Q-Learning: A Data Analysis Method for Constructing Adaptive Interventions

Inbal Nahum-Shani
The Methodology Center, Penn State

Susan Murphy; Min Qian
Dept. of Statistics & Institute for Social Research, U of Michigan

William E. Pelham; Beth Gnagy; Greg Fabiano; Jim Waxmonsky; Jihnhee Yu
Center for Children and Families, SUNY Buffalo

Fixed vs. Adaptive Interventions

- Fixed Intervention strategies: “one size fits all”
  - The same dose or type of services are offered to all clients.
  - No adjustment over time.

- Adaptive interventions: sequential processes
  - The dose or type of services are individualized based on clients’ characteristics or clinical presentation.
  - Adjustment over time in response to ongoing performance.

Decision Rules Operationalize Adaptive Decision Making

- **Tailoring variables:** subjects characteristics and intermediate outcomes (e.g., response or adherence to past treatment).
- Link subjects' values on the tailoring variables with specific levels and types of intervention components
- Example (intervention for improving perceived social support):

First stage intervention = {social skill}

IF evaluation = {non-response}

THEN at Step t+1 apply decision {intensify first stage intervention}

ELSE IF evaluation = {response}

THEN at Step t+1 continue on present intervention

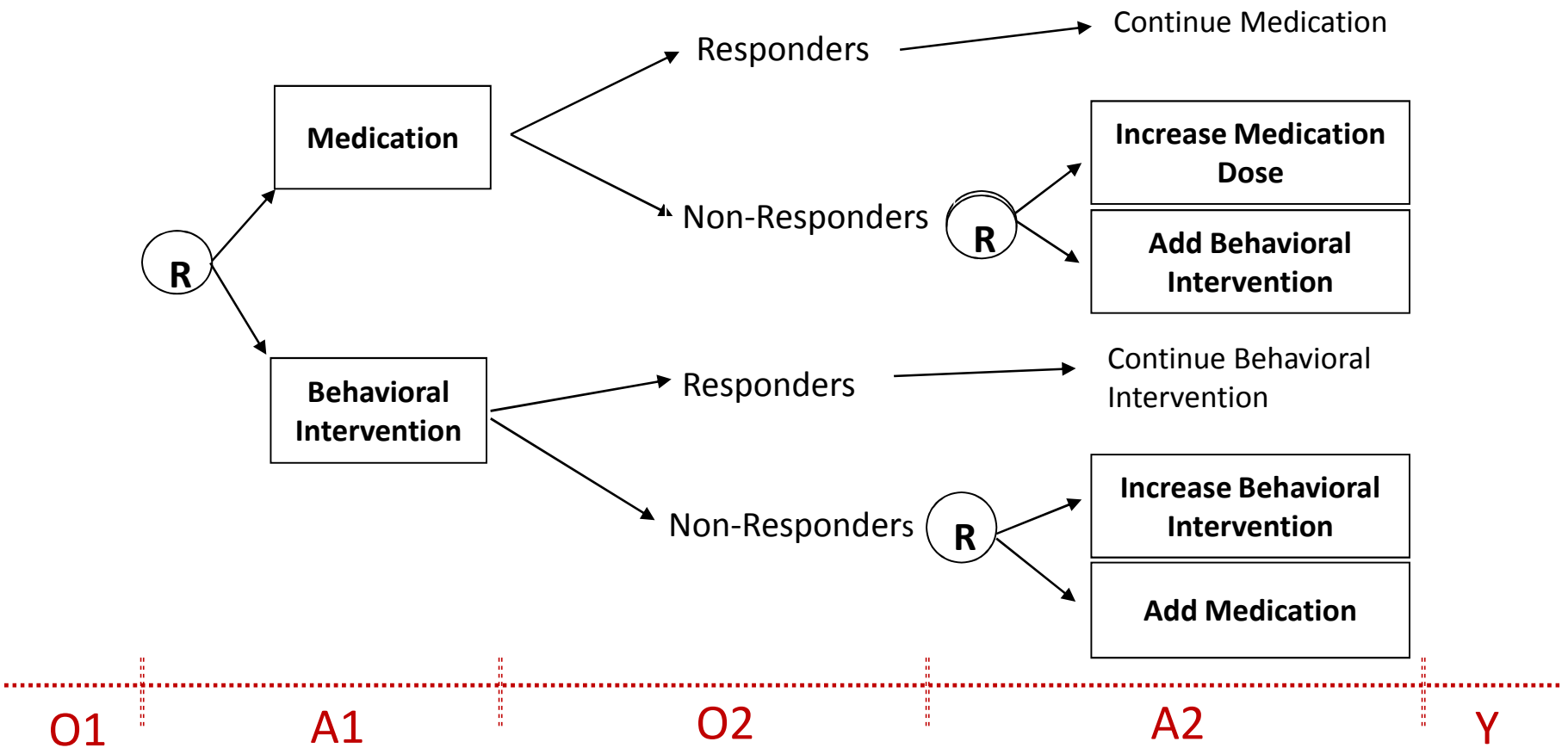
Building a High Quality Adaptive Intervention

- Selecting good decision rules.
- Research methods for finding the best sequence of decision rules are relatively new.
- Introducing Q-learning (Watkins, 1989) – a novel methodology that can be used for the construction of adaptive interventions from data.
- Example: "Adaptive Interventions for Children with Attention Deficit Hyperactivity Disorder (ADHD) study": *Center for Children and Families, SUNY at Buffalo:* **http://ccf.buffalo.edu/default.php**

Data for Constructing Adaptive Interventions

- Sequential Multiple Assignment Randomized Trial.
  - Two intervention stages.
  - Observable data {O1, A1, O2, A2, Y}.
  - A1: stage 1 intervention;
  - A2: stage 2 intervention;
    both binary (coded -1/1), 0.5 randomization probability.
  - O1: observations at the beginning of stage 1.
  - O2: observations at the beginning of stage 2.
  - Y→ primary outcome (after the end of stage 2)

'Adaptive Interventions for Children with ADHD' study—*simplified version*



- The only tailoring variable in the original experiment is response/non-response to initial intervention
- The experiment was not sized to detect other tailoring variables.
- We want to use the data to develop a more deeply tailored adaptive intervention.
- Q-learning is a secondary data analysis method

Measures

- **Medication prior to stage 1 intervention (O1):** whether (=1) or not (=0) medicated at school before stage 1 intervention.
- **First-stage intervention (A1):** 1=BMOD; -1=MED
- **Adherence to stage 1 intervention (O2):** 1 =high; 0=low.
- **Second-stage intervention (A2):** 1=Enhance; -1=Add.
- **Primary outcome (Y):** classroom performance at the end of the school year (range 1-5, the higher the better)

Q-learning (Watkins, 1989; Murphy, 2005)

- Popular method from computer science.
- Regression-based: one regression for each stage.
- Backwards induction: moving backwards in time from the last stage to the first stage.

Stage 2 Regression (non-responders only)

- We want to find out what is the best stage-2 intervention, given the child's history up until this stage.
- Regress Y on O1, A1, O2, A2, A1*A2, O2*A2
- Potential tailoring variables:
  - O2: adherence to stage 1 intervention
  - A1: stage 1 intervention
- We obtain:

$$\hat{\beta}_{20} + \hat{\beta}_{21}O_1 + + \hat{\beta}_{22}A_1 + \hat{\beta}_{23}A_1O_1 + \hat{\beta}_{24}O_2 + (\hat{\gamma}_{21} + \hat{\gamma}_{22}A_1 + \hat{\gamma}_{23}O_2)A_2$$

If $(\hat{\gamma}_{21} + \hat{\gamma}_{22}A_1 + \hat{\gamma}_{23}O_2) > 0$ then A2=1 is the best

If $(\hat{\gamma}_{21} + \hat{\gamma}_{22}A_1 + \hat{\gamma}_{23}O_2) < 0$ then A2=-1 is the best

Create a New Variable: $\hat{Y}$

- **Calculated based on stage 2 regression:**

  For non-responders:

$$\hat{Y} = \hat{\beta}_{20} + \hat{\beta}_{21}O_1 + \hat{\beta}_{22}A_1 + \hat{\beta}_{23}A_1O_1 + \hat{\beta}_{24}O_2 + max(\hat{\gamma}_{21} + \hat{\gamma}_{22}A_1 + \hat{\gamma}_{23}O_2)$$

  For responders: $\hat{Y} = Y$

- **$\hat{Y}$ is the:**
  - Estimated mean outcome as a function of variables that may include or be affected by stage 1 intervention; setting the stage 2 intervention equal to the "best" intervention.
  - The dependent variable in the stage 1 regression for children moving to stage 2.

Stage 1 Regression:

- We want to find out what is the best stage-1 intervention, given
  - the child's history up until this stage;
  - that we intend to take the best stage 2 intervention in the future.
- Regress $\hat{Y}$ on O1, A1, A1*O1
- Potential tailoring variable→ O1: Medication prior to stage 1.
- Obtain:

$$\hat{\beta}_{10} + \hat{\beta}_{11} O_1 + (\hat{\gamma}_{11} + \gamma_{12} \hat{O}_1) A_1$$

If $(\hat{\gamma}_{11} + \hat{\gamma}_{12} O_1) > 0$ then A1=1 the best stage 1 option

If $(\hat{\gamma}_{11} + \hat{\gamma}_{12} O_1) < 0$ then A1=-1 the best stage 1 option

Challenges for Inference

- The maximization operation $\max(\hat{\gamma}_{21} + \hat{\gamma}_{22}A_1 + \hat{\gamma}_{23}O_{22})$
- $\hat{\Upsilon}$ is a non smooth function of $\hat{\gamma}_{21}, \hat{\gamma}_{22}, \hat{\gamma}_{23}$ -- non-differentiable at $(\hat{\gamma}_{21} + \hat{\gamma}_{22}A_1 + \hat{\gamma}_{23}O_{22} = 0)$
- Since $\hat{\gamma}_{11}, \hat{\gamma}_{12}$ are functions of $\hat{\Upsilon}$, we need new methodologies for constructing confidence intervals for these estimates.
- Confidence intervals for $\hat{\gamma}_{11}, \hat{\gamma}_{12}$ using the soft thresholding operation (Chakraborty et al., 2009).

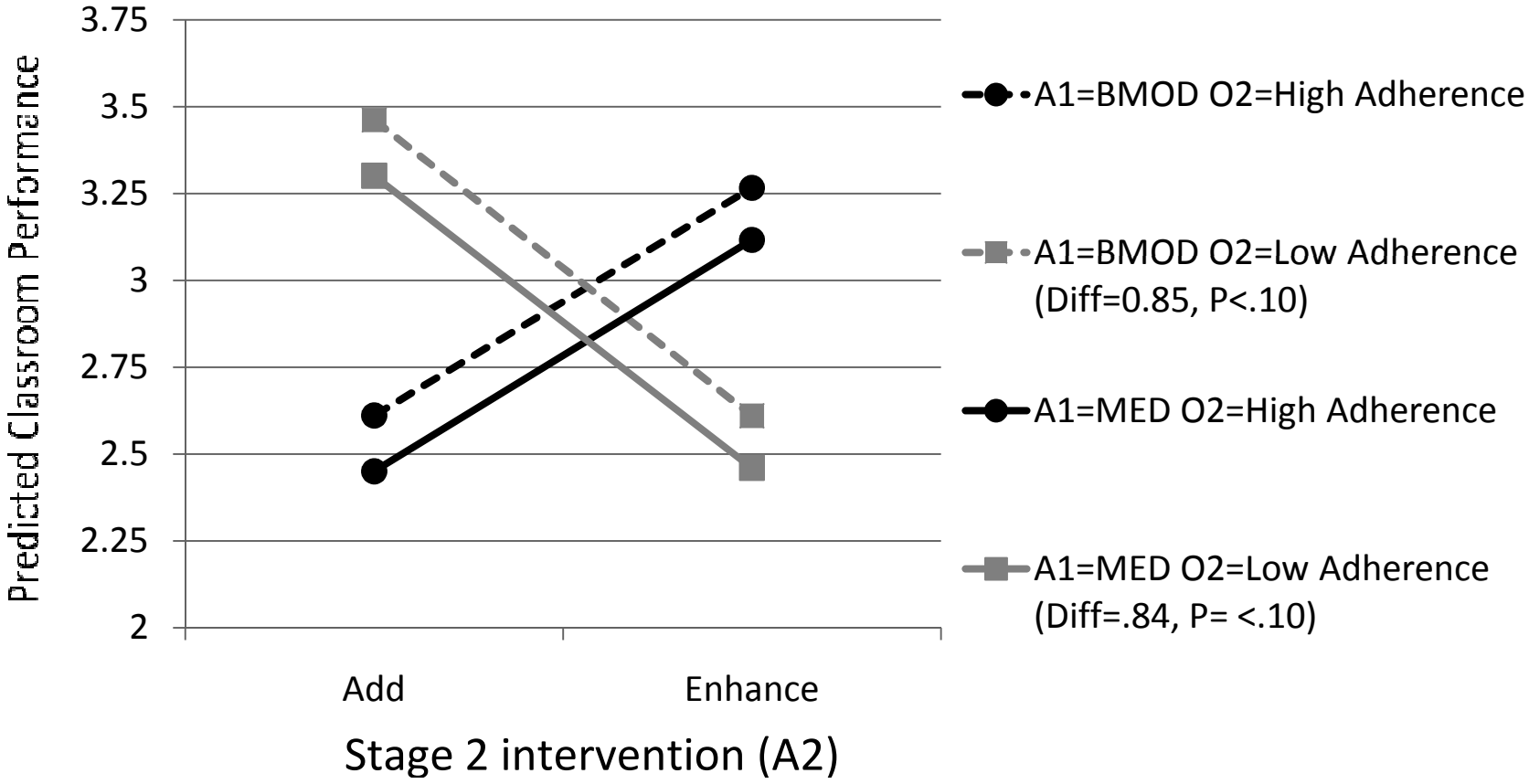
Results: Stage 2 Regression

A1 (stage 1 intervention) 1 = BMOD -1= MED	O2 (adherence) 1=high 0=low	Estimated $(\hat{\gamma}_{21} + \hat{\gamma}_{22}A_1 + \hat{\gamma}_{23}O_2)$	SE	Lower limit 90% CI	Upper limit 90% CI
-1	1	0.33	0.24	-0.07	0.73
-1	**0**	**-0.42**	**0.27**	**-0.87**	**0.02**
1	1	0.33	0.22	-0.04	0.70
1	**0**	**-0.43**	**0.25**	**-0.85**	**-0.01**

If Adherence is low, A2=-1 (Add) would maximize the term $(\hat{\gamma}_{21} + \hat{\gamma}_{22}A_1 + \hat{\gamma}_{23}O_2)A_2$

Regardless of the stage 1 intervention

Interaction Plot for Stage 2 Regression

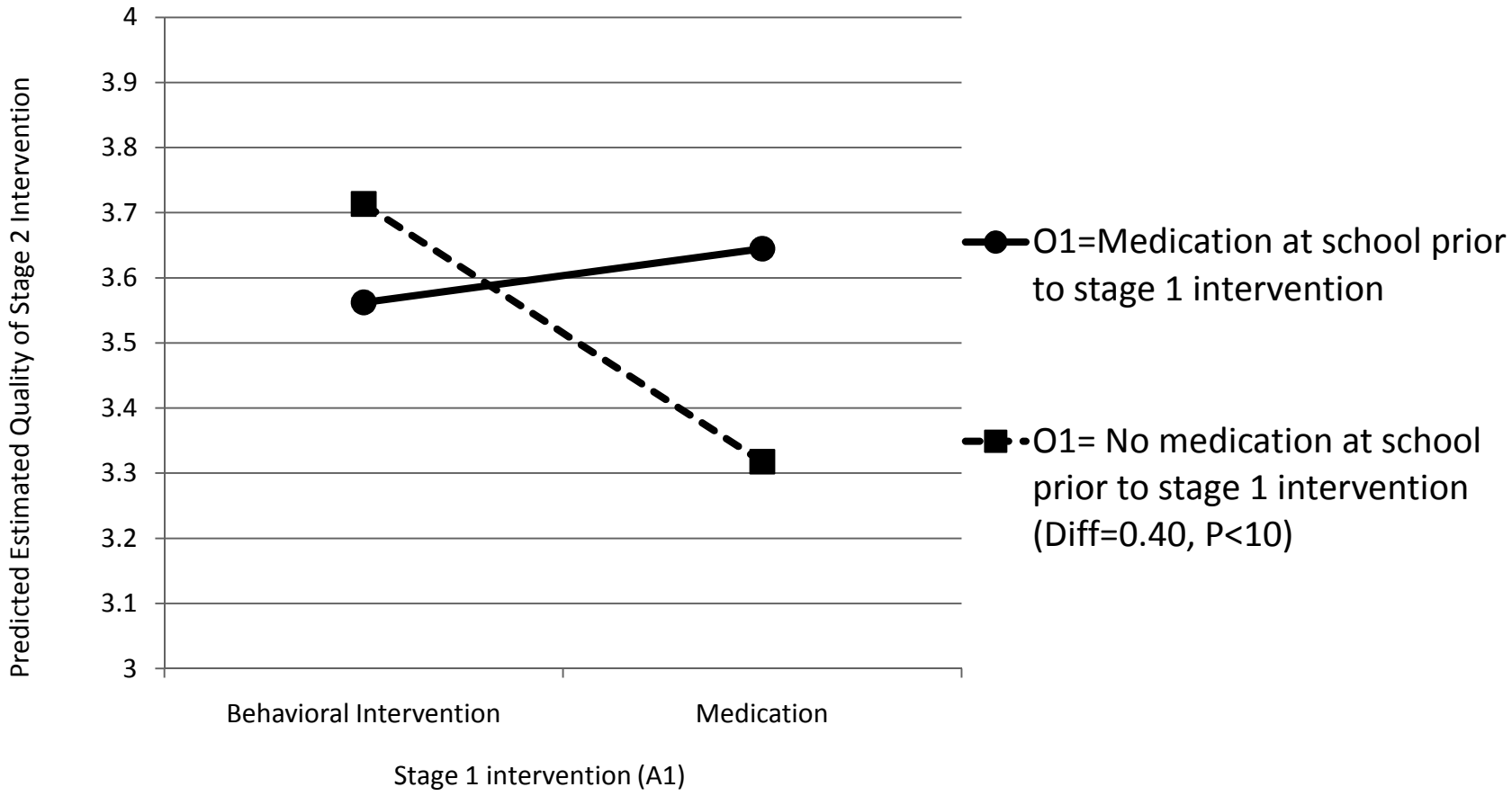


Results: Stage 1 Regression

O1 (medicated prior to stage 1 intervention) 1= yes 0= no	Estimated $(\hat{\gamma}_{11} + \hat{\gamma}_{12} O_1)$	SE	Lower limit 90% CI	Upper limit 90% CI
1	-0.04	0.11	-0.33	0.24
0	**0.20**	**0.06**	**0.002**	**0.38**

If medicated prior to stage 1, A1=1 (BMOD) maximizes the term $(\hat{\gamma}_{11} + \gamma_{12}\hat{O}_1)A_1$

Interaction Plot for Stage 1 Regression



Decision Rule:

IF *medication at school prior to first stage intervention={no}*

THEN *stage 1 intervention= {BMOD}.*

ELSE *stage 1 intervention= {MED} or {BMOD}.*

Then,

IF *response to stage 1 intervention ={inadequate}*

THEN IF *adherence to first stage intervention={low},*

THEN *stage 2 intervention= {ADD}.*

ELSE *stage 2 intervention={ADD} or {ENHANCE}.*

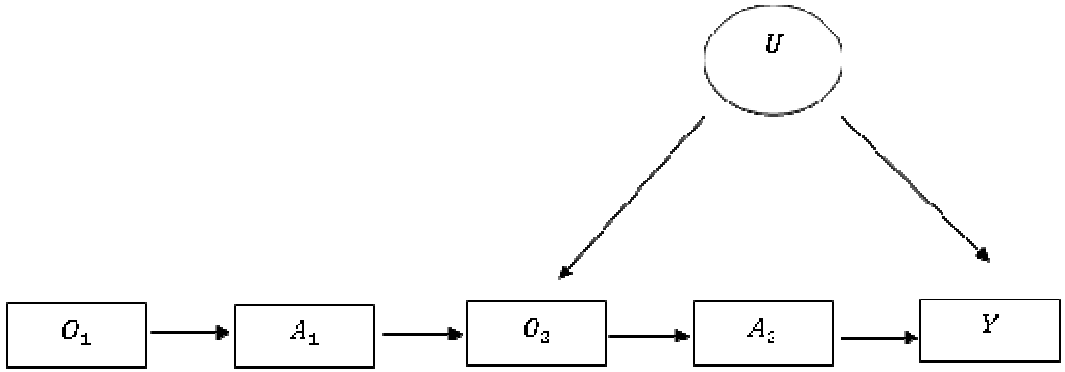
ELSE *continue first stage intervention.*

Advantages of Q-learning Approach

- Compared with a single regression-based approach

  $Y \sim \theta_0 + \theta_1 O_1 + \theta_2 A_1 + \theta_3 O_1 A_1 + \theta_4 O_2 + \theta_5 A_2 + \theta_6 A_1 A_2 + \theta_7 A_2 O_2$

- Reduces potential bias resulting from mediators of the relationship between the first stage intervention and the primary outcome.
- Reduces potential bias resulting from unmeasured causes (U) of both the tailoring variables and the primary outcome.



- This seminar can be found at:
  http://www.stat.lsa.umich.edu/~samurphy/Seminars/SPR2010.pdf

- e-Mail Inbal Nahum-Shani or Susan Murphy for questions: **iun2@psu.edu; samurphy@umich.edu**

- We acknowledge support for this project from the National Institutes of Health grant P50 DA10075, and the Institute for Education Sciences grant R324B060045.

References

- Chakraborty, B., Murphy, S.A., & Strecher, V. (2009). Inference for non-regular parameters in optimal dynamic treatment regimes. Statistical Methods in Medical Research, In Press.

- Collins, L.M., Murphy, S.A., & Bierman, K.A. (2004), A Conceptual Framework for Adaptive Preventive Interventions, Prevention Science, 5, 185-196.

- Murphy, S.A. (2005). An experimental design for the development of adaptive treatment strategies. Statistics in Medicine, 24, 455-1481.

- Murphy, S.A., Lynch, K.G., Oslin, D., Mckay, J.R. & TenHave, T. (2007). Developing adaptive treatment strategies in substance abuse research. Drug and Alcohol Dependence, 88s, s24-s30.

- Nahum-Shani et al., (2010). Q-Learning: A Data Analysis Method for Constructing Adaptive Interventions. Technical Report, the Methodology Center, Penn State University

Two general methods for constructing adaptive interventions

- **Modeling the system followed by constructing the intervention**
  - First estimate a multivariate distribution for $\{O1, A1, O2, \ldots, A_k, O_k+1\}$.
  - Second use (approximate) "dynamic programming" to construct the policy.
- **Modeling & constructing simultaneously**
  - Q-learning (Watkins, 1989), Temporal Difference Methods (Sutton & Barto, 1998), LSPI (Lagoudakis & Parr, 2003).