Analysis of **Myelodysplastic Syndrome**

Applied Statistics | 26/07/2021
Manfred Nesti, Luca Caivano,
Taguhi Mesropyan, Michele Precuzzi

Tutors: Prof. F. Ieva, Dr. M., Spreafico, Dr. C. Gregorio

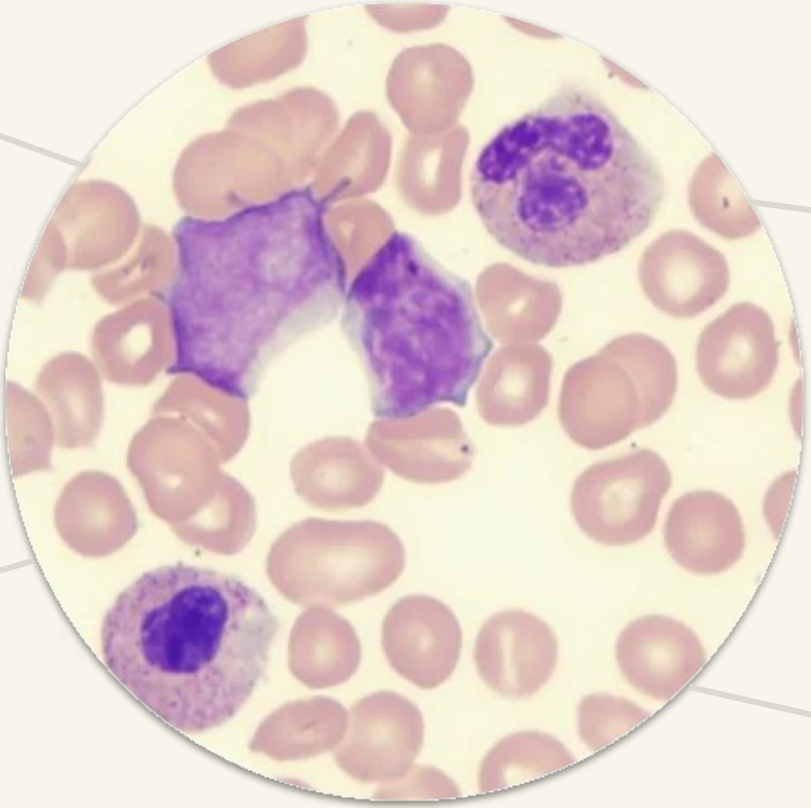
POLITECNICO MILANO 1863

HUMANITAS UNIVERSITY

IRCCS HUMANITAS RESEARCH HOSPITAL

Myelodysplastic Syndrome

very rare disease



can progress to AML

very high mortality

only cure is transplantation

Roadmap


Exploration and Preprocessing
Survival Analysis
Survival Cluster Analysis

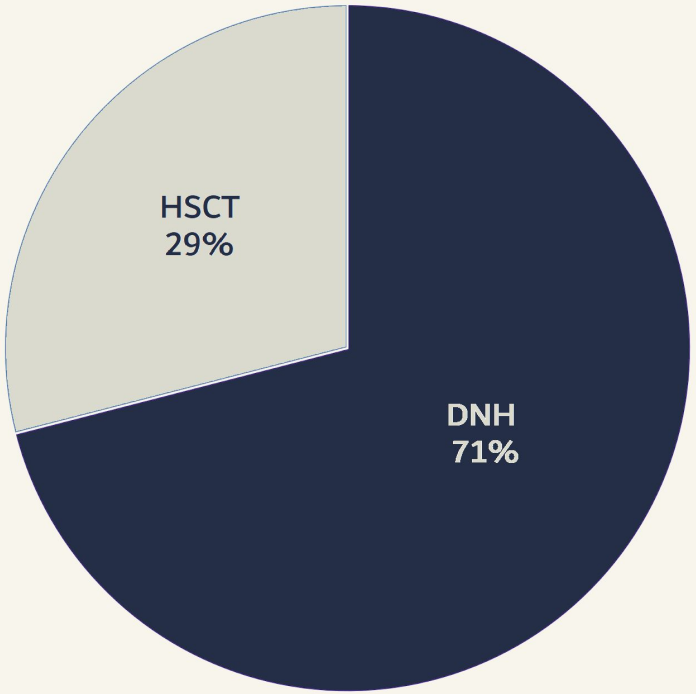
Exploration & Preprocessing

Dataset

	Cohort	IPSSR	Gender	Age	Genomic group	Hemoglobin (g / dL)	Leucocytes ($\times 10^9$ / L)	...
1	HSCT	*NA*	M	56	0	10.6	5.50	. . .
2	DNH	2	F	46	6	9.90	5.0	. . .
3	DNH	4	M	73	3	9.40	3.63	. . .

Dataset

Number of patients: 2876



HSCT = “Transplanted”

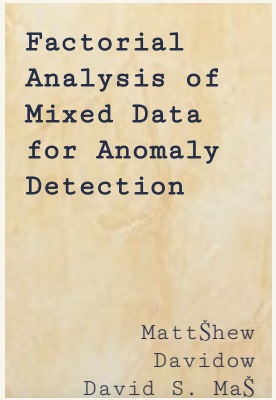
DNH = “Not Transplanted”

Factorial Analysis of Mixed Data (FAMD)

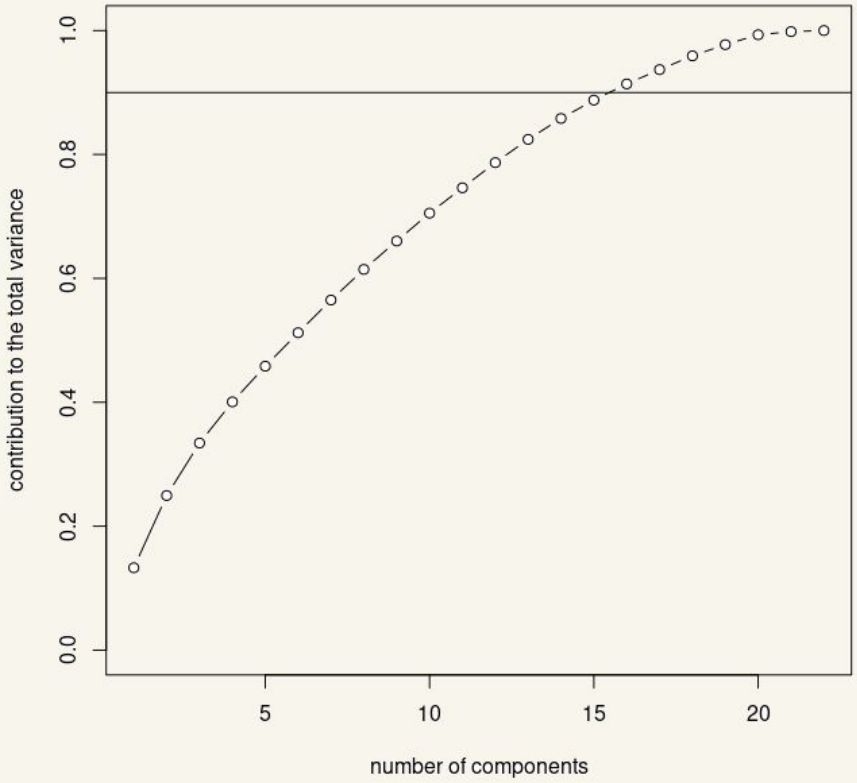
PCA

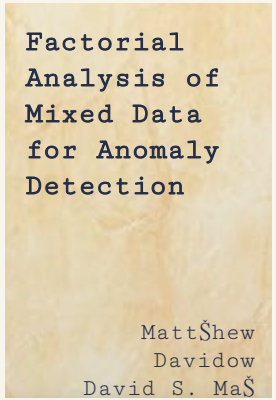
MCA

Takes into account different levels of categorical variables



Factorial Analysis of Mixed Data (FAMD)

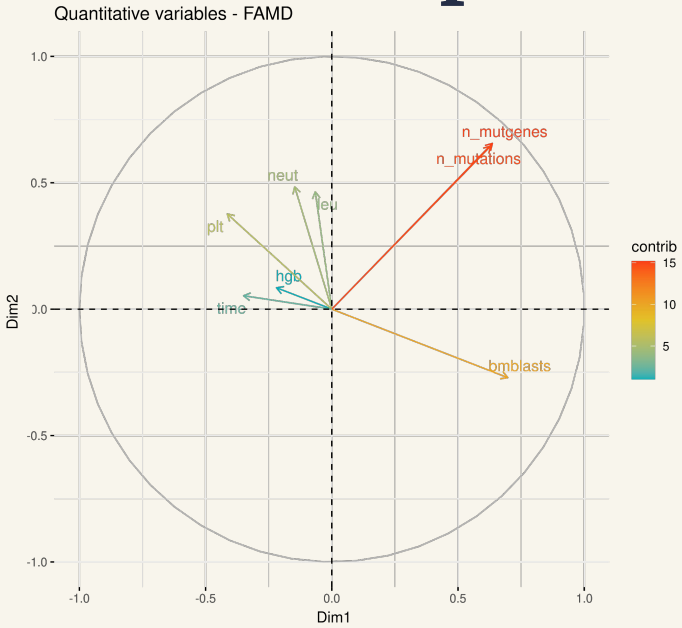


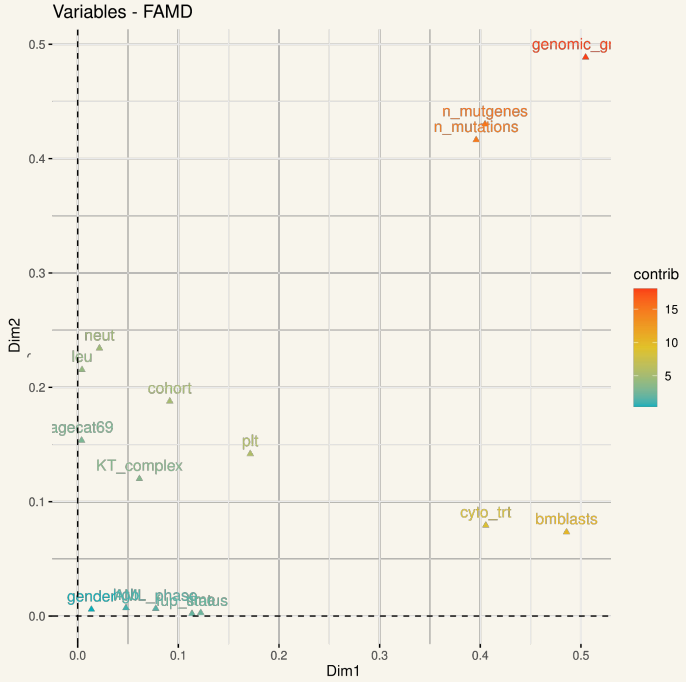


No clear elbow

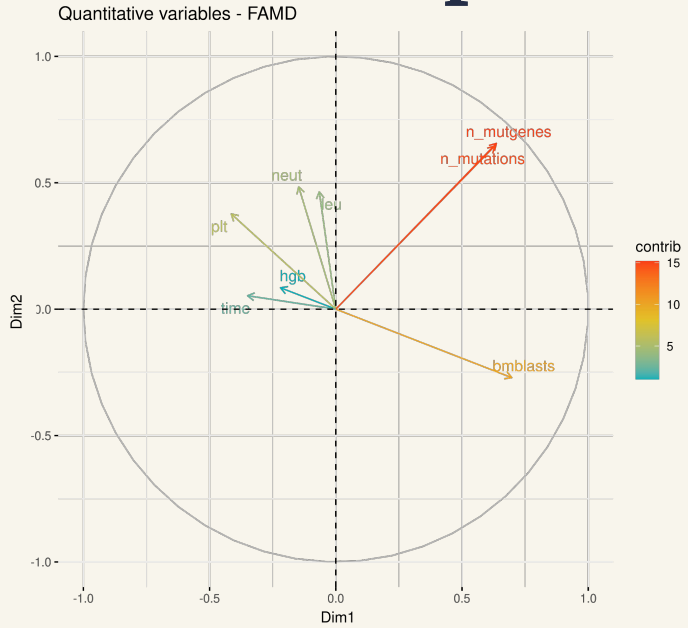
16 components out of 22 explain **90%** of variance

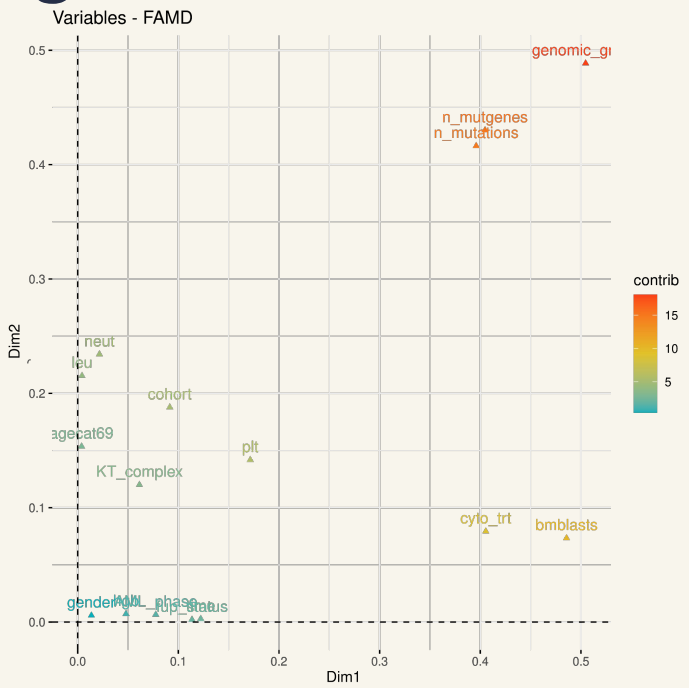
Exploration through FAMD





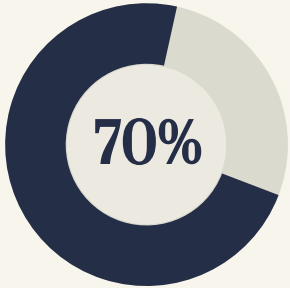
Exploration through FAMD





	Platelets	Blast Cells	N. Mutations	Surv. Time	Genomic group: 6	Cytoreduction
1st Component	−	+	+	−	−	+
2nd Component	−	+	−	·	·	·

SVM Classifier for IPSS-R



IPSS-R

2013 / 2876

Missing values

Available data

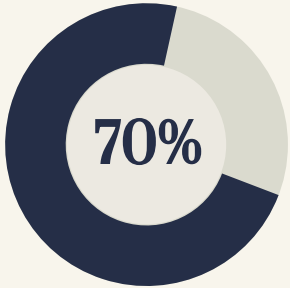
SVM Classifier for IPSS-R

SVM

Linear kernel

Cost: **10**

- Hemoglobin
- Gender
- Survival time
- Age
- Genomic group
- Leucocytes
- Platelets
- . . .



IPSS-R

2013 / 2876

Missing values

Available data

SVM Classifier for IPSS-R

SVM

Linear kernel

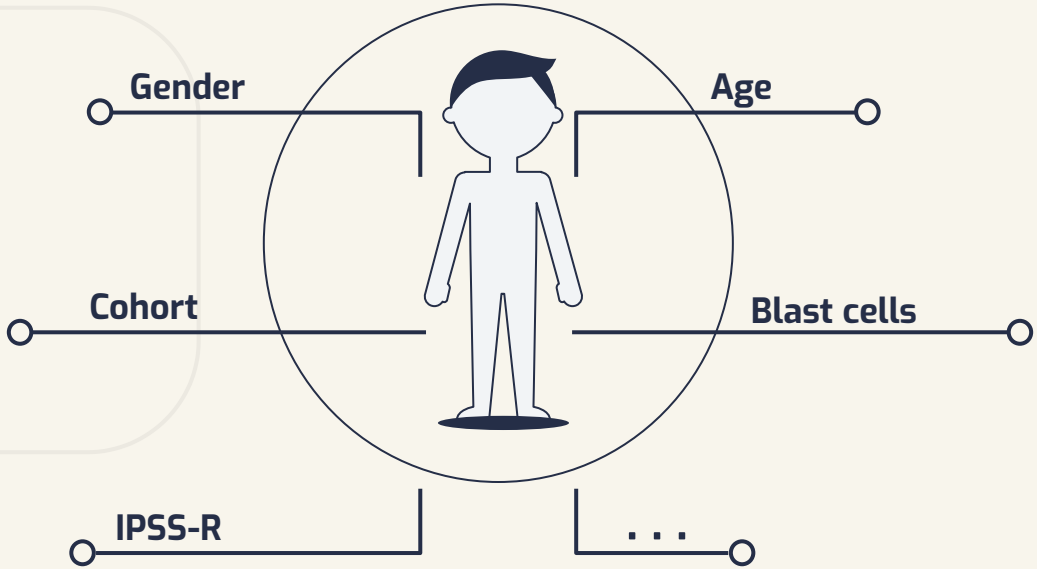
Cost: **10**

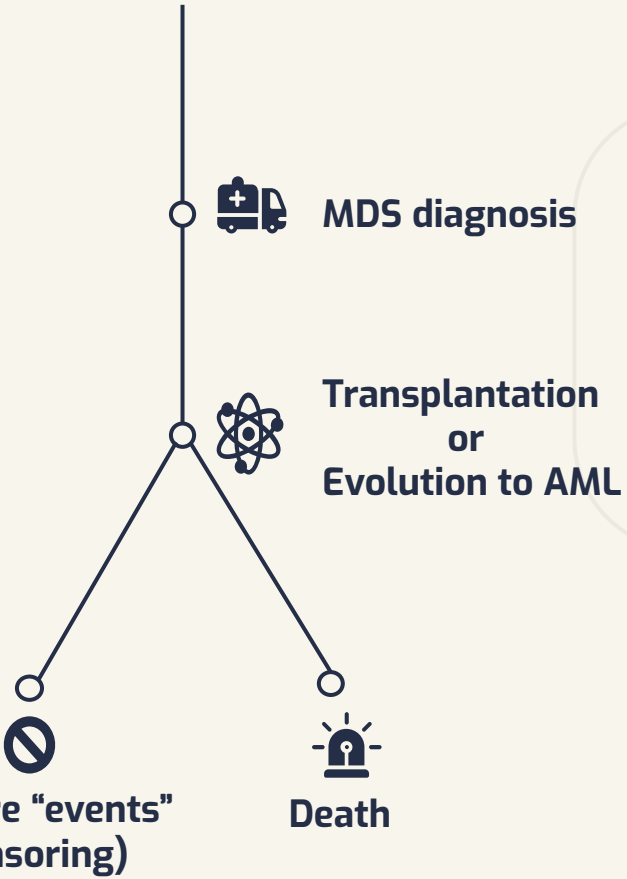
- Hemoglobin
- Gender
- Survival time
- Age
- Genomic group
- Leucocytes
- Platelets
- . . .

Predicted True	Very low	Low	Intermediate	High	Very high
Very low	**4**	5	0	0	0
Low	1	**25**	6	0	0
Intermediate	0	7	**10**	3	0
High	0	0	9	**10**	3
Very high	0	0	0	5	**18**

Survival Analysis

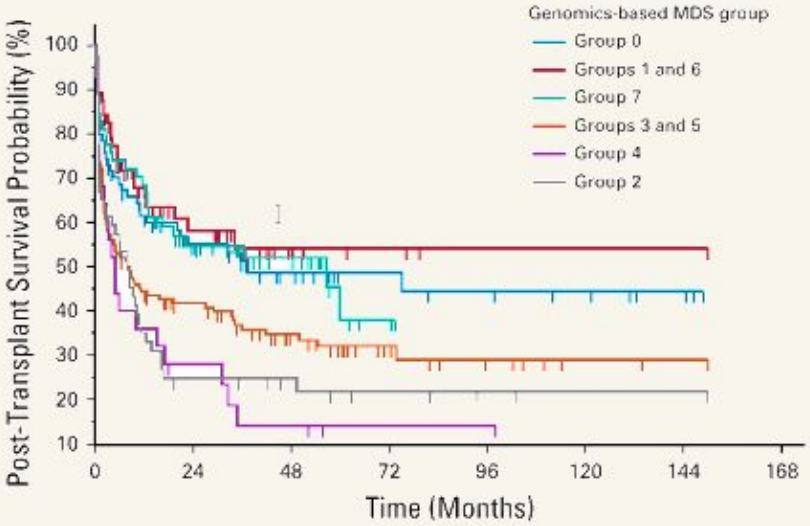
Patient Medical History





Kaplan-Meier Survival Curves

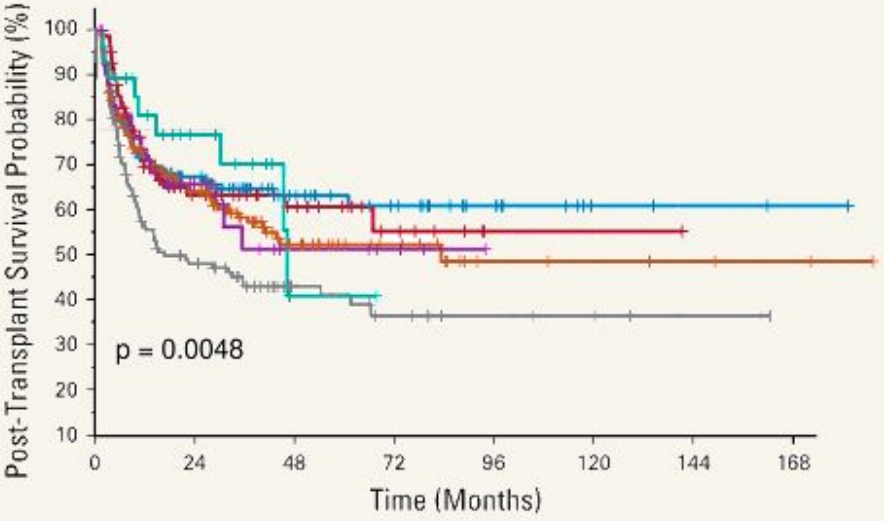
- **Non-parametric** method
- Estimates the **survival probability** from observed survival times
- **Step-function** with survival probabilities with respect to time
- Estimator of **median survival**
- Takes into account **censoring** of data

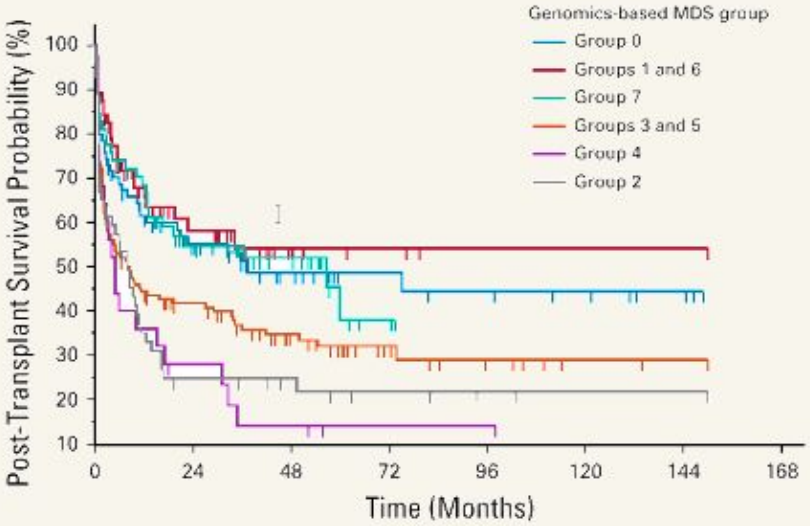


Classification and Personalized Prognostic Assessment on the Basis of Clinical and Genomic Features in MDS

Matteo Bersanelli et al.

Kaplan-Meier Survival Curves





0.0048

P-Value of **Log-Rank Test**

H_0 : all survival curves are equivalent

vs

H_1 : at least one differs

Classification and Personalized Prognostic Assessment on the Basis of Clinical and Genomic Features in MDS

Matteo Bersanelli et al.

Cox Regression

Cox Regression

$$h(t) = h_0(t)e^{\sum_{i=1}^{n} \beta_i X_i}$$

- $h(t)$ = hazard ratio
- $h_0(t)$ = baseline hazard
- β_i = model coefficients
- X_i = covariates

T = time of event occurrence

$$h(t) := \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t}$$

Cox Regression

$$h(t) = h_0(t)e^{\sum_{i=1}^{n} \beta_i X_i}$$

- $h(t)$ = hazard ratio
- $h_0(t)$ = baseline hazard
- β_i = model coefficients
- X_i = covariates

Linear regression
for the exponent

T = time of event occurrence

$$h(t) := \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t}$$

Cox Regression

$$h(t) = h_0(t)e^{\sum_{i=1}^{n} \beta_i X_i}$$

- $h(t)$ = hazard ratio
- $h_0(t)$ = baseline hazard
- β_i = model coefficients
- X_i = covariates

Linear regression
for the exponent

T = time of event occurrence

$$h(t) := \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t}$$

	Coef	e Coef	SE (Coef)	P-value	
Gender	0.270	1.310	0.074	0.000	***
Cohort	0.597	1.816	0.095	3.68e-10	***
IPSS-R	0.276	1.317	0.055	6.85e-07	***
Hemoglobin	-0.108	0.898	0.020	7.85e-08	***
Leucocytes	0.043	1.043	0.008	1.46e-08	***
Platelets	0.001	0.999	2.91e-4	0.005	**
Blast cells	0.026	1.026	0.007	2.41e-5	***
Karyotype	0.744	2.105	0.128	6.28e-09	***
Age	0.664	1.942	0.078	< 2e-16	***
Genomic group: 2, 4	0.3121	1.366	0.110	0.005	**
Genomic group: 3,5,7	0.270	1.310	0.074	3.42e-4	***

Cox Regression

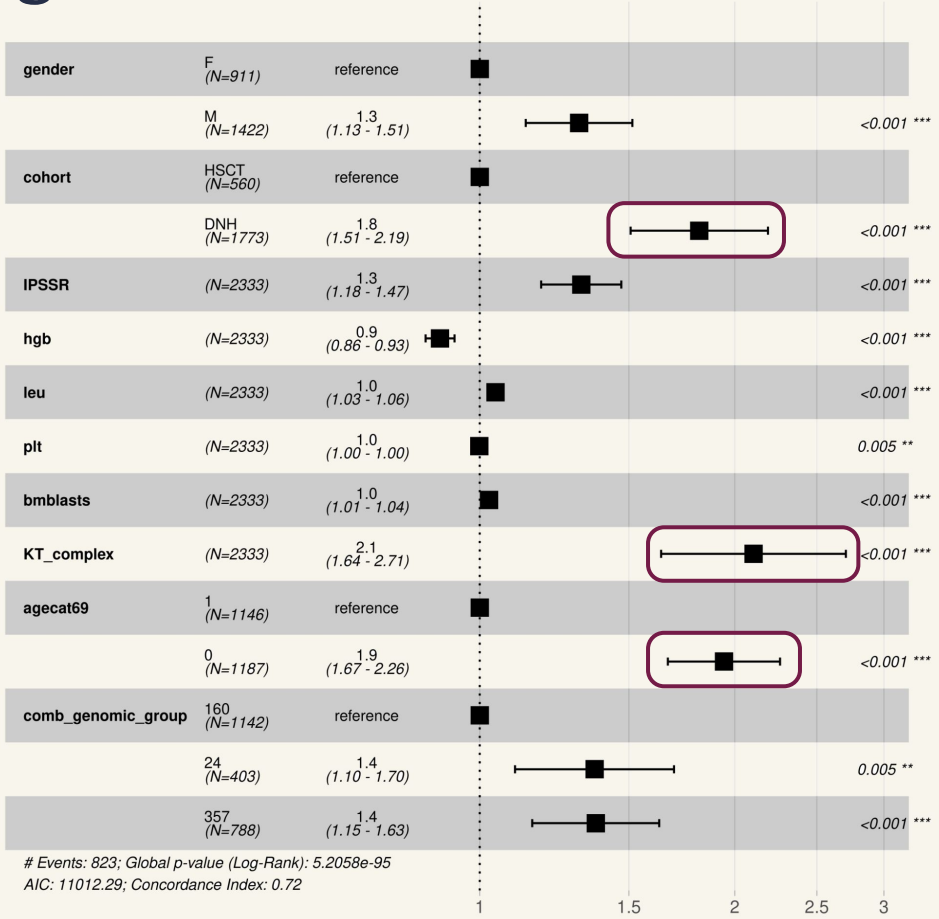
$$h(t) = h_0(t)e^{\sum_{i=1}^{n} \beta_i X_i}$$

- $h(t)$ = hazard ratio
- $h_0(t)$ = baseline hazard
- β_i = model coefficients
- X_i = covariates

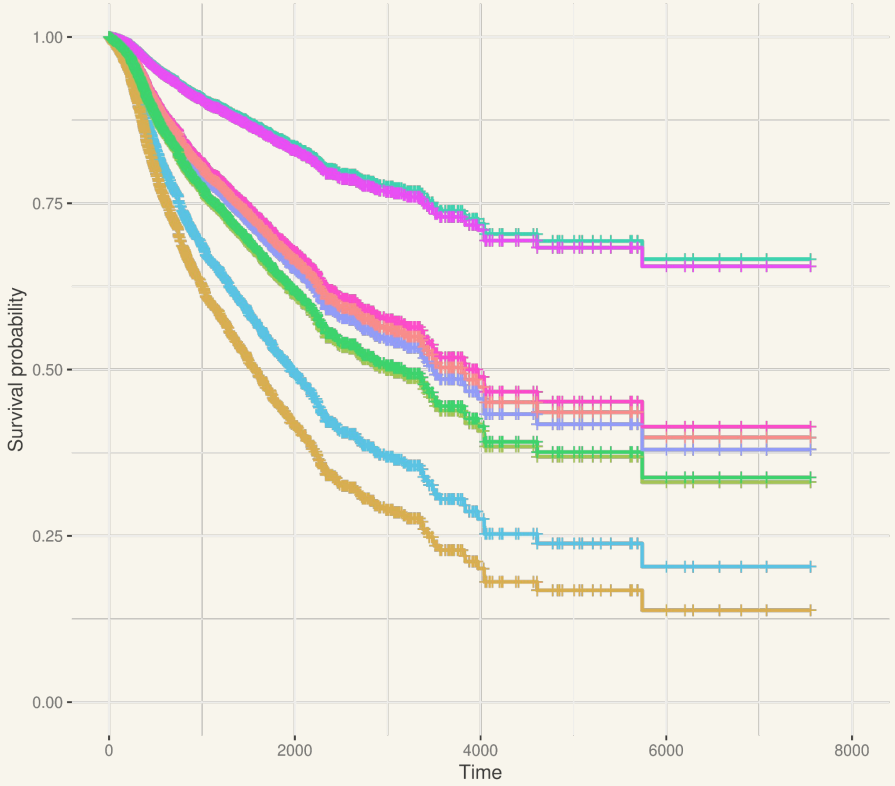
Linear regression
for the exponent

T = time of event occurrence

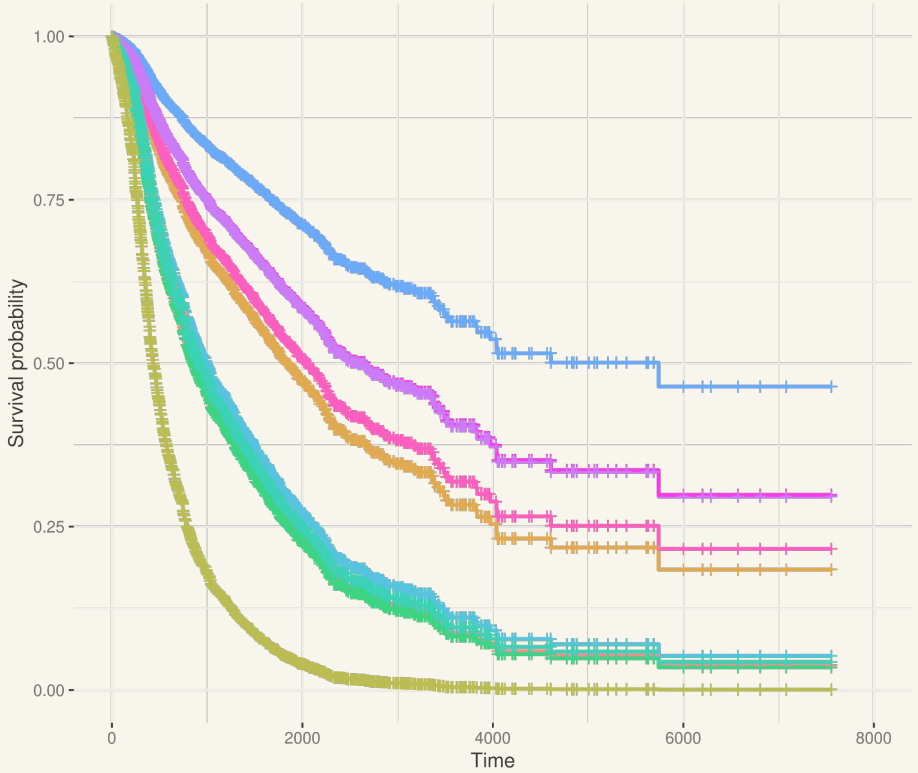
$$h(t) := \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t}$$



Survival Probability Prediction



Patients with low IPSS-R



Patients with high IPSS-R

Time Dependent Cox Models

Rely on **proportional hazard assumption**

Maybe some non-transplanted patients were supposed to get transplantation

Evolution of MDS into AML occurs in the future

Time Dependent Cox Models

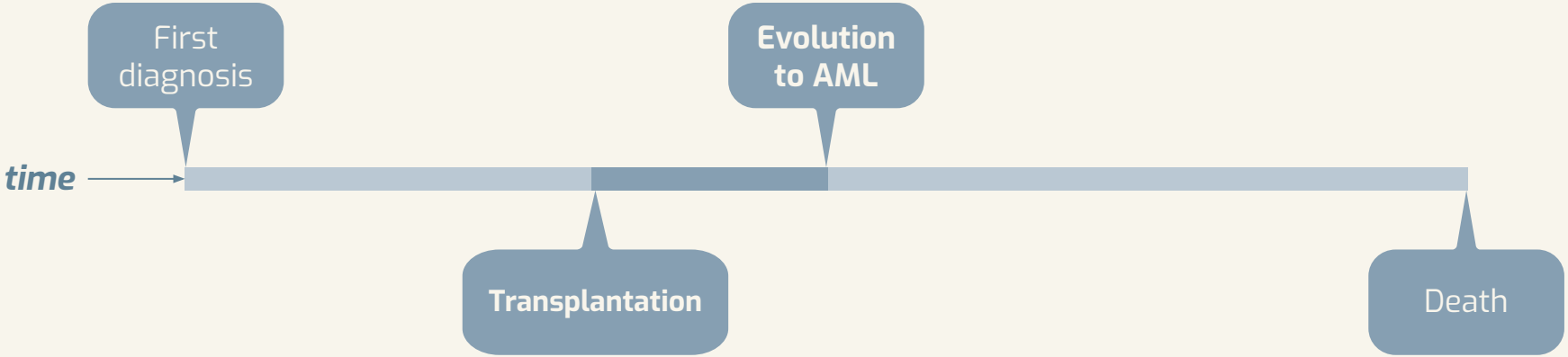
$$h(t) = h_0(t)e^{\sum_{i=1}^{n} \beta_i X_i(t)}$$

Time dependent
covariates

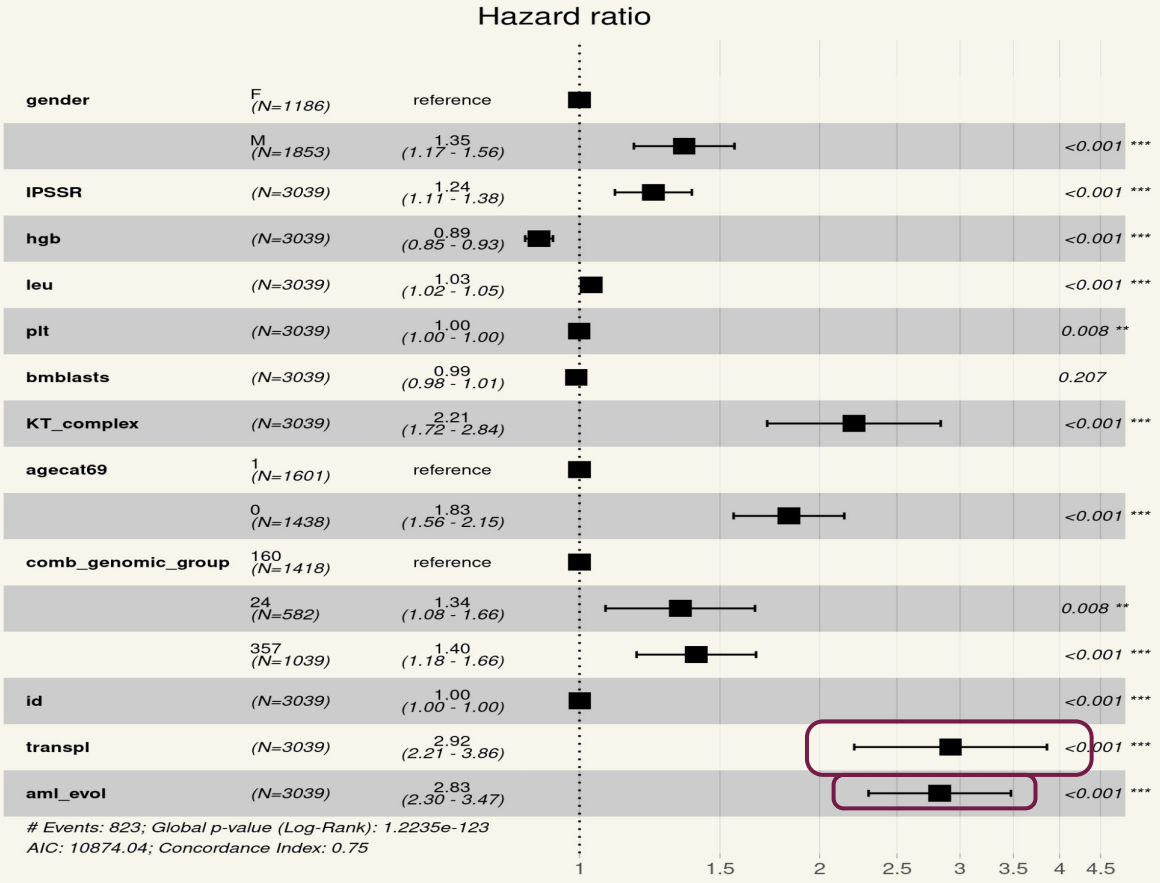
Time Dependent Cox Models

$$h(t) = h_0(t)e^{\sum_{i=1}^{n} \beta_i X_i(t)}$$

Time dependent
covariates

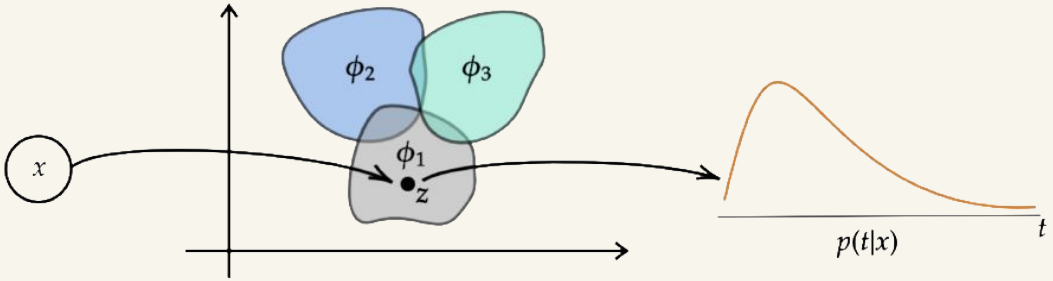


Time Dependent Cox Models



Survival Cluster Analysis

Survival Cluster Analysis



Clustering

Identifying population with distinct risk profiles

Prediction

Accurate individualized time-to-event predictions

Survival Cluster Analysis

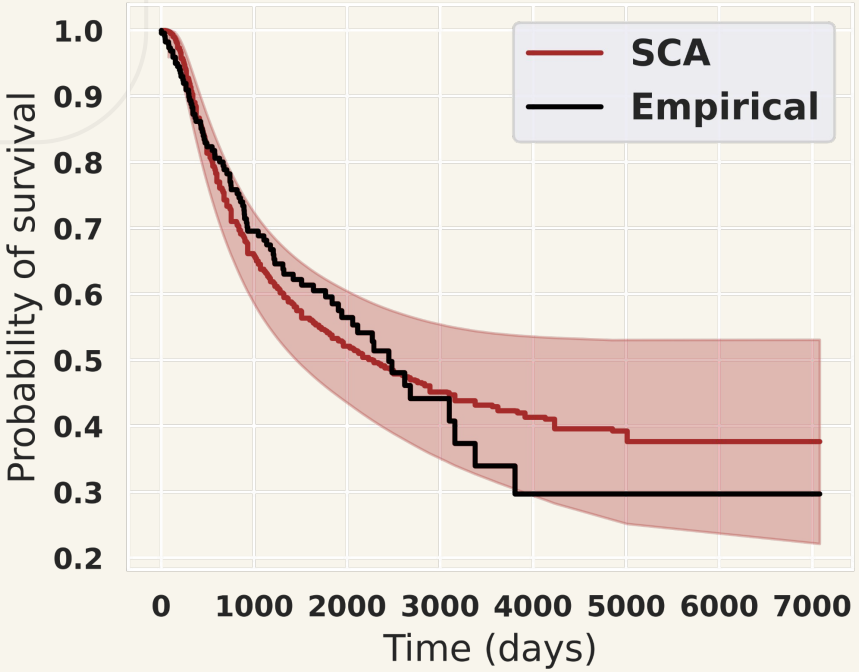
P. Chapfuwa et al.

Goodness of Fit

0.724

Concordance Index

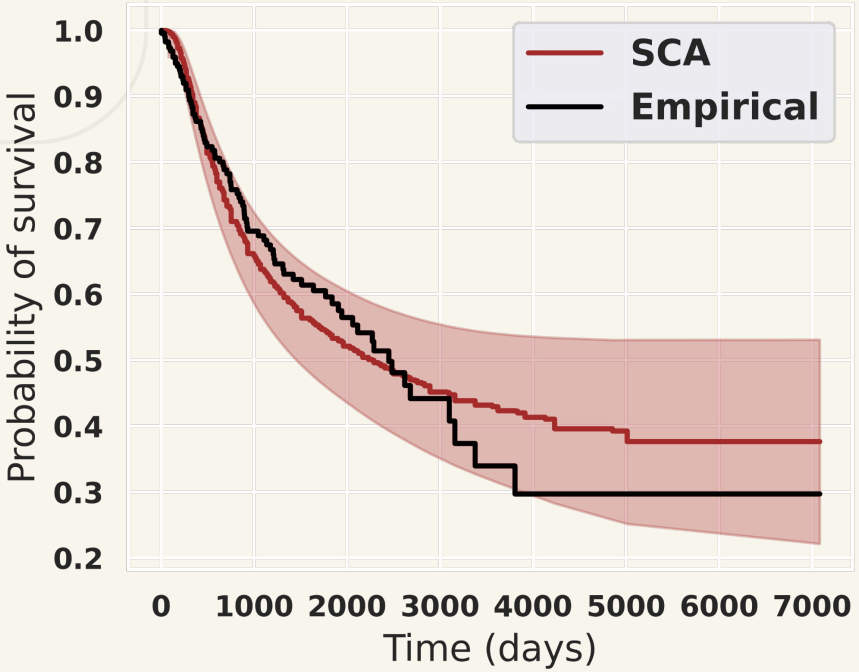
Goodness of Fit

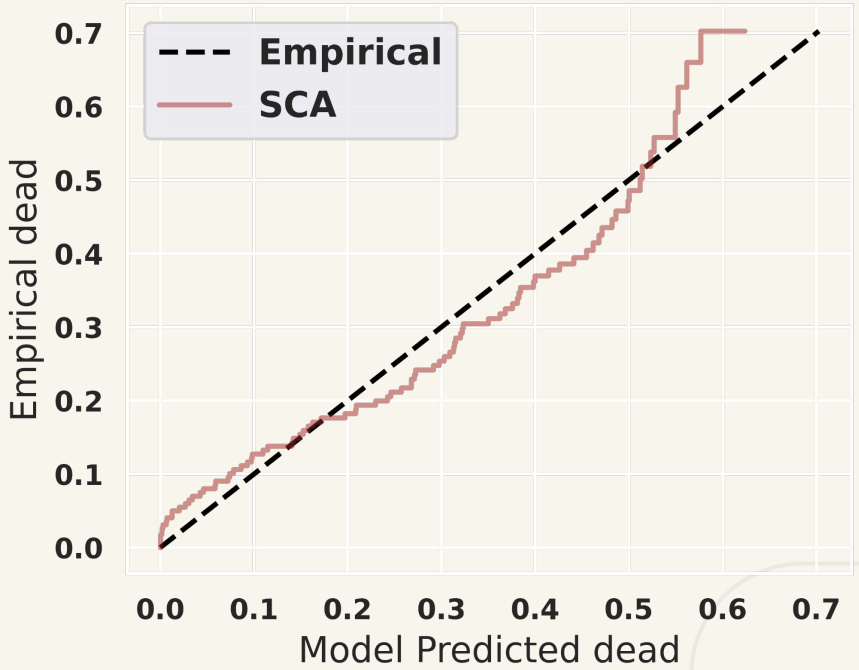


0.724

Concordance Index

Goodness of Fit



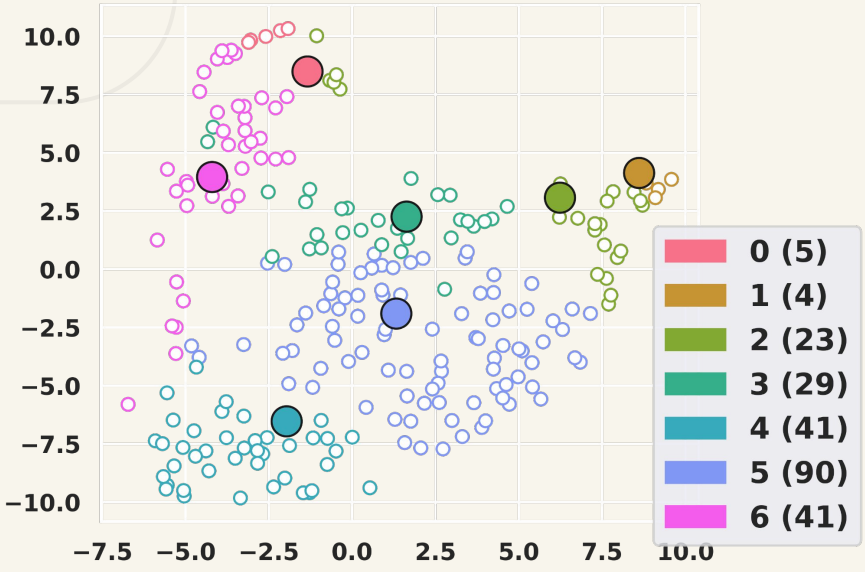


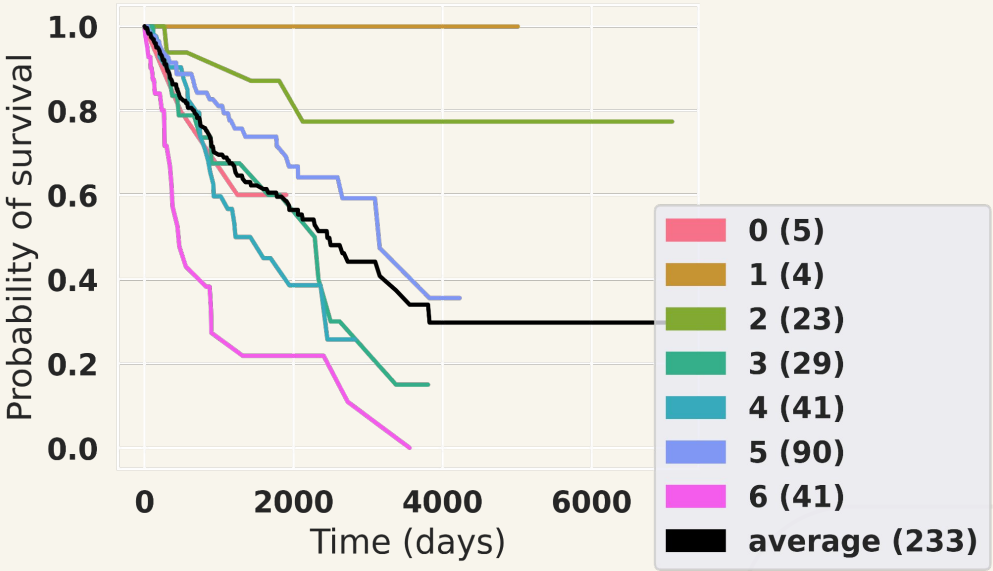
0.724
Concordance Index

0.0017
P-Value of Test H_0 : slope = 1 **vs** H_1 : slope ≠ 1

0.958
Regression slope

Survival Clusters



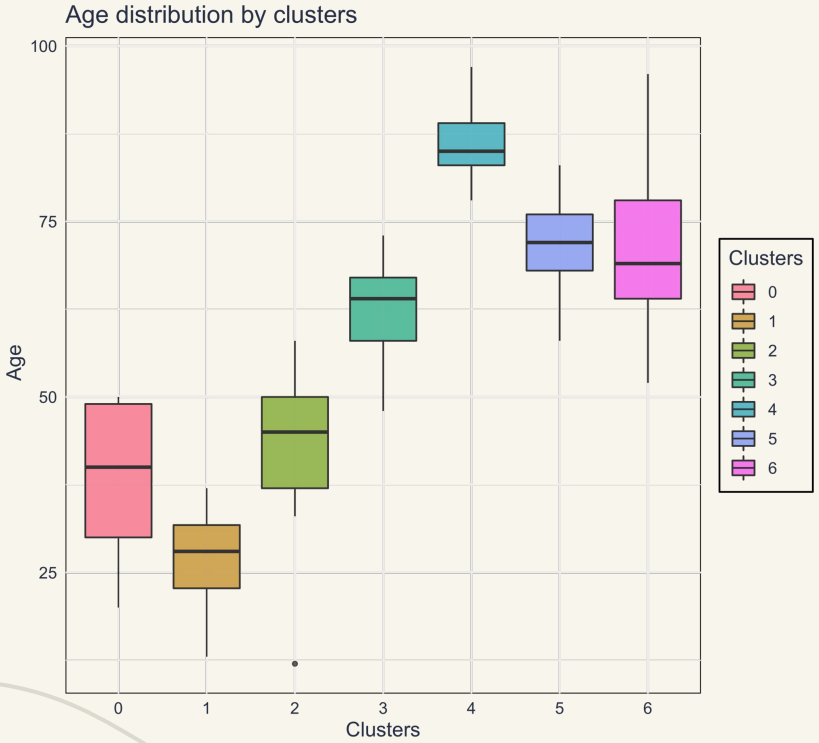


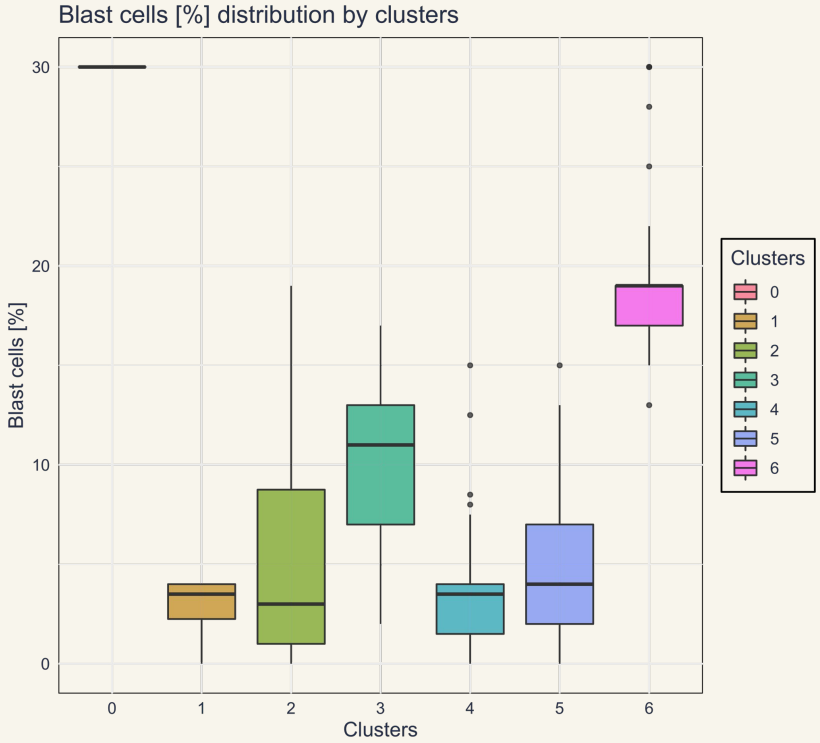
t-SNE: non-linear dimensionality reduction technique for visualizing data

0.0000

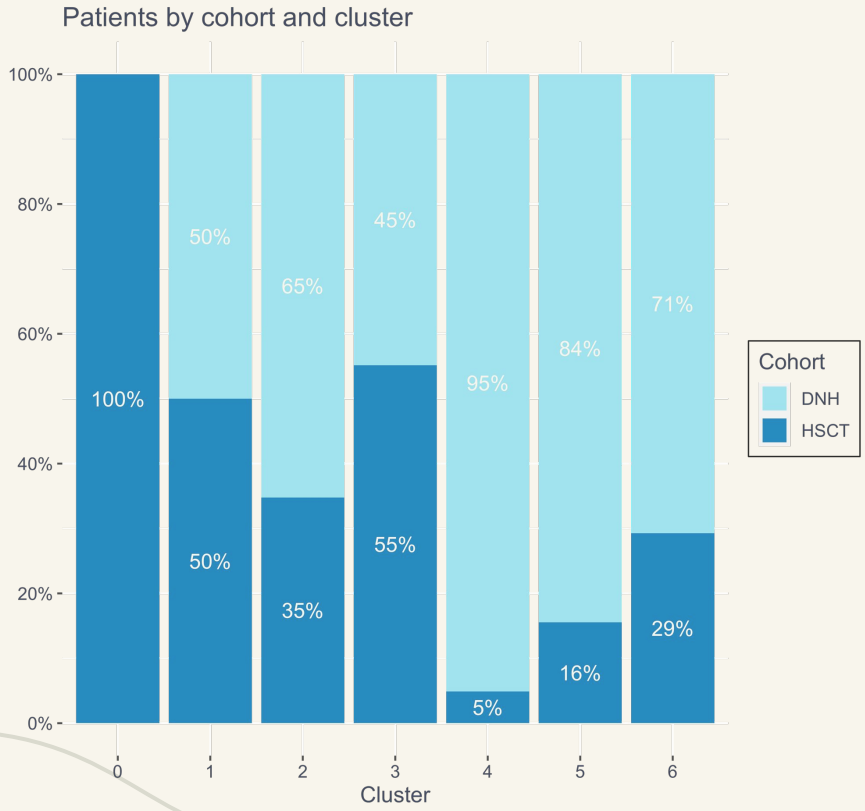
P-Value of Log-Rank Test

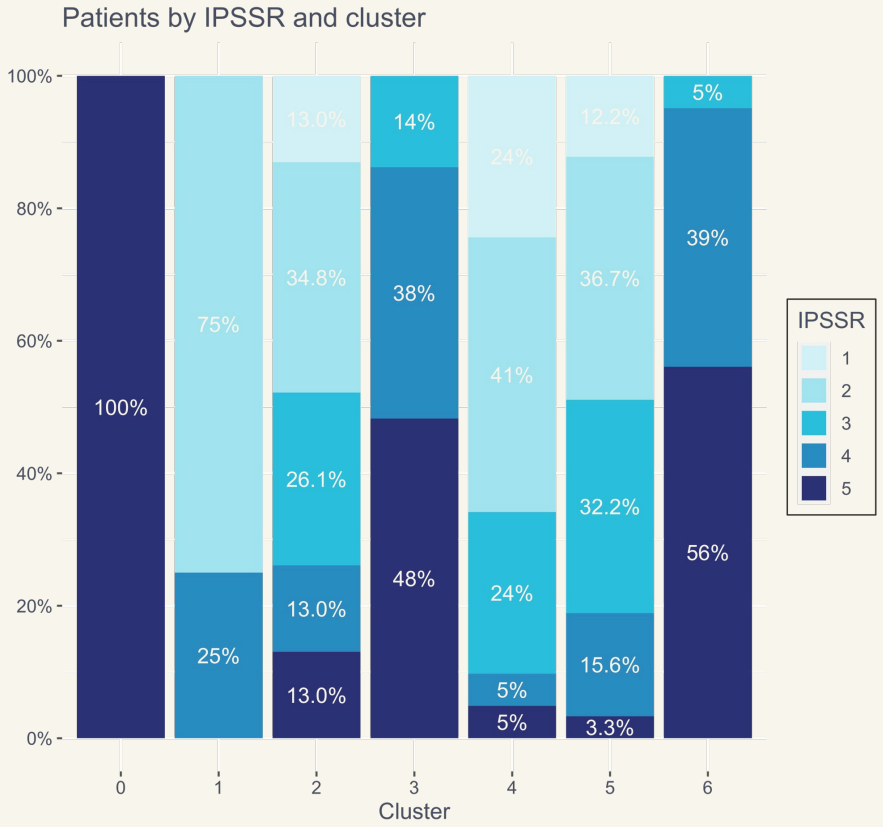
Clusters characterization



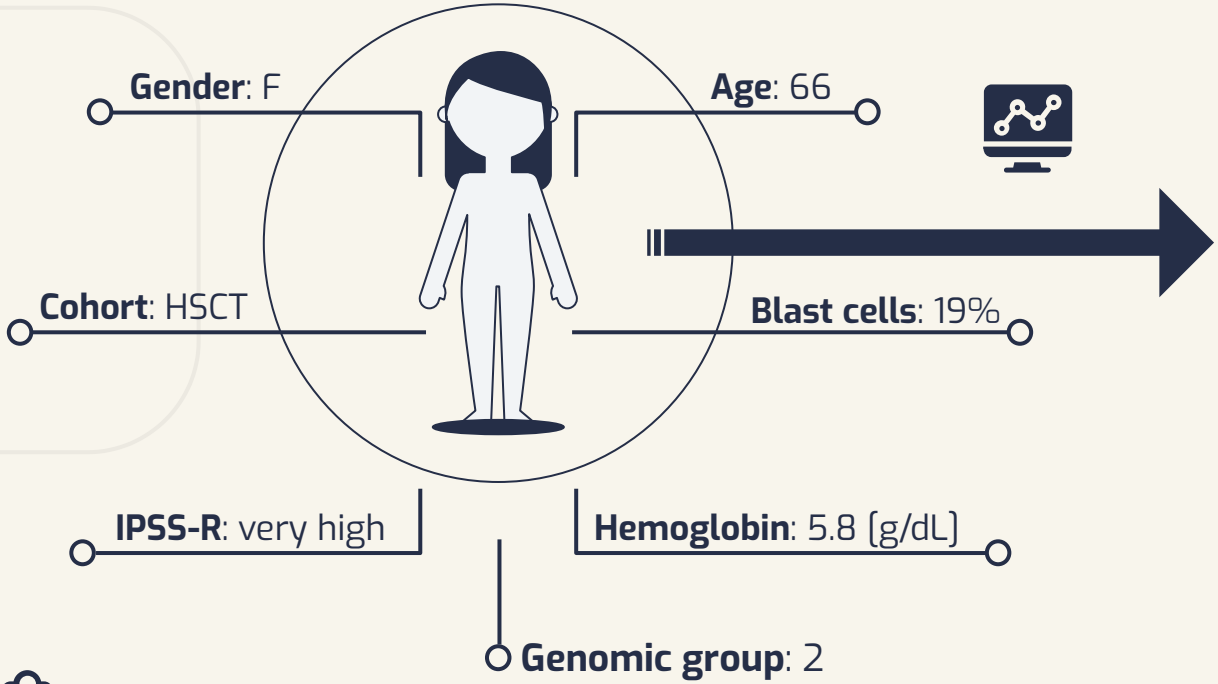


Clusters characterization





… but what's that for?



Cluster:

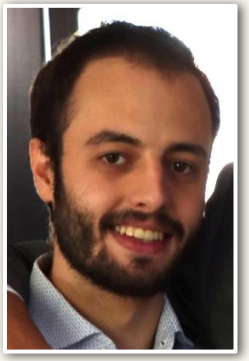
very high risk

Predicted survival time:

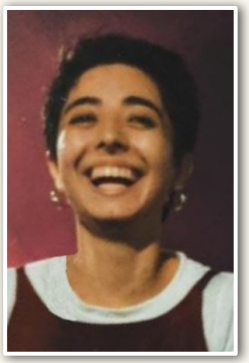
354 days

Thank you!

Manfred



Luca



Taguhi

Michele