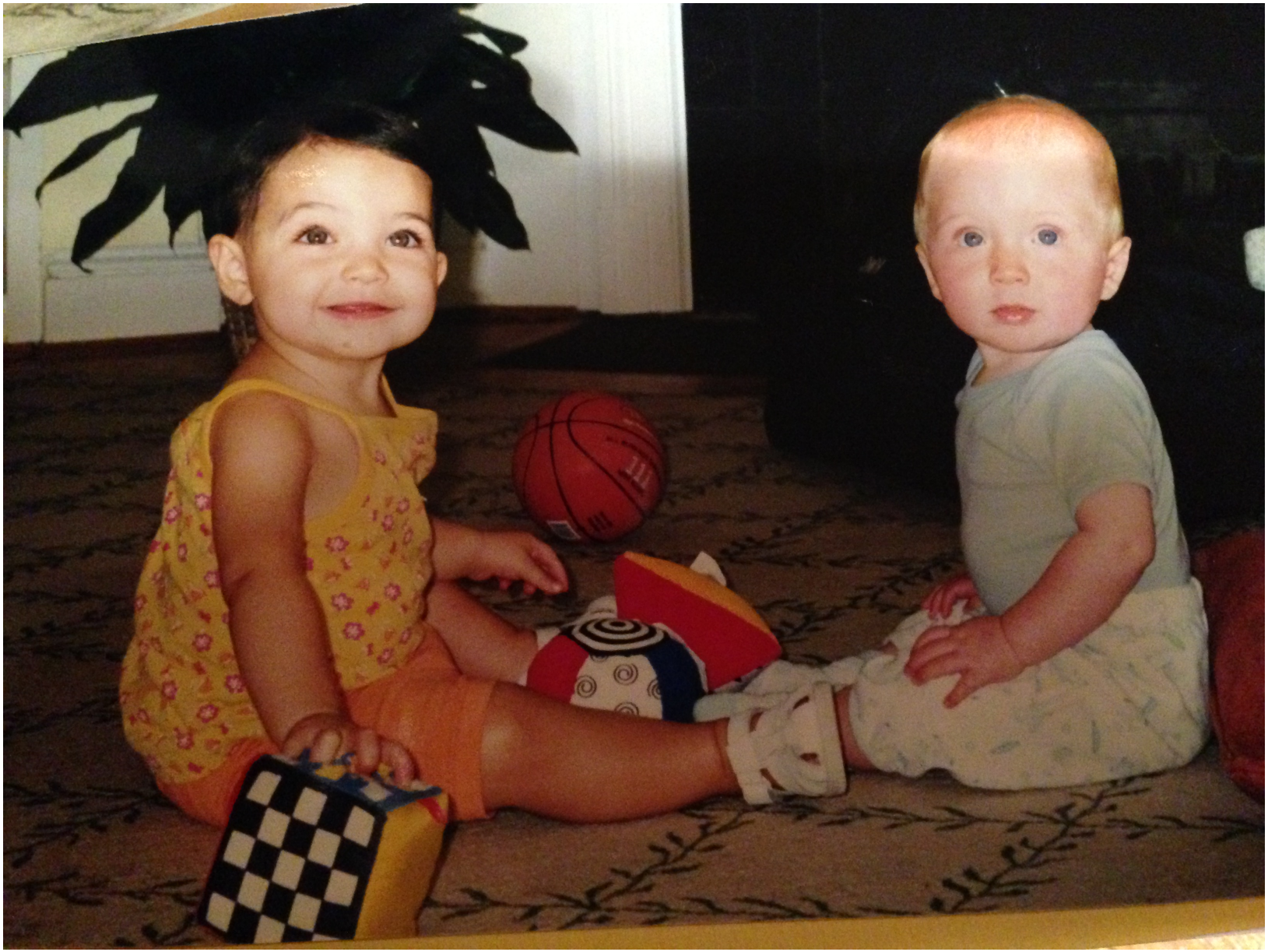
Inference in Randomized Trials with Death and Missingness

Daniel Scharfstein
Johns Hopkins University
dscharf@jhu.edu

April 4, 2016



HT-ANAM 302 Study

- Anamorelin is a drug developed for the treatment of cancer cachexia and anorexia.
- HT-ANAM 302 was a randomized, double-blind, placebo-controlled Phase III study designed to evaluate the efficacy of anamorelin in patients with advanced non-small cell lung cancer.
- Lean body mass (LBM) was scheduled to be measured at baseline (Y_0), 6 weeks (Y_1) and 12 weeks (Y_2)
- Primary functional endpoint: $Z = \frac{(Y_2+Y_1)}{2} - Y_0$

Death and missingness

	Placebo $n = 157$	Anamorelin $n = 322$
Died Prior to Wk 12	24 (15.3%)	54 (16.8%)
Survivors with complete data	93 (59.2%)	185 (57.5%)
Survivors missing only Wk 6	3 (1.9%)	17 (5.3%)
Survivors missing only Wk 12	17 (10.8%)	31 (9.6%)
Survivors missing both Wks 6, 12	20 (12.7%)	35 (10.9%)

Central Question

How should data from studies like HT-ANAM 302 be analyzed to evaluate the effect of treatment on the functional outcome?

Key Issue

- Distinction between missing data and data truncated by death
  - Missing data: exist but not collected
  - Data truncated by death: does not exist and undefined
- Can't just treat as a missing data problem.

Common Approaches

1. Evaluate treatment effect on functional outcome conditional on survival
   - Conditioning on post-baseline factor
2. Joint modeling survival and functional outcomes
   - Allows extrapolation of outcomes after death
3. Principal stratification
   - Applies to a subset of patients who are not identifiable at baseline
4. Composite endpoint combining survival and functional outcomes
   - May be hard to separate effect on function.

Bottom Line

NO PERFECT SOLUTIONS

Not a fan of Approaches 1 and 2.

Goal

To construct a composite endpoint approach that handles both death and missing data

Notation

- $T = 0, 1$: treatment assignment
- X vector baseline covariates
- Y_0: baseline functional measure at t_0
- $Y_1, \ldots, Y_K$: functional outcomes at $t_1, \ldots, t_K$
- L: survival time
- $A_k = I(L > t_k)$: survival status at t_k
- $Z = g(Y_0, \ldots, Y_K)$: primary functional endpoint
  - e.g. $K = 2$, $Z = (Y_2 + Y_1)/2 - Y_0$
  - only defined when $A_K = 1$

Composite Outcome

Finite-valued random variable U which assigns a score to each patient such that

- each patient who dies prior to t_K is assigned a score according to their survival time (L), with shorter survival times assigned lower scores
- each patient who survives past t_K is assigned a score (higher than those who died prior to t_K) according to their functional status (Z), with lower functional status assigned lower scores.

Only the ordering of U is important, not the actual score assignments.

Treatment Effect

Treatment effect (θ) is measured by the probability that the outcome for an individual with $T = 0$ is less than the outcome of an individual with $T = 1$ minus the probability that the outcome for an individual with $T = 0$ is greater than the outcome of an individual with $T = 1$

- $\theta = 0$ under the null
- $\theta > 0$ favors $T = 1$; $\theta < 0$ favors $T = 0$
- First part: Mann-Whitney
- Second part: needed to handle ties

Can also compare the treatment-specific quantiles of U.

Estimation of θ

In the absence of missing data,

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

where $n_0 = \sum_i (1 - T_i)$ and $n_1 = \sum_i T_i$.

Missing Data

- R_k: missing data indicator (defined when $A_k = 1$)
- $S = (R_1, \ldots, R_K)$ (defined when $A_K = 1$)
  - $Y_{obs}^{(s)} = \{Y_k : R_k = 1, k \geq 1, S = s\}$
  - $Y_{mis}^{(s)} = \{Y_k : R_k = 0, k \geq 1, S = s\}$
  - Z is unobserved when $S \neq \mathbf{1}$.

To estimate θ, need to impute Z or equivalently $Y_{mis}^{(s)}$ for $s \neq \mathbf{1}$

Missing Data Assumptions

$$f(Y_{mis}^{(s)} | A_K = 1, Y_{obs}^{(s)}, Y_0, X, T, S = s)$$
$$\propto \exp(\beta_T Z) \underbrace{f(Y_{mis}^{(s)} | A_K = 1, Y_{obs}^{(s)}, Y_0, X, T, S = \mathbf{1})}_{\text{Reference Distribution}}$$

for all $s \neq \mathbf{1}$,

- β_T is a treatment-specific sensitivity parameter.
- $\beta_T = 0$ (i.e., benchmark assumption) reduces to the complete case missing value (CCMV) restrictions applied to the missing data patterns for patients alive at t_K.
- CCMV is different than missing at random (MAR) assumption.

HT-ANAM 302 Study

- $K = 2$, $Z = (Y_1 + Y_2)/2 - Y_0$.
- $\beta'_T = 2\beta_T$

$$f(Y_2|A_2 = 1, Y_1, Y_0, X, T, S = (1,0)) \propto \exp(\beta'_T Y_2) \underbrace{f(Y_2|A_2 = 1, Y_1, Y_0, X, T, S = \mathbf{1})}_{\text{Reference Distribution}}$$

For subjects alive at t_2, who are observed at time t_1, who share the same functional measure at t_1 and who share the same baseline factors, the distribution of Y_2 for those whose functional measure at t_2 is missing is, when $\beta'_T > 0$ (< 0), more heavily weighted toward higher (lower) values of Y_2 than those whose functional measure at t_2 is observed.

HT-ANAM 302 Study

$$f(Y_1|A_2 = 1, Y_2, Y_0, X, T, S = (0,1))$$
$$\propto \exp(\beta_T' Y_1) \underbrace{f(Y_1|A_2 = 1, Y_2, Y_0, X, T, S = \mathbf{1})}_{\text{Reference Distribution}}$$

For subjects alive at t_2, who are observed at time t_2, who share the same functional measure at t_2 and who share the same baseline factors, the distribution of Y_1 for those whose functional measure at t_1 is missing is, when $\beta_T' > 0$ (< 0), more heavily weighted toward higher (lower) values of Y_1 than those whose functional measure at t_1 is observed.

HT-ANAM 302 Study

$$f(Y_1, Y_2 | A_2 = 1, Y_0, X, T, S = (0,0))$$
$$\propto \exp\left(\beta_T'(Y_1 + Y_2)\right) \underbrace{f(Y_1, Y_2 | A_2 = 1, Y_0, X, T, S = \mathbf{1})}_{\text{Reference Distribution}}$$

For subjects alive at t_2 and who share the same baseline factors, the joint distribution of Y_1 and Y_2 for those whose functional measures at t_1 and t_2 are missing is, when $\beta_T' > 0$ (< 0), more heavily weighted toward higher (lower) values of Y_1 and Y_2 than those whose measures are fully observed.

HT-ANAM 302 Study

- Ignore conditioning on Y_0 and X and suppose $f(Y_1, Y_2|A_2 = 1, T, S = \mathbf{1})$ is multivariate normal with mean $(\mu_{T,1}, \mu_{T,2})$ and variance-covariance matrix

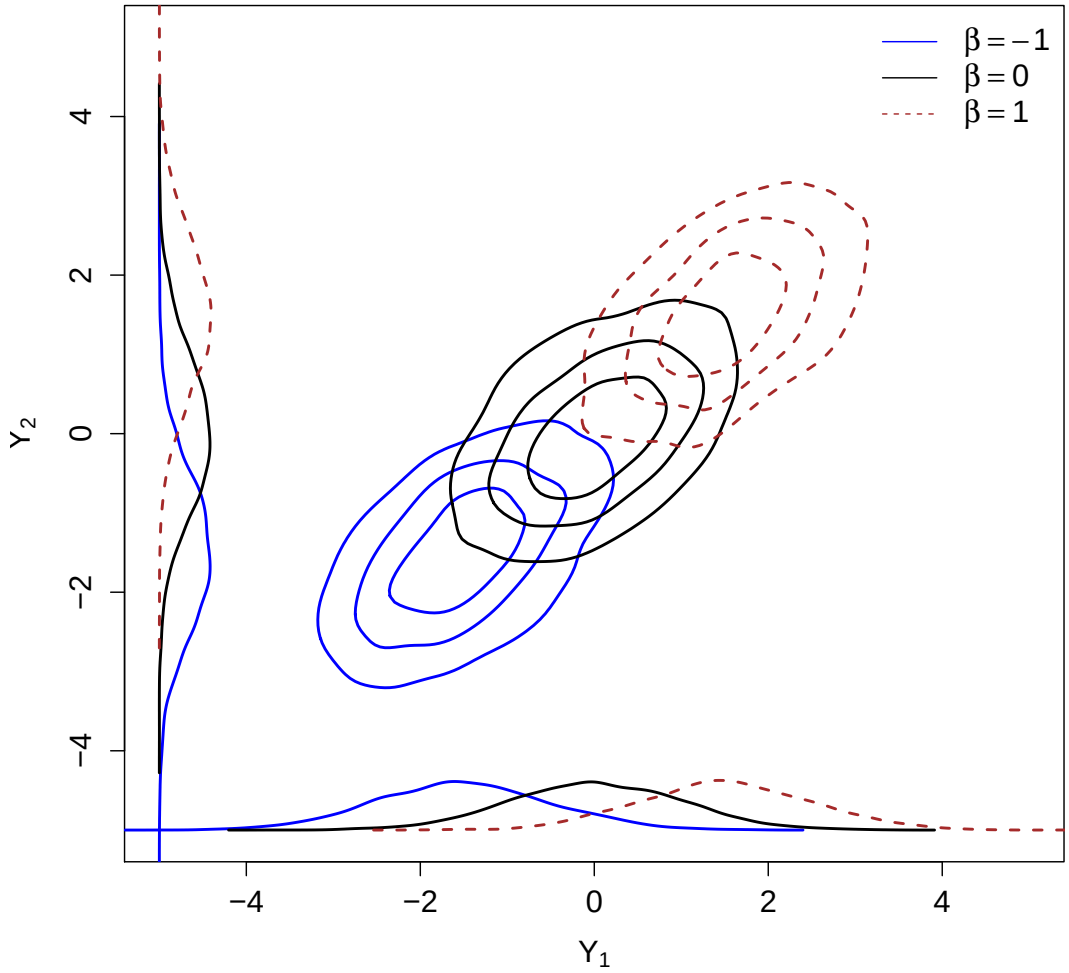
$$\Sigma_T = \begin{bmatrix} \sigma^2_{T,1} & \rho_T \sigma_{T,1} \sigma_{T,2} \\ \rho_T \sigma_{T,1} \sigma_{T,2} & \sigma^2_{T,2} \end{bmatrix}$$

- $f(Y_2|A_2 = 1, Y_1, T, S = (1, 0))$ is normal with mean $\mu_{T,2} + \beta'_T(1 - \rho^2_T)\sigma^2_{T,2} + \rho_T \frac{\sigma_{T,2}}{\sigma_{T,1}}(Y_1 - \mu_{T,1})$ and variance $(1 - \rho^2_T)\sigma^2_{T,2}$
- $f(Y_1|A_2 = 1, Y_2, T, S = (0, 1))$ is normal with mean $\mu_{T,1} + \beta'_T(1 - \rho^2_T)\sigma^2_{T,1} + \rho_T \frac{\sigma_{T,1}}{\sigma_{T,2}}(Y_2 - \mu_{T,2})$ and variance $(1 - \rho^2_T)\sigma^2_{T,1}$

HT-ANAM 302 Study

- $f(Y_1, Y_2|A_2 = 1, T, S = (0,0))$ is multivariate normal with mean $(\mu_{T,1} + \beta'_T\sigma^2_{T,1} + \beta'_T\rho_T\sigma_{T,1}\sigma_{T,2}, \mu_{T,2} + \beta'_T\sigma^2_{T,2} + \beta'_T\rho_T\sigma_{T,1}\sigma_{T,2})$ and variance-covariance matrix Σ_T.
- If $\rho_T > 0$, then the means increase linearly in β'_T
- β'_T has no impact on the variances and covariances.
- $\beta'_T > 0$ ($\beta'_T < 0$) implies that the non-identified distributions have more (less) mass at higher values than their reference distributions.

Example: Exponential tilting



Modeling

Need to specify of a model for

$$f(\overline{Y}_K | A_K = 1, Y_0, X, T, S = \mathbf{1})$$

- To respect bounds, define

$$\phi(y_k) = \log\left\{\frac{y_k - B_L}{B_U - y_k}\right\},$$

- $Y_k^\dagger = \phi(Y_k)$ and $\overline{Y}_k^\dagger = (Y_1^\dagger, \ldots, Y_k^\dagger)$.
- One-to-one mapping between

$$h(\overline{Y}_K^\dagger | A_K = 1, Y_0, X, T, S = \mathbf{1})$$

and

$$f(\overline{Y}_K | A_K = 1, Y_0, X, T, S = \mathbf{1})$$

.

Modeling

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

- Posit a model for each component of the product.

Modeling

$$h(Y_k^\dagger | A_K = 1, \overline{Y}_{k-1}^\dagger, Y_0, X, T = t, S = \mathbf{1})$$
$$= h_{k,t}(Y_k^\dagger - \mu_{k,t}(\overline{Y}_{k-1}^\dagger, Y_0, X; \boldsymbol{\alpha}_{k,t}))$$

- $\mu_{k,t}(\overline{Y}_{k-1}^\dagger, Y_0, X; \boldsymbol{\alpha}_{k,t})$ is a specified function
- $\boldsymbol{\alpha}_{k,t}$ is an unknown parameter vector
- $h_{k,t}$ is an unspecified time/treatment-specific density function.

Estimation

- The parameter vectors $\boldsymbol{\alpha}_{k,t}$ can be estimated by minimizing the least squares objective function

$$\sum_{i=1}^{n} I(T_i = t) A_{K,i} \left(\prod_{k=1}^{K} R_{k,i} \right) \{ Y_{k,i}^{\dagger} - \mu_{k,t}(\overline{Y}_{k-1}^{\dagger}, Y_0, X; \boldsymbol{\alpha}_{k,t}) \}^2$$

- The density function $h_{k,t}$ can be estimated by kernel density estimation based on the residuals $\{ Y_{k,i}^{\dagger} - \mu_{k,t}(\overline{Y}_{k-1,i}^{\dagger}, Y_{0,i}, X_i; \widehat{\boldsymbol{\alpha}}_{k,t}) : T_i = t, A_{K,i} = 1, R_{1,i} = \ldots, R_{K,i} = 1, i = 1, \ldots, n \}$
- $f(\overline{Y}_K | A_K = 1, Y_0, X, T, S = \mathbf{1})$ is estimated by

$$\prod_{k=1}^{K} \widehat{h}_{k,t}(Y_k^{\dagger} - \mu_{k,t}(\overline{Y}_{k-1}^{\dagger}, Y_0, X; \widehat{\boldsymbol{\alpha}}_{k,t})) \left| \frac{d\phi(Y_k)}{dY_k} \right|.$$

Imputation/Estimation

- For each individual i alive at t_K and who is in a stratum $s \neq \mathbf{1}$, impute the missing functional outcomes by drawing (using Metropolis-Hastings algorithm) from the density that is proportional to

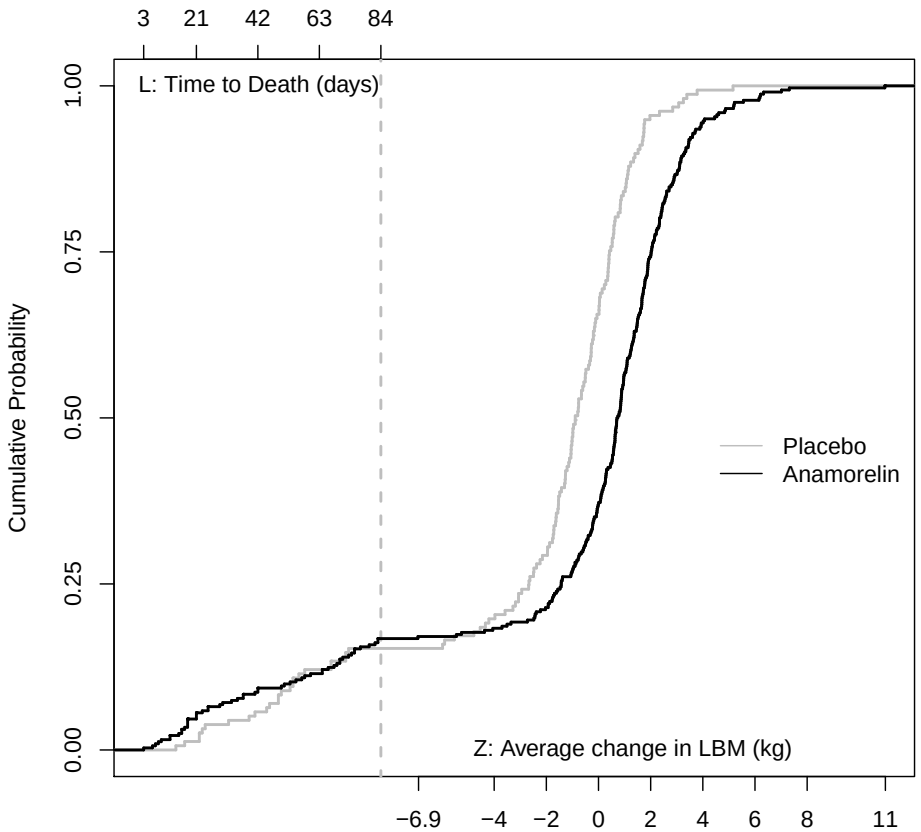
$$\exp(\beta_T Z)\widehat{f}(Y_{mis}^{(s)} | A_K = 1, Y_{obs}^{(s)} = Y_{obs,i}, Y_0 = Y_{0,i}, X = X_i, T = T_i, S = \mathbf{1})$$

- Draw M copies of the missing functional outcomes to create M complete datasets.
- For each complete dataset m, estimate θ by $\widehat{\theta}_m$.
- Overall estimator of θ is $\tilde{\theta} = \frac{1}{M}\sum_{m=1}^{M}\widehat{\theta}_m$.
- Confidence intervals can be constructed by non-parametric bootstrap

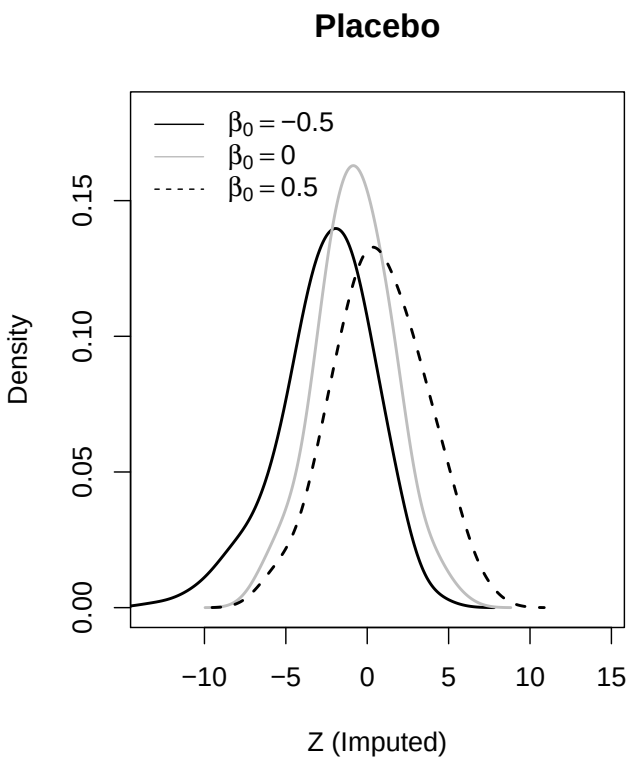
HT-ANAM 302 Study

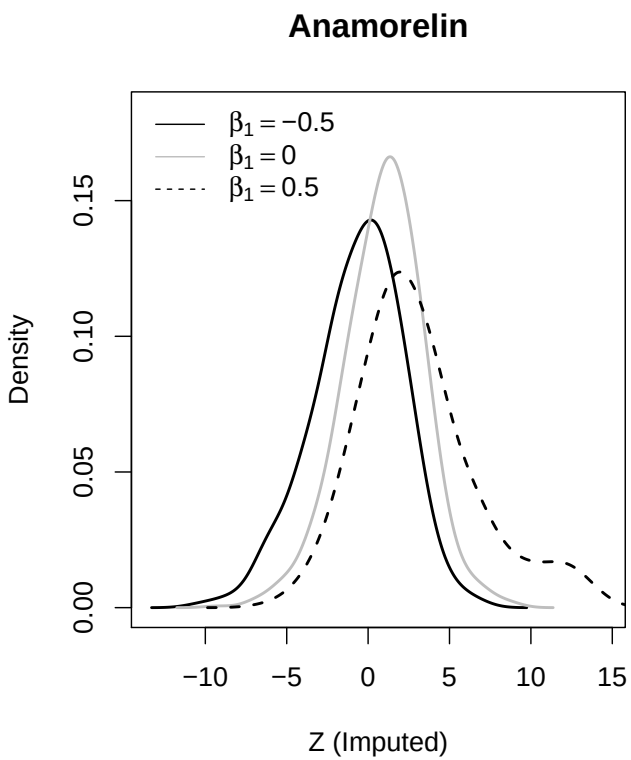
- Baseline covariates: ECOG performance status, age, gender, BMI, weight loss in prior 6 months
- LBM is bounded between 24 and 140
- 10 imputed datasets
- Under benchmark assumptions,
  - $\widehat{\theta} = 0.30$ (95% CI: 0.16 to 0.37, $p < 0.0001$)
  - Placebo: Median -0.98 kg (95% CI: -1.27 kg to -0.28 kg).
  - Anamorelin: Median 0.69 kg (95% CI: 0.43 kg to 0.93 kg).

HT-ANAM 302 Study

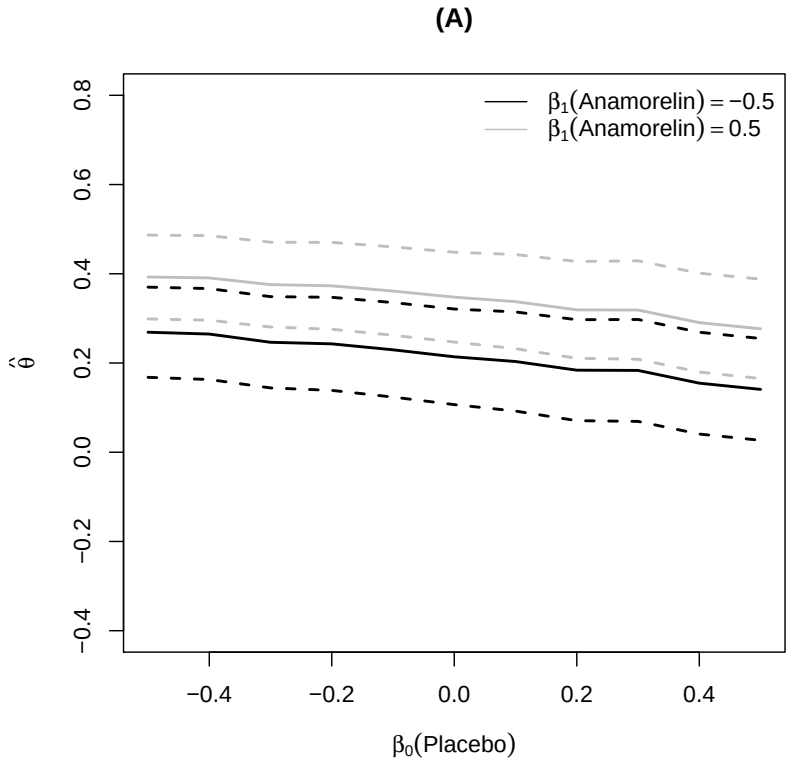


HT-ANAM 302 Study

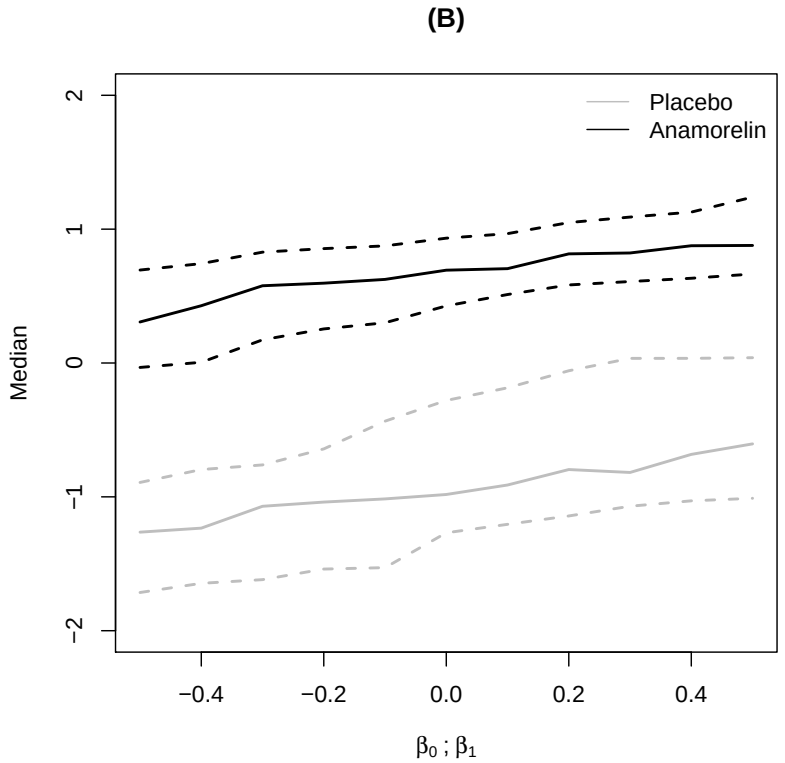




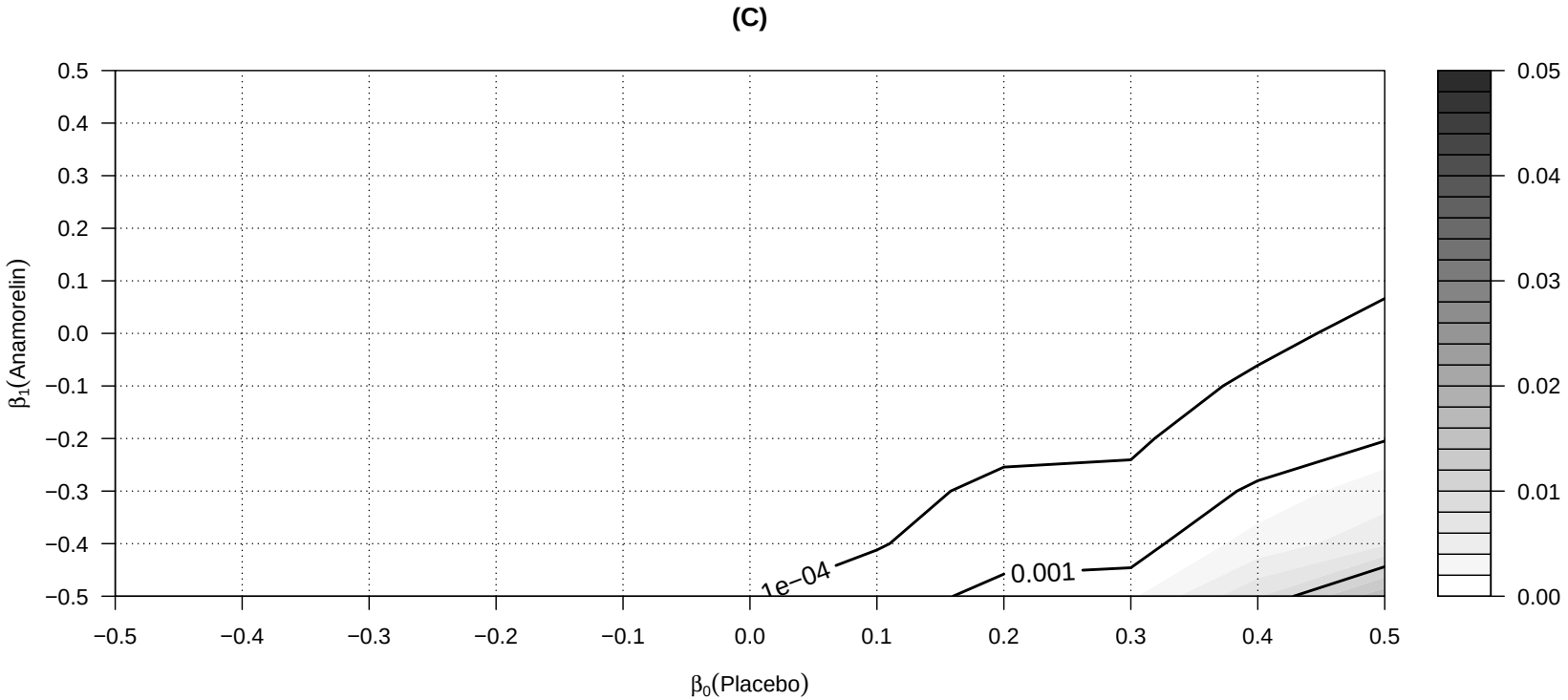
HT-ANAM 302 Study



HT-ANAM 302 Study



HT-ANAM 302 Study



Discussion

- Method presumes that death and the functional outcome can be ordered in a scientifically meaningful way.
- Use mixed methods to confirm that ordering is consistent with the health preferences of patient population.
- Ranking scheme is similar to 'untied worst-rank score analysis" for missing data of Lachin (1999).
- The "worst-rank score analysis" ranks all the patients who died ($A_K = 0$) the same and is also commonly used.
- CCMV is a strong benchmark assumption.
- Assumed survival time is always known, need to extend methods to handle censoring.
- Software is available at `http://sow.familyds.com/shiny/composite/`.