Author: Dr. Bondor Cosmina-Ioana

Survival data analysis

A ALWAYS

S SEEK

K KNOWLEDGE

Objectives

- Kaplan –Meier survival curve analysis
- Hazard function
- Comparison of two survival curves
- prediction with Cox regression
- Exercises

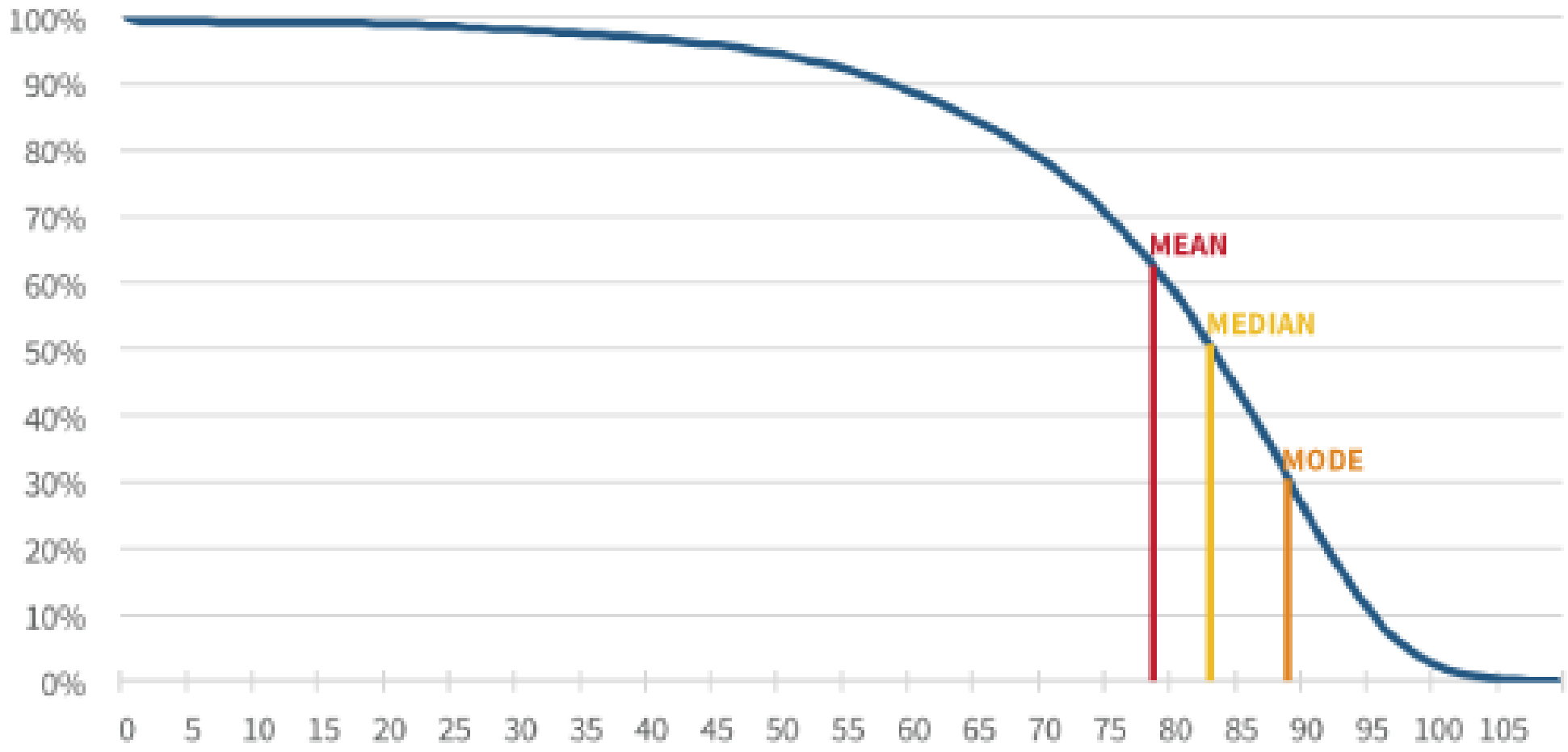
Survival time

- from birth to death

in medicine

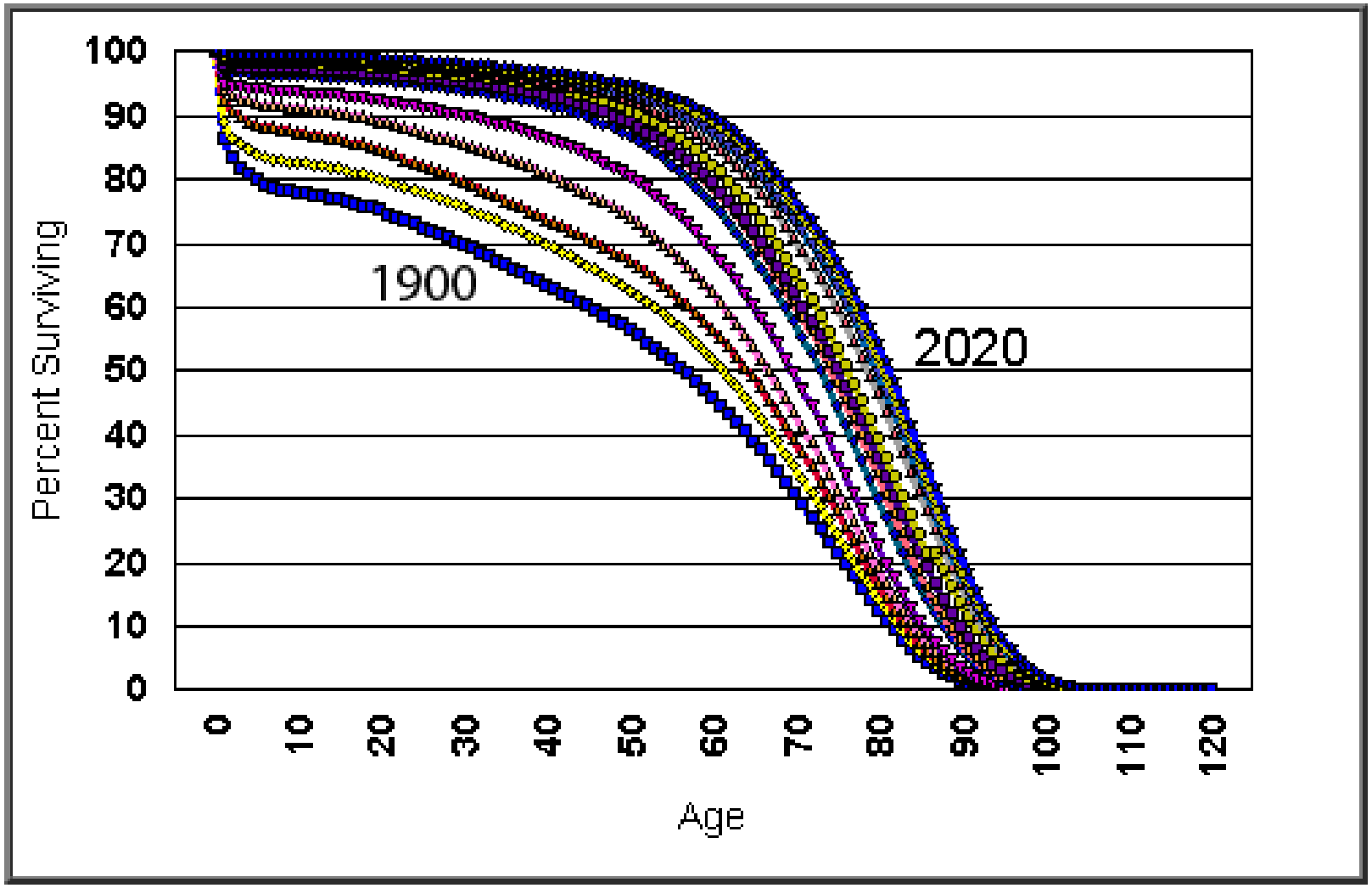
- from symptoms to death
- from presentation at medical doctor with the symptoms to death

PROJECTED SURVIVAL CURVE FOR CHILDREN BORN IN 2014



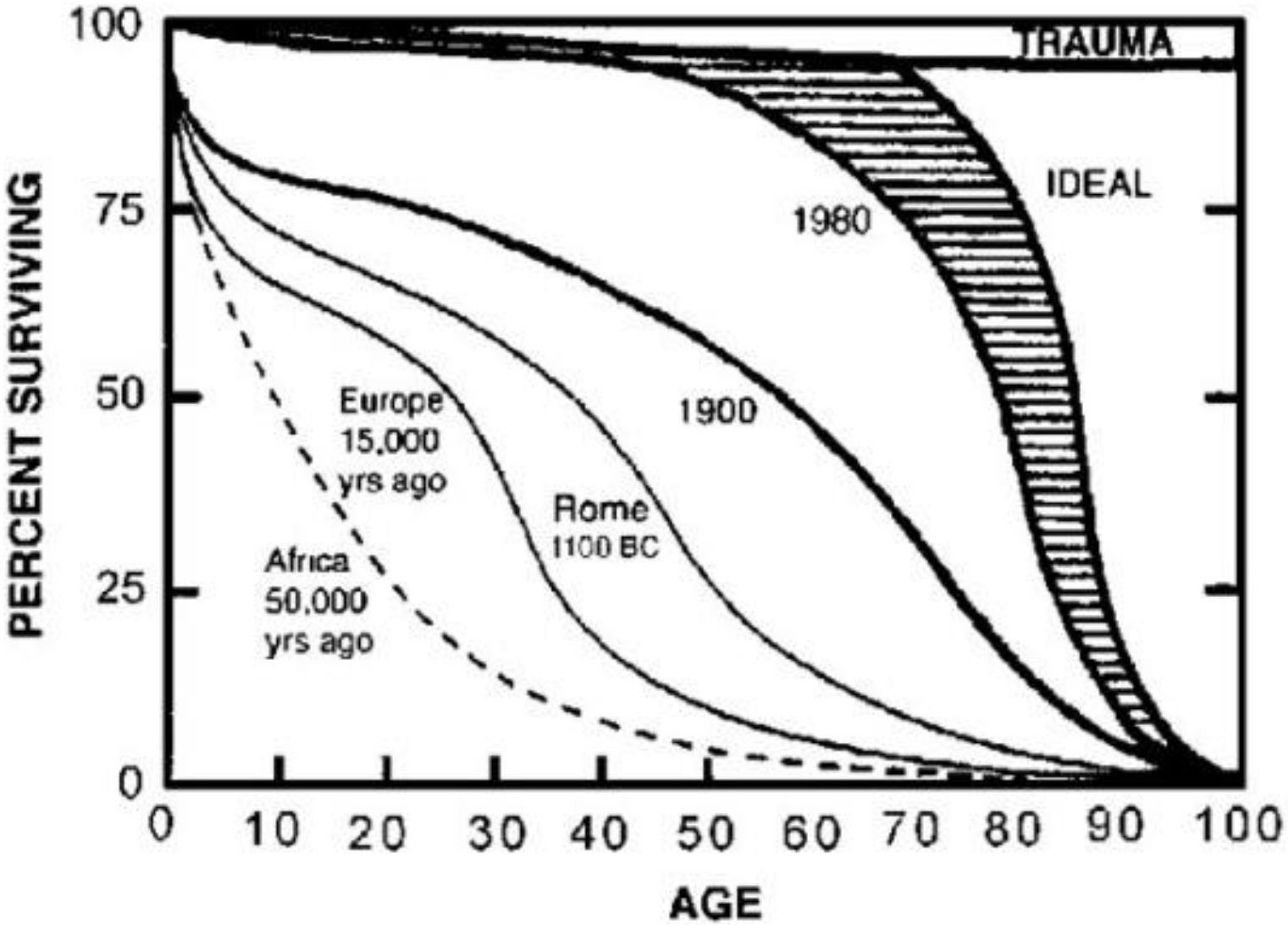
Source: Human Mortality Database. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org. (Data downloaded on 12/1/2016)

The Survival Curve, Life-Expectancy, and Control of Ageing



The above graphs are comparative Survival Curves for over 100 years, from 1900 and as projected into the near future.

https://www.fis.org/LE-Calc/form_A_Basic_criteria.php



Troen BR. The Biology of aging. Mount Sinai Journal of Medicine A Journal of Translational and Personalized Medicine 2003;70(1):3-22

The question

What influences survival time?

a procedure

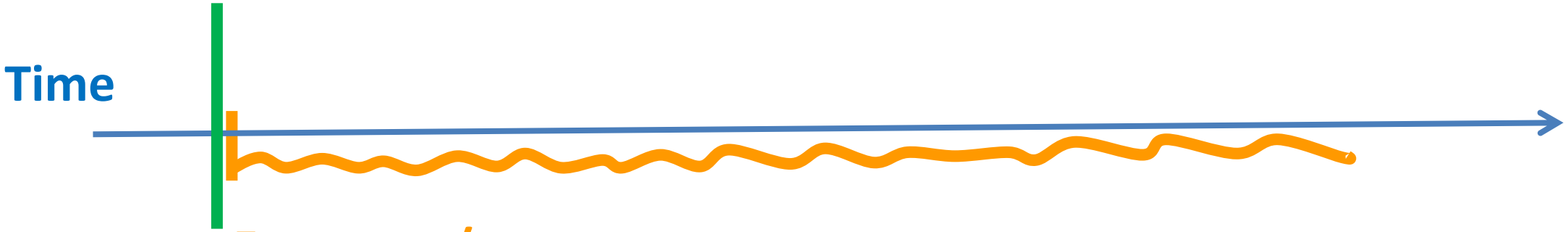
a treatment

a risk factor

a protective factor

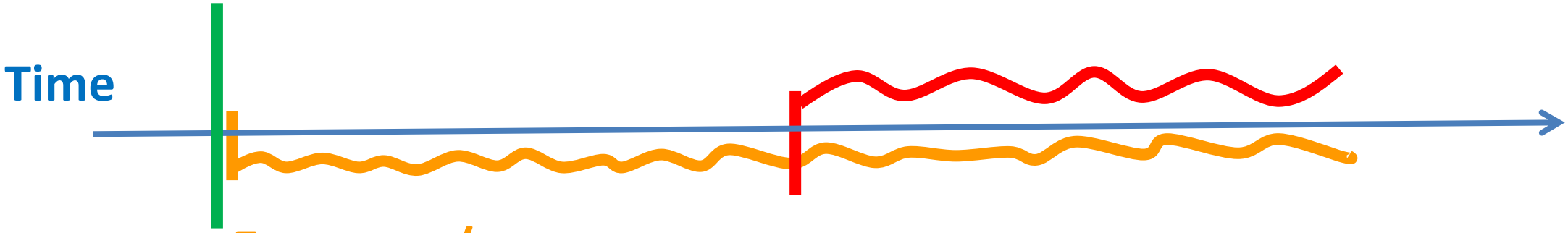
start of the study – random recruitment of patients

start of the study – random recruitment of patients



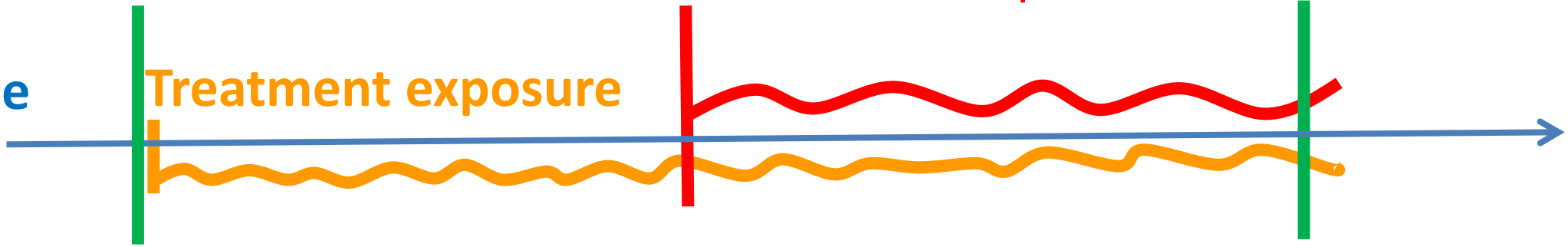
start of the study – random recruitment of patients

The occurrence of a predefined event



start of the study –
random recruitment
of patient

The occurrence of a predefined event



Follow-up – Evaluation

Survival time - Dependent variable – our interest

Survival study - terms

- survival time – time elapsed between the inclusion of a subject in the study and the onset of a predefinit event
  - death
    - overall survival – time until death
  - onset of a disease, complication, relapse
    - progression-free survival
  - remission, recovery, healing
    - one-minus survival
  - other

How to recognize a survival analysis?

- there is an **event** whose occurrence is variable in time and time to the onset of the event is of interest
- event
  - positive decrease in time
    - Ex. event - the patient recover
    - Researchers are interested in identifying the factors that reduce recovery time
  - negative increase in time
    - Ex. event - metastasis
    - Researchers are interested in identifying the factors that increase the time to metastasis onset

LELC Scenario

- Lung cancer with lymphoepithelial carcinoma (LELC) – a rare type of cancer. All patients were extracted from US databases from 1973 to 2011. A total of 62 patients.
- Possible objectives
  - how long they survived from the presentation at the doctor
  - how do various risk factors influence the survival time

He J, Shen J, Pan H, Huang J, Liang W, He J. Pulmonary lymphoepithelioma-like carcinoma: a Surveillance, Epidemiology, and End Results database analysis. J Thorac Dis. 2015 Dec;7(12):2330-8.

PTX scenario

26 dialysis patients with parathyroidectomy (PTX)

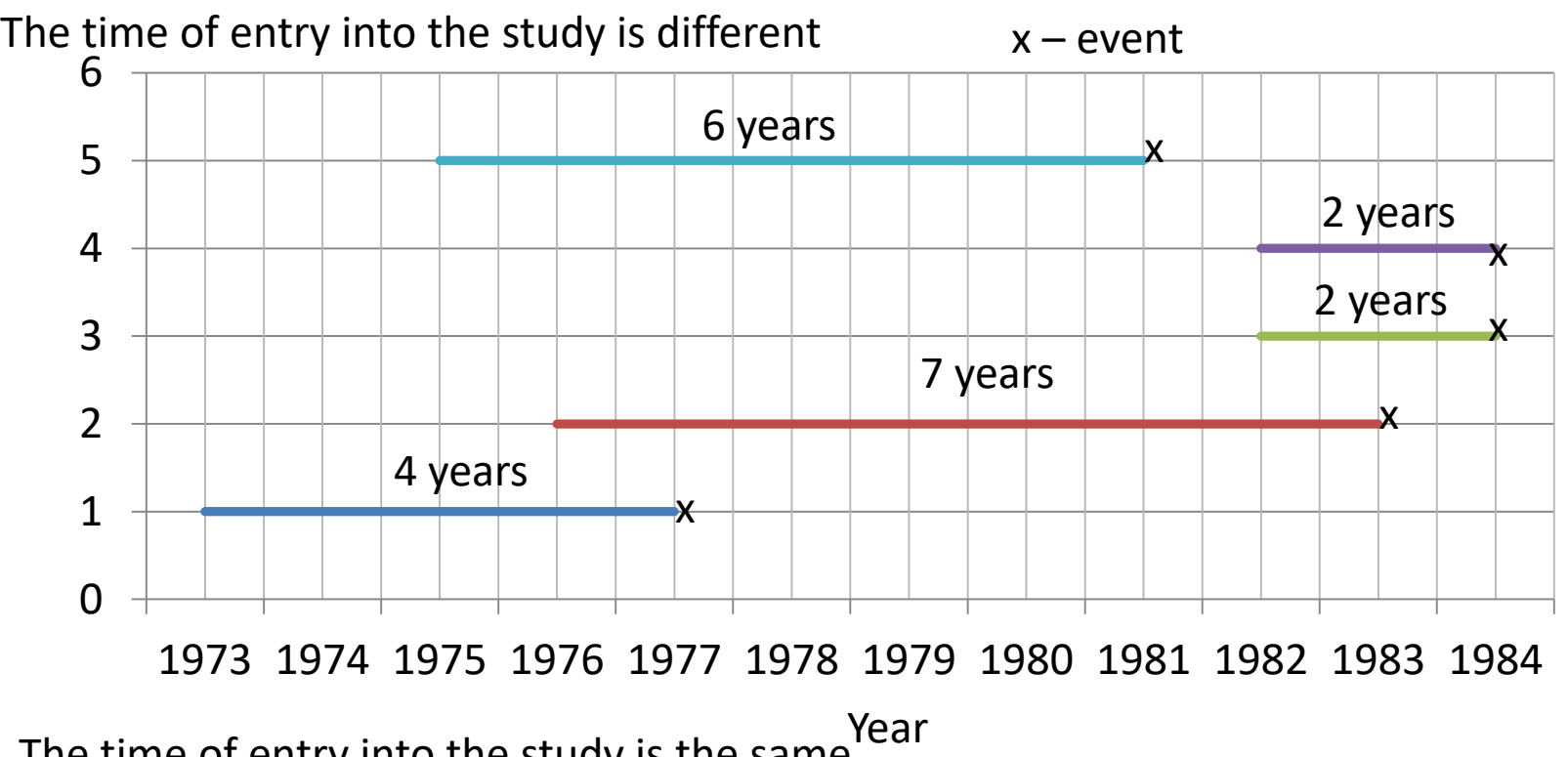
26 dialysis patients without PTX

event: mortality

interest: survival time

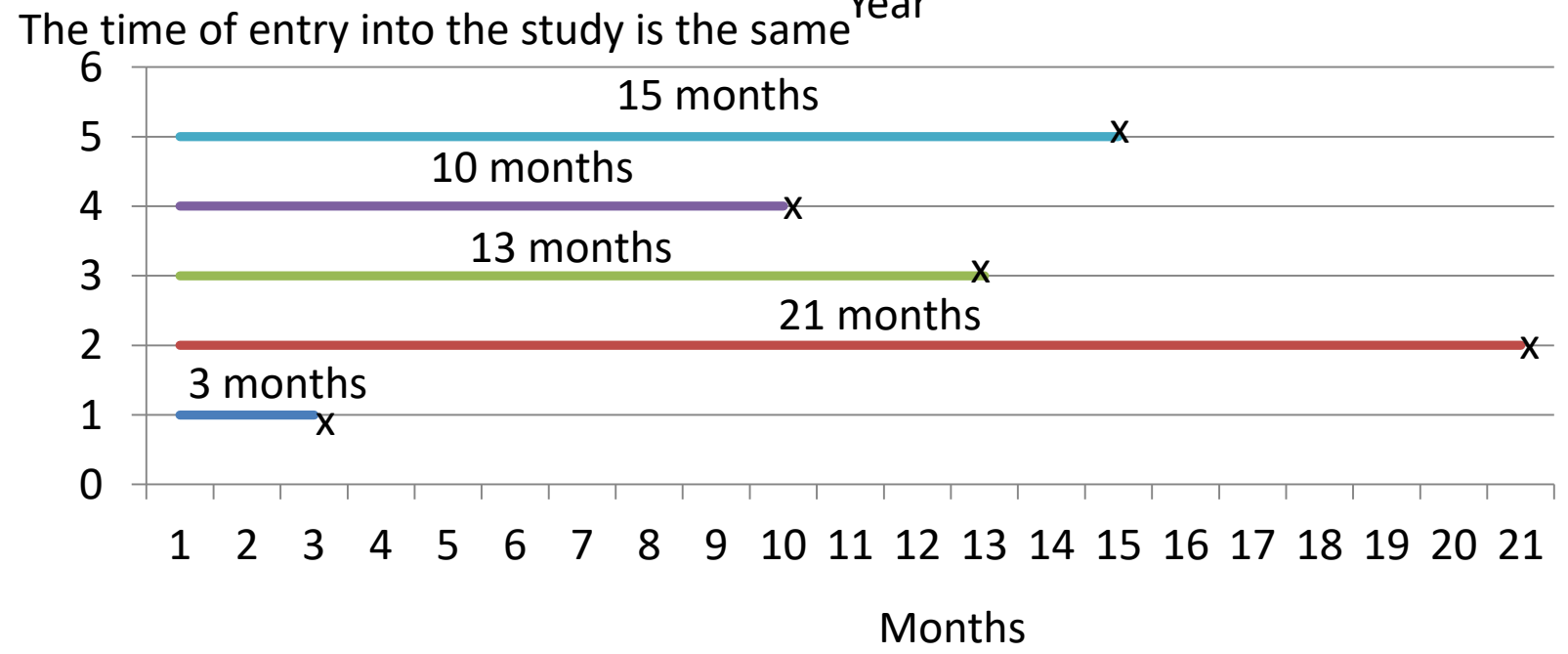
Question: Is PTX influencing the survival time?

Moldovan D, Racasan S, et al. Survival after parathyroidectomy in chronic hemodialysis patients with severe secondary hyperparathyroidism. Int Urol Nephrol. 2015 Nov;47(11):1871-7.



LELC scenario

patient A, B, C, E, F



PTX scenario

patient A, B, C, E, F

Survival study

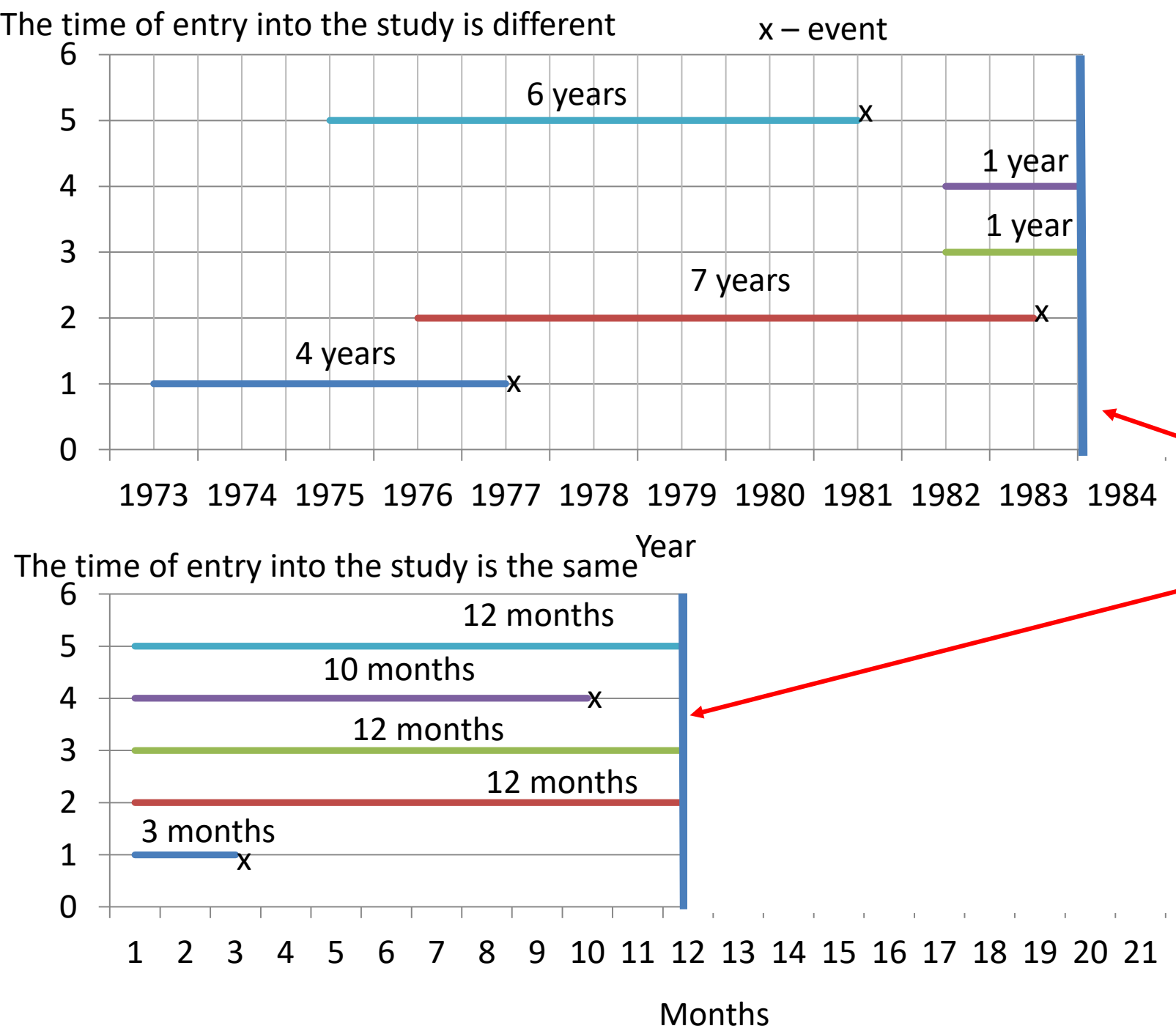
dependent variable

- time to event

  = survival time

Event

- negative
  - e.g. death, metastasis, relapse
- positive
  - e.g. cure, remission



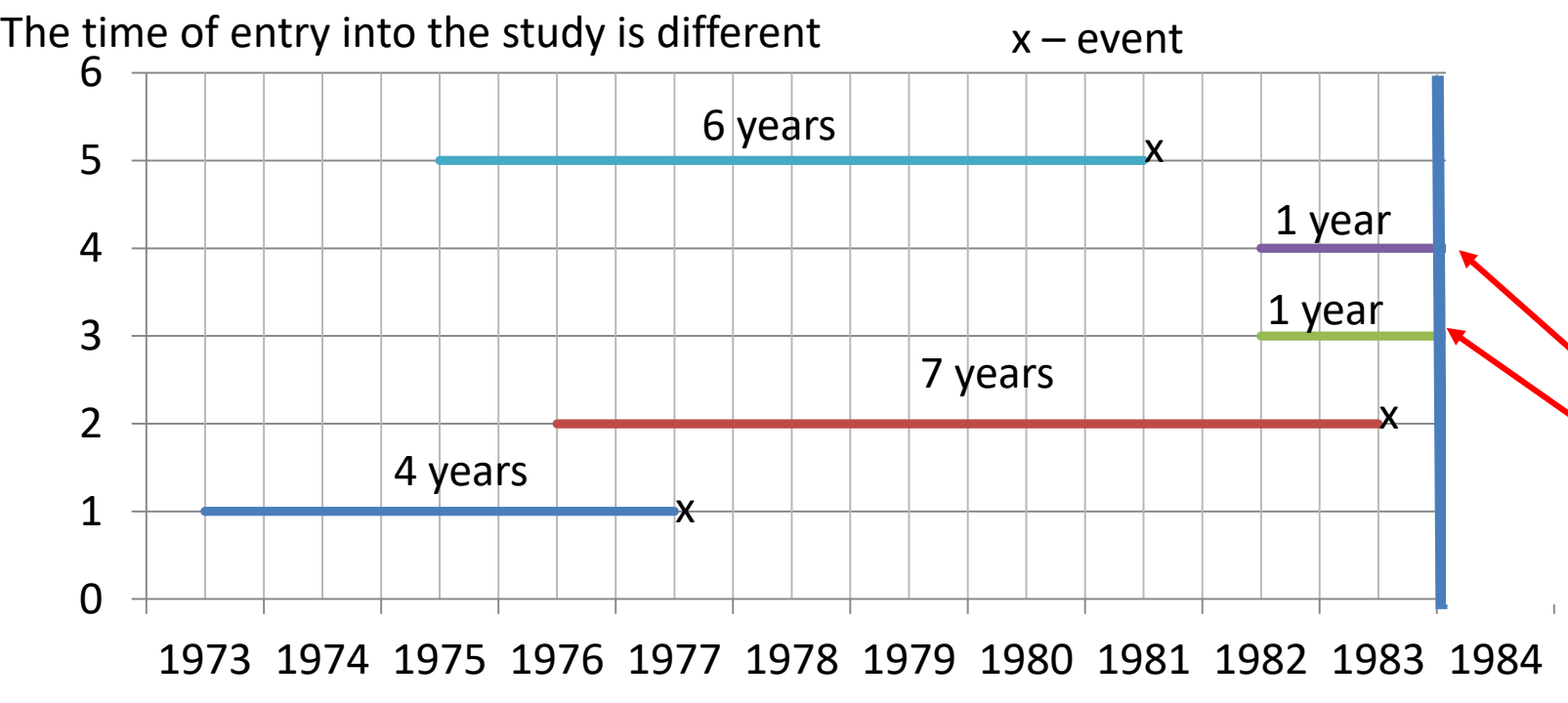
LELC scenario

patient A, B, C, E, F

researcher decide when to finish the study

PTX scenario

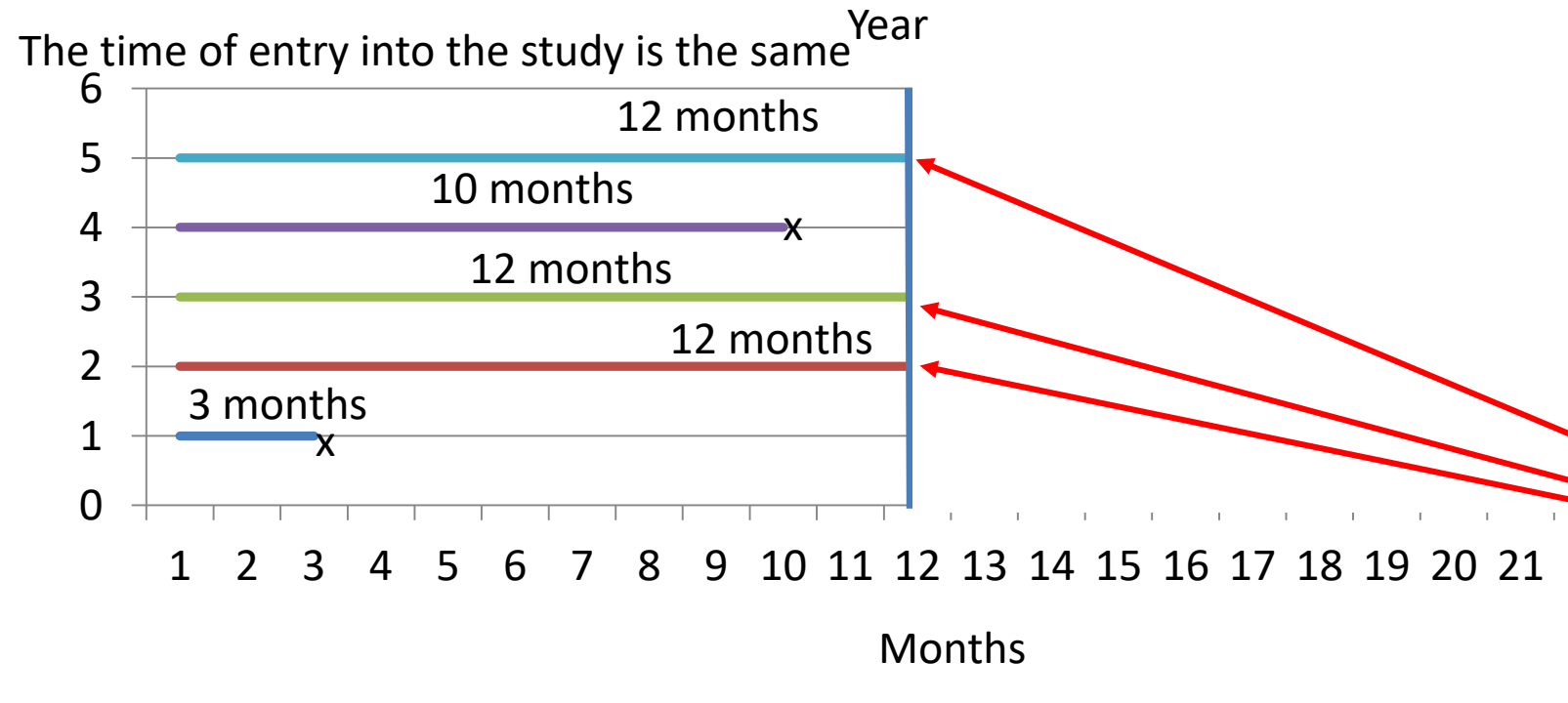
patient A, B, C, E, F



LELC scenario

patient A, B, C, E, F

2 subjects are alive



PTX scenario

patient A, B, C, E, F

3 subjects are alive

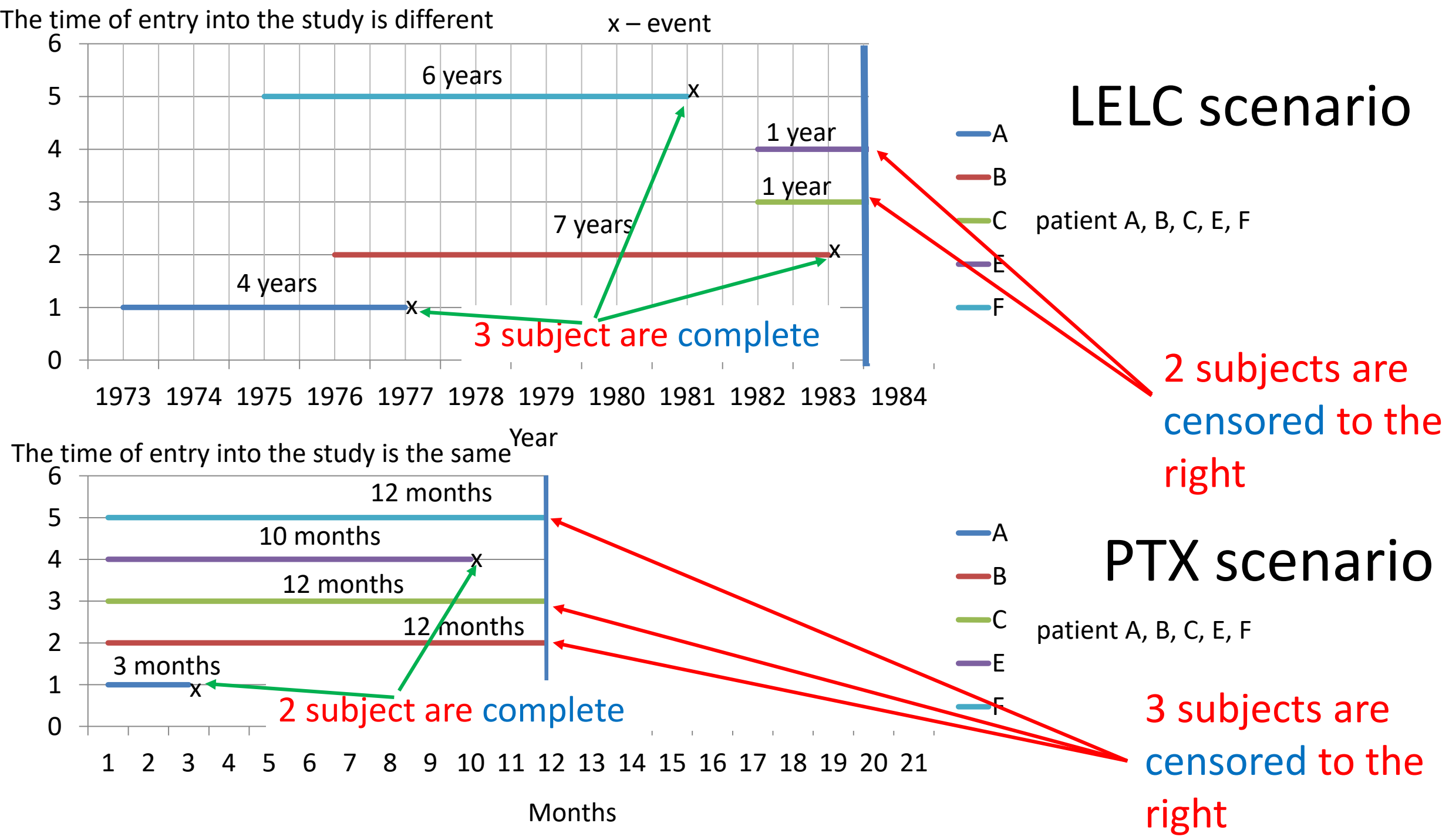
Event

complete status

- Complete observation
  - a person who experienced the event

incomplete status

- Censored observation
  - on the right
    - a person alive at the end of the study
    - a person lost from the study
      - lost from follow-up
      - whishes to quit
      - died from other causes
  - on the left
    - the event happen before the start of the study
  - interval-censoring
    - the exact time of an event is unknown, but it is known to have occurred within a specific time interval


The time of entry into the study is different
x – event
6 years
1 year
1 year
7 years
4 years
3 subject are complete
1973 1974 1975 1976 1977 1978 1979 1980 1981 1982 1983 1984
Year
A
B
C
E
F
patient A, B, C, E, F
LELC scenario
2 subjects are censored to the right
The time of entry into the study is the same
12 months
10 months
12 months
12 months
3 months
2 subject are complete
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
Months
A
B
C
E
F
patient A, B, C, E, F
PTX scenario
3 subjects are censored to the right

Data required for the analysis

	A	B	C	D	E	F	G
1	Id patient	Enter data	Event data	Lost from the study data	Event until May 1, 2023 (0-no, 1-with event, 2- lost from study)	Survival time to event/lost from the study	Censored
2	1	10/10/2020			0	31	right
3	2	5/4/2021	5/4/2022		1	12	completed
4	3	2/5/2020			0	36	right
5	4	4/11/2020	4/11/2022		1	24	completed
6	5	22/01/2023			0	4	right
7	6	10/3/2020			0	22	right
8	7	24/4/2021		24/4/2022	2	12	right
9	8	10/5/2022			0	11	right
10	9	9/9/2020	9/9/2021		1	12	completed
11	10	14/01/2023			0	4	right
12							
13	Study finish		1/5/2023				

- green – the doctor
- white – the statistician

Validity condition

- The moment of subject inclusion
  - the same for all the subjects
    - Ex. date of treatment initiation
    - Ex. date of diagnosis
- Drop from the study did not differ from those who remain
  - The probability of survival for the subjects who drop out is the same as for those who remain in the study

Survival analysis

a descriptive technique

- parameters
  - life table
  - median/mean survival time
  - 1-year, 3-year, 5-year survival rate
- graph
  - Kaplan–Maier survival curves

an inferential technique

- evaluate/predict the influence of a factor/therapy on the survival time
  - log-rank test
  - Cox regression analysis
- predict the influence of multiple factors/therapies on the survival time
  - Cox regression analysis

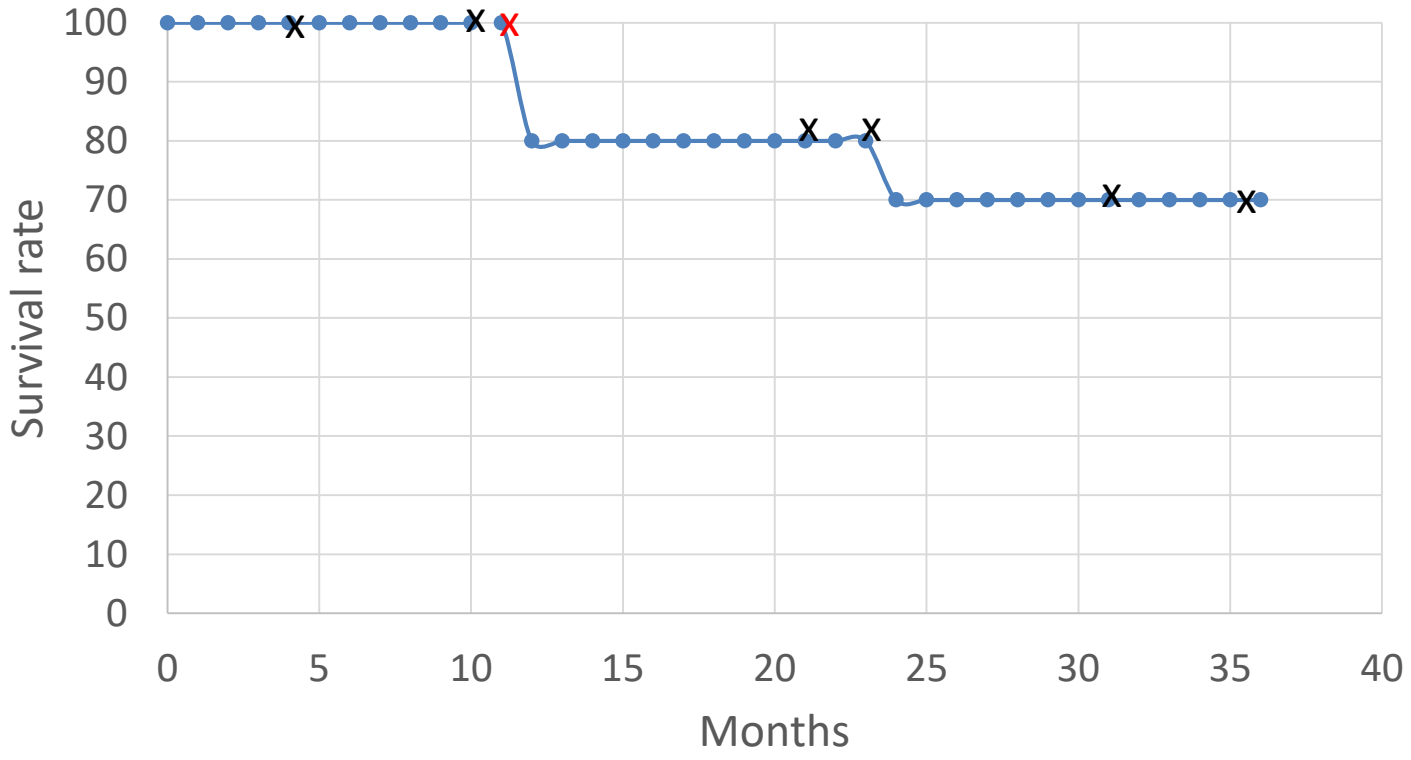
Survival curve

- the proportion of people alive at different times

	A	B	C	D	E	F	G
1	Id patient	Enter data	Event data	Lost from the study data	Event until May 1, 2023 (0-no, 1-with event, 2- lost from study)	Survival time to event/lost from the study	Censored
2	1	10/10/2020			0	31	right
3	2	5/4/2021	5/4/2022		1	12	completed
4	3	2/5/2020			0	36	right
5	4	4/11/2020	4/11/2022		1	24	completed
6	5	22/01/2023			0	4	right
7	6	10/3/2020			0	22	right
8	7	24/4/2021		24/4/2022	2	12	right
9	8	10/5/2022			0	11	right
10	9	9/9/2020	9/9/2021		1	12	completed
11	10	14/01/2023			0	4	right
12							
13	Study finish		1/5/2023				

Month	Survival rate (%)	
0	100	
1	100	
2	100	
3	100	
4	100	2 not completed
5	100	
6	100	
7	100	
8	100	
9	100	
10	100	
11	100	1 not completed
12	80	2 not completed
13	80	
14	80	
15	80	
16	80	
17	80	
18	80	
19	80	
20	80	
21	80	
22	80	1not completed
23	80	
24	70	1not completed

Survival curve

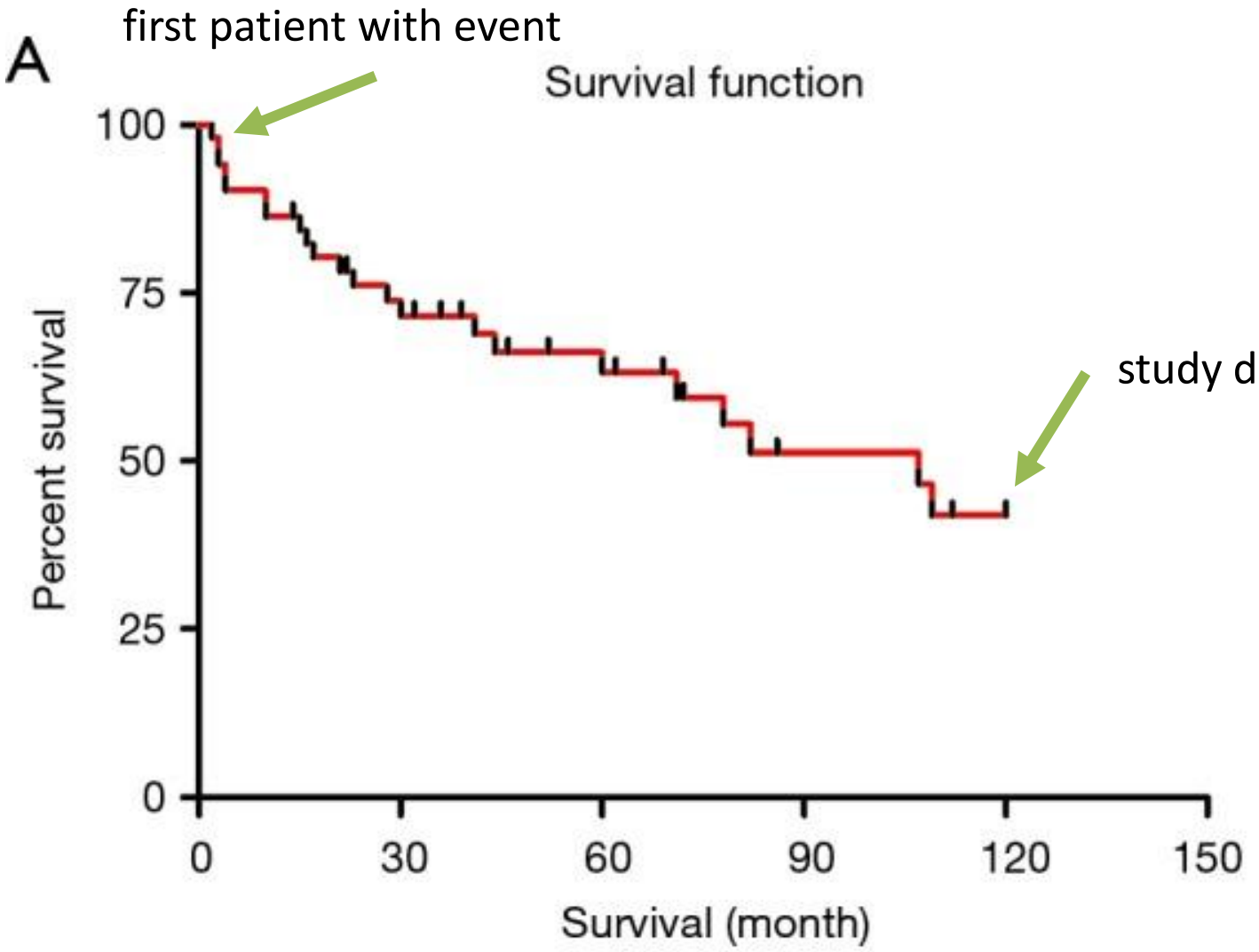


Month	Survival rate (%)	
0	100	
1	100	
2	100	
3	100	
4	100	2 not completed
5	100	
6	100	
7	100	
8	100	
9	100	
10	100	
11	100	1 not completed
12	80	2 not completed
13	80	
14	80	
15	80	
16	80	
17	80	
18	80	
19	80	
20	80	
21	80	
22	80	1not completed
23	80	
24	70	1not completed

...

x – status incomplet

Study LELC



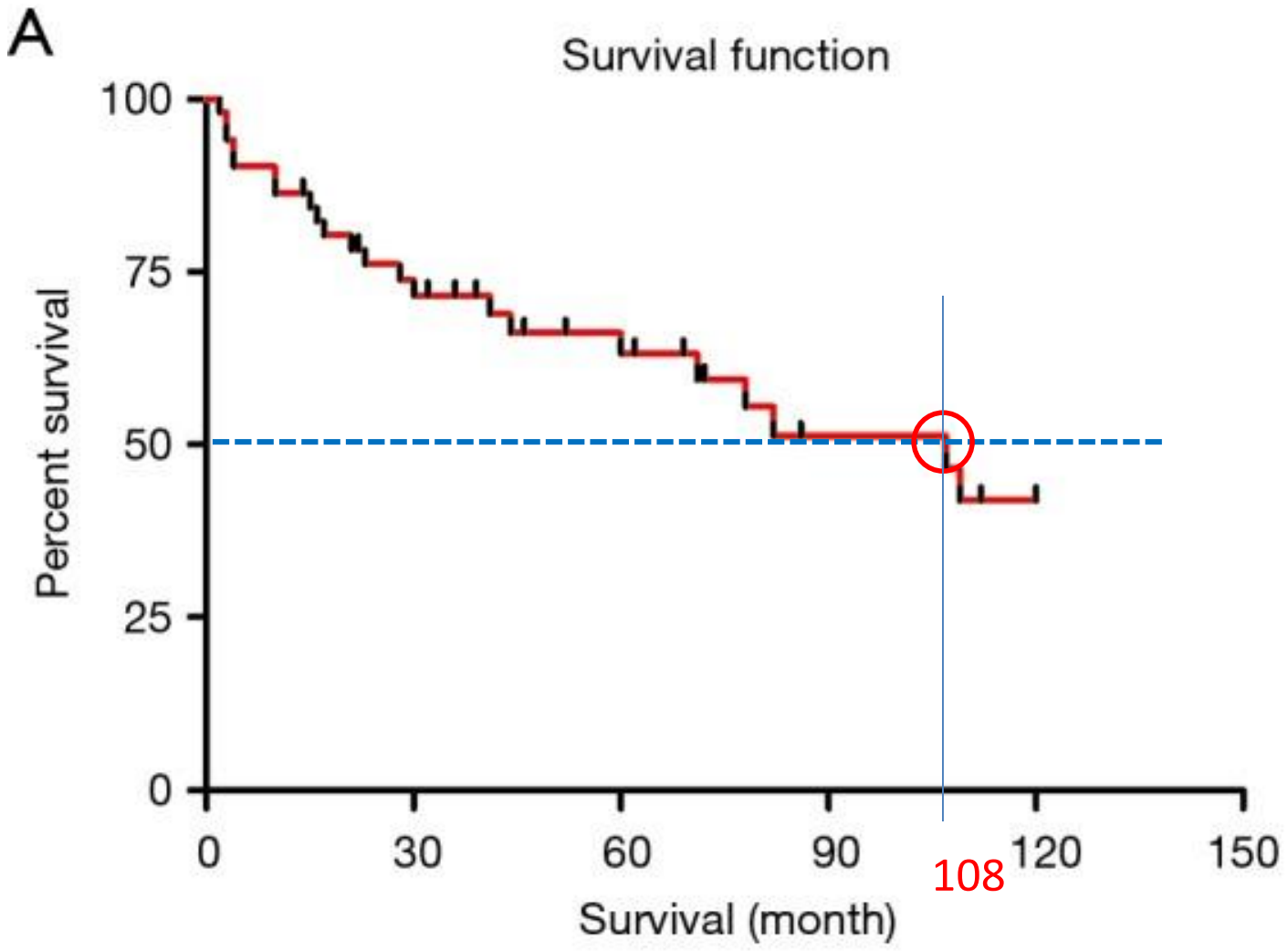
- (A) Survival curve for the patients with LELC

He J, Shen J, Pan H, Huang J, Liang W, He J. Pulmonary lymphoepithelioma-like carcinoma: a Surveillance, Epidemiology, and End Results database analysis. J Thorac Dis. 2015 Dec;7(12):2330-8.

Median time of survival

A value in time such that

- 50% of individuals survive less than this time,
- 50% of individuals survive more than this time



Median survival time

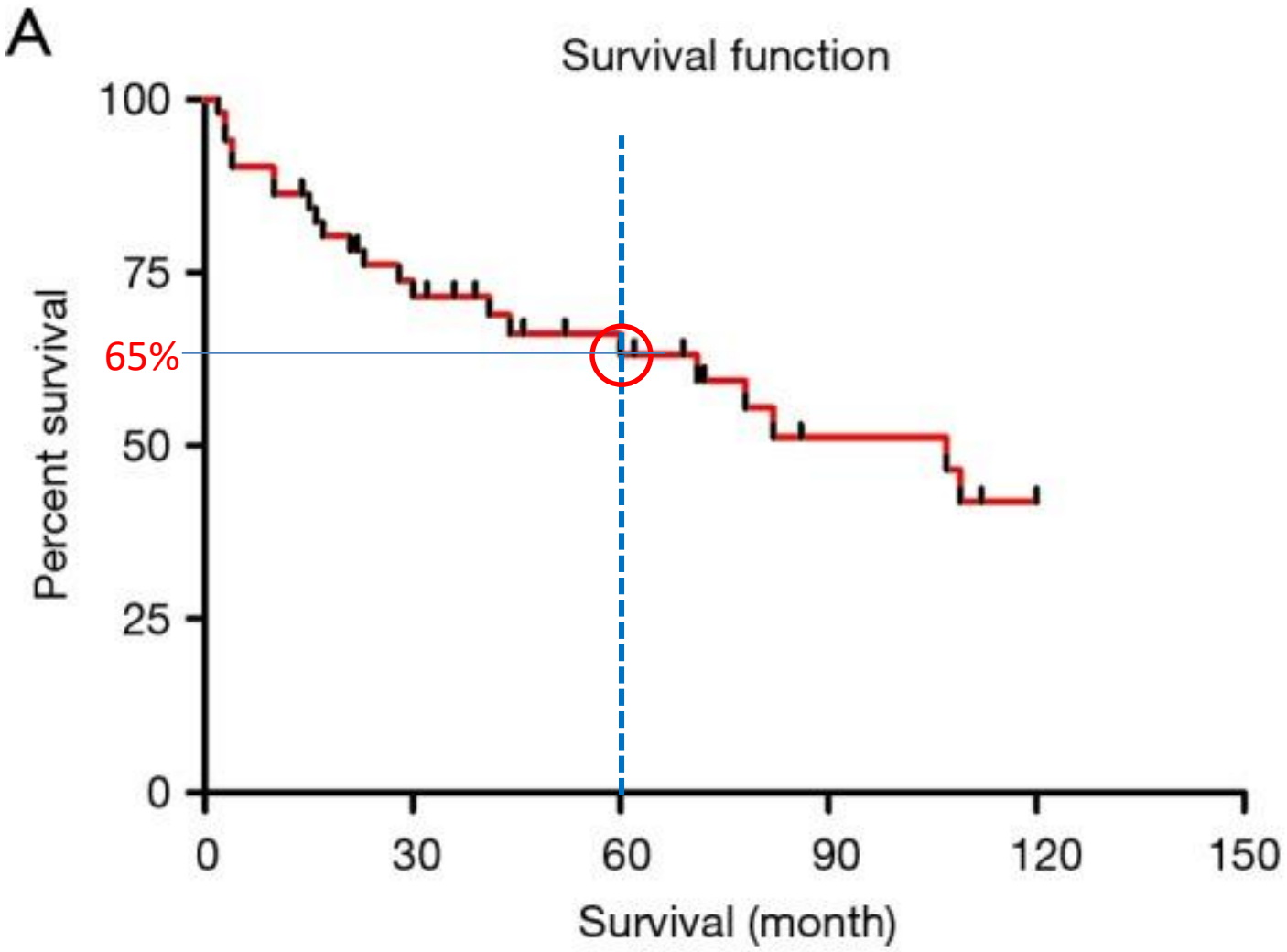
- Step 1. Draw a line parallel to the Ox axis at the 50% survival rate
- Step 2. Where the curve intersects the line, read the corresponding value from the Ox axis: in this case 108 months

- (A) Survival curve for the patients with LELC

He J, Shen J, Pan H, Huang J, Liang W, He J. Pulmonary lymphoepithelioma-like carcinoma: a Surveillance, Epidemiology, and End Results database analysis. J Thorac Dis. 2015 Dec;7(12):2330-8.

Survival rate at X months

- Percent of sujects without event at x months



Survival rate at 5 years = 60 months

- Step 1. Draw a line parallel to the Oy axis at 60 months of survival.
- Step 2. Where the curve intersects, read the corresponding value from the Oy axis: in this case 65%

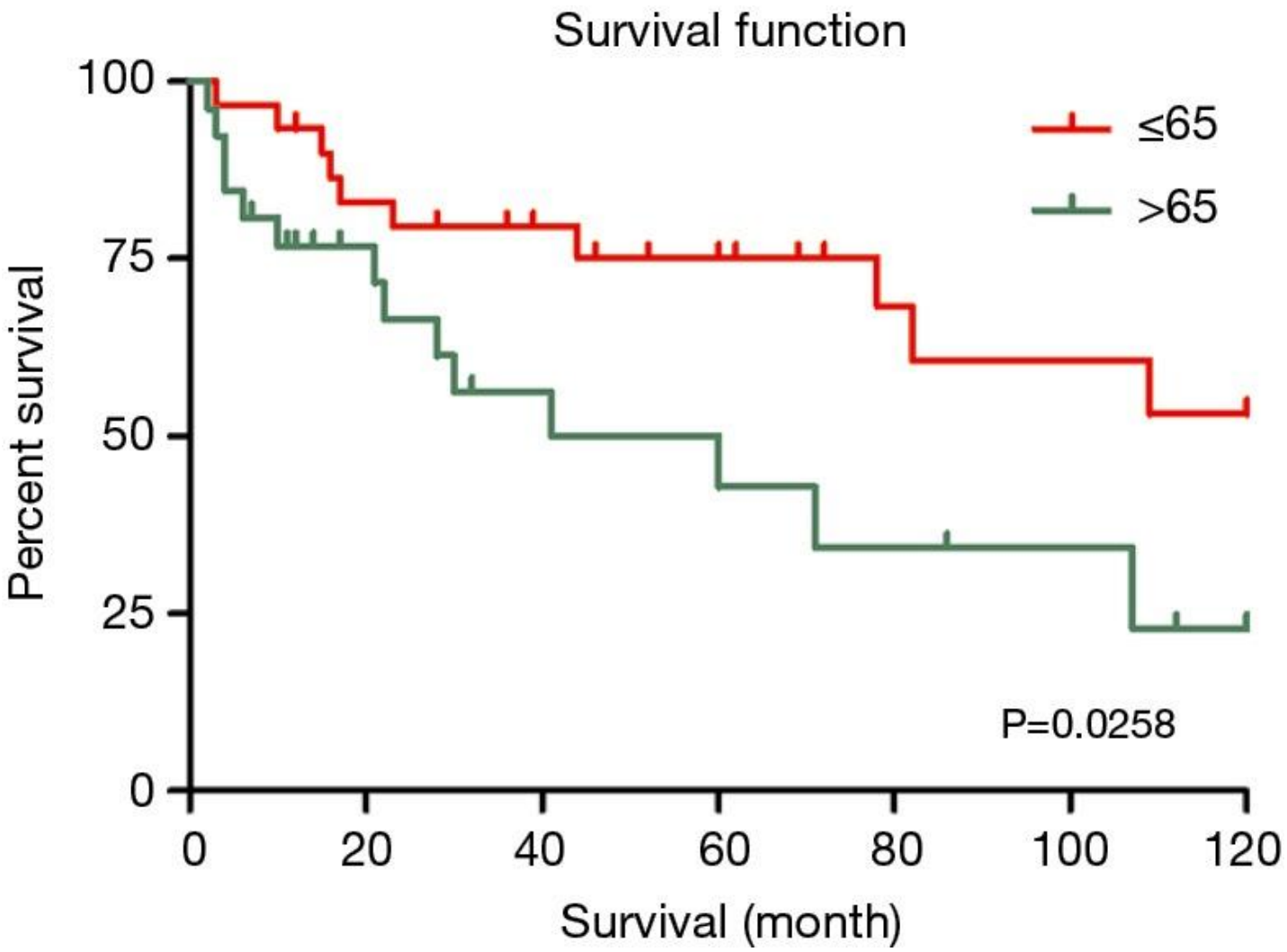
- (A) Survival curve for the patients with LELC

He J, Shen J, Pan H, Huang J, Liang W, He J. Pulmonary lymphoepithelioma-like carcinoma: a Surveillance, Epidemiology, and End Results database analysis. J Thorac Dis. 2015 Dec;7(12):2330-8.

Comparing two survival curves

- who survived more?
  - in case of
    - treatment with/without
    - risk factor present/absent

Comparing two survival curves

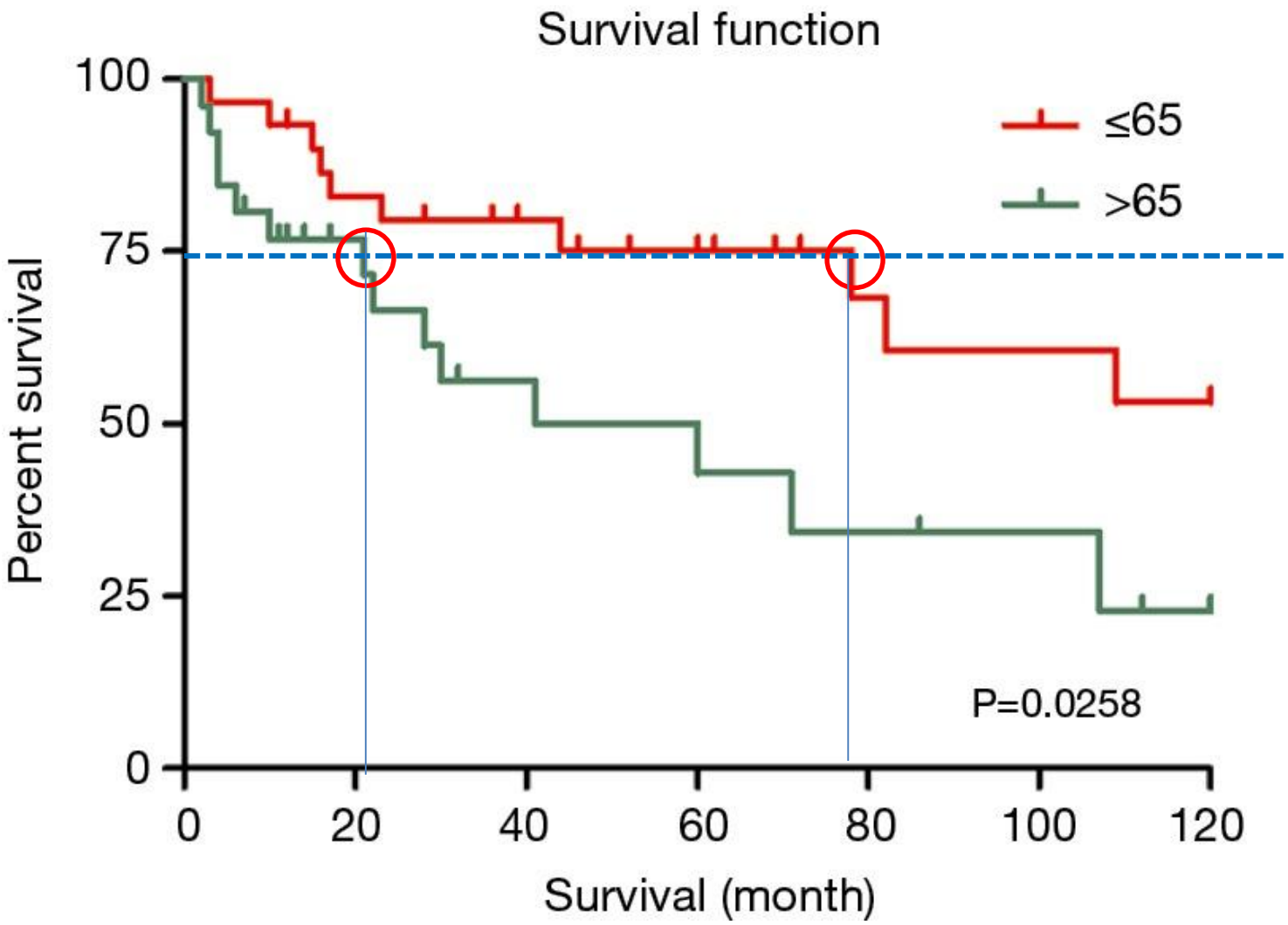


- ~percentile 25th: Those over 65 years survive 20 months compared to 80 months for those above 65 years old

- Survival curve for the patients with LELC for subjects ≥65 years old or less

He J, Shen J, Pan H, Huang J, Liang W, He J. Pulmonary lymphoepithelioma-like carcinoma: a Surveillance, Epidemiology, and End Results database analysis. J Thorac Dis. 2015 Dec;7(12):2330-8.

Comparing two survival curves

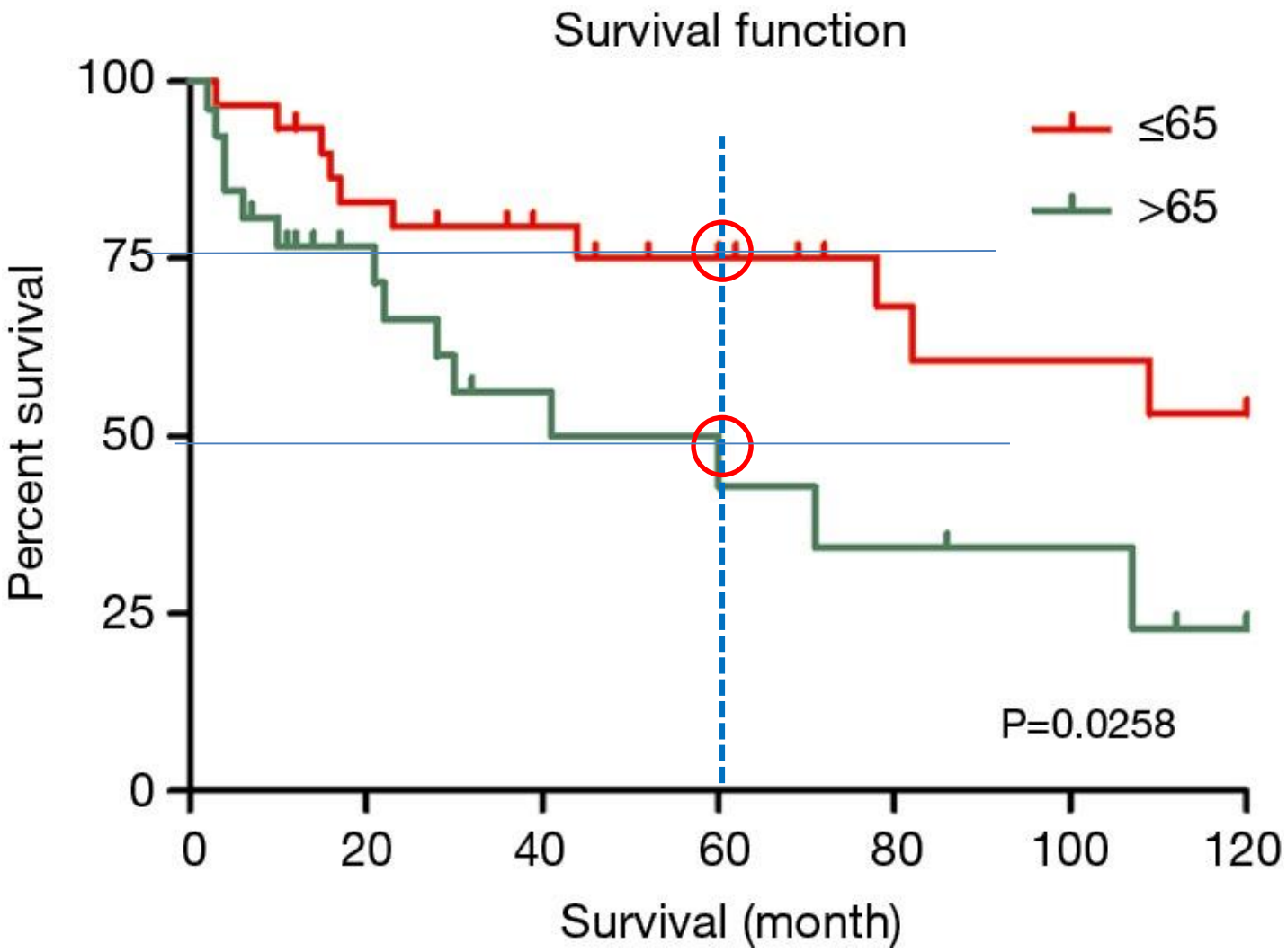


- ~percentile 25th: Those over 65 years survive 20 months compared to 79 months for those above 65 years old

- Survival curve for the patients with LELC for subjects ≥65 years old or less

He J, Shen J, Pan H, Huang J, Liang W, He J. Pulmonary lymphoepithelioma-like carcinoma: a Surveillance, Epidemiology, and End Results database analysis. J Thorac Dis. 2015 Dec;7(12):2330-8.

Comparing two survival curves



- ~at 60 months 50% of those over 65 survive compared to 75% of those under 65

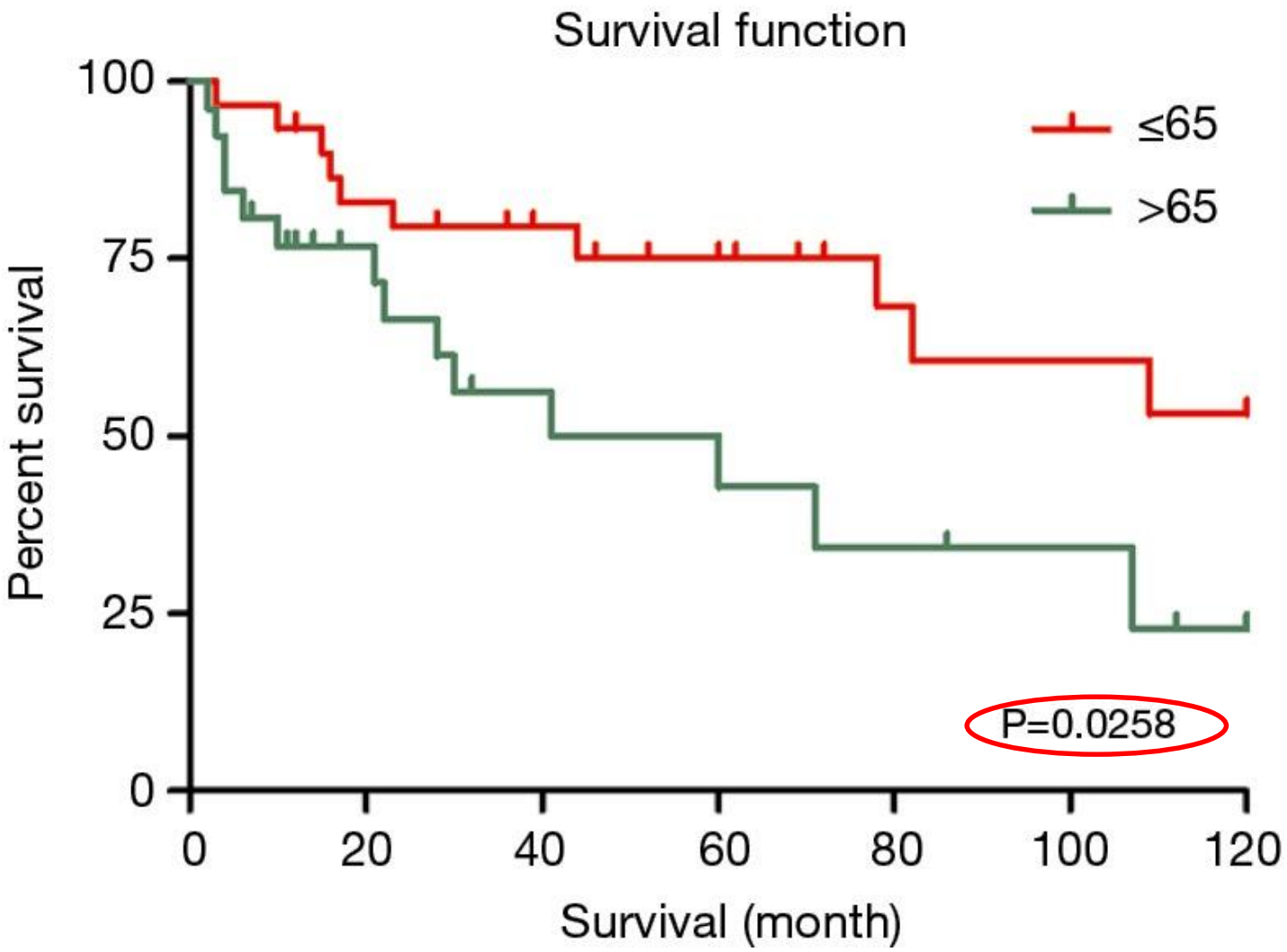
- Survival curve for the patients with LELC for subjects ≥65 years old or less

He J, Shen J, Pan H, Huang J, Liang W, He J. Pulmonary lymphoepithelioma-like carcinoma: a Surveillance, Epidemiology, and End Results database analysis. J Thorac Dis. 2015 Dec;7(12):2330-8.

Statistical tests used to compare survival curves

- Wilcoxon rank sum test
- The **Logrank Test**
- the Mantel–Haenszel chi-square statistic

Comparing two survival curves



- p=0.02<0.05 there is significant statistically difference between the survival time of the two groups

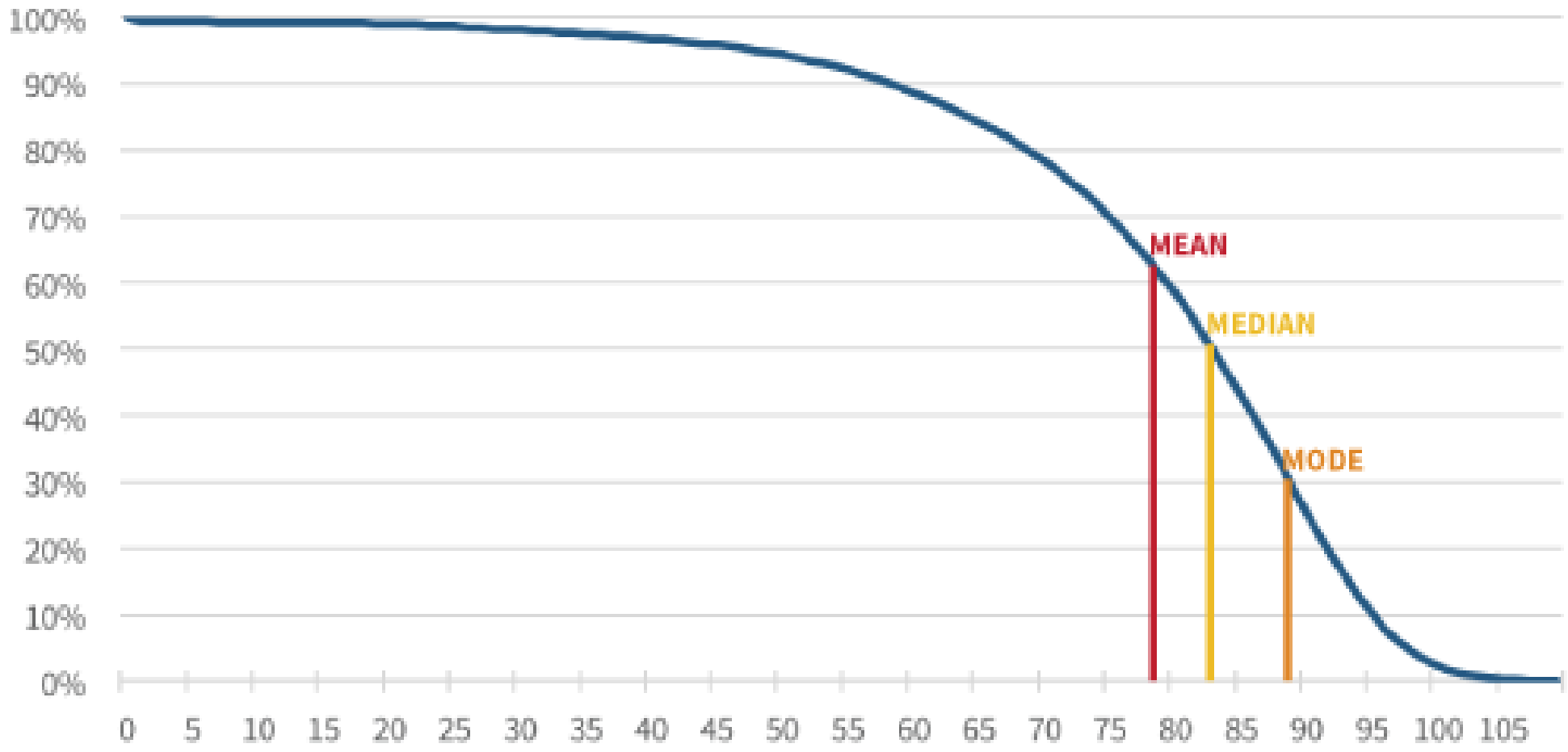
- Survival curve for the patients with LELC for subjects ≥65 years old or less

He J, Shen J, Pan H, Huang J, Liang W, He J. Pulmonary lymphoepithelioma-like carcinoma: a Surveillance, Epidemiology, and End Results database analysis. J Thorac Dis. 2015 Dec;7(12):2330-8.

Cox regression

- procedure for estimating survival time in a population

PROJECTED SURVIVAL CURVE FOR CHILDREN BORN IN 2014



Source: Human Mortality Database. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org. (Data downloaded on 12/1/2016)

Prediction: Cox Proportional Hazard Model

- If there are censored observations, then we do not know the survival time for all patients, but we can predict it
- A multivariate analysis (we analyze the influence of several factors on survival)
- Adjusting survival for confounding factors

Hazard rate

- $H=\frac{d}{\sum f+\sum c}$,

d = number of deaths

f = survival time to death

c = survival time to the end of the study for those who did not die

	A	B	C	D
1	Month	Number of patients	Number of events	Survival months
2	0	52	0	0
3	1	52	0	0
4	2	52	1	2
5	3	51	2	6
6	4	49	2	8
7	5	47	1	5
8	6	46	0	0
9	7	46	1	7
10	8	45	0	0
11	9	45	0	0
12	10	45	1	10
13	11	44	0	0
14	12	44	1	12
15	13	43	1	13
16	14	42	1	14
17	15	41	1	15

Hazard rate

- 40 (Column B at the end) patients alive at the end of the study
- Sum d (column C) =0+0+1+2+...+1=12 patients died
- Sum f (column D) = 0+0+2+6+…+14+15 = 92
- Sum c = 40*16 = 640

- $H=\frac{12}{92+640}$ =0.02 average deaths per month

d = number of deaths

f = survival time to death for persons who died

c = survival time to the end of the study for those who did not die

Hazard ratio

- the risk ratio at a given time for two groups, one exposed the other unexposed
- can be calculated for multiple risk factors

Hazard ratio – HR interpretation

- HR=1
  - The hazard rate is the same in both groups.
  - there is no difference in survival
- HR> 1
  - The hazard rate of the group coded 1 is greater than the hazard rate of the group coded 0 by x% = HR-1
- HR <1
  - The hazard rate of the group coded 1 is less than the hazard rate of the group coded 0 by x% = 1-HR

Regression equation

- Regression Cox :

$$h(t, X_1, X_2 \ldots X_5) = h_0(t)e^{b_1X_1 + b_2X_2 + \cdots + b_5X_5}$$

- the probability of dying in the next time interval given that the patient has lived up to this date
- X1, ... X5 the independent variables

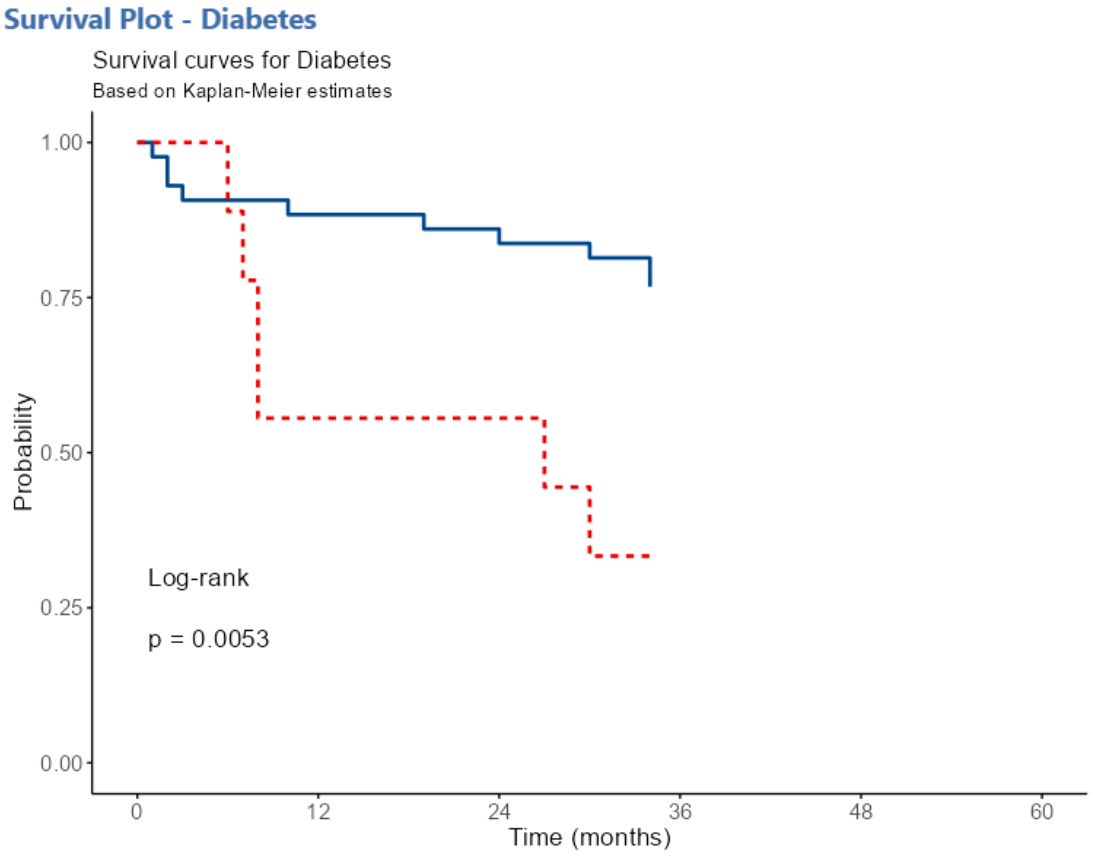
Months	Event	Age	Sex	Dialysis ti...	Diabetes
30	1	34	1	33	0
34	0	39	2	66	0
34	0	40	2	100	0
34	0	41	1	91	0
34	0	44	2	41	0
34	0	44	2	24	0
34	0	44	2	117	0
30	1	45	1	23	1
7	1	46	2	14	1
2	1	47	2	23	0
34	0	47	1	17	0
34	0	49	1	38	0
34	0	49	1	32	0
34	0	50	1	15	0
34	1	50	1	100	0
8	1	52	1	25	1
34	0	52	1	39	0
27	1	54	2	10	1
34	1	54	1	39	0
34	0	54	1	44	0
34	0	54	1	13	1
24	1	55	1	168	0
2	1	55	1	29	0
19	1	55	2	55	0

Scenario

- patient with diabetes and dyalisis
- which factors influence the survival time

Cox Table- Diabetes

Explanatory	Levels	all	HR (Univariable)
Diabetes	0	43 (82.7)	-
	1	9 (17.3)	3.91 (1.40-10.89, p=0.009)



HR=3.91 95%CI 1.40-10.89, p=0.009

The hazard of coded group 1 is higher than the hazard of coded group 0 with HR-1=2.91; 290% ~ 3 times higher

The risk to die in the group with diabetes is 3 times higher than in the group without diabetes

Confidence interval is large – study is imprecise

The risk is of unclear importance (1.40 is near 1)

There is significant statistically difference between the survival time in subjects with diabetes and without

Hazard ratio

Cox Table- Diabetes

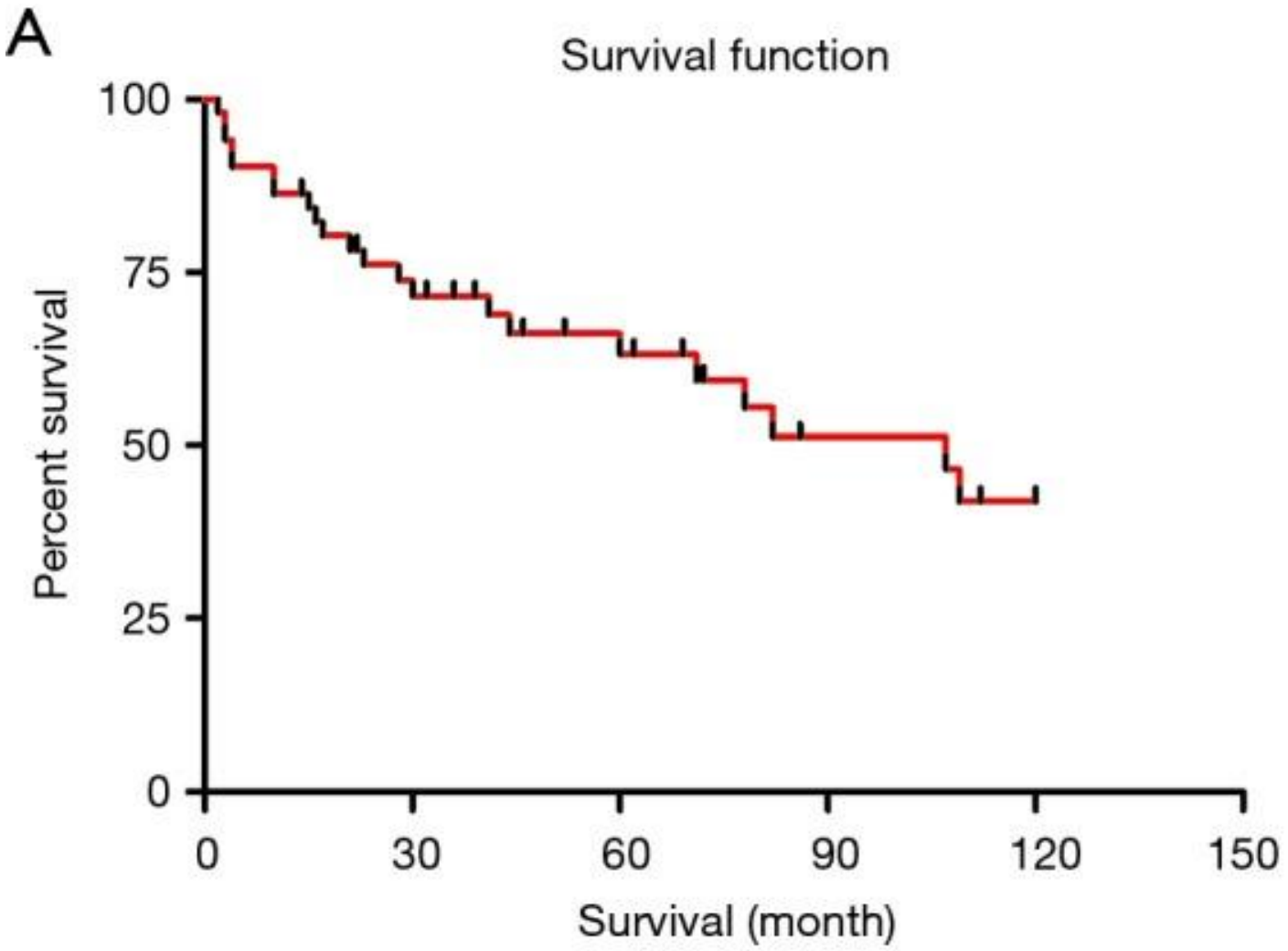
Explanatory	Levels	all	HR (Univariable)
Diabetes	0	43 (82.7)	-
	1	9 (17.3)	3.91 (1.40-10.89, p=0.009)

- The risk of death at any time in diabetes group is approximately 3 times the risk in group without diabetes.
- We assumed that the hazard or risk of death is the same over time (our starting assumption – the main requirement in Cox regression).

Scenario 1

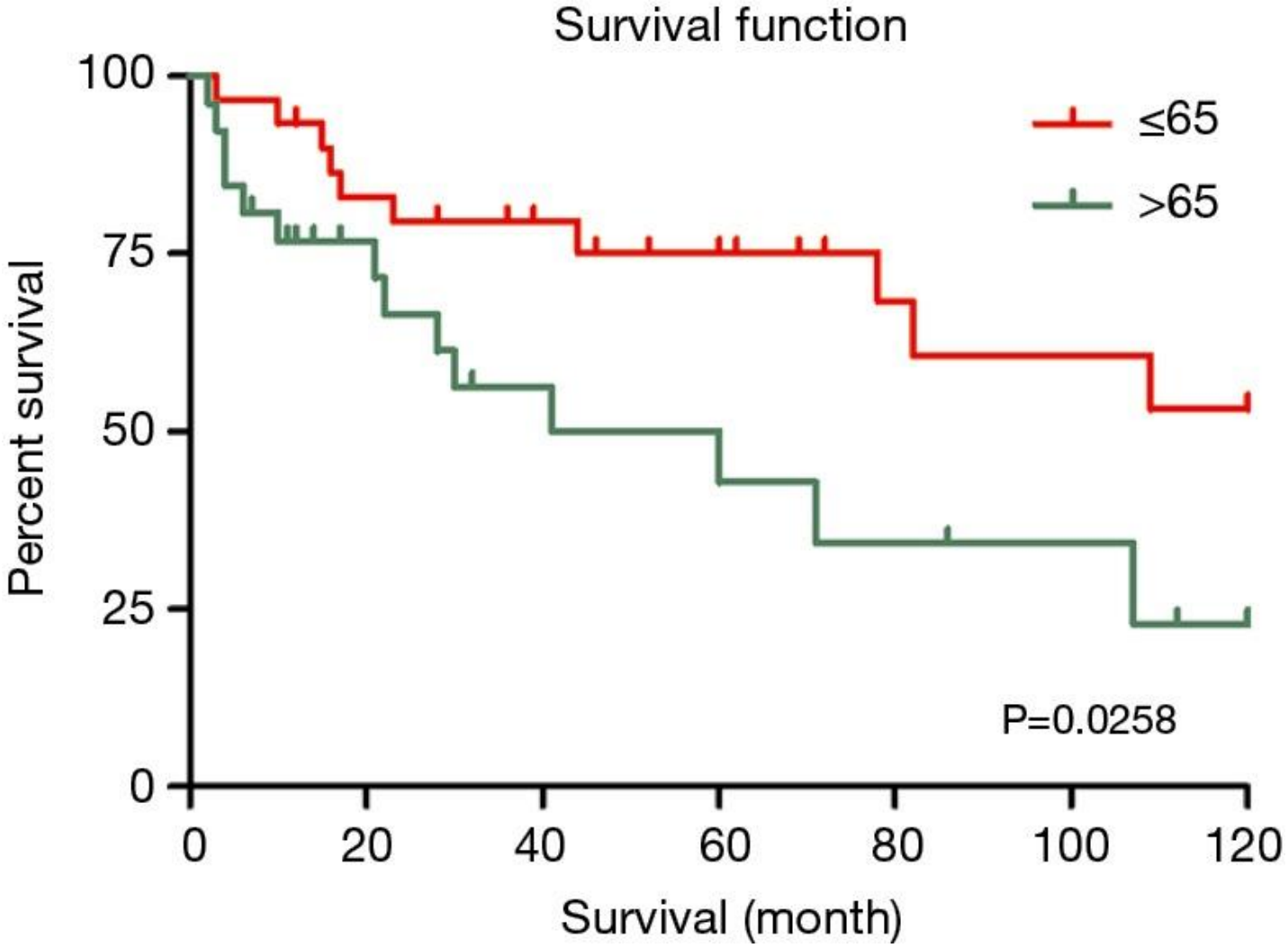
- Lung cancer with lymphoepithelial carcinoma (LELC) – a rare type of cancer. All patients were extracted from US databases from 1973 to 2011. A total of 62 patients. Median survival was 107 months. Survival rates at 1, 3 and 5 years were 85.6%, 74.5% and 55.2%.

He J, Shen J, Pan H, Huang J, Liang W, He J. Pulmonary lymphoepithelioma-like carcinoma: a Surveillance, Epidemiology, and End Results database analysis. J Thorac Dis. 2015 Dec;7(12):2330-8.



- (A) Survival of patients with LELC

He J, Shen J, Pan H, Huang J, Liang W, He J. Pulmonary lymphoepithelioma-like carcinoma: a Surveillance, Epidemiology, and End Results database analysis. J Thorac Dis. 2015 Dec;7(12):2330-8.



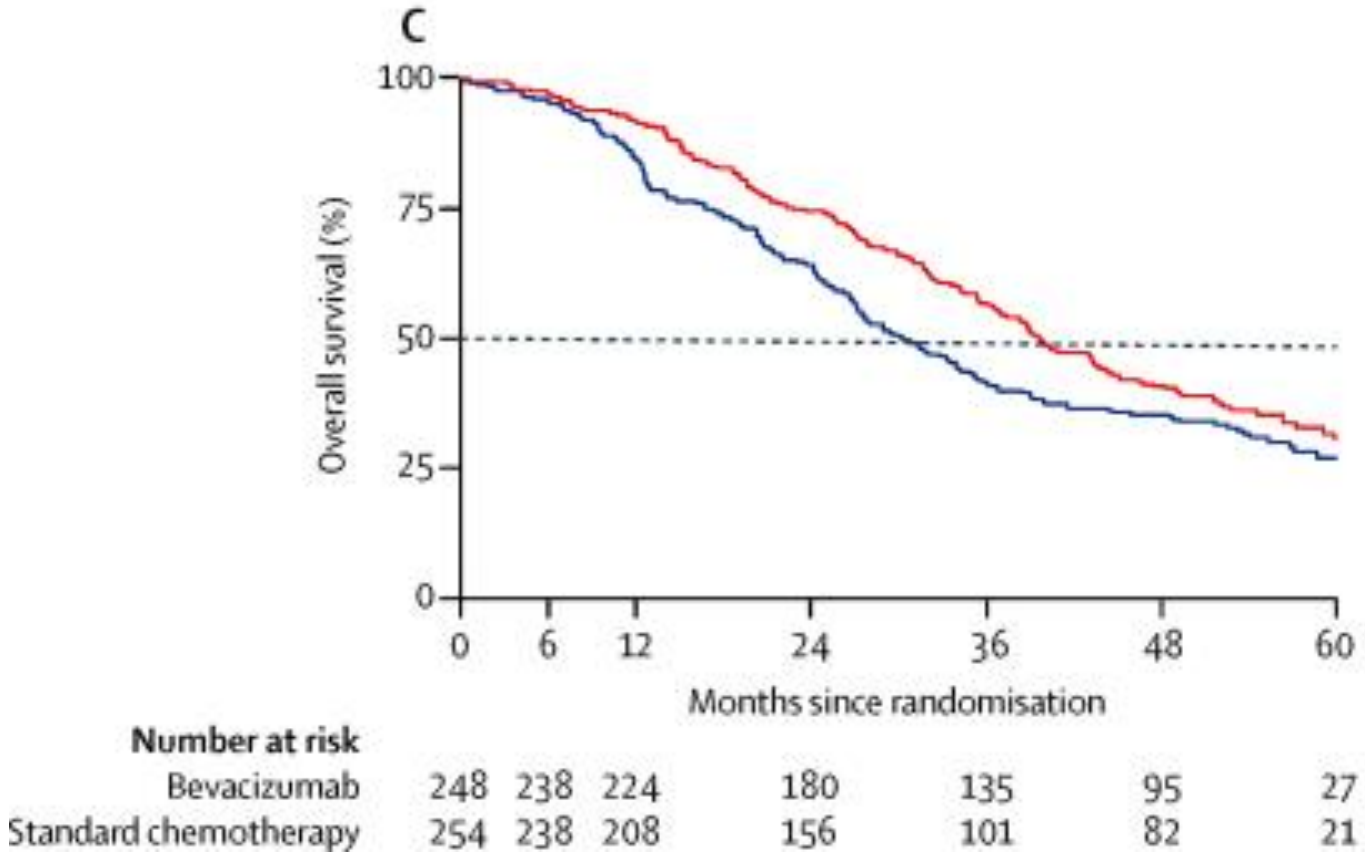
- Survival of patients with LELC by age >65 versus ≤65 years

He J, Shen J, Pan H, Huang J, Liang W, He J. Pulmonary lymphoepithelioma-like carcinoma: a Surveillance, Epidemiology, and End Results database analysis. J Thorac Dis. 2015 Dec;7(12):2330-8.

Scenario 2

- ICON7 – international phase 3 clinical trial, in 11 countries. 1528 women with newly diagnosed ovarian cancer were enrolled in the study. They were assigned to standard chemotherapy or standard chemotherapy plus bevacizumab. They were followed for an average of 48 months

Oza AM, Cook AD, Pfisterer J, Embleton A, Ledermann JA, Pujade-Lauraine E, et al. Standard chemotherapy with or without bevacizumab for women with newly diagnosed ovarian cancer (ICON7): overall survival results of a phase 3 randomised trial. Lancet Oncol. 2015 Aug;16(8):928-36.



- Kaplan-Meier survival curves for the two groups

Oza AM, Cook AD, Pfisterer J, Embleton A, Ledermann JA, Pujade-Lauraine E, et al. Standard chemotherapy with or without bevacizumab for women with newly diagnosed ovarian cancer (ICON7): overall survival results of a phase 3 randomised trial. Lancet Oncol. 2015 Aug;16(8):928-36.

Scenario 3

The objective of the study was to compare mortality between a group of 26 dialysis patients with hyperparathyroidism and parathyroidectomy (PTX) with a group of 26 dialysis patients with hyperparathyroidism, without PTX intervention, but with drug treatment. The study was observational.

Moldovan D, Racasan S, Kacso IM, Rusu C, Potra A, Bondor C, Patiu IM, Gherman-Căprioară M. Survival after parathyroidectomy in chronic hemodialysis patients with severe secondary hyperparathyroidism. Int Urol Nephrol. 2015 Nov;47(11):1871-7.

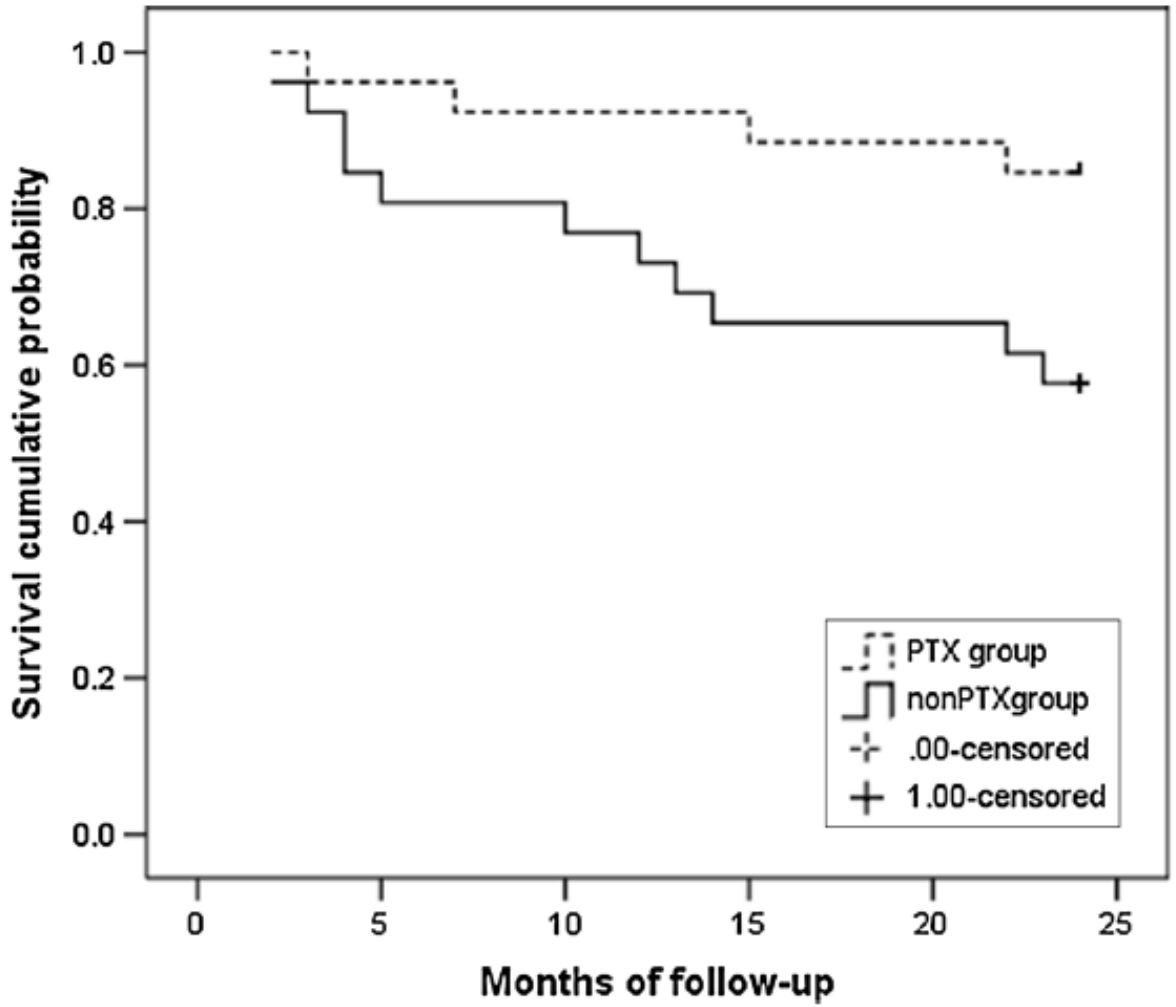


Fig. 1 Kaplan–Meier survival curve illustrates 24-month survival in PTX versus non-PTX group. Survival rate was 84.6 % in PTX group and 53.8 % in non-PTX group ($p = 0.01$)

Table 5 Difference in mortality between PTX and non-PTX groups: multivariate analysis

Parameter	Univariate analysis[a]		Multivariate analysis[b]	
	HR (95 % CI)	*p*	HR (95 % CI)	*p*
Intervention (HR for non-PTX)	3.27 (1.04–10.3)	**0.043**	3.50 (1.09–11.18)	**0.035**
Age (years)	1.03 (0.98–1.08)	0.310	1.06 (0.90–1.18)	0.341
Gender	3.37 (0.95–11.95)	0.060	2.97 (0.98–10.45)	0.122
Presence of DM	3.63 (1.31–10.03)	**0.013**	3.85 (1.34–11.02)	**0.012**
Presence of HTN	1.44 (0.46–4.53)	0.531	1.22 (0.42–3.63)	0.643
Presence of cardiac failure	0.92 (0.33–2.53)	0.869	0.82 (0.45–1.5)	0.688
HD vintage (months)	1.00 (0.99–1.00)	0.334	1.00 (0.99–1.80)	0.243
iPTH (pg/ml)	1.00 (1.00–1.00)	0.168	1.02 (0.90–1.10)	0.230
Ca (mg/dl)	0.73 (0.44–1.22)	0.231	0.71 (0.38–1.18)	0.308
P (mg/dl)	1.20 (0.92–1.56)	0.180	1.24 (0.89–1.52)	0.228
ALP (U/l)	1.00 (1.00–1.00)	0.112	1.00 (1.00–1.00)	0.112
Albumin (g/dl)	0.43 (0.09–2.08)	0.294	0.51 (0.12–2.12)	0.314
Hb (g/dl)	0.74 (0.52–1.07)	0.111	0.70 (0.50–1.14)	0.123
CRP (mg/dl)	1.87 (1.23–2.83)	**0.030**	1.88 (1.11–3.19)	**0.019**

HR hazard ratio, *CI* confidence interval, *PTX* parathyroidectomy, *DM* diabetes mellitus, *HTN* hypertension, *HD* hemodialysis, *Ca* calcium, *P* phosphorus, *ALP* alkaline phosphatase, *Hb* hemoglobin, *CRP* C-reactive protein

[a] Cox univariate regression

[b] Cox multivariate regression (ENTER method) with intervention, DM and CRP as independent predictors

- Multumesc !