Proportions in medical research

Stephen Kerr

Summary

- How do we commonly use proportions in medical research?

- Relative risk
- Risk difference
- Odds ratio
- NNT/NNH

- Chi-square and Fisher’s exact test – Formal comparisons

- Multivariate regression models to compare proportions

Summary Survival (time to event) analysis

- What is survival analysis?
- Estimating the survival function (life tables and Kaplan-Meier curves)
- The log rank test
- Cox Proportional Hazard regression (multivariate analysis)
- Univariate screening
- Building up a multivariate model
- (Assessing the proportional hazard's assumption)
- Poisson regression
- Competing risks regression

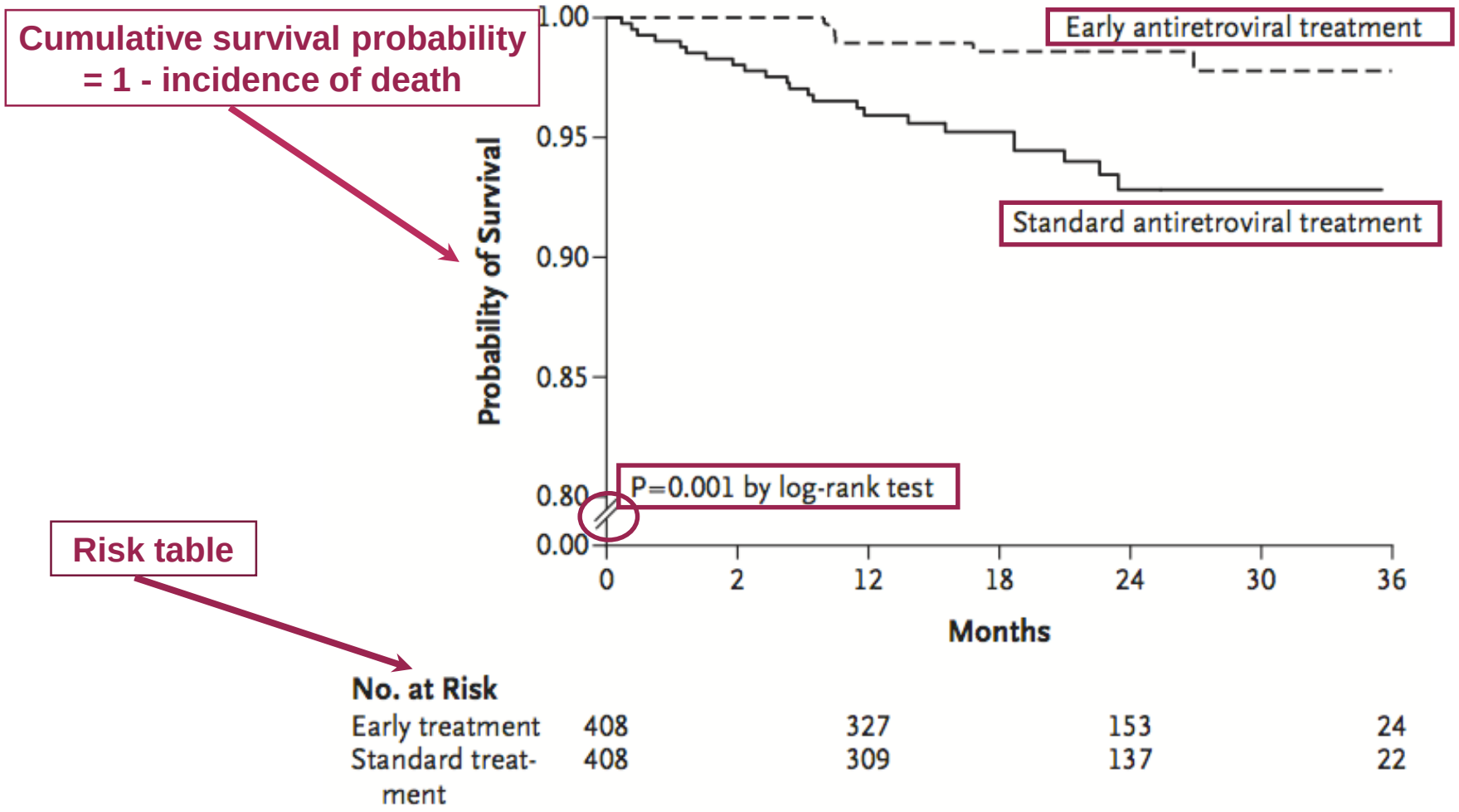
What is survival analysis?

- Statistical methods for analysing longitudinal data on the occurrence of events.
- Events (Binary variables) may include:
  - death, injury, onset/recovery from illness, drug toxicity or changing above or below a clinical threshold of a meaningful continuous variable (eg CD4 counts)
- Often used for data from randomised clinical trials or cohort studies

What are we trying to estimate?

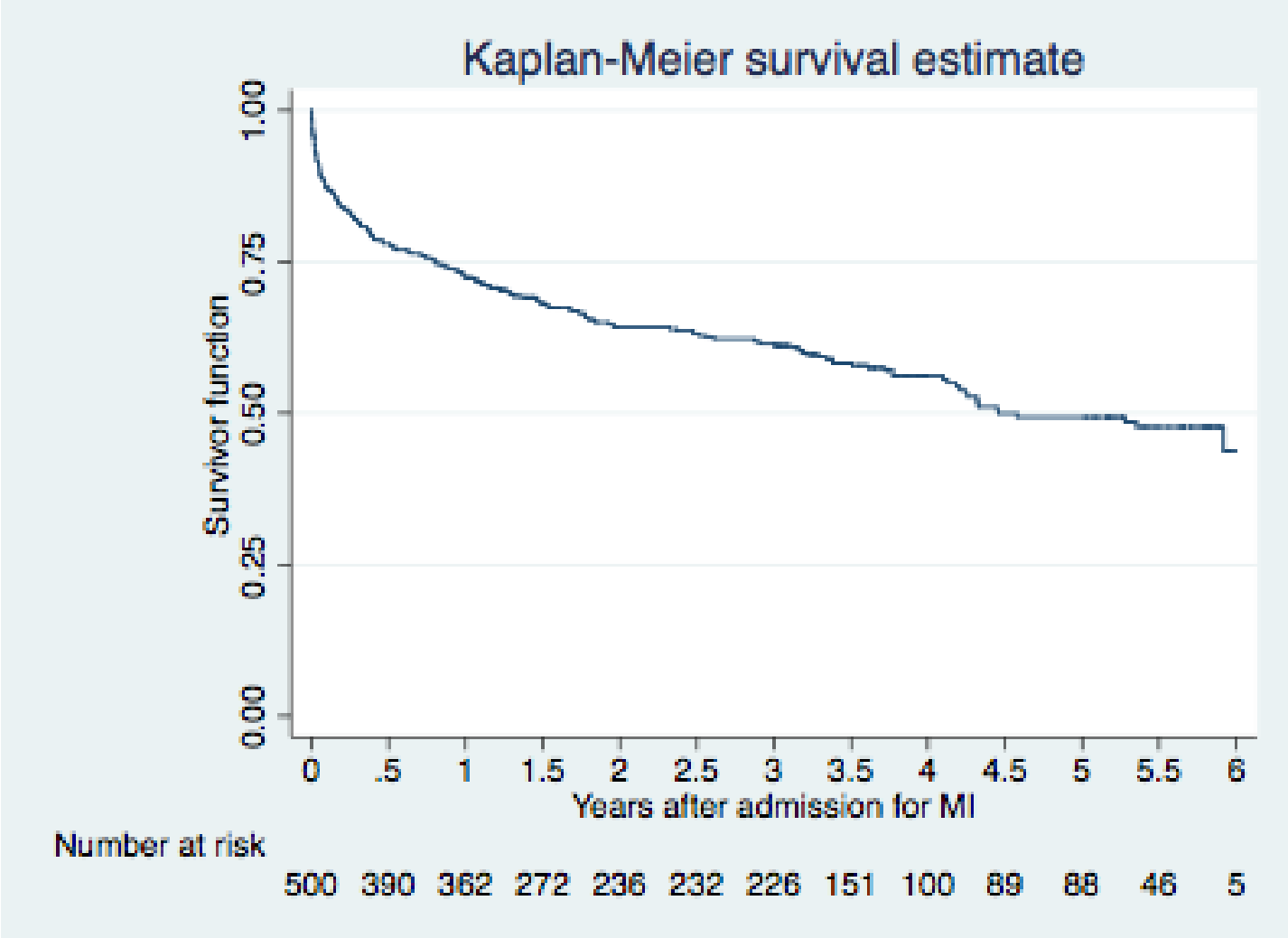
- S(t) = Survivor function
- S(t) = Proportion surviving at least to time t or beyond

Early versus Standard Antiretroviral Therapy for HIV-Infected Adults in Haiti - NEJM 2010; 363:257-265



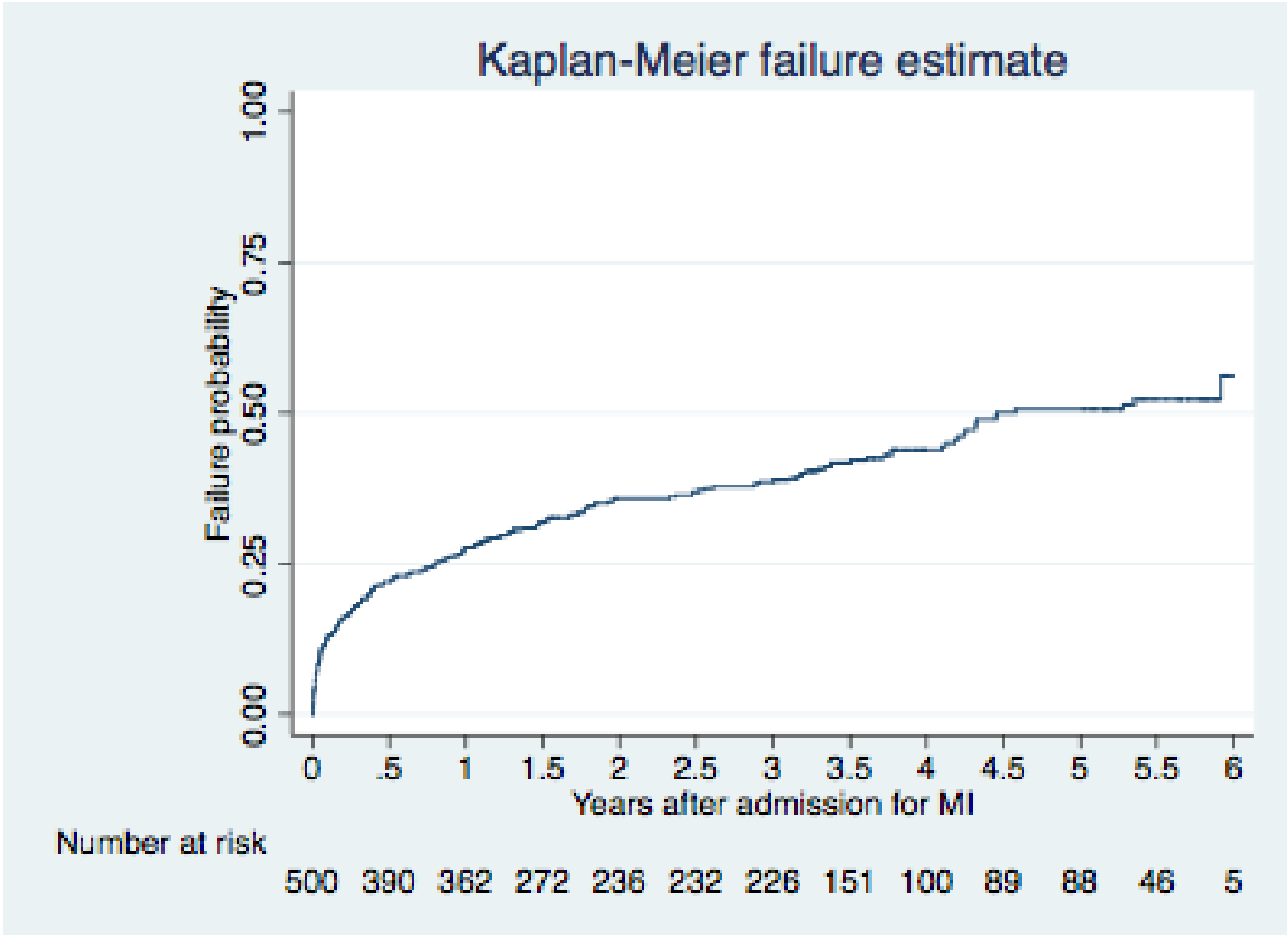
Survivor function

S(t)



Failure function

1 - S(t)

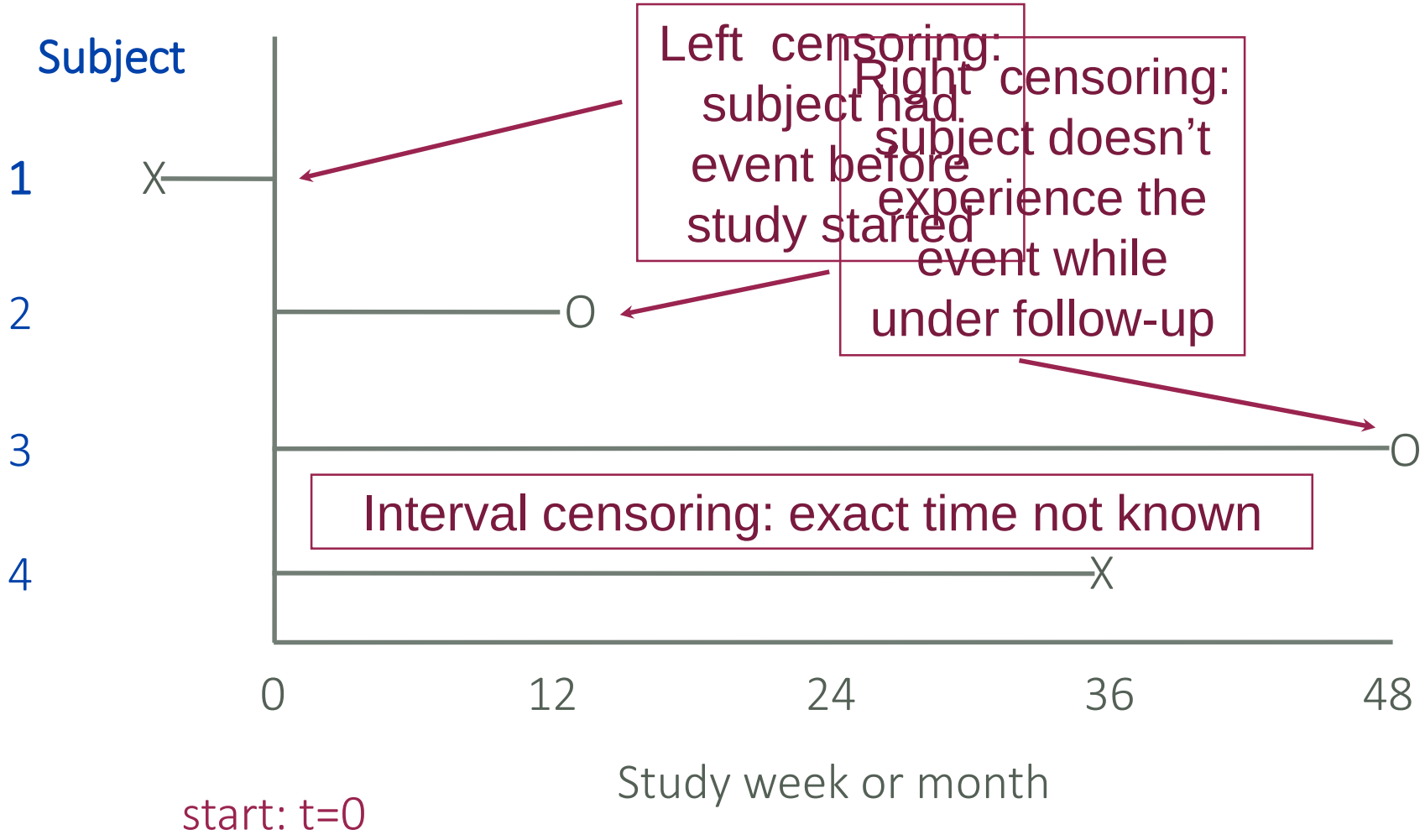


Survival analysis terms

- Time-to-event: Time from study entry until a participant has a particular outcome

- Censoring: Participant are ‘censored’ if they are lost to follow up or drop out of the study, or if the study ends before they die or have an outcome of interest.
  - They are counted as ‘event’-free for the time they were enrolled in the study.

Types of censoring/truncation

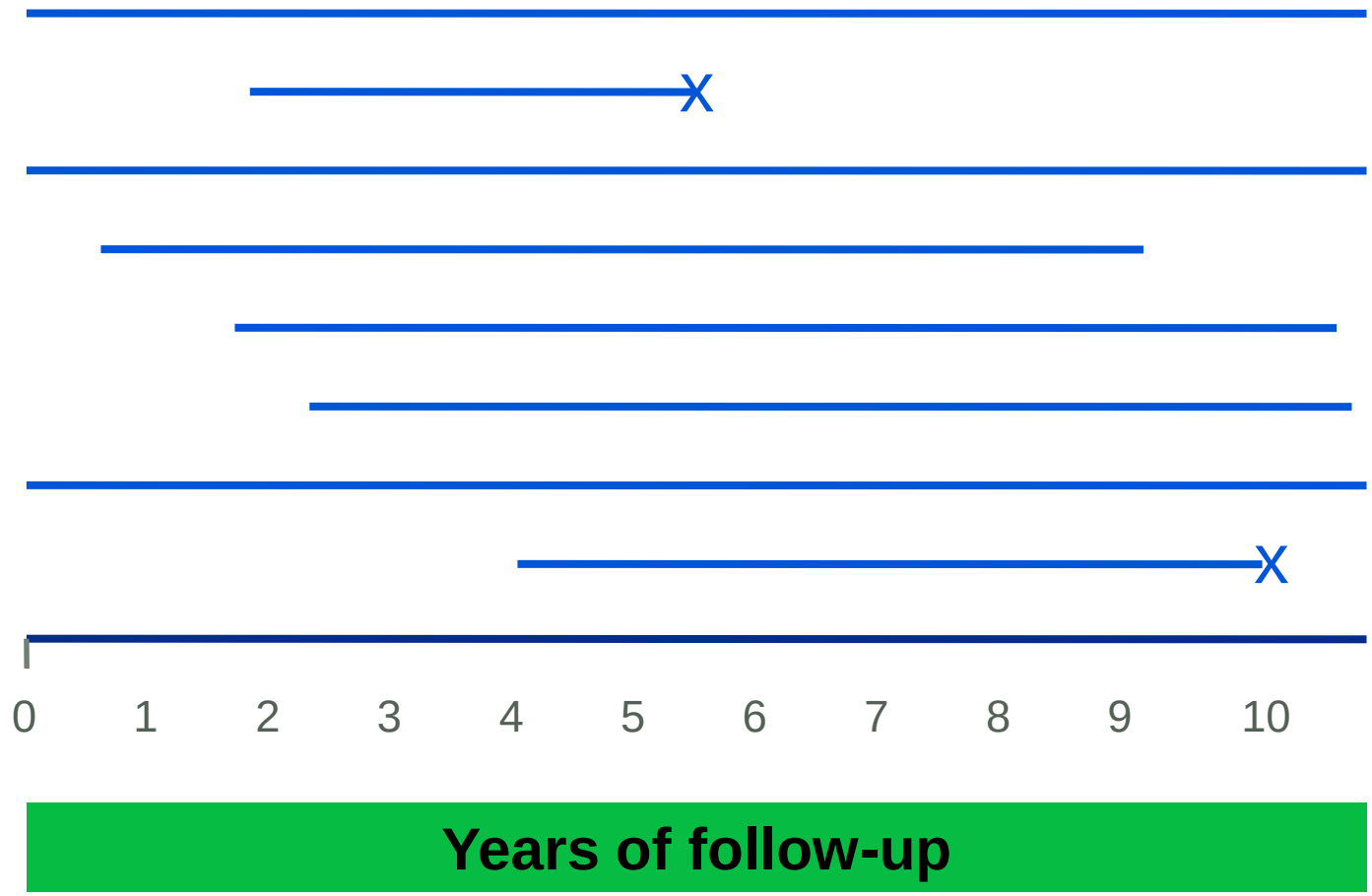


Why do we want to censor?

- When we make absolute and relative comparisons between groups (relative and absolute risks) we base these comparisons on ‘incidence density’ between groups. Subjects who do not experience the event, contribute person time until while they are in the study.

- Rate expressed with a relevant denominator
  - (eg X/100 PYFU)

Longitudinal studies, rates and censoring



Rates and incidence

- Subjects are followed for different amounts of time (drop outs, enter at different times)
- Subjects who drop out are ‘censored’ in the analysis (they contribute time until they drop out)

$$\textbf{Rate} = \frac{\textbf{Number of events occurring}}{\textbf{Total person years of follow-up (PYFU)}}$$

Survival analysis

- 1. Why not compare proportion of events in your groups using risk/odds ratios or logistic regression?

Doesn’t account for time

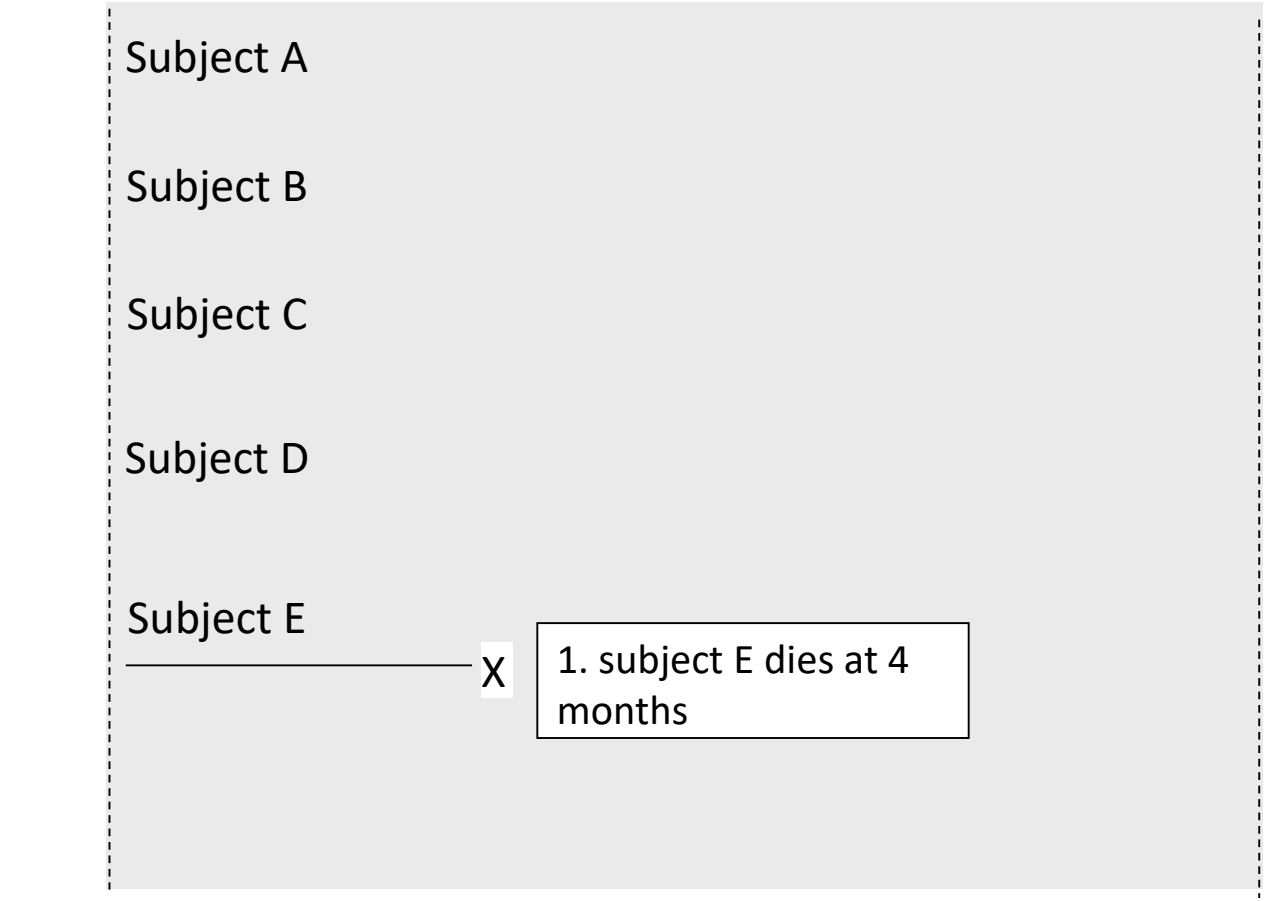
- 2. Why not compare mean time-to-event between your groups using a t-test or linear regression?

Doesn’t account for follow-up time of people who didn’t have an event

Setting up a survival dataset

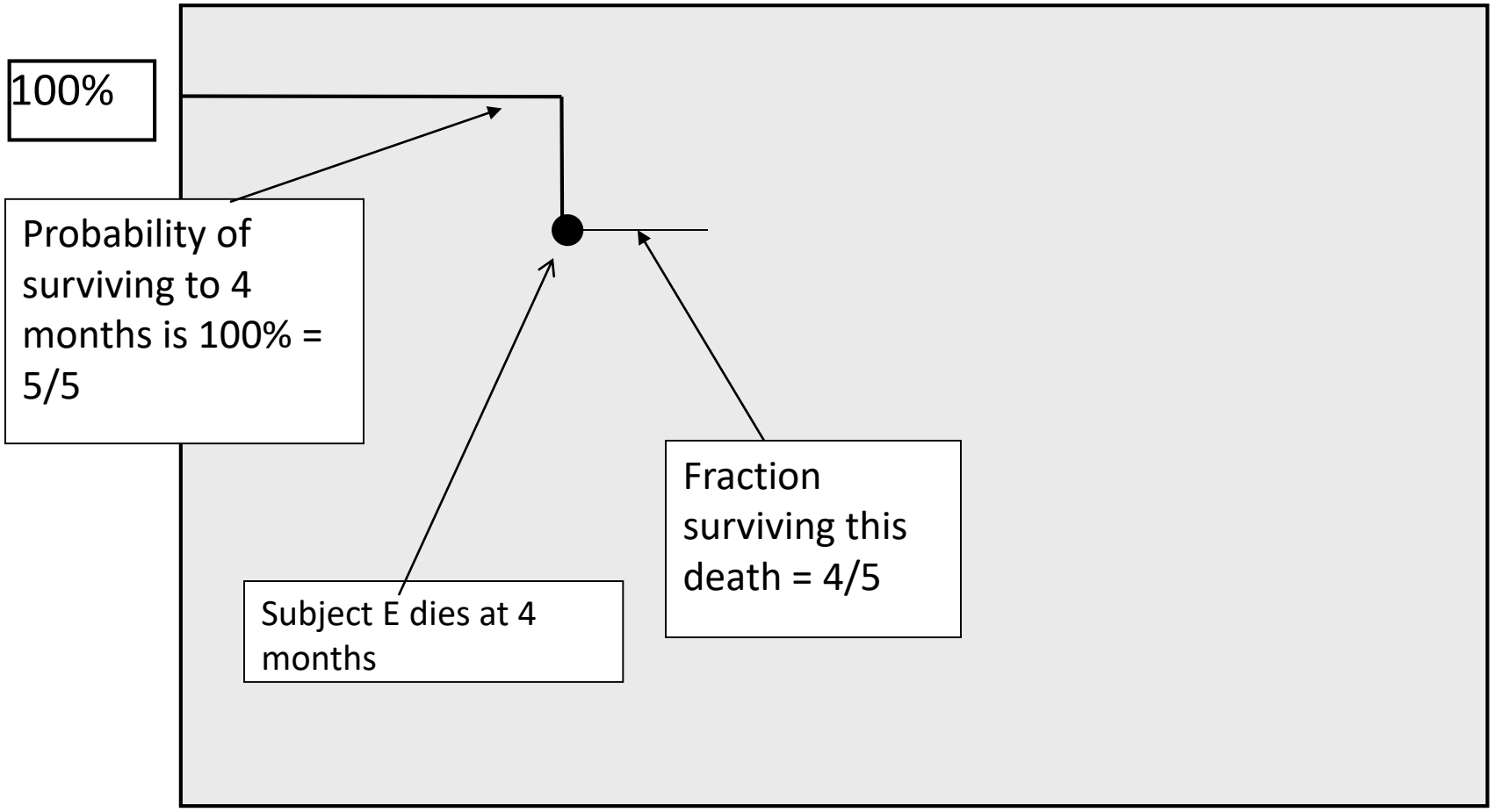
- Variables required:
  - 1) Subject ID
  - 2) Did the subject experience the event?
    - 0 = censored; 1 = experienced event
  - 3) Time from study start to the event, or being censored
  - 4) Other potentially important co-variates

Survival Data



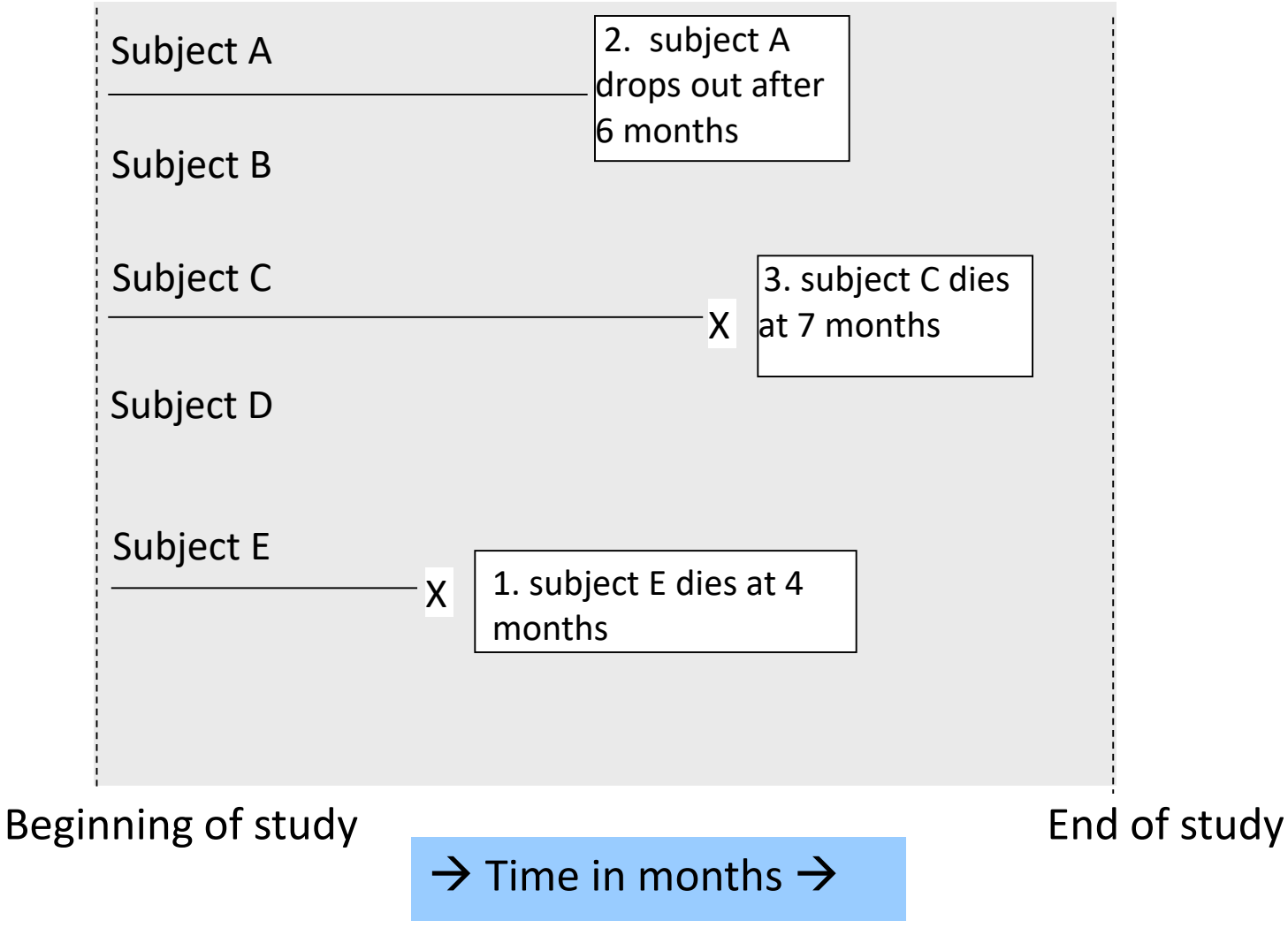
Slide from Kristen Cobb, Stanford

Corresponding Kaplan-Meier Curve



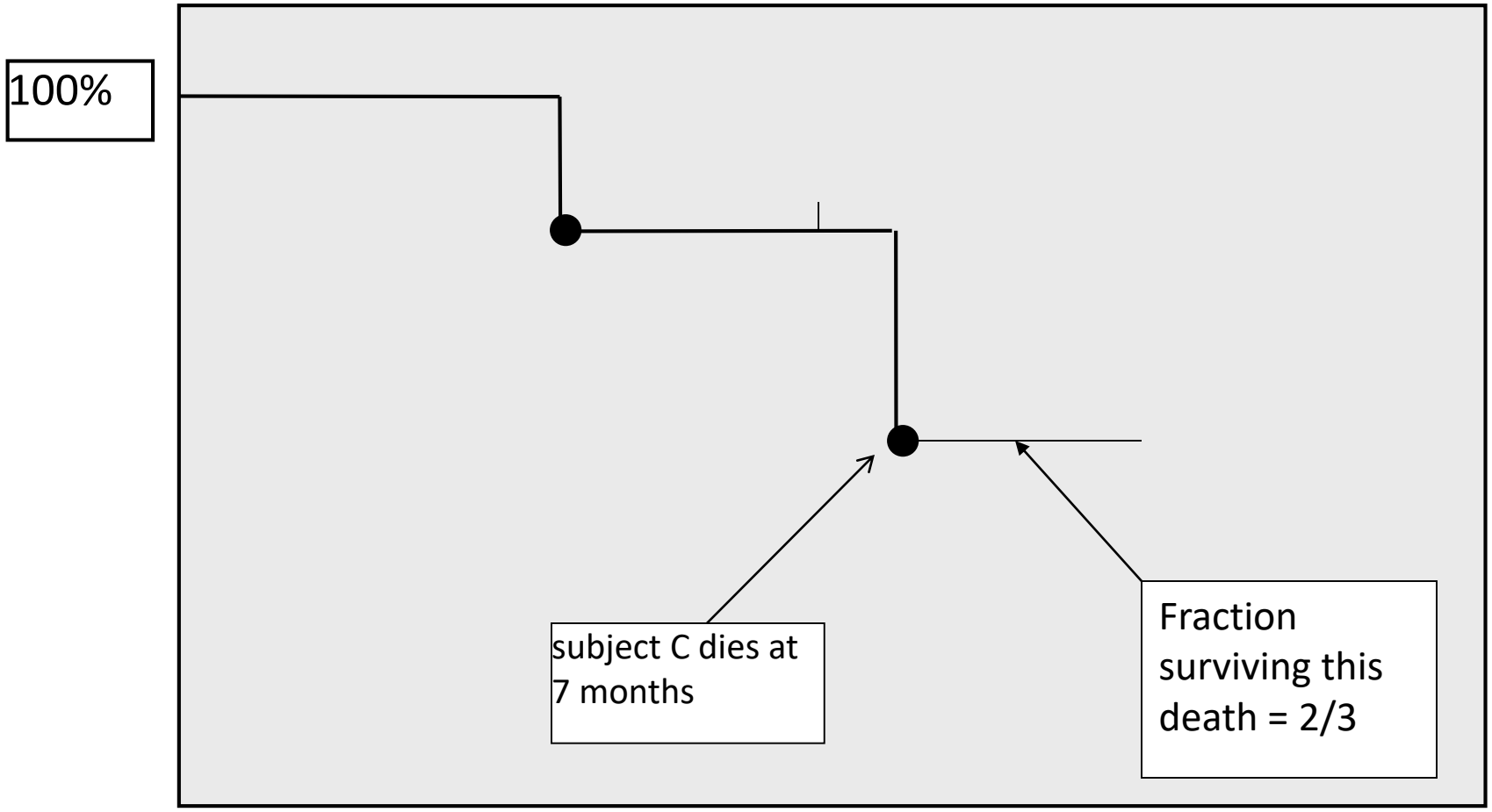
Slide from Kristen Cobb, Stanford

Survival Data



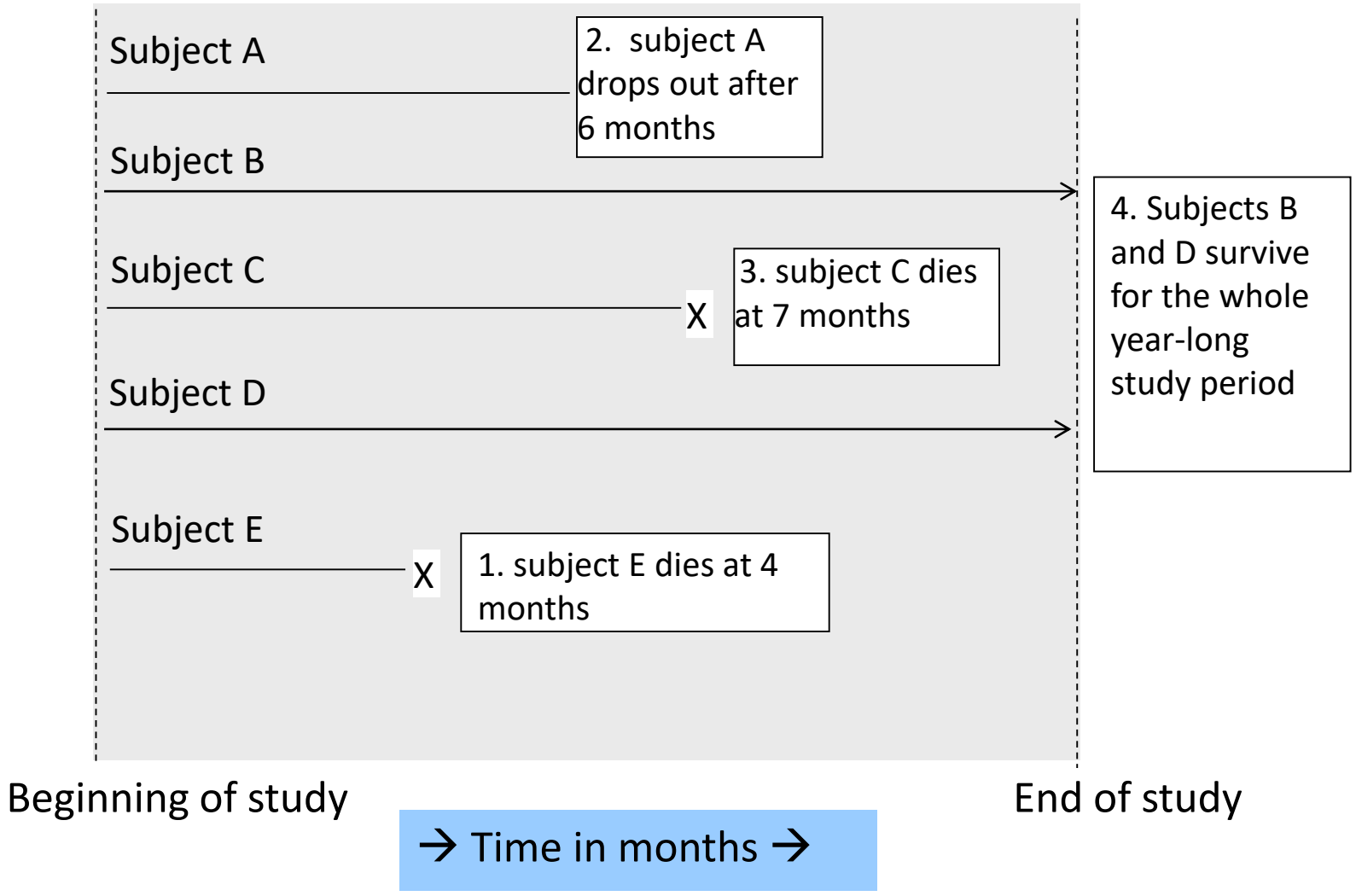
Slide from Kristen Cobb, Stanford

Corresponding Kaplan-Meier Curve



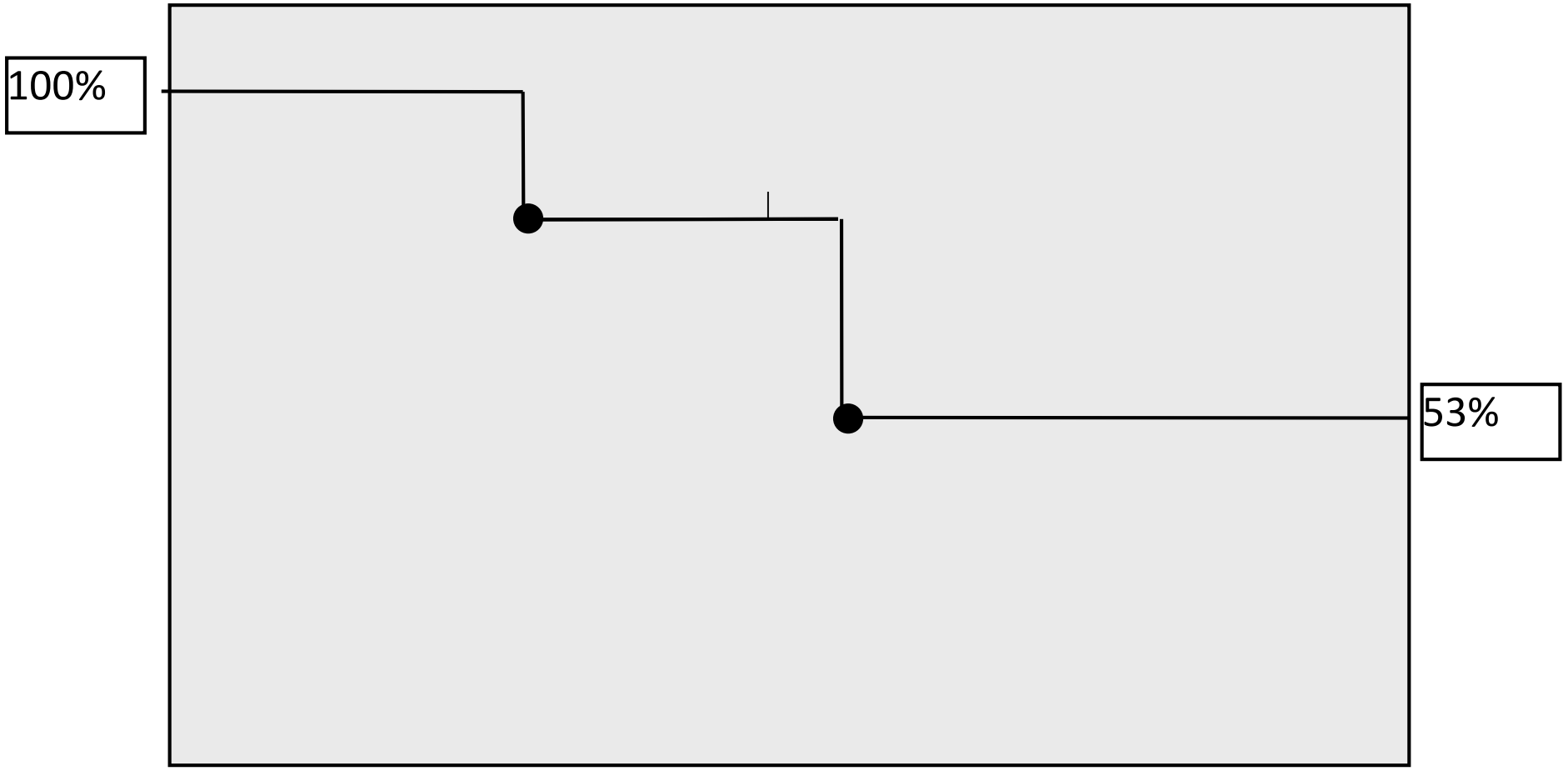
Slide from Kristen Cobb, Stanford

Survival Data



Slide from Kristen Cobb, Stanford

Corresponding Kaplan-Meier Curve



Slide from Kristen Cobb, Stanford

The product limit estimate

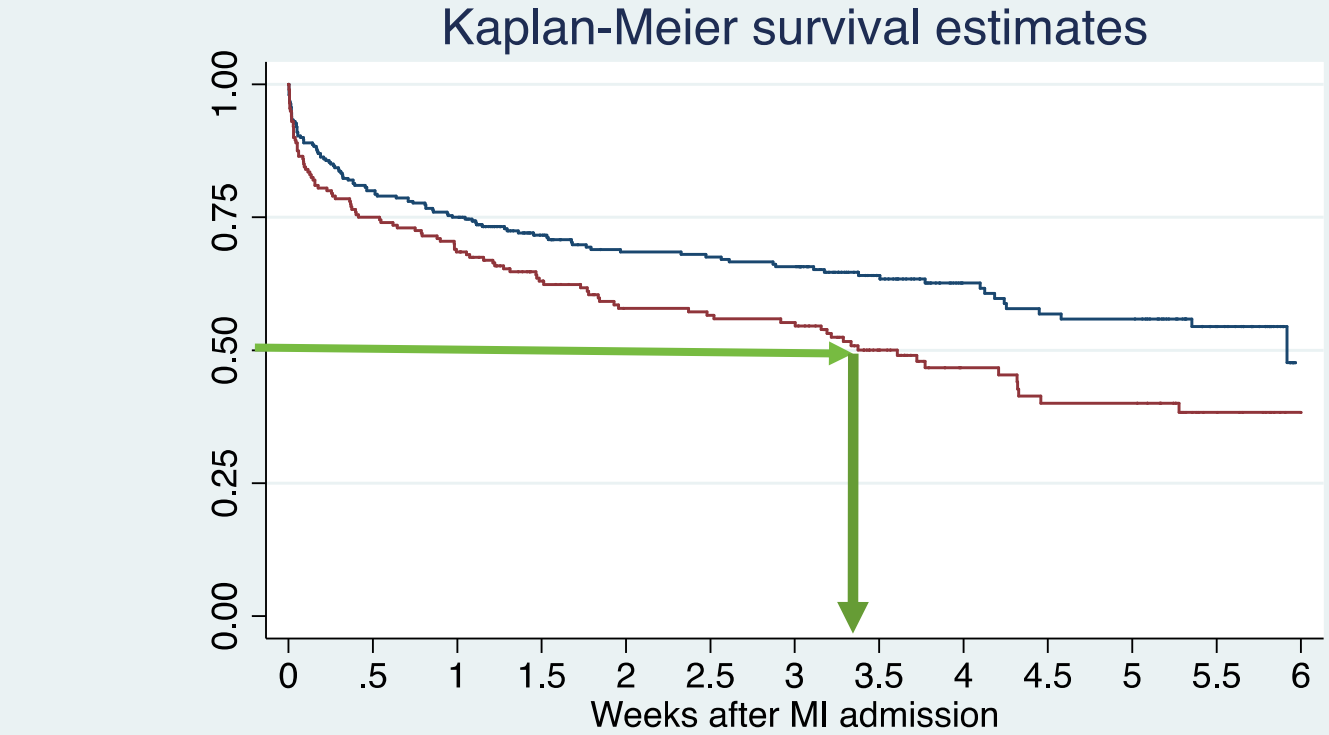
- The probability of surviving in the entire year, taking into account censoring
- = (4/5) (2/3) = 53% (conditional probability)

- > 40% (2/5) because the one drop-out survived at least a *some* of the year.
- <60% (3/5) because we don’t know if the one drop-out would have survived until the end of the year.

Kaplan-Meier function

- Non-parametric estimate of the survival function
- The probability of surviving past certain times in the sample
- Commonly used to compare study populations and screen continuous variables for multivariate analysis

Is survival equivalent by sex?




```
. sts test sex, logrank

        failure _d:  died == 1
  analysis time _t:  years_fu
```

Log-rank test for equality of survivor functions

sex	Events observed	Events expected
Male	**111**	**130.73**
Female	**104**	**84.27**
Total	**215**	**215.00**

chi2(1) = **7.79**
Pr>chi2 = **0.0053**

HR = 1.46 (95%CI 1.11 – 1.91); P=0.006

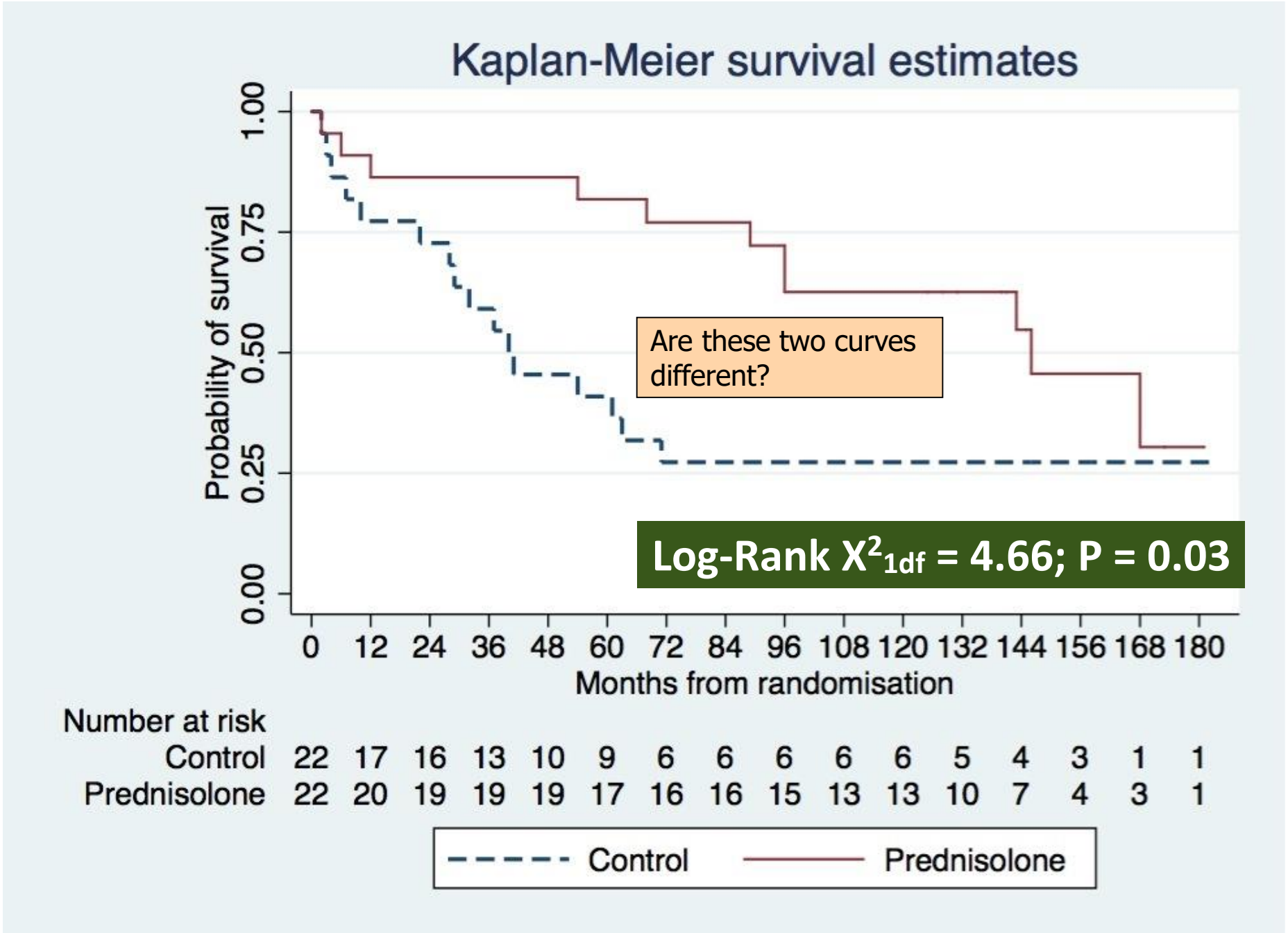
Kaplan-Meier example

Researchers randomised 44 patients with chronic active hepatitis were to receive prednisolone or no treatment (control), then compared survival curves.

Example from: BMJ 1998;317:468-469 (15 August)

Survival times (months) of 44 patients with chronic active hepatitis randomised to receive prednisolone or no treatment. (*=censored)

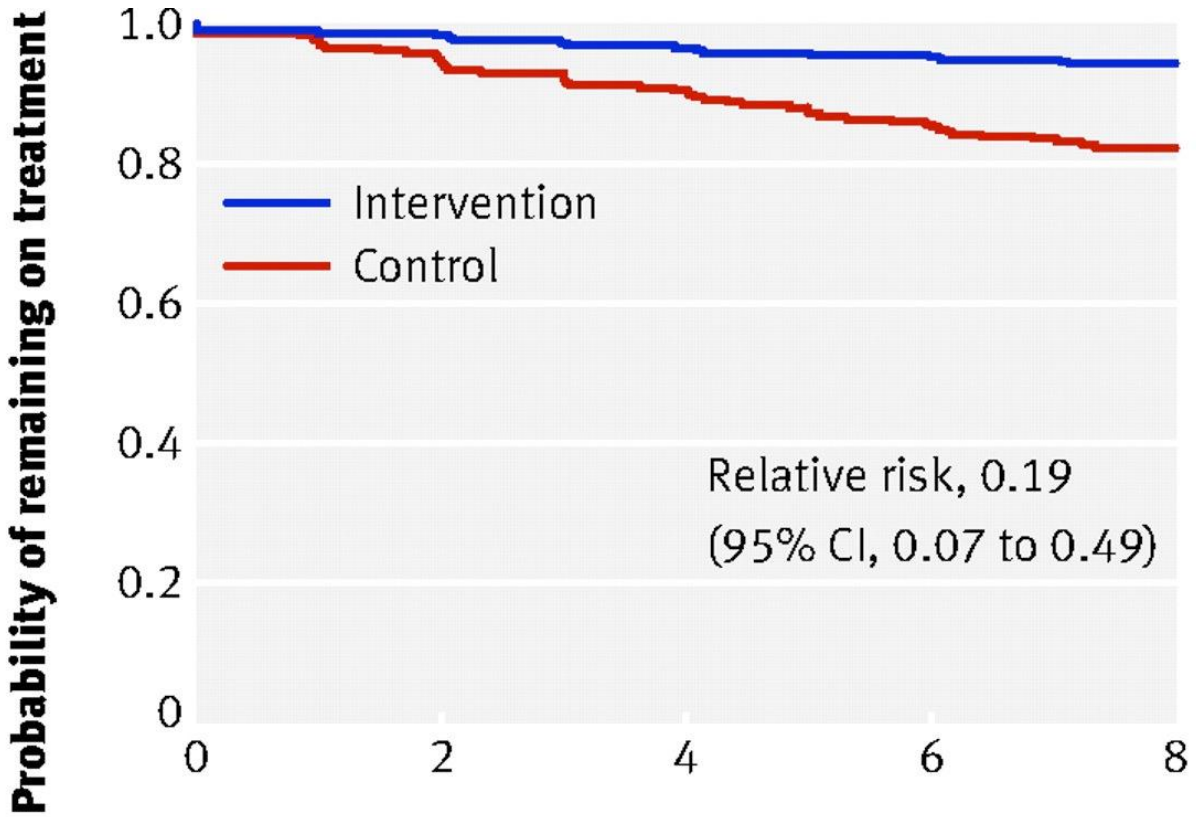
Prednisolone (n=22)	Control (n=22)
2	2
6	3
12	4
54	7
56*	10
68	22
89	28
96	29
96	32
125*	37
128*	40
131*	41
140*	54
141*	61
143	63
145*	71
146	127*
148*	140*
162*	146*
168	158*
173*	167*
181*	182*



Slide from Kristen Cobb, Stanford

Good practices with KM curves

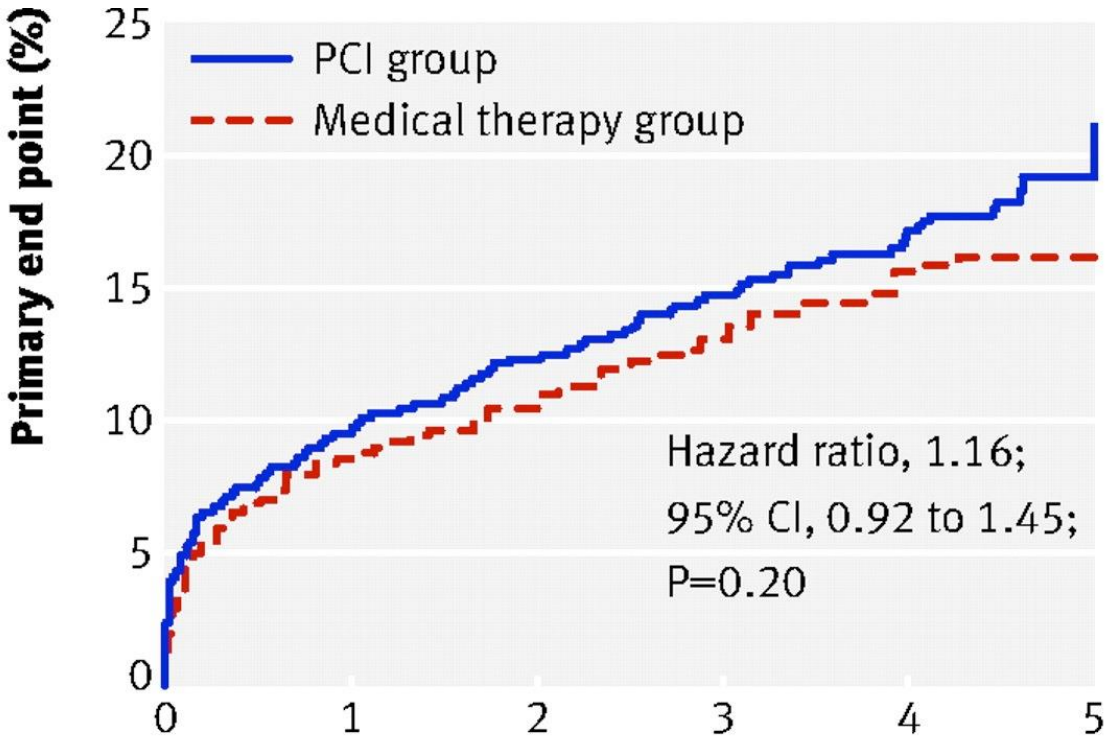
Fig 4 Kaplan-Meier plot from trial of strategies to improve adherence to tuberculosis treatment (redrawn from Thiam et al16).



No at risk					
Intervention	778	744	719	694	497
Control	744	683	635	590	392

Pocock S J et al. BMJ 2008;336:1166-1169

Fig 3 Kaplan-Meier curve from trial of percutaneous coronary intervention (PCI) for persistent occlusion after myocardial infarction.



No at risk

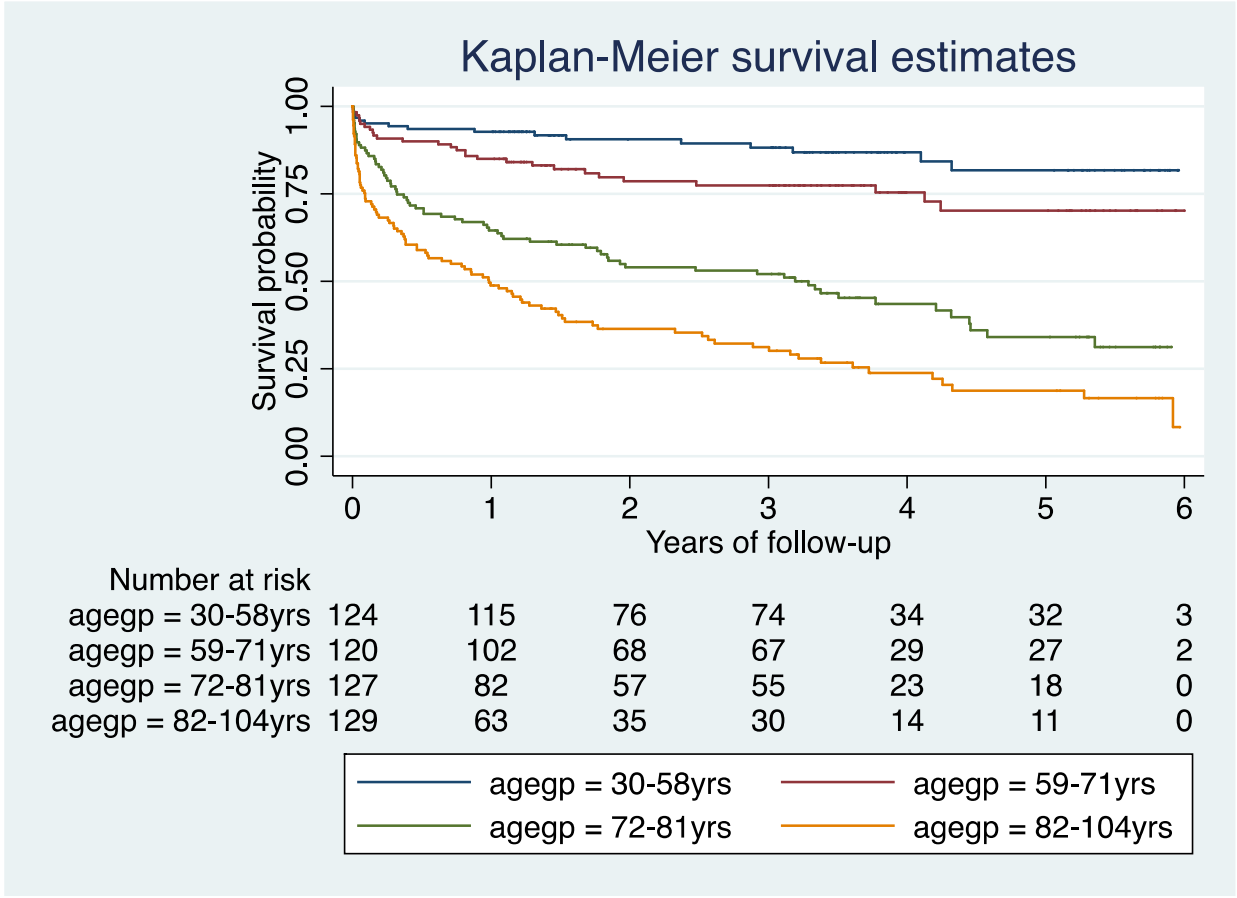
PCI group	1082	895	719	482	265	85
Medical therapy group	1084	909	714	474	268	78

Pocock S J et al. BMJ 2008;336:1166-1169

Variables in dataset

Variable	Name	Description	Codes / Values
1	id	Identification Number	1 - 500
2	age	Age at Hospital Admission	Years
3	gender	Gender	0 = Male, 1 = Female
4	hr	Initial Heart Rate	Beats per minute
5	sysbp	Initial Systolic Blood Pressure	mmHg
6	diasbp	Initial Diastolic Blood Pressure	mmHg
7	bmi	Body Mass Index	kg/m^2
8	cvd	History of Cardiovascular Disease	0 = No, 1 = Yes
9	afb	Atrial Fibrillation	0 = No, 1 = Yes
10	sho	Cardiogenic Shock	0 = No, 1 = Yes
11	chf	Congestive Heart Complications	0 = No, 1 = Yes
12	av3	Complete Heart Block	0 = No, 1 = Yes
13	miord	MI Order	0 = First, 1 = Recurrent
14	mitype	MI Type	0 = non Q-wave, 1 = Q-wave
15	year	Cohort Year	1 = 1997, 2 = 1999, 3 = 2001
16	admitdate	Hospital Admission Date	mm/dd/yyyy
17	disdate	Hospital Discharge Date	mm/dd/yyyy
18	fdate	Date of last Follow Up	mm/dd/yyyy
19	los	Length of Hospital Stay	Days from Hospital Admission to Hospital Discharge
20	dstat	Discharge Status from Hospital	0 = Alive, 1 = Dead
21	lenfol	Total Length of Follow-up	Days from Hospital Admission Date to Date of Last Follow-up
22	fstat	Vital Status at Last Follow-up	0 = Alive 1 = Dead

Age groups



_t	Haz. Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
agegp						
59-71yrs	**2.015295**	**.6283153**	**2.25**	**0.025**	**1.093847**	**3.712963**
72-81yrs	**5.741304**	**1.599474**	**6.27**	**0.000**	**3.325629**	**9.911682**
82-104yrs	**9.7058**	**2.647717**	**8.33**	**0.000**	**5.686248**	**16.56673**

Age groups

_t	Haz. Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
agegp						
59-71yrs	2.015295	.6283153	2.25	0.025	1.093847	3.712963
72-81yrs	5.741304	1.599474	6.27	0.000	3.325629	9.911682
82-104yrs	9.7058	2.647717	8.33	0.000	5.686248	16.56673

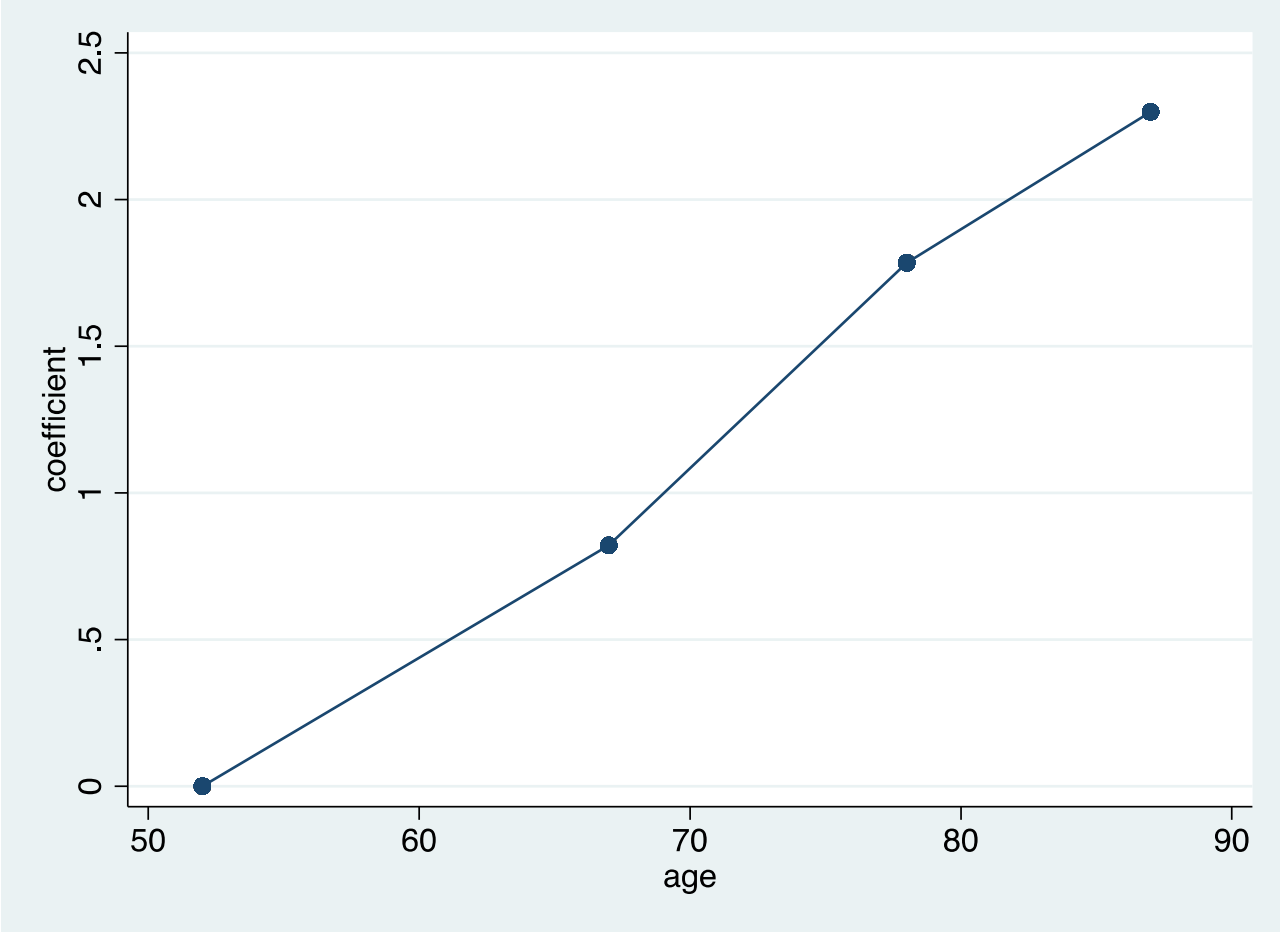
Age as continuous

_t	Haz. Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
age	1.068491	.0064943	10.90	0.000	1.055838	1.081295

```
. di exp(0.066347*20)
3.7694911
```

20 year increase in hazard

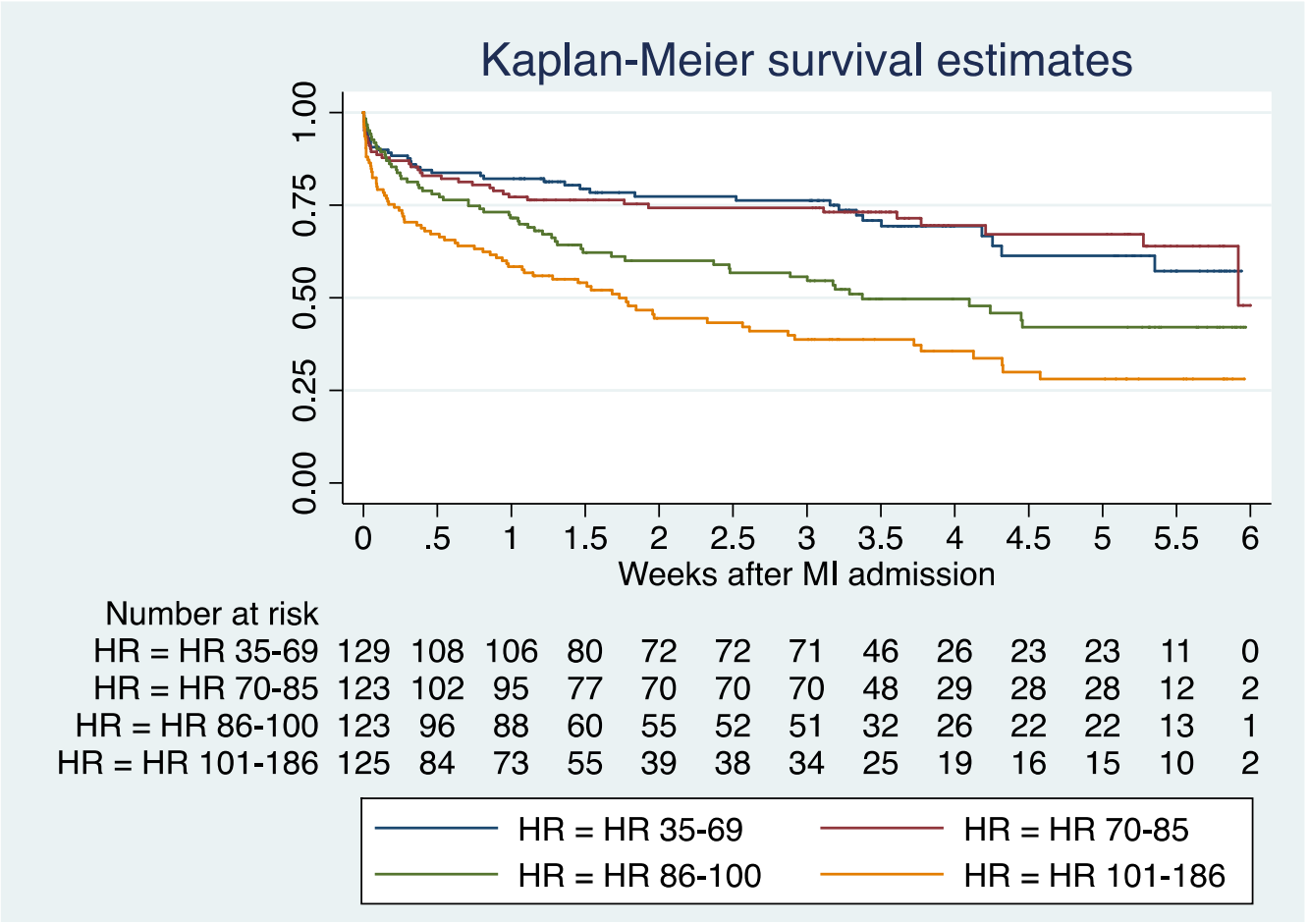
Plotting the midpoint of the quartile against the coefficient



Multivariate model for age and sex

	Univariate		Multivariate	
Covariate	HR (95%CI)	P	HR (95%CI)	P
Female vs male	1.46 (1.11 = 1.91)	0.006	0.94 (0.71 – 1.23)	0.64
Age (per year)	1.07 (1.06 – 1.08)	<0.001	1.07 (1.06 – 1.08)	<0.001

HR groups



_t	Haz. Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
HR						
HR 70-85	**1.006349**	**.2325183**	**0.03**	**0.978**	**.6398492**	**1.582778**
HR 86-100	**1.809909**	**.3744508**	**2.87**	**0.004**	**1.206569**	**2.714946**
HR 101-186	**2.635273**	**.5241768**	**4.87**	**0.000**	**1.784493**	**3.891673**

Antiviral Therapy: in press (doi: 10.3851/IMP3121)

Original article

Life expectancy after initiation of combination antiretroviral therapy in Thailand

Sirinya Teeraananchai[1,2], Suchada Chaivooth[3], Stephen J Kerr[1,2,4], Sorakij Bhakeecheep[3], Anchalee Avihingsanon[1,5], Achara Teeraratkul[6], Petchsri Sirinirund[7], Matthew G Law[2], Kiat Ruxrungtham[1,5]*

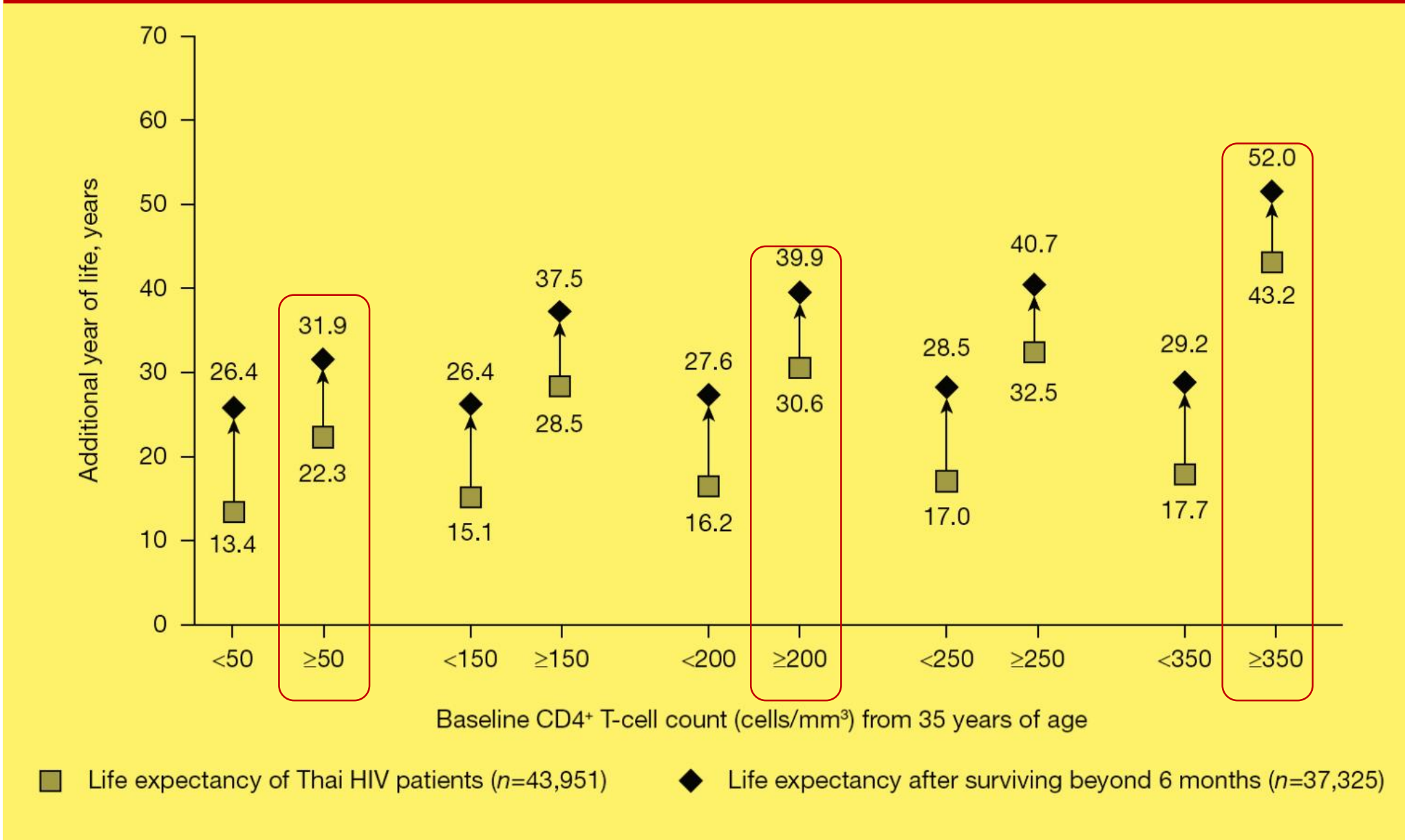
[1]HIV-NAT, Thai Red Cross AIDS Research Centre, Bangkok, Thailand
[2]Kirby Institute, University of New South Wales, Sydney, Australia

Thai National UC AIDS Program (NAP): 2008-2014

N =201,688 with **618,837** person-years of follow up. Median age **37** yrs.
Median **CD4+** T-cell count was **109** cells/mm3 and
ART first-line: **NNRTI 91%** (NPV 59%, EFV 32%), bPI 7 %

Average Life-expectancy from 35 yo

National AIDS Program, Thailand



J Acquir Immune Defic Syndr 2017;75:219–225

First-Line Antiretroviral Treatment Outcomes and Durability in HIV-Infected Children Treated Through the Universal Coverage Health Program in Thailand

Sirinya Teeraananchai, MS,† Torsak Bunupuradah, MD,* Thanyawee Puthanakit, MD,*‡§*
Stephen J. Kerr, PhD,†‖ Kiat Ruxrungtham, MD,*¶ Suchada Chaivooth, MD,#*
*Sorakij Bhakeecheep, MD,# Matthew G. Law, PhD,† and Kulkanya Chokephaibulkit, MD***

- N=4120
- Children receiving ART through UC program had good treatment outcomes, although **a fifth (20%)** required switching regimen by **3 years.**
- Avoiding **NNRTI first-line** regimens in **high-risk children** may help **prevent treatment failure**.

National AIDS Program, Thailand

Children and Adolescents Antiretorviral Treatment Outcomes

N=4120

Median age of **9.3 yrs** (IQR 5.8–12.0)

Median duration of ART = **3.7 years**

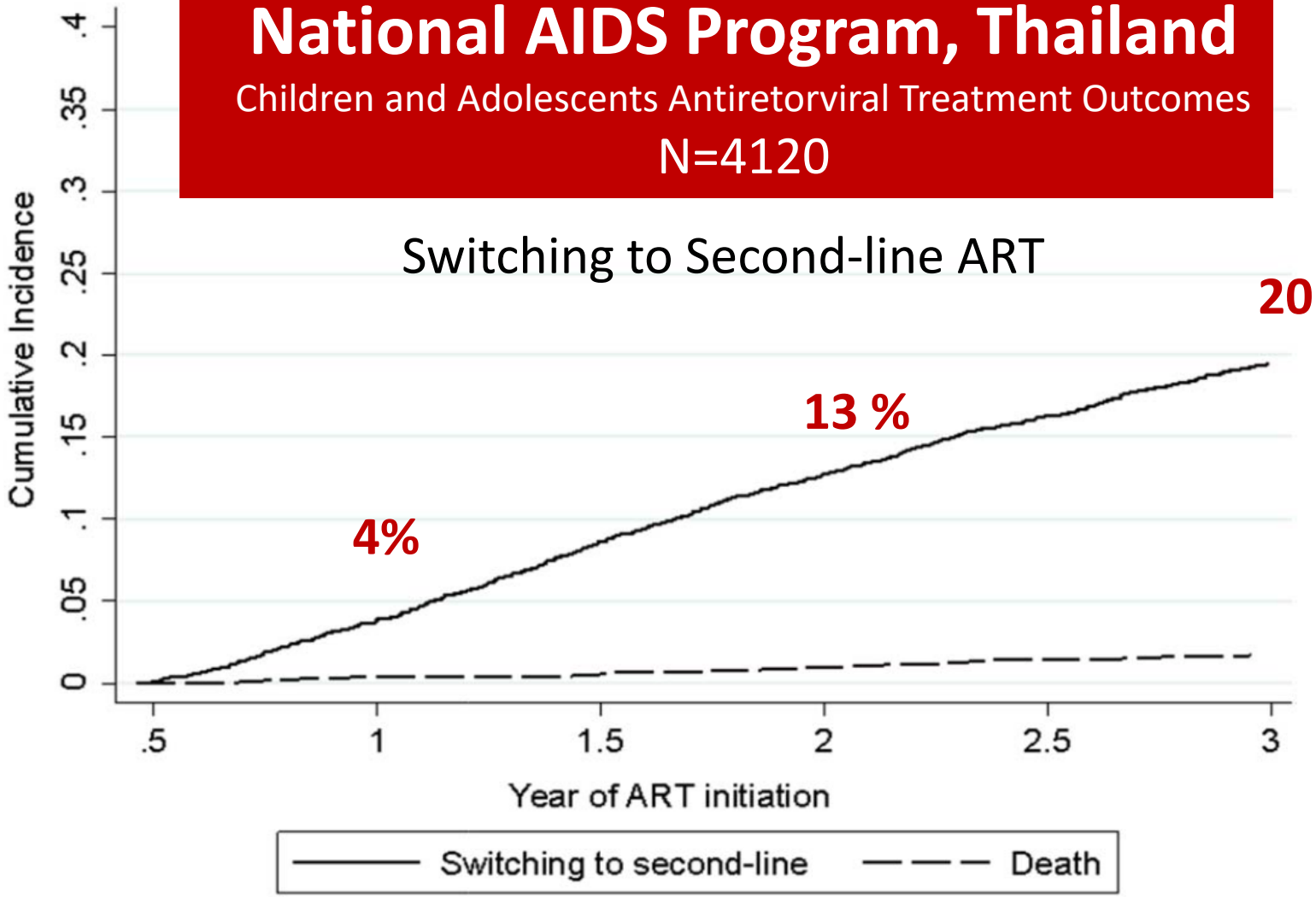


FIGURE 2. Cumulative incidence of switching to second-line regimens showed the estimates incidence of switching to second-line regimen.

Switching due to failure

(N=1054)

1. 84% had VF
2. 19% had immunological failure

Risk factors

1. Children **aged ≥12 years & <5 years** at ART initiation
2. starting with **NNRTIs**, and
3. baseline **CD4% <10%**

J AIDS 2017;75:219–225

TABLE 3. Characteristics Associated With Switching to Second-Line Regimens

Characteristics (N = 4120)	No. of SW	PY	SW Rate (95% CI) per 100 FU	Univariate		Multivariate	
				SHR (95% CI)	*P*	aSHR (95% CI)	*P*
Sex					0.140		
Male	490	6855	7 (6.54 to 7.81)	1.10 (0.97 to 1.24)			
Female	564	8632	6.53 (6.02 to 7.10)	Ref			
Age at ART initiation, yrs					0.004		0.004
<5	206	3001	6.86 (5.99 to 7.87)	1.18 (0.98 to 1.42)		1.29 (1.07 to 1.55)	
5 to <9	259	4456	5.81 (5.15 to 6.56)	Ref		Ref	
9 to <12	297	4398	6.75 (6.03 to 7.57)	1.15 (0.97 to 1.35)		1.14 (0.96 to 1.34)	
≥12	292	3632	8.04 (7.17 to 9.02)	1.36 (1.15 to 1.61)		1.33 (1.12 to 1.57)	
Year of ART initiation					0.004		0.039
2008	392	6863	5.71 (5.17 to 6.31)	Ref		Ref	
2009	259	3593	7.21 (6.38 to 8.14)	1.21 (1.03 to 1.42)		1.14 (0.97 to 1.34)	
2010	179	2193	8.16 (7.05 to 9.45)	1.32 (1.10 to 1.58)		1.24 (1.03 to 1.49)	
2011	125	1556	8.03 (6.74 to 9.57)	1.30 (1.06 to 1.60)		1.26 (1.02 to 1.56)	
2012	76	892	8.52 (6.80 to 10.66)	1.48 (1.15 to 1.92)		1.43 (1.11 to 1.85)	
2013–2014*	23	389	5.91 (3.93 to 8.89)	1.43 (0.93 to 2.21)		1.35 (0.88 to 2.09)	
Baseline CDC clinical stage, N (%)					0.632		
N or A	815	11,788	6.91 (6.46 to 7.41)	1.09 (0.89 to 1.35)			
B	137	2086	6.57 (5.55 to 7.76)	1.04 (0.80 to 1.34)			
C	102	1612	6.33 (5.21 to 7.68)	Ref			
First regimen					<0.001		<0.001
NNRTI-based ART	996	14,135	7.05 (6.62 to 7.50)	1.64 (1.25 to 2.15)		1.61 (1.23 to 2.12)	
PI-based ART	58	1352	4.29 (3.32 to 5.55)	Ref		Ref	
Baseline CD4%					<0.001		0.001
<10%	404	4856	8.32 (7.55 to 9.17)	1.58 (1.32 to 1.90)		1.48 (1.22 to 1.79)	
10 to <16%	147	2180	6.74 (5.74 to 7.93)	1.29 (1.03 to 1.62)		1.22 (0.97 to 1.53)	
≥16%	158	3105	5.09 (4.35 to 5.95)	Ref		Ref	
Unknown	345	5346	6.45 (5.81 to 7.17)	1.25 (1.03 to 1.51)		1.24 (1.02 to 1.50)	
Region					0.471		
Bangkok	89	1541	5.77 (4.69 to 7.11)	Ref			
Central	106	1443	7.34 (6.07 to 8.88)	1.24 (0.93 to 1.65)			
Northern	234	3672	6.37 (5.61 to 7.24)	1.09 (0.85 to 1.39)			
Southern	137	2068	6.62 (5.60 to 7.83)	1.11 (0.85 to 1.46)			
Eastern	121	1611	7.51 (6.28 to 8.98)	1.28 (0.98 to 1.69)			
Western	44	664	6.63 (4.93 to 8.91)	1.13 (0.79 to 1.62)			
Northeastern	323	4487	7.20 (6.46 to 8.03)	1.22 (0.96 to 1.54)			

*There were no children who switched ART regimens in 2014 and then this year was combined with 2013.
aSHR, adjusted subdistribution hazard ratios; CDC, the Centers for Disease Control and Prevention classification; FU, follow-up; PY, person-years; SW, switching to second-line regimens.

European Journal of Preventive Cardiology

EUROPEAN SOCIETY OF CARDIOLOGY®

Original scientific paper

An updated prediction model of the global risk of cardiovascular disease in HIV-positive persons: The Data-collection on Adverse Effects of Anti-HIV Drugs (D:A:D) study

Nina Friis-Møller[1], Lene Ryom[2], Colette Smith[3], Rainer Weber[4], Peter Reiss[5], F Dabis[6], Stephane De Wit[7], Antonella D'Arminio Monforte[8], Ole Kirk[2], Eric Fontas[9], Caroline Sabin[3], Andrew Phillips[3], Jens Lundgren[2] and Matthew Law[10] on behalf of the D:A:D study group

European Journal of Preventive Cardiology
0(00) 1–10

DOI: 10.1177/2047487315579291
ejpc.sagepub.com

SAGE

Table 2. Multivariable risk factor models.

	Full model				Reduced model			
Predictor	HR	(95% CI)	*p*	β	HR	(95% CI)	*p*	β
Ln age	22.0	(16.3, 29.6)	<0.001	3.090	24.0	(17.9, 32.1)	<0.001	3.178
Male *vs.* female	1.37	(1.13, 1.66)	0.001	0.314	1.41	(1.16, 1.71)	<0.001	0.344
Diabetes (yes *vs.* no)	1.96	(1.59, 2.42)	<0.001	0.675	2.08	(1.69, 2.56)	<0.001	0.731
Family history (yes *vs.* no)	1.37	(1.14, 1.64)	0.001	0.314	1.39	(1.16, 1.67)	<0.001	0.330
Smoke								
Current *vs.* never	2.25	(1.91, 2.63)	<0.001	0.809	2.26	(1.93, 2.65)	<0.001	0.816
Former *vs.* never	1.24	(1.01, 1.51)	0.038	0.213	1.27	(1.04, 1.55)	0.019	0.239
Ln cholesterol (mmol/l)	2.58	(2.04, 3.27)	<0.001	0.948	2.98	(2.35, 3.78)	<0.001	1.092
Ln HDL (mmol/l)	0.61	(0.51, 0.72)	<0.001	−0.501	0.59	(0.50, 0.71)	<0.001	−0.519
Ln systolic blood pressure (mmHg)	4.59	(2.84, 7.42)	<0.001	1.523	4.56	(2.82, 7.39)	<0.001	1.518
Ln2 CD4 count (cells/mm^3)	0.89	(0.84, 0.94)	<0.001	−0.119	0.89	(0.84, 0.94)	<0.001	−0.114
Receiving abacavir (yes *vs.* no)	1.47	(1.26, 1.71)	<0.001	0.384	–			
PI exposure (per year)	1.048	(1.009, 1.088)	0.015	0.0467	–			
NRTI exposure (per year)	1.028	(1.003, 1.054)	0.028	0.0278	–			
Framingham model (2008)								
	Women				Men			
Ln age (years)	10.3	(5.6, 18.6)	<0.001	2.329	21.4	(14.0, 32.5)	<0.001	3.061
Diabetes (yes *vs.* no)	2.00	(1.49, 2.67)	<0.001	0.692	1.78	(1.43, 2.20)	<0.001	0.574
Smoker current *vs.* no	1.70	(1.40, 2.06)	<0.001	0.529	1.92	(1.65, 2.24)	<0.001	0.655
Ln cholesterol	3.35	(2.00, 5.62)	<0.001	1.209	3.08	(2.05, 4.62)	<0.001	1.124
Ln HDL	0.49	(0.35, 0.69)	<0.001	−0.708	0.39	(0.30, 0.52)	<0.001	−0.933
Ln SBP if not treated	15.8	(7.9, 31.9)	<0.001	2.762	6.91	(3.91, 12.20)	<0.001	1.933
Ln SBP if treated	16.8	(8.5, 33.5)	<0.001	2.823	7.38	(4.22, 12.92)	<0.001	1.999

β corresponds to ln(HR). HR: hazard ratio; CI: confidence interval; Ln: log (base e); Ln2: log (base 2); HDL: high-density lipoprotein; PI: protease inhibitor; NRTI: nucleoside reverse transcriptase inhibitor; SBP: systolic blood pressure