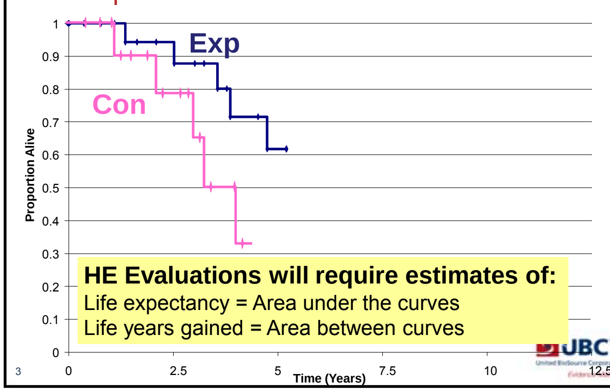
Parametric Survival Analysis Overview

Jack Ishak, PhD
United BioSource Corporation

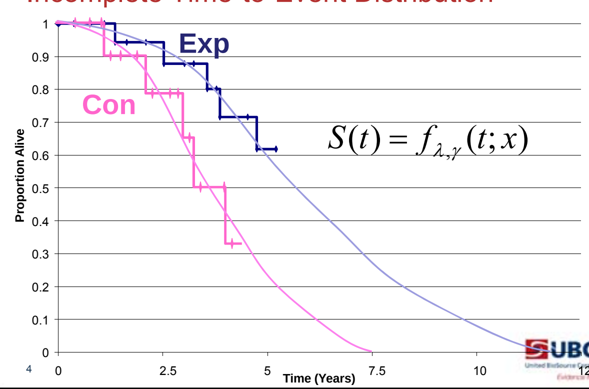
Background

- Health economic evaluations require data on the costs and benefits of treatments over extended periods of time, possibly over the lifetime of patients
- The data used to support the evaluations cover a relatively short time window (no more than a few years)
- Observed time to event distributions must be projected

Incomplete Time-to-Event Distribution



Incomplete Time-to-Event Distribution



Survival Analysis

- The methods that are most often employed to analyze time-to-event data are
  - Kaplan-Meier + Log-Rank/Wilcoxon Test
    - Produces empirical estimate of the time-to-event distribution and compare between groups
  - Cox (proportional hazard) regression
    - Measure the effect of multiple predictors without modeling underlying distribution
    - Assuming proportional hazards between levels of predictors
- Neither of these methods produce an estimate of the functional form of the underlying distribution

UBC United BioSource Corporation *Evidence Matters*

Parametric Survival Analysis

- Explicitly models the functional form of the event times using various statistical distributions
- Most commonly used
  - Exponential
  - Weibull
  - Gompertz
  - Log-Logistic
  - Log-Normal
  - Gamma (not covered)
- Generally involve two parameters
  - Scale (λ) and Shape (γ) parameters
- Shape generally assumed constant across individuals
- Scale related to determinants via regression
  - Can quantify the effect of predictors, particularly treatment

$$g(\lambda_i, \gamma) = \beta_{CON} + \beta_{EXP-CON} \times Ind(EXP)_i$$

UBC United BioSource Corporation *Evidence Matters*

Parametric Survival Analysis

- Conceptually same as linear case, but *Normal* replaced by appropriate distribution
- It is implemented in a regression framework, estimated by maximizing the likelihood of the data
  - For patients observed to have event at time *t*
    - Likelihood contribution: P(T=t) = f(t) (density function)
  - For patients censored at time t
    - Likelihood contribution: Prob = P(T> t) = S(t) (survival function)

UBC United BioSource Corporation *Evidence Matters*

Functions Characterizing Parametric Distributions

- Density Function: $f(t) = P[T=t]$
- Cumulative Incidence: $F(t) = P[T \leq t]$
- Survival Distribution: $S(t) = P[T > t\,]$
- Hazard Function
  $h(t) = P\,[\,t \leq T < t + \Delta \mid T > t] \div \Delta \quad \text{for } \Delta \rightarrow 0$

UBC United BioSource Corporation *Evidence Matters*

Relationship between Survival and Hazard Functions

- Survival as a function of hazard

$$S(t) = \exp\left[-\int_0^t h(s)\ ds\right]$$

- Hazard as a function of survival

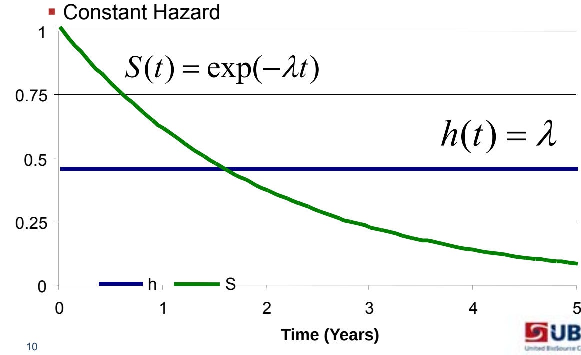
$$h(t) = -\frac{d}{dt}\log S(t)$$

- Example: constant hazard $h(t) = \lambda$

$$S(t) = \exp[-\lambda \times t]$$

Exponential Distribution

- Constant Hazard



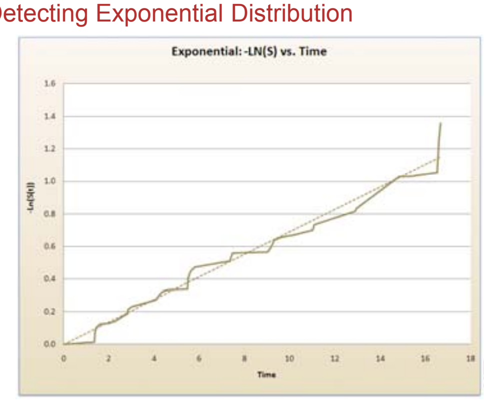
Exponential Distribution: Constant Hazard

- Graphical test for exponential distribution

$$S(t) = \exp(-\lambda t) \qquad \log(S(t)) = -\lambda t$$

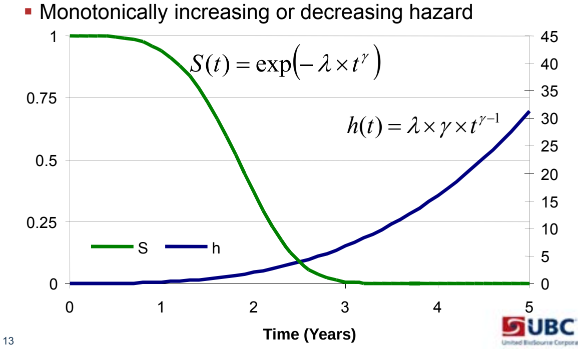
- Process
  - –Obtain estimates of S(t) using a KM procedure
  - –Calculate log(S(t)) and plot against time
- If log(S) vs. time is relatively linear, underlying distribution likely linear
- Slope of the line is the hazard of the event

Detecting Exponential Distribution



Weibull Distribution

- Monotonically increasing or decreasing hazard



13

Weibull Distribution

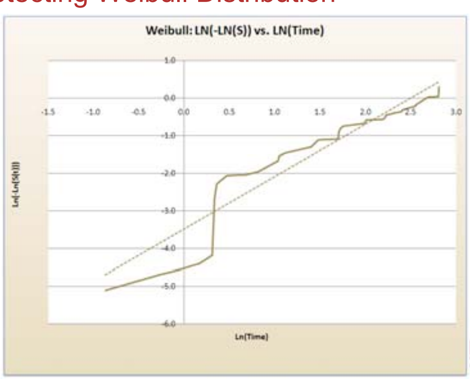
- Shape
  - –γ > 1: Hazards increase over time
  - –γ = 1: Hazard is constant (Exponential Distribution)
  - –γ < 1: Hazards decreases over time
- Graphical test

$$\log\left(-\log(S(t))\right) = \log(\lambda) + \gamma \log(t)$$

- If log(-log(S)) vs. log(time) is relatively linear, underlying distribution likely Weibull

14

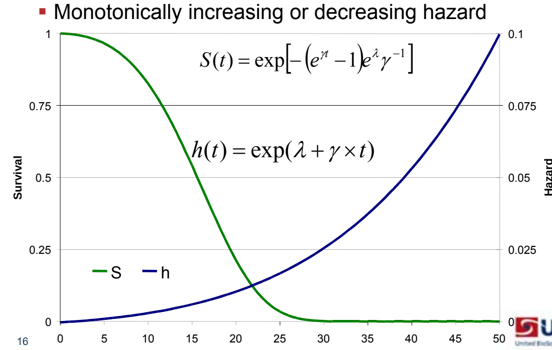
Detecting Weibull Distribution



15

Gompertz Distribution

- Monotonically increasing or decreasing hazard



16

Gompertz Distribution

- Shape
  - $\gamma > 0$: Hazards increase exponentially over time
  - $\gamma = 0$: Hazard is constant (Exponential Distribution)
  - $\gamma < 0$: Hazards decreases exponentially over time
- Graphical Test

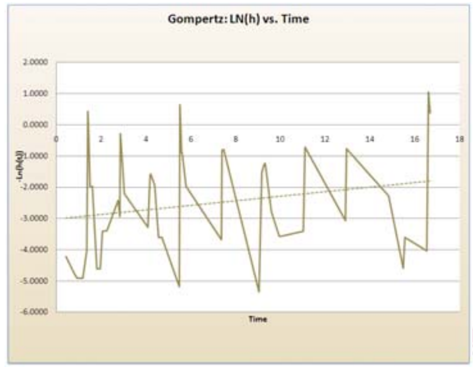
$$\log(h(t)) = \lambda + \gamma \times t$$

- If log of empirical hazards vs. time is linear, underlying distribution may be Gompertz
  - Empirical hazards:

$$\hat{h}(t_i) = \frac{-\left[\log\left(S(t_i)\right) - \log\left(S(t_{i-1})\right)\right]}{t_i - t_{i-1}}$$

17

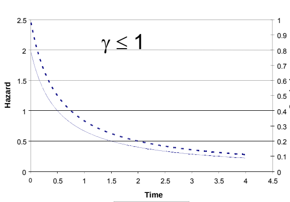
Detecting Gompertz Distribution

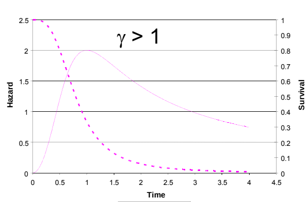


18

Log-Logistic Distribution

$$S(t) = \frac{1}{1 + \lambda \times t^{\gamma}} \quad h(t) = \frac{\lambda \times \gamma \times t^{\gamma - 1}}{\left(1 + \lambda \times t^{\gamma}\right)^2}$$





19

Log-Logistic Distribution

- Graphical test

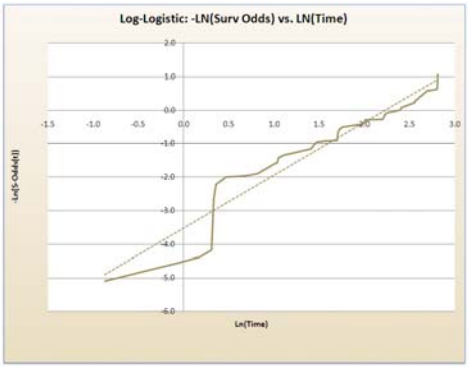
$$Odds = \frac{S(t)}{1 - S(t)} = \lambda \times t^{\gamma}$$

$$\log(Odds) = \log(\lambda) + \gamma \log(t)$$

- So if log(S/1-S) vs. log(time) is relatively linear, underlying distribution likely log-logistic

20

Detecting Log-Logistics Distribution



Log-Normal Distribution

- Very similar to log-logistic distribution with $\gamma > 1$
- Specification:

  Log(T) ~ Normal(μ,σ)

- Survival Distribution

$$S(t) = 1 - \Phi\left[\frac{\log(t) - \mu}{\sigma}\right]$$

- Where Φ is the standard cumulative normal distribution

Log-Normal Distribution in SAS

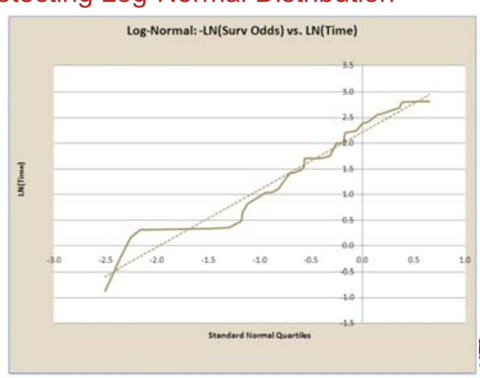
- Graphical Test

$$\log(t) = \mu + \Phi^{-1}\left[1 - S_i(t)\right] \times \sigma$$

  – Φ^{-1} available in standard software

- If the underlying distribution is truly normal, the observed log-times should line up with normal quartiles corresponding to the observed failure probabilities

Detecting Log-Normal Distribution

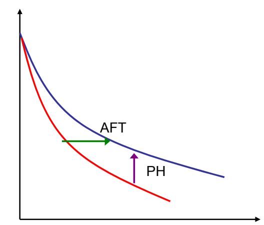


Identifying Best Fitting Distribution

- Start with exploratory analyses graphs to indentify plausible fits
- Use fit statistics to help distinguish between fits
  - Log-Likelihood, AIC, BIC
- Factor in contextual (e.g., clinical) knowledge about the outcome
  - What shape of hazard makes most sense?
  - Do projections based on the distributions make sense? Which is the most clinically plausible?
- Does a single distribution work for the entire sample (e.g., for experimental vs. control groups)
  - Important to look at groups separately at first
  - Even if both look Weibull (for example), if the shape parameter estimates are different, a single model is inappropriate

Parametric Survival Regression:
PH vs. AFT Models

- PH Models
  - Effect of determinants is proportional (multiplicative) on hazards
  - Determinants increase or decrease hazards
  - Assumed to be constant over time
- AFT Models
  - Effect of determinants is proportional (multiplicative) on survival time
  - Determinants accelerate or delay the occurrence of the event
  - Assumed to be constant for any survival percentile
    - Median time accelerated or delayed by



- Some models can be parameterized either way
  - Weibull, exponential, log-logistic

Regression Formulations

- Exponential, Weibull, Gompertz, Log-Logistic

$$\frac{-\log(\lambda_i)}{\gamma} = -X_i\beta$$

  - Hazard ratio (or odds ratio of LL): exp(-β×γ)
  - AFT Factor: exp(β)
- Log Normal

$$\lambda_i = X_i\beta$$

  - λ is mean log-time of event
  - σ is the std deviation
  - AFT Factor: exp(β)

Regression Analysis Steps

- Identify distribution that appears to fit best to the data
  - If have important groups (e.g., treatment), look at them separately
- Identify potential predictors
  - Look at each on it's own, or in presence of key variables (e.g., treatment)
  - Look at significance and clinical importance
- Build multi-variable regression model
  - Iteratively, using potential predictors identified previously
  - Consider both clinical and statistical significance
  - Keep only what is necessary (parsimonious model)