



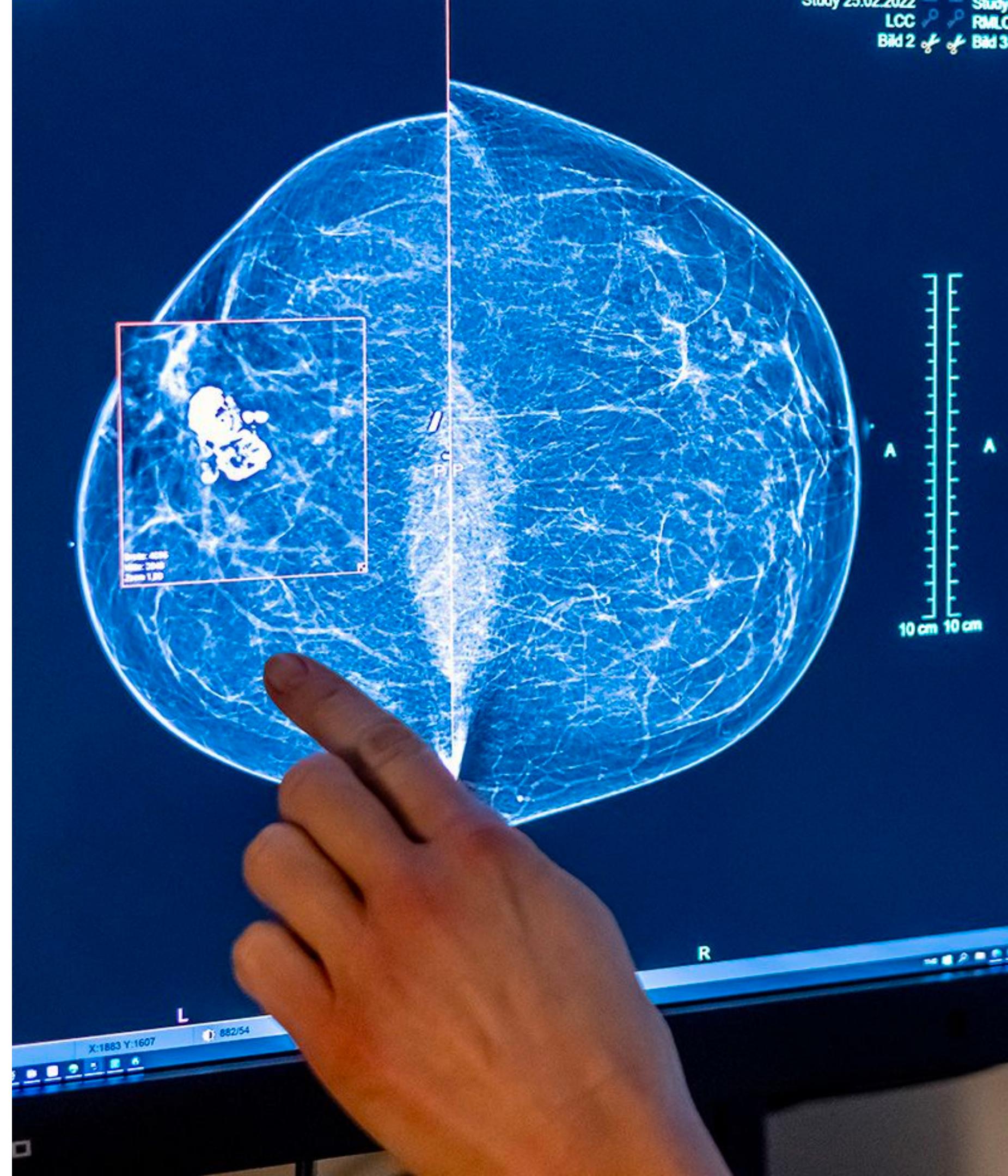
POLITECNICO
MILANO 1863



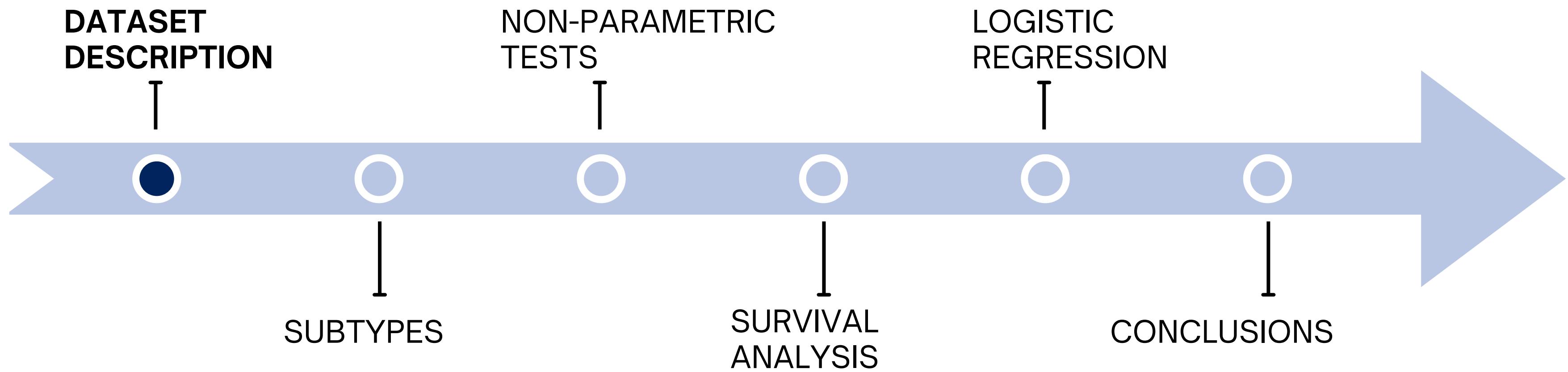
UNIVERSITÀ
DEGLI STUDI
DI MILANO

Analysis on Breast Cancer

Erika Ballabio
Kevin Del Gaudio
Isabella Rossi
Giulio Vidotto



STEPS



DATASET HISTORY



The Molecular Taxonomy of Breast Cancer International Consortium (**METABRIC**) database is a Canada-UK Project which contains ***targeted sequencing data of primary breast cancer samples***.

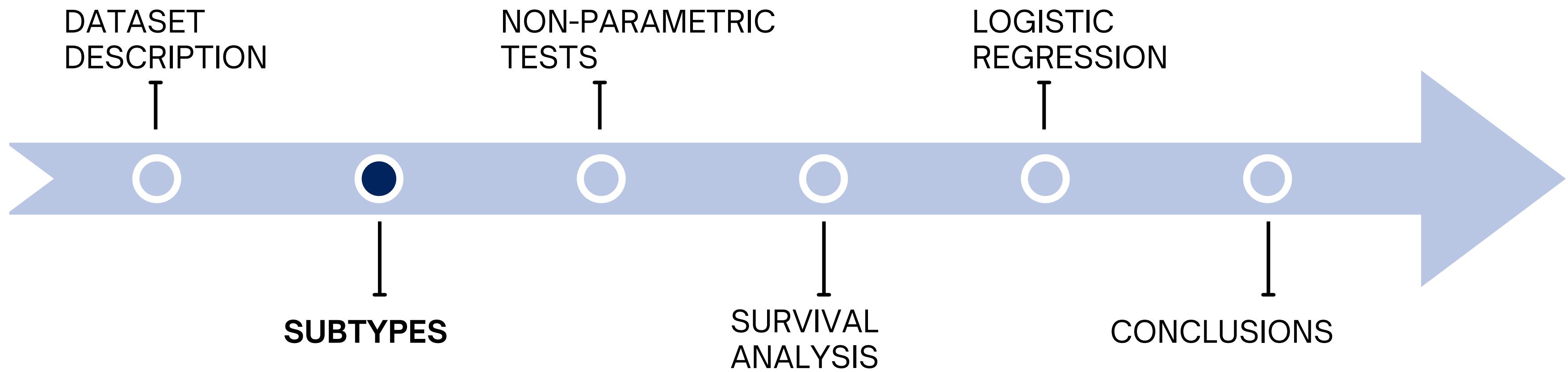
Clinical and genomic data was downloaded from cBioPortal.

This dataset provides the following informations:

- **Age** of the patient at the diagnosis of the tumor
- The **number of lymph nodes** found positive
- If the tumor is **metastatic**
- **Size, grade, spread** of the tumor
- If the tumor **expresses certain receptors** (ER and HER2)

The dataset was collected by Professor Carlos Caldas from Cambridge Research Institute and Professor Sam Aparicio from the British Columbia Cancer Centre in Canada and published on Nature Communications (Pereira et al., 2016).

STEPS



RECEPTOR STATUS IN BREAST CANCER

Receptors are proteins that bind to external messengers, such as hormones or growth factors, to create a series of downstream effects that mediate a specific response in the cell. Breast Cancers are characterised by an **abnormal expression of some key receptors that significantly increase the proliferation of its cells.**

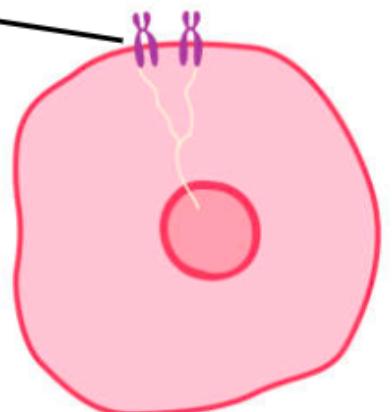
ER

Estrogen receptors are activated by the hormone estrogen and then are able to bind the DNA to regulate the activity of different genes that induce the cells to enter the cell cycle.

About 55% to 60% of all breast cancers have positive estrogen receptors. [1]

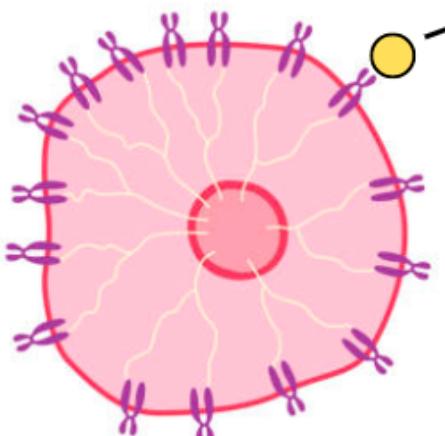
Normal Cell

Protein receptor



Cancer Cell

Hormone or growth factor



HER2

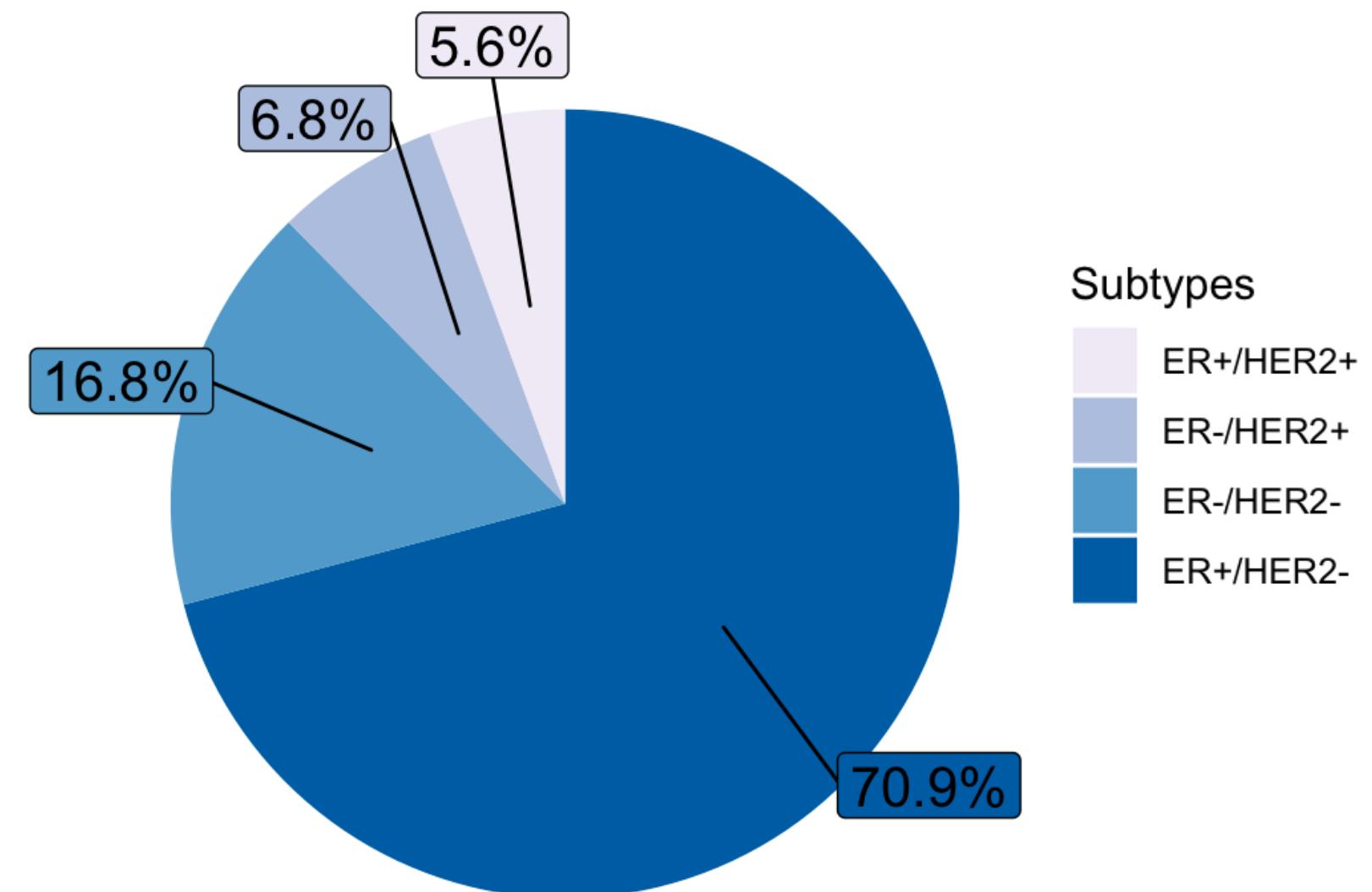
Human epidermal growth factor receptor 2 belongs to the epidermal growth factor receptor family and it stimulates cell proliferation via receptor tyrosine kinase intracellular signaling.

Generally, it is over-expressed in 13–22% of breast cancers. [2]

POPULATION SUBTYPES

Breast cancer is a genetically and clinically heterogeneous disease with multiple subtypes and their classification has evolved over the years. We have considered a simplified grouping that only considers whether there is an expression of the receptors ER and HER2.

- Generally speaking patients with **ER+** tumors have a better prognosis than patients with ER- tumors[3].
- It has been found that **HER2+** breast cancer has better survival outcomes compared to HER2- breast cancer[4].



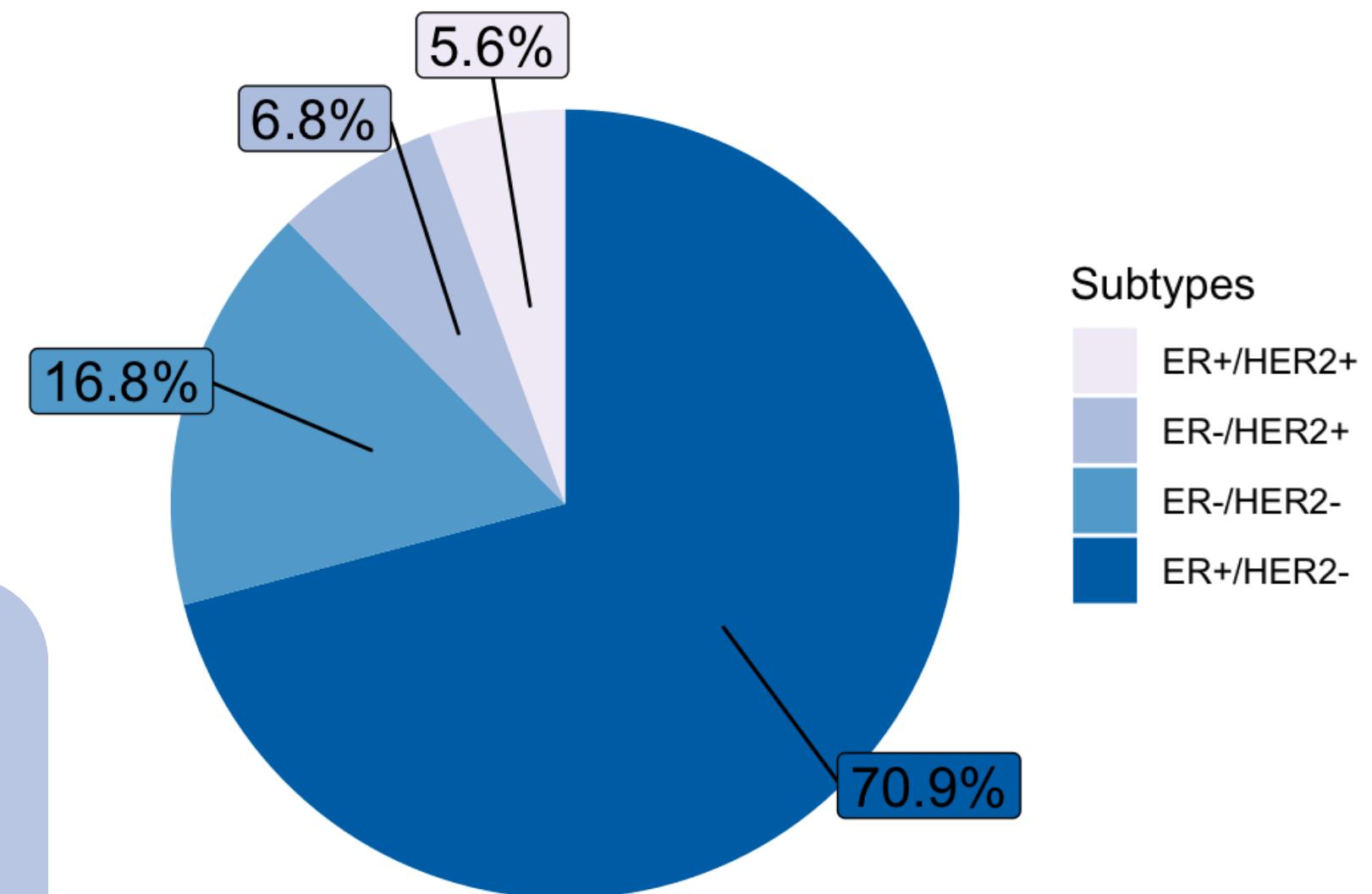
POPULATION SUBTYPES

Breast cancer is a genetically and clinically heterogeneous disease with multiple subtypes and their classification has evolved over the years. We have considered a simplified grouping that only considers whether there is an expression of the receptors ER and HER2.

- Generally speaking patients with **ER+** tumors have a better prognosis than patients with ER- tumors[3].
- It has been found that **HER2+** breast cancer has better survival outcomes compared to HER2- breast cancer[4].

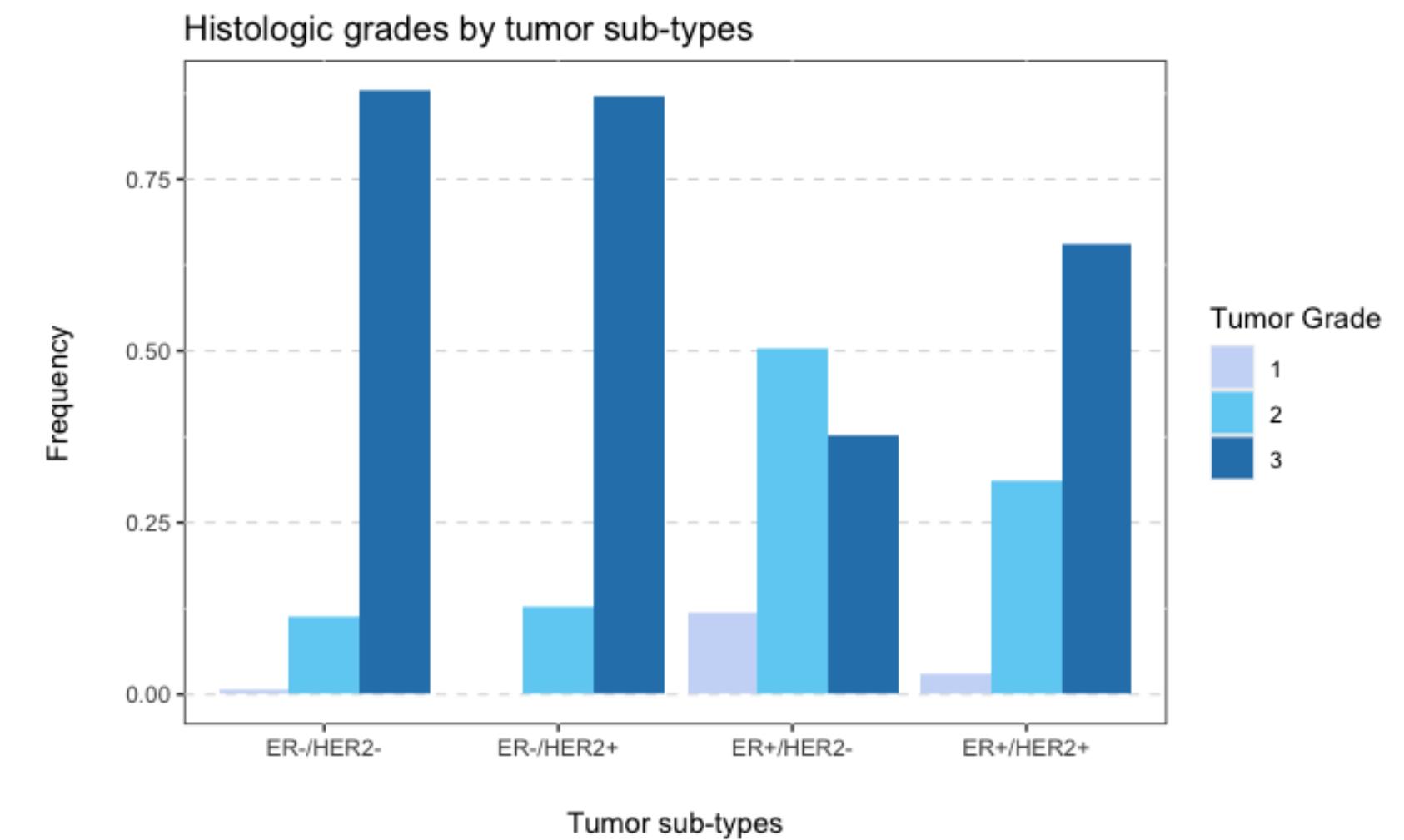
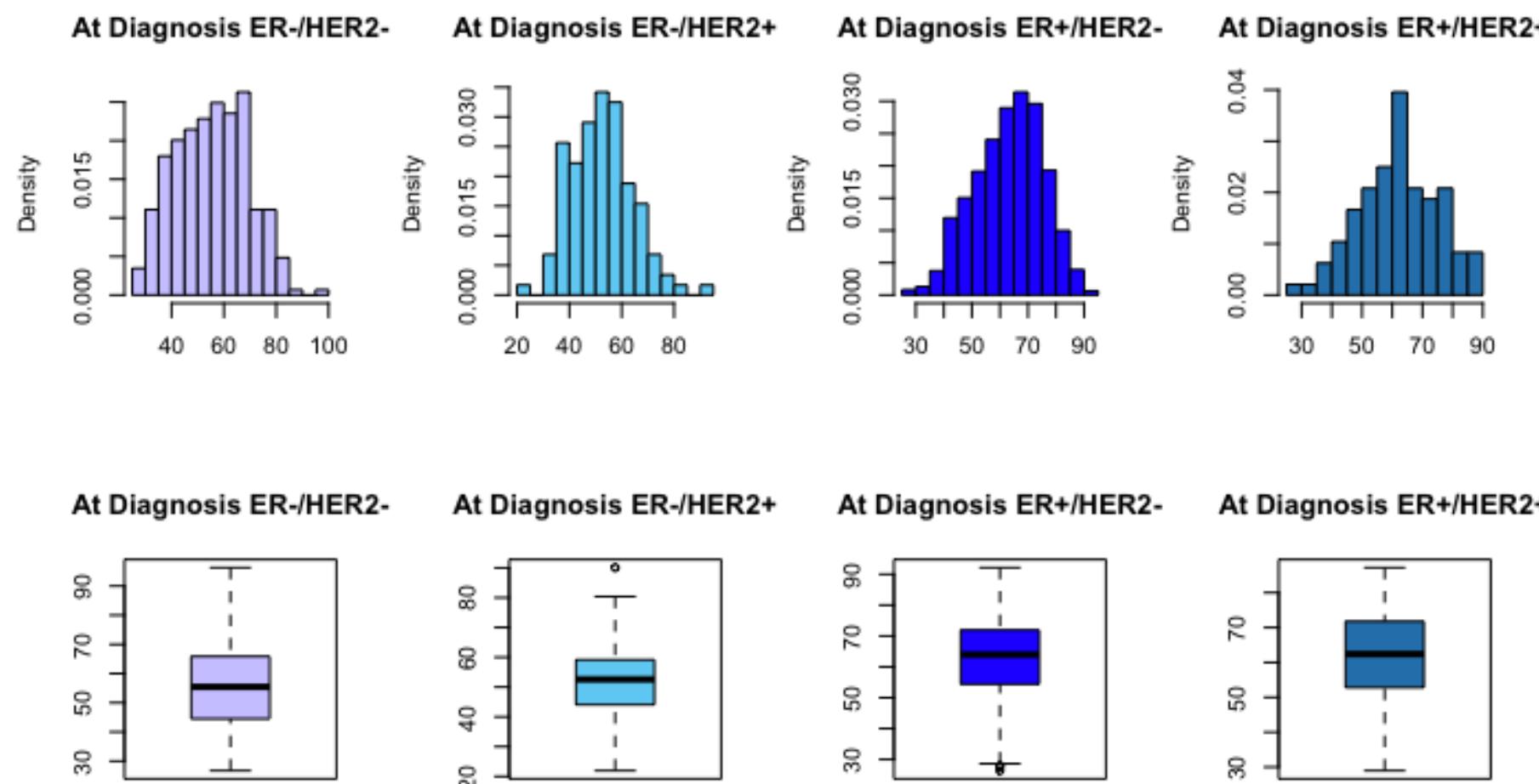
GOAL:

we would like to investigate the **survival** of Breast Cancer considering these subgroups and several other covariants that will be later discussed.

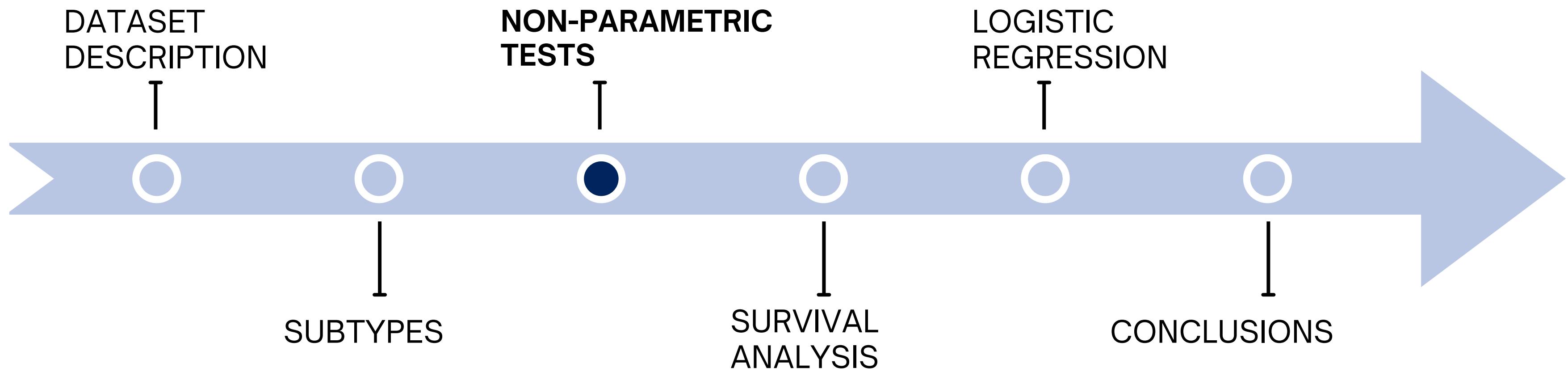


DESCRIPTIVE ANALYSIS

We further investigated the characteristics of our subgroups by studying their distribution by Age at Diagnosis and the Tumor Grade they present.



STEPS



TUMOR CHARACTERISTICS

Before moving forward, we need to clarify some of the variables used in our dataset.

Tumor SIZE

Tumor size is measured based on an imaging scan or surgical removal of the tumor and are divided in:

- **T1** : the tumor is < 20 mm in diameter .
- **T2**: the tumor is > 20 mm and < 50 mm
- **T3** : the tumor is > 50 mm in diameter.

Tumor SPREAD

The spread of a cancer describes how far it has spread from where it originated:

- If the **value is 0** the tumor has not spread.
- If the **value is 1** the tumor may have spread to the surrounding tissues, the lymph nodes or, even worse, to at least one other body organ.

Tumor GRADE

The grade of a cancer depends on what the cells look like under a microscope.

In general:

- A **lower grade (1)** indicates a *lower-growing cancer with more normal-looking cells*
- A **higher grade (3)** indicates a *faster-growing cancer with numerous undifferentiated cells*.

NON-PARAMETRIC TESTS

To perform a better Survival Analysis we decided to carry out a series of **chi-squared tests** to highlight if there is an association between certain variables and the overall survival status (Dead, Alive).

TEST I

There is an association between the **subtypes** and the overall survival (p-value = 1.72e-09)

TEST II

There is an association between the **tumor grade** and the overall survival (p-value = 3.99e-11)

TEST III

There is an association between **tumor size** and overall survival (p-value = 9.42e-11)

TEST IV

There is no association between the **intervals of age** (under 60 e over 60) and overall survival (p-value = 0.3575)

NON-PARAMETRIC TESTS

To perform a better Survival Analysis we decided to carry out a series of **chi-squared tests** to highlight if there is an association between certain variables and the overall survival status (Dead, Alive).

TEST I

There is an association between the **subtypes** and the overall survival (p-value = 1.72e-09)



TEST II

There is an association between the **tumor grade** and the overall survival (p-value = 3.99e-11)



TEST III

There is an association between **tumor size** and overall survival (p-value = 9.42e-11)

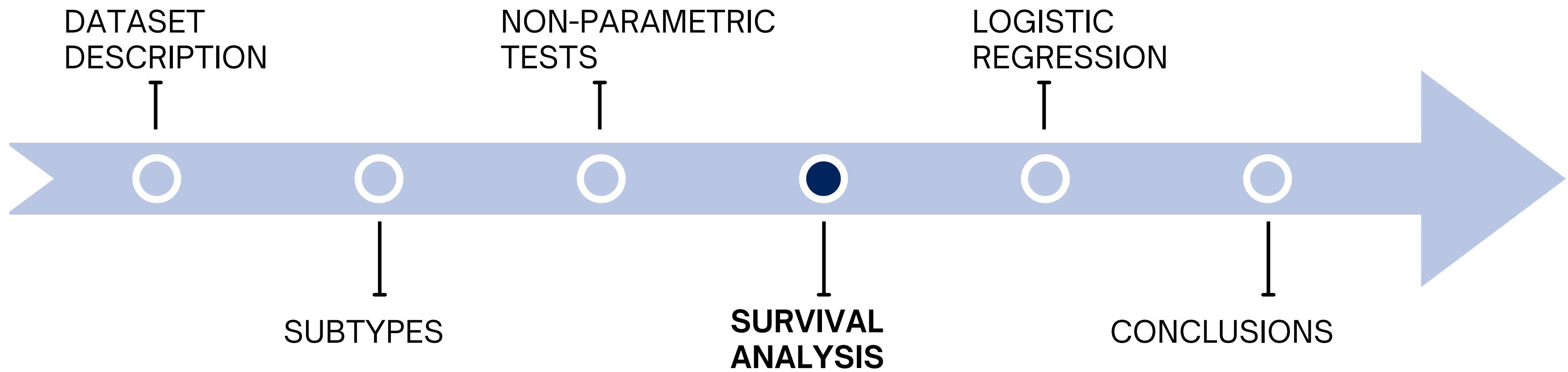


TEST IV

There is no association between the **intervals of age** (under 60 e over 60) and overall survival (p-value = 0.3575)

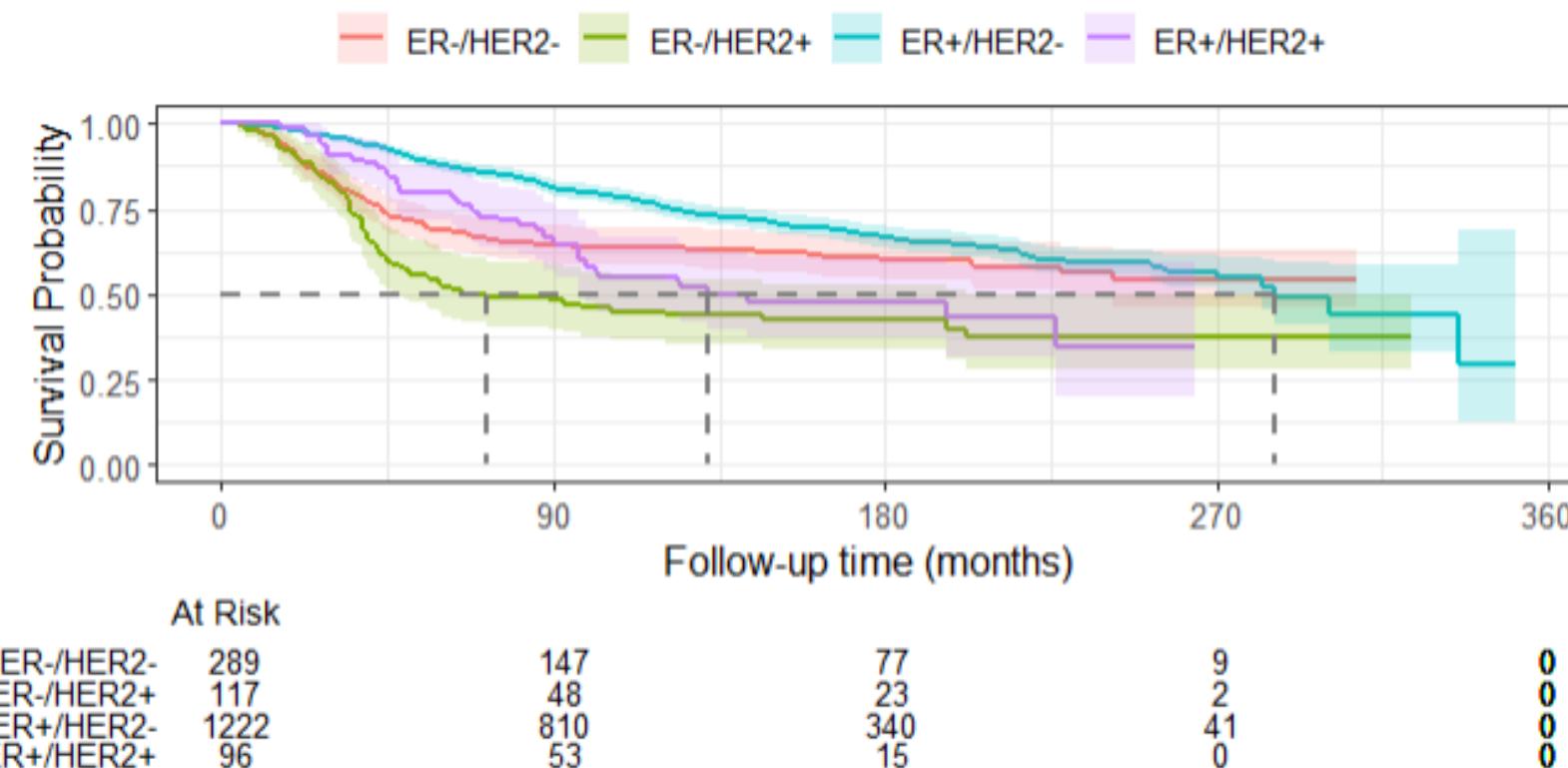


STEPS

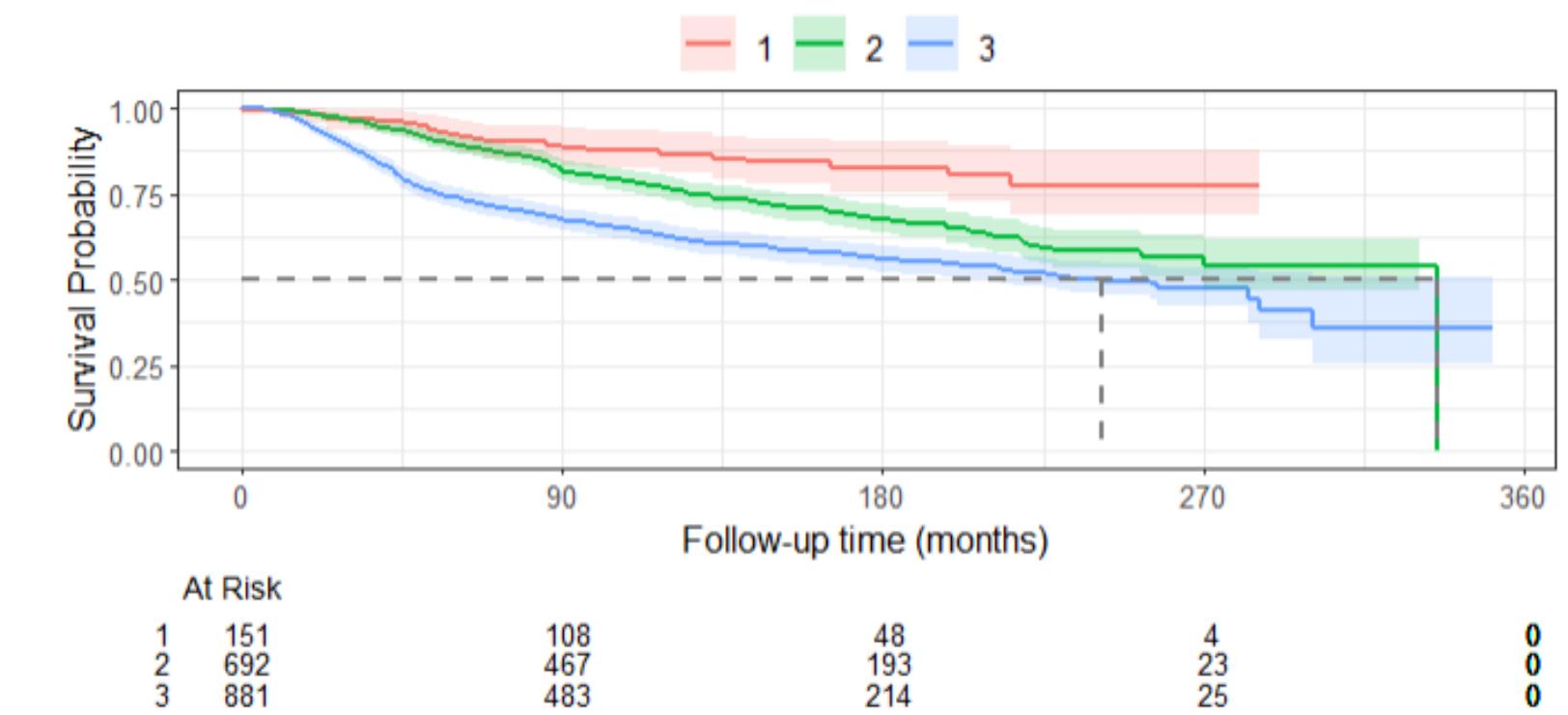


SURVIVAL PROBABILITY PLOT

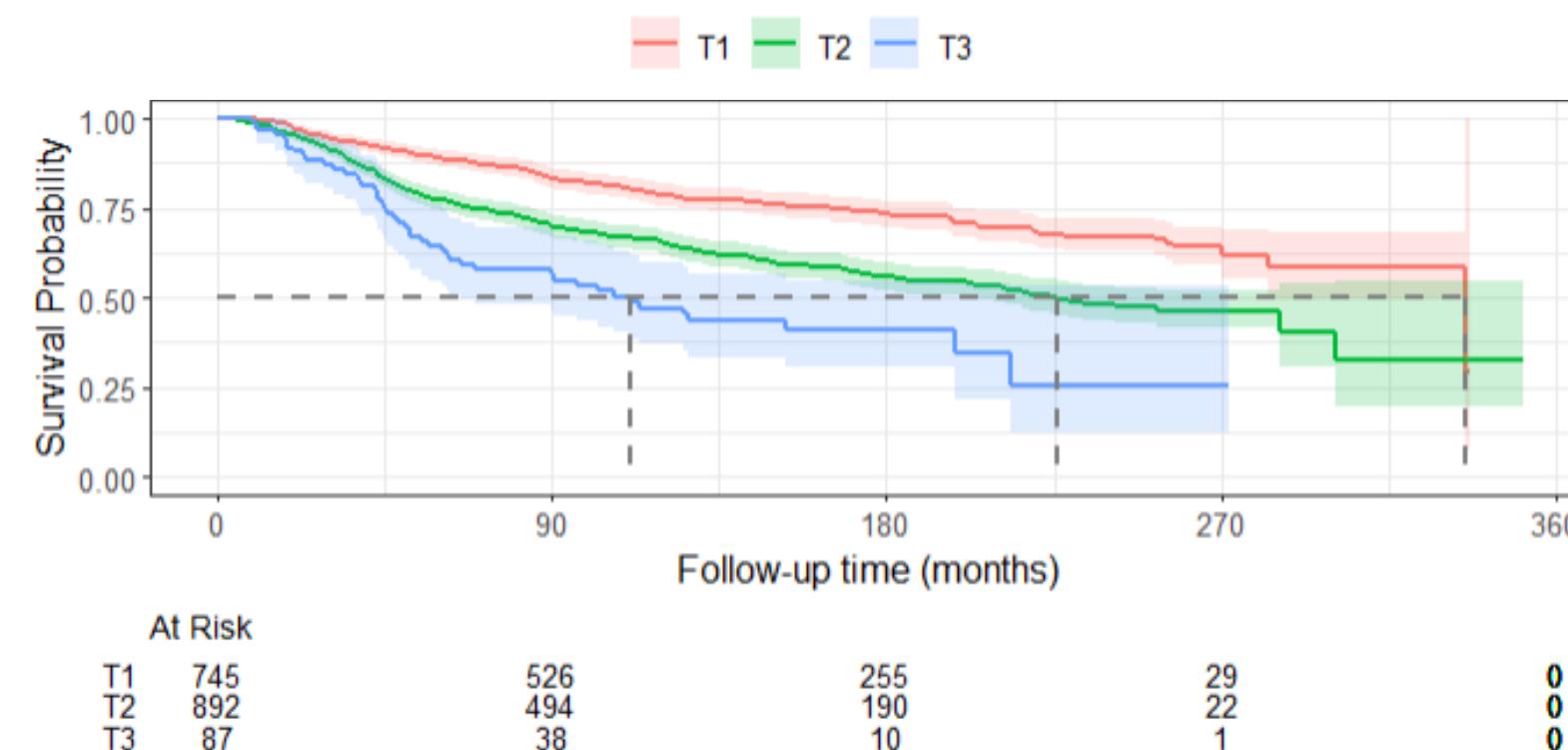
ER+/HER2- → higher survival



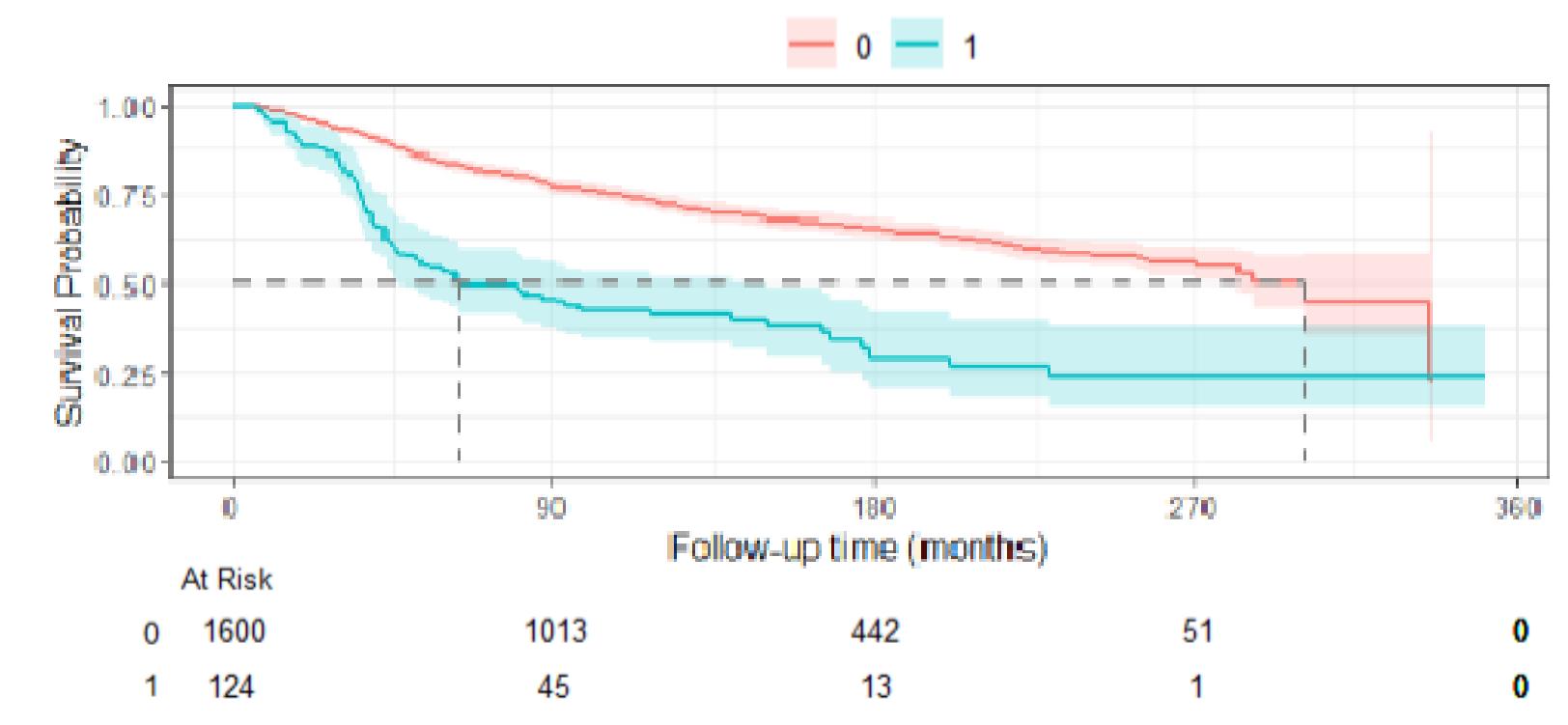
Higher neoplasm Histologic Grade → lower survival



Higher tumor size grade → lower survival

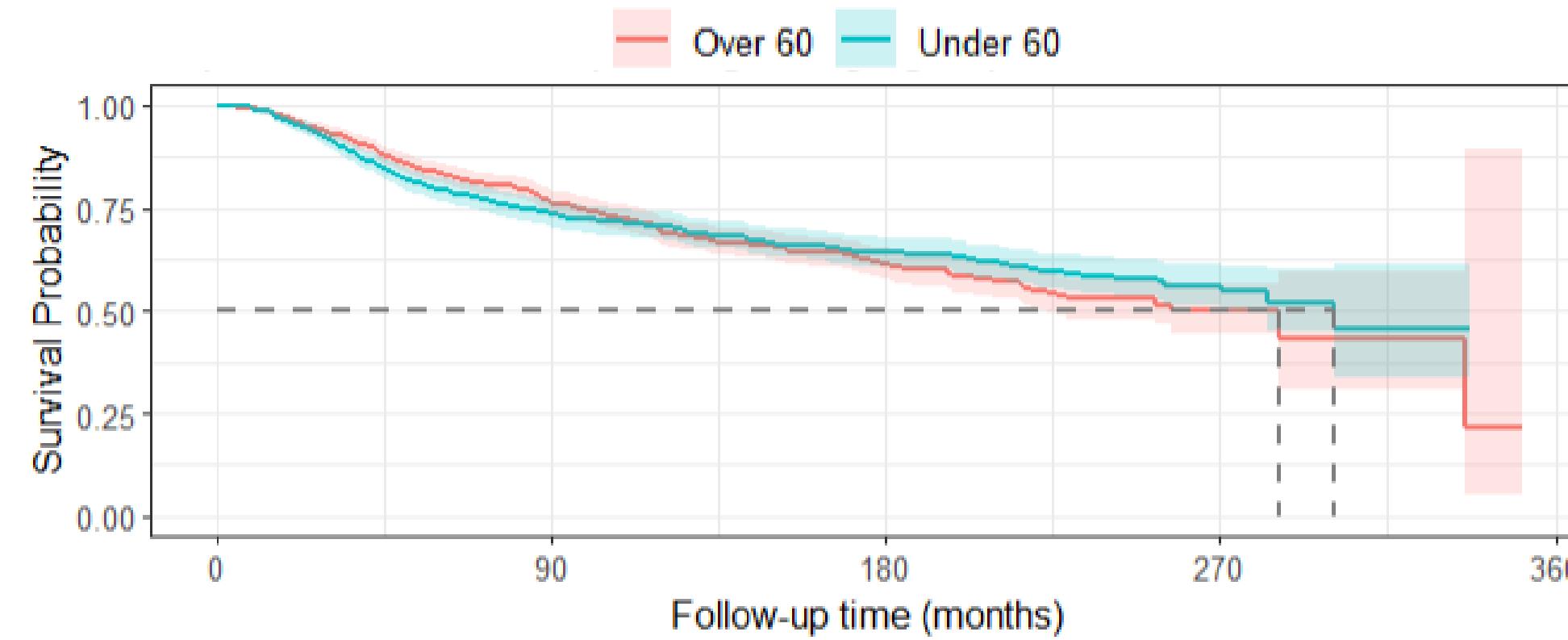


Higher tumor spread → lower survival



SURVIVAL PROBABILITY PLOT

No effect due to age

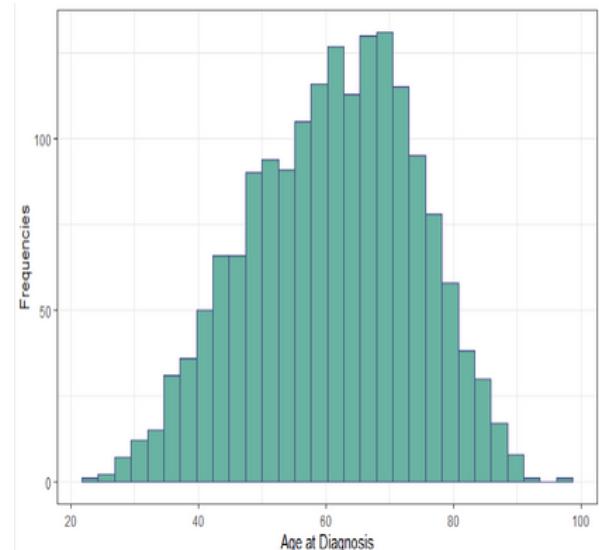


At Risk					
Over 60	956	567	212	17	0
Under 60	768	491	243	35	0

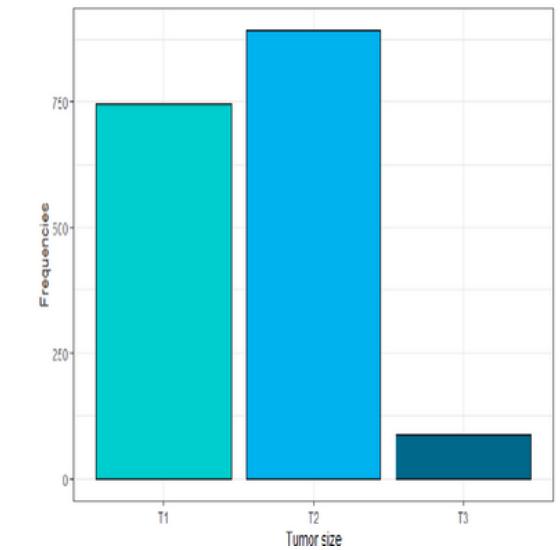
COX MODEL FOR OVERALL SURVIVAL

Using a **time dependent coefficient cox model**, we study how the survival is affected by the following covariates:

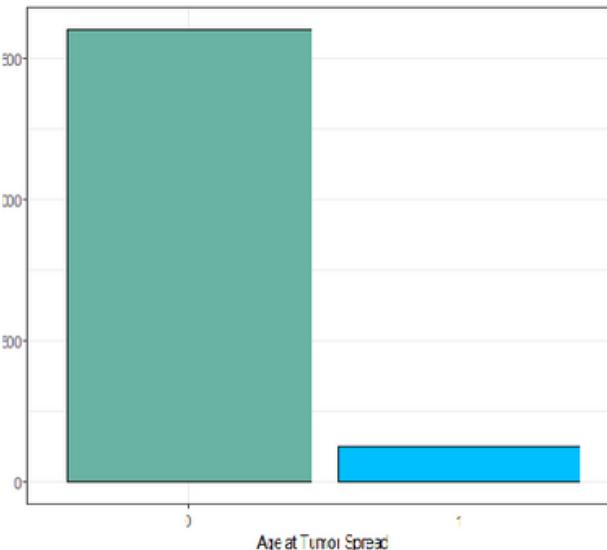
Age at Diagnosis



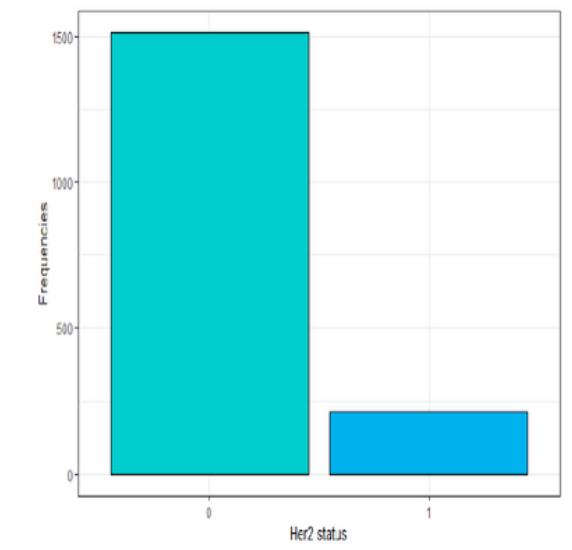
Tumor size



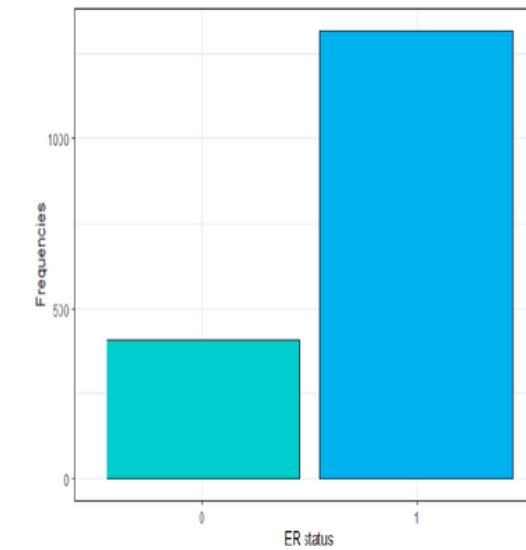
Tumor spread



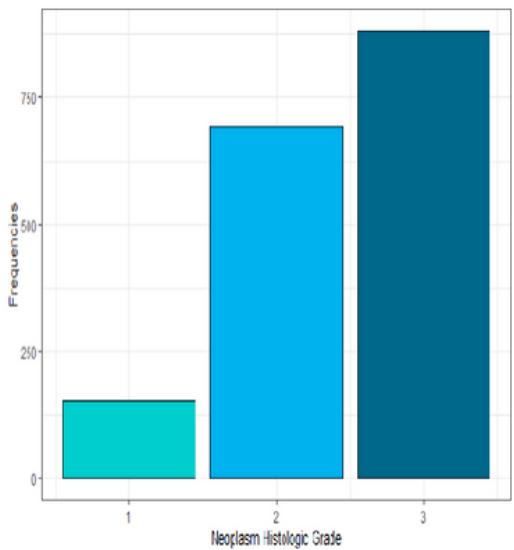
HER2 status



ER status



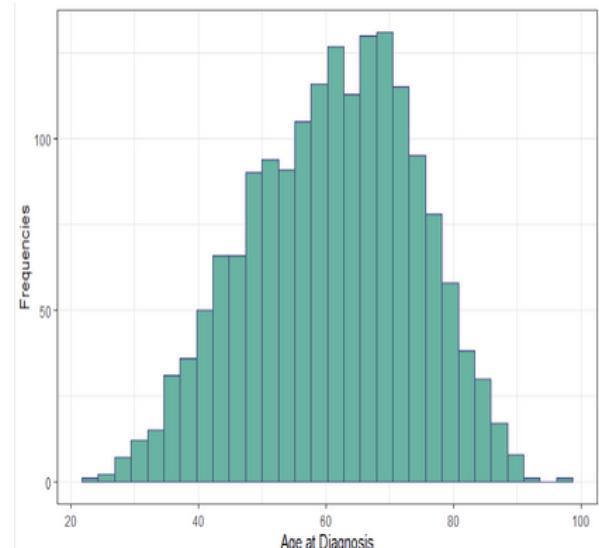
Neoplasm Histologic Grade



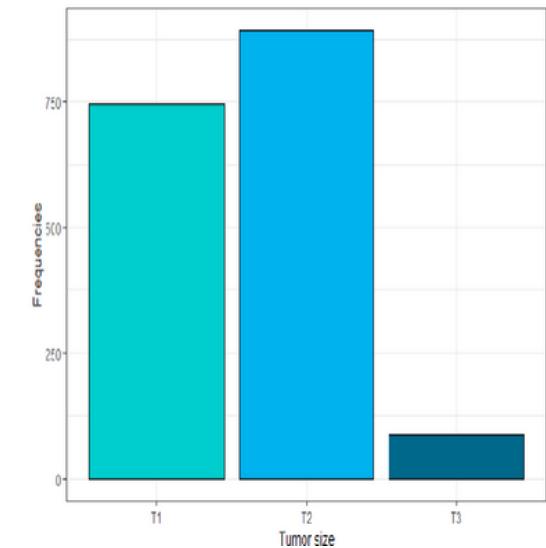
COX MODEL FOR OVERALL SURVIVAL

Using a **time dependent coefficient cox model**, we study how the survival is affected by the following covariates:

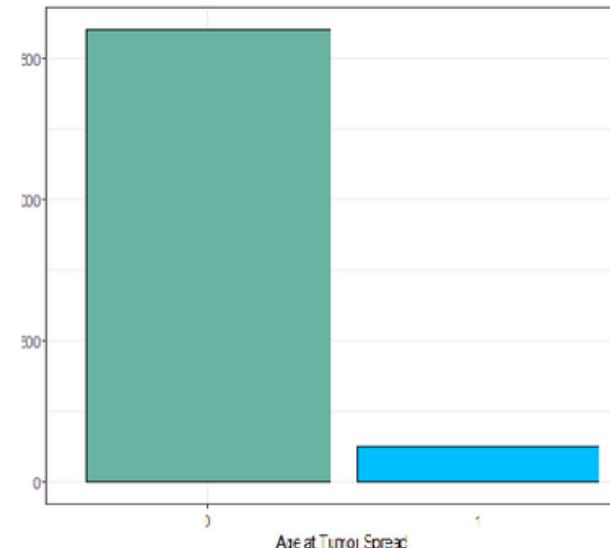
Age at Diagnosis



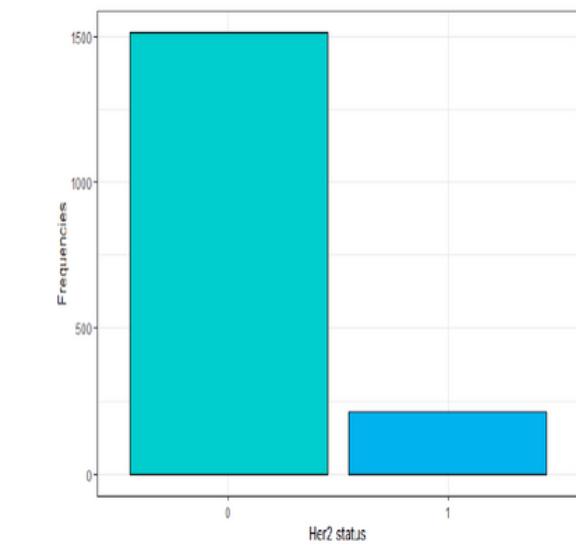
Tumor size



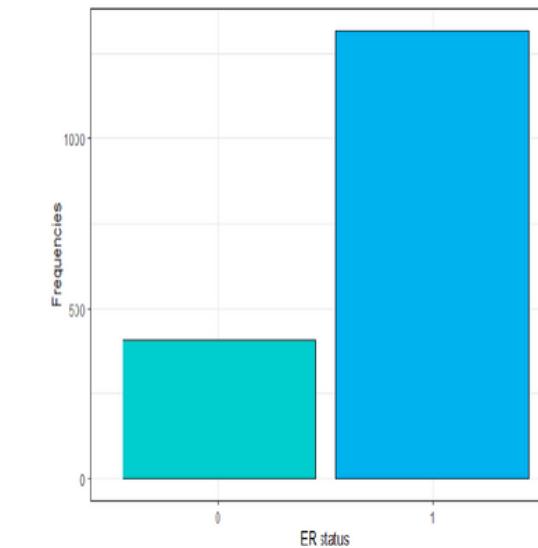
Tumor spread



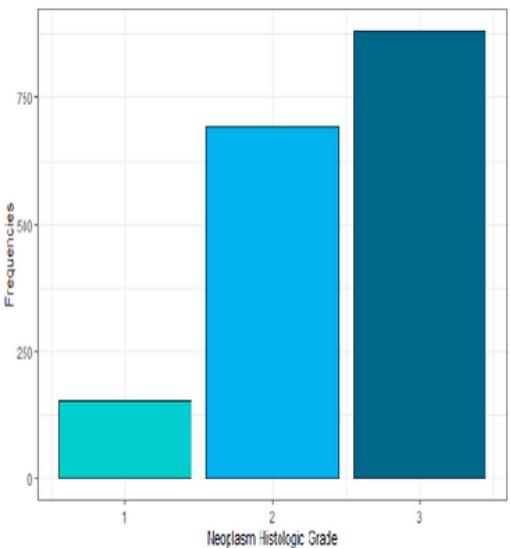
HER2 status



ER status



Neoplasm Histologic Grade



First interval

0 - 48 months or 0 - 4 years

$T_{i,0}$

Second interval

48 - 96 months or 4 - 8 years

$T_{i,48}$

Third interval

over 96 months or over 8 years

Time

$T_{i,96}$

COX MODEL - RESULTS

Variable Name	HR = $\exp(\beta)$	P-values	Confidence Intervals
Age At Diagnosis	1,01	0.0132	1.01 (1.00 - 1.01)
Tumor size (T2)	1,69	1.33e-08	1.69 (1.41 - 2.03)
Tumor size (T3)	2,67	3.43e-09	2.67 (1.93 - 3.70)
HER2 Status	1,74	2.73e-06	1.74 (1.40 - 2.17)
Tumor spread	2,07	1.82e-08	2.11 (1.64 - 2.7)

The forest plot displays the hazard ratio (HR) for each variable, along with its P-value and 95% confidence interval. The x-axis represents the hazard ratio, ranging from 1 to 3.5. The variables are ordered by their P-values: Age At Diagnosis (P=0.0132), Tumor size (T2) (P=1.33e-08), Tumor size (T3) (P=3.43e-09), HER2 Status (P=2.73e-06), and Tumor spread (P=1.82e-08). All variables show a significant increase in hazard ratio compared to the reference value of 1.

COX MODEL - RESULTS

First time
interval
year ≤ 4

Variable Names	HR = $\exp(\beta)$	P-values	Confidence Intervals
ER Status (year ≤ 4)	0,37	6,57E-12	[0,2757 - 0,4886]
Neoplasm Histologic Grade 2 (year ≤ 4)	1,25	0,5746	[0,5673 - 2,7782]
Neoplasm Histologic Grade 3 (year ≤ 4)	1,28	0,01758	[1,1775 - 5,5271]

Second time
interval
 $4 < \text{year} \leq 8$

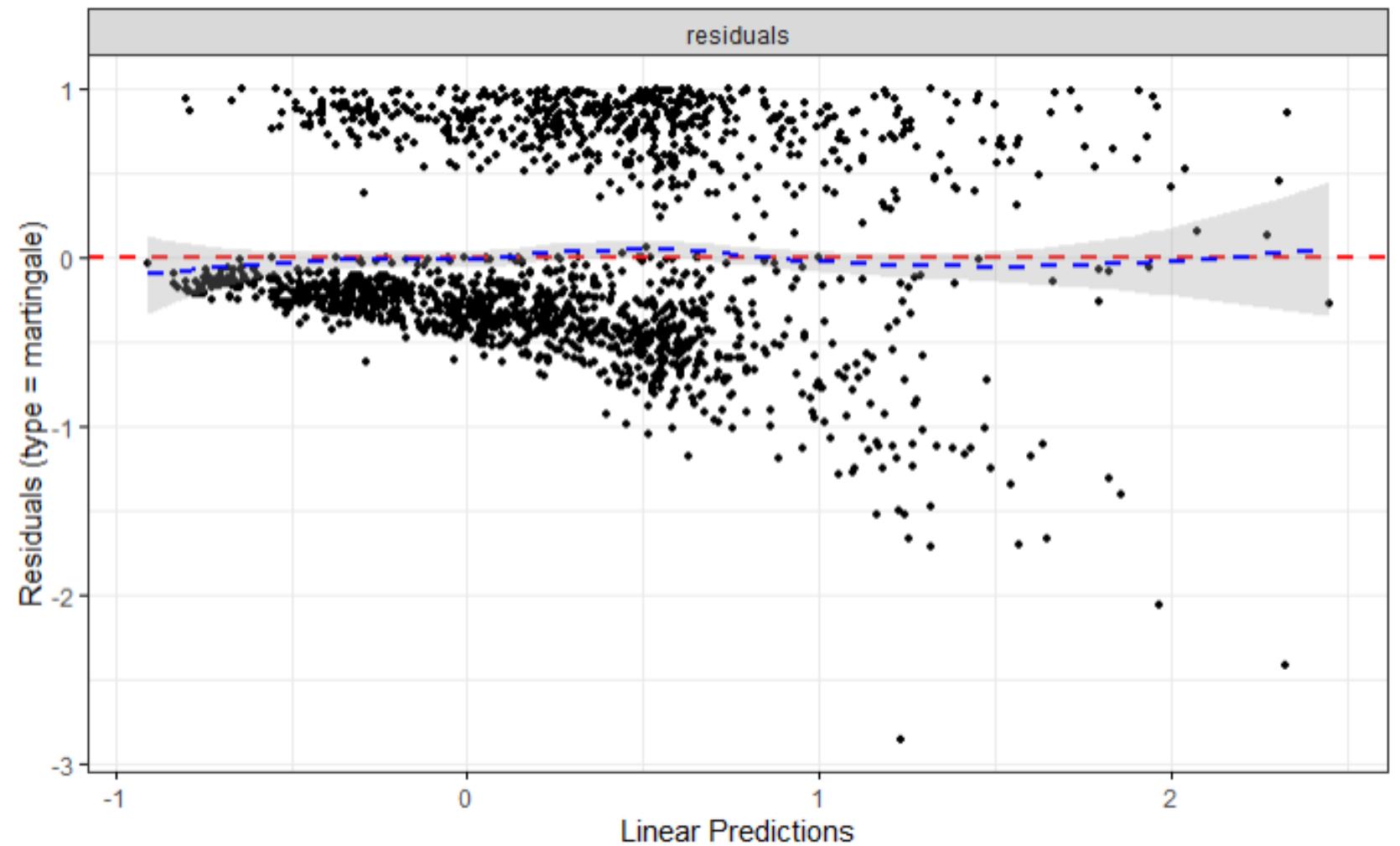
Variable Names	HR = $\exp(\beta)$	P-values	Confidence Intervals
ER Status ($4 < \text{year} \leq 8$)	1,05	0,81104	[0,6981 - 1,5830]
Neoplasm Histologic Grade 2 ($4 < \text{year} \leq 8$)	1,69	0,13506	[0,8484 - 3,3861]
Neoplasm Histologic Grade 3 ($4 < \text{year} \leq 8$)	1,76	0,11152	[0,8766 - 3,5555]

Third time
interval
 $\text{year} > 8$

Variable Names	HR = $\exp(\beta)$	P-values	Confidence Intervals
ER Status ($\text{year} > 8$)	2,95	8,63E-05	[1,7186 - 5,0586]
Neoplasm Histologic Grade 2 ($\text{year} > 8$)	2,5	0,02026	[1,1537 - 5,4334]
Neoplasm Histologic Grade 3 ($\text{year} > 8$)	2,83	0,00898	[1,2966 - 6,1700]

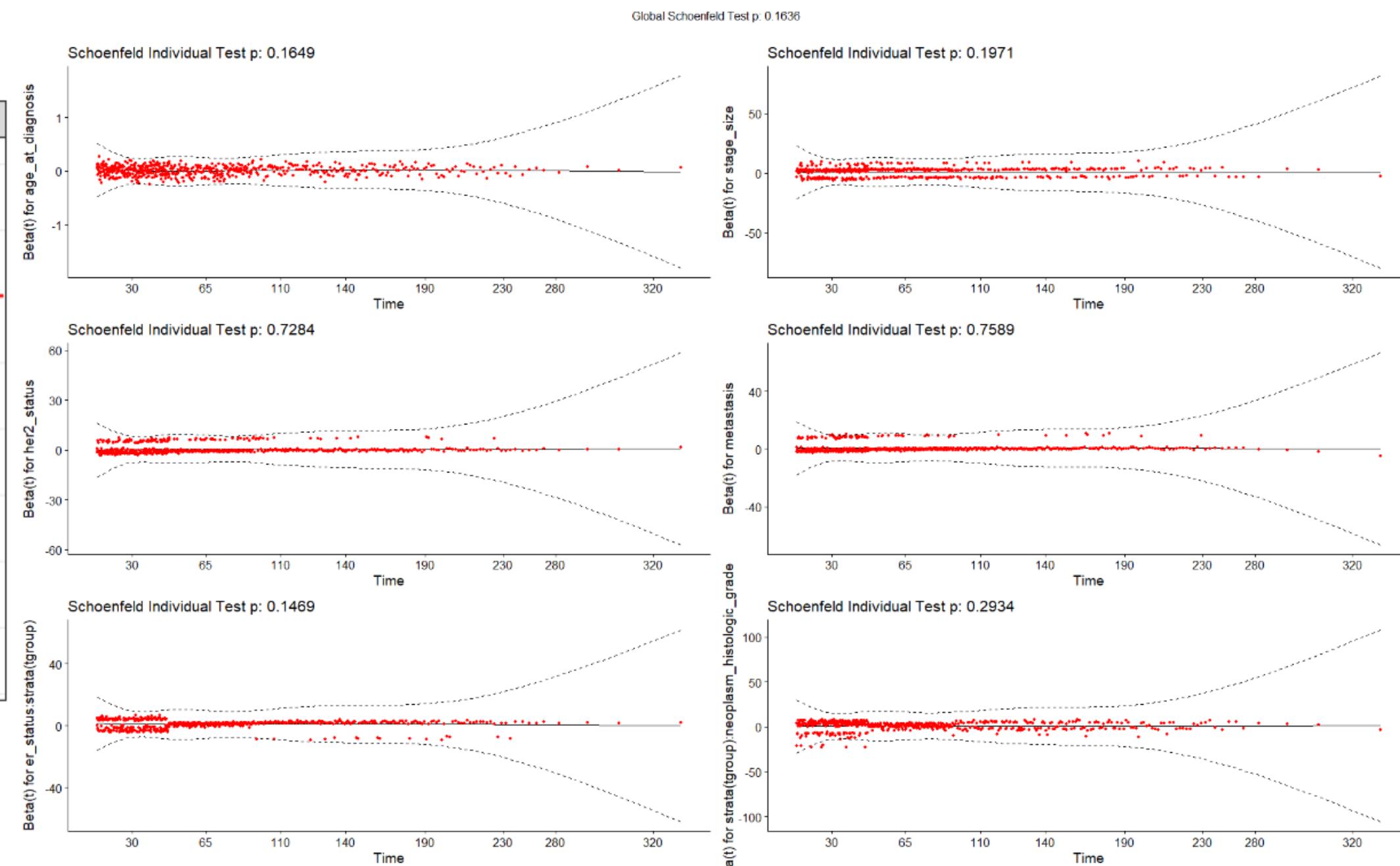
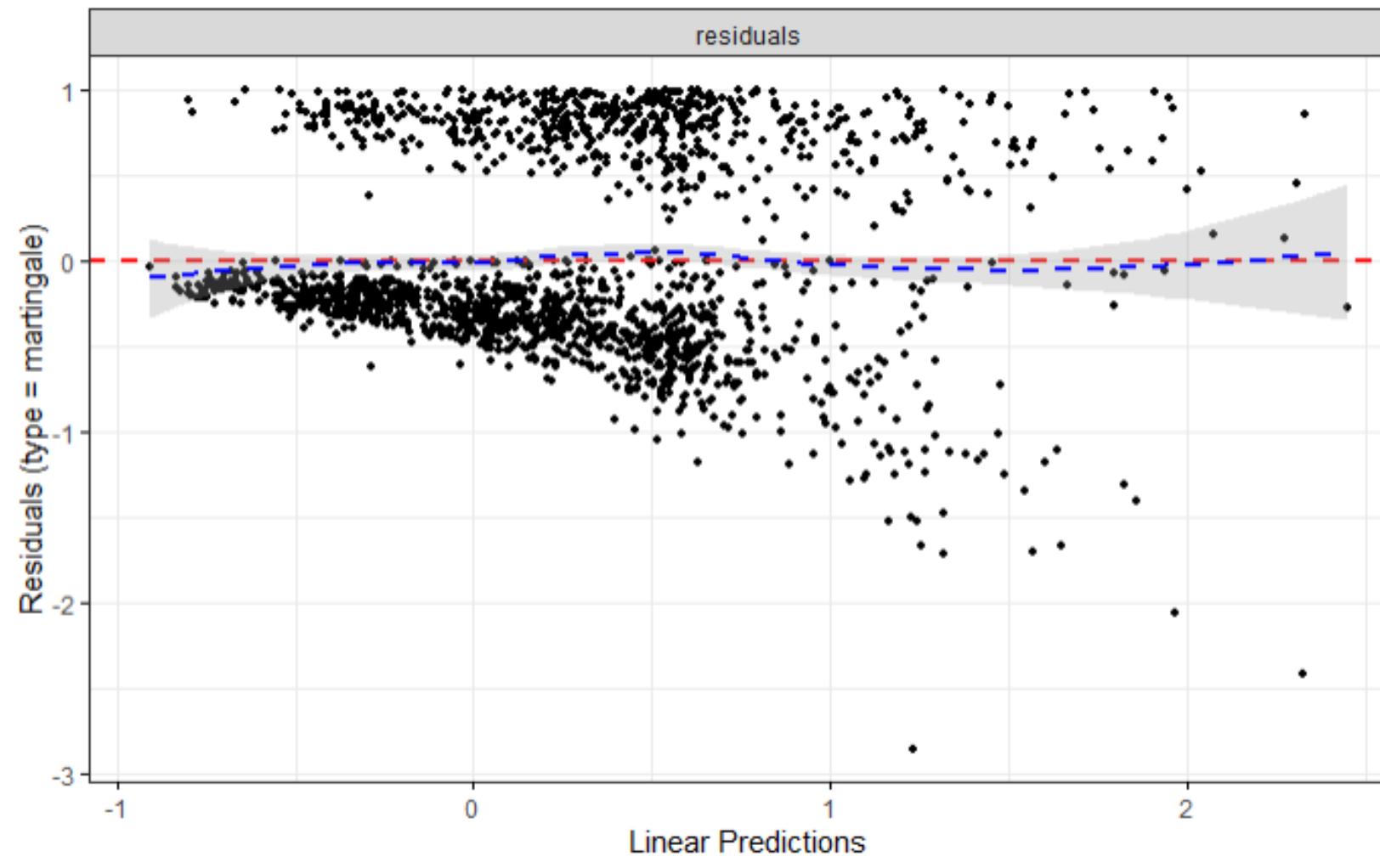
COX MODEL - ASSUMPTION CHECK

- We checked the **Goodness of fit** of the model by plotting the residuals:

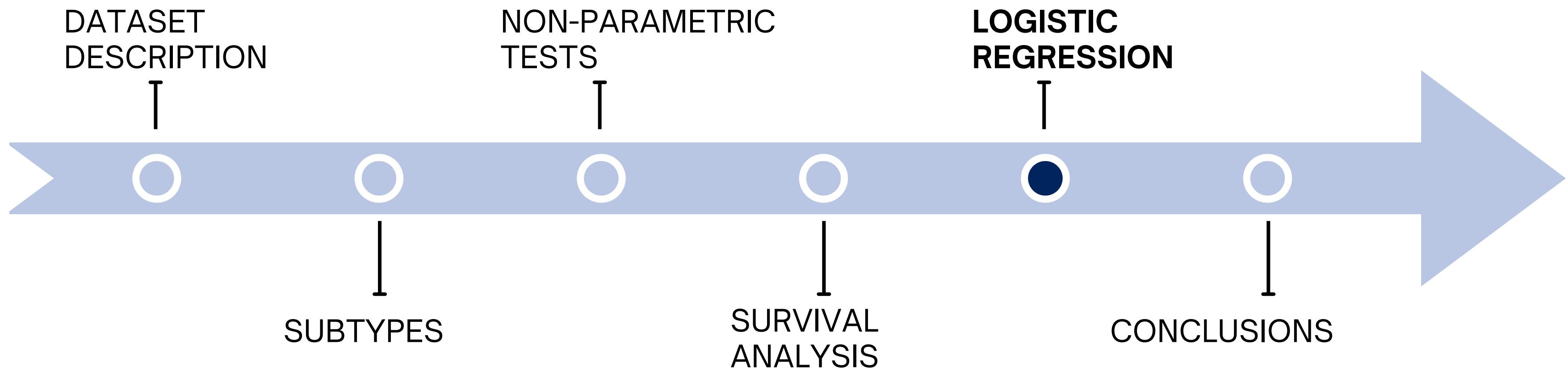


COX MODEL - ASSUMPTION CHECK

- We checked the **Goodness of fit** of the model by plotting the residuals:
- We checked the **Proportional Hazard Assumption** of each covariate by plotting the Schoenfeld residuals:



STEPS



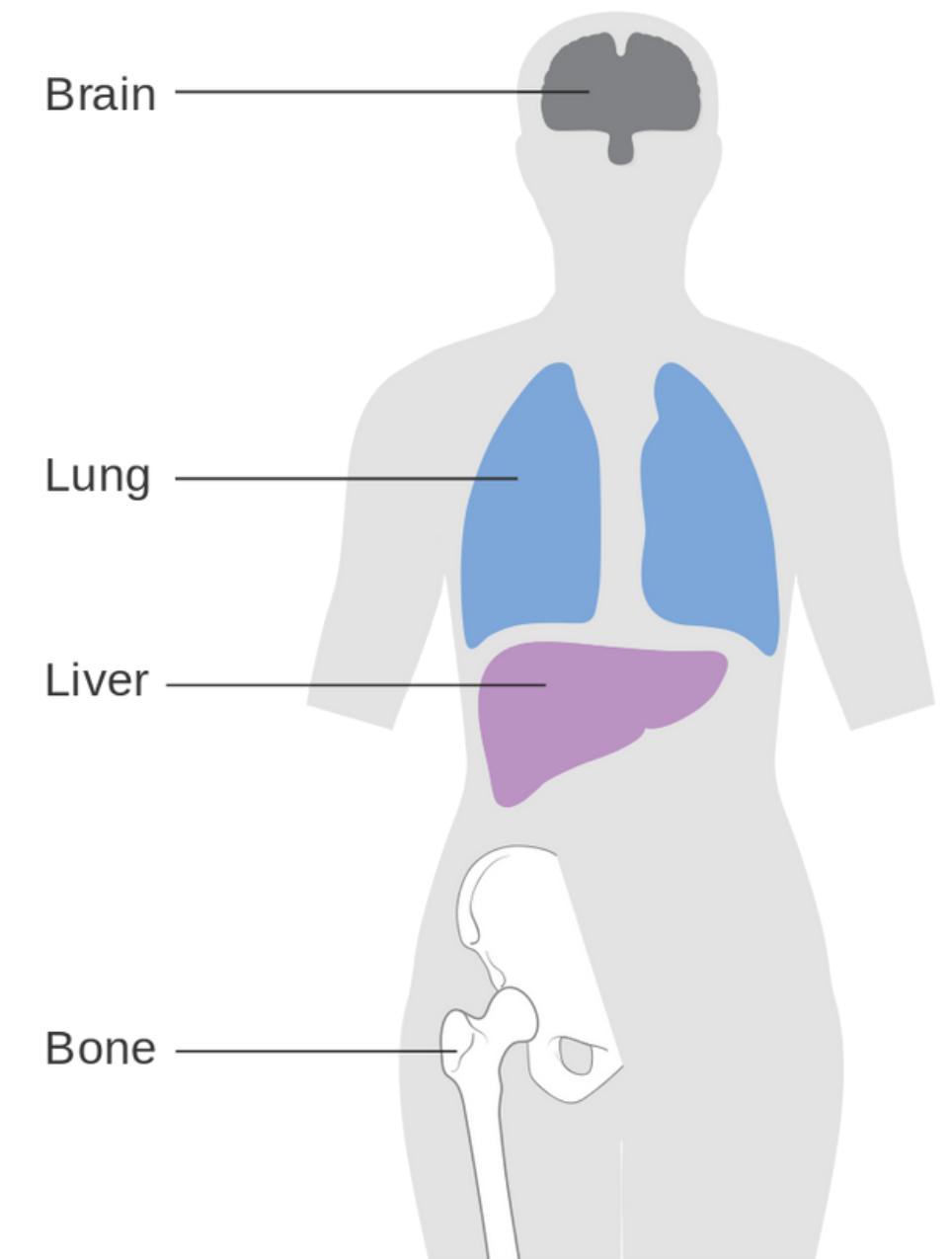
WHAT ARE METASTASIS?

- ▶ Metastatic breast cancer refers to the spread of cancer cells from the site where a primary tumor originated to other parts of the body. In our case, beyond the **axillary lymph nodes**.

▶ Where does breast cancer metastasize?

Breast cancer can metastasize anywhere in the body but primarily metastasizes to:

- the **bone** (most common site for metastasis)
- the **lungs**,
- the **liver**,
- the **brain**.



LOGISTIC REGRESSION MODEL

Two main classes:

- Class 1 → Patients with ***metastasis*** (observations = 124)
- Reference Class → Patient without ***metastasis*** (observations = 1600)

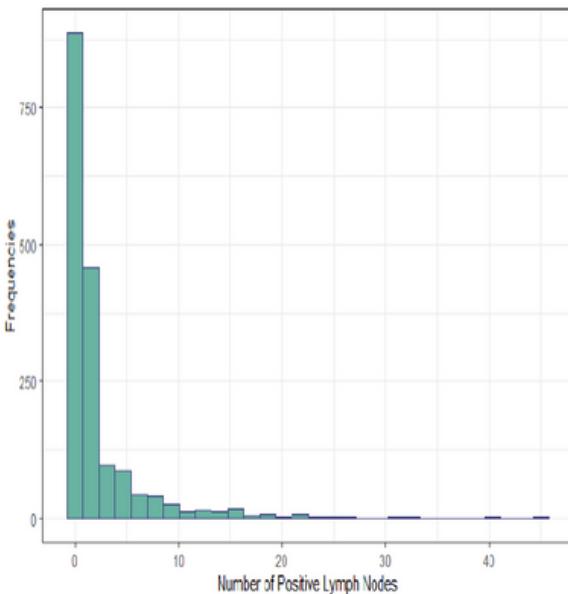
LOGISTIC REGRESSION MODEL

Two main classes:

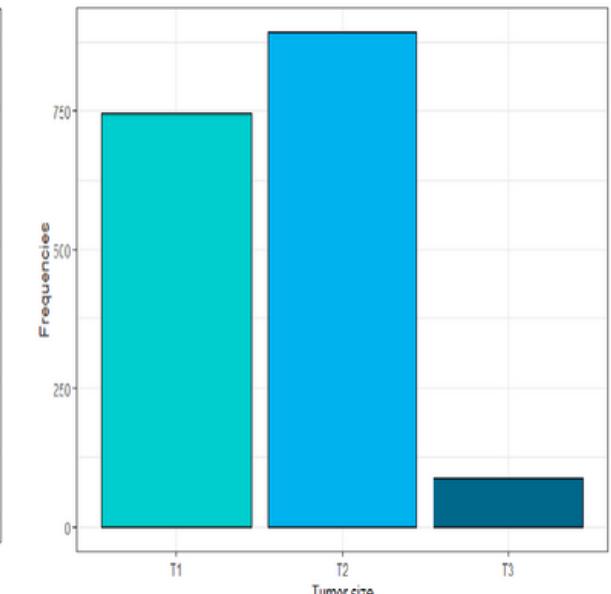
- Class 1 → Patients with metastasis (observations = 124)
- Reference Class → Patient without metastasis (observations = 1600)

To study the probability of belonging to **Class 1** or to the **Reference Class** we used these predictors:

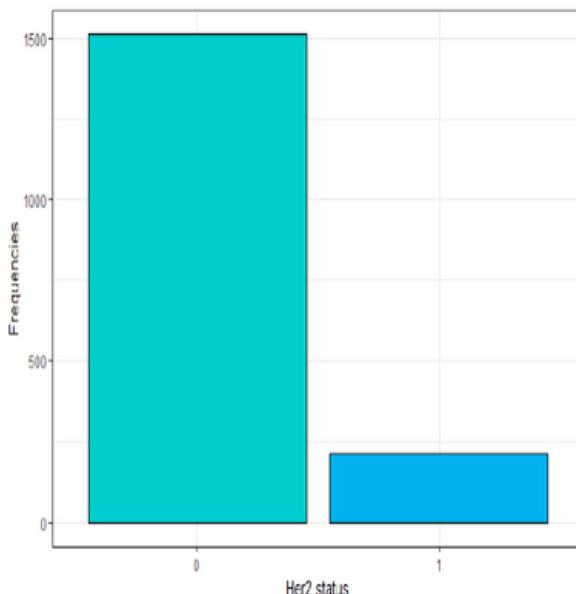
Nº of Positive Lymph Nodes



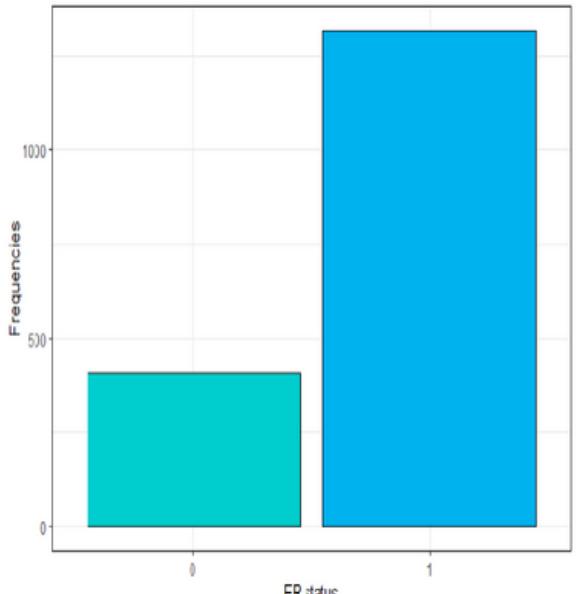
Tumor size



HER2 status



ER status



TP53

TP53 is a tumor suppressor gene. **Mutant TP53** can exert oncogenic effects and enhance metastasis in diverse cancers. [5]

PIK3CA

Mutations on **PIK3CA** lead to a hyperactivation of the PI3K signaling pathway, which can promote uncontrolled cell growth and survival.

LOGISTIC REGRESSION MODEL - RESULTS AND VALIDATION

- After fitting our model on the data, we obtained these results:

Variable Name	OR	P-values
Number of Positive Lymph Nodes	1.75	< 2e-16
Tumor size (T2)	0.52	0.0621
Tumor size (T3)	0.35	0.0938
HER2 Status	0.78	0.5622
ER Status	0.27	6.62e-05
TP53	0.49	2.36e-05
PIK3CA	2.87	< 2e-16

LOGISTIC REGRESSION MODEL - RESULTS AND VALIDATION

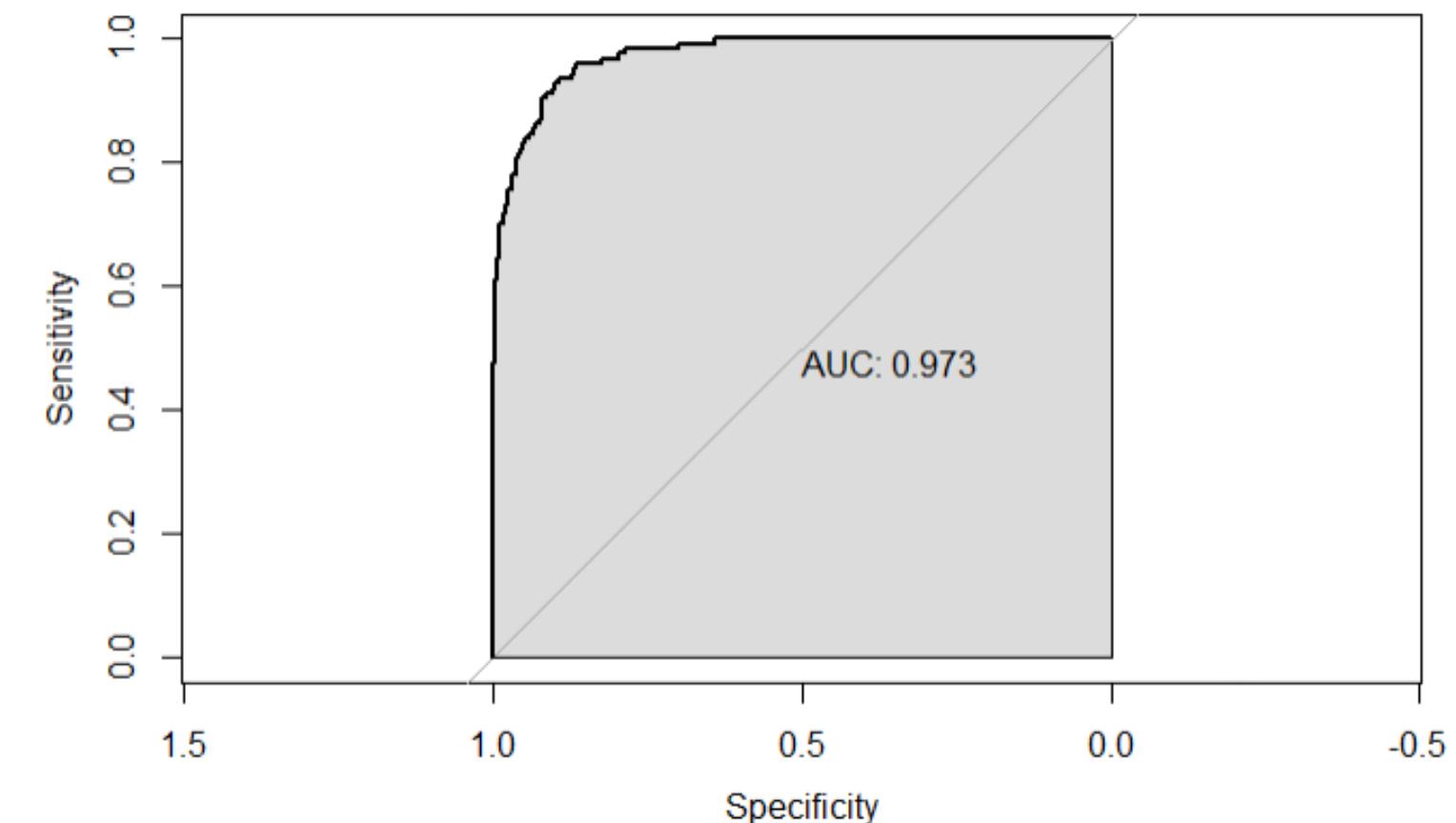
- After fitting our model on the data, we obtained these results:

Variable Name	OR	P-values
Number of Positive Lymph Nodes	1.75	< 2e-16
Tumor size (T2)	0.52	0.0621
Tumor size (T3)	0.35	0.0938
HER2 Status	0.78	0.5622
ER Status	0.27	6.62e-05
TP53	0.49	2.36e-05
PIK3CA	2.87	< 2e-16

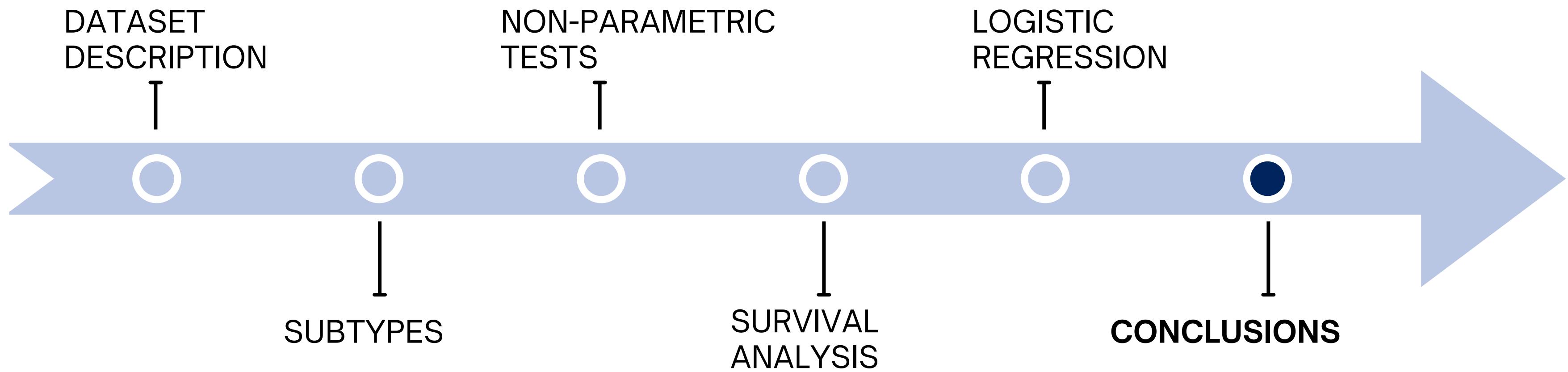
- To evaluate the model we calculated the specificity and the sensitivity by using a 10 folds cross validation with an empirical threshold of 7%.

- Sensitivity: 0.87
- Specificity: 0.91

To see the relationship between specificity and sensitivity we plotted the ROC curve



STEPS



CONCLUSION

- There is an association between overall survival and **tumor size, grade** and the **different sub-populations** while there is no association with the intervals of age (under/over 60);
- Patients with ER+/HER2- , lower tumor size or lower tumor grade have a better survival.
- Tumor grades 2 and 3 and the ER expression status are considered risk factors after 8 years from diagnosis even if they were not significant in past intervals (or even more peculiar, if they were protective factors, like ER in the first interval).
- Finally the logistic regression showed how in predicting if a patient has metastasis ER status and TP53 are protective factors while the number of lymph nodes and in particular PIK3CA are risk factors; tumor size T2 and T3 as well as HER2 status aren't significant.

CONCLUSION

- There is an association between overall survival and **tumor size, grade** and the **different sub-populations** while there is no association with the intervals of age (under/over 60);
- Patients with ER+/HER2- , lower tumor size or lower tumor grade have a better survival.
- Tumor grades 2 and 3 and the ER expression status are considered risk factors after 8 years from diagnosis even if they were not significant in past intervals (or even more peculiar, if they were protective factors, like ER in the first interval).
- Finally the logistic regression showed how in predicting if a patient has metastasis ER status and TP53 are protective factors while the number of lymph nodes and in particular PIK3CA are risk factors; tumor size T2 and T3 as well as HER2 status aren't significant.

ER+ tumors have generally better prognosis, even if after 8 years from the diagnosis the ER status can be considered a risk factor

ER status can be considered a predictor of survival and metastasis but it cannot be used as the only one[6].

BIBLIOGRAPHY

- [1] Pasić R, Dulbegović B, Softić D, Wittlif JL. Steroidni receptori i karcinom dojke [Steroid receptors and breast carcinoma]. Med Pregl. 1991;44(1-2):48-51.
- [2] Pegram, M., Jackisch, C. & Johnston, S.R.D. Estrogen/HER2 receptor crosstalk in breast cancer: combination therapies to improve outcomes for patients with hormone receptor-positive/HER2-positive breast cancer. npj Breast Cancer 9, 45 (2023).
- [3] Allen M Gown, Current issues in ER and HER2 testing by IHC in breast cancer, Modern Pathology, Volume 21, Supplement 2, 2008, Pages S8-S15, ISSN 0893-3952.
- [4] Yang C, Zhang X, Chen Y, Li P, Zhang J, Xu A, Huang N, Liang M, Chen Y, Wang K. Survival differences between HER2-0 and HER2-low-expressing breast cancer - A meta-analysis of early breast cancer patients. Crit Rev Oncol Hematol. 2023 May;185:103962. doi: 10.1016/j.critrevonc.2023.103962. Epub 2023 Mar 14. PMID: 36921780.
- [5] Tang Q, Su Z, Gu W, Rustgi AK. Mutant p53 on the Path to Metastasis. Trends Cancer. 2020 Jan;6(1):62-73. doi: 10.1016/j.trecan.2019.11.004. Epub 2019 Dec 16. PMID: 31952783; PMCID: PMC7485681.
- [6] Lower EE, Glass EL, Bradley DA, Blau R, Heffelfinger S. Impact of metastatic estrogen receptor and progesterone receptor status on survival. Breast Cancer Res Treat. 2005 Mar;90(1):65-70. doi: 10.1007/s10549-004-2756-z. PMID: 15770528

**THANK YOU
FOR YOUR ATTENTION**