Censoring, Kaplan-Meier, Cox regression

Håkon K. Gjessing

Professor/Principal Investigator

Centre for Fertility and Health, Norwegian Institute of Public Health, Oslo

Department of Global Public Health and Primary Care, University of Bergen

Makerere

Tuesday 13 June, 2023

CENTRE FOR FERTILITY AND HEALTH

Event History Analysis: What and Why?

New England Journal of Medicine

Editorial, Jan. 6, 2000, p. 42-49

The eleven most important developments in medicine over the past millennium

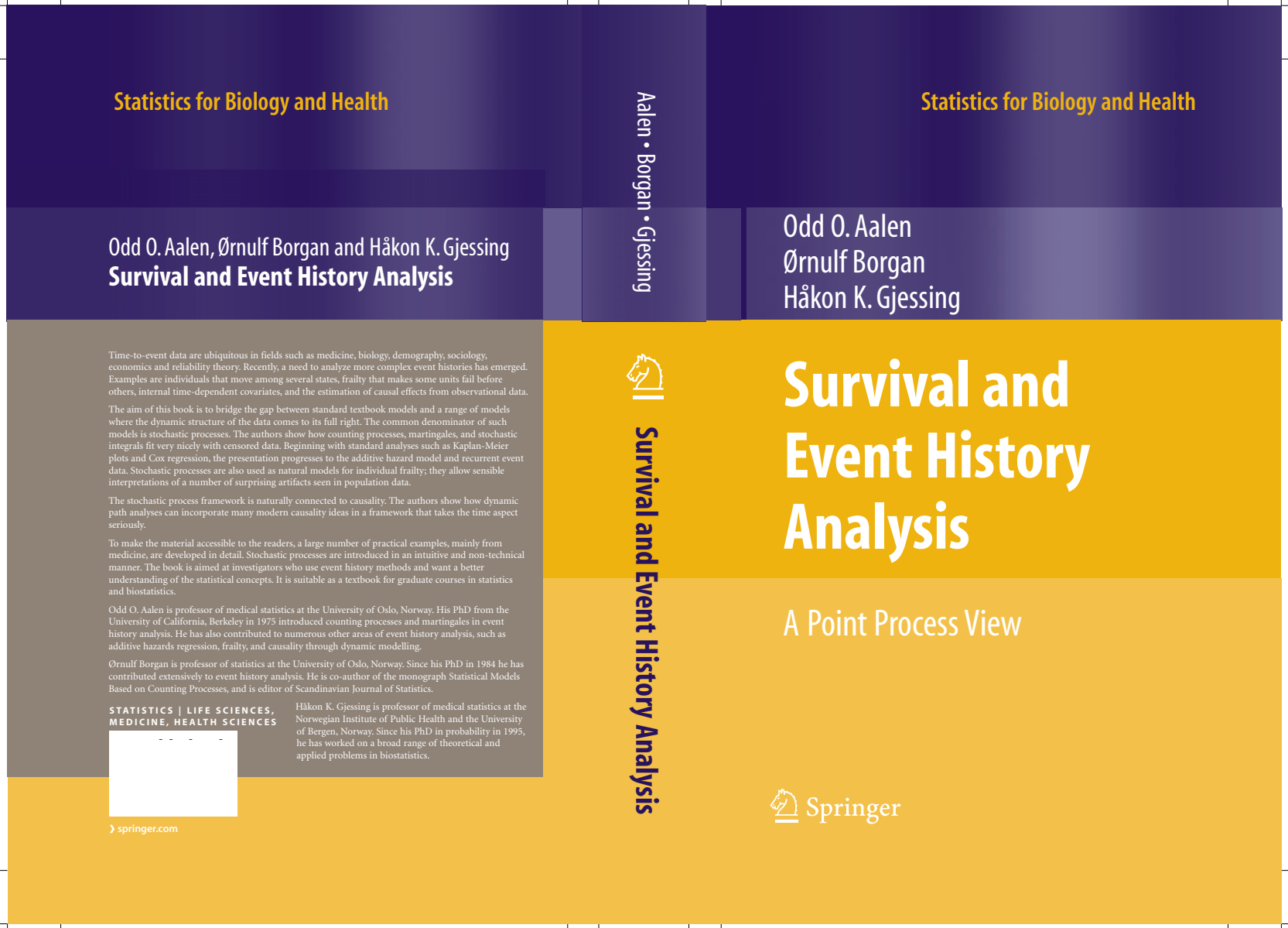
- Elucidation of human anatomy and physiology
- Discovery of cells and their substructures
- Elucidation of the chemistry of life
- *Application of statistics to medicine*
- Development of anesthesia
- Discovery of the relation of microbes to disease
- Elucidation of inheritance and genetics
- Knowledge of the immune system
- Development of body imaging
- Discovery of antimicrobial agents
- Development of molecular pharmacotherapy

Event History Analysis: What and Why?

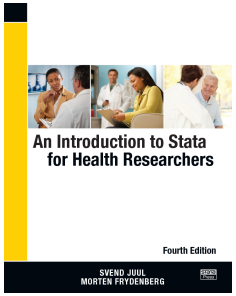
Application of statistics to medicine

> *"Sir David Cox's 1972 paper on proportional-hazards regression ignited the fields of survival analysis and semiparametric inference (using partial specification of the probability distribution of the outcomes under investigation). Rapid improvements in computer support were essential to the growing role of empirical investigation and statistical inference."*

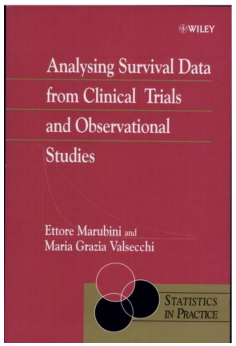
Totally Indecent Self-Promotion



A SAMPLE OF BOOKS

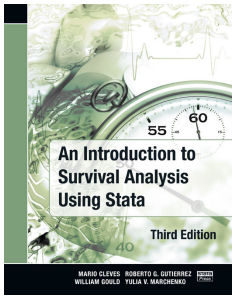


An Introduction to Stata for Health Researchers, Fourth Edition
Svend Juul and Morten Frydenberg
Stata Press, 2014



Analysing Survival Data from Clinical Trials and Observational Studies
Ettore Marubini, Maria Grazia Valsecchi
Wiley, 2004

Survival Analysis and Epidemiological Tables Reference Manual
Stata Press, 2013



An Introduction to Survival Analysis Using Stata, Third Edition
Mario Cleves, William Gould, Roberto G. Gutierrez, and Yulia V. Marchenko
Stata Press, 2010

Event History Analysis: What and Why?

- Outcome: Time to "event"
- Additional problem: Censoring (and truncation)
- For example:

1. Time from cancer diagnosis to death
   Censoring: Cancer patients get transferred to another hospital (loss-to-followup)
2. Time from started malaria treatment to cured
   Censoring: Patients leave treatment/followup when most severe symptoms end
3. Time from inserting a dental filling to when it fails
   Censoring: Study ends after 5 years
4. Time from first birth to the second (for the same mother)
   Censoring(?): Mother too high age, or decides not to have more children
5. Time from hip prosthesis operation to failure/re-operation
   Censoring: The prosthesis lasts for the rest of the patient's life
6. Time from entering marriage to divorce
   Censoring: The couple moves abroad (loss-to-followup), or never get divorced!

NOTE: Is censoring *independent*?
(not necessarily the case in all the examples above)

Independent Censoring

Intuitively:

- The censored individual is not “special” in any way, i.e.:
  - at any time before censoring, the risk of an event is the same as for everybody else.
  - after the time of censoring, the individual would have had the same risk profile as everybody else if it could have been observed.

Examples:

- Patients leave treatment/followup when most severe symptoms end. **Probably not independent**
- End of study. **Usually independent**

Event history analysis: Choice of time scale

Three most common time scales:

1. **Time from inclusion to event (study time)**

   Example: Time from cancer diagnosis to death

   Zero: Date of inclusion (individual)

2. **Calendar time**

   Example: Time from a fixed date (e.g. 01 Jan 2020) to infection with the SARS-CoV-2 virus

   Zero: Start date (common)

3. **Age**

   Example: Age at death

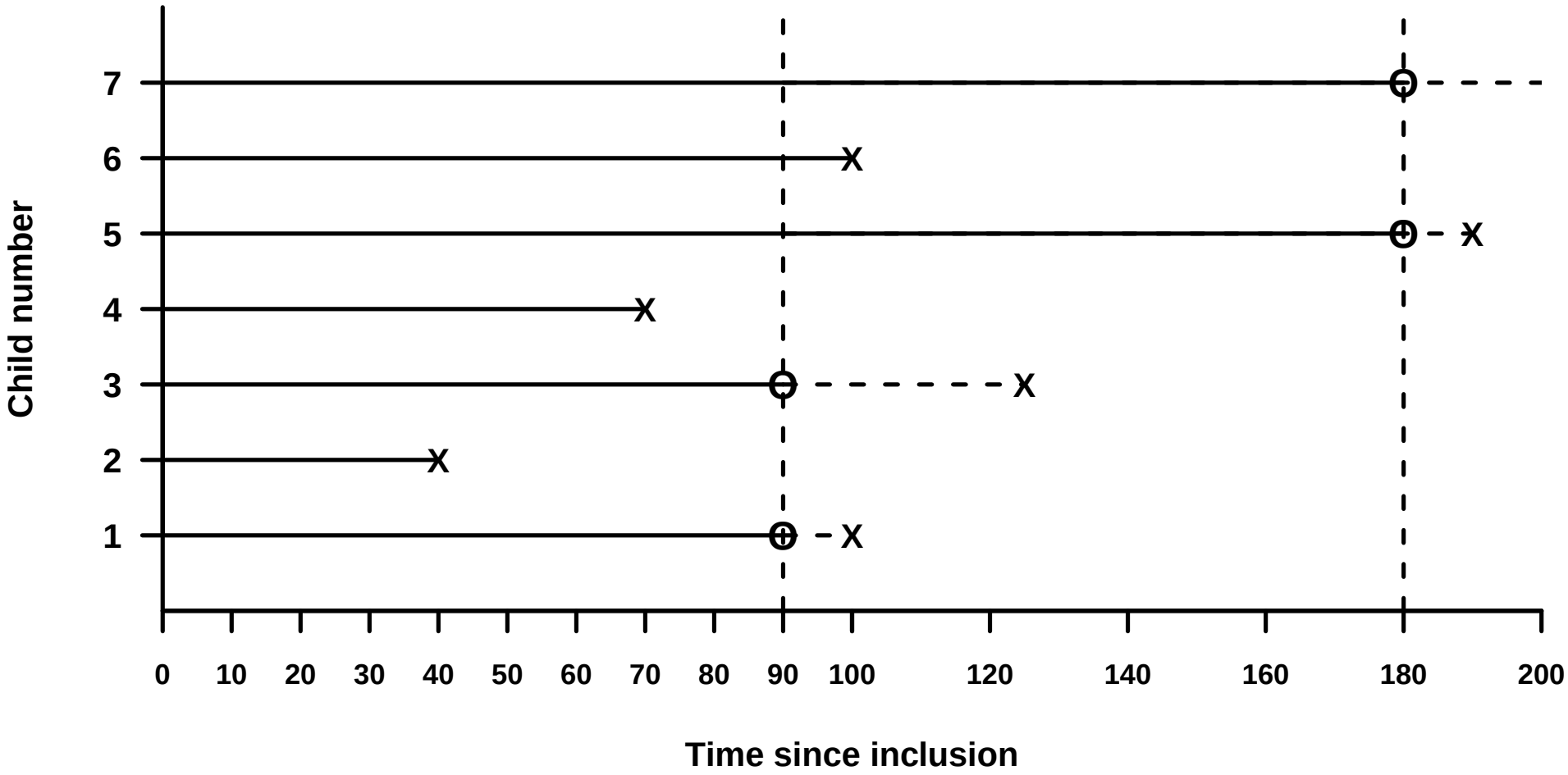
   Zero: Date of birth (individual)

Nepal Study: Time from first pneumonia admission to next

- **Outcome:**
  Time from first admission with pneumonia until *next admission*
- Age range: 2 months to 3 years
- **Dates:** November 2003 to December 2007
- **Main exposure:** Zinc versus placebo
- A total of 719 + 350 = 1069 children
- 719 children have two admissions
- 350 children have only the first admission (during follow-up period):
  Censoring! Forget these (for the time being!!)

(Data from Tor Strand, Maria Mathisen, and others,
Centre for International Health, UiB)

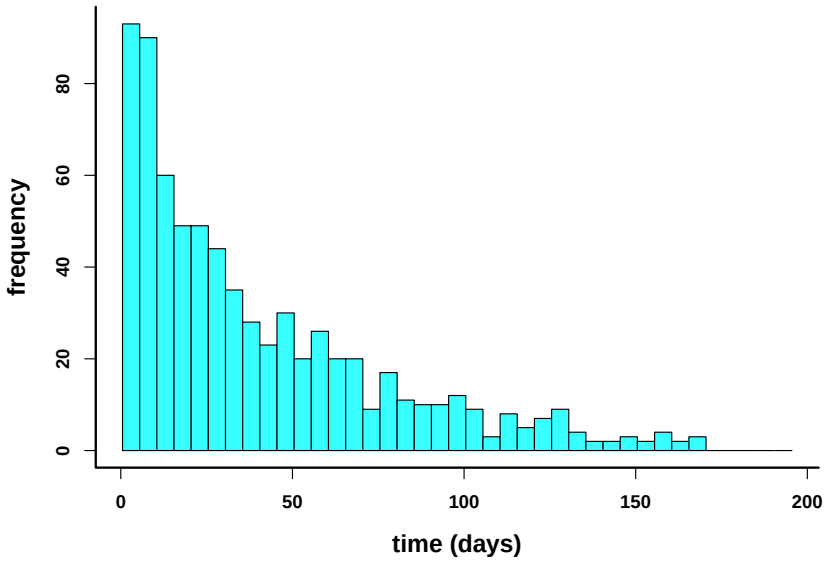
TIMELINE FOR EVENTS AND CENSORING



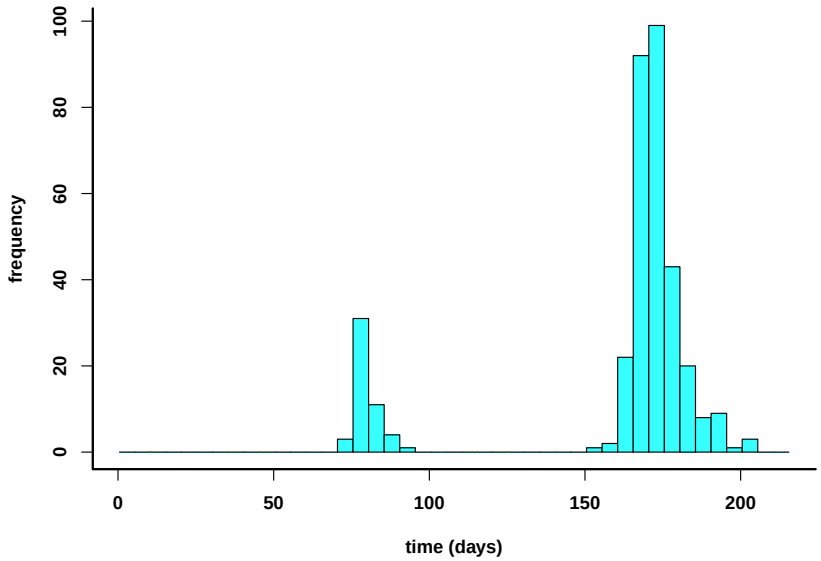
X = event O = censoring

TIME DISTRIBUTIONS

Distribution of 719 new infections



Follow-up time of the 350 censored



DATA, SELECTED VARIABLES

	id	date	age	sex	treat.orig	time	event	time.14	treat
1	1	2004-01-25	10	1	1	69	1	56	1
2	2	2004-03-22	13	2	1	123	1	110	0
3	6	2003-11-30	7	1	1	190	0	177	1
4	8	2003-12-02	5	2	0	185	0	172	1
5	9	2003-12-03	4	1	0	93	0	80	1
6	13	2003-12-24	6	2	0	183	0	170	1
:		:			:				:
:		:			:				:

`date`: date of inclusion

`age`: age (at inclusion) in months

`sex`: boys = 1, girls = 2

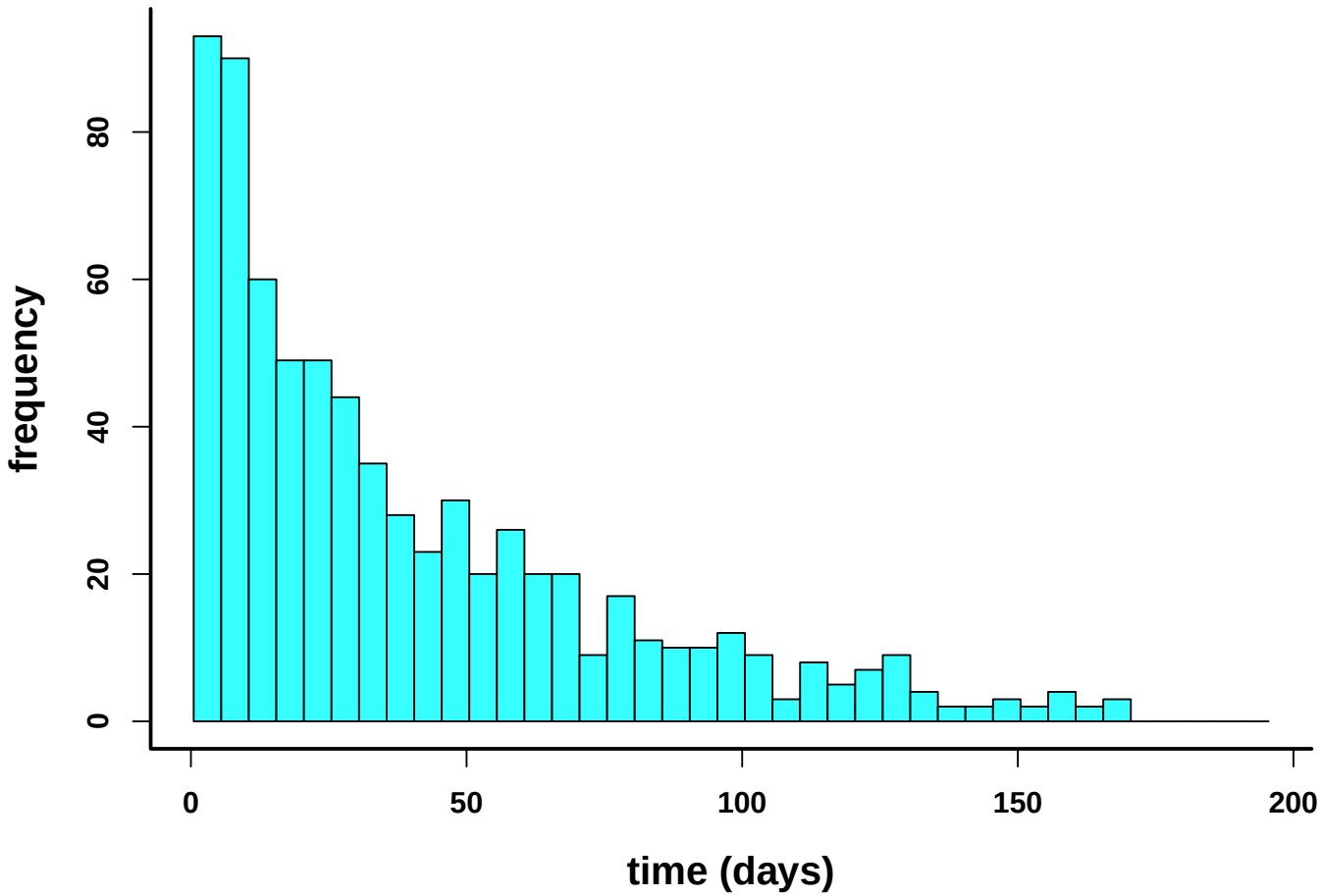
`time`: time since inclusion to event or censoring

`event`: new episode = 1, censored = 0

`time.14 = time - 13`: Starts counting after 14 days

`treat`: zinc = 1, placebo = 0

DAYS TO NEXT OCCURRENCE OF PNEUMONIA: HISTOGRAM



(Categories of 5 days, with 14 days “latency”)

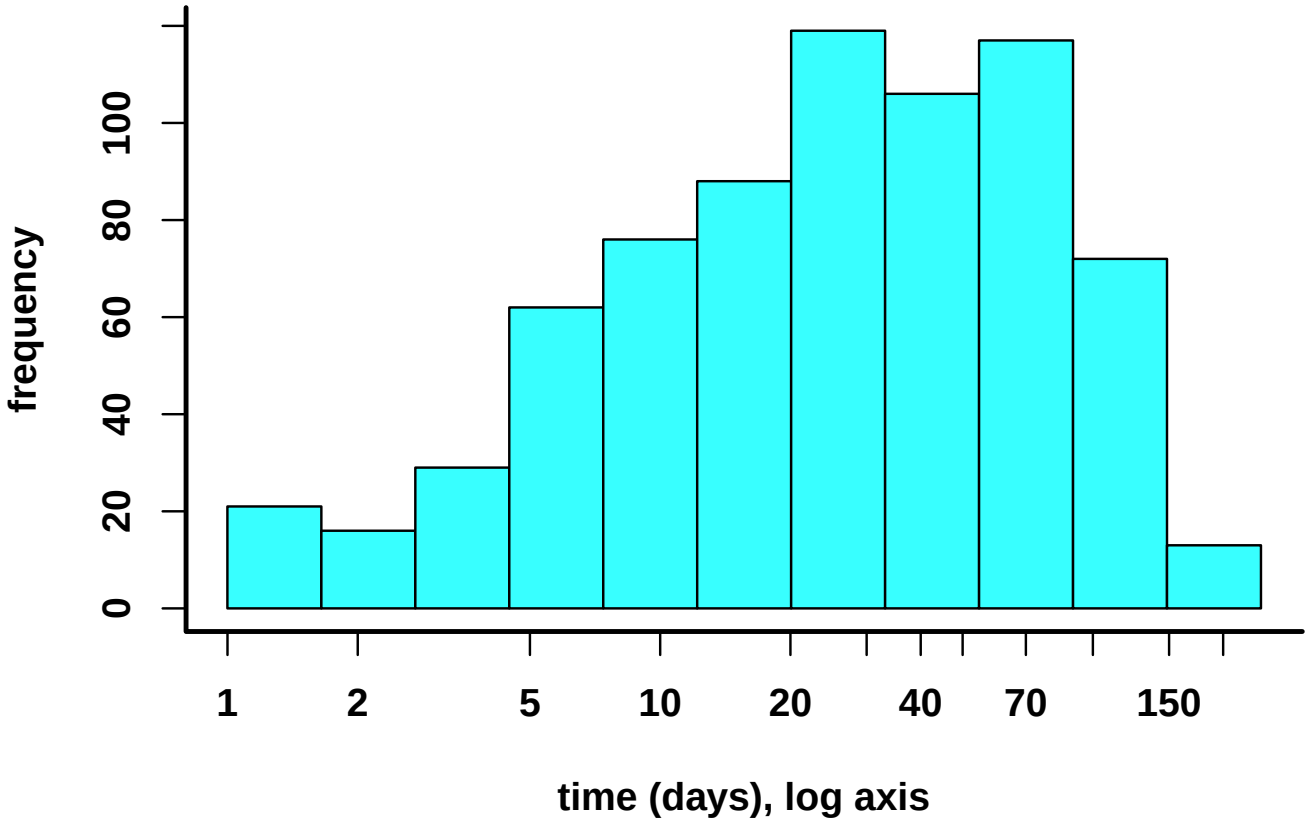
NOTE!

Typical for a time-to-event study

- Always positive values
- NOT a normal distribution, i.e. no t-test nor ordinary regression
- Often skewed distribution, with a tail to the right
- What to do with the 350 that never got a new infection? (censoring)

TRICK?? TRY A LOG TRANSFORM

Histogram with log scale:



DID IT HELP??

- Much closer to a normal distribution.

  Can use ordinary t-test to compare zinc and placebo

- Problem...: we have still not dealt with the 350 censored

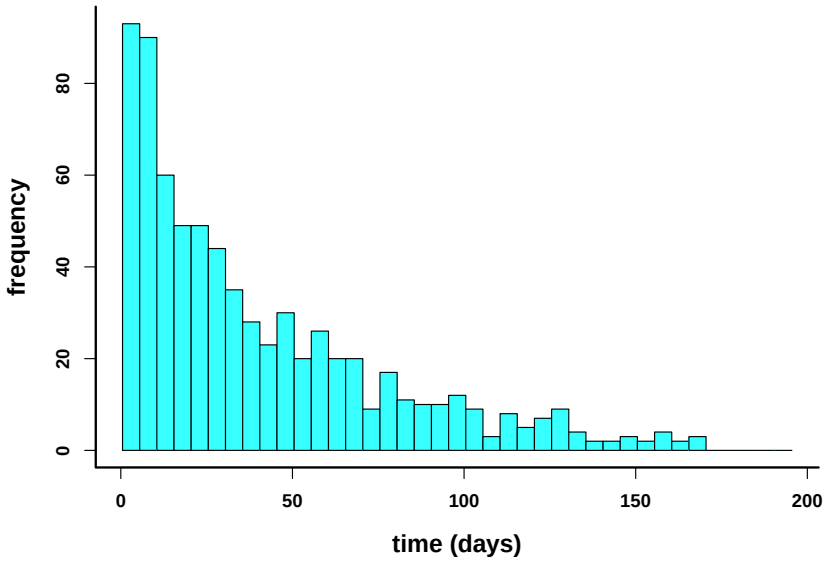
- Ordinary regression/t-test do NOT deal with censoring

- ... we need something better...

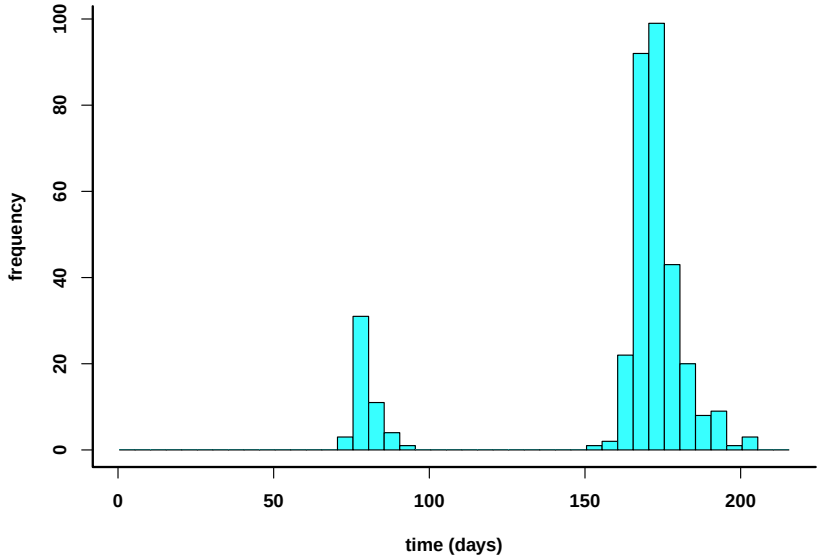
TRICK?? TRY DICHOTOMIZATION

- New infection before 50 days:
  - 1: New infection before 50 days
  - 0: Infection after 50 days, or censored
- This would actually work (using logistic regression)!
- ... but somewhat crude (looks only at a single time point)

Distribution of 719 new infections



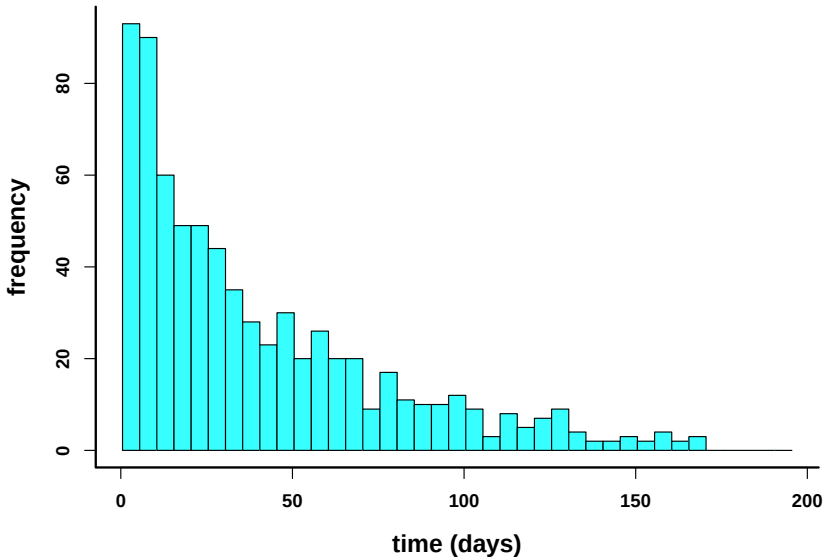
Follow-up time of the 350 censored



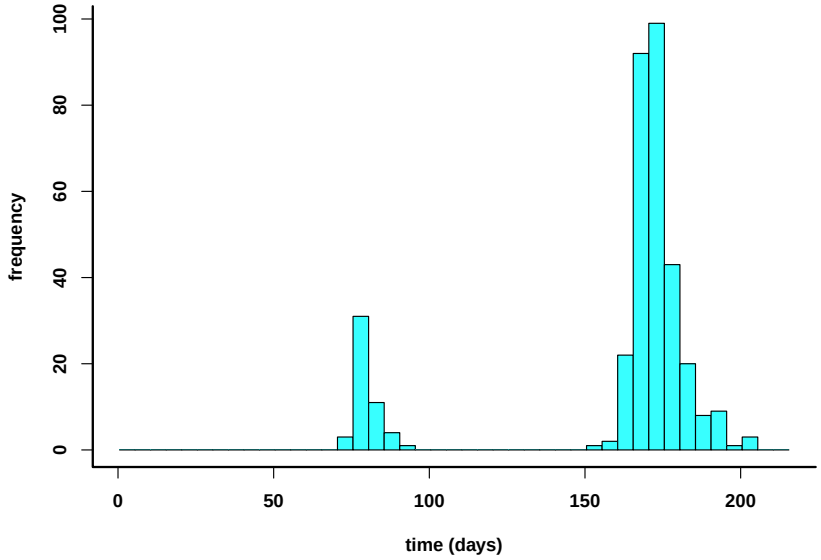
TRICK?? TRY DICHOTOMIZATION

- New infection before 100 days:
  - 1: New infection before 100 days
  - 0: Infection after 100 days, or censored??????
- This would NOT work!
- Censoring happens before cutoff value (100 days)!

Distribution of 719 new infections



Follow-up time of the 350 censored

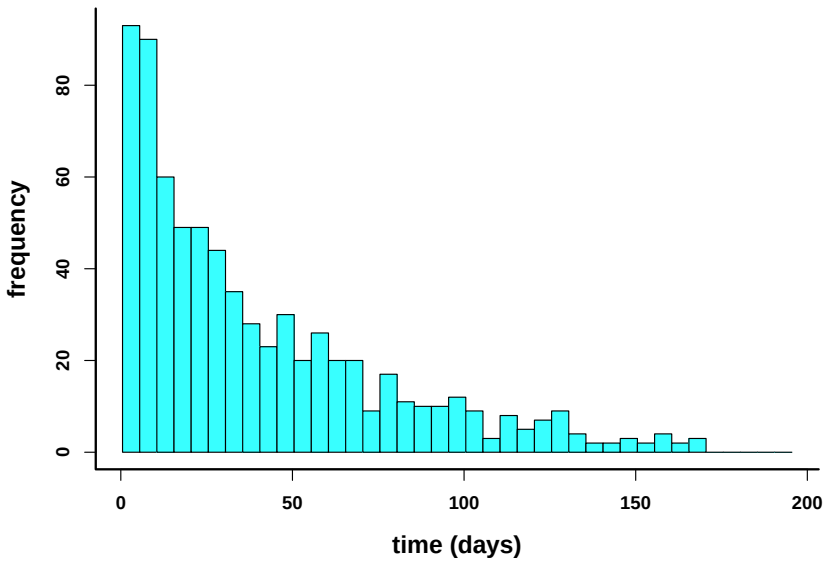


The "workhorse" of Survival Analysis

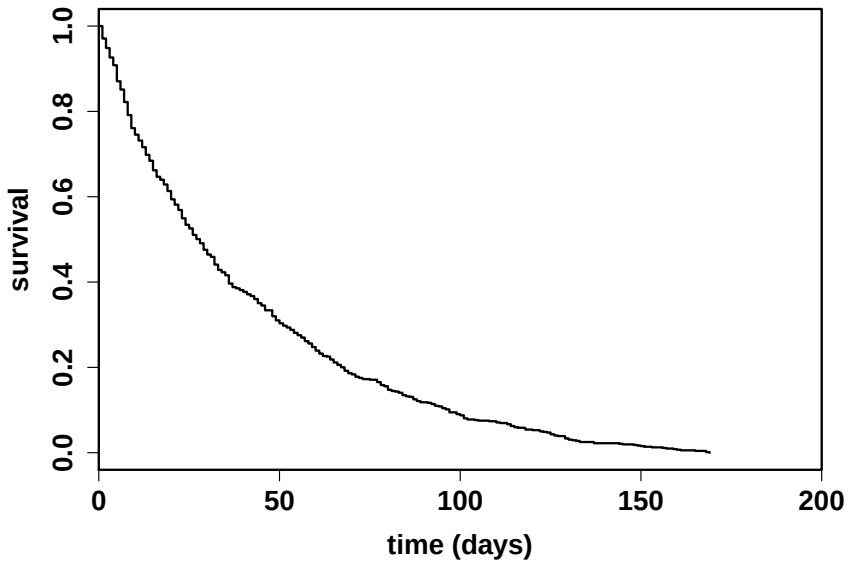
Survival curve

- T is time to event. Survival curve: $S(t) = P(T > t)$.
- I.e., the probability of "surviving" more than t days.

Distribution

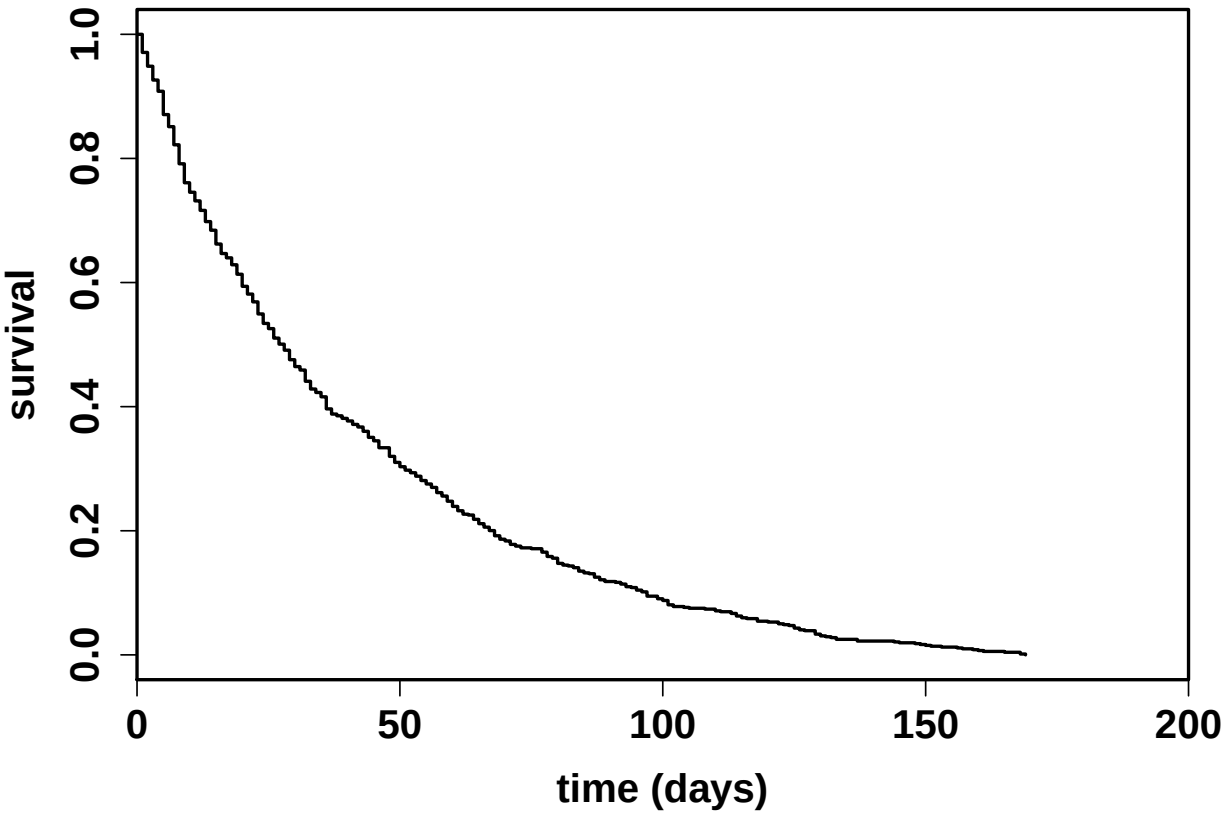


Survival curve



(Note: they look alike, but that's a coincidence)

KAPLAN-MEIER ESTIMATE OF THE SURVIVAL CURVE



Shows the proportion that still have *not* had a new infection

E.g.: After 50 days there are about 30% that still haven't had a new infection

KAPLAN-MEIER: COMPUTATION WITHOUT CENSORING

Compute S(t) *without* censoring:

$$S(t) = \frac{\text{the number of children without new infection at time } t}{\text{total number included}}$$

Survival day-to-day (from software):

time	n.risk	n.event	survival
1	719	21	0.97079
2	698	16	0.94854
3	682	16	0.92629
4	666	13	0.90821
5	653	27	0.87065

Survival first day:

$$\frac{719-21}{719} = 0.97079$$

Survival first two days:

$$\text{survival first day} \times \text{survival second day} = 0.97079 \times \frac{698-16}{698} = 0.94854$$

etc., same result as the simple rule (above).

KAPLAN-MEIER: COMPUTATION WITH CENSORING

Compute S(t) *with* censoring:

As an illustration: Assume 100 children were censored at day 3:

Survival day-to-day:

```
time n.risk n.event n.censored survival
   1    719      21          0  0.97079
   2    698      16          0  0.94854
   3    682      16        100  0.92629        new
   4    566      13          0  0.90821 -->  0.90501
   5    553      27          0  0.87065 -->  0.86083
```

Survival first 4 days:

$$\text{survival first 3 days} \times \frac{566 - 13}{566} = 0.92629 \times 0.9770318 = 0.90501$$

Survival first 5 days:

$$0.9050148 \times 0.9511754 = 0.86083$$

Effect of censoring accumulates!

KAPLAN-MEIER: NUMBERS "AT RISK"

- **The important issue is:**

  THE NUMBER OF CHILDREN
  "AT RISK"
  FOR A NEW INFECTION AT A GIVEN DAY

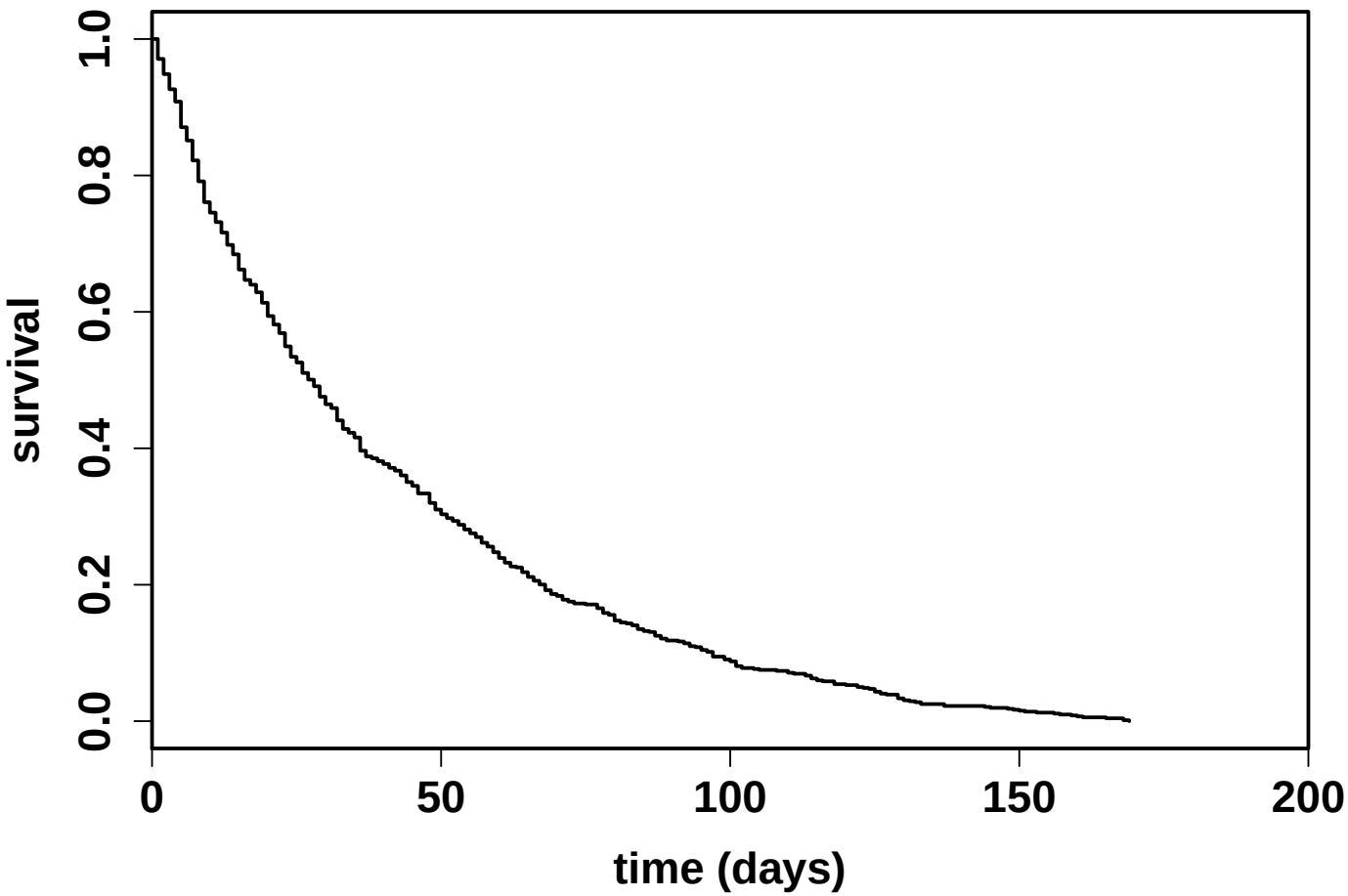
- That is, the number of children who – at a given day – are

  WITHOUT A NEW INFECTION AND ALSO NOT CENSORED

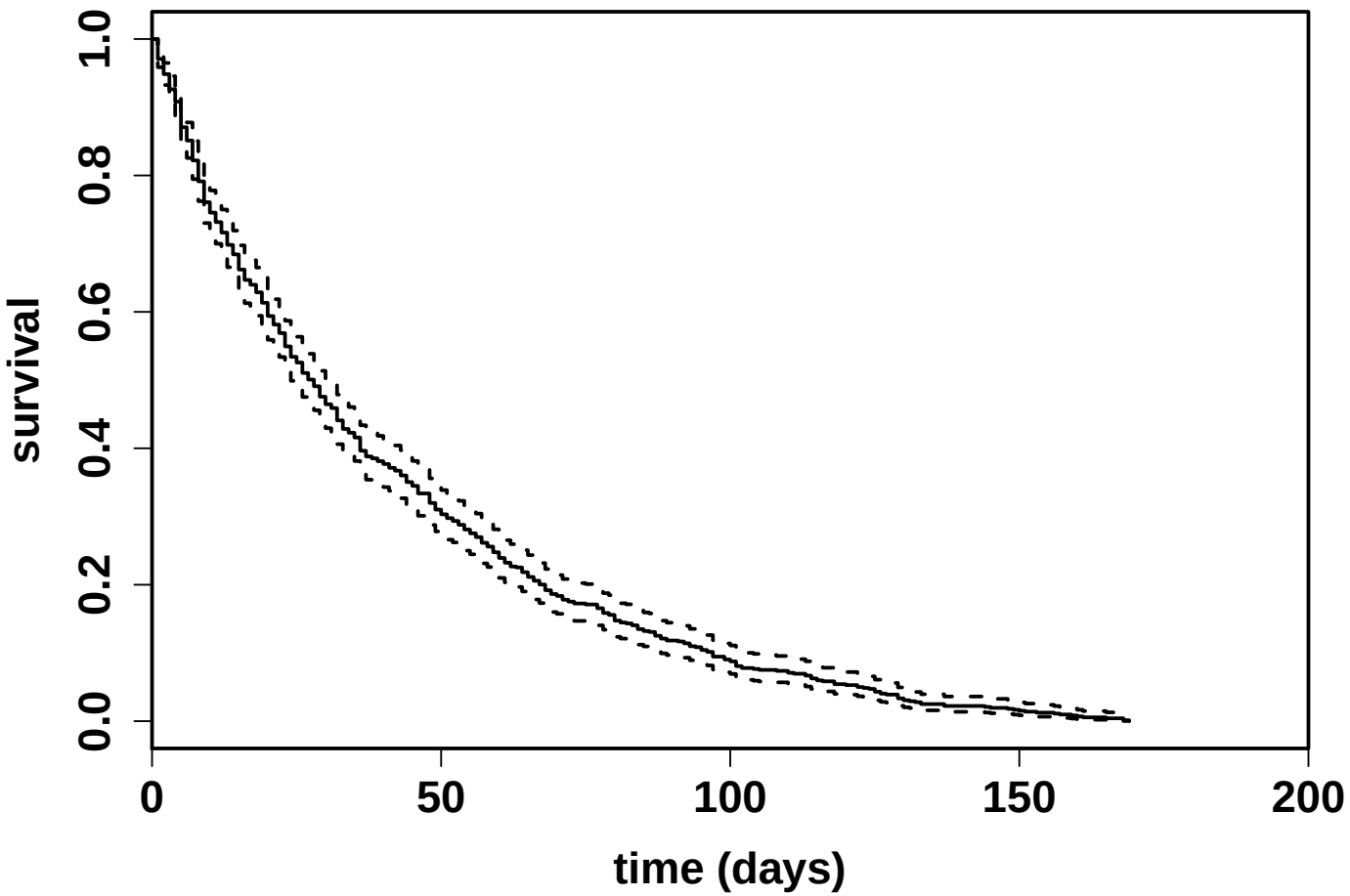
- In other words, the number of children who still *can* experience an event
- Kaplan-Meier uses children *as long as they are at risk*,

  then removes them from the computation

KAPLAN-MEIER ESTIMATE OF THE SURVIVAL CURVE



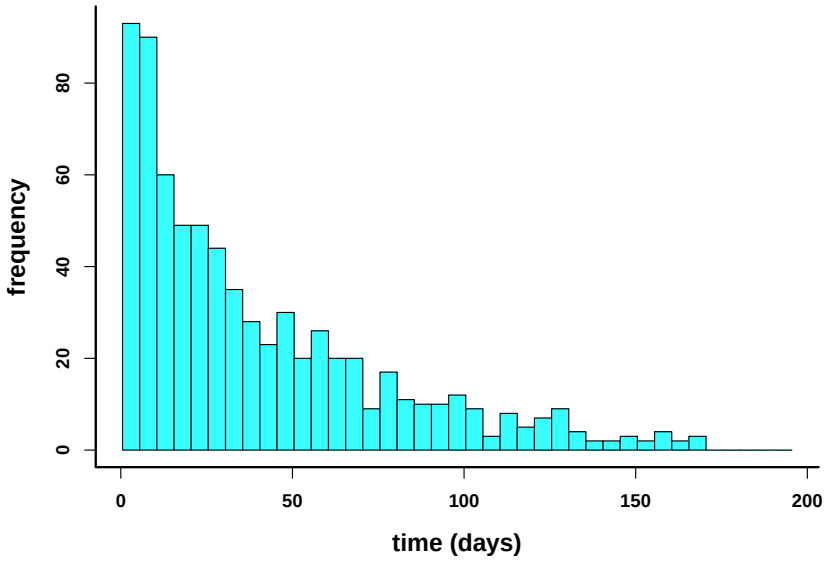
KAPLAN-MEIER ESTIMATE OF THE SURVIVAL CURVE



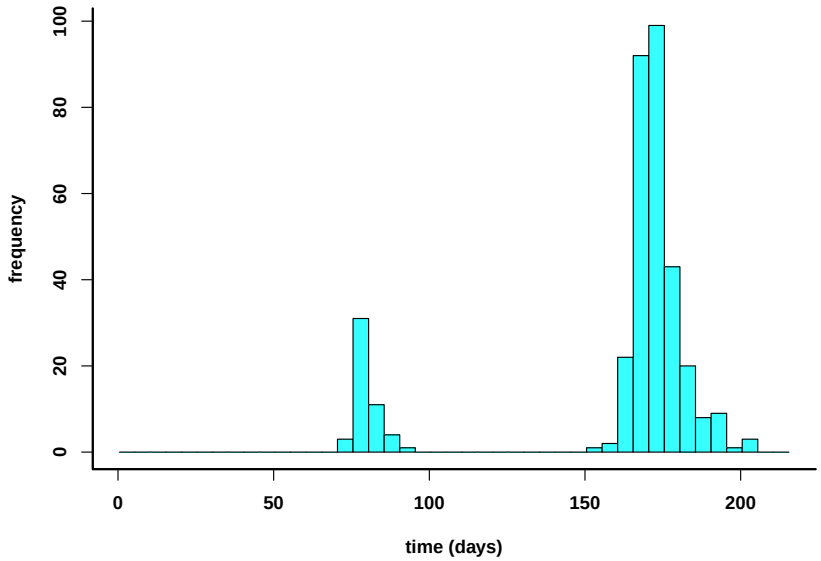
(With 95% “pointwise” confidence intervals)

NOW, HOW ABOUT THE 350 CENSORED CHILDREN?

Distribution of new infections

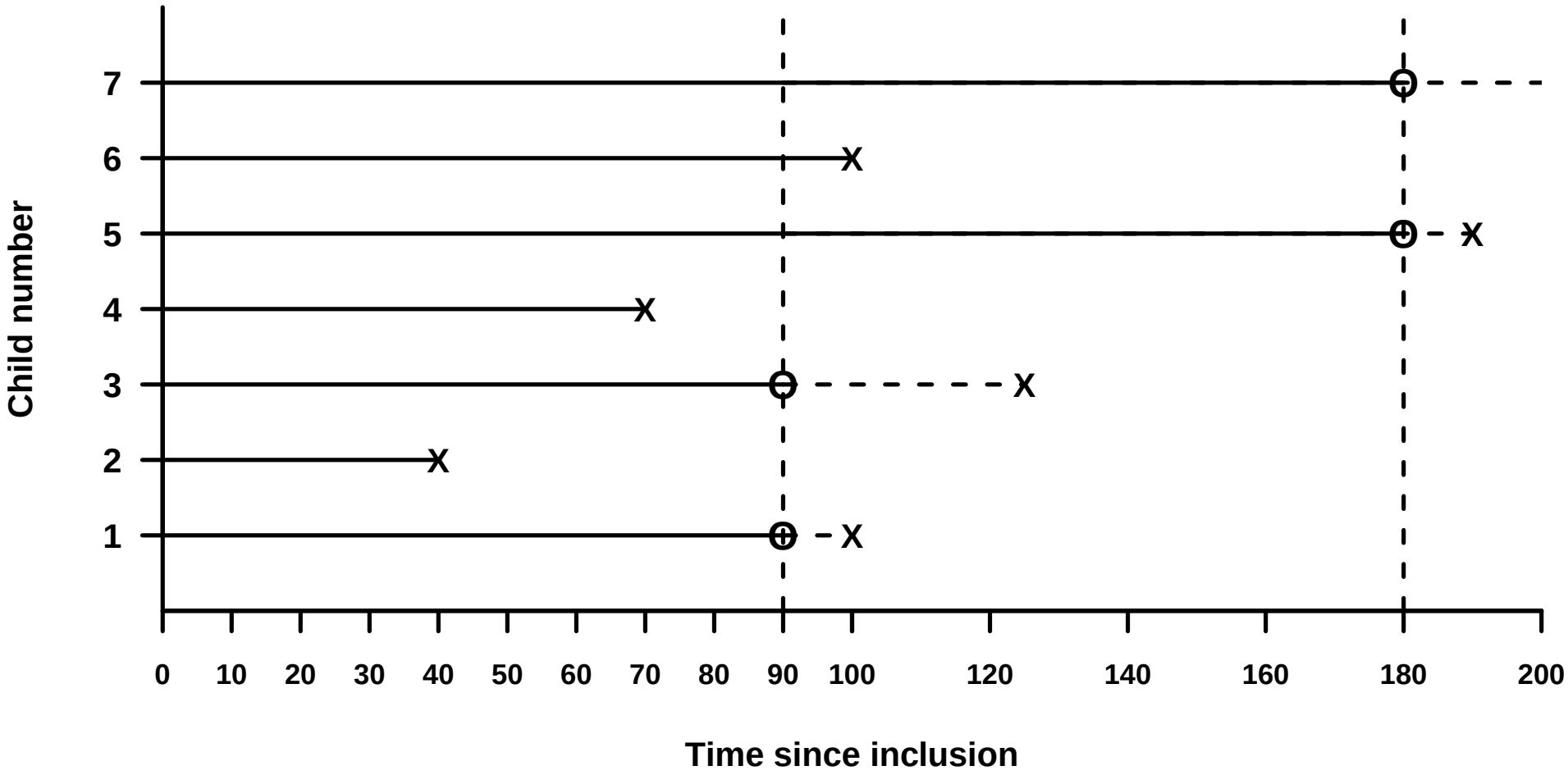


Follow-up time of the 350 censored



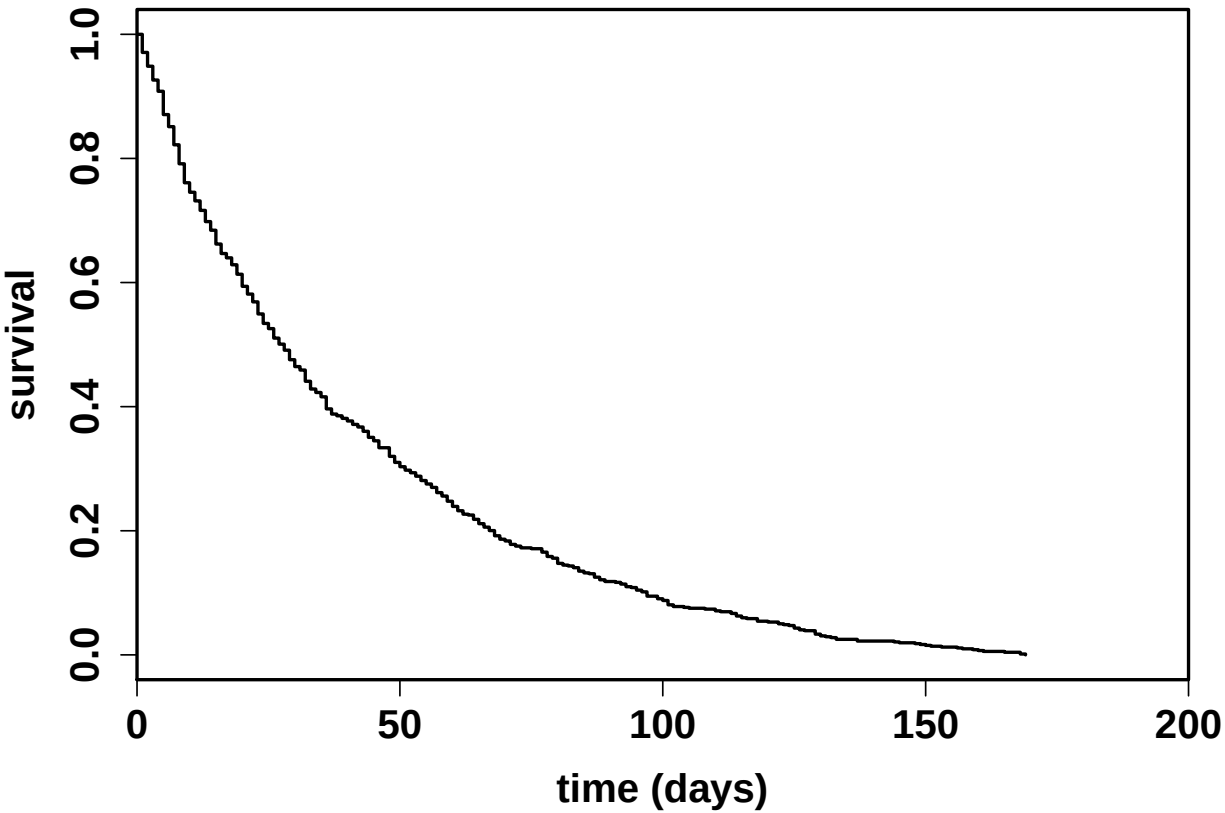
- Censoring has been ignored so far, but NOW we can deal with it....

TIMELINE FOR EVENTS AND CENSORING



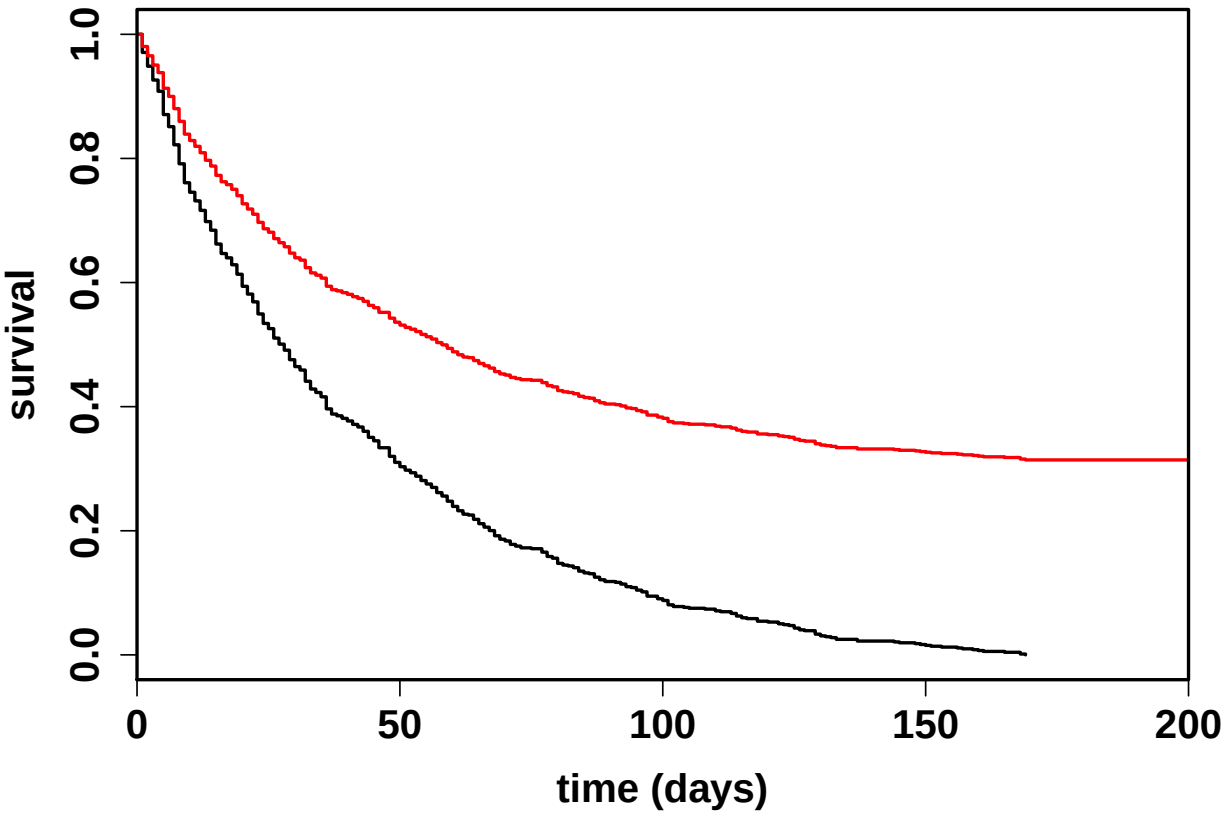
X = event O = censoring

KAPLAN-MEIER, *without* THE CENSORED



(Ignores censoring)

KAPLAN-MEIER, *with* THE CENSORED!



Red line: Kaplan-Meier with censoring

Independent expert opinion

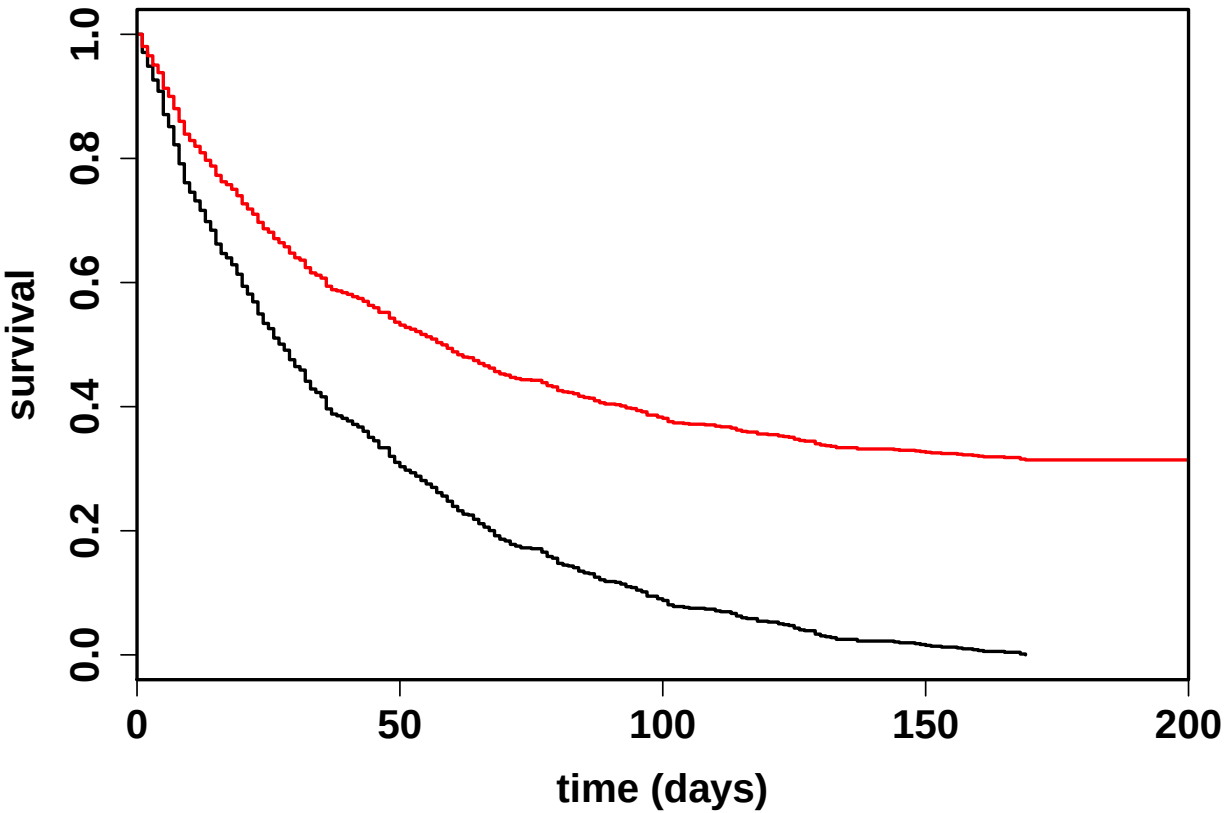
What to avoid

Two BAD approaches:

1. Remove all censored before analysis.
   Under-estimates the survival curve (as we have just seen)

2. Treat them as “real” endpoints, i.e. new admissions.
   Under-estimates the survival curve

KAPLAN-MEIER, *with* THE CENSORED!



Red line: Kaplan-Meier with censoring

Mean versus Median: "The final showdown"

Compute:

- Mean time to next infection
- Median time to next infection

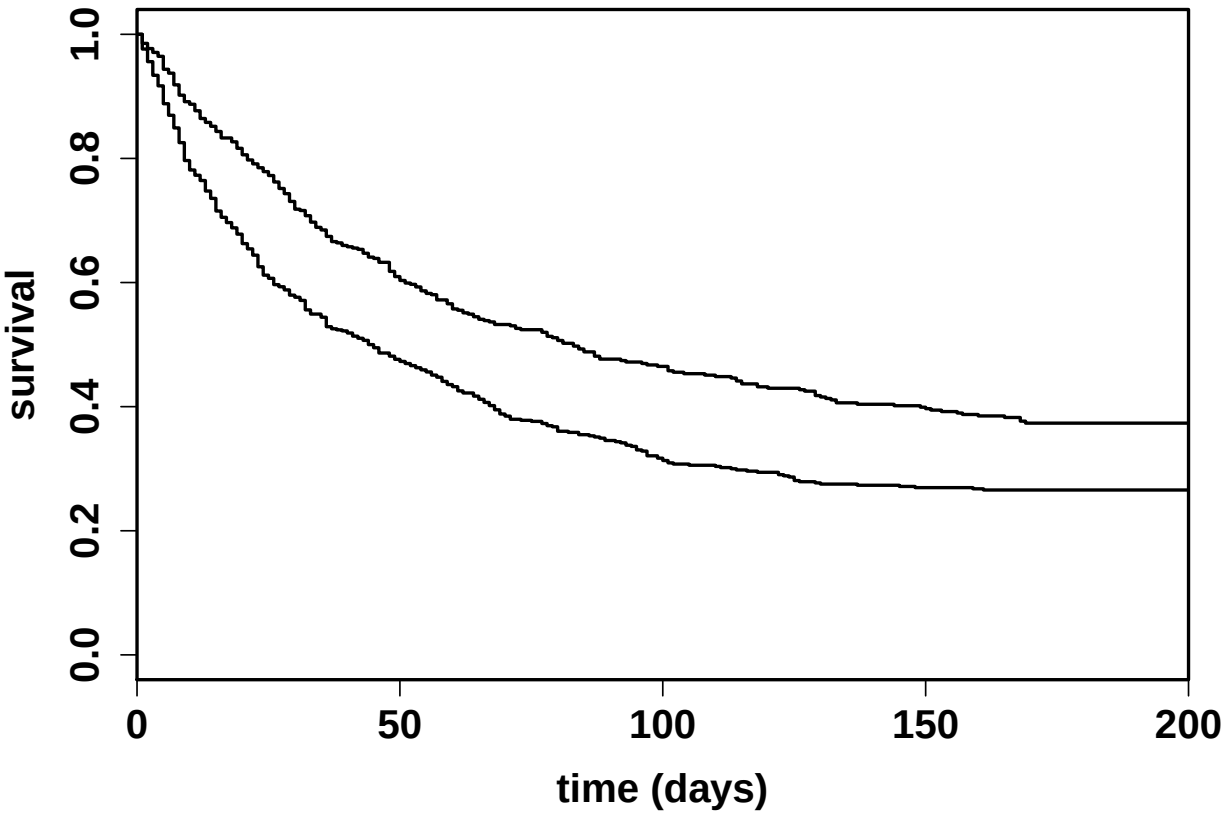
Results without censoring (Wrong!)

```
     n  events  median 0.95LCL 0.95UCL
   719     719      28      24      31

                 mean
                 40.1    37.3    43.1
```

Results with censoring (Correct!)

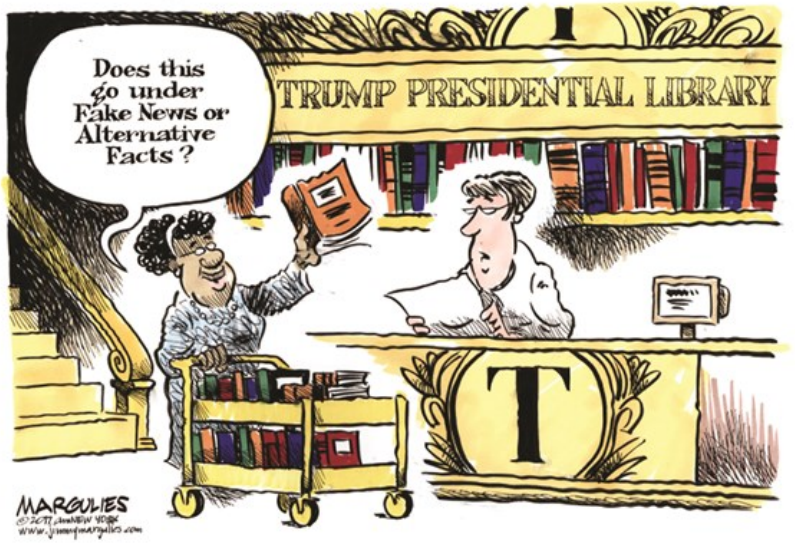
```
     n  events  median 0.95LCL 0.95UCL
  1069     719      58      51      66

                 mean
                 ???     ???     ???
```

KAPLAN-MEIER: COMPARE GROUPS

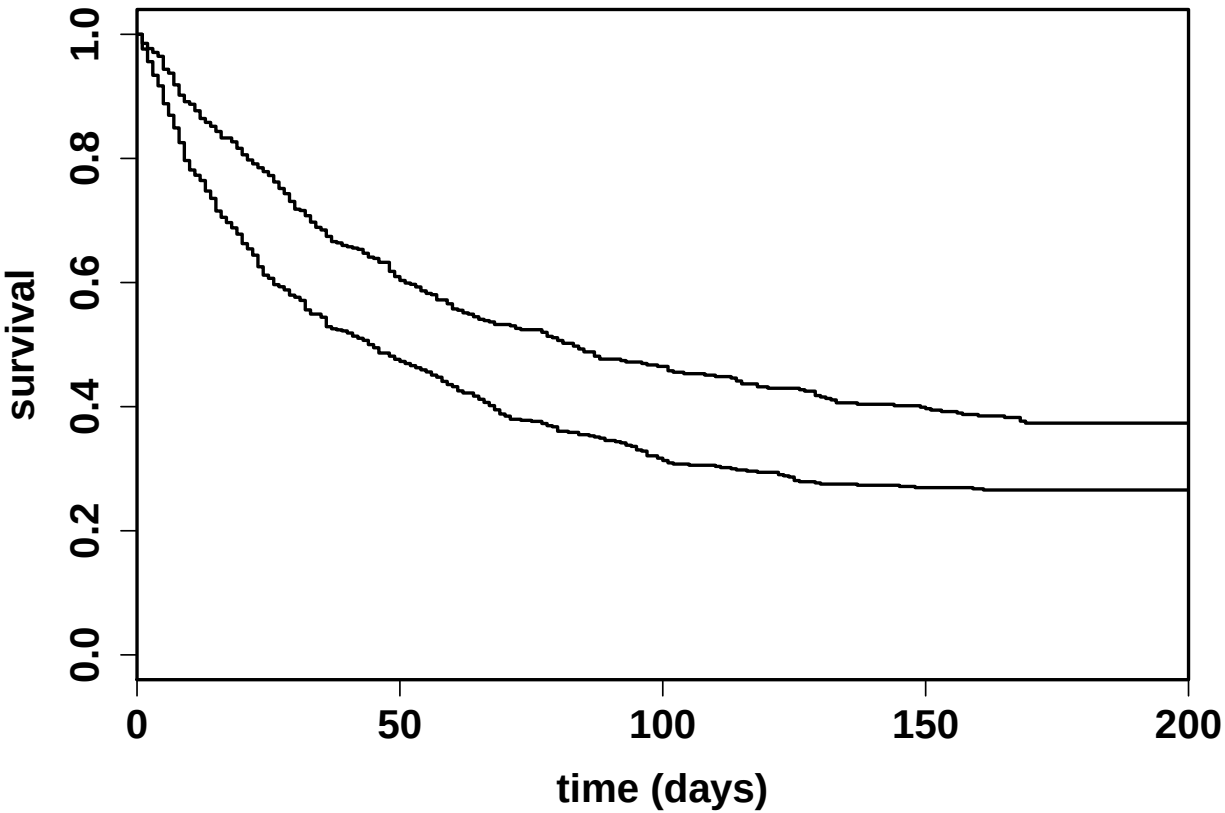


- Compare effect of zinc treatment with placebo
- Top curve is treatment, bottom curve is placebo

THE RESULTS ARE "FAKE NEWS" ABOUT ZINC



KAPLAN-MEIER: COMPARE GROUPS



- Compare effect of zinc treatment with placebo
- Top curve is treatment, bottom curve is placebo (we might wish!)

DATA, SELECTED VARIABLES

	id	date	age	sex	treat.orig	time	event	time.14	treat
1	1	2004-01-25	10	1	1	69	1	56	1
2	2	2004-03-22	13	2	1	123	1	110	0
3	6	2003-11-30	7	1	1	190	0	177	1
4	8	2003-12-02	5	2	0	185	0	172	1
5	9	2003-12-03	4	1	0	93	0	80	1
6	13	2003-12-24	6	2	0	183	0	170	1
:		:			:				:
:		:			:				:

`date`: date of inclusion

`age`: age (at inclusion) in months

`sex`: boys = 1, girls = 2

`time`: time since inclusion to event or censoring

`event`: new episode = 1, censored = 0

`time.14 = time - 13`: Starts counting after 14 days

`treat`: zinc = 1, placebo = 0

Define data as survival data

```
. stset time_14, failure(event==1)
```

```
     failure event:  event == 1
obs. time interval:  (0, time_14]
 exit on or before:  failure

--------------------------------------------------------------------------
     1069  total observations
        0  exclusions
--------------------------------------------------------------------------
     1069  observations remaining, representing
      719  failures in single-record/single-failure data
    84794  total analysis time at risk and under observation
                                                at risk from t =         0
                                     earliest observed entry t =         0
                                          last observed exit t =       205
```

TESTING THE DIFFERENCE

(We don't really need to test.... difference is obvious here!)

(But still... a p-value might be useful)

Log-rank test

- Preferred when hazards are (roughly) proportional

```
. sts test treat
```

```
time.14 = time - 13
event = 0 (censoring) or 1 (new episode)
treat = 0 (placebo) or 1 (zinc)
```

Log-rank test

```
. sts test treat
```

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

      |    Events      Events
treat |  observed    expected
------+-------------------------
0     |       427      360.79
1     |       292      358.21
------+-------------------------
Total |       719      719.00

          chi2(1) =       24.75
          Pr>chi2 =      0.0000
```

Wilcoxon-type test

- Preferred when hazards are non-proportional

```
. sts test treat, wilcoxon
```

Wilcoxon (Breslow) test for equality of survivor functions

treat	Events observed	Events expected	Sum of ranks
0	427	360.79	54718
1	292	358.21	-54718
Total	719	719.00	0

chi2(1) = 30.89
Pr>chi2 = 0.0000

COX REGRESSION

The group difference should be measured, not only tested!

Cox regression:

- Can *test* difference (more or less like log-rank)
- Can *measure* difference (as “Hazard Ratio”, HR)
- Can produce confidence intervals for difference
- Can adjust for other variables/confounders (multiple regression)

... one of the most frequently used methods in medical statistics....

But remember that it assumes:

- “Independent” censoring
- Proportional hazards

Hazard rate α

- "Instantaneous" probability of new event
- Same as incidence rate, but used in different settings:

  INCIDENCE RATE: - Estimated directly from data
  - Over a time interval of some length
  - Often with an "open population"

  HAZARD RATE: - A mathematical concept, estimated from the model
  - Instantaneous, i.e. over a "very short" time interval
  - Often with a "closed population" or at the individual level

- In our data:

  Events first 5 days: 21 + 16 + 16 + 13 + 27 = 93

  Total (to begin with): 1069

  Very roughly, $\alpha(0) = \frac{93}{1069 \times 5} = 0.017$

Thus, the hazard rate is about 1.7% new events per day to begin with

FOUR WAYS TO DESCRIBE SURVIVAL

Survival time $T \geqslant 0$ (disregard censoring for the moment...)

Cumulative distribution function:

$$F(t) = P(T \leqslant t)$$

Survival function:

$$S(t) = P(T > t) = 1 - F(t)$$

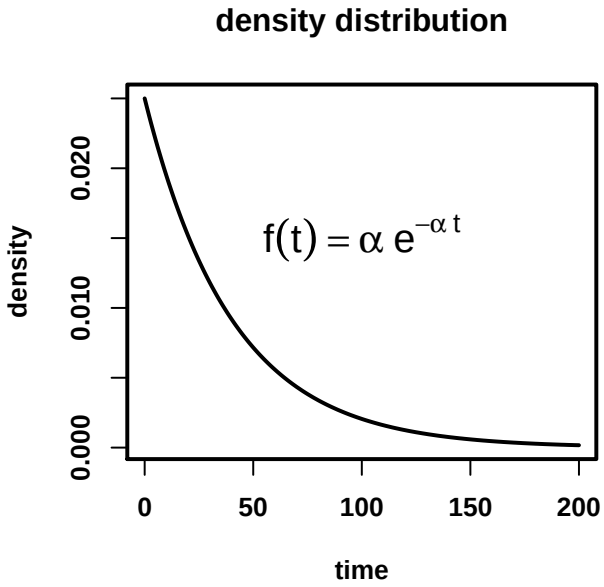
Distribution function (aka density):

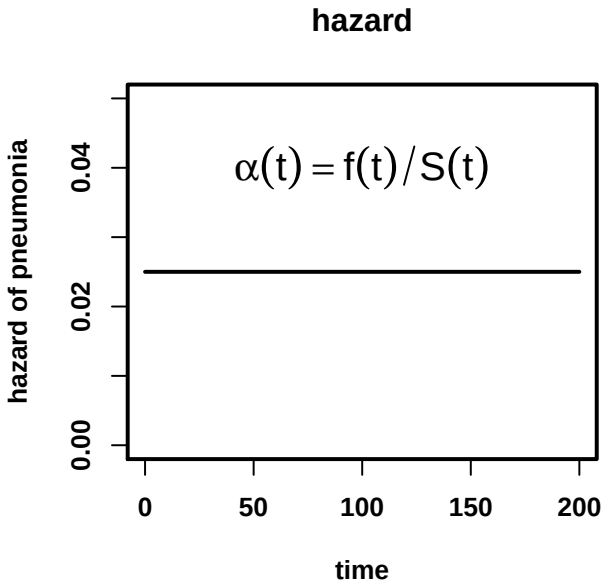
$$f(t) = F'(t)$$

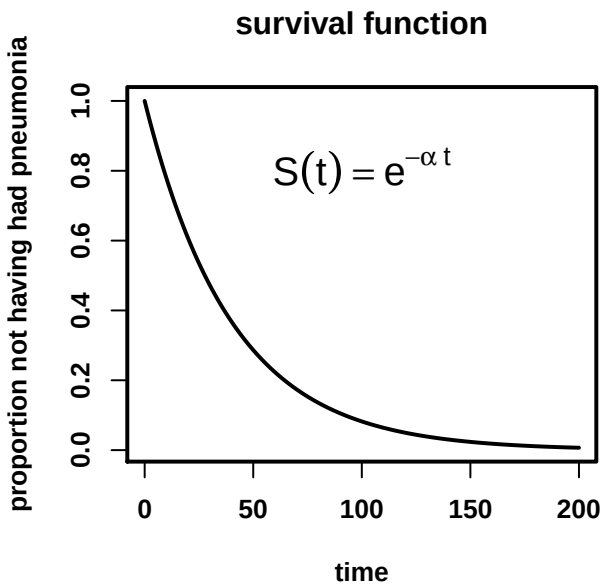
Hazard:

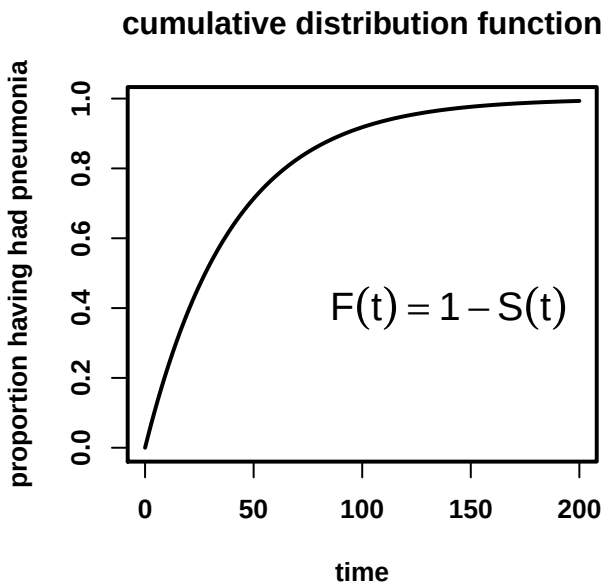
$$\alpha(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} P(t \leqslant T < t + \Delta t | T \geqslant t)$$

FOUR WAYS TO DESCRIBE SURVIVAL

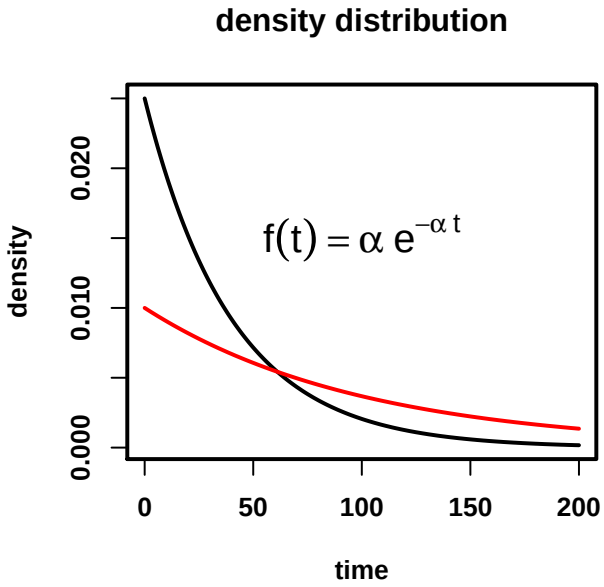


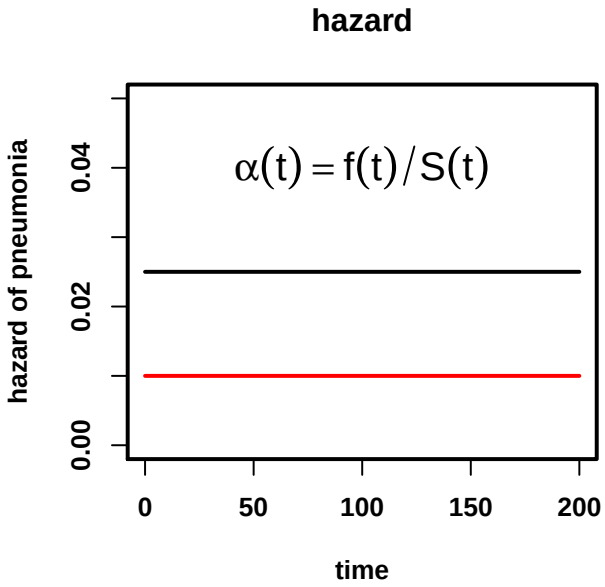


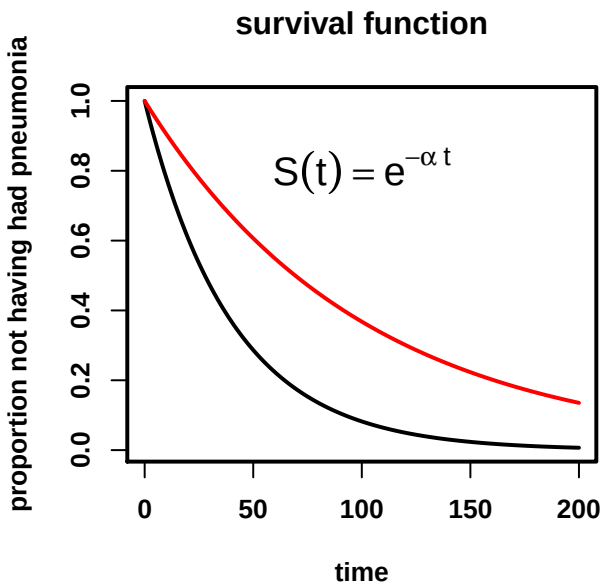


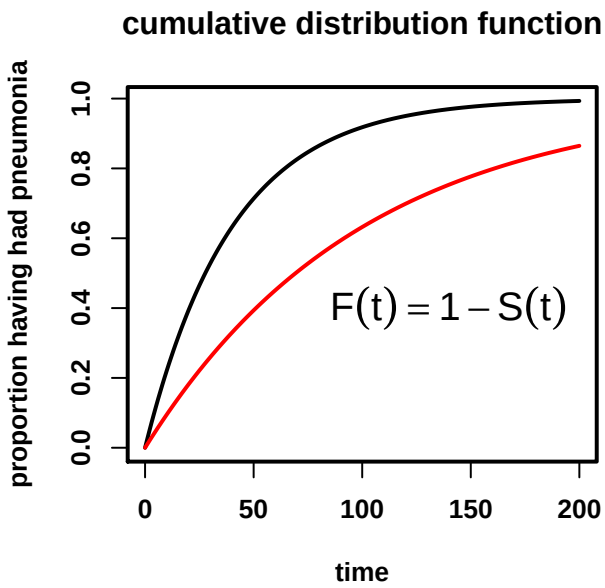


FOUR WAYS TO DESCRIBE SURVIVAL

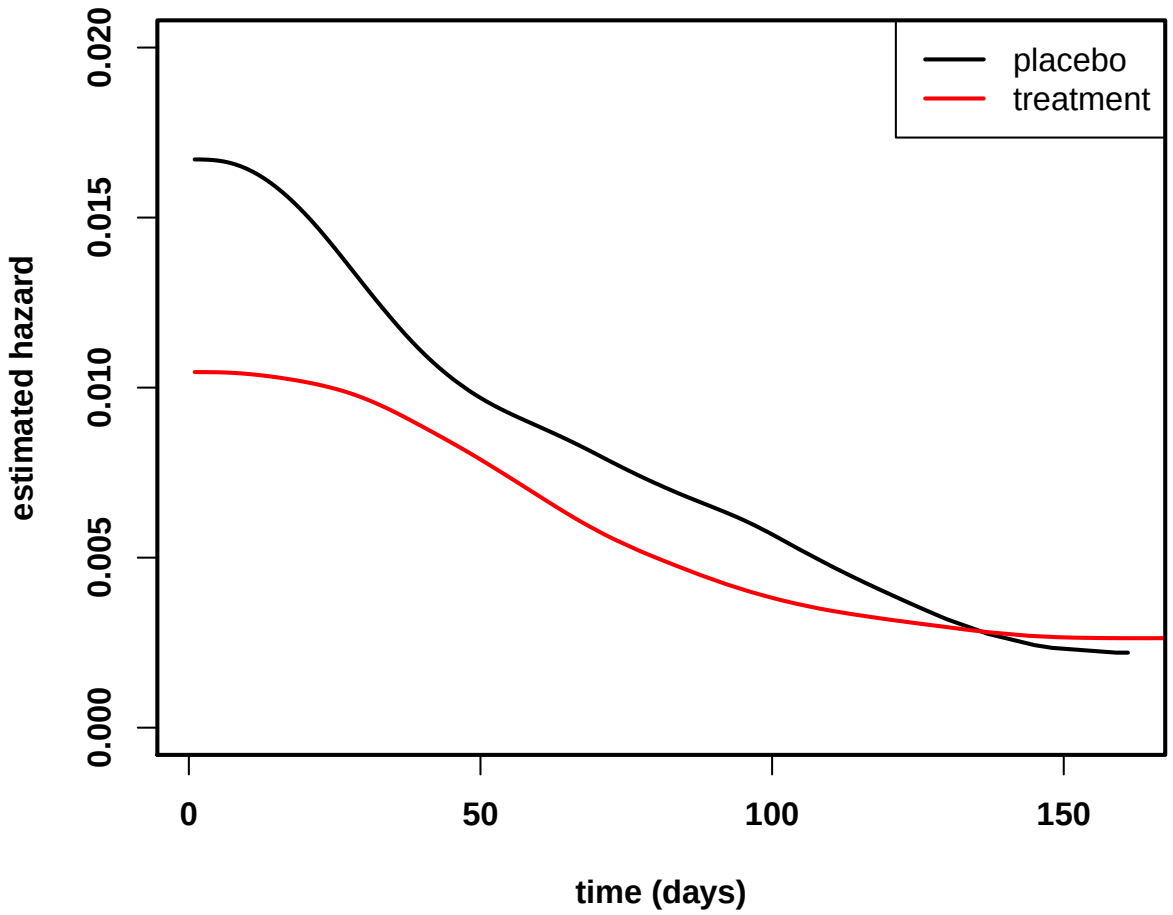








WHAT DO THE HAZARDS ACTUALLY LOOK LIKE?



Note: Hazards are notoriously difficult to estimate!
... but almost never needed

COX (PROPORTIONAL HAZARDS) REGRESSION

HAZARD:

$$\alpha(t) = \alpha_0(t) \exp(\beta_1 x_1 + \beta_2 x_2 + \cdots)$$

- $\alpha_0(t)$ is the *baseline hazard*
- $x_1, x_2, \ldots$ are the covariates
- $\beta_1, \beta_2, \ldots$ are the corresponding parameters

Covariates:

$x_1, x_2, \ldots$ are covariates as in any other regression,

continuous or categorical (using dummy variables)

Baseline hazard:

β_0 not needed, α_0 takes its role

$\alpha_0(t)$ is thus the hazard (at time t) when all $x_1 = x_2 = \cdots = 0$

COX (PROPORTIONAL HAZARDS) REGRESSION

FOR EXAMPLE:

$$\alpha(t) = \alpha_0(t) \exp(\beta_1 x_1)$$

$x_1 = 0$ (placebo) and $x_1 = 1$ (treatment)

$$\begin{aligned} \alpha_{\text{placebo}}(t) &= \alpha_0(t) \\ \alpha_{\text{treatment}}(t) &= \alpha_0(t) \exp(\beta_1) \end{aligned}$$

Hazard (rate) ratio

$$\text{HRR} = \frac{\alpha_{\text{treatment}}(t)}{\alpha_{\text{placebo}}(t)} = \frac{\alpha_0(t) \exp(\beta_1)}{\alpha_0(t)} = \exp(\beta_1)$$

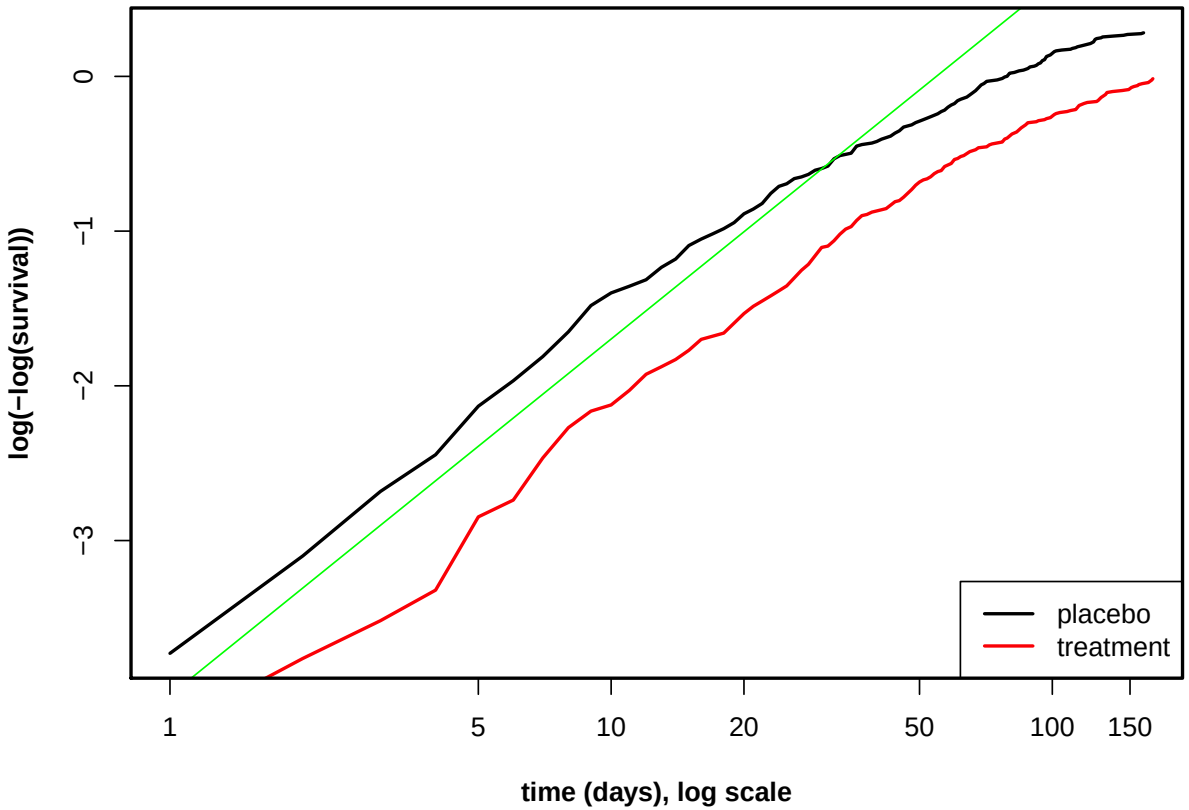
COX REGRESSION (OUR DATA)

```
. stcox treat
```

```
Cox regression -- Breslow method for ties

No. of subjects =         1069                  Number of obs    =      1069
No. of failures =          719
Time at risk    =        84794
                                                LR chi2(1)       =     24.61
Log likelihood  =   -4663.7197                  Prob > chi2      =    0.0000

------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       treat |   .6880192   .0523209    -4.92   0.000     .5927483    .7986028
------------------------------------------------------------------------------
```

Hazard is reduced to about 69% relative to no treatment

ARE HAZARDS PROPORTIONAL? "LOG-MINUS-LOG" PLOT



- **x-axis:** log(time)
- **y-axis:** log(-log S(t))

Rationale:

- Constant vertical distance $\rightarrow$ proportional hazards
- Lines with slope 1 $\rightarrow$ constant hazard (compare with green line)

Cox regression, remarks on Relative Risk

→ Separate presentation, Cox

COX REGRESSION, MULTIVARIATE

```
. stcox treat sex
```

```
Cox regression -- Breslow method for ties

No. of subjects =         1069                  Number of obs    =      1069
No. of failures =          719
Time at risk    =        84794
                                                LR chi2(2)       =     27.65
Log likelihood  =   -4662.2008                  Prob > chi2      =    0.0000

------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       treat |   .6890169   .0524052    -4.90   0.000     .5935936      .79978
         sex |   .8772892   .0660674    -1.74   0.082     .7569026    1.016824
------------------------------------------------------------------------------
```

- Adjustment for `sex` has little effect on `treat`
- `sex` is in itself not significant

(Note: In a randomized study you would not necessarily adjust for covariates)

COX REGRESSION, STRATIFIED

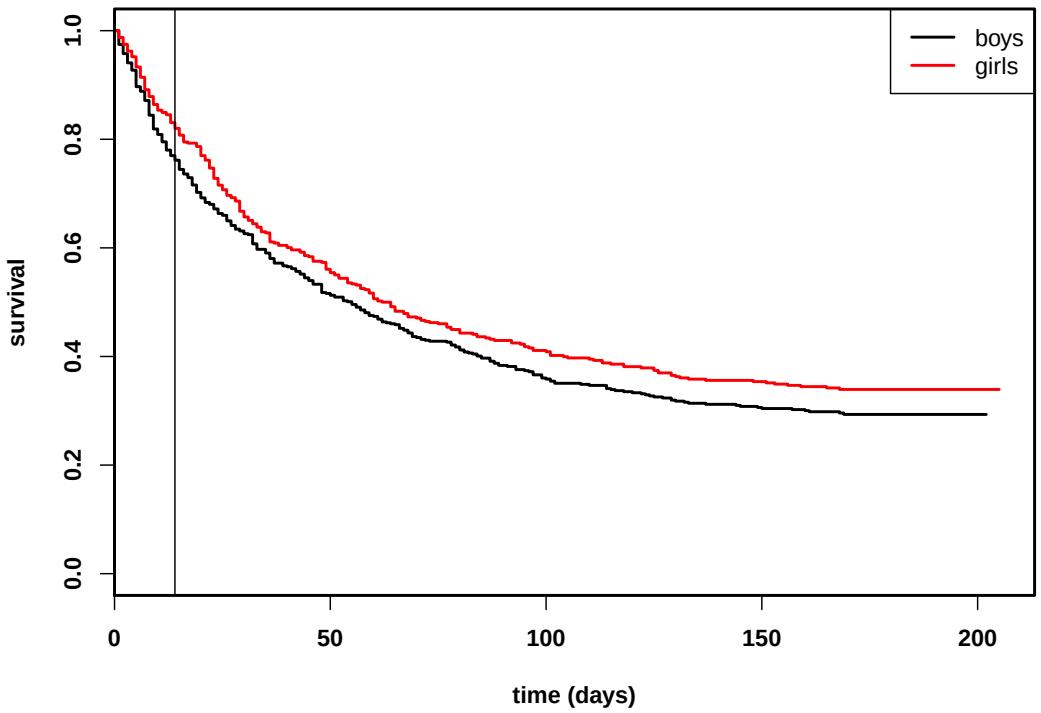
```
. stcox treat, strata(sex)
```

```
Stratified Cox regr. -- Breslow method for ties

No. of subjects =         1069                  Number of obs    =      1069
No. of failures =          719
Time at risk    =        84794
                                                LR chi2(1)       =     24.26
Log likelihood  =   -4171.5604                  Prob > chi2      =    0.0000

------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       treat |   .6896468   .0524831    -4.88   0.000     .5940856    .8005794
------------------------------------------------------------------------------
                                                        Stratified by sex
```

- Treatment effect is assumed equal in both strata of `sex`
- But separate baselines are allowed for boys and girls

(Note: This is just as an illustration; not really necessary)

SURVIVAL FOR BOYS AND GIRLS



Somewhat peculiar. Does something happen the first two weeks?

Cox regression, interaction

Is the effect of treatment *different* for boys and girls?

Warning: When dealing with interactions, the exact coding of variables is very important.

- Dummy variables make interpretations much easier.
- `treat` is already a dummy.
- R can (could!) do this more simply than Stata,
  but the following works for both.

Create dummy:

$$\texttt{sex01} = \texttt{sex} - 1$$

that is, `sex01` is a dummy variable for girl, with boys = 0, girls = 1

Create interaction term:

$$\texttt{treat.sex} = \texttt{treat} \times \texttt{sex01}$$

→ Separate presentation, Coding of interactions

COX REGRESSION, STANDARD INTERACTION

```
. stcox treat sex01 treat_sex
```

```
Cox regression -- Breslow method for ties

No. of subjects =         1069                     Number of obs    =      1069
No. of failures =          719
Time at risk    =        84794
                                                   LR chi2(3)       =     31.36
Log likelihood  =   -4660.3458                     Prob > chi2      =    0.0000

------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       treat |   .6047332   .0621173    -4.90   0.000     .4944582    .7396019
       sex01 |   .7767475   .0768479    -2.55   0.011     .6398318    .9429615
   treat_sex |   1.343316   .2059711     1.92   0.054     .9946334    1.814233
------------------------------------------------------------------------------
```

(Interaction is borderline significant, should probably just be dumped)

Interpretation:

$$\text{Treatment HRR for girls} = 1.34 \times \text{treatment HRR for boys}$$

Create two interaction terms (and leave out `treat` from equation):

$$\texttt{treat\_sex\_0} = \texttt{treat} \times (1 - \texttt{sex01}) \quad \text{(Effect among boys)}$$

$$\texttt{treat\_sex\_1} = \texttt{treat} \times \texttt{sex01} \quad \text{(Effect among girls)}$$

```
. stcox sex01 treat_sex_0 treat_sex_1
```

COX REGRESSION, INTERACTION, ALTERNATIVE CODING

```
Cox regression -- Breslow method for ties

No. of subjects =         1069                  Number of obs    =      1069
No. of failures =          719
Time at risk    =        84794
                                                LR chi2(3)       =     31.36
Log likelihood  =   -4660.3458                  Prob > chi2      =    0.0000
```

_t	Haz. Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
sex01	.7767475	.0768479	-2.55	0.011	.6398318	.9429615
treat_sex_0	.6047332	.0621173	-4.90	0.000	.4944582	.7396019
treat_sex_1	.8123476	.0925445	-1.82	0.068	.6497871	1.015577

- Risk of new pneumonia in boys is reduced to 60.5% with treatment
- Risk of new pneumonia in girls is reduced to 81.2% with treatment