Analyse Biostatistique Démo

Jadot Cédric

2026-02-17

Baseline Characteristics

```
table1 <- clinical_data %>%
  select(age, sex, biomarker_baseline, treatment) %>%
  tbl_summary(
    by = treatment,
    statistic = list(
      all_continuous() ~ "{mean} ({sd})",
      all_categorical() ~ "{n} ({p}%)"
    ),
    digits = all_continuous() ~ 2,
    missing = "no"
  ) %>%
  add_p() %>%
  add_overall() %>%
  modify_header(label = "**Variable**") %>%
  modify_caption("**Table 1. Baseline characteristics by treatment group**")

table1 %>% as_gt()
```
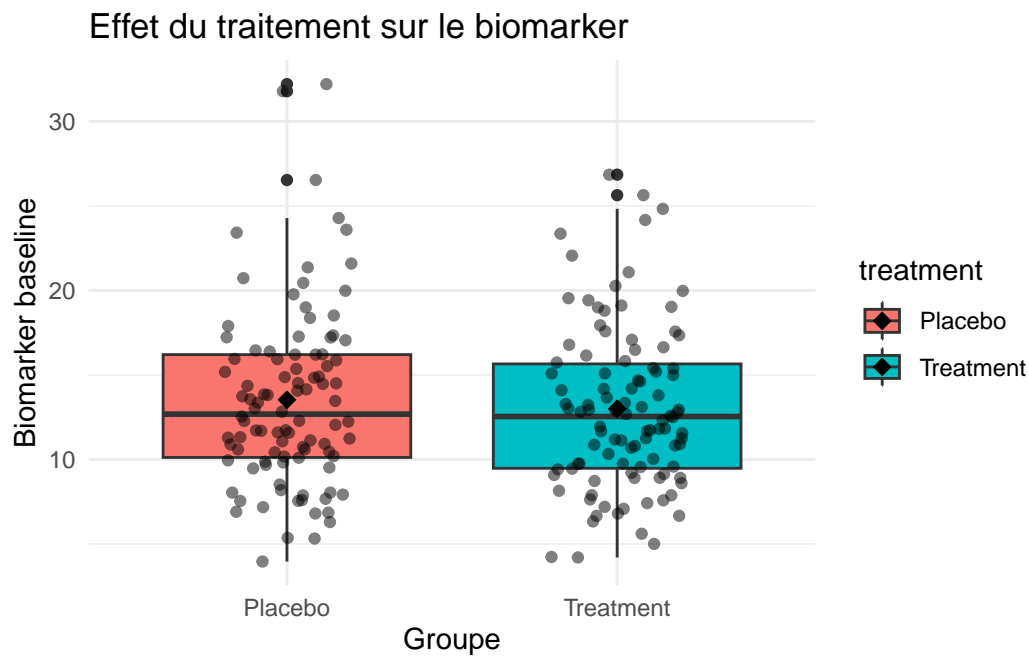
Analyse du biomarker

```
ggplot(clinical_data, aes(x = treatment, y = biomarker_baseline, fill = treatment)) +
  geom_boxplot() +
  geom_jitter(width = 0.2, alpha = 0.5) +
  stat_summary(fun = mean, geom = "point", shape = 18, size = 3, color = "black") +
  theme_minimal() +
```

Variable	Overall N = 200[1]	Placebo N = 98[1]	Treatment N = 102[1]	p-value[2]
age	64.92 (9.44)	64.03 (9.30)	65.76 (9.55)	0.4
sex				0.5
Male	93 (47%)	43 (44%)	50 (49%)	
Female	107 (54%)	55 (56%)	52 (51%)	
biomarker_baseline	13.26 (5.05)	13.53 (5.29)	12.99 (4.83)	0.5

[1] Mean (SD); n (%)

[2] Wilcoxon rank sum test; Pearson's Chi-squared test

```r
labs(title = "Effet du traitement sur le biomarker",
     x = "Groupe",
     y = "Biomarker baseline")
```


Effet du traitement sur le biomarker

```r
lm_model <- lm(biomarker_baseline ~ treatment + age + sex, data = clinical_data)
summary(lm_model)
```

```
Call:
lm(formula = biomarker_baseline ~ treatment + age + sex, data = clinical_data)
```

```
Residuals:
    Min      1Q  Median      3Q     Max
-8.9879 -3.3506 -0.6639  2.4768 18.5158

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)        13.76022    2.50108   5.502 1.17e-07 ***
treatmentTreatment -0.46324    0.72042  -0.643    0.521
age                -0.01213    0.03836  -0.316    0.752
sexFemale           0.97147    0.72238   1.345    0.180
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.062 on 196 degrees of freedom
Multiple R-squared:  0.01208,   Adjusted R-squared:  -0.003037
F-statistic: 0.7992 on 3 and 196 DF,  p-value: 0.4957
```

```r
broom::tidy(lm_model, conf.int = TRUE)
```

```
# A tibble: 4 x 7
  term               estimate std.error statistic    p.value conf.low conf.high
  <chr>                 <dbl>     <dbl>     <dbl>      <dbl>    <dbl>     <dbl>
1 (Intercept)          13.8       2.50       5.50  0.000000117    8.83      18.7
2 treatmentTreatment   -0.463     0.720     -0.643 0.521         -1.88       0.958
3 age                  -0.0121    0.0384    -0.316 0.752         -0.0878     0.0635
4 sexFemale             0.971     0.722      1.34  0.180         -0.453      2.40
```

Analyse de survie

```r
surv_obj <- Surv(time = clinical_data$survival_time,
                 event = clinical_data$status)

km_fit <- survfit(surv_obj ~ treatment, data = clinical_data)

ggsurvplot(
  km_fit,
  data = clinical_data,
  risk.table = TRUE,
  pval = TRUE,
```
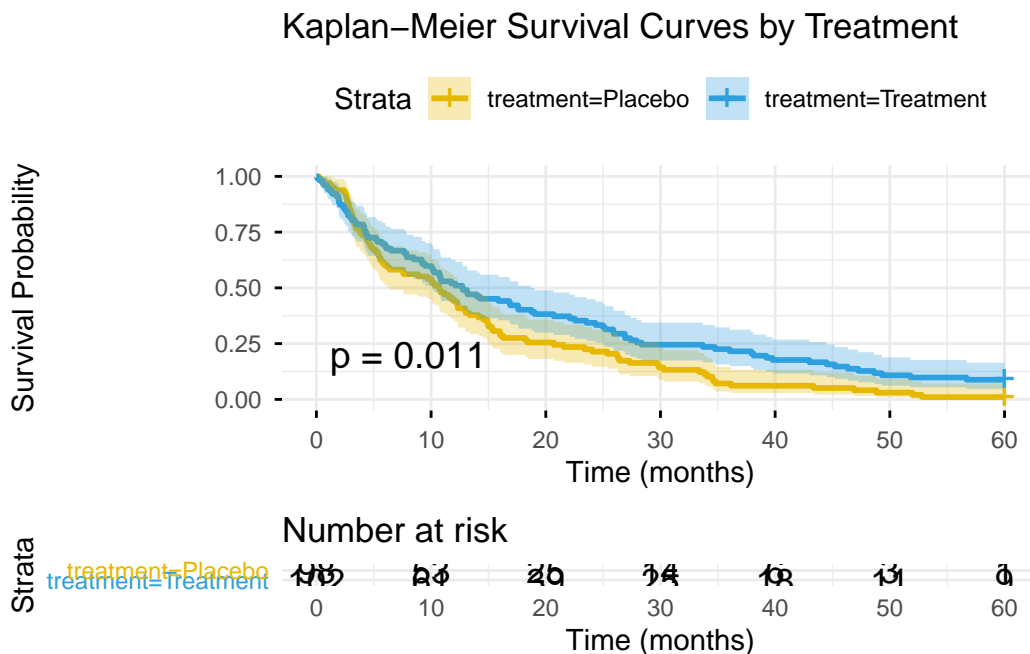
```
  conf.int = TRUE,
  palette = c("#E7B800", "#2E9FDF"),
  ggtheme = theme_minimal(),
  title = "Kaplan-Meier Survival Curves by Treatment",
  xlab = "Time (months)",
  ylab = "Survival Probability"
)
```

```
Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
i Please use `linewidth` instead.
i The deprecated feature was likely used in the ggpubr package.
  Please report the issue at <https://github.com/kassambara/ggpubr/issues>.
```

```
Ignoring unknown labels:
* colour : "Strata"
```


Kaplan–Meier Survival Curves by Treatment

```
logrank_test <- survdiff(surv_obj ~ treatment, data = clinical_data)
logrank_test
```

```
Call:
survdiff(formula = surv_obj ~ treatment, data = clinical_data)
```

```
                   N Observed Expected (O-E)^2/E (O-E)^2/V
treatment=Placebo   98       97       80      3.61       6.43
treatment=Treatment 102      93      110      2.63       6.43

 Chisq= 6.4  on 1 degrees of freedom, p= 0.01
```

Modèle de Cox ajusté

```
cox_model <- coxph(Surv(survival_time, status) ~ treatment + age + sex + biomarker_baseline,
                   data = clinical_data)

broom::tidy(cox_model, exponentiate = TRUE, conf.int = TRUE)
```

```
# A tibble: 4 x 7
  term              estimate std.error statistic    p.value conf.low conf.high
  <chr>                <dbl>     <dbl>     <dbl>      <dbl>    <dbl>     <dbl>
1 treatmentTreatment   0.639    0.150     -2.98   0.00289     0.476     0.858
2 age                  1.04     0.00825    4.70   0.00000255  1.02      1.06
3 sexFemale            1.09     0.149      0.588  0.557       0.815     1.46
4 biomarker_baseline   0.999    0.0142    -0.0971 0.923       0.971     1.03
```

```
cox_table <- tidy(cox_model, exponentiate = TRUE, conf.int = TRUE) %>%
  mutate(term = str_replace_all(term, "`", "")) %>%
  arrange(desc(estimate))

ggplot(cox_table, aes(x = reorder(term, estimate), y = estimate,
                      ymin = conf.low, ymax = conf.high)) +
  geom_pointrange(color = "#2E9FDF", size = 0.8) +
  geom_hline(yintercept = 1, linetype = "dashed") +
  coord_flip() +
  theme_minimal(base_size = 12) +
  labs(
    x = "",
    y = "Hazard Ratio (95% CI)",
    title = "Forest plot - Cox proportional hazards model"
  ) +
  theme(axis.text.y = element_text(size = 10))
```

Forest plot – Cox proportional hazards model