

Datos armonizados

Curvas de supervivencia

Elena Eyre Sánchez, PhD

2024-12-29

Contents

1	Introducción y Objetivo	1
2	Paquetes y datos	1
2.1	Metadata original	1
3	Overall survival	2
3.1	Todos los datos	2
3.2	Reducción del período de seguimiento	3
3.3	Inmunoterapias	5
3.4	Por terapia disponible	6

1 Introducción y Objetivo

2 Paquetes y datos

2.1 Metadata original

Los datos a comparar son los datos armonizados de los resultados obtenidos de los análisis previos de deconvolución:

- GSE22155 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi>) con dos plataformas: GPL6102 y GPL6947.
- GSE35640 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE35640>)
- GSE50509 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE50509>)
- GSE61992 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE61992>)
- GSE54467 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE54467>)
- TCGA-SKCM (<https://portal.gdc.cancer.gov/projects/TCGA-SKCM>)

De manera excepcional, tenemos GSE120575 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE120575>), que consiste en un dataset scRNA-seq en el que han analizado muestras tumorales y controles.

3 Overall survival

3.1 Todos los datos

Inicialmente examino todos los estudios que aportaban información de supervivencia (tanto los estudios que pasaron la deconvolución, como el estudio de inmunoterapias de datos scRNA-seq).

Debido a que hay estudios sin datos sobre supervivencia, hago también un pre-procesamiento para eliminarlos de este análisis. Estudios con información útil para el análisis de supervivencia:

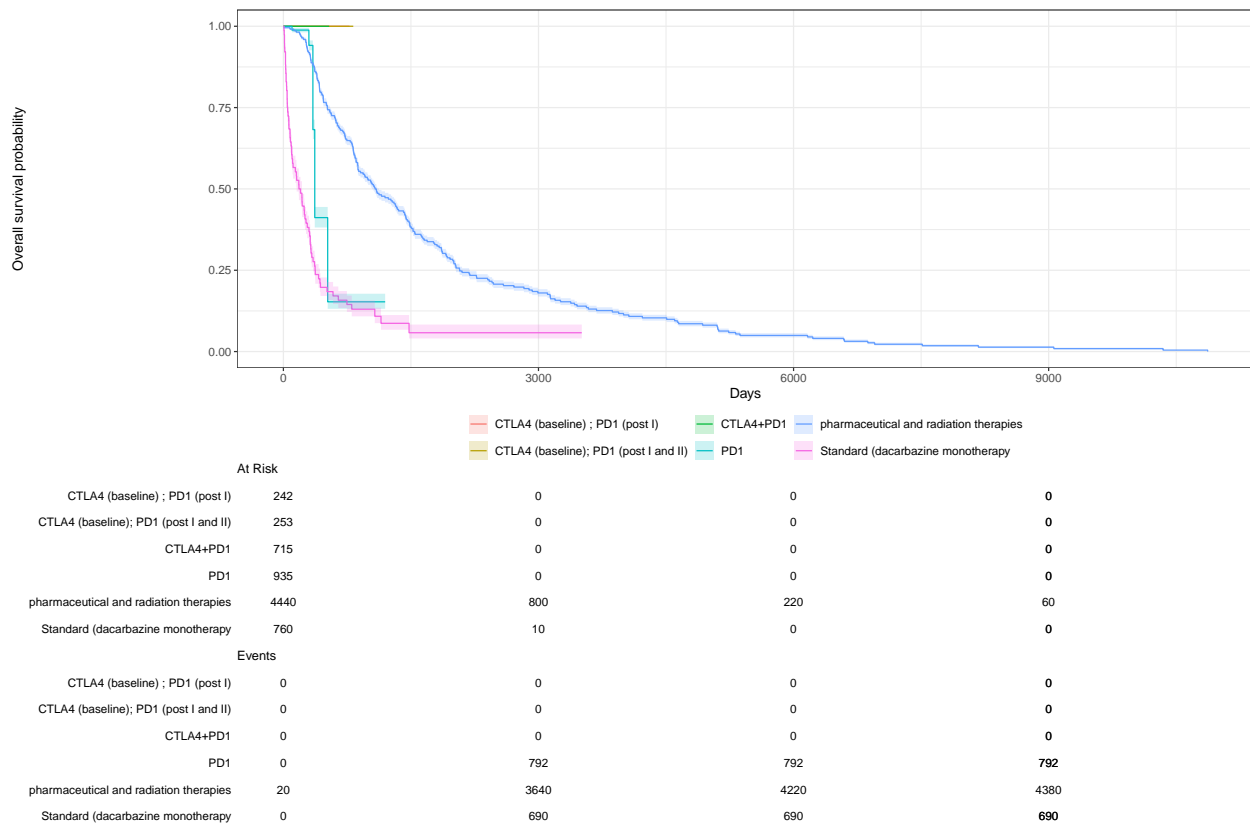
```
## [1] "GSE120575"          "tcga_skcm"          "GSE22155_GPL6102" "GSE22155_GPL6947"
```

Para generar gráficos Kaplan-Meier, primero creo un survival object con la función `survfit()`, y el gráfico se genera con la función `ggsurvfit()`:

```
## List of 18
## $ n      : int [1:6] 242 253 715 935 4440 760
## $ time   : num [1:295] 777 822 339 388 539 101 300 347 369 521 ...
## $ n.risk : num [1:295] 242 253 715 473 242 935 924 880 638 385 ...
## $ n.event : num [1:295] 0 0 0 0 0 11 44 242 253 242 ...
## $ n.censor : num [1:295] 242 253 242 231 242 0 0 0 0 0 ...
## $ surv     : num [1:295] 1 1 1 1 1 ...
## $ std.err  : num [1:295] 0 0 0 0 0 ...
## $ cumhaz   : num [1:295] 0 0 0 0 0 ...
## $ std.chaz : num [1:295] 0 0 0 0 0 ...
## $ strata   : Named int [1:6] 1 1 3 6 215 69
## $.- attr(*, "names")= chr [1:6] "treatment=CTLA4 (baseline) ; PD1 (post I)" "treatment=CTLA4 (base
## $ type     : chr "right"
## $ logse    : logi TRUE
## $ conf.int : num 0.95
## $ conf.type: chr "log"
## $ lower    : num [1:295] 1 1 1 1 1 ...
## $ upper    : num [1:295] 1 1 1 1 1 ...
## $ t0       : num 0
## $ call     : language survfit(formula = Surv(os_days, os_event) ~ treatment, data = df2)
## $.- attr(*, "class")= chr "survfit"
```

Characteristic	HR ¹	95% CI ¹	p-value
treatment			
CTLA4 (baseline) ; PD1 (post I)	—	—	
CTLA4 (baseline); PD1 (post I and II)	1.01	0.00, Inf	>0.9
CTLA4+PD1	1.00	0.00, Inf	>0.9
PD1	47,127,473	0.00, Inf	>0.9
pharmaceutical and radiation therapies	16,839,495	0.00, Inf	>0.9
Standard (dacarbazine monotherapy)	78,576,300	0.00, Inf	>0.9

¹HR = Hazard Ratio, CI = Confidence Interval



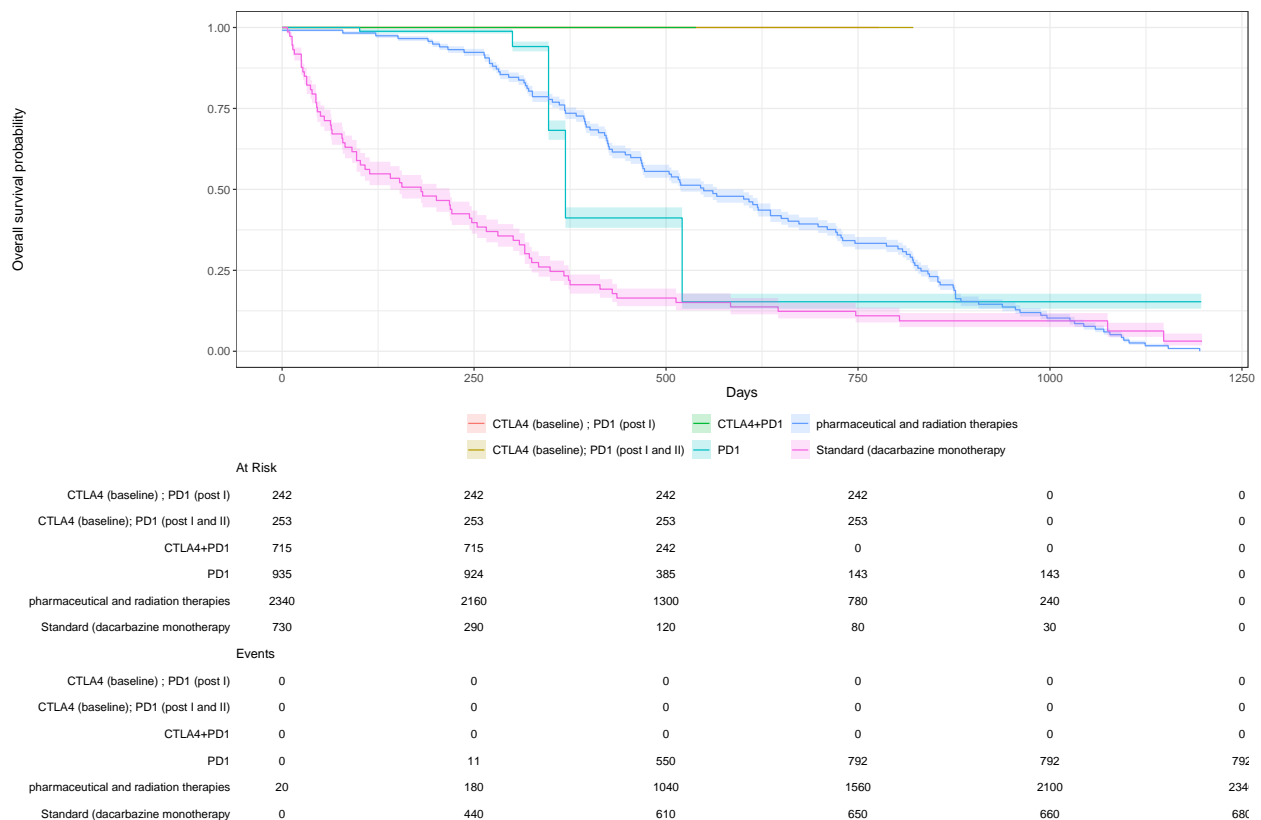
El resumen estadístico con la función `coxph()` devuelve p-valores que sugiere la no significancia de la comparación. Probablemente esto se deba a la desigualdad entre las diferentes curvas de los estudios.

3.2 Reducción del período de seguimiento

Para confirmar si la falta de comparaciones significativas se debe a diferencia de las curvas debido a que los seguimientos fueron prolongados únicamente en pocos estudios, examino un período más reducido.

```
## List of 18
## $ n      : int [1:6] 242 253 715 935 2340 730
## $ time   : num [1:187] 777 822 339 388 539 101 300 347 369 521 ...
## $ n.risk : num [1:187] 242 253 715 473 242 935 924 880 638 385 ...
## $ n.event : num [1:187] 0 0 0 0 0 11 44 242 253 242 ...
## $ n.censor : num [1:187] 242 253 242 231 242 0 0 0 0 0 ...
```

```
## $ surv      : num [1:187] 1 1 1 1 1 ...
## $ std.err    : num [1:187] 0 0 0 0 0 ...
## $ cumhaz     : num [1:187] 0 0 0 0 0 ...
## $ std.chaz   : num [1:187] 0 0 0 0 0 ...
## $ strata     : Named int [1:6] 1 1 3 6 110 66
##   .. attr(*, "names")= chr [1:6] "treatment=CTLA4 (baseline) ; PD1 (post I)" "treatment=CTLA4 (base
## $ type       : chr "right"
## $ logse      : logi TRUE
## $ conf.int    : num 0.95
## $ conf.type   : chr "log"
## $ lower       : num [1:187] 1 1 1 1 1 ...
## $ upper       : num [1:187] 1 1 1 1 1 ...
## $ t0         : num 0
## $ call       : language survfit(formula = Surv(os_days, os_event) ~ treatment, data = df3)
## - attr(*, "class")= chr "survfit"
```



La apariencia de los gráficos mejora por permitir observar todos los tratamientos con más claridad, pero no mejoran los resultados estadísticos:

```
## Call:
## survival::coxph(formula = Surv(os_days, os_event) ~ treatment,
##   data = df3)
##
##               coef exp(coef) se(coef)
## treatmentCTLA4 (baseline); PD1 (post I and II) 1.448e-02 1.015e+00 8.591e+02
## treatmentCTLA4+PD1                             1.591e-02 1.016e+00 8.315e+02
## treatmentPD1                                    1.857e+01 1.164e+08 6.330e+02
## treatmentpharmaceutical and radiation therapies 1.862e+01 1.215e+08 6.330e+02
## treatmentStandard (dacarbazine monotherapy      1.950e+01 2.948e+08 6.330e+02
```

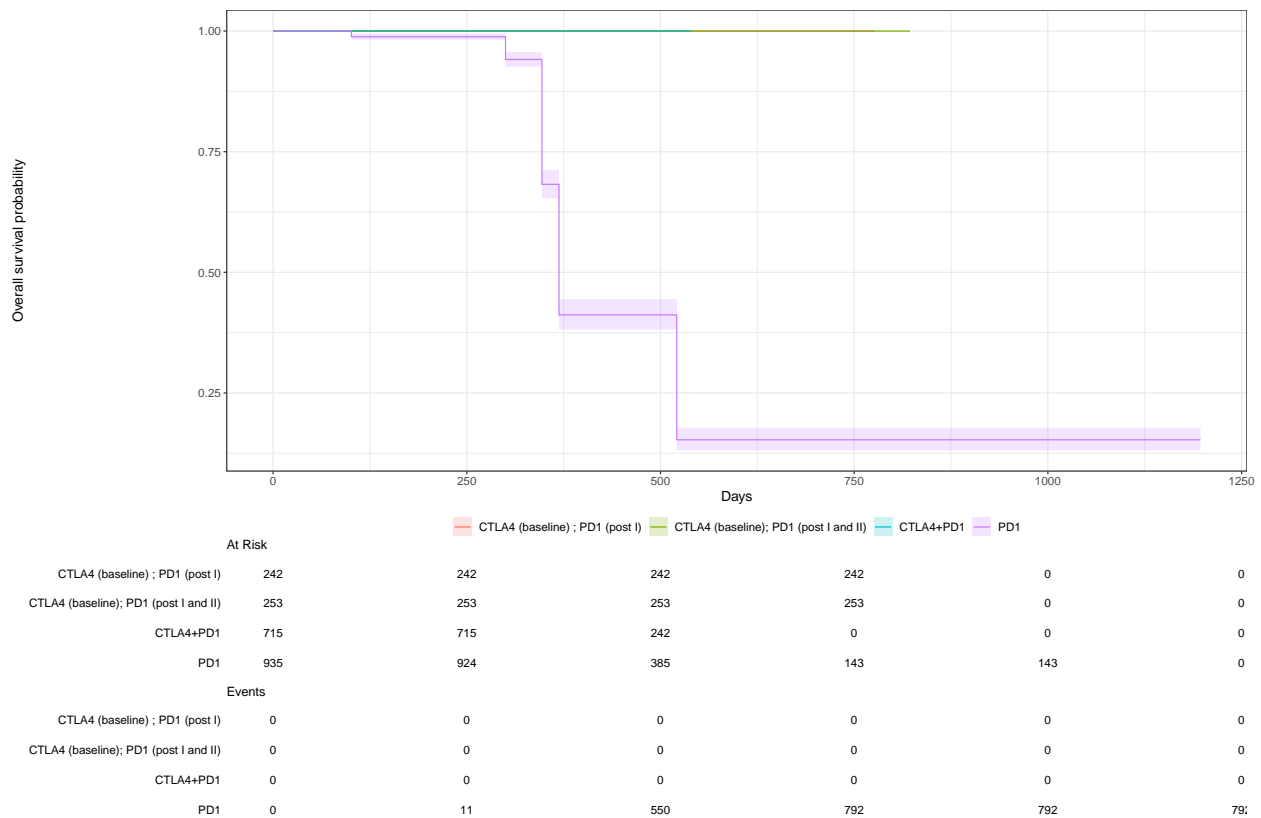
Characteristic	HR [†]	95% CI [†]	p-value
treatment			
CTLA4 (baseline) ; PD1 (post I)	—	—	
CTLA4 (baseline); PD1 (post I and II)	1.01	0.00, Inf	>0.9
CTLA4+PD1	1.02	0.00, Inf	>0.9
PD1	116,413,635	0.00, Inf	>0.9
pharmaceutical and radiation therapies	121,491,275	0.00, Inf	>0.9
Standard (dacarbazine monotherapy)	294,840,245	0.00, Inf	>0.9

[†]HR = Hazard Ratio, CI = Confidence Interval

```
##                               z      p
## treatmentCTLA4 (baseline); PD1 (post I and II) 0.000 1.000
## treatmentCTLA4+PD1                          0.000 1.000
## treatmentPD1                                0.029 0.977
## treatmentpharmaceutical and radiation therapies 0.029 0.977
## treatmentStandard (dacarbazine monotherapy      0.031 0.975
##
## Likelihood ratio test=2373  on 5 df, p=< 2.2e-16
## n= 5215, number of events= 3812
```

3.3 Inmunoterapias

```
## List of 18
## $ n      : int [1:4] 242 253 715 935
## $ time   : num [1:11] 777 822 339 388 539 101 300 347 369 521 ...
## $ n.risk  : num [1:11] 242 253 715 473 242 935 924 880 638 385 ...
## $ n.event : num [1:11] 0 0 0 0 0 11 44 242 253 242 ...
## $ n.censor : num [1:11] 242 253 242 231 242 0 0 0 0 0 ...
## $ surv    : num [1:11] 1 1 1 1 1 ...
## $ std.err  : num [1:11] 0 0 0 0 0 ...
## $ cumhaz   : num [1:11] 0 0 0 0 0 ...
## $ std.chaz : num [1:11] 0 0 0 0 0 ...
## $ strata   : Named int [1:4] 1 1 3 6
##   ..- attr(*, "names")= chr [1:4] "treatment=CTLA4 (baseline) ; PD1 (post I)" "treatment=CTLA4 (base
## $ type      : chr "right"
## $ logse     : logi TRUE
## $ conf.int  : num 0.95
## $ conf.type: chr "log"
## $ lower     : num [1:11] 1 1 1 1 1 ...
## $ upper     : num [1:11] 1 1 1 1 1 ...
## $ t0        : num 0
## $ call      : language survfit(formula = Surv(os_days, os_event) ~ treatment, data = df3)
## - attr(*, "class")= chr "survfit"
```



La apariencia de los gráficos mejora por permitir observar todos los tratamientos con más claridad, pero no mejoran los resultados estadísticos:

```
## Call:
## survival::coxph(formula = Surv(os_days, os_event) ~ treatment,
##   data = df3)
##
##               coef exp(coef) se(coef)
## treatmentCTLA4 (baseline); PD1 (post I and II) -2.637e-07  1.000e+00  2.559e+03
## treatmentCTLA4+PD1                             8.458e-02  1.088e+00  2.328e+03
## treatmentPD1                                   2.114e+01  1.518e+09  1.829e+03
##               z      p
## treatmentCTLA4 (baseline); PD1 (post I and II) 0.000 1.000
## treatmentCTLA4+PD1                             0.000 1.000
## treatmentPD1                                   0.012 0.991
##
## Likelihood ratio test=1696 on 3 df, p=< 2.2e-16
## n= 2145, number of events= 792
```

3.4 Por terapia disponible

3.4.1 En perspectiva de género

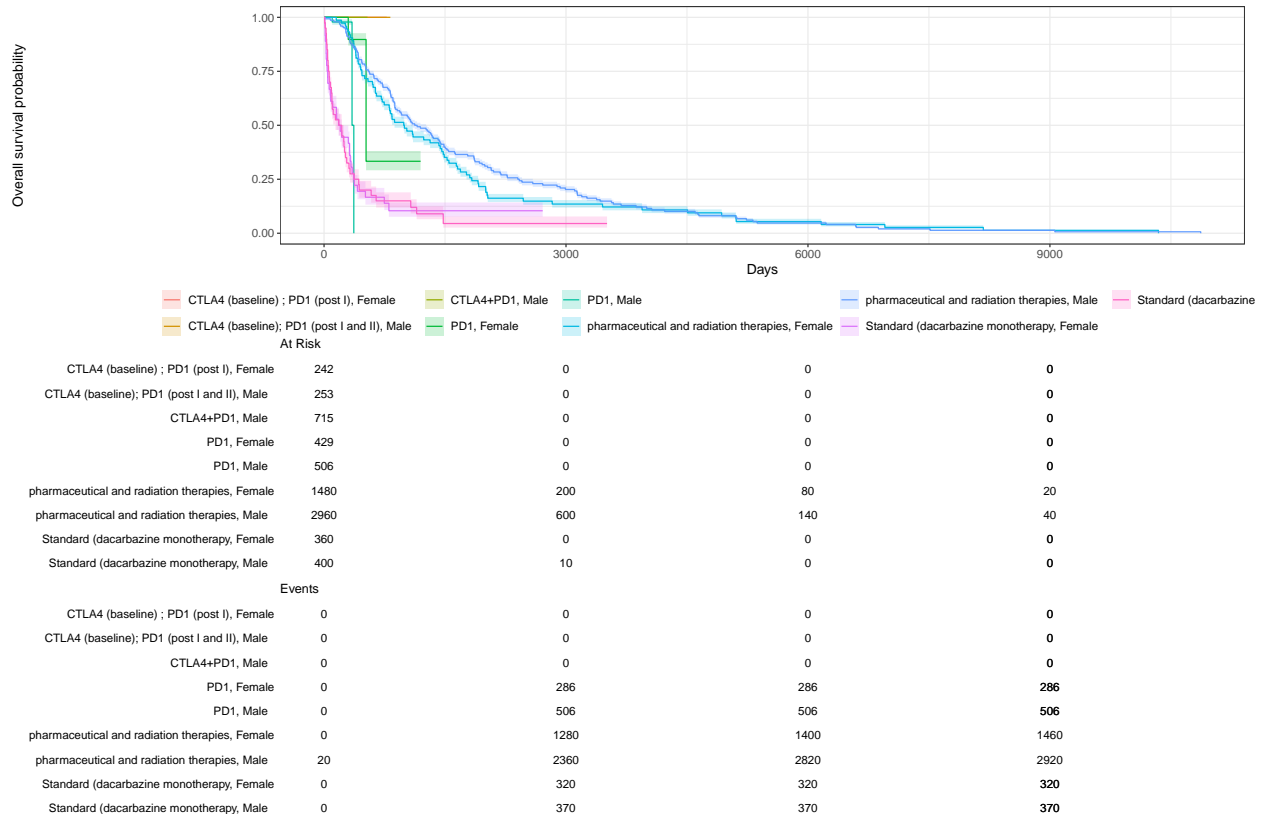
- Todas las terapias con información disponible: Se aprecian evoluciones diferentes.

```
## List of 18
## $ n      : int [1:9] 242 253 715 429 506 1480 2960 360 400
## $ time    : num [1:299] 777 822 339 388 539 ...
## $ n.risk  : num [1:299] 242 253 715 473 242 429 385 143 506 495 ...
## $ n.event : num [1:299] 0 0 0 0 0 44 242 0 11 242 ...
```

Characteristic	HR ¹	95% CI ¹	p-value
treatment			
CTLA4 (baseline) ; PD1 (post I)	—	—	
CTLA4 (baseline); PD1 (post I and II)	1.00	0.00, Inf	>0.9
CTLA4+PD1	1.09	0.00, Inf	>0.9
PD1	1,518,155,284	0.00, Inf	>0.9

¹HR = Hazard Ratio, CI = Confidence Interval

```
## $ n.censor : num [1:299] 242 253 242 231 242 0 0 143 0 0 ...
## $ surv      : num [1:299] 1 1 1 1 1 ...
## $ std.err    : num [1:299] 0 0 0 0 0 ...
## $ cumhaz     : num [1:299] 0 0 0 0 0 ...
## $ std.chaz   : num [1:299] 0 0 0 0 0 ...
## $ strata     : Named int [1:9] 1 1 3 3 3 73 144 33 38
##   .. attr(*, "names")= chr [1:9] "treatment=CTLA4 (baseline) ; PD1 (post I), sex=Female" "treatment=CTLA4 (baseline); PD1 (post I and II), sex=Male" "treatment=CTLA4+PD1, sex=Female" "treatment=PD1, sex=Female" "treatment=PD1, sex=Male" "treatment=pharmaceutical and radiation therapies, sex=Female" "treatment=pharmaceutical and radiation therapies, sex=Male" "treatment=Standard (dacarbazine monotherapy), sex=Female" "treatment=Standard (dacarbazine monotherapy), sex=Male"
## $ type       : chr "right"
## $ logse      : logi TRUE
## $ conf.int    : num 0.95
## $ conf.type   : chr "log"
## $ lower       : num [1:299] 1 1 1 1 1 ...
## $ upper       : num [1:299] 1 1 1 1 1 ...
## $ t0         : num 0
## $ call        : language survfit(formula = Surv(os_days, os_event) ~ treatment + sex, data = df2)
## - attr(*, "class")= chr "survfit"
```



Characteristic	HR ¹	95% CI ¹	p-value
treatment			
CTLA4 (baseline) ; PD1 (post I)	—	—	
CTLA4 (baseline); PD1 (post I and II)	0.90	0.00, Inf	>0.9
CTLA4+PD1	0.90	0.00, Inf	>0.9
PD1	45,537,977	0.00, Inf	>0.9
pharmaceutical and radiation therapies	15,554,361	0.00, Inf	>0.9
Standard (dacarbazine monotherapy)	73,682,760	0.00, Inf	>0.9
sex			
Female	—	—	
Male	1.12	1.06, 1.19	<0.001

¹HR = Hazard Ratio, CI = Confidence Interval

La apariencia de los gráficos mejora por permitir observar todos los tratamientos con más claridad, y confirman la curva de hombres como significativa:

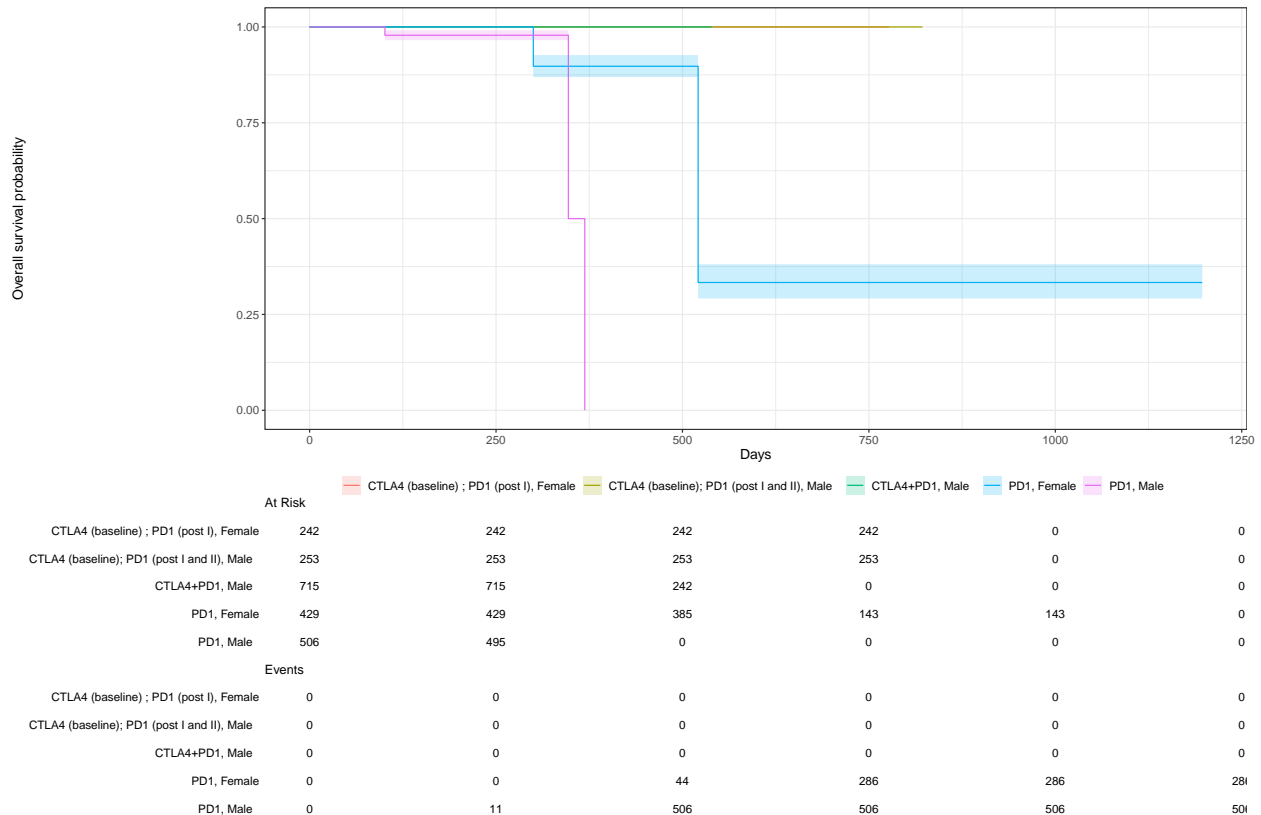
```
## Call:
## survival::coxph(formula = Surv(os_days, os_event) ~ treatment +
##     sex, data = df2)
##
##               coef exp(coef)
## treatmentCTLA4 (baseline); PD1 (post I and II) -1.058e-01 8.996e-01
## treatmentCTLA4+PD1                             -1.078e-01 8.978e-01
## treatmentPD1                                     1.763e+01 4.554e+07
## treatmentpharmaceutical and radiation therapies 1.656e+01 1.555e+07
## treatmentStandard (dacarbazine monotherapy)    1.812e+01 7.368e+07
## sexMale                                         1.149e-01 1.122e+00
##               se(coef)      z      p
## treatmentCTLA4 (baseline); PD1 (post I and II) 5.147e+02 0.000 1.000
## treatmentCTLA4+PD1                             4.864e+02 0.000 1.000
## treatmentPD1                                     3.759e+02 0.047 0.963
## treatmentpharmaceutical and radiation therapies 3.759e+02 0.044 0.965
## treatmentStandard (dacarbazine monotherapy)    3.759e+02 0.048 0.962
## sexMale                                         2.848e-02 4.035 5.46e-05
##
## Likelihood ratio test=2418 on 6 df, p=< 2.2e-16
## n= 7345, number of events= 5922
```

- Inmunoterapias (estudio GSE120575): el tratamiento con PD1 no parece tener eficacia en hombres y muy baja en mujeres.

```
## List of 18
## $ n      : int [1:5] 242 253 715 429 506
## $ time   : num [1:11] 777 822 339 388 539 ...
## $ n.risk  : num [1:11] 242 253 715 473 242 429 385 143 506 495 ...
## $ n.event : num [1:11] 0 0 0 0 0 44 242 0 11 242 ...
## $ n.censor : num [1:11] 242 253 242 231 242 0 0 143 0 0 ...
## $ surv    : num [1:11] 1 1 1 1 1 ...
## $ std.err  : num [1:11] 0 0 0 0 0 ...
## $ cumhaz   : num [1:11] 0 0 0 0 0 ...
## $ std.chaz : num [1:11] 0 0 0 0 0 ...
```



```
## $ strata : Named int [1:5] 1 1 3 3 3
## .. attr(*, "names")= chr [1:5] "treatment=CTLA4 (baseline) ; PD1 (post I), sex=Female" "treatment=
## $ type : chr "right"
## $ logse : logi TRUE
## $ conf.int : num 0.95
## $ conf.type: chr "log"
## $ lower : num [1:11] 1 1 1 1 1 ...
## $ upper : num [1:11] 1 1 1 1 1 ...
## $ t0 : num 0
## $ call : language survfit(formula = Surv(os_days, os_event) ~ treatment + sex, data = df3)
## - attr(*, "class")= chr "survfit"
```



La apariencia de los gráficos mejora por permitir observar todos los tratamientos con más claridad, y confirman la curva de hombres como significativa:

```
## Call:
## survival::coxph(formula = Surv(os_days, os_event) ~ treatment +
##     sex, data = df3)
##
##               coef exp(coef) se(coef)
## treatmentCTLA4 (baseline); PD1 (post I and II) -3.522e+00 2.953e-02 2.343e+03
## treatmentCTLA4+PD1 -3.342e+00 3.535e-02 2.221e+03
## treatmentPD1 2.043e+01 7.483e+08 1.675e+03
## sexMale 3.522e+00 3.386e+01 1.732e-01
##
##               z p
## treatmentCTLA4 (baseline); PD1 (post I and II) -0.002 0.999
## treatmentCTLA4+PD1 -0.002 0.999
## treatmentPD1 0.012 0.990
## sexMale 20.337 <2e-16
```

Characteristic	HR [†]	95% CI [†]	p-value
treatment			
CTLA4 (baseline) ; PD1 (post I)	—	—	
CTLA4 (baseline); PD1 (post I and II)	0.03	0.00, Inf	>0.9
CTLA4+PD1	0.04	0.00, Inf	>0.9
PD1	748,320,917	0.00, Inf	>0.9
sex			
Female	—	—	
Male	33.9	24.1, 47.5	<0.001

[†]HR = Hazard Ratio, CI = Confidence Interval

##

Likelihood ratio test=2544 on 4 df, p=< 2.2e-16

n= 2145, number of events= 792

- Estudio GSE22155: Los pacientes correspondientes a la plataforma GPL6102 sobreviven más tiempo pero el tratamiento tiene baja eficacia en ambos sexos.

List of 18

\$ n : int [1:4] 250 110 290 110

\$ time : num [1:73] 10 13 16 27 45 46 64 78 82 108 ...

\$ n.risk : num [1:73] 250 240 230 220 210 200 190 180 170 160 ...

\$ n.event : num [1:73] 10 10 10 10 10 10 10 10 10 10 ...

\$ n.censor : num [1:73] 0 0 0 0 0 0 0 0 0 0 ...

\$ surv : num [1:73] 0.96 0.92 0.88 0.84 0.8 0.76 0.72 0.68 0.64 0.6 ...

\$ std.err : num [1:73] 0.0129 0.0187 0.0234 0.0276 0.0316 ...

\$ cumhaz : num [1:73] 0.04 0.0817 0.1251 0.1706 0.2182 ...

\$ std.chaz : num [1:73] 0.0126 0.0183 0.0229 0.027 0.0309 ...

\$ strata : Named int [1:4] 24 10 29 10

\$.- attr(*, "names")= chr [1:4] "treatment=Standard (dacarbazine monotherapy, sex=Female, study=GSE22155)"

\$ type : chr "right"

\$ logse : logi TRUE

\$ conf.int : num 0.95

\$ conf.type: chr "log"

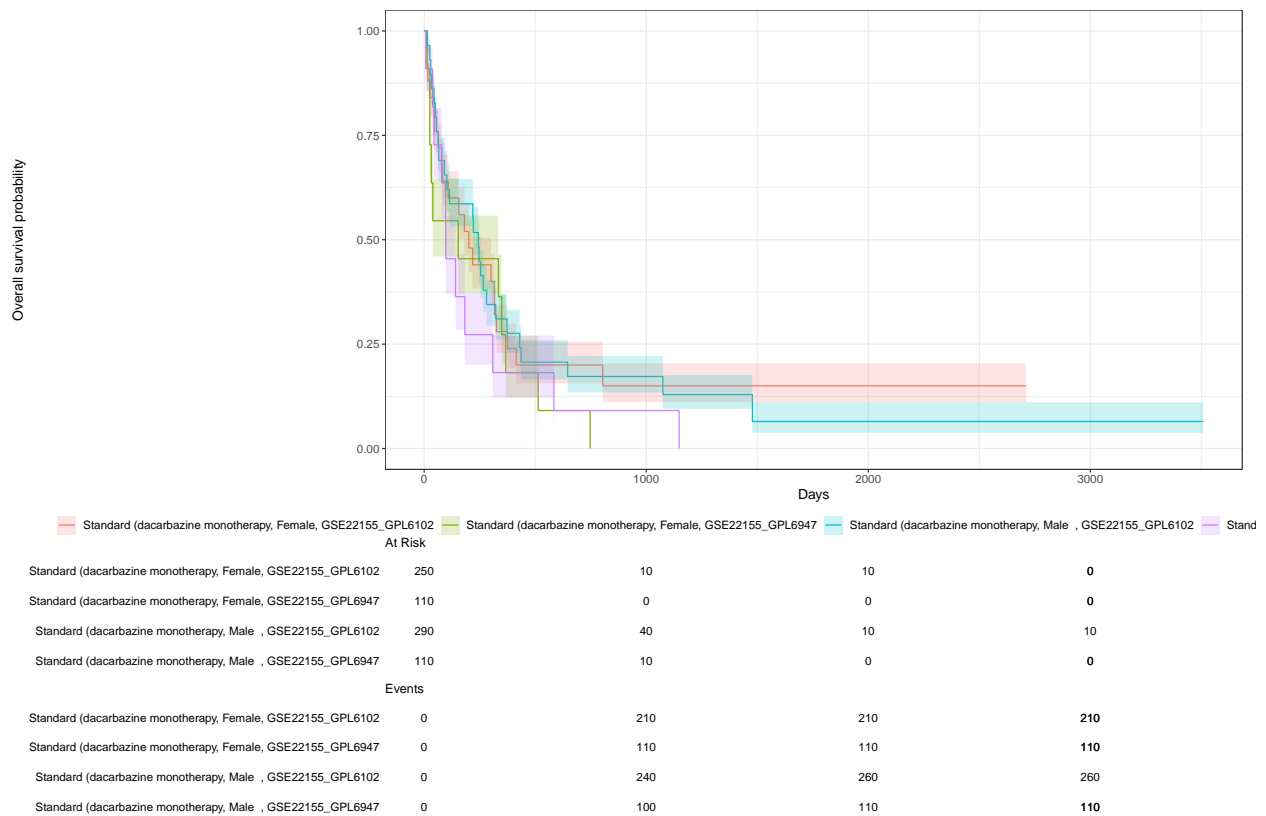
\$ lower : num [1:73] 0.936 0.887 0.841 0.796 0.752 ...

\$ upper : num [1:73] 0.985 0.954 0.921 0.887 0.851 ...

\$ t0 : num 0

\$ call : language survfit(formula = Surv(os_days, os_event) ~ treatment + sex + study, data = d)

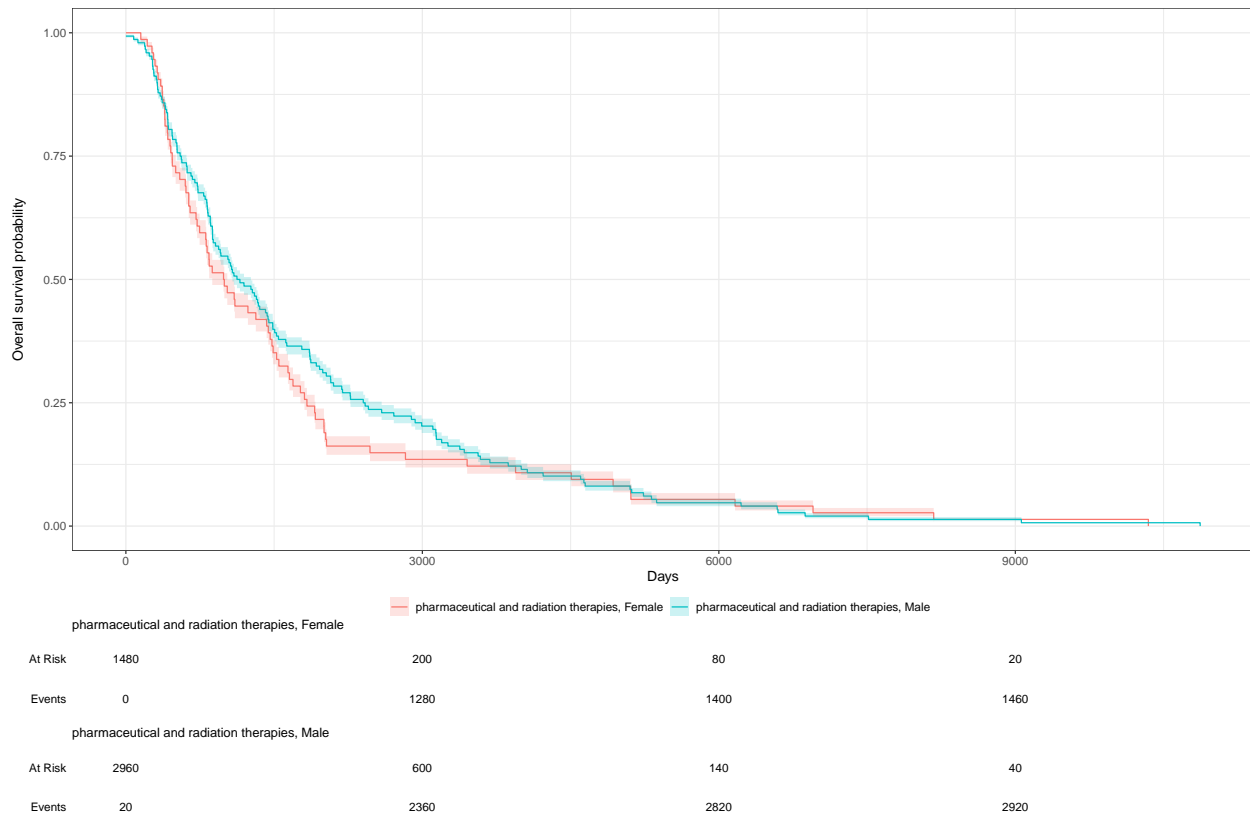
\$.- attr(*, "class")= chr "survfit"



La apariencia de los gráficos mejora por permitir observar todos los tratamientos con más claridad, y da error por no tener suficientes niveles para comparar:

- Estudio tcga_skcm: Los pacientes tienen mal pronóstico con este tratamiento, tanto si se es hombre o mujer.

```
## List of 18
## $ n      : int [1:2] 1480 2960
## $ time   : num [1:217] 151 216 264 279 295 317 326 352 368 369 ...
## $ n.risk : num [1:217] 1480 1460 1440 1420 1400 1380 1360 1340 1320 1300 ...
## $ n.event : num [1:217] 20 20 20 20 20 20 20 20 20 20 ...
## $ n.censor : num [1:217] 0 0 0 0 0 0 0 0 0 0 ...
## $ surv    : num [1:217] 0.986 0.973 0.959 0.946 0.932 ...
## $ std.err : num [1:217] 0.00304 0.00433 0.00534 0.00621 0.007 ...
## $ cumhaz  : num [1:217] 0.0135 0.0272 0.0411 0.0552 0.0695 ...
## $ std.chaz : num [1:217] 0.00302 0.0043 0.00531 0.00617 0.00695 ...
## $ strata  : Named int [1:2] 73 144
## .. attr(*, "names")= chr [1:2] "treatment=pharmaceutical and radiation therapies, sex=Female" "tr
## $ type    : chr "right"
## $ logse   : logi TRUE
## $ conf.int : num 0.95
## $ conf.type: chr "log"
## $ lower    : num [1:217] 0.981 0.965 0.949 0.934 0.92 ...
## $ upper    : num [1:217] 0.992 0.981 0.97 0.958 0.945 ...
## $ t0       : num 0
## $ call     : language survfit(formula = Surv(os_days, os_event) ~ treatment + sex, data = df3)
## - attr(*, "class")= chr "survfit"
```

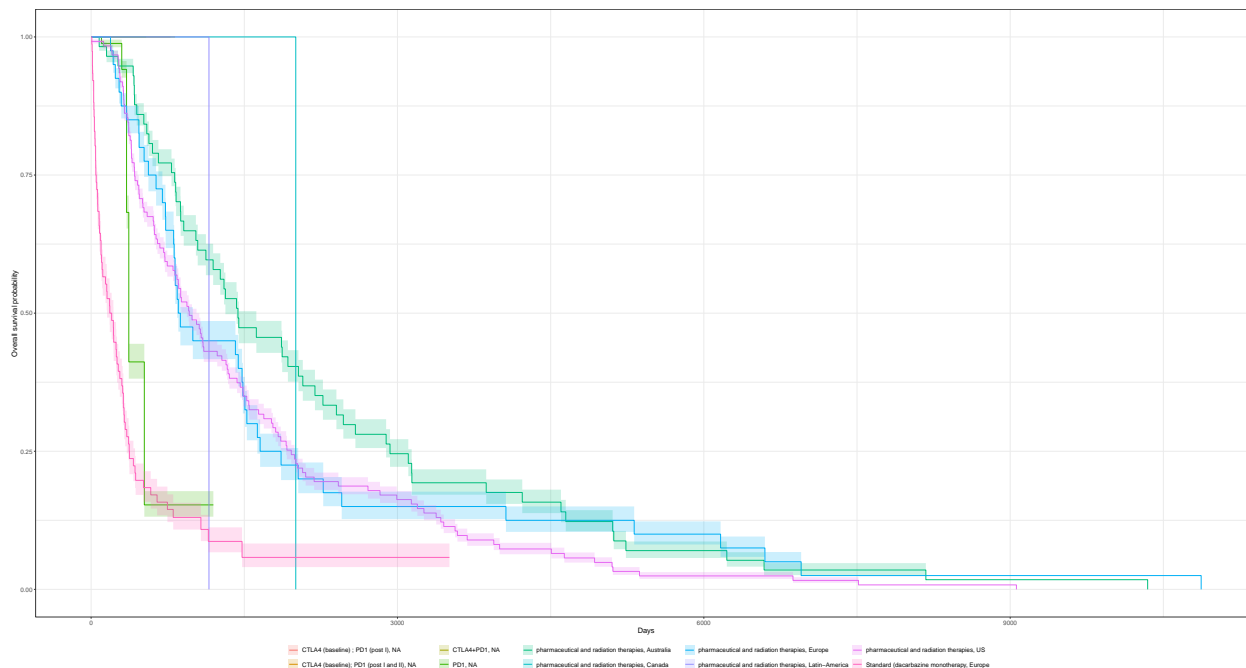


La apariencia de los gráficos mejora por permitir observar todos los tratamientos con más claridad, y da error por not tener suficientes niveles para comparar:

3.4.2 País de origen

Dado que varios autores han puesto el centro de investigación en lugar del país (hay estudios que en los que han colaborado centros de diversos países), lo harmonizo con una nueva variable para los continentes.

```
## List of 18
## $ n      : int [1:10] 242 253 715 935 1140 20 800 20 2460 760
## $ time   : num [1:298] 777 822 339 388 539 101 300 347 369 521 ...
## $ n.risk  : num [1:298] 242 253 715 473 242 935 924 880 638 385 ...
## $ n.event : num [1:298] 0 0 0 0 0 11 44 242 253 242 ...
## $ n.censor : num [1:298] 242 253 242 231 242 0 0 0 0 0 ...
## $ surv    : num [1:298] 1 1 1 1 1 ...
## $ std.err  : num [1:298] 0 0 0 0 0 ...
## $ cumhaz   : num [1:298] 0 0 0 0 0 ...
## $ std.chaz : num [1:298] 0 0 0 0 0 ...
## $ strata   : Named int [1:10] 1 1 3 6 57 1 40 1 119 69
##   .. attr(*, "names")= chr [1:10] "treatment=CTLA4 (baseline) ; PD1 (post I), region=NA
## $ type     : chr "right"
## $ logse    : logi TRUE
## $ conf.int : num 0.95
## $ conf.type: chr "log"
## $ lower    : num [1:298] 1 1 1 1 1 ...
## $ upper    : num [1:298] 1 1 1 1 1 ...
## $ t0       : num 0
## $ call     : language survfit(formula = Surv(os_days, os_event) ~ treatment + region, data = df2)
## - attr(*, "class")= chr "survfit"
```

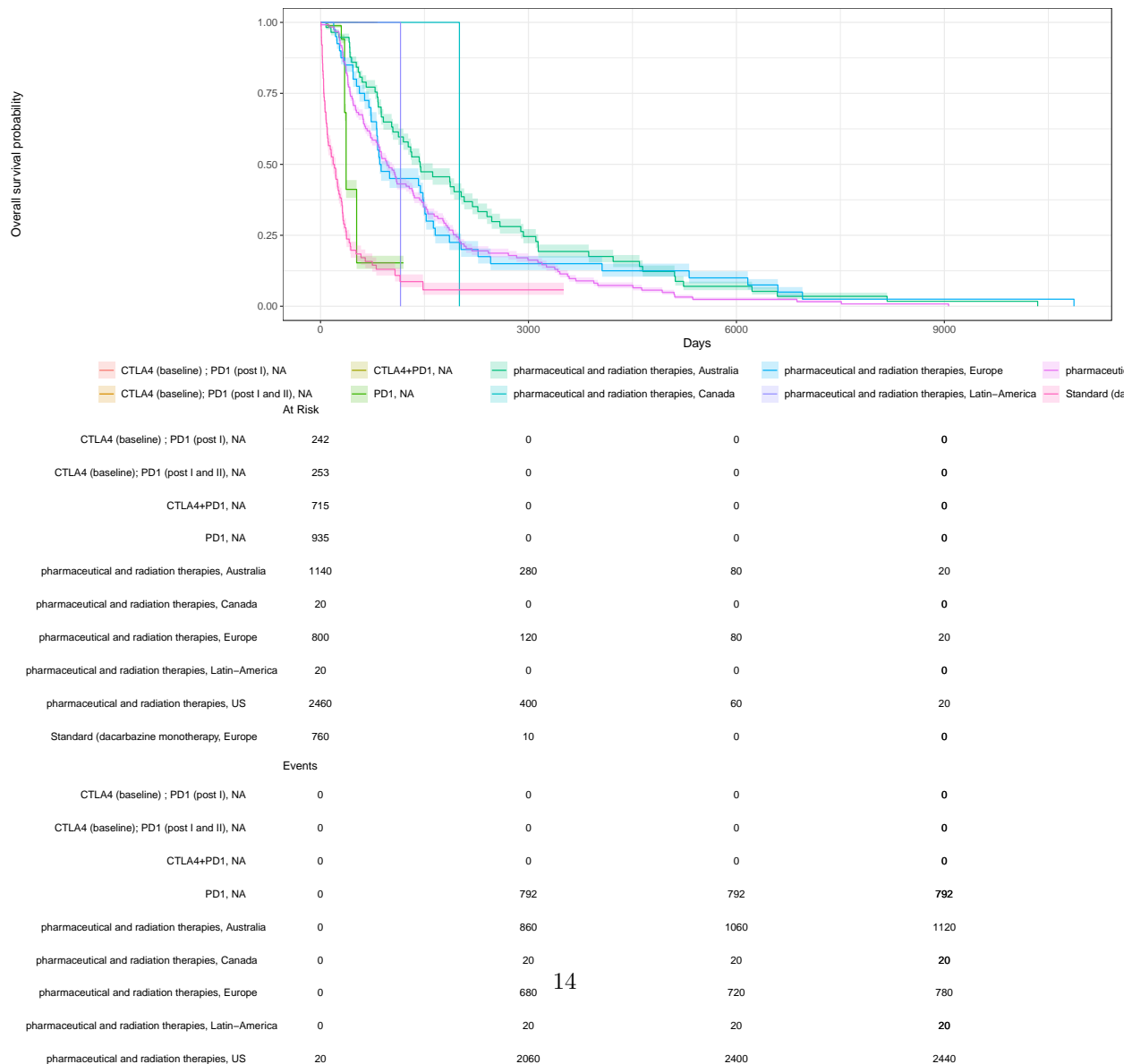


La apariencia de los gráficos mejora por permitir observar todos los tratamientos con más claridad, y sugieren las regiones como un factor que cambiaría significativamente:

```
## Call:
## survival::coxph(formula = Surv(os_days, os_event) ~ treatment +
##     region, data = df2)
##
##
##               coef exp(coef)
## treatmentCTLA4 (baseline); PD1 (post I and II)  8.831e-03  1.009e+00
## treatmentCTLA4+PD1                             8.342e-03  1.008e+00
## treatmentPD1                                    1.767e+01  4.722e+07
## treatmentpharmaceutical and radiation therapies  1.642e+01  1.356e+07
## treatmentStandard (dacarbazine monotherapy)     1.803e+01  6.775e+07
## regionCanada                                   -1.489e-02  9.852e-01
## regionEurope                                   1.475e-01  1.159e+00
## regionLatin-America                           5.239e-01  1.689e+00
## regionNA                                       NA          NA
## regionUS                                       3.373e-01  1.401e+00
##
##               se(coef)      z      p
## treatmentCTLA4 (baseline); PD1 (post I and II)  5.151e+02  0.000  0.99999
## treatmentCTLA4+PD1                             4.865e+02  0.000  0.99999
## treatmentPD1                                    3.762e+02  0.047  0.96253
## treatmentpharmaceutical and radiation therapies  3.762e+02  0.044  0.96518
## treatmentStandard (dacarbazine monotherapy)     3.762e+02  0.048  0.96177
## regionCanada                                   2.259e-01 -0.066  0.94743
## regionEurope                                   4.675e-02  3.154  0.00161
## regionLatin-America                           2.261e-01  2.317  0.02048
## regionNA                                       0.000e+00    NA    NA
## regionUS                                       3.605e-02  9.356 < 2e-16
##
## Likelihood ratio test=2498 on 9 df, p=< 2.2e-16
## n= 7345, number of events= 5922
```

Characteristic	HR [†]	95% CI [†]	p-value
treatment			
CTLA4 (baseline) ; PD1 (post I)	—	—	
CTLA4 (baseline); PD1 (post I and II)	1.01	0.00, Inf	>0.9
CTLA4+PD1	1.01	0.00, Inf	>0.9
PD1	47,217,761	0.00, Inf	>0.9
pharmaceutical and radiation therapies	13,562,682	0.00, Inf	>0.9
Standard (dacarbazine monotherapy	67,746,205	0.00, Inf	>0.9
region			
Australia	—	—	
Canada	0.99	0.63, 1.53	>0.9
Europe	1.16	1.06, 1.27	0.002
Latin-America	1.69	1.08, 2.63	0.020
NA			
US	1.40	1.31, 1.50	<0.001

[†]HR = Hazard Ratio, CI = Confidence Interval



Characteristic	HR [†]	95% CI [†]	p-value
treatment			
CTLA4 (baseline) ; PD1 (post I)	—	—	
CTLA4 (baseline); PD1 (post I and II)	1.01	0.00, Inf	>0.9
CTLA4+PD1	1.01	0.00, Inf	>0.9
PD1	47,217,761	0.00, Inf	>0.9
pharmaceutical and radiation therapies	13,562,682	0.00, Inf	>0.9
Standard (dacarbazine monotherapy)	67,746,205	0.00, Inf	>0.9
region			
Australia	—	—	
Canada	0.99	0.63, 1.53	>0.9
Europe	1.16	1.06, 1.27	0.002
Latin-America	1.69	1.08, 2.63	0.020
NA			
US	1.40	1.31, 1.50	<0.001

[†]HR = Hazard Ratio, CI = Confidence Interval

La apariencia de los gráficos mejora por permitir observar todos los tratamientos con más claridad, y las regiones cambian de manera significativa:

```
## Call:
## survival::coxph(formula = Surv(os_days, os_event) ~ treatment +
##     region, data = df2)
##
##               coef exp(coef)
## treatmentCTLA4 (baseline); PD1 (post I and II) 8.831e-03 1.009e+00
## treatmentCTLA4+PD1                             8.342e-03 1.008e+00
## treatmentPD1                                   1.767e+01 4.722e+07
## treatmentpharmaceutical and radiation therapies 1.642e+01 1.356e+07
## treatmentStandard (dacarbazine monotherapy     1.803e+01 6.775e+07
## regionCanada                                   -1.489e-02 9.852e-01
## regionEurope                                   1.475e-01 1.159e+00
## regionLatin-America                           5.239e-01 1.689e+00
## regionNA                                       NA          NA
## regionUS                                       3.373e-01 1.401e+00
##               se(coef)      z      p
## treatmentCTLA4 (baseline); PD1 (post I and II) 5.151e+02 0.000 0.99999
## treatmentCTLA4+PD1                             4.865e+02 0.000 0.99999
## treatmentPD1                                   3.762e+02 0.047 0.96253
## treatmentpharmaceutical and radiation therapies 3.762e+02 0.044 0.96518
## treatmentStandard (dacarbazine monotherapy     3.762e+02 0.048 0.96177
## regionCanada                                   2.259e-01 -0.066 0.94743
## regionEurope                                   4.675e-02 3.154 0.00161
## regionLatin-America                           2.261e-01 2.317 0.02048
## regionNA                                       0.000e+00    NA    NA
## regionUS                                       3.605e-02 9.356 < 2e-16
##
## Likelihood ratio test=2498 on 9 df, p=< 2.2e-16
## n= 7345, number of events= 5922
```

Si nos centramos por cada región:

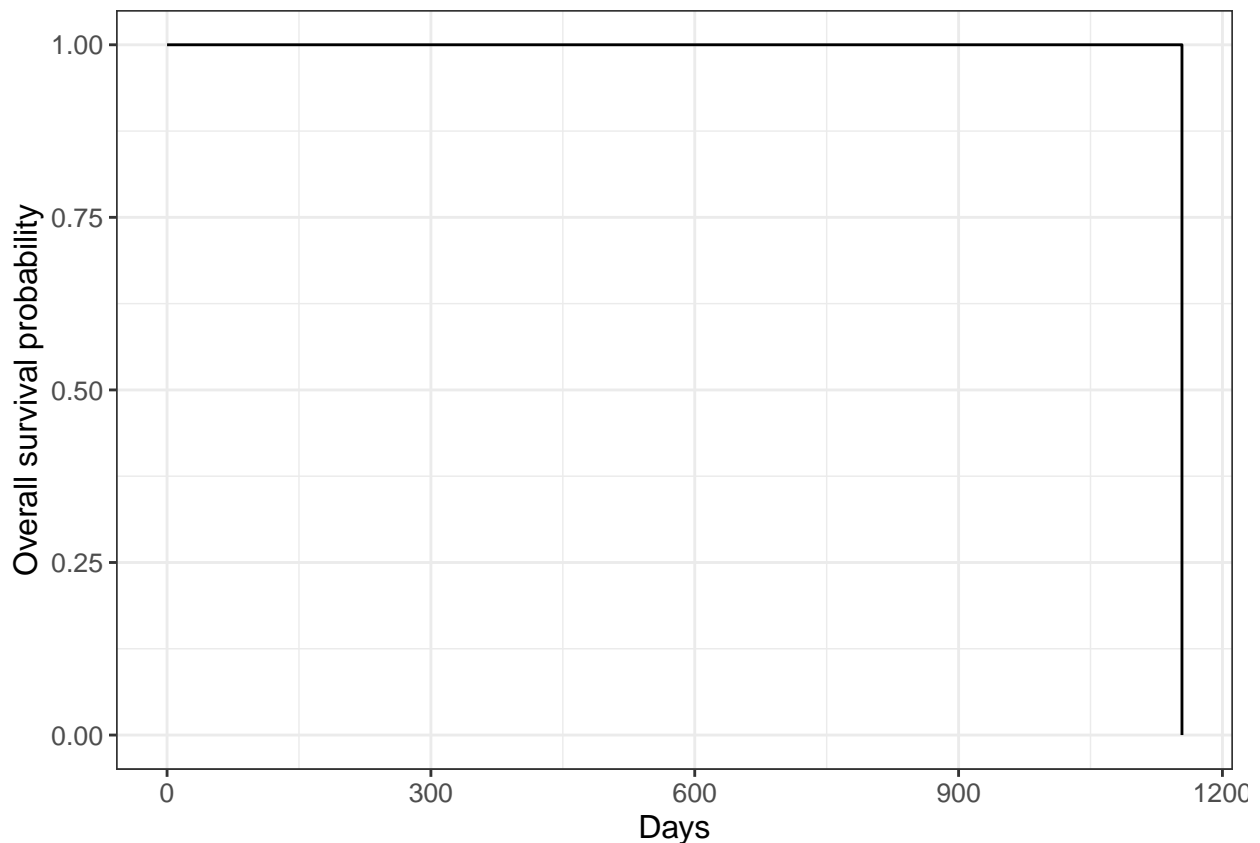
- Latino-América: Sólo hay un punto de seguimiento en el que todos los pacientes fallecieron.

```
#unique(df2$region)
#[1] "NA" "Australia" "US" "Europe" "Canada" "Latin-America"
df3 <- df2[df2$region == "Latin-America",]

library(ggsurvfit)
# https://www.emilyzabor.com/tutorials/survival_analysis_in_r_tutorial.html
s1 <- survfit(Surv(os_days, os_event) ~ treatment, data = df3) # Creación de un survival object
str(s1)

## List of 17
## $ n : int 20
## $ time : num 1154
## $ n.risk : num 20
## $ n.event : num 20
## $ n.censor : num 0
## $ surv : num 0
## $ std.err : num Inf
## $ cumhaz : num 1
## $ std.chaz : num 0.224
## $ type : chr "right"
## $ logse : logi TRUE
## $ conf.int : num 0.95
## $ conf.type: chr "log"
## $ lower : num NA
## $ upper : num NA
## $ t0 : num 0
## $ call : language survfit(formula = Surv(os_days, os_event) ~ treatment, data = df3)
## - attr(*, "class")= chr "survfit"

# Kaplan-Meier plots:
survfit2(Surv(os_days, os_event) ~ treatment, data = df3) %>%
  ggsurvfit() +
  labs(
    x = "Days",
    y = "Overall survival probability"
  ) +
  add_confidence_interval() #+ # Para añadir intervalo de confianza en las curvas
```

```
# add_risktable() # Tablas de pacientes supervivientes y eventos en cada intervalo contemplado
```

La apariencia de los gráficos mejora por permitir observar todos los tratamientos con más claridad, y no da resultado por no tener suficientes niveles:

- Canada: Sólo hay un punto de seguimiento en el que todos los pacientes fallecieron.

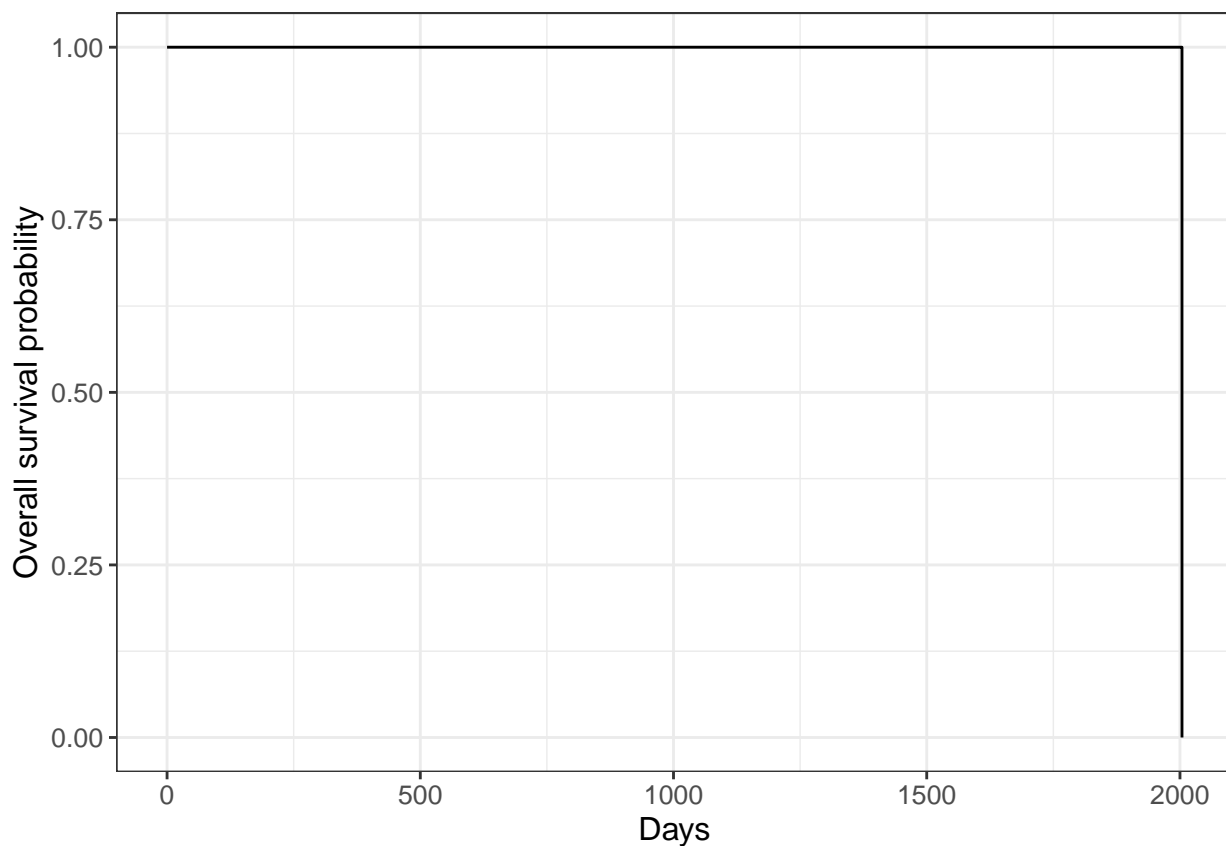
```
#unique(df2$region)
#[1] "NA" "Australia" "US" "Europe" "Canada" "Latin-America"
df3 <- df2[df2$region == "Canada",]
```

```
library(ggsurvfit)
# https://www.emilyzabor.com/tutorials/survival_analysis_in_r_tutorial.html
s1 <- survfit(Surv(os_days, os_event) ~ treatment, data = df3) # Creación de un survival object
str(s1)
```

```
## List of 17
## $ n      : int 20
## $ time   : num 2004
## $ n.risk  : num 20
## $ n.event : num 20
## $ n.censor : num 0
## $ surv    : num 0
## $ std.err  : num Inf
## $ cumhaz   : num 1
## $ std.chaz : num 0.224
## $ type     : chr "right"
## $ logse    : logi TRUE
## $ conf.int : num 0.95
```

```
## $ conf.type: chr "log"
## $ lower : num NA
## $ upper : num NA
## $ t0 : num 0
## $ call : language survfit(formula = Surv(os_days, os_event) ~ treatment, data = df3)
## - attr(*, "class")= chr "survfit"
```

```
# Kaplan-Meier plots:
survfit2(Surv(os_days, os_event) ~ treatment, data = df3) %>%
  ggsurvfit() +
  labs(
    x = "Days",
    y = "Overall survival probability"
  ) +
  add_confidence_interval() #+ # Para añadir intervalo de confianza en las curvas
```



```
# add_risktable() # Tablas de pacientes supervivientes y eventos en cada intervalo contemplado
```

La apariencia de los gráficos mejora por permitir observar todos los tratamientos con más claridad, y no da resultado por no tener suficientes niveles:

- Europa: Hay más supervivencia dependiendo más del centro de investigación que de la terapia en sí, o del sexo del paciente.

```
#unique(df2$region)
#[1] "NA" "Australia" "US" "Europe" "Canada" "Latin-America"
df3 <- df2[df2$region == "Europe",]

library(ggsurvfit)
```

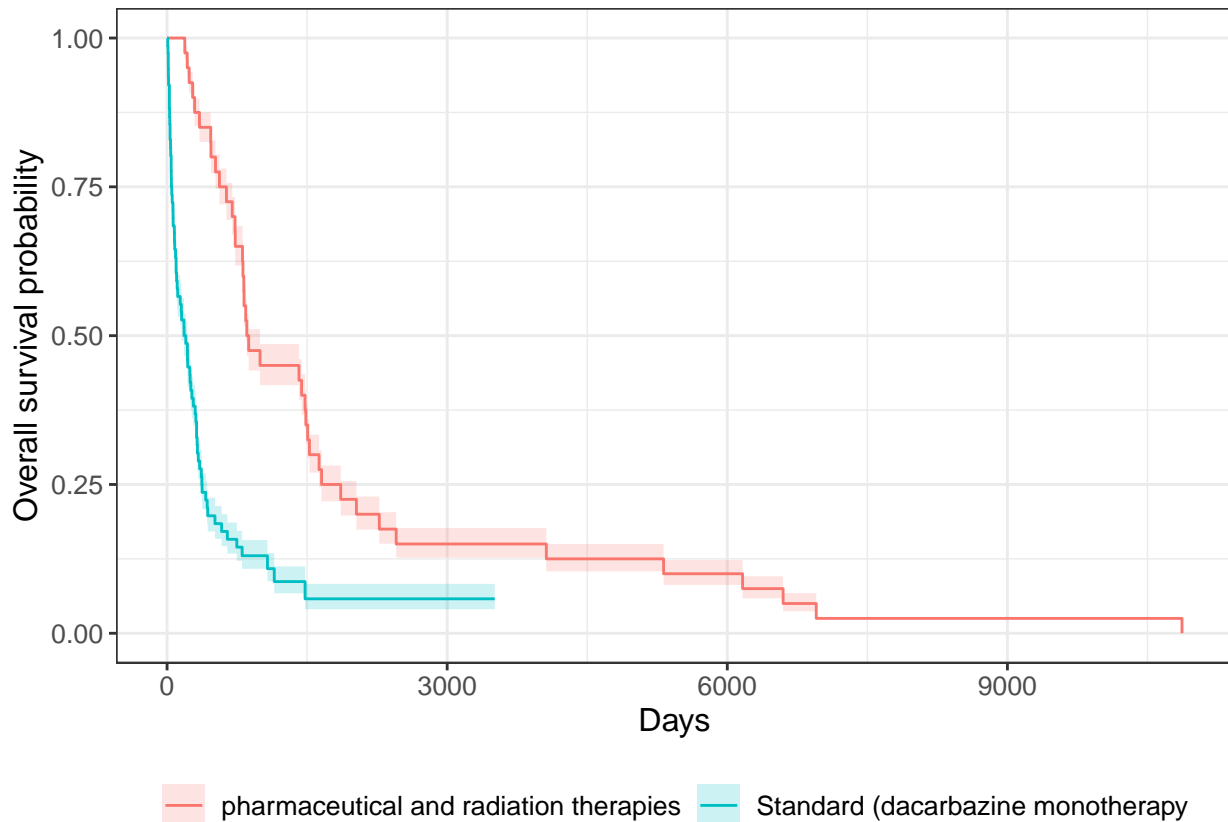
```

# https://www.emilyzabor.com/tutorials/survival\_analysis\_in\_r\_tutorial.html
s1 <- survfit(Surv(os_days, os_event) ~ treatment, data = df3) # Creación de un survival object
str(s1)

## List of 18
## $ n      : int [1:2] 800 760
## $ time   : num [1:109] 190 216 237 274 295 347 468 470 519 561 ...
## $ n.risk  : num [1:109] 800 780 760 740 720 700 680 660 640 620 ...
## $ n.event : num [1:109] 20 20 20 20 20 20 20 20 20 20 ...
## $ n.censor : num [1:109] 0 0 0 0 0 0 0 0 0 0 ...
## $ surv    : num [1:109] 0.975 0.95 0.925 0.9 0.875 0.85 0.825 0.8 0.775 0.75 ...
## $ std.err  : num [1:109] 0.00566 0.00811 0.01007 0.01179 0.01336 ...
## $ cumhaz   : num [1:109] 0.025 0.0506 0.077 0.104 0.1318 ...
## $ std.chaz : num [1:109] 0.00559 0.00801 0.00994 0.01163 0.01319 ...
## $ strata   : Named int [1:2] 40 69
## ..- attr(*, "names")= chr [1:2] "treatment=pharmaceutical and radiation therapies" "treatment=Stan
## $ type      : chr "right"
## $ logse     : logi TRUE
## $ conf.int  : num 0.95
## $ conf.type : chr "log"
## $ lower     : num [1:109] 0.964 0.935 0.907 0.879 0.852 ...
## $ upper     : num [1:109] 0.986 0.965 0.943 0.921 0.898 ...
## $ t0        : num 0
## $ call      : language survfit(formula = Surv(os_days, os_event) ~ treatment, data = df3)
## - attr(*, "class")= chr "survfit"

# Kaplan-Meier plots:
survfit2(Surv(os_days, os_event) ~ treatment, data = df3) %>%
  ggsurvfit() +
  labs(
    x = "Days",
    y = "Overall survival probability"
  ) +
  add_confidence_interval() #+ # Para añadir intervalo de confianza en las curvas

```

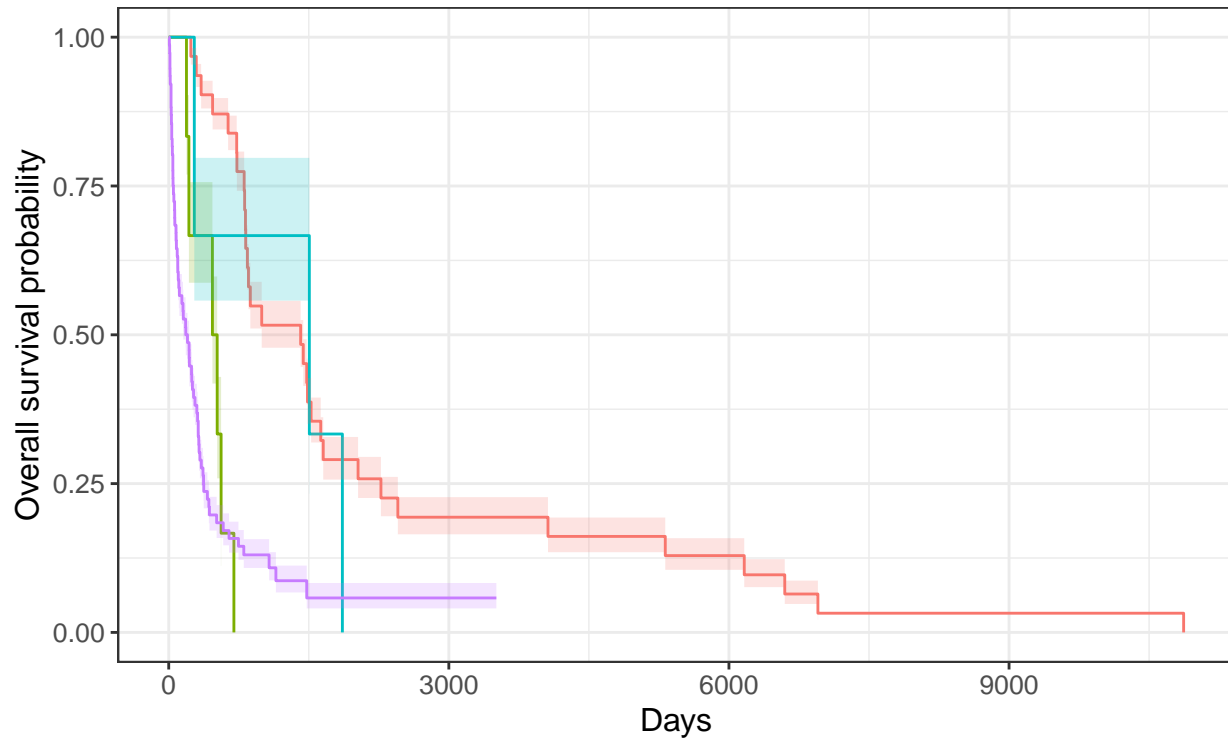


```
# add_risktable() # Tablas de pacientes supervivientes y eventos en cada intervalo contemplado
```

```
s1 <- survfit(Surv(os_days, os_event) ~ treatment + contact_country, data = df3) # Creación de un survi
str(s1)
```

```
## List of 18
## $ n      : int [1:4] 620 120 60 760
## $ time   : num [1:109] 237 295 347 470 636 728 730 808 813 822 ...
## $ n.risk : num [1:109] 620 600 580 560 540 520 500 480 460 440 ...
## $ n.event : num [1:109] 20 20 20 20 20 20 20 20 20 20 ...
## $ n.censor : num [1:109] 0 0 0 0 0 0 0 0 0 0 ...
## $ surv    : num [1:109] 0.968 0.935 0.903 0.871 0.839 ...
## $ std.err  : num [1:109] 0.00733 0.01055 0.01315 0.01546 0.01761 ...
## $ cumhaz   : num [1:109] 0.0323 0.0656 0.1001 0.1358 0.1728 ...
## $ std.chaz : num [1:109] 0.00721 0.01037 0.01292 0.01519 0.0173 ...
## $ strata   : Named int [1:4] 31 6 3 69
## .. attr(*, "names")= chr [1:4] "treatment=pharmaceutical and radiation therapies, contact_country"
## $ type     : chr "right"
## $ logse    : logi TRUE
## $ conf.int : num 0.95
## $ conf.type: chr "log"
## $ lower    : num [1:109] 0.954 0.916 0.88 0.845 0.81 ...
## $ upper    : num [1:109] 0.982 0.955 0.927 0.898 0.868 ...
## $ t0       : num 0
## $ call     : language survfit(formula = Surv(os_days, os_event) ~ treatment + contact_country,
## - attr(*, "class")= chr "survfit"
```

```
# Kaplan-Meier plots:
survfit2(Surv(os_days, os_event) ~ treatment + contact_country, data = df3) %>%
  ggsurvfit() +
  labs(
    x = "Days",
    y = "Overall survival probability"
  ) +
  add_confidence_interval() #+ # Para añadir intervalo de confianza en las curvas
```



cal and radiation therapies, Greater Poland Cancer Center — pharmaceutical and radiation therapies, N

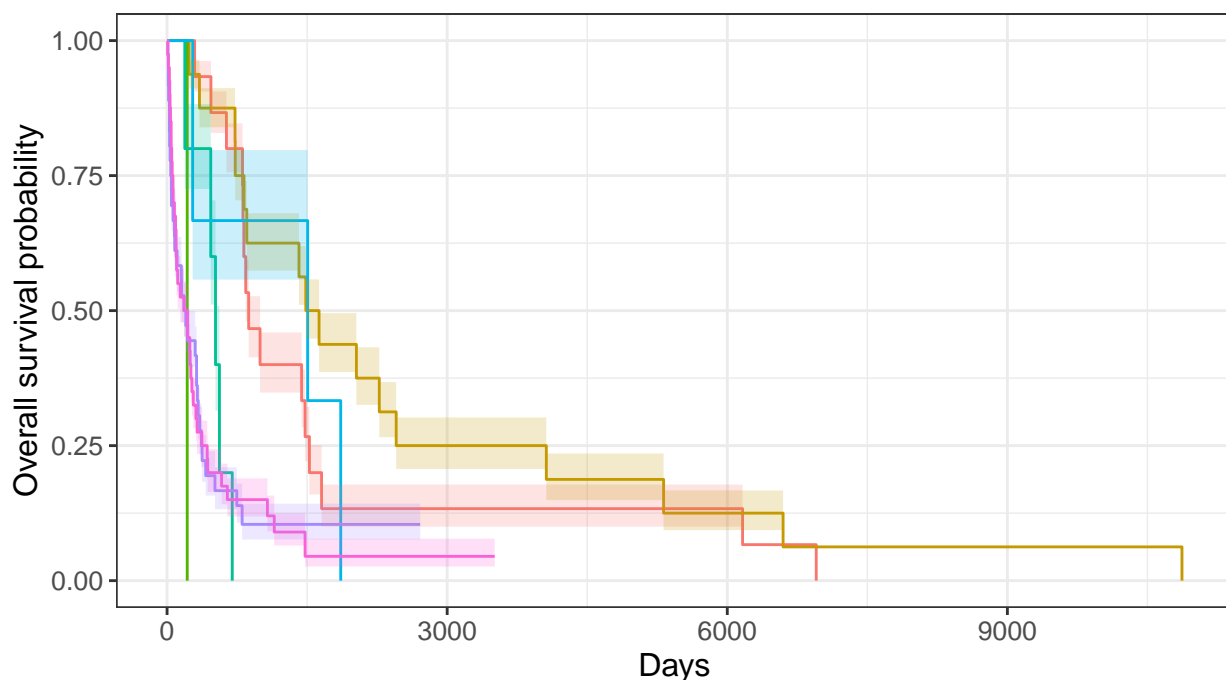
```
# add_risktable() # Tablas de pacientes supervivientes y eventos en cada intervalo contemplado
```

```
s1 <- survfit(Surv(os_days, os_event) ~ treatment + contact_country + sex, data = df3) # Creación de un
str(s1)
```

```
## List of 18
## $ n      : int [1:7] 300 320 20 100 60 360 400
## $ time    : num [1:111] 295 470 636 808 813 ...
## $ n.risk  : num [1:111] 300 280 260 240 220 200 180 160 140 120 ...
## $ n.event : num [1:111] 20 20 20 20 20 20 20 20 20 20 ...
## $ n.censor : num [1:111] 0 0 0 0 0 0 0 0 0 0 ...
## $ surv    : num [1:111] 0.933 0.867 0.8 0.733 0.667 ...
## $ std.err  : num [1:111] 0.0154 0.0226 0.0289 0.0348 0.0408 ...
## $ cumhaz   : num [1:111] 0.0667 0.1381 0.215 0.2984 0.3893 ...
## $ std.chaz : num [1:111] 0.0149 0.0218 0.0278 0.0335 0.0392 ...
## $ strata   : Named int [1:7] 15 16 1 5 3 33 38
## .. attr(*, "names")= chr [1:7] "treatment=pharmaceutical and radiation therapies, contact_country"
## $ type     : chr "right"
## $ logse    : logi TRUE
```

```
## $ conf.int : num 0.95
## $ conf.type: chr "log"
## $ lower    : num [1:111] 0.906 0.829 0.756 0.685 0.615 ...
## $ upper    : num [1:111] 0.962 0.906 0.847 0.785 0.722 ...
## $ t0       : num 0
## $ call      : language survfit(formula = Surv(os_days, os_event) ~ treatment + contact_country +
## - attr(*, "class")= chr "survfit"
```

```
# Kaplan-Meier plots:
survfit2(Surv(os_days, os_event) ~ treatment + contact_country + sex, data = df3) %>%
  ggsurvfit() +
  labs(
    x = "Days",
    y = "Overall survival probability"
  ) +
  add_confidence_interval() #+ # Para añadir intervalo de confianza en las curvas
```



radiation therapies, Greater Poland Cancer Center, Female — pharmaceutical and radiation therapies,
 radiation therapies, Greater Poland Cancer Center, Male — Standard (dacarbazine monotherapy, Sv

```
# add_risktable() # Tablas de pacientes supervivientes y eventos en cada intervalo contemplado
```

La apariencia de los gráficos mejora por permitir observar todos los tratamientos con más claridad, y obtenemos varios resultados significativos:

```
## Call:
## survival::coxph(formula = Surv(os_days, os_event) ~ treatment,
## data = df3)
##
##               coef exp(coef) se(coef)      z
## treatmentStandard (dacarbazine monotherapy 1.24366   3.46828  0.05742 21.66
##                                     p
```

Characteristic	HR ¹	95% CI ¹	p-value
treatment			
pharmaceutical and radiation therapies	—	—	
Standard (dacarbazine monotherapy)	3.47	3.10, 3.88	<0.001

¹HR = Hazard Ratio, CI = Confidence Interval

Characteristic	HR ¹	95% CI ¹	p-value
treatment			
pharmaceutical and radiation therapies	—	—	
Standard (dacarbazine monotherapy)	4.31	3.81, 4.87	<0.001
contact_country			
Essen	—	—	
Greater Poland Cancer Center	3.96	3.21, 4.89	<0.001
Norfolk and Norwich Hospital	1.35	1.03, 1.76	0.030
Sweden			

¹HR = Hazard Ratio, CI = Confidence Interval

```
## treatmentStandard (dacarbazine monotherapy <2e-16
##
## Likelihood ratio test=451.4 on 1 df, p=< 2.2e-16
## n= 1560, number of events= 1490
## Call:
## survival::coxph(formula = Surv(os_days, os_event) ~ treatment +
##   contact_country, data = df3)
##
##               coef exp(coef) se(coef)      z
## treatmentStandard (dacarbazine monotherapy 1.46103   4.31040  0.06271 23.298
## contact_countryGreater Poland Cancer Center 1.37627   3.96009  0.10785 12.761
## contact_countryNorfolk and Norwich Hospital 0.29868   1.34807  0.13732  2.175
## contact_countrySweden                      NA         NA  0.00000    NA
##
##               p
## treatmentStandard (dacarbazine monotherapy <2e-16
## contact_countryGreater Poland Cancer Center <2e-16
## contact_countryNorfolk and Norwich Hospital 0.0296
## contact_countrySweden                      NA
##
## Likelihood ratio test=580.8 on 3 df, p=< 2.2e-16
## n= 1560, number of events= 1490
## Call:
## survival::coxph(formula = Surv(os_days, os_event) ~ treatment +
##   contact_country + sex, data = df3)
##
##               coef exp(coef) se(coef)      z
## treatmentStandard (dacarbazine monotherapy 1.46026   4.30707  0.06263 23.316
## contact_countryGreater Poland Cancer Center 1.46265   4.31739  0.11039 13.249
## contact_countryNorfolk and Norwich Hospital 0.39149   1.47919  0.13973  2.802
```

Characteristic	HR ¹	95% CI ¹	p-value
treatment			
pharmaceutical and radiation therapies	—	—	
Standard (dacarbazine monotherapy)	4.31	3.81, 4.87	<0.001
contact_country			
Essen	—	—	
Greater Poland Cancer Center	4.32	3.48, 5.36	<0.001
Norfolk and Norwich Hospital	1.48	1.12, 1.95	0.005
Sweden			
sex			
Female	—	—	
Male	0.81	0.73, 0.90	<0.001

¹HR = Hazard Ratio, CI = Confidence Interval

```
## contact_countrySweden      NA      NA 0.00000      NA
## sexMale                    -0.21018 0.81044 0.05525 -3.804
##                             p
## treatmentStandard (dacarbazine monotherapy < 2e-16
## contact_countryGreater Poland Cancer Center < 2e-16
## contact_countryNorfolk and Norwich Hospital 0.005083
## contact_countrySweden      NA
## sexMale                    0.000142
##
## Likelihood ratio test=595.2 on 4 df, p=< 2.2e-16
## n= 1560, number of events= 1490
```

- US: se registra el mismo tipo de terapia, y el tiempo de supervivencia se puede asociar al centro, pero siempre con mal pronóstico.

```
#unique(df2$region)
#[1] "NA"      "Australia" "US"      "Europe"    "Canada"    "Latin-America"
df3 <- df2[df2$region == "US",]

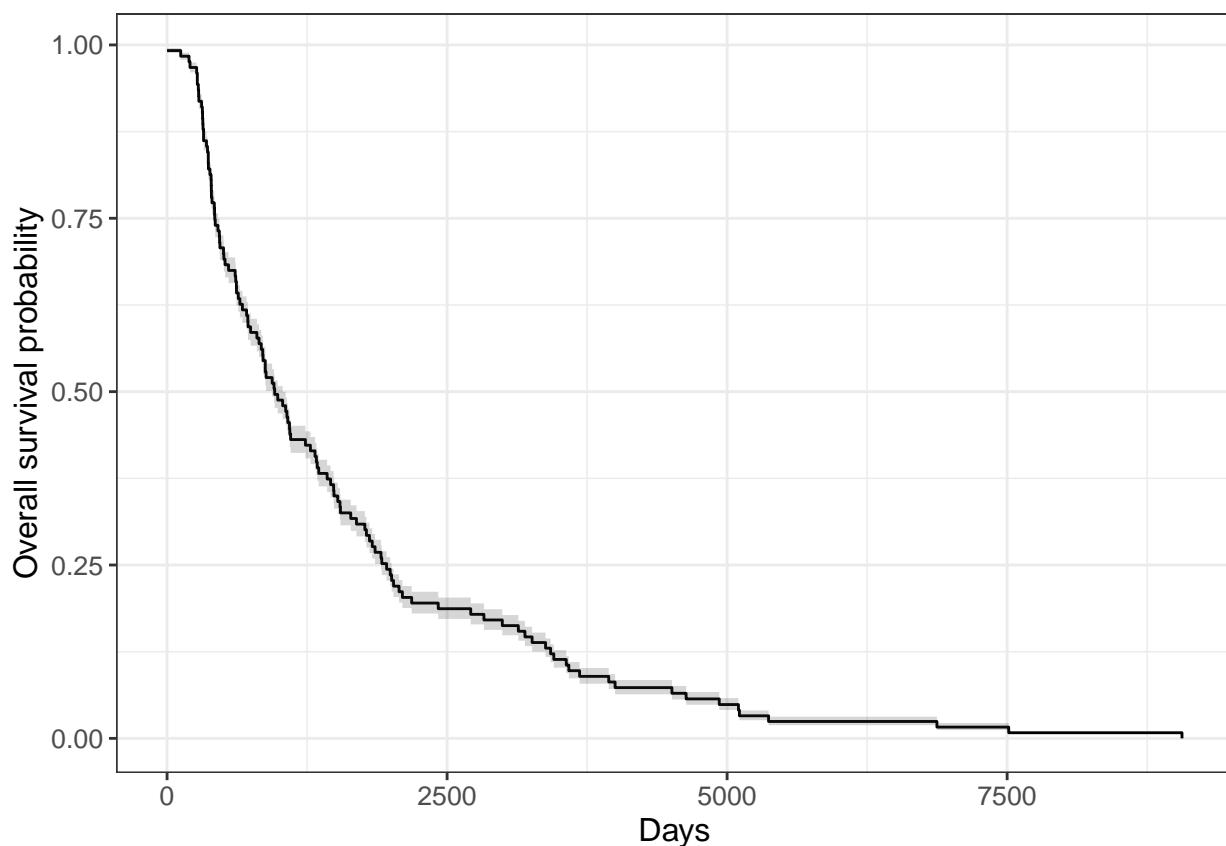
library(ggsurvfit)
# https://www.emilyzabor.com/tutorials/survival_analysis_in_r_tutorial.html
s1 <- survfit(Surv(os_days, os_event) ~ treatment, data = df3) # Creación de un survival object
str(s1)
```

```
## List of 17
## $ n      : int 2460
## $ time    : num [1:119] 0 122 196 205 264 270 279 282 284 308 ...
## $ n.risk  : num [1:119] 2460 2440 2420 2400 2380 2360 2320 2300 2280 2260 ...
## $ n.event : num [1:119] 20 20 20 20 20 40 20 20 20 20 ...
## $ n.censor : num [1:119] 0 0 0 0 0 0 0 0 0 0 ...
## $ surv     : num [1:119] 0.992 0.984 0.976 0.967 0.959 ...
## $ std.err  : num [1:119] 0.00183 0.00259 0.00319 0.0037 0.00415 ...
## $ cumhaz   : num [1:119] 0.00813 0.01633 0.02459 0.03292 0.04133 ...
## $ std.chaz : num [1:119] 0.00182 0.00258 0.00317 0.00368 0.00413 ...
## $ type     : chr "right"
## $ logse    : logi TRUE
```



```
## $ conf.int : num 0.95
## $ conf.type: chr "log"
## $ lower    : num [1:119] 0.988 0.979 0.97 0.96 0.952 ...
## $ upper    : num [1:119] 0.995 0.989 0.982 0.975 0.967 ...
## $ t0       : num 0
## $ call     : language survfit(formula = Surv(os_days, os_event) ~ treatment, data = df3)
## - attr(*, "class")= chr "survfit"
```

```
# Kaplan-Meier plots:
survfit2(Surv(os_days, os_event) ~ treatment, data = df3) %>%
  ggsurvfit() +
  labs(
    x = "Days",
    y = "Overall survival probability"
  ) +
  add_confidence_interval() # Para añadir intervalo de confianza en las curvas
```



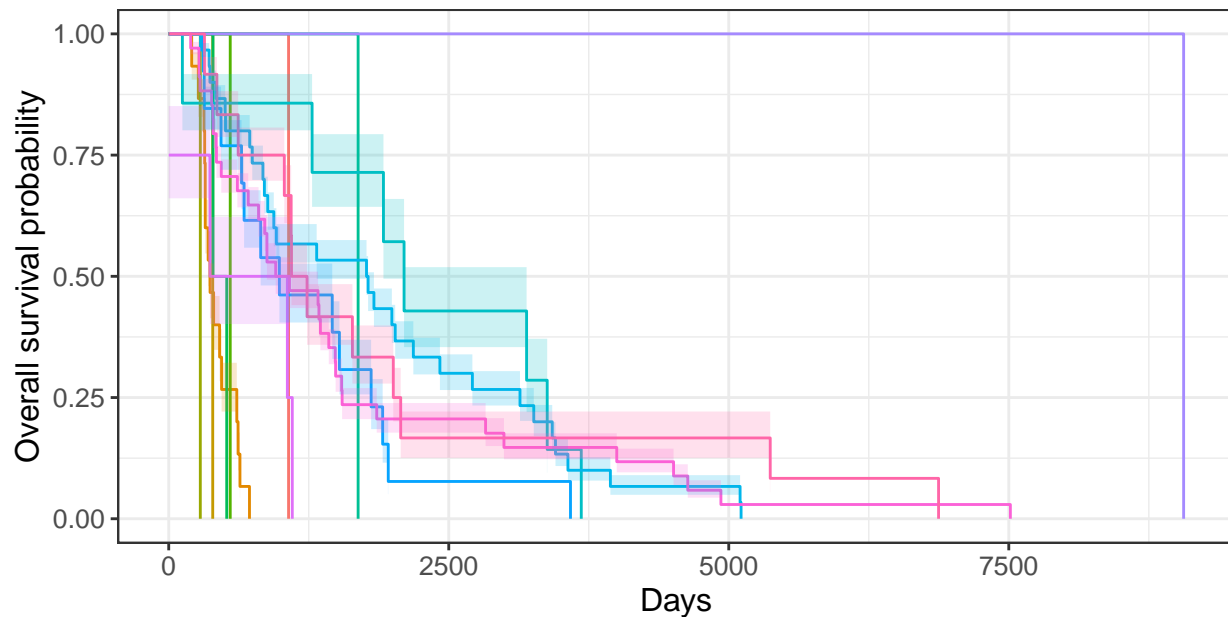
```
# add_risktable() # Tablas de pacientes supervivientes y eventos en cada intervalo contemplado
```

```
s1 <- survfit(Surv(os_days, os_event) ~ treatment + contact_country, data = df3) # Creación de un survi
str(s1)
```

```
## List of 18
## $ n      : int [1:14] 20 300 20 20 20 40 20 140 600 260 ...
## $ time   : num [1:120] 1070 205 264 315 321 326 352 369 401 454 ...
## $ n.risk : num [1:120] 20 300 280 260 240 220 180 160 140 120 ...
## $ n.event: num [1:120] 20 20 20 20 20 40 20 20 20 20 ...
## $ n.censor: num [1:120] 0 0 0 0 0 0 0 0 0 0 ...
```

```
## $ surv      : num [1:120] 0 0.933 0.867 0.8 0.733 ...
## $ std.err   : num [1:120] Inf 0.0154 0.0226 0.0289 0.0348 ...
## $ cumhaz    : num [1:120] 1 0.0667 0.1381 0.215 0.2984 ...
## $ std.chaz   : num [1:120] 0.2236 0.0149 0.0218 0.0278 0.0335 ...
## $ strata    : Named int [1:14] 1 14 1 1 1 2 1 7 30 13 ...
## .. attr(*, "names")= chr [1:14] "treatment=pharmaceutical and radiation therapies, contact_country=pharmaceutical and radiation therapies, International Genomics Consortium"
## $ type      : chr "right"
## $ logse     : logi TRUE
## $ conf.int   : num 0.95
## $ conf.type  : chr "log"
## $ lower      : num [1:120] NA 0.906 0.829 0.756 0.685 ...
## $ upper      : num [1:120] NA 0.962 0.906 0.847 0.785 ...
## $ t0        : num 0
## $ call      : language survfit(formula = Surv(os_days, os_event) ~ treatment + contact_country, data = df3)
## - attr(*, "class")= chr "survfit"
```

```
# Kaplan-Meier plots:
survfit2(Surv(os_days, os_event) ~ treatment + contact_country, data = df3) %>%
  ggsurvfit() +
  labs(
    x = "Days",
    y = "Overall survival probability"
  ) +
  add_confidence_interval() #+ # Para añadir intervalo de confianza en las curvas
```



	— pharmaceutical and radiation therapies, International Genomics Consortium	— pl
ty	— pharmaceutical and radiation therapies, John Wayne Cancer Center	— pl
th System	— pharmaceutical and radiation therapies, MD Anderson	— pl

```
# add_risktable() # Tablas de pacientes supervivientes y eventos en cada intervalo contemplado
s1 <- survfit(Surv(os_days, os_event) ~ treatment + sex, data = df3) # Creación de un survival object
```

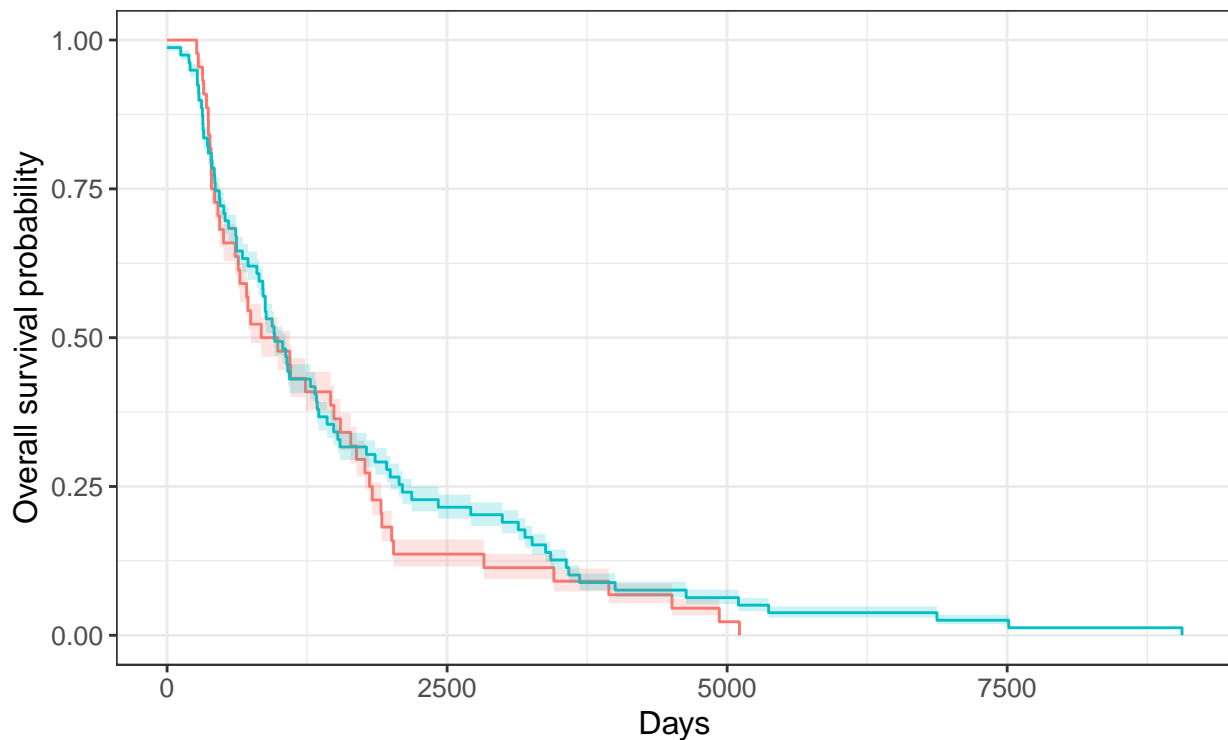
```

str(s1)

## List of 18
## $ n      : int [1:2] 880 1580
## $ time   : num [1:121] 264 279 317 326 352 368 369 383 393 394 ...
## $ n.risk  : num [1:121] 880 860 840 820 800 780 760 740 720 700 ...
## $ n.event : num [1:121] 20 20 20 20 20 20 20 20 20 20 ...
## $ n.censor : num [1:121] 0 0 0 0 0 0 0 0 0 0 ...
## $ surv    : num [1:121] 0.977 0.955 0.932 0.909 0.886 ...
## $ std.err  : num [1:121] 0.00514 0.00736 0.00912 0.01066 0.01207 ...
## $ cumhaz   : num [1:121] 0.0227 0.046 0.0698 0.0942 0.1192 ...
## $ std.chaz : num [1:121] 0.00508 0.00727 0.00901 0.01053 0.01193 ...
## $ strata   : Named int [1:2] 44 77
## ..- attr(*, "names")= chr [1:2] "treatment=pharmaceutical and radiation therapies, sex=Female" "tr
## $ type     : chr "right"
## $ logse    : logi TRUE
## $ conf.int : num 0.95
## $ conf.type: chr "log"
## $ lower    : num [1:121] 0.967 0.941 0.915 0.89 0.866 ...
## $ upper    : num [1:121] 0.987 0.968 0.949 0.928 0.908 ...
## $ t0       : num 0
## $ call     : language survfit(formula = Surv(os_days, os_event) ~ treatment + sex, data = df3)
## - attr(*, "class")= chr "survfit"

# Kaplan-Meier plots:
survfit2(Surv(os_days, os_event) ~ treatment + sex, data = df3) %>%
  ggsurvfit() +
  labs(
    x = "Days",
    y = "Overall survival probability"
  ) +
  add_confidence_interval() #+ # Para añadir intervalo de confianza en las curvas

```



— pharmaceutical and radiation therapies, Female — pharmaceutical and radiation therapies, M

add_risktable() # Tablas de pacientes supervivientes y eventos en cada intervalo contemplado

La apariencia de los gráficos mejora por permitir observar todos los tratamientos con más claridad, y no da significativo por no tener suficientes niveles:

- Australia: Las mujeres suelen tener más tiempo de supervivencia, pero siempre todos con mal pronóstico.

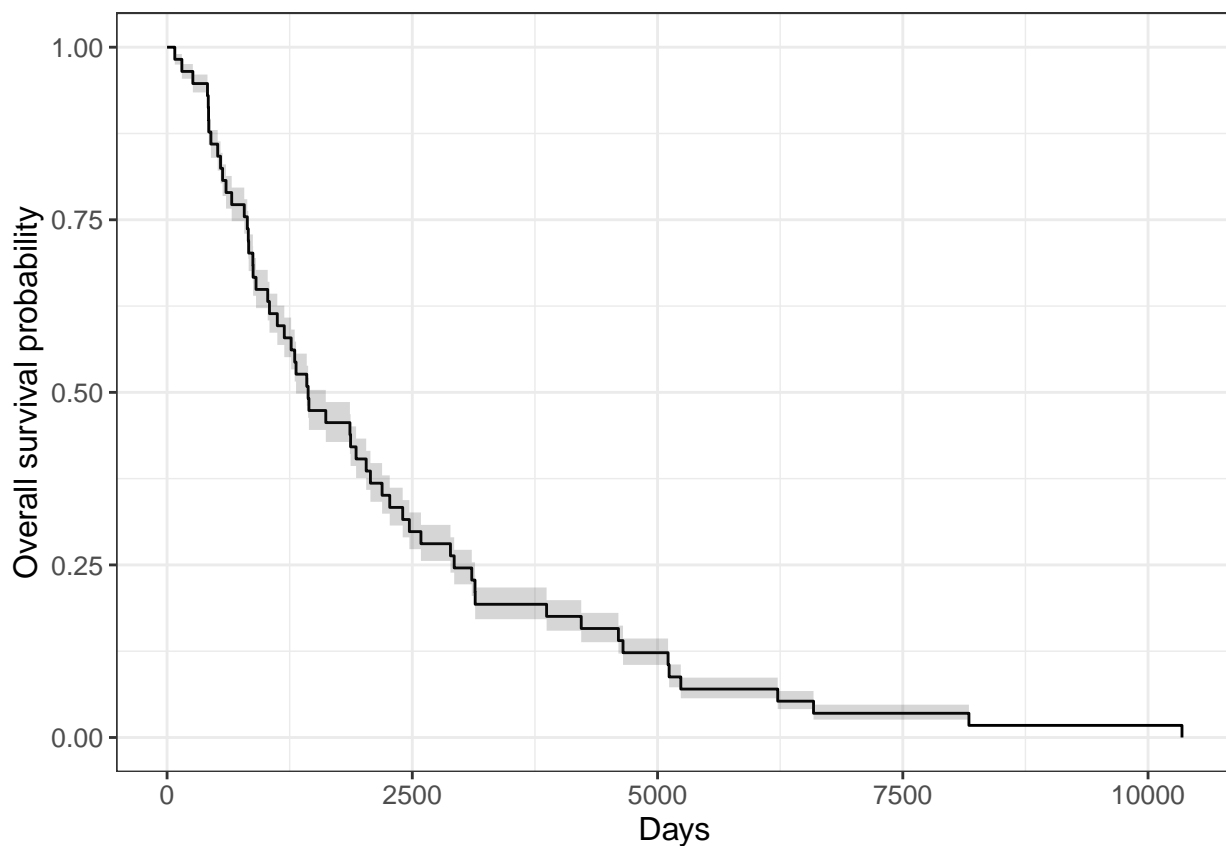
```
#unique(df2$region)
#[1] "NA" "Australia" "US" "Europe" "Canada" "Latin-America"
df3 <- df2[df2$region == "Australia",]

library(ggsurvfit)
# https://www.emilyzabor.com/tutorials/survival\_analysis\_in\_r\_tutorial.html
s1 <- survfit(Surv(os_days, os_event) ~ treatment, data = df3) # Creación de un survival object
str(s1)
```

```
## List of 17
## $ n : int 1140
## $ time : num [1:57] 79 151 263 412 420 423 425 447 516 545 ...
## $ n.risk : num [1:57] 1140 1120 1100 1080 1060 1040 1020 1000 980 960 ...
## $ n.event : num [1:57] 20 20 20 20 20 20 20 20 20 20 ...
## $ n.censor : num [1:57] 0 0 0 0 0 0 0 0 0 0 ...
## $ surv : num [1:57] 0.982 0.965 0.947 0.93 0.912 ...
## $ std.err : num [1:57] 0.00396 0.00565 0.00698 0.00814 0.00918 ...
## $ cumhaz : num [1:57] 0.0175 0.0354 0.0536 0.0721 0.091 ...
## $ std.chaz : num [1:57] 0.00392 0.0056 0.00692 0.00806 0.0091 ...
## $ type : chr "right"
## $ logse : logi TRUE
## $ conf.int : num 0.95
```

```
## $ conf.type: chr "log"
## $ lower : num [1:57] 0.975 0.954 0.934 0.915 0.896 ...
## $ upper : num [1:57] 0.99 0.976 0.96 0.945 0.929 ...
## $ t0 : num 0
## $ call : language survfit(formula = Surv(os_days, os_event) ~ treatment, data = df3)
## - attr(*, "class")= chr "survfit"
```

```
# Kaplan-Meier plots:
survfit2(Surv(os_days, os_event) ~ treatment, data = df3) %>%
  ggsurvfit() +
  labs(
    x = "Days",
    y = "Overall survival probability"
  ) +
  add_confidence_interval() ## # Para añadir intervalo de confianza en las curvas
```



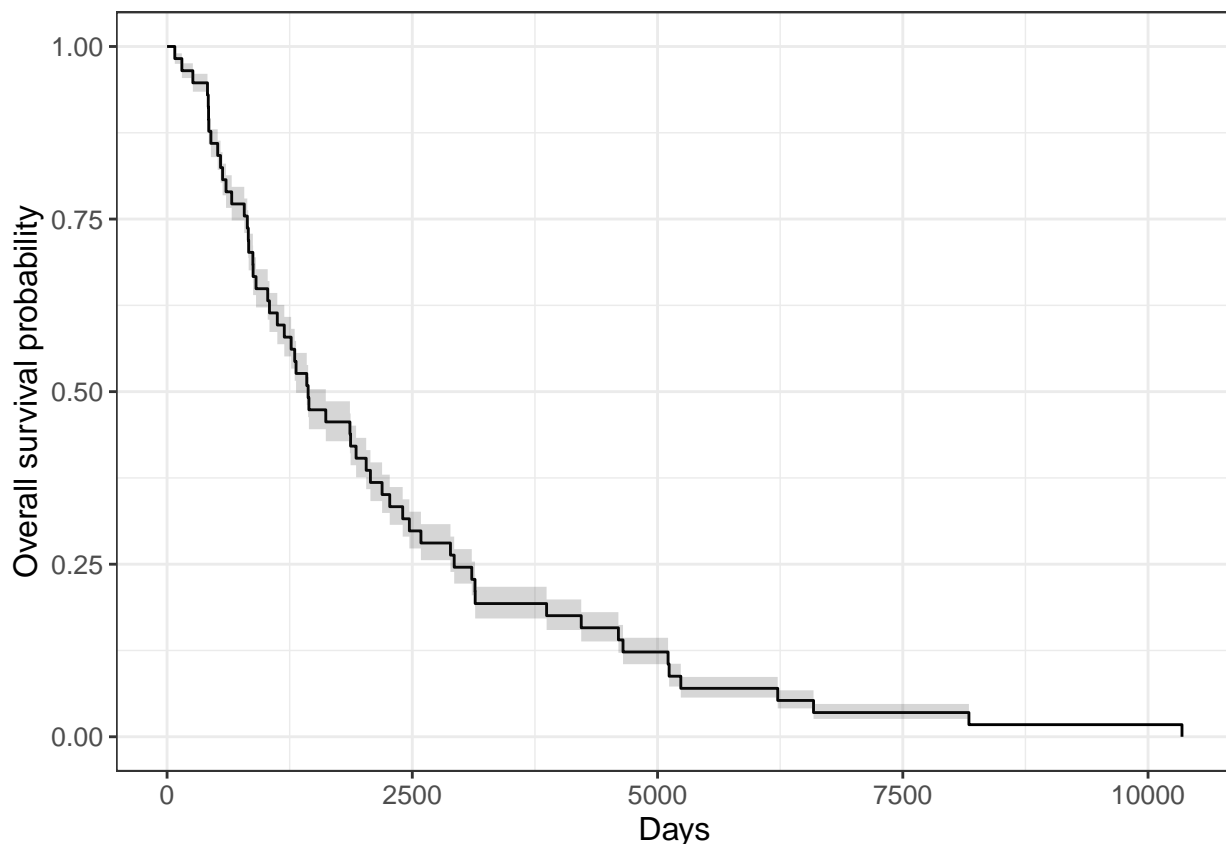
```
# add_risktable() # Tablas de pacientes supervivientes y eventos en cada intervalo contemplado
```

```
s1 <- survfit(Surv(os_days, os_event) ~ treatment + contact_country, data = df3) # Creación de un survi
str(s1)
```

```
## List of 17
## $ n : int 1140
## $ time : num [1:57] 79 151 263 412 420 423 425 447 516 545 ...
## $ n.risk : num [1:57] 1140 1120 1100 1080 1060 1040 1020 1000 980 960 ...
## $ n.event : num [1:57] 20 20 20 20 20 20 20 20 20 20 ...
## $ n.censor : num [1:57] 0 0 0 0 0 0 0 0 0 0 ...
## $ surv : num [1:57] 0.982 0.965 0.947 0.93 0.912 ...
```

```
## $ std.err : num [1:57] 0.00396 0.00565 0.00698 0.00814 0.00918 ...
## $ cumhaz : num [1:57] 0.0175 0.0354 0.0536 0.0721 0.091 ...
## $ std.chaz : num [1:57] 0.00392 0.0056 0.00692 0.00806 0.0091 ...
## $ type : chr "right"
## $ logse : logi TRUE
## $ conf.int : num 0.95
## $ conf.type: chr "log"
## $ lower : num [1:57] 0.975 0.954 0.934 0.915 0.896 ...
## $ upper : num [1:57] 0.99 0.976 0.96 0.945 0.929 ...
## $ t0 : num 0
## $ call : language survfit(formula = Surv(os_days, os_event) ~ treatment + contact_country,
## - attr(*, "class")= chr "survfit"
```

```
# Kaplan-Meier plots:
survfit2(Surv(os_days, os_event) ~ treatment + contact_country, data = df3) %>%
  ggsurvfit() +
  labs(
    x = "Days",
    y = "Overall survival probability"
  ) +
  add_confidence_interval() #+ # Para añadir intervalo de confianza en las curvas
```



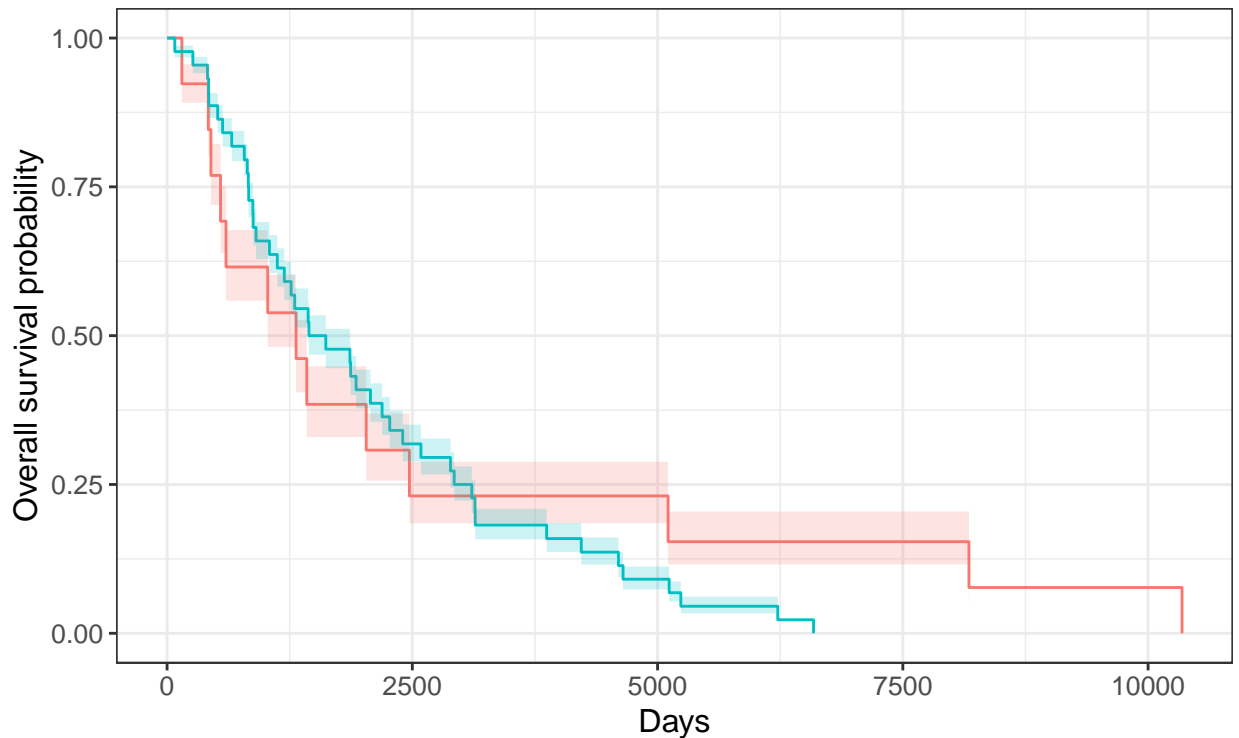
```
# add_risktable() # Tablas de pacientes supervivientes y eventos en cada intervalo contemplado

s1 <- survfit(Surv(os_days, os_event) ~ treatment + sex, data = df3) # Creación de un survival object
str(s1)
```

```
## List of 18
```

```
## $ n      : int [1:2] 260 880
## $ time   : num [1:57] 151 420 447 545 601 ...
## $ n.risk : num [1:57] 260 240 220 200 180 160 140 120 100 80 ...
## $ n.event : num [1:57] 20 20 20 20 20 20 20 20 20 20 ...
## $ n.censor : num [1:57] 0 0 0 0 0 0 0 0 0 0 ...
## $ surv    : num [1:57] 0.923 0.846 0.769 0.692 0.615 ...
## $ std.err  : num [1:57] 0.0179 0.0264 0.034 0.0413 0.049 ...
## $ cumhaz   : num [1:57] 0.0769 0.1603 0.2512 0.3512 0.4623 ...
## $ std.chaz : num [1:57] 0.0172 0.0254 0.0325 0.0395 0.0466 ...
## $ strata   : Named int [1:2] 13 44
## ..- attr(*, "names")= chr [1:2] "treatment=pharmaceutical and radiation therapies, sex=Female" "tr
## $ type      : chr "right"
## $ logse     : logi TRUE
## $ conf.int  : num 0.95
## $ conf.type : chr "log"
## $ lower     : num [1:57] 0.891 0.803 0.72 0.638 0.559 ...
## $ upper     : num [1:57] 0.956 0.891 0.822 0.751 0.677 ...
## $ t0        : num 0
## $ call      : language survfit(formula = Surv(os_days, os_event) ~ treatment + sex, data = df3)
## - attr(*, "class")= chr "survfit"

# Kaplan-Meier plots:
survfit2(Surv(os_days, os_event) ~ treatment + sex, data = df3) %>%
  ggsvrfit() +
  labs(
    x = "Days",
    y = "Overall survival probability"
  ) +
  add_confidence_interval() #+ # Para añadir intervalo de confianza en las curvas
```



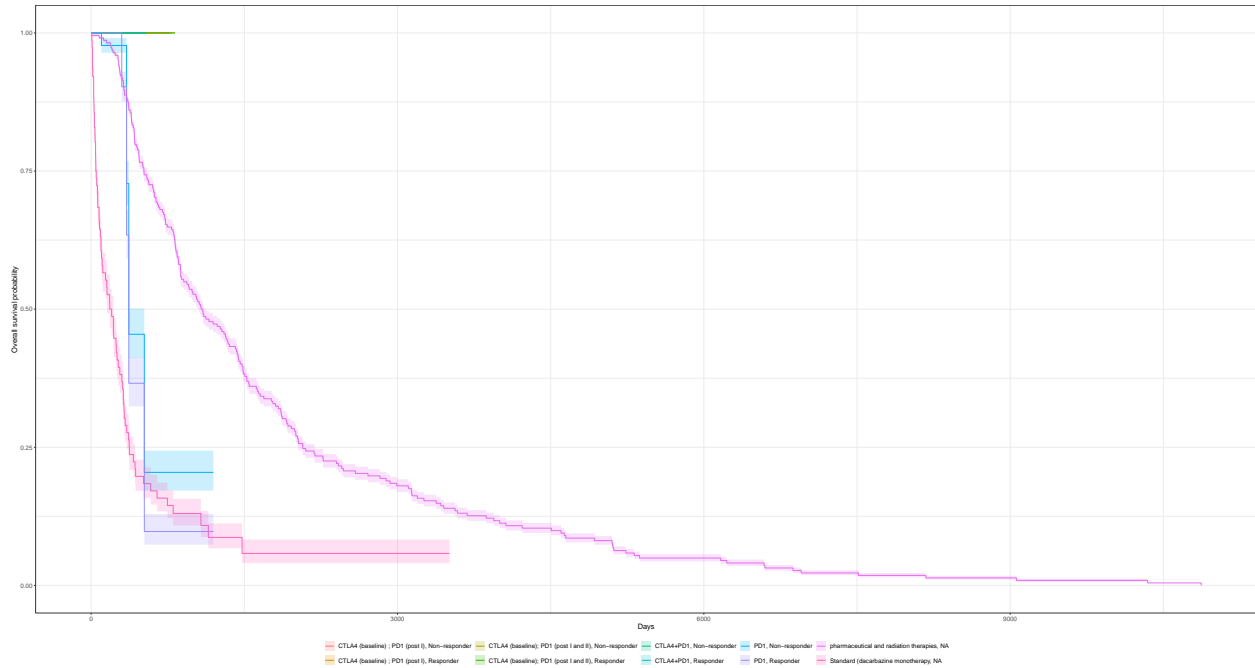
— pharmaceutical and radiation therapies, Female — pharmaceutical and radiation therapies, M

add_risktable() # Tablas de pacientes supervivientes y eventos en cada intervalo contemplado

La apariencia de los gráficos mejora por permitir observar todos los tratamientos con más claridad, y no da significativo por no tener suficientes niveles:

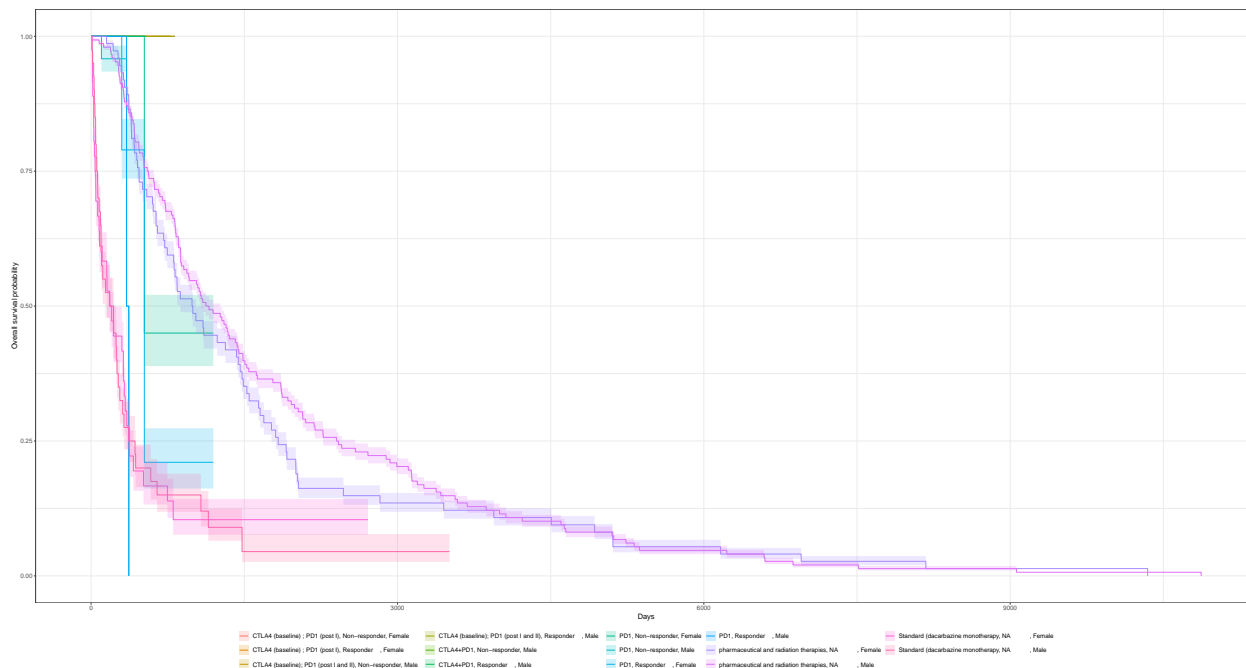
3.4.3 Respuesta al tratamiento

```
## List of 18
## $ n      : int [1:10] 121 121 132 121 363 352 484 451 4440 760
## $ time   : num [1:304] 777 777 822 822 339 388 539 339 388 539 ...
## $ n.risk : num [1:304] 121 121 132 121 363 242 121 352 231 121 ...
## $ n.event : num [1:304] 0 0 0 0 0 0 0 0 0 0 ...
## $ n.censor : num [1:304] 121 121 132 121 121 121 121 121 110 121 ...
## $ surv    : num [1:304] 1 1 1 1 1 1 1 1 1 1 ...
## $ std.err  : num [1:304] 0 0 0 0 0 0 0 0 0 0 ...
## $ cumhaz   : num [1:304] 0 0 0 0 0 0 0 0 0 0 ...
## $ std.chaz : num [1:304] 0 0 0 0 0 0 0 0 0 0 ...
## $ strata   : Named int [1:10] 1 1 1 1 3 3 5 5 215 69
## .. attr(*, "names")= chr [1:10] "treatment=CTLA4 (baseline) ; PD1 (post I), response=Non-responder"
## $ type     : chr "right"
## $ logse    : logi TRUE
## $ conf.int : num 0.95
## $ conf.type: chr "log"
## $ lower     : num [1:304] 1 1 1 1 1 1 1 1 1 1 ...
## $ upper     : num [1:304] 1 1 1 1 1 1 1 1 1 1 ...
## $ t0       : num 0
## $ call      : language survfit(formula = Surv(os_days, os_event) ~ treatment + response, data = df2)
## - attr(*, "class")= chr "survfit"
```

List of 18

```
## $ n      : int [1:14] 121 121 132 121 363 352 220 264 209 242 ...
## $ time    : num [1:308] 777 777 822 822 339 388 539 339 388 539 ...
## $ n.risk   : num [1:308] 121 121 132 121 363 242 121 352 231 121 ...
## $ n.event  : num [1:308] 0 0 0 0 0 0 0 0 0 ...
## $ n.censor : num [1:308] 121 121 132 121 121 121 121 121 110 121 ...
## $ surv     : num [1:308] 1 1 1 1 1 1 1 1 1 ...
## $ std.err   : num [1:308] 0 0 0 0 0 0 0 0 0 ...
## $ cumhaz    : num [1:308] 0 0 0 0 0 0 0 0 0 ...
## $ std.chaz  : num [1:308] 0 0 0 0 0 0 0 0 0 ...
## $ strata    : Named int [1:14] 1 1 1 1 3 3 2 3 3 2 ...
##   ..- attr(*, "names")= chr [1:14] "treatment=CTLA4 (baseline) ; PD1 (post I), response=Non-responder"
## $ type      : chr "right"
## $ logse     : logi TRUE
## $ conf.int  : num 0.95
## $ conf.type : chr "log"
## $ lower     : num [1:308] 1 1 1 1 1 1 1 1 1 ...
## $ upper     : num [1:308] 1 1 1 1 1 1 1 1 1 ...
## $ t0        : num 0
## $ call      : language survfit(formula = Surv(os_days, os_event) ~ treatment + response + sex,
## - attr(*, "class")= chr "survfit"
```



La apariencia de los gráficos mejora por permitir observar todos los tratamientos con más claridad, y obtenemos varios resultados significativos:

```
## Call:
## survival::coxph(formula = Surv(os_days, os_event) ~ treatment,
##   data = df2)
##
##
```

	coef	exp(coef)	se(coef)
## treatmentCTLA4 (baseline); PD1 (post I and II)	9.467e-03	1.010e+00	5.148e+02
## treatmentCTLA4+PD1	4.902e-03	1.005e+00	4.864e+02
## treatmentPD1	1.767e+01	4.713e+07	3.760e+02
## treatmentpharmaceutical and radiation therapies	1.664e+01	1.684e+07	3.760e+02
## treatmentStandard (dacarbazine monotherapy	1.818e+01	7.858e+07	3.760e+02

```
##
##
```

	z	p
## treatmentCTLA4 (baseline); PD1 (post I and II)	0.000	1.000
## treatmentCTLA4+PD1	0.000	1.000
## treatmentPD1	0.047	0.963
## treatmentpharmaceutical and radiation therapies	0.044	0.965
## treatmentStandard (dacarbazine monotherapy	0.048	0.961

```
##
## Likelihood ratio test=2401 on 5 df, p=< 2.2e-16
## n= 7345, number of events= 5922
## Call:
## survival::coxph(formula = Surv(os_days, os_event) ~ treatment +
##   response, data = df2)
##
##
```

	coef	exp(coef)
## treatmentCTLA4 (baseline); PD1 (post I and II)	1.663e-02	1.017e+00
## treatmentCTLA4+PD1	8.811e-03	1.009e+00
## treatmentPD1	1.768e+01	4.784e+07
## treatmentpharmaceutical and radiation therapies	1.647e+01	1.428e+07
## treatmentStandard (dacarbazine monotherapy	1.802e+01	6.672e+07

Characteristic	HR [†]	95% CI [†]	p-value
treatment			
CTLA4 (baseline) ; PD1 (post I)	—	—	
CTLA4 (baseline); PD1 (post I and II)	1.01	0.00, Inf	>0.9
CTLA4+PD1	1.00	0.00, Inf	>0.9
PD1	47,127,473	0.00, Inf	>0.9
pharmaceutical and radiation therapies	16,839,495	0.00, Inf	>0.9
Standard (dacarbazine monotherapy)	78,576,300	0.00, Inf	>0.9

[†]HR = Hazard Ratio, CI = Confidence Interval

Characteristic	HR [†]	95% CI [†]	p-value
treatment			
CTLA4 (baseline) ; PD1 (post I)	—	—	
CTLA4 (baseline); PD1 (post I and II)	1.02	0.00, Inf	>0.9
CTLA4+PD1	1.01	0.00, Inf	>0.9
PD1	47,842,506	0.00, Inf	>0.9
pharmaceutical and radiation therapies	14,278,040	0.00, Inf	>0.9
Standard (dacarbazine monotherapy)	66,722,989	0.00, Inf	>0.9
response			
NA	—	—	
Non-responder	0.71	0.62, 0.82	<0.001
Responder			

[†]HR = Hazard Ratio, CI = Confidence Interval

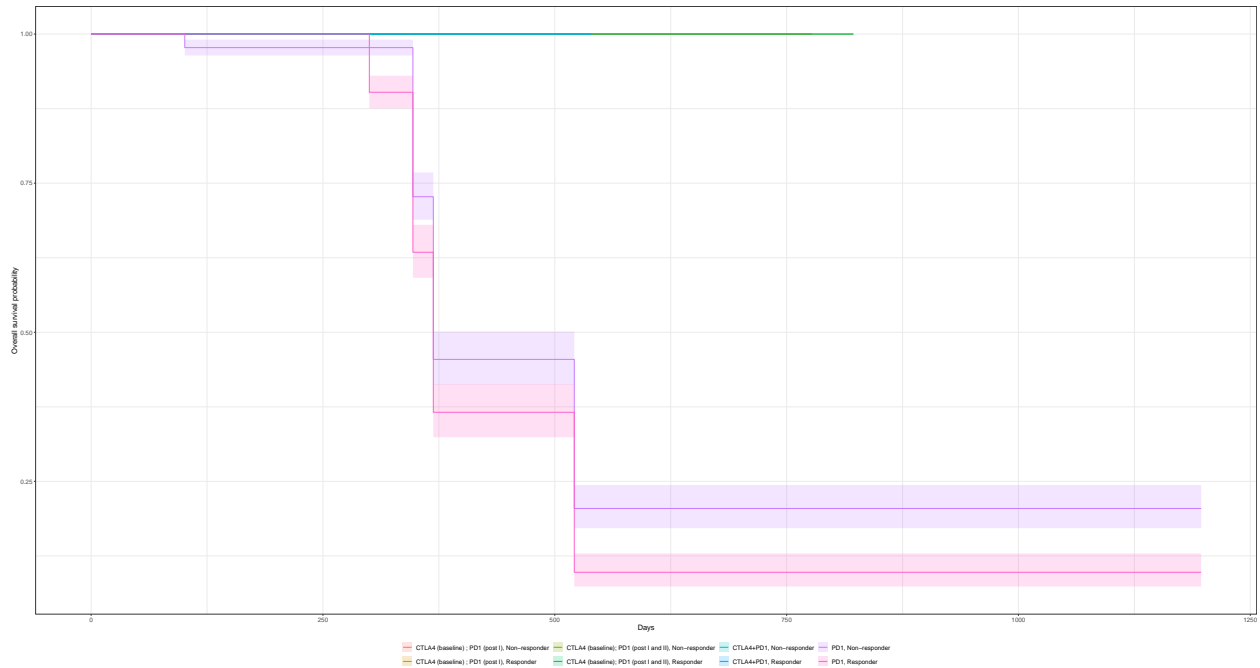
```
## responseNon-responder      -3.367e-01  7.141e-01
## responseResponder           NA           NA
##               se(coef)      z      p
## treatmentCTLA4 (baseline); PD1 (post I and II)  5.123e+02  0.000  1.000
## treatmentCTLA4+PD1                4.842e+02  0.000  1.000
## treatmentPD1                      3.742e+02  0.047  0.962
## treatmentpharmaceutical and radiation therapies  3.742e+02  0.044  0.965
## treatmentStandard (dacarbazine monotherapy)    3.742e+02  0.048  0.962
## responseNon-responder      7.120e-02 -4.729  2.26e-06
## responseResponder          0.000e+00    NA      NA
##
## Likelihood ratio test=2423  on 6 df, p=< 2.2e-16
## n= 7345, number of events= 5922
## Call:
## survival::coxph(formula = Surv(os_days, os_event) ~ treatment +
##     response + sex, data = df2)
##
##               coef exp(coef)
## treatmentCTLA4 (baseline); PD1 (post I and II) -9.548e-02  9.089e-01
## treatmentCTLA4+PD1                -1.008e-01  9.041e-01
## treatmentPD1                      1.765e+01  4.624e+07
```

Characteristic	HR [†]	95% CI [†]	p-value
treatment			
CTLA4 (baseline) ; PD1 (post I)	—	—	
CTLA4 (baseline); PD1 (post I and II)	0.91	0.00, Inf	>0.9
CTLA4+PD1	0.90	0.00, Inf	>0.9
PD1	46,242,754	0.00, Inf	>0.9
pharmaceutical and radiation therapies	13,261,023	0.00, Inf	>0.9
Standard (dacarbazine monotherapy	62,881,351	0.00, Inf	>0.9
response			
NA	—	—	
Non-responder	0.72	0.63, 0.83	<0.001
Responder			
sex			
Female	—	—	
Male	1.12	1.06, 1.18	<0.001

[†]HR = Hazard Ratio, CI = Confidence Interval

```
## treatmentpharmaceutical and radiation therapies 1.640e+01 1.326e+07
## treatmentStandard (dacarbazine monotherapy 1.796e+01 6.288e+07
## responseNon-responder -3.298e-01 7.191e-01
## responseResponder NA NA
## sexMale 1.116e-01 1.118e+00
## se(coef) z p
## treatmentCTLA4 (baseline); PD1 (post I and II) 5.123e+02 0.000 1.000
## treatmentCTLA4+PD1 4.842e+02 0.000 1.000
## treatmentPD1 3.742e+02 0.047 0.962
## treatmentpharmaceutical and radiation therapies 3.742e+02 0.044 0.965
## treatmentStandard (dacarbazine monotherapy 3.742e+02 0.048 0.962
## responseNon-responder 7.122e-02 -4.630 3.65e-06
## responseResponder 0.000e+00 NA NA
## sexMale 2.846e-02 3.922 8.80e-05
##
## Likelihood ratio test=2439 on 7 df, p=< 2.2e-16
## n= 7345, number of events= 5922
## List of 18
## $ n : int [1:8] 121 121 132 121 363 352 484 451
## $ time : num [1:20] 777 777 822 822 339 388 539 339 388 539 ...
## $ n.risk : num [1:20] 121 121 132 121 363 242 121 352 231 121 ...
## $ n.event : num [1:20] 0 0 0 0 0 0 0 0 0 0 ...
## $ n.censor : num [1:20] 121 121 132 121 121 121 121 121 110 121 ...
## $ surv : num [1:20] 1 1 1 1 1 1 1 1 1 1 ...
## $ std.err : num [1:20] 0 0 0 0 0 0 0 0 0 0 ...
## $ cumhaz : num [1:20] 0 0 0 0 0 0 0 0 0 0 ...
## $ std.chaz : num [1:20] 0 0 0 0 0 0 0 0 0 0 ...
## $ strata : Named int [1:8] 1 1 1 1 3 3 5 5
## ..- attr(*, "names")= chr [1:8] "treatment=CTLA4 (baseline) ; PD1 (post I), response=Non-responder"
## $ type : chr "right"
## $ logse : logi TRUE
```

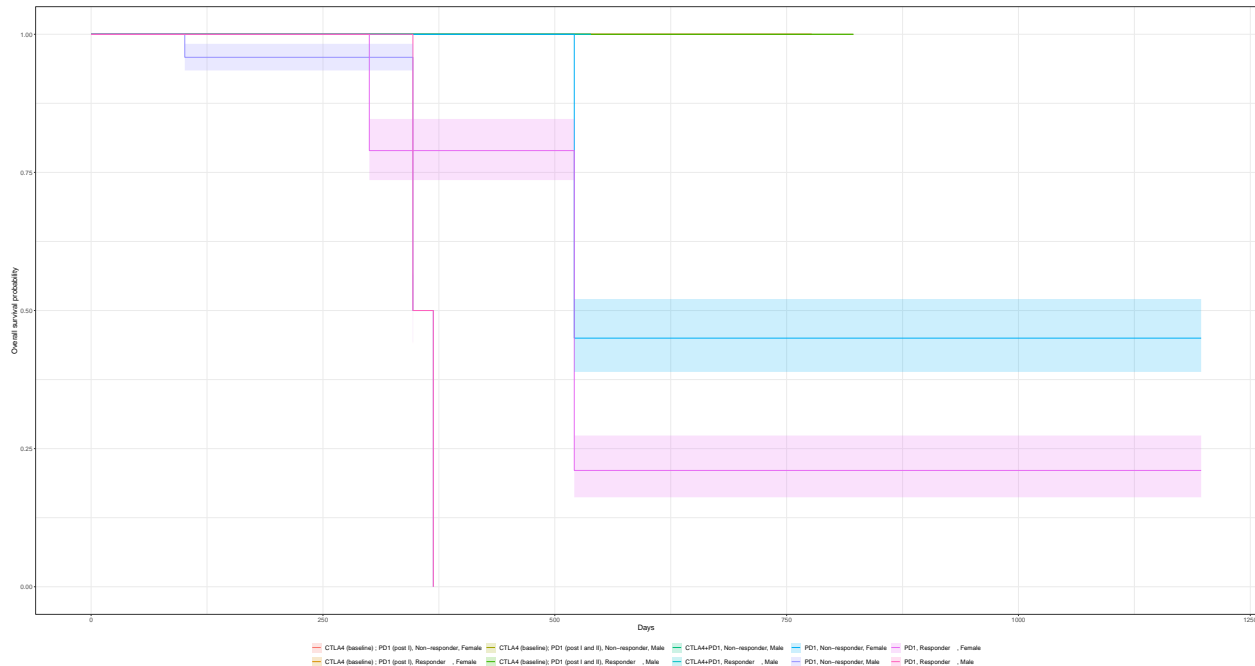
```
## $ conf.int : num 0.95
## $ conf.type: chr "log"
## $ lower    : num [1:20] 1 1 1 1 1 1 1 1 1 1 ...
## $ upper    : num [1:20] 1 1 1 1 1 1 1 1 1 1 ...
## $ t0       : num 0
## $ call     : language survfit(formula = Surv(os_days, os_event) ~ treatment + response, data = df2)
## - attr(*, "class")= chr "survfit"
```



```
## List of 18
## $ n      : int [1:10] 121 121 132 121 363 352 220 264 209 242
## $ time    : num [1:20] 777 777 822 822 339 388 539 339 388 539 ...
## $ n.risk  : num [1:20] 121 121 132 121 363 242 121 352 231 121 ...
## $ n.event : num [1:20] 0 0 0 0 0 0 0 0 0 0 ...
## $ n.censor: num [1:20] 121 121 132 121 121 121 121 121 110 121 ...
## $ surv    : num [1:20] 1 1 1 1 1 1 1 1 1 1 ...
## $ std.err  : num [1:20] 0 0 0 0 0 0 0 0 0 0 ...
## $ cumhaz   : num [1:20] 0 0 0 0 0 0 0 0 0 0 ...
## $ std.chaz : num [1:20] 0 0 0 0 0 0 0 0 0 0 ...
## $ strata   : Named int [1:10] 1 1 1 1 3 3 2 3 3 2
## ..- attr(*, "names")= chr [1:10] "treatment=CTLA4 (baseline) ; PD1 (post I), response=Non-responder"
## $ type     : chr "right"
## $ logse    : logi TRUE
## $ conf.int : num 0.95
## $ conf.type: chr "log"
## $ lower    : num [1:20] 1 1 1 1 1 1 1 1 1 1 ...
## $ upper    : num [1:20] 1 1 1 1 1 1 1 1 1 1 ...
## $ t0       : num 0
## $ call     : language survfit(formula = Surv(os_days, os_event) ~ treatment + response + sex,
## - attr(*, "class")= chr "survfit"
```

Characteristic	HR ¹	95% CI ¹	p-value
treatment			
CTLA4 (baseline) ; PD1 (post I)	—	—	
CTLA4 (baseline); PD1 (post I and II)	1.00	0.00, Inf	>0.9
CTLA4+PD1	1.09	0.00, Inf	>0.9
PD1	1,518,155,284	0.00, Inf	>0.9

¹HR = Hazard Ratio, CI = Confidence Interval



La apariencia de los gráficos mejora por permitir observar todos los tratamientos con más claridad, y obtenemos varios resultados significativos:

```
## Call:
## survival::coxph(formula = Surv(os_days, os_event) ~ treatment,
##   data = df2)
##
##               coef exp(coef) se(coef)
## treatmentCTLA4 (baseline); PD1 (post I and II) -2.637e-07  1.000e+00  2.559e+03
## treatmentCTLA4+PD1                             8.458e-02  1.088e+00  2.328e+03
## treatmentPD1                                    2.114e+01  1.518e+09  1.829e+03
##
##               z      p
## treatmentCTLA4 (baseline); PD1 (post I and II) 0.000 1.000
## treatmentCTLA4+PD1                             0.000 1.000
## treatmentPD1                                    0.012 0.991
##
## Likelihood ratio test=1696 on 3 df, p=< 2.2e-16
## n= 2145, number of events= 792
##
## Call:
## survival::coxph(formula = Surv(os_days, os_event) ~ treatment +
##   response, data = df2)
```

Characteristic	HR ¹	95% CI ¹	p-value
treatment			
CTLA4 (baseline) ; PD1 (post I)	—	—	
CTLA4 (baseline); PD1 (post I and II)	1.01	0.00, Inf	>0.9
CTLA4+PD1	1.09	0.00, Inf	>0.9
PD1	1,557,281,659	0.00, Inf	>0.9
response			
Non-responder	—	—	
Responder	1.39	1.21, 1.60	<0.001

¹HR = Hazard Ratio, CI = Confidence Interval

```
##
##                                coef exp(coef) se(coef)
## treatmentCTLA4 (baseline); PD1 (post I and II) 7.223e-03 1.007e+00 2.553e+03
## treatmentCTLA4+PD1                        8.937e-02 1.093e+00 2.323e+03
## treatmentPD1                             2.117e+01 1.557e+09 1.825e+03
## responseResponder                        3.327e-01 1.395e+00 7.129e-02
##                                z          p
## treatmentCTLA4 (baseline); PD1 (post I and II) 0.000      1.000
## treatmentCTLA4+PD1                        0.000      1.000
## treatmentPD1                             0.012      0.991
## responseResponder                        4.666 3.07e-06
##
## Likelihood ratio test=1717 on 4 df, p=< 2.2e-16
## n= 2145, number of events= 792
## Call:
## survival::coxph(formula = Surv(os_days, os_event) ~ treatment +
##     response + sex, data = df2)
##
##                                coef exp(coef) se(coef)
## treatmentCTLA4 (baseline); PD1 (post I and II) -3.504e+00 3.008e-02 2.343e+03
## treatmentCTLA4+PD1                        -3.327e+00 3.590e-02 2.221e+03
## treatmentPD1                             2.046e+01 7.714e+08 1.675e+03
## responseResponder                        2.567e-01 1.293e+00 7.143e-02
## sexMale                                3.509e+00 3.343e+01 1.732e-01
##                                z          p
## treatmentCTLA4 (baseline); PD1 (post I and II) -0.001 0.998807
## treatmentCTLA4+PD1                        -0.001 0.998805
## treatmentPD1                             0.012 0.990252
## responseResponder                        3.594 0.000326
## sexMale                                20.258 < 2e-16
##
## Likelihood ratio test=2557 on 5 df, p=< 2.2e-16
## n= 2145, number of events= 792
```

Characteristic	HR ^I	95% CI ^I	p-value
treatment			
CTLA4 (baseline) ; PD1 (post I)	—	—	
CTLA4 (baseline); PD1 (post I and II)	0.03	0.00, Inf	>0.9
CTLA4+PD1	0.04	0.00, Inf	>0.9
PD1	771,359,903	0.00, Inf	>0.9
response			
Non-responder	—	—	
Responder	1.29	1.12, 1.49	<0.001
sex			
Female	—	—	
Male	33.4	23.8, 46.9	<0.001

^IHR = Hazard Ratio, CI = Confidence Interval