Inference in Randomized Trials with Death and Missingness

Chenguang Wang   Daniel O. Scharfstein   Ying Yan

FDA-Industry Workshop 2014

JOHNS HOPKINS
UNIVERSITY

Outline

1. Motivation
2. Method
3. Software
4. Analysis results
5. Summary

Motivating example

- Randomized, double-blind, placebo-controlled phase III study
- Intent-to-treat population: advanced non-small cell lung cancer subjects
- Functional outcomes scheduled to be measured at baseline, 6 weeks and 12 weeks

Death and missingness

	Arm A	Arm B
	$n = 157$	$n = 322$
Died Prior to Wk 12	15%	17%
Survivors with complete data	59%	57%
Survivors missing only Wk 6	2%	5%
Survivors missing only Wk 12	11%	10%
Survivors missing both Wk 6 and 12	13%	11%

overall: 16% deaths; 30% survivors with missing data

Data truncated by death

Common analysis methods:

- Evaluate treatment effects conditional on survival
- Joint modeling survival and functional outcome
- Evaluate causal treatment effects for principal stratum
- Composite endpoint combining survival and functional outcomes

Goal

To propose a composite outcome approach that handles missing clinical evaluation data among subjects alive at the assessment times.

Outline

General setting

- Consider a two arm randomized study with $T = 0, 1$
- Outcomes $Y_0, \ldots, Y_K$ collected at $t_0, \ldots, t_K$, respectively
- Functional endpoint defined by $Z = f(Y_0, \ldots, Y_K)$
  - example: $Z = Y_K$
  - example: $Z = Y_K - Y_0$
  - motivating study: $Z = (Y_2 + Y_1)/2 - Y_0$
- Survival time denoted by L
- Baseline covariates denoted by X
- Life status at t_K denoted by $\delta = I(L > t_K)$
- Composite endpoint: $C(L, \delta Z)$

Ranking

- Assume that higher values of Z denote better outcomes
- Assume no missing data at this moment
- Consider two subjects i and j with composite endpoint C_i and C_j, respectively
- $C_i > C_j$ (i better than j) only if
  - $\delta_i = \delta_j = 1$ and $Z_i > Z_j$, or
  - $\delta_i = \delta_j = 0$ and $L_i > L_j$, or
  - $\delta_i > \delta_j$
- Ranking may incorporate clinical meaningful differences in Z and L

Hypothesis testing

- Consider observing $C_{i,0}$ from subject i with $T = 0$, $C_{j,1}$ from subject j with $T = 1$
- Parameter of interest: $\theta = P(C_{i,0} > C_{j,1}) - P(C_{i,0} < C_{j,1})$
- $\theta = 0$ if no treatment effect
- Hypothesis: $H_0 : \theta = 0$ vs. $H_0 : \theta \neq 0$

Hypothesis testing

- Estimate θ by

$$\widehat{\theta} = \frac{1}{n_0 n_1} \sum_{i:T_i=0} \sum_{j:T_j=1} \{\mathrm{I}(C_i < C_j) - \mathrm{I}(C_i > C_j)\}$$

- Variance of $\widehat{\theta}$ available in closed form
- Consider the Wald test

Treatment effect size

- θ quantifies treatment effect size
- Recommend to compare quantiles (e.g. median) of $C(L, \delta Z)$ from each arm

Missingness

- For survivors ($\delta = 1$)
  - Denote τ_k to be the missingness indicator of Y_k
  - Denote $S = (\tau_1, \ldots, \tau_K)$ to be the missing pattern
- Intermittent missingness

Assumptions

- Denote $s_c = (\tau_1 = \ldots = \tau_K = 1)$, missing pattern for "completers"
- Let Y_{obs} and Y_{mis} denote the observed and missing outcomes
- Benchmark assumptions
  - CCMV: Complete case missing-variable restrictions
  - For all s,
    $f(Y_{\text{mis}} | Y_{\text{obs}}, X, T, S = s) = f(Y_{\text{mis}} | Y_{\text{obs}}, X, T, S = s_c)$

Step 1: Model completers

- Denote $(Y_1, \ldots, Y_k)$ by $\overline{Y}_k$
- Factorize the joint distribution of $\overline{Y}_K$ as

$$f(\overline{Y}_K | Y_0, X, T, S = s_c) = \prod_{k=1}^{K} f(Y_k | \overline{Y}_{k-1}, Y_0, X, T, S = s_c)$$

- Specify

$$Y_k | \overline{Y}_{k-1}, Y_0, X, T, S = s_c = \alpha_{0,k}^T + \alpha_{1,k}^T \overline{Y}_{k-1} + \alpha_{2,k}^T Y_0 + \alpha_{3,k}^T X + \epsilon.$$

- Allow ϵ to be non-parametrically distributed

Step 2: Impute missing data

- Under normality assumptions, $f(Y_{\text{mis}}|Y_{\text{obs}}, X, T, S = s_c)$ available in closed form
- Under non-parametric distribution assumptions, $f(Y_{\text{mis}}|Y_{\text{obs}}, X, T, S = s_c)$ can be numerically evaluated

Example: $K = 2$

S	τ_1	τ_2
s_1	0	0
s_2	0	1
s_3	1	0
s_4	1	1

Models for completers

$$f(Y_1 | Y_0, X, T, S = s_4)$$
$$f(Y_2 | Y_1, Y_0, X, T, S = s_4)$$

Imputation

$$f(Y_2, Y_1 | Y_0, X, T, S = s_1) = f(Y_2, Y_1 | Y_0, X, T, S = s_4)$$
$$f(Y_1 | Y_2, Y_0, X, T, S = s_2) = f(Y_1 | Y_2, Y_0, X, T, S = s_4)$$
$$f(Y_2 | Y_1, Y_0, X, T, S = s_3) = f(Y_2 | Y_1, Y_0, X, T, S = s_4)$$

Example: $K = 2$

- Consider a subject with Y_0, X, T, and $S = s_2$ (only Y_2 observed)
- Need to impute Y_1 from $f(Y_1|Y_2, Y_0, X, T, S = s_2)$
- By CCMV, $f(Y_1|Y_2, Y_0, X, T, S = s_2) = f(Y_1|Y_2, Y_0, X, T, S = s_4)$
- $f(Y_1|Y_2, Y_0, X, T, S = s_4)$ not available under non-parametric error distribution assumption

Example: $K = 2$

To sample from $f(Y_1|Y_2)$ (omit the condition on $Y_0, X, T, S = s_4$ for compactness), note

$$f(Y_1|Y_2) \propto f(Y_2|Y_1)f(Y_1)$$

- $f(Y_2|Y_1)$ bounded by M obtained by kernel density estimation
- rejection sampling using $f(Y_1)$ as an instrumental distribution

Sensitivity analysis

- Introduce sensitivity parameters Δ in a parsimonious way
- Alternative assumptions: for all s,

$$f(Y_{\text{mis}} | Y_{\text{obs}}, X, T, S = s) \propto \exp\{\Delta Z\} f(Y_{\text{mis}} | Y_{\text{obs}}, X, T, S = s_c)$$

Example

- $K = 2$
- $Z = (Y_2 + Y_1)/2 - Y_0$
- Given Δ, imputation assumption:

$$f(Y_2, Y_1 | Y_0, X, T, S = s_1) \propto e^{\Delta \frac{Y_2}{2}} e^{\Delta \frac{Y_1}{2}} f(Y_2, Y_1 | Y_0, X, T, S = s_4)$$

$$f(Y_1 | Y_2, Y_0, X, T, S = s_2) \propto e^{\Delta \frac{Y_1}{2}} f(Y_1 | Y_2, Y_0, X, T, S = s_4)$$

$$f(Y_2 | Y_1, Y_0, X, T, S = s_3) \propto e^{\Delta \frac{Y_2}{2}} f(Y_2 | Y_1, Y_0, X, T, S = s_4)$$

Exponential tilting model

Consider an exponential tilting model

$$f_{Y'}(y) \propto e^{\Delta y} f_Y(y)$$

- Under normality
  - $Y \sim N(\mu, \sigma^2)$
  - $Y' \sim N(\mu + \Delta\sigma^2, \sigma^2)$
- Under non-parametric assumption
  - $\widehat{f}_Y(y) = \sum_{i=1}^{n} \frac{1}{n} K_h(y - Y_i)$
  - $\widehat{f}_{Y'}(y) = \sum_{i=1}^{n} \frac{e^{\Delta Y_i}}{\sum_{j=1}^{n} e^{\Delta Y_j}} K_h(y - Y_i)$

Outline

Web application

- Currently available at
  http://ebayes.synology.me/shiny/composite/
- Major components
  - upload and review data
  - specify endpoints and imputation model
  - basic graphics
  - specify ranking rule
  - generate imputed dataset
  - bootstrap analysis

Outline

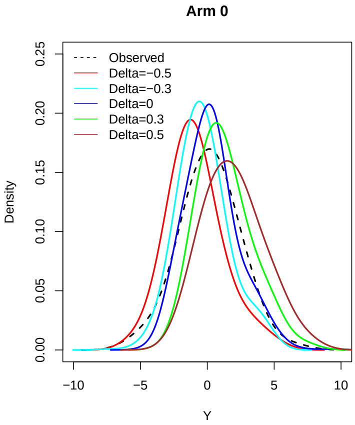
Analysis

- Covariates

Covariates	Levels
ECOG	0:$\{0, 1\}$, 1:$\{2\}$
AGE	0:≤ 65, 1:> 65
SEX	0:M, 1:F
BMI	0:≤ 18.5, 1:> 18.5
WEIGHT LOSS	0:$\leq 10\%$, 1:$> 10\%$
Y0	Continuous

- 500 bootstrap samples, 15 imputed datasets for each bootstrap sample
- Sensitivity parameters $\Delta = \{-0.5, -0.4, \ldots, 0.5\}$

Imputed data



Hypothesis testing

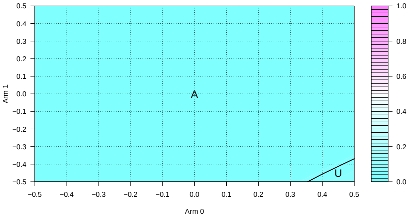
Normality	$\widehat{\theta}(95\%CI)$	p-value
Without Normality	0.28(0.17, 0.38)	< 0.0001
With Normality	0.24(0.13, 0.35)	< 0.0001

Median

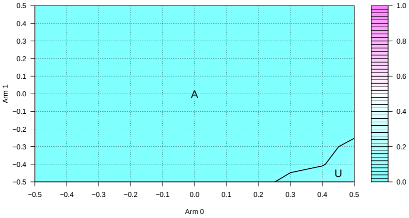
Normality	Arm 0 (95%CI)	Arm 1 (95%CI)
Without Normality	-0.44(-0.88, 0.20)	1.10(0.76, 1.42)
With Normality	-0.49(-1.09, 0.22)	1.03(0.62, 1.36)

Sensitivity analysis

Without normality



With normality



Outline

Summary

- Propose a composite endpoint approach for evaluating treatment effects in randomized clinical trials with death and missingness
- Apply complete case missing-variable restrictions (CCMV) for handling missing data in survivors
- Exponential tilting model for sensitivity analysis
- Online web application developed

THE END