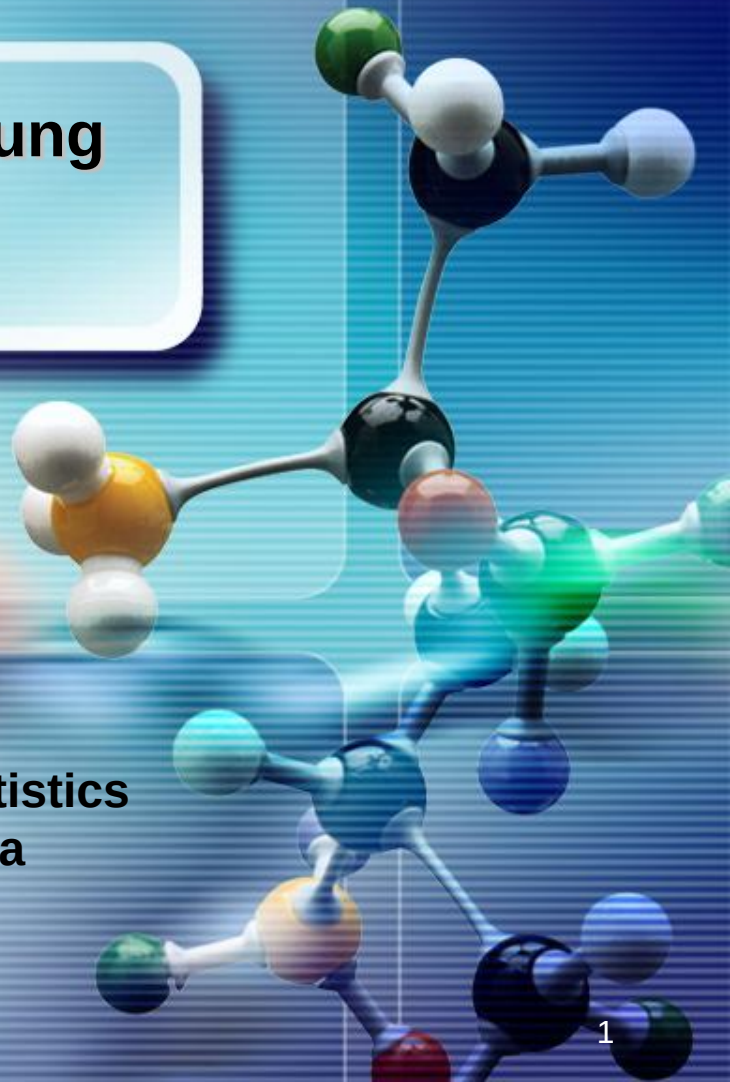
Statistical Modeling of Breast and Lung Cancers

Cancer Research Team

Department of Mathematics and Statistics
University of South Florida




Thank You!

Outline

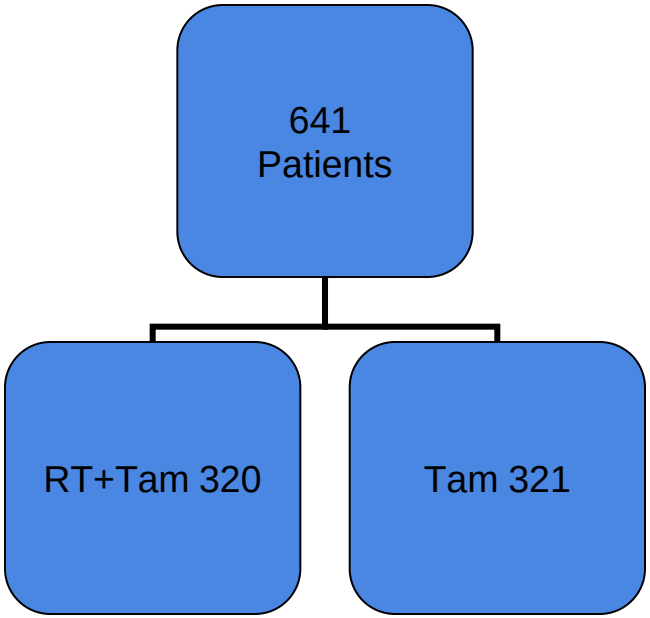
- Nonparametric and parametric analysis of treatment effectiveness of breast cancer
- Statistical modeling of relapse time of breast cancer with different treatments
- Markov modeling of breast cancer states
- Statistical analysis of lung cancer mortality time
- Sensitivity analysis of breast cancer doubling time

Objective 1.

Using data which was collected from 1992 to 2000 on breast cancer, a total of 641 women were randomized: 320 in the combined radiation and tamoxifen arm (RT + Tam), and 321 in the tamoxifen-alone arm (Tam). The objective of this presentation is to investigate whether treatment RT + Tam is more effective than Tam alone with respect to the relapse time of a given patient, in which both parametric and nonparametric methods as well as decision tree technique are used.

Reference: Decision tree for competing risks survival probability in breast cancer study by N.A Ibrahim, et al. International Journal of Biomedical Sciences Volume 3 Number 1, 2008

Data



- For the 641 patients, the data includes censored and uncensored observations. Since there are only 77 uncensored observations (26 in RT + Tam arm and 51 in Tam arm) which means nearly 90% are censored observations, the data is analyzed separately for the uncensored data (77) and censored data (641). For simplicity, we call these two datasets dataset 1 and dataset 2 for later use.

- Response variable: relapse time
- Attributable variables:
  1. pathsize: size of tumor
  2. age: age of patient
  3. hgb: haemoglobin
  4. hist: histology (DUC,LOB,MED,MIX,OTH)
  5. hrlevel: hormone receptor level (NEG,POS)
  6. nodediss: Whether axillary node dissection was done (Y,N)

Nonparametric and parametric analysis of treatments

- Nonparametric: Log-rank Test are used to test the difference of mean relapse time of RT + Tam and Tam arms.
- Parametric: Goodness-of-fit tests are used to find the classical distribution for RT + Tam arm and Tam arm with respect to the two datasets.

Results

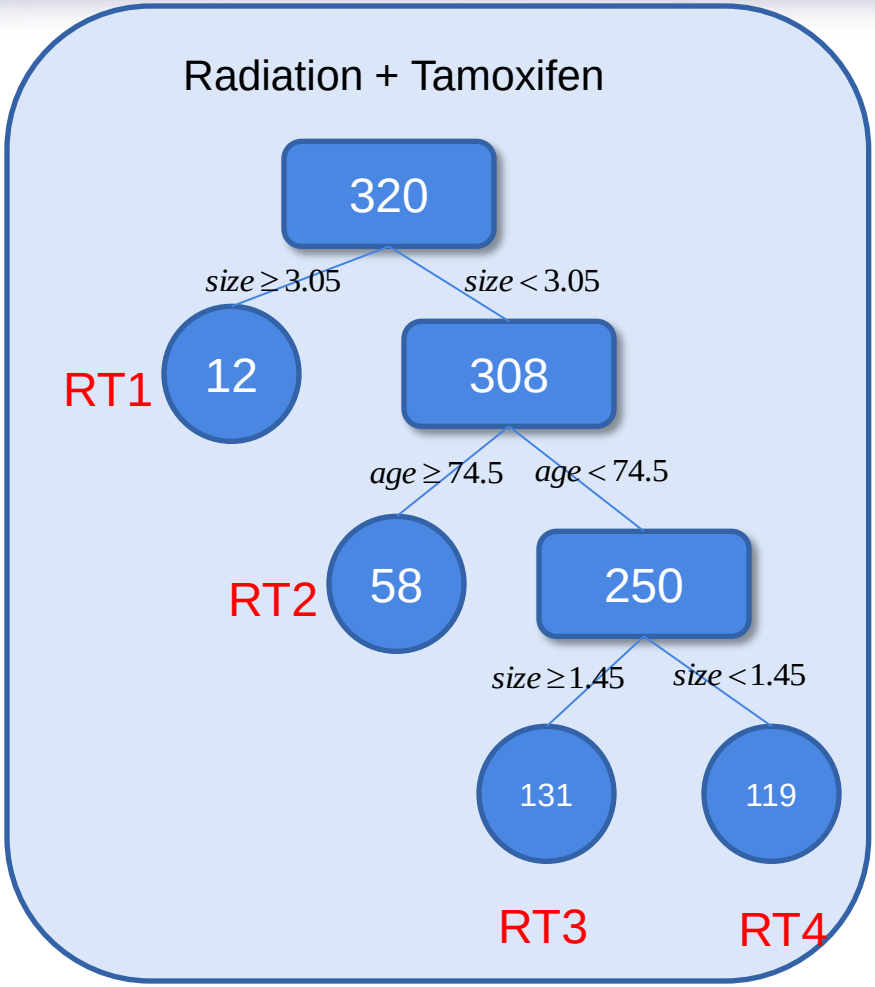
	Nonparametric	Parametric
320 RT+Tam 321 Tam	Log-rank test p-value = 0.0017 $\mu_1 > \mu_2$	RT + Tam (320): **Log-normal** $\mu = 0.903$ $\sigma = 5.148$ Tam (321): **Log-normal** $\mu = 0.583$ $\sigma = 3.491$ Likelihood ratio test : p-value = 0.001~0.05 $\mu_1 > \mu_2$

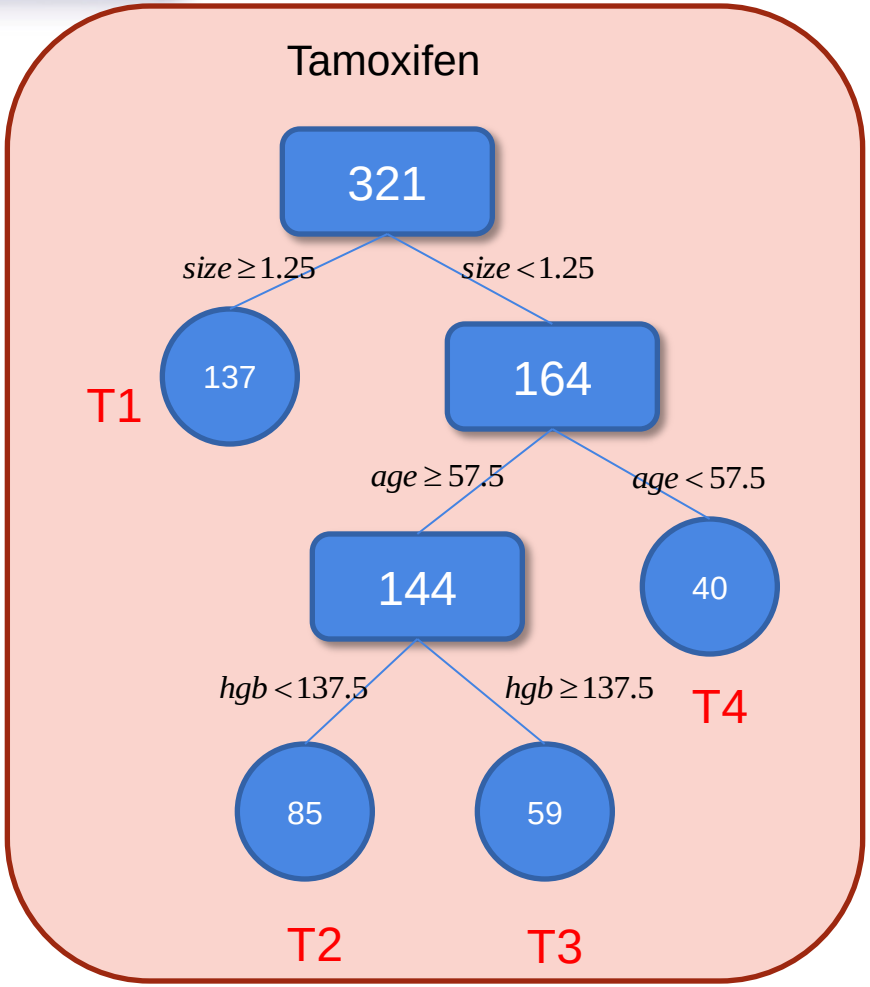
Decision Tree

- The clinicopathological characters of breast cancer patients are heterogeneous. Consequently, the survival times are different in subgroups of patients. Decision tree is used to homogenize the data by separating the data into different subgroups.

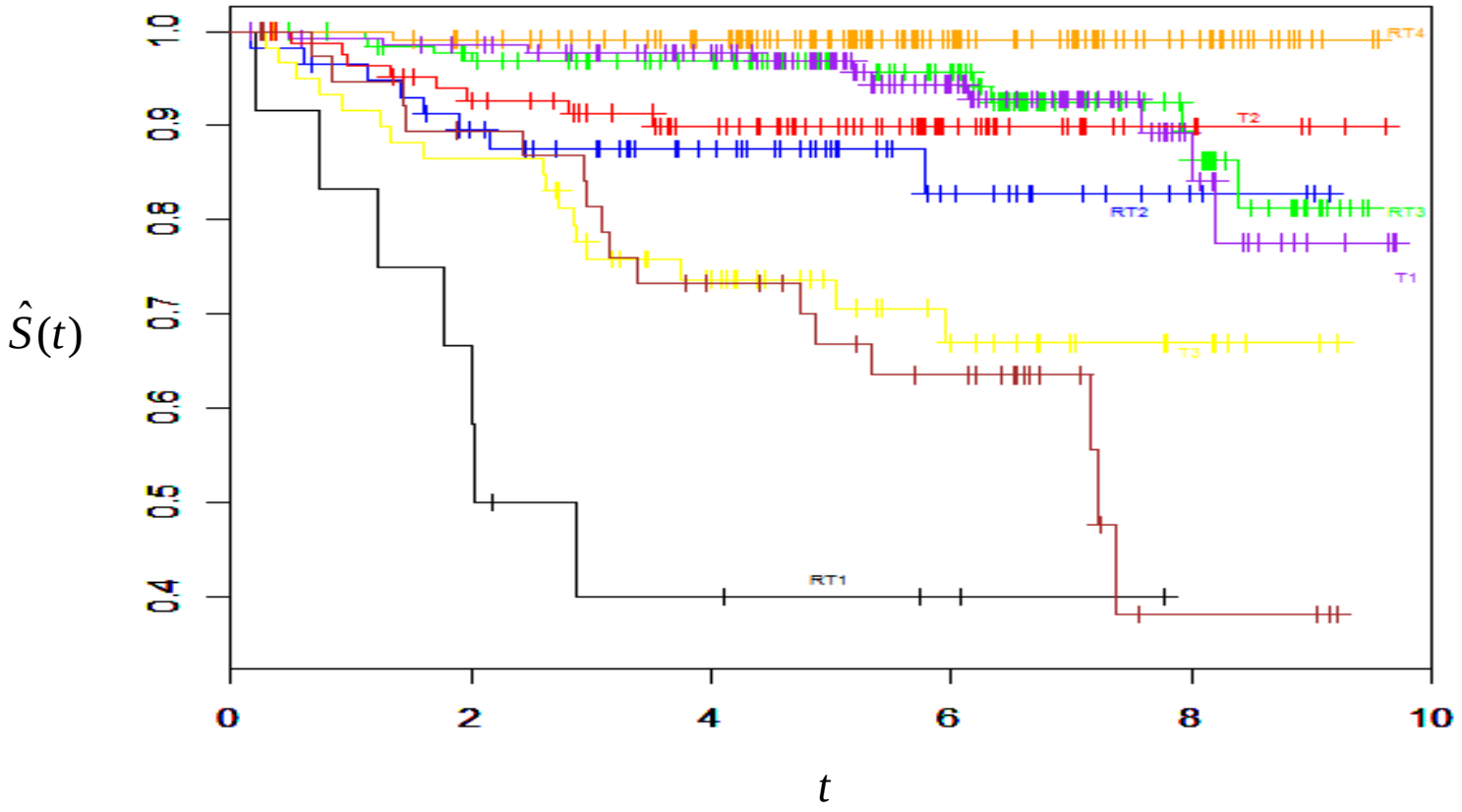
- As can be seen from the following graph. RT+Tam arm is divided into 3 groups denoted by RT1,RT2,RT3,RT4 from the left to the right; Tam arm is divided into 4 groups denoted by T1,T2,T3,T4 from the left to the right.

Decision tree





Decision tree



Conclusion

- As can be seen from the above, although RT+Tam arm has overall better performance than Tam arm, after the partition of each arm, both the best (RT4) and worst (RT1) survival curve come from RT+Tam.

- Overall comparison of the 7 subgroup survival curves shows some clustering
  - ★ RT2,RT3,RT4,T1,T2
  - ★★ T3,T4
  - ★★★ RT1

Objective 2

Using data which was collected from 1992 to 2000 on breast cancer, a total of 641 women were randomized: 320 in the combined radiation and tamoxifen arm (RT + Tam), and 321 in the tamoxifen-alone arm (Tam). The objective of this presentation is to identify the significant factors and possible interactions of those factors that contributes to the reoccurrence of breast cancer, moreover, statistical modeling of relapse time as a response variable based on other attributable variables are presented for predicting purpose. Furthermore, cure rate model is used to compare the effectiveness of different treatments with respect to the cure rate of breast cancer patients.

Reference: Decision tree for competing risks survival probability in breast cancer study by N.A Ibrahim, et al. International Journal of Biomedical Sciences Volume 3 Number 1, 2008

AFT Model and Cox-PH model

- An accelerated failure time model (AFT model) is a parametric model that assumes that the effect of a covariate is to multiply the predicted event time by some constant. Consider a random variable W with a standard distribution and generate a family of survival distributions by introducing location and scale parameter.By adding covariates to the location parameter ,we obtain the AFT model

$$\log T = x^{'}\beta + \sigma W$$

An alternative approach to modeling survival data is to assume the effect of covariates is to multiply the hazard by a constant

$$\lambda(t, x) = \lambda_0(t)e^{x^{'}\beta}$$

Significant Prognostic Factors

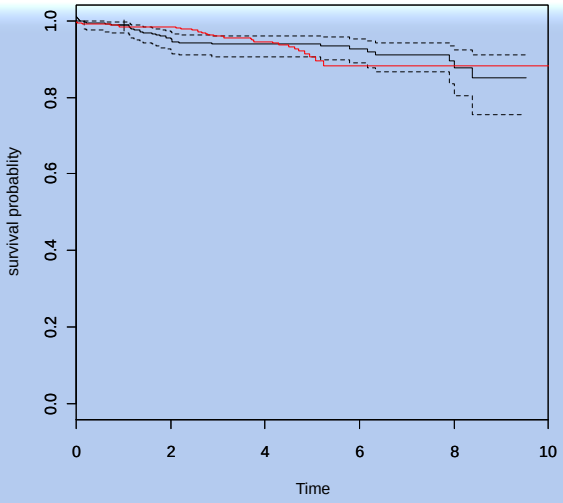
Attributable variables:
pathsize, age, hgb, hist, hrlevel, nodediss

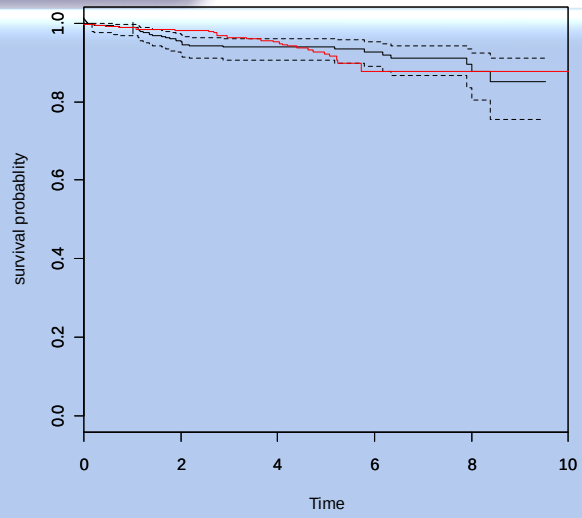
RT+Tam	lognormal	exponential	Weibull	Cox-PH
age	0.002*	0.008*	0.011*	0.01*
pathsize	0.01*	0.0002*	0.0002*	0.00086*
nodediss	0.021*	0.009*	0.012*	0.012*
hrlevel	0.027*	0.010*	0.008*	0.016*
age:nodediss	0.037*	0.022*	0.026*	0.028*
nodediss:hrlevel	0.009*	0.0005*	0.0008*	0.00067*
pathsize:hrlevel	0.078	0.060	0.041*	0.099

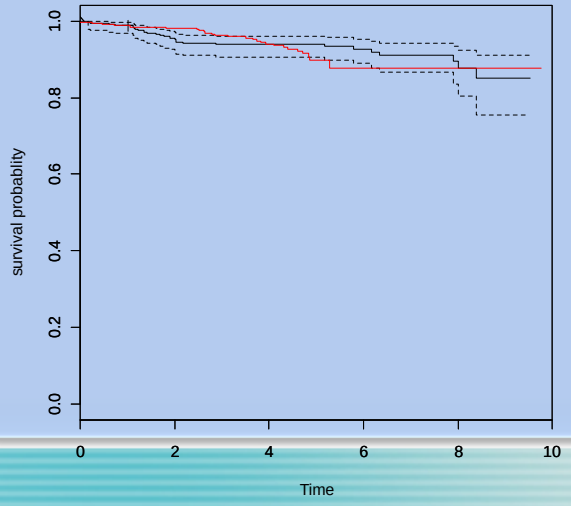
Significant Prognostic Factors

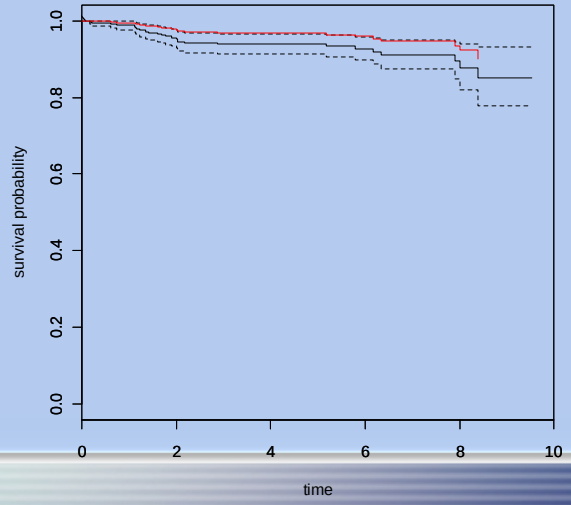
Tam	lognormal	exponential	Weibull	Cox-PH
age	0.343	0.294	0.287	0.32
hgb	0.037*	0.645	0.630	0.68
pathsize	0.339	0.316	0.300	0.33
nodediss	0.025*	0.017*	0.020*	0.018*
hrlevel	0.006*	0.002*	0.003*	0.002*
age:pathsize	0.143	0.112	0.106	0.120
age:nodediss	0.038*	0.006*	0.007*	0.0065*
hgb:nodediss	0.054	0.077	0.079	0.075
age:hgb	NA	0.131	0.128	0.150

Survival Curve for RT+Tam

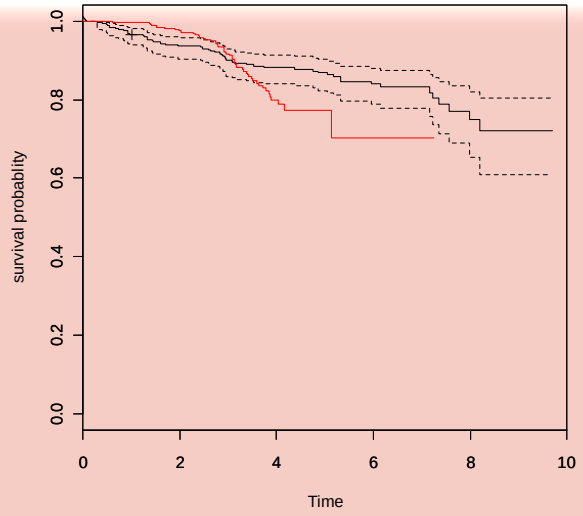


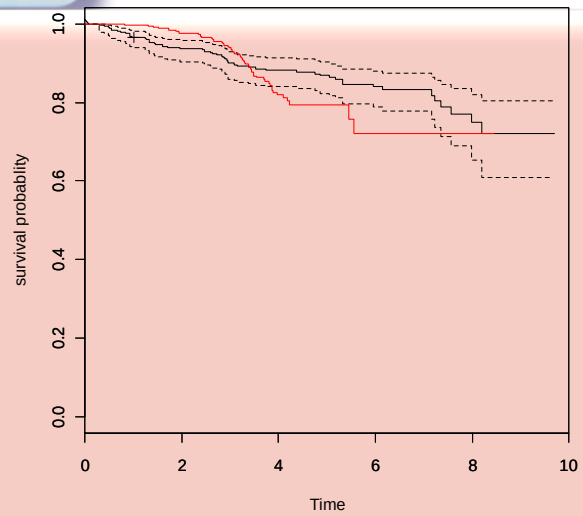


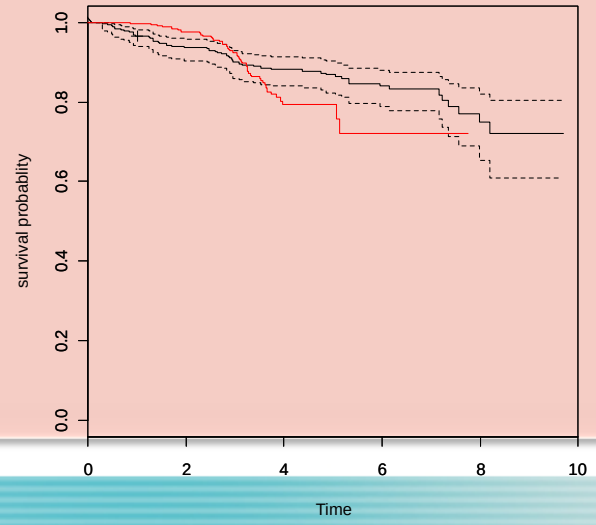


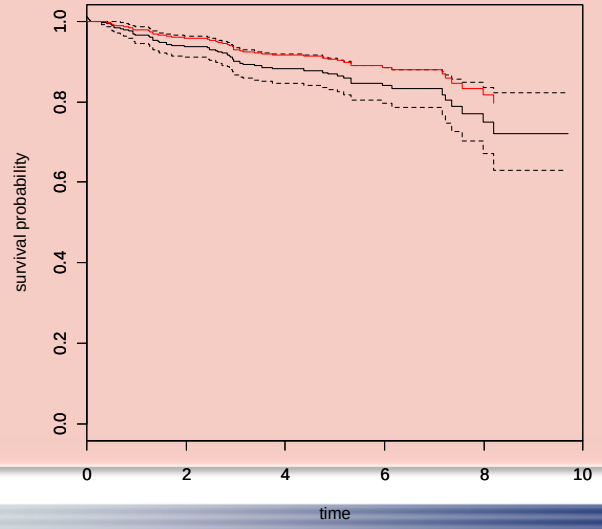


Survival Curve for Tam









Cure Rate Model

- Any clinical trial consists of heterogeneous group of patients that can be divided into two groups. Those who respond favorably to the treatment and subsequently become insusceptible to the disease are called cured. The others that do not respond to the treatment remain uncured or susceptible to reoccurrence of disease.

- Let π denote the proportion of cured patients and 1-π is the proportion of uncured patients, and then the survival function for the group is

$$S(t) = \pi + (1-\pi)S_u(t)$$

  where $S_u(t)$ is the survival function of the uncured group

- Covariates can be included to uncured survival function using AFT model, and they can also be included to cure rate using logistic regression

$$\log(\frac{\pi}{1-\pi}) = \exp(x'\beta)$$

Results

RT+Tam	No covariates	Covariates in survival function	Covariates in survival function and cure rate	Covariates and interactions in survival function
Weibull	0.2471	0.1	Not fixed	0.1
L.normal	0.2593	0.0057	Not fixed	0.1
Gamma	0.8053	0.0064	Not fixed	0.1
G.L.L	0.6118	0.1	Not fixed	0.1
L.logistic	0.2903	0.1	Not fixed	0.1
G.F	0.0065	0.002152	Not fixed	0.1
E.G.G	0.0043	0.0038	Not fixed	0.1
Rayleigh	0.8799	0.1	Not fixed	0.1

Results

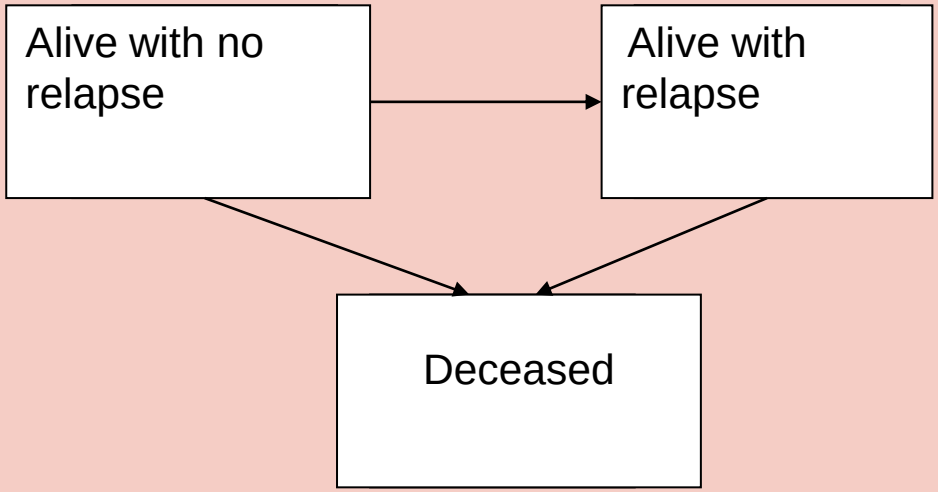
Tam	No covariate	Covariates in survival function	Covariates in survival function and cure rare	Covariates and interactions in survival function
Weibull	0.1980	0.0748	Not fixed	0.0748
L.normal	0.1127	0.0748	Not fixed	0.0748
Gamma	0.1695	0.449	Not fixed	0.0748
G.L.L	0.1515	0.0748	Not fixed	0.0748
L.logistic	0.1874	0.0748	Not fixed	0.0748
G.F	0.0166	0.0748	Not fixed	0.0748
E.G.G	0.0186	0.5582	Not fixed	0.0748
Rayleigh	0.7572	0.0748	Not fixed	0.0748

Objective 3

- To investigate the progression of breast cancer patients between three different states in different treatment groups (RT+Tam and Tam). Transition intensity between different states and transition probabilities among the three states during different time periods such as 2-year, 4-year,5-year, and 10-year are calculated.

Data

- Same data is used and the three stages that we are interested in the study are: alive with no relapse, alive with relapse, and deceased.



Markov Model

- Markov Chain if the conditional probabilities between the stages at different times satisfy the Markov property: the conditional probability of future one-step-event conditioned on the entire past of the process is just conditioned on the present stage of the process.

- The transition probability from stage to stage at time and transition intensity are defined by

$$p_{ij}(t) = p(X_{t+1} = j \mid X_t = i)$$

and

$$q_{ij}(t) = \lim_{h \to 0} \frac{P(X(t+h) = j \mid X(t) = i)}{h}$$

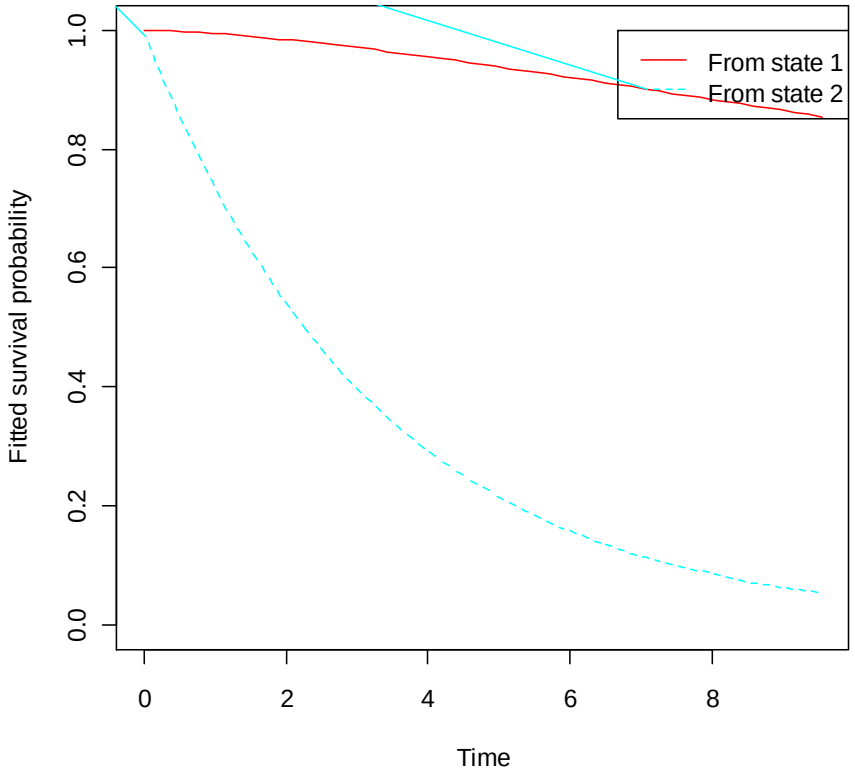
,

Results (transition intensity)

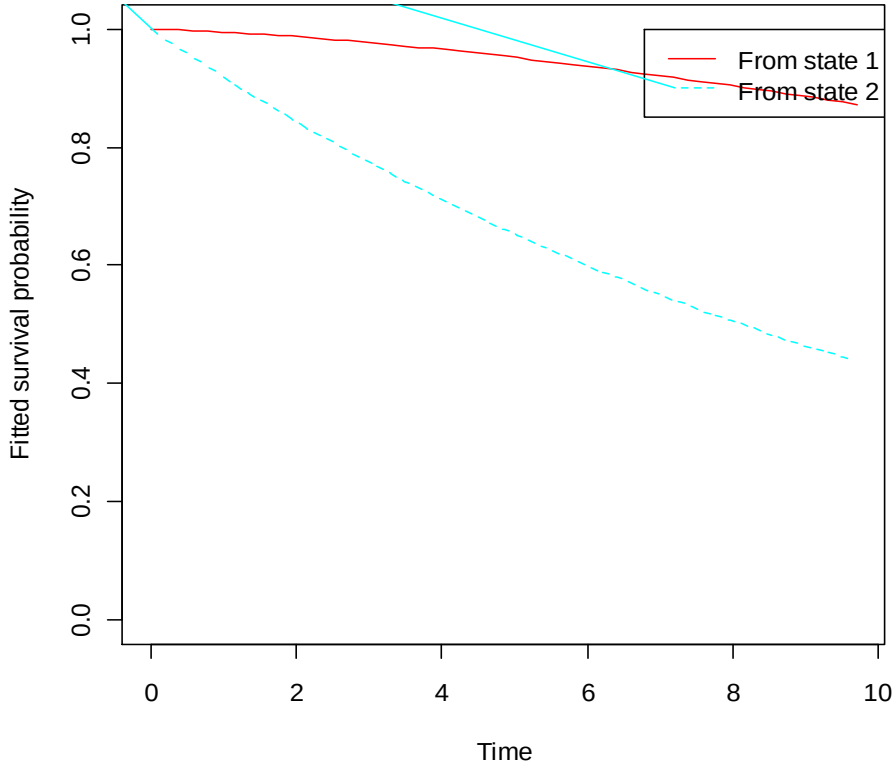
RT+Tam	State 1	State 2	State 3
State 1	-0.02301	0.01957	0.0034
State 2	0	-0.3074	0.3074
State 3	0	0	0

Tam	State 1	State 2	State 3
State 1	-0.03917	0.03528	0.003889
State 2	0	-0.08553	0.08553
State 3	0	0	0

Results (graph)



RT+Tam



Tam

Results (5-year transition probability)

RT+Tam	State 1	State 2	State 3
State 1	0.8913	0.0466	0.0621
State 2	0	0.2151	0.7849
State 3	0	0	0

Tam	State 1	State 2	State 3
State 1	0.8221	0.1295	0.0484
State 2	0	0.6527	0.3473
State 3	0	0	0

Conclusion

- Combined treatment of **tamoxifen** and **radiation** is more effective than single treatment of **tamoxifen** in preventing the recurrence of breast cancer. However, for patients who had relapse of breast cancer, single treatment of **tamoxifen** proves to be more effective than combined treatment with respect to the survival probability.

- Transition probabilities between different stages during 2 years, 4 years, 5 years and 10 years are also calculated for predicting purpose.

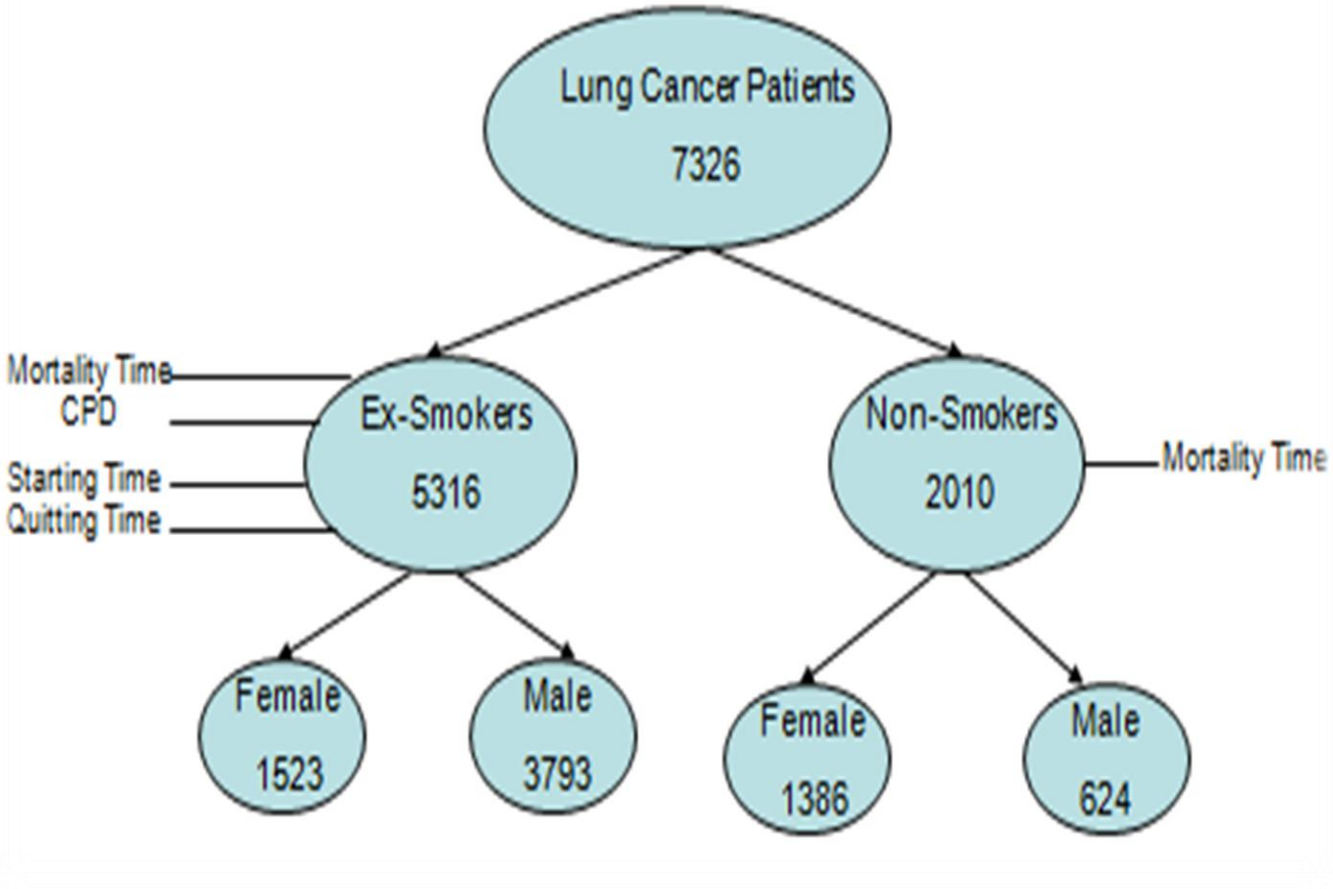
Objective 4

- The objective of the subject study is to conduct parametric analysis to address the basic probabilistic behavior of mortality time of both female and male lung cancer patients of ex-smokers and non-smokers, respectively.

- Mean mortality times are compared between non-smokers and ex-smokers, female non-smokers and male non-smokers, and female ex-smokers and male non-smokers. Meanwhile, important entities related to lung cancer mortality time such as cigarettes per day (CPD), and duration of smoking (DUR) are compared between female and male ex-smoker lung cancer patients.

- Finally, we developed a model to predict the mortality time of ex-smokers with a high degree of accuracy.

Data and Variables



Parametric Analysis

More than 40 different classical distributions are fitted to the data and three goodness-of-fit tests including Kolmogorov-Smirnov, Anderson-Darling, and Chi-Square are conducted for the mortality time of lung cancer patients for female ex-smokers, male ex-smokers, all ex-smokers, female non-smokers, male non-smokers, all non-smokers, respectively

Results

	Johnson SB	Beta	Three - parameter Weibulll
Female ex-smokers	NA	73.995 (8.9577)	74.007 (8.9365)
Male ex-smokers	NA	74.543 (8.1875)	74.542 (8.2119)
Ex-smokers	NA	74.384 (8.4155)	74.387 (8.428)
Female non-smokers	NA	76.117 (10.213)	76.148 (10.165)
Male non-smokers	NA	76.011 (9.6368)	76.015 (9.6551)
Non-smokers	NA	76.085 (10.041)	76.103 (10.022)

Mean and Standard Deviation of Fitted Distributions

Results

	Johnson SB	Beta	Three - parameter Weibulll
Female ex-smokers	(59.9, 87.622) (55.419, 90.07)	(58.586, 88.016) (55.364, 90.327)	(58.456, 87.916) (55.371, 90.168)
Male ex-smokers	(60.527, 87.493) (57.827, 89.55)	(60.557, 87.52) (57.849, 89.591)	(60.304, 87.386) (57.519, 89.482)
Ex-smokers	(59.9, 87.622) (57.061, 89.701)	(59.98, 87.659) (57.057, 89.85)	(59.751, 87.541) (56.871, 89.68)
Female non-smokers	(58.145, 91.777) (54.682, 93.933)	(58.304, 91.886) (54.794, 94.142)	(58.277, 91.742) (54.582, 94.205)
Male non-smokers	(58.888, 90.419) (55.014, 92.541)	(58.87, 90.391) (55.045, 92.434)	(58.671, 90.298) (54.727, 92.425)
Non-smokers	(58.367, 91.373) (54.761, 93.541)	(58.461, 91.428) (54.811, 93.643)	(58.354, 91.302) (54.567, 93.657)

90% and 95% Confidence Interval

Nonparametric Comparison

To test if there is significant difference of mortality time with respect to gender and smoking status, and also difference of entities like cigarettes per day and duration of smoking with respect to gender.(for ex-smokers only since they are all zeros for non-smokers). Wilcoxon two - sample test is performed to test the difference of the means.

(1) Mortality time between ex-smokers and non-smokers

(2) Ex-Smokers mortality time between female and male

(3) Non-Smokers Mortality time between female and male

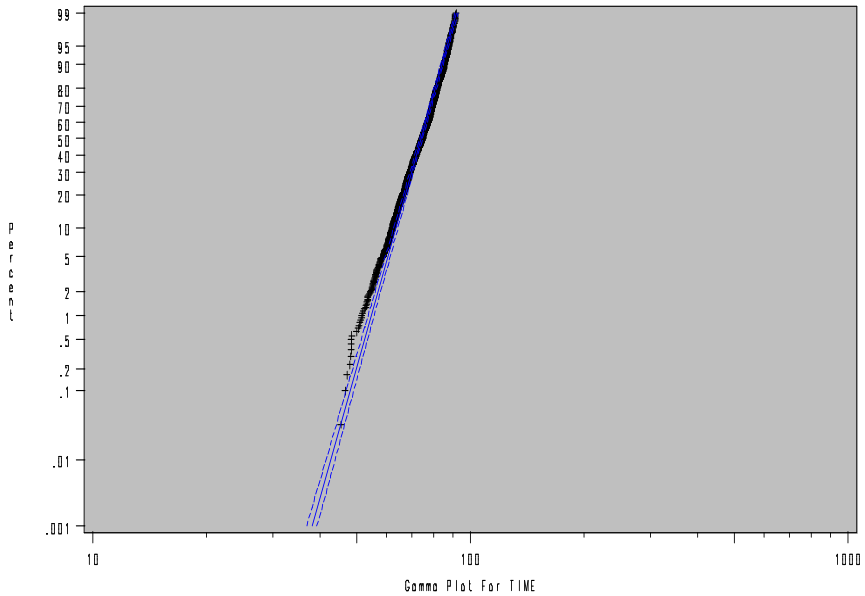
(4) Ex-smokers CPD between female and male

(5) Ex-smokers DUR between female and male

Results

H_o	$\bar{t}_{m(ex)} \geq \bar{t}_{m(non)}$	$\bar{t}_{m(\,female-ex)} = \bar{t}_{m(male-ex)}$	$\bar{t}_{m(\,female-non)} = \bar{t}_{m(male-non)}$	$\bar{CPD}_{male} \leq \bar{CPD}_{female}$	$\bar{DUR}_{male} \leq \bar{DUR}_{female}$
P-value	0.0018	0.1180	0.8106	<0.0001	0.0001
Conclusion	$\bar{t}_{m(ex)} < \bar{t}_{m(non)}$	$\bar{t}_{m(\,female-ex)} = \bar{t}_{m(male-ex)}$	$\bar{t}_{m(\,female-non)} = \bar{t}_{m(male-non)}$	$\bar{CPD}_{male} > \bar{CPD}_{female}$	$\bar{DUR}_{male} > \bar{DUR}_{female}$

AFT Model Results

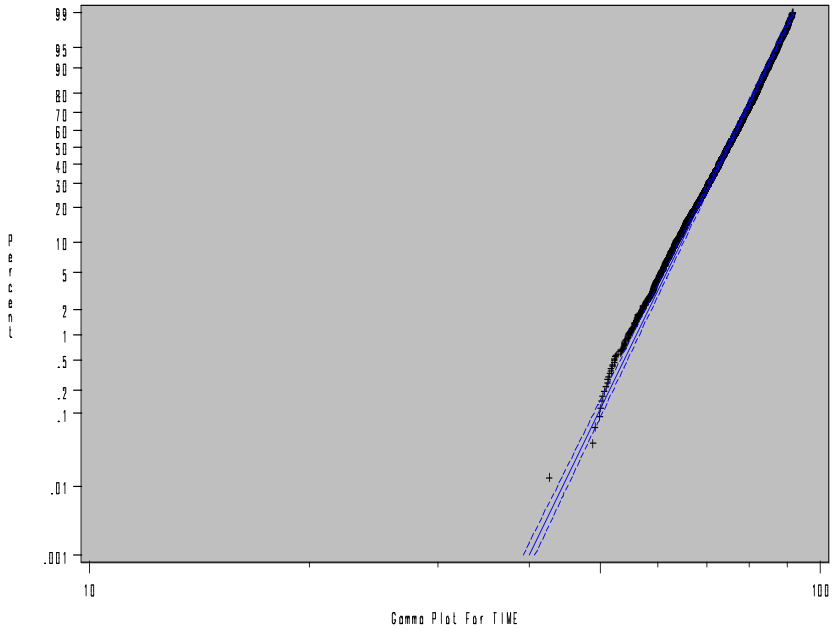


Percentage Plot for Female Ex-Smokers

mean: 0.008175027
variance: 0.1108183
mean: 0.1378148
variance: 7.911363

Percentage Plot of Male Ex-Smokers

mean: 0.007539421
variance: 0.1054347
mean: 0.1439845
variance: 7.651358



Conclusions

Probabilistic behavior of lung cancer patients mortality time is investigated. And comparison of key entities in lung cancer such as mortality time, CPD, and duration of smoking are conducted between different race groups and between different smoking status groups. Also, generalized gamma AFT model is found to be the bet model in predicting mortality time of ex-smokers lung cancer patients.

Objective 5

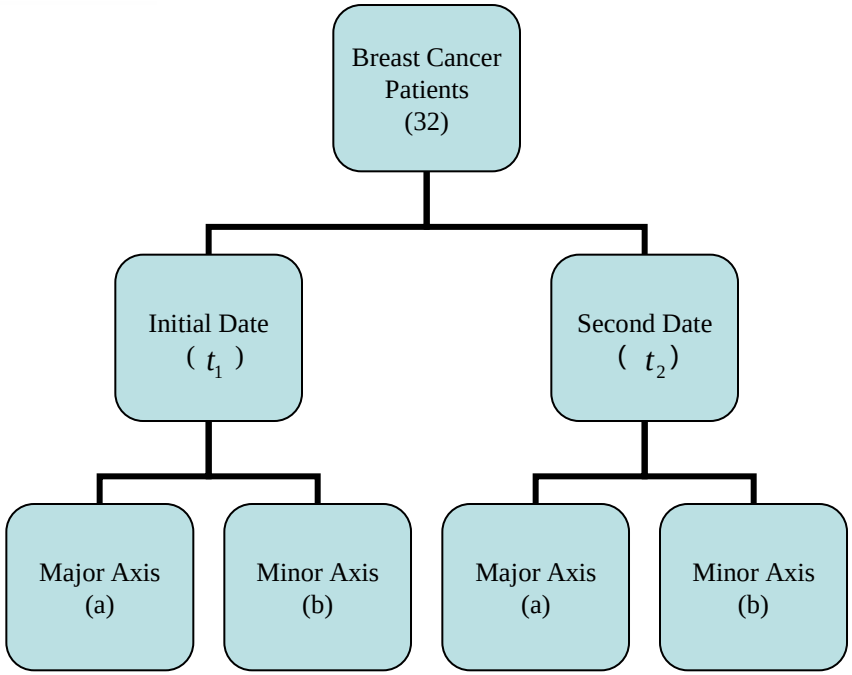
- The objective of the subject study is to illustrate the sensitivity of the probabilistic behavior of breast cancer tumor growth (doubling time) with respect to different volume and growth model assumptions.

Definition of Doubling Time and Assumptions

- Doubling Time: The time it takes for a tumor to double in volume. It depends on two assumptions: how the volume is calculated the way tumor grows.

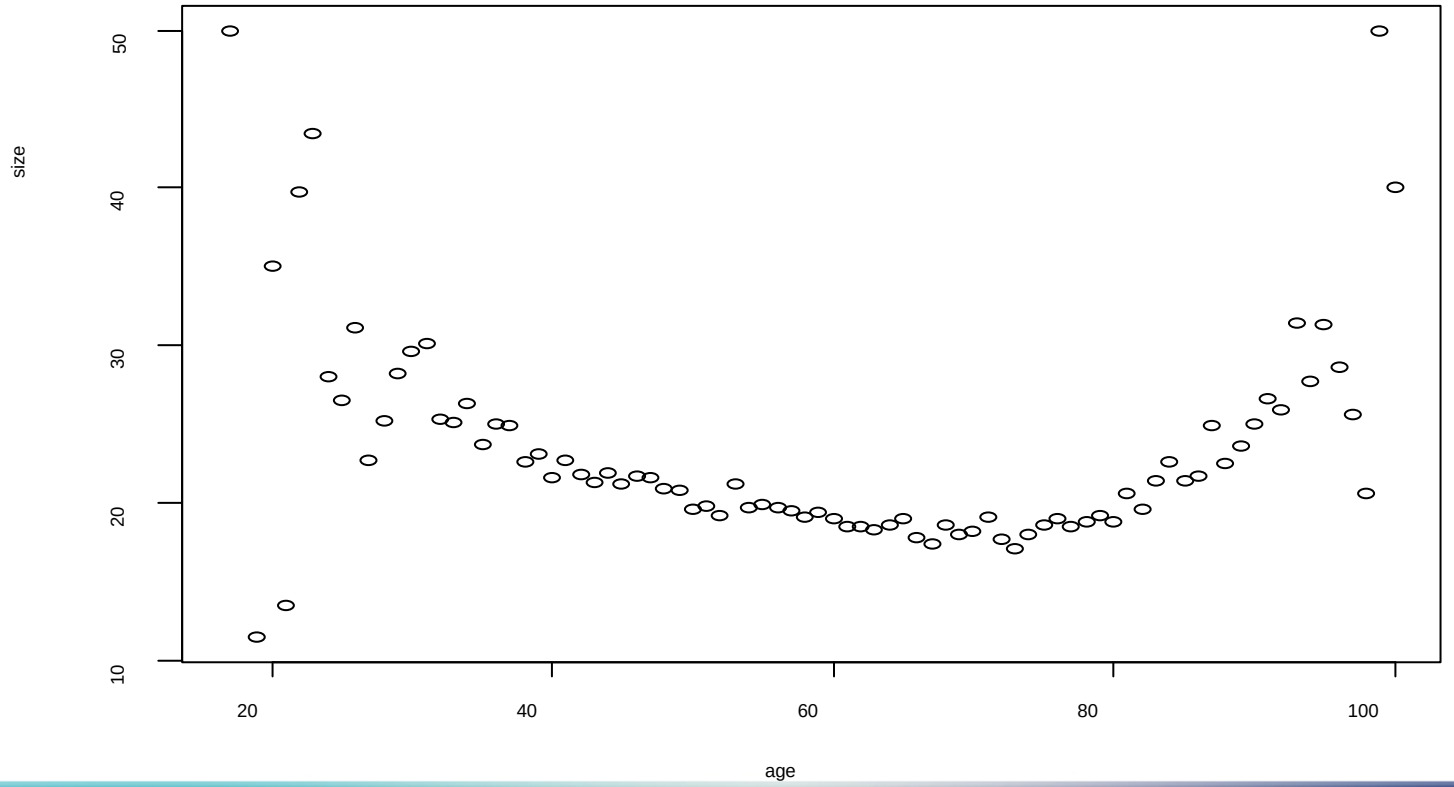
- Tumor Growth Assumption
- 1. Exponential Growth
- 2. Linear Growth
- 3. Quadratic Growth

- Tumor Volume Assumption
- 1. $V = \frac{4}{3} \cdot \pi \cdot r^3 \qquad r = a/2$
- 2. $V = \frac{4}{3} \cdot \pi \cdot r^3 \qquad r = \frac{2a+b}{12}$
- 3. $V = \frac{4}{3}\pi \cdot (\frac{a}{2})^2 \cdot (\frac{b}{2})$
- 4. $V = \frac{4}{3}\pi \cdot \frac{1}{2}a \cdot \frac{1}{2}b \cdot \frac{1}{2} \cdot (\frac{1}{2}a + \frac{1}{2}b)$

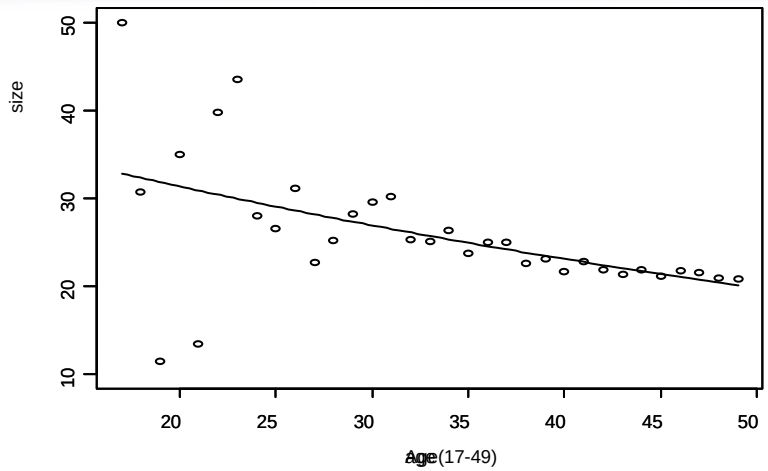


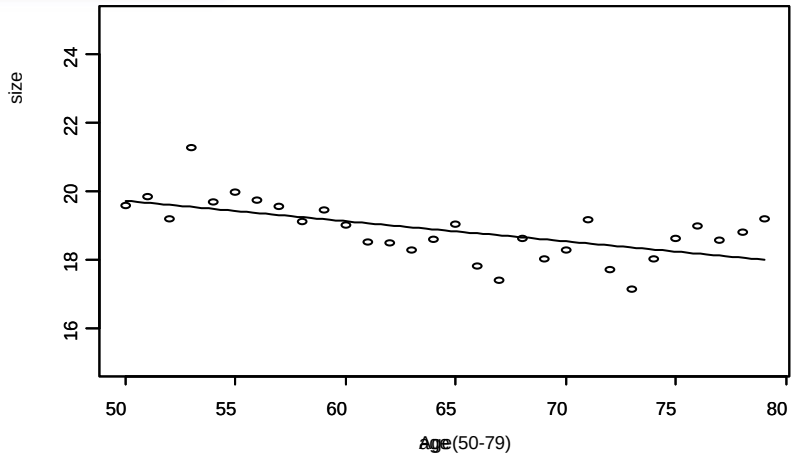
Tumor Growth Assumption

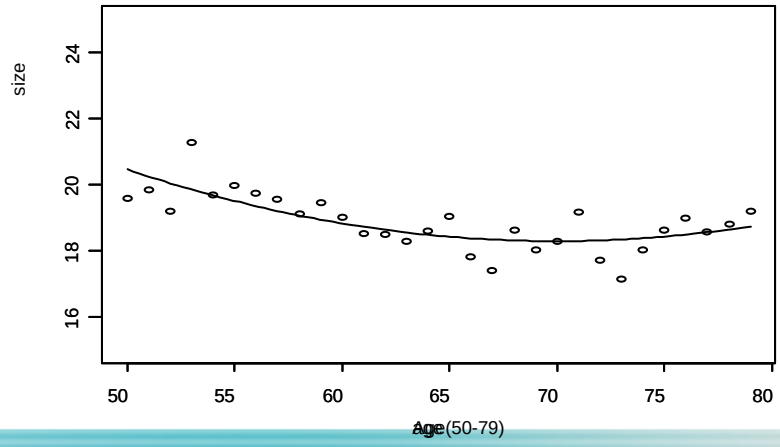
- Plot the average tumor size of breast cancer patients in each age as follows (age 17-100). Partition the age into three groups: 17-49, 50-79, 80-100, exponential decay function was used to fit the curve from 17-49, linear and quadratic from 50-79, and exponential growth from 80-100.

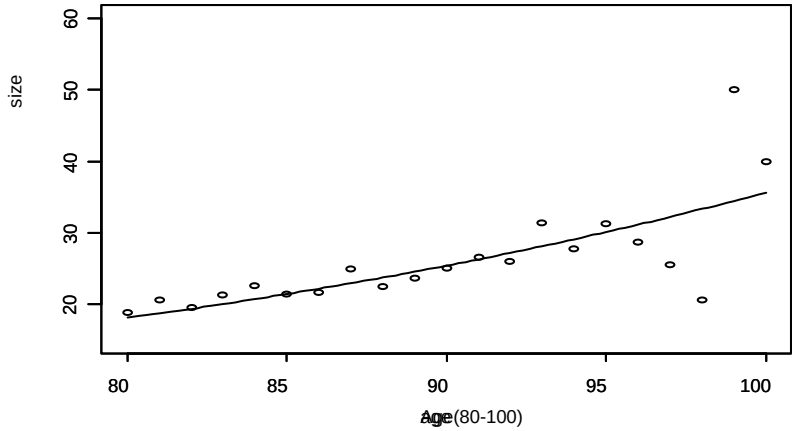


Tumor Growth Plots









Results (Growth Assumption)

For fixed volume assumption: spherical volume. Using three different growth assumptions (*exponential, linear, quadratic*). The distributions and related properties are as follows:

Growth	Distribution	Mean	S.D	95% lower limit	95% upper limit
exponential	Lognormal(3-parameter)	628.61	2459.9	66.894	3824.6
exponential	Lognormal	419.36	551.35	35.631	1809.0
linear	Fatigue Life	466.69	726.68	11.84	2580.4
linear	Lognormal	573.99	1741.8	9.0575	3563.2
quadratic	Johnson SB	NA	NA	65.101	994.63
quadratic	Lognormal	336.95	345.11	44.756	1238.0

Rank of lognormal distribution using goodness-of-fit tests under spherical, exponential assumptions, 15, 12,11.

Results (Volume Assumption)

- For fixed growth assumption: exponential growth while using three different volume assumptions (spherical, averaged spherical, oblate spherical, averaged oblate spherical)

Growth	Distribution	Mean	S.D	95% lower limit	95% upper limit
Spherical	Lognormal (3-parameter)	628.61	2459.9	66.894	3824.6
Spherical	Lognormal	419.36	551.35	35.631	1809.0
Ave. Spherical	Lognormal (3-parameter)	778.0	2590.5	102.05	4542.0
Ave. Spherical	Lognormal	588.58	793.94	47.653	2578.4
Oblate Spherical	Frechet (3-parameter)	NA	NA	110.75	9073.2
Oblate Spherical	Lognormal	685.92	983.2	49.467	3113.7
Ave. Oblate Spherical	Fatigue Life (3-parameter	563.05	744.44	96.523	2728.1
Ave. Oblate Spherical	Lognormal	543.93	699.01	48.221	2313.48

Conclusions

- Different combination of volume assumption and growth model could result in different tumor doubling time, and thus the probability distribution of doubling time. Necessary work should be done to determine which volume and growth assumption is the best to describe breast cancer tumor before simply assuming doubling time assumes Lognormal distribution.


Thank you
www.wallcoo.com