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 24:

Missing Time to Event Data: Sensitivity Analyses

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

1

Methods of Analyzing Data

An Example

Where am I going?

We consider a simple (simplistic?) approach that can be used to explore sensitivity to MAR assumptions

We have investigated the robustness to semi-parametric assumptions used in the sensitivity analysis

2

2

Example: Basic Approach

- Consider the analysis we would do with complete data
- Derive a (semi)parametric model to impute data under MAR
  - Multiple imputation to obtain inference
- Create MNAR model by couching MAR model in a larger family
  - Additional parameters model the departures from MAR
  - Parameters specific to each treatment group
- By MNAR assumption, there is nothing in the data that can estimate the additional parameters that model MNAR
  - Conduct a series of multiple imputation analyses conditional on assumed values for the additional MNAR parameters
- Find the "tipping point": the values of the MNAR parameters that substantially change inference relative to MAR model
  - Must account for "burden of proof": pivotal RCT, noninferiority, etc
  - Secondarily assess reasonableness of that tipping point

3

3

Example: Time to Event Analysis

- Setting of time to event examined first, because
  - The typical analysis method with noninformative censoring (complete data in a sense) is relatively standard
    - Unadjusted: logrank test
    - Adjusted: proportional hazards regression
  - There are no nuisance parameters
    - (With means of continuous data, we will have to also consider the variability of measurements)
- Mechanisms for missingness
  - Administrative censoring from times of accrual and data analysis
    - MAR that is handled well by KM
  - Potentially informative censoring due to loss of follow-up
    - (Competing risks could be handled providing consistent with the estimand of greatest interest)

4

4

Potential Methods

- Many proposals varying in
  - Analysis models: (semi)parametric vs nonparametric
  - Modeling of missingness: assumptions, predictive markers
  - Goals: estimation, inference
    - Fisher, Kanarek, *Rel and Biometry, Stat Analysis of Lifelength*,1974.
    - Lagakos, Williams, *Biometrika*, 1978.
    - Slud, Rubinstein, *Biometrika*, 1983.
    - Klein, Moeschberger, *Biometrics*, 1988.
    - Robins, Rotnitzky, *Aids Epi Meth*, 1992.
    - Robins, *Proc Biopharm, ASA*, 1993.
    - Zheng, Klein, *Biometrika*, 1995.
    - Scharfstein, Robins, Eddings, Rotnitzky, *Biometrics*, 2001.
    - Scharfstein, Robins, *Biometrika*, 2002.
    - Siannis, Copas, Lu, *Biostatistics*, 2005.
    - Zhang, Heitjan, *Clin Trials*, 2005.
    - Rotnitzky, Farall, Bergesio, Scharfstein, *JRSS Series B*, 2007.
    - Liu, Heitjan, *StatMed*, 2011.

5

5

Example: Logrank Test

- Estimating equation from score function of partial likelihood

$$U(\theta_0) = \sum_t \left(d_{1t} - d_{\bullet t} \frac{n_{1t}}{n_{\bullet t}} \right) = \sum_t \frac{n_{1t} n_{0t}}{n_{1t} + n_{0t}} \left(\hat{\lambda}_{1t} - \hat{\lambda}_{0t} \right)$$

- Under the strong null hypothesis (no treatment effect on any aspect of the distribution), PH holds for the treatment parameter
- Under the weak null hypothesis we are examining some sort of weighted time average of the hazard ratio, and presuming that average HR is 1
  - The weights will depend both on the underlying survival distribution and the censoring distribution
  - But with only administrative censoring, we typically accept that

6

6

Example: Approach

- We use a pattern mixture model to reproduce an analysis that would only have administrative censoring
  - We presume we were happy with interpretation of HR in presence of administrative censoring
- The accrual time and data analysis time is known for all subjects
  - We thus compute an administrative censoring time
- We will ultimately impute the minimum of a survival time and the administrative censoring time

7

7

Example: Pattern Mixture Model

$$\begin{aligned}[Y_{obs}, Y_{mis}, M \mid X] &= [Y_{obs}, Y_{mis} \mid M, X] \times [M \mid X] \\ &= [Y_{mis} \mid Y_{obs}, M, X] \times [Y_{obs} \mid M, X] \times [M \mid X] \\ &\overset{MAR}{=} [Y_{mis} \mid Y_{obs}, X] \times [Y_{obs} \mid M, X] \times [M \mid X]\end{aligned}$$

$[M \mid X]$ distribution of missingness within each treatment arm

$[Y_{obs} \mid M, X]$ estimated by hazard among subjects who are at most administratively censored within each treatment arm

$[Y_{mis} \mid Y_{obs}, X]$ estimated by proportionally increased / decreased hazard after time of potentially informative censoring separately for each treatment arm

8

8

Example: Summary

- Time to event analysis from RCT with
  - Administrative censoring
  - Potentially informative censoring
- Primary analysis: A standard KM or PH analysis (MAR)
  - Assumes imputation of missing data from all subjects still at risk
- Explore sensitivity to change in hazard at time of informative censoring (MNAR)
  - Multiply impute administratively censored data
  - Estimate treatment effect for each hypothesized change in hazard
- Display contour plot of inference as change in hazard varies
  - Consider bias of missing data varies by treatment group
  - HR estimates, CI, p values

9

9

Imputation Probability Model

- Distinguish subjects with
  - Observed event : $(Y_{ki}, \delta_{ki} = 1);\ M_{ki}(t) \equiv 0$
  - Administrative censoring : $(Y_{ki}, \delta_{ki} = 0);\ M_{ki}(t) \equiv 0$
  - Potential informative censoring: $(Y_{ki}, \delta_{ki} = 0);\ M_{ki}(t) = \mathbf{1}_{[t>Y_{ki}]}$
- Presume conditional hazard model for each arm using PH

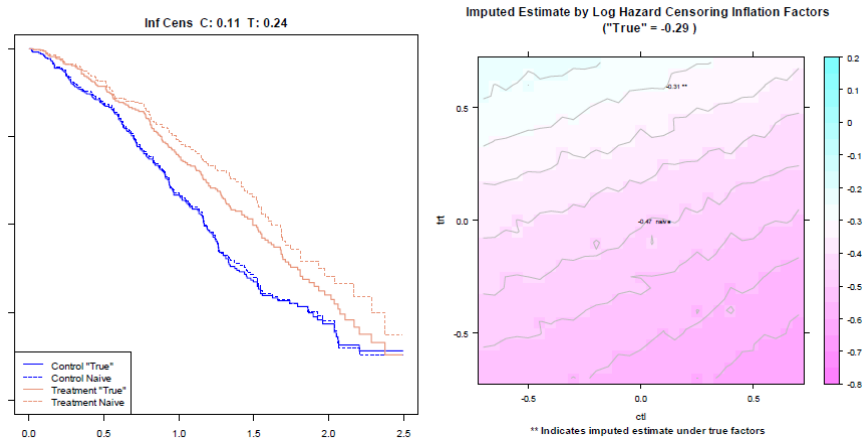
$$\lambda_{ki}\left(t \mid M_{ki}(t)\right) = \lambda_{k0}(t) \times \Delta^{M_{ki}(t)}$$

- Note
  - $\lambda_{k0}(t)$ is estimable using complete data from all subjects
  - Impute administratively censored event times using presumed Δ
  - Untestable use of PH here is motivated by dimension reduction

10

10

Example: Contour Plots



	survival	informative censoring	log hazard inflation factor
control:	Weibull(B=2.0, k=1.5)	Exponential(B=6)	0.1
treatment:	Weibull(B=2.6, k=1.5)	Exponential(B=4)	0.6

11

Interpretation

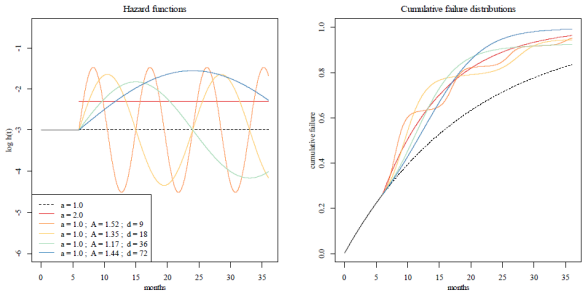
- Primary analysis might be based on MAR: $\Delta = 0$ for both arms
- Tipping point might identify $\Delta's$ when p value or CI bounds meet regulatory burden of proof
  - E.g., for pivotal or noninferiority trials
- Adequacy of tipping point is then judged subjectively by considering for each treatment arm the patterns missingness by
  - Baseline characteristics: age, sex, concomitant disease
  - Putative reasons for dropout: early response, lack of response, AEs, general health status

12

12

Example: Impact of PH Assumption

- This simplistic model presumes all potentially informative censoring shares common constant HR within treatment arms
- Is modeling an average effect adequate?
  - Various more complicated models that have same average
  - Consider hazard functions of varying shape after potentially informative censoring



13

Figure 14. Sinusoidal perturbations equivalent to $\alpha_t = 2.0$

13

Example: Impact of PH Assumption

- Generally reasonable (though slightly low) coverage probability across a wide variety of scenarios

	Estimated Treatment log(HR)								
Scenario	Mean "True"	"True" CI Coverage Rate	Mean "True" CI Width	Mean Naïve	Naïve CI Coverage Rate	Mean Naïve CI Width	Mean Imputed	Imputed CI Coverage Rate	Mean Imputed CI Width
base	-0.272	0.950	0.422	-0.392	0.834	0.480	-0.273	0.930	0.458
a	-0.276	0.961	0.422	-0.393	0.846	0.480	-0.273	0.941	0.458
b	-0.280	0.948	0.423	-0.393	0.849	0.480	-0.272	0.932	0.458
c	-0.280	0.946	0.423	-0.393	0.849	0.480	-0.272	0.932	0.458
d	-0.267	0.954	0.421	-0.392	0.826	0.480	-0.273	0.930	0.458
e	-0.278	0.951	0.423	-0.392	0.845	0.480	-0.273	0.929	0.458

(Reference: 2012 MS Thesis, Eric Meier – www.RCTdesign.org)

14

14

Extension to Other Settings - 1

- Adjusted time to event analyses
  - Using estimated hazards from (possibly stratified) PH regression in imputation relatively straightforward
- Binary outcomes
  - Model treatment arm (and baseline covariate) specific MNAR odds ratios
  - Impact of departures from common OR needs to be explored
    - Mean-variance relationship may have greater impact, though PH regression can be viewed as stratified Mantel-Haenszel, so may generalize

15

15

Extension to Regression Setting

- Adjusted time to event analyses
  - Using estimated hazards from (possibly stratified) PH regression in imputation relatively straightforward
- Using the Cox PH regression model estimates for the complete data, we can estimate the baseline survival curve
  - Covariates and covariate parameter estimates can be used to estimate the survival curve assuming MAR
- We then use the estimated hazard function to impute residual survival under MNAR models, again finding the tipping points
  - In RCT setting, this could be effected by estimating each treatment arm separately for the imputation

16

16

Further Dimension Reduction

- I have parameterized the effect of missingness separately for each treatment arm
- This is only important if there is no (semi)parametric model that is valid across treatment and control groups
  - Under the strong null hypothesis such separate treatment is not necessary, but under alternatives it may be more important
- If there are not large departures from a (semi)parametric model, it is likely sufficient to report contrasts across the MNAR parameters
  - Data analyst can explore the richer parameterization and only report the lower dimension tipping point if relatively constant
  - One dimension for odds or hazards; two dimensions for means

17

17

Final Comments

- Careful design of RCT to minimize missing data is all important
- Protocol should anticipate problems and pre-specify how they will be handled
- Sensitivity analyses should be included to quantify the possible impact of the missing data
  - Frequentist vs Bayesian vs minimax
  - How many researchers have we convinced vs the ”average” researcher
- There is some hope that simple sensitivity analyses are possible
  - But it is not clear that they are ready for prime time, because the intended audience is still highly skeptical

18

18

Bottom Line

"An ounce of prevention is worth a pound of cure"

19

19

Really Bottom Line

"You better think (think) about what you're trying to do…"

-Aretha Franklin, "Think"

20

20