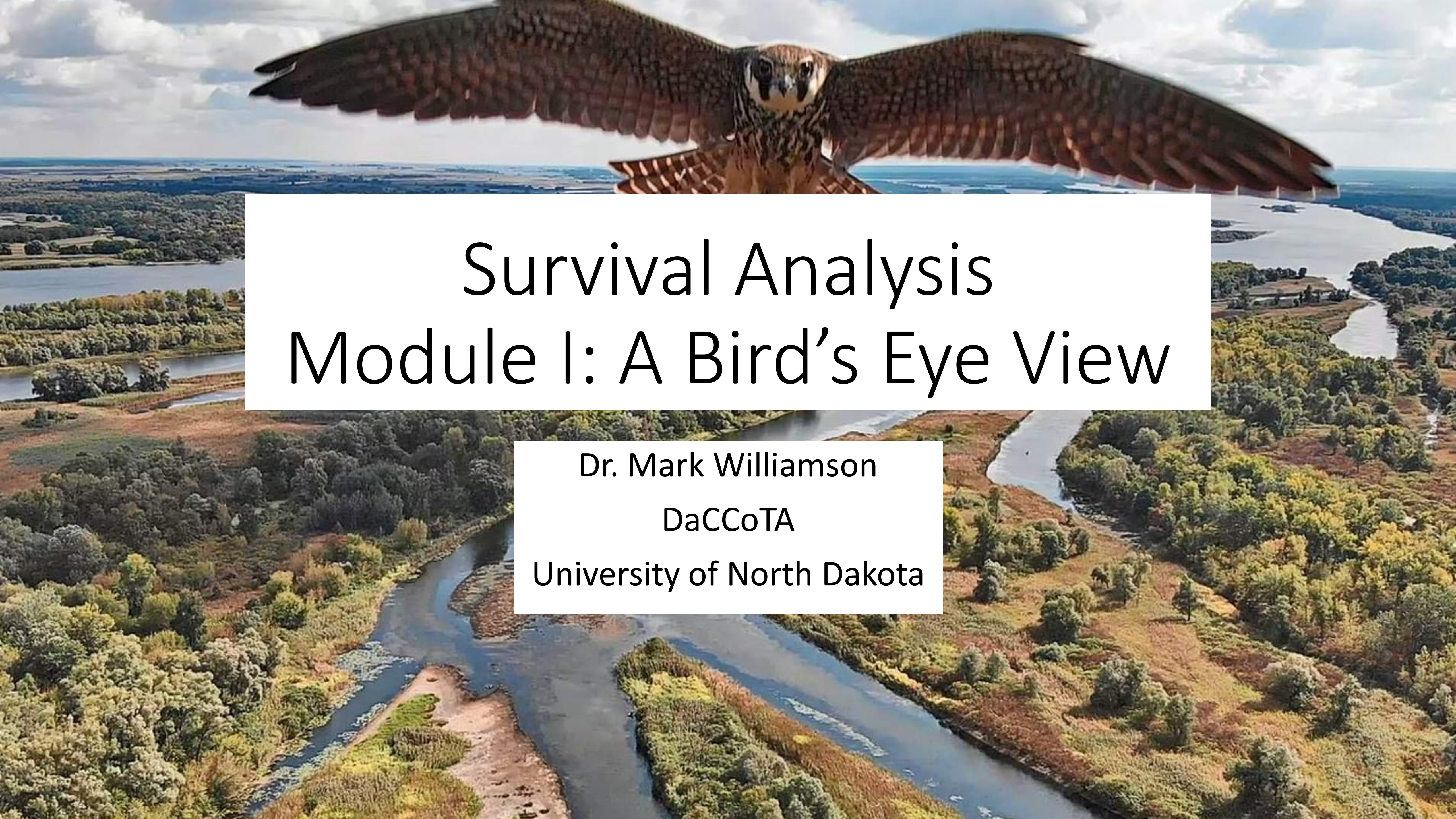


Survival Analysis
Module I: A Bird's Eye View

Dr. Mark Williamson

DaCCoTA

University of North Dakota

DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

UND UNIVERSITY OF NORTH DAKOTA

Introduction

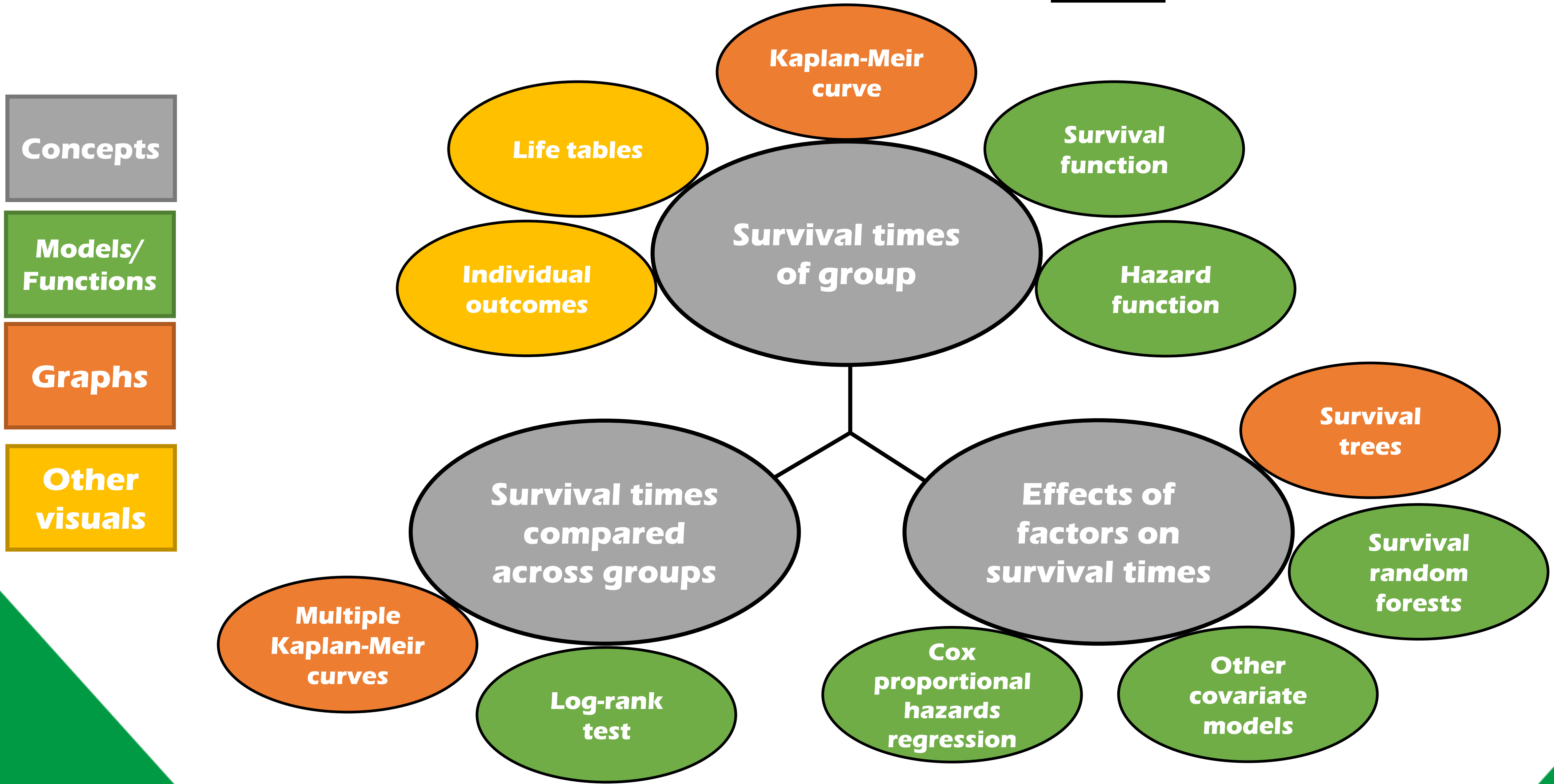
What is Survival Analysis?

- Fundamentally time-to-event data [1,2]
  - Cohort is measured over time for event
  - Event may be death/disease/failure/bankruptcy/etc.
  - Important in fields such as epidemiology, engineering, economics, etc.
- Survival analysis is used in three typical ways:
  - Survival time of a group
  - Survival time across groups
  - Survival time based on factors
- Typical approaches are visualization through survival curves and modeling using methods such as the log-rank test or Cox regression
- Important terms include **event**, **time**, **censoring**, **survival function**, and **hazard function** [1]
- Basic rules of survival [3]
  - "No one starts out dead" -> $S(t)=1$ @ $t=0$
  - "Everyone dies eventually" -> $S(t)=0$ @ $t=\infty$
  - "Once you are dead, you stay dead" -> $S(t)$ never increases

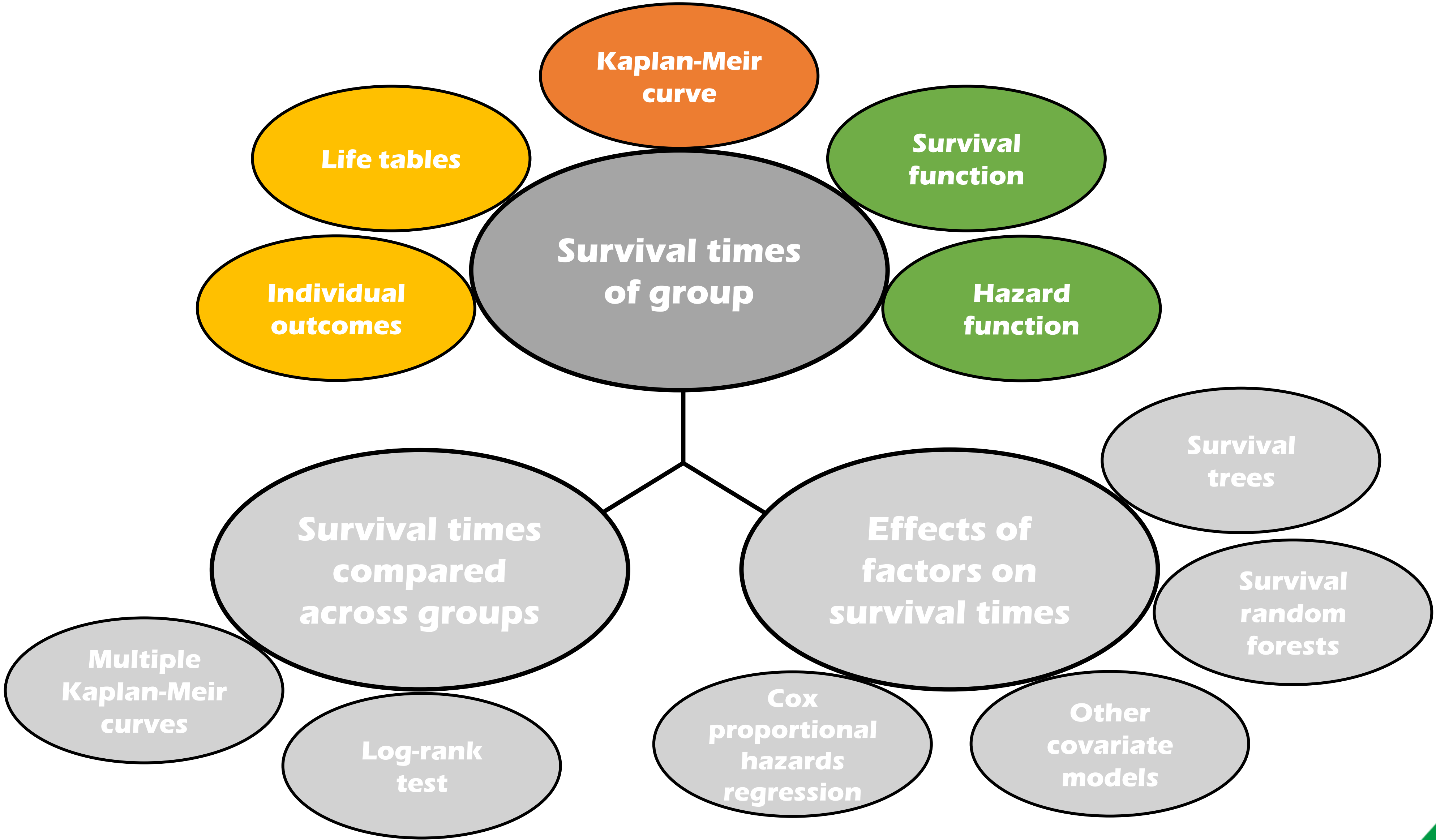
DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

UND
UNIVERSITY OF NORTH DAKOTA

Landscape

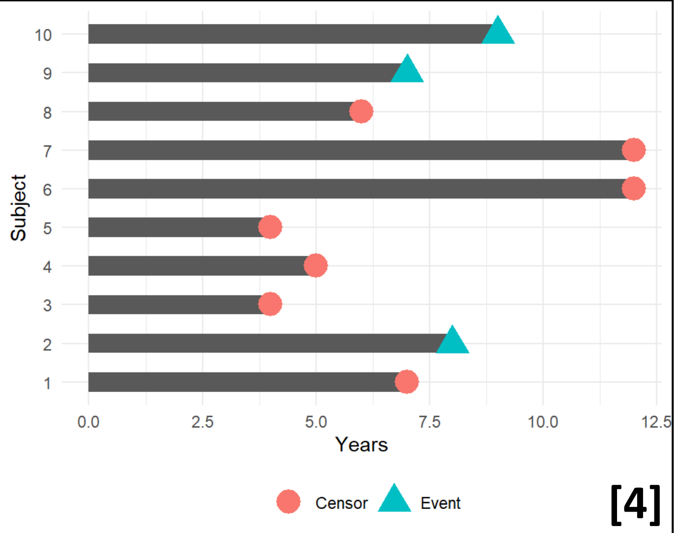


Structure and Uses



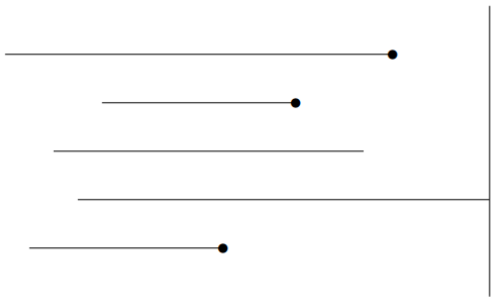
Structure and Uses 2

- **Survival times of group**
  - **Individual outcomes**
  - **Life tables**
  - **Kaplan-Meir curve**
  - **Survival function**
  - **Hazard function**



[4]

Example



T_i	C_i	Y_i	δ_i
80	100	80	1
40	80	40	1
74+	74	74	0
85+	85	85	0
40	95	40	1

Termination of study

[2]

	A	B	C	D	E	F	G	H
2								
3	Yrs in Trial	Died		t	d	n	1-d/n	S(t)
4	3	1		0		18		1
5	5	1		2	1	18	0.944444	0.944444
6	8	1		3	1	17	0.941176	0.888889
7	10	0		5	3	16	0.8125	0.722222
8	5	1		8	3	12	0.75	0.541667
9	5	0		9	0	9	1	0.541667
10	8	1		10	0	8	1	0.541667
11	12	1		11	2	6	0.666667	0.361111
12	15	0		12	2	4	0.5	0.180556
13	14	0						
14	2	1						
15	11	1						
16	10	0						
17	9	0						
18	12	1						
19	5	1						
20	8	1						
21	11	1						

[5]

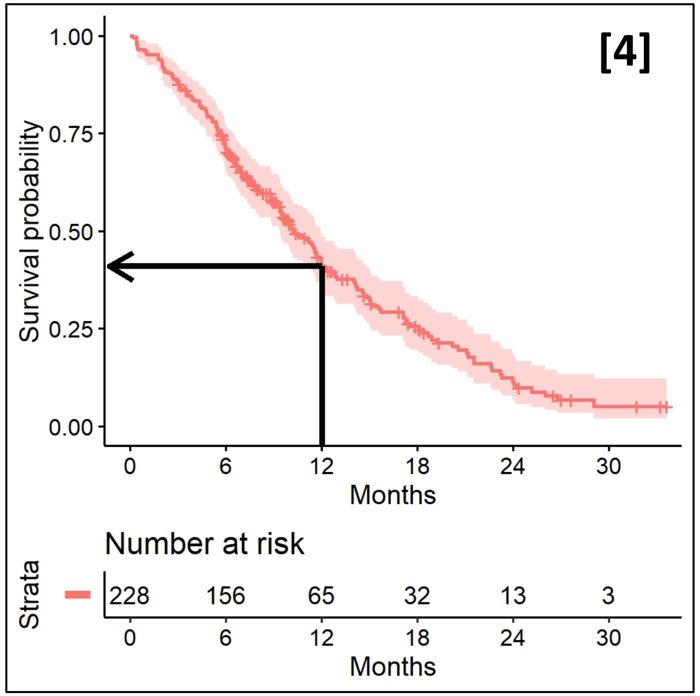
H	I	J	K	L	M	N	O	P	Q	R	S
x	nx	bx	mx	lx	lx mx	lx mx x			coeff.	decription	equation
0	30	0	0	1	0	0			x	number of days	
1	30	0	0	1	0	0			nx	Number of living individuals	
2	30	18	0.6	1	0	0			bx	number of births	
3	29	425	14.65517	0.966667	0.6	1.8			mx	number of births/number of survivers	bx/nx
4	29	340	11.72414	0.966667	14.16667	56.66667			lx	survivership	nx/N (total number of ind)
5	29	0	0	0.966667	11.33333	56.66667			dx	deaths	
6	27	325	12.03704	0.9	0	0			lxmx		
7	26	493	18.96154	0.866667	10.83333	75.83333					
8	24	217	9.041667	0.8	16.43333	131.4667			Ro	net reproductive rate	SUM(lxmx)*
9	23	130	5.652174	0.766667	7.233333	65.1			G	generation time	SUM(lx*mx*x)/SUM(lx*mx)
10	15	126	8.4	0.5	4.333333	43.33333			r	intrinsice rate of increase	ln(Ro)/G
11	15	0	0	0.5	4.2	46.2					
12	15	0	0	0.5	0	0					
13	15	3	0.2	0.5	0	0					
14	8	3	0.375	0.266667	0.1	1.4					
15	8	1	0.125	0.266667	0.1	1.5					
16	7	0	0	0.233333	0.033333	0.533333					
17	0	0	0	0	0	0					
18	0	0	0	0	0	0					

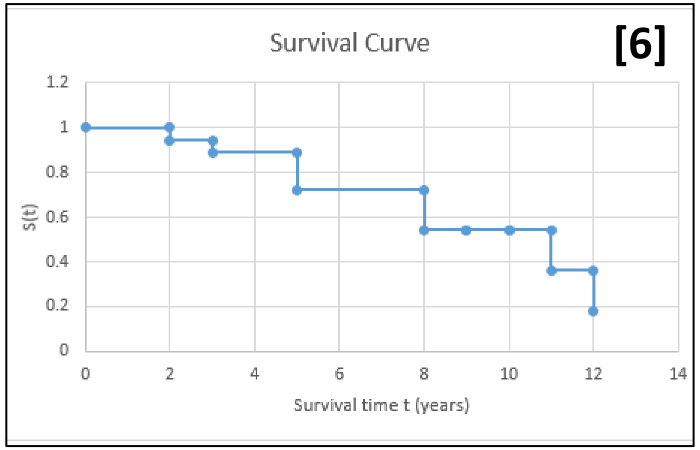
DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

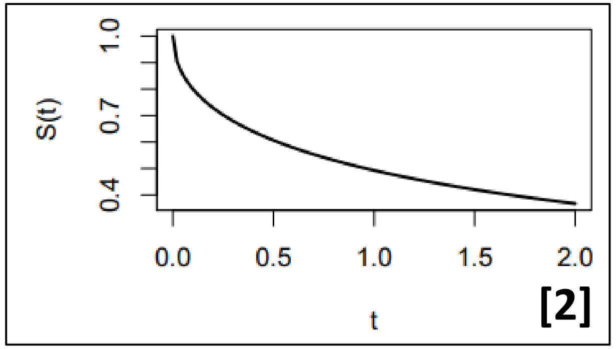
UNIVERSITY OF NORTH DAKOTA

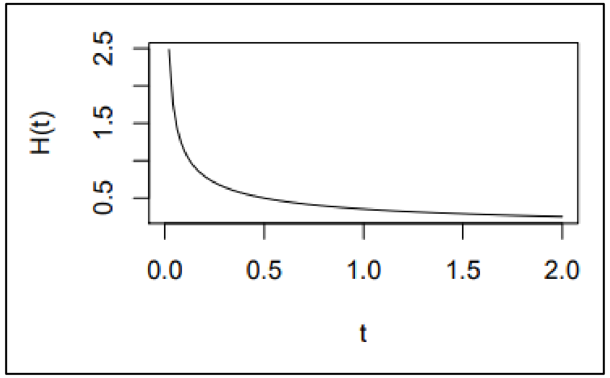
Structure and Uses 2

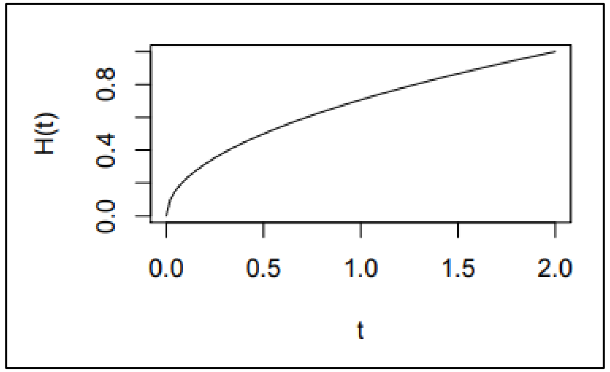
- Survival times of group
- Individual outcomes
- Life tables
- Kaplan-Meir curve
- Survival function
- Hazard function



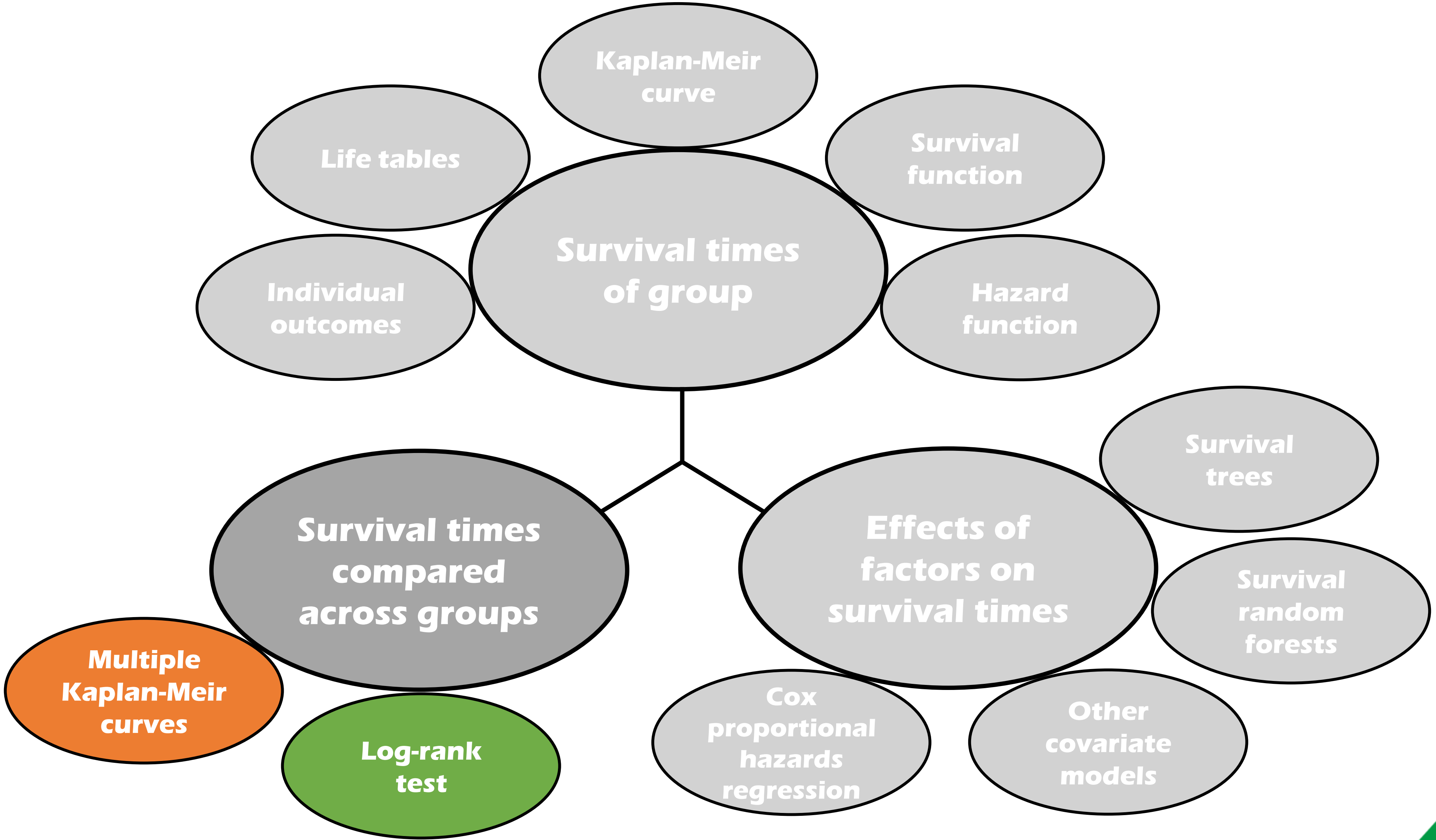








Structure and Uses 3

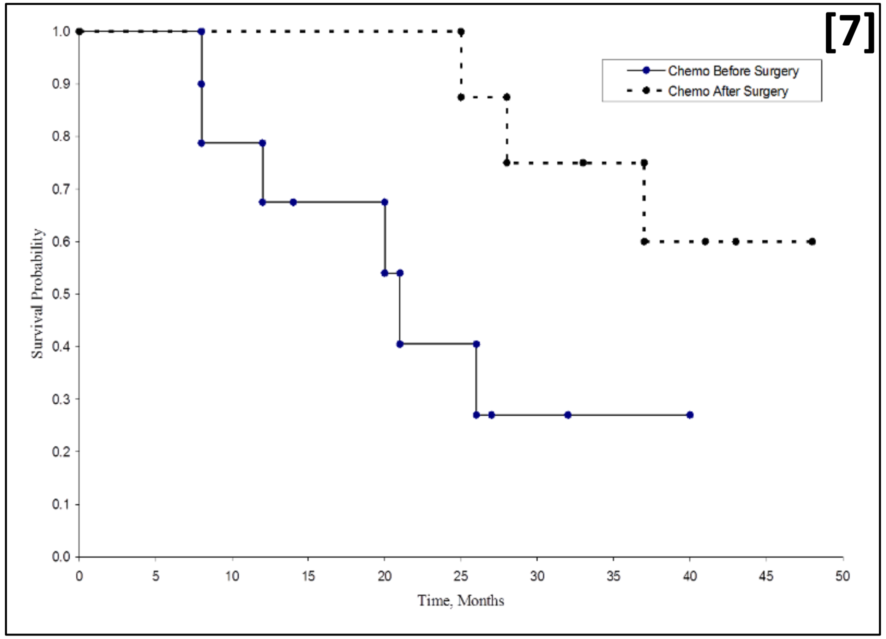


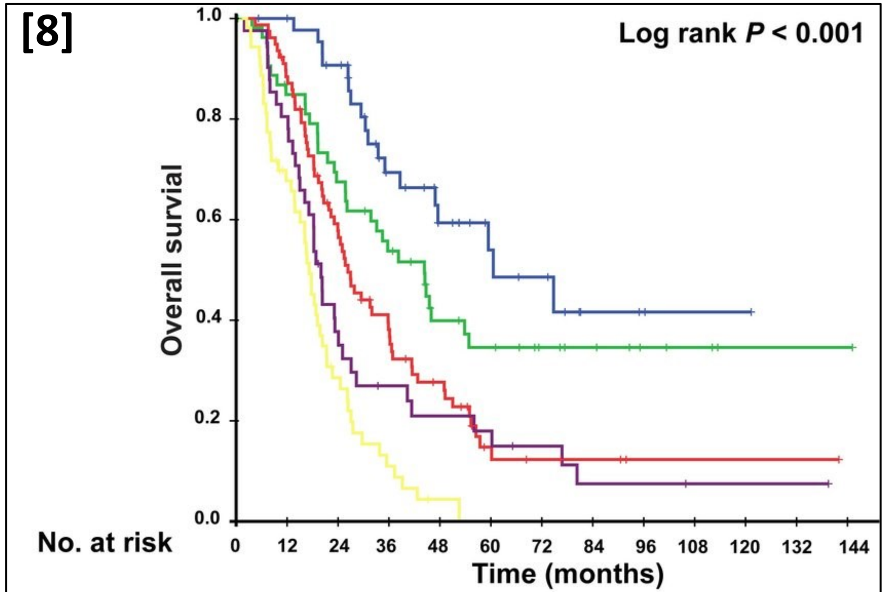
Structure and Uses 4

Survival times compared across groups

Multiple Kaplan-Meir curves

Log-rank test

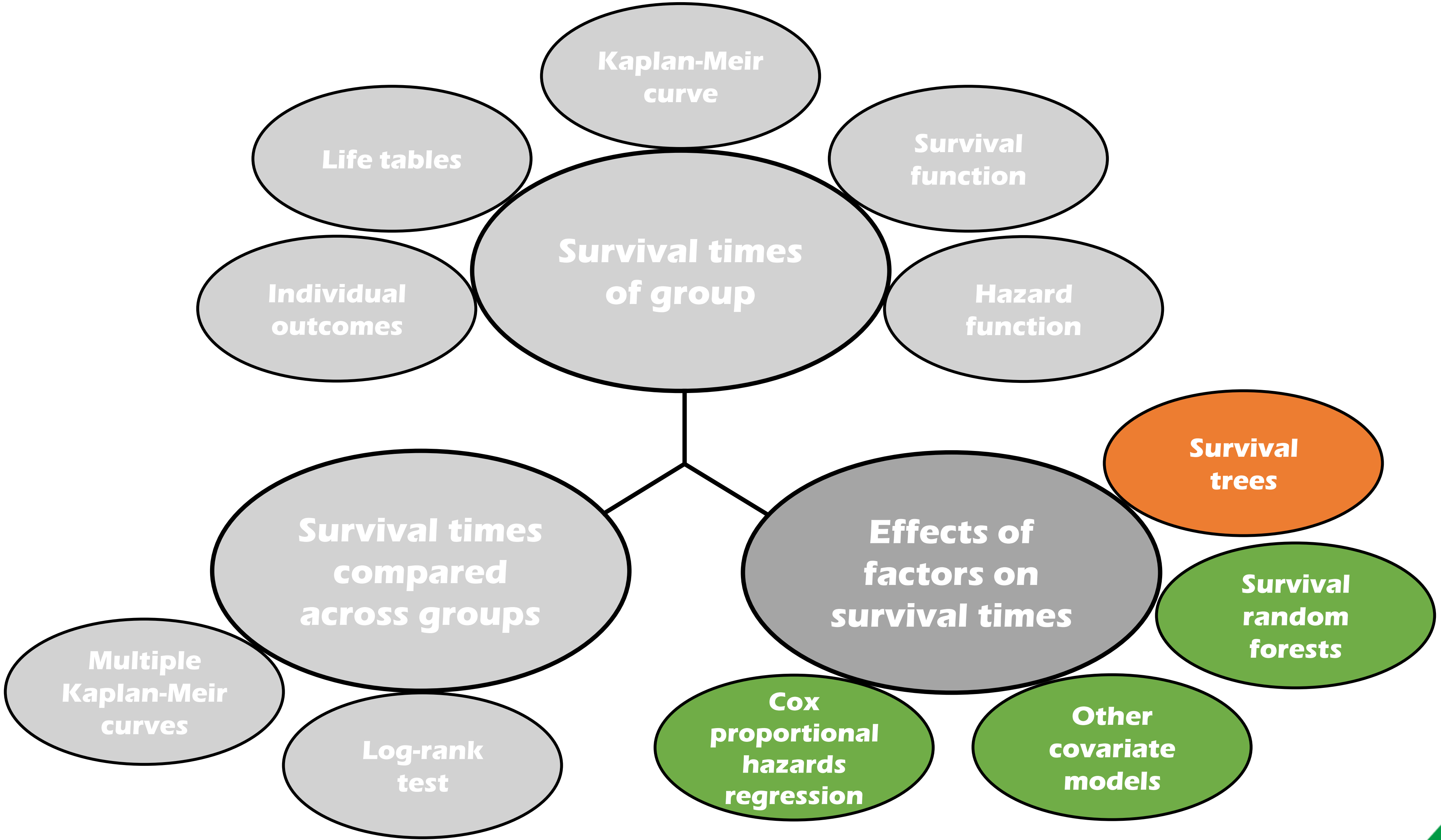




No. at risk	0	12	24	36	48	60	72	84	96	108	120	132	144
Group A	45	44	38	23	16	10	8	3	2	1	1	0	0
Group B	54	44	35	27	16	13	9	7	4	3	1	1	1
Group C	78	68	42	26	17	6	4	3	1	1	1	1	0
Group D	41	33	14	9	7	6	4	2	2	1	1	1	0
Group E	53	33	13	5	1	0	0	0	0	0	0	0	0

Log-rank statistic for two groups $= \dfrac{(O_2 - E_2)^2}{Var(O_2 - E_2)} \sim \chi_1^2$ [9]

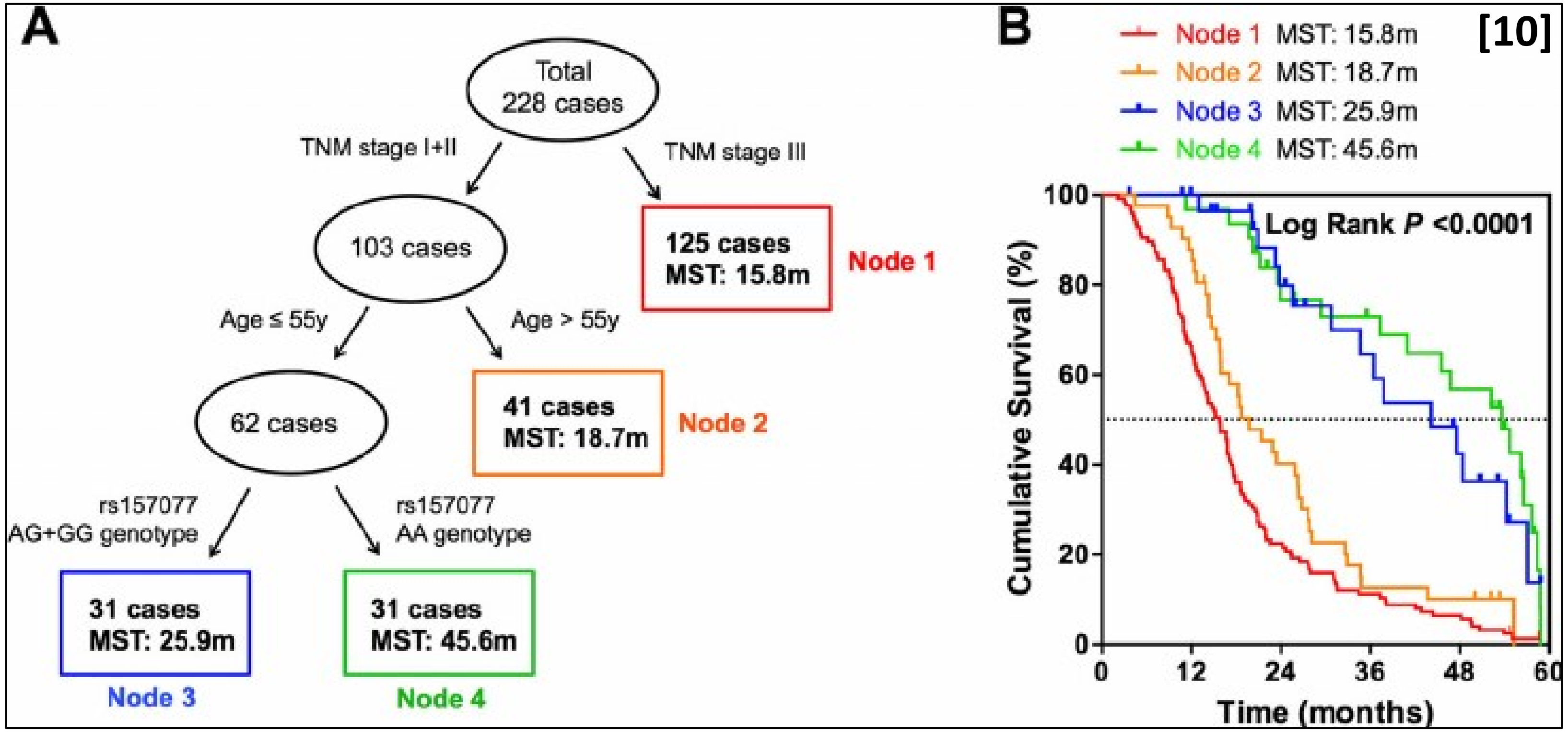
Structure and Uses 5



DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

Structure and Uses 6

- Effects of factors on survival times
- **Survival trees**
- Survival random forests
- Cox proportional hazards regression
- Other covariate models



[10]

UNIVERSITY OF NORTH DAKOTA

Structure and Uses 6

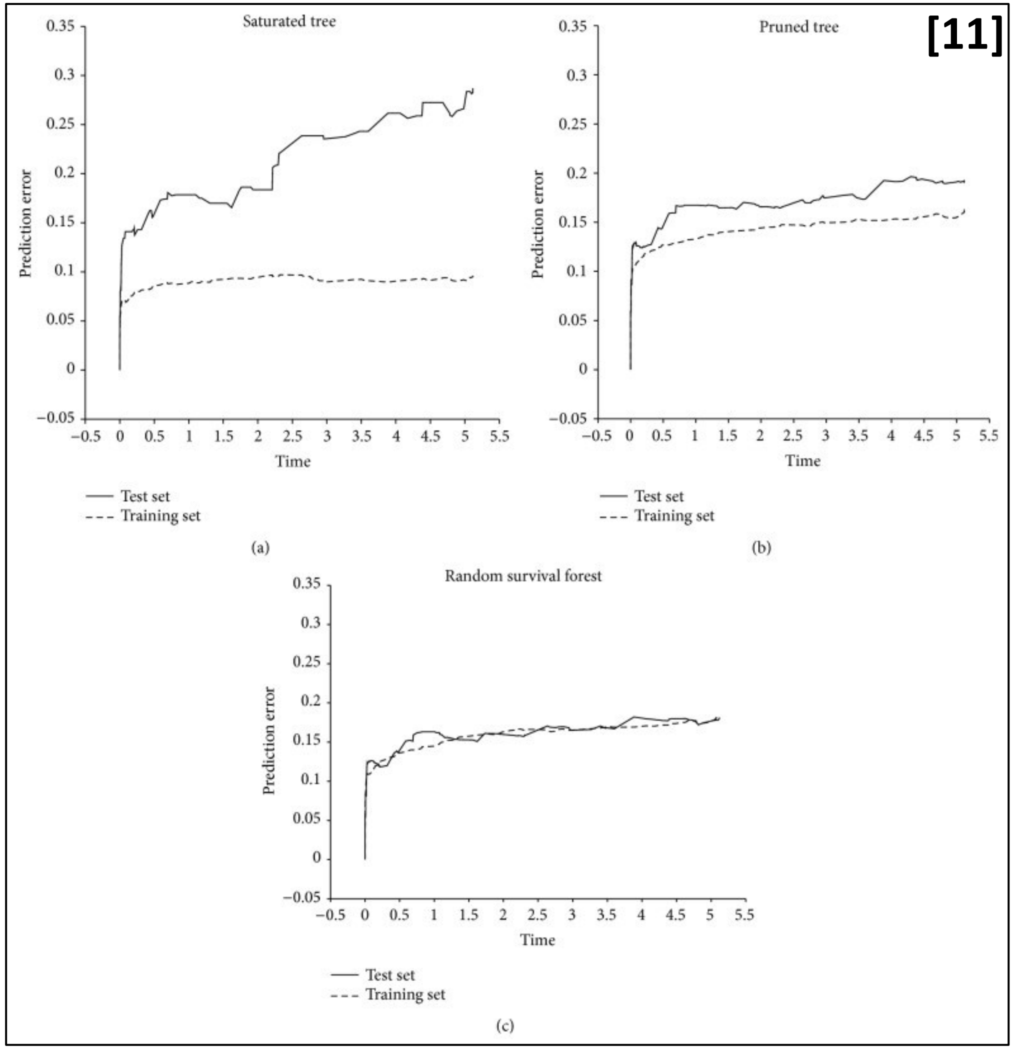
Effects of factors on survival times

Survival trees

Survival random forests

Cox proportional hazards regression

Other covariate models



DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

UNIVERSITY OF NORTH DAKOTA

Structure and Uses 6

Effects of factors on survival times

Survival trees

Survival random forests

Cox proportional hazards regression

Other covariate models

[12]

Risk Factor	Parameter Estimate	P-Value	Hazard Ratio (HR) (95% CI for HR)
Age, years	0.11691	0.0001	1.124 (1.111-1.138)
Male Sex	0.40359	0.0002	1.497 (1.215-1.845)
Systolic Blood Pressure	0.01645	0.0001	1.017 (1.012-1.021)
Current Smoker	0.76798	0.0001	2.155 (1.758-2.643)
Total Serum Cholesterol	-0.00209	0.0963	0.998 (0.995-2.643)
Diabetes	-0.02366	0.1585	0.816 (0.615-1.083)

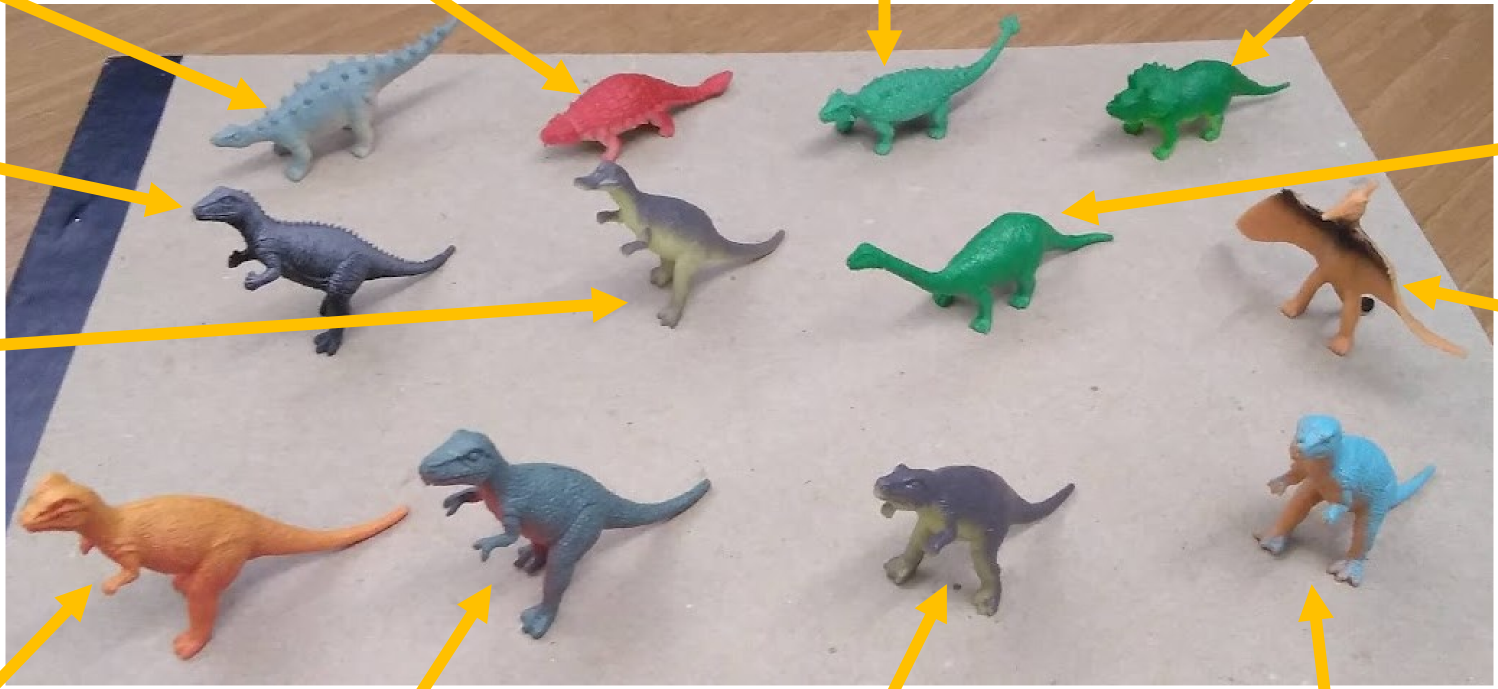
There are four ways to introduce covariates in parametric survival models

1. Parametric families, where the parameters of a distribution, such as λ and p in a Weibull, depend on covariates
2. Accelerated life, where the log of survival time follows a linear model
3. Proportional hazards, where the log of the hazard function follows a linear model
4. Proportional odds, where the logit of the survival function follows a linear model

[13]

Examples

Generating Data



DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

UND
UNIVERSITY OF NORTH DAKOTA

Examples

Generating Data 2

Shaking for 60 seconds. Falling off means counted as Dead (1).

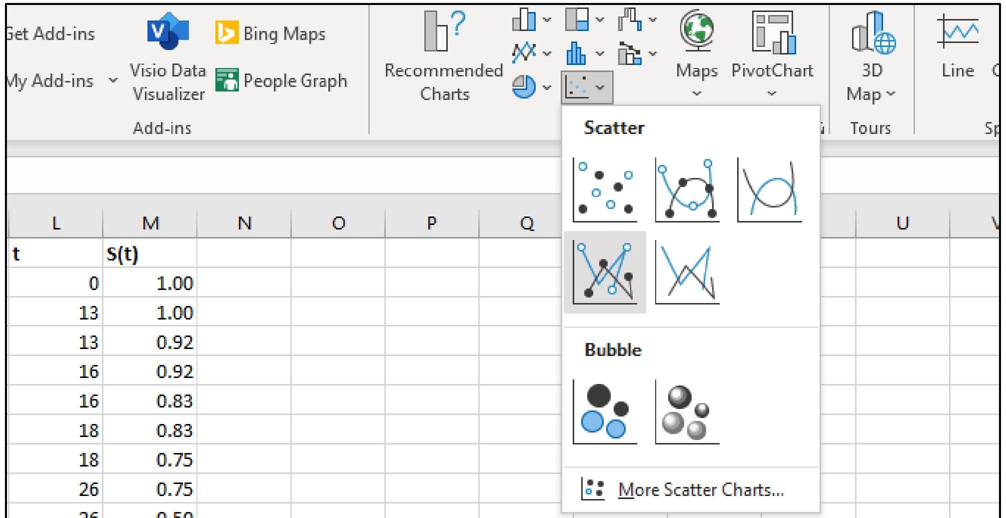
Time	Dead	N
0		12
13	1	12
16	1	11
18	1	10
26	3	9
27	1	6
30	1	5
31	1	4
38	1	3
41	1	2
50	1	1
60	0	0

Examples

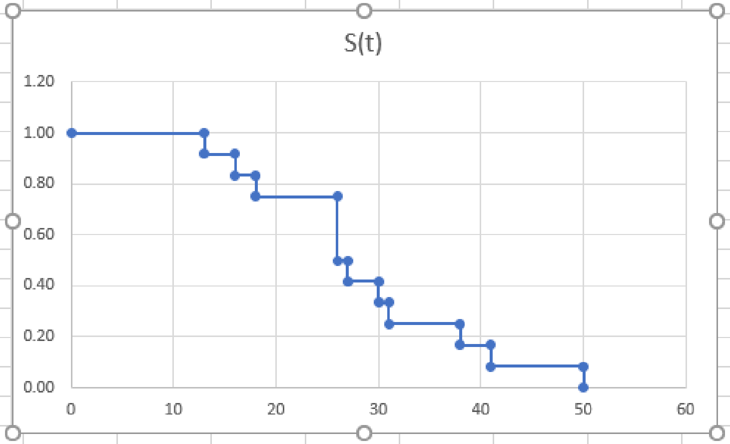
Excel

	A	B	C	D
1	type	id	time	status
2	Blue Anklo	1	26	1
3	Red Anklo	2	26	1
4	Green Anklo	3	31	1
5	Green Tricer	4	18	1
6	Black Iguan	5	38	1
7	Grey Hadr	6	13	1
8	Green Sauro	7	30	1
9	Orang Pter	8	26	1
10	Orange Carno	9	50	1
11	Blue Red Carno	10	27	1
12	Grey Carno	11	16	1
13	Light Blue Carno	12	41	1

F	G	H	I	J
Time	Dead	N	1-d/n	S(t)
0		12		1
13	1	12	0.916667	0.916667
16	1	11	0.909091	0.833333
18	1	10	0.9	0.75
26	3	9	0.666667	0.5
27	1	6	0.833333	0.416667
30	1	5	0.8	0.333333
31	1	4	0.75	0.25
38	1	3	0.666667	0.166667
41	1	2	0.5	0.083333
50	1	1	0	0



L	M
t	S(t)
0	1.00
13	1.00
13	0.92
16	0.92
16	0.83
18	0.83
18	0.75
26	0.75
26	0.50
27	0.50
27	0.42
30	0.42
30	0.33
31	0.33
31	0.25
38	0.25
38	0.17
41	0.17
41	0.08
50	0.08
50	0.00



Examples

R

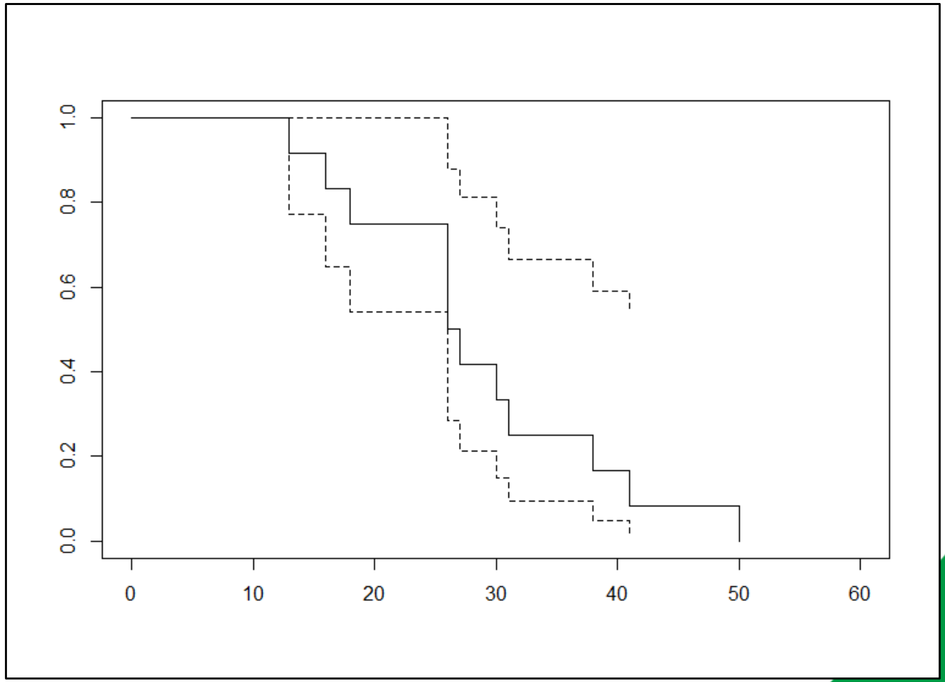
```
#Load and View Data
library("survival")
setwd("C:/Users/Mark.Williamson.2/Desktop/Williamson Data/R/R_data")
Dino <-read.csv("Dino_Survival.csv")
print(Dino)
```

```
#Survival Curve
dinofit <-survfit(Surv(time, status==1)~1,
                  data=Dino)
summary(dinofit)
plot(dinofit, xlim=c(0,60))
```

```
#Smooth Survival Curve Attempt 1
dinofit2 <-survreg(Surv(time, status==1)~1,
                   data=Dino, dist='exponential')
summary(dinofit2) #3.350

T<-seq(0,60,0.1)

plot(dinofit, xlim=c(0,60))
points(T, 1-pexp(T, exp(-3.350)),
       cex=0.3, pch=16, col="red")
```

```
#Smooth Survival Curve Attempt 3
dinofit3 <-survreg(Surv(time, status==1)~1,
                   data=Dino, dist='weibull', scale=0)
summary(dinofit3) #3.464; 0.331

plot(dinofit, xlim=c(0,60))
points(T, 1-pweibull(T,1/0.331, exp(3.464)),
       cex=0.3, pch=16, col="purple")
```

	type	id	time	status
1	Blue_Anklo	1	26	1
2	Red_Anklo	2	26	1
3	Green_Anklo	3	31	1
4	Green_Tricer	4	18	1
5	Black_Iguan	5	38	1
6	Grey_Hadr	6	13	1
7	Green_Sauro	7	30	1
8	Orang_Pter	8	26	1
9	Orange_Carno	9	50	1
10	Blue_Red_Carno	10	27	1
11	Grey_Carno	11	16	1
12	Light_Blue_Carno	12	41	1



DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

UNIVERSITY OF NORTH DAKOTA

Examples

R

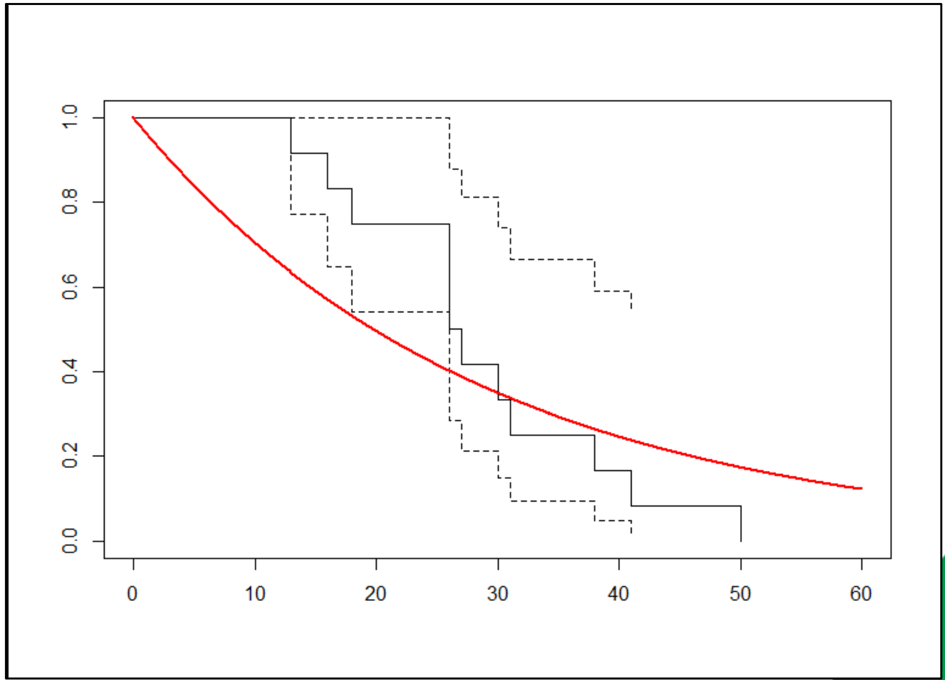
```
#Load and View Data
library("survival")
setwd("C:/Users/Mark.Williamson.2/Desktop/Williamson Data/R/R_data")
Dino <-read.csv("Dino_Survival.csv")
print(Dino)
```

```
#Survival Curve
dinofit <-survfit(Surv(time, status==1)~1,
                  data=Dino)
summary(dinofit)
plot(dinofit, xlim=c(0,60))
```

```
#Smooth Survival Curve Attempt 1
dinofit2 <-survreg(Surv(time, status==1)~1,
                   data=Dino, dist='exponential')
summary(dinofit2) #3.350

T<-seq(0,60,0.1)

plot(dinofit, xlim=c(0,60))
points(T, 1-pexp(T, exp(-3.350)),
       cex=0.3, pch=16, col="red")
```

```
#Smooth Survival Curve Attempt 3
dinofit3 <-survreg(Surv(time, status==1)~1,
                   data=Dino, dist='weibull', scale=0)
summary(dinofit3) #3.464; 0.331

plot(dinofit, xlim=c(0,60))
points(T, 1-pweibull(T,1/0.331, exp(3.464)),
       cex=0.3, pch=16, col="purple")
```

	type	id	time	status
1	Blue_Anklo	1	26	1
2	Red_Anklo	2	26	1
3	Green_Anklo	3	31	1
4	Green_Tricer	4	18	1
5	Black_Iguan	5	38	1
6	Grey_Hadr	6	13	1
7	Green_Sauro	7	30	1
8	Orang_Pter	8	26	1
9	Orange_Carno	9	50	1
10	Blue_Red_Carno	10	27	1
11	Grey_Carno	11	16	1
12	Light_Blue_Carno	12	41	1



DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

UND UNIVERSITY OF NORTH DAKOTA

Examples

R

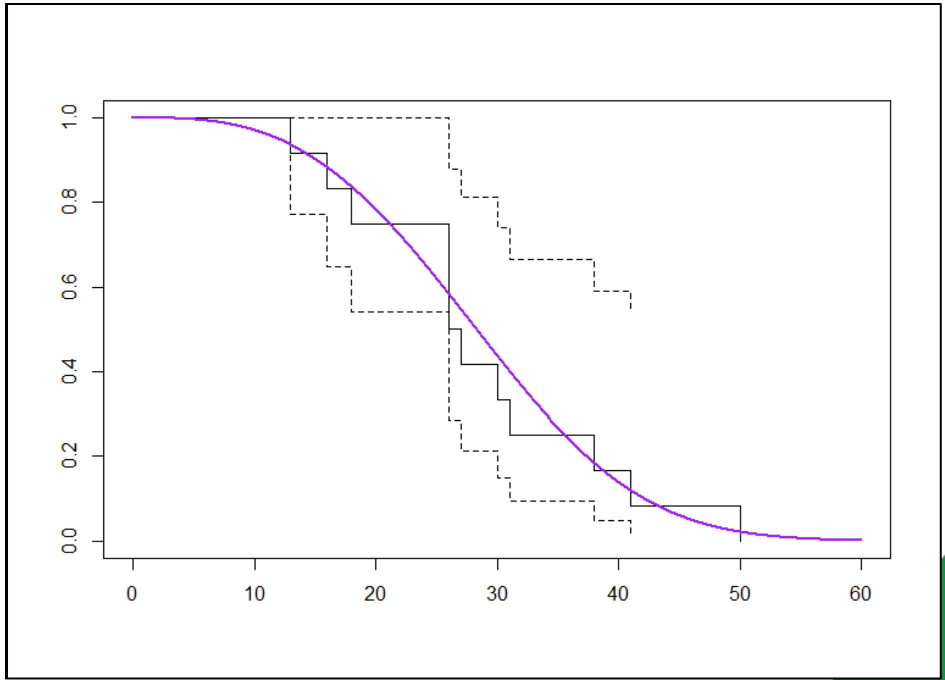
```
#Load and View Data
library("survival")
setwd("C:/Users/Mark.Williamson.2/Desktop/Williamson Data/R/R_data")
Dino <-read.csv("Dino_Survival.csv")
print(Dino)
```

	type	id	time	status
1	Blue_Anklo	1	26	1
2	Red_Anklo	2	26	1
3	Green_Anklo	3	31	1
4	Green_Tricer	4	18	1
5	Black_Iguan	5	38	1
6	Grey_Hadr	6	13	1
7	Green_Sauro	7	30	1
8	Orang_Pter	8	26	1
9	Orange_Carno	9	50	1
10	Blue_Red_Carno	10	27	1
11	Grey_Carno	11	16	1
12	Light_Blue_Carno	12	41	1

```
#Survival Curve
dinofit <-survfit(Surv(time, status==1)~1,
                  data=Dino)
summary(dinofit)
plot(dinofit, xlim=c(0,60))
```

```
#Smooth Survival Curve Attempt 1
dinofit2 <-survreg(Surv(time, status==1)~1,
                   data=Dino, dist='exponential')
summary(dinofit2) #3.350

T<-seq(0,60,0.1)

plot(dinofit, xlim=c(0,60))
points(T, 1-pexp(T, exp(-3.350)),
       cex=0.3, pch=16, col="red")
```

```
#Smooth Survival Curve Attempt 3
dinofit3 <-survreg(Surv(time, status==1)~1,
                   data=Dino, dist='weibull', scale=0)
summary(dinofit3) #3.464; 0.331

plot(dinofit, xlim=c(0,60))
points(T, 1-pweibull(T,1/0.331, exp(3.464)),
       cex=0.3, pch=16, col="purple")
```



Examples

SAS

```
*Load and Format Data;
PROC IMPORT datafile='/home/markwilliamson20/my_courses/markwilliamson0/MW_Datasets_2021/Presentation and Module Datasets 2021/Dino_Survival.csv'
    dbms=csv out=dino replace; getnames=yes; guessingrows=10000;
DATA dino; set dino;
    if id >7 then do; diet="Carnivore"; end;
    else do; diet="Herbivore"; end;
PROC PRINT data=dino;
```

```
*Survival Function;
PROC LIFETEST data=dino plots=survival(atrisk cb);
    time time*status(0);
```

```
*Survival Function (two groups);
PROC LIFETEST data=dino plots=survival(atrisk cb);
    strata diet;
    time time*status(0);
```

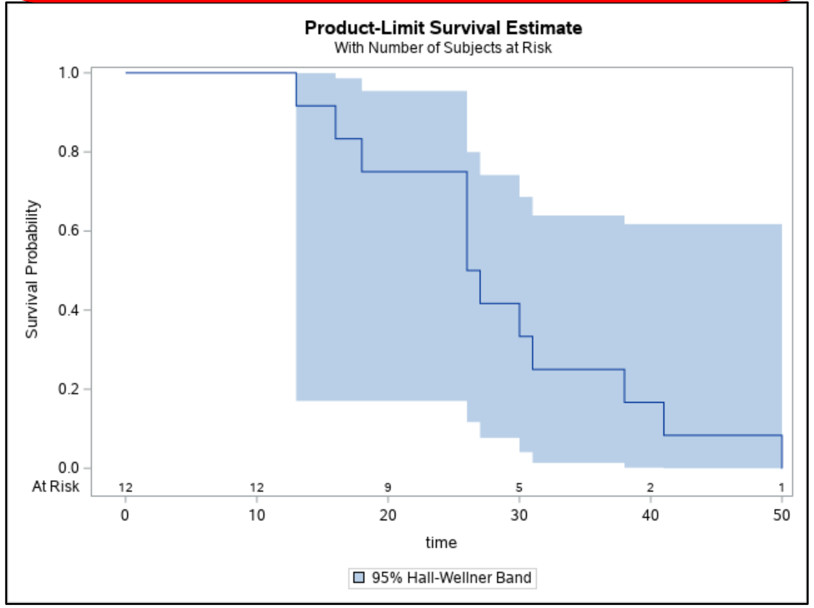
```
*Hazard Function;
PROC LIFETEST data=dino plots=hazard;
    time time*status(0);
```

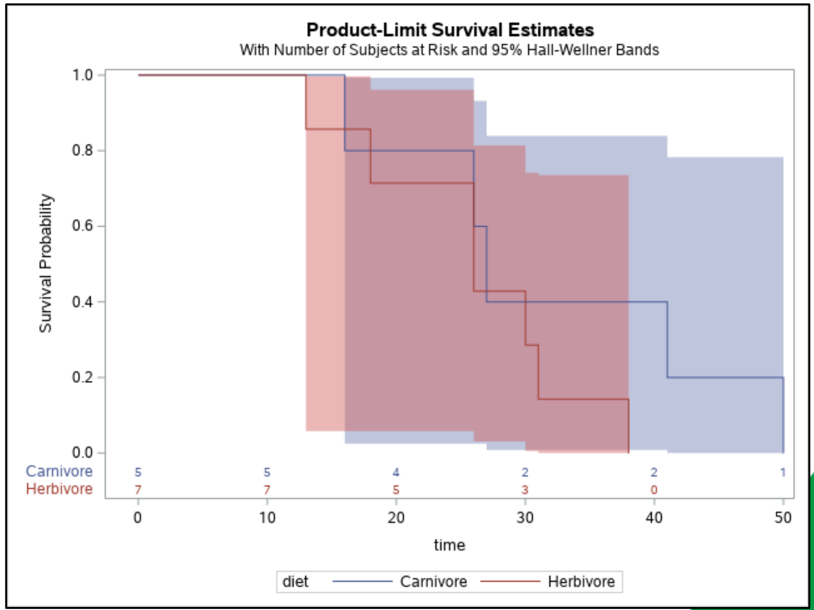
```
*Culmulative Hazard Function;
PROC LIFETEST data=dino nelson method=pl;
    time time*status(0);
    ods output ProductLimitEstimates=dino_ple;
PROC PRINT data=dino_ple;
PROC SGPLOT data =dino_ple;
    series x=time y=CumHaz;
```

Obs	type	id	time	status	diet
1	Orang_Pter	8	26	1	Carnivore
2	Orange_Carno	9	50	1	Carnivore
3	Blue_Red_Carno	10	27	1	Carnivore
4	Grey_Carno	11	16	1	Carnivore
5	Light_Blue_Carno	12	41	1	Carnivore
6	Blue_Anklo	1	26	1	Herbivore
7	Red_Anklo	2	26	1	Herbivore
8	Green_Anklo	3	31	1	Herbivore
9	Green_Tricer	4	18	1	Herbivore
10	Black_Iguan	5	38	1	Herbivore
11	Grey_Hadr	6	13	1	Herbivore
12	Green_Sauro	7	30	1	Herbivore

Test of Equality over Strata

Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	1.2420	1	0.2651
Wilcoxon	0.5226	1	0.4697
-2Log(LR)	0.1270	1	0.7216





Examples

SAS

```
    *Load and Format Data;
PROC IMPORT datafile='/home/markwilliamson20/my_courses/markwilliamson0/MW_Datasets_2021/Presentation and Module Datasets 2021/Dino_Survival.csv'
    dbms=csv out=dino replace; getnames=yes; guessingrows=10000;
DATA dino; set dino;
    if id >7 then do; diet="Carnivore"; end;
    else do; diet="Herbivore"; end;
PROC PRINT data=dino;
```

```
    *Survival Function;
PROC LIFETEST data=dino plots=survival(atrisk cb);
    time time*status(0);
```

```
    *Survival Function (two groups);
PROC LIFETEST data=dino plots=survival(atrisk cb);
    strata diet;
    time time*status(0);
```

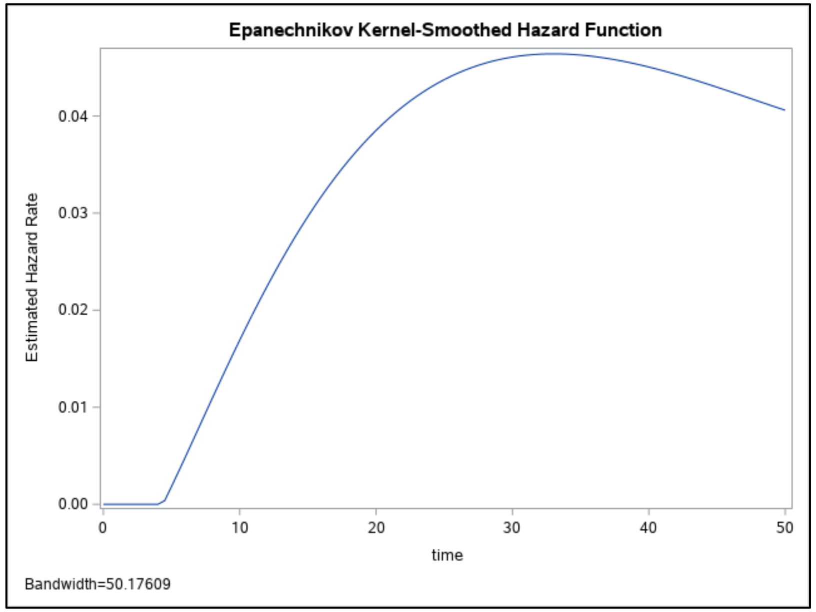
```
    *Hazard Function;
PROC LIFETEST data=dino plots=hazard;
    time time*status(0);
```

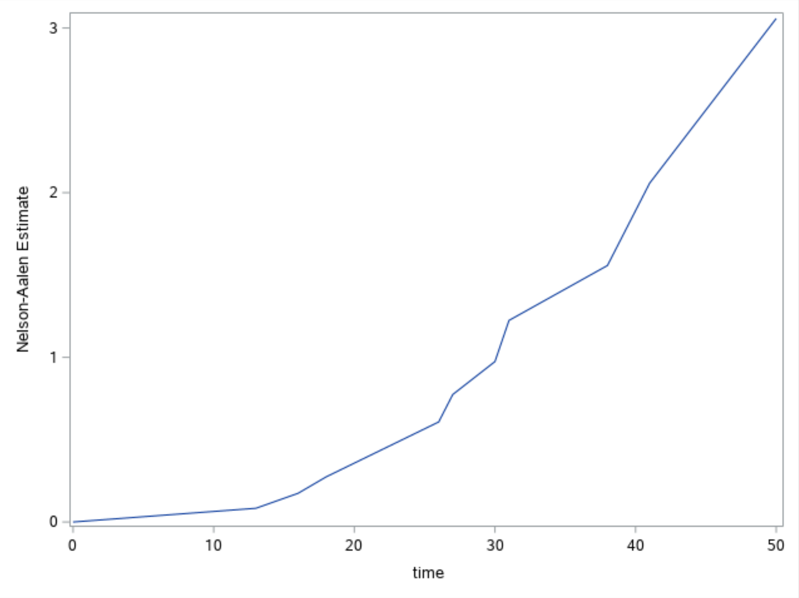
```
    *Culmulative Hazard Function;
PROC LIFETEST data=dino nelson method=pl;
    time time*status(0);
    ods output ProductLimitEstimates=dino_ple;
PROC PRINT data=dino_ple;
PROC SGPLOT data =dino_ple;
    series x=time y=CumHaz;
```

Obs	type	id	time	status	diet
1	Orang_Pter	8	26	1	Carnivore
2	Orange_Carno	9	50	1	Carnivore
3	Blue_Red_Carno	10	27	1	Carnivore
4	Grey_Carno	11	16	1	Carnivore
5	Light_Blue_Carno	12	41	1	Carnivore
6	Blue_Anklo	1	26	1	Herbivore
7	Red_Anklo	2	26	1	Herbivore
8	Green_Anklo	3	31	1	Herbivore
9	Green_Tricer	4	18	1	Herbivore
10	Black_Iguan	5	38	1	Herbivore
11	Grey_Hadr	6	13	1	Herbivore
12	Green_Sauro	7	30	1	Herbivore

Test of Equality over Strata

Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	1.2420	1	0.2651
Wilcoxon	0.5226	1	0.4697
-2Log(LR)	0.1270	1	0.7216





DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

UNIVERSITY OF NORTH DAKOTA

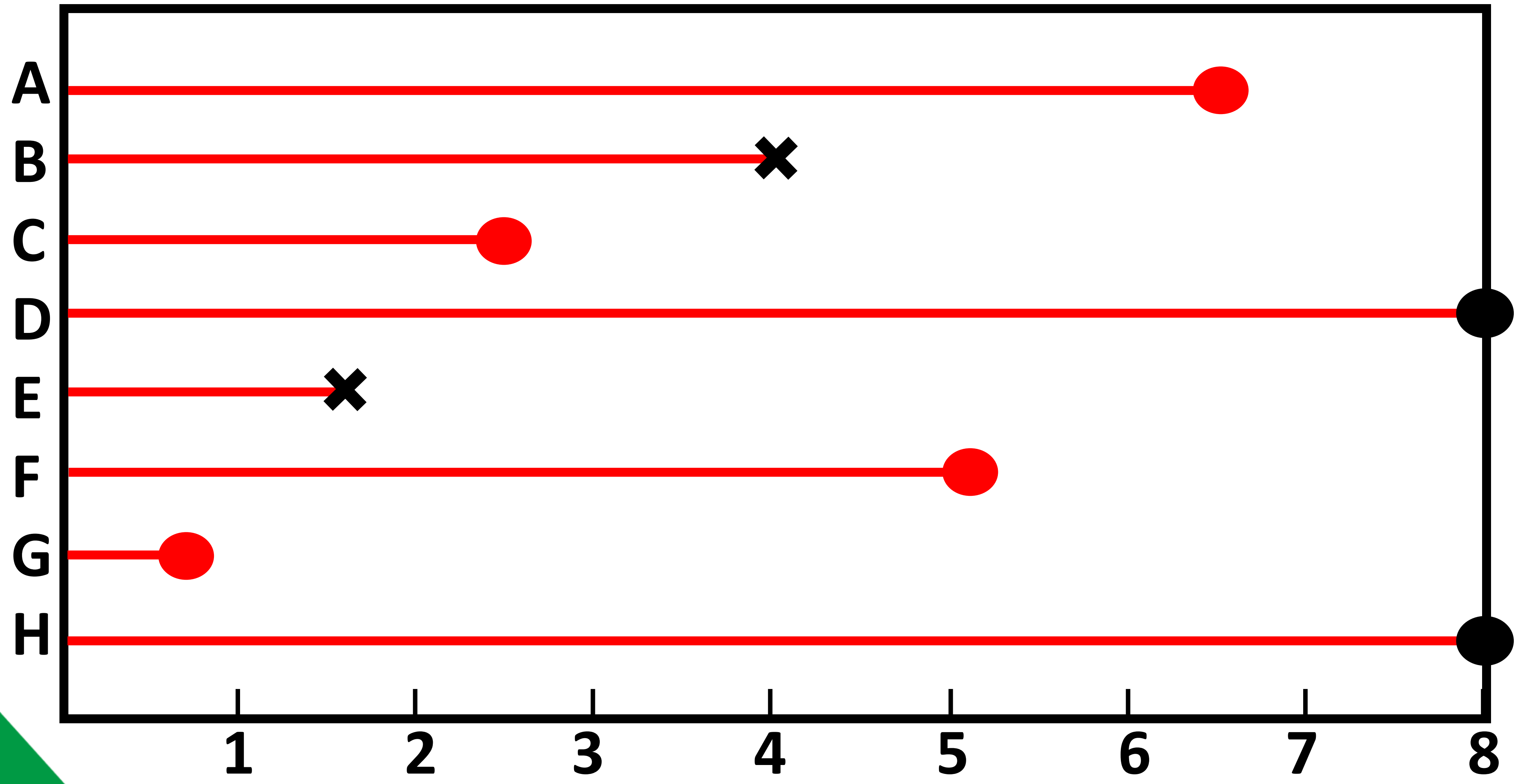
Assessment

qualtrics XM

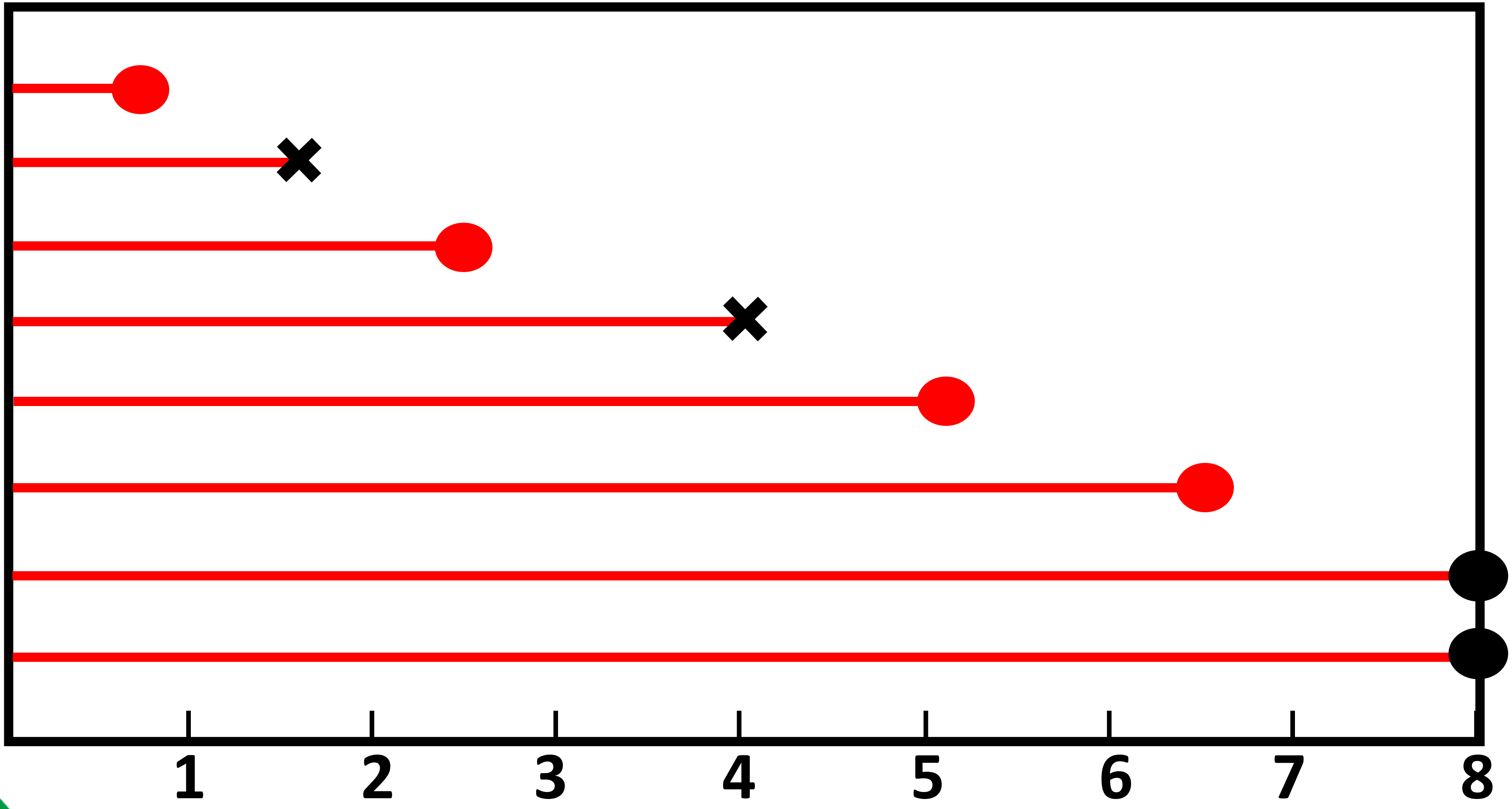
UNIVERSITY OF NORTH DAKOTA UND

https://und.qualtrics.com/jfe/form/SV_3Q4QXbdVLMGTva6

An Illustration



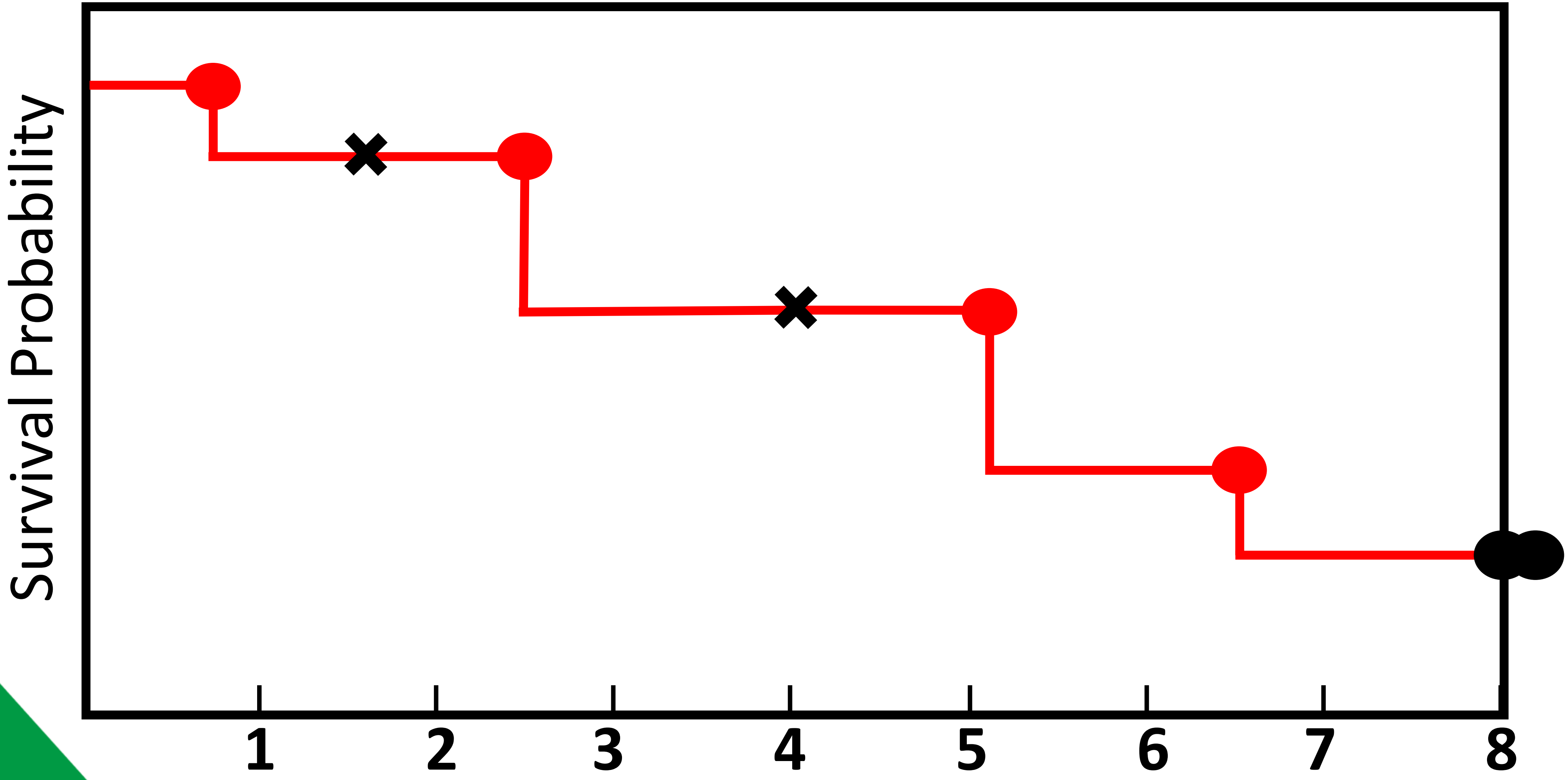
An Illustration



DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

UNIVERSITY OF NORTH DAKOTA

An Illustration



Summary and Conclusion

- Survival analysis is used for time-to-event data
- Major types are 1) survival time of group, 2) survival time across groups, and 3) survival time based on factors
- Survival curves and life tables are common visualization methods
- Log-rank tests and Cox regressions are common analysis methods
- Tune in next time for a more detailed look in Survival Analysis Module II: Leaves and Trees

DaCCoTA
DAKOTA CANCER COLLABORATIVE ON TRANSLATIONAL ACTIVITY

UND UNIVERSITY OF NORTH DAKOTA

References

[1] https://en.wikipedia.org/wiki/Survival_analysis

[2] http://www.stat.columbia.edu/~madigan/W2025/notes/survival.pdf

[3] https://www.real-statistics.com/survival-analysis/survival-analysis-basic-concepts/

[4] https://www.emilyzabor.com/tutorials/survival_analysis_in_r_tutorial.html

[5] https://www.real-statistics.com/survival-analysis/kaplan-meier-procedure/kaplan-meier-overview/

[6] https://www.real-statistics.com/survival-analysis/kaplan-meier-procedure/survival-curve/

[7] https://sphweb.bumc.bu.edu/otlt/mph-modules/bs/bs704_survival/BS704_Survival5.html

[8] https://www.researchgate.net/publication/340989649_Establishment_of_an_integrated_model_for_predicting_survival_and_guiding_treatment_in_local_recurrence_nasopharyngeal_carcinoma/figures?lo=1

[9] https://www.slideshare.net/zhe1/kaplan-meier-survival-curves-and-the-logrank-test

[10] https://www.researchgate.net/publication/274396208_Glutathione_S-transferase_O2_gene_rs157077_polymorphism_predicts_response_to_transarterial_chemoembolization_in_hepatocellular_carcinoma/figures?lo=1

[11] https://pubmed.ncbi.nlm.nih.gov/26858773/#&gid=article-figures&pid=figure-1-uid-0

[12] https://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/BS704_Survival/BS704_Survival6.html

[13] https://data.princeton.edu/pop509/pop509slides1.pdf

DaCCoTA
DAKOTA CANCER COLLABORATIVE
ON TRANSLATIONAL ACTIVITY

UNIVERSITY OF NORTH DAKOTA

Acknowledgements

- The DaCCoTA is supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award Number U54GM128729.
- For the labs that use the Biostatistics, Epidemiology, and Research Design Core in any way, including this Module, please acknowledge us for publications. ***"Research reported in this publication was supported by DaCCoTA (the National Institute of General Medical Sciences of the National Institutes of Health under Award Number U54GM128729)".***

DaCCoTA
DAKOTA CANCER COLLABORATIVE
ON TRANSLATIONAL ACTIVITY