

Use of personalized Dynamic Treatment Regimes (DTRs) and Sequential Multiple Assignment Randomized Trials (SMARTs) in mental health studies

Ying LIU¹, Donglin ZENG², Yuanjia WANG^{1*}

Summary: Dynamic treatment regimens (DTRs) are sequential decision rules tailored at each point where a clinical decision is made based on each patient's time-varying characteristics and intermediate outcomes observed at earlier points in time. The complexity, patient heterogeneity, and chronicity of mental disorders call for learning optimal DTRs to dynamically adapt treatment to an individual's response over time. The Sequential Multiple Assignment Randomized Trial (SMARTs) design allows for estimating causal effects of DTRs. Modern statistical tools have been developed to optimize DTRs based on personalized variables and intermediate outcomes using rich data collected from SMARTs; these statistical methods can also be used to recommend tailoring variables for designing future SMART studies. This paper introduces DTRs and SMARTs using two examples in mental health studies, discusses two machine learning methods for estimating optimal DTR from SMARTs data, and demonstrates the performance of the statistical methods using simulated data.

Keywords: SMART; dynamic treatment regimes; personalized medicine; O-learning; Q-learning; double robust estimation

[*Shanghai Arch Psychiatry*. 2014; **26**(6): 376-383. doi: <http://dx.doi.org/10.11919/j.issn.1002-0829.214172>]

1. Dynamic Treatment Regimens (DTRs)

Sequential treatments, a sequence of interventions in which the treatment decisions are adapted to the time-varying clinical status of the patient, are useful in treating many complex chronic mental disorders. For instance, existing clinical literature reports on the potential benefit of behavioral or pharmacological interventions, but patients' heterogeneous responses to each modality of treatment may call for sequential, individualized treatments, especially in cases where the patient is non-responsive to monotherapy. Dynamic Treatment Regimes (DTRs) operationalize the sequential

process of medical decision making and closely reflect actual clinical practice. DTRs are sequential decision rules, tailored at each stage to patients' time-varying features and intermediate outcomes. They are also known as adaptive treatment strategies^[1], multi-stage treatment strategies,^[2,3] and treatment policies.^[4-6] Examples of clinical trials involving sequential treatments and DTRs in mental health include the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) trial for treating depression,^[7,8] the Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) trial for treating schizophrenia,^[9] Managing Alcoholism in People Who

¹ Department of Biostatistics, Mailman School of Public Health, Columbia University, New York, NY, United States

² Department of Biostatistics, University of North Carolina at Chapel Hill, United States

*correspondence: yuanjia.wang@columbia.edu

Do Not Respond to Naltrexone (EXTEND) for treating alcohol dependence,^[10] the Reinforcement-Based Treatment for Pregnant Drug Abusers (HOME III) trial,^[11] Adaptive Pharmacological and Behavioral Treatments for Children with Attention Deficit/Hyperactivity Disorder (ADHD) trial,^[12,13] and the Adaptive Autism Spectrum Disorder (ASD) Developmental and Augmented Intervention.^[14]

Compared to conventional interventions in which all patients in each arm of the trial are offered the same treatment with the same dosage, DTRs have several important advantages.^[15] (a) Treatment can be assigned to patients according to their personal features and, thus, maximize potential benefits. (b) If the effectiveness of an intervention changes over time, DTRs allow patients to be switched to other more promising treatments. (c) When there are comorbid conditions – as is often the case for mental disorders – DTRs can help decide which disorder should be treated primarily and when simultaneous treatment of multiple conditions is necessary. (d) When relapse occurs, DTRs can be used to make the optimal clinical decisions about resumption or alteration of the treatment strategy. (e) DTRs can be used to identify the lowest effective dose and, thus, minimize risk of adverse effects. And (f) the option of switching medications when using DTRs increases participant adherence during a clinical trial.

1.1 Sequential Multiple Assignment Randomized Trials (SMARTs)

Valid evaluations of the effectiveness of DTRs are based on the notion of potential outcomes, defined as the outcome of a subject had he followed a particular treatment regime, possibly different from the observed regime for the subject. Two assumptions are required to estimate the causal effect of a dynamic regime in this framework:^[16,17]

1. Stable unit treatment value assumption: A subject’s outcome is not influenced by other subjects’ treatment allocations.^[18]
2. No unmeasured confounders assumption: The newly assigned treatments are conditional on the history up to the current time but independent of potential future outcomes from the treatment.^[19]

Sequential Multiple Assignment Randomized Trials (SMARTs) are used to generate data that can be used to make causal inferences of specific treatment sequences and to compare the expected outcomes of different sequences. SMARTs randomize treatments at each critical decision point and, thus, provide the best possible data for making causal interpretations of the different DTRs. Below we use two examples to illustrate SMARTs.

1.2 Examples of SMARTs

We first illustrate a SMART using a trial for pregnant drug abusers^[11] as an example. The goal of the trial is to study how the intensity and scope of reinforcement based treatment (RBT) might be adapted to a pregnant woman’s progress in treatment. There are four types of RBT (in order of intensity of the intervention): abbreviated RBT (aRBT), reduced RBT (rRBT), treatment-as-usual RBT (tRBT), and enhanced RBT (eRBT). At the first stage of the trial, each participant is randomized to one of the two intermediate intensity interventions (tRBT or rRBT). In the second stage after two weeks, non-responders are re-randomized to continue the original intervention or use the next more intensive intervention, and responders are re-randomized to continue with the same intervention or to use the next less intensive intervention. This trial is illustrated in Figure 1.

A second example is a SMART study of treatments for children with attention deficit/hyperactivity disorder (ADHD).^[12,13] The study lasted for a school year (i.e., 8 months). Interventions include differing doses of methamphetamine and differing intensities of a behavioral modification intervention. As demonstrated in Figure 2, children were randomly assigned to begin with low-intensity behavioral modification or with low-

Figure 1. Design of adaptive reinforcement-based treatment for pregnant drug abusers

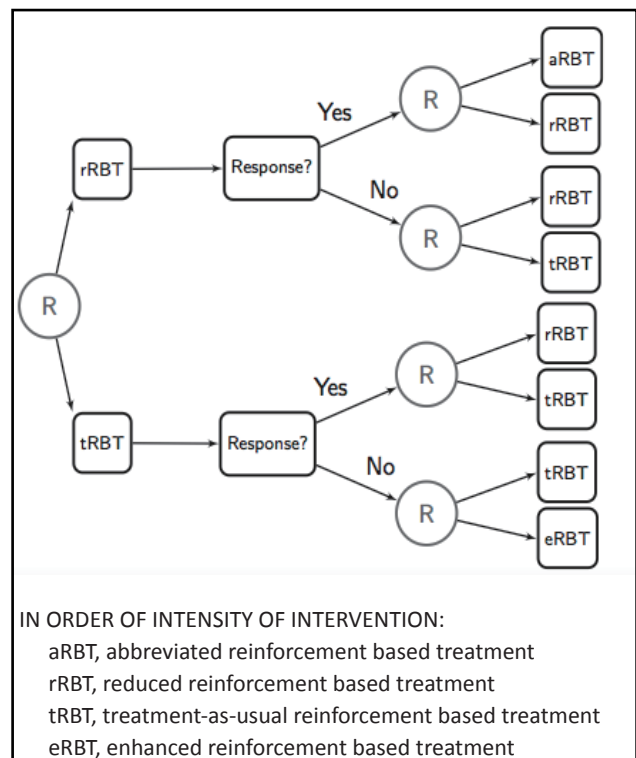
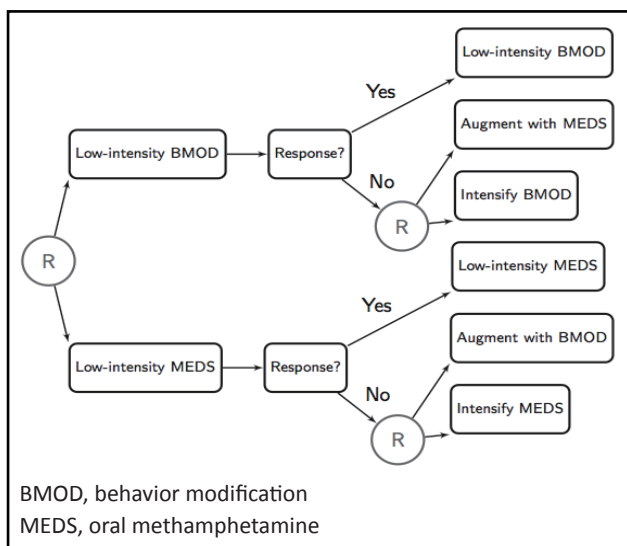


Figure 2. Design of trial on adaptive pharmacological and behavioral treatments for children with Attention Deficit/Hyperactivity Disorder (ADHD)



dose medication. This stage lasts for two months, after which the Impairment Rating Scale (IRS) [20] and the individualized List of Target Behaviors (LTB) measure [21] were used to assess each child’s response to initial treatment. Children who responded would continue to receive the initial low intensity treatment. Children who did not respond would be re-randomized to either intensify the initial treatment or to receive adjunctive treatment with the alternative type of treatment. The target outcome of the study was school performance score at the end of study. The primary aim of the study was to test the main effect of beginning with low-

dose medication versus beginning with low-intensity behavioral modification on the rate of non-response by the end of the school year. Secondary aims included (a) how baseline variables (e.g., prior medication history, ADHD impairment score, the comorbid presence of an oppositional defiance disorder [ODD] diagnosis, race, etc.) influence the choice of treatments in the first and second stage; and (b) differences in the effect between the four adaptive interventions embedded in the design.

2. Statistical analysis of data collected in SMARTs

2.1. Primary analysis

The primary aims of the above ADHD SMART study are listed in table 1. Comparisons of first-stage and second-stage intervention options can be made using a two-sample t-test for the two groups of patients. When comparing the imbedded adaptive intervention options in the last row of Table 1, it is necessary to compare weighted averages that adjust for the response rate of the initial treatment and randomization probabilities; inverse probability weighting [22] generates weighted averages that reflect the response rate in the population. A more detailed description of the primary analyses of SMART studies and specifically for this ADHD trial can be found in Nahum and Shani. [23] The sample size estimation for the primary analysis can be found in Otting. [24]

2.2. Finding the optimal DTR

Besides comparison of two initial regimes, it is also of interest to find the optimal regime (i.e., resulting in the best final outcome) using the rich data collected from SMARTs. One benefit of the optimal regime is that it

Table 1. Primary analysis questions and example in the ADHD study

Type of primary question	Example in the Attention Deficit/Hyperactivity Disorder (ADHD) study
Comparing first-stage intervention options	Compare the potential outcomes for patients beginning with low-intensity behavior modification (BMOD) and low-dose oral methamphetamine (MEDS)
Comparing second-stage intervention options	Among patients who do not respond to the first stage treatment, compare intensifying the initial intervention versus augmenting the initial intervention with the alternative intervention
Comparing adaptive intervention options	<p>There are four imbedded adaptive interventions:</p> <ol style="list-style-type: none"> 1. Begin with BMOD, augment with MEDS if not responding 2. Begin with BMOD, intensify BMOD if not responding 3. Begin with MEDS, augment with BMOD if not responding 4. Begin with MEDS, intensify MEDS if not responding <p>The goal is to compare the mean outcomes for all pairs of these adaptive interventions.</p>

assigns individualized treatments at each stage based on a patient's personal characteristics and intermediate outcomes; this approach is likely to produce better overall outcomes compared to 'one-size-fits-all' regimes that are not tailored to patients' personal features. The optimal DTR also provides insights about the effects of patients' characteristics on the choice of treatment and eventual outcome; based on this information, researchers can design future confirmatory SMART trials.

Estimating optimal DTR from SMART data has recently received considerable attention in the statistics community; several statistical methods have been developed to achieve this goal.^[25] Here we focus on two machine-learning methods which are flexible, computational efficient, and applicable to handling large numbers of patient-specific characteristics (including genomic and imaging characteristics) as potential tailoring variables,

Q-learning, first proposed in Watkins,^[26] was implemented to analyze SMART data by Murphy and colleagues^[27] and Zhao and colleagues.^[28] It is a regression-based method to identify optimal multi-stage decision rules, where the optimal treatment at each stage is discovered by a backward induction to maximize the estimated Q-function ("Q" stands for "quality of action"). Q-learning is based on simple linear regression model and can be implemented by a SAS procedure known as PROC QLEARN.^[29] For single-stage studies when the assumptions hold and the regression model is correctly specified, Q-learning is efficient. Thus it is widely used to analyze SMART studies with a limited number of tailoring variables. However, regression based Q-learning may suffer from incorrect model assumptions when the number of tailoring variables is large. Even if using nonparametric learning algorithms, the Q-learning approach selects the optimal treatment by modeling the Q-function and its contrasts that are not explicitly related to the optimization of the objective function (i.e., value function^[30]). The mismatch between maximizing the Q-function and the value function potentially leads to suboptimal regimes due to overfitting of the regression model.

Recent advances in statistical methodology avoid these problems. Outcome-weighted learning (O-learning) which was first introduced by Zhao and colleagues^[31] to choose optimal treatment rules by directly optimizing the expected clinical outcome at the end of the study for single-stage trials. The resulting optimal treatment regimen is found by weighted supportive vector machines (SVM) and can take any unconstrained nonparametric functional form. Their simulation studies demonstrate that O-learning

outperforms Q-learning, especially in small-sample settings with a large number of tailoring variables. Zhao, and colleagues^[32] generalized the developed O-learning to multiple-stage trials by a backward iterative method.

Most recently, Zeng and colleagues,^[33] proposed Augmented Multi-stage Outcome-weighted Learning (AMOL), which integrates Q-learning under the O-learning framework and, thus, improves the performance of O-learning. This method incorporates doubly robust augmentation which is also referred as augmented inverse probability weighting originally proposed in the missing data literature^[34] into O-learning by drawing information from regression model-based Q-learning at each stage in the decision tree. Thus, it combines the robustness of O-learning with the imputation ability of Q-learning.

AMOL has three new features not reported in the studies by Zhao and colleagues.^[31,32] Firstly, for single-stage trials, AMOL generalizes the original O-learning^[31] to allow for negative outcome values instead of adding an arbitrarily large constant^[31] which leads to numeric instability. This feature is useful when there are both positive and negative outcomes observed in a clinical study (e.g., rate of change of clinical symptoms). Secondly, by using residuals from a regression on variables other than the treatment assignment as outcome values, AMOL is able to reduce the variability of weights in O-learning to achieve numeric stability and efficiency gain. Thirdly, and most importantly, for multiple-stage trials, AMOL estimates optimal DTRs via a backward induction learning procedure^[32] which starts from the last stage and propagates backwards to the first stage to boost efficiency through augmentation and integration with Q-learning. At each stage of the study of interest, the optimal treatment regimes are obtained using only subjects whose treatment assignments coincide with the optimal rule for all the future stages in the study. Thus, one major limitation of O-learning is that the number of subjects used for inferring optimal treatment rules decreases geometrically with the increasing number of stages, so their method may be inefficient. In contrast, at each stage, AMOL uses robustly weighted O-learning for estimating the optimal DTRs; the weights are based on the observed outcome and a conditional expectation term for subjects who follow the optimal treatment rules in future stages or – for those who do not follow optimal rules in future stages – weights imputed from regression models obtained from Q-learning. Therefore, AMOL, as a hybrid approach, simultaneously takes advantage of the robustness of nonparametric O-learning and also makes use of the model-based Q-learning which uses data from all subjects.

2.3 Example of Q-learning and O-learning based analyses of ADHD data

The ADHD data analysis we present here was simulated by investigators at the University of Michigan based on an ongoing two-stage SMART trial on ADHD^[12] that has been used in a workshop about SMART that can be downloaded at: (http://www-personal.umich.edu/~dalmiral/software/mw_workshop_files/SAS%20Code/adhd_simulated_data.txt). The primary outcome of the study is the school performance score (ranging from 1 to 5) measured at the end of the study. There are 150 subjects, four baseline covariates (e.g. prior medication history, ADHD impairment score, ODD diagnosis, race) and two time-varying covariates

including adherence to the initial treatment and months to remission. There were 99 participants who did not respond to first stage intervention and are re-randomized in the second stage.

We present the estimated coefficients of the optimal DTR estimated by Q-learning and AMOL in Table 2. AMOL gives a sparse set of variables with non-important variables yielding coefficients near zero. In contrast, Q-learning leads to many more variables with non-zero coefficients. We can rank the importance of standardized covariates by the magnitude of their coefficients. In stage 1, medication prior to enrollment has the largest magnitude coefficient estimated by AMOL (-0.001557, Table 2), which is more than 3-fold

Table 2. Standardized coefficients for the optimal dynamic treatment rule estimated by various methods using data from the Attention Deficit/Hyperactivity Disorder (ADHD) study^a

	stage 1		stage 2		
	Q-L	AMOL ^b	Q-L	AMOL ^b	
Intercept	3.454	0	Intercept	2.889	0
ODD diagnosis	-0.199	-0.229	ODD Diagnosis	-0.144	0
Baseline ADHD score	-0.357	0.276	ADHD score	-0.28	0
Prior medication	-0.028	-1.557	Prior medication	0.012	0
White race	0.211	0.456	White race	0.247	0.088
trt1 (1 for BMOD; -1 for MED)	0.225		trt1	0.273	-0.043
ODD diagnosis* trt1	-0.068		ODD diagnosis* trt1	-0.141	0
ADHD *trt1	0.163		ADHD *trt1	0.075	0
Prior medication*trt1	-0.348		Prior medication *trt1	-0.049	0
race*trt1	0.086		White race*trt1	0.11	0.088
			Months to non-response	-0.015	0
			Adherence to trt1	0.003	0.999
			Months to non-response*trt1	-0.33	0
			Adherence to trt1*trt1	0.09	0
			trt2	-0.385	
			...		
			Adherence to trt1*trt2	0.633	

Q-L, Q-learning O-L, O-learning AMOL, Augmented Multi-stage Outcome-weighted Learning
 ODD, Oppositional Defiant Disorder BMOD, Behavioral Modification MED, Medication
 trt1: first stage treatment, 1=use BMOD; -1=use MED
 trt2: second stage treatment, 1=intensify current treatment; -1=add alternative treatment

^a Q-learning also included other interaction terms with trt2 which are omitted in the table

^b The reported coefficients were obtained from fitting a linear prediction rule for the outcome with listed variables included as covariates in AMOL. The estimated coefficients were the numbers displayed in this column multiplied by 0.001 for the ease to show relative magnitude of each variable (e.g., the estimated coefficient for prior medication was -0.001557).

the magnitude of the second largest covariate (race). The fitted optimal DTR suggests that patients who previously took medication before the trial would be better off starting with medication, and those who did not take medication before the trial should start with behavioral modification. In stage 2, adherence to treatment in stage 1 has the largest magnitude coefficient (0.999, Table 2). The AMOL fitted optimal DTR suggests that patients who adhered to their initial treatment should be assigned to continue with the same treatment, while patients who did not adhere to the first treatment should switch.

3. Discussion

This paper has introduced the design of SMARTs for assessment of DTRs in psychiatric research, the statistical methods used to make inference about the primary goal in such studies, and the most recently introduced machine learning methods for identifying the best treatment and for identifying potential tailoring variables for future confirmative trials. A few core issues about the statistical analyses of SMART and DTR merit further research. Most methods on identifying optimal DTR from SMART are targeted on continuous outcomes; further work will be need to extend this approach to deal with ordinal or categorical outcomes and censored survival events. Moreover, in mental health research there is often interest in a combination of outcomes (to comprehensively assess potential benefit); for example, alleviation of symptoms may be considered in conjunction with increased quality of life and functioning, time to response, and reduction

of side effects. In this situation it may be insufficient to represent all information in a single dimensional outcome. Further work will be needed to develop machine-learning methods for handling such multi-dimensional outcomes. Another issue is that in many clinical studies there may be multiple options – not just two – at each stage of the study; current machine-learning methods need to be extended to identify optimal DTRs when multiple treatment options are possible at each stage of the study. Future research is also needed to develop methods for selecting the feature variables from observational studies that will best maximize interpretability of constructed DTR. Finally, one practical challenge is that multiple-stage randomized clinical trials require prolonged commitment and compliance from all participants. Missing data in SMARTs is often a rule rather than an exception, so continued effort is needed to find creative ways for reducing missing data and for statistically dealing with missing data. Shortreed and colleagues^[35] recently discussed imputation methods for handling missing data in SMART.

Conflict of Interest

The authors report no conflict of interest related to this manuscript.

Funding

This research is sponsored by the United States National Institute of Health (NS082062, NS073671).

个体化动态治疗方案和多重方案随机序贯试验在精神卫生研究中的应用

Liu Y, Zeng DL, Wang YJ

概述: 动态治疗方案 (Dynamic treatment regimens, DTRs) 是一种序贯决策规则, 是根据每个患者随时间变化而变化的特征和先前观察到的中间结果而量身定制的临床决策。精神障碍具有慢性和复杂性的特点, 精神障碍患者具有异质性特点。这就要求随时间推移, 根据个体对治疗反应的不同而分析出最佳的治疗方案, 并动态地应用到患者之后的治疗中。多重方案随机序贯试验 (Sequential Multiple Assignment Randomized Trial, SMARTs) 的设计可以估计 DTRs 的治疗效应。SMARTs 收集到大量的个体化变量和中间结果, 在此基础上应用已有的现代统计工具可以优化 DTRs。这些统

计方法也可为今后的 SMARTs 研究设计推荐量身定制的变量。本文通过两个精神卫生研究案例介绍了 DTRs 和 SMARTs, 讨论了从 SMARTs 数据估算出最佳 DTR 的两种不同的计算机自动分析方法, 并使用模拟数据演示这两种统计方法的性能。

关键词: 多重方案随机序贯试验, 动态治疗方案, 个体化医疗, O 型学习, Q 型学习, 双稳健估计

本文全文中文版从 2015 年 01 月 25 日起在 www.shanghaiarchivesofpsychiatry.org 可供免费阅读下载

References

- Lavori PW, Dawson R. A design for testing clinical strategies: biased adaptive within-subject randomization. *J R Stat Soc Ser A Stat Soc.* 2000; **163**(1): 29-38. doi: <http://dx.doi.org/10.1111/1467-985X.00154>
- Thall PF, Sung HG, Estey EH. Selecting therapeutic strategies based on efficacy and death in multicourse clinical trials. *J Am Stat Assoc.* 2002; **97**(457): 29-39
- Thall PF, Wathen JK. Covariate-adjusted adaptive randomization in a sarcoma trial with multi-stage treatments. *Stat Med.* 2005; **24**(13): 1947-1964. doi: <http://dx.doi.org/10.1002/sim.2077>
- Lunceford JK, Davidian M, Tsiatis AA. Estimation of survival distributions of treatment policies in two-stage randomization designs in clinical trials. *Biometrics.* 2002; **58**(1): 48-57. doi: <http://dx.doi.org/10.1111/j.0006-341X.2002.00048.x>
- Wahed AS, Tsiatis A A. Optimal estimator for the survival distribution and related quantities for treatment policies in two-stage randomization designs in clinical trials. *Biometrics.* 2004; **60**(1): 124-133. doi: <http://dx.doi.org/10.1111/j.0006-341X.2004.00160.x>
- Wahed AS, Tsiatis AA. Semiparametric efficient estimation of survival distributions in two-stage randomisation designs in clinical trials with censored data. *Biometrika.* 2006; **93** (1): 163-177
- Rush AJ, Fava M, Wisniewski SR, Lavori PW, Trivedi MH, Sackeim HA, et al. Sequenced treatment alternatives to relieve depression (STAR*D): rationale and design. *Controlled clin trials.* 2004; **25**(1): 119-142
- Rush A, Trivedi M, Wisniewski S, Nierenberg A, Stewart J, Warden D, et al. Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR*D report. *Am J Psychiatry.* 2006; **163**(11): 1905-1917
- Schneider LS, Ismail MS, Dagerman K, Davis S, Olin J, McManus D, et al. Clinical antipsychotic trials of intervention effectiveness (CATIE): Alzheimer's disease trial. *Schizophr bull.* 2003; **29**(1): 57. doi: <http://dx.doi.org/10.1093/schbul/sbj026>
- McKay J, Horn DV, Oslin D, Lynch K, Ivey M. A randomized trial of extended telephone-based continuing care for alcohol dependence: within-treatment substance use outcomes. *J Consult Clin Psychol.* 2010; **78**: 912-923. doi: <http://dx.doi.org/10.1037/a0020700>
- Jones H, O'Grady K, Tuten M. Reinforcement-based treatment improves the maternal treatment and neonatal outcomes of pregnant patients enrolled in comprehensive care treatment. *Am J Addict.* 2011; **20**: 196-204. doi: <http://dx.doi.org/10.1111/j.1521-0391.2011.00119.x>
- Pelham WE, Hoza B, Pillow DR, Gnagy EM, Kipp HL, Greiner AR, et al. Effects of methylphenidate and expectancy on children with ADHD: Behavior, academic performance, and attributions in a summer treatment program and regular classroom setting. *J Consult Clin Psychol.* 2002; **70**: 320-335
- Pelham W, Fabiano G A. Evidence-based psychosocial treatments for attention-deficit/hyperactivity disorder. *J Clin Child Adolesc Psychol.* 2008; **37**(1): 184-214. doi: <http://dx.doi.org/10.1080/15374410701818681>
- Kasari C, Freeman S, Paparella T. Joint attention and symbolic play in young children with autism:a randomized controlled intervention study. *J Child Psychol Psychiatry.* 2006; **47**: 611-620. doi: <http://dx.doi.org/10.1111/j.1469-7610.2005.01567.x>
- Lei H, Nahum-Shani I, Lynch K, Oslin D, Murphy S. (2012). A "SMART" design for building individualized treatment sequences. *Annu Rev Clin Psychol.* 2012; **8**(1): 21-48. Epub 2011 Dec 12. doi: <http://dx.doi.org/10.1146/annurev-clinpsy-032511-143152>
- Murphy SA, Van Der Laan MJ, Robins JM. Marginal mean models for dynamic regimes. *J Am Stat Assoc.* 2001; **96**(456): 1410-1423
- Moodie EE, Richardson TS, Stephens DA. Demystifying optimal dynamic treatment regimes. *Biometrics.* 2007; **63**(2): 447-455. doi: <http://dx.doi.org/10.1111/j.1541-0420.2006.00686.x>
- Rubin DB. Bayesian inference for causal effects: The role of randomization. *Ann Stat.* 1978; **6**: 34-58
- Berkane M. *Latent variable modeling and applications to causality.* New York: Springer New York; 1997
- Fabiano G, Pelham E, Waschbusch D, Gnagy M, Lahey B. A practical measure of impairment: psychometric properties of the impairment rating scale in samples of children with attention deficit hyperactivity disorder and two school-based samples. *J Clin Child Adolesc Psychol.* 2006; **35**: 369-385
- Pelham WE Jr, Gnagy EM, Greenslade KE, Milich R. Teacher ratings of DSM-III-R symptoms for the disruptive behavior disorders. *J Am Acad Child Adolesc Psychiatry.* 1992; **31**(2): 210-218. doi: <http://dx.doi.org/10.1097/00004583-199203000-00006>
- Hernán MÁ, Brumback B, Robins JM. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology.* 2000; **11**(5): 561-570
- Nahum-Shani IM, Qian M, Almirall D. Experimental design and primary data analysis methods for comparing adaptive interventions. *Psychological methods.* 2012; **17**(4): 457. doi: <http://dx.doi.org/10.1037/a0029372>
- Oetting AI, Levy JA, Weiss RD, Murphy SA. Statistical methodology for a SMART design in the development of adaptive treatment strategies. In: *Causality and Psychopathology: Finding the Determinants of Disorders and their Cures Arlington.* VA: American Psychiatric Publishing, Inc; 2007
- Chakraborty B, Moodie EE. *Statistical Methods for Dynamic Treatment Regimes.* New York: Springer; 2013
- Watkins CJ. *Learning from delayed rewards* (Ph.D. dissertation). UK: University of Cambridge; 1989

27. Murphy SA, Collins LM, Rush AJ. Customizing treatment to the patient: Adaptive treatment strategies. *Drug Alcohol Depend.* 2007; **88**(Suppl 2): S1-S3. doi: <http://dx.doi.org/10.1016/j.drugalcdep.2007.02.001>
28. Zhao Y, Kosorok MR, Zeng D. Reinforcement learning design for cancer clinical trials. *Stat Med.* 2009; **28**(26): 3294-3315. doi: <http://dx.doi.org/10.1002/sim.3720>
29. Ertefaie A, Almirall DA, Huang L, Dziak JJ, Wagner AT, Murphy SA. *SAS PROCQLEARN users' guide (Version 1.0)*. University Park: The Methodology Center, Penn State; 2012. Retrieved from <http://methodology.psu.edu>
30. Qian M, Murphy SA. Performance guarantees for individualized treatment rules. *Ann Stat.* 2011; **39** (2): 1180
31. Zhao Y, Zeng D, Rush AJ, Kosorok MR. Estimating individualized treatment rules using outcome weighted learning. *J Am Stat Assoc.* 2012; **107** (499): 1106-1118. doi: <http://dx.doi.org/10.1080/01621459.2012.695674>
32. Zhao Y, Zeng D, Laber E, Kosorok MR. New statistical learning methods for estimating optimal dynamic treatment regimes. *J Am Stat Assoc.* 2014; in press. doi: <http://dx.doi.org/10.1080/01621459.2014.937488>
33. Zeng D, Wang Y, Liu Y, Kosorok M. *Improved Outcome Weighted Learning for Dynamic Treatment Regimes*. Massachusetts, Boston: Joint Statistical Meetings; 2014
34. Robins JM, Rotnitzky A, Zhao LP. Estimation of regression coefficients when some regressors are not always observed. *J Am Stat Assoc.* 1994; **89**: 846-866
35. Shortreed SM, Laber E, Stroup TS, Pineau J. A multiple imputation strategy for sequential multiple assignment randomized trials. *Stat Med.* 2014; **33**(24): 4202-4214. Epub 11 Jun 2014. doi: <http://dx.doi.org/10.1002/sim.6223>

(received, 2014-10-22; accepted, 2014-11-20)



Ying Liu (M. Phil M.S.) graduated from the Department of Mathematics at Peking University and is currently a Ph.D. candidate in the Department of Biostatistics at Columbia University. Her dissertation research is focused on merging statistical modeling and medical domain knowledge with machine learning algorithms to make personalized medical decisions using complex biomedical data. She also works with clinical collaborators and psychiatrists at the New York State Psychiatric Institutes on innovative clinical trials and epidemiological studies in anorexia nervosa, schizophrenia and other mental disorders.