Functional and imaging data in precision medicine

R. Todd Ogden

Department of Biostatistics, Columbia University

with Adam Ciarleglio and Eva Petkova (New York University),
Thaddeus Tarpey (Wright State University),
Aaron Ma and Hyung Park (Columbia University)

February 4, 2016

Motivation

- Major depressive disorder (MDD) affects approximately 5% of the worldwide population each year.
- It is the leading cause of disability worldwide (in terms of years lost due to disability).
- Standard treatments take (at least) 6–8 weeks to take effect, during which time patients have poor quality of life and are at high risk of suicide.
- Treatment assignment is done by “trial and error.”

A motivating dataset: The EMBARC study

The EMBARC (Establishing Moderators And Biosignatures of Antidepressant Response for Clinical Care) study is an ongoing multi-site randomized placebo-controlled clinical trial.

- 400 subjects are randomized to placebo or citalopram
- At baseline, measure
  - clinical characteristics — diagnostic measures, treatment history, comorbidity, ...
  - neurophychological assessments — word fluency, emotion processing and regulation, ...
  - brain structure — structural MRI
  - integrity of white matter tracks in the brain — DTI
  - "resting state" EEG and fMRI
  - brain function while performing certain tasks — fMRI and EEG

Can imaging (and other) data be used in making patient-specific treatment decisions?

EMBARC goals

The primary goals of the EMBARC project are:

1. To select measurements that can be made at baseline that will help predict patient response to treatment.
2. To determine a rule, based on these measurements, that will assign the treatment that is best for each patient.

Baseline measurements consist of *scalar* quantities and *functional* data.

Setup

- Continuous response Y (large values are better)
- Treatment $A = -1$ or 1
- Scalar covariates $\boldsymbol{Z} = (1, Z_1, \ldots, Z_p)^\intercal$ (age, severity, clinical/cognitive measures, etc.)
- Functional observations $\boldsymbol{Z} = \{X_1(t), t \in \mathcal{T}_1\}, \ldots, \{X_q(t), t \in \mathcal{T}_q\}$ (can be 1-D, 2-D, 3-D, ...)

Setup

- Potential outcomes: $Y^*(-1), Y^*(1)$ but we observe only $Y = Y^*(1)(A+1)/2 + Y^*(-1)(1-A)/2$.
- "Treatment regime": $g : (\boldsymbol{Z}, \boldsymbol{X}) \to \{-1, 1\}$
- We want to choose g to make $\mathrm{E}[Y^*(g(Z,X))]$ as large as possible.
- The "value" of the decision rule is $\int_{\boldsymbol{Z}} \int_{\boldsymbol{X}} \mathrm{E}[Y^*(g(Z,X))] d\boldsymbol{X} d\boldsymbol{Z}$

Some general approaches

- A-learning ("Advantage learning"): Murphy (2003); Robins (2004); Blatt, *et al.* (2004)
- Q-learning ("Quality learning"): Watkins and Dayan (1992); Nahum-Shani *et al.* (2010)
- OWL ("Outcome weighted learning"): Zhao, *et al.* (2012)

These generally rely on a relatively small number of scalar covariates to make decisions.

Techniques using functional data also exist: McKeague and Qian (2014); Ciarleglio *et al.* (2015).

These all rely on some type of model (but they try to make the methods robust to model misspecification).

Data model

$$E[Y|\boldsymbol{Z},\boldsymbol{X},A] = h_{\boldsymbol{\alpha},\boldsymbol{\beta}}(\boldsymbol{Z},\boldsymbol{X}) + \frac{A}{2} f_{\boldsymbol{\gamma},\boldsymbol{\omega}}(\boldsymbol{Z},\boldsymbol{X})$$

- $h_{\boldsymbol{\alpha},\boldsymbol{\beta}}(\boldsymbol{Z},\boldsymbol{X}) = \boldsymbol{\alpha}^\intercal \boldsymbol{Z} + \sum_{\ell=1}^{p} \int \beta_\ell(s) X_\ell(s) ds$
- $f_{\boldsymbol{\gamma},\boldsymbol{\omega}}(\boldsymbol{Z},\boldsymbol{X}) = \boldsymbol{\gamma}^\intercal \boldsymbol{Z} + \sum_{\ell=1}^{p} \int \omega_\ell(s) X_\ell(s) ds$
- $\boldsymbol{\beta} = \{\beta_1, \ldots, \beta_q\}$ and $\boldsymbol{\omega} = \{\omega_1, \ldots, \omega_q\}$

$$f_{\boldsymbol{\gamma},\boldsymbol{\omega}}(\boldsymbol{Z},\boldsymbol{X}) = E[Y|\boldsymbol{Z},\boldsymbol{X},A=1] - E[Y|\boldsymbol{Z},\boldsymbol{X},A=-1]$$

Therefore:

$$\begin{aligned} g^{opt}(\boldsymbol{Z},\boldsymbol{X}) &= sign\{f_{\boldsymbol{\gamma},\boldsymbol{\omega}}(\boldsymbol{Z},\boldsymbol{X})\} \\ &= sign\left\{\boldsymbol{\gamma}^\intercal \boldsymbol{Z} + \sum_{\ell=1}^{p} \int \omega_\ell(s) X_\ell(s) ds\right\} \end{aligned}$$

Fitting the model

Elements of fitting procedure (Ciarleglio, *et al.*, 2015):

- Express functional observations in terms of eigenfunctions of smoothed estimated covariance operator (Goldsmith, *et al.*, 2011)
- Express the objective function for fitting the data as a loss function in the framework of A-learning (Murphy, 2003)
- Smoothing parameters may be chosen by REML

We would like to consider a procedure for variable selection also.

Modified covariates method (Tian *et al.*, 2014)

Note that: $E(2YA|\boldsymbol{Z},\boldsymbol{X}) = f_{\boldsymbol{\gamma},\boldsymbol{\omega}}(\boldsymbol{Z},\boldsymbol{X})$

Estimate $\boldsymbol{\gamma}$ and $\boldsymbol{\omega}$ by minimizing:

$$\frac{1}{n}\sum_{i=1}^{n}\left(2Y_iA_i - f_{\boldsymbol{\gamma},\boldsymbol{\omega}}(\boldsymbol{Z}_i,\boldsymbol{X}_i)\right)^2 \propto \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - f_{\boldsymbol{\gamma},\boldsymbol{\omega}}(\boldsymbol{Z}_i,\boldsymbol{X}_i)\frac{A_i}{2}\right)^2$$

So we can estimate $\boldsymbol{\gamma}$ and $\boldsymbol{\omega}$ by fitting

$$\begin{aligned} Y &= f_{\boldsymbol{\gamma},\boldsymbol{\omega}}(\boldsymbol{Z},\boldsymbol{X}) \cdot \frac{A}{2} + \varepsilon \\ &= \boldsymbol{\gamma}^{\mathsf{T}}\left\{\boldsymbol{Z} \cdot \frac{A}{2}\right\} + \sum_{\ell=1}^{p}\int \omega_\ell(s)\left\{X_\ell(s)\frac{A}{2}\right\} ds + \varepsilon \end{aligned}$$

Combining estimation with variable selection and roughness penalization

Define $L_n(\boldsymbol{\gamma}, \boldsymbol{\omega}) = \frac{1}{n} \sum_{i=1}^{n} \left(Y_i - f_{\boldsymbol{\gamma},\boldsymbol{\omega}}(\boldsymbol{Z}_i, \boldsymbol{X}_i) \frac{A_i}{2} \right)^2$

Express functional components in terms of spline basis functions, choose $\boldsymbol{\gamma}$ and $\boldsymbol{\omega}$ to minimize (Gertheiss, *et al.*, 2013)

$$L_n(\boldsymbol{\gamma}, \boldsymbol{\omega}) + \lambda \left\{ \sum_{j=2}^{p+1} J(|\gamma_j|) + \sum_{\ell=1}^{q} P_{\rho_\ell}(\omega_\ell) \right\}$$

$J(|\gamma_j|) = |\gamma_j| \qquad P_{\rho_\ell}(\omega_\ell) = \left\{ ||\omega_\ell||^2 + \rho_\ell Q(\omega_\ell) \right\}^{1/2} \qquad ||\omega_\ell|| = \int \omega_\ell^2(s) ds$

1D: $Q(\omega_\ell) = \left|\left| \frac{\partial^2 \omega_\ell}{\partial s^2} \right|\right|^2$ 2D: $Q(\omega_\ell) = \left|\left| \frac{\partial^2 \omega_\ell}{\partial s^2} \right|\right|^2 + \left|\left| \frac{\partial^2 \omega_\ell}{\partial s \partial t} \right|\right|^2 + \left|\left| \frac{\partial^2 \omega_\ell}{\partial t^2} \right|\right|^2$

Tuning parameters:

- λ (complexity)
- $\rho_\ell, \ell = 1, \ldots, q$ (smoothness)

Group lasso

For fixed $\lambda, \rho_1, \ldots, \rho_q$, this objective function can be optimized using procedures for the "group lasso" (Yuan and Lin, 2006).

- Scalar covariates are regarded as groups of size 1.
- Can fit using `grplasso` in `R`.
- Tuning parameters may be chosen by cross-validation (although there are $q + 1$ of them ...).

Other variations

- Augmentation (Tian *et al.*, 2014)
- Adaptive lasso (Zhou, 2006)
- SCAD (Fan and Li, 2001)
- MCP (Zhang 2010)

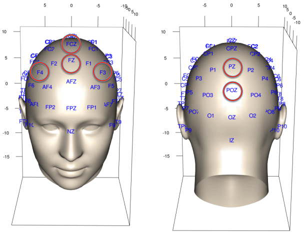
Application: EMBARC data (1D functions)

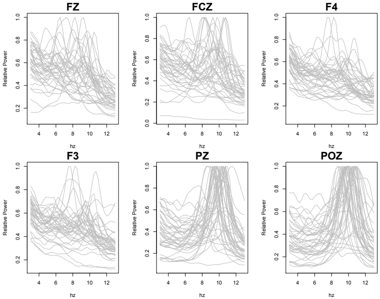
Treatment: Placebo ($n_P = 58$), Sertraline ($n_S = 54$)

Scalars:

Age at MDE	bp1831	bp1822	Sex
Dur of MDE	bp1844	Chronicity	Age at Eval.
bp 1827	bp1837	Severity	Years Educ.

Functions:

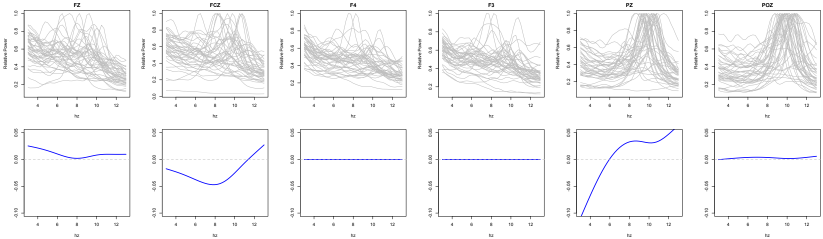




Response: S-S slope from LMEM with time, site, time×site

Results: Contrast coefficient estimates

Variable	$\hat{\gamma}$	Variable	$\hat{\gamma}$
Treatment	-0.001	bp1822	-0.009
Age at MDE	-0.021	Chron.	0.012
Dur of MDE	0.043	Severity	-0.012
bp1827	0	Sex	-0.066
bp1831	0	Age at Eval.	-0.055
bp1844	0.005	Years Ed.	0.074
bp1837	-0.029		

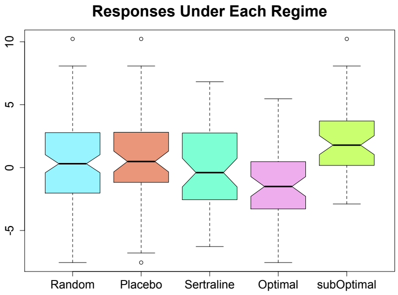


Results: In-sample expected response

Optimal	**Received** Placebo	Sertraline
Placebo	25	22
Sertraline	33	32

Estimated mean change in HAMD

Random	$0.038 \times 8 =$	0.304
Placebo	$0.078 \times 8 =$	0.624
Sertraline	$-0.005 \times 8 =$	-0.040
Optimal	$-0.171 \times 8 =$	-1.368
subOptimal	$0.255 \times 8 =$	2.040



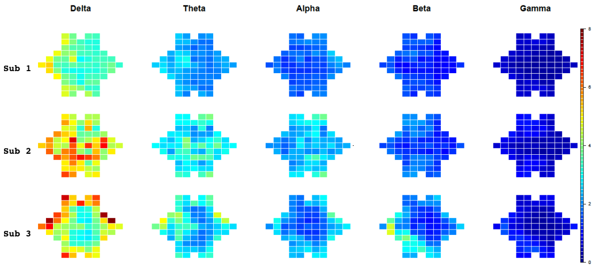
Application: EMBARC data (2D functions)

Treatment: Placebo ($n_P = 58$), Sertraline ($n_S = 54$)

Scalars:

Age at MDE	bp1831	bp1822	Sex
Dur of MDE	bp1844	Chronicity	Age at Eval.
bp 1827	bp1837	Severity	Years Educ.

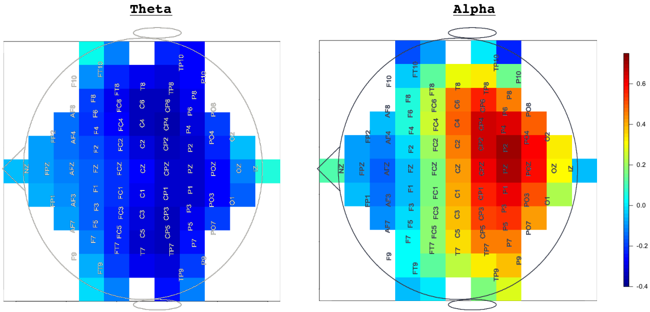
Functions:



Response: S-S slope from LMEM with time, site, time×site

Results: Contrast coefficient estimates

Variable	$\hat{\gamma}$	Variable	$\hat{\gamma}$
Treatment	-0.158	bp1822	0
Age at MDE	0	Chron.	0
Dur of MDE	0	Severity	0
bp1827	0	Sex	0
bp1831	0	Age at Eval.	0
bp1844	0	Years Ed.	0
bp1837	0		

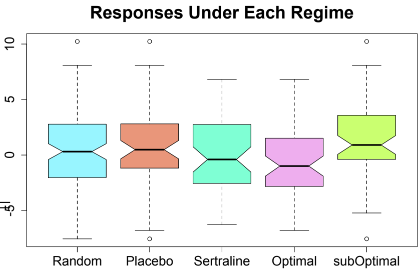


Results: In-sample expected response

	Received	
Optimal	Placebo	Sertraline
Placebo	15	12
Sertraline	33	42

Estimated Mean Change in HAMD

Random	$0.038 \times 8 =$	0.304
Placebo	$0.078 \times 8 =$	0.624
Sertraline	$-0.005 \times 8 =$	-0.040
Optimal	$-0.095 \times 8 =$	-0.760
subOptimal	$0.176 \times 8 =$	1.408



Generated Effect Modifiers

Redefining the treatment effect as $A = 0$ or 1, the linear model with a single predictor W is

$$Y = \nu_0 + \nu_1 A + \nu_2 W + \nu_3 (AW) + \epsilon$$

W is called a "treatment effect modifier" if $\nu_3 \neq 0$.

For assigning treatment, the only important term is the interaction term.

A "generated effect modifier" (GEM) is a linear combination of the available predictors $W = \boldsymbol{\gamma}^\intercal \boldsymbol{Z} + \sum_{\ell=1}^{p} \int \omega_\ell(s) X_\ell(s) ds$.

There are several criteria by which $\boldsymbol{\gamma}$ and $\omega_1, \ldots, \omega_p$ may be chosen (Petkova, *et al.*, 2016).

Nonparametric generated effect modifiers

Single index model:
$E[Y|A=1] - E[Y|A=0] = \nu_1 A + h(AW) + \epsilon$, where h is unspecified but constrained to be smooth and again,

$$W = \boldsymbol{\gamma}^\intercal \boldsymbol{Z} + \sum_{\ell=1}^{p} \int \omega_\ell(s) X_\ell(s) ds. \tag{1}$$

We fit this model by expressing h in terms of B-splines, applying a smoothness penalty and iterating between estimation of the parameters in (1) and the coefficients of h (Park, *et al.*, 2016).

Multiple index model:

$$E[Y|A=1] - E[Y_A=0] = \nu_1 A + h_1(AW_1) + \ldots + h_r(AW_r) + \epsilon,$$

Distance based methods

Define $m_a(\boldsymbol{x}) = E[Y|\boldsymbol{X} = \boldsymbol{x}, A = a]$ One way to estimate $m_a(x)$ nonparametrically is with the generalization of the Nadaraya-Watson estimator (Ferraty and Vieu, 2006)

$$\hat{m}_a(x) = \frac{\sum_{i=1}^n K(d(x, X_i)/h) 1(A_i = a) Y_i}{\sum_{i=1}^n K(d(x, X_i)/h) 1(A_i = a)}$$

- K is a kernel
- h is a bandwidth
- d is a *semi-metric*

Potential semi-metrics:

- Euclidean
- PCA-based
- wavelet-based

Discussion

- Difficult to do well when n is "moderate"
- Inference on estimated coefficient functions
- Dynamic treatments
- Accounting for side effects
- More than two treatments

References

Ciarleglio A, Petkova E, Ogden RT, and Tarpey, T (2015). Treatment decisions based on scalar and functional covariates. *Biometrics* **71**: 884–894.

Fan J and Li R (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association* **96**: 1348–1360.

Gertheiss J, Maity A, and Staicu A-M (2013). Variable selection in generalized functional linear models. *Stat* **2**:86–101.

Murphy S (2003). Optimal dynamic treatment regimes (with discussion). *Journal of the Royal Statistical Society, Series B* **65**: 331–336.

Goldsmith J,, Bobb J, Crainiceanu C, and Reich, D (2011). Penalized functional regression. *Journal of Computational and Graphical Statistics* **20**:830–851.

Park H, Tarpey T, Petkova E, and Ogden RT (2016). Nonparametric generated effect modifier regression. In preparation.

Petkova E, Tarpey T., Su Z., and Ogden RT. (2016). Generated effect modifiers (GEMs) in randomized clinical trials. Submitted for publication.

Tian L, Alizadeh A, Gentles A, and Tibshirani, R. (2014) A simple method for estimating interactions between a treatment and a large number of covariates. *Journal of the American Statistical Association* **508**: 1517–1532.

Tibshirani R (1996). Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society, Series B* **58**: 267–288.

Yuan M and Lin Y (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society, Series B* **68**: 49–67.

Zhang, C-H (2010). Nearly unbiased variable selection under minimax concave penalty. *Annals of Statistics* **38**: 894–942.