ENAR 2021

Extracting Scalar Measures from Functional Data with Missingness

Lanqiu Yao & Thaddeus Tarpey

NYU School of Medicine

Department of Population Health

Division of Biostatistics

Mar 17, 2021

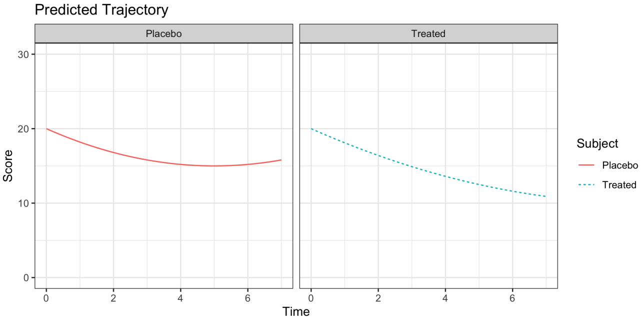
Introduction

- In functional data analysis, we want to get a scalar value to represent the trajectory.
- In Precision medicine, how could we make a treatment decision rule (TDR) when the outcomes are curves.

Introduction

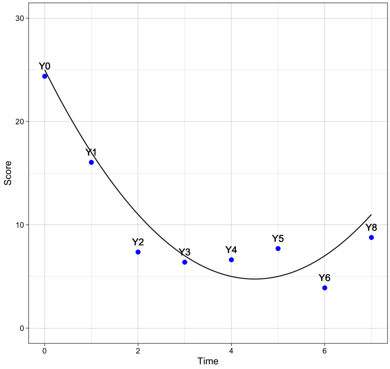
- In functional data analysis, we want to get a scalar value to represent the trajectory.
- In Precision medicine, how could we make a treatment decision rule (TDR) when the outcomes are curves.

- Scalars: A v.s. B
- Curves:



Scalar Outcome from a Functional Trajectory

Figure: Example Trajectory



Longitudinal Outcomes

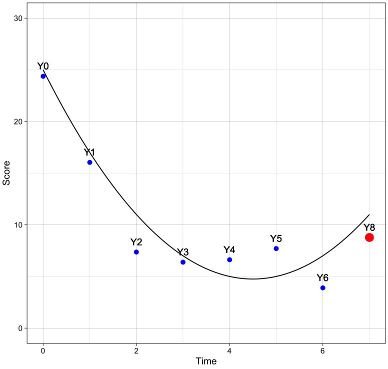
- Y_i: Observed outcomes
- Trajectory: Outcome generation function

Extract a scalar

-
-
-

Scalar Outcome from a Functional Trajectory

Figure: Example Trajectory



Longitudinal Outcomes

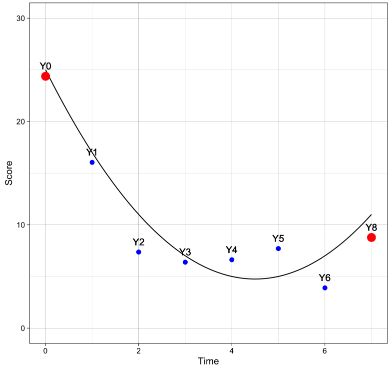
- Y_i: Observed outcomes
- Trajectory: Outcome generation function

Extract a scalar

- Last observed outcome: Y_8
-
-

Scalar Outcome from a Functional Trajectory

Figure: Example Trajectory



Longitudinal Outcomes

- Y_i: Observed outcomes
- Trajectory: Outcome generation function

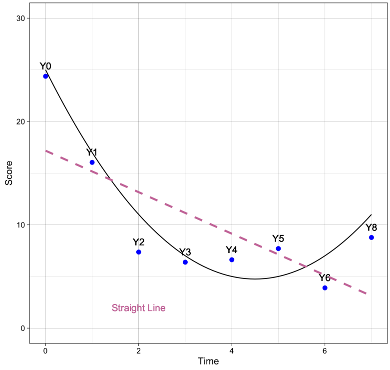
Extract a scalar

- Last observed outcome: Y_8
- Change Score: $Y_8 - Y_1$ or $\frac{Y_8 - Y_1}{t_8 - t_1}$
-

NYU School of Medicine
NYU LANGONE MEDICAL CENTER

Scalar Outcome from a Functional Trajectory

Figure: Example Trajectory



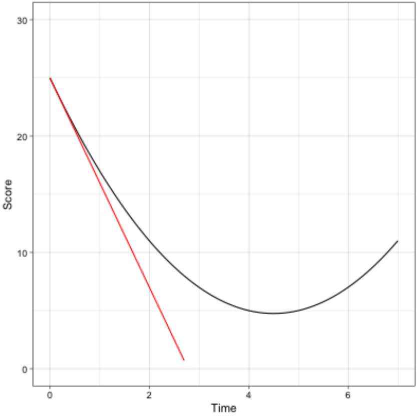
Longitudinal Outcomes

- Y_i: Observed outcomes
- Trajectory: Outcome generation function

Extract a scalar

- Last observed outcome: Y_8
- Change Score: $Y_8 - Y_1$ or $\frac{Y_8 - Y_1}{t_8 - t_1}$
- Slope of straight line

Average Tangent Slope: Average rate of change (average derivative)



Method

Observed Outcome Function

$$\tilde{y}_i(t) = y_i(t) + \epsilon_{it} \tag{1}$$

Method

Observed Outcome Function

$$\tilde{y}_i(t) = y_i(t) + \epsilon_{it} \tag{1}$$

Subject-specific Trajectory Function:

$$y_i(t) = \sum_{u=0}^{p} (\beta_u + \gamma_{iu}) b_u(t) \tag{2}$$

Method

Observed Outcome Function

$$\tilde{y}_i(t) = y_i(t) + \epsilon_{it} \tag{1}$$

Subject-specific Trajectory Function:

$$y_i(t) = \sum_{u=0}^{p} (\beta_u + \gamma_{iu}) b_u(t) \tag{2}$$

Outcome Trajectory Function:

$$y(t) = \sum_{u=0}^{p} \beta_u b_u(t) \tag{3}$$

Method

Observed Outcome Function

$$\tilde{y}_i(t) = y_i(t) + \epsilon_{it} \tag{1}$$

Subject-specific Trajectory Function:

$$y_i(t) = \sum_{u=0}^{p} (\beta_u + \gamma_{iu}) b_u(t) \tag{2}$$

Outcome Trajectory Function:

$$y(t) = \sum_{u=0}^{p} \beta_u b_u(t) \tag{3}$$

Average Tangent Slope (ATS) [1]

$$\frac{1}{b-a} \int_a^b y'(t) dt = \frac{y(b) - y(a)}{b-a} \tag{4}$$

For a functional data, the observed outcome:

$$\tilde{\boldsymbol{Y}}_i = \boldsymbol{X}_i(\boldsymbol{\beta} + \boldsymbol{\gamma}_i) + \boldsymbol{\epsilon}_i \tag{5}$$

where

- $i \in \{1, ..., n\}$, n subjects in total;
- $\boldsymbol{X}_i$ is the design matrix that contains time information.

$$\boldsymbol{X}_i = \begin{pmatrix} b_0(t_1) & \cdots & b_p(t_1) \\ \vdots & \vdots & \vdots \\ b_0(t_{m_i}) & \cdots & b_p(t_{m_i}) \end{pmatrix} \tag{6}$$

- $\tilde{\boldsymbol{Y}}_i$ is a $m_i \times 1$ vector presenting the observed outcome at time $t_1, t_2, ..., t_{m_i}$;
- $\boldsymbol{\beta} = (\beta_0, ..., \beta_p)^\mathsf{T}$ is vector of the fixed effect;
- $\boldsymbol{\gamma}_i$ is the ith subject's random effect and follows a MVN

$$\boldsymbol{b}_i \sim N(\boldsymbol{0}, \boldsymbol{D})$$

- $\boldsymbol{\epsilon}_i$ is the random error follows $\boldsymbol{\epsilon}_i \sim N(0, \sigma^2)$

Example: Quadratic Trajectory

For example, if the outcome trajectory is quadratic in a longitudinal study, then the design matrix is

$$\boldsymbol{X}_i = \begin{pmatrix} 1 & t_1 & t_1^2 \\ \vdots & \vdots & \vdots \\ 1 & t_{m_i} & t_{m_i}^2 \end{pmatrix} \tag{7}$$

and the outcome trajectory function is

$$\tilde{y}_i(t) = (\beta_0 + \gamma_{i0}) + (\beta_1 + \gamma_{i1})t + (\beta_2 + \gamma_{i2})t^2 + \epsilon_{ij} \tag{8}$$

The ATS:

$$\frac{1}{t_m - t_1}\int_{t_1}^{t_m} y'(t)dt = \frac{y(t_m) - y(t_1)}{t_m - t_1} = \beta_1 + \beta_2(t_1 + t_m) \tag{9}$$

ATS estimators

If the data is balanced and no missing values:

ATS estimators

If the data is balanced and no missing values:

- Crude Estimator: Taking the difference in the observed first and last outcomes divided by the difference in the corresponding times and average these over all participants:

$$\widehat{\text{ATS}}^{\text{Crude}} = \frac{1}{n}\sum_{i=1}^{n}\frac{\tilde{y}_i(t_m) - \tilde{y}_i(t_1)}{t_m - t_1} \quad (10)$$

ATS estimators

If the data is balanced and no missing values:

- Crude Estimator: Taking the difference in the observed first and last outcomes divided by the difference in the corresponding times and average these over all participants:

$$\widehat{\text{ATS}}^{\text{Crude}} = \frac{1}{n}\sum_{i=1}^{n}\frac{\tilde{y}_i(t_m)-\tilde{y}_i(t_1)}{t_m-t_1} \tag{10}$$

- Linear Estimator: Taking the slope by fitting a linear mixed effects model

$$\widehat{\text{ATS}}^{\text{Linear}} = \hat{\alpha} \ \text{(the estimated slope of linear term)} \tag{11}$$

ATS estimators

If the data is balanced and no missing values:

- Crude Estimator: Taking the difference in the observed first and last outcomes divided by the difference in the corresponding times and average these over all participants:
$$\widehat{\text{ATS}}^{\text{Crude}} = \frac{1}{n}\sum_{i=1}^{n} \frac{\tilde{y}_i(t_m) - \tilde{y}_i(t_1)}{t_m - t_1} \tag{10}$$

- Linear Estimator: Taking the slope by fitting a linear mixed effects model
$$\widehat{\text{ATS}}^{\text{Linear}} = \hat{\alpha} \ \text{ (the estimated slope of linear term)} \tag{11}$$

- Formula Estimator: Pluging in the estimated $\hat{\beta}$ in
$$\begin{aligned}\widehat{\text{ATS}}^{\text{Formula}} =& \frac{\hat{y}(t_m) - \hat{y}(t_1)}{t_m - t_1} \\ =& \hat{\beta}_1 + \hat{\beta}_2(t_1 + t_m) \ \text{ (if quadratic)}\end{aligned} \tag{12}$$

NYU School of Medicine
NYU LANGONE MEDICAL CENTER

Properties of ATS estimators

Expectations:

- Crude Estimator:

$$E(\widehat{\text{ATS}}^{\text{Crude}}) = E(\frac{1}{n}\sum_{i=1}^{n}\frac{\tilde{y}_i(t_m) - \tilde{y}_i(t_1)}{t_m - t_1}) = \frac{y(t_m) - y(t_1)}{t_m - t_1} \tag{13}$$

- Linear Estimator:

$$\begin{aligned} E(\widehat{\text{ATS}}^{\text{Linear}}) =& E((\boldsymbol{X}_*^{\mathsf{T}}\boldsymbol{X}_*)^{-1}\boldsymbol{X}_*^{\mathsf{T}}\tilde{\boldsymbol{y}})_{[2,1]} = [(\boldsymbol{X}_*^{\mathsf{T}}\boldsymbol{X}_*)^{-1}\boldsymbol{X}_*^{\mathsf{T}}\boldsymbol{X}\boldsymbol{\beta}]_{[2,1]} \\ =& \beta_1 + \beta_2\left[\frac{\sum(t_j - \bar{t})^3}{\sum(t_j - \bar{t})^2} + 2\bar{t}\right] \quad \text{(if linear or quadratic)} \end{aligned} \tag{14}$$

- Formula Estimator:

$$E(\widehat{\text{ATS}}^{\text{Formula}}) = E(\frac{\widehat{y}(t_m) - \widehat{y}(t_1)}{t_m - t_1}) = \frac{y(t_m) - y(t_1)}{t_m - t_1} \tag{15}$$

NYU School of Medicine
NYU LANGONE MEDICAL CENTER

Properties of ATS estimators

Variances:

- Crude Estimator:

$$\mathrm{Var}(\widehat{\mathrm{ATS}}^{\mathrm{Crude}}) = \frac{1}{n}\boldsymbol{h}^{\mathsf{T}}\left(\sigma^2\boldsymbol{I} + \boldsymbol{X}_i\boldsymbol{D}\boldsymbol{X}_i^{\mathsf{T}}\right)\boldsymbol{h} \tag{16}$$

where $\boldsymbol{h} = (\frac{1}{t_m - t_1}, 0, ..., 0, \frac{1}{t_m - t_1})^{\mathsf{T}}$

.

- Formula Estimator:

$$\mathrm{Var}(\widehat{\mathrm{ATS}}^{\mathrm{Formula}}) = \frac{1}{n}\boldsymbol{g}^{\mathsf{T}}\left(\sigma^2(\boldsymbol{X}_i^{\mathsf{T}}\boldsymbol{X}_i)^{-1} + \boldsymbol{D}\right)\boldsymbol{g} \tag{17}$$

where $\boldsymbol{g} = \frac{\boldsymbol{b}(t_m) - \boldsymbol{b}(t_1)}{t_m - t_1} = (0, 1, t_1 + t_m)^{\mathsf{T}}$ (if quadratic)

Estimation of ATS with Missingness

Missingness Mechanism

- Missing Completely at Random (MCAR): The missingness is a completely random process so that the probability of a missing outcome is independent of any covariates and outcomes;
- Missing at Random (MAR): If the cause of the missingness depends on some observed variable or variables that have been collected, the missing data mechanism is assumed to be MAR.
- Missing not at Random (MNAR): The missingness is dependent on some unobserved covariate/outcome values.

Estimation of ATS with Missingness

Missingness Handling Approaches

- Complete cases analysis (CCA): deletes all participants with missing values
- Weighting: adjust for unequal sampling fractions
- Single imputation: mean, median, last observation carried forward (LOCF)
- Multiple imputation (MI)

Formula Estimator with Missingness

Let S denote the time at the last observation. Suppose $S = t_j$ with probability p_j ($\sum_{j=2}^{m} p_j = 1$).

The formula estimator:

$$\widehat{\text{ATS}}^{\text{Formula}} = \frac{\widehat{y}(t_m) - \widehat{y}(t_1)}{t_m - t_1} = \hat{\beta}_1 + \hat{\beta}_2(t_1 + t_m) \ \ \text{(if quadratic)} \tag{18}$$

The $\widehat{y}(t_1), \widehat{y}(t_m)$ or more specifical, the $\hat{\beta}_1$,
hatβ_2 are estimated from mixed effect model, which is ”robust” to missing data.

Crude Estimator with Missingness

For the crude estimator:

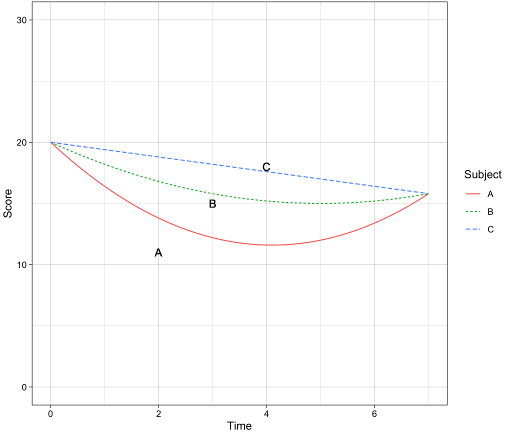
$$\begin{aligned} E(\widehat{\text{ATS}}_i^{\text{Crude}}) =& E\big(E(\widehat{\text{ATS}}_i^{\text{Crude}}|S)\big) \\ =& \sum_{j=2}^{m} p_j E(\widehat{\text{ATS}}_i^{\text{Crude}}|S = t_j) \\ =& \beta_1 + \beta_2 t_0 + \beta_2 \sum_{j=2}^{m} p_j t_j \quad \text{(if quadratic)} \\ \neq& \beta_1 + \beta_2 t_0 + \beta_2 t_m \end{aligned} \tag{19}$$

which is biased.

ATS and ATS with Missingness

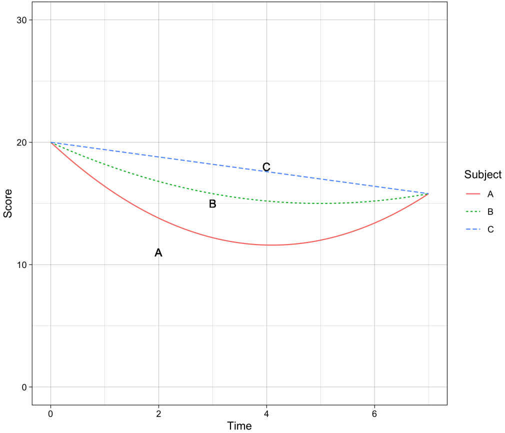
- When there is no missing data, the Crude Estimator (Change Score) is an unbiased estimator for the ATS.
- The Linear Estimator is only unbiased when the trajectory is quadratic
- The Formula Estimator is also unbiased, and has more power than the Crude Estimator
- When there is missing data (MCAR, MAR), the Crude Estimator is biased while the Formula Estimator still has good performance.
- Missing data handling approaches could help

Weighted Average Tangent Slope



NYU School of Medicine
NYU LANGONE MEDICAL CENTER

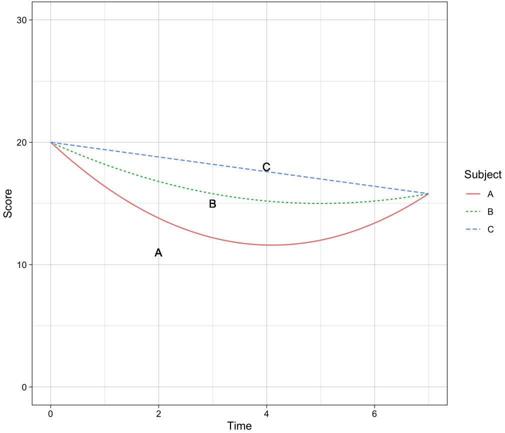
Weighted Average Tangent Slope



ATS

$$\frac{1}{b-a}\int_a^b y'(t)dt = \frac{y(b)-y(a)}{b-a} \quad (20)$$

Weighted Average Tangent Slope



ATS

$$\frac{1}{b-a}\int_a^b y'(t)dt = \frac{y(b)-y(a)}{b-a} \quad (20)$$

WATS

$$\int_a^b w(t)y'(t)dt \quad (21)$$

where $\int_a^b w(t)dt = 1$.

WATS: Weight Function Selection

Could be arbitrary: $w(t) = c\exp(y'_{\text{placebo}}(t))$

WATS: Weight Function Selection

Could be arbitrary: $w(t) = c\exp(y'_{\text{placebo}}(t))$

Optimize the classification:

The mean outcome trajectory is

$$y_k(t) = \boldsymbol{\beta}_k^{\mathsf{T}} \boldsymbol{b}(t) \tag{22}$$

Then the WATS can be expressed as

$$\begin{aligned} \textbf{WATS}_k &= \int_T w(t) y'_k(t) dt \\ &= \int_T w(t) [\boldsymbol{b}'(t)]^{\mathsf{T}} \boldsymbol{\beta}_k dt \\ &= [E_w(\boldsymbol{b}')]^{\mathsf{T}} \boldsymbol{\beta}_k \end{aligned} \tag{23}$$

The estimated WATS is

$$\widehat{\text{WATS}}_k = [E_w(\boldsymbol{b}')]^{\mathsf{T}} \widehat{\boldsymbol{\beta}}_k \tag{24}$$

where $[E_w(\boldsymbol{b}')]^{\mathsf{T}} = \int_T w(t) [\boldsymbol{b}'(t)]^{\mathsf{T}} dt = (E_w(b'_1), E_w(b'_2), ..., E_w(b'_p))^{\mathsf{T}}$.

Optimization:

$$\widehat{w(t)} = \underset{w(t)}{\operatorname{argmax}} \frac{\left[\widehat{\text{WATS}}_1 - \widehat{\text{WATS}}_2\right]^2}{\text{Var}(\widehat{\text{WATS}}_1) + \text{Var}(\widehat{\text{WATS}}_2)} = \frac{[E_w(\boldsymbol{b}')]^\mathsf{T}(\widehat{\boldsymbol{\beta}}_1 - \widehat{\boldsymbol{\beta}}_2)(\widehat{\boldsymbol{\beta}}_1 - \widehat{\boldsymbol{\beta}}_2)^\mathsf{T} E_w(\boldsymbol{b}')}{[E_w(\boldsymbol{b}')]^\mathsf{T}\left(\text{Var}(\widehat{\boldsymbol{\beta}}_1) + \text{Var}(\widehat{\boldsymbol{\beta}}_2)\right)[E_w(\boldsymbol{b}')]} \tag{25}$$

with $\int_T w(t)dt = 1$.

We just want to find the vector of expectation of basis functions, $[E_w(\boldsymbol{b}')]^\mathsf{T} = (E_w(b_1), E_w(b_2), ..., E_w(b_p))^\mathsf{T}$, that optimizes Eq(25) and ensures each expectation is non negative.

$$\widehat{w(t)} = \underset{w(t)}{\text{argmax}} \frac{[E_w(\boldsymbol{b}')]^{\mathsf{T}} (\widehat{\boldsymbol{\beta}}_1 - \widehat{\boldsymbol{\beta}}_2)(\widehat{\boldsymbol{\beta}}_1 - \widehat{\boldsymbol{\beta}}_2)^{\mathsf{T}} E_w(\boldsymbol{b}')}{[E_w(\boldsymbol{b}')]^{\mathsf{T}} \left(\text{Var}(\widehat{\boldsymbol{\beta}}_1) + \text{Var}(\widehat{\boldsymbol{\beta}}_2)\right) [E_w(\boldsymbol{b}')]}$$

Optimization Process:

- 1. Choose the basis functions $\boldsymbol{b}(t)$ (any basis functions should be able to get the same result)
- 2. Fit the mixed effect model and get estimations of $\widehat{\boldsymbol{\beta}}_k$ and $\hat{\boldsymbol{\gamma}}_{ki}$ ($k \in \{1, 2\}$)

$$\text{Subject trajectory: } \; y_{ki}(t) = \boldsymbol{b}^{\mathsf{T}}(t)(\boldsymbol{\beta}_k + \boldsymbol{\gamma}_{ki})$$

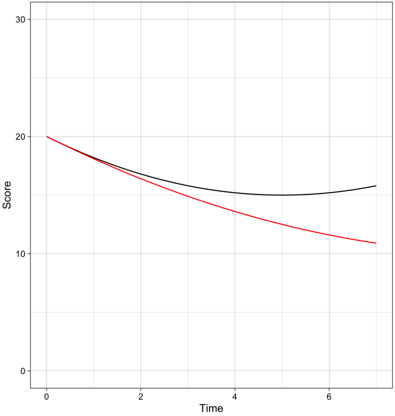
- 3. Estimate the variance of $\widehat{\boldsymbol{\beta}}_k$ with $(\sum_{i=1}^{n} \boldsymbol{b}^{\mathsf{T}} \widehat{\boldsymbol{V}}_{ki}^{-1} \boldsymbol{b})^{-1}$, and $\widehat{\boldsymbol{V}}_{ki} = \boldsymbol{b}\widehat{\boldsymbol{D}}\boldsymbol{b}^{\mathsf{T}} + \hat{\sigma}^2 \boldsymbol{I}$, where $\widehat{\boldsymbol{D}}$ is the estimated covariance matrix.
- 4. Plug in $\widehat{\boldsymbol{\beta}}_k$ and Var($\widehat{\boldsymbol{\beta}}_k$) in the Eq(25). Apply optimization algorithms (e.g. Nelder and Mead) to find the $\boldsymbol{G}^* = \boldsymbol{E}_w(\boldsymbol{u}')$ that maximizes the Eq(25).
- 5. Plug the $\boldsymbol{G}^*$ back in Eq(23) and get the estimated $\widehat{\text{WATS}}_k = \boldsymbol{G}^{*\mathsf{T}} \widehat{\boldsymbol{\beta}}_k$

$$\widehat{\text{WATS}}_k = [E_w(\boldsymbol{b}')]^{\mathsf{T}} \widehat{\boldsymbol{\beta}}_k$$

Simulation Setting

Simulation Setting

Quadratic Cruves



Trajectory Functions:

- Placebo Group:

$$y_{\text{pbo}}(t) = 20 - 2t + 0.2t^2$$

- Treatment Group:

$$y_{\text{pbo}}(t) = 20 - 2t + 0.1t^2$$

Simulation Setting

- Two groups, each group contains $n = 100$ subjects.
- $\boldsymbol{t} = (0, 1, ..., 7)^{\mathsf{T}}$

$$\boldsymbol{X}_i = \begin{pmatrix} 1 & t_1 & t_1^2 \\ \vdots & \vdots & \vdots \\ 1 & t_{m_i} & t_{m_i}^2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 7 & 7^2 \end{pmatrix}$$

- Random effect $\boldsymbol{\gamma}_{1,i} \sim N(\mathbf{0}, \boldsymbol{D}_1)$, $\boldsymbol{\gamma}_{2,i} \sim N(\mathbf{0}, \boldsymbol{D}_2)$, where

$$\boldsymbol{D}_1 = \boldsymbol{D}_2 = \begin{pmatrix} 8 & 3 & -0.4 \\ 3 & 1.5 & -0.16 \\ -0.4 & -0.16 & 0.03 \end{pmatrix}$$

- Random error $\boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_2 \sim N(0, \sigma^2)$, and $\sigma \in \{1, 2, 10\}$.

NYU School of Medicine
NYU LANGONE MEDICAL CENTER

Simulation Setting

Missingness (30 %):

- MCAR
- Drop-off: 50% no missing; 30% missing only the last observation; 10% missing the last 2 observations; 5% missing the last 3 observations, and 5% missing the last 4 observations.

Method:

ATS estimation:

- Crude estimator, Linear estimator, Formula estimator

Power analysis:

- Crude estimator, Linear estimator, Formula estimator, Weighted average tangent slope (WATS), ANOVA (adjusted baseline)

Simulation Results

Quadratic Cruves

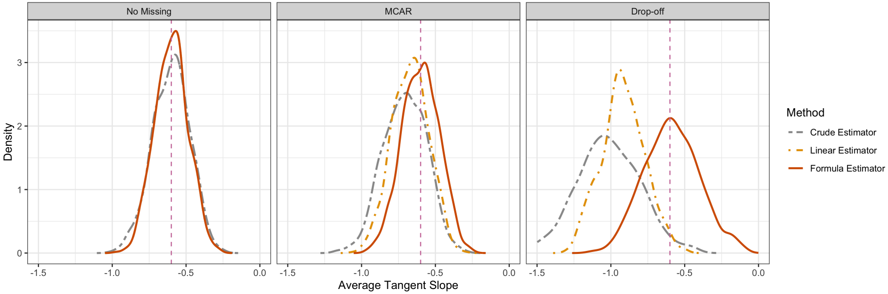


Figure: Densities for the ATS estimation methods

Simulation Results

Quadratic Cruves

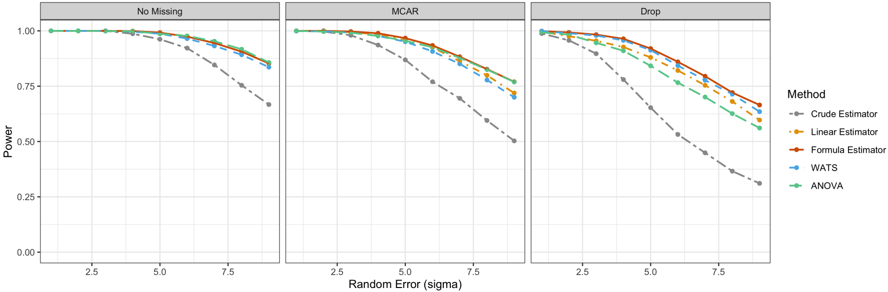
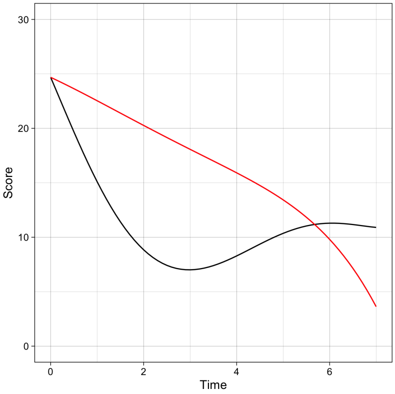


Figure: Power Comparison

Complex Case

Non-Quadratic Curves



Missingness:

- MCAR: 30%
- Drop off: 50% Drop off at time 6

ATS estimation:

- Crude estimator, Linear estimator, Spline

Power analysis:

- Crude estimator, Linear estimator, Spline, Weighted average tangent slope (WATS), ANOVA (adjusted baseline)

Simulation Results

Non-Quadratic Cruves

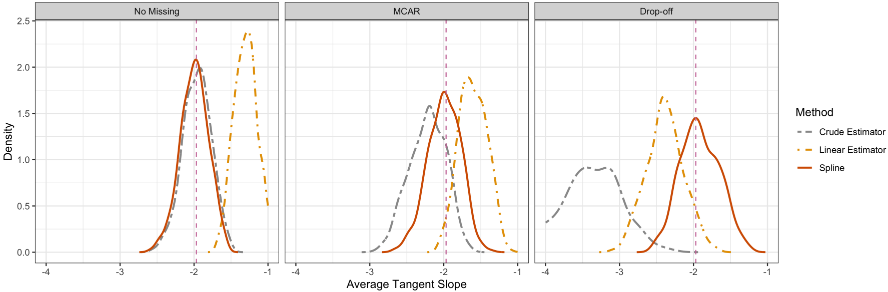


Figure: Power Comparison

Simulation Results

Non-Quadratic Cruves

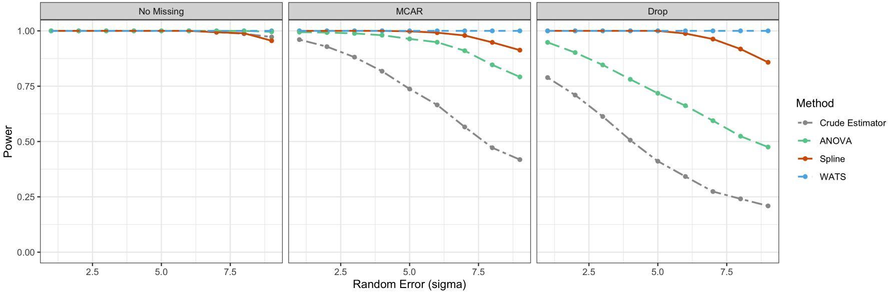


Figure: Power Comparison

Discussion

- Average tangent slope can provide a meaningful scalar summary of a functional trajectory.
- The weighted average tangent slope allows additional flexibility in extracting a scalar summary statistic.
- Both methods have outstanding performances than the other scalar measures and are robust to missing values (MCAR, Drop-off)
- In precision medicine, the ATS and WATS can help the derivation of the optimal treatment decision.

Thank you!

Q & A

Reference

Thaddeus Tarpey et al. "Extracting scalar measures from functional data with applications to placebo response". In: *Statistics and Its Interface* 14.3 (2021), pp. 255–265.