Discovering Linear Biosignatures for Treatment Response Based on Maximizing Kullback-Leibler Divergence in Linear mixed-effect Models

Lanqiu Yao & Thaddeus Tarpey

NYU School of Medicine

Department of Population Health

Division of Biostatistics

October 25, 2019

Outline:

NYU School of Medicine
NYU LANGONE MEDICAL CENTER

Introduction: Background

The World Health Organization has predicted that by 2020, depression will be the second-leading cause of disease burden globally.

- 18.57 % of adults are experiencing a mental health illness, equivalent to 45 million Americans.

NYU School of Medicine
NYU LANGONE MEDICAL CENTER

Introduction: Background

The World Health Organization has predicted that by 2020, depression will be the second-leading cause of disease burden globally.

- 18.57 % of adults are experiencing a mental health illness, equivalent to 45 million Americans.
- 4.38 % are experiencing a severe mental health illness.

NYU School of Medicine
NYU LANGONE MEDICAL CENTER

Introduction: Background

The World Health Organization has predicted that by 2020, depression will be the second-leading cause of disease burden globally.

- 18.57 % of adults are experiencing a mental health illness, equivalent to 45 million Americans.
- 4.38 % are experiencing a severe mental health illness.
- The state prevalence of adult mental illness ranges from 16.19% in New Jersey to 25.03 % in Idaho.

have serious thoughts of suicide in the U.S.

NYU School of Medicine
NYU LANGONE MEDICAL CENTER

Introduction: Challenges

- Challenges in mental health research
  - Diagnosis

Introduction: Challenges

- You feel sad or irritable most of the day, nearly every day.
- You are less interested in most activities you once enjoyed.
- You suddenly lose or gain weight or have a change in appetite.
- You have trouble falling asleep or want to sleep more than usual.
- You experience feelings of restlessness.
- You feel unusually tired and have a lack of energy.
- You feel worthless or guilty, often about things that wouldn't normally make you feel that way.
- You have difficulty concentrating, thinking, or making decisions.
- You think about harming yourself or committing suicide.

To be diagnosed with major depressive disorder (MDD), you must have 5 or more of the symptoms.

NYU School of Medicine
NYU LANGONE MEDICAL CENTER

Introduction: Challenges

- You feel sad or irritable most of the day, nearly every day.
- You are less interested in most activities you once enjoyed.
- You suddenly lose or gain weight or have a change in appetite.
- You have trouble falling asleep or want to sleep more than usual.
- You experience feelings of restlessness.
- You feel unusually tired and have a lack of energy.
- You feel worthless or guilty, often about things that wouldn't normally make you feel that way.
- You have difficulty concentrating, thinking, or making decisions.
- You think about harming yourself or committing suicide.

To be diagnosed with major depressive disorder (MDD), you must have 5 or more of the symptoms.

$$\binom{9}{5} + \binom{9}{6} + \binom{9}{7} + \binom{9}{8} + \binom{9}{9} = 247$$

NYU School of Medicine
NYU LANGONE MEDICAL CENTER

Introduction: Challenges

- Challenge in mental health research
  - Diagnosis

Introduction: Challenges

- Challenge in mental health research
  - Diagnosis
  - Treatment

Introduction: Challenges

- Challenge in mental health research
  - Diagnosis
  - Treatment
  - Classification

Introduction: Motivation Data

The EMBARC Study

Establishing moderators and biosignatures of antidepressant response in clinical care (EMBARC)

- A multi-site, placebo-controlled randomized clinical trial

NYU School of Medicine
NYU LANGONE MEDICAL CENTER

Introduction: Motivation Data

The EMBARC Study

Establishing moderators and biosignatures of antidepressant response in clinical care (EMBARC)

- A multi-site, placebo-controlled randomized clinical trial
- Participants with early onset ($\leq$30 years) MDD are included

NYU School of Medicine
NYU LANGONE MEDICAL CENTER

Introduction: Motivation Data

The EMBARC Study

Establishing moderators and biosignatures of antidepressant response in clinical care (EMBARC)

- A multi-site, placebo-controlled randomized clinical trial
- Participants with early onset ($\leq$30 years) MDD are included
- 8-week trial

NYU School of Medicine
NYU LANGONE MEDICAL CENTER

Introduction: Motivation Data

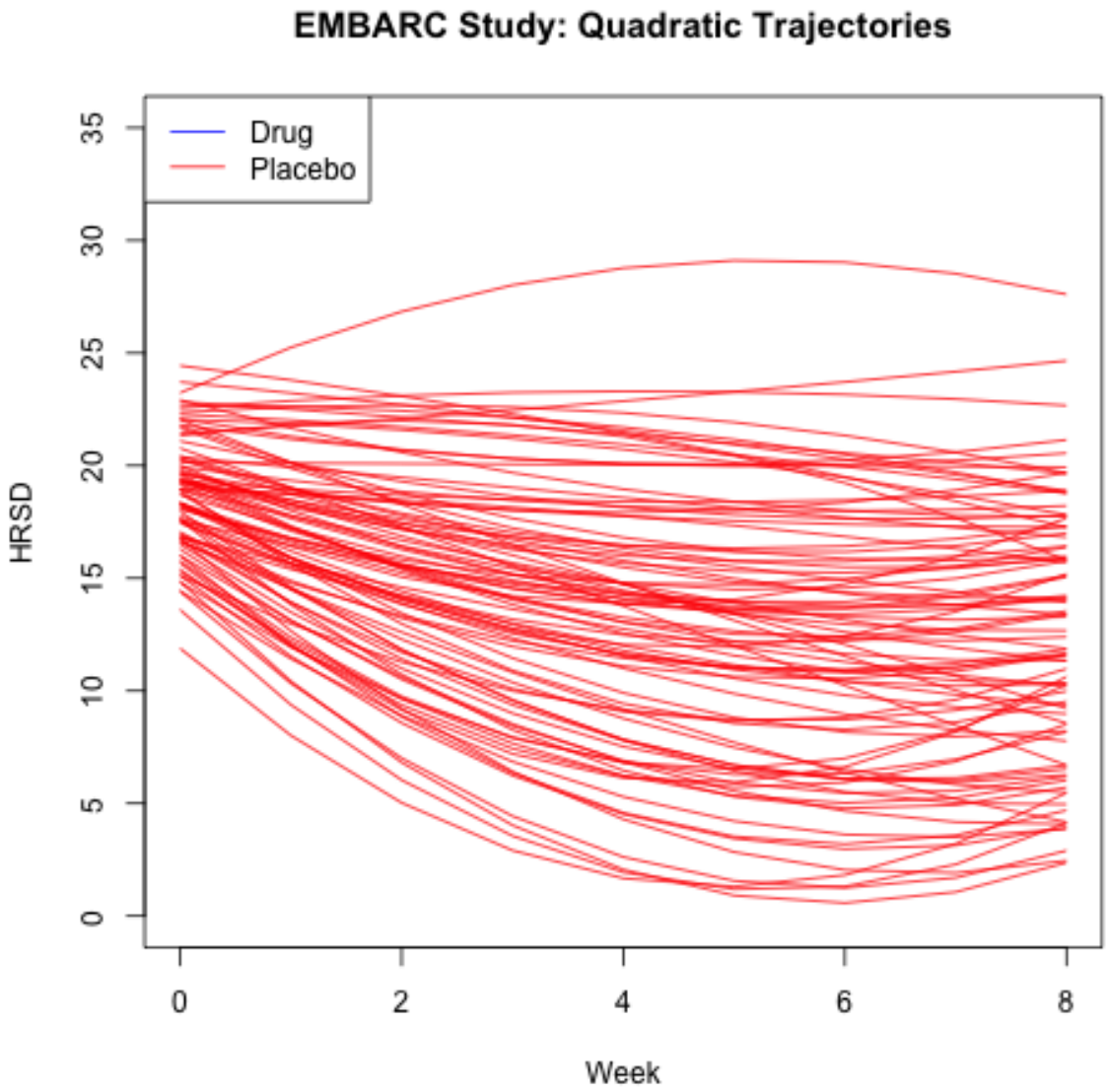
The EMBARC Study

Establishing moderators and biosignatures of antidepressant response in clinical care (EMBARC)

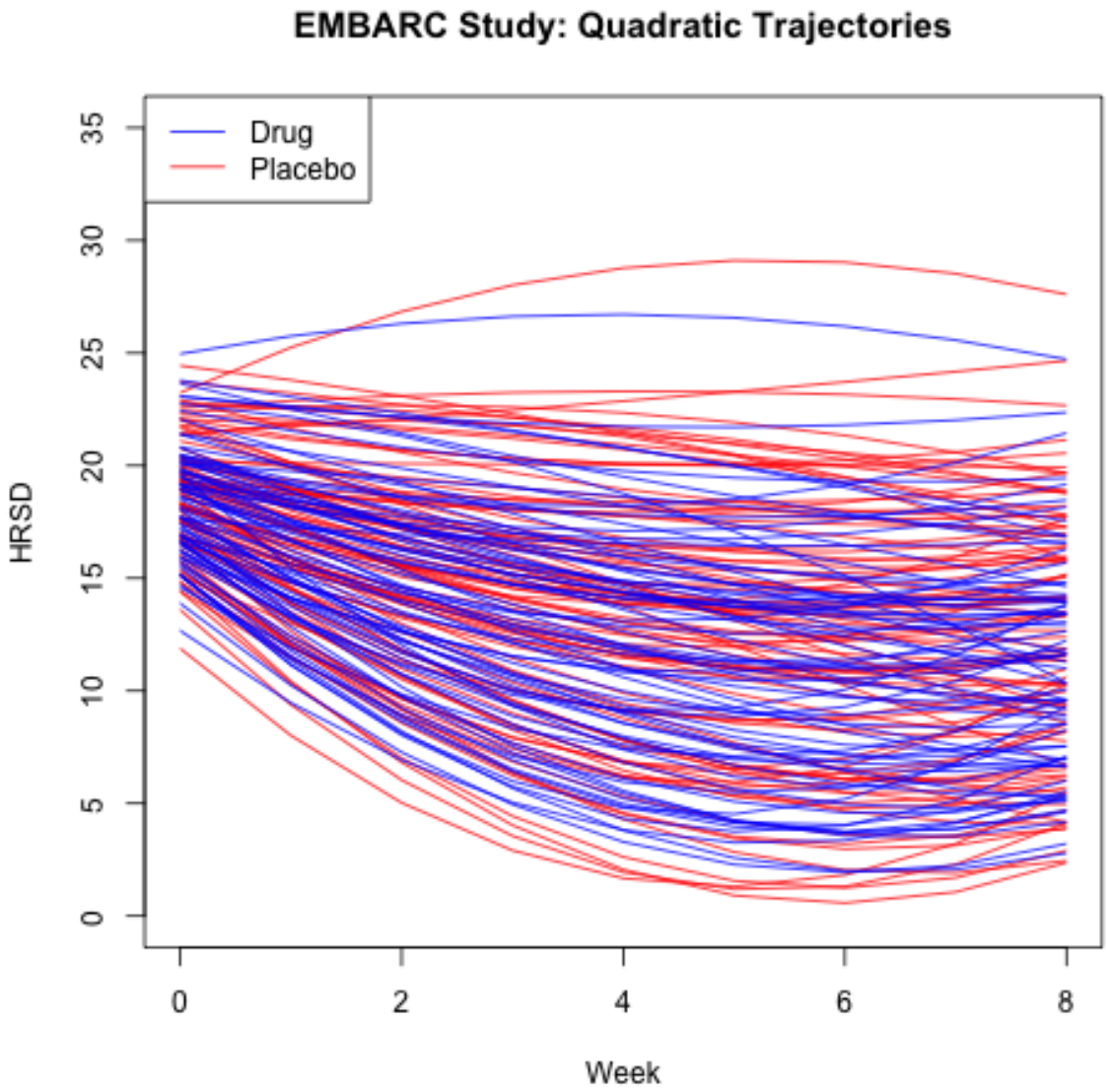
- A multi-site, placebo-controlled randomized clinical trial
- Participants with early onset ($\leq$30 years) MDD are included
- 8-week trial
- Hamilton Depression Rating Scale (HDRS) is used to measure a patient's level of depression

NYU School of Medicine
NYU LANGONE MEDICAL CENTER

Motivation Data: EMBARC

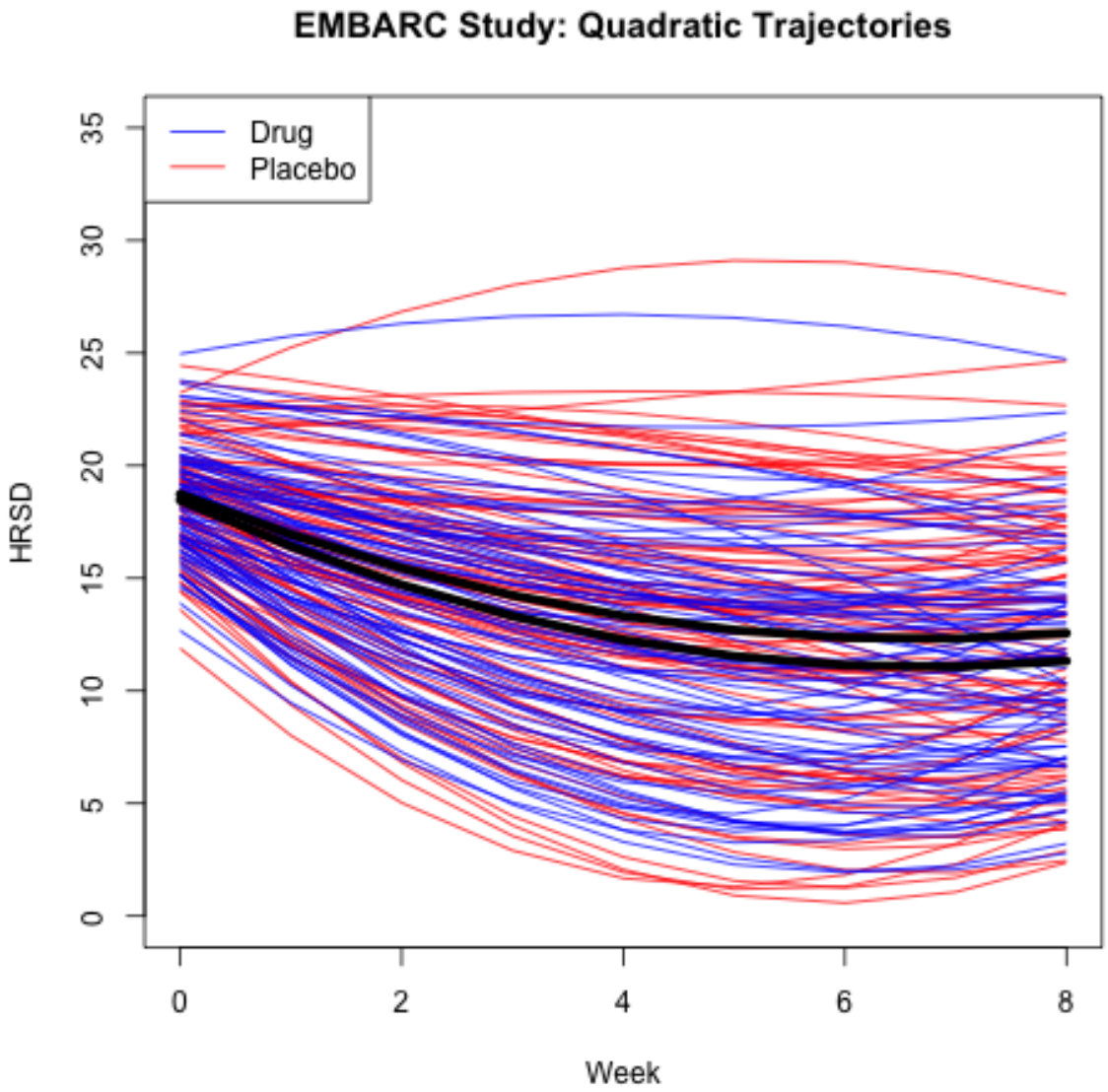


Motivation Data: EMBARC



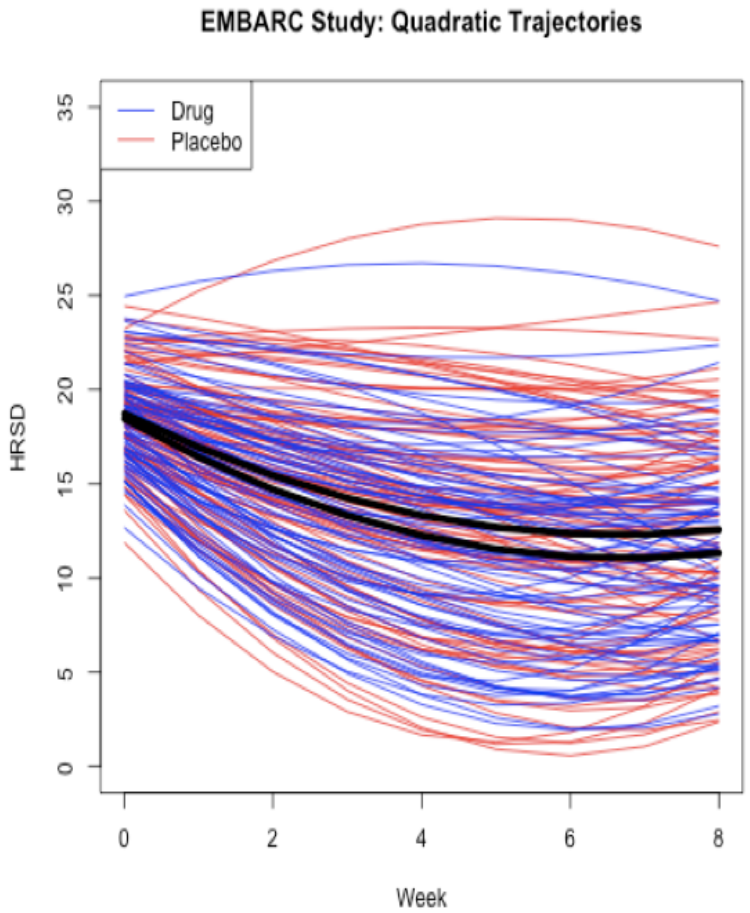
School of Medicine
NYU LANGONE MEDICAL CENTER

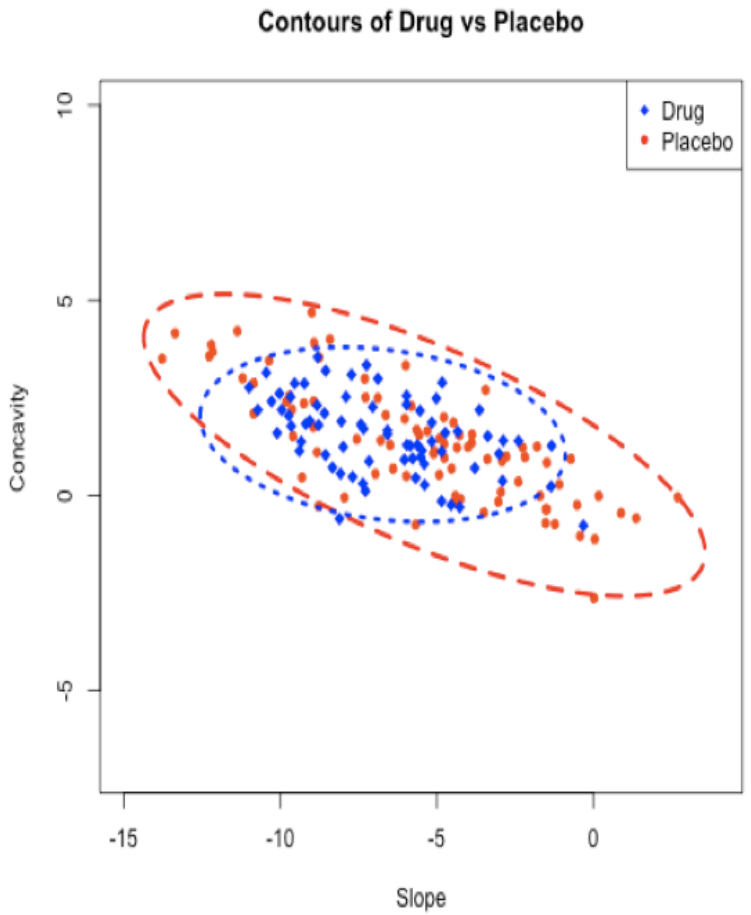
Motivation Data: EMBARC



EMBARC

Slope vs Concavity



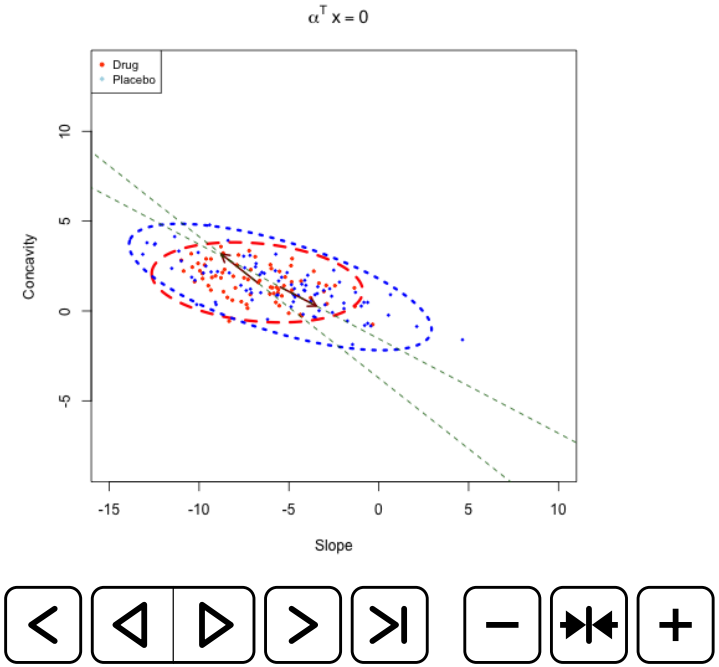


Medicine
NYU LANGONE MEDICAL CENTER

EMBARC

Slope vs Concavity

- link
- Separate the two distributions
- How far the two distributions are: **Purity**



Kullback-Leibler divergence

Recall Kullback-Leibler divergence:

In statistics, the Kullback-Leibler (KL) divergence is a measure of how one probability distribution F_1 is different from a second reference probability distribution F_2. For distributions F_1 and F_2 of a continuous random variable, the KL divergence is defined as:

$$D_{KL}(F_1||F_2) = \int_{-\infty}^{+\infty} f_1(x) \log(\frac{f_1(x)}{f_2(x)}) dx \tag{1}$$

where f_1 and f_2 denote the probability density of F_1 and F_2.

NYU School of Medicine
NYU LANGONE MEDICAL CENTER

Kullback-Leibler divergence

KL divergence

$$D_{KL}(F_1 \| F_2) = \int_{-\infty}^{+\infty} f_1(x) \log(\frac{f_1(x)}{f_2(x)}) dx$$

It has properties:

- The Kullback–Leibler Divergence is always non-negative:

$$D_{KL}(F_1 \| F_2) \geq 0, \ D_{KL}(F_2 \| F_1) \geq 0$$

- The KL Divergence $D_{KL}(F_1 \| F_2)$ can be thought of as something like a measurement of how far the distribution F_1 is from the distribution F_2.

NYU School of Medicine
NYU LANGONE MEDICAL CENTER

Model

In our setting, we assume the outcomes are from a linear mixed model:

$$Y = S\beta + b + \epsilon,$$

-
-
-
-
-
-
-

Model

In our setting, we assume the outcomes are from a linear mixed model:

$$Y = S(\beta + b + \Gamma(w)) + \epsilon,$$

-
-
-
-
-
-
-

Model

In our setting, we assume the outcomes are from a linear mixed model:

$$Y = S(\beta + b + \Gamma(w)) + \epsilon, \tag{2}$$

- $w = \alpha' x$ is the combination of the input baseline covariates.
- $\mathbf{S}$ is the matrix of times (intercept, linear, and quadratic term)
- β is the vector of covariates for fixed effects of S
- b is the vector of random effects
- Γ is the vector of fixed effects of the baseline covariates.
- $\alpha' x$ is the combination of the input baseline covariates.
- α has the restriction that $||\alpha|| = 1$

Model

In our setting, we assume the outcomes are from a linear mixed model:

$$Y = S(\beta + b + \Gamma(w)) + \epsilon,$$

Define the covariate matrix of S as z. The z contains both fixed effects and random effects.

$$z = \beta + b + \Gamma w$$

That is, we have distributions for the mixed-effect model coefficients z given $w = \alpha' x$, where

$$z|w \sim N(\beta_j + \Gamma_j w, D_j),$$

for treatment $j = 1, 2$.

NYU School of Medicine
NYU LANGONE MEDICAL CENTER

Assumptions

Based on the KL divergence, we define the *purity* of the data, which represents how much the differences between the treatment group distribution $f_1(x)$ and the placebo group distribution $f_2(x)$.

The assumptions:

- $f_1(z|w) \sim N(\mu_1, D_1),\ \mu_1 = \beta_1 + \Gamma_1 w$
- $f_2(z|w) \sim N(\mu_2, D_2),\ \mu_2 = \beta_2 + \Gamma_2 w$
- $X \sim MVN(\mu_x, \Sigma_x)$

Purity Functions

Subject Purity Function $g(\alpha' x)$ Define the **purity function** regard to a subject with baseilne biosignature x (i.e. the **purity function** given α and the baseline biosignature x) as:

$$\begin{aligned} g(\alpha' x) =& D_{KL}(F_1 || F_2) + D_{KL}(F_2 || F_1) \\ =& \int \log(f_1(z|\alpha' x)) f_1(z|\alpha' x) dz - \int \log(f_2(z|\alpha' x)) f_1(z|\alpha' x) dz \\ &+ \int \log(f_2(z|\alpha' x)) f_2(z|\alpha' x) dz - \int \log(f_1(z|\alpha' x)) f_2(z|\alpha' x) dz \end{aligned} \tag{3}$$

where

- $f_1(z|w) \sim N(\mu_1, D_1)$, $\mu_1 = \beta_1 + \Gamma_1 w$
- $f_2(z|w) \sim N(\mu_2, D_2)$, $\mu_2 = \beta_2 + \Gamma_2 w$

NYU School of Medicine
NYU LANGONE MEDICAL CENTER

Purity Functions

Data Purity Function purity(α) Define

- f_w as the distribution of the combination of baseline signature
- $w = \alpha' x$

The purity function regards to the whole data set is defined as:

$$\begin{aligned} \text{purity}(\alpha) = & \int g(\alpha' x) f_w(\alpha' x) d\alpha' x \\ = & E(g(\alpha' x)) \end{aligned} \tag{4}$$

Estimate:

$$\hat{\text{purity}}(\alpha) = \bar{g}(\alpha' x) = \frac{1}{n} \sum_i^n g(\alpha' x_i)$$

NYU School of Medicine
NYU LANGONE MEDICAL CENTER

Purity Calculation

$$g(\alpha' x) = \int f_1 \log f_1 - \int f_1 \log f_2 + \int f_2 \log f_2 - \int f_2 \log f_1$$

- $\int f_1 \log f_1 = -\frac{p}{2}\log(2\pi) - \frac{1}{2}\log(|D_1|) - \frac{p}{2}$
- $\int f_2 \log f_2 = -\frac{p}{2}\log(2\pi) - \frac{1}{2}\log(|D_2|) - \frac{p}{2}$
- $\int f_1 \log f_2 =$
  $-\frac{p}{2}\log(2\pi) - \frac{1}{2}\log(|D_2|) - \frac{1}{2}\left(tr(D_2^{-1}D_1) + (\mu_1 - \mu_2)' D_2^{-1}(\mu_1 - \mu_2)\right)$
- $\int f_2 \log f_1 =$
  $-\frac{p}{2}\log(2\pi) - \frac{1}{2}\log(|D_1|) - \frac{1}{2}\left(tr(D_1^{-1}D_2) + (\mu_1 - \mu_2)' D_1^{-1}(\mu_1 - \mu_2)\right)$

NYU School of Medicine
NYU LANGONE MEDICAL CENTER

Purity Calculation

The dataset's purity, which is the expectation of the $g()$ function is:

$$\begin{aligned} \text{purity}(\alpha) =& E(g(\alpha' x)) \\ =& A_0 + \frac{A_1}{2} + A_2 \mu_x' \alpha + \frac{A_3}{2}[\alpha' \Sigma_x \alpha + \alpha' \mu_x \mu_x' \alpha] \end{aligned} \tag{5}$$

where

- $A_0 = -p + \frac{1}{2} tr(D_2^{-1} D_1) + \frac{1}{2} tr(D_1^{-1} D_2)$
- $A_1 = (\beta_1 - \beta_2)'(D_1^{-1} + D_2^{-1})(\beta_1 - \beta_2)$
- $A_2 = (\beta_1 - \beta_2)'(D_1^{-1} + D_2^{-1})(\Gamma_1 - \Gamma_2)$
- $A_3 = (\Gamma_1 - \Gamma_2))'(D_1^{-1} + D_2^{-1})((\Gamma_1 - \Gamma_2)$

All A_0, A_1, A_2, A_3 are scalars.

NYU School of Medicine
NYU LANGONE MEDICAL CENTER

Purity Optimization

Data Purity

$$\text{purity}(\alpha) = A_0 + \frac{A_1}{2} + A_2\mu_x'\alpha + \frac{A_3}{2}[\alpha'\Sigma_x\alpha + \alpha'\mu_x\mu_x'\alpha]$$

Goal:

Find the α that maximizes the purity function.

- Method to find the extreme value:
  - Derivation

    A_0, A_1, A_2, A_3 are also functions of α
  - Newton-Raphson

NYU School of Medicine
NYU LANGONE MEDICAL CENTER

Algorithm

We could summarize the above purity calculation function as

Algorithm 1 Purity Calculation and Optimization

1: Select baseline covariates x, with dimension p.
2: Initial an α. Calculate $w = \alpha' x$
3: Fit the models $Y = S(\beta + b + \Gamma(w)) + \epsilon$ from data in group 1 and group 2, separately.
4: Estimate $\beta_1, \beta_2, \Gamma_1, \Gamma_2, D_1, D_2$
5: Calculate purity based on the function

$$\text{purity}(\alpha) = A_0 + \frac{A_1}{2} + A_2 \mu_x' \alpha + \frac{A_3}{2}[\alpha' \Sigma_x \alpha + \alpha' \mu_x \mu_x' \alpha]$$

6: Optimize the purity(α) function with Newton Raphson algorithm. Obtain the $\hat{\alpha}^*$ that maximizes the purity function. =0

NYU School of Medicine
NYU LANGONE MEDICAL CENTER

Simulation Setting

Model

$$Y_j = S_j(\beta_j + b_j + \Gamma_j(w)) + \epsilon,\ W = \alpha' x,\ j \in \{1, 2\}$$

- Time points matrix S: week $0, 1, 2, ..., 7$

$$S = \begin{pmatrix} 1 & t_0 & t_0^2 \\ 1 & t_1 & t_1^2 \\ ... & ... & ... \\ 1 & t_7 & t_7^2 \end{pmatrix}$$

- $\beta_1 = (0, 3, 0.9)', \beta_2 = (0, 3.1, 1)'$
- $\Gamma_1 = (0, 1, 0)'$, angle between the two Γ_1 and Γ_2 directions range between 0, 30, 60, 90, 120, 150, and 180 degrees.
- $||\Gamma_1|| = ||\Gamma_2|| = 1$

NYU School of Medicine
NYU LANGONE MEDICAL CENTER

Simulation Setting

- $b_i \sim N(0, D_i)$, D_1, D_2 (randomly generated matrices):

$$D_1 = \begin{pmatrix} 1.45 & -0.11 & 0.20 \\ -0.11 & 0.17 & -0.08 \\ 0.20 & -0.08 & 0.22 \end{pmatrix}, \; D_2 = \begin{pmatrix} 1.02 & -0.23 & 0.15 \\ -0.23 & 0.68 & 0.25 \\ 0.15 & 0.25 & 1.36 \end{pmatrix}$$

- The covariates are equally correlated ($\rho = 0.5$) and normal: $X \sim N(0, \Sigma_X)$
- True $\alpha_0 = (0.5, 0.5, 0.5, 0.5)'_4$ for $p = 4$ dimensional covariates
- True $\alpha_0 = (0.25, 0.25, ..., 0.25)'_1 6$ for $p = 16$ dimensional covariates
- $\|\alpha\| = 1$

NYU School of Medicine
NYU LANGONE MEDICAL CENTER

Simulation Setting

Table: Scenario settings

Scenario	N	p	D matrix	Initial α
1	400	4	Estiamted: $\hat{D}_1, \hat{D}_2$	α_0
2	400	4	True D_1, D_2	α_0
3	400	4	Estiamted: $\hat{D}_1, \hat{D}_2$	$\alpha_0 + N(0, 0.1)$
4	400	4	True D_1, D_2	$\alpha_0 + N(0, 0.1)$
5	2000	16	Estiamted: $\hat{D}_1, \hat{D}_2$	α_0
6	2000	16	True D_1, D_2	α_0
7	2000	16	Estiamted: $\hat{D}_1, \hat{D}_2$	$\alpha_0 + N(0, 0.1)$
8	2000	16	True D_1, D_2	$\alpha_0 + N(0, 0.1)$

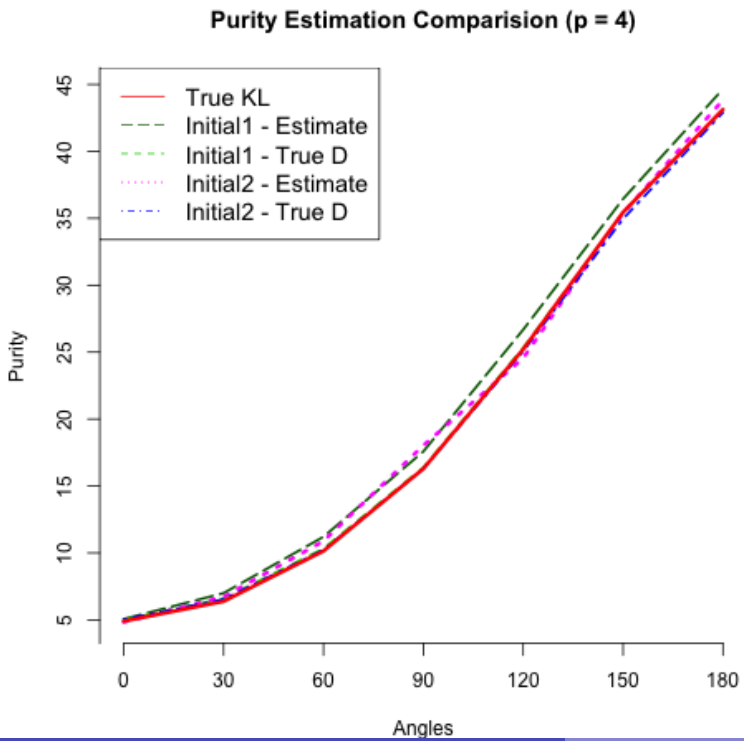
# of simulated data: 100

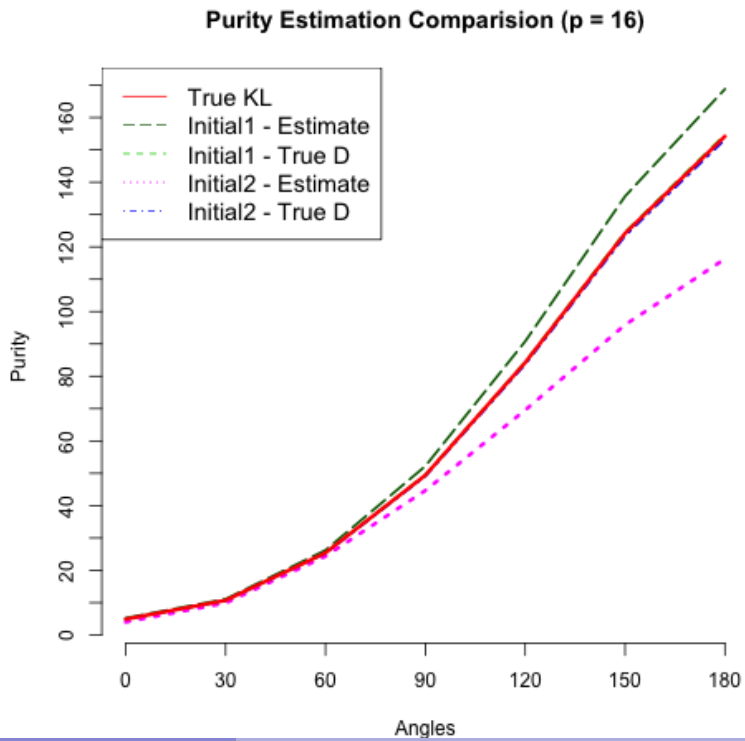
NYU School of Medicine
NYU LANGONE MEDICAL CENTER

Simulation Results: Purity Estimation

Simulation results for $p = 4$ and $p = 16$ predictors, sample size $n = 200$ per treatment arm.

- Initial 1: true α used as initial starting point in search
- Initial 2: Random starting value for α
- Estimate: Randome effect covariance matrix estimated
- True D: True random effect covariance matrix used

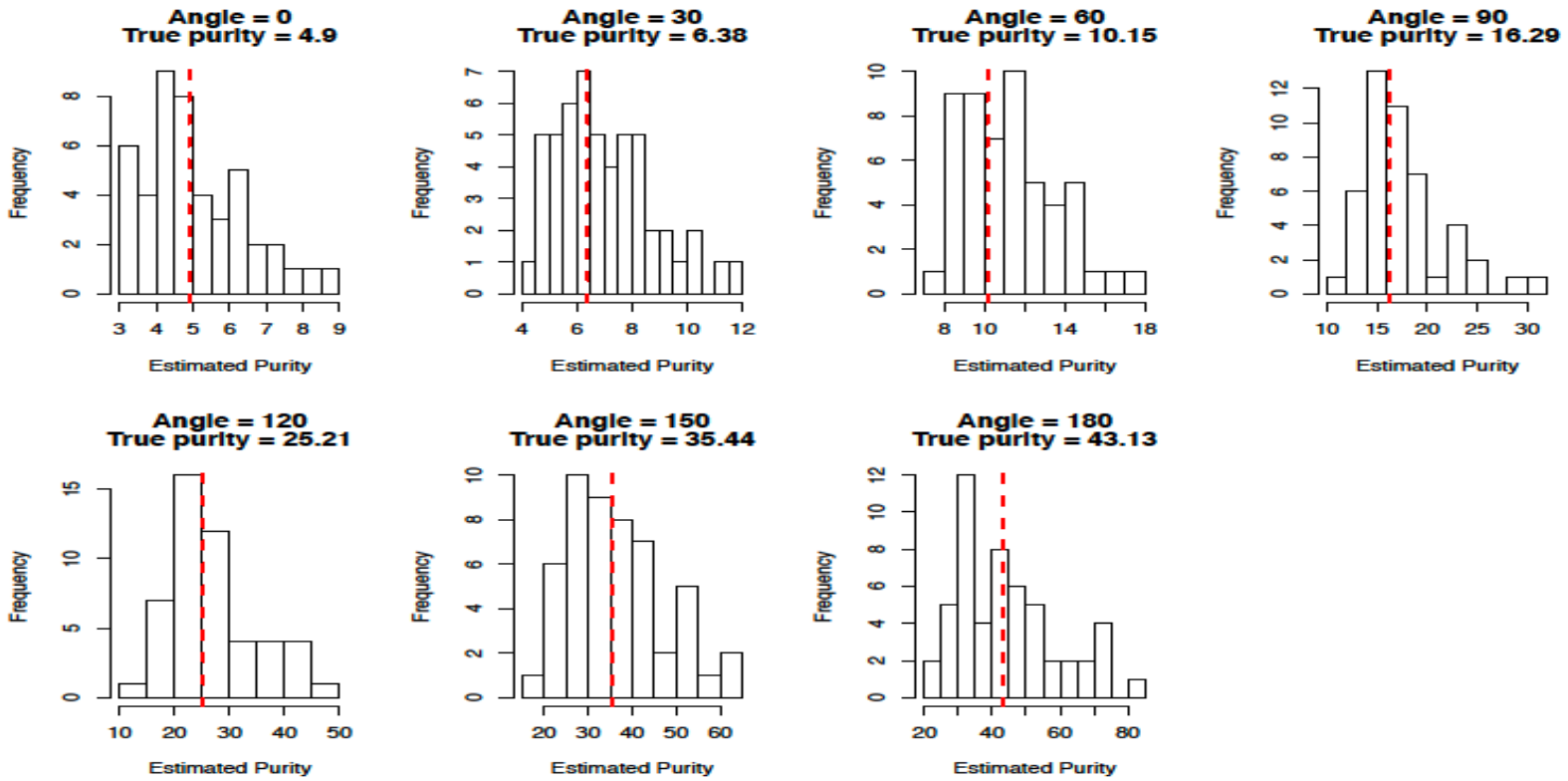




Purity Histogram

Purity Distributions as angle between Γ_1 and Γ_2 varies

- $p = 4$ predictors, sample size $n = 200$ per treatment arm

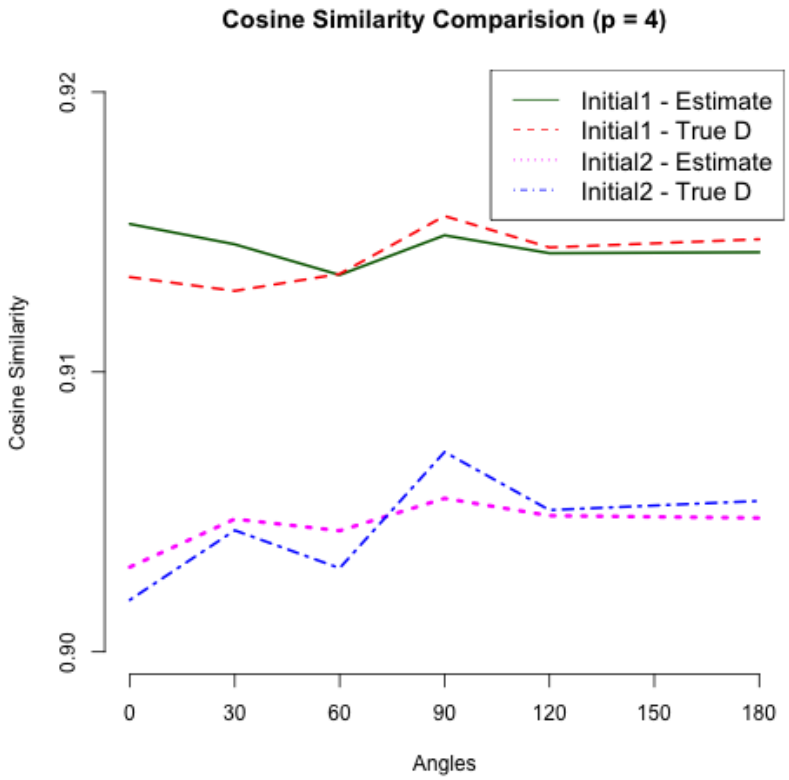


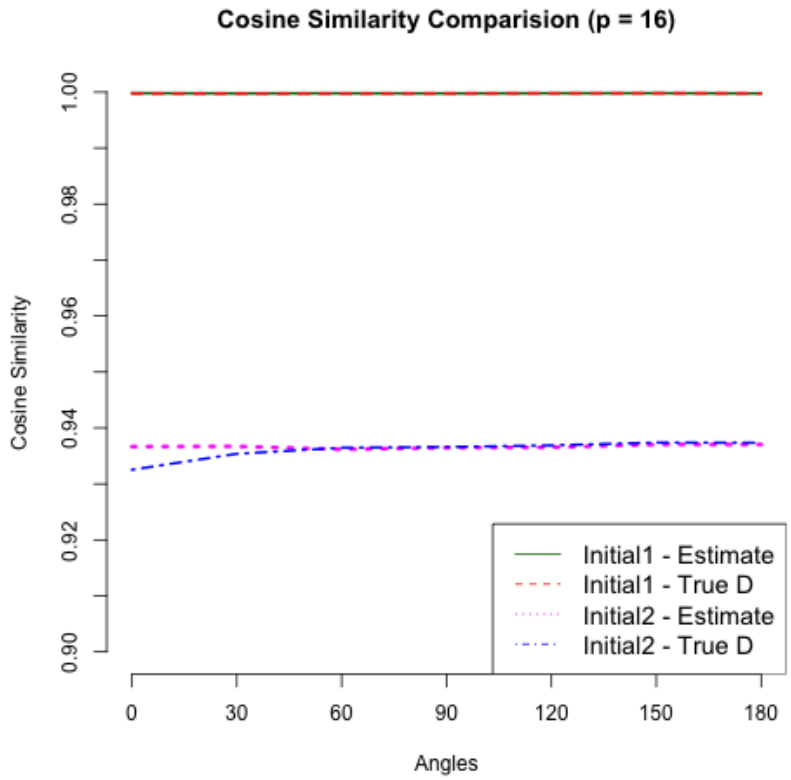
NYU School of Medicine
NYU LANGONE MEDICAL CENTER

Simulation Results

- Cosine Similarity: a measure of similarity between two (high-dimensional) vectors

$$\text{Similarity} = \cos(\theta) = \alpha' \hat{\alpha}, \quad \|\alpha\| = \|\hat{\alpha}\| = 1$$

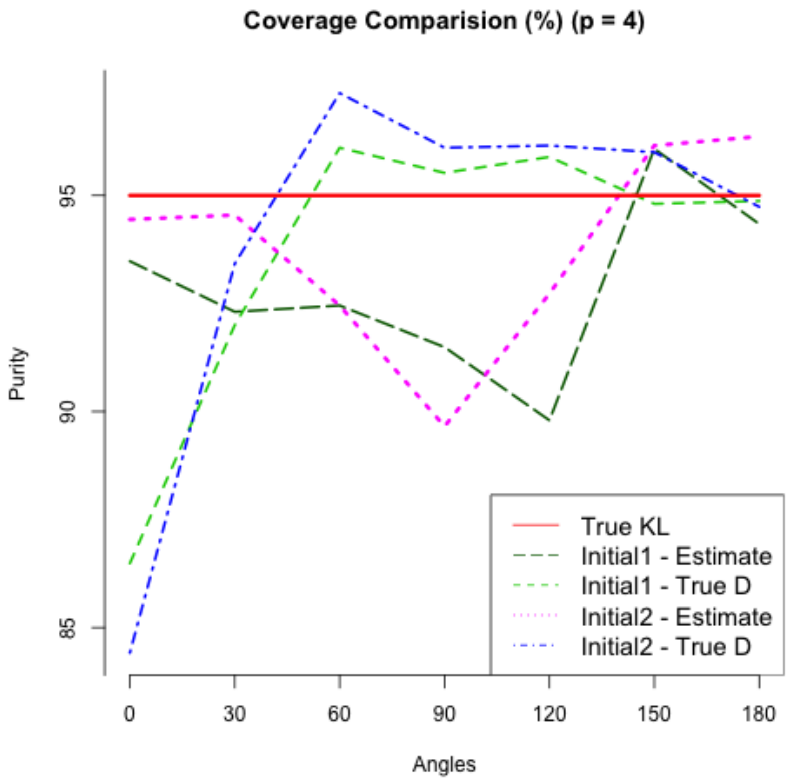


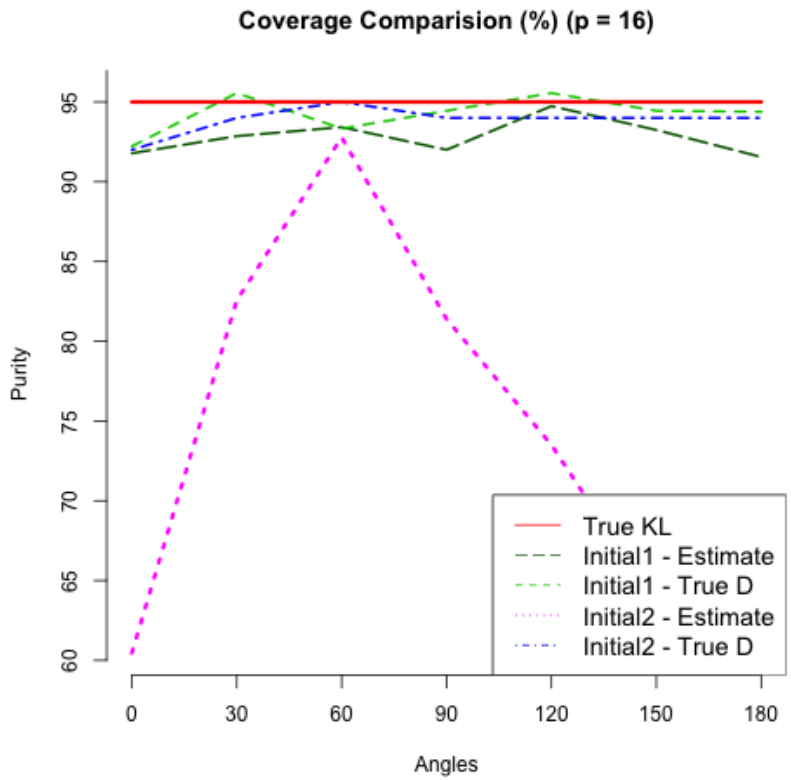


Simulation Results

- Coverage (95 % confidence interval)

$$\mathrm{KL}_0 \in (\hat{KL} - 1.96\hat{\sigma}_{KL}, \hat{KL} + 1.96\hat{\sigma}_{KL})$$





Example: EMBARC data analysis

The EMBARC Study

Establishing moderators and biosignatures of antidepressant response in clinical care (EMBARC)

- A multi-site, placebo-controlled randomized clinical trial
- Participants with early onset ($\leq$30 years) MDD are included
- 8-week trial
- Hamilton Depression Rating Scale (HDRS) is used to measure a patient's level of depression
- $N = 160$: 87 in control arm and 73 in intervention arm

NYU School of Medicine
NYU LANGONE MEDICAL CENTER

Covariates (baseline biosignature)

Covariates (baseline biosignature)

Covariate name	Description
age_evaluation	Age at baseline
dur_MDE	Duration of current major depressive episode
age_MDE	Age of first major depressive episode
axis2	Severity of the most severy Axis II diagnosis at baseline
anger_attack	Severity of anger attacks at baseline
anxious	Severity of anxiety at baseline

NYU School of Medicine
NYU LANGONE MEDICAL CENTER

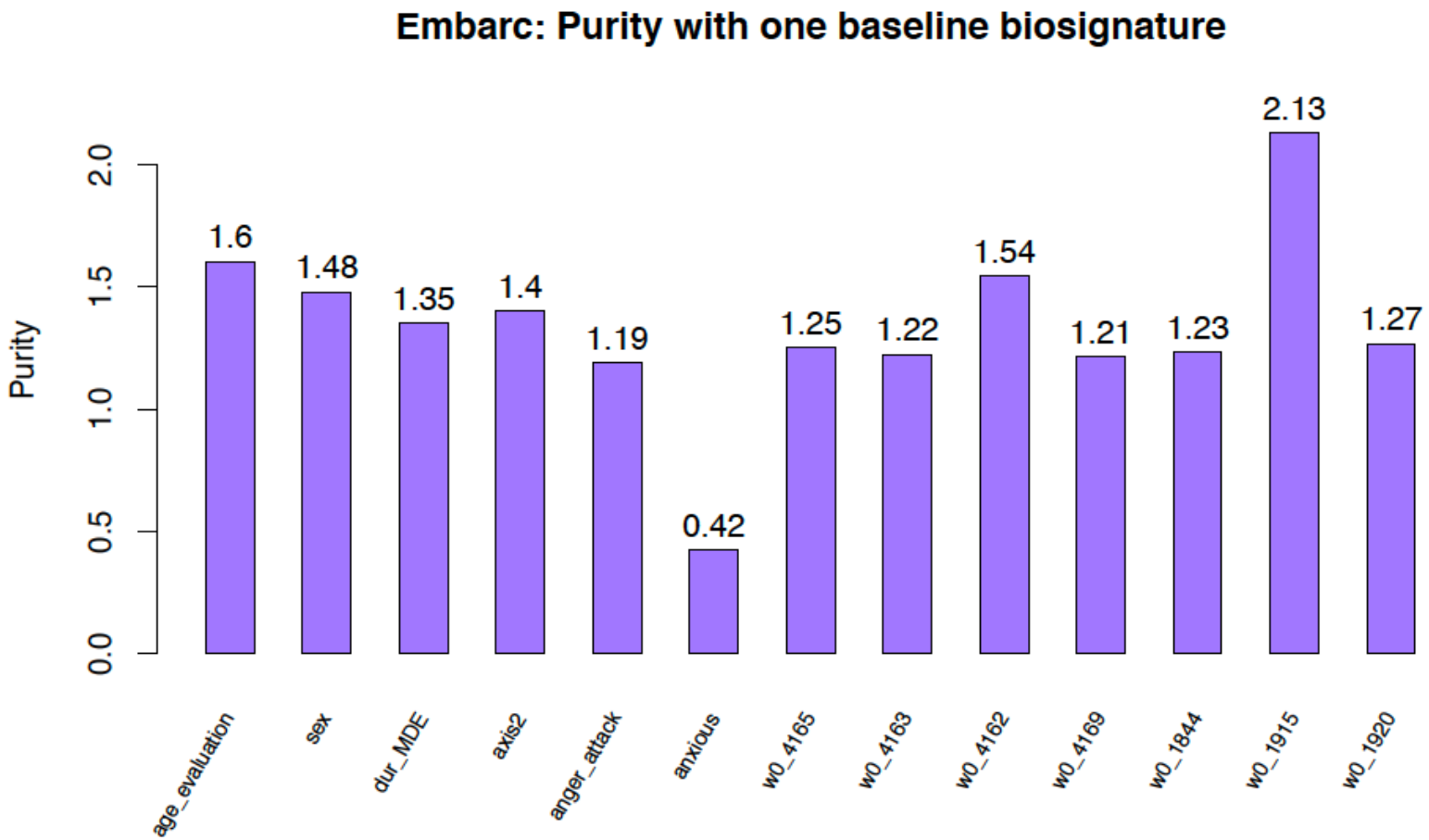
Covariates (baseline bio-signature)

Behavior Measures

Covariate name	Description
w0_4165	Interference Reaction Time in negative trials
w0_4167	Interference Reaction Time in non-negative trials
w0_4163	Interference Reaction Time in all trials
w0_4162	Total number of correct trials
w0_4169	Median Reaction time for correct trials
w0_1844	Number of valid recalled words in the Word Fluency task
w0_1916	Flanker Accuracy
w0_1915	Flanker Reaction Time
w0_1920	Accuracy effect

NYU School of Medicine
NYU LANGONE MEDICAL CENTER

EMBARC: Purity with one baseline biosignature



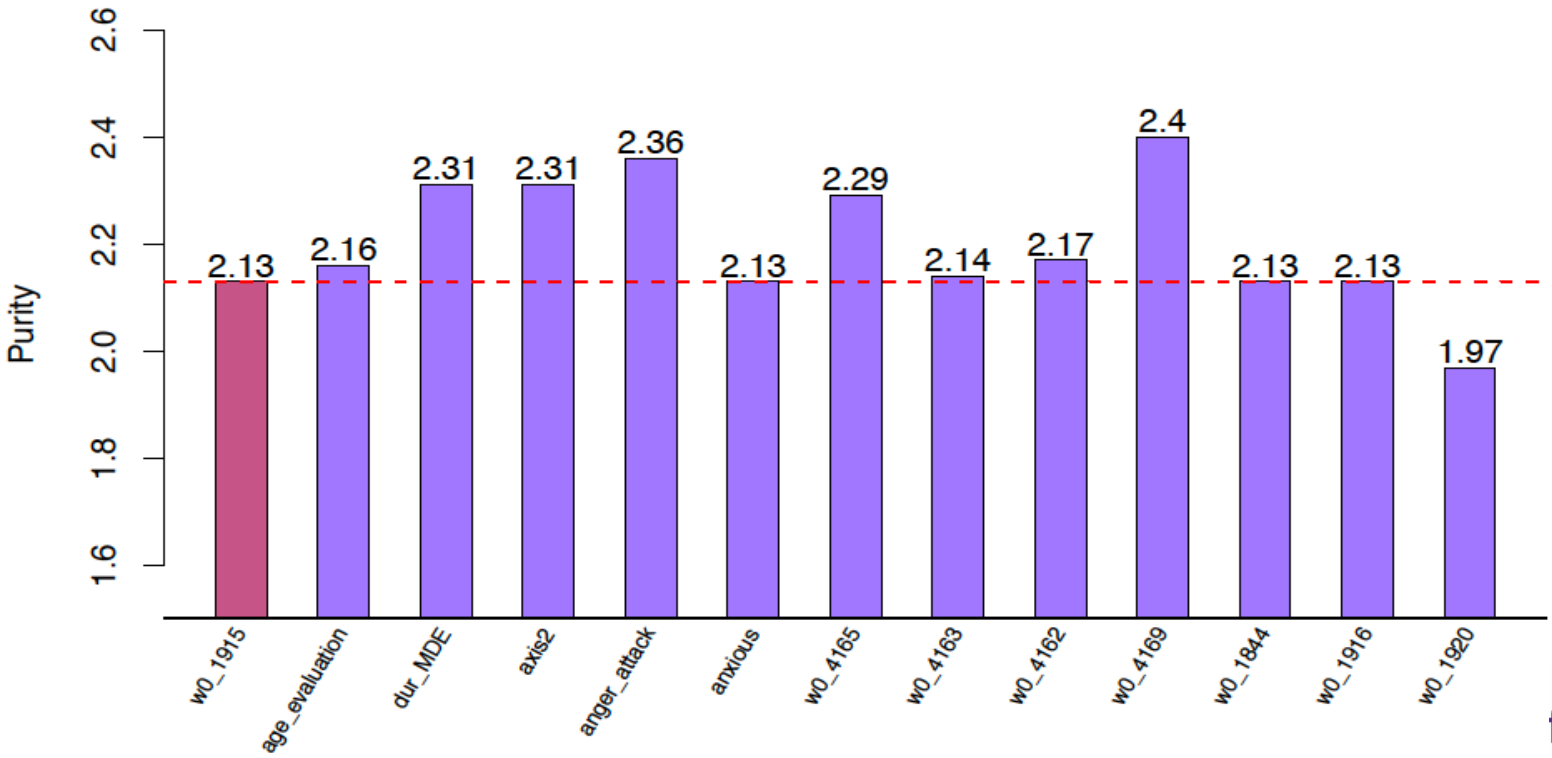
NYU School of Medicine
NYU LANGONE MEDICAL CENTER

EMBARC: Purity with two baseline biosignature

w0_1915: Flanker Reaction Time

Purity is calculated with w0_1915 and one of the other covariates.

Figure: Purity calculated with w0_1915 and another covariate

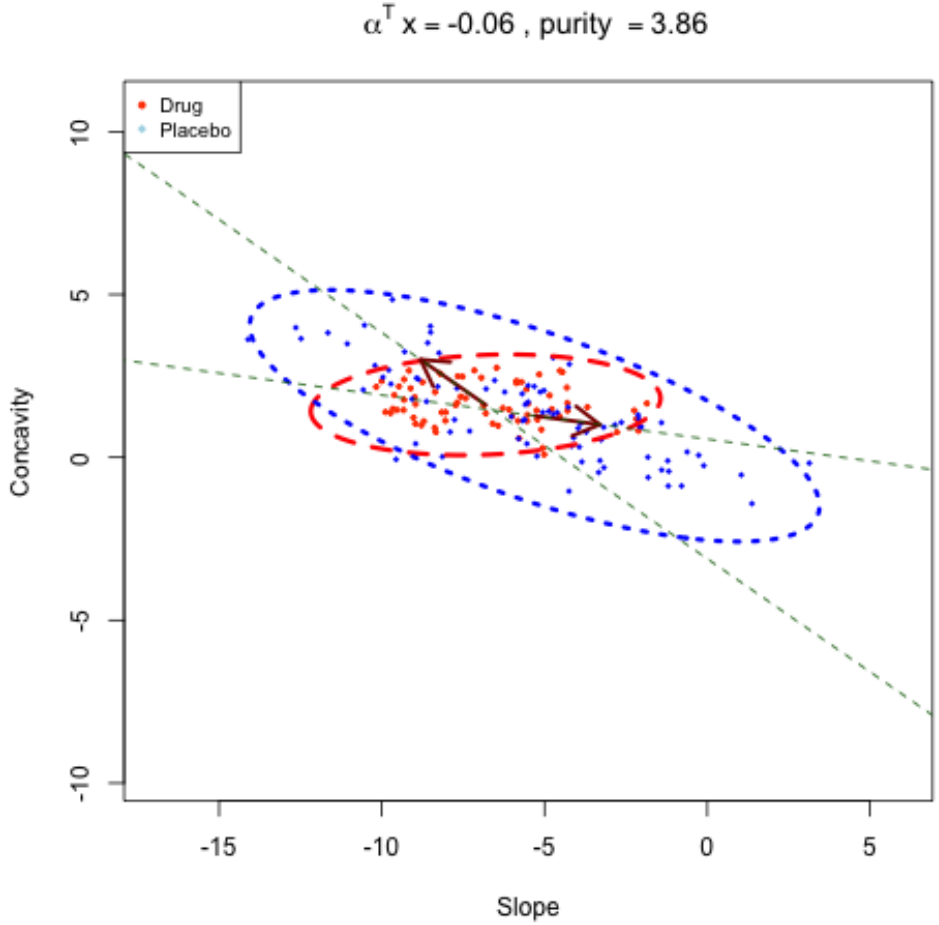


EMBARC: Purity with All Baseline Variables

Covariates	α_j
Age at baseline	-0.05
Duration of current major depressive episode	0.22
Age of first major depressive episode	0.21
Severity Axis II diagnosis at baseline	0.18
Severity of anger attacks at baseline	0.03
Severity of anxiety at baseline	0.27
Interference Reaction Time (negative trials)	0.32
Interference Reaction Time (non-negative trials)	-0.60
Interference Reaction Time (all trials)	-0.13
Total number of correct trials	-0.18
Median Reaction time for correct trials	0.09
# of valid recalled words (Word Fluency task)	-0.10
Flanker Accuracy	0.22
Flanker Reaction Time	-0.43
Accuracy effect	-0.17

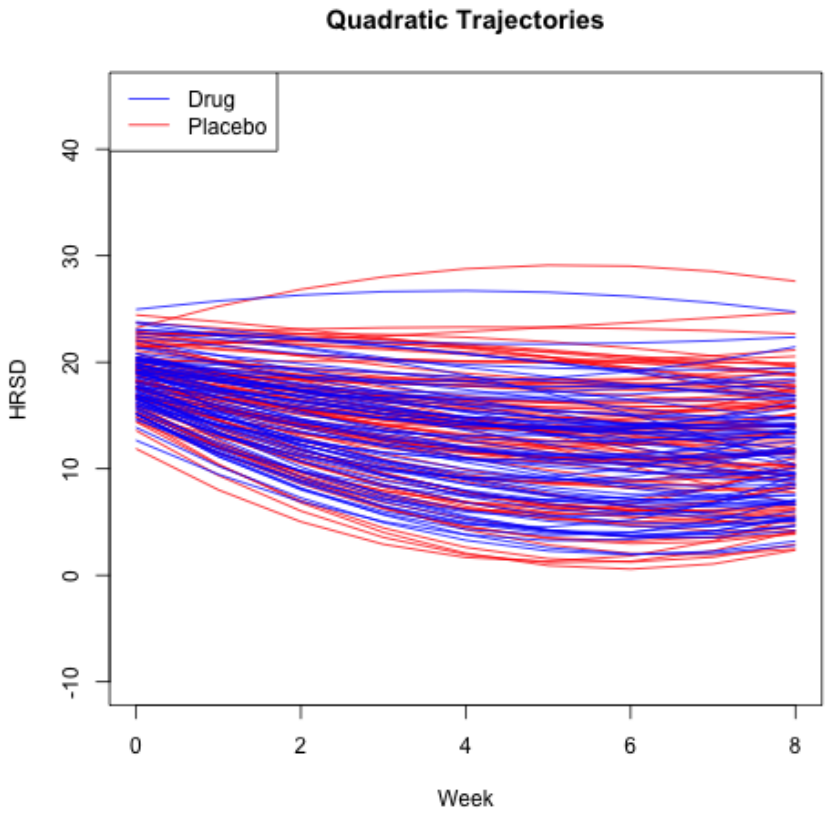
The Purity = 5.45

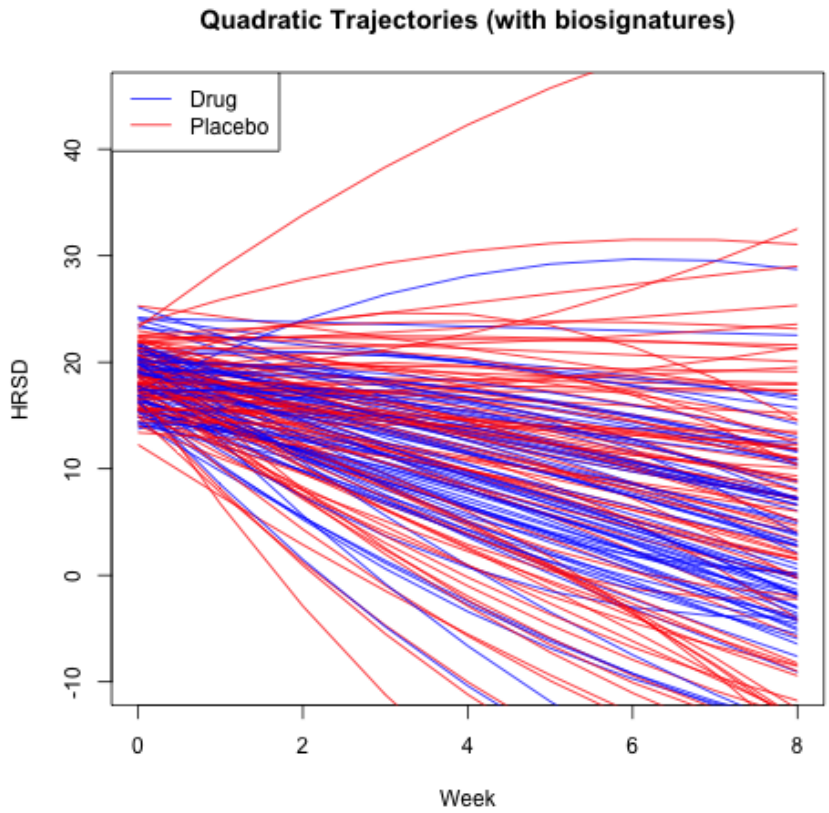
EMBARC: Individual Purity



EMBARC: Individual Purity Trajectory

Fit model without or with biosignatures (used the estimated α)



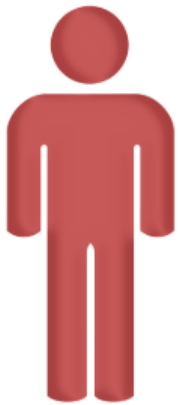


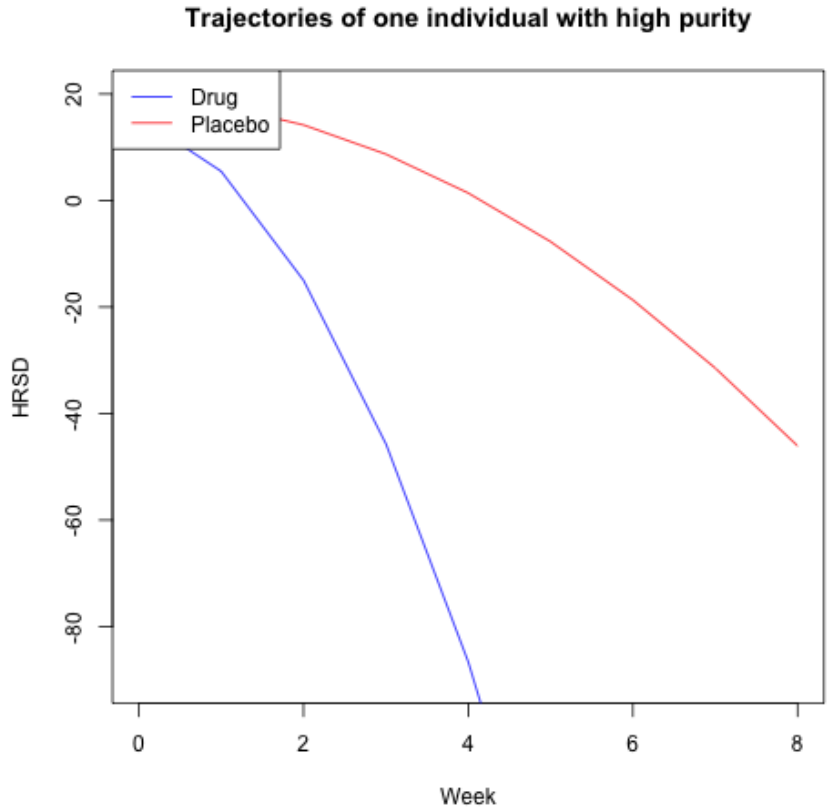
NYU School of Medicine
NYU LANGONE MEDICAL CENTER

EMBARC: Individual Purity

One subject:

- $\alpha' x = -2.76$
- Individual purity: 26.16





Conclusion

- EMBARC study had two arms: Placebo and Treatment
- Methodology described here finds a biosignature that separates the two treatment groups conditional on baseline covariate biosignature.
- Simulations showed that the linear combination of biosignatures can be estimated accurately
- Comparison of purity within a study and comparison of purity with different studies.

Reference

- Tarpey T, Petkova E, Zhu L. Stratified psychiatry via convexity-based clustering with applications towards moderator analysis. Statistics and its interface. 2016 Jul 1;9(3):255.
- Tarpey T, Ogden RT, Petkova E, Christensen R. A paradoxical result in estimating regression coefficients. The American Statistician. 2014 Oct 2;68(4):271-6.
- Petkova E, Tarpey T, Su Z, Ogden RT. Generated effect modifiers (GEM's) in randomized clinical trials. Biostatistics. 2016 Jul 27;18(1):105-18.
- Bock HH. Convexity-based clustering criteria: theory, algorithms, and applications in statistics. Statistical Methods and Applications. 2004 Feb 1;12(3):293-317.

NYU School of Medicine
NYU LANGONE MEDICAL CENTER

Thank you!