

# ANALISI DELLA SOPRAVVIVENZA: Dataset SOFA

Confronto tra diversi indicatori  
SOFA e CCSCORE

## IL DATASET

- Composto da **64 righe** (pazienti) e **20 variabili**:
  - Variabili d'interesse: SOFA Score (*SOFAING*) e CC Score (*CCSCORE*)
  - Codice identificativo del Paziente
  - Date: Data di Nascita, Data dell'Intervento, Data di Dimissione, Data di Decesso
  - Variabili socio-demografiche: Genere, Peso, Altezza, Stato Civile
  - Variabili relative ad analisi di laboratorio: Calcio, Albumina, Emoglobina, Vitamina D
  - Variabili relative all'intervento: Tipo di Anestesia, Durata dell'Intervento, Tempo di Ricovero
  - Altre Variabili: Numero di Cadute, MMSE
- **Obiettivo:** Dopo aver pulito e descritto il DB, analizzare la mortalità tra chi ha un punteggio SOFA elevato vs basso e ripetere per il CC Score. Valutare quale tra questi è il più predittivo.

# STEP DI LAVORO

1

## PULIZIA DATI

- Gestione date
- Gestione formattazione mista
- Gestione valori non plausibili

2

## GESTIONE DATI MANCANTI

- Rimozione Vitamina D
- Imputazione tramite funzione "MICE"

3

## ANALISI DESCRITTIVA

- Ricategorizzazione SOFA e CCSCORE
- Analisi univariata

4

## ANALISI DELLA SOPRAVVIVENZA

- Kaplan-Meier
- Modello di Cox
- RMST

# PULIZIA DATI

1

## Gestione Date e Formati

- 1800 in 1900 (cambio origine)
- Gestione dati con lo slash
- Ricodifica per passare da virgola a punto in certi valori di alcune variabili continue

2

## Gestione Valori Non Plausibili

Resi NA valori non plausibili nella realtà

# PULIZIA DATI

## Gestione Date

- Date con errori di battitura
  - Assenza del separatore "/"
  - 1800 al posto di 1900: data di origine per il calcolo delle date 01.01.1970

	PAZIENTE	NASCITA
33	79	5307
34	80	13/07/1829
35	82	13728

```

library(readxl)
date_with_text <- function(x, origin = "1899-12-30"){
  v <- rep(0:63)
  num <- suppressWarnings(as.numeric(x))
  for(i in 1:length(x)){
    if(x[i]<0){
      v[i] <- as.Date(num[i], origin = "1970-01-01")
    if(anyNA(num)){
      y <- as.character(y)
      y[is.na(num)] <- as.character(x[is.na(num)])
    }
    else{v[i] <- as.Date(num[i], origin = "1899-12-30")}
  }
  return(v)
}
SOFA$NASCITA <- date_with_text(SOFA$NASCITA)
SOFA$NASCITA <- as.Date(SOFA$NASCITA,origin = "1970-01-01")

```

	PAZIENTE	NASCITA
33	79	1914-07-12
34	80	1929-07-13
35	82	1937-08-01

- Alcune celle codificate in caratteri e non in numeri.  
Nell'importazione in R la formattazione non è corretta  
Variabili coinvolte: MMSE, Albumina, Calcio, Vitamina D, Emoglobina

```

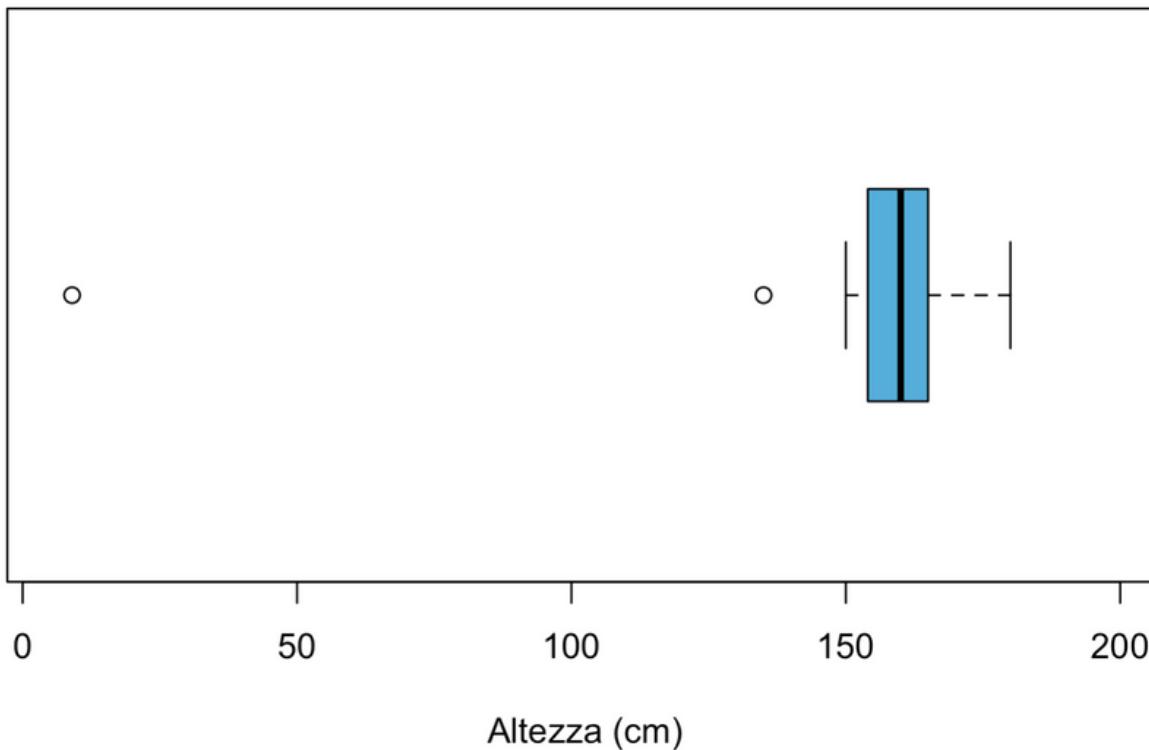
for(i in 1:length(SOFA$MMSE)){
  if(grepl(", ",SOFA$MMSE[i])){
    SOFA$MMSE[i] <- gsub(", ",".",SOFA$MMSE[i])
  }
}
SOFA$MMSE <- round(as.numeric(SOFA$MMSE),1)

```

# PULIZIA DATI

## Gestione Valori Non Plausibili

- Per il **PESO** e l'**ALTEZZA**:



- Per il **CALCIO** [¹] esclusi valori > 12 mg/dl

- Per l'**ALBUMINA** [²] in un paziente sano valori compresi tra 3.5 e 5.5 g/dl

n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
<int>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	0.984375	3.387937	0.4220272	2.56	3.095	3.39	3.625	4.5

- Per la data del decesso un soggetto presenta "si è rifiutata" ricodificata con la data di fine dello studio

```
require(skimr)
skim_without_charts(SOFA$PESO)

for(i in 1:length(SOFA$PESO)){
  if (is.na(SOFA$PESO[i])){ i = i + 1}
  else if(SOFA$PESO[i] == 33){SOFA$PESO[i] = NA}}
```

```
for(i in 1:length(SOFA$`DATA DECESSO`)){
  if (is.na(SOFA$`DATA DECESSO`[i]))
  { i = i + 1}
  else if(SOFA$`DATA DECESSO`[i] == "SI è RIFIUTATA")
  {SOFA$`DATA DECESSO`[i] = "2013-01-31"}}
```

# GESTIONE DATI MANCANTI

1

## Rimozione Vitamina D [<sup>3</sup>]

Il 25% dei dati riferiti a questa variabile è mancante, dunque si è deciso di eliminare la colonna dal dataset

2

## Imputazione con funzione "MICE"

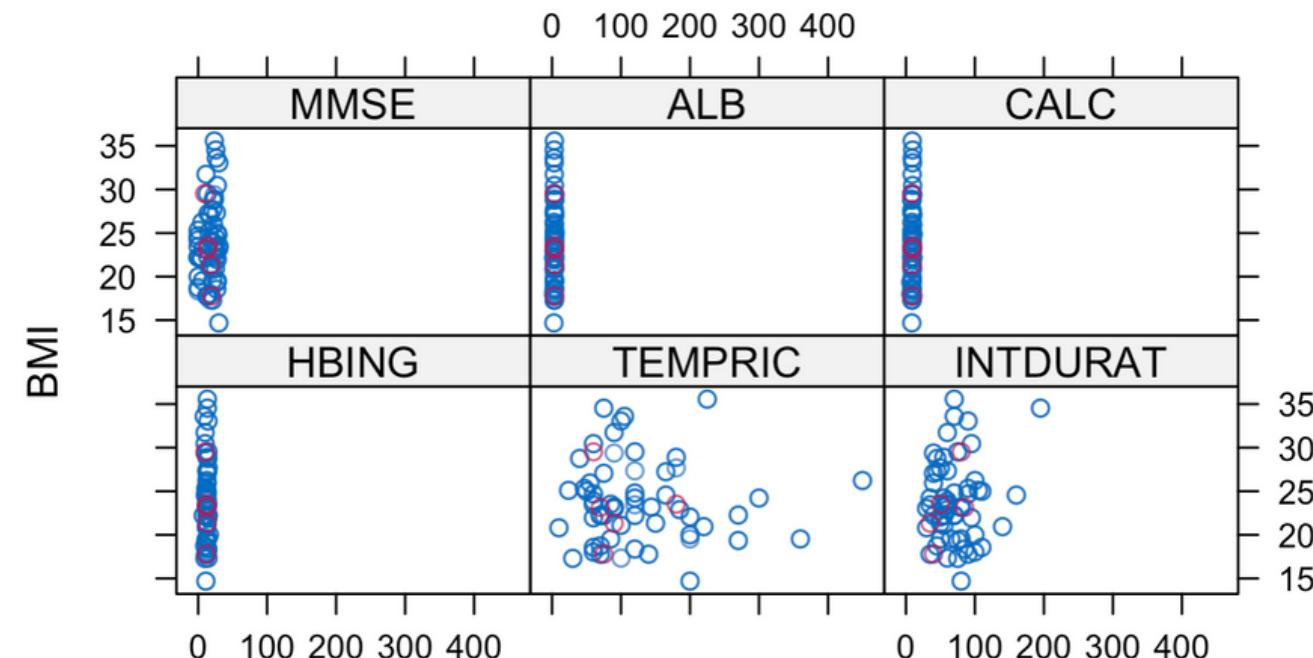
Per le variabili con una percentuale di dati mancanti limitata, è stata eseguita un'imputazione

# GESTIONE DATI MANCANTI

## Imputazione Tramite Mice

### Perché l'imputazione MICE?

- Imputazione **MICE** (*Multiple Imputation by Chained Equations*): metodo efficace perché consente di gestire i dati mancanti in modo rigoroso e statisticamente valido.
- Genera più versioni dei dati originali, ciascuna con valori imputati diversi, ed esegue analisi separate su ogni versione dei dati.
- I risultati delle analisi vengono combinati per produrre risultati finali più precisi e robusti rispetto ai metodi di imputazione più semplici.
- Tiene conto delle relazioni tra le variabili e delle incertezze nei dati mancanti, per cui è adatta ad una vasta gamma di situazioni.



```
require(mice)
SOFA_subset <- subset(SOFA,
                       select = c(BMI, MMSE, ALB, CALC, HBING, TEMPRIC,
                                  INTDURAT, CADUTE))
imputati <- mice(data = SOFA_subset, method = 'pmm', m = 1,
                  maxit = 100, seed = 1234)
```

# ANALISI DESCRITTIVA

1

## Variabili di Interesse

- SOFA Score
- CC Score

2

## Analisi univariata

- Test di Wilcoxon e Fisher
- Data visualization

# ANALISI DESCRITTIVA

## Variabili di Interesse

### SOFA

- *Sequential Organ Failure Assessment*: valuta la funzionalità degli organi e permette di identificare precocemente situazioni critiche/fatali.
- Utilizzato nelle unità di terapia intensiva per seguire i progressi dei pazienti e guidare le decisioni terapeutiche.

### CC SCORE

- *Charlson comorbidity score*: metodo utilizzato per prevedere l'aspettativa di vita di un paziente con condizioni patologiche coesistenti.
- Consente di misurare la comorbidità e correlarla alla probabilità di sopravvivenza e al consumo di risorse sanitarie, assegnando un punteggio a 19 condizioni.

# ANALISI DESCRITTIVA

## Variabili di Interesse

### SOFA

- *Sequential Organ Failure Assessment*: valuta la funzionalità degli organi e permette di identificare precocemente situazioni critiche/fatali.
- Utilizzato nelle unità di terapia intensiva per seguire i progressi dei pazienti e guidare le decisioni terapeutiche.
- La variabile è stata ricodificata in base ai dati

```
SOFA$SOFA_cat1 <- 0
for(i in 1:length(SOFA$SOFAING)){
  if (SOFA$SOFAING[i] == 0 | SOFA$SOFAING[i] == 1)
    {SOFA$SOFA_cat1[i] = 0}
  if (SOFA$SOFAING[i] == 2 | SOFA$SOFAING[i] == 3 |
      SOFA$SOFAING[i] == 4)
    {SOFA$SOFA_cat1[i] = 1}}
```

### DATI NON IMPUTATI

	Evento		
SOFA	0	1	
Nessuna/Lieve disfunzione	36	6	
Moderata/Elevata disfunzione	2	5	$\frac{5}{7} = 71\%$

### DATI IMPUTATI

	Evento		
SOFA	0	1	
Nessuna/Lieve disfunzione	47	7	
Moderata/Elevata disfunzione	3	6	$\frac{6}{9} = 67\%$

# ANALISI DESCRITTIVA

## Variabili di Interesse

### CC SCORE

- *Charlson comorbidity score*: metodo utilizzato per prevedere l'aspettativa di vita di un paziente con condizioni patologiche coesistenti.
- Consente di misurare la comorbidità e correlarla alla probabilità di sopravvivenza e al consumo di risorse sanitarie, assegnando un punteggio a 19 condizioni.
- La variabile è stata ricodificata in base ai dati

```

SOFA$CCSCORE_cat1 <- 0
for(i in 1:length(SOFA$CCSCORE)){
  if (is.na(SOFA$CCSCORE[i])){
    SOFA$CCSCORE_cat1[i] <- NA}
  else if(SOFA$CCSCORE[i] == 0 | SOFA$CCSCORE[i] == 1 |
          SOFA$CCSCORE[i] == 2){
    SOFA$CCSCORE_cat1[i] <- 0}
  else if(SOFA$CCSCORE[i] == 3 | SOFA$CCSCORE[i] == 4 |
          SOFA$CCSCORE[i] == 5 | SOFA$CCSCORE[i] == 6){
    SOFA$CCSCORE_cat1[i] <- 1}
}
```

### DATI NON IMPUTATI

	Evento	
CC SCORE	0	1
Comorbidità non severa	21	2
Severa comorbidità	17	9

$\frac{9}{26} = 35\%$

### DATI IMPUTATI

	Evento	
CC SCORE	0	1
Comorbidità non severa	27	2
Severa comorbidità	23	11

$\frac{11}{34} = 32\%$

# ANALISI DESCRITTIVA

## Analisi univariata

```

rndr <- function(x, name, ...) {
  if (length(x) == 0) {
    y <- SOFA_pulito[[name]]
    s <- rep("", length(render.default(x=y, name=name, ...)))
    if (is.numeric(y)) {
      p <- wilcox.test(y ~ SOFA_pulito$SOFA_cat1)$p.value
    } else if (is.factor(y) || is.character(y)) {
      p <- fisher.test(table(y, droplevels(SOFA_pulito$SOFA_cat1)))$p.value
    }
    s[3] <- sub("<", "<", format.pval(p, digits=3, eps=0.001))
    s
  } else {
    if (is.numeric(x)){
      render.default(x=x, name=name, ...)
      parse.abbrev.render.code(c("", "Median [IQR]"))(x)
    } else {render.default(x=x, name=name, ...)}
  }
}

rndr.strat <- function(label, n, ...) {
  ifelse(n==0, label, render.strat.default(label, n, ...))
}

table1(~ BMI + Età + Albumina + Calcio + Emoglobina + `Durata intervento` | SOFA_cat1,
       data=SOFA_pulito, overall="Total", render = rndr,
       render.strat=rndr.strat, droplevels=F, topclass="Rtable1-grid")

table1(~ Sesso +`Stato Civile` + `Num. cadute` + `CC Score` + Anestesia + MMSE | SOFA_cat1,
       data=SOFA_pulito, overall="Total", render = rndr,
       render.strat=rndr.strat, droplevels=F, topclass="Rtable1-grid")

```

### TEST DI FISHER

Utilizzato per verificare associazione tra due variabili categoriche in alternativa al **test Chi-quadrato** quando la dimensione del campione è piccola (numerosità di cella < 5).

### TEST DI WILCOXON

Utilizzato per confrontare le mediane di due campioni indipendenti, alternativa non parametrica al test T (non soddisfatte ipotesi di normalità)

**H<sub>0</sub>**: **mediana<sub>1</sub> = mediana<sub>2</sub>**

# ANALISI DESCRITTIVA

## Analisi univariata

	Nessuna/Lieve disfunzione (N=54)	Moderata/Elevata disfunzione (N=9)	P-value	Total (N=63)
<b>Sesso</b>				
Maschi	6 (11.1%)	4 (44.4%)		10 (15.9%)
Femmine	48 (88.9%)	5 (55.6%)	0.0288	53 (84.1%)
<b>Stato Civile</b>				
Coniugato	16 (29.6%)	5 (55.6%)		21 (33.3%)
Non sposato/Vedovo	38 (70.4%)	4 (44.4%)	0.146	42 (66.7%)
<b>Num. cadute</b>				
1	18 (33.3%)	3 (33.3%)		21 (33.3%)
2	36 (66.7%)	6 (66.7%)	1	42 (66.7%)
<b>CC Score</b>				
Comoborbidità non severa	27 (50.0%)	2 (22.2%)		29 (46.0%)
Severa comoborbidità	27 (50.0%)	7 (77.8%)	0.16	34 (54.0%)
<b>Anestesia</b>				
Generale	20 (37.0%)	4 (44.4%)		24 (38.1%)
Spinale/Plessica/Combinata	34 (63.0%)	5 (55.6%)	0.721	39 (61.9%)
<b>MMSE</b>				
Cognitivamente sano	17 (31.5%)	4 (44.4%)		21 (33.3%)
Lieve deterioramento cognitivo	9 (16.7%)	1 (11.1%)	0.887	10 (15.9%)
Con demenza	28 (51.9%)	4 (44.4%)		32 (50.8%)

## VARIABILI CATEGORICHE

- Variabile significativa: genere
- Non risultano altre significatività
- Analisi con dati non imputati: risultati analoghi

# ANALISI DESCRITTIVA

## Analisi univariata

	<b>Nessuna/Lieve disfunzione (N=54)</b>	<b>Moderata/Elevata disfunzione (N=9)</b>	<b>P-value</b>	<b>Total (N=63)</b>
<b>BMI</b>				
Median [IQR]	23.3 [5.18]	23.0 [4.87]	0.536	23.2 [5.22]
<b>Età (anni)</b>				
Median [IQR]	85.0 [8.75]	87.0 [6.00]	0.694	85.0 [8.50]
<b>Albumina (g/dl)</b>				
Median [IQR]	3.42 [0.600]	3.20 [0.270]	0.242	3.39 [0.530]
<b>Calcio (mg/dl)</b>				
Median [IQR]	9.05 [0.875]	9.00 [0.200]	0.548	9.00 [0.800]
<b>Emoglobina (g/dl)</b>				
Median [IQR]	12.8 [1.25]	12.9 [2.90]	0.516	12.8 [1.40]
<b>Durata intervento (minuti)</b>				
Median [IQR]	60.0 [38.8]	70.0 [15.0]	0.898	60.0 [35.0]

## VARIABILI CONTINUE

- **BMI :** 
$$\frac{\text{Peso (kg)}}{(\text{Altezza (m)})^2}$$

- **Età :** (all'intervento)

*Data Intervento – Data Nascita*

- Nessuna variabile significativa
- Analisi con dati non imputati : risultati analoghi

# ANALISI DELLA SOPRAVVIVENZA

1

## Creazione variabili TEMPO ed EVENTO

- Variabili necessarie per analisi della sopravvivenza

2

## Stimatore di Kaplan-Meier

- Confronto sopravvivenza SOFA vs. CCScore
- Analisi test log-Rank

3

## Modello di Cox

- Verifica degli assunti
- Modello di Cox senza covariate
- Model Selection

4

## RMST

- Nuova metrica di sintesi [6]

# ANALISI DELLA SOPRAVVIVENZA

## Creazione variabili

### TEMPO:

- Differenza tra decesso e data dell'intervento in giorni.
- In assenza di data del decesso: differenza tra data di fine dello studio (31.01.2013) e data dell'intervento.

```
SOFA$time <- as.integer(SOFA$`DATA DECESSO` - SOFA$DATINT)
for(i in 1:length(SOFA$`DATA DECESSO`)){
  if (SOFA$time[i] < 0){SOFA$time[i] <- NA}
}

SOFA <- SOFA[!is.na(SOFA$time), ]
```

### EVENTO:

- L'evento è 0 se la data del decesso è quella della fine dello studio (scelta arbitraria)
- L'evento è 1 se è presente una data del decesso antecedente il 31.01.2013 (fine studio)

```
SOFA$evento <- 1
for(i in 1:length(SOFA$`DATA DECESSO`)){
  if (is.na(SOFA$`DATA DECESSO`[i])){
    {SOFA$`DATA DECESSO`[i] <- "2013-01-31"}
    if(SOFA$`DATA DECESSO`[i] == "2013-01-31")
      {SOFA$evento[i] <- 0}
  }
}

SOFA$`DATA DECESSO` <- as.Date(SOFA$`DATA DECESSO`, format="%Y-%m-%d")
```

# ANALISI DELLA SOPRAVVIVENZA

## Stimatore di Kaplan-Meier

Per il calcolo della funzione di sopravvivenza

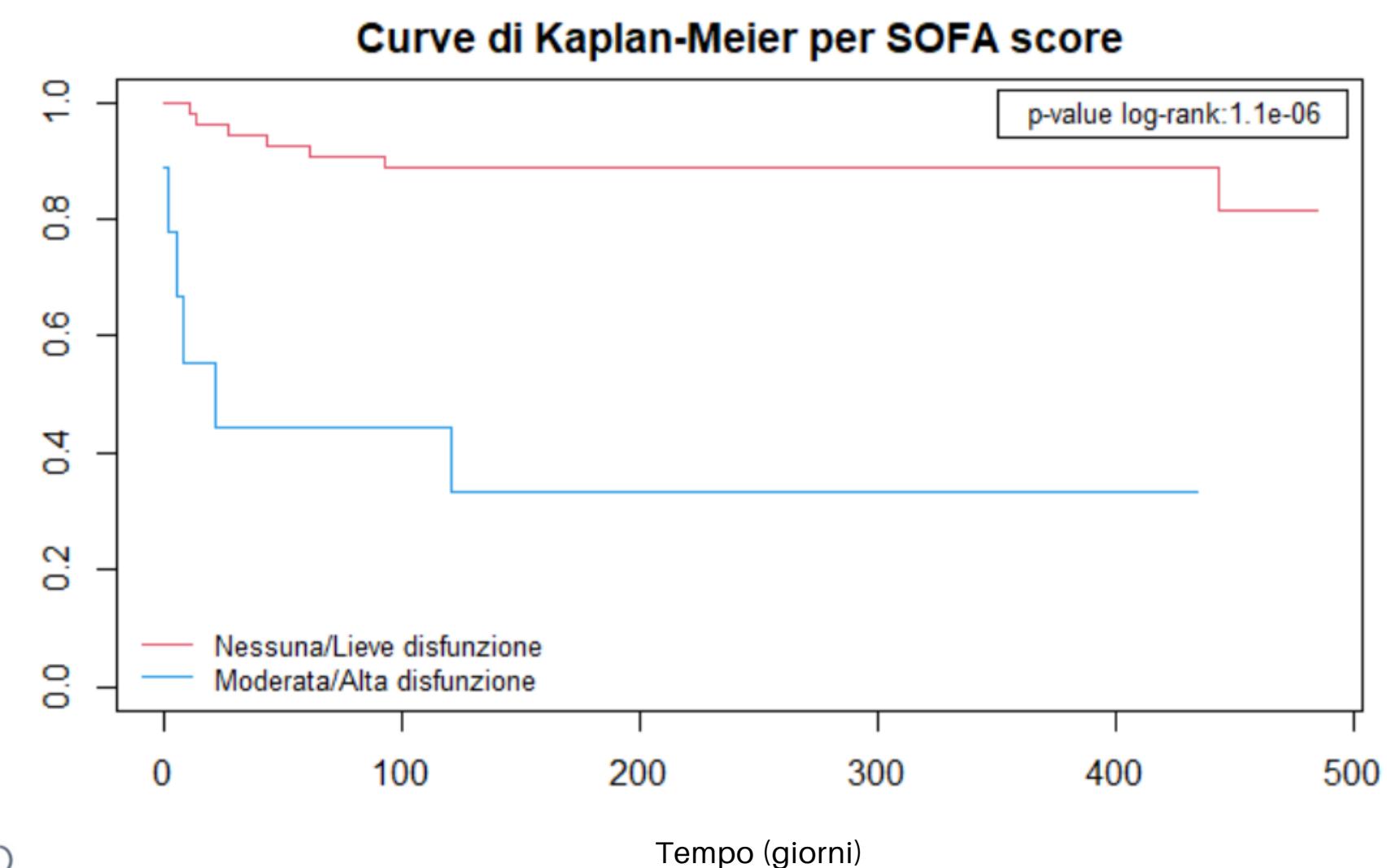
$$S(t_j) = \prod_{i=1}^j \left( \frac{n_i - d_i}{n_i} \right) \quad j = 1, \dots, J \quad D_i: \text{numero eventi} \\ J \leq N \quad n_i: \text{numero di soggetti a rischio}$$

- Stima correttamente la curva di sopravvivenza anche con dati censurati tenendo conto del numero di soggetti che sono effettivamente a rischio di sviluppare l'evento.
- I soggetti censurati entrano nel conteggio degli *a rischio* di evento solo fino alla censura.

```
library(survival)
require(Rcpp)

km_sofacat <- survfit(Surv(time, evento) ~ SOFA_cat1, data = SOFA_pulito)
rank <- survdiff(Surv(time, evento) ~ SOFA_cat1, data=SOFA_pulito)

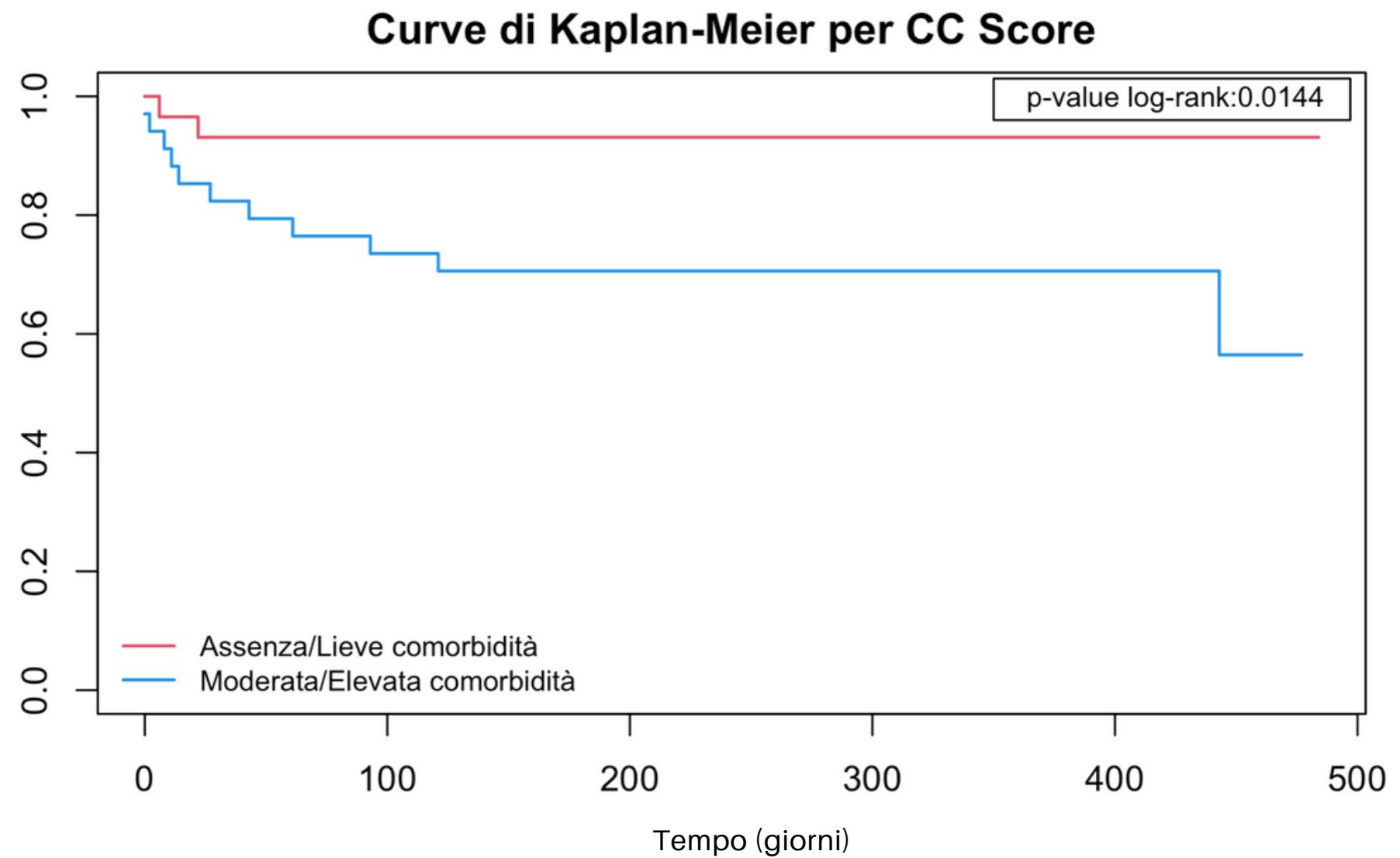
plot(km_sofacat, col = c(2,4), lwd=c(1.5,1.5), main='Curve di Kaplan-Meier per SOFA score')
legend("bottomleft", legend = c("Nessuna/Lieve disfunzione", "Moderata/Alta disfunzione"),
       col = c(2,4), lwd = c(1.5,1.5), cex = 0.8,bty="n")
text(x = 425, y = 0.98, paste0("p-value log-rank:", round(rank$pvalue,7)), cex = 0.8)
rect(xleft = 350, ybottom = 0.94, xright = 497, ytop = 1.02, lwd = 1,
      lty="solid" , col = "transparent")
```



# ANALISI DELLA SOPRAVVIVENZA

## Stimatore di Kaplan-Meier

```
km_ccscorecat <- survfit(Surv(time, evento) ~ SOFA_pulito$CCSCORE_cat1,  
                           data = SOFA_pulito)  
rank <- survdiff(Surv(time, evento) ~ CCSCORE_cat1, data=SOFA_pulito)  
plot(km_ccscorecat, col = c(2,4), lwd=c(1.5,1.5))  
legend("bottomleft", legend = c("Assenza/Lieve comorbidità",  
                               "Moderata/Elevata comorbidità"),  
      col = c(2,4), lwd=c(1.5,1.5), cex = 0.8, bty = "n")  
  
text(x = 425, y = 0.996, paste0("p-value log-rank:", round(rank$pvalue,4)),  
     cex = 0.8)  
rect(xleft = 350, ybottom = 0.96, xright = 497, ytop = 1.03, lwd = 1,  
     lty="solid" , col = "transparent")
```



## Predittività SOFA vs CC score:

- Dalle curve di Kaplan-Meier emerge che il punteggio SOFA sia più predittivo del CC Score (test Log Rank)
- La situazione varia in base al campione considerato [4]
- Il CC Score è stato creato con lo scopo di predire la mortalità di pazienti anziani in relazione a determinate comorbidità

# ANALISI DELLA SOPRAVVIVENZA

## Modello di Cox

- Il modello a rischi proporzionali proposto da Cox (1972) è il modello di regressione più utilizzato per l'analisi dei dati di sopravvivenza.

Il modello assume che la funzione hazard per un individuo  $i$  caratterizzato dal vettore di covariate  $x_i = (x_{i1}, \dots, x_{ik})$ , sia

$$h_i(t) = \lambda_0(t) \exp(\beta_1 x_{i1} + \dots + \beta_k x_{ik})$$

- L'hazard per un individuo  $i$  al tempo  $t$  è il prodotto di due fattori:
  - Una funzione  $\lambda_0(t)$  che non viene specificata
  - Una combinazione lineare di un insieme di  $k$  covariate fisse, che viene poi esponenziata  $\lambda_0(t)$  è la funzione hazard per un individuo le cui covariate hanno tutte valore pari a 0. Viene definito baseline hazard

$$HR = \frac{h_i(t)}{h_j(t)} = \exp\{\beta_1(x_{i1} - x_{j1}) + \dots + \beta_k(x_{ik} - x_{jk})\}$$

- La misura dell'effetto relativo (HR) richiede solo la stima dei  $\beta$ , e non di  $\lambda_0(t)$

# ANALISI DELLA SOPRAVVIVENZA

Modello di Cox - solo Covariata di Interesse

## DATI NON IMPUTATI

```
SOFA      coef exp(coef)  se(coef)      z Pr(>|z|)  
2.3064   10.0381  0.6377  3.617  0.000298 ***  
---  
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
CCSCORE    coef exp(coef)  se(coef)      z Pr(>|z|)  
1.5141   4.5454  0.7827  1.934  0.0531 .  
---  
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

## DATI IMPUTATI

```
SOFA      coef exp(coef)  se(coef)      z Pr(>|z|)  
2.3053   10.0271  0.5818  3.962  7.43e-05 ***  
---  
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
CCSCORE    coef exp(coef)  se(coef)      z Pr(>|z|)  
1.6808   5.3697  0.7696  2.184  0.029 *  
---  
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

- Risultati pressoché analoghi tra modello con dati imputati e modello con dati mancanti
- Il modello di Cox che considera soltanto SOFA predice meglio la mortalità rispetto al modello con solo il CC Score come covariata

# ANALISI DELLA SOPRAVVIVENZA

Modello di Cox: Model Selection - Dati imputati

```
library(survival)
library(MASS)
surv <- Surv(SOFA_pulito$time, SOFA_pulito$evento)
model <- coxph(surv ~ SOFA+Sesso+`Stato Civile`+`Num. cadute`+`CC Score`+
    MMSE+Albumina+Calcio+Emoglobina+`Durata intervento`+
    `Tempo ricovero`+Anestesia+Età+BMI, data = SOFA_pulito)
step_model <- stepAIC(model, direction = "backward")
summary(step_model)
```

Il modello con AIC minore risulta essere quello con:

- SOFA
- Durata Intervento
- BMI
- Emoglobina
- Tipologia di anestesia
- MMSE (Mini-Mental State Examination)

```
Step: AIC=86.72
surv ~ SOFA + MMSE + Emoglobina + `Durata intervento` + Anestesia +
    BMI
```

	Df	AIC
<none>		86.721
- `Durata intervento`	1	87.250
- BMI	1	87.774
- Emoglobina	1	88.042
- Anestesia	1	89.131
- MMSE	2	89.348
- SOFA	1	103.819

```
Call:
coxph(formula = surv ~ SOFA + MMSE + Emoglobina + `Durata intervento` +
    Anestesia + BMI, data = SOFA_pulito)

n= 63, number of events= 13
```

# ANALISI DELLA SOPRAVVIVENZA

Modello di Cox: Modello migliore

	coef	exp(coef)	se(coef)	z	Pr(> z )	
SOFA - 1 vs 0	3.21888	25.00009	0.74704	4.309	1.64e-05	***
MMSE - 1 vs 0	1.06308	2.89526	1.57728	0.674	0.5003	
MMSE - 2 vs 0	2.22708	9.27278	1.01383	2.197	0.0280	*
Emoglobina	-0.40859	0.66458	0.22028	-1.855	0.0636	.
`Durata intervento`	0.02043	1.02064	0.01273	1.605	0.1085	
Anestesia - 2 vs 1	1.64686	5.19066	0.89465	1.841	0.0657	.
BMI	-0.14071	0.86874	0.08510	-1.654	0.0982	.
---						
Signif. codes:	0	'***'	0.001	'**'	0.01	'*'
	0.05	'.'	0.1	'.'	1	

Coloro che hanno *Moderata/Alta disfunzione* hanno un rischio **27** volte superiore di morire rispetto a coloro che hanno *Nessuna/Lieve disfunzione*

→ Coloro che hanno *Moderata/Alta disfunzione* hanno un rischio **25** volte superiore di morire rispetto a coloro che hanno *Nessuna/Lieve disfunzione*

## CON VARIABILI SOCIO-DEMOGRAFICHE (AIC=90.20)

	coef	exp(coef)	se(coef)	z	Pr(> z )	
SOFA - 1 vs 0	3.31036	27.39487	0.87758	3.772	0.000162	***
MMSE - 1 vs 0	0.78001	2.18150	1.71027	0.456	0.648334	
MMSE - 2 vs 0	2.05096	7.77536	1.12249	1.827	0.067677	.
Emoglobina	-0.37811	0.68515	0.22328	-1.693	0.090364	.
`Durata intervento`	0.01730	1.01745	0.01400	1.236	0.216565	
Anestesia - 2 vs 1	1.77282	5.88741	0.95557	1.855	0.063562	.
BMI	-0.14305	0.86671	0.09164	-1.561	0.118521	
Età	0.03382	1.03440	0.05343	0.633	0.526786	
Sesso - 2 vs 1	0.08393	1.08756	0.83574	0.100	0.920004	
---						
Signif. codes:	0	'***'	0.001	'**'	0.01	'*'
	0.05	'.'	0.1	'.'	1	

# ANALISI DELLA SOPRAVVIVENZA

Modello di Cox: Model Selection - Dati non imputati

Step: AIC=64.95

surv ~ SOFA + `Num. cadute` + Albumina + Calcio + Età + BMI

Aggiungendo anche il genere nell'analisi:

AIC=65.88

Il modello con AIC minore risulta essere quello con:

- SOFA
- Numero di cadute
- Calcio
- Albumina
- Età }
- BMI }



2 Variabili Socio-Demografiche significative

		coef	exp(coef)	se(coef)	z	Pr(> z )	
	SOFA - 1 vs 0	3.72932	41.65089	1.09265	3.413	0.000642	***
	`Num. cadute`	-1.51764	0.21923	0.90556	-1.676	0.093754	.
	Albumina	-2.18877	0.11205	1.10701	-1.977	0.048021	*
	Calcio	0.82605	2.28428	0.48233	1.713	0.086778	.
	Età	0.14526	1.15634	0.07181	2.023	0.043075	*
	BMI	-0.26920	0.76399	0.10293	-2.615	0.008914	**
	Sesso - 2 vs 1	0.97219	2.64372	0.96966	1.003	0.316050	
	---						
	Signif. codes:	0	***	0.001	**	0.01	*, 0.05 ., 0.1 , , 1



Coloro che hanno *Moderata/Alta disfunzione* hanno un rischio quasi 42 volte **superiore** di morire rispetto a coloro che hanno *Nessuna/Lieve disfunzione*

# ANALISI DELLA SOPRAVVIVENZA

## Modello di Cox: Verifica degli Assunti

### • Assunto centrale del modello di Cox:

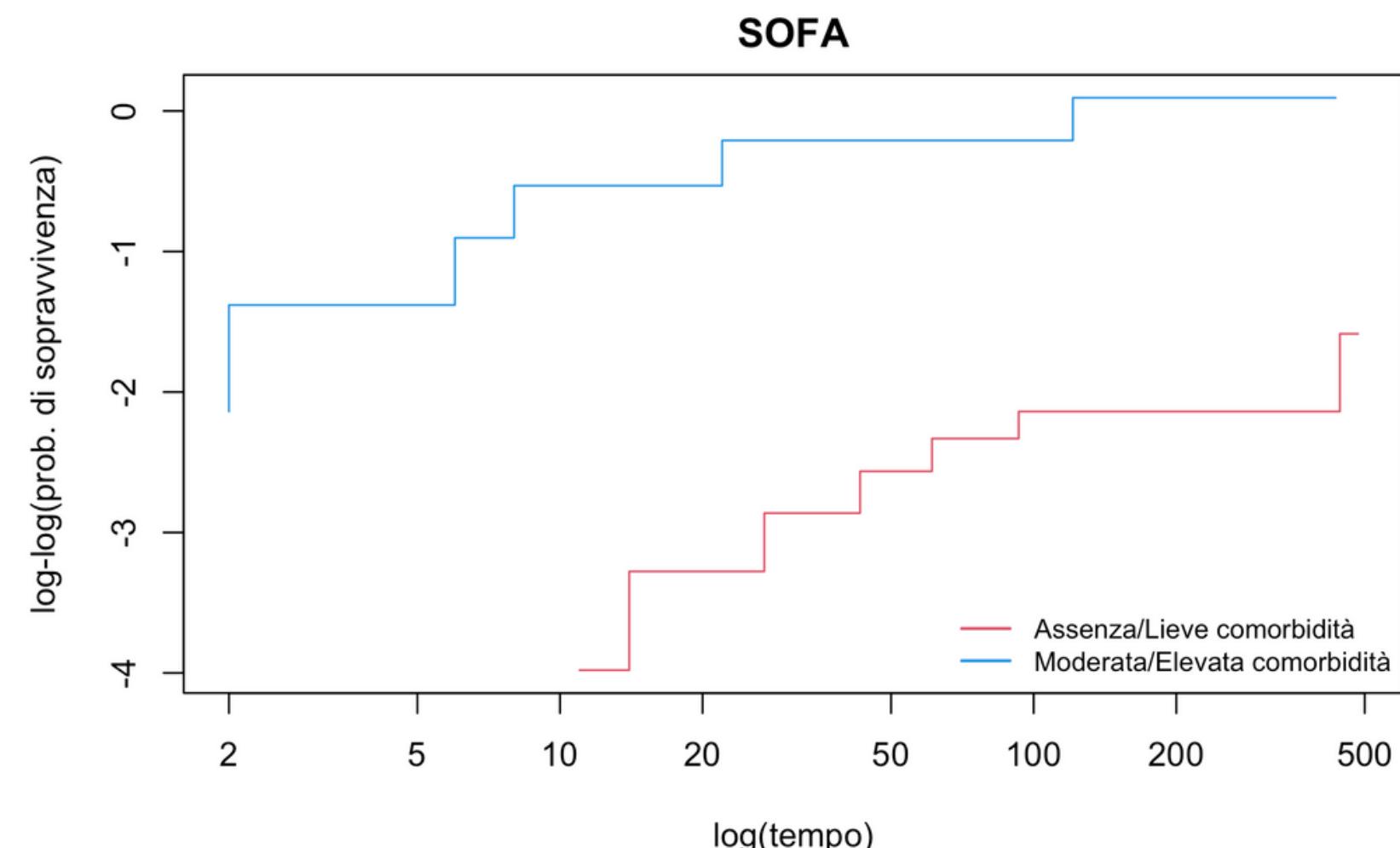
Gli effetti delle covariate sugli hazard ratio rimangano costanti nel tempo (HR proporzionali nel tempo).

- In R si usa la funzione 'cox.zph()'

	chisq	df	p
SOFA	1.3508	1	0.25
MMSE	1.7631	2	0.41
Emoglobina	0.7974	1	0.37
`Durata intervento`	0.0178	1	0.89
Anestesia	3.2726	1	0.07
BMI	2.0018	1	0.16
GLOBAL	8.8264	7	0.27

- Rispettato assunto di proporzionalità per ogni variabile del modello, anche con dati non imputati

```
km_home <- survfit(surv ~ SOFA, data = SOFA_pulito)
plot(km_home, fun = "cloglog", xlab = "log(tempo)", ylab = "log-log(prob. di sopravvivenza)",
     main = "SOFA", col=c(2,4))
legend("bottomright", legend = c("Assenza/Lieve comorbidità", "Moderata/Elevata comorbidità"),
       col = c(2,4), lwd=c(1.5,1.5), cex = 0.8, bty = "n")
```



# ANALISI DELLA SOPRAVVIVENZA

## RMST

- L'**RMST** (*Restricted Mean Survival Time* - tempo medio di sopravvivenza ristretto) è il tempo medio di sopravvivenza fino ad un determinato tempo  $\tau$ , corrisponde all'area sotto la curva di sopravvivenza calcolata fino al tempo  $\tau$  ed costituisce l'aspettativa di vita in un orizzonte temporale di  $\tau$  giorni. [5]
- La differenza (o rapporto) tra **RMST** è una misura alternativa all'Hazard Ratio per sintetizzare l'effetto di una covariata, ma è più semplice da comunicare a pazienti e clinici. Corrisponde ai *giorni* di vita guadagnati dagli individui con *lieve/nessuna disfunzione* rispetto agli individui con *moderata/alta disfunzione* in un orizzonte temporale di  $\tau$  giorni

### Sopravvivenza a 6 mesi

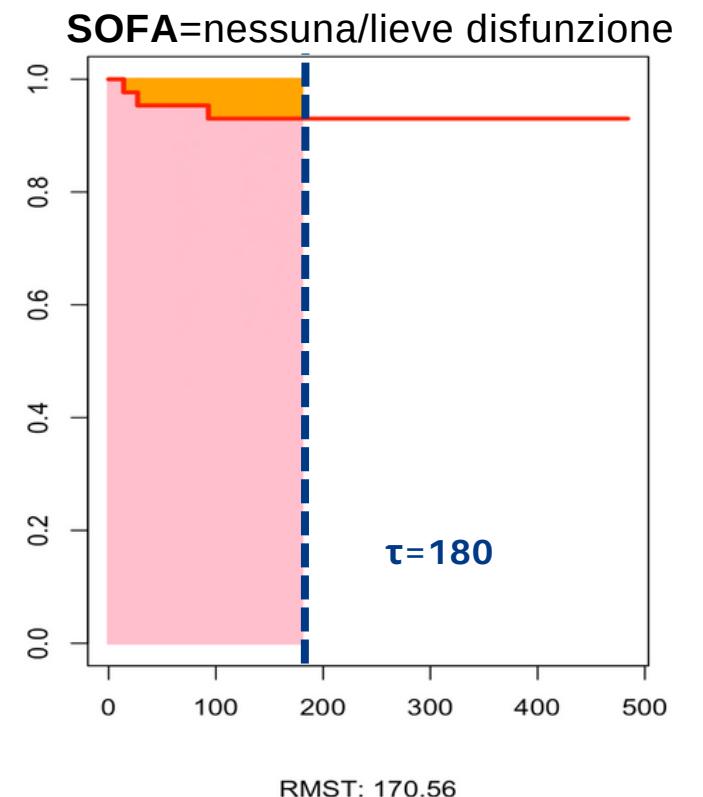
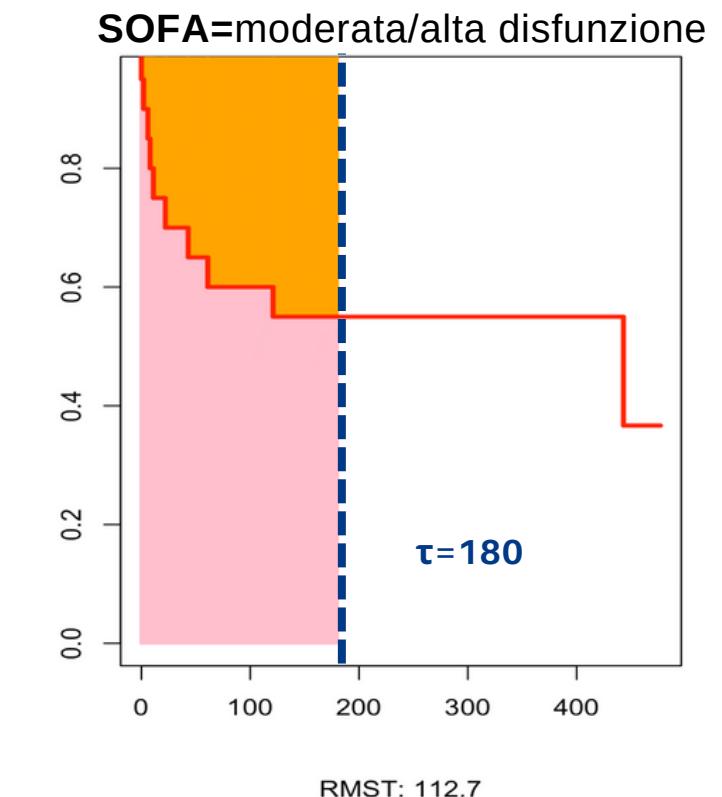
The truncation time: tau = 180 was specified.

Restricted Mean Survival Time (RMST) by arm

	Est.	se	lower .95	upper .95
RMST (arm=1)	112.700	17.557	78.290	147.110
RMST (arm=0)	170.558	5.439	159.898	181.219

Between-group contrast

	Est.	lower .95	upper .95	p
RMST (arm=1)-(arm=0)	-57.858	-93.882	-21.834	0.002
RMST (arm=1)/(arm=0)	0.661	0.484	0.902	0.009



# CRITICITA' ED ASPETTI DA APPROFONDIRE

## 1 Numerosità

- Conclusioni non del tutto coerenti con la letteratura

## 2 Contattare clinici

- Contattare i clinici per ricavare alcuni dati mancanti

## 3 Vitamina D

- Imputare la variabile "Vitamina D" tramite la funzione "Mice"
- Si pone  $m=25$  (% di dati mancanti)
- Si effettua stima pulled su tutti i dataset ottenuti dall'imputazione

## 4 Imputazione MICE

- L'imputazione MICE non sembra essere risolutiva nella gestione dei dati mancanti
- Si sarebbe potuto approfondire l'analisi con ulteriori metodi di imputazione
- E' favorita l'eliminazione delle osservazioni, in disaccordo con la letteratura [6]
- I dati mancanti potrebbero essere MNAR (Missing Not At Random)

# BIBLIOGRAFIA

---

- [1] Goyal A, Anastasopoulou C, Ngu M, Singh S. Hypocalcemia. 2022 Jul 24. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2022 Jan-. PMID: 28613662.
- [2] Campos Munoz A, Jain NK, Gupta M. Albumin Colloid. [Updated 2022 Jun 27]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2022 Jan-. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK534241/>
- [3] Balvers MG, Brouwer-Brolsma EM, Endenburg S, de Groot LC, Kok FJ, Gunnewiek JK. Recommended intakes of vitamin D to optimise health, associated circulating 25-hydroxyvitamin D concentrations, and dosing regimens to treat deficiency: workshop report and overview of current literature. *J Nutr Sci.* 2015 May 25;4:e23. doi: 10.1017/jns.2015.10. PMID: 26090099; PMCID: PMC4463009.
- [4] Comparison of APACHE III, Charlson Comorbidity Index, and SOFA Score for Prediction of Mortality Following Acute Lung Injury  
Singh, Balwinder et al. *CHEST*, Volume 142, Issue 4, 301A
- [5] Liao JJZ, Liu GF, Wu WC. Dynamic RMST curves for survival analysis in clinical trials. *BMC Med Res Methodol.* 2020 Aug 27;20(1):218. doi: 10.1186/s12874-020-01098-5. PMID: 32854619; PMCID: PMC7534804
- [6] Lipkovich I, Kadziola Z, Xu L, Sugihara T, Mallinckrodt CH. Comparison of several multiple imputation strategies for repeated measures analysis of clinical scales: to truncate or not to? *J Biopharm Stat.* 2014;24(4):924-43. doi: 10.1080/10543406.2014.901344. PMID: 24697735.