



Stat 6227 final project

Survival Analysis of German Breast Cancer

Group 3


Rui Li, Ruqian Qin, Jian Sun, Hanbing Zhou

The George Washington University





Outline

- Background and Dataset
 - Model and Analysis
 1. Data Analysis and Model Selection
 2. Regression Diagnostics
 3. Analyze Specific Problems
 - Finale
- 

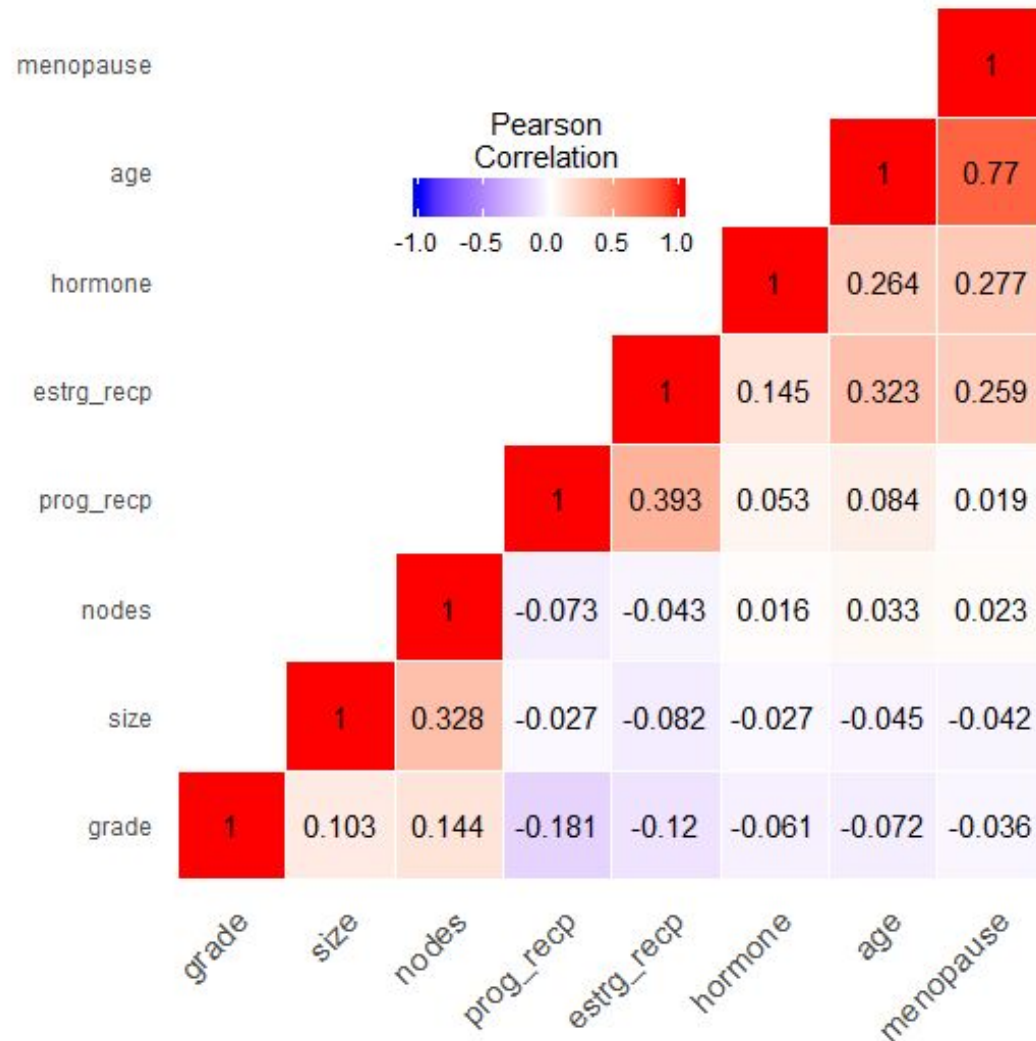
Background and Dataset

- One prospective study of node-positive breast cancer of the German Breast Cancer Study Group (GBSG)
- Total 686 patients, of whom 299 had an event for recurrence-free survival and 171 died
- Primary outcome variable: overall survival and recurrence-free survival time
- Standard prognostic factors – age, menopausal status, tumour size, tumour grade, number of positive lymph nodes, progesterone (PR) and estrogen receptor (ER) concentrations – were investigated

Data Explanation

Variable	Codes/Values	Mean	Sd	Quartiles		
Numeric:				25%	50%	75%
Age at diagnosis	Years	53.05	10.12	46	53	61
Tumor Size	mm	29.33	14.30	20	25	35
# of Nodes involved	1-51	5.01	5.48	1	3	7
# of Progesterone Receptors	0-2380	110.00	202.33	7	33	132
# of Estrogen Receptors	0-1144	96.25	152.08	8	36	114
Categorical:			Codes: Number			
Menopausal status	1=Yes, 2=No		Yes: 290; No:396			
Hormone Therapy	1=Yes; 2=No		Yes: 440; No: 246			
Tumor Grade	1-3(G-1, G-2, G-3)		G-1:81; G-2:444; G-3:161			

Correlation Heat Map



Study Target:

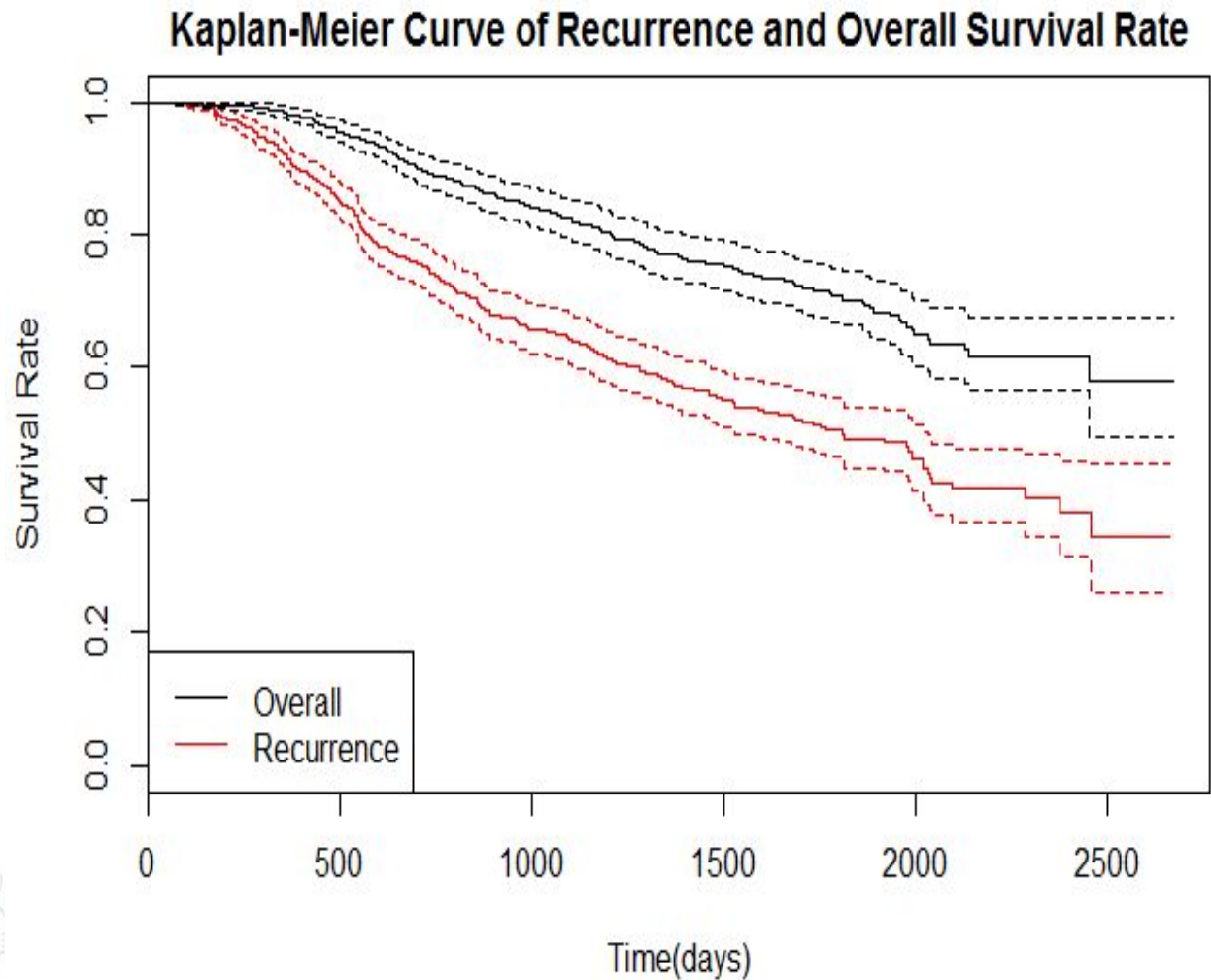
- Build a prognostic model that predicts the clinical course of a breast cancer patient.
- Modelling the effects of standard prognostic factors
- Evaluate efficacy of hormone therapy and provide treatment advice based on research

A decorative network diagram in the top-left corner, featuring a complex web of interconnected nodes and lines. The nodes are represented by circles of varying sizes, some with concentric rings, and the lines are thin and grey. The diagram is partially cut off by the left edge of the frame.

Data Analysis and Model Selection



K-M estimator for survival rate



Hypothesis test

- **Menopause:**

$$H_0 : S_{\text{pre-menopause}}(t) = S_{\text{post-menopause}}(t)$$

$$H_a : S_{\text{pre-menopause}}(t) \neq S_{\text{post-menopause}}(t)$$

- For overall survival:

	Q	Var	Z	pNorm
1	4.5157e+00	4.1591e+01	0.70020	0.48380

- For recurrence free survival:

	Q	Var	Z	pNorm
1	4.5020e+00	7.2351e+01	0.529281	0.59661



Hypothesis test

- Hormone therapy:**

$$H_0: S_{\text{treated}}(t) = S_{\text{non-treated}}(t)$$
$$H_a: S_{\text{treated}}(t) \neq S_{\text{non-treated}}(t)$$

- For overall survival:

	Q	Var	Z	pNorm
1	-1.0166e+01	4.0347e+01	-1.60045	0.109498

- For recurrence free survival:

Call:

```
survdifff(formula = Surv(rectime, censrec) ~ hormone, data = gbcs,  
rho = 0)
```

	N	Observed	Expected	(O-E) ² /E	(O-E) ² /V
hormone=0	440	205	180	3.37	8.56
hormone=1	246	94	119	5.12	8.56

Chisq= 8.6 on 1 degrees of freedom, p= 0.00343

Stepwise Cox Model Selection

	Least AIC	Kept Variables
Survtime	1953.89	Size, grade2, grade3, nodes and prog_recp
Rectime	3485.68	Hormone0, size, grade2, grade3, nodes and prog_recp

- For overall survival, size, grade2, grade3, nodes have positive effect on survival rate.
- For recurrence event, all variables have positive effect.





Regression Diagnostics

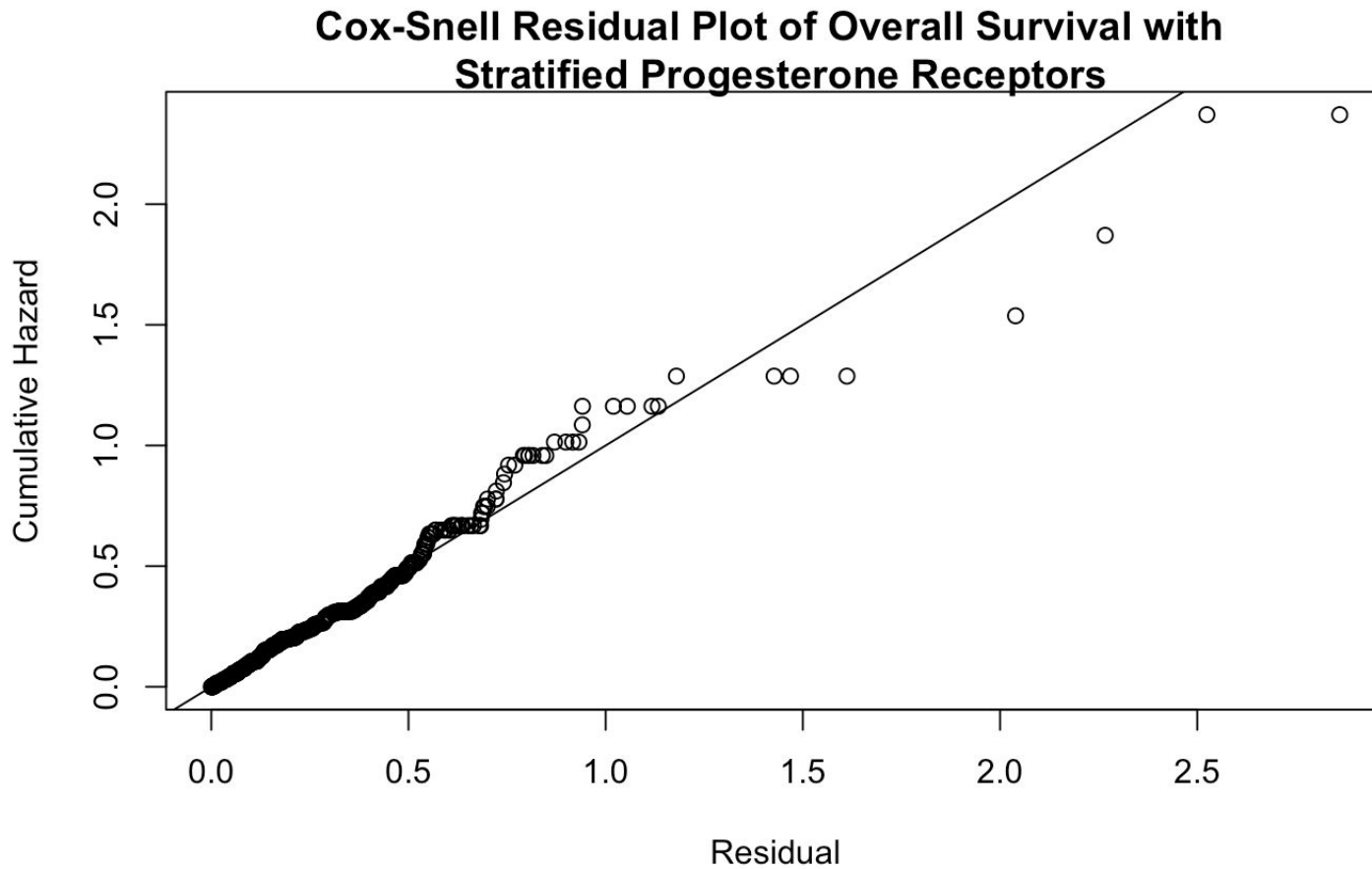
Overall Survival Model Diagnosis - Proportional Assumption

	rho	chisq	p
size	-0.03166	0.17473	0.6759
grade2	0.00706	0.00861	0.9261
grade3	-0.02751	0.13326	0.7151
nodes	0.09030	0.91654	0.3384
prog_recp	0.17272	8.47595	0.0036
GLOBAL	NA	12.64064	0.0270

The prog_recp (PR) has p-value < 0.05 , it might violate the proportional assumption, so we applied stratified model with

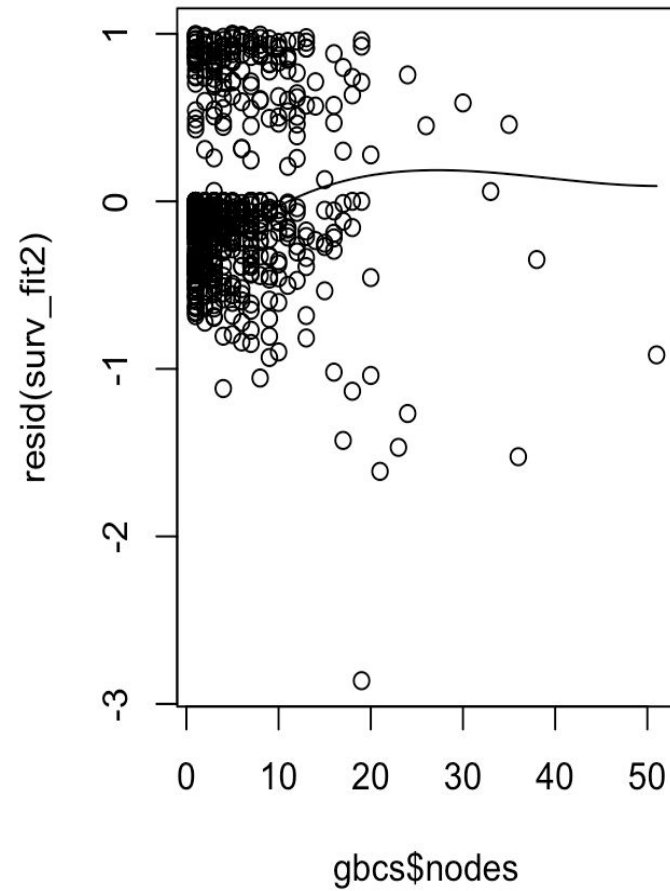
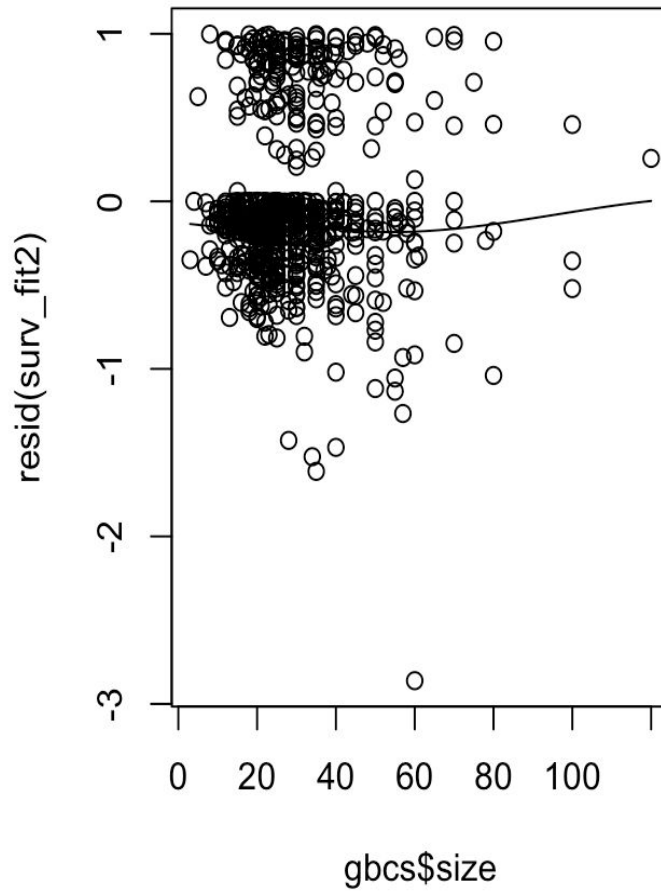
PR level 1: < 20 ; level 2: $[21, 90]$; level 3: > 91

After log-likelihood test, we could not reject the stratified model.

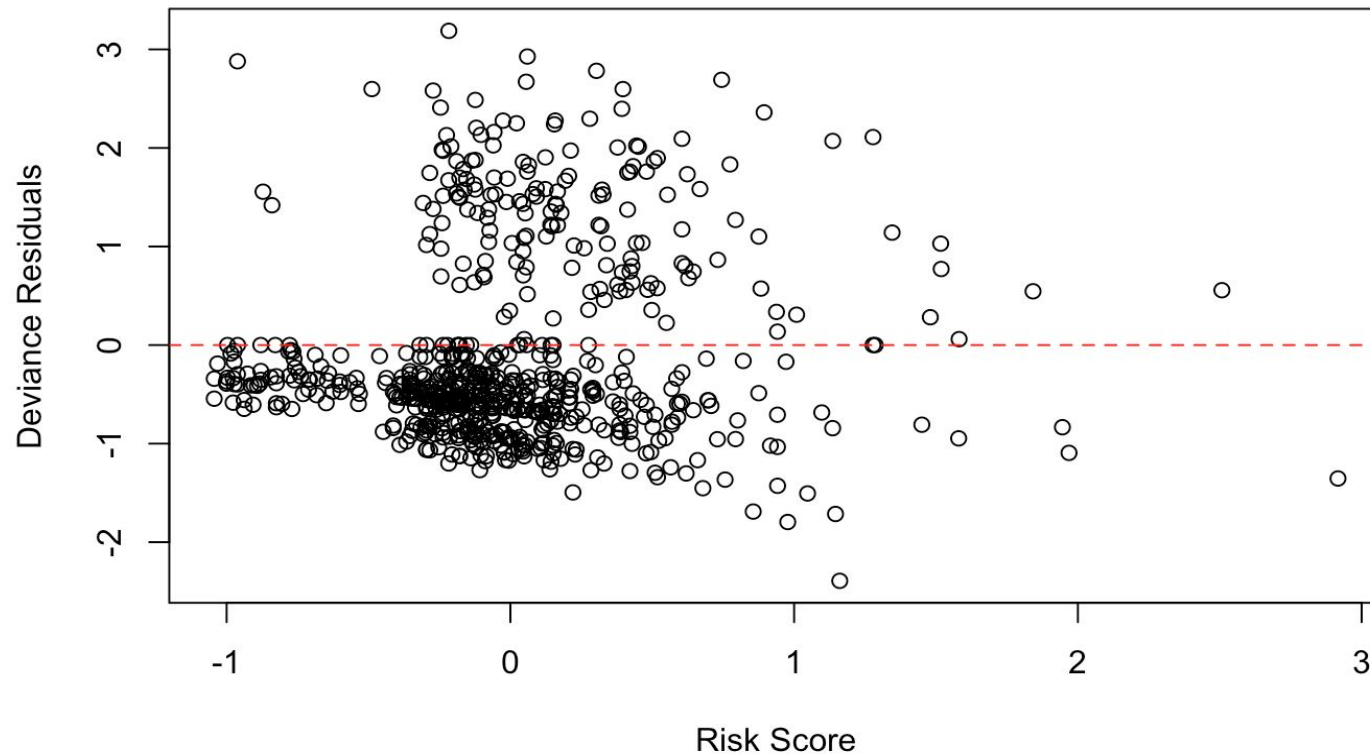


Might consider variable transformation ?

Martingale Residual Plot



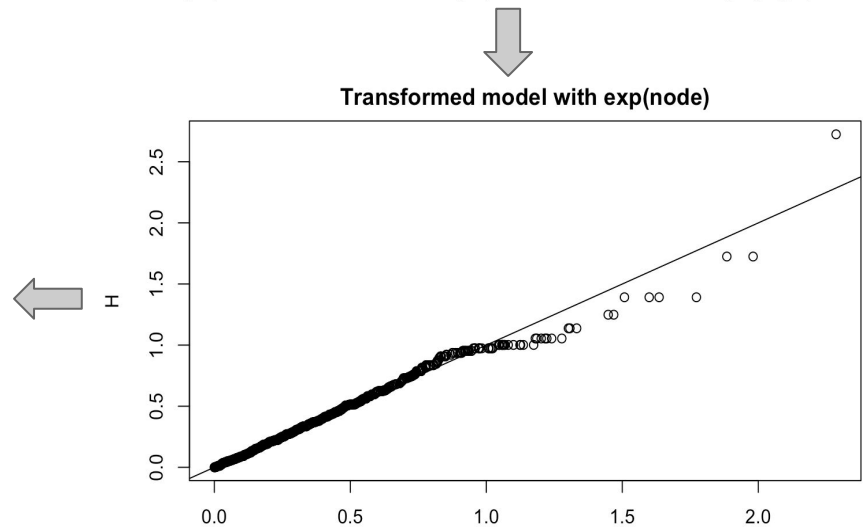
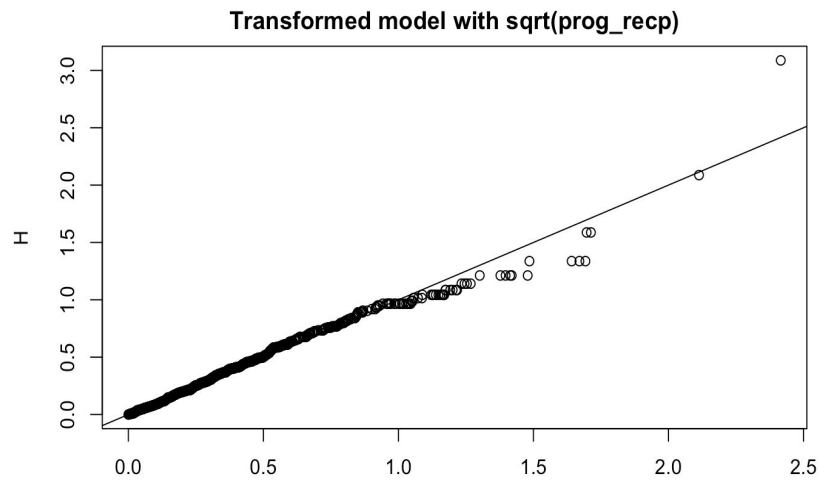
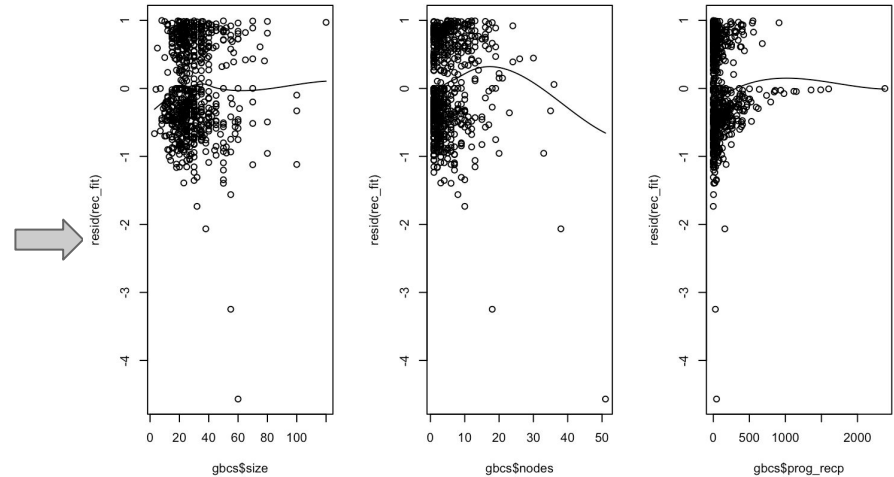
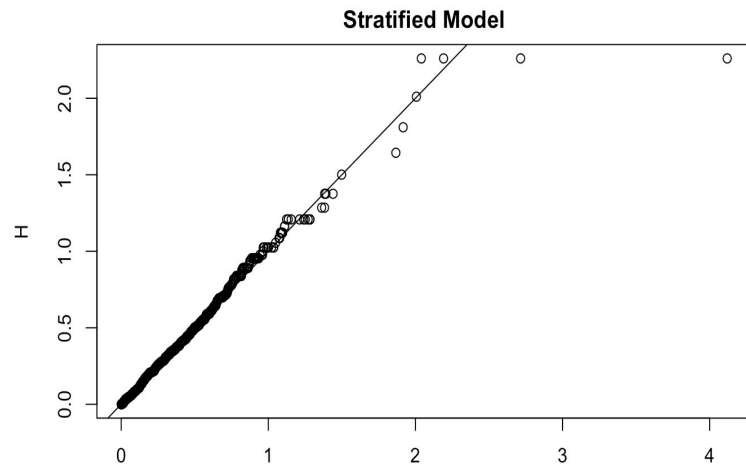
Deviance Residuals Plot



Not very symmetrix:

- Too much censoring data (>50%)
- Transformation might needed???

Recurrence Model Diagnosis




A decorative background featuring a network diagram with nodes and connecting lines, primarily located in the top-left and bottom-right corners. The nodes are represented by circles of varying sizes, some with concentric rings, and the lines are thin and gray.

Analyze Specific Problems



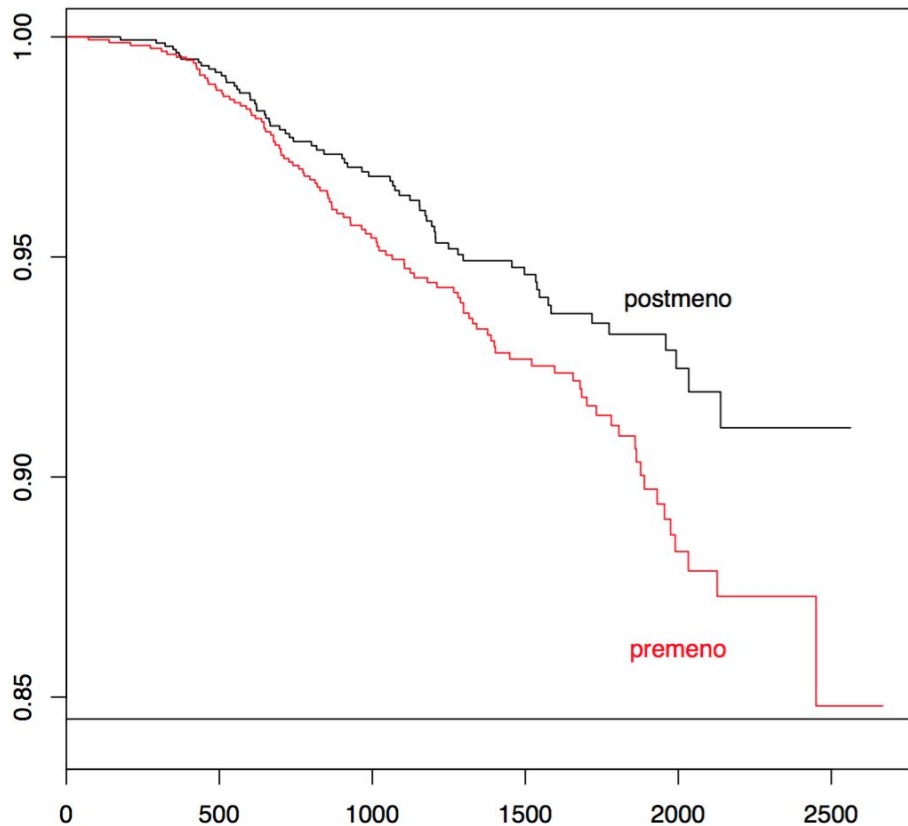
1.

**What is the effect of
late menopause on
patients who are older
than 55 ?**

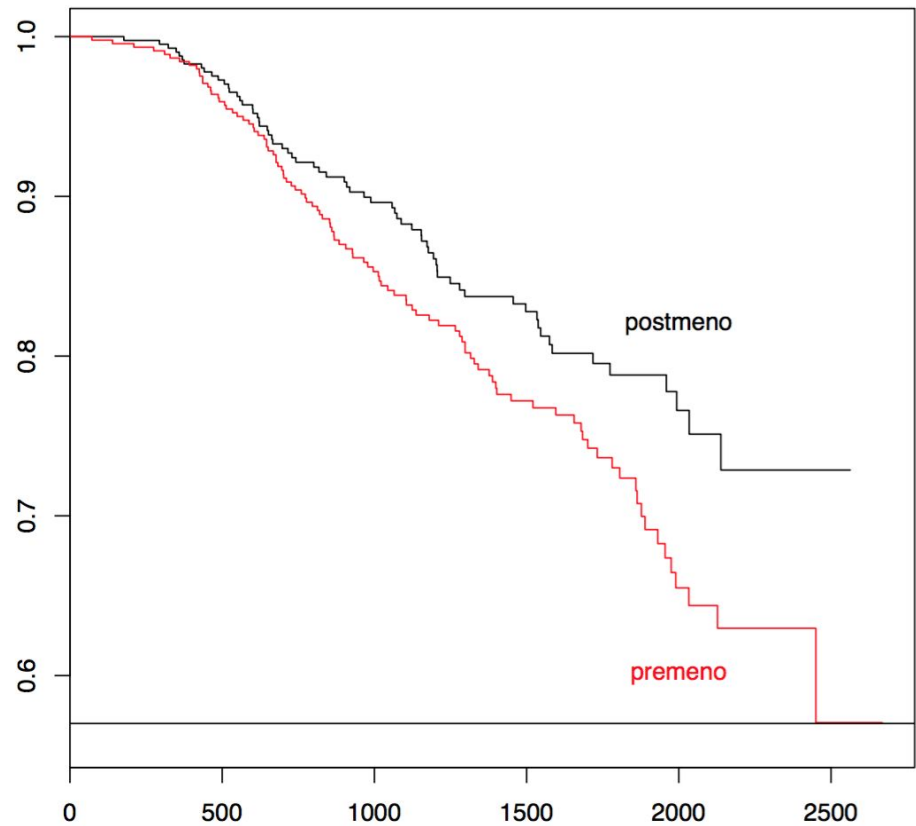


Age: 55 or older Tumor Grade: 1, 2, 3

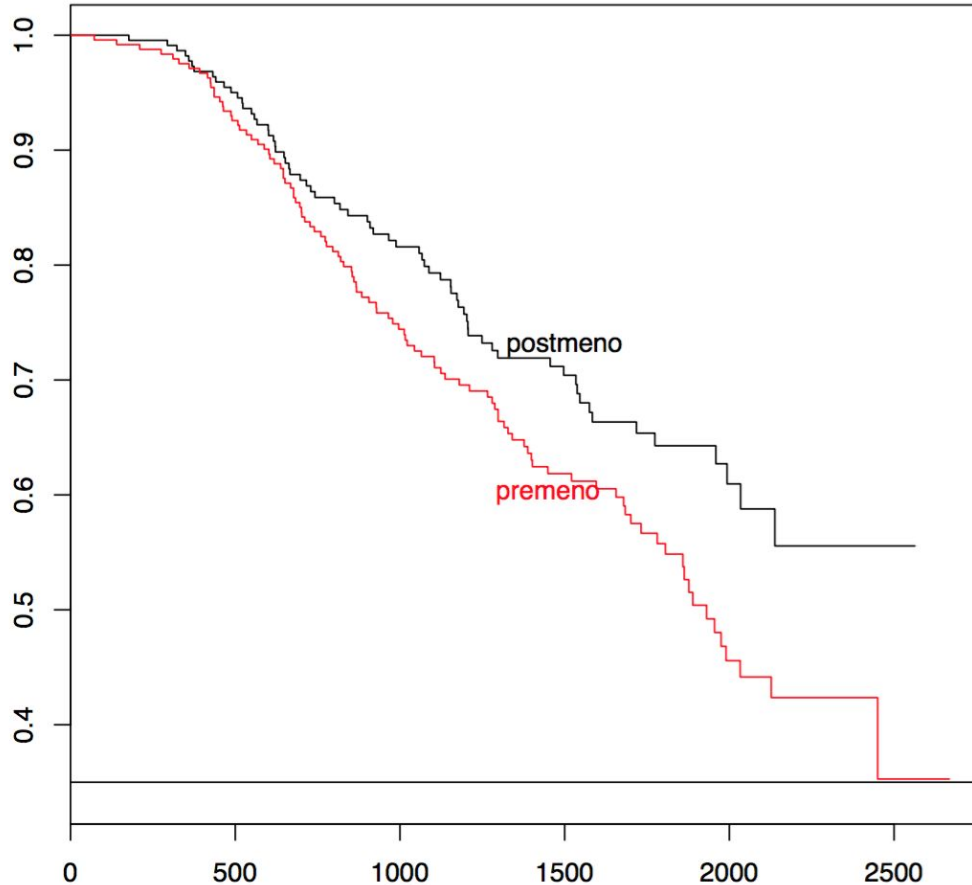
Survival Rate for people over 55 in tumor grade 1,
stratified by menopause



Survival Rate for people over 55 in tumor grade 2,
stratified by menopause



**Survival Rate for people over 55 in tumor grade 3,
stratified by menopause**




1. The higher tumor grade is, the lower survival rate is.
2. Post Menopause women are safer.

A decorative network diagram in the top-left corner, consisting of interconnected nodes and lines, with some nodes highlighted in blue.

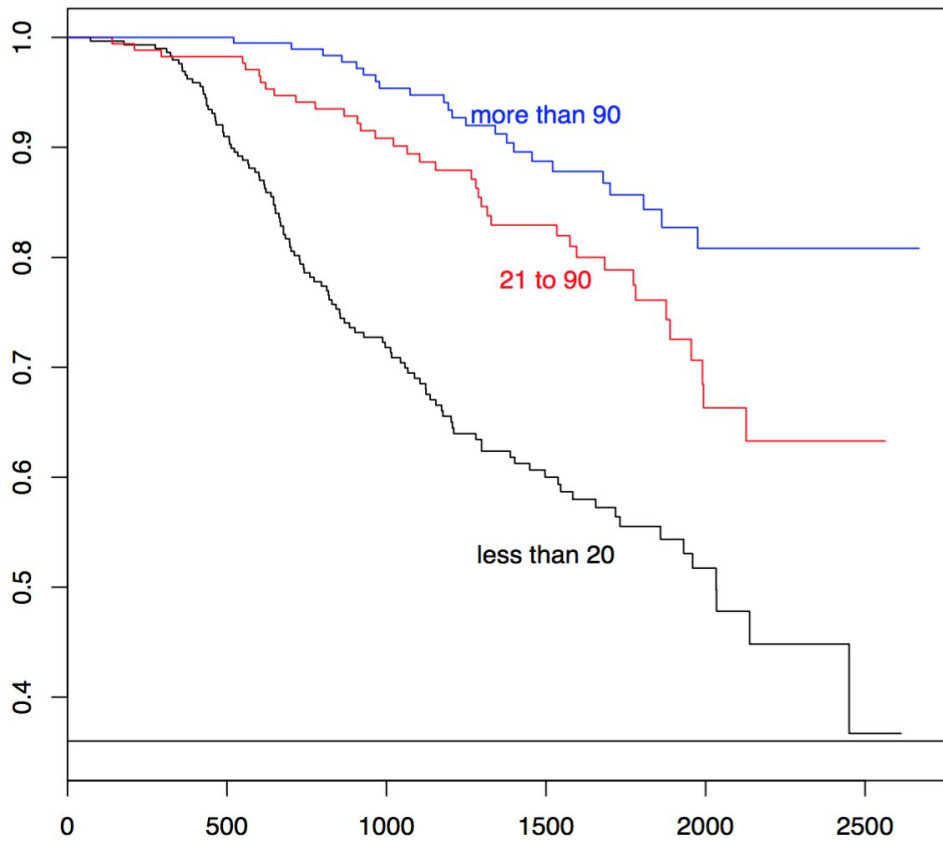
2.

**Does PR or ER work
well on curbing Breast
Cancer ?**

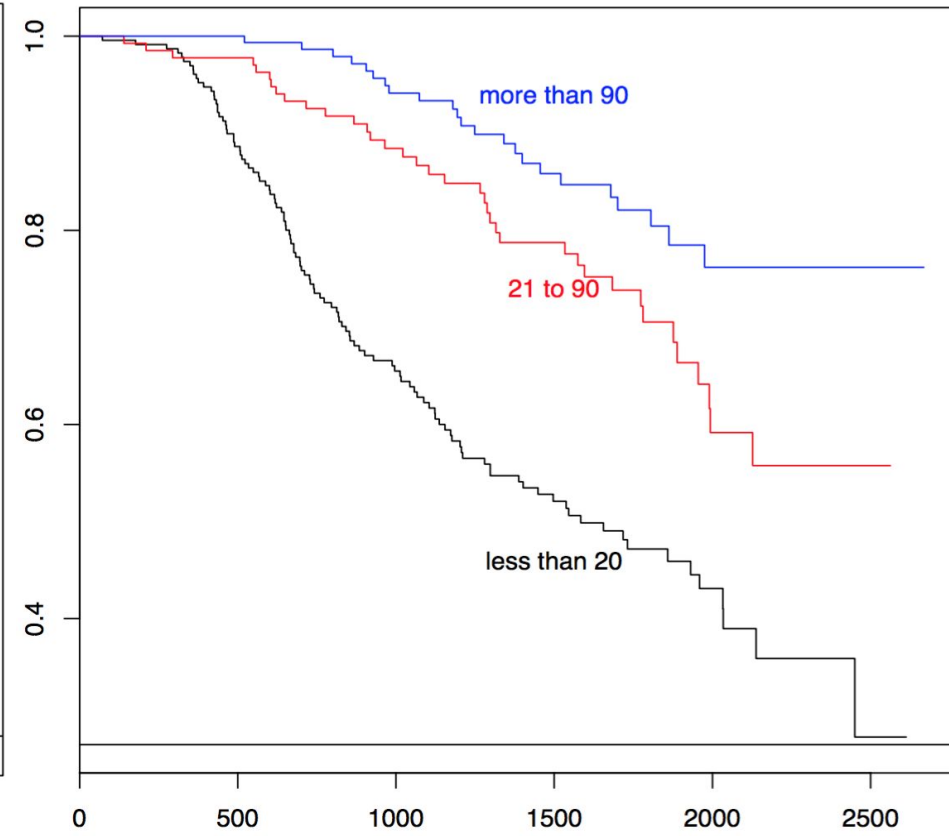
A decorative network diagram in the bottom-right corner, consisting of interconnected nodes and lines, with some nodes highlighted in blue.

Age: 55 or older Tumor Grade: 2, 3

**Survival Rate for people over 55 in tumor grade 2,
stratified by Progesterone Receptor**

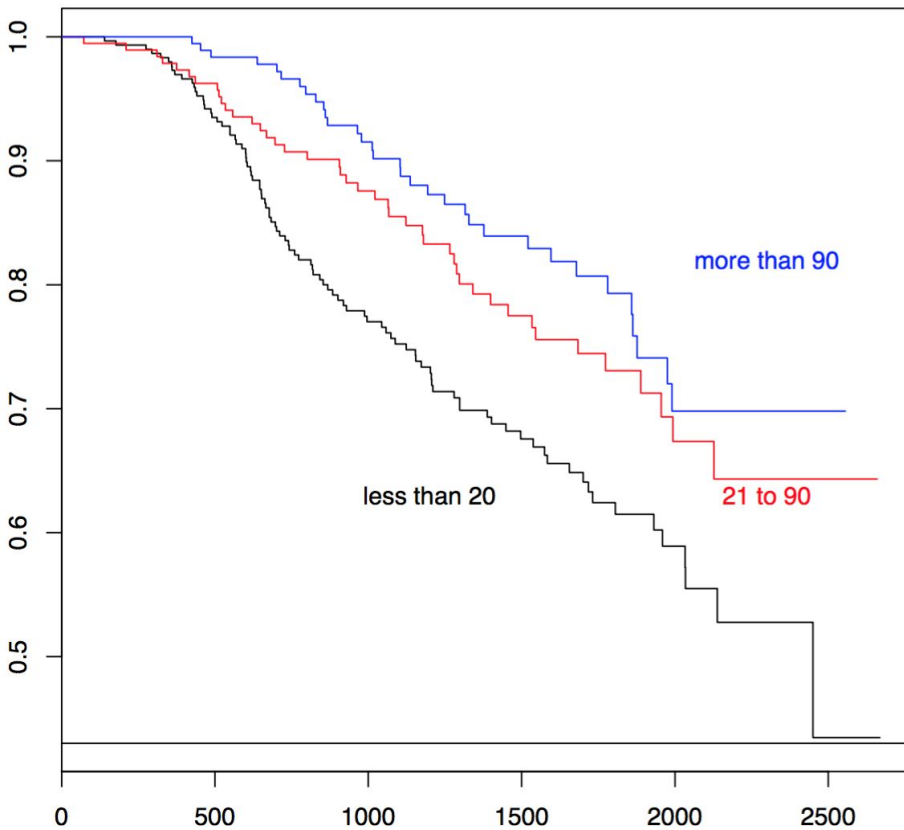


**Survival Rate for people over 55 in tumor grade 3,
stratified by Progesterone Receptor**

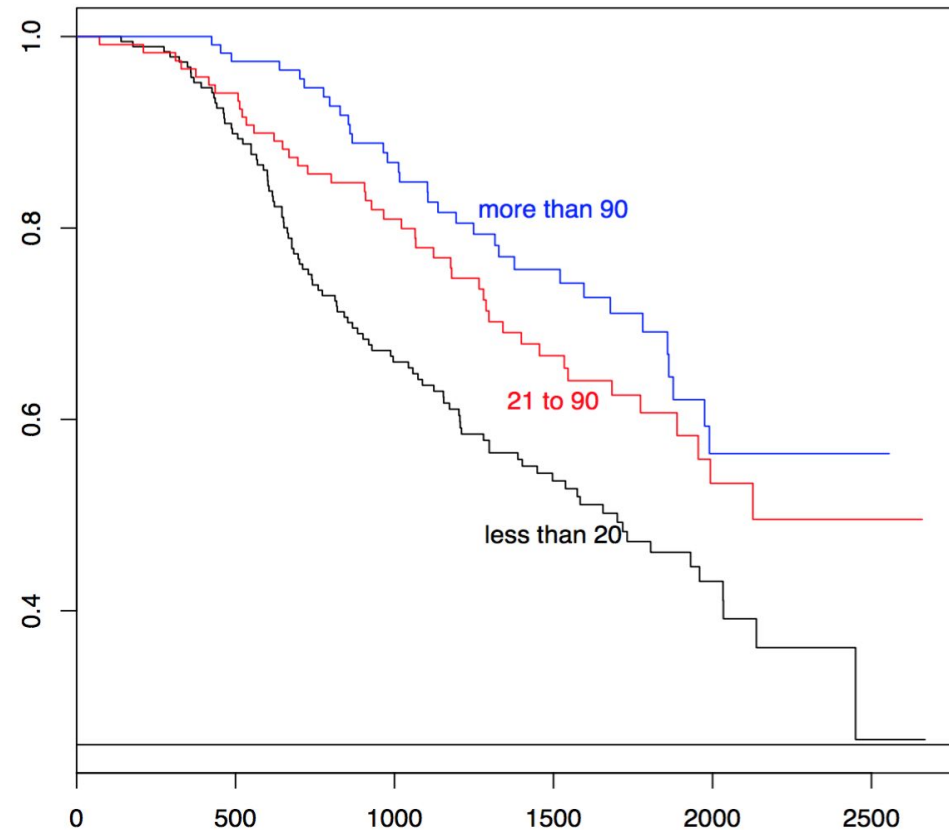


Higher receptors repress cancer and improve survival rate.
Both receptors study lead to same results.

Survival Rate for people over 55 in tumor grade 2, stratified by Estrogen Receptor




Survival Rate people over 55 in tumor grade 3, stratified by Estrogen Receptor



A decorative network diagram in the top-left corner, consisting of interconnected nodes and lines, rendered in a light gray color. The nodes are represented by small circles, some of which are double-lined, and the lines are thin and gray.

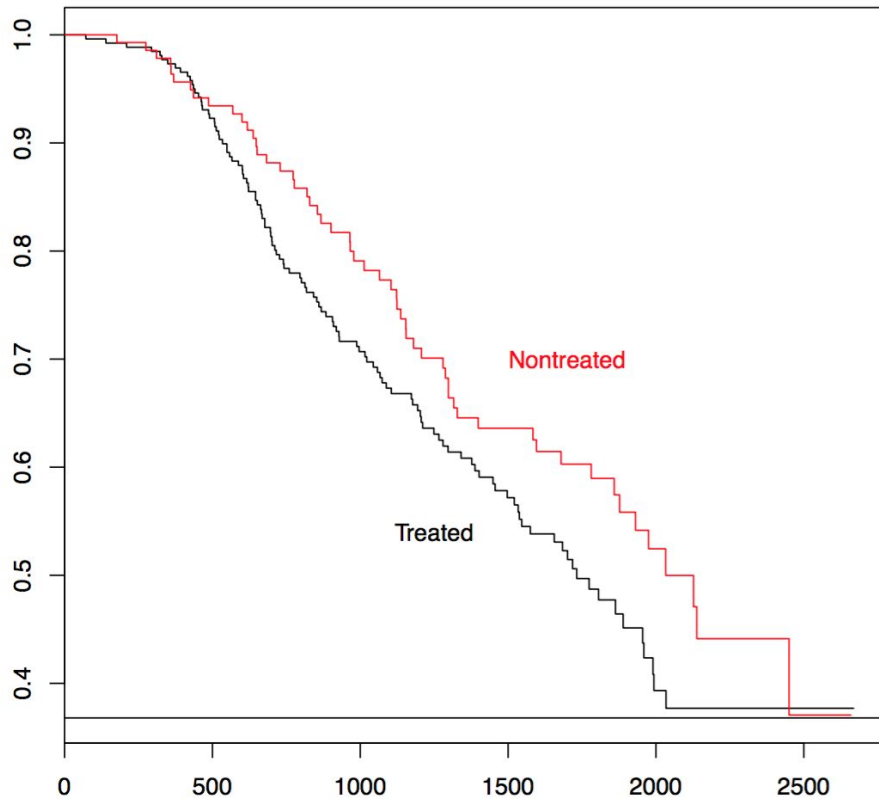
3.

**Is it useful for patients
to use hormone
therapy on controlling
disease ?**

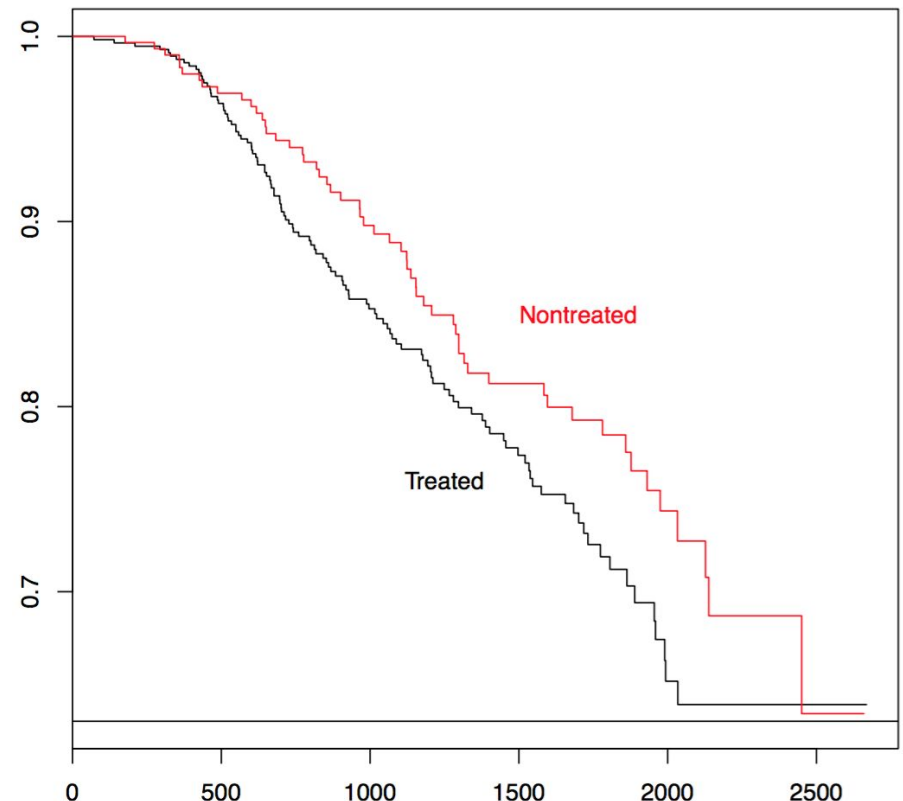
A decorative network diagram in the bottom-right corner, consisting of interconnected nodes and lines, rendered in a light gray color. The nodes are represented by small circles, some of which are double-lined, and the lines are thin and gray.

Age: under 45; over 55 Tumor Grade: 2,3
Progesterone Receptor: lo:20; 21:90

Survival Rate for women over 55 in tumor grade 2 and lower receptors, stratified by hormone



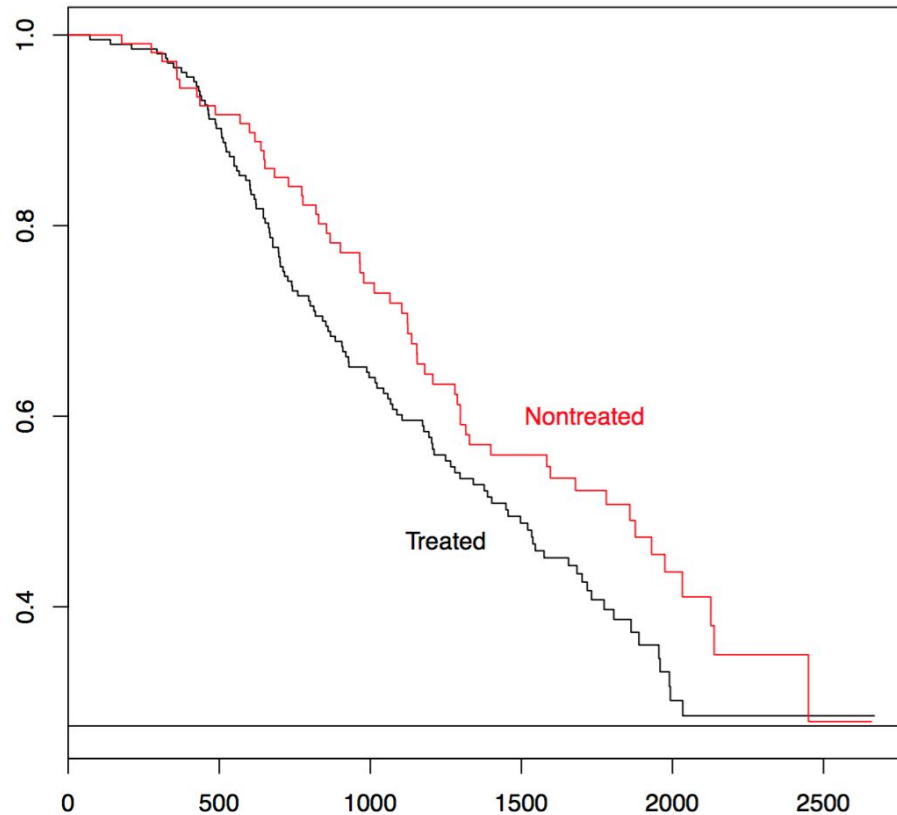
Survival Rate for women over 55 in tumor grade 2 and middle receptors, stratified by hormone



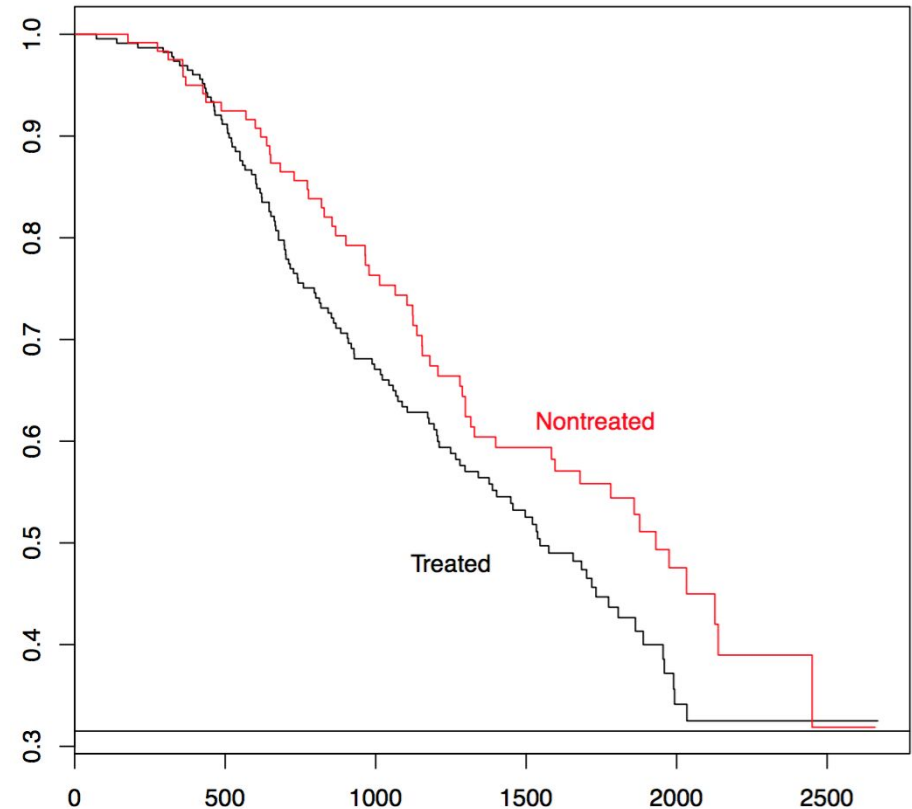
Higher Progesterone Receptor Number does not make hormone therapy work better.

Serious Tumor Grade and Younger Age do not make the hormone therapy work better

Survival Rate for women over 55 in tumor grade 3 and lower receptors, stratified by hormone



Survival Rate for women under 45 in tumor grade 3 and lower receptors, stratified by hormone



On different conditions,
getting hormone therapy reduces survival rate.

Finale

1. For the overall survival rate, significant factors are:
 - Tumor size
 - Grade of tumor
 - Number of lymph nodes
 - PR concentrations;
2. For recurrence patients, besides the factors above, hormone therapy might also affect survival rate.

Tip1

Inquire if premenopause over 55, otherwise hard to cure cancer;

Tip2

For receptor-test result positive patients, the hormone therapy is recommended, otherwise they can do chemotherapy.

Tip3

The efficacy of hormone therapy might vary among different case scenario



Q & A