDNA.Dynamics, data = circ_data)

	n	events	median	0.95LCL	0.95UCL
ctDNA.Dynamics=1	263	10	NA	NA	NA
ctDNA.Dynamics=2	16	1	NA	NA	NA
ctDNA.Dynamics=3	25	11	28.4	15.77	NA
ctDNA.Dynamics=4	20	15	10.6	8.05	NA

Hide

```

event_summary <- circ_data %>%
  group_by(ctDNA.Dynamics) %>%
  summarise(
    Total = n(),
    Events = sum(DFS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)

```

ctDNA.Dynamics <dbl>	Total <int>	Events <int>	Fraction <dbl>	Percentage <dbl>
1	263	10	0.03802281	3.802281
2	16	1	0.06250000	6.250000
3	25	11	0.44000000	44.000000
4	20	15	0.75000000	75.000000

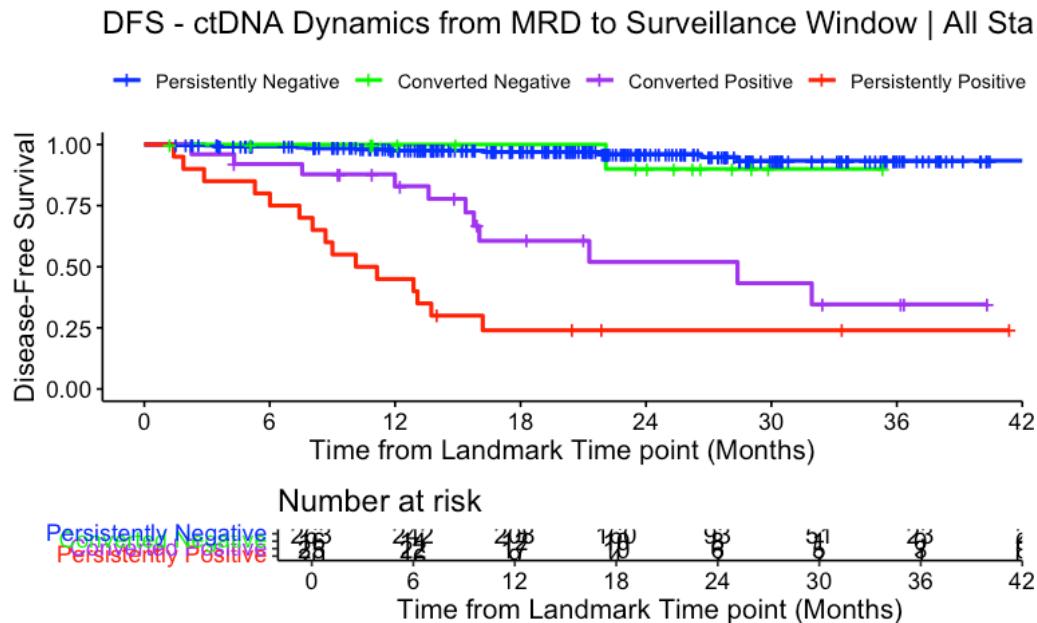
4 rows

Hide

```

surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
KM_curve <- survfit(surv_object ~ ctDNA.Dynamics, data = circ_data, conf.int=0.95,conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette = c("blue","green","purple", "red"), title="DFS - ctDNA Dynamics from MRD to Surveillance Window | All Stages", ylab= "Disease-Free Survival", xlab="Time from Landmark Time point (Months)", legend.labs=c("Persistently Negative", "Converted Negative", "Converted Positive", "Persistently Positive"), legend.title="")

```



Hide

```
summary(KM_curve, times= c(24))
```

```
Call: survfit(formula = surv_object ~ ctDNA.Dynamics, data = circ_data,
  conf.int = 0.95, conf.type = "log-log")

  ctDNA.Dynamics=1
  time      n.risk      n.event      survival      std.err      lower      95% CI      upper      95% CI
  24.000    93.000     8.0000     0.961        0.014      0.922      0.981

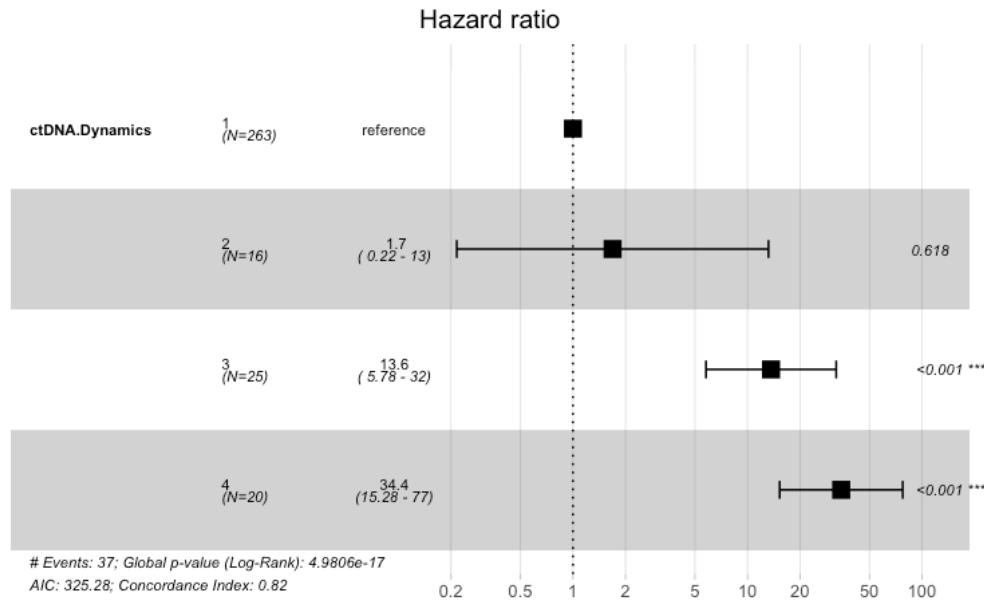
  ctDNA.Dynamics=2
  time      n.risk      n.event      survival      std.err      lower      95% CI      upper      95% CI
  24.0000   8.0000     1.0000     0.9000      0.0949      0.4730      0.9853

  ctDNA.Dynamics=3
  time      n.risk      n.event      survival      std.err      lower      95% CI      upper      95% CI
  24.000    6.0000     9.0000     0.519        0.125      0.260      0.727

  ctDNA.Dynamics=4
  time      n.risk      n.event      survival      std.err      lower      95% CI      upper      95% CI
  24.0000   2.0000    15.0000     0.2400      0.0980      0.0821      0.4428
```

Hide

```
circ_data$ctDNA.Dynamics <- factor(circ_data$ctDNA.Dynamics, levels=c("1","2","3","4"))
cox_fit <- coxph(surv_object ~ ctDNA.Dynamics, data=circ_data)
ggforest(cox_fit, data = circ_data)
```



Hide

```
summary(cox_fit)
```

Call:
 coxph(formula = surv_object ~ ctDNA.Dynamics, data = circ_data)

n= 324, number of events= 37

	coef	exp(coef)	se(coef)	z	Pr(> z)
ctDNA.Dynamics2	0.5229	1.6869	1.0493	0.498	0.618
ctDNA.Dynamics3	2.6136	13.6480	0.4383	5.963	2.48e-09 ***
ctDNA.Dynamics4	3.5381	34.4018	0.4141	8.545	< 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95
ctDNA.Dynamics2	1.687	0.59281	0.2157	13.19
ctDNA.Dynamics3	13.648	0.07327	5.7804	32.22
ctDNA.Dynamics4	34.402	0.02907	15.2803	77.45

Concordance= 0.822 (se = 0.04)
 Likelihood ratio test= 79.02 on 3 df, p=<2e-16
 Wald test = 77.71 on 3 df, p=<2e-16
 Score (logrank) test = 168.8 on 3 df, p=<2e-16

Hide

```

rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA CRC_Clinical Data 122023.csv")
circ_data <- circ_data[circ_data$CLIA.CRC==TRUE,]
circ_data <- circ_data[circ_data$ctDNA.MRD!="",]
circ_data$DFS.months=circ_data$DFS.months-2.5
circ_data <- circ_data[circ_data$DFS.months>=0,]

circ_data <- circ_data %>%
  mutate(ctDNA.Dynamics = case_when(
    ctDNA.MRD == "NEGATIVE" & ctDNA.Surveillance == "NEGATIVE" ~ 1,
    ctDNA.MRD == "POSITIVE" & ctDNA.Surveillance == "NEGATIVE" ~ 2,
    ctDNA.MRD == "NEGATIVE" & ctDNA.Surveillance == "POSITIVE" ~ 3,
    ctDNA.MRD == "POSITIVE" & ctDNA.Surveillance == "POSITIVE" ~ 4
  )) %>%
  filter(!is.na(ctDNA.Dynamics))

circ_data$DFS.Event <- as.logical(circ_data$DFS.Event)
circ_data$DFS.months <- as.numeric(circ_data$DFS.months)
circ_data <- circ_data[circ_data$ctDNA.Dynamics!="",]
survfit(Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)~ctDNA.Dynamics, data = circ_data)

```

```

Call: survfit(formula = Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event) ~
  ctDNA.Dynamics, data = circ_data)

      n  events median 0.95LCL 0.95UCL
ctDNA.Dynamics=1 263      10      NA      NA
ctDNA.Dynamics=2  16       1      NA      NA
ctDNA.Dynamics=3  25      11  28.4  15.77      NA
ctDNA.Dynamics=4  20      15  10.6   8.05      NA

```

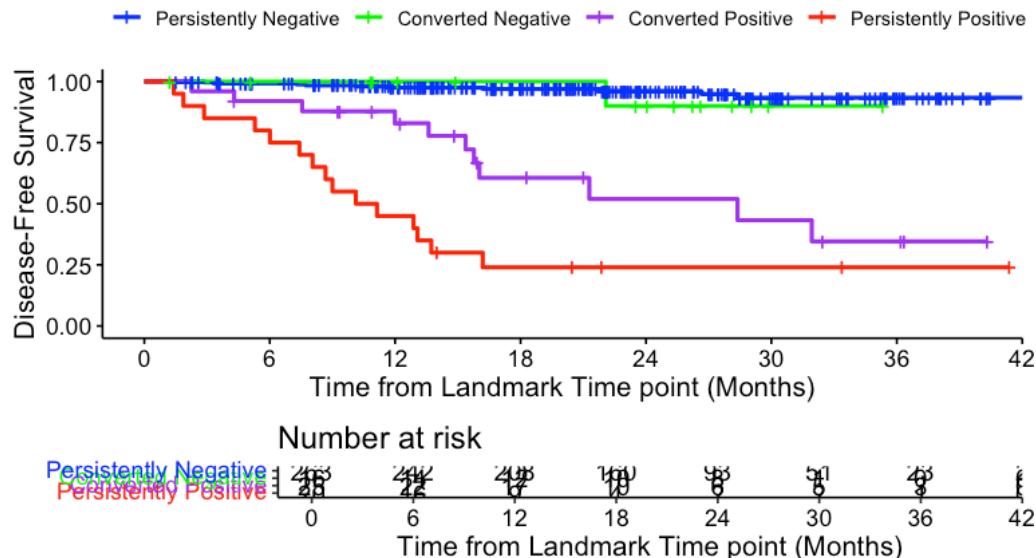
Hide

```

surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
KM_curve <- survfit(surv_object ~ ctDNA.Dynamics, data = circ_data, conf.int=0.95, conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette
e=c("blue","green","purple","red"), title="DFS - ctDNA Dynamics from MRD to Surveillance Window | All Stages", y
lab= "Disease-Free Survival", xlab="Time from Landmark Time point (Months)", legend.labs=c("Persistently Negativ
e", "Converted Negative", "Converted Positive", "Persistently Positive"), legend.title="")

```

DFS - ctDNA Dynamics from MRD to Surveillance Window | All Sta



Hide

```
summary(KM_curve, times= c(24))
```

```
Call: survfit(formula = surv_object ~ ctDNA.Dynamics, data = circ_data,
conf.int = 0.95, conf.type = "log-log")

ctDNA.Dynamics=1
  time    n.risk    n.event    survival    std.err lower 95% CI upper 95% CI
 24.000  93.000     8.000     0.961     0.014    0.922    0.981

ctDNA.Dynamics=2
  time    n.risk    n.event    survival    std.err lower 95% CI upper 95% CI
24.0000  8.0000    1.0000     0.9000    0.0949    0.4730    0.9853

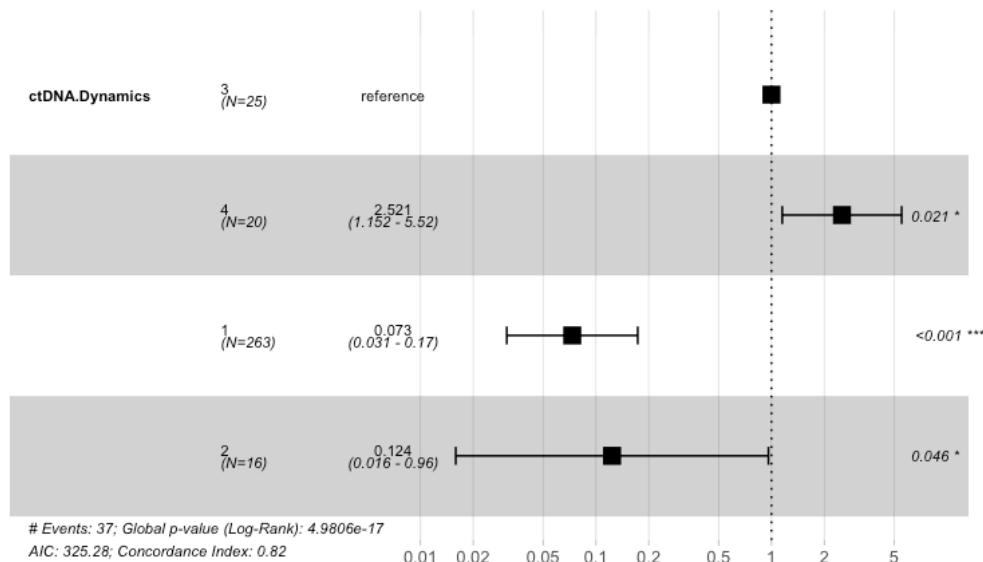
ctDNA.Dynamics=3
  time    n.risk    n.event    survival    std.err lower 95% CI upper 95% CI
 24.000   6.000     9.000     0.519     0.125    0.260    0.727

ctDNA.Dynamics=4
  time    n.risk    n.event    survival    std.err lower 95% CI upper 95% CI
24.0000  2.0000   15.0000     0.2400    0.0980    0.0821    0.4428
```

Hide

```
circ_data$ctDNA.Dynamics <- factor(circ_data$ctDNA.Dynamics, levels=c("3","4","1","2"))
cox_fit <- coxph(surv_object ~ ctDNA.Dynamics, data=circ_data)
ggforest(cox_fit, data = circ_data)
```

Hazard ratio



Hide

```
summary(cox_fit)
```

```
Call:
coxph(formula = surv_object ~ ctDNA.Dynamics, data = circ_data)

n= 324, number of events= 37

            coef exp(coef) se(coef)      z Pr(>|z|)
ctDNA.Dynamics4  0.92452  2.52065  0.39952  2.314  0.0207 *
ctDNA.Dynamics1 -2.61359  0.07327  0.43834 -5.963 2.48e-09 ***
ctDNA.Dynamics2 -2.09071  0.12360  1.04642 -1.998  0.0457 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

            exp(coef) exp(-coef) lower .95 upper .95
ctDNA.Dynamics4  2.52065     0.3967   1.15197    5.515
ctDNA.Dynamics1  0.07327    13.6480   0.03103     0.173
ctDNA.Dynamics2  0.12360     8.0907   0.01590     0.961

Concordance= 0.822 (se = 0.04 )
Likelihood ratio test= 79.02 on 3 df,  p=<2e-16
Wald test          = 77.71 on 3 df,  p=<2e-16
Score (logrank) test = 168.8 on 3 df,  p=<2e-16
```

#DFS by ctDNA Dynamics from MRD to Surveillance Window - High Risk Stage II or Stage III

Hide

```

rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA_CRC_Clinical Data 122023.csv")
circ_data <- circ_data[circ_data$CLIA.CRC==TRUE,]
circ_data <- circ_data[circ_data$Risk.Stage==TRUE,]
circ_data <- circ_data[circ_data$ctDNA.MRD!=""]
circ_data$DFS.months=circ_data$DFS.months-2.5
circ_data <- circ_data[circ_data$DFS.months>=0,]

circ_data <- circ_data %>%
  mutate(ctDNA.Dynamics = case_when(
    ctDNA.MRD == "NEGATIVE" & ctDNA.Surveillance == "NEGATIVE" ~ 1,
    ctDNA.MRD == "POSITIVE" & ctDNA.Surveillance == "NEGATIVE" ~ 2,
    ctDNA.MRD == "NEGATIVE" & ctDNA.Surveillance == "POSITIVE" ~ 3,
    ctDNA.MRD == "POSITIVE" & ctDNA.Surveillance == "POSITIVE" ~ 4
  )) %>%
  filter(!is.na(ctDNA.Dynamics))

circ_data$DFS.Event <- as.logical(circ_data$DFS.Event)
circ_data$DFS.months <- as.numeric(circ_data$DFS.months)
circ_data <- circ_data[circ_data$ctDNA.Dynamics!=""]
survfit(Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)~ctDNA.Dynamics, data = circ_data)

```

Call: survfit(formula = Surv(time = circ_data\$DFS.months, event = circ_data\$DFS.Event) ~ ctDNA.Dynamics, data = circ_data)

	n	events	median	0.95LCL	0.95UCL
ctDNA.Dynamics=1	224	10	NA	NA	NA
ctDNA.Dynamics=2	14	1	NA	NA	NA
ctDNA.Dynamics=3	22	10	28.4	16.03	NA
ctDNA.Dynamics=4	19	15	10.1	8.05	NA

Hide

```

event_summary <- circ_data %>%
  group_by(ctDNA.Dynamics) %>%
  summarise(
    Total = n(),
    Events = sum(DFS.Event),
    Fraction = Events / n(),
    Percentage = (Events / n()) * 100
  )
print(event_summary)

```

ctDNA.Dynamics	Total	Events	Fraction	Percentage
<dbl>	<int>	<int>	<dbl>	<dbl>
1	224	10	0.04464286	4.464286
2	14	1	0.07142857	7.142857
3	22	10	0.45454545	45.454545
4	19	15	0.78947368	78.947368

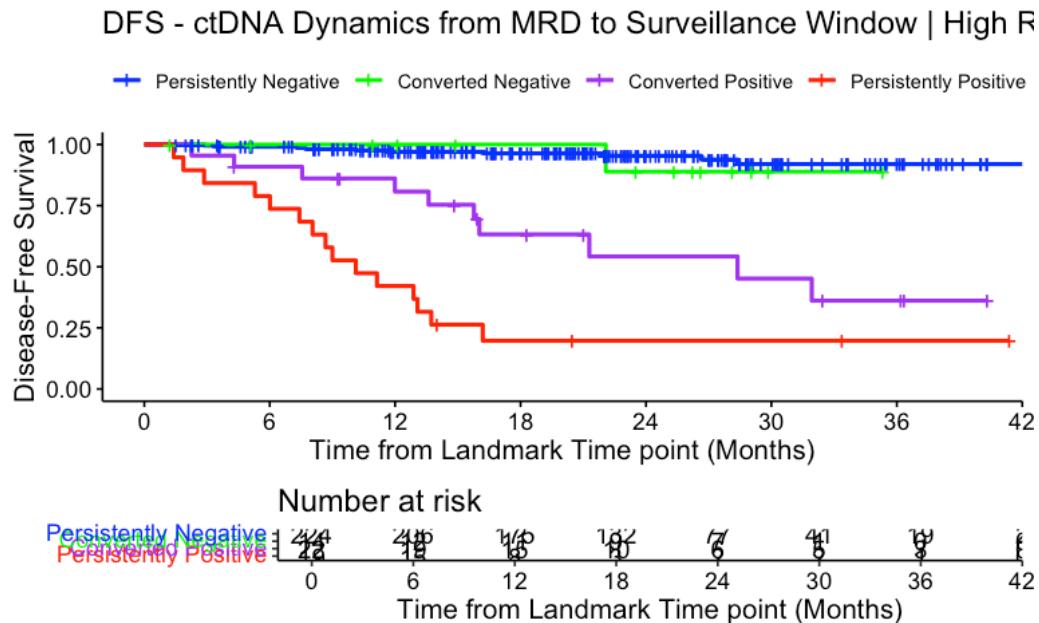
4 rows

Hide

```

surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
KM_curve <- survfit(surv_object ~ ctDNA.Dynamics, data = circ_data, conf.int=0.95, conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette = c("blue", "green", "purple", "red"), title="DFS - ctDNA Dynamics from MRD to Surveillance Window | High Risk Stage II or Stage III", ylab= "Disease-Free Survival", xlab="Time from Landmark Time point (Months)", legend.labs=c("Persistently Negative", "Converted Negative", "Converted Positive", "Persistently Positive"), legend.title="")

```



Hide

```
summary(KM_curve, times= c(24))
```

```
Call: survfit(formula = surv_object ~ ctDNA.Dynamics, data = circ_data,
  conf.int = 0.95, conf.type = "log-log")

  ctDNA.Dynamics=1
  time      n.risk      n.event      survival      std.err      lower      95% CI      upper      95% CI
  24.0000    77.0000     8.0000    0.9539      0.0166      0.9073      0.9774

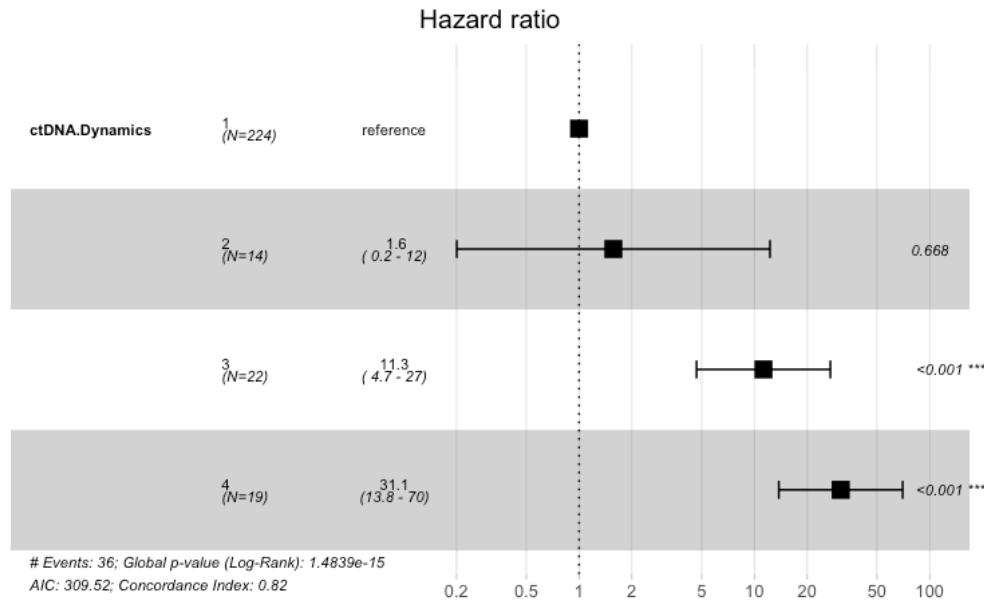
  ctDNA.Dynamics=2
  time      n.risk      n.event      survival      std.err      lower      95% CI      upper      95% CI
  24.0000    7.0000      1.0000    0.8890      0.1050      0.4330      0.9840

  ctDNA.Dynamics=3
  time      n.risk      n.event      survival      std.err      lower      95% CI      upper      95% CI
  24.0000    6.0000      8.0000    0.5420      0.1280      0.2720      0.7500

  ctDNA.Dynamics=4
  time      n.risk      n.event      survival      std.err      lower      95% CI      upper      95% CI
  24.0000    2.0000     15.0000    0.1974      0.0948      0.0551      0.4032
```

Hide

```
circ_data$ctDNA.Dynamics <- factor(circ_data$ctDNA.Dynamics, levels=c("1","2","3","4"))
cox_fit <- coxph(surv_object ~ ctDNA.Dynamics, data=circ_data)
ggforest(cox_fit, data = circ_data)
```



Hide

```
summary(cox_fit)
```

Call:
 coxph(formula = surv_object ~ ctDNA.Dynamics, data = circ_data)

n= 279, number of events= 36

	coef	exp(coef)	se(coef)	z	Pr(> z)
ctDNA.Dynamics2	0.4503	1.5688	1.0494	0.429	0.668
ctDNA.Dynamics3	2.4212	11.2599	0.4483	5.401	6.64e-08 ***
ctDNA.Dynamics4	3.4366	31.0803	0.4147	8.286	< 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95
ctDNA.Dynamics2	1.569	0.63743	0.2006	12.27
ctDNA.Dynamics3	11.260	0.08881	4.6765	27.11
ctDNA.Dynamics4	31.080	0.03217	13.7872	70.06

Concordance= 0.816 (se = 0.04)
 Likelihood ratio test= 72.14 on 3 df, p=1e-15
 Wald test = 72.35 on 3 df, p=1e-15
 Score (logrank) test = 150.9 on 3 df, p=<2e-16

Hide

```

rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA CRC_Clinical Data 122023.csv")
circ_data <- circ_data[circ_data$CLIA.CRC==TRUE,]
circ_data <- circ_data[circ_data$Risk.Stage==TRUE,]
circ_data <- circ_data[circ_data$ctDNA.MRD!=""]
circ_data$DFS.months=circ_data$DFS.months-2.5
circ_data <- circ_data[circ_data$DFS.months>=0,]

circ_data <- circ_data %>%
  mutate(ctDNA.Dynamics = case_when(
    ctDNA.MRD == "NEGATIVE" & ctDNA.Surveillance == "NEGATIVE" ~ 1,
    ctDNA.MRD == "POSITIVE" & ctDNA.Surveillance == "NEGATIVE" ~ 2,
    ctDNA.MRD == "NEGATIVE" & ctDNA.Surveillance == "POSITIVE" ~ 3,
    ctDNA.MRD == "POSITIVE" & ctDNA.Surveillance == "POSITIVE" ~ 4
  )) %>%
  filter(!is.na(ctDNA.Dynamics))

circ_data$DFS.Event <- as.logical(circ_data$DFS.Event)
circ_data$DFS.months <- as.numeric(circ_data$DFS.months)
circ_data <- circ_data[circ_data$ctDNA.Dynamics!=""]
survfit(Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)~ctDNA.Dynamics, data = circ_data)

```

```

Call: survfit(formula = Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event) ~
  ctDNA.Dynamics, data = circ_data)

      n events median 0.95LCL 0.95UCL
ctDNA.Dynamics=1 224      10      NA      NA      NA
ctDNA.Dynamics=2  14       1      NA      NA      NA
ctDNA.Dynamics=3  22      10    28.4   16.03      NA
ctDNA.Dynamics=4  19      15    10.1    8.05      NA

```

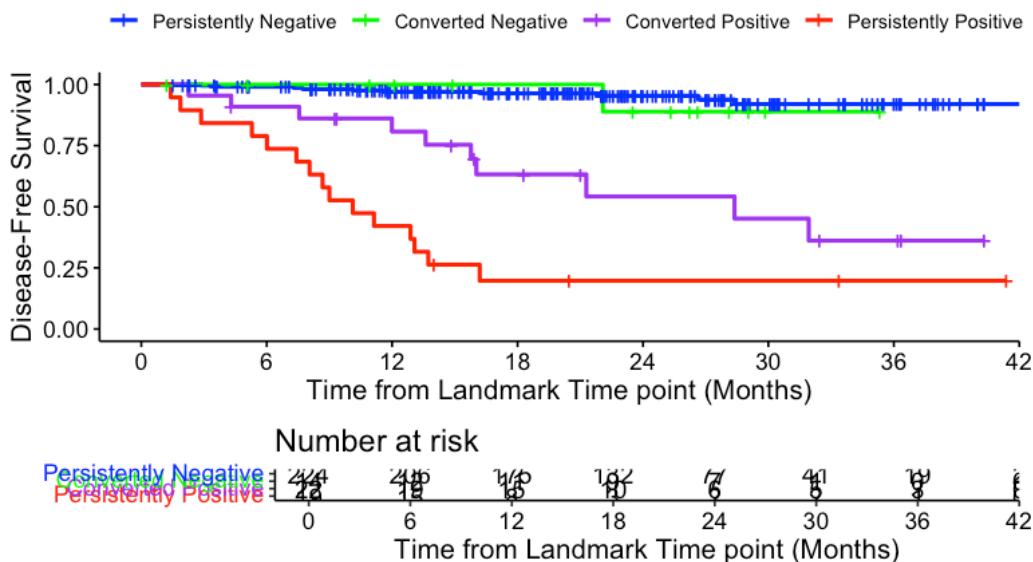
Hide

```

surv_object <- Surv(time = circ_data$DFS.months, event = circ_data$DFS.Event)
KM_curve <- survfit(surv_object ~ ctDNA.Dynamics, data = circ_data, conf.int=0.95, conf.type="log-log")
ggsurvplot(KM_curve, data = circ_data, pval = FALSE, conf.int = FALSE, risk.table = TRUE, break.time.by=6, palette = c("blue", "green", "purple", "red"), title="DFS - ctDNA Dynamics from MRD to Surveillance Window | High Risk Stage II or Stage III", ylab = "Disease-Free Survival", xlab="Time from Landmark Time point (Months)", legend.labs=c("Persistently Negative", "Converted Negative", "Converted Positive", "Persistently Positive"), legend.title="")

```

DFS - ctDNA Dynamics from MRD to Surveillance Window | High R



Hide

```
summary(KM_curve, times= c(24))
```

```
Call: survfit(formula = surv_object ~ ctDNA.Dynamics, data = circ_data,
conf.int = 0.95, conf.type = "log-log")

      ctDNA.Dynamics=1
  time   n.risk   n.event   survival   std.err lower 95% CI upper 95% CI
24.000  77.000    8.0000   0.9539    0.0166   0.9073   0.9774

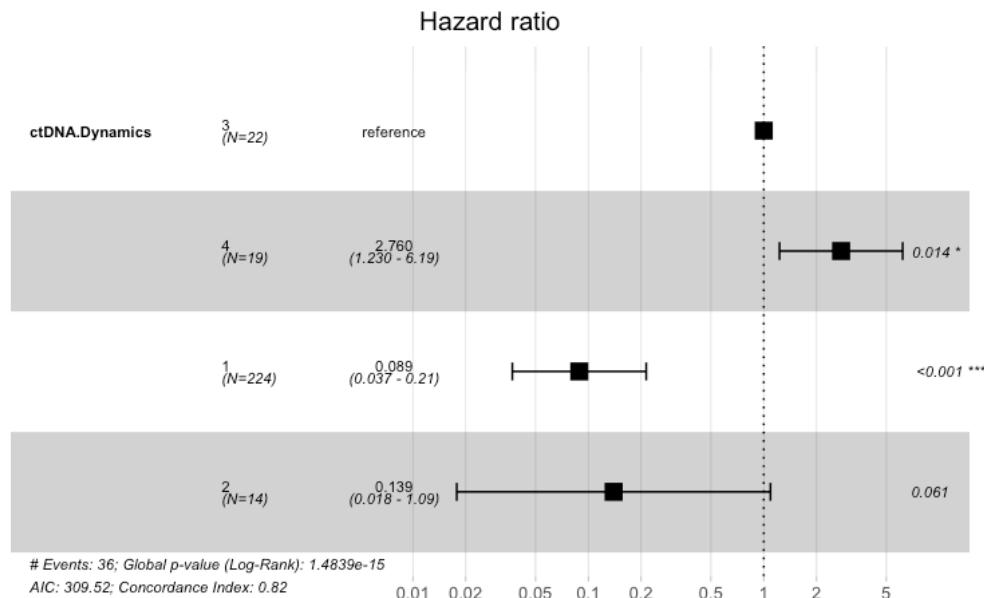
      ctDNA.Dynamics=2
  time   n.risk   n.event   survival   std.err lower 95% CI upper 95% CI
24.000    7.000    1.0000   0.889     0.105   0.433    0.984

      ctDNA.Dynamics=3
  time   n.risk   n.event   survival   std.err lower 95% CI upper 95% CI
24.000    6.000    8.0000   0.542     0.128   0.272    0.750

      ctDNA.Dynamics=4
  time   n.risk   n.event   survival   std.err lower 95% CI upper 95% CI
24.000   2.0000   15.0000   0.1974    0.0948   0.0551   0.4032
```

Hide

```
circ_data$ctDNA.Dynamics <- factor(circ_data$ctDNA.Dynamics, levels=c("3","4","1","2"))
cox_fit <- coxph(surv_object ~ ctDNA.Dynamics, data=circ_data)
ggforest(cox_fit, data = circ_data)
```



Hide

```
summary(cox_fit)
```

```
Call:
coxph(formula = surv_object ~ ctDNA.Dynamics, data = circ_data)

n= 279, number of events= 36

      coef exp(coef) se(coef)      z Pr(>|z|)
ctDNA.Dynamics4  1.01533  2.76027  0.41237  2.462  0.0138 *
ctDNA.Dynamics1 -2.42124  0.08881  0.44832 -5.401 6.64e-08 ***
ctDNA.Dynamics2 -1.97094  0.13933  1.05059 -1.876  0.0607 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef) exp(-coef) lower .95 upper .95
ctDNA.Dynamics4  2.76027     0.3623   1.23012   6.1938
ctDNA.Dynamics1  0.08881    11.2599   0.03689   0.2138
ctDNA.Dynamics2  0.13933     7.1774   0.01777   1.0922

Concordance= 0.816 (se = 0.04 )
Likelihood ratio test= 72.14 on 3 df,  p=1e-15
Wald test          = 72.35 on 3 df,  p=1e-15
Score (logrank) test = 150.9 on 3 df,  p=<2e-16
```

#Median cfDNA concentration across different windows post-surgery

Hide

```
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA CRC cfDNA data_ctDNA timepoints.csv")
circ_data <- circ_data[circ_data$cfDNA.Include == TRUE,]
circ_data$Timepoint <- factor(circ_data$Timepoint, levels = c("0-2 weeks", "2-4 weeks", "4-6 weeks", "6-8 weeks",
"8-10 weeks", "On-treatment", "Surveillance"))
circ_data$cfDNAconc <- as.numeric(circ_data$cfDNAconc)
median_cfDNAconc <- circ_data %>%
  group_by(Timepoint) %>%
  dplyr::summarize(
    median_cfDNAconc = median(cfDNAconc, na.rm = TRUE),
    n_observations = n(),
    .groups = "drop"
  )
print(median_cfDNAconc)
```

Timepoint	median_cfDNAconc	n_observations
<fctr>	<dbl>	<int>
0-2 weeks	8.695588	11
2-4 weeks	8.000202	140
4-6 weeks	5.692500	173
6-8 weeks	5.619716	84
8-10 weeks	5.283333	57
On-treatment	8.361000	514
Surveillance	5.030206	2912
7 rows		

Hide

```

boxplot(cfDNAconc~Timepoint, data=circ_data, main="median cfDNA Concentration", xlab="Intervals from Surgery", yl
ab="cfDNA Concentration", col="white", border="black", ylim = c(0, 40))

#Pairwise Wilcoxon-test p values
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA CRC cfDNA data_ctDNA timepoints.csv")
circ_data <- circ_data[circ_data$cfDNA.Include == TRUE,]
circ_data$Timepoint <- factor(circ_data$Timepoint, levels = c("0-2 weeks", "2-4 weeks", "4-6 weeks", "6-8 weeks",
"8-10 weeks", "On-treatment", "Surveillance"))
circ_data$cfDNAconc <- as.numeric(circ_data$cfDNAconc)

# Initialize a matrix to store p-values for manual Wilcoxon tests
timepoints <- levels(circ_data$Timepoint)
p_value_matrix <- matrix(NA, nrow = length(timepoints), ncol = length(timepoints))
rownames(p_value_matrix) <- timepoints
colnames(p_value_matrix) <- timepoints

# Perform pairwise Wilcoxon tests manually and store p-values
for (i in 1:length(timepoints)) {
  for (j in i:length(timepoints)) {
    if (i != j) {
      # Subset data for the two Timepoints
      data1 <- circ_data %>% filter(Timepoint == timepoints[i]) %>% pull(cfDNAconc)
      data2 <- circ_data %>% filter(Timepoint == timepoints[j]) %>% pull(cfDNAconc)

      # Perform Wilcoxon test and store the p-value
      test_result <- wilcox.test(data1, data2, exact = FALSE)
      p_value_matrix[i, j] <- test_result$p.value
      p_value_matrix[j, i] <- test_result$p.value # Symmetric matrix
    } else {
      p_value_matrix[i, j] <- 1 # Set diagonal to 1 for clarity
    }
  }
}

# Replace NA values with 1 in the p-value matrix for completeness
p_value_matrix[is.na(p_value_matrix)] <- 1.00

# Melt the p-value matrix for plotting
p_value_data <- melt(p_value_matrix)
colnames(p_value_data) <- c("Timepoint1", "Timepoint2", "p_value")

# Add asterisks based on p-value thresholds
p_value_data <- p_value_data %>%
  mutate(
    asterisk = case_when(
      p_value < 0.001 ~ "***",
      p_value < 0.01 ~ "**",
      p_value < 0.05 ~ "*",
      TRUE ~ ""
    )
  )

# Create the heatmap with asterisks for significance levels
ggplot(p_value_data, aes(x = Timepoint1, y = Timepoint2, fill = p_value)) +
  geom_tile(color = "white") + # Background tiles for grid
  geom_text(aes(label = asterisk), color = "black", vjust = -0.5) + # Add asterisks
  scale_fill_gradient(low = "lightgreen", high = "red") + # Gradient for p-values
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  labs(title = "Pairwise Wilcoxon-Test P-Values",
       fill = "P-Value")

```

#ctDNA positivity across different windows post-surgery

```

rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA CRC cfDNA data_ctDNA timepoints.csv")
circ_data <- circ_data[circ_data$cfDNA.Include == TRUE,]
circ_data$Timepoint <- factor(circ_data$Timepoint, levels = c("0-2 weeks", "2-4 weeks", "4-6 weeks", "6-8 weeks",
"8-10 weeks", "On-treatment", "Surveillance"))
circ_data$cfDNAconc <- as.numeric(circ_data$cfDNAconc)

rate_by_timepoint <- circ_data %>%
  group_by(Timepoint) %>%
  summarise(
    n_total = n(), # Total number of patients in each Timepoint
    n_positive = sum(biomarker_status == "POSITIVE"), # Number of positive cases
    n_negative = sum(biomarker_status == "NEGATIVE"), # Number of negative cases
    percentage_positive = mean(biomarker_status == "POSITIVE") * 100, # Positivity rate
    percentage_negative = mean(biomarker_status == "NEGATIVE") * 100 # Negativity rate
  )
print(rate_by_timepoint)

```

Timepoint	n_total	n_positive	n_negative	percentage_positive	percentage_negative
<fctr>	<int>	<int>	<int>	<dbl>	<dbl>
0-2 weeks	11	1	10	9.090909	90.90909
2-4 weeks	140	17	123	12.142857	87.85714
4-6 weeks	173	32	141	18.497110	81.50289
6-8 weeks	84	22	62	26.190476	73.80952
8-10 weeks	57	11	46	19.298246	80.70175
On-treatment	514	64	450	12.451362	87.54864
Surveillance	2912	160	2752	5.494505	94.50549

7 rows

Hide

```

# Create the stacked bar plot for positivity and negativity rates by Timepoint
bar_midpoints <- barplot(
  t(as.matrix(rate_by_timepoint[, c("percentage_positive", "percentage_negative")])), # Transpose to get the correct format
  names.arg = rate_by_timepoint$Timepoint,
  col = c("red", "blue"), # Colors: red for positive, blue for negative
  main = '% ctDNA Positive and Negative Samples by Timepoint',
  xlab = 'Timepoint',
  ylab = '% ctDNA Samples',
  ylim = c(0, 100),
  legend = c("% Positive", "% Negative"), # Adding a legend for clarification
  args.legend = list(x = "topright")
)
par(new = TRUE)

```

Hide

```

plot(bar_midpoints, rate_by_timepoint$n_total, type = "b", col = "black", pch = 19, axes = FALSE, xlab = "", ylab = "", lwd = 2)
axis(side = 4) # Add the secondary y-axis on the right

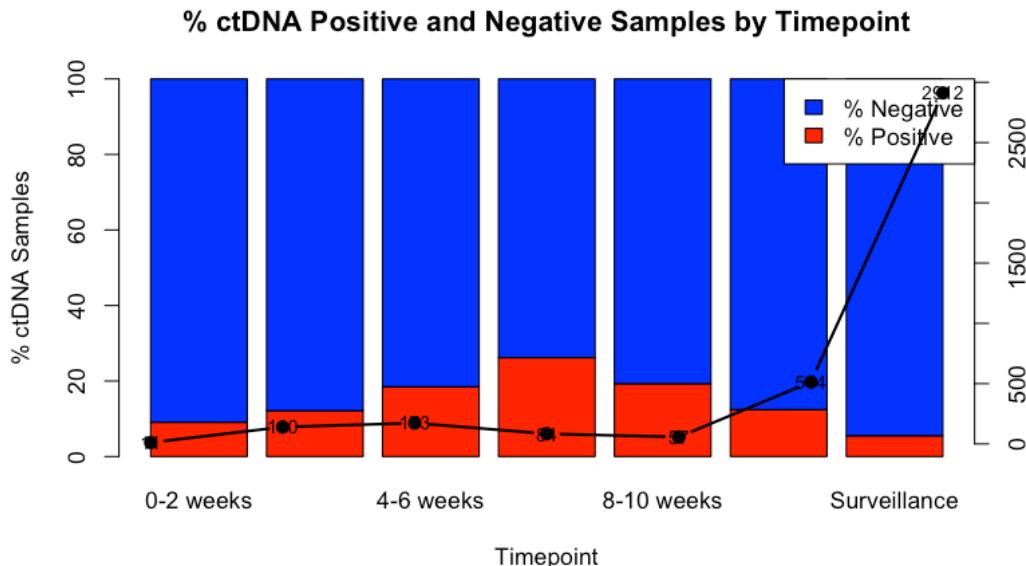
```

Hide

```

mtext("Total Number of Samples", side = 4, line = 3) # Label for the secondary y-axis
text(bar_midpoints, rate_by_timepoint$n_total + 3, labels = rate_by_timepoint$n_total, col = "black", cex = 0.8)

```



Hide

```
#Perform fisher's exact test matrix and heatmap
rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA CRC cfDNA data_ctDNA timepoints.csv")
circ_data <- circ_data[circ_data$cfDNA.Include == TRUE,]
circ_data$Timepoint <- factor(circ_data$Timepoint, levels = c("0-2 weeks", "2-4 weeks", "4-6 weeks", "6-8 weeks",
"8-10 weeks", "On-treatment", "Surveillance"))
circ_data$cfDNAconc <- as.numeric(circ_data$cfDNAconc)
timepoints <- levels(circ_data$Timepoint) # Use ordered levels
p_value_matrix <- matrix(NA, nrow = length(timepoints), ncol = length(timepoints))
rownames(p_value_matrix) <- timepoints
colnames(p_value_matrix) <- timepoints

for (i in 1:length(timepoints)) {
  for (j in i:length(timepoints)) {
    if (i != j) {
      # Subset data for the two Timepoints
      subset1 <- circ_data %>% filter(Timepoint == timepoints[i])
      subset2 <- circ_data %>% filter(Timepoint == timepoints[j])

      # Create contingency table
      contingency_table <- matrix(c(
        sum(subset1$biomarker_status == "POSITIVE"), sum(subset1$biomarker_status == "NEGATIVE"),
        sum(subset2$biomarker_status == "POSITIVE"), sum(subset2$biomarker_status == "NEGATIVE")
      ), nrow = 2, byrow = TRUE)

      # Perform Fisher's exact test
      test_result <- fisher.test(contingency_table)

      # Store p-value in the matrix
      p_value_matrix[i, j] <- test_result$p.value
      p_value_matrix[j, i] <- test_result$p.value # Symmetric matrix
    } else {
      p_value_matrix[i, j] <- 1 # Set diagonal to 1 for clarity
    }
  }
}

print(p_value_matrix)
```

	0-2 weeks	2-4 weeks	4-6 weeks	6-8 weeks	8-10 weeks	On-treatment	Surveillance
0-2 weeks	1.0000000	1.000000000	6.919353e-01	2.857000e-01	0.6738454083	1.000000e+00	4.643687e-01
2-4 weeks	1.0000000	1.000000000	1.58861559	1.000000e+00	1.913544e-01	1.000000000	5.680924e-02
4-6 weeks	0.6919353	0.158861559	1.000000e+00	1.913544e-01	1.000000000	5.680924e-02	8.035880e-09
6-8 weeks	0.2857000	0.010358245	1.913544e-01	1.000000e+00	0.4193440867	2.126873e-03	1.724838e-09
8-10 weeks	0.6738454	0.259489475	1.000000e+00	4.193441e-01	1.000000000	1.498403e-01	3.034774e-04
On-treatment	1.0000000	1.000000000	5.680924e-02	2.126873e-03	0.1498402610	1.000000e+00	6.108761e-08
Surveillance	0.4643687	0.002706764	8.035880e-09	1.724838e-09	0.0003034774	6.108761e-08	1.000000e+00

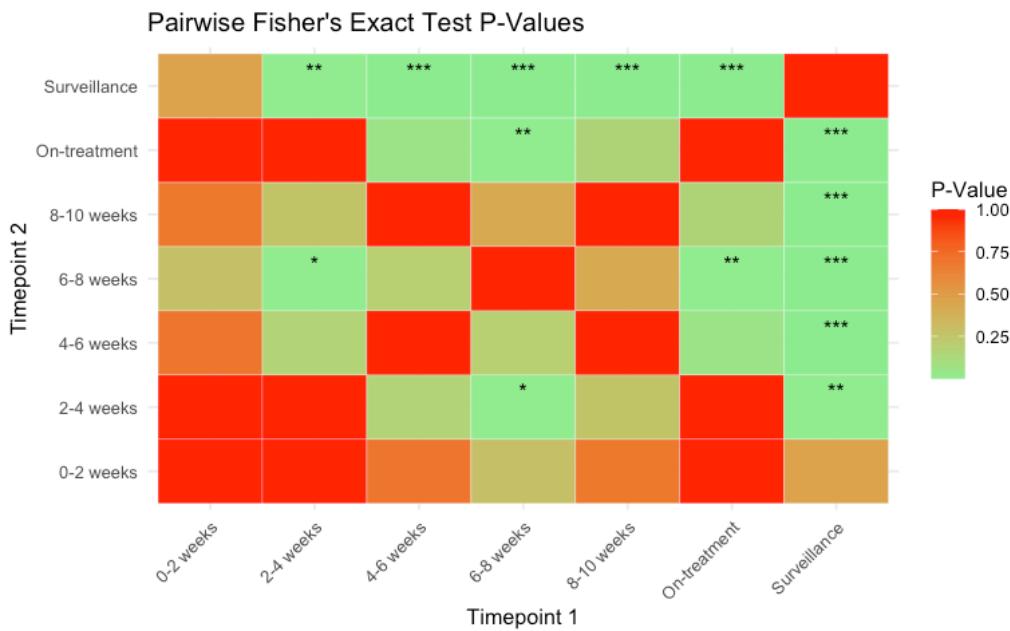
Hide

```

p_value_data <- melt(p_value_matrix)
colnames(p_value_data) <- c("Timepoint1", "Timepoint2", "p_value")
p_value_data <- p_value_data %>%
  mutate(
    asterisk = case_when(
      p_value < 0.001 ~ "***",
      p_value < 0.01 ~ "**",
      p_value < 0.05 ~ "*",
      TRUE ~ ""
    )
  )

# Create the heatmap with asterisks for significance levels
ggplot(p_value_data, aes(x = Timepoint1, y = Timepoint2, fill = p_value)) +
  geom_tile(color = "white") + # Background tiles for grid
  geom_text(aes(label = asterisk), color = "black", vjust = -0.5) + # Add asterisks
  scale_fill_gradient(low = "lightgreen", high = "red") + # Gradient for p-values
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  labs(title = "Pairwise Fisher's Exact Test P-Values",
       x = "Timepoint 1",
       y = "Timepoint 2",
       fill = "P-Value")

```



#Logistic regression for ctDNA positivity

Hide

```

rm(list=ls())
setwd("~/Downloads")
circ_data <- read.csv("CLIA CRC cfDNA data_ctDNA timepoints.csv")
circ_data <- circ_data[circ_data$cfDNA.Include == TRUE,]

# Calculate the 20th, 40th, 60th, 80th, and 100th percentiles for cfDNAconc
thresholds <- quantile(circ_data$cfDNAconc, probs = seq(0, 1, 0.2), na.rm = TRUE)
print(thresholds)

```

0%	20%	40%	60%	80%	100%
1.125000	3.566842	4.764706	6.244382	9.357955	265.185000

Hide

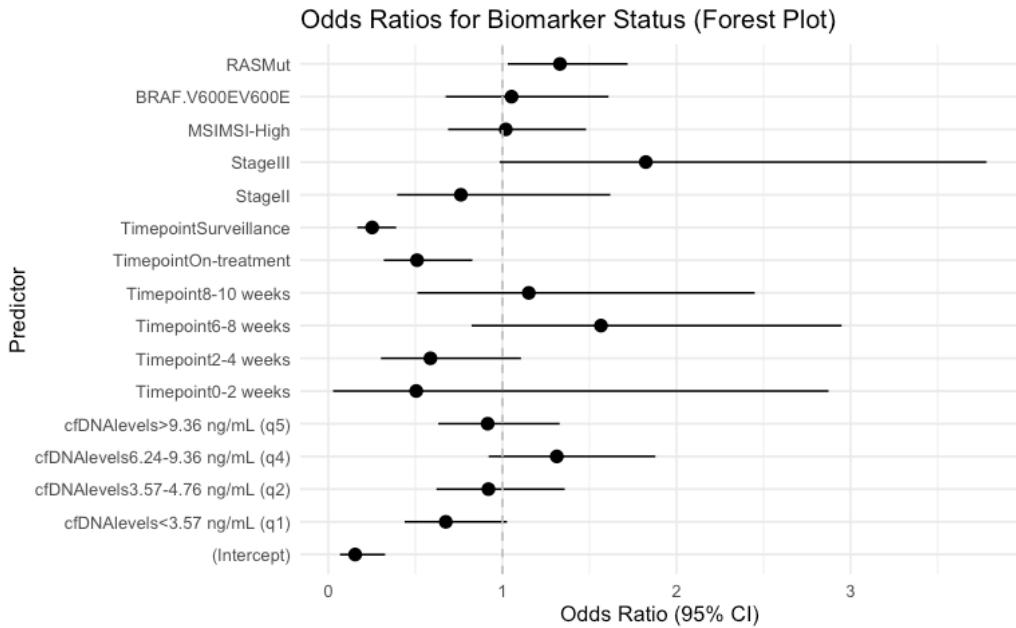
```

circ_data <- circ_data %>%
  mutate(cfDNAlevels = cut(cfDNAconc,
                           breaks = thresholds,
                           include.lowest = TRUE,
                           labels = c("Q1", "Q2", "Q3", "Q4", "Q5")))

circ_data$cfDNAconc <- as.numeric(circ_data$cfDNAconc)
circ_data$biomarker_status <- factor(circ_data$biomarker_status, levels = c("NEGATIVE", "POSITIVE"))
circ_data$cfDNAlevels <- factor(circ_data$cfDNAlevels, levels = c("Q3", "Q1", "Q2", "Q4", "Q5"), labels = c("4.76
-6.24 ng/mL (q3)", "<3.57 ng/mL (q1)", "3.57-4.76 ng/mL (q2)", "6.24-9.36 ng/mL (q4)", ">9.36 ng/mL (q5)"))
circ_data$Timepoint <- factor(circ_data$Timepoint, levels = c("4-6 weeks", "0-2 weeks", "2-4 weeks", "6-8 weeks",
"8-10 weeks", "On-treatment", "Surveillance"))
circ_data$Stage <- factor(circ_data$Stage, levels = c("I", "II", "III"))
circ_data$MSI <- factor(circ_data$MSI, levels = c("MSS", "MSI-High"))
circ_data$BRAF.V600E <- factor(circ_data$BRAF.V600E, levels = c("WT", "MUT"), labels = c("WT", "V600E"))
circ_data$RAS <- factor(circ_data$RAS, levels = c("WT", "MUT"), labels = c("WT", "Mut"))
logistic_model <- glm(biomarker_status ~ cfDNAlevels + Timepoint + Stage + MSI + BRAF.V600E + RAS,
                       data = circ_data,
                       family = binomial)

# Extract odds ratios and confidence intervals
results <- tidy(logistic_model, conf.int = TRUE, exponentiate = TRUE) %>%
  mutate(term = factor(term, levels = term)) # Preserve term ordering
ggplot(results, aes(x = term, y = estimate, ymin = conf.low, ymax = conf.high)) +
  geom_pointrange() +
  geom_hline(yintercept = 1, linetype = "dashed", color = "gray") + # Reference line at OR = 1
  coord_flip() +
  labs(title = "Odds Ratios for Biomarker Status (Forest Plot)",
       x = "Predictor",
       y = "Odds Ratio (95% CI)") +
  theme_minimal()

```



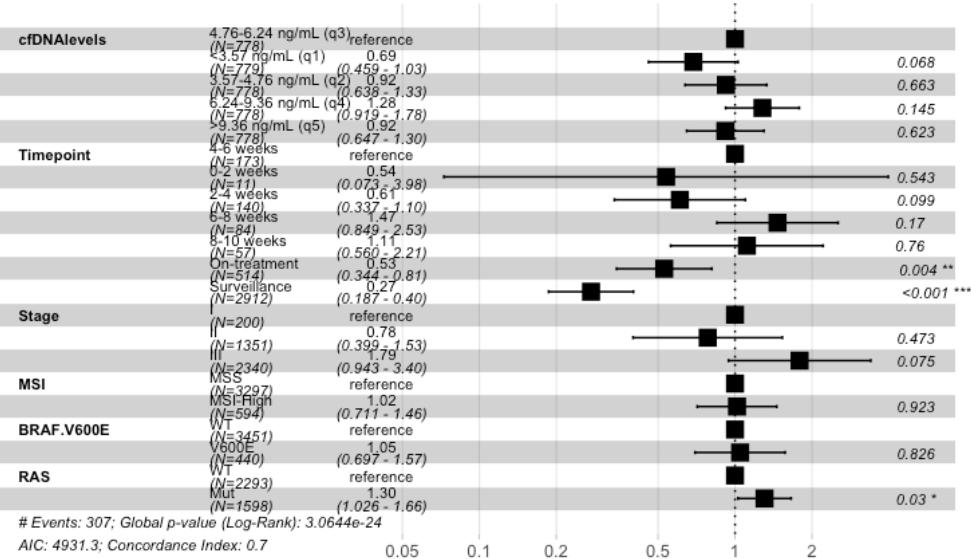
Hide

```

cox_model <- coxph(Surv(rep(1, nrow(circ_data)), biomarker_status == "POSITIVE") ~ cfDNAlevels + Timepoint + Stag
e + MSI + BRAF.V600E + RAS,
                      data = circ_data)
ggforest(cox_model, data = circ_data, main = "Odds Ratios for Biomarker Status", cpositions = c(0.02, 0.22, 0.4))

```

Odds Ratios for Biomarker Status



Hide

```
# Adjust p-values using False Discovery Rate (FDR) adjustment (Benjamini-Hochberg method)
p_values <- summary(cox_model)$coefficients[, 5]
adjusted_p_values <- p.adjust(p_values, method = "fdr")
results <- data.frame(
  Variable = rownames(summary(cox_model)$coefficients),
  Original_P_Value = p_values,
  FDR_Adjusted_P_Value = adjusted_p_values
)
print(results)
```

Variable	Original_P_Value
<chr>	<dbl>
cfDNAlevels<3.57 ng/mL (q1)	6.788475e-02
cfDNAlevels3.57-4.76 ng/mL (q2)	6.627780e-01
cfDNAlevels6.24-9.36 ng/mL (q4)	1.445142e-01
cfDNAlevels>9.36 ng/mL (q5)	6.229291e-01
Timepoint0-2 weeks	5.431864e-01
Timepoint2-4 weeks	9.882970e-02
Timepoint6-8 weeks	1.696129e-01
Timepoint8-10 weeks	7.604623e-01
TimepointOn-treatment	3.612323e-03
TimepointSurveillance	3.008260e-11

1-10 of 15 rows | 1-3 of 3 columns

Previous 1 2 Next