

# Association between poor immune function and cancer risk in adult people with HIV following enrollment

Anuradha Vyas, MD,MPH

## Introduction

**Study Population:** The analytic study population consisted of adults (>18 years old) living with HIV (PWH) who were enrolled in the study.

### Exposure Definition

The primary exposure in this study was poor immune function, defined by:

1. **AIDS-defining illness (ADI):** A history of any condition classified as an AIDS-defining illness, such as certain cancers, opportunistic infections, or other serious HIV-related conditions.
2. **CD4 count less than 200 cells/mmB3:** A CD4 count of less than 200 cells/mmB3 prior to enrollment, reflecting advanced immune suppression.
3. The exposure was further stratified by **viral suppression status** at enrollment:

**Viral suppression:** Defined as having a viral load of less than 200 copies/mL at the time of enrollment.

**No viral suppression:** Defined as having a viral load of 200 copies/mL or higher.

## Packages

```
library(data.table)
library(flowchart)
library(tidyverse)
library(gridExtra)
library(plotrix)
library(gtsummary)
library(gt)
library(survival)
library(survminer)
library(patchwork)
```

# Load data

```
CD4 <- fread("cd4_file.csv")
Patient <-fread("patient_file.csv")
VL<- fread("vl_file.csv")
```

## Data Cleaning

### Immune function

Poor immune function is defined as having an AIDS-defining illness or having a CD4 count less than 200 prior to enrollment. This association likely differs between those who are able to maintain HIV viral suppression and those who cannot. Viral suppression is defined as having <200 copies/mL at enrollment. Therefore, the exposure of immune function must be stratified into four groups.

- Immune\_functionB
  - 0=poor immune function and virally suppressed;
  - 1=poor immune function and not virally suppressed;
  - 2=good immune function and virally suppressed;
  - 3=good immune function and not virally suppressed
- The lowest **CD4 count** recorded prior to enrollment was used to assess immune function. This ensures that we are capturing the most critical period of immune suppression leading up to the time of enrollment.
- The last recorded **viral load** prior to enrollment was used to determine viral suppression status. This value reflects the individual's HIV viral load before the study began, providing a snapshot of their ability to control the virus.

```
Patient <- Patient |>
  mutate(enrollment_date = as.Date(paste(enroll_year,
                                           enroll_month,
                                           enroll_day, sep = "-")))

CD4_filtered <- CD4 |>
  mutate(CD4_date = as.Date(paste(CD4_year, CD4_month, CD4_day, sep = "-")),
         CD4_date_ym =as.Date(sprintf("%d-%02d-01",CD4_year, CD4_month))) |>
  left_join(Patient, by = "ID") |>
  filter((!is.na(CD4_date) & CD4_date < enrollment_date) |
         (!is.na(CD4_date_ym) & CD4_date_ym < enrollment_date) |
         (is.na(CD4_date) & CD4_year < enroll_year)) |>
  group_by(ID) |>
  slice_min(order_by = CD4, with_ties = FALSE) |>
  mutate(CD4_count = case_when(
```

```

  is.na(CD4) ~ NA_character_,
  CD4 < 200 ~ "<200",
  TRUE ~ ">200")) |>
select(ID, CD4, CD4_count)

VL_filtered <- VL |>
mutate(VL_date = as.Date(paste(VL_year, VL_month, VL_day, sep = "-"))) |>
left_join(Patient, by = "ID") |>
filter(VL_date <= enrollment_date) |>
group_by(ID) |>
slice_max(order_by = VL_date, with_ties = FALSE) |>
mutate(viral_load = case_when(
  is.na(vload) ~ NA_character_,
  vload < 200 ~ "suppressed",
  TRUE ~ "unsuppressed")) |>
select(ID, vload, viral_load)

Immune_function_data <- Patient |>
left_join(CD4_filtered, by = "ID") |>
left_join(VL_filtered, by = "ID") |>
mutate(
  Immune_function = case_when(
    is.na(CD4_count) | is.na(viral_load) | is.na(adiflag) ~ NA_integer_,
    (CD4_count == "<200" | adiflag == 1) & viral_load == "suppressed" ~ 0,
    (CD4_count == "<200" | adiflag == 1) & viral_load == "unsuppressed" ~ 1,
    (CD4_count == ">200" & adiflag == 0) & viral_load == "suppressed" ~ 2,
    TRUE ~ 3))

Patient <- Patient |>
left_join(Immune_function_data, by = "ID") |>
mutate(Immune_function = factor(Immune_function,
  levels = c(0, 1, 2, 3),
  labels = c("Poor Immune, Suppressed",
    "Poor Immune, Not Suppressed",
    "Good Immune, Suppressed",
    "Good Immune, Not Suppressed")))

```

## Time- t

- Time from enrollment date to first of cancer diagnosis or loss to followup. Numeric, measured in years (with fractions of a year rounded to the thousandth decimal – e.g., 48 days should be coded 0.131).
- **Handling Missing Cancer Diagnosis Date:** For individuals who did not have a cancer diagnosis by the end of the study, **time (T)** was calculated as the time from **enrollment to the censoring date** (i.e., the last date the individual was followed). This approach filled missing data by assigning the time from enrollment to the censoring date for those who were cancer-free by the study's end.
- **Missing Date Information:** In cases where only the year of date was available, but the month and day were missing, it was assumed that the event (e.g., enrollment or diagnosis) occurred on the

1st day of the year. This approach maximized the use of available data while minimizing the impact of missing dates.

```
Patient <- Patient |>
mutate(
  LTFU_date = ifelse(
    complete.cases(LTFU_year.x, LTFU_month.x, LTFU_day.x),
    sprintf("%d-%02d-%02d", LTFU_year.x, LTFU_month.x, LTFU_day.x),
    NA_character_
  ),
  LTFU_YM = ifelse(
    is.na(LTFU_day.x) & complete.cases(LTFU_year.x, LTFU_month.x),
    sprintf("%d-%02d-01", LTFU_year.x, LTFU_month.x),
    NA_character_
  ),
  cancerdx_date = ifelse(
    complete.cases(cancerdx_year.x, cancerdx_month.x, cancerdx_day.x),
    sprintf("%d-%02d-%02d", cancerdx_year.x, cancerdx_month.x, cancerdx_day.x),
    NA_character_
  ),
  cancerdx_YM = ifelse(
    is.na(cancerdx_day.x) & complete.cases(cancerdx_year.x, cancerdx_month.x),
    sprintf("%d-%02d-01", cancerdx_year.x, cancerdx_month.x),
    NA_character_
  ),
  censoring_date = as.Date("2020-12-31", format = "%Y-%m-%d"),

t =round(
  pmax(0,
    ifelse(
      !is.na(cancerdx_date) | !is.na(LTFU_date) | !is.na(censoring_date),
      as.numeric(difftime(
        pmin(as.Date(cancerdx_date, format = "%Y-%m-%d"),
          as.Date(LTFU_date, format = "%Y-%m-%d"),
          censoring_date, na.rm = TRUE),
        as.Date(enrollment_date.x, format = "%Y-%m-%d"),
        units = "days"
      )) / 365.25,
      # If year and month are available but not day, calculate based on that
      ifelse(
        !is.na(cancerdx_YM) | !is.na(LTFU_YM) | !is.na(censoring_date),
        as.numeric(difftime(
          pmin(as.Date(cancerdx_YM, format = "%Y-%m-%d"),
            as.Date(LTFU_YM, format = "%Y-%m-%d"),
            censoring_date, na.rm = TRUE),
          as.Date(enrollment_date.x, format = "%Y-%m-%d"), units = "days"
        )) / 365.25,
        # If only year is available, calculate time based on year
```

```

    ifelse(
      !is.na(cancerdx_year.x) | !is.na(LTFU_year.x) | !is.na(censoring_date),
      as.numeric(difftime(
        pmin(as.Date(paste(cancerdx_year.x, "-12-31", sep = ""),
                      format = "%Y-%m-%d"),
              as.Date(paste(LTFU_year.x, "-12-31", sep = ""),
                      format = "%Y-%m-%d"),
              censoring_date, na.rm = TRUE),
        as.Date(paste(enroll_year.x, "-12-31", sep = ""),
                  format = "%Y-%m-%d"), units = "days"
      )) / 365.25,
      # If no dates available, use censoring date
      ifelse(
        !is.na(enrollment_date.x) | !is.na(censoring_date),
        as.numeric(difftime(censoring_date,
                              as.Date(enrollment_date.x,
                                      format = "%Y-%m-%d"),
                              units = "days")) / 365.25))))),3))

```

## Cancer

Indicator for cancer diagnosisB

1=patient diagnosed with cancer during the study period (prior to loss to follow up);

0=patient was censored before cancer diagnosis.

- **Cancer Diagnosis Prior to Enrollment:** If a cancer diagnosis was recorded prior to enrollment, the individual was labeled as having cancer at the time of enrollment (coded as 0). This indicates that the patient had an established diagnosis before entering the study, allowing for accurate stratification of cancer status at baseline.

```

Patient <- Patient |>
mutate(
  Cancer_D = ifelse(is.na(cancerdx_year.x), 0,
    ifelse(
      (is.na(cancerdx_date)), 0,
      ifelse(
        cancerdx_date < enrollment_date.x, 0,
        ifelse(
          (!is.na(LTFU_date) & !is.na(cancerdx_date) & LTFU_date < cancerdx_date), 0,
          ifelse((enrollment_date.x < cancerdx_date & (is.na(LTFU_date) |
                                                    cancerdx_date < LTFU_date)) |
                (enrollment_date.x < cancerdx_YM & (is.na(LTFU_YM) |
                                                    cancerdx_YM < LTFU_YM)) |
                (enroll_year.x < cancerdx_year.x & (is.na(LTFU_year.x) |
                                                    cancerdx_year.x < LTFU_year.x))),
            1,0))))))

```

# Final Data

## Inclusion criteria:

- Adults aged 18 years or older
- Individuals who had complete data on Immune function.

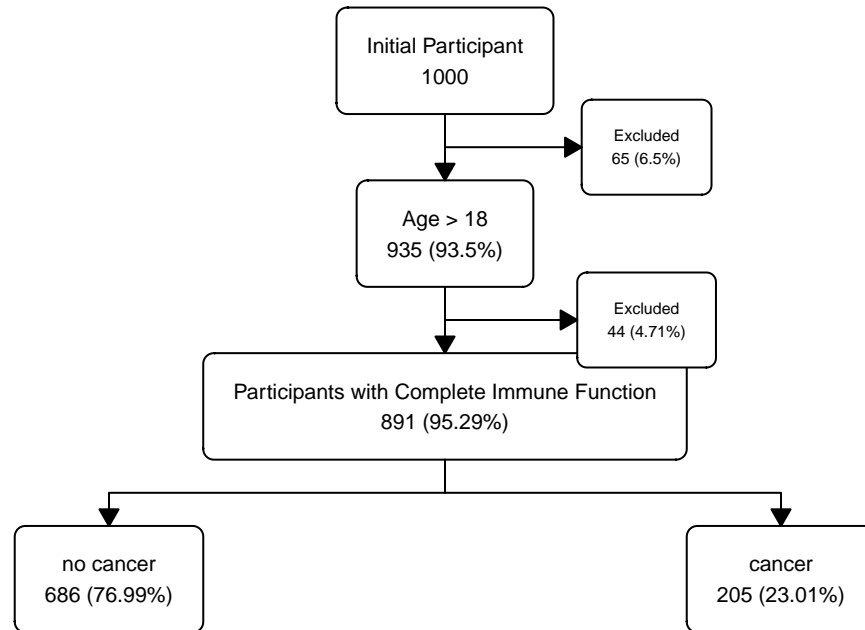
```
analytic_file <- Patient |>
  filter(age.x > 18,
         !is.na(Immune_function)) |>
  select(ID, Immune_function, t, Cancer_D, age.x) |>
  rename(Age = age.x,
         Cancer=Cancer_D)

write.csv(analytic_file, "analytic_file.csv", row.names = FALSE)
```

## Flowchart: Participant Selection

```
Patient |>
  as_fc(label = "Initial Participant") |>
  fc_filter(age.x > 18, label = "Age > 18", show_exc = TRUE) |>
  fc_filter(!is.na(Immune_function), label = "Participants with Complete Immune Function", show_exc = TRUE) |>
  fc_split(Cancer_D, label = c("0" = "no cancer", "1" = "cancer")) |>
  fc_draw(title = "Flow chart for Participant selection")
```

## Flow chart for Participant selection

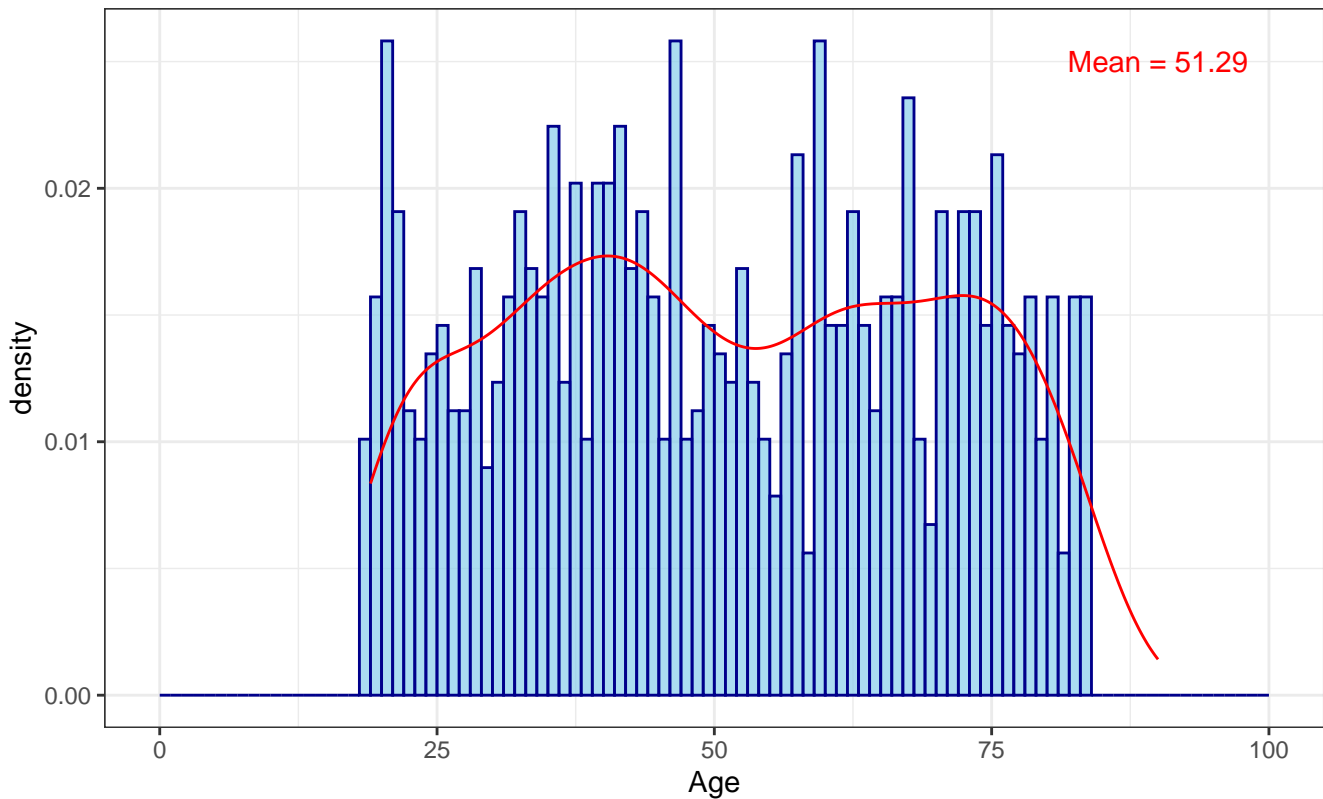


## Descriptive Statistics

### Age

```
ggplot(analytic_file, aes(Age)) +  
  geom_histogram(aes(y = after_stat(density)),  
    breaks = seq(0, 100, 1),  
    color = "darkblue",  
    fill = "skyblue",  
    alpha = 0.7,  
    binwidth = 4,) +  
  labs(title = "Fig 1. Distribution of Age ") +  
  geom_density(color = "red") +  
  theme_bw() +  
  theme(plot.title = element_text(size = 12, face = "bold", hjust = 0.5)) +  
  annotate("text", x = 90, y = 0.025,  
    label = paste("Mean =", round(mean(analytic_file$Age, na.rm = TRUE), 2)),  
    color = "red")
```

**Fig 1. Distribution of Age**

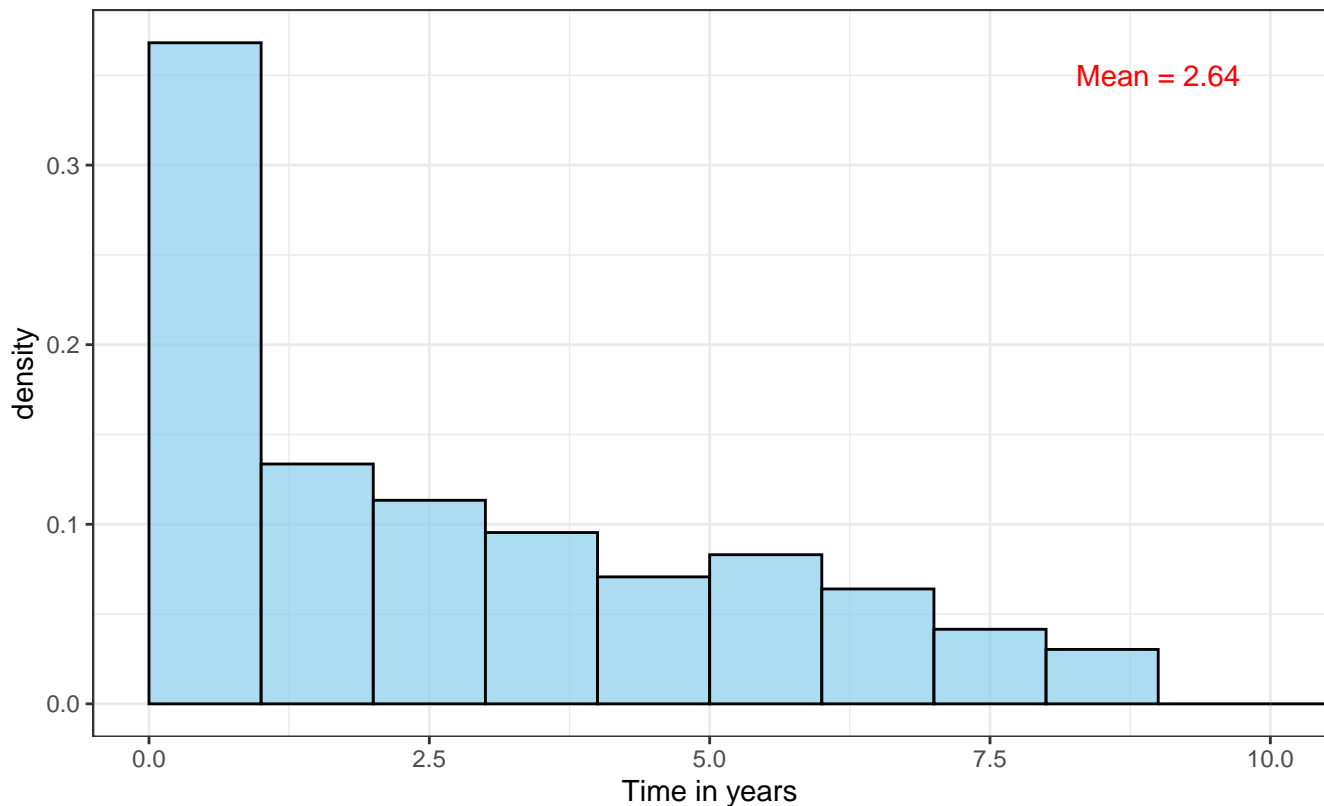


## Time

```
ggplot(analytic_file, aes(t)) +  
  geom_histogram(aes(y = after_stat(density)),  
    breaks = seq(0, 100, 1),  
    color = "black",  
    fill = "skyblue",  
    alpha = 0.7,  
    binwidth = 4) +  
  labs(title = "Fig 2. Time from Enrollment to Cancer Diagnosis or Censoring",  
    x = "Time in years") +  
  theme_bw() +  
  theme(plot.title = element_text(size = 9, face = "bold", hjust = 0.5))+  
  coord_cartesian(xlim = c(0, 10)) +  
  annotate("text",  
    x = 9, y = 0.35,  
    label = paste("Mean =",  
      round(mean(analytic_file$t, na.rm = TRUE), 2)),  
    color = "red")
```



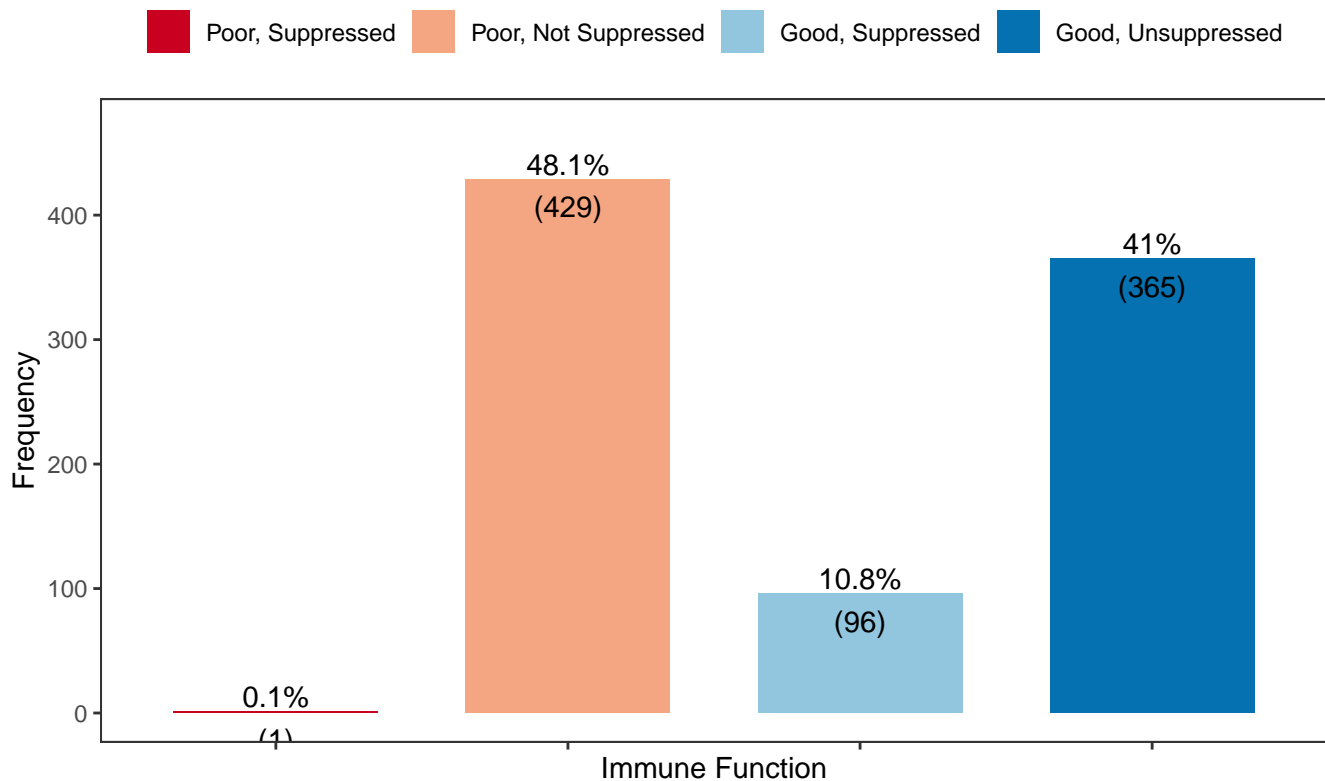
Fig 2. Time from Enrollment to Cancer Diagnosis or Censoring



## Immune function

```
ggplot(analytic_file, aes(x = Immune_function, fill = Immune_function)) +
  geom_bar(width = 0.7, show.legend = TRUE) +
  geom_text(stat = "count",
            aes(label = paste0(round(prop.table(after_stat(count)) * 100,1),
                                "%\n(", ..count.., ")"), vjust = 0.6)) +
  labs(title = "Fig 3. Distribution for Immune Function",
        y = "Frequency", x = "Immune Function") +
  scale_fill_brewer(palette = "RdBu",
                    labels = c("Poor, Suppressed",
                                "Poor, Not Suppressed",
                                "Good, Suppressed",
                                "Good, Unsuppressed"),
                    name = NULL) +
  theme_bw() +
  theme(panel.grid = element_blank(), axis.text.x = element_blank(),
        legend.position = "top",
        plot.title = element_text(size = 12, face = "bold", hjust = 0.5)) +
  coord_cartesian(ylim = c(0, 470))
```

**Fig 3. Distribution for Immune Function**



## Cancer

```
f <- table(analytic_file$Cancer)
colors <- c("darkblue", "lightblue")
labels <- paste(f, "(", round(prop.table(f) * 100, 1), "%)", sep = "")

par(mar = c(1,1,0.3,1))
pie3D(f, labels = labels, main = "Fig 4. Cancer Distribution",
      col = colors,
      explode = 0.1,
      theta = 0.7,
      start = pi / 2)

mtext("Fig 4. Cancer Distribution",
      side = 3,
      line = 1,
      cex = 1.0,
      font = 1)
abline(h = par("usr")[4] + 0.1, col = "black", lwd = 1)

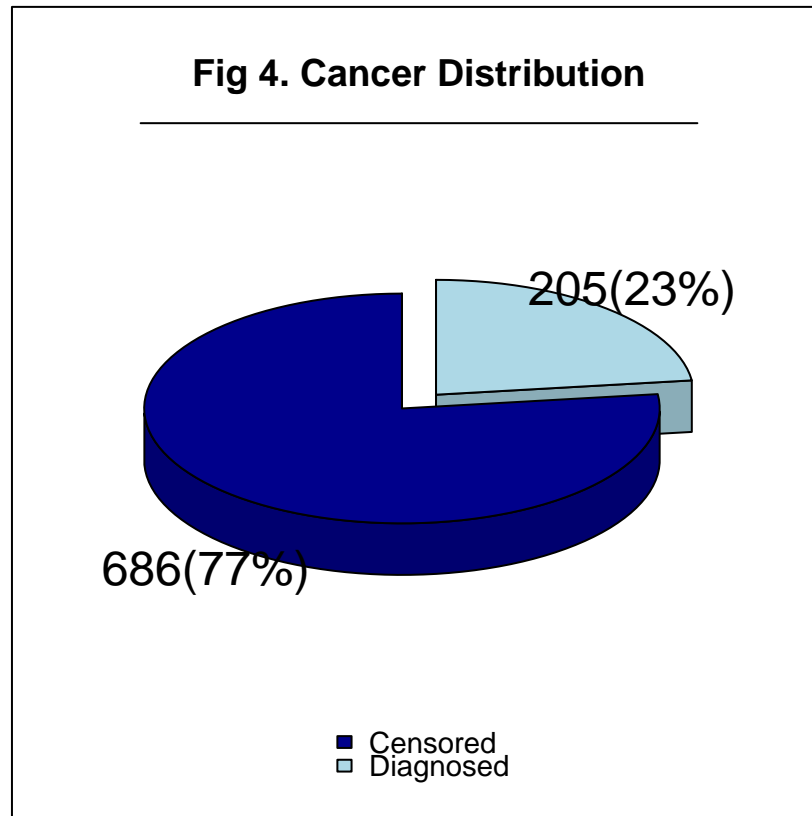
legend("bottom",
      legend = c("Censored", "Diagnosed"),
      fill = colors,
      bty = "n",
```

```

cex = 0.9,
inset = c(0, -0.15))

box(lwd = 1)

```



## Analysis

### Bivariate analysis

```

analytic_file |>
  select(-ID) |>
  tbl_summary(by = Immune_function) |>
  add_n() |>
  add_p() |>
  as_gt() |>
  tab_header(title = md("Participant Characteristics by Immunity")) |>
  cols_width(
    everything() ~ px(100))

```

### Plotting Function

## Participant Characteristics by Immunity

Characteristic	N	Poor Immune, Sup- pressed N = 1 <sup>1</sup>	Poor Immune, Not Sup- pressed N = 429 <sup>1</sup>	Good Immune, Sup- pressed N = 96 <sup>1</sup>	Good Immune, Not Sup- pressed N = 365 <sup>1</sup>	p-value <sup>2</sup>
t	891	0.00 (0.00, 0.00)	1.13 (0.00, 3.27)	2.96 (0.98, 5.27)	2.98 (0.91, 5.56)	<0.001
Cancer	891	0 (0%)	155 (36%)	4 (4.2%)	46 (13%)	<0.001
Age	891	33 (33, 33)	52 (37, 69)	48 (34, 67)	48 (35, 68)	0.2

<sup>1</sup>Median (Q1, Q3); n (%)

<sup>2</sup>Kruskal-Wallis rank sum test; Fisher's exact test

```
km_plot <- function(N, km_fit, plot_type, plot_title) {

  # Determine if legend should be shown based on plot_type
  show_legend <- ifelse(plot_type == "pct", TRUE, FALSE)

  # Create the base plot
  plot <- ggsurvplot(km_fit,
    fun = plot_type,
    palette = c("#483D8B", "#DC143C", "#6B8E23", "#4682B4"),
    xlim = c(0, 7),
    title = paste("Fig", N, ".", plot_title, "curve"),
    legend.title = " ",
    legend.labs = c("Poor, Suppressed",
                    "Poor, Not Suppressed",
                    "Good, Suppressed",
                    "Good, Not Suppressed"),
    conf.int = TRUE,
    pval = TRUE,
    risk.table = TRUE,
    tables.y.text = FALSE,
    tables.height = 0.28,
    tables.col = "strata",
    tables.theme = theme_cleantable(),
    ggtheme = theme_minimal() +
      theme(plot.title = element_text(size = 9,
                                       face = "bold",
                                       hjust = 0.5,
                                       vjust = -2),
            panel.border = element_rect(color = "black",
                                       fill = NA,
                                       size = 1),
            legend.text = element_text(size = 6)),
    # Show legend only for "pct" plot type (Survival Curve)
```

```

        legend.position = ifelse(show_legend, "right", "none"))
    )

    # If the plot_type is not "pct", remove the legend from the main plot
    if (plot_type != "pct") {
      plot$plot <- plot$plot + theme(legend.position = "none")
    }

    # Explicitly remove the legend from the risk table if necessary
    plot$table <- plot$table + theme(legend.position = "none")

    # Return the plot with conditional legend visibility
    return(plot)
  }

```

## Kaplan-Meier survival curve

```

km_fit <- survfit(Surv(t, Cancer) ~ Immune_function,
                  data = analytic_file)

Survival_curve <- km_plot(5, km_fit, "pct",
                          "Kaplan-Meier Time to Cancer Diagnosis for 7 year risk of cancer")
Risk_curve <- km_plot(6, km_fit, "event", "Cumulative Risk")
Hazard_curve <- km_plot(7, km_fit, "cumhaz", "Cumulative Hazard")

pdf("Kaplan_Meier_Combined_Curves.pdf", width = 8, height = 4)
grid.arrange(Survival_curve$plot,
              Survival_curve$table,
              ncol = 1,                # One column for Survival plot and table
              heights = c(1, 0.35),    # Adjust height of Survival plot and table
              widths = c(6))           # Adjust width of Survival plot and table

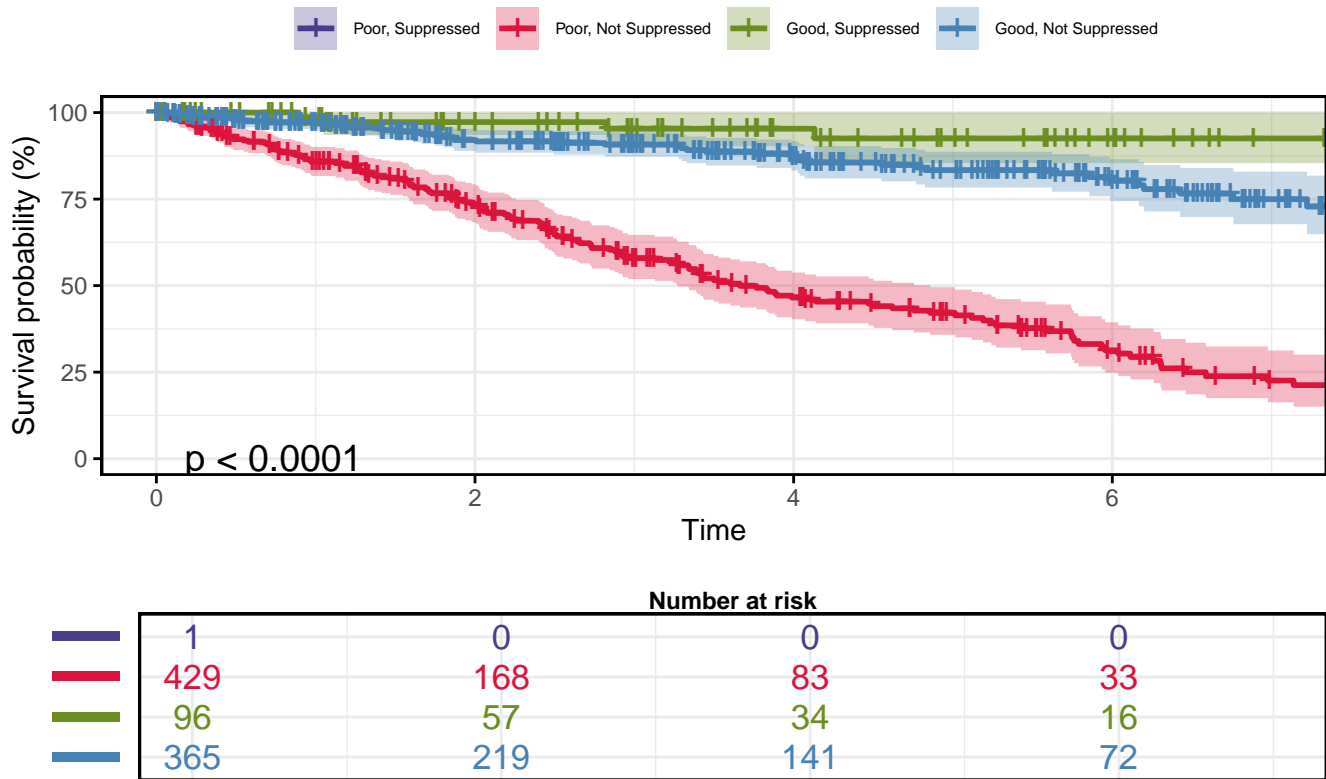
grid.arrange(Risk_curve$plot,
              Hazard_curve$plot,
              ncol = 2,                # Two columns for Risk and Hazard curves
              heights = c(3),          # Adjust height of both plots
              widths = c(6, 6))        # Adjust width of both plots
dev.off()

```

pdf  
2

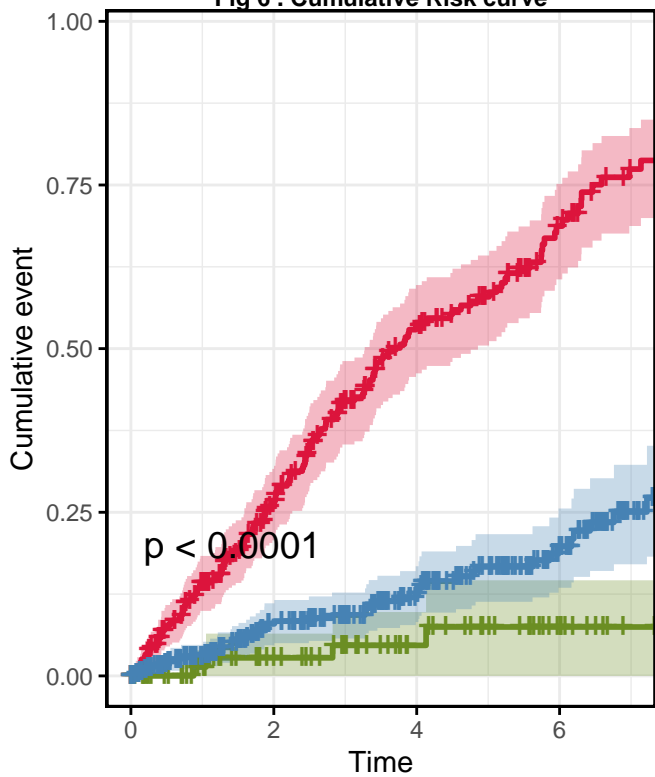
Survival\_curve

**Fig 5 . Kaplan–Meier Time to Cancer Diagnosis for 7 year risk of cancer curve**

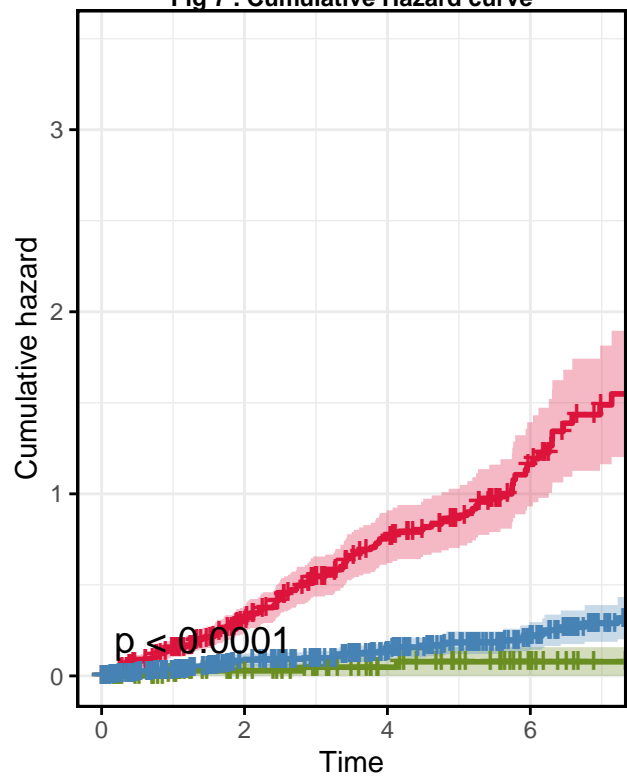


`Risk_curve$plot` | `Hazard_curve$plot`

**Fig 6 . Cumulative Risk curve**



**Fig 7 . Cumulative Hazard curve**



## Cancer outcome at 0, 1 and 7 years

Characteristic	0 Year	1 Year	7 Year
Immune_function			
Poor Immune, Suppressed	100% (100%, 100%)	— (—, —)	— (—, —)
Poor Immune, Not Suppressed	100% (100%, 100%)	86% (82%, 90%)	23% (16%, 31%)
Good Immune, Suppressed	100% (100%, 100%)	99% (96%, 100%)	93% (85%, 100%)
Good Immune, Not Suppressed	100% (100%, 100%)	97% (95%, 99%)	75% (68%, 83%)

The Kaplan-Meier estimates revealed the following trends:

- **Individuals with poor immune function and no viral suppression** had the highest cumulative risk of developing cancer, with a steep increase in the probability of cancer diagnosis over time.
- **Individuals with poor immune function but viral suppression** showed a moderate risk, with a slower rise in the probability of cancer compared to the first group.
- **Individuals with good immune function and no viral suppression** exhibited the lowest cumulative risk, with a gradual increase in cancer risk over time.
- **Individuals with poor immune function with viral suppression** had no representation since only 1 subject was there.

```
# Create the KM summary at specific times
tbl_survfit(km_fit,
            times = c(0,1,7),
            label_header = "**{time} Year**")|>
as_gt()|>
tab_header(title = md("Cancer outcome at 0, 1 and 7 years"))
```

The Kaplan-Meier survival analysis reveals notable differences in cancer survival across the immune function groups. Individuals with **poor immune function and not suppressed viral load** began with a survival rate of **64.1%** but experienced a dramatic decline over seven years, with survival dropping to just **3.96%** by year 7, indicating a significantly high risk of cancer over time. In contrast, individuals with **good immune function and suppressed viral load** started with a survival rate of **90.62%**, but their survival also decreased to **7.29%** by year 7, suggesting an increased risk despite initial better survival. Similarly, those with **good immune function and not suppressed viral load** had a survival rate of **90.4%** at baseline, which decreased to **11.0%** after seven years, reflecting a substantial rise in cancer risk over time. Overall, these results emphasize the strong association between immune function and cancer survival, with those having poor immunity and unsuppressed viral load exhibiting the highest risk of cancer progression.

```
# Fit the survival model using survdiff
surv_fit <- survdiff(Surv(t, Cancer) ~ Immune_function, data = analytic_file)

data.frame(
  N = surv_fit$n,
  Observed = surv_fit$obs,
  Expected = round(surv_fit$exp)
)|> gt()|> tab_header(title = "Cancer outcome: Observed vs. Expected")
```

## Cancer outcome: Observed vs. Expected

N.groups	N.Freq	Observed	Expected
Immune_function=Poor Immune, Suppressed	1	0	0
Immune_function=Poor Immune, Not Suppressed	429	155	72
Immune_function=Good Immune, Suppressed	96	4	26
Immune_function=Good Immune, Not Suppressed	365	46	107

The log-rank test results indicate significant differences in cancer survival distributions across the immune function groups. For the group with **poor immune function and not suppressed viral load**, there were 429 individuals, and 155 cancer events were observed, much higher than the expected 72.2 events, highlighting an increased risk of cancer in this group. In contrast, the **good immune function and suppressed viral load** group, consisting of 96 individuals, had only 4 observed cancer events, which is significantly lower than the expected 26.2, suggesting a protective effect of suppressed viral load in individuals with good immune function. The **good immune function and not suppressed viral load** group showed fewer cancer events than expected, with 46 observed out of an expected 106.6. The **chi-square statistic** of **149** with **2 degrees of freedom** and a **p-value**  $< 2e-16$  indicates a highly significant difference in survival distributions, confirming that immune function plays a critical role in cancer risk among individuals living with HIV.

## Cox proportional hazards model

```
Unadjusted <- tbl_uvregression(
  analytic_file,
  method = coxph,
  y = Surv(t, Cancer),
  exponentiate = TRUE,
  hide_n = TRUE,
  include = c("Immune_function"),
  pvalue_fun = label_style_pvalue(digits = 2)
)

Adjusted <- tbl_uvregression(
  analytic_file,
  method = coxph,
  y = Surv(t, Cancer),
  exponentiate = TRUE,
  hide_n = TRUE,
  include = c("Immune_function", "Age"),
  pvalue_fun = label_style_pvalue(digits = 2)
)

tbl_merge(
  tbls = list( Unadjusted, Adjusted),
  tab_spanner = c( "**Unadjusted**", "**Adjusted**" )
)|>
```



## Time to Cancer

Characteristic	Unadjusted			Adjusted		
	HR <sup>1</sup>	95% CI <sup>1</sup>	p-value	HR <sup>1</sup>	95% CI <sup>1</sup>	p-value
Immune_function						
Poor Immune, Suppressed	—	—		—	—	
Poor Immune, Not Suppressed	5.06	3.64, 7.05	<0.001	5.06	3.64, 7.05	<0.001
Good Immune, Suppressed	0.36	0.13, 0.99	0.048	0.36	0.13, 0.99	0.048
Good Immune, Not Suppressed						
Age				1.00	1.00, 1.01	0.44

<sup>1</sup>HR = Hazard Ratio, CI = Confidence Interval

```
as_gt()|>
tab_header(title = "Time to Cancer")
```

The Cox Proportional Hazards model shows that individuals with **poor immune function and not suppressed viral load** have a significantly higher cancer risk, with a hazard ratio of **5.06**, meaning they are over five times more likely to develop cancer than the reference group. Conversely, those with **good immune function and suppressed viral load** have a **lower risk** of cancer, with a hazard ratio of **0.36**, indicating they are about 64% less likely to develop cancer. These results highlight the impact of immune function on cancer risk in individuals living with HIV.

## Results

At the onset of the study, 891 adults (b %18 years) living with HIV (PWH) were enrolled to examine the association between immune function and cancer risk. Participants were categorized based on their immune status into two groups: those with poor immune function (e.g., CD4 count <200 cells/mmB3 or other indicators of immune compromise) and those with good immune function (e.g., CD4 count b %500 cells/mmB3). Participants were followed from enrollment until the occurrence of one of the following events: 1) cancer diagnosis, 2) loss to follow-up, or 3) administrative censoring, which occurred on December 31, 2020.

At baseline, individuals with poor immune function and suppressed viral load (1 out of 1) had cancer with a 0% survival rate. In contrast, individuals with poor immune function and un-suppressed viral load (154 out of 429) had cancer with a 3% survival rate. Among individuals with good immune function, those with suppressed viral load (9 out of 96) also had cancer, while those with un-suppressed viral load (35 out of 365) demonstrated cancer incidence as well.

At the 7-year follow-up, cancer events were recorded across the groups. Among those with poor immune function and un-suppressed viral load, 250 cancer events occurred out of 17 individuals. Among those with good immune function and suppressed viral load, 80 cancer events were recorded out of 7 individuals, while 290 cancer events were observed in individuals with good immune function and un-suppressed viral load, out of 40 individuals. These findings highlight a strong association between poor immune function and increased cancer risk, with a clear difference in cancer incidence between those with poor versus good immune function. This emphasizes the critical role immune status plays in cancer progression among adults living with HIV.

The Cox Proportional Hazards model further supports these findings, showing that individuals with poor immune function and un-suppressed viral load face a significantly higher cancer risk, with a hazard ratio of 5.06, indicating they are more than five times as likely to develop cancer compared to the reference group. On the other hand, individuals with good immune function and suppressed viral load have a lower cancer risk, with a hazard ratio of 0.36, suggesting they are about 64% less likely to develop cancer. These results underscore the importance of immune function in determining cancer risk in this population.

When adjusted for age, the Cox Proportional Hazards model reveals that individuals with poor immune function and un-suppressed viral load continue to have a substantially higher cancer risk (hazard ratio 5.07,  $p\text{-value} < 2e-16$ ). Individuals with good immune function and suppressed viral load remain at lower risk (hazard ratio 0.36,  $p\text{-value} = 0.0477$ ). However, age did not significantly influence cancer risk in this model (hazard ratio 0.9997,  $p\text{-value} = 0.9313$ ), indicating no meaningful association between age and cancer risk once immune function was accounted for.

In conclusion, immune function is a significant predictor of cancer risk in individuals with HIV, while age does not appear to have a notable impact in this analysis.