Mixed effect models to compare dynamic treatment regimens with SMART data

Brook Luers
Department of Statistics, University of Michigan
luers@umich.edu
brookluers.com

ENAR
March 26, 2019

Summary

Mixed models can be used to make causal comparisons of dynamic treatment regimens using longitudinal data from a SMART.

SMART: multi-stage randomized trial; enables causal comparison of dynamic treatment regimens (sequences of treatment decisions)

Outline

- **–** Background: DTRs, SMARTs

- **–** A mixed model for longitudinal SMARTs
    - ○ estimation: weighted, pseudo-likelihood
    - ○ inverse probability weights; but they are random
    - ○ mean estimator is robust to misspecified random effects
    - ○ random effects prediction

- **–** Data analysis using example SMART in autism

Background: DTRs, SMARTs

What is a dynamic treatment regimen (DTR)?

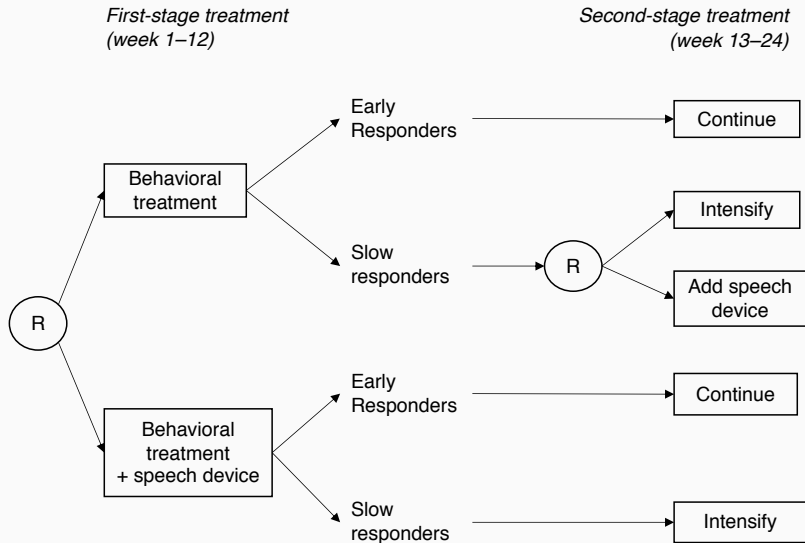**A sequence of rules for deciding how to provide treatment over time**
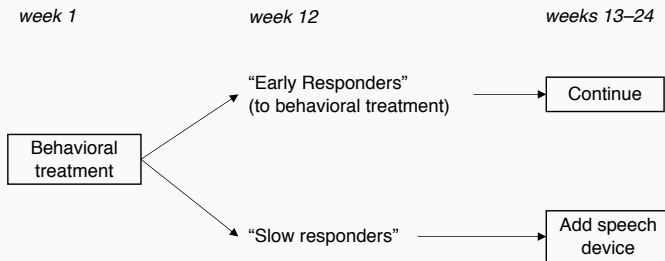
Kasari et al. [2014]

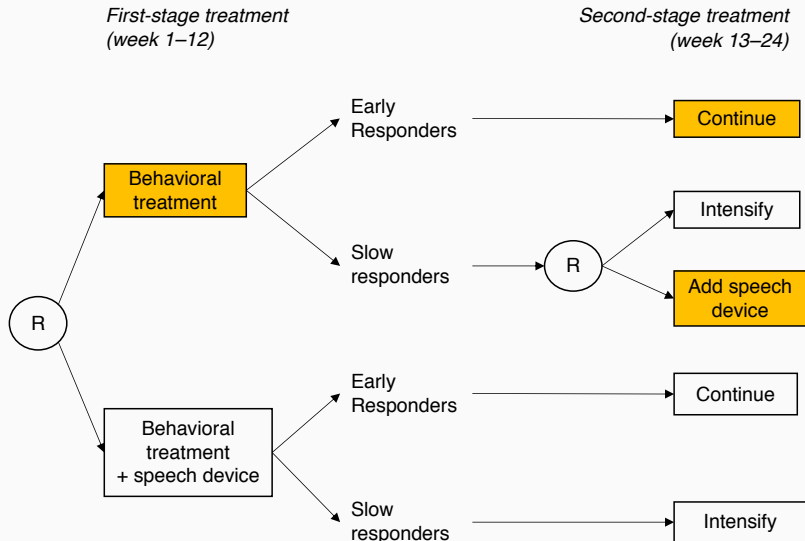Given current patient information, what treatment is provided next?

One treatment regimen, multiple treatment sequences

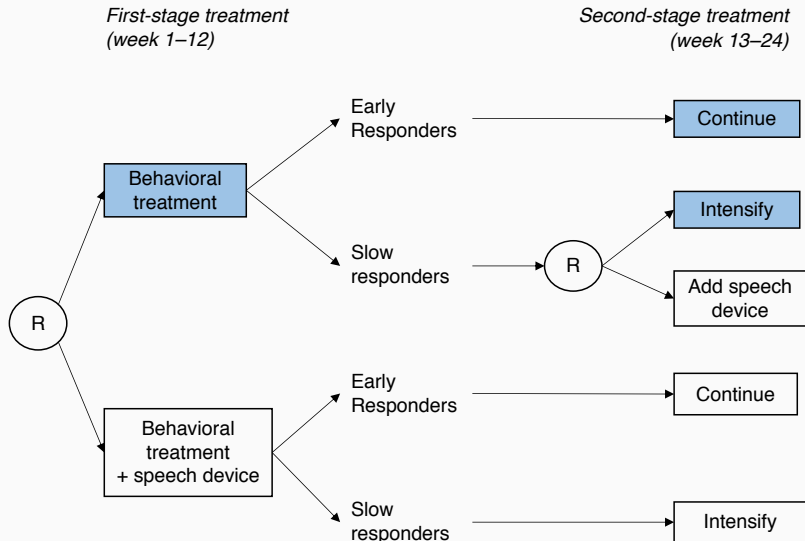# SMART: a trial design for comparing DTRs

SMART: a trial design for comparing DTRs
Three DTRs are embedded in this trial

First-stage treatment
(week 1–12)

Second-stage treatment
(week 13–24)

Early Responders → Continue

Behavioral treatment

Slow responders → R → Intensify

Add speech device

Early Responders → Continue

Behavioral treatment + speech device
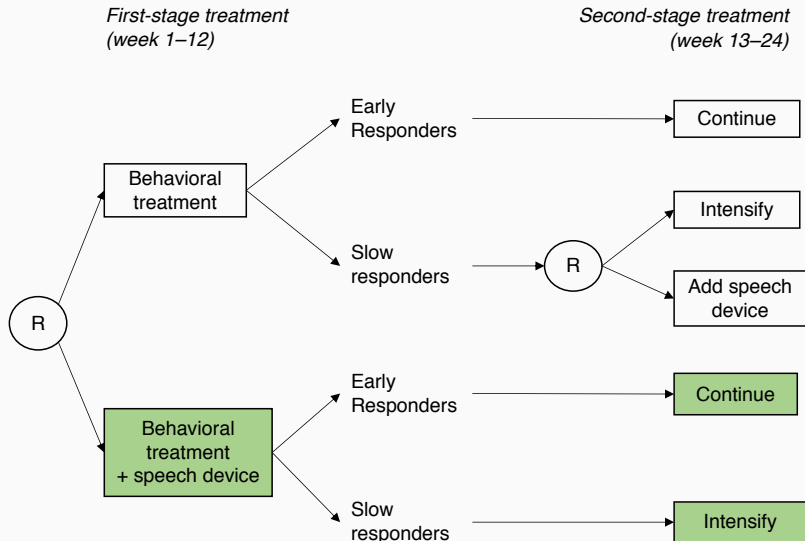
Slow responders → Intensify

SMART: a trial design for comparing DTRs

Three DTRs are embedded in this trial
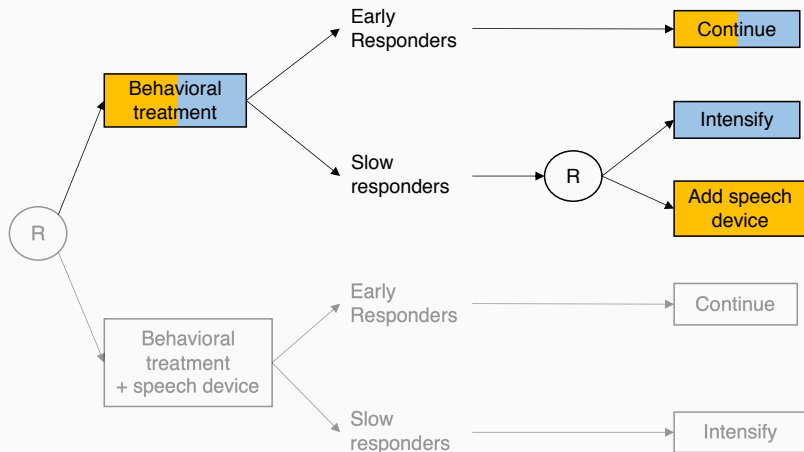
SMART: a trial design for comparing DTRs
Three DTRs are embedded in this trial

A treatment sequence observable under two DTRs

Two DTRs start with behavioral treatment

…both assign the same second-stage treatment to early responders

Scientific aim: causal comparisons of DTRs

Identify the best dynamic treatment regimen using longitudinal data from a SMART.

Other scientific aims are possible
- e.g. Which first-stage treatment is best?

Statistical framework
for making causal comparisons among DTRs

$Y_i^{(a_1, a_2)}$: vector of longitudinal potential outcomes under (a_1, a_2)

(a_1, a_2): fixed treatment regimen, not a random variable

Compare DTRs by estimating, for each (a_1, a_2),

$$\mathbb{E}\left(Y_i^{(a_1, a_2)} \mid L_i \right),$$

the mean outcome had the entire population followed (a_1, a_2).

Marginal over $R_i^{(a_1)}$: "slow/early responder"

$L_i =$ vector of baseline covariates

Statistical framework: modeling assumptions

Focus on linear models for continuous $Y_i^{(a_1,a_2)}$

$$\mathbb{E}\left(Y_i^{(a_1,a_2)} \mid L_i\right) = X_i^{(a_1,a_2)}\beta$$

$X_i^{(a_1,a_2)}$: $n_i \times p$ design matrix for the ith participant
with n_i observation times $t_{i1}, \ldots, t_{in_i}$

a function of $t_{ij}, (a_1, a_2),$ and L_i (baseline covariates)

Causal comparisons, e.g.

$$\mathbb{E}\left(Y_i^{(a_1,a_2)} \mid L_i\right) - \mathbb{E}\left(Y_i^{(b_1,b_2)} \mid L_i\right)$$

are functions of β.

A mixed model for longitudinal SMARTs

The model

$$Y_i^{(a_1,a_2)} = X_i^{(a_1,a_2)}\beta + Z_i b_i + \epsilon_i$$

$Z_i : n_i \times q$ random effects design matrix

$b_i : q$-dimensional random effects vector

$\epsilon_i \sim N(0, \sigma^2 I_{n_i})$ is independent of $b_i \sim N(0, G)$

i.e.

$$Y_i^{(a_1,a_2)} \mid L_i, b_i \sim N\left(X_i^{(a_1,a_2)}\beta + Z_i b_i, \sigma^2 I_{n_i}\right)$$

$$Y_i^{(a_1,a_2)} \mid L_i \sim N\left(X_i^{(a_1,a_2)}\beta, Z_i G Z_i^{\mathsf{T}} + \sigma^2 I_{n_i}\right)$$

Why mixed models?

- **–** Familiar tool for longitudinal data analysis, not yet available for SMARTs

- **–** Flexible parametrization of $\mathrm{Var}\left(Y_i^{(a_1, a_2)} \mid L_i \right)$
  - ○ function of covariates
  - ○ parsimonious: # parameters does not depend on n_i
  - ○ potential efficiency gains

- **–** Assess subject-to-subject variation
  - ○ subject-specific trajectories via random effects prediction ("BLUP")
  - ○ separate variance parameters for subject-level variation

Model parameters

$$Y_i^{(a_1, a_2)} = X_i^{(a_1, a_2)} \beta + Z_i b_i + \epsilon_i$$

$$\implies V_i := \mathrm{Var}\left(Y_i^{(a_1, a_2)} \mid L_i \right) = Z_i G Z_i^\mathsf{T} + \sigma^2 I_{n_i}$$

Parameters:

– β: p-dimensional vector of mean parameters

– α: vector of unique parameters in $V_i = V_i(\alpha)$.

Notation for observed data

Random treatment assignments: A_{1i}, A_{2i}

By design of the SMART,

$A_{1i} = a_1 \in \{1, -1\}$ with prob. $\mathbb{P}(A_{1i} = a_1)$

$A_{2i} = a_2 \in \{1, -1\}$ with prob. $\mathbb{P}(A_{2i} = a_2 \mid A_{1i}, R_i = 0)$
(only non-responders are randomized twice)

Denote

$R_i \in \{0, 1\}$: observed response status

Y_i : observed longitudinal outcome

$i = 1, \dots, N$ participants

Estimation: causal assumptions

We do not observe the potential outcomes

$$Y_i^{(a_1, a_2)},\ R_i^{(a_1)} \text{ (binary response status)}$$

for all (a_1, a_2).

So we assume

$$R_i = \sum_{a_1} \mathbb{1}_{[A_{1i} = a_1]} R_i^{(a_1)}$$

$$Y_i = R_i Y_i^{(A_{1i})} + (1 - R_i) Y_i^{(A_{1i}, A_{2i})}$$

$$\left. \begin{array}{c} Y_i^{(a_1, a_2)} \perp\!\!\!\perp A_{1i} \\ R_i^{(a_1)} \perp\!\!\!\perp A_{1i} \\ Y_i^{(a_1, a_2)} \perp\!\!\!\perp A_{2i} \mid A_{1i}, R_i \end{array} \right\} \text{ satisfied by design}$$

Estimation: weighted, pseudo-likelihood

If we observed $Y_i^{(a_1, a_2)}$ for all (a_1, a_2),
we could compute MLEs using likelihood for $Y_i^{(a_1, a_2)}$

Since we only observe Y_i, compute

$$\hat{\alpha}, \hat{\beta} = \arg \max_{\alpha, \beta} -\frac{1}{2} \sum_i \sum_{a_1, a_2} \tilde{W}_i(A_{1i}, R_i, A_{2i}, a_1, a_2) f_i(\alpha, \beta, a_1, a_2)$$

$$f_i(\alpha, \beta, a_1, a_2) := \log \det(V_i(\alpha)) \\ + (Y_i - X_i^{(a_1, a_2)} \beta)^\intercal V_i(\alpha)^{-1} (Y_i - X_i^{(a_1, a_2)} \beta)$$

Estimation: weighted, pseudo-likelihood

Can solve for $\hat{\beta}$

$$\hat{\beta}(\alpha) = \left(\sum_i \sum_{a_1, a_2} \tilde{W}_i(A_{1i}, R_i, A_{2i}, a_1, a_2) X_i^{(a_1, a_2)\intercal} V_i(\alpha)^{-1} X_i^{(a_1, a_2)} \right)^{-1}$$

$$\left(\sum_i \sum_{a_1, a_2} \tilde{W}_i(A_{1i}, R_i, A_{2i}, a_1, a_2) X_i^{(a_1, a_2)\intercal} V_i(\alpha)^{-1} Y_i \right)$$

In practice,

$$\hat{\alpha} = \arg \max_{\alpha} -\frac{1}{2} \sum_i \sum_{a_1, a_2} \tilde{W}_i(A_{1i}, R_i, A_{2i}, a_1, a_2) f_i(\alpha, \hat{\beta}(\alpha), a_1, a_2)$$

$$\hat{\beta} = \hat{\beta}(\hat{\alpha})$$

Weights: a function of $(A_{1i}, R_i, A_{2i}, a_1, a_2)$

The function is known by design, but $\tilde{W}_i$ is a random variable

$$\tilde{W}_i(A_{1i}, R_i, A_{2i}, a_1, a_2) = W_i(A_{1i}, R_i, A_{2i}, a_1, a_2) C_i(A_{1i}, R_i, A_{2i}, a_1, a_2)$$

$$W_i(A_{1i}, R_i, A_{2i}, a_1, a_2)$$
$$:= \frac{1}{\mathbb{P}(A_{1i} = a_1)} \left(R_i + \frac{(1 - R_i)}{\mathbb{P}(A_{2i} = a_2 \mid A_{1i}, R_i = 0)} \right)$$

$$C_i(A_{1i}, R_i, A_{2i}, a_1, a_2)$$
$$:= \begin{cases} 1 & \text{if } A_{1i}, R_i, A_{2i} \text{ observable under } (a_1, a_2) \\ 0 & \text{otherwise} \end{cases}$$
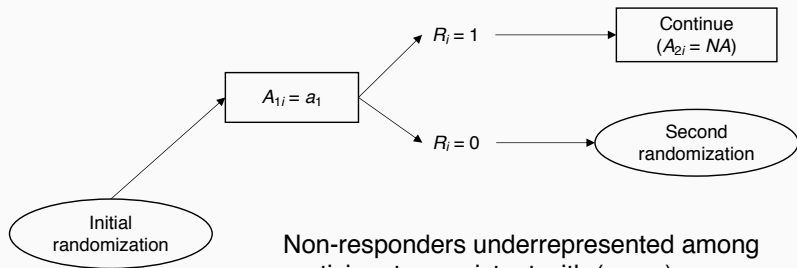
Weights: a function of $(A_{1i}, R_i, A_{2i}, a_1, a_2)$

Suppose $a_1, a_2 \in \{1, -1\}$, and

$$\mathbb{P}(A_{1i} = a_1) = 0.5, \qquad \mathbb{P}(A_{2i} = a_2 \mid A_{1i}, R_i = 0) = 0.5.$$

For the fixed regimen (a_1, a_2),

$$W_i(A_{1i}, R_i, A_{2i}, a_1, a_2) = \begin{cases} 4 & \text{if } R_i = 0 \\ 2 & \text{if } R_i = 1 \end{cases}$$

Non-responders underrepresented among participants consistent with (a_1, a_2)

Properties of $\hat{\beta}$

consistent, asymptotically Gaussian even if V_i misspecified

i.e. random effects $Z_i b_i$ can be misspecified

Estimate of $\mathrm{Var}\left(\hat{\beta}\right)$: $\frac{1}{N}\hat{J}^{-1}\hat{I}\hat{J}^{-1}$

$$\hat{J} = \frac{1}{N} \sum_i \sum_{a_1, a_2} \tilde{W}_i(A_{1i}, R_i, A_{2i}, a_1, a_2) X_i^{(a_1, a_2)\intercal} \hat{V}_i(\hat{\alpha})^{-1} X_i^{(a_1, a_2)}$$

$$\hat{I} = \frac{1}{N} \sum_i \hat{U}_i \hat{U}_i^{\intercal}$$

$$\hat{U}_i = \sum_{a_1, a_2} \tilde{W}_i(A_{1i}, R_i, A_{2i}, a_1, a_2) X_i^{(a_1, a_2)\intercal} \hat{V}_i(\hat{\alpha})^{-1} (Y_i - X_i^{(a_1, a_2)} \hat{\beta}).$$

Random effects prediction

Compute a prediction based on the weighted pseudo-likelihood:

$$
\begin{aligned}
\hat{b}_i &:= \hat{b}_i(\hat{\alpha}, \hat{\beta}) \\
&= \frac{\sum_{a_1, a_2} \tilde{W}_i(A_{1i}, R_i, A_{2i}, a_1, a_2) \hat{G} Z_i^{\mathsf{T}} \hat{V}_i^{-1} (Y_i - X_i^{(a_1, a_2)} \hat{\beta})}{\sum_{a_1, a_2} \tilde{W}_i(A_{1i}, R_i, A_{2i}, a_1, a_2)}
\end{aligned}
$$

Motivated by

$$
\mathbb{E}\left(b_i \mid Y_i^{(a_1, a_2)}, X_i^{(a_1, a_2)} \right) = G Z_i^{\mathsf{T}} V_i^{-1} (Y_i^{(a_1, a_2)} - X_i^{(a_1, a_2)} \beta)
$$

Other statistical methods for SMARTs
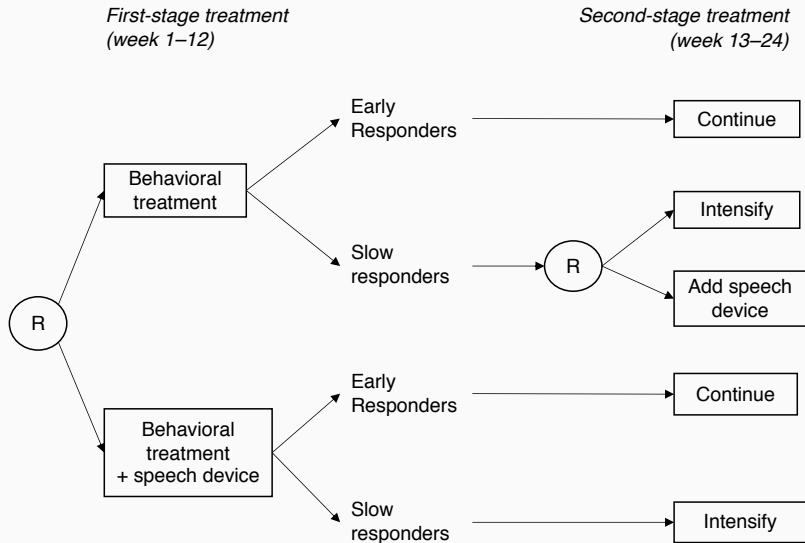
Longitudinal SMARTs: weighted, GEE-like estimators

- **–** Lu et al. [2016], Li [2017], Seewald et al. [2018], Dziak et al. [2019]

- **–** Working model for variance-covariance, sandwich standard errors

General references

- **–** Murphy [2005], Nahum-Shani et al. [2012], Almirall et al. [2014]

Data analysis: example SMART in autism

Recall the autism SMART design

A piecewise linear model with random intercept

Three DTRs: $(1, 1), (1, -1), (-1, 1)$

Y_{ij} : number of "socially communicative utterances"

$t_{ij} \in \{0, 12, 24, 36\}$ months, $\quad N = 61$

$$
\begin{aligned}
\mathbb{E}\left(Y_{ij}^{(a_1, a_2)} \mid X_{ij}^{(a_1, a_2)}, b_i \right) = \beta_0 &+ t_j^{[0,12]} \left(\beta_1 + \beta_2 a_1 \right) \\
&+ t_j^{(12,36]} \left(\beta_3 + \beta_4 a_1 + \beta_5 \mathbb{1}_{[a_1=1]} a_2 \right) \\
&+ \beta_6 \text{age}_i + b_i,
\end{aligned}
$$

where

$$
t_j^{[0,12]} = \left(t_j \mathbb{1}_{[t_j \leq 12]} + 12 \mathbb{1}_{[t_j > 12]} \right)
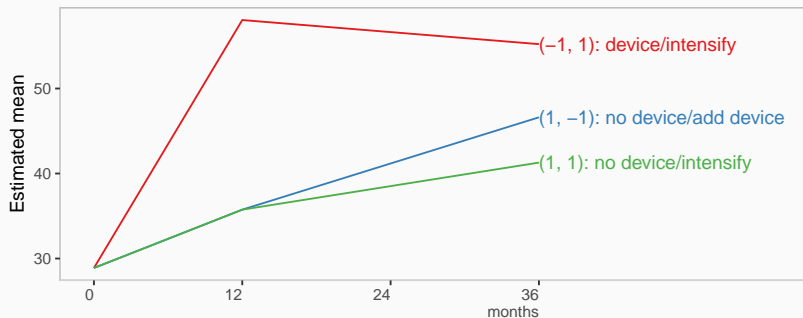$$
$$
t_j^{(12,36]} = (t_j - 12) \mathbb{1}_{[t_j > 12]}
$$
$$
\text{age}_i = \text{age at baseline}
$$
$$
b_i \sim N(0, \sigma_b^2) : \text{ random intercept}
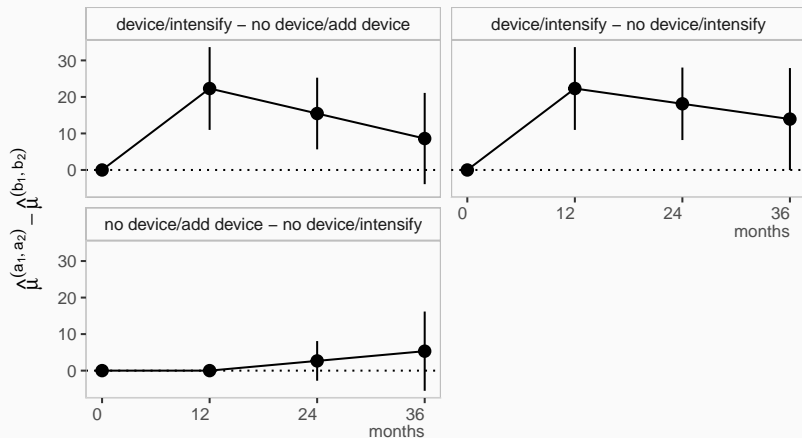$$

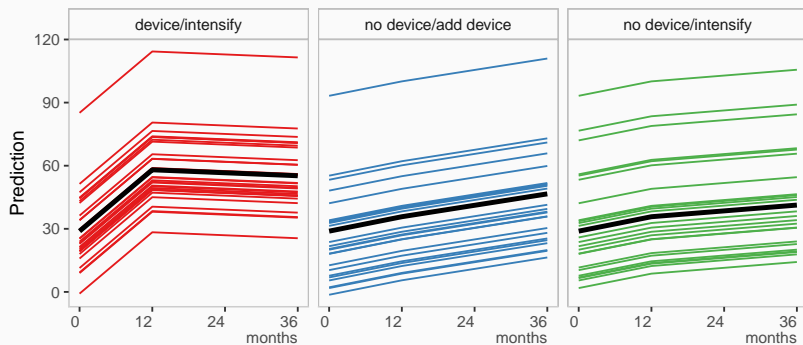# Primary aim: compare DTRs

Estimated mean under each DTR

Primary aim: compare DTRs

Pairwise contrasts with 90% confidence intervals

Person-specific predictions

$$X_i^{(a_1,a_2)}\hat{\beta} + Z_i\hat{b}_i$$

$\hat{\sigma} = 21$

$\hat{\sigma}_b = 24$

Next steps

– Prove $\hat{b}_i$ has minimum MSE
– Software implementation
  ○ currently using "tricks" valid only with integer weights
– Scenarios when Z_i can include (a_1, a_2), so that V_i depends on (a_1, a_2)
– Advantages of mixed models for missing data in SMARTs

Thank you

Download slides: `brookluers.com/research`

Coauthors

Min Qian (Columbia University)

Inbal Nahum-Shani (University of Michigan)

Connie Kasari (University of California, Los Angeles)

Daniel Almirall (University of Michigan)

References

D. Almirall, I. Nahum-Shani, N. Sherwood, and S.A. Murphy. Introduction to SMART designs for the development of adaptive interventions: with application to weight loss research. *Translational Behavioral Medicine*, 4(3):260–274, 2014.

John J. Dziak, Jamie Yap, Daniel Almirall, James R. McKay, Kevin G. Lynch, and Inbal Nahum-Shani. A data analysis method for using longitudinal binary outcome data from a SMART to compare adaptive interventions. *Multivariate Behavioral Research*, pages 1–24, 2019. URL https://doi.org/10.1080/00273171.2018.1558042.

Connie Kasari, Ann Kaiser, Kelly Goods, Jennifer Nietfeld, Pamela Mathy, Rebecca Landa, Susan Murphy, and Daniel Almirall. Communication interventions for minimally verbal children with autism: A sequential multiple assignment randomized trial. *Journal of the American Academy of Child & Adolescent Psychiatry*, 53(6): 635–646, 2014.

Zhiguo Li. Comparison of adaptive treatment strategies based on longitudinal outcomes in sequential multiple assignment randomized trials. *Statistics in Medicine*, 36(3):403–415, 2017.

Xi Lu, Inbal Nahum-Shani, Connie Kasari, Kevin G. Lynch, David W. Oslin, William E. Pelham, Gregory Fabiano, and Daniel Almirall. Comparing dynamic treatment regimes using repeated-measures outcomes: modeling considerations in smart studies. *Statistics in Medicine*, 35(10):1595–1615, 2016.

References (continued)

S.A. Murphy. An experimental design for the development of adaptive treatment strategies. *Statistics in Medicine*, 24(10):1455–1481, May 2005.

I. Nahum-Shani, M. Qian, D. Almirall, W.E. Pelham, B. Gnagy, G. Fabiano, J. Waxmonsky, J. Yu, and S.A. Murphy. Experimental design and primary data analysis methods for comparing adaptive interventions. *Psychological Methods*, 17: 457–477, 2012.

Nicholas J. Seewald, Kelley M. Kidwell, Inbal Nahum-Shani, Tianshuang Wu, James R. McKay, and Daniel Almirall. Sample size considerations for comparing dynamic treatment regimens in a sequential multiple-assignment randomized trial with a continuous longitudinal outcome. *arXiv e-prints*, art. arXiv:1810.13094, Oct 2018.