Survival analysis (Kaplan-Meier and Cox Proportional Hazard Analysis)

Dr Aye Mya Chan Thar
Assistant Director
Expanded Programme on Immunization
Department of Medical Research

Randomised controlled trials: Time to event

Severe P et al. Early versus Standard Antiretroviral Therapy for HIV Infected Adults in Haiti NEJM 2010;363:257

- Eligibility criteria – confirmed CD4+ T-cell count > 200 and < 350 per mL at baseline and no history of an acquired immunodeficiency syndrome (AIDS) illness
- Between 2005 and 2008, a total of 816 participants — 408 per group — were enrolled and were followed for a median of 21 months
- The primary study end point was survival (death).
- End of follow up was 1 May 2009
- Therefore, participants had variable follow-up times

- How does one analyse these data?
- Need methods that take account of the variable lengths of follow-up
- Cannot use methods such as logistic regression

Typical method of analysis: Kaplan-Meier estimates of survival

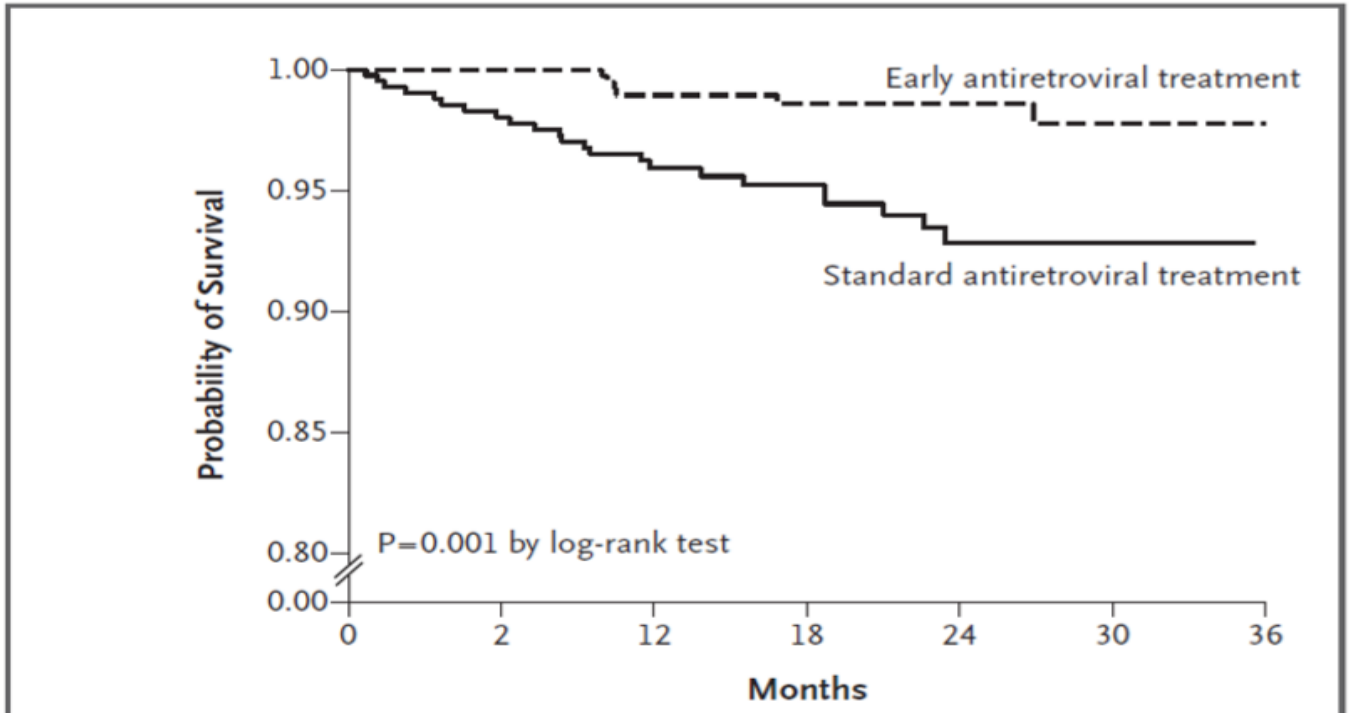


Figure 2. Kaplan–Meier Estimates of Survival in the Early-Treatment and Standard-Treatment Groups.

Severe P et al. NEJM 2010;363:257

Typical method of analysis: Cox regression to estimate hazard ratio

- The unadjusted hazard ratio for the risk of death with standard treatment as compared with early treatment was 4.0 (95% confidence interval [CI], 1.6 to 9.8).

Severe P et al. NEJM 2010;363:257

Longitudinal studies

Studies where participants are followed over time

- Cohort studies – People initially free of disease followed over time to see who develops outcome (disease)
- Survival studies – People followed from the time of an event such as the diagnosis of disease to see who dies or has disease recurrence
- Intervention studies – people randomised to two or more treatment regimens and followed to see who develops pre-specified outcome

➢ In this subject, we assume that subjects experience only one disease endpoint (it is always possible to examine time until the first occurrence).

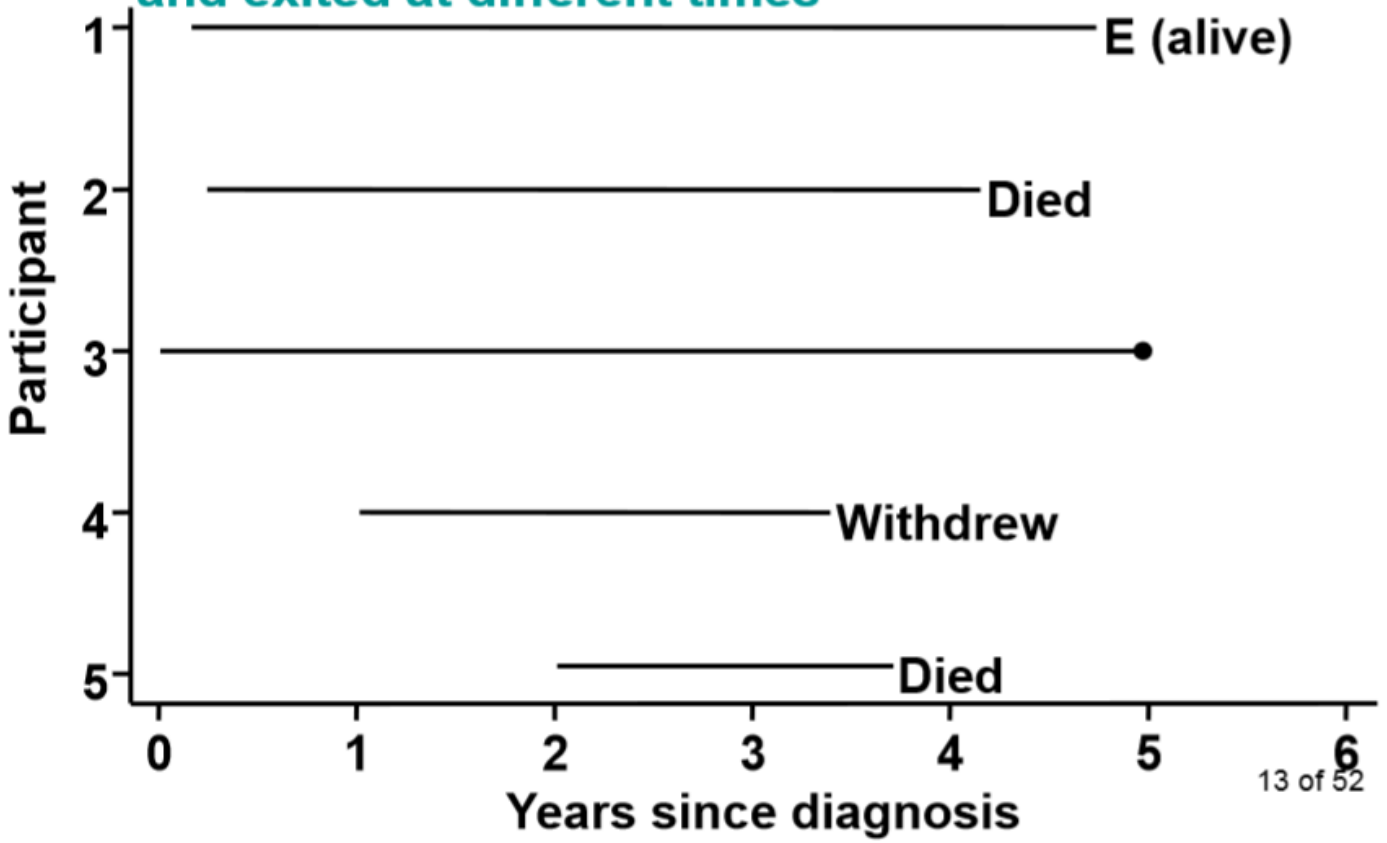
Variable follow-up times common

Some possible reasons:

- People recruited over time, but same end date
- People move into study area after study began
- Delay between diagnosis & recruitment
- Loss to follow up: e.g. emigration or withdrawal
- Death from other causes
- If the population of interest is defined by their age: e.g. women of child bearing age (ie.15-44 years)

Example: 5 yr study of prostate cancer

participants recruited at varying times after diagnosis, and exited at different times



Outcome data for participants

Two pieces of information for each person:

- Whether they experience the event – Death, disease diagnosis, injury,…
- The time under observation – (length of time they were followed) – We will call it observation time

Observation time for each person

- Starts when they join the study
- Stops at the first of the following times:
  - the time they develop the outcome of interest
  - the time they die from another disease
  - the time they are lost to follow-up □
  - the time the follow-up period ends

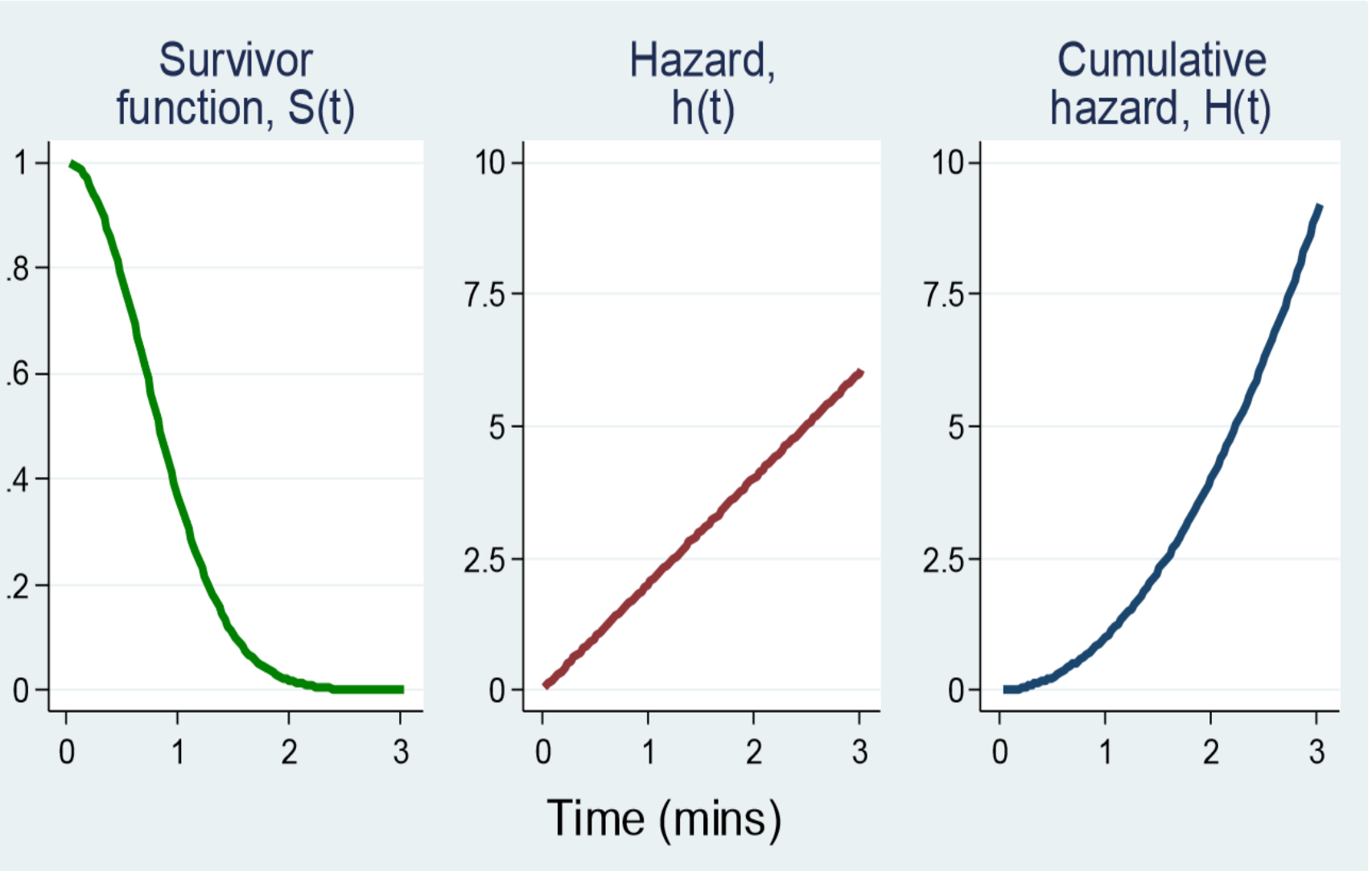
i.e. the time during which, if the person experienced an event, the event would be recorded in the study

Kaplan-Meier curves and the log rank test

Survival analysis: Overview

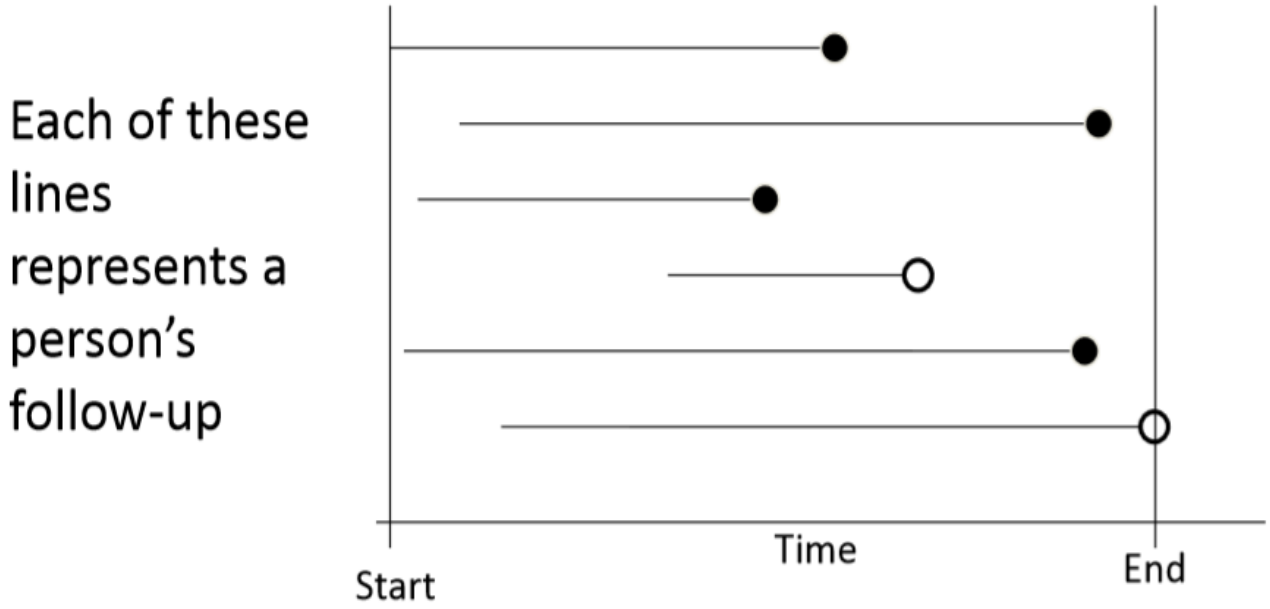
- Outcome/response: Time to an event (survival time)
  - An individual's survival time is a random variable denoted by T
  - T can take values t (≥0) with probabilities determined by the survivor or hazard functions
- Rate is not assumed to be constant over time
- Concentrates on:
  - Survivor function, S(t): probability of survival to time t
  - Hazard function h(t): Instantaneous rate of the event at time t
  - Cumulative hazard function, H(t): cumulative hazard by time t estimated by the sum of the risks at each time i at which an event occurs

    = Σ (di / ni) (up to and including t)
- Kaplan-Meier survival curve is an estimate of the survivor function
- Survival analysis closely related to analysis of rates (e.g. Mantel - Haneszel & Poisson regression)
  - Methods can examine confounding and effect modification by time


Survivor function, S(t)
Hazard, h(t)
Cumulative hazard, H(t)
1
.8
.6
.4
.2
0
10
7.5
5
2.5
0
0
1
2
3
Time (mins)

Censoring

Experience of 6 individuals followed over time in a study



Closed circles – dead

Open circles – censored

Censored: did not experience event during study period, so exact survival time unknown

Types of censoring

- Non-informative: Probability of being censored unrelated to probability of event
  - End of follow-up
  - Accidental death in CHD study
- Informative: Probability of being censored related to probability of event –
  - Too ill to attend follow-up
  - Emigration after full recovery
  - Suicide after cancer recurrence

➢ Most commonly used survival analysis methods (e.g. Kaplan-Meier survival curve estimate, log-rank test & Cox regression) assume non-informative censoring

Early vs standard antiretroviral therapy in Haiti

- Eligibility criteria:
  - Infected with HIV and ≥ 18 years old
  - Confirmed CD4+ T-cell count > 200 and < 350 per cubic millimeter at baseline
  - No history of acquired immunodeficiency syndrome (AIDS)
- Recruitment: Between 2005 and 2008, 816 participants (408 per group) were enrolled and followed for median of 21 months
- The primary study end point was survival (death)
- End of follow-up: 1st May 2009

Severe P et al. Early versus Standard Antiretroviral Therapy for HIV-Infected Adults in Haiti NEJM 2010;363:257

Kaplan-Meier survival curve

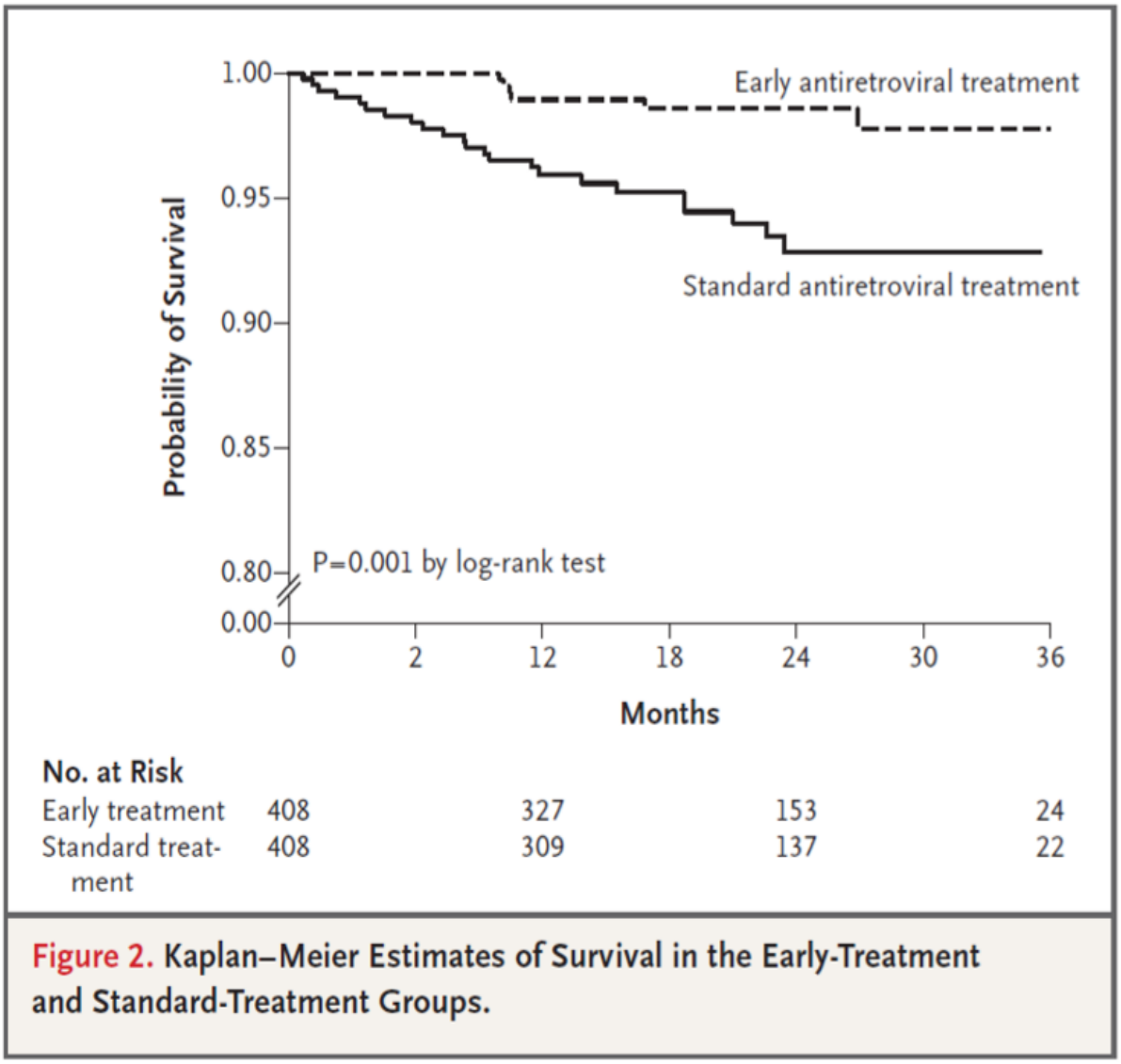


Figure 2. Kaplan–Meier Estimates of Survival in the Early-Treatment and Standard-Treatment Groups.

Severe P et al. NEJM 2010;363:257

Survival analysis terminology: Risk set

- Exact times at which events occur referred to as t
  - t1, t2, t3, …
- Risk set: Set of individuals still being studied at each time, t, that an event occurs

Kaplan-Meier survival curve: Example dataset

- Hypothetical RCT comparing two treatments for uncomplicated falciparum malaria
- Recruitment: 50 eligible patients (25 per group) with uncomplicated falciparum malaria were randomisation to two treatment groups :
  - Combination therapy: artesunate+mefloquine; or
  - Monotherapy (standard treatment): artesunate
- Primary study end point: Reappearance of original falciparum malaria infection (referred to as recrudescence)
- Follow-up: Patients assessed daily until day 28 after enrollment

Kaplan-Meier survival curve: Calculation

- Calculations will be done using Excel

Kaplan-Meier survival curve: Calculation Part 1: Risk at time t, given alive at that time

At each time (t) that someone dies, calculate the risk (r_t) of dying at that time:

$$r_t = d_t/n_t$$

d_t = number of deaths at time t

n_t = number of people still being studied at time t

This is the conditional probability of dying at time t, having survived to that point

Kaplan-Meier survival curve: Calculation Part 2: Probability of surviving at time t

- Estimated probability of surviving at time t:

$$s_t = 1 - r_t = \frac{n_t - d_t}{n_t}$$

- At any time t where $d_t = 0$, the probability of survival equals 1
- This is the conditional probability of surviving beyond time t, having survived to time t

Kaplan-Meier survival curve: Calculation Part 3: Kaplan-Meier estimate of the survivor function S(t)

- Calculation of survivor function $S(t)$ uses rules of conditional probabilities:

$$S(t_1) = \Pr(\text{surviving beyond time } t_1)$$
$$= \Pr(\text{surviving until just before } t_1)$$
$$\times \Pr(\text{surviving at time } t_1 \text{ having survived until just before } t_1)$$
$$= S(t_0) \times s_{t1}$$

- The estimated survival probability from start time t_0 until just before first failure time t_1 is 1, and so:

$$S(t_1) = S(t_0) \times s_{t1} = 1 \times s_{t1} = s_{t1}$$
$$S(t_2) = S(t_1) \times s_{t2} = s_{t1} \times s_{t2}$$
$$S(t_j) = S(t_{j-1}) \times s_{tj} = s_{t1} \times s_{t2} \times s_{t3} \ldots \times s_{tj}$$

Known as the product-limit formula

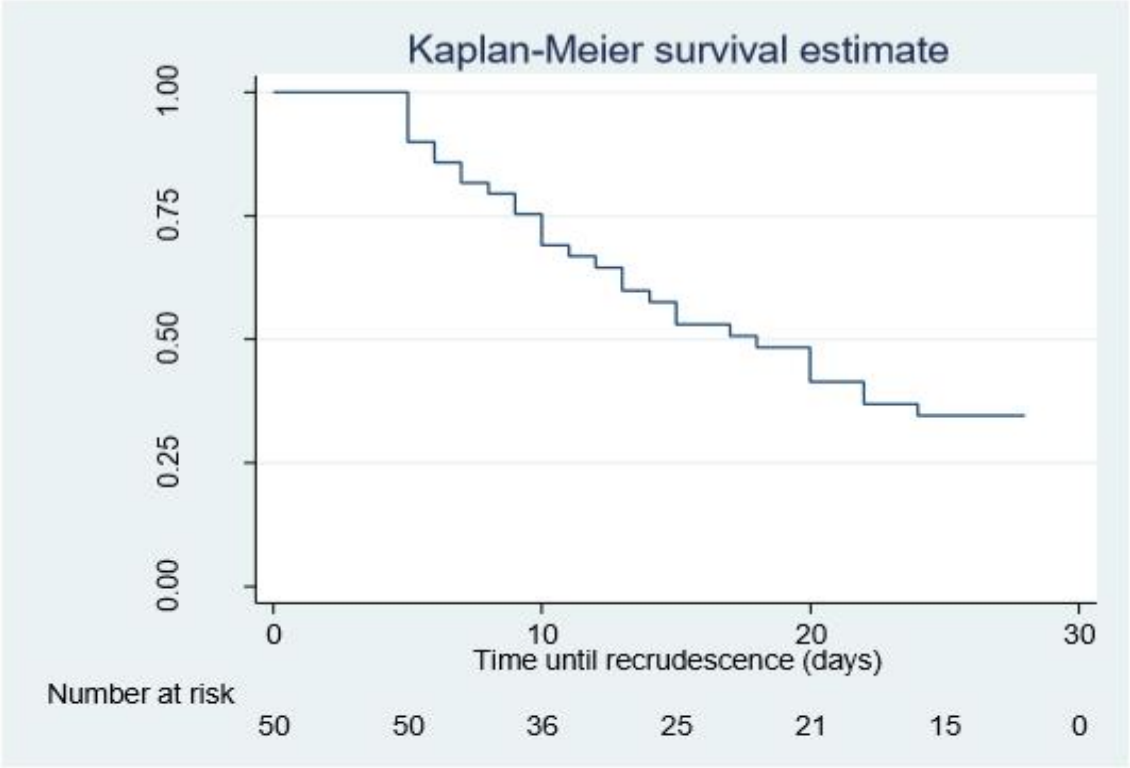
Kaplan-Meier survival curve: Calculation
STATA command: K-M estimate of *S(t)*

```
stset time, fail(recrud)
sts list
```

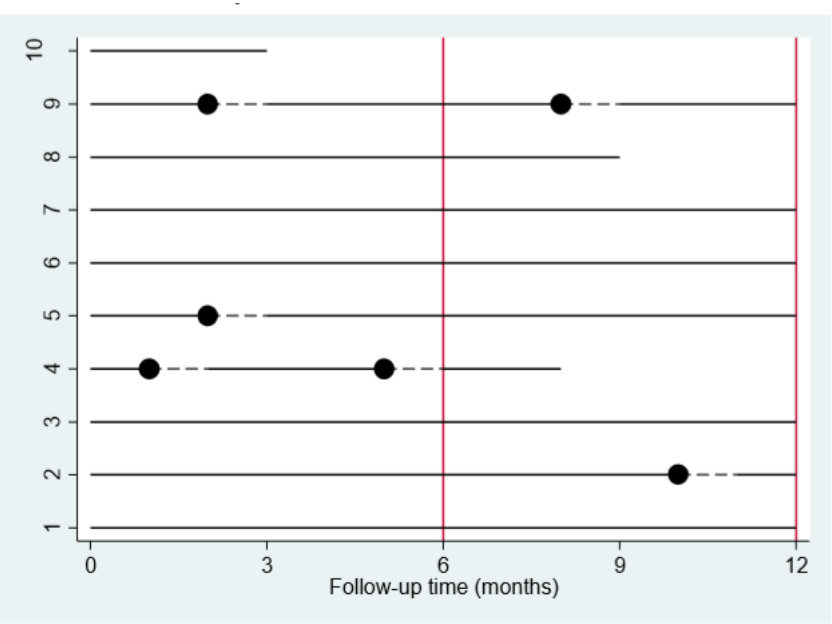
Time	Beg. Total	Fail	Net Lost	Survivor Function	Std. Error	[95% Conf.	Int.]
5	50	5	2	0.9000	0.0424	0.7763	0.9571
6	43	2	0	0.8581	0.0497	0.7252	0.9298
7	41	2	0	0.8163	0.0554	0.6764	0.9000
8	39	1	0	0.7953	0.0578	0.6528	0.8843
9	38	2	0	0.7535	0.0619	0.6066	0.8519
10	36	3	3	0.6907	0.0665	0.5400	0.8007
11	30	1	0	0.6677	0.0682	0.5154	0.7818
12	29	1	0	0.6447	0.0696	0.4912	0.7625
13	28	2	0	0.5986	0.0718	0.4442	0.7229
14	26	1	0	0.5756	0.0727	0.4213	0.7026
15	25	2	0	0.5295	0.0738	0.3765	0.6612
17	23	1	0	0.5065	0.0741	0.3546	0.6400
18	22	1	0	0.4835	0.0742	0.3331	0.6185
20	21	3	0	0.4144	0.0735	0.2705	0.5524
22	18	2	0	0.3684	0.0722	0.2305	0.5068
24	16	1	0	0.3453	0.0713	0.2111	0.4835
28	15	0	15	0.3453	0.0713	0.2111	0.4835

Kaplan-Meier survival curve: Calculation
STATA command: Plot K-M estimate of *S(t)*

```
stset time, fail(recrud)
sts graph, risktable(0(5)30) ///
        xtitle("Time until recrudescence (days)") ///
        ytitle("Survival Probability")
```



- K-M survival curve (& cox regression) suitable for analysis of time to 1st event
- Recurrent events survival analysis
  - Events occur more than once per subject during follow-up period
  - Example: vivax malaria infections



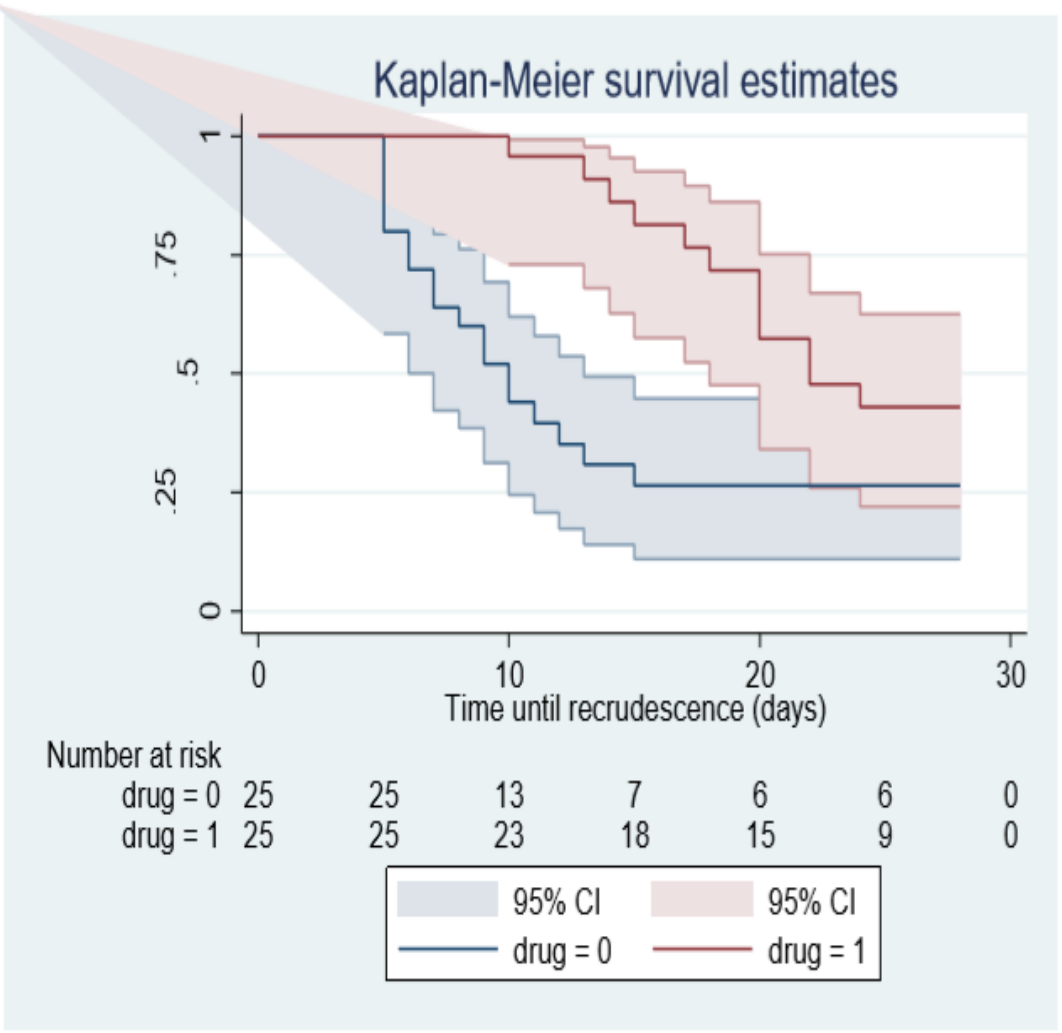
Follow-up of 10 patients with recurrent vivax malaria infections

• Black dot – patient had the event (recrudescence)

• Dashed line – period when patient is not a risk for vivax recurrence

• Vertical red line – end of follow-up (either 6 or 12 months)

Graphical comparison of K-M survival curves



- Combination therapy: drug=~~0~~ 1
- Monotherapy: drug=~~1~~ 0

sts graph, ci by(drug) risktable(0(5)30) /// xtitle("Time until recrudescence (days)") /// ytitle("Survival Probability")

Graphical comparison of K-M survival curves

- Patients that received combination therapy have longer time to recrudescence
- Differences in survival curves is not constant over time
  - Same until day 5 after treatment, then they deviate
- Can compare survival in the two groups at a single time point
  - Inefficient, do not use information on survival at remaining time points
- How do we compare the entire survival curve between groups?

Statistical comparison of K-M survival curves

- Compare hazard functions, h(t), not survivor functions
- Why?
  - Can assume ratio of hazards in the two groups is constant over time
  - Known as the **proportional hazards assumption**

Comparing groups: Proportional hazards assumption

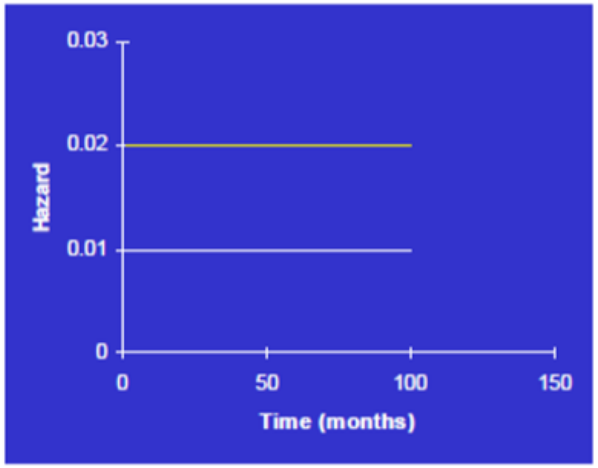
$$HR = \frac{h_1(t)}{h_0(t)} = \text{constant at all times } t$$

Hazard ratio HR

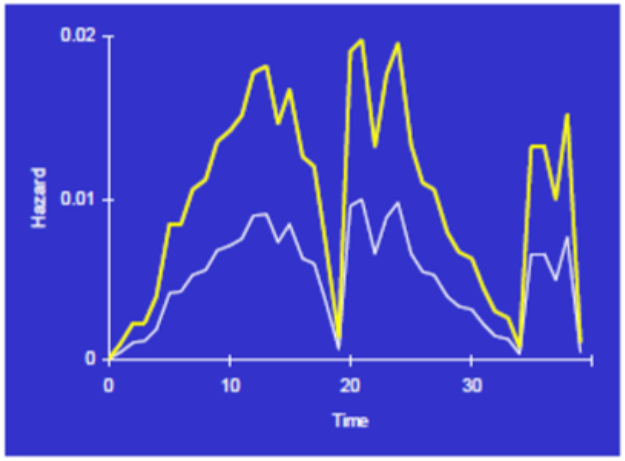
- Reporting a single HR is only useful if it does not vary with time
- If HR varies over time (i.e. HR(t)), then time is an effect modifier of the association between the exposure and outcome (referred to as non-proportional hazards)

Proportional Hazards

- Constant hazard

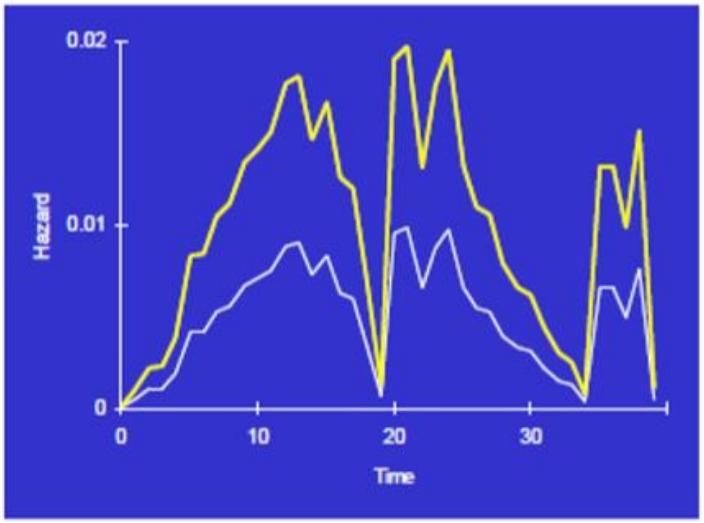


- Hazard varies over time

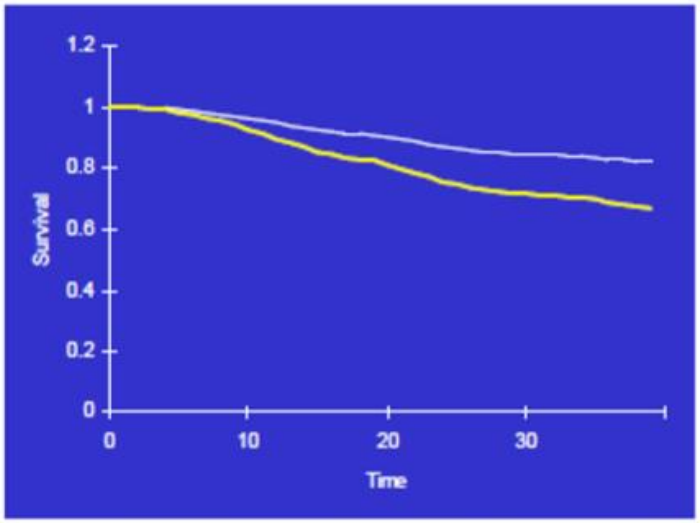


Proportional Hazards

- Hazard function *h(t)*



- Survivor function *S(t)*



Statistical comparison of hazards: Log rank (Mantel-Cox) test

- Log-rank test:
  - Evaluates whether or not K-M survival curves for two or more groups are statistically equivalent
  - Only valid if proportional hazards assumption is satisfied
- Null Hypothesis: HR = 1
  - No difference in population hazard functions between groups (equivalently, no difference between population survivor functions)
- Log-rank compares observed failures with expected failures calculated under the null hypothesis

Log rank test: STATA commands

```
sts test drug

         failure _d:  recrud
   analysis time _t:  time
```

Null Hypothesis H_0: $HR = 1$

```
Log-rank test for equality of survivor functions

      |   Events          Events
drug  |   observed        expected
------+-------------------------
0     |         18           11.07
1     |         12           18.93
------+-------------------------
Total |         30           30.00

           chi2(1) =         7.41
           Pr>chi2 =       0.0065
```

- **Events observed:** $O_k = \sum_i O_{ki}$, for k=0,1
- **Events expected:** $E_k = \sum_i E_{ki}$, for k = 0,1
- **chi2(1)** = U^2/V (log-rank test statistic)

Mantel-Cox estimate of hazard ratio: STATA commands

```
stmc drug

        failure _d:  recrud
  analysis time _t:  time

Mantel-Cox comparisons

Mantel-Haenszel estimates of the rate ratio
   comparing drug==1 vs. drug==0
   controlling for time (by clicks)

Overall Mantel-Haenszel estimate, controlling for time

------------------------------------------------------------
     RR       chi2       P>chi2      [95% Conf. Interval]
------------------------------------------------------------
  0.335       7.09       0.0078         0.144       0.781
------------------------------------------------------------
```

This value is slightly different to χ^2 statistic obtained using `sts test drug`

HR_{MC}

Log-rank test and Mantel-Cox estimate of hazard ratio: Summary

- Log-rank test assumes proportional hazards
- Null Hypothesis H0: Population survivor/hazard functions are equal or, equivalently, HR = 1
- sts test: STATA command to perform the log-rank test
- stmc: STATA command to calculate the Mantel-Cox estimate of the HR & log-rank test
- Only need to know how to interpret output from sts test and stmc

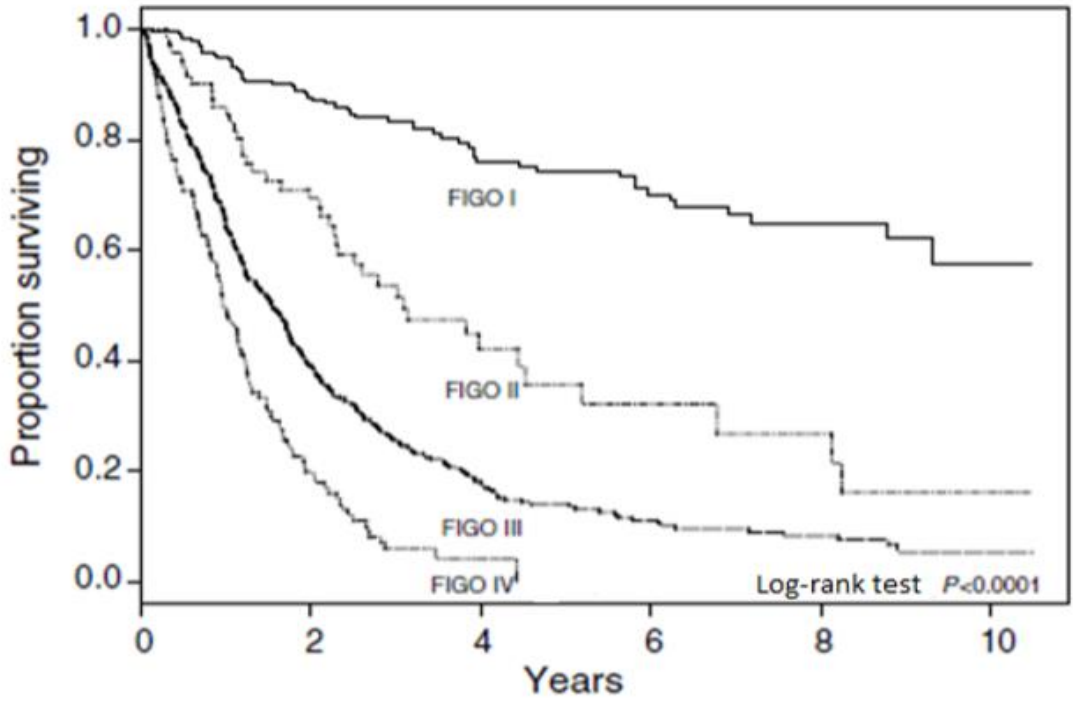
Example: Ovarian cancer survival according to FIGO stage

- Eligibility criteria:
  - Diagnosed with primary epithelial ovarian carcinoma at Western General Hospital in Edinburgh
- Recruitment: Between 1990 and 1999, 825 patients were enrolled
- The primary study end point was survival (death)
- End of follow-up: December 2000
- Aim: Compare survival according to FIGO (Federation of Gynecologists and Obstetricians) staging of ovarian cancer
  - Categorical variable with 4 stages
  - Severity of ovarian cancer increases from Stage I (e.g. confined to ovaries and fallopian tubes) to IV (cancer spread to other organs)

Clark TG et al. Survival Analysis Part 1: Basic concepts and first analyses. BJC, 89:232-238

K-M survival curve

Survival in ovarian cancer study according to FIGO stage



Number at risk at time (years)	0	2	4	6	8
FIGO I	200	151	93	66	44
FIGO II	72	45	17	10	6
FIGO III	414	142	48	23	14
FIGO IV	123	22	3	1	1

Questions on previous K-M survival curves

- Which FIGO stage has the worst survival?
- What is the null hypothesis of the log-rank test?
- What is the median survival for patients with FIGO stage III and IV ovarian cancer?
- Can the median survival be calculated for FIGO stage I?
- What is the 5 year survival for FIGO stage II?
- Comparing the curves and the number at risk table would you say there is:
  - a) very little censoring;
  - b) moderate amount of censoring; or
  - c) a great deal of censoring?

Summary

- Survival analysis: – Outcome is time to an event
  - Focuses on survivor function S(t) and hazard function h(t)
  - Does not assume the hazard is constant over time
- Life tables:
  - Exact time of event is not known
- Kaplan-Meier survival curve:
  - Exact time of event is known
- Compare survival using log-rank (Mantel-Cox) test
  - Requires hazards are proportional (constant hazard ratio)

Cox Proportional Hazard Analysis (Cox Regression)

Cox proportional hazard Analysis (Cox Regression)

- Assess effects of exposures with three or more levels
- Control for confounding variables
- Test the proportional hazards assumption

Bowel cancer in Victoria, 1995-2004 follow-up censored five years after diagnosis

- We are interested in the influence of period of diagnosis on survival for people diagnosed with bowel cancer
- Population: Victorian adults, diagnosed with bowel cancer between 1995 and 2004.
- Exposure: Period of diagnosis (1995-1999 vs 2000-2004)
- Outcome: (Time to) Death
- Potential confounders

  Measured:

  Age at diagnosis (< 55, 55-64, 65-74, 75+)

  Sex

  Place of residence (Melbourne versus the bush),

  Country of birth (Aus./NZ, UK/Ire., South Europe, other)

  Unmeasured: Stage at diagnosis

Bowel cancer in Victoria, 1995-2004

follow-up censored five years after diagnosis

```
. stset survdays, fail(dead) scale(365.25)
   id(id)

        failure event:  dead != 0 & dead < .
obs. time interval:  (0, survdays]
 exit on or before:  failure
    t for analysis:  time/365.25

--------------------------------------------------------------------------
     1948  total obs.
        0  exclusions
--------------------------------------------------------------------------
     1948  obs. remaining, representing
     1948  subjects
      817  failures in single record/single failure data
 4639.379  total analysis time at risk, at risk from t =         0
                              earliest observed entry t =         0
                                   last observed exit t =         5
```

ALWAYS check the output carefully.

- Is anyone excluded?
- Do the times make sense?
- Is the number of failures correct?

Cox regression – risk sets

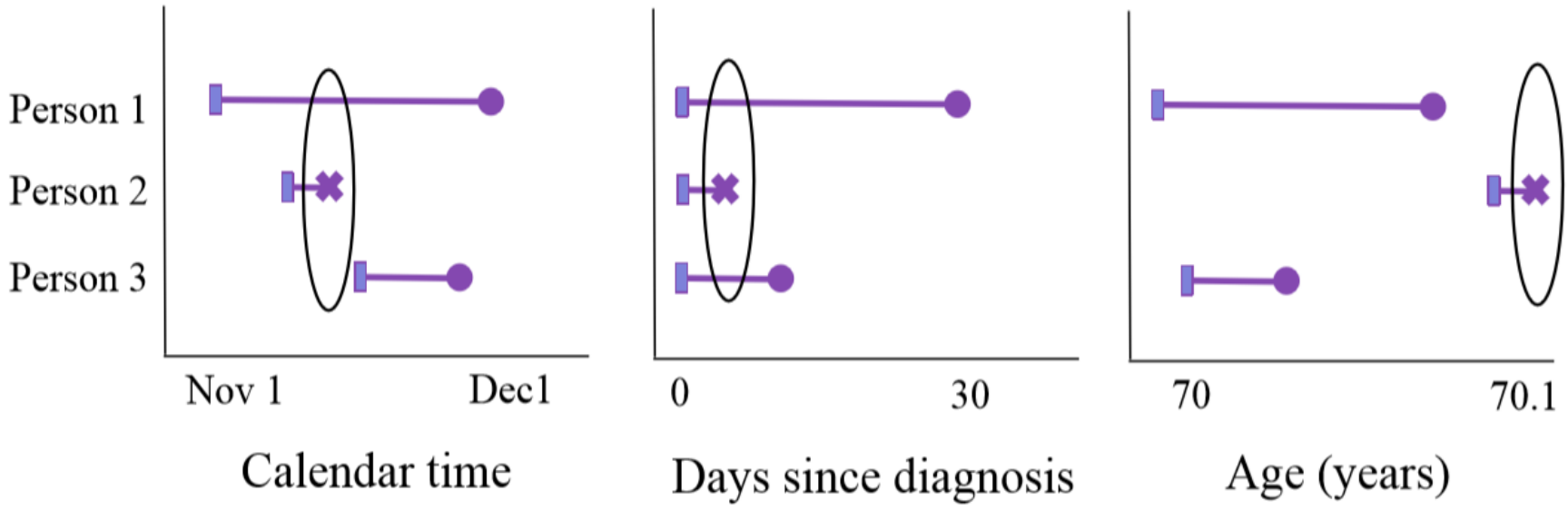
Cox regression is based on comparing people within risk sets:

- Each failure time (i.e. time of death) has its own risk set, comprising every study participant who is at risk of the event (death) at that time.
- E.g. Day 2: 1 death (2 days after diagnosis)
  Day 3: 1 subject censored (due to emigration)
  Day 5: 1 death
  - 1st failure time (day 2): the risk set contains every study participant.
  - 2nd failure time (day 5): the risk set contains every study participant except the person who has already died, and the person who has been censored.

What do we mean when we say another person was at risk "at that time"?

- On the same calendar day?
- At the same number of days since a particular event?
- At the same age?

Choice of time axis



Choice of time axis

On which timescale are the hazards (likely to be) proportional?
Best choice usually scale over which hazard changes most

Survival study

- Time since diagnosis often most relevant
- Risk set at time t contains people surviving to time t after diagnosis

RCT

- Time since randomization natural scale
- Risk set at time t contains people surviving to time t after randomization

Cohort study (long term)

- Age is often the natural scale
  - Risk set at age t contains people surviving to age t
- Time since first exposed
  - Risk set at time t contains people surviving to time t after initial exposure

Common time scales in epidemiology

Origin	Time scale
Birth	Age
Any fixed date	Calendar time
First exposure	Time exposed
Entry into study	Time in study
Disease onset	Time since onset
Diagnosis	Time since diagnosis
Start of treatment	Time on treatment

Cox regression

Key Features of Cox Regression

- We do not make any assumptions about how the hazard function varies over time.
- Instead, we make the Proportional Hazards assumption; that the exposure effects (hazard ratios) are constant over time.

- At each event time, we compare exposure values for the participant experiencing the event with values for all other participants who are at risk of the event at that time (in the risk set for that event time).
- Under the proportional hazards assumption, we can estimate the hazard ratio by combining information across all event times.

Cox regression: mathematical form

- Let's consider a single binary variable x (0=unexposed, 1=exposed)

$$\boxed{h(t) = h_0(t)\exp(\beta x)}$$

If $x=0$, then $h(t) = h_0(t) \times \exp(\beta \times 0) = h_0(t) \times \exp(0) = h_0(t)$

If $x=1$, then $h(t) = h_0(t) \times \exp(\beta \times 1) = h_0(t) \times \exp(\beta)$

- The hazard ratio for exposure x is

$$HR = \frac{\text{hazard}(x=1)}{\text{hazard}(x=0)} = \frac{h_0(t)\exp(\beta)}{h_0(t)} = \exp(\beta)$$

so β is the logarithm of the hazard ratio for the exposure x.

- $h_0(t)$ is called the baseline hazard: the hazard for a subject with $x=0$.

Cox regression: mathematical form

The Cox model is:

$$h(t) = h_0(t) \times \exp(\beta_1 x_1 + \beta_2 x_2 + \ldots \beta_p x_p)$$

- $h(t)$ = hazard at time t
- $h_0(t)$ is the **baseline hazard**: the hazard for a subject with all x's = 0.

- x_1, x_2, etc are the predictor variables ("exposures")
- β_1, β_2, etc are the logarithms of the hazard ratios
- These parameters are estimated using (partial) maximum likelihood techniques.

We sometimes write the Cox model on the log scale:

$$\log(h(t)) = \log(h_0(t)) + \beta_1 x_1 + \beta_2 x_2 + \ldots \beta_p x_p$$

Bowel cancer

Influence of period of diagnosis on survival

```
. stset survdays, fail(dead) scale(365.25) id(id)
. stmc period
```

RR	chi2	P>chi2	[95% Conf.	Interval]
0.841	5.76	0.0164	0.731	0.969

```
. stcox period
No. of subjects =         1948                     Number of obs    =      1948
No. of failures =          817
Time at risk    =  4639.379192
                                                   LR chi2(1)       =      5.79
Log likelihood  =     -5823.9                      Prob > chi2      =    0.0161
```

_t	Haz. Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
period	.8397592	[illegible]97	-2.40	0.017	.7279876	.9686917

The interpretation is similar to that for Poisson regression

Cox regression: controlling for confounders

```
. xi: stcox period i.agegp sex metro i.country
No. of subjects =         1947                 Number of obs    =      1947
No. of failures =          816
Time at risk    =  4635.428474
                                               LR chi2(9)       =    152.42
Log likelihood  =   -5743.6063                 Prob > chi2      =    0.0000
------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      period |   .8545394   .0626477    -2.14   0.032      .740166    .9865863
   _Iagegp_2 |   1.349034   .2143243     1.88   0.060     .9880761    1.841854
   _Iagegp_3 |   1.496459   .2154028     2.80   0.005     1.128604    1.984213
   _Iagegp_4 |   2.519621   .3469547     6.71   0.000      1.92364    3.300248
         sex |   .9066998   .0639988    -1.39   0.165     .7895544    1.041226
       metro |   1.194406   .0887854     2.39   0.017     1.032472    1.381737
 _Icountry_2 |    1.14059   .1453568     1.03   0.302     .8884897    1.464221
 _Icountry_3 |   .9931371   .1240303    -0.06   0.956      .777508    1.268568
 _Icountry_4 |   .4944224   .0582203    -5.98   0.000     .3925234    .6227744
------------------------------------------------------------------------------
```

- Interpretation?
- Additionally adjusting for measured confounders had almost no effect on the hazard ratio for period of diagnosis. But we haven't included stage of diagnosis.

Proportional hazards (PH) assumption

- PH assumption (for two exposure groups):

$$\frac{h_1(t)}{h_0(t)} = \text{constant}, \text{ or } h_1(t) = h_0(t) \times \text{constant}$$

- Non-proportional hazards correspond to interaction between exposure variable & time
  - i.e., the hazard ratio changes with time
  - e.g. the relative effect of treatment weakens over time

- Ways to assess the PH assumption:
  i. Visually
  ii. Using a statistical test

Visual check of PH assumption

- Easier to check using the cumulative hazards
- We can estimate the cumulative hazards using the Nelson-Aalen estimate.

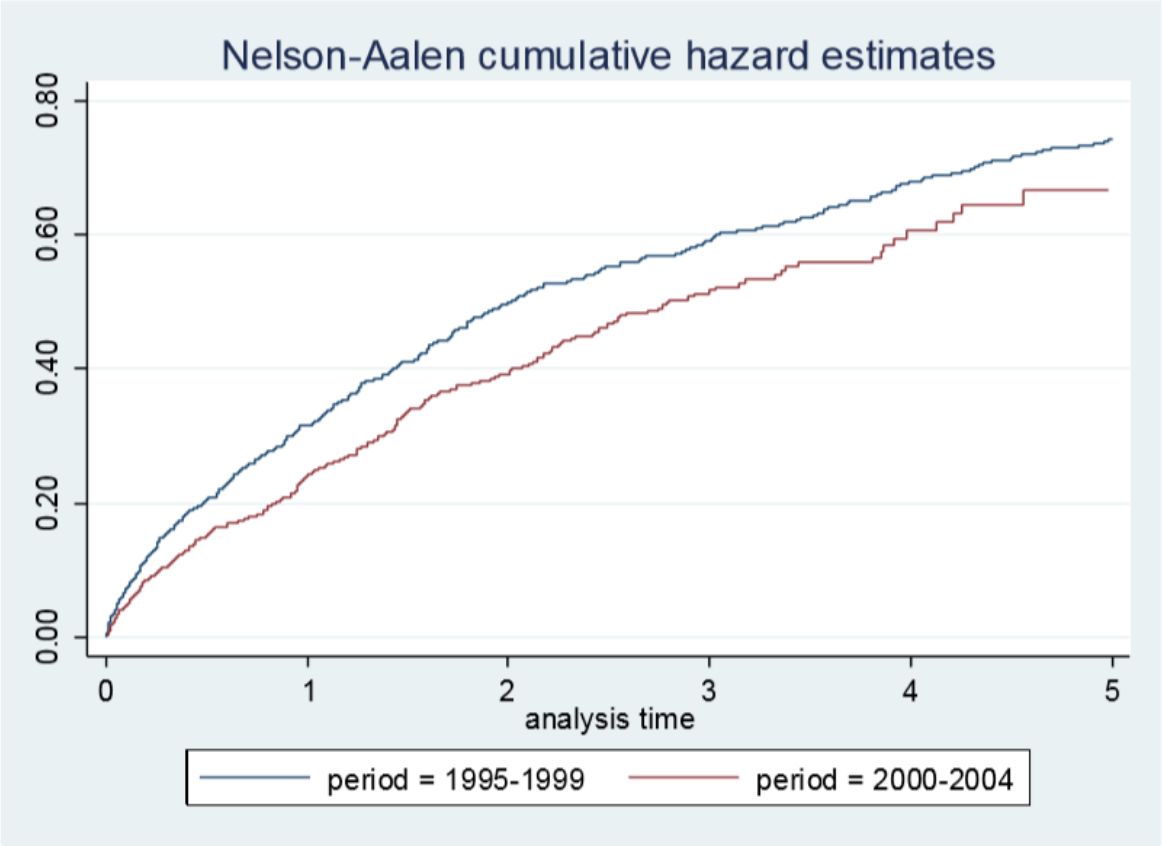
- If $h_1(t) = h_0(t) \times \text{constant}$

  then also $H_1(t) = H_0(t) \times \text{constant}$

- Therefore, a plot of **cumulative hazard** against **time** for each exposure category should be proportional.

Cumulative hazard versus time
Are the curves proportional?

```
. sts graph, cumhaz by(period)
```



Easier visual check of PH assumption

- Under the PH assumption, $H_1(t) = H_0(t) \times \text{constant}$

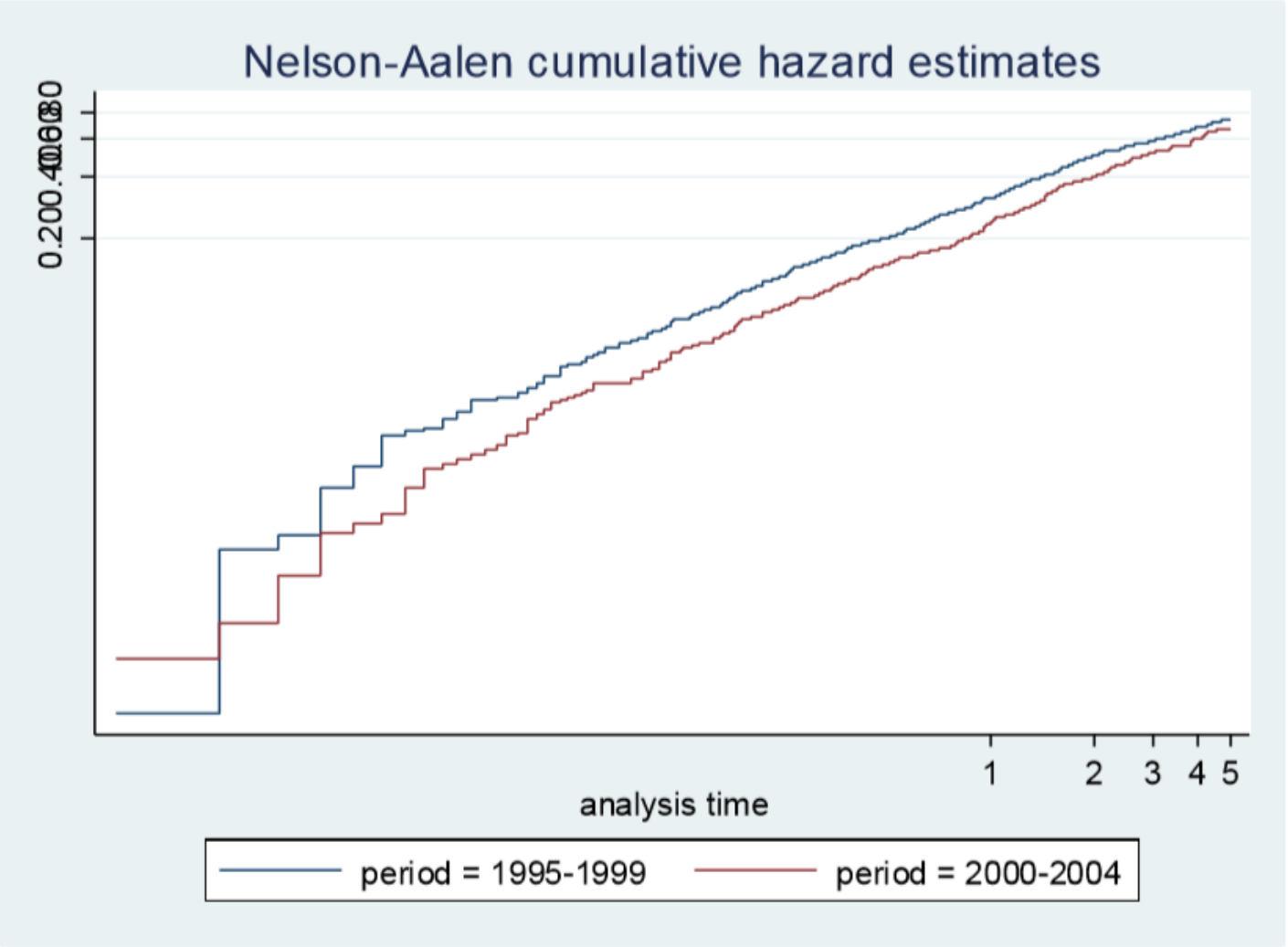
  Taking logs, this is equivalent to

$$\log(H_1(t)) = \log(H_0(t)) + \log(\text{constant})$$

- Therefore, a plot of **log(cumulative hazard)** against **time** for each exposure category should have parallel lines
- We often **log(cumulative hazard)** against **log(time)** to obtain a nicer graph
  (this stretches out the left-hand side of the graph).

Log(cumulative hazard) versus log(time)
Are the curves parallel?

```
. sts graph, cumhaz by(period) yscale(log) xscale(log)
```



Statistical tests for proportional hazards assumption

- We want to test whether the hazard ratio varies with time. Is there an interaction between exposure and time?

- We will cover two ways of testing this:
  - Stata command: estat phtest (especially estat phtest, detail) based on Schoenfeld residuals

Testing proportional hazards assumption using estat phtest

```
xi: stcox period i.agegp sex metro i.country
```

```
. estat phtest

Test of proportional-hazards assumption

------------------------------------------------------------------
             |                  chi2       df       Prob>chi2
-------------+----------------------------------------------------
global test  |                 10.79        9          0.2901
------------------------------------------------------------------
```

Null hypothesis:

Hazards for each variable are proportional (e.g. for period, $h_1(t)/h_0(t)$ = constant)

Dealing with non-proportional hazards

If variable is a confounder:

- Stratify on it
  - This creates separate risk sets according to the values of the variable
  - We assume that the hazard ratios for other exposures are the same within each strata
  - We no longer calculate a hazard ratio for the stratifying variable

If the variable is of interest:

- Fit an interaction between the variable and time – Split the follow-up time into different periods then fit interaction – Stata's "tvc" option fits the interaction without having to split the data

Summary

- For a binary exposure, Mantel-Cox (log rank) method & Cox regression give similar results
- Cox regression is a flexible regression method for dealing with multiple variables (like Poisson & logistic regression)
- Can examine the proportional hazards assumption:
  - by plotting log(H(t)) against time (or log(time))
  - by plotting log(-log(S(t))) against time (or log(time))
  - using statistical tests based on Schoenfeld residuals or...
  - by fitting an interaction between the exposure and time

THANK YOU!