

# Introducción al análisis de la supervivencia con



**Sara Tous i Belmonte**

*Estadística*

Unitat d'Infeccions i Càncer - Molecular, Programa d'Epidemiologia en Recerca del Càncer

Tel. +34 932607812 - ext 3185

[stous@iconcologia.net](mailto:stous@iconcologia.net)

IDIBELL  
Institut d'Investigació Biomèdica de Bellvitge

Salut/

ICO  
Institut Català d'Oncologia

OBJECTIUS GLOBAIS  
DE DESARROLLO SOSTENIBLE

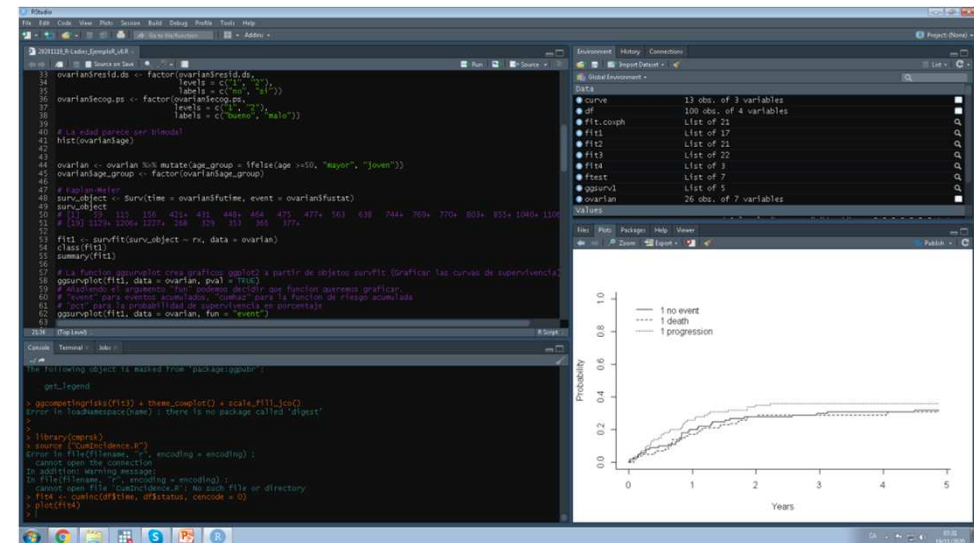
**Institut Català d'Oncologia | Institut d'Investigació Biomèdica de Bellvitge**

Avda. Granvia de l'Hospitalet 199-203, 08908 L'Hospitalet de Llobregat, Barcelona

<http://ico.gencat.cat> | [www.idibell.cat](http://www.idibell.cat)

# Índice de la presentación

- Un poco de teoría
- Un poco de sintaxis
- Ejemplos prácticos



# Un poco de teoría



- En el **análisis de supervivencia** se estudia el tiempo hasta un evento específico de interés, como por ejemplo la muerte o la recurrencia de una enfermedad, y se comparan dos (o más) grupos de pacientes respecto a ese tiempo.
- Sea  $T$  el tiempo hasta un suceso de interés,  $\varepsilon$ :
  - **Función de supervivencia,  $S(t)$** : Probabilidad de que un individuo sobreviva más de  $t$  unidades de tiempo, es decir, que  $\varepsilon$  ocurra después de  $t$  años:

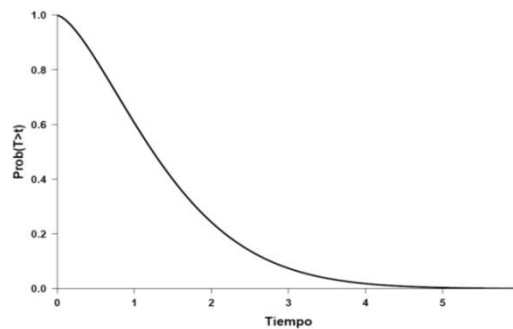
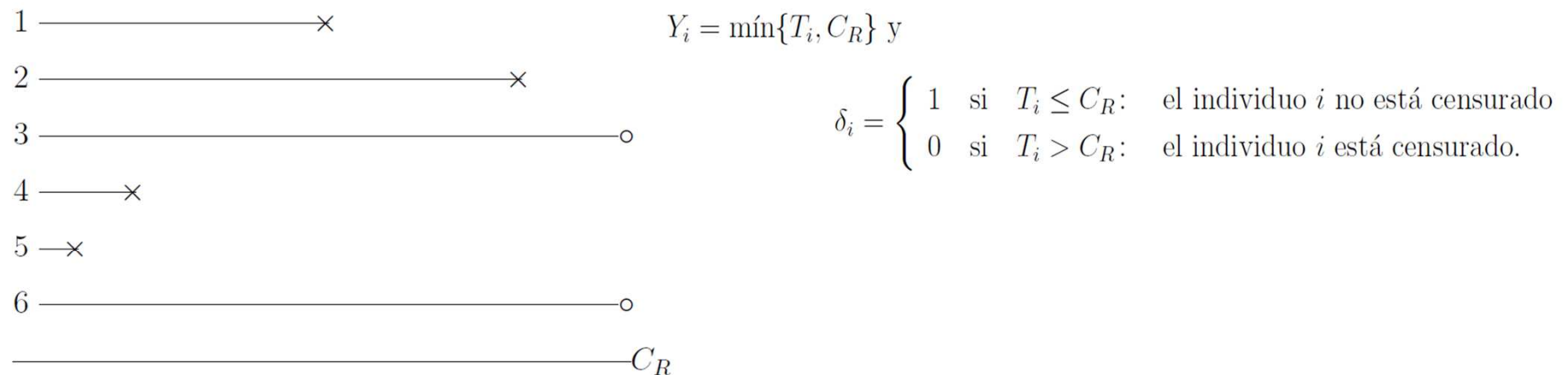


Figura 1.1: Función de supervivencia correspondiente a una distribución de Weibull con parámetro de escala  $\rho = 0,63$  y de forma  $k = 1,5$

- El término "**censura**" se refiere a datos incompletos. Nos centraremos en esta presentación en la censura por la derecha, por ser la más común en este tipo de análisis:
  - Pacientes que se pierden
  - Pacientes que abandonan el estudio
  - Pacientes que no han experimentado el evento de interés al finalizar el tiempo de seguimiento del estudio



**Figura 2.2:** Datos con censura por la derecha, de tipo I. Los individuos 3 y 6 están censurados por la derecha

La censura tiene que ser **no informativa** y nunca causada por el evento de interés (***competing risk analysis***)

- El **estimador Kaplan-Meier** es un estadístico no paramétrico que nos permite estimar la función de supervivencia:
  - No se basa en la asunción de una distribución de probabilidad subyacente
  - Proporciona la probabilidad que un individuo sobreviva pasado un determinado tiempo  $t$ .
  - Para  $t = 0$ , el estimador Kaplan-Meier vale 1 y a medida que  $t$  tiende a infinito el estimador tiende a 0
  - Se basa en la asunción de que la probabilidad de sobrevivir a un cierto tiempo  $t$  es igual al producto de las tasas observadas de supervivencia hasta  $t$  (probabilidades condicionadas)
- Estimador de Nelson y Aalen:
  - Se usa para la estimación de la función de riesgo y riesgo acumulado
  - Se comporta mejor cuando las muestras son pequeñas
  - Permite seleccionar entre modelos paramétricos

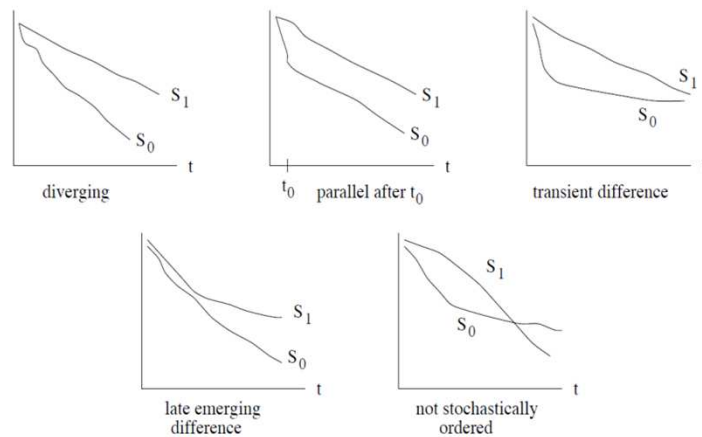


Figura 4.1: Diferentes formas en que pueden diferir dos curvas de supervivencia (Apuntes de Lagakos). Aquí usamos 0 y 1 en lugar de 1 y 2.

En función de cómo sean de diferentes las dos curvas se pueden aplicar pesos distintos para una mejor estimación.

Pruebas:

- Prueba de Gehan
- Prueba de Tarone y Ware
- Prueba de Peto
- Familia de pruebas de Fleming y Harrington

¿Cómo escoger la prueba?

- Prueba de Gehan → detecta diferencias a corto plazo (si las censuras no son muy distintas)
- **Prueba de logrank** → cuando las exposiciones tienen un riesgo constante. No afecta la censura
- Prueba de Tarone Ware → eficiente para un rango muy amplio de alternativas
- Familia de pruebas de Fleming y Harrington → mediante la elección de  $p$  y  $q$  permiten identificar diferencias a corto o a largo plazo ( $p=q=0 \rightarrow \text{logrank}$ )

## — El modelo de Cox, riesgos proporcionales:

- Equivale al modelo de regresión lineal en análisis de la supervivencia
- Sea  $T$  una variable aleatoria que indica el tiempo hasta un evento  $\mathcal{E}$ , y sea  $\mathbf{Z} = (Z_1, Z_2, \dots, Z_p)$  el perfil que recoge los factores pronósticos (covariantes) recogido al inicio del estudio
- Se basa en la función de riesgo

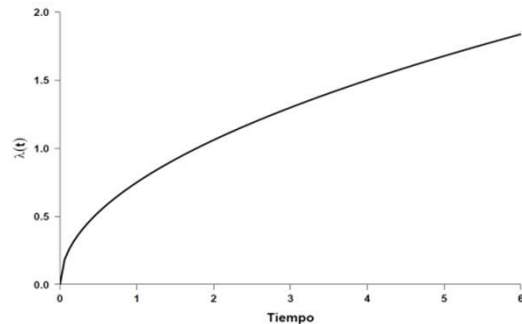


Figura 1.4: Función de riesgo correspondiente a una distribución de Weibull con parámetro de escala  $\rho = 0,63$  y de forma  $k = 1,5$

- Puede ajustarse por covariables en un modelo de regresión
- Asume que el riesgo de las covariables incluidas en el modelo es constante en el tiempo para todos sus niveles.
  - Si no se cumple → extensiones del modelo de Cox

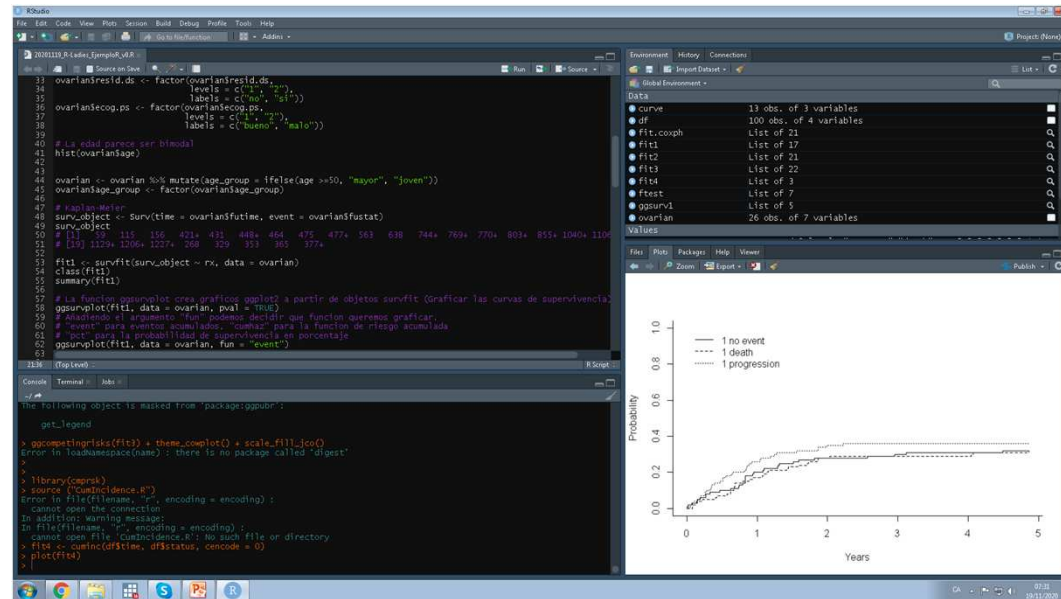
# Bibliografía

- Análisis de la supervivencia. Guadalupe Gómez (UPC), Olga Julià (UB) y Klaus Langohr(UPC).
- Klein, J.P. y Moeschberger, M.L. (1997). *Survival Analysis. Techniques for Censored and Truncated Data*. New York: Wiley.



# Un poco de sintaxis

- Librerías:
  - survival
  - survminer
  - dplyr



- Base de datos de ejemplo:
  - ovarian dataset (Edmunson J.H. et al., 1979). Está en la librería *survival*

Referencia: <https://www.datacamp.com/community/tutorials/survival-analysis-R>

# Ejemplos prácticos

# Referencias:

# <https://www.aridhia.com/blog/r-for-researchers-8-essential-cheatsheets-for-research-data-analysis/>

#

<https://www.datacamp.com/community/tutorials/survival-analysis-R#comments>

# <https://rpkgs.datanovia.com/survminer/index.html>

# <http://www.stat.unipg.it/luca/R/>

## # Cargamos las librerías

`#install.packages()` si no se han instalado antes las librerías

```
library(survival)
```

```
#devtools::install_github("kassambara/survminer",  
build_vignettes = FALSE)
```

```
library(survminer)
```

```
library(dplyr)
```

# # Importamos la base de datos de cancer de ovario ("ovarian") y le echamos un vistazo

```
data(ovarian)
glimpse(ovarian)
# Rows: 26
# Columns: 6
# $ futime  <dbl> 59, 115, 156, 421, 431, 448, 464, 475, 477, 563, 638, 744, 769, 770,
803, 855, 1040, 1106, 1129...
# $ fustat  <dbl> 1, 1, 1, 0, 1, 0, 1, 1, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 0
# $ age     <dbl> 72.3315, 74.4932, 66.4658, 53.3644, 50.3397, 56.4301, 56.9370,
59.8548, 64.1753, 55.1781, 56.75...
# $ resid.ds <dbl> 2, 2, 2, 2, 2, 1, 2, 2, 2, 1, 1, 1, 2, 2, 1, 1, 2, 1, 1, 2, 1, 2, 2, 1, 2, 1
# $ rx      <dbl> 1, 1, 1, 2, 1, 1, 2, 2, 1, 2, 1, 2, 2, 2, 1, 1, 1, 1, 2, 2, 2, 1, 1, 2, 2, 2
# $ ecog.ps <dbl> 1, 1, 2, 1, 1, 2, 2, 2, 1, 2, 2, 1, 2, 1, 1, 2, 2, 1, 1, 1, 2, 2, 1, 2, 1, 1

help(ovarian)
```

## # Dicotomizamos la edad y cambiamos las etiquetas

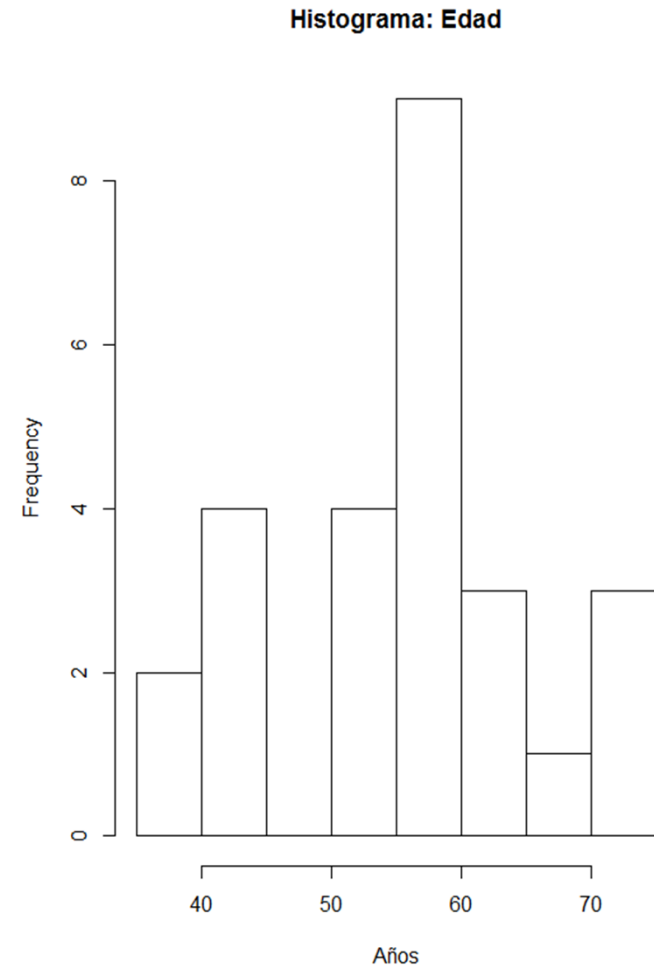
```
ovarian$rx <- factor(ovarian$rx,  
                     levels = c("1", "2"),  
                     labels = c("A", "B"))  
ovarian$resid.ds <- factor(ovarian$resid.ds,  
                           levels = c("1", "2"),  
                           labels = c("no", "si"))  
ovarian$ecog.ps <- factor(ovarian$ecog.ps,  
                           levels = c("1", "2"),  
                           labels = c("bueno",  
                                       "malo"))
```

## # La edad parece ser bimodal

```
hist(ovarian$age)
```

```
ovarian <- ovarian %>%  
  mutate(age_group =  
    ifelse(age >= 50,  
      "mayor", "joven"))
```

```
ovarian$age_group <-  
factor(ovarian$age_group)
```



# # Kaplan-Meier

```
surv_object <- Surv(time = ovarian$futime, event =  
ovarian$fustat)
```

```
surv_object
```

```
#[1] 59 115 156 421+ 431 448+ 464 475 477+ 563 638 744+ 769+ 770+ 803+ 855+ 1040+ 1106+  
#[19] 1129+ 1206+ 1227+ 268 329 353 365 377+
```

```
fit1 <- survfit(surv_object ~ rx, data = ovarian)
```

```
class(fit1)
```

```
Call: survfit(formula = surv_object ~ rx, data = ovarian)
```

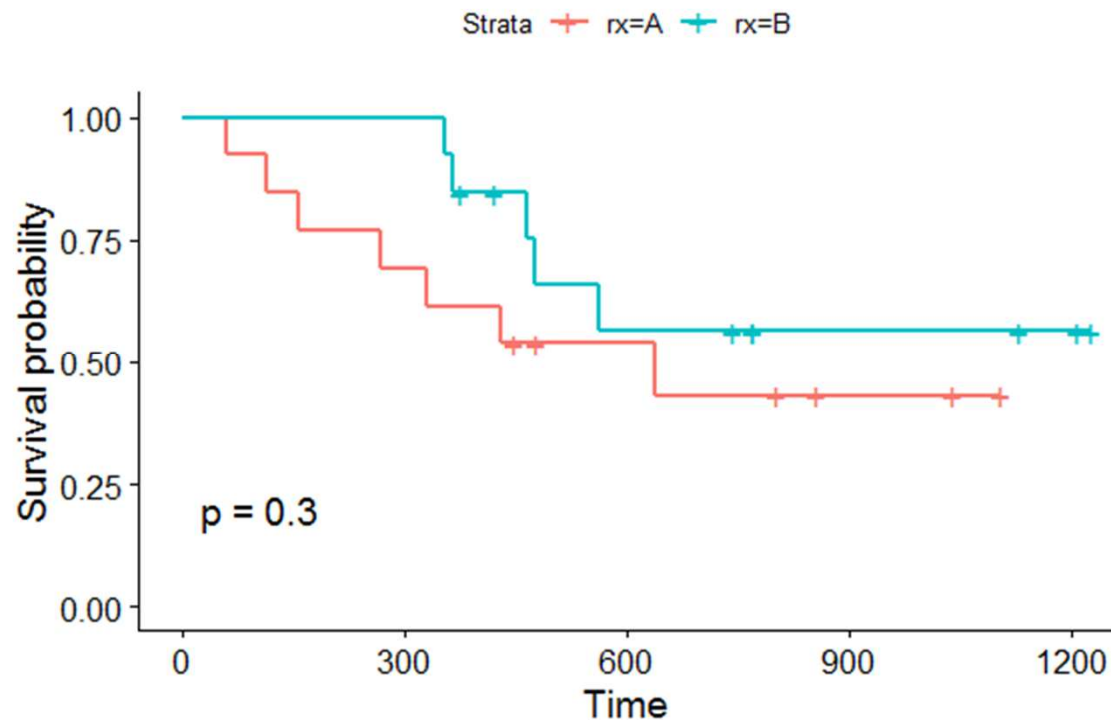
```
summary(fit1)
```

rx=A								
time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
59	13	1	0.923	0.0739		0.789		1.000
115	12	1	0.846	0.1001		0.671		1.000
156	11	1	0.769	0.1169		0.571		1.000
268	10	1	0.692	0.1280		0.482		0.995
329	9	1	0.615	0.1349		0.400		0.946
431	8	1	0.538	0.1383		0.326		0.891
638	5	1	0.431	0.1467		0.221		0.840

rx=B								
time	n.risk	n.event	survival	std.err	lower	95% CI	upper	95% CI
353	13	1	0.923	0.0739		0.789		1.000
365	12	1	0.846	0.1001		0.671		1.000
464	9	1	0.752	0.1256		0.542		1.000
475	8	1	0.658	0.1407		0.433		1.000
563	7	1	0.564	0.1488		0.336		0.946

# La funcion ggsurvplot crea graficos ggplot2 a partir de objetos survfit (Graficar las curvas de supervivencia)

```
ggsurvplot(fit1, data = ovarian, pval = TRUE)
```





#Añadiendo el argumento "fun" podemos decidir que funcion queremos graficar.

```
ggsurvplot(fit1, data =  
ovarian, fun = "event")
```

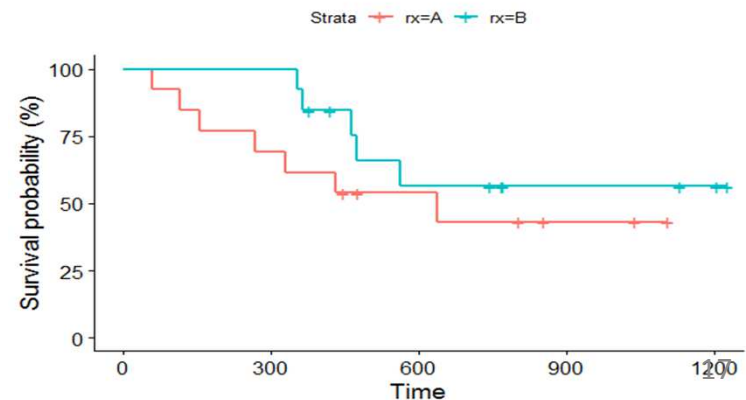
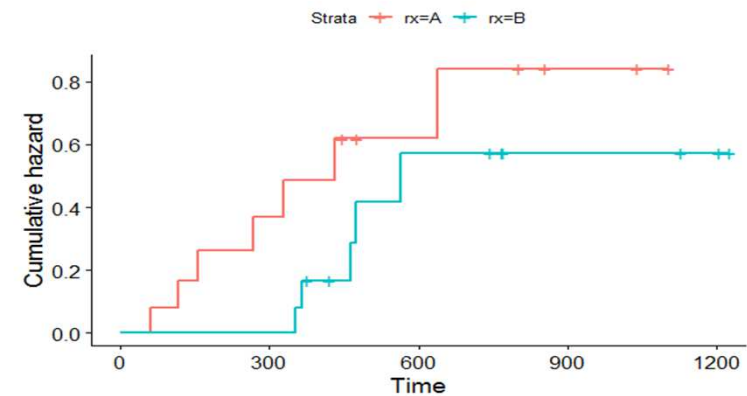
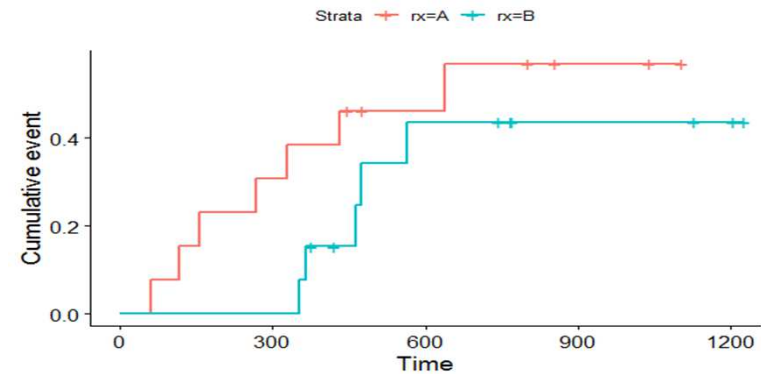
"event" para eventos  
acumulados

```
ggsurvplot(fit1, data =  
ovarian, fun = "cumhaz")
```

"cumhaz" para la funcion  
de riesgo acumulada

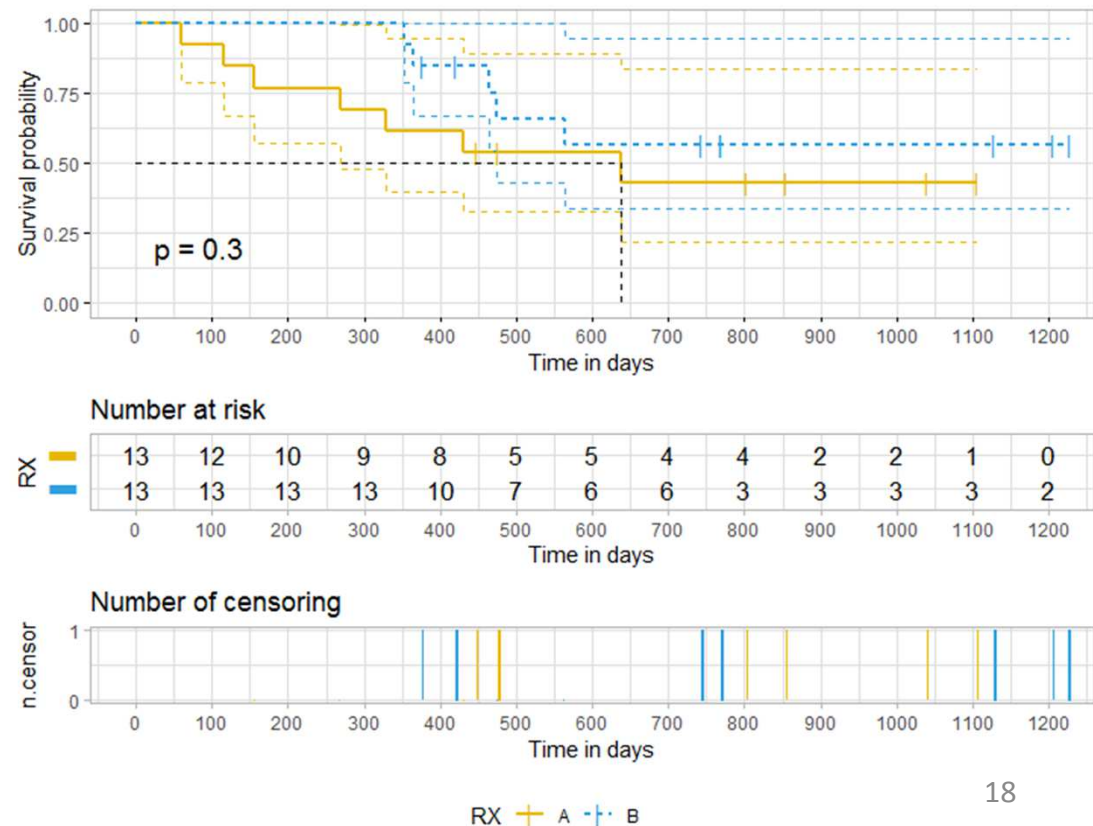
```
ggsurvplot(fit1, data =  
ovarian, fun = "pct")
```

"pct" para la probabilidad de  
supervivencia en porcentaje



# ggsurvplot tiene muchos parámetros gráficos para **escoger**: intervalos de confianza, mostrar la tabla de individuos a riesgo, la posición de la leyenda, anotaciones adicionales, añadir el p-valor del test de logrank), titulo, subtítulo, ...

```
ggsurv1 <- ggsurvplot(fit1, data = ovarian, conf.int = TRUE,
  conf.int.style = "step", pval = TRUE, risk.table = TRUE,
  risk.table.height = 0.25, risk.table.y.text.col = TRUE,
  risk.table.y.text = FALSE, size = 1, linetype = "strata",
  palette = c("#E7B800", "#2E9FDF"), legend =
  "bottom", legend.title = "RX", legend.labs =
  c("A", "B"), censor.shape = "|", censor.size = 4,
  ncensor.plot = TRUE, ncensor.plot.height =
  0.25, ggtheme = theme_
  light(), xlim = c(0,1200),
  , xlab = "Time in days",
  break.time.by = 100,
  surv.median.line = "hv")
```

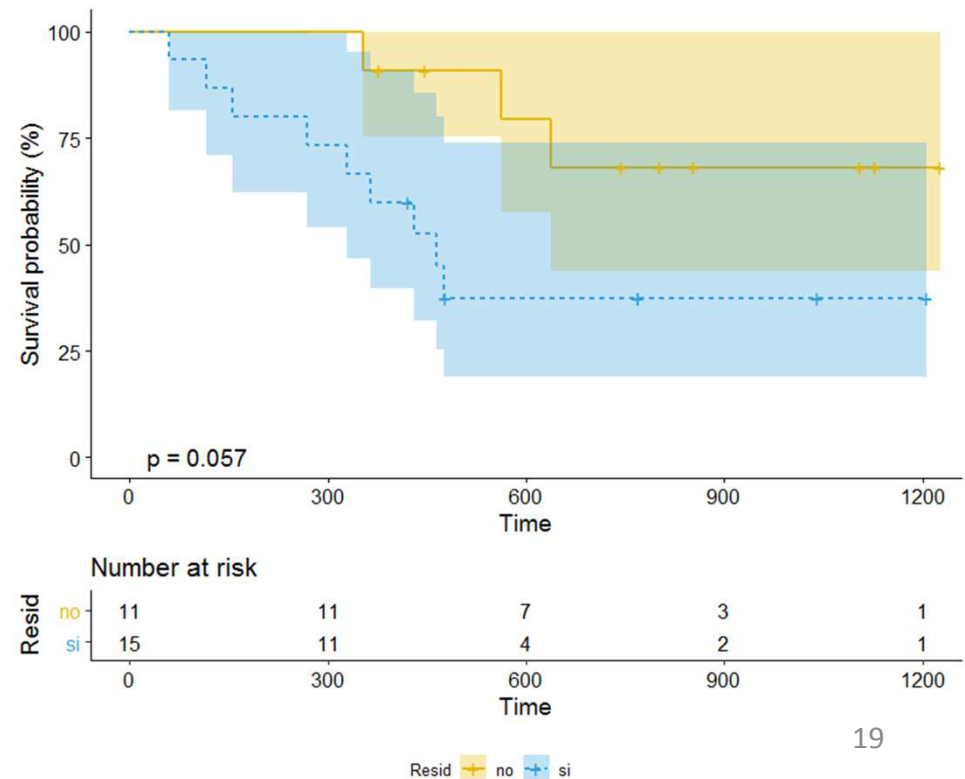
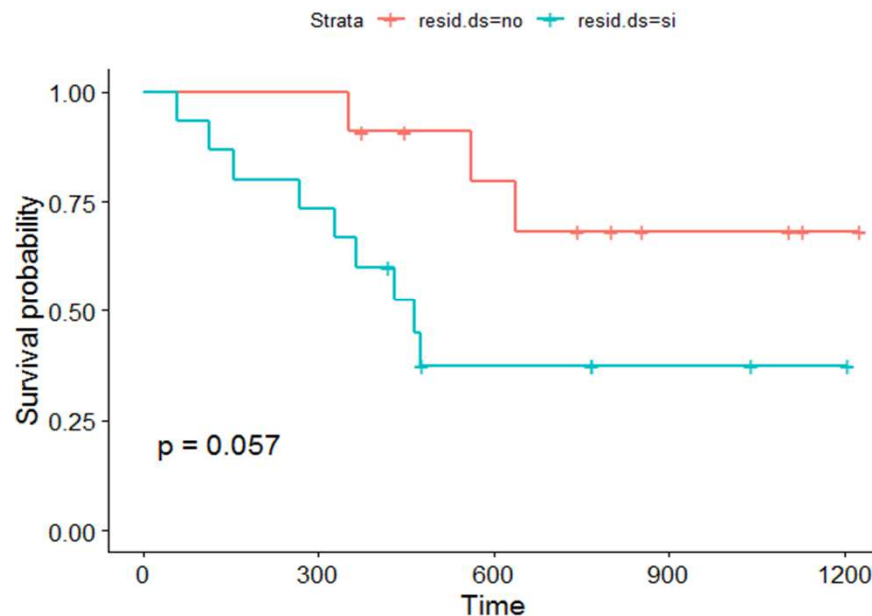


ggsurv1

# # Valores predictivos del status residual de la enfermedad

```
fit2 <- survfit(surv_object ~ resid.ds, data = ovarian)
```

```
ggsurvplot(fit2, data = ovarian, pval = TRUE)  
ggsurvplot(fit2, data = ovarian, conf.int = TRUE,  
  pval = TRUE, fun = "pct", risk.table = TRUE,  
  size = 1, linetype = "strata",  
  palette = c("#E7B800", "#2E9FDF"),  
  legend = "bottom", legend.title = "Resid",  
  legend.labs = c("no", "si"))
```



# # Ajustamos un modelo de riesgos proporcionales de Cox

```
fit.coxph <- coxph(surv_object ~ rx + resid.ds + age_group +  
                  ecog.ps, data = ovarian)  
summary(fit.coxph)
```

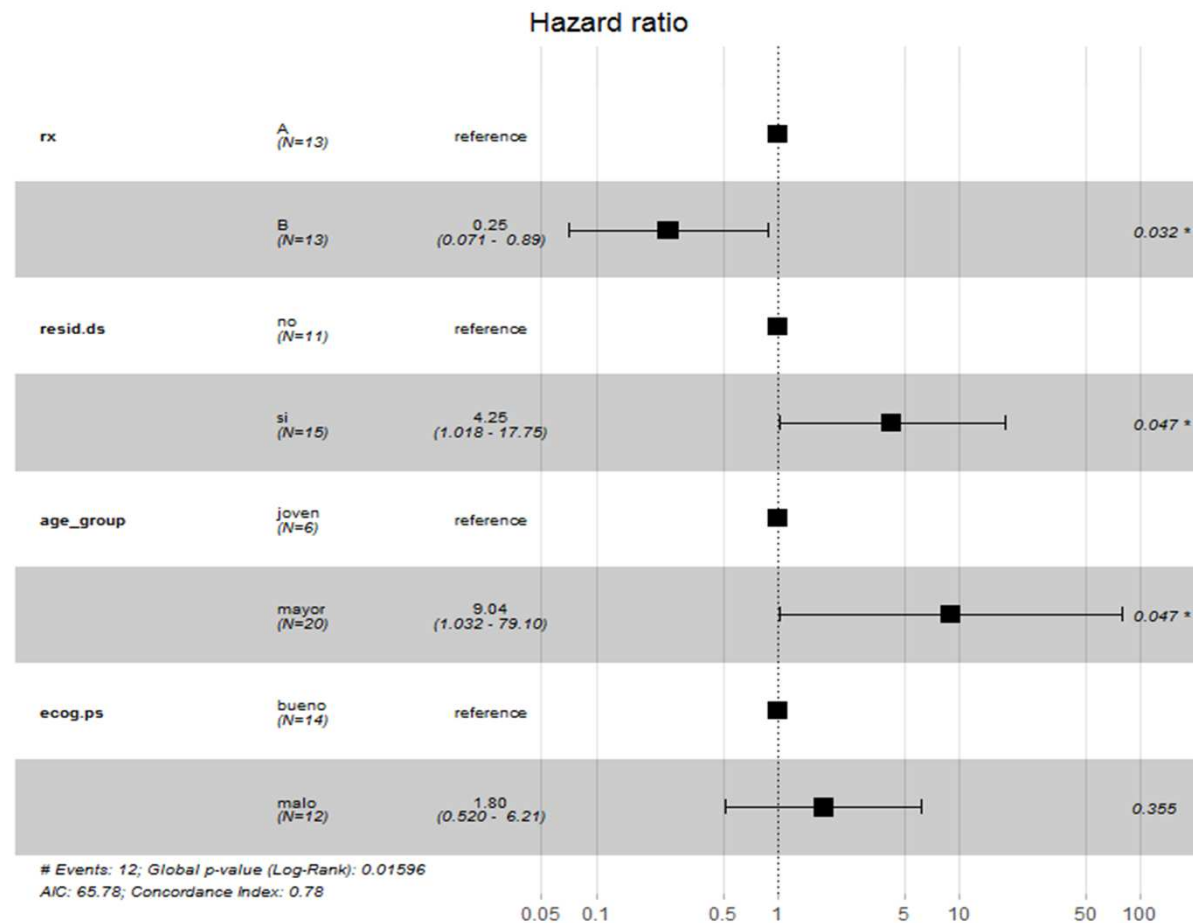
```
> summary(fit.coxph)  
Call:  
coxph(formula = surv_object ~ rx + resid.ds + age_group + ecog.ps,  
      data = ovarian)  
  
n= 26, number of events= 12  
  
              coef exp(coef) se(coef)      z Pr(>|z|)  
rxB           -1.3814    0.2512   0.6448 -2.142  0.0322 *  
resid.dssi      1.4470    4.2503   0.7292  1.984  0.0472 *  
age_groupmayor  2.2013    9.0368   1.1069  1.989  0.0467 *  
ecog.psmalo     0.5859    1.7966   0.6329  0.926  0.3546  
---  
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
  
              exp(coef) exp(-coef) lower .95 upper .95  
rxB              0.2512    3.9803    0.0710    0.8891  
resid.dssi        4.2503    0.2353    1.0180   17.7453  
age_groupmayor    9.0368    0.1107    1.0324   79.1031  
ecog.psmalo       1.7966    0.5566    0.5197    6.2113  
  
Concordance= 0.782 (se = 0.065 )  
Likelihood ratio test= 12.19 on 4 df,  p=0.02  
Wald test              = 9.02 on 4 df,  p=0.06  
Score (logrank) test = 11.97 on 4 df,  p=0.02
```

# # La funcion ggforest() de la libreria survminer crea un foresplot para el modelo de Cox ajustado

```
ggforest(fit.coxph, data = ovarian)
```

# Para cada covariable,  
muestra la estimacion de:

1. Hazard ratios,
2. Intervalos de confianza
3. P-valores



# La funcion `cox.zph()` de la libreria "survival" se puede usar para verificar la asuncion de riesgos proporcionales de un modelo de Cox ajustado

# La funcion `ggcoxzph()` de la libreria "survminer" podemos hacer una verificacion grafica

# Produce un grafico del tiempo vs los residuos de Scheonfeld para cada covariable

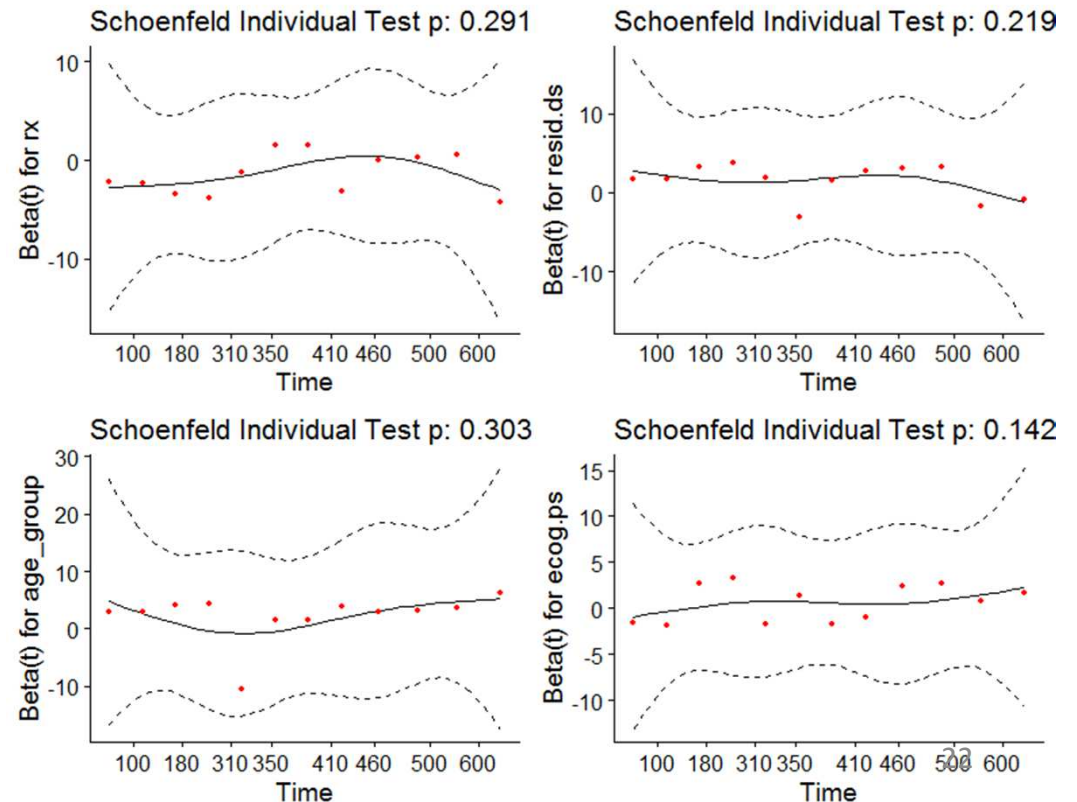
```
ftest <- cox.zph(fit.coxph)
fctest
```

```
ggcoxzph(fctest)
```

```
> fctest
```

	chisq	df	p
rx	1.11	1	0.29
resid.ds	1.51	1	0.22
age_group	1.06	1	0.30
ecog.ps	2.15	1	0.14
GLOBAL	5.31	4	0.26

Global Schoenfeld Test p: 0.2571

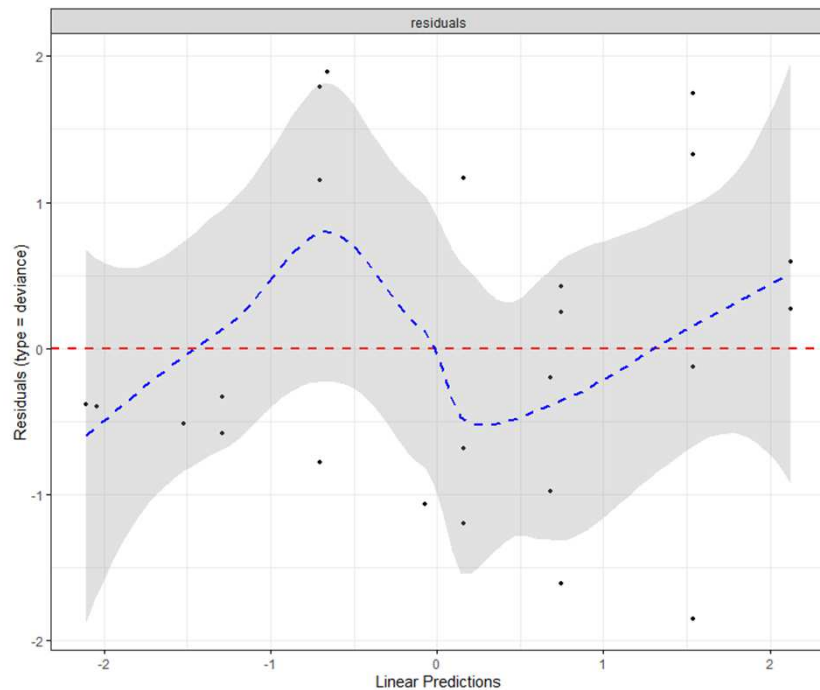


# La funcion `ggcoxdiagnostics()` de la libreria "survminer" grafica diferentes residuos como funciones del tiempo, predictores lineales o las propias observaciones

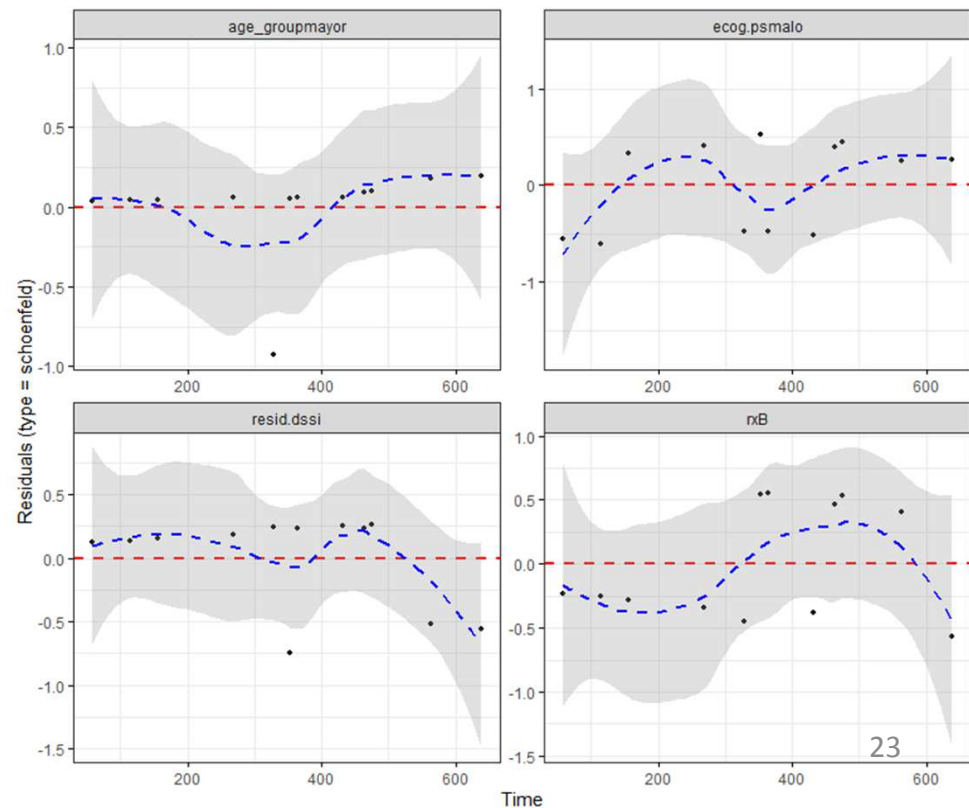
# El parametro `type` permite escoger el tipo de residuos: "martingale", "deviance", "score", "schoenfeld", "dfbeta", "dfbetas", "scaledsch"

# El parametro `ox.scale` permite escoger que graficar en el eje OX: "linear.predictions", "observation.id" o "time"

```
ggcoxdiagnostics(fit.coxph, type =  
"deviance", ox.scale = "linear.predictions")
```



```
ggcoxdiagnostics(fit.coxph, type =  
"schoenfeld", ox.scale = "time")
```

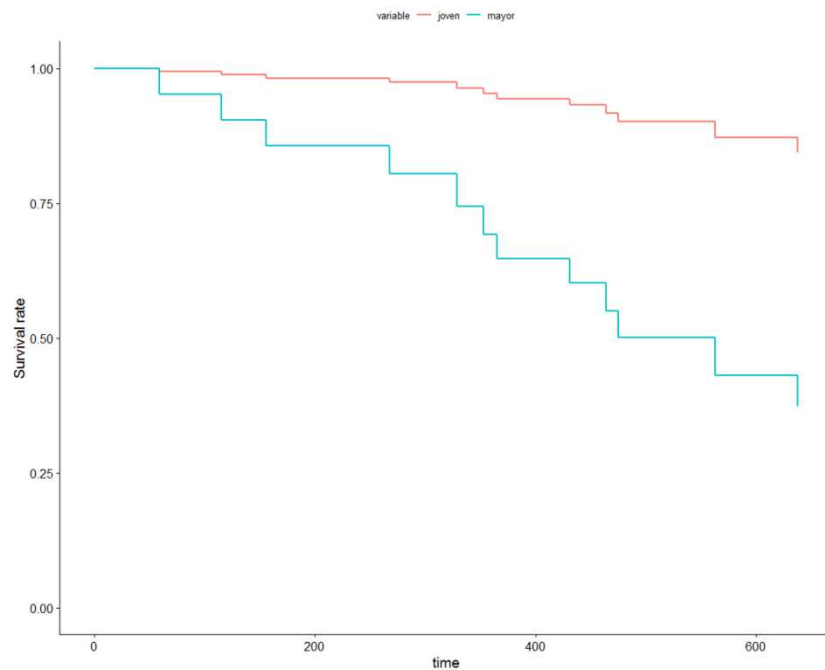




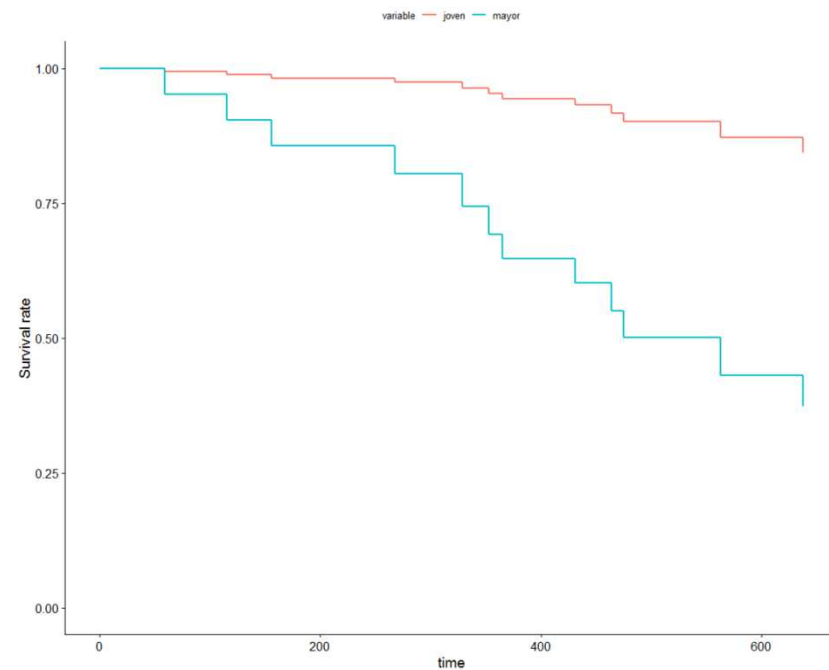
# La funcion `ggadjustedcurves()` de la libreria `survimer` grafica las curvas de supervivencia ajustadas por el modelo de riesgos proporcionales de Cox

# Estas curvas difieren de las creadas con Kaplan-Meier ya que contienen la informacion del modelo de Cox ajustado

```
ggadjustedcurves(fit.coxph, data  
= ovarian, variable='age_group')
```



```
ggadjustedcurves(fit.coxph,  
data = ovarian, variable='rx')
```





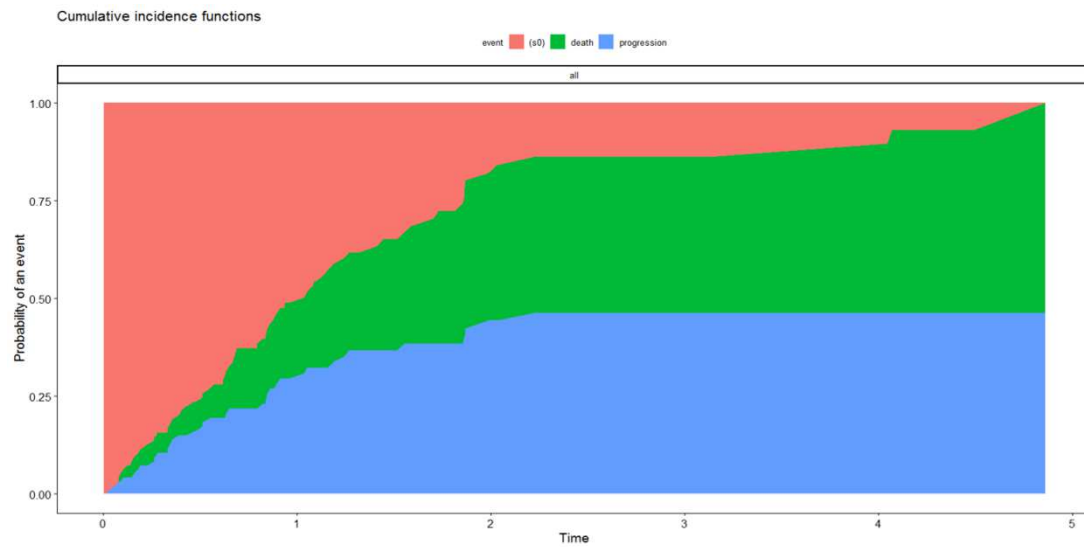
## #Competing risks

### # Computaciones preliminares:

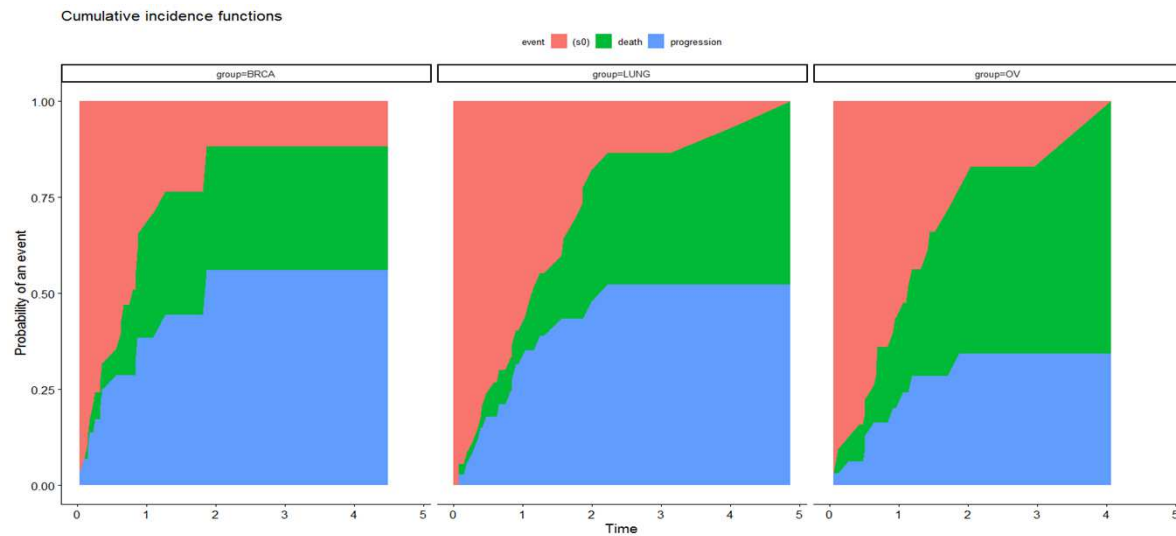
```
set.seed(2)
ss <- rexp(100)
gg <-
factor(sample(1:3,100,replace=TRUE),1:3,c('BRCA','LUNG','OV'))
cc <- factor(sample(0:2,100,replace=TRUE),0:2,c('no
event', 'death', 'progression'))
strt <- sample(1:2,100,replace=TRUE)

df <- data.frame(time = ss, group = gg, status = cc,
strt)
```

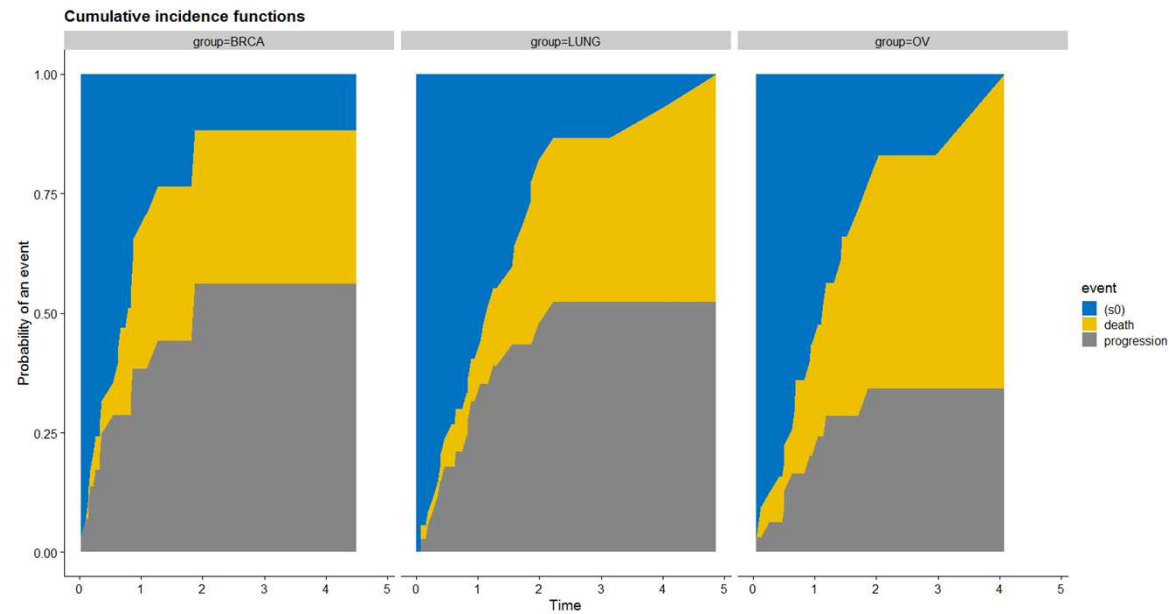
```
fit2 <- survfit(Surv(time, status, type="mstate") ~ 1, data=df)
ggcompetingrisks(fit2)
```



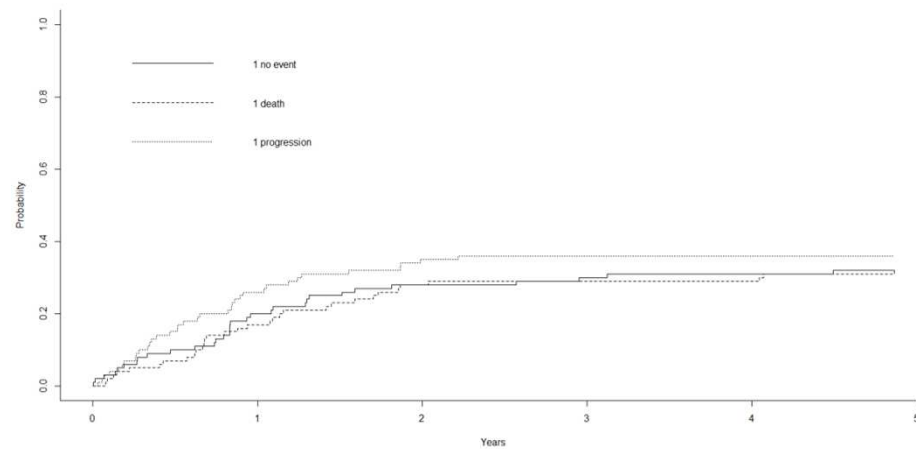
```
fit3 <- survfit(Surv(time, status, type="mstate") ~ group, data=df)
ggcompetingrisks(fit3)
```



```
library(ggsci);library(cowplot)
ggcompetingrisks(fit3) + theme_cowplot() + scale_fill_jco()
```



```
library(cmprsk)
fit4 <- cuminc(df$time, df$status, cencode = 0)
plot(fit4)
```



# Introducción al análisis de la supervivencia con



**Sara Tous i Belmonte**

*Estadística*

Unitat d'Infeccions i Càncer - Molecular, Programa d'Epidemiologia en Recerca del Càncer

Tel. +34 932607812 - ext 3185

[stous@iconcologia.net](mailto:stous@iconcologia.net)

IDIBELL  
Institut d'Investigació Biomèdica de Bellvitge

Salut/

ICO  
Institut Català d'Oncologia

OBJECTIUS GLOBAIS  
DE DESARROL·L SOSTENIBLE

**Institut Català d'Oncologia | Institut d'Investigació Biomèdica de Bellvitge**

Avda. Granvia de l'Hospitalet 199-203, 08908 L'Hospitalet de Llobregat, Barcelona

<http://ico.gencat.cat> | [www.idibell.cat](http://www.idibell.cat)