Regression modeling in survival analysis – Part 1

Michael Otterstatter

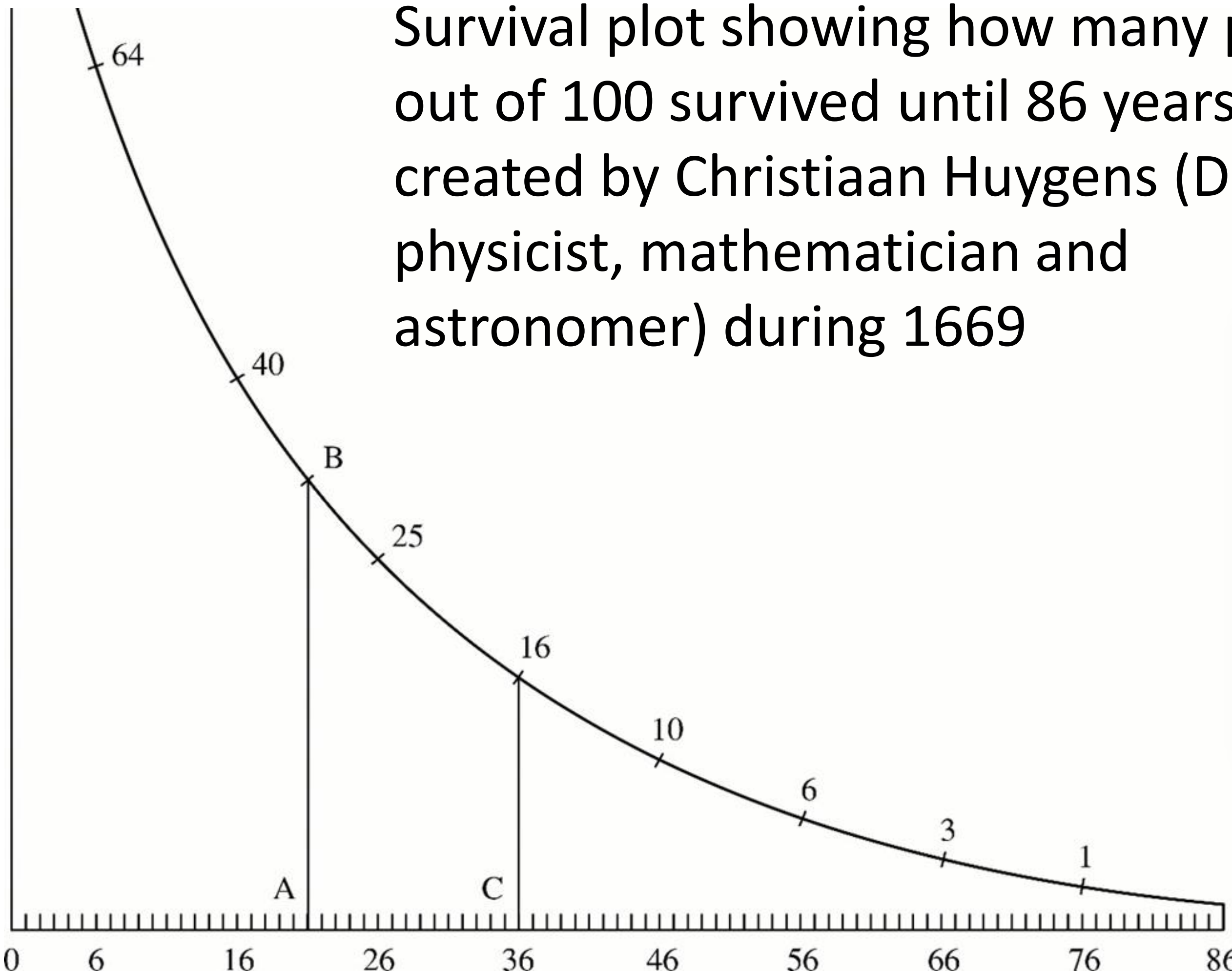
BCCDC Biostats Session

October 25, 2019

Session overview

- In this session we will discuss
  - basic concepts of survival analysis
  - components of survival data
  - regression modeling for survival data

Survival analysis, circa 1669



Survival plot showing how many people out of 100 survived until 86 years of age, created by Christiaan Huygens (Dutch physicist, mathematician and astronomer) during 1669

Wainer, H. Annual Review of Psychology. Vol. 52: 305-335.

Background

- Simply put, ‘survival analysis’ is the analysis of longitudinal event data, specifically the <u>time-to-event</u>
- Often, and historically, these analyses focussed on the survival, or time-to-death, of people
- But, the same models apply to the time to injury, illness, admission, readmission, recovery, or any definable health or disease state, and even the time to failure of machines!

Background

- Consider typical survival data, where individuals are followed over time as they move from one state (alive) to another (dead)

- Typically we would model binary states (alive/dead) with logistic regression and continuous variables (time) with ordinary linear regression

- But neither model is appropriate when we have binary state change occurring over time, i.e., we care about both the state change *and* when it occurred (the ‘time-to-event’)

Background

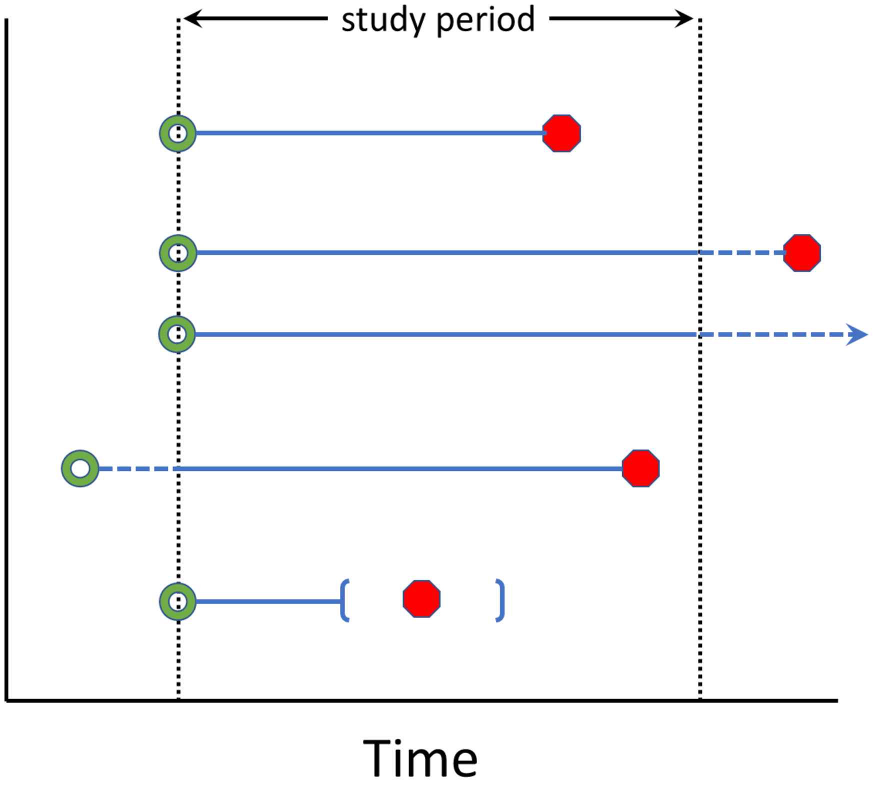
- As with other regression problems, we wish to
  - *estimate* (on average, how long did group A survive?)
  - *compare* (did group A survive longer than group B?)
  - *model* if a particular set of covariates predicts survival time
- One new concept particularly relevant for survival analysis: censoring

Concepts: Censoring

- Censoring: not all individuals will have the event of interest by the end of study (or, may be lost to follow-up or have a competing outcome)
- Note that censored individuals still provide information for our analysis because they have some follow-up time without an event

Concepts: Censoring

- *Right censoring*: event occurs unknown time after end of follow-up
- *Left censoring*: event occurs unknown time before start of follow-up
- *Interval censoring*: event occurs at unknown time within a study interval

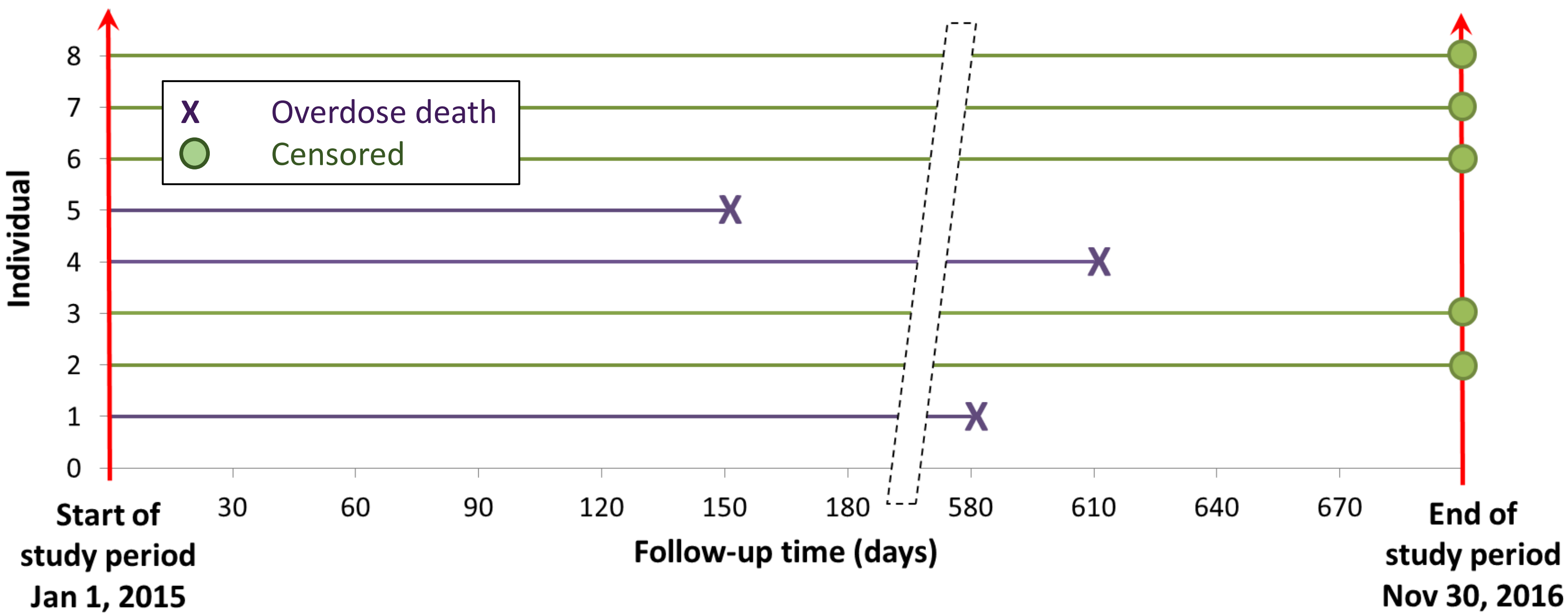


Survival analysis

1. Define event, time zero, time scale and exit
   - Event of interest: typically a single, clear-cut event (e.g., death, diagnosis, re-admission) but could be repeated (recurrent) events or multiple different (competing) events
   - Time zero (origin): beginning of follow-up, e.g., a fixed point in calendar time, a baseline age, a time of exposure or diagnosis, etc.
   - Time scale: usually calendar time, but could be age
   - How participants exit study: typically, when they have the event of interest or are censored (end of study, or lost to follow-up)

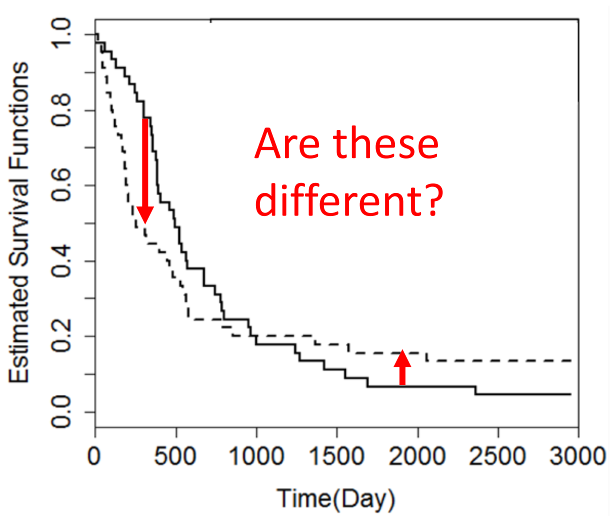
Example – analysis design

- *Event of interest*: death due to drug overdose
- *Time zero*: Jan 1, 2015
- *Time scale*: calendar time in days
- *Exit*: end of study (Nov 30, 2016) or overdose death

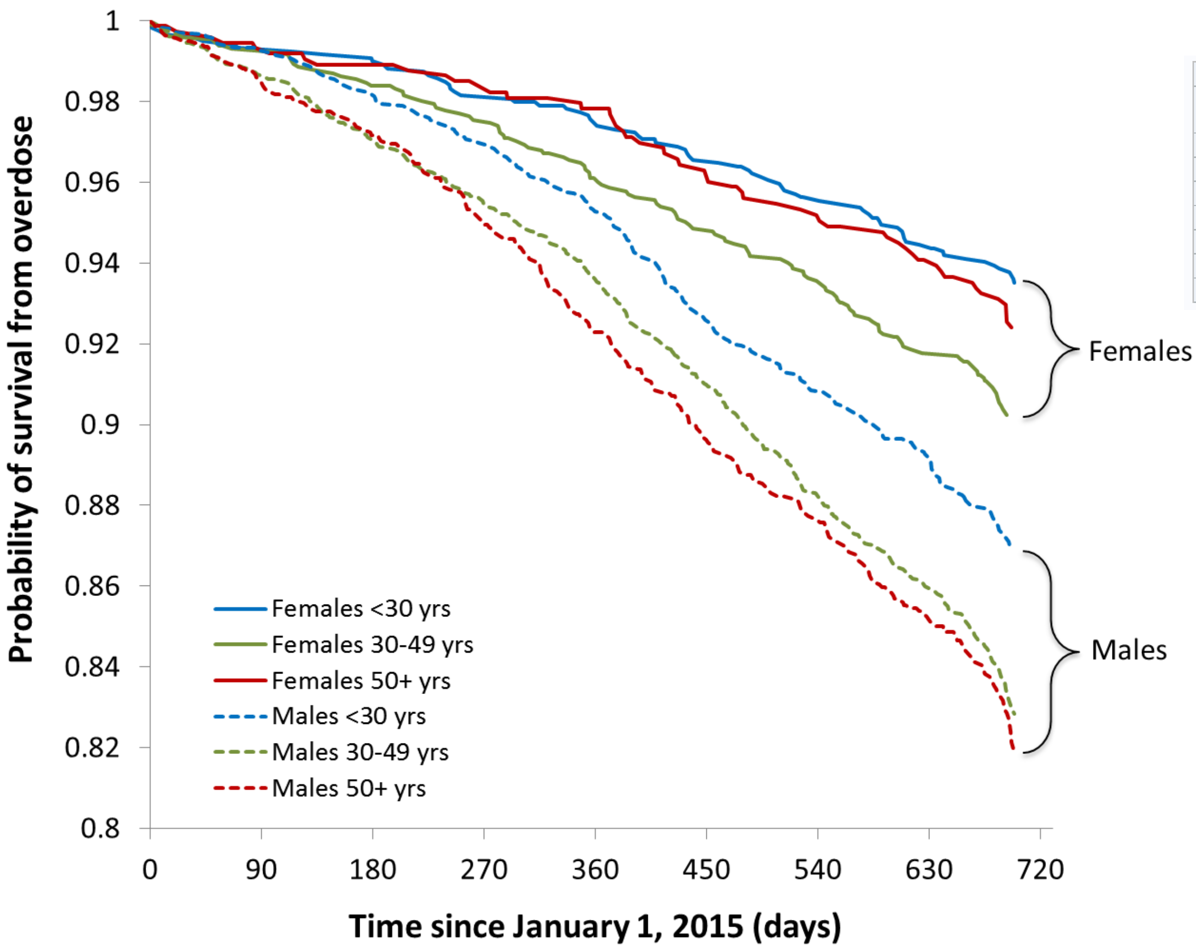


Survival analysis

2. Descriptive analysis: univariate modeling
   - non-parametric Kaplan-Meier curves describe survival distribution; provide proper median & quartile statistics
   - provide simple univariate comparisons between groups
   - statistical tests (e.g., log rank, Wilcoxon) can be used to compare curves but must be interpreted cautiously



Example – KM curves



Summary of the Number of Censored and Uncensored Values

Stratum	gender	age_grp	Total	Failed	Censored	Percent Censored
1	F	30-50	1312	127	1185	90.32
2	F	50+	744	55	689	92.61
3	F	<30	1192	77	1115	93.54
4	M	30-50	3118	531	2587	82.97
5	M	50+	1425	249	1176	82.53
6	M	<30	2082	272	1810	86.94
Total			9873	1311	8562	86.72

Test of Equality over Strata

Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	146.9066	5	<.0001
Wilcoxon	146.6612	5	<.0001
-2Log(LR)	153.6749	5	<.0001

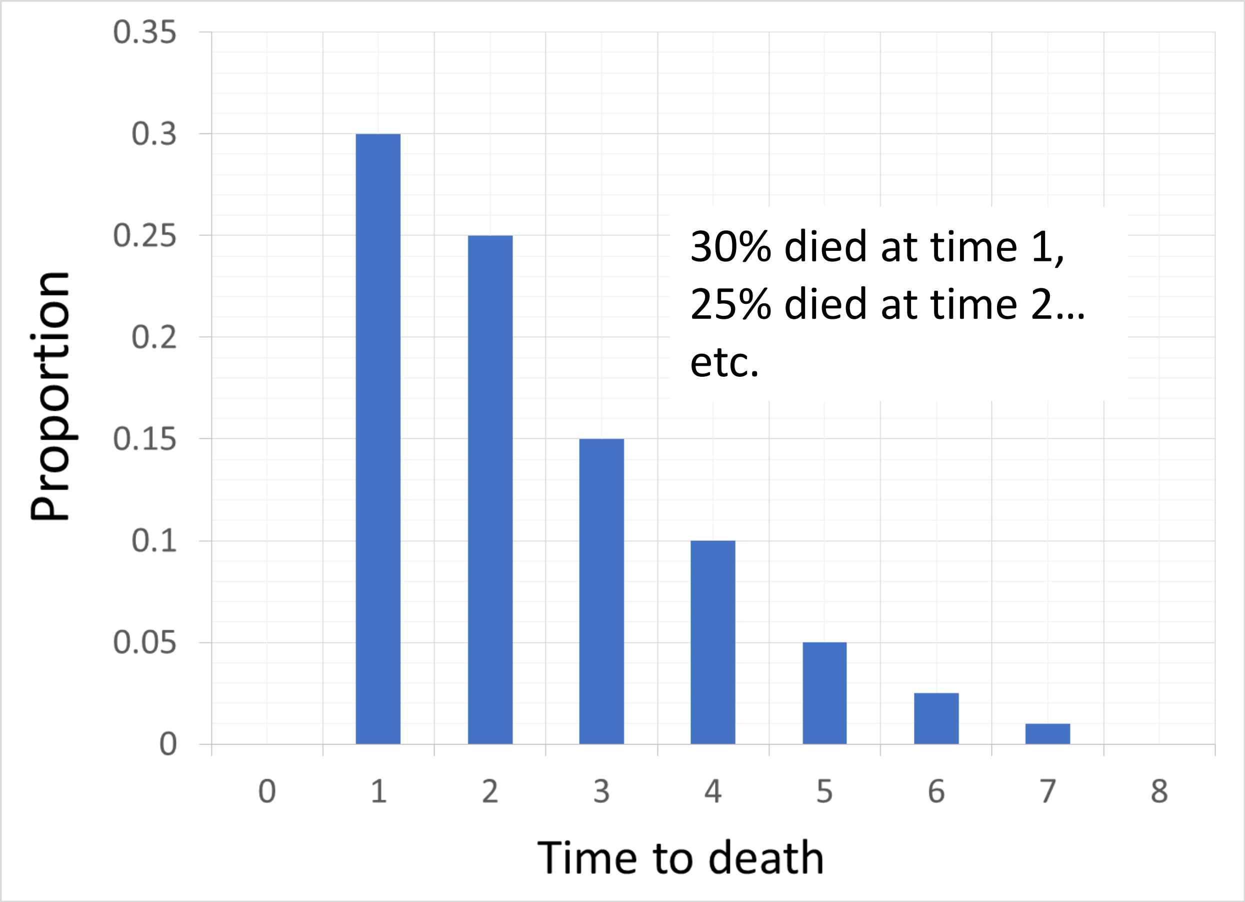
Survival analysis

3. Inferential analysis: multivariate modeling
    - Cox regression (semi-parametric)

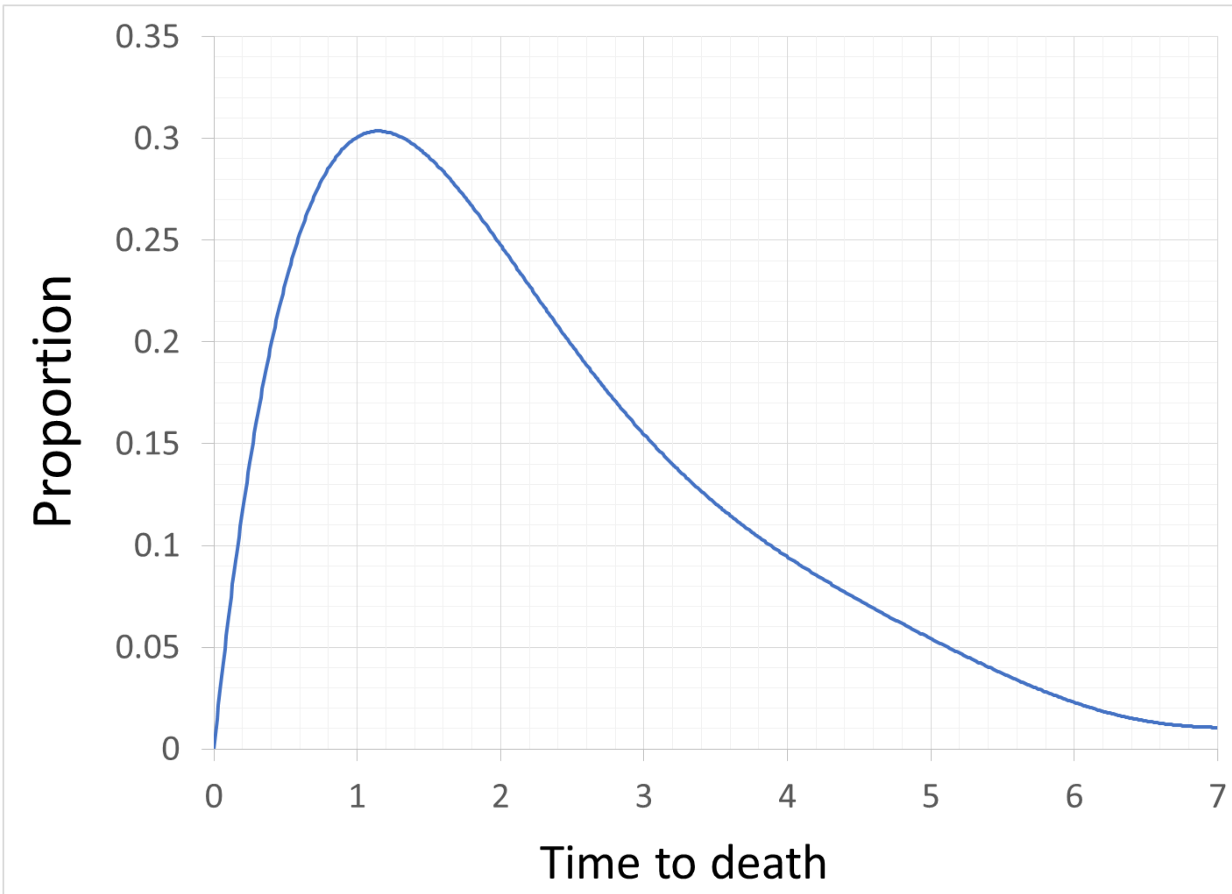
$$h_i(t) = h_0(t)e^{\beta_1 x_1 + \beta_2 x_2 \dots}$$

Wait, we need some more concepts first…

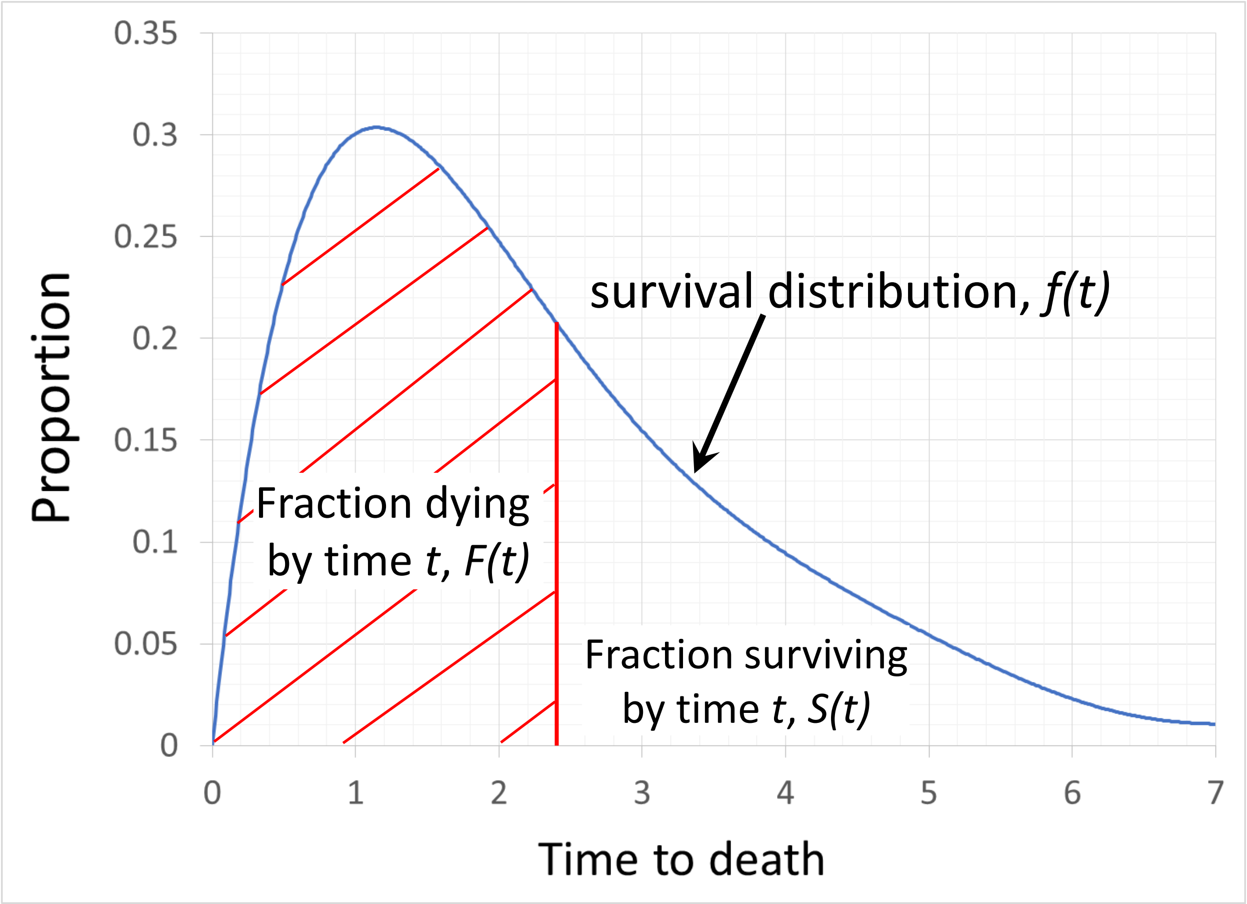
Concepts: Survival and hazard



Concepts: Survival and hazard



Concepts: Survival and hazard



Concepts: Survival and hazard

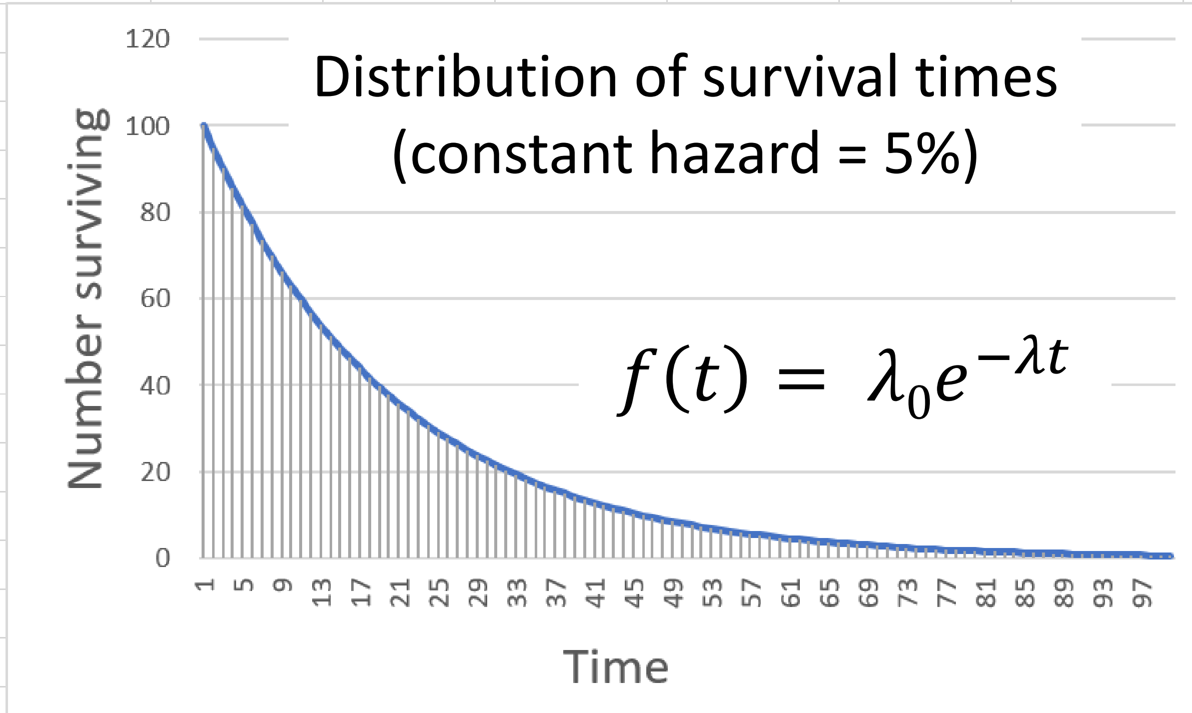
- *f(t)*: distribution (probability density) of survival times
- *F(t)*: proportion of population dying by time *t* (cumulative distribution of *f(t)*)
- Survival function 1 – *F(t)* or *S(t)*: proportion of population surviving by time t
- Hazard function *h(t)*: instantaneous risk of death at time *t* (or, probability of death in the next small interval)

Concepts: Survival and hazard

- Often our interest is in modeling the hazard (e.g., risk of death), but what form should it take?
  - the simplest would be to assume a constant hazard (i.e., risk of death remains the same over time)
  - What would survival times look like if we have a constant hazard?

Constant hazard

time	prop dying	num surviving
0	0.05	100
1	0.05	95
2	0.05	90
3	0.05	86
4	0.05	81
5	0.05	77
6	0.05	74
7	0.05	70
8	0.05	66
9	0.05	63
10	0.05	60
11	0.05	57
12	0.05	54
13	0.05	51
14	0.05	49

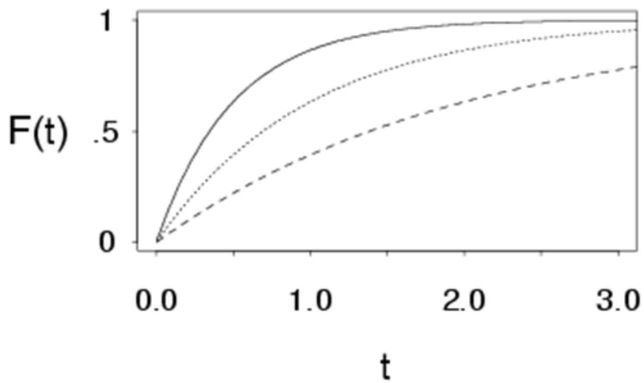


$$f(t) = \lambda_0 e^{-\lambda t}$$

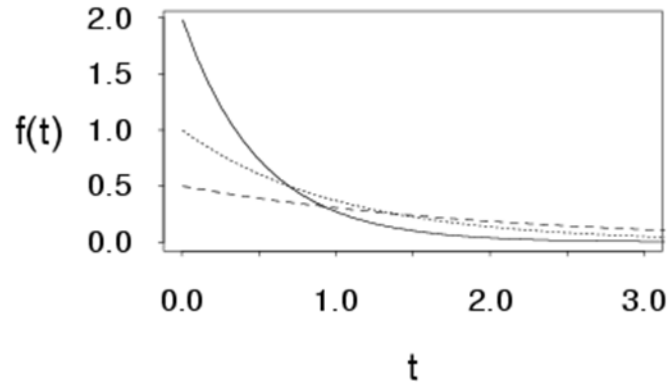
Assuming a constant hazard results in an exponential distribution of survival times

Constant hazard

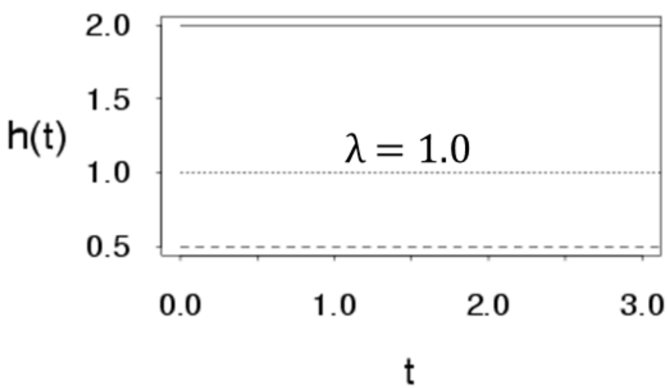
Fraction dying at *t*



Time to death



Hazard at *t*



Next time…

- Regression models for survival data

References

- Columbia University Mailman School of Public Health. Population Health Methods. Time to event data analysis. https://www.mailman.columbia.edu/research/population-health-methods/time-event-data-analysis
- George H. Dunteman & Moon-Ho R. Ho. 2011. Survival Analysis. *In*, An Introduction to Generalized Linear Models. SAGE Publications, Inc.
- McCullagh P, Nelder JA. 1989. *Generalized Linear Models*. Chapman & Hall.
- O'Quigley, J., 2008. *Proportional hazards regression* (Vol. 542). New York: Springer.
- Sainani, K.L. Introduction to Survival Analysis. Stanford University Department of Health Research and Policy. https://web.stanford.edu/~kcobb/index.html