General Linear Models

Survival Analysis

– O.Örsan Özener

Survival Analysis

- We are interested in “**time to event outcome variable”**
- Different type of outcome variable compared to other models
- Time to event variable ~ time until a certain event
  - Heart Attack
  - Customer Churn
  - Machine Breakdown
  - Death ???

Survival Analysis

- Time to event ~
  - failure time
  - survival time
  - event time
- Typical questions of interest
  - What is the probability that a participant survives 5 years?
  - Are there differences in survival between groups (placebo vs treatment)?
  - When will the customer most likely to churn?
  - How do certain personal, behavioral or clinical characteristics affect participants' chances of survival?

Time to Event Variables

- Have some unique features
- Always positive
- Distributions are often skewed.
  - E.g. In high risk patients, relapses may occur early
- Normality assumptions do not apply.
- Nonparametric procedures can be used but there are additional issues (discussed later).
- Complete data (actual time to event data) is not always available

Real Life Application

- Consider the Churn data in Telecom sector
- What do you infer from the data about churn times?

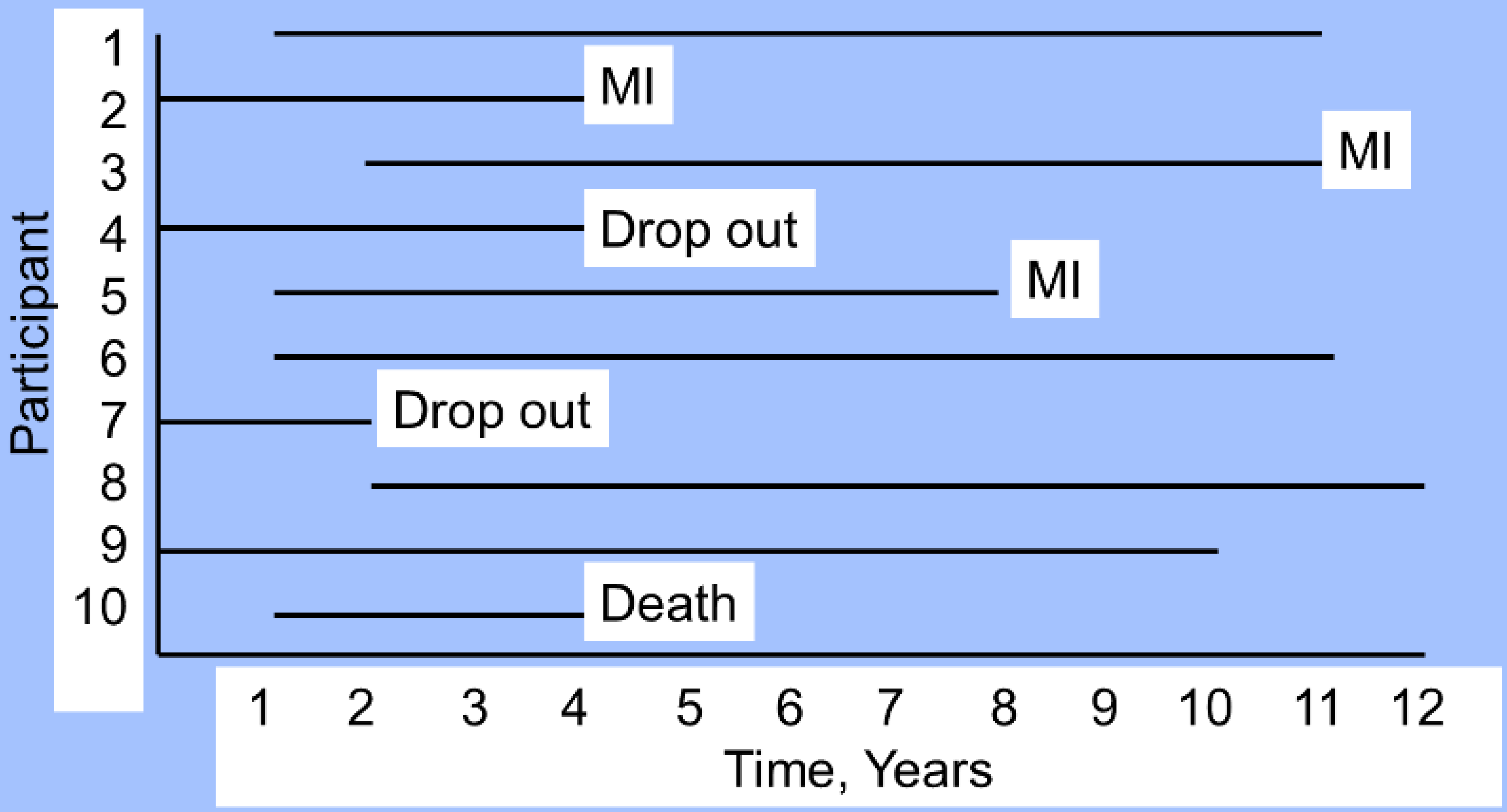
Censored times

- May have incomplete follow-up information
- True survival time cannot be measured
  - Participants drops out of study
  - May enroll late to the study
- Survival Time > Follow-up time

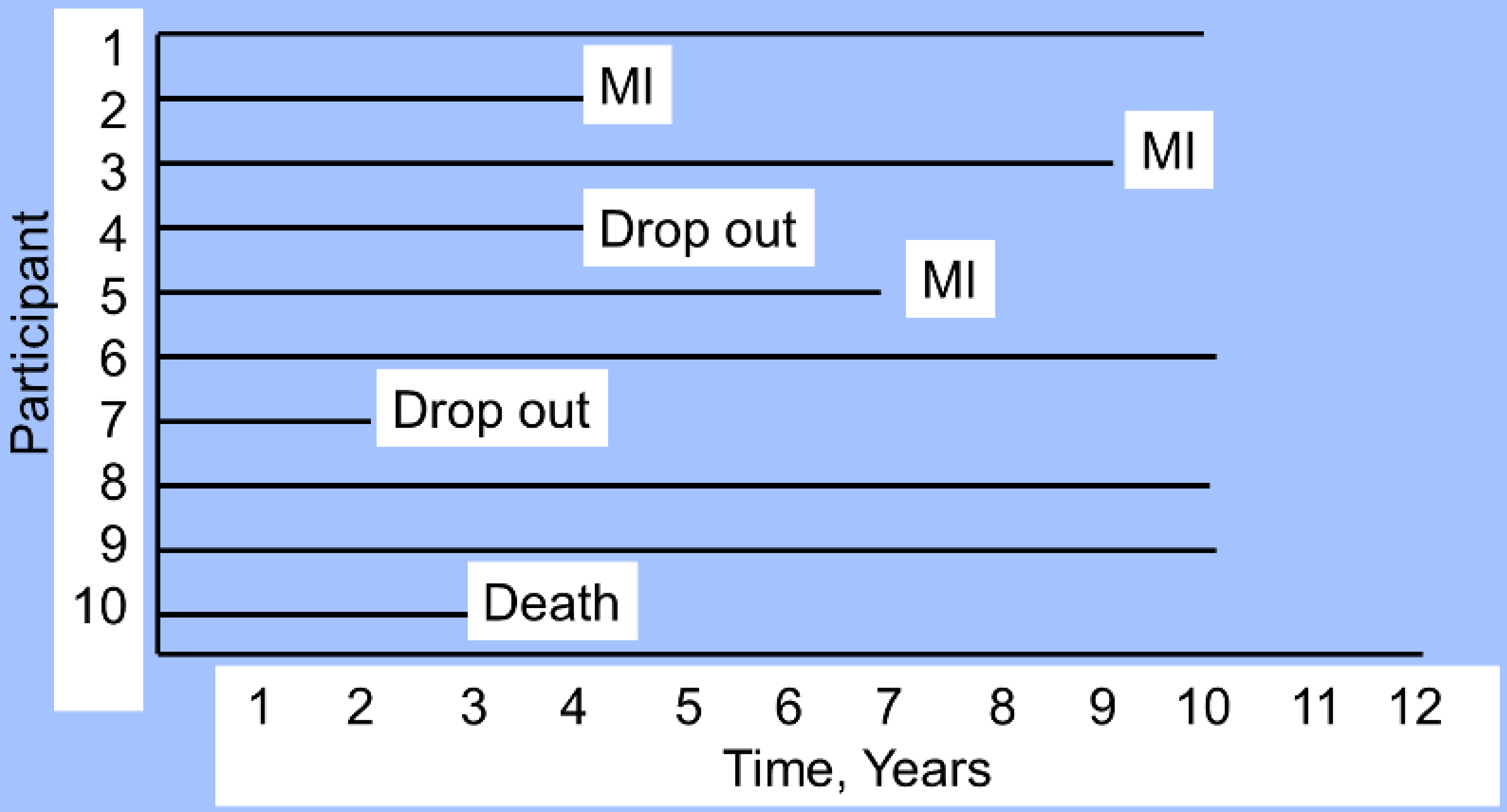
Censoring

- **Right Censoring**
  - Participant does not experience the event during the study
    - Participants drops out of study
    - Participant is event free
- Survival Time > Follow-up time
  - Follow-up time of a participant may be equal to or less than the duration of the study

Example of Censoring



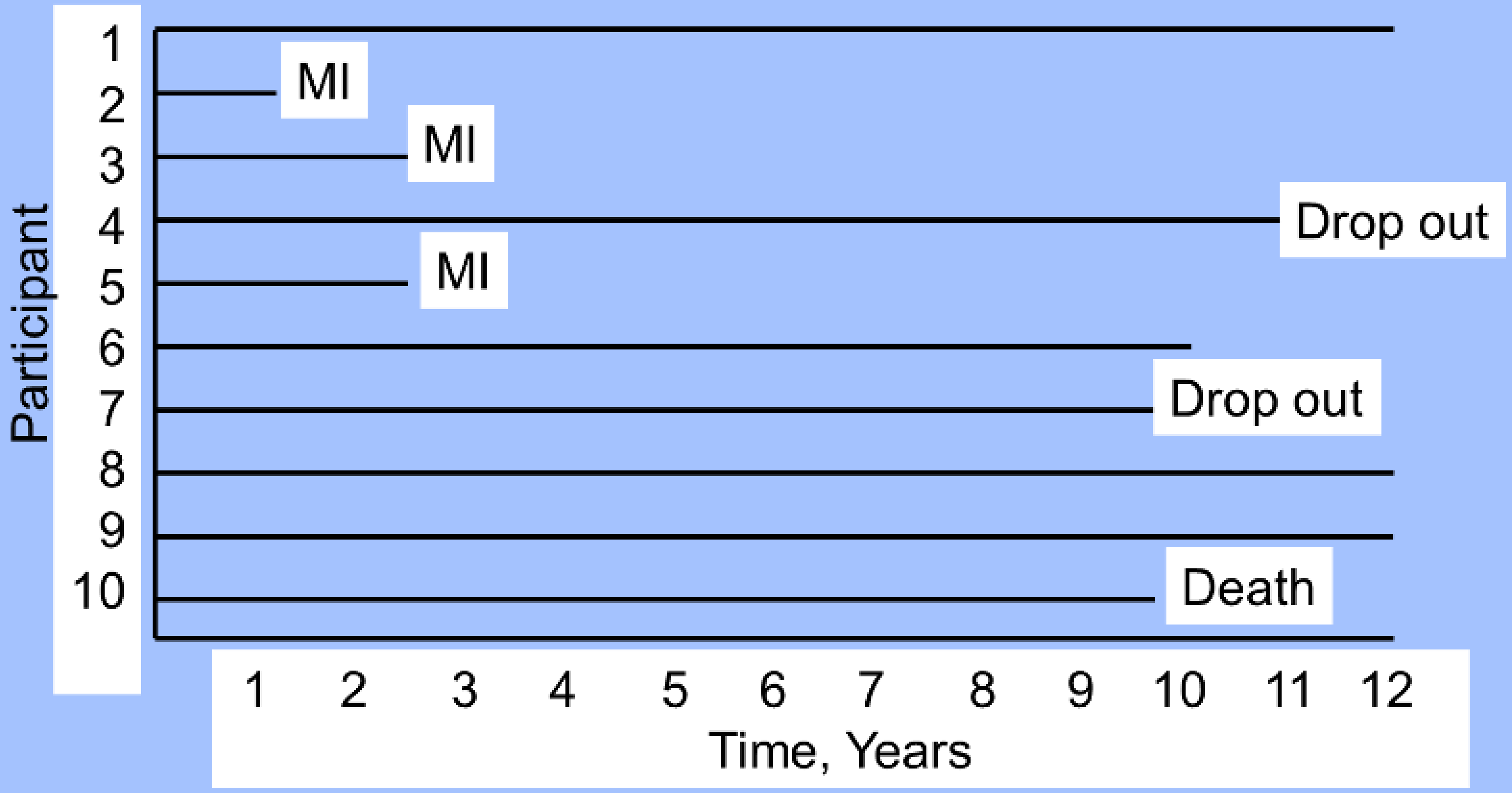
Example of Censoring



Censoring - Assumption

- **Censoring**
  - **independent** or unrelated to the likelihood of developing the event of interest.
- **Non-informative censoring**
  - assumes that the participants whose data are censored would have the same distribution of failure times if they were actually observed.
- Why do we need this assumption?
- Is it a realistic assumption? In what context?

Example of Censoring



Do you think that the survival times are the same for both studies?

Follow-up time

- Follow up time is measured from time zero
  - the start of the study
  - the point at which the participant is considered to be at risk
- until the event occurs,
  - the study ends
  - the participant is lost
  - whichever comes first
- Entry-Participation Time is important

Survival Function

- t ~ a random variable of survival times
- f(t) ~ probability density function of t
- F(t) ~ probability of observing survival time less than or equal to some time t

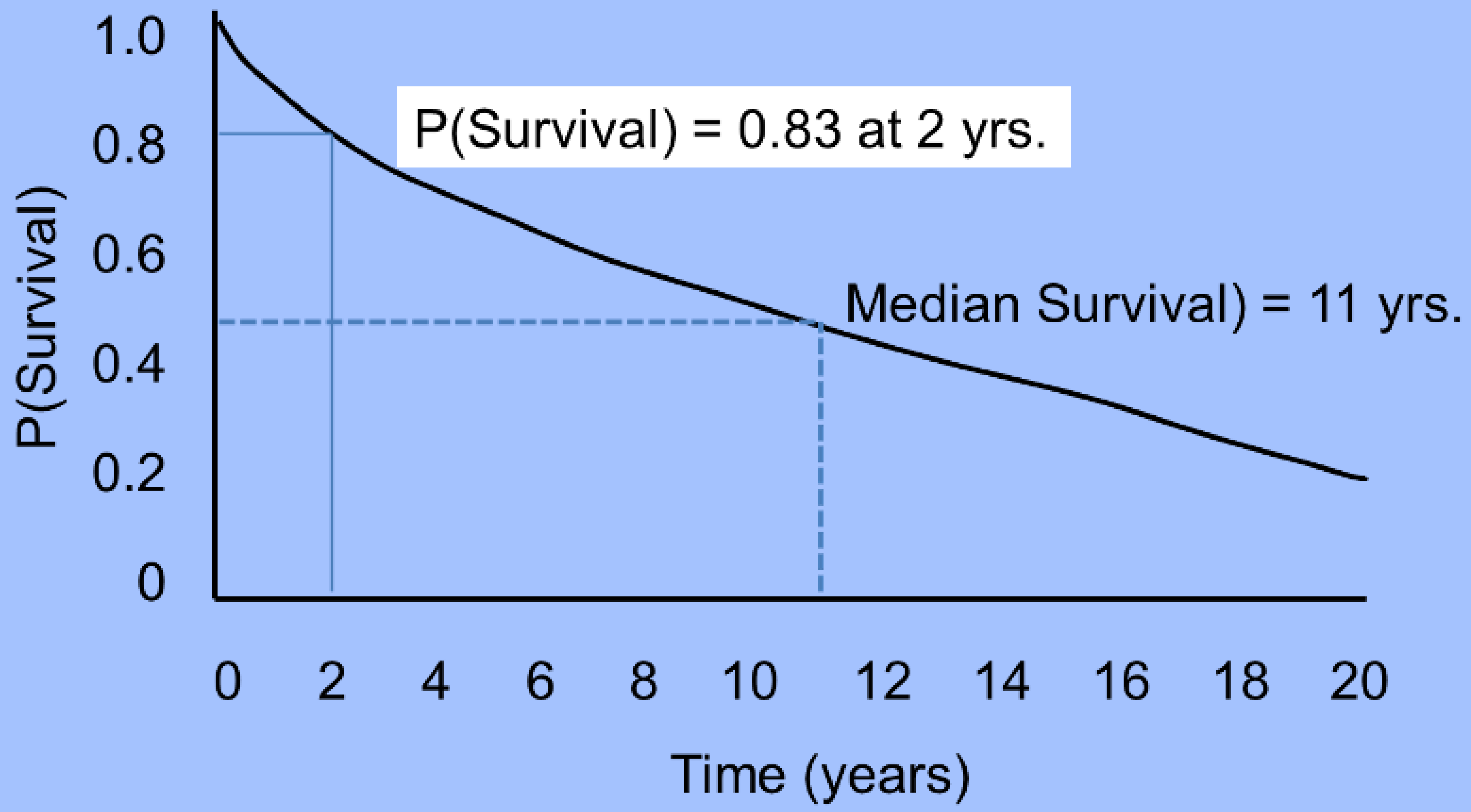
$$F(t) = \int_0^t f(u)du$$

- S(t) ~ The survivor function, the probability of surviving past time t,

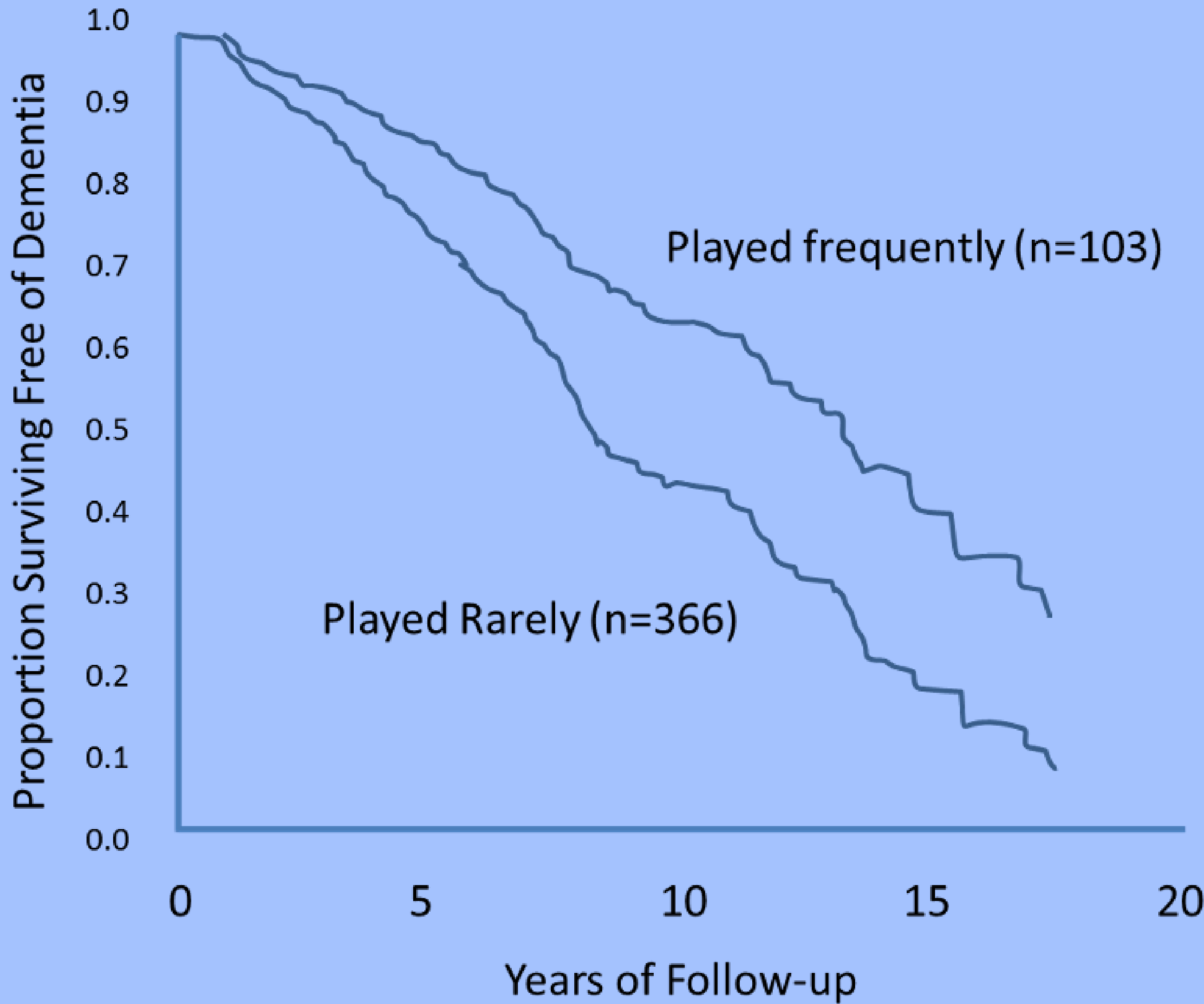
$$S(t) = 1 - F(t)$$

Survival Function

- Survival function ~ S(t)



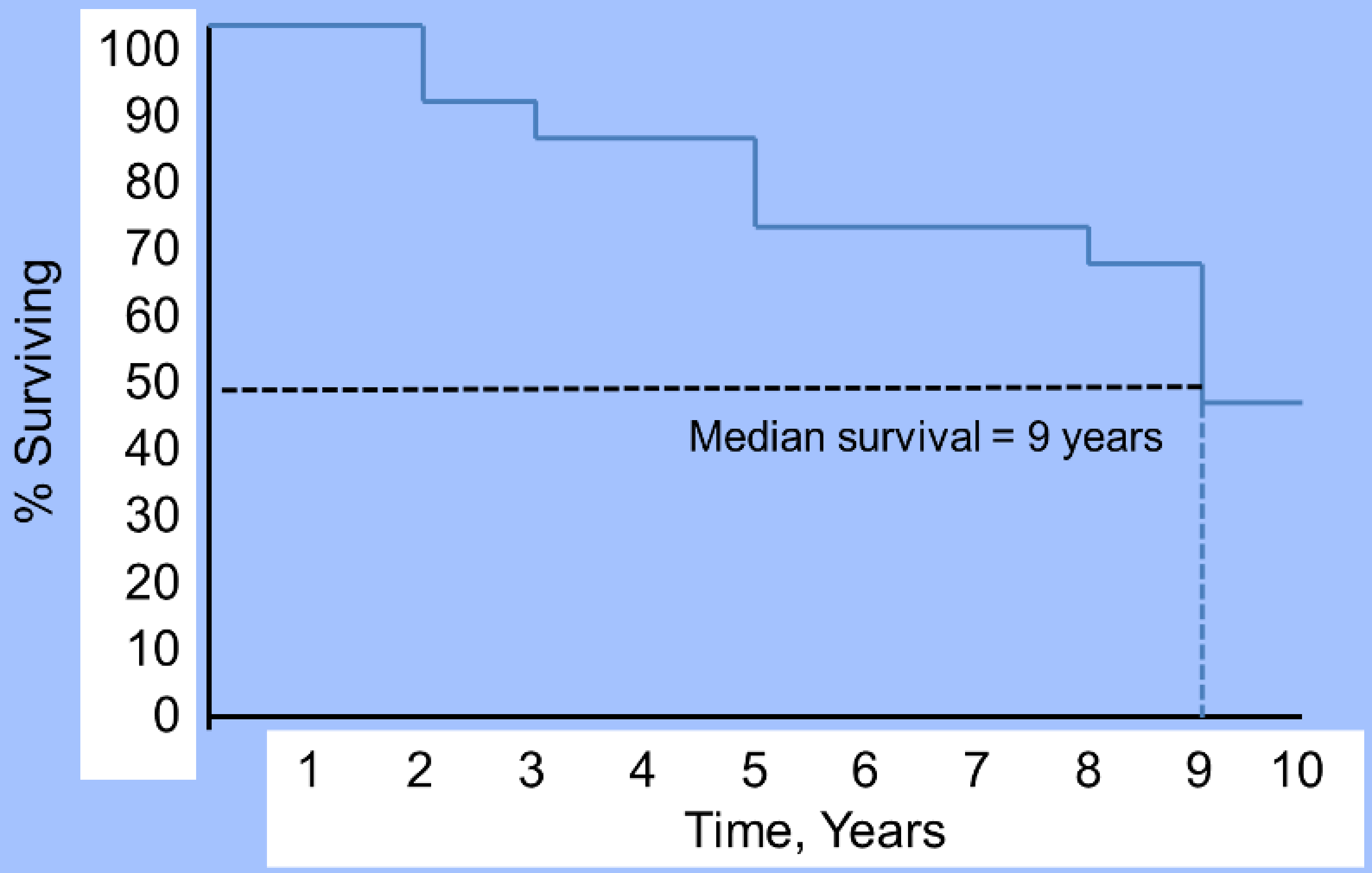
Estimating the Survival Function



Estimating the Survival Function

- Parametric Methods
  - make some assumptions about the survival times
  - exponential, Weibull, Gompertz and log-normal distributions
  - Exponential’s assumption~ a participant's likelihood of suffering the event of interest is independent of how long that person has been event-free.
- Non-parametric Methods
  - Usually step functions by nature
  - **Life Table (Actuarial Table)**
  - **Kaplan-Meier (Product Limit) Approach**

Estimating the Survival Function



Estimating the Survival Function

ID	Death	Last Contact	ID	Death	Last Contact
1		24	11		24
2	3		12		21
3		11	13		12
4		19	14	1	
5		24	15		10
6		13	16	23	
7	14		17		6
8		2	18	5	
9		18	19		9
10		17	20	17	

Life Table (Actuarial Table)

Interval in Years	Number Alive at Beginning of Interval	Number of Deaths During Interval	Number Censored
0-4	20	2	1
5-9	17	1	2
10-14	14	1	4
15-19	9	1	3
20-24	5	1	4

Life Table (Actuarial Table)

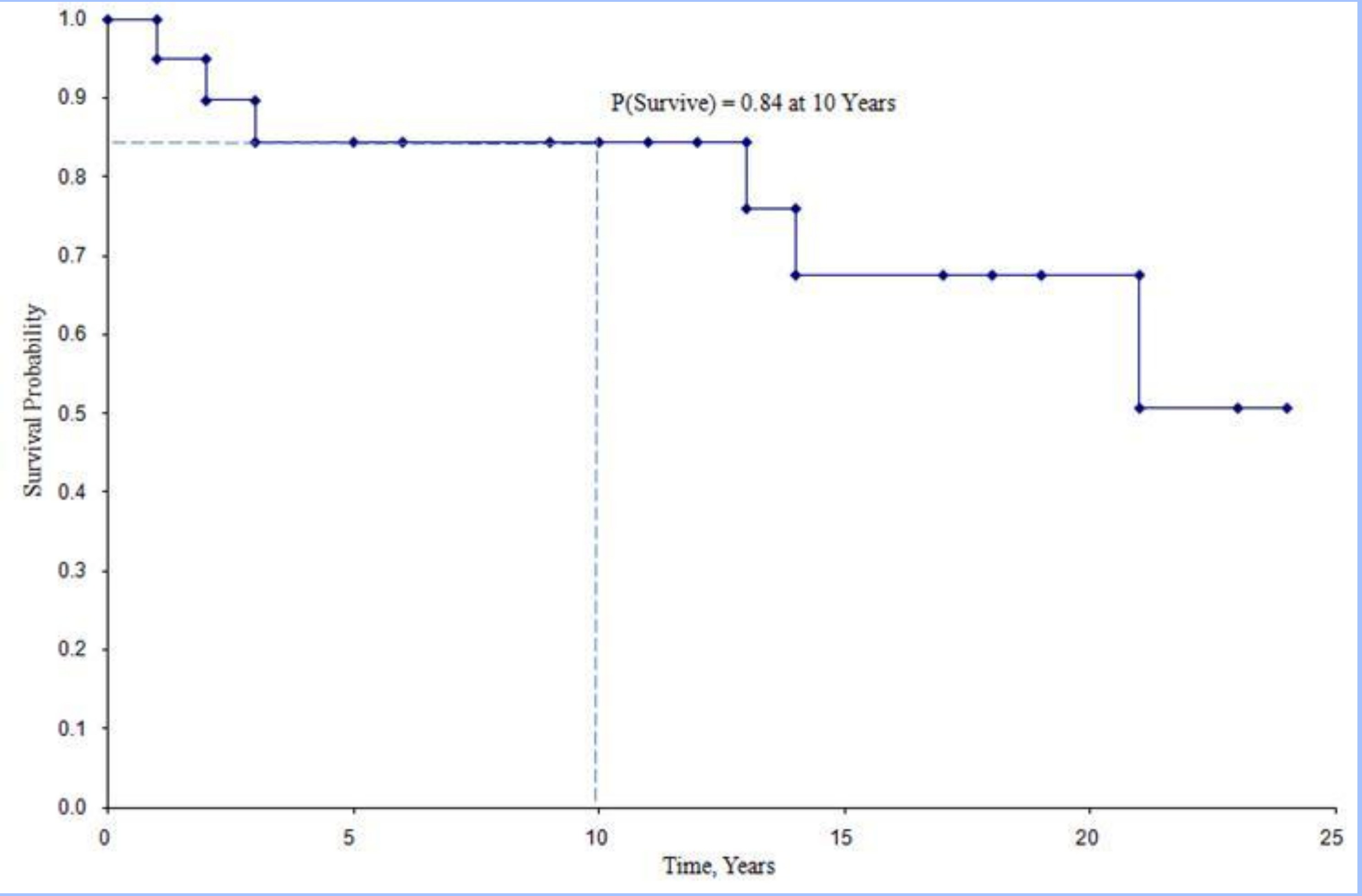
Interval in Years	Number At Risk	Average Number At Risk	Number of Deaths	Lost to Follow-Up, C_t	Prop. Dying, q_t	Prop. Surv, p_t	Surv. Prob. S_t
0-4	20	19.5	2	1	0.103	0.897	0.897
5-9	17	16.0	1	2	0.063	0.937	0.840
10-14	14	12.0	1	4	0.083	0.917	0.770
15-19	9	7.5	1	3	0.133	0.867	0.668
20-24	5	3.0	1	4	0.333	0.667	0.446

Time is divided into equally spaced intervals.

Kaplan-Meier Approach

Time	# at Risk Nt	# of Deaths Dt	#Censored Ct	Sur. Prob. $S_{t+1} = S_t*((N_{t+1}-D_{t+1})/N_{t+1})$
0	20			1
1	20	1		1*((20-1)/20) = 0.950
2	19		1	0.950*((19-0)/19)=0.950
3	18	1		0.950*((18-1)/18) = 0.897
5	17	1		0.897*((17-1)/17) = 0.844
6	16		1	0.844
9	15		1	0.844
10	14		1	0.844
11	13		1	0.844
12	12		1	0.844
13	11		1	0.844
14	10	1		0.760
17	9	1	1	0.676
18	7		1	0.676
19	6		1	0.676
21	5		1	0.676
23	4	1		0.507
24	3		3	0.507

Kaplan-Meier Approach



SE & CI

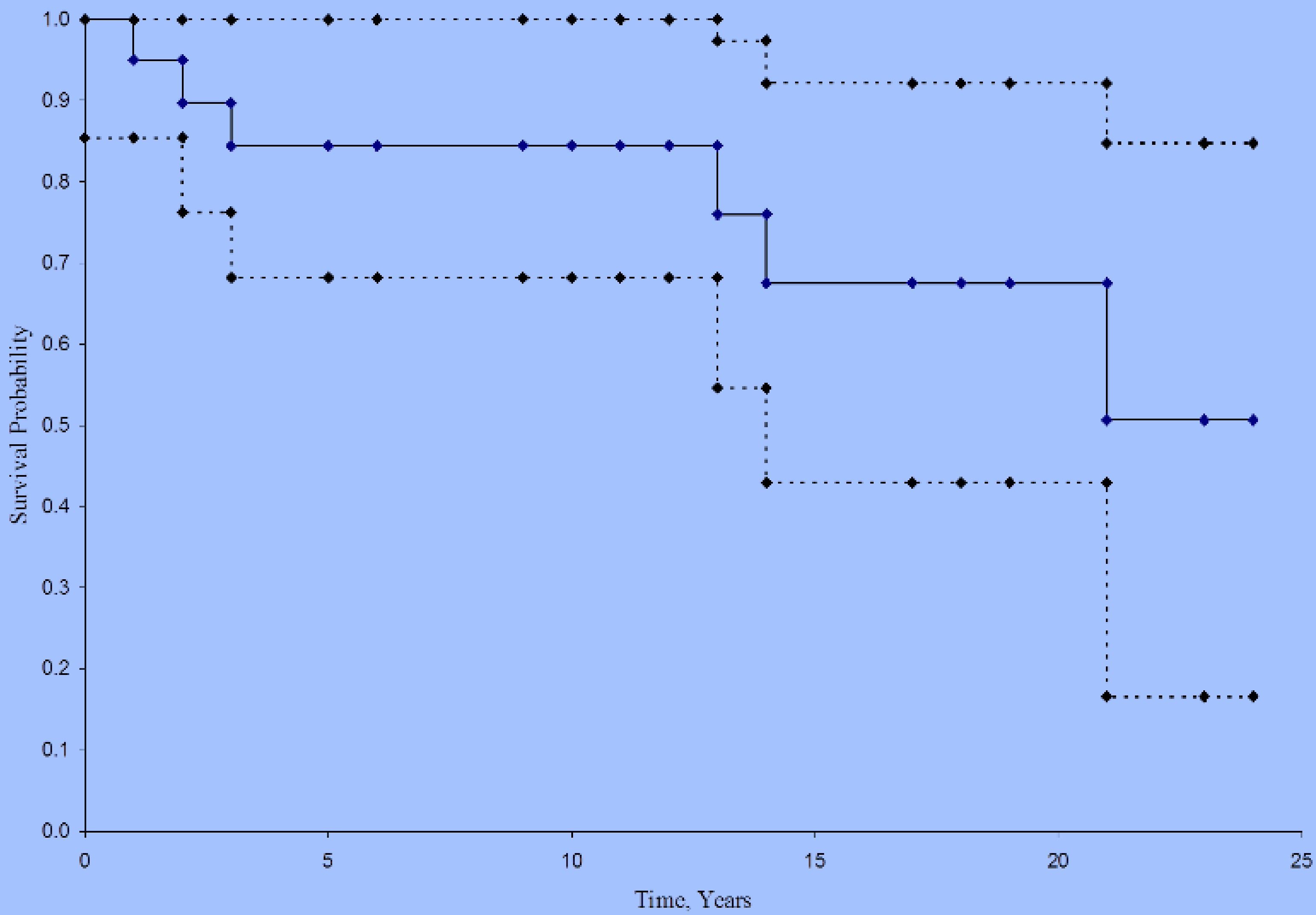
- Standard error of the survival estimates

$$SE\left(S_t\right) = S_t\sqrt{\sum\frac{D_t}{N_t(N_t - D_t)}}$$

SE & CI

Time	Nt	Dt	St	Dt/Nt(Nt-Dt)	∑Dt/Nt(Nt-Dt)	St√∑Dt/Nt(Nt-Dt)	1.96*SE(St)
0	20		1				
1	20	1	0.950	0.003	0.003	0.049	0.096
2	19		0.950	0.000	0.003	0.049	0.096
3	18	1	0.897	0.003	0.006	0.069	0.135
5	17	1	0.844	0.004	0.010	0.083	0.162
6	16		0.844	0.000	0.010	0.083	0.162
9	15		0.844	0.000	0.010	0.083	0.162
10	14		0.844	0.000	0.010	0.083	0.162
11	13		0.844	0.000	0.010	0.083	0.162
12	12		0.844	0.000	0.010	0.083	0.162
13	11		0.844	0.000	0.010	0.083	0.162
14	10	1	0.760	0.011	0.021	0.109	0.214
17	9	1	0.676	0.014	0.035	0.126	0.246
18	7		0.676	0.000	0.035	0.126	0.246
19	6		0.676	0.000	0.035	0.126	0.246
21	5		0.676	0.000	0.035	0.126	0.246
23	4	1	0.507	0.083	0.118	0.174	0.341
24	3		0.507	0.000	0.118	0.174	0.341

SE & CI



1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Survival Probability
0
5
10
15
20
25
Time, Years

Comparing Survival Curves

- Comparison of two or more groups
  - Treatment vs Placebo
  - A Type vs B Type
- Log Rank Test
  - test the null hypothesis of no difference in survival between two or more independent groups
  - Checks whether the survival curves are identical
  - Several variations of the test
  - Similar to chi-square test

Comparing Survival Curves

Chemotherapy Before Surgery			Chemotherapy After Surgery	
Month of Death	Month of Last Contact		Month of Death	Month of Last Contact
8	8		33	48
12	32		28	48
26	20		41	25
14	40			37
21				48
27				25
				43

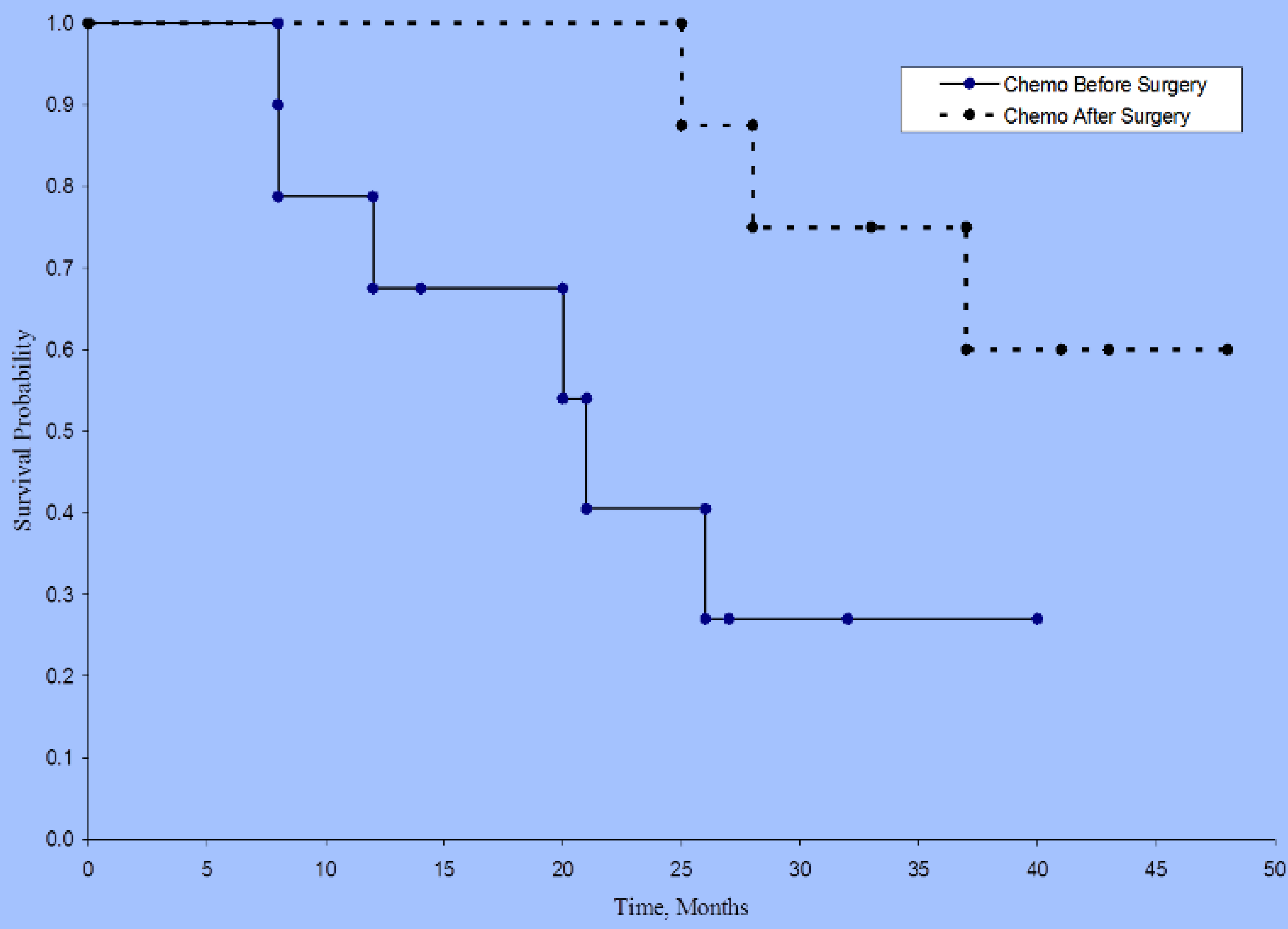
Kaplan-Meier ~ Group Receiving Chemotherapy Before Surgery

Time	N_t	D_t	C_t	S_t
0	10			1
8	10	1	1	0.900
12	8	1		0.788
14	7	1		0.675
20	6		1	0.675
21	5	1		0.540
26	4	1		0.405
27	3	1		0.270
32	2		1	0.270
40	1		1	0.270

Kaplan-Meier ~ Group Receiving Chemotherapy After Surgery

Time	N_t	D_t	C_t	S_t
0	10			1
25	10		2	1.000
28	8	1		0.875
33	7	1		0.750
37	6		1	0.750
41	5	1		0.600
43	4		1	0.600
48	3		3	0.600

Comparing Survival Curves



Comparing Survival Curves

- Null hypothesis: The two survival curves are identical
- The log rank statistic is approximately distributed as a chi-square test statistic

$$\chi^2 = \sum \frac{(\Sigma O_{jt} - \Sigma E_{jt})^2}{\Sigma E_{jt}}$$

Comparing Survival Curves

Time	# at Risk in G1 N_{1t}	# at Risk in G2 N_{2t}	# of Events in G1 O_{1t}	# of Events in G2 O_{2t}
8	10	10	1	0
12	8	10	1	0
14	7	10	1	0
21	5	10	1	0
26	4	8	1	0
27	3	8	1	0
28	2	8	0	1
33	1	7	0	1
41	0	5	0	1

Comparing Survival Curves

Time	N_{1t}	N_{2t}	N_t	O_{1t}	O_{2t}	O_t	E_{1t}	E_{2t}
8	10	10	20	1	0	1	0.500	0.500
12	8	10	18	1	0	1	0.444	0.556
14	7	10	17	1	0	1	0.412	0.588
21	5	10	15	1	0	1	0.333	0.667
26	4	8	12	1	0	1	0.333	0.667
27	3	8	11	1	0	1	0.273	0.727
28	2	8	10	0	1	1	0.200	0.800
33	1	7	8	0	1	1	0.125	0.875
41	0	5	5	0	1	1	0.000	1.000
				6	**3**		**2.620**	**6.380**

Comparing Survival Curves

- Computing the test statistic

$$\chi^2 = \Sigma \frac{(\Sigma O_{jt} - \Sigma E_{jt})^2}{\Sigma E_{jt}} = \frac{(6-2.620)^2}{2.620} + \frac{(3-6.380)^2}{6.380} = \mathbf{4.360} + \mathbf{1.791} = \mathbf{6.151}$$

- The test statistic is approximately distributed as chi-square with 1 degree of freedom.
- Reject H_0 if $X^2 > 3.84$
- Rejected with α = 0.05

Cox (Proportional Hazards) Regression

- Cox regression
  - Relate several risk factors or exposures, considered simultaneously, to survival time
- **Hazard rate**
  - the risk of failure (i.e., the risk or probability of suffering the event of interest), given that the participant has survived up to a specific time
- Hazard represents the expected number of events per one unit of time
  - Thus, hazard in a group can exceed 1

Cox (Proportional Hazards) Regression

- Hazard Function
  - The primary focus of survival analysis is typically to model the hazard rate

$$h(t) = \frac{f(t)}{S(t)}$$

  - The hazard rate thus describes the instantaneous rate of failure at time t and ignores the accumulation of hazard up to time t

Cox (Proportional Hazards) Regression

- Hazard Function
  - cumulative hazard function

$$H(t) = \int_0^t h(u)du$$

  - One interpretation of the cumulative hazard function is the expected number of failures over time interval [0,t]

Cox (Proportional Hazards) Regression

- Function Relations

$$h(t) = \frac{f(t)}{S(t)}$$

$$f(t) = -\frac{dS}{dt}$$

$$S(t) = \exp(-H(t))$$

$$F(t) = 1 - \exp(-H(t))$$

$$f(t) = h(t)\exp(-H(t))$$

Parametric Estimation of S(t)

We will use maximum likelihood estimation to estimate the unknown parameters of the parametric distributions.

- If Y_i is uncensored, the ith subject contributes $f(Y_i)$ to the likelihood
- If Y_i is censored, the ith subject contributes $Pr(y > Y_i)$ to the likelihood.

The joint likelihood for all n subjects is

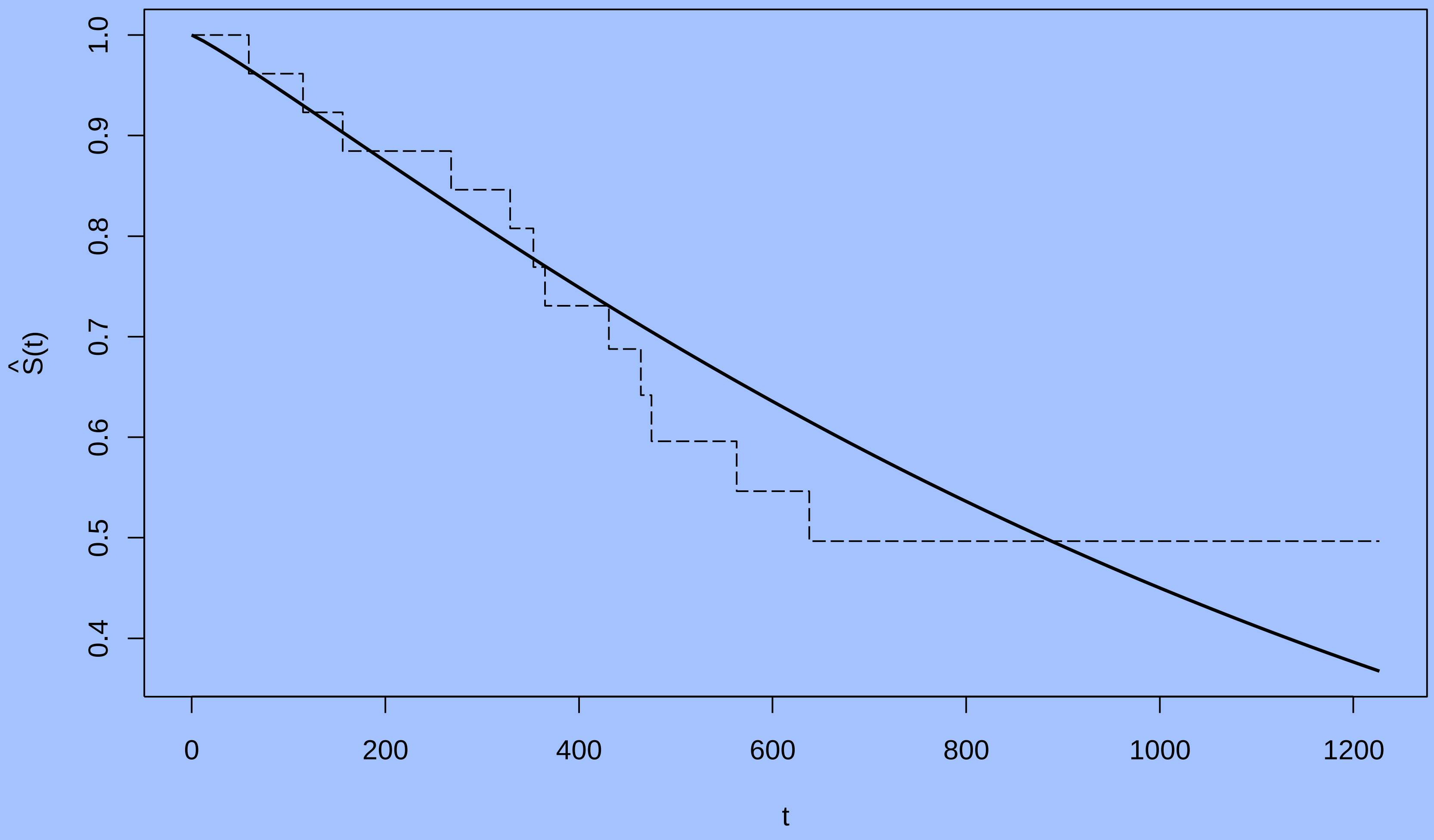
$$L = \prod_{i:\delta_i=1}^{n} f(Y_i) \prod_{i:\delta_i=0}^{n} S(Y_i).$$

Parametric Estimation of S(t)

The log-likelihood can be written as

$$\log L = \sum_{i:\delta_i=1}^{n} \log(h(Y_i)) - \sum_{i=1}^{n} H(Y_i).$$

Parametric Estimation of S(t)



Cox (Proportional Hazards) Regression

- **Hazard ratio**
  - Similar to odds ratio in multiple logistic regression
  - Ratio of the total number of observed to expected events in two independent comparison groups:

$$HR = \frac{\Sigma O_{Exp,t} / \Sigma E_{Exp,t}}{\Sigma O_{Unex,t} / \Sigma E_{Unex,t}} = \frac{\Sigma O_{treated,t} / \Sigma E_{treated,t}}{\Sigma O_{control,t} / \Sigma E_{control,t}}$$

- Hazard ratio in Chemotherapy example

$$HR = \frac{O_t / E_{t,\ Chemo\ before\ Surgery}}{O_t / E_{t\ Chemo\ After\ Surgery}} = \frac{6/2.620}{3/6.380} = \mathbf{4.870}$$

Cox (Proportional Hazards) Regression

- **Single vs Multiple Factors**
  - it is often of interest to assess the association between several risk factors, considered simultaneously, and survival time
  - Thus, Cox regression analysis is used for that purpose
- Assumptions:
  - independence of survival times between distinct individuals in the sample,
  - a multiplicative relationship between the predictors and the hazard (as opposed to a linear one)
  - a constant hazard ratio over time.

Cox (Proportional Hazards) Regression

- The Cox proportional hazards regression model

$$h(t) = h_0(t)\,\mathbf{exp}\left(b_1 X_1 + b_2 X_2 + \ldots + b_p X_p\right)$$

- A simple model with one predictor, X_1.

$$h(t) = h_0(t)\,\mathbf{exp}\left(b_1 X_1\right)$$

- Compare two participants first with $X_1 = a$, second $X_1 = b$
  - The hazard ratio is the ratio of these two expected hazards:
  - $h_0(t)\exp(b_1 a) / h_0(t)\exp(b_1 b) = \exp(b_1(a-b))$

Cox (Proportional Hazards) Regression

- Hazard Ratio

$$\frac{h(t)}{h_0(t)} = \mathbf{exp}\left(b_1X_1 + b_2X_2 + \ldots b_pX_p\right)$$

- Log of the relative hazard ~ linear function of the predictors

$$\mathbf{ln}\ \left\{\frac{h(t)}{h_0(t)}\right\} = b_1X_1 + b_2X_2 + \ldots b_pX_p$$

- Cox model ~ **semi-parametric model**
  - no assumptions about the shape of the baseline hazard function
  - Other assumptions (independence, proportional relations)

Cox (Proportional Hazards) Regression

- Cox models
  - maximum likelihood methods
  - what is the probability of observing subject i fail at time tj?
  - At the beginning of a given time interval tj, say there are Rj subjects still at-risk, each with their own hazard rates:

$$h(t_j|x_i) = h_0(t_j)\exp(x_i\beta)$$

  - The probability of observing subject j fail out of all Rj remaining at-risk subjects is the proportion of the sum total of hazard rates of all Rj subjects that is made up by subject j's hazard rate.

Cox (Proportional Hazards) Regression

- Cox models
  - For example, if there were three subjects still at risk at time tj, the probability of observing subject 2 fail at time tj would be:

$$\Pr(\text{subject=2}|\text{failure=tj}) = \frac{h\left(t_j \middle| x_2\right)}{h\left(t_j \middle| x_1\right) + h\left(t_j \middle| x_2\right) + h\left(t_j \middle| x_3\right)}$$

  - All of those hazard rates are based on the same baseline hazard rate $h_0(t_i)$ so we can simplify the above expression to:

$$\Pr(\text{subject=2}|\text{failure=tj}) = \frac{\exp(x_2\beta)}{\exp(x_1\beta) + \exp(x_2\beta) + \exp(x_3\beta)}$$

Cox (Proportional Hazards) Regression

- Cox models
  - We can similarly calculate the joint probability of observing each of the n subject's failure times, or the likelihood of the failure times, as a function of the regression parameters, β, given the subject's covariates values x_j:

$$L(\beta) = \prod_{j=1}^{n} \left\{ \frac{\exp(x_j \beta)}{\sum_{i \in R_j} \exp(x_i \beta)} \right\}$$

Example

	Die (n=402)	Do Not Die (n=4778)
Mean (SD) Age, years	65.6 (8.7)	56.1 (7.5)
N (%) Male	221 (55%)	2145 (45%)

Risk Factor	Parameter Estimate	P-Value
Age, years	0.11149	0.0001
Male Sex	0.67958	0.0001

Example

Risk Factor	Parameter Est.	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)

Example 2

Group	Number of Participants	Number (%) of CVD Events	Group
Normal Weight	1651	202 (12.2%)	Normal Weight
Overweight	1648	241 (14.6%)	Overweight
Obese	638	100 (15.7%)	Obese

Example 2

	Overweight			Obese		
Model	Param. Est.	P- Value	HR (95% CI for HR)	Param. Est.	P- Value	HR (95% CI for HR)
Unadjusted or Crude Model	0.19484	0.0411	1.215 (1.008-1.465)	0.27030	0.0271	1.310 (1.031-1.665)
Age and Sex Adjusted	0.06525	0.5038	1.067 (0.882-1.292)	0.28960	0.0188	1.336 (1.049-1.701)
Adjusted for Clinical Risk Factors*	0.07503	0.4446	1.078 (0.889-1.307)	0.24944	0.0485	1.283 (1.002-1.644)

Time-Dependent Covariates

- Risk factors or predictors may change over time
- The Cox proportional hazards regression model with time dependent covariates

$$\ln\left\{\frac{h(t)}{h_0(t)}\right\} = b_1 X_1(t) + b_2 X_2(t) + \ldots + b_p X_p(t)$$

- Survival analysis models can include both time dependent and time independent predictors simultaneously

Time-dependent covariates

- Covariate values for an individual may change over time
- For example, if you are evaluating the effect of taking the drug for a cancer risk in an observational study, subject may start and stop the drug at will. Subject A may be taking the drug at the time of the first event, but may have stopped taking it by the time the 15th event happens.
- If you are evaluating the effect of weight on diabetes risk over a long study period, subjects may gain and lose large amounts of weight, making their baseline weight a less than ideal predictor.
- If you are evaluating the effects of smoking on the risk of pancreatic cancer, study participants may change their smoking habits throughout the study.
- Cox regression can handle these time-dependent covariates!

Time-dependent covariates

- For example, evaluating the effect of taking oral contraceptives (OCs) on stress fracture risk in women athletes over two years—many women switch on or off OCs .
- If you just examine risk by a woman's OC-status at baseline, can't see much effect for OCs. But, you can incorporate times of starting and stopping OCs.

Time-dependent covariates

- Ways to look at OC use:
- Not time-dependent
  - Ever/never during the study
  - Yes/no use at baseline
  - Total months use during the study
- Time-dependent
  - Using OCs at event time t (yes/no)
  - Months of OC use up to time t

Time-dependent covariates: Example data

4 events

ID	Time	Fracture	StartOC	StopOC
1	12	1	0	12
2	11	0	10	11
3	20	1	.	.
4	24	0	0	24
5	19	0	0	11
6	6	1	.	.
7	17	1	1	7

1. Time independent predictor…

- Baseline use (yes/no)

Time-dependent covariates

Order by Time…

ID	Time	Fracture	StartOC	StopOC
6	6	1	.	.
2	11	0	10	11
1	12	1	0	12
7	17	1	1	7
5	19	0	0	11
3	20	1	.	.
4	24	0	0	24

Time-dependent covariates

3 OC users at baseline

ID	Time	Fracture	StartOC	StopOC
6	6	1	.	.
2	11	0	10	11
1	12	1	0	12
7	17	1	1	7
5	19	0	0	11
3	20	1	.	.
4	24	0	0	24

Time-dependent covariates

4 non-users at baseline

ID	Time	Fracture	StartOC	StopOC
6	6	1	.	.
2	11	0	10	11
1	12	1	0	12
7	17	1	1	7
5	19	0	0	11
3	20	1	.	.
4	24	0	0	24

Time-dependent covariates

First event is in a non-OC user at baseline. (risk set: 3 users/4 non)

Next is a censoring (non-user)

Second event is in a baseline

Third event is in a non-user at

Next is a censoring (baseline

Fourth and last event is in a non-

Censoring.

ID	Time	Frac		
6	6	1		
2	11	0		
1	12	1		
7	17	1		
5	19	0	0	11
3	20	1	.	.
4	24	0	0	24

The PL using baseline value of OC use

$$L_p(\beta_{oc}) = \frac{e^{\beta(0)}}{3e^{\beta(1)} + 4e^{\beta(0)}} \, x \, \frac{e^{\beta(1)}}{3e^{\beta(1)} + 2e^{\beta(0)}} \, x \, \frac{e^{\beta(0)}}{2e^{\beta(1)} + 2e^{\beta(0)}} \, x \, \frac{e^{\beta(0)}}{e^{\beta(1)} + e^{\beta(0)}}$$

The PL using ever/never value of OC use

A second time-independent option would be to use the variable "ever took OCs" during the study period…

Time-dependent covariates

First event is in a never-user. (risk set: 5 ever users/2 never)

Next is a censoring (ever user)

Second event is in an ever-user.

Third event is in an ever-user(risk set: 3 users/1 non)

Next is a censoring (ever user).

Fourth and last event is in a never-

Censoring.

ID	Time	Fracture	StartOC	StopOC
6	6	1		
2	11	0		
1	12	1		
7	17	1		
5	19	0	0	11
3	20	1	.	.
4	24	0	0	24

The PL using ever/never value of OC use

"Ever took OCs" during the study period

$$L_p(\beta_{oc}) = \frac{e^{\beta(0)}}{5e^{\beta(1)} + 2e^{\beta(0)}} \, x \, \frac{e^{\beta(1)}}{4e^{\beta(1)} + e^{\beta(0)}} \, x \, \frac{e^{\beta(1)}}{3e^{\beta(1)} + e^{\beta(0)}} \, x \, \frac{e^{\beta(0)}}{e^{\beta(1)} + e^{\beta(0)}}$$

Time-dependent…

Time-dependent covariates

First event at time 6

ID	Time	Fracture	StartOC	StopOC
6	6	1	.	.
2	11	0	10	11
1	12	1	0	12
7	17	1	1	7
5	19	0	0	11
3	20	1	.	.
4	24	0	0	24

The PL at t=6

$$L_p(\beta_{oc}) = \frac{e^{\beta x_6(t=6)}}{e^{\beta x_1(6)} + e^{\beta x_2(6)} + e^{\beta x_3(6)} + e^{\beta x_4(6)} + e^{\beta x_5(6)} + e^{\beta x_6(6)} + e^{\beta x_7(6)}}$$

X is time-dependent

Time-dependent covariates

At the first event-time (6), there are 4 not on OCs and 3 on OCs.

ID	Time	Fractur		
6	6	1	.	.
2	11	0	10	11
1	12	1	0	12
7	17	1	1	7
5	19	0	0	11
3	20	1	.	.
4	24	0	0	24

The PL at t=6

$$L_p(\beta_{oc}) = \frac{e^{\beta x_6(t=6)}}{e^{\beta x_1(6)} + e^{\beta x_2(6)} + e^{\beta x_3(6)} + e^{\beta x_4(6)} + e^{\beta x_5(6)} + e^{\beta x_6(6)} + e^{\beta x_7(6)}}$$

$$= \frac{e^{\beta(0)}}{3e^{\beta(0)} + 4e^{\beta(1)}}$$

Time-dependent covariates

Second event at time 12

ID	Time	Fracture	StartOC	StopOC
[redacted]				
1	12	1	0	12
7	17	1	1	7
5	19	0	0	11
3	20	1	.	.
4	24	0	0	24

The PL at t=12

$$L_p(\beta_{oc}) = \frac{e^{\beta(0)}}{3e^{\beta(0)} + 4e^{\beta(1)}} x \frac{e^{\beta(1)}}{2e^{\beta(1)} + 3e^{\beta(0)}}$$

Time-dependent covariates

Third event at time 17

ID	Time	Fracture	StartOC	StopOC
[redacted]				
7	17	1	1	7
5	19	0	0	11
3	20	1	.	.
4	24	0	0	24

The PL at t=17

$$L_p(\beta_{oc}) = \frac{e^{\beta(0)}}{3e^{\beta(0)} + 4e^{\beta(1)}} \, x \, \frac{e^{\beta(1)}}{2e^{\beta(1)} + 3e^{\beta(0)}} \, x \, \frac{e^{\beta(0)}}{e^{\beta(1)} + 3e^{\beta(0)}}$$

Time-dependent covariates

Fourth event at time 20

ID	Time	Fracture	StartOC	StopOC
[redacted]				
3	20	1	.	.
4	24	0	0	24

The PL at t=20

$$L_p(\beta_{oc}) = \frac{e^{\beta(0)}}{3e^{\beta(0)} + 4e^{\beta(1)}} \, x \, \frac{e^{\beta(1)}}{2e^{\beta(1)} + 3e^{\beta(0)}} \, x \, \frac{e^{\beta(0)}}{e^{\beta(1)} + 3e^{\beta(0)}} \, x \, \frac{e^{\beta(0)}}{e^{\beta(1)} + e^{\beta(0)}}$$

vs. PL for OC-status at baseline (from before):

$$L_p(\beta_{oc}) = \frac{e^{\beta(0)}}{4e^{\beta(0)} + 3e^{\beta(1)}} \, x \, \frac{e^{\beta(1)}}{3e^{\beta(1)} + 2e^{\beta(0)}} \, x \, \frac{e^{\beta(0)}}{2e^{\beta(1)} + 2e^{\beta(0)}} \, x \, \frac{e^{\beta(0)}}{e^{\beta(1)} + e^{\beta(0)}}$$