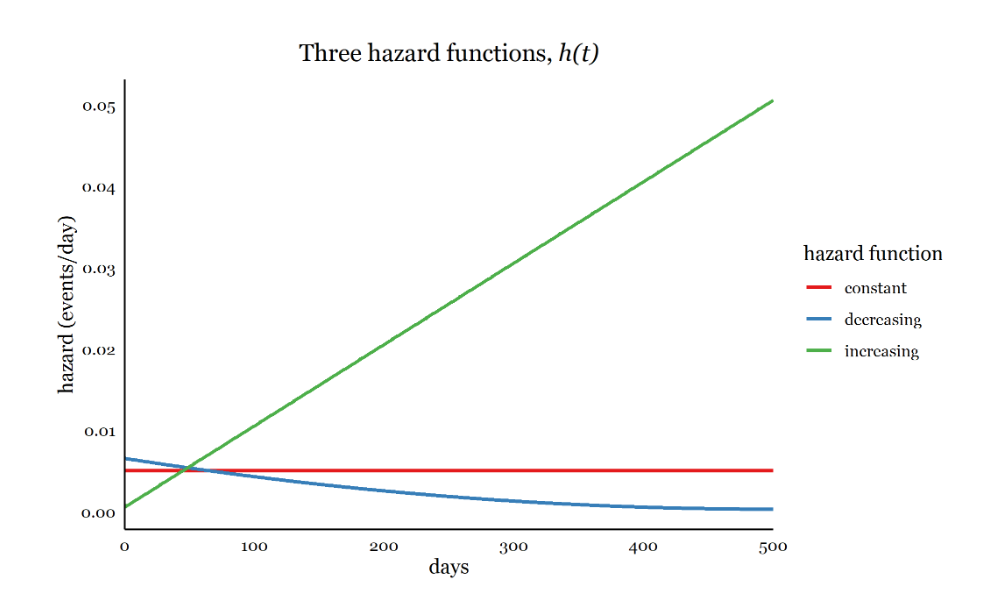
Survival analysis

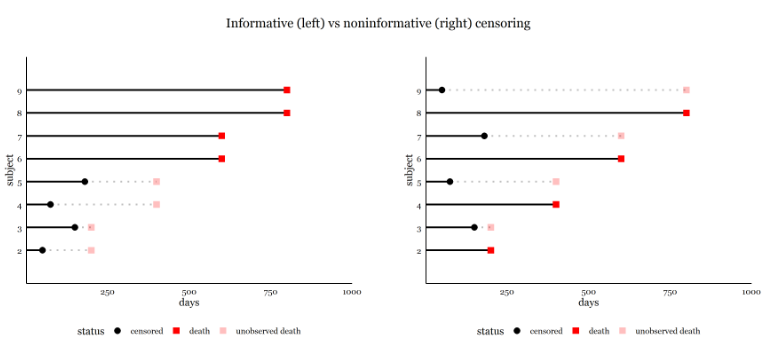
* Models how much time before an event occurs
* Outcome variable = length of time to an event/survival time/failure time/time to event
* Event = often referred to as failure
* One goal is to estimate probability that a subject survives without experiencing the event past some time
* Sometimes true survival time (T) can’t be observed (censoring)
* Follow-up-time = true survival time or censoring time
* The survival function, S(t), expresses the probability that a subject’s true survival time (T) will exceed time, t, i.e. that the subject will survive beyond time, t
* A graph with a green line

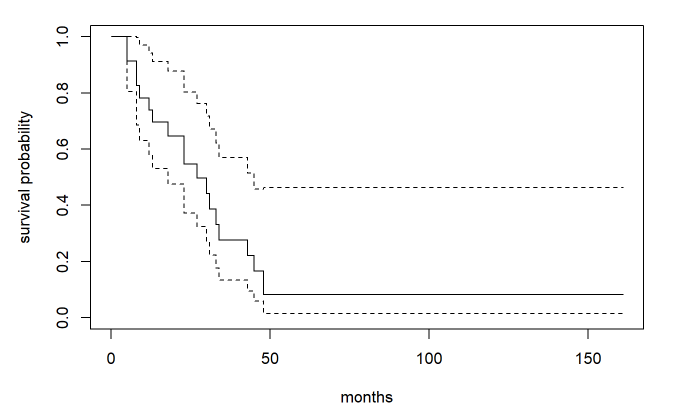
  AI-generated content may be incorrect.
  + Assumptions
    - S(0) = 1: all subjects survive the very first moment
    - S(∞) = 0: all subjects fail after infinite time
  + Median survival time = the time, t, at which half of the population is expected to still be surviving
  + A graph with a green line

    AI-generated content may be incorrect.
* Hazard function
  + Kaplan-Meier estimator can be used to estimate S(t) directly
  + Cox model focuses on the hazard function or rate, h(t) which is inversely related to the S(t)
  + Hazard function at time t, h(t) = instantaneous rate of events at time t, given that the subject has survived until time, t
  + 
  + Increasing hazard function means more events per unit of time, meaning decreasing survival
  + h(t) ≥ 0, so the hazard function can never be negative
* Cumulative hazard
  + Cumulative hazard function, H(t) is how much hazard a subject has accumulated over time up to time, t
  + The probability that a subject will fail over time increases as the hazard accumulates
  + Because h(t) ≥ 0, H(t) can never decrease with time
  + The survival function is inversely related to the cumulative hazard function as a subject’s cumulative hazard grows, the survival probability decreases
  + S(t) = exp(-H(t))
  + So by modelling either the S(t) of h(t) we can infer the other
* Censoring
  + Many times the exact time when the event occurs is unknow or censored
  + Right censoring = subjects’ actual survival time is greater than their observed time
    - Study ends before event
    - Subject lost to follow up
    - Subject no longer at risk after study begins
    - A graph with black lines and red dots

      AI-generated content may be incorrect.
  + Left censoring = subjects’ actual survival time is less than their observed time
    - Common in diseases infection – positive tests might be delayed by days or even years
    - Interval censoring means that a subject’s survival time is unknown but known to lie between two observed time points
    - A screenshot of a graph

      AI-generated content may be incorrect.
  + Assumption of noninformative censoring
    - A subject’s censoring time should not be related to the unobserved survival time
    - Distribution of censoring times and survival times are unrelated



* Data for survival analysis
  + Simplest structure
    - Single row per subject
    - A status variable coding whether the subject experienced the event or not (censored)
    - A single time variable measuring T, time to event (or censoring time, time of last observation)
    - Variables for covariates, assumed to be time-constant in this structure
* Kaplan-Meier estimation with survfit()
  + n = total number at risk
  + Events = total number of events that occurred
  + Median = survival time, t, at which S(t) = .5 or the time at which 50% of those at risk are expected to be alive
  + 0.95LCL, 0.95UCL = 95% confidence limits for median survival
  + KM survival function can be graphed
  + It estimates the survival function as a step function, where S(t) only changes at timepoints when an event occurs
    - So the true underlying survival curve may be smooth – but the KM curve is limited by the number of event times observed
  + 
  + Dotted lines are the confidence intervals
* The Cox proportional hazards model
  + Estimates changes to h(t) instead of s(t)
  + The Cox model an estimate the effects of multiple predictors/covariates
    - No distribution assumed for survival times
    - Naturally accommodates right-censoring and time-varying covariates
    - Can be extended in many ways:
      * Time-varying covariates
      * Random effect frailties for recurrent events or clustering
      * Competing risks modelling
  + - : the hazard at time, t, for a subject with predictor X1 equal to the value of x1
    - : the baseline hazard at time, t, the hazard for a subject with all predictors equal to zero
    - : the hazard ration comparing the hazard for a subject with X1=x1 to a subject with X1 = 0
  + If 0 = control, and 1 = treatment
    - A hazard ratio of .25 means that treatment has a quarter the hazard of control, or a 75% decrease
    - A HR of 2 means that treatment has twice the hazard of control, or a 100% increase – worse survival
  + In general exp(b1) expresses the hazard ratio for a one unit increase in the associated covariate
  + Proportional hazards
    - Standard Cox model assumes proportional hazards, meaning that the effects of covariates are constant over time
    - The Cox model is easily extended to accommodate multiple predictors, each of whose effects are assumed to be proportional over time
    - h(t|X1, X2, … Xp) = h0(t)exp(b1X1 + b2X2 + … + bpXp)
    - each coefficient bi can be exponentiated to calculate a hazard ratio