

reportRmd

2023-01-13

Contents

Overview	2
Summary statistics	3
Univariate regression	8
Multivariable analysis	10
Combining univariate and multivariable models	11
Changing the output	13
Combining tables	13
Simple Survival Summaries	14
Survival Times in Long Format	15
Stratified Survival Summary	16
Plotting Functions	17
Plotting bivariate relationships	17
Plotting odds ratios	18
Plotting survival curves	19
PDF Output	20
Data Sets	20
pembrolizumab	20
ctDNA	20

Overview

reportRmd is a package designed to facilitate the reporting of common statistical outputs easily in RMarkdown documents. The package supports pdf, html and word output without any changes to the body of the report. The main features are Table 1 style summaries, combining multiple univariate regression models into a single table, tidy multivariable model output and combining univariate and multivariable regressions into a single table. Single table summaries of median survival times and survival probabilities are also provided. A highly customisable survival curve function, based on ggplot2 can be used to create publication-quality plots. Visualisation plots are also available for bivariate relationships and logistic regression models.

A word of caution:

The reportRmd package is designed to facilitate statistical reporting and does not provide any checks regarding the suitability of the models fit.

Summary statistics

Basic summary statistics

```
rm_covsum(data=pembrolizumab,  
covs=c('age', 'sex'))
```

	n=94
age	
Mean (sd)	57.9 (12.8)
Median (Min,Max)	59.1 (21.1, 81.8)
sex	
Female	58 (62)
Male	36 (38)

Set IQR = T for interquartile range instead of Min/Max

```
rm_covsum(data=pembrolizumab,  
covs=c('age', 'sex'), IQR=TRUE)
```

	n=94
age	
Mean (sd)	57.9 (12.8)
Median (Q1,Q3)	59.1 (49.5, 68.7)
sex	
Female	58 (62)
Male	36 (38)

Or all.stats=T for both

```
rm_covsum(data=pembrolizumab,  
covs=c('age', 'sex'), all.stats = TRUE)
```

	n=94
age	
Mean (sd)	57.9 (12.8)
Median (Q1,Q3)	59.1 (49.5, 68.7)
Range (min, max)	(21.1, 81.8)
sex	
Female	58 (62)
Male	36 (38)

This will produce summary statistics by Sex

```
rm_covsum(data=pembrolizumab, maincov = 'sex',
covs=c('age', 'pdl1', 'change_ctdna_group'))
```

	Full Sample (n=94)	Female (n=58)	Male (n=36)	p-value
age				0.30
Mean (sd)	57.9 (12.8)	56.9 (12.6)	59.3 (13.1)	
Median (Min,Max)	59.1 (21.1, 81.8)	56.6 (34.1, 78.2)	61.2 (21.1, 81.8)	
pdl1				0.76
Mean (sd)	13.9 (29.2)	15.0 (30.5)	12.1 (27.3)	
Median (Min,Max)	0 (0, 100)	0.5 (0.0, 100.0)	0 (0, 100)	
Missing	1	0	1	
change ctdna group				0.81
Decrease from baseline	33 (45)	19 (48)	14 (42)	
Increase from baseline	40 (55)	21 (52)	19 (58)	
Missing	21	18	3	

To indicate which statistical test was used use `show.tests=TRUE`

```
rm_covsum(data=pembrolizumab, maincov = 'sex',
covs=c('age', 'pdl1', 'change_ctdna_group'),
show.tests=TRUE)
```

	Full Sample (n=94)	Female (n=58)	Male (n=36)	p-value	StatTest
age				0.30	Wilcoxon Rank Sum
Mean (sd)	57.9 (12.8)	56.9 (12.6)	59.3 (13.1)		
Median (Min,Max)	59.1 (21.1, 81.8)	56.6 (34.1, 78.2)	61.2 (21.1, 81.8)		
pdl1				0.76	Wilcoxon Rank Sum
Mean (sd)	13.9 (29.2)	15.0 (30.5)	12.1 (27.3)		
Median (Min,Max)	0 (0, 100)	0.5 (0.0, 100.0)	0 (0, 100)		
Missing	1	0	1		
change ctdna group				0.81	Fisher Exact
Decrease from baseline	33 (45)	19 (48)	14 (42)		
Increase from baseline	40 (55)	21 (52)	19 (58)		
Missing	21	18	3		

Effect sizes (Cramer V and η^2) can be added with `effSize=TRUE`

```
rm_covsum(data=pembrolizumab, maincov = 'sex',
covs=c('age', 'pd11', 'change_ctdna_group'),
effSize=TRUE, show.tests=TRUE)
```

	Full Sample (n=94)	Female (n=58)	Male (n=36)	p-value	Effect Size	
age				0.30	0.008	Wilcoxon R
Mean (sd)	57.9 (12.8)	56.9 (12.6)	59.3 (13.1)			
Median (Min,Max)	59.1 (21.1, 81.8)	56.6 (34.1, 78.2)	61.2 (21.1, 81.8)			
pd11				0.76	0.002	Wilcoxon R
Mean (sd)	13.9 (29.2)	15.0 (30.5)	12.1 (27.3)			
Median (Min,Max)	0 (0, 100)	0.5 (0.0, 100.0)	0 (0, 100)			
Missing	1	0	1			
change ctdna group				0.81	0.02	Fish
Decrease from baseline	33 (45)	19 (48)	14 (42)			
Increase from baseline	40 (55)	21 (52)	19 (58)			
Missing	21	18	3			

Group comparisons are non-parametric by default, specify `testcont='ANOVA'` for t-tests/ANOVA

```
rm_covsum(data=pembrolizumab, maincov = 'sex',
covs=c('age', 'pd11'),
testcont='ANOVA',
show.tests=TRUE)
```

	Full Sample (n=94)	Female (n=58)	Male (n=36)	p-value	StatTest
age				0.39	t-test
Mean (sd)	57.9 (12.8)	56.9 (12.6)	59.3 (13.1)		
Median (Min,Max)	59.1 (21.1, 81.8)	56.6 (34.1, 78.2)	61.2 (21.1, 81.8)		
pd11				0.63	t-test
Mean (sd)	13.9 (29.2)	15.0 (30.5)	12.1 (27.3)		
Median (Min,Max)	0 (0, 100)	0.5 (0.0, 100.0)	0 (0, 100)		
Missing	1	0	1		

The default is to indicate percentages by columns (ie percentages within columns add to 100)

```
rm_covsum(data=pembrolizumab, maincov = 'sex',  
covs=c('cohort'),  
pvalue = FALSE)
```

	Full Sample (n=94)	Female (n=58)	Male (n=36)
cohort			
A	16 (17)	3 (5)	13 (36)
B	18 (19)	18 (31)	0 (0)
C	18 (19)	18 (31)	0 (0)
D	12 (13)	7 (12)	5 (14)
E	30 (32)	12 (21)	18 (50)

But you can also specify to show by row instead

```
rm_covsum(data=pembrolizumab, maincov = 'sex',  
covs=c('cohort'),  
pvalue = FALSE,  
percentage='row')
```

	Full Sample (n=94)	Female (n=58)	Male (n=36)
cohort			
A	16	3 (19)	13 (81)
B	18	18 (100)	0 (0)
C	18	18 (100)	0 (0)
D	12	7 (58)	5 (42)
E	30	12 (40)	18 (60)

Univariate regression

Combining multiple univariate regression analyses into a single table

```
rm_uvsum(data=pembrolizumab, response='orr',
covs=c('age', 'pd11', 'change_ctdna_group'))
```

	OR(95%CI)	p-value	N
age	0.96 (0.91, 1.00)	0.09	94
pd11	0.97 (0.95, 0.98)	< 0.001	93
change ctdna group		0.002	73
Decrease from baseline	Reference		33
Increase from baseline	28.74 (5.20, 540.18)		40

If the response is continuous linear regression is the default

```
rm_uvsum(data=pembrolizumab, response='l_size',
covs=c('age', 'cohort'))
```

	Estimate(95%CI)	p-value	Global p-value	N
age	-0.58 (-1.54, 0.38)	0.23		94
cohort			< 0.001	94
A	Reference			16
B	-38.04 (-74.95, -1.13)	0.04		18
C	20.35 (-16.56, 57.26)	0.28		18
D	-24.79 (-65.82, 16.23)	0.23		12
E	31.69 (-1.56, 64.95)	0.06		30

... unless two variables are specified and then survival analysis is run

```
rm_uvsum(data=pembrolizumab, response=c('os_time', 'os_status'),
covs=c('age', 'pd11', 'change_ctdna_group'))
```

	HR(95%CI)	p-value	N
age	0.99 (0.97, 1.01)	0.16	94
pd11	0.99 (0.98, 1.00)	0.03	93
change ctdna group		< 0.001	73
Decrease from baseline	Reference		33
Increase from baseline	3.06 (1.62, 5.77)		40

Correlated observations can be handled using GEE

```
rm_uvsum(response = 'size_change',
covs=c('time', 'ctdna_status'),
gee=TRUE,
id='id', corstr="exchangeable",
family=gaussian("identity"),
data=ctDNA, showN=TRUE)
```

	Estimate(95%CI)	p-value	Global p-value	N
time	-0.12 (-0.44, 0.19)	0.44		262
ctdna status			<0.001	262
Clearance	Reference			134
No clearance, decrease from baseline	61.29 (37.38, 85.20)	<0.001		42
No clearance, increase from baseline	82.52 (64.67, 100.37)	<0.001		86

If you want to check the underlying models, set `returnModels = TRUE`

```
rm_uvsum(response = 'orr',
covs=c('age', 'sex'),
data=pembrolizumab, returnModels = TRUE)
```

```
## $age
##
## Call:  glm(formula = orr ~ age, family = binomial, data = data)
##
## Coefficients:
## (Intercept)      age
##  4.12269      -0.04231
##
## Degrees of Freedom: 93 Total (i.e. Null);  92 Residual
## Null Deviance:      85.77
## Residual Deviance: 82.53      AIC: 86.53
##
## $sex
##
## Call:  glm(formula = orr ~ sex, family = binomial, data = data)
##
## Coefficients:
## (Intercept)    sexMale
##  1.9859      -0.8873
##
## Degrees of Freedom: 93 Total (i.e. Null);  92 Residual
## Null Deviance:      85.77
## Residual Deviance: 83.21      AIC: 87.21
```

Multivariable analysis

To create a nice display for multivariable models the model needs to first be fit. By default, the variance inflation factor will be shown to check for multicollinearity. To suppress this column set `vif=FALSE`. Note: variance inflation factors are not computed (yet) for multilevel or GEE models.

```
glm_fit <- glm(orr~change_ctdna_group+pd11+age,
              family='binomial',
              data = pembrolizumab)
rm_mvsum(glm_fit)
```

	OR(95%CI)	p-value	VIF
change ctdna group		0.006	1.03
Decrease from baseline	Reference		
Increase from baseline	23.92 (2.49, 229.77)		
pd11	0.97 (0.95, 0.99)	0.01	1.24
age	0.94 (0.87, 1.01)	0.08	1.23

p-values can be adjusted for multiple comparisons using any of the options available in the `p.adjust` function. This argument is also available for univariate models run with `rm_uvsum`.

```
rm_mvsum(glm_fit, showN = TRUE, vif=TRUE, p.adjust = 'holm')
```

	OR(95%CI)	p-value	N	VIF	raw p-value
change ctdna group		0.02	73	1.03	0.006
Decrease from baseline	Reference		33		
Increase from baseline	23.92 (2.49, 229.77)		40		
pd11	0.97 (0.95, 0.99)	0.02	73	1.24	0.01
age	0.94 (0.87, 1.01)	0.08	73	1.23	0.08

Combining univariate and multivariable models

To display both models in a single table run the `rm_uvsum` and `rm_mvsum` functions with `tableOnly=TRUE` and combine.

```
uvsumTable <- rm_uvsum(data=pembrolizumab, response='orr',
covs=c('age','sex','pd11','change_ctdna_group'),tableOnly = TRUE)

glm_fit <- glm(orr~change_ctdna_group+pd11,
              family='binomial',
              data = pembrolizumab)
mvsumTable <- rm_mvsum(glm_fit, showN = TRUE,tableOnly = TRUE)

rm_uv_mv(uvsumTable,mvsumTable)
```

	Unadjusted OR(95%CI)	N	p	Adjusted OR(95%CI)	N (adj)	p (adj)
age	0.96 (0.91, 1.00)	94	0.09			
sex		94	0.11			
Female	Reference	58				
Male	0.41 (0.13, 1.22)	36				
pd11	0.97 (0.95, 0.98)	93	<0.001	0.98 (0.96, 1.00)	73	0.02
change ctdna group		73	0.002		73	0.004
Decrease from baseline	Reference	33		Reference	33	
Increase from baseline	28.74 (5.20, 540.18)	40		24.71 (2.87, 212.70)	40	

This can also be done with adjusted p-values, but when combined the raw p-values are dropped

```
uvsumTable <- rm_uvsum(data=pembrolizumab, response='orr',
covs=c('age', 'sex', 'pd11', 'change_ctdna_group'), tableOnly = TRUE, p.adjust='holm')

glm_fit <- glm(orr~change_ctdna_group+pd11,
              family='binomial',
              data = pembrolizumab)
mvsumTable <- rm_mvsum(glm_fit, tableOnly = TRUE, p.adjust='holm')

rm_uv_mv(uvsumTable, mvsumTable)
```

	Unadjusted OR(95%CI)	N	p	Adjusted OR(95%CI)	p (adj)
age	0.96 (0.91, 1.00)	94	0.18		
sex		94	0.18		
Female	Reference	58			
Male	0.41 (0.13, 1.22)	36			
pd11	0.97 (0.95, 0.98)	93	<0.001	0.98 (0.96, 1.00)	0.02
change ctdna group		73	0.005		0.007
Decrease from baseline	Reference	33		Reference	
Increase from baseline	28.74 (5.20, 540.18)	40		24.71 (2.87, 212.70)	

Changing the output

If you need to make changes to the tables, setting `tableOnly=TRUE` will return a data frame for any of the `rm_` functions. Changes can be made, and the table output using `outTable()`

```
mvsumTable <- rm_mvsum(glm_fit, showN = TRUE, tableOnly = TRUE)
names(mvsumTable)[1] <- 'Predictor'
outTable(mvsumTable)
```

Predictor	OR(95%CI)	p-value	N	VIF
change ctdna group		0.004	73	1.01
Decrease from baseline	Reference		33	
Increase from baseline	24.71 (2.87, 212.70)		40	
pd11	0.98 (0.96, 1.00)	0.02	73	1.01

Combining tables

Tables can be nested with the `nestTable()` function

```
cohortA <- rm_uvsum(data=subset(pembrolizumab,cohort=='A'),
  response = 'pd11',
  covs=c('age','sex'),
  tableOnly = T)
cohortA$Cohort <- 'Cohort A'
cohortE <- rm_uvsum(data=subset(pembrolizumab,cohort=='E'),
  response = 'pd11',
  covs=c('age','sex'),
  tableOnly = T)
cohortE$Cohort <- 'Cohort E'
nestTable(rbind(cohortA,cohortE),head_col = 'Cohort',to_col = 'Covariate')
```

	Estimate(95%CI)	p-value	N
Cohort A			
age	2.94 (-0.70, 6.58)	0.10	15
sex		0.14	15
Female	Reference		3
Male	-40.25 (-96.25, 15.75)		12
Cohort E			
age	-0.44 (-1.02, 0.15)	0.14	30
sex		0.10	30
Female	Reference		12
Male	-14.86 (-32.57, 2.85)		18

Simple Survival Summaries

Displaying survival probabilities at different times by sex using Kaplan Meier estimates

```
rm_survsum(data=pembrolizumab,time='os_time',status='os_status',
group="sex",survtimes=seq(12,36,12),survtimeunit='months')
```

sex	Events/Total	Median (95%CI)	12months (95% CI)	24months (95% CI)	36months (95% CI)
Female	39/58	14.29 (9.69, 23.82)	0.55 (0.44, 0.69)	0.34 (0.24, 0.50)	0.29 (0.18, 0.45)
Male	25/36	11.24 (6.14, 25.33)	0.50 (0.36, 0.69)	0.31 (0.18, 0.52)	0.27 (0.15, 0.48)
			Log Rank Test	ChiSq	0.5 on 1 df
				p-value	0.46

Survival Times in Long Format

Displaying survival probabilities at different times by sex using Cox PH estimates

```
rm_survtime(data=pembrolizumab,time='os_time',status='os_status',  
            strata="sex",survtimes=c(12,24),survtimeunit='mo',type='PH')
```

Time (mo)	At Risk	Events	Censored	Survival Rate (95% CI)
Overall	94			
12	48	44	2	0.53 (0.44, 0.64)
24	24	17	7	0.33 (0.25, 0.45)
Female	58			
12	31	26	1	0.55 (0.44, 0.70)
24	16	11	4	0.35 (0.24, 0.50)
Male	36			
12	17	18	1	0.51 (0.37, 0.70)
24	8	6	3	0.32 (0.19, 0.52)

Displaying survival probabilities at different times by sex, adjusting for age using Cox PH estimates

```
rm_survtime(data=pembrolizumab,time='os_time',status='os_status', covs='age',  
            strata="sex",survtimes=c(12,24),survtimeunit='mo',type='PH')
```

Time (mo)	At Risk	Events	Censored	Survival Rate (95% CI)
Overall	94			
12	48	44	2	0.54 (0.44, 0.65)
24	24	17	7	0.33 (0.25, 0.45)
Female	58			
12	31	26	1	0.56 (0.44, 0.70)
24	16	11	4	0.35 (0.24, 0.50)
Male	36			
12	17	18	1	0.51 (0.37, 0.70)
24	8	6	3	0.31 (0.19, 0.53)

Stratified Survival Summary

```
rm_survdiff(data=pembrolizumab,time='os_time',status='os_status',  
            covs='sex',strata='cohort',digits=1)
```

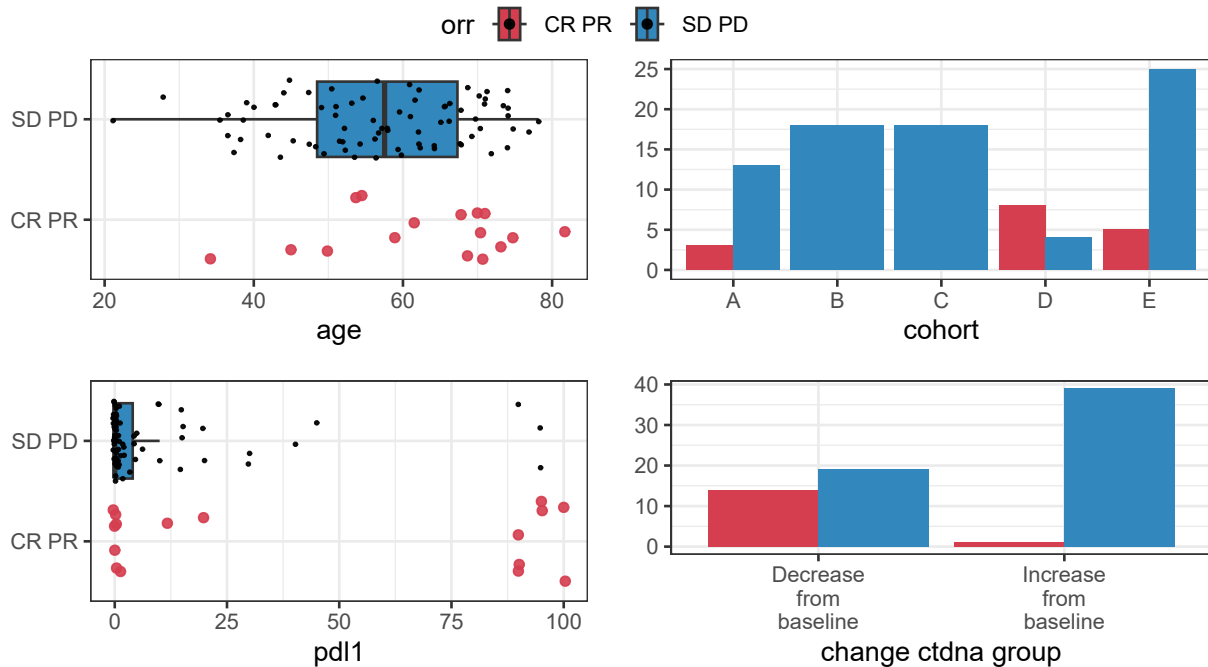
group	N	Observed	Expected	Median (95%CI)
Overall	94	64		14.0 (9.0, 20.1)
Female	58	39	43.0	14.3 (9.7, 23.8)
Male	36	25	21.0	11.2 (6.1, 25.3)
Log Rank Test stratified by cohort				ChiSq = 1.9 on 1 df p-value = 0.17

Plotting Functions

Plotting bivariate relationships

These plots are designed for quick inspection of many variables, not for publication. This is the plotting version of `rm_uvsum`

```
plotuv(data=pembrolizumab, response='orr',  
covs=c('age', 'cohort', 'pd11', 'change_ctdna_group'))
```



Plotting odds ratios

This will default to a log scale, but can be set to linear using `logscale=FALSE`

```
require(ggplot2)
glm_fit <- glm(orr~age+sex+change_ctdna_group+pd11,
              family='binomial',
              data = pembrolizumab)
forestplot2(glm_fit)
```

change ctdna group

Decrease from baseline (Reference)

Increase from baseline 39.29 (3.37, 458.44)

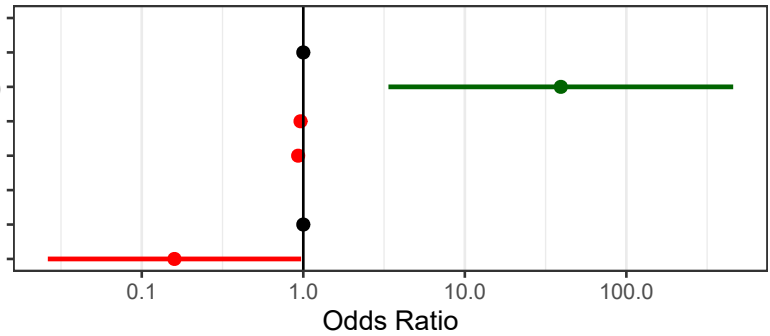
pd11 0.96 (0.93, 0.99)

age 0.93 (0.86, 1.01)

sex

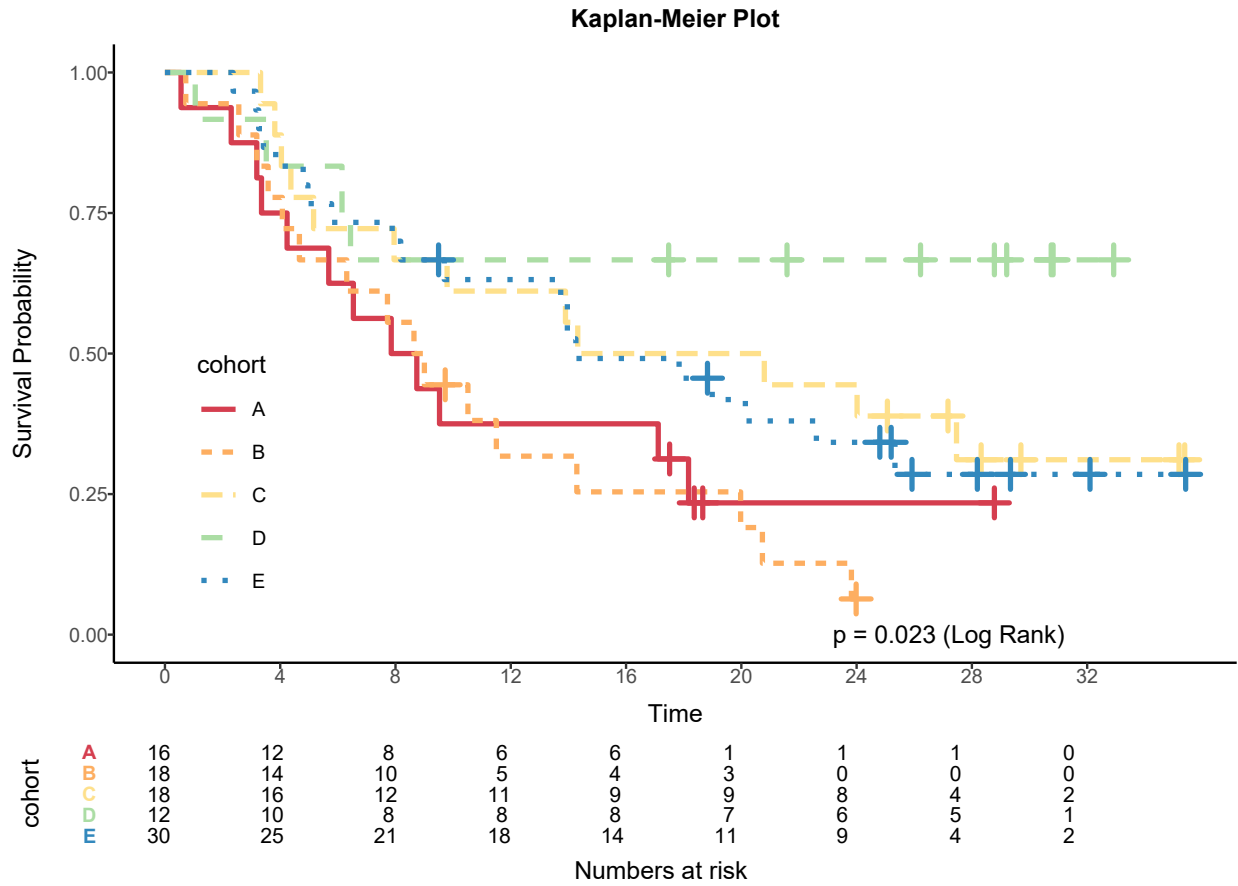
Female (Reference)

Male 0.16 (0.03, 0.97)



Plotting survival curves

```
ggkmCIF(response = c('os_time', 'os_status'),
cov='cohort',
data=pembrolizumab)
```



PDF Output

For pdf to be correctly generate when using survival curves it is recommended that the cairo format be used. This can be specified with the following command in the setup code chunk:

```
knitr::opts_chunk$set(message = FALSE, warning = FALSE, dev="cairo_pdf")
```

Data Sets

pembrolizumab

Survival status and ctDNA levels for patients receiving pembrolizumab

A data frame with 94 rows and 15 variables:

- **id** Patient ID
- **age** Age at study entry
- **sex** Patient Sex
- **cohort** Study Cohort: A = Squamous cell carcinoma of soft pallate, B = Triple negative breast cancer, C = Ovarian, high grade serous, D = Melanoma, E = Other Solid Tumor
- **l_size** Target lesion size at baseline
- **pd11** PD L1 percent
- **tmb** log of TMB
- **baseline_ctdna** Baseline ctDNA
- **change_ctdna_group** Did ctDNA increase or decrease from baseline to cycle 3
- **orr** Objective Response
- **cbr** Clinical Beneficial Response
- **os_status** Overall survival status, 0 = alive, 1 = deceased
- **os_time** Overall survival time in months
- **pfs_status** Progression free survival status, 0 = progression free, 1 = progressed
- **pfs_time** Progression free survival time in months

ctDNA

Longitudinal changes in tumour size since baseline for patients by changes in ctDNA status (clearance, decrease or increase) since baseline.

A data frame with 270 rows and 5 variables:

- **id** Patient ID
- **cohort** Study Cohort: A = Squamous cell carcinoma of soft pallate, B = Triple negative breast cancer, C = Ovarian, high grade serous, D = Melanoma, E = Other Solid Tumor
- **ctdna_status** Change in ctDNA since baseline
- **time** Number of weeks on treatment
- **size_change** Percentage change in tumour measurement