2024 Summer Institute In Statistics for Clinical & Epidemiological Research

Module 3:

Design, Conduct, and Analysis of Randomized Clinical Trials with Time to Event Primary Endpoints

Lecture 10:

Parametric Inference with Time to Event Data

Scott S. Emerson, M.D., Ph.D.
Professor Emeritus of Biostatistics
University of Washington

1

General Analysis Models

Choice of Summary Measures

Used for Inference

2

Summarizing Effect

- Based on marginal distributions
  - Difference / ratio of means (arithmetic, geometric, …)
  - Difference / ratio of proportion exceeding some threshold
  - Difference / ratio of medians (or other quantiles)
  - Ratio of odds of exceeding some threshold
  - Ratio of hazard (averaged across time?)
  - …
- Based on joint distribution
  - Median difference / ratio of paired observations
  - Probability that a randomly chosen measurement from one population might exceed that from the other
  - …

3

Statistical Models

- Options for inference
  - Parametric models
    - Weibull, lognormal, etc.
  - Semiparametric models
    - Proportional hazards, etc.
  - Nonparametric
    - Weighted rank tests: logrank, Wilcoxon, etc.
    - Comparison of Kaplan-Meier estimates

4

(Semi)parametric vs Nonparametric

- Choice of statistical model can affect
  - Computational methods for estimating the summary measure
  - Precision of summary measure estimates
  - Robustness of inference about the summary measure
  - Ability to estimate the summary measure

5

General Analysis Models

Probability Models

6

Right Censored Data

- Notation:

Unobserved:

True times to event: $\{T_1^0, T_2^0, \ldots, T_n^0\}$

Censoring Times: $\{C_1, C_2, \ldots, C_n\}$

Observed data:

Observation Times: $T_i = \min(T_i^0, C_i)$

Event indicators: $D_i = \begin{cases} 1 & \text{if } T_i = T_i^0 \\ 0 & \text{otherwise} \end{cases}$

7

Probability Distributions

Failure time $T > 0$ measures time to an event:

Cumulative distribution function: $F(t) = \Pr(T \le t)$

Survivor function: $S(t) = \Pr(T > t) = 1 - F(t)$

Density: $f(t) = \frac{d}{dt} F(t)$

Hazard function: $\lambda(t) = \lim_{h \downarrow 0} \frac{\Pr(T \in [t, t+h) \mid T \ge t)}{h}$

Cumulative hazard function: $\Lambda(t) = \int_0^t \lambda(u)\, du$

Censoring variable distribution: C has cdf $G(\cdot)$; pdf $g(\cdot)$

8

Parametric Models

- F is known up to some finite dimensional parameter vectors

$$F(t) = \Psi(t, \vec{\Phi})$$

where:

$\Psi(\cdot,\cdot)$ has known form

$\vec{\Phi}$ is finite dimensional and unknown

9

Parametric Inference

- Parametric inference generally proceeds through likelihood methods
  - MLE found by Newton-Raphson iteration
- Asymptotic distributions from theory of regular problems

$$L(\vec{\Phi}; \vec{T}, \vec{D}) = \prod_{i=1}^{n} \left(f(T_i; \vec{\Phi})(1 - G(T_i)) \right)^{D_i} \left(S(T_i; \vec{\Phi}) g(T_i) \right)^{1-D_i}$$

- Under independent censoring

$$L(\vec{\Phi}; \vec{T}, \vec{D}) \propto \prod_{i=1}^{n} \left[\left(f(T_i; \vec{\Phi}) \right) \right]^{D_i} \left[\left(S(T_i; \vec{\Phi}) \right) \right]^{1-D_i}$$

10

Parametric Summary Measures

Mean: $\hat{\theta} = \int_0^\infty u \, f(u; \hat{\vec{\Phi}}) \, du$

Median: $\hat{\theta} = F^{-1}(0.5; \hat{\vec{\Phi}})$

Proportion above threshold: $\hat{\theta} = \int_a^\infty f(u; \hat{\vec{\Phi}}) \, du$

Weighted average of hazard: $\hat{\theta} = \int_0^\infty w(u) \lambda(u; \hat{\vec{\Phi}}) \, du$

11

Parametric Survival Models

- Commonly used parametric survival models are generally accelerated failure time models
  - Exponential (constant hazard)
  - Weibull (monotonic hazards)
  - Gamma (monotonic hazards)
  - Lognormal (increasing, then decreasing hazard)
  - Log logistic (increasing, then decreasing hazard)
  - Families joining several of the above

12

Exponential Survival Models

- For $T \sim \mathcal{E}(\lambda)$, $f(t;\lambda) = \lambda e^{\{-\lambda t\}}$, $S(t;\lambda) = e^{\{-\lambda t\}}$
- Mean-variance relationship $E[T] = \frac{1}{\lambda}$, $\text{Var}[T] = \frac{1}{\lambda^2}$
- The exponential model corresponds to a constant hazard λ
  - It is thus a proportional hazards family
- The exponential model also corresponds to an accelerated failure time model: $S(t;\lambda) = S(t\lambda;1)$
  - Hence, when comparing distributions with parameters λ_1, λ_2, all quantiles are proportional
  - $Mdn[T] = \frac{\log 2}{\lambda}$

13

13

Exponential Regression Models

- For covariate vector $\vec{X}$, we model the hazard (mean) according to
$$log(\lambda) = \vec{X}\ \vec{\beta}$$
- Estimating equations for one sample

$$L(\vec{\Phi}, \vec{T}, \vec{D}) \propto \prod_{i=1}^{n} \lambda^{\{D_i\}} e^{-\lambda T_i}$$

$$\frac{\partial}{\partial \lambda} L(\vec{\Phi}, \vec{T}, \vec{D}) = \frac{\sum D_i}{\lambda} - \sum T_i$$

$$\hat{\lambda} = \frac{\sum D_i}{\sum T_i} \ \dot{\sim}\ N\left(\lambda, \frac{\sum D_i}{(\sum T_i)^2}\right)$$

14

14

Exponential Distribution

- If the hazard can be approximated as nearly constant over the support of the censoring distribution, we can estimate the hazard from the
  - number of observations,
  - mean observation time, and
  - number of observed events
- The exponential distribution is memoryless, which is a rather strong assumption in human lifetimes, but sometimes is not too bad an assumption for residual lifetime in very serious disease

$$\Pr(T \geq s + t \mid T \geq s) = \Pr(T \geq t)$$

15

15

Weibull Survival Models

- For $T \sim \mathcal{W}(\lambda, p)$, $f(t; \lambda, p) = p\lambda(\lambda t)^{p-1} e^{-(\lambda t)^p}$, $S(t; \lambda) = e^{-(\lambda t)^p}$
- Mean-variance $E[T] = \frac{1}{\lambda}\Gamma\left(1 + \frac{1}{p}\right)$

$$Var[T] = \frac{1}{\lambda^2}\left[\Gamma\left(1 + \frac{2}{p}\right) - \left\{\Gamma\left(1 + \frac{1}{p}\right)\right\}^2\right]$$

- Hazard function $\mathrm{h}(t; \lambda, p) = p\lambda(\lambda t)^{p-1}$
- Log hazard function is linear in log time

$$\log \mathrm{h}(t; \lambda, p) = (\log p + p \log \lambda) + (p - 1)\log t$$

- Weibull can be good approximation for any distribution that is log linear in log t over the support of the censoring distribution

16

Weibull Survival Models

- If shape parameter $p = 1$, then Exponential
  - $p > 1$ has monotonically increasing hazard
  - $p < 1$ has monotonically decreasing hazard
- For fixed shape parameter p, then
  - Proportional hazards family, and
  - Accelerated failure time model (all quantiles proportional)
    - $Mdn(T) = \frac{(\log 2)^{1/p}}{\lambda}$
- Assuming constant p, regression models: $\log \lambda = \boldsymbol{X}\vec{\beta}$
  - β_j either log hazard ratio or log median ratio for $\Delta X_j = 1$

17

Weibull Survival Models

- Graphical diagnostics for Weibull distribution:
  $$\log(-\log S(t)) = p(\log t + \log \lambda)$$
- The logarithm of a Weibull random variable is related to an extreme value distribution
- The Weibull distribution can thus be motivated as relating to the failure of a system when components are in series
  - "A chain is as strong as its weakest link"

18

Parametric Accelerated Failure Time Models

- The generation of AFT regression models is based on

$$\log T = Y = \vec{X}\,\vec{\beta} + \sigma W,\ W\ some\ error\ distribution$$

- When $e^W \sim F(m_1, m_2)$ an F distribution, then
  - $(m_1, m_2) = (1{,}1)$ $T \sim$ log logistic
  - $(m_1, m_2) = (1, \infty)$ $T \sim$ Weibull (and Exponential if σ=1)
  - $(m_1, m_2) = (m_1, \infty)$ $T \sim$ Generalized Gamma (and gamma if σ=1)
  - $(m_1, m_2) = (\infty, \infty)$ $T \sim$ log normal

- These models have not seen much use
  - Difficulty in getting them to converge with Newton Raphson
  - My view: Distributions that have some scientific validity involve infinite parameters

19

Parametric Inference

- Parametric inference generally proceeds through likelihood methods
  - MLE found by Newton-Raphson iteration

- Asymptotic distributions from theory of regular problems

$$L\left(\vec{\Phi}; \vec{T}, \vec{D}\right) = \prod_{i=1}^{n} \left(f(T_i; \vec{\Phi})(1 - G(T_i))\right)^{D_i} \left(S(T_i; \vec{\Phi}) g(T_i)\right)^{1-D_i}$$

- Under independent censoring

$$L(\vec{\Phi}; \vec{T}, \vec{D}) \propto \prod_{i=1}^{n} \left[\left(f\left(T_i; \vec{\Phi}\right)\right)\right]^{D_i} \left[\left(S\left(T_i; \vec{\Phi}\right)\right)\right]^{1-D_i}$$

20

Parametric vs Distribution-Free

- Choice of summary measures
  - Parametric: Use the natural parameterization of the parametric family
  - Distribution-free: Use a summary measure chosen by clinical issues
- Estimation of arbitrary functionals of distribution
  - Parametric: Use natural parameters of the parametric family
    - For instance, 5 year survival in Weibull: $\exp\left(-(5\hat{\lambda})^{\hat{p}}\right)$
  - Distribution-free: Use estimates derived from Kaplan-Meier ***when they can be estimated***
- Testing for differences between distributions
  - Parametric: Test natural parameters which usually correspond to stochastic ordering (i.e., one curve dominates the other)
  - Distribution-free: Test for differences in the robustly estimated summary measure derived from Kaplan-Meier and its SE

21

21

Parametric Models: Issues

- Advantages
  - Can estimate any of the summary measures
  - Can handle sparse data
  - Can extrapolate beyond support of censoring distribution
- Disadvantages
  - Not robust to other distributions
    - Parametric estimates with censoring do not generally have easy nonparametric interpretation
      - E.g., lognormal model is not particularly robust
  - Little reason to suggest particular distribution
    - But motivation does exist for Weibull and Gamma

22