Lecture 2

Monday, September 11, 2023 10:01

BIOS522_Slides2

EMORY ROLLINS SCHOOL OF PUBLIC HEALTH | Department of Biostatistics and Bioinformatics

BIOS 522: Survival Analysis Methods

Lecture 2:

The survival function

Previously

- Examples of time-to-event analyses in practice
- Introduction to the time-to-event data format
- Review the analysis of continuous data, cumulative incidence data, and incidence rate data
- Right censoring and censoring notation

2

Today's learning objectives

- *Define the survival function*
- *Estimate the survival function in the absence of censoring*
- *Estimate the survival function in the presence of censoring*
- *Interpret estimated survival functions in the literature*
- *Characterize whether censoring is informative or non-informative*

Cumulative distribution function (CDF)

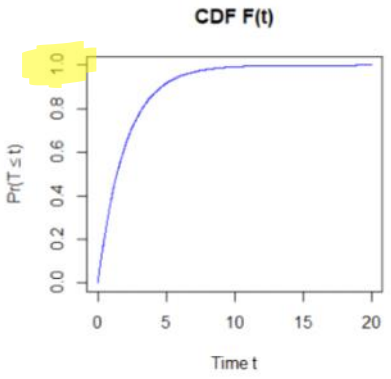
- Measures the probability of failing by time t

true survival variable

time t

$$F(t) = \Pr(T \leq t)$$

- For time-to-event data, the CDF starts at 0 at time 0, and then increases to maximum value of 1



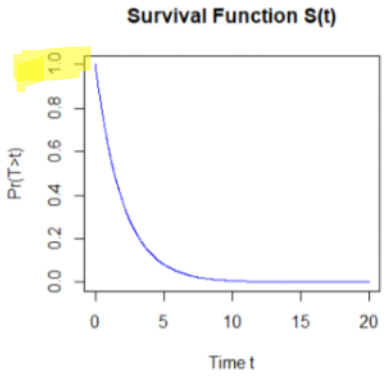
Survival function $S(t)$

- Measures the probability of surviving past time t

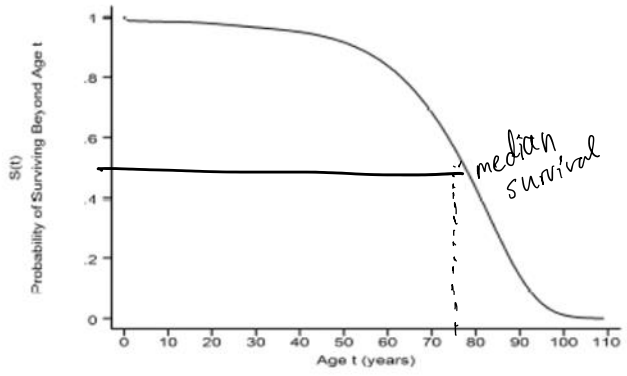
$$S(t) = \Pr(T > t)$$

- The survival function is like the CDF but flipped

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



Survival curve example - Human survival

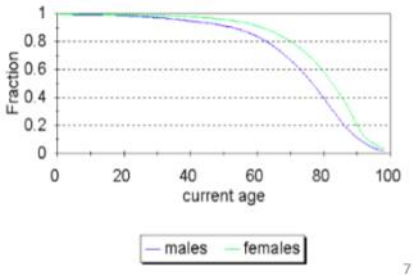


- Probability of surviving beyond 10 years of age, $S(10) = 0.99$
- Probability of surviving beyond 80 years of age, $S(80) = 0.39$
- Probability of surviving beyond 100 years of age, $S(100) = 0.01$
- Median survival is 75 years of age: $S(75) = 0.50$

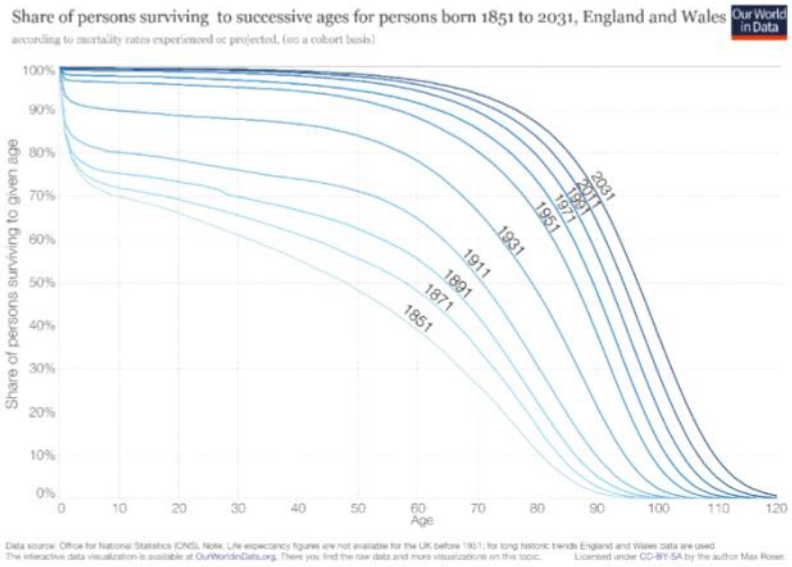
Why are we interested in $S(t)$?

- Data from a single population (**one-sample problem**)
  - Estimate the survival probability at a **landmark time** (e.g. 5 years)
  - Estimate median survival

- Data from two independent populations (**two-sample problem**)
  - Estimate survival for each group separately
  - Test if survival is equal across the two groups



7



Observed and predicted human mortality data for persons born between 1851 and 2031 in England and Wales.

can look at how survival has improved over time based on these separate curves

(cohorts born at diff. times)

8

Source: https://ourworldindata.org/life-expectancy-how-is-it-calculated-and-how-should-it-be-interpreted

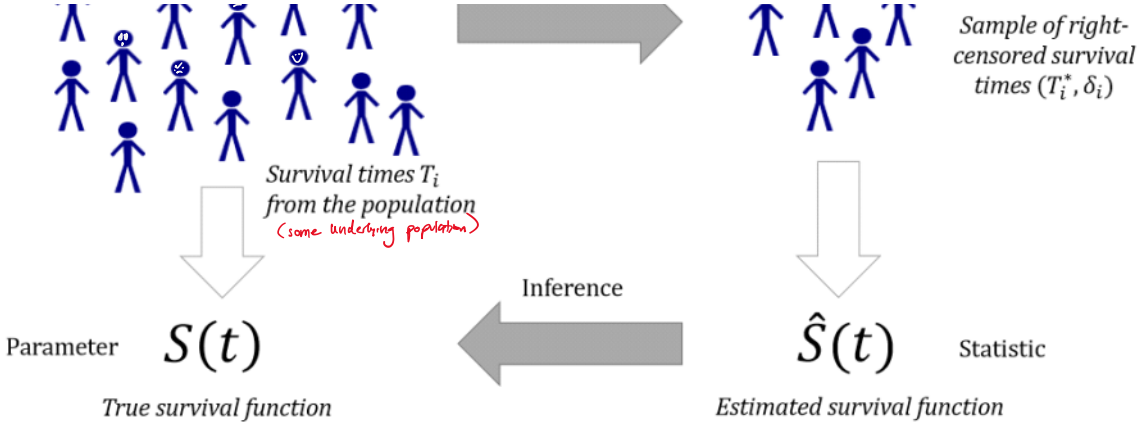
Statistical inference

We want to know about these

Random selection

We have these to work with

Sample of right-censored survival times (T_i^*, δ_i)



Adapted from: https://www.cliffsnotes.com/study-guides/statistics/sampling/populations-samples-parameters-and-statistics

9

Estimators of $S(t)$

In the absence of censoring:

- One minus the **empirical CDF**

In the presence of censoring:

- **Kaplan-Meier estimator**

10

Empirical CDF

- To estimate the probability of having failed at time t, we measure the proportion of the sample that has failed by time t

$$\hat{F}(t) = \frac{\sum_{i=1}^{n} I[T_i \leq t]}{n}$$

sum of indicators

← makes it a proportion

- No censoring allowed

- No censoring allowed
- Estimate $\hat{S}(t) = 1 - \hat{F}(t)$

11

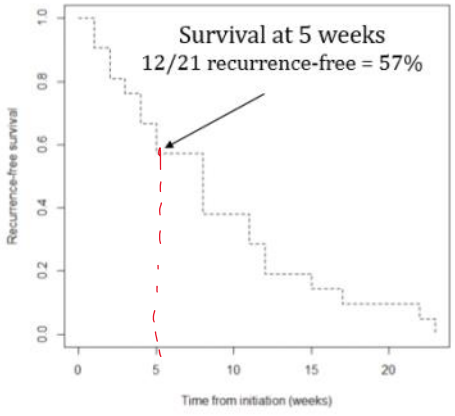
Leukemia data

Time from initiation to remission/recurrence (weeks)

1, 1, 2, 2, 3, 4, 4, 5, 5, 8, 8, 8, 8, 11, 11, 12, 12, 15, 17, 22, 23

12

Leukemia example



- Empirical CDF "Step function"
  - Starts at 0
  - Increases when a failure occurs
  - Reaches 1 when all have failed
- For $\hat{S}(t) = 1 - \hat{F}(t)$
  - Starts at 1
  - Decreases when a failure occurs
  - Reaches 0 when all have failed

13

Another calculation approach

Time 0	Time 1	Time 2	Time 3	Time 4	Time 5
21 at risk	*21 at risk* *19 survived*	*19 at risk* *17 survived*	*17 at risk* *16 survived*	*16 at risk* *14 survived*	*14 at risk* *12 survived*
X X X X X X X X X X X X X X X X X X X X X	X X X X X X X X X X X X X X X X X X X X X	X X X X X X X X X X X X X X X X X X X X X	X X X X X X X X X X X X X X X X X	X X X X X X X X X X X X X X X X	X X X X X X X X X X X X X X
$\left(\frac{21}{21}\right)$ = 100%	$\left(\frac{19}{21}\right)$ = 90.5%	$\left(\frac{19}{21}\right)\left(\frac{17}{19}\right)$ $= \left(\frac{17}{21}\right) = 81.0\%$	$\left(\frac{19}{21}\right)\left(\frac{17}{19}\right)\left(\frac{16}{17}\right)$ $= \left(\frac{16}{21}\right) = 76.2\%$	$\left(\frac{19}{21}\right)\left(\frac{17}{19}\right)\left(\frac{16}{17}\right)\left(\frac{14}{16}\right)$ $= \left(\frac{14}{21}\right) = 66.7\%$	$\left(\frac{19}{21}\right)\left(\frac{17}{19}\right)\left(\frac{16}{17}\right)\left(\frac{14}{16}\right)\left(\frac{12}{14}\right)$ $= \left(\frac{12}{21}\right) = 57.1\%$

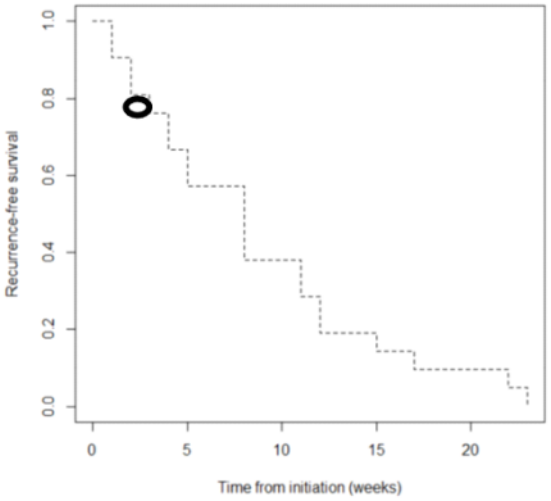
Leukemia failure times (in weeks)

1, 1, 2, 2, 3, 4, 4, 5, 5, 8, 8, 8, 8, 11, 11, 12, 12, 15, 17, 22, 23

14

What happens when there is censoring?

- After someone is censored, we do not know when/if they fail
- How do we modify our analyses accordingly?



15

Rethinking our calculations with censoring

Time 0	Time 1	Time 2	Time 3	Time 4	Time 5
21 at risk	*21 at risk* *19 survived*	*19 at risk* *17 survived* one is censored	*16 at risk* *15 survived*	*15 at risk* *13 survived*	*13 at risk* *11 survived*
X X X X X X X X X X X X X X X X X X X X X	X X X X X X X X X X X X X X X X X X X X X	X X X X X X X X X X X X X X X X X X X	X X X X X X X X X X X X X X X X	X X X X X X X X X X X X X X X	X X X X X X X X X X X X X
$\left(\frac{21}{21}\right)$ = 100%	$\left(\frac{19}{21}\right)$ = 90.5%	$\left(\frac{19}{21}\right)\left(\frac{17}{19}\right)$ = 81.0%	$\left(\frac{19}{21}\right)\left(\frac{17}{19}\right)\left(\frac{15}{16}\right)$ = 75.9%	$\left(\frac{19}{21}\right)\left(\frac{17}{19}\right)\left(\frac{15}{16}\right)\left(\frac{13}{15}\right)$ = 65.8%	$\left(\frac{19}{21}\right)\left(\frac{17}{19}\right)\left(\frac{15}{16}\right)\left(\frac{13}{15}\right)\left(\frac{11}{13}\right)$ = 55.7%

Leukemia failure times (in weeks)

1, 1, 2, 2, 2+, 3, 4, 4, 5, 5, 8, 8, 8, ~~8~~, 11, 11, 12, 12, 15, 17, 22, 23

someone who failed at time 8 was actually censored at time 2

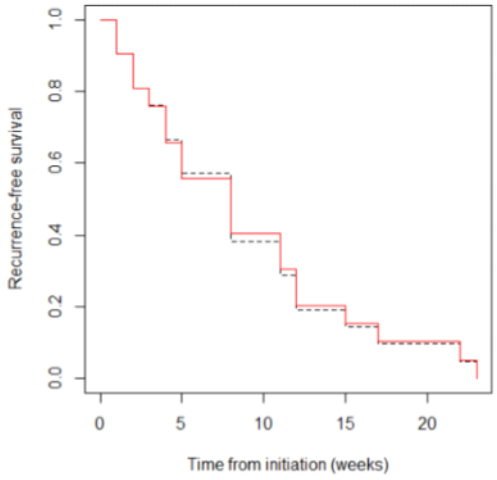
Rethinking our calculations with censoring

Time 0	Time 1	Time 2	Time 3	Time 4	Time 5
21 at risk	*21 at risk* *19 survived*	*19 at risk* *17 survived*	*16 at risk* *15 survived*	*15 at risk* *13 survived*	*13 at risk* *11 survived*
X X X X X X X X X X X X X X X X X X X X X	X X X X X X X X X X X X X X X X X X X X X	X X X X X X X X X X X X X X X X X X X	X X X X X X X X X X X X X X X X	X X X X X X X X X X X X X X X	X X X X X X X X X X X X X
$\left(\frac{21}{21}\right)$ = 100%	$\left(\frac{19}{21}\right)$ = 90.5%	$\left(\frac{19}{21}\right)\left(\frac{17}{19}\right)$ = 81.0%	$\left(\frac{19}{21}\right)\left(\frac{17}{19}\right)\left(\frac{15}{16}\right)$ = 75.9% (compare to 76.2%)	$\left(\frac{19}{21}\right)\left(\frac{17}{19}\right)\left(\frac{15}{16}\right)\left(\frac{13}{15}\right)$ = 65.8% (compare to 66.7%)	$\left(\frac{19}{21}\right)\left(\frac{17}{19}\right)\left(\frac{15}{16}\right)\left(\frac{13}{15}\right)\left(\frac{11}{13}\right)$ = 55.7% (compare to 57.1%)

Leukemia failure times (in weeks)

1, 1, 2, 2, 2+, 3, 4, 4, 5, 5, 8, 8, 8, ~~8~~, 11, 11, 12, 12, 15, 17, 22, 23

Note the slight difference in estimates…



- Black dotted line is with full data and empirical CDF
- Red solid line is adjusting for censoring

Kaplan-Meier estimator

for censored data

$$\hat{S}(t) = \prod_{j:t_j \le t} \left(1 - \frac{d_j}{n_j}\right)$$

failures

interval?

AT RISK

Where n_j is the number at risk at the start of the interval, and d_j is the number who fail at the end of the interval

Also known as the **product-limit estimator**

Kaplan-Meier construction

Unique failure/ censoring time t_j	Number at risk n_j during $(t_{j-1}, t_j]$	Number of deaths d_j at t_j	Number censored c_j at t_j	Conditional survival probability $\hat{q}_j$	Kaplan-Meier estimate $[t_j, t_{j+1})$
$t_0 = 0$					$t = [0,1)$ $\hat{S}(t) = 1$
$t_1 = 1$	$t = (0,1]$ $n_1 = 21$	$d_1 = 2$	$c_1 = 0$	$\hat{q}_1 = \left(1 - \frac{2}{21}\right)$	$t = [1,2)$ $\hat{S}(t) = \widehat{q_1}$
$t_2 = 2$	$t = (1,2]$ $n_2 = 19$	$d_2 = 2$	$c_2 = 1$	$\hat{q}_2 = \left(1 - \frac{2}{19}\right)$	$t = [2,3)$ $\hat{S}(t) = \widehat{q_1 q_2}$
$t_3 = 3$	$t = (2,3]$ $n_3 = 16$	$d_3 = 1$	$c_3 = 0$	$\hat{q}_3 = \left(1 - \frac{1}{16}\right)$	$t = [3,4)$ $\hat{S}(t) = \widehat{q_1 q_2 q_3}$

Gaining insight into Kaplan-Meier

- It is instructive to think about how the Kaplan-Meier estimator places mass at observed failure times
- Efron (1967) describes the "Redistribution of Mass" algorithm

Redistribution of mass algorithm

- **Step 1:** Arrange data in increasing order, with censored observations to the right of uncensored observations in the case of ties
- **Step 2:** Put mass 1/n at each observation
- **Step 3:** Start from the smallest observation and move "right". Each time a censored observation is reached, redistribute its mass evenly to all observations to the right.
- **Step 4:** Repeat Step 3 until all censored observations (except largest observations) have no mass. If the largest time is censored, regard this mass as placed beyond the largest time.

Ten participants in the study

Step 1	2	2	3^+	5	5^+	7	9	16	16	18^+
Step 2	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$
Step 3	↓	↓	↪	$\frac{1}{70}$	$\frac{1}{70}$	$\frac{1}{70}$	$\frac{1}{70}$	$\frac{1}{70}$	$\frac{1}{70}$	$\frac{1}{70}$
	↓	↓	↓	↓	↪	$\frac{1}{5}(\frac{8}{70})$	$\frac{1}{5}(\frac{8}{70})$	$\frac{1}{5}(\frac{8}{70})$	$\frac{1}{5}(\frac{8}{70})$	$\frac{1}{5}(\frac{8}{70})$
Total Mass	$\frac{2}{10}$		0	$\frac{8}{70}$	0	$\frac{24}{175}$	$\frac{24}{175}$	$\frac{48}{175}$		Assume this is somewhere > 18

Split this person's mass among remaining ppl (+ 1/70 for everyone)

$\frac{1}{10} + \frac{1}{70}$

Each observation gets an equal amount of mass (1/10)

Step 1	2	2	3^+	5	5^+	7	9	16	16	18^+
Step 2	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$
Step 3	$\downarrow$	$\downarrow$	$\hookrightarrow$	$\frac{1}{70}$	$\frac{1}{70}$	$\frac{1}{70}$	$\frac{1}{70}$	$\frac{1}{70}$	$\frac{1}{70}$	$\frac{1}{70}$
	$\downarrow$	$\downarrow$		$\downarrow$	$\hookrightarrow$	$\frac{1}{5}(\frac{8}{70})$	$\frac{1}{5}(\frac{8}{70})$	$\frac{1}{5}(\frac{8}{70})$	$\frac{1}{5}(\frac{8}{70})$	$\frac{1}{5}(\frac{8}{70})$
Total Mass	$\frac{2}{10}$		0	$\frac{8}{70}$	0	$\frac{24}{175}$	$\frac{24}{175}$	$\frac{48}{175}$		Assume this is somewhere > 18

24

When someone fails, we allocate their mass to that time point

Step 1	2	2	3^+	5	5^+	7	9	16	16	18^+
Step 2	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$
Step 3	$\downarrow$	$\downarrow$	$\hookrightarrow$	$\frac{1}{70}$	$\frac{1}{70}$	$\frac{1}{70}$	$\frac{1}{70}$	$\frac{1}{70}$	$\frac{1}{70}$	$\frac{1}{70}$
	$\downarrow$	$\downarrow$		$\downarrow$	$\hookrightarrow$	$\frac{1}{5}(\frac{8}{70})$	$\frac{1}{5}(\frac{8}{70})$	$\frac{1}{5}(\frac{8}{70})$	$\frac{1}{5}(\frac{8}{70})$	$\frac{1}{5}(\frac{8}{70})$
Total Mass	$\frac{2}{10}$		0	$\frac{8}{70}$	0	$\frac{24}{175}$	$\frac{24}{175}$	$\frac{48}{175}$		Assume this is somewhere > 18

25

When an individual is censored, their mass (1/10) is split across the 7 remaining participants

	2	2	3^+	5	5^+	7	9	16	16	18^+
Step 1										
Step 2	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$
Step 3	$\downarrow$	$\downarrow$	$\hookrightarrow$	$\frac{1}{70}$	$\frac{1}{70}$	$\frac{1}{70}$	$\frac{1}{70}$	$\frac{1}{70}$	$\frac{1}{70}$	$\frac{1}{70}$
	$\downarrow$	$\downarrow$		$\downarrow$	$\hookrightarrow$	$\frac{1}{5}(\frac{8}{70})$	$\frac{1}{5}(\frac{8}{70})$	$\frac{1}{5}(\frac{8}{70})$	$\frac{1}{5}(\frac{8}{70})$	$\frac{1}{5}(\frac{8}{70})$
Total Mass	$\frac{2}{10}$		0	$\frac{8}{70}$	0	$\frac{24}{175}$	$\frac{24}{175}$	$\frac{48}{175}$		Assume this is somewhere > 18

	2	2	3^+	5	5^+	7	9	16	16	18^+
Step 1										
Step 2	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$
Step 3	$\downarrow$	$\downarrow$	$\hookrightarrow$	$\frac{1}{70}$	$\frac{1}{70}$	$\frac{1}{70}$	$\frac{1}{70}$	$\frac{1}{70}$	$\frac{1}{70}$	$\frac{1}{70}$
	$\downarrow$	$\downarrow$		$\downarrow$	$\hookrightarrow$	$\frac{1}{5}(\frac{8}{70})$	$\frac{1}{5}(\frac{8}{70})$	$\frac{1}{5}(\frac{8}{70})$	$\frac{1}{5}(\frac{8}{70})$	$\frac{1}{5}(\frac{8}{70})$
Total Mass	$\frac{2}{10}$		0	$\frac{8}{70}$	0	$\frac{24}{175}$	$\frac{24}{175}$	$\frac{48}{175}$		Assume this is somewhere > 18

The next time a participant fails, we add their mass (including any reallocated mass) to that time point

This continues, here splitting this censored person's mass of 8/70 across the 5 remaining participants

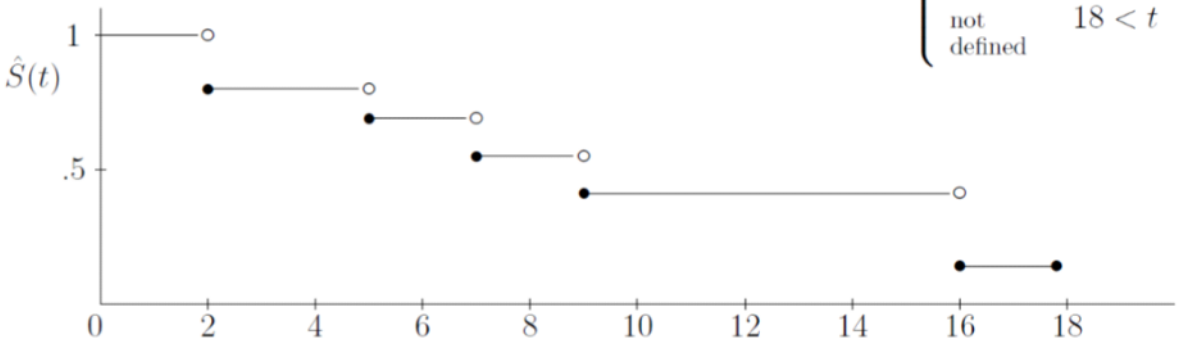
	2	2	3^+	5	5^+	7	9	16	16	18^+
Step 1										
Step 2	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$
Step 3	$\downarrow$	$\downarrow$	$\hookrightarrow$	$\frac{1}{70}$	$\frac{1}{70}$	$\frac{1}{70}$	$\frac{1}{70}$	$\frac{1}{70}$	$\frac{1}{70}$	$\frac{1}{70}$
	$\downarrow$	$\downarrow$		$\downarrow$	$\hookrightarrow$	$\frac{1}{5}(\frac{8}{70})$	$\frac{1}{5}(\frac{8}{70})$	$\frac{1}{5}(\frac{8}{70})$	$\frac{1}{5}(\frac{8}{70})$	$\frac{1}{5}(\frac{8}{70})$
Total Mass	$\frac{2}{10}$		0	$\frac{8}{70}$	0	$\frac{24}{175}$	$\frac{24}{175}$	$\frac{48}{175}$		Assume this is somewhere > 18

28

We sum the masses over time to estimate the CDF, and subtract this from 1 to get the estimated survival function

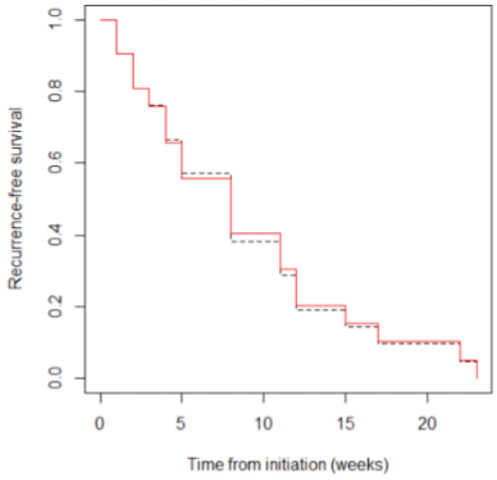
This is another way of calculating the Kaplan-Meier estimator

$$\hat{S}(t) = \begin{cases} 1 & 0 \le t < 2 \\ .8 & 2 \le t < 5 \\ .69 & 5 \le t < 7 \\ .55 & 7 \le t < 9 \\ .41 & 9 \le t < 16 \\ .14 & 16 \le t \le 18 \\ \text{not defined} & 18 < t \end{cases}$$



29

A second look at our earlier plots…



- Black dotted line is with full data and empirical CDF
- Red solid line is adjusting for censoring

Censoring assumptions

- Kaplan-Meier estimation relies on the assumption of **non-informative** or **independent** censoring:
  - We can use the observed failures to model what we would have seen in the censored individuals if they had not been censored
  - Thus, censored people must be similar to who are not censored
  - Rigorously, it means that T and C are independent random variables (or can be conditionally independent given covariates)
- This is an **untestable assumption** given the observed data, but there are some exceptional settings…

Non-informative censoring

- **Administrative censoring**
  - The censoring time is the end of the study
  - This is set by the researchers and is the same for all individuals
  - Since C is the same for everyone, there is no association between T and C

- **Random censoring**
  - Censoring times vary across individuals, but they are random and thus unassociated with T

e.g. someone dies of a car crash in a study about cancer survival

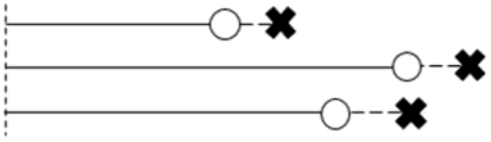
32

Informative censoring

- What about when censoring is **informative** or **dependent**?

- Imagine a study in which the people who are lost from the study are the people who are too sick to make it to their appointments
  - They are the people most likely to fail soon afterwards
  - This is a classic example of informative censoring

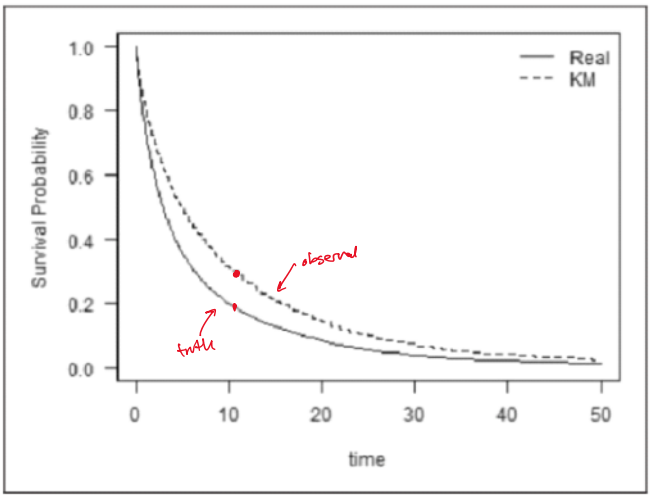
or, ppl who feel better don't feel the need to show up



T and C are dependent

33

Dependent censoring undermines validity



KM shows higher survival rate bc ppl that are missing/censored are too sick to go

We see KM is w/ "healthier" people

What does it mean in practice?

- A study with a lot of administrative censoring can still be a high quality study (*indicates a rare event!*)
- A study with a lot of loss to follow-up is likely a lower quality study
- The concern is that these participants are different from those who are retained
- Investigators should examine the potential reasons why these participants do not complete follow-up, and assess whether there is potential for bias

Today's activity

- Calculate a Kaplan-Meier curve

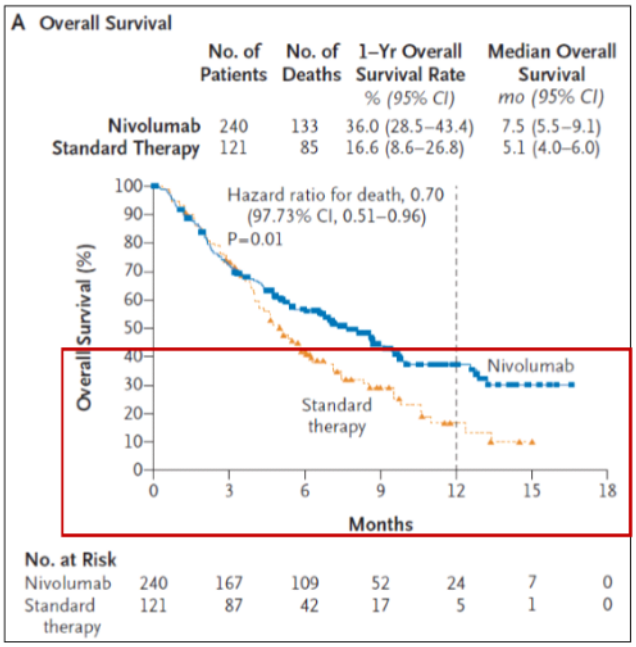
- Study examples of published Kaplan-Meier figures
  - Notice the details
  - *I will show three examples, and then it is your turn!*

Example 1: Carcinoma of the head and neck

Abbreviated summary

- A trial was conducted to evaluate a novel treatment for carcinoma of the head and neck.
- Patients with recurrent squamous-cell carcinoma of the head and neck were randomized to receive either nivolumab or standard therapy.
- The primary endpoint of the trial was overall survival (death by any cause) from the time of randomization.

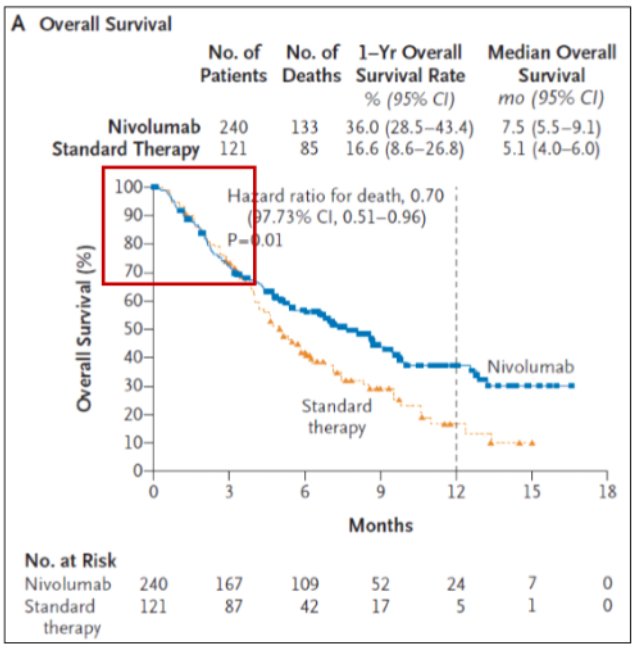
Source: Ferris et al. (2016) *NEJM*, DOI: 10.1056/NEJMoa1602252



Kaplan–Meier curves for overall survival among all the patients who underwent randomization and were assigned to receive either nivolumab or standard therapy.

What do I notice?

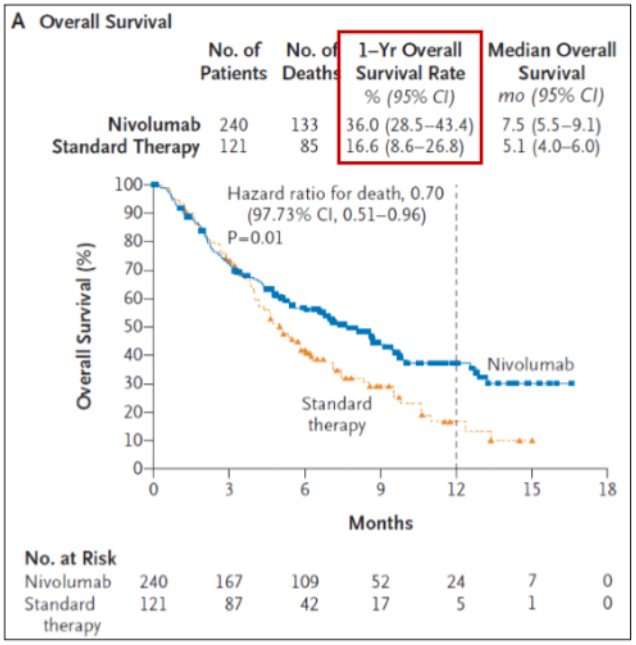
1. Poor overall survival in both groups (only 30% survival with novel treatment)



Kaplan–Meier curves for overall survival among all the patients who underwent randomization and were assigned to receive either nivolumab or standard therapy.

What do I notice?

2. Similar survival across treatment groups during the first 3 months after randomization, separating after 3 months



Kaplan–Meier curves for overall survival among all the patients who underwent randomization and were assigned to receive either nivolumab or standard therapy.

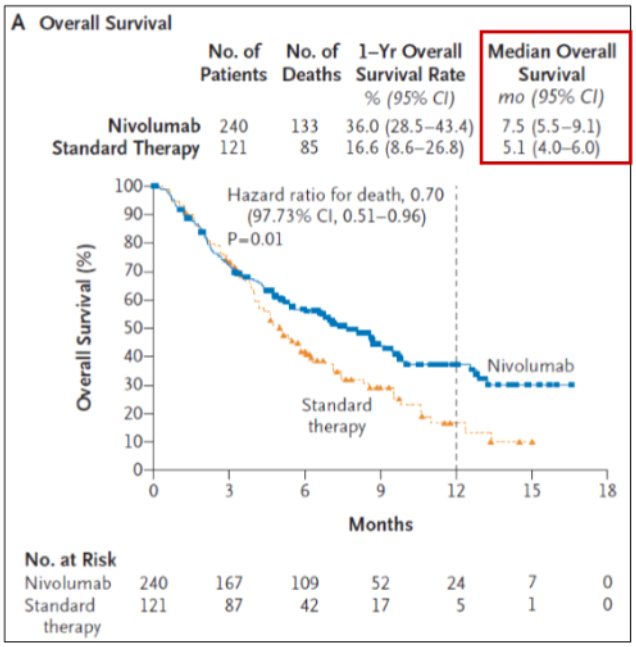
What do I notice?

3. Estimated 1-year overall survival rates (probabilities) reported.

Estimated 1-year survival probability of 36.0% (28.5 to 43.4%) for the nivolumab group.

Estimated 1-year survival probability of 16.6% (8.6 to 26.8%) for the standard therapy group.

40



Kaplan–Meier curves for overall survival among all the patients who underwent randomization and were assigned to receive either nivolumab or standard therapy.

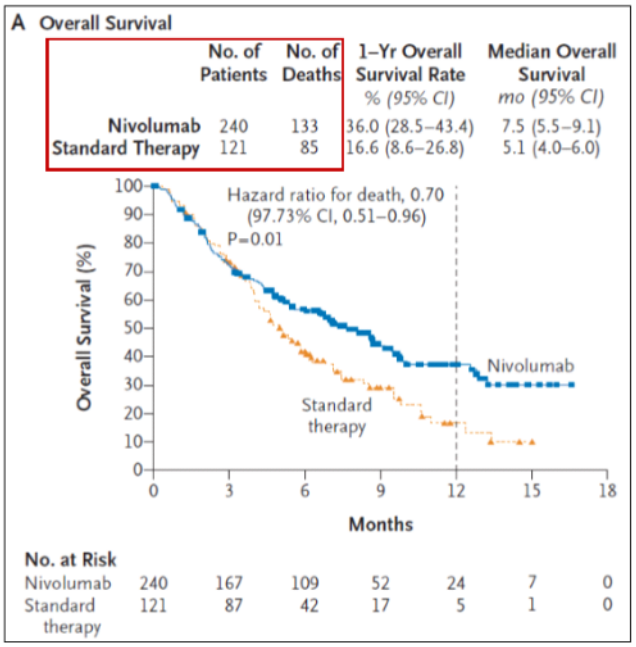
What do I notice?

4. Median overall survival reported.

Median survival is 7.5 months (5.5 to 9.1) for the nivolumab group.

Median survival is 5.1 months (4.0 to 6.0) for the standard therapy group.

41



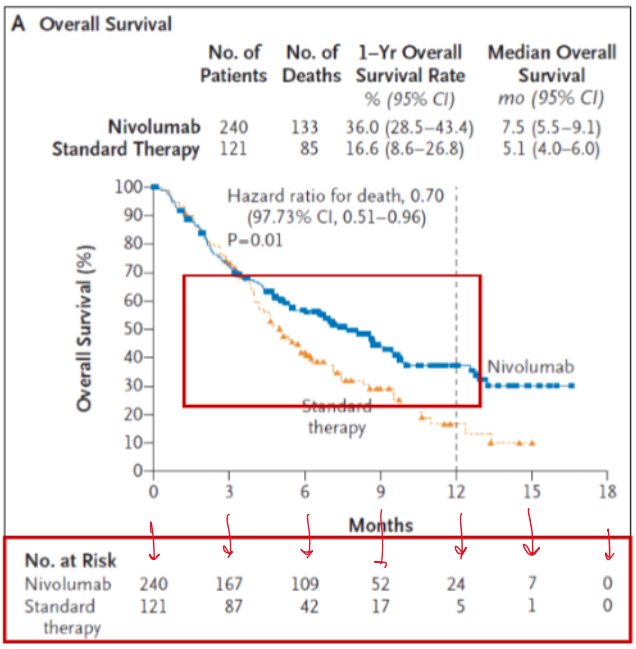
Kaplan–Meier curves for overall survival among all the patients who underwent randomization and were assigned to receive either nivolumab or standard therapy.

What do I notice?

5. Sample sizes and numbers of events reported.

240 nivolumab patients with 133 deaths observed.

121 standard therapy patients with 85 deaths observed.



Kaplan–Meier curves for overall survival among all the patients who underwent randomization and were assigned to receive either nivolumab or standard therapy.

What do I notice?

6. From figure caption, symbols indicate censored observations. → here, □ or △
Heavy censoring occurred because the database was locked for an interim analysis before follow-up was completed on all participants.
(they ended the study based on interim analysis → wanted to make the treatment available to everyone)

The table at the bottom allows us to monitor how many participants remained at risk at later time points.

(mosquito borne)

Example 2: Dengue vaccine trial

Abbreviated summary

- A trial was conducted to evaluate the efficacy of a dengue vaccine.
- Children aged 2-14 in five countries across the Asia-Pacific region were randomized in a 2:1 ratio to receive three doses of vaccine or placebo.
- The primary outcome was date of symptom onset of virologically confirmed clinical dengue.
- The time origin for the intention-to-treat analysis (including all randomized participants regardless of protocol violations) was calculated from the time of randomization.

Source: Capeding et al. (2014) *The Lancet*, DOI: 10.1016/S0140-6736(14)61060-6

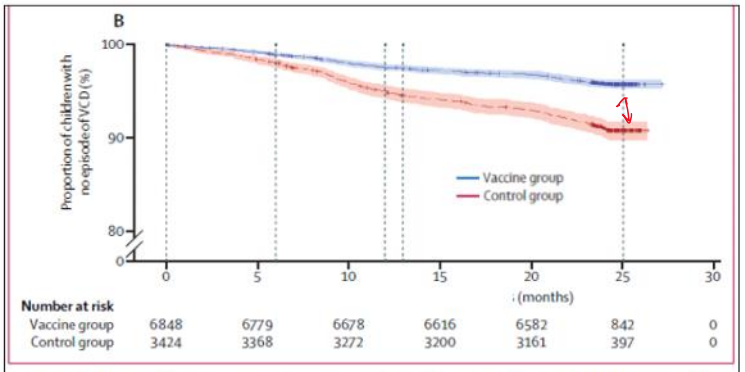


Figure 2: **Kaplan-Meier curve for symptomatic virologically-confirmed dengue (VCD) due to any serotypes taking place from 28 days after the third injection (ie, from month 13) in the per-protocol population (A) and at any time during the trial from day 0, irrespective of protocol compliance, in the intention-to-treat population (B)**

Dashed vertical lines show major study milestones: injections (at months 0, 6, 12); the start of the period for the primary, per-protocol analysis (month 13); and the end of the active phase of surveillance (month 25). Error bars show 95% CIs. Note breaks in y axes.

What do I notice?

1. Widening difference in dengue-free survival between the vaccine and control groups, with more control group participants developing dengue during the study.

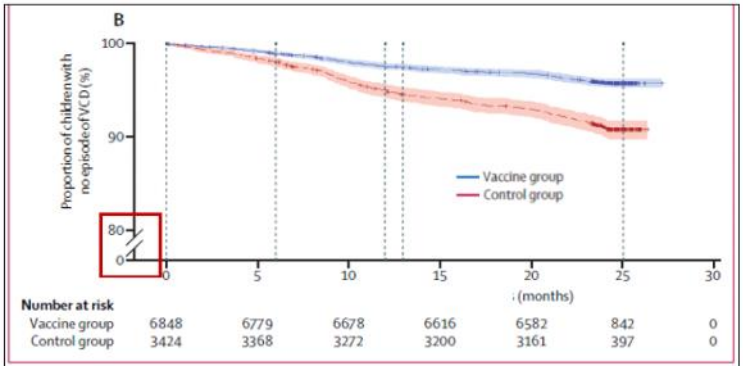


***Figure 2*: Kaplan-Meier curve for symptomatic virologically-confirmed dengue (VCD) due to any serotypes taking place from 28 days after the third injection (ie, from month 13) in the per-protocol population (A) and at any time during the trial from day 0, irrespective of protocol compliance, in the intention-to-treat population (B)**

Dashed vertical lines show major study milestones: injections (at months 0, 6, 12); the start of the period for the primary, per-protocol analysis (month 13); and the end of the active phase of surveillance (month 25). Error bars show 95% CIs. Note breaks in y axes.

What do I notice?

2. Break in y-axis, with only survival of 80-100% plotted.

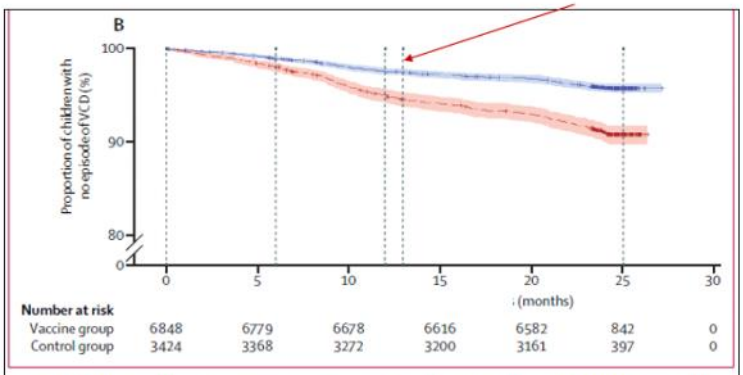


***Figure 2*: Kaplan-Meier curve for symptomatic virologically-confirmed dengue (VCD) due to any serotypes taking place from 28 days after the third injection (ie, from month 13) in the per-protocol population (A) and at any time during the trial from day 0, irrespective of protocol compliance, in the intention-to-treat population (B)**

Dashed vertical lines show major study milestones: injections (at months 0, 6, 12); the start of the period for the primary, per-protocol analysis (month 13); and the end of the active phase of surveillance (month 25). Error bars show 95% CIs. Note breaks in y axes.

What do I notice?

3. Dashed vertical lines show major study milestones: injections (at months 0, 6, 12); the start of follow-up for the primary analysis (month 13); and the end of the active phase of surveillance (month 25)

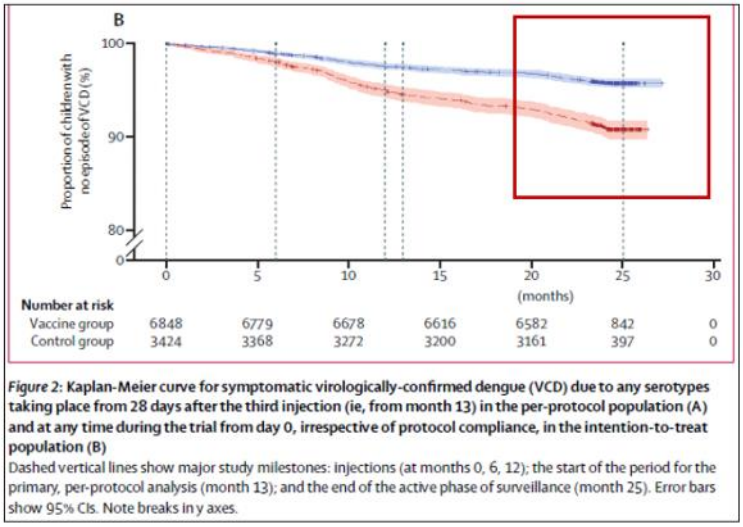


Figure 2: **Kaplan-Meier curve for symptomatic virologically-confirmed dengue (VCD) due to any serotypes taking place from 28 days after the third injection (ie, from month 13) in the per-protocol population (A) and at any time during the trial from day 0, irrespective of protocol compliance, in the intention-to-treat population (B)**

Dashed vertical lines show major study milestones: injections (at months 0, 6, 12); the start of the period for the primary, per-protocol analysis (month 13); and the end of the active phase of surveillance (month 25). Error bars show 95% CIs. Note breaks in y axes.

What do I notice?

4. Heavy censoring at the end of the follow-up period. Pre-specified time is 25 months, but exact timing varies across participants.

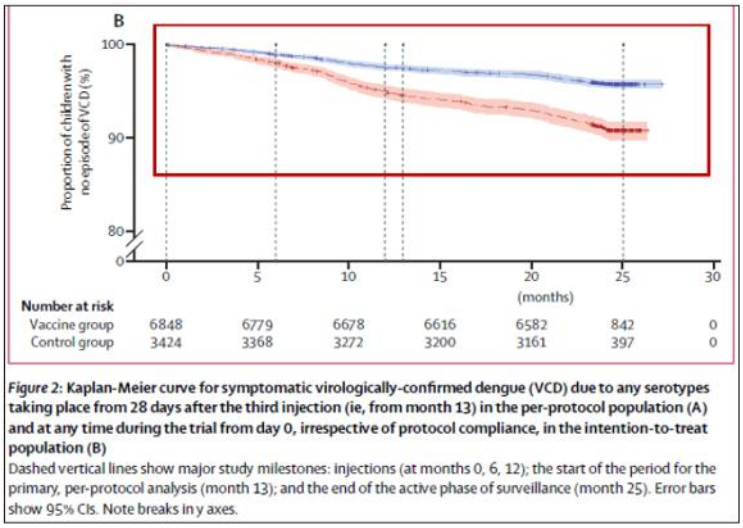


Figure 2: **Kaplan-Meier curve for symptomatic virologically-confirmed dengue (VCD) due to any serotypes taking place from 28 days after the third injection (ie, from month 13) in the per-protocol population (A) and at any time during the trial from day 0, irrespective of protocol compliance, in the intention-to-treat population (B)**

Dashed vertical lines show major study milestones: injections (at months 0, 6, 12); the start of the period for the primary, per-protocol analysis (month 13); and the end of the active phase of surveillance (month 25). Error bars show 95% CIs. Note breaks in y axes.

What do I notice?

5. 95% confidence intervals are shown with shaded bands.

Why is the CI narrower?
Larger n

Example 3: Acute coronary syndrome trial

Abbreviated summary

- A trial was conducted to compare the efficacy of ticagrelor or prasugrel.
- Patients presenting with acute coronary syndrome were randomized to receive either ticagrelor or prasugrel.
- The primary outcome was a composite of death, myocardial infarction.
- Patients were followed for one year from the time of randomization.

Source: Schüpke et al. (2019) *NEJM*, DOI: 10.1056/NEJMoa1908973

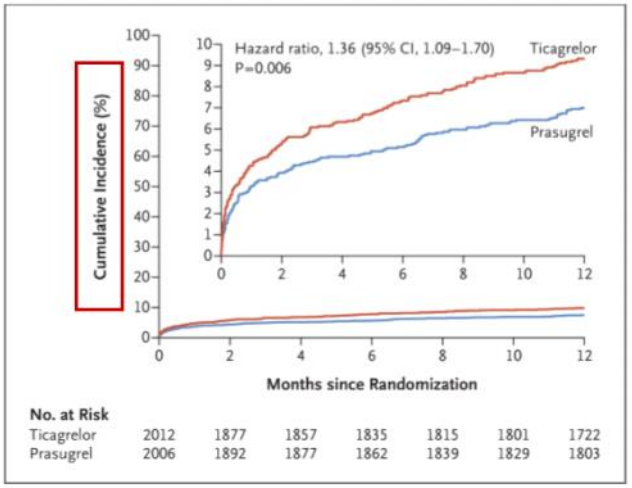


Figure 2. Cumulative Incidence of the Primary End Point at 1 Year.

The Kaplan–Meier curves show the cumulative incidence of the primary end point, which was a composite of death, myocardial infarction, or stroke at 1 year. The inset shows the same data on an enlarged y axis.

What do I notice?

1. The Kaplan-Meier method is used to generate cumulative incidence curves (CDF). These start from zero and increase over time.

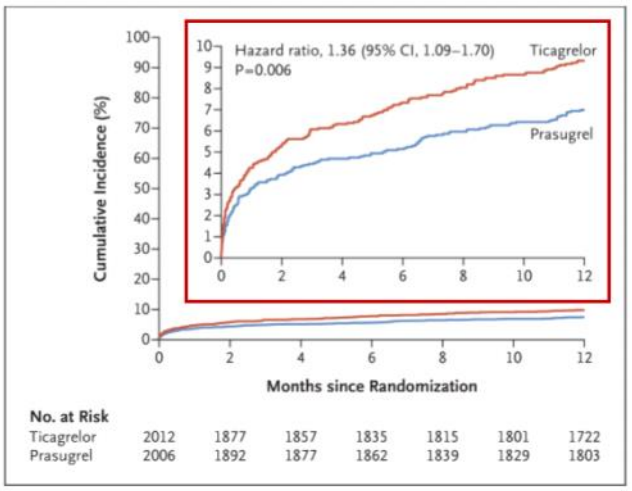


Figure 2. Cumulative Incidence of the Primary End Point at 1 Year.

The Kaplan–Meier curves show the cumulative incidence of the primary end point, which was a composite of death, myocardial infarction, or stroke at 1 year. The inset shows the same data on an enlarged y axis.

What do I notice?

2. Because the cumulative incidence at one year is <10%, an inset is shown with an enlarged y-axis.

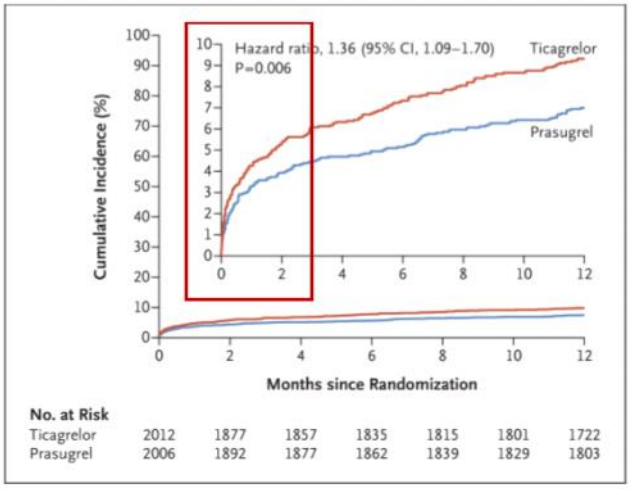


Figure 2. Cumulative Incidence of the Primary End Point at 1 Year.

The Kaplan–Meier curves show the cumulative incidence of the primary end point, which was a composite of death, myocardial infarction, or stroke at 1 year. The inset shows the same data on an enlarged y axis.

What do I notice?

3. There is an early period of elevated risk, immediately after presenting with acute coronary syndrome.