PHARMACOEPIDEMIOLOGY AND DRUG SAFETY 2014; **23**: 580–585
Published online 3 April 2014 in Wiley Online Library (wileyonlinelibrary.com) **DOI**: 10.1002/pds.3606

REVIEW

Personalizing medicine: a review of adaptive treatment strategies

Michael P. Wallace* and Erica E. M. Moodie

Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Canada

ABSTRACT

Much of current pharmacological practice focuses on identifying the single 'best' treatment (or course of treatments) for a particular disease. Recently, however, focus has begun to shift towards a more patient-centric rather than disease-centric approach, where personal characteristics are used to identify the optimal treatment for an individual. Adaptive treatment strategies (also known as dynamic treatment regimes) are part of a rapidly expanding area of research whereby such personalized treatments can be identified. These methods can lead to improved results over standard 'one size fits all' approaches, as well as provide a route to formalizing a common practice of using ad hoc approaches when deciding or updating management plans. Here, we provide an introduction to adaptive treatment strategies, explaining their background, their purpose, and how they can be employed in practice. Copyright © 2014 John Wiley & Sons, Ltd.

KEY WORDS—adaptive treatment strategies; dynamic treatment regimes; personalized medicine; Q-learning; sequential randomization; pharmacoepidemiology

Received 29 November 2013; Revised 4 February 2014; Accepted 4 February 2014

INTRODUCTION

Traditionally, pharmacological practice has aimed to identify the single 'best' treatment for a particular illness. This often involves comparing two treatments on a cohort of subjects and seeing which, on average, performs better. It is well known, however,[1] that two patients given the same treatment may well respond differently, with these concerns particularly evident in the context of chronic illnesses. With an ageing global population,[2] such conditions are becoming more prevalent[3] and now account for over 60% of global mortality.[4] Treatments for chronic conditions may have unanticipated or poorly understood long-term effects, while differing responses to treatments must be considered not only across individuals but also within them over time.

Personalized medicine has emerged as a catchall term for medical management plans that encompass patient-level characteristics. Rather than directly focusing on a disease or diagnosis, a more personalized approach instead looks at patient characteristics that inform which treatment options will work best for a particular individual at a particular time. Such approaches, particularly in the chronic care setting, have long been employed by physicians on an ad hoc basis.[5,6] Strategies such as 'play the winner, drop the loser',[7] where treatments that appear to be working are maintained while others dropped, are relatively commonplace despite often relying on limited evidence or even mere intuition. Interest has therefore begun to grow in developing more algorithmic approaches to personalizing treatment.[8,9]

Adaptive treatment strategies (ATSs) are a rapidly expanding area that can be viewed as a more formal means of implementation of personalized approaches. A 'strategy' in this context is a sequence of decision rules where, at each point a treatment decision is made, it may be informed by a patient's *history* of covariates and past treatments. While, in its simplest form, one can consider an adaptive strategy for a 'single-rule' management plan ('prescribe drug A if the patient is overweight; prescribe drug B otherwise', perhaps), more generally, they can be employed over a lengthy follow-up where decisions are undertaken at the beginning of each of a number of *stages* of treatment. Our primary interest lies in these long-term settings,

*Correspondence to: M. P. Wallace, Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Purvis Hall, 1020 Pine Avenue West, Montreal, Quebec, Canada H3A 1A2. Email: michael.wallace@mcgill.ca

where our management plans *tailor* treatment to changes in patient characteristics and their response to previous treatments. The time-varying nature of such approaches is reflected in another popular term—*dynamic treatment regimes*[10,11]—while other authors have referred to 'treatment policies' in this context.[12–14]

In this paper, we review the concept of multi-stage treatments and describe Sequential Multiple Assignment Randomized Trials (SMARTs), a design specifically aimed at collecting high-quality data so as to learn about ATSs. We also briefly touch on analyses of data arising from such studies. Of course, in an ever-developing field, there are a huge number of estimation and inferential methods available for a variety of settings, which we do not cover in detail here. Discussion and description of these are instead best left to a dedicated resource.[11] We endeavour nonetheless to demonstrate that adaptive strategies are, in some cases, the only means of achieving optimal outcomes for individuals, and learning about such strategies must be done through studies of the sequences as a whole rather than in a series of studies of components of the strategy.

TWO-STAGE TREATMENTS (AND BEYOND)

To illustrate ATSs more concretely, we take inspiration from the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) trial. This multi-level randomized controlled trial (RCT) was designed to compare treatment regimes for patients with major depressive disorder (MDD).[15,16] The sequential nature of the study was characterized by recurring clinic visits whereat depression severity was assessed using Quick Inventory of Depressive Symptomatology (QIDS) scores.[17] Participants progressed through a sequence of treatment stages where the treatment options available depended on their previous management and patient preference. Treatment decisions were centred on the dichotomy of *switching* from the current therapy to a different one or *augmenting* it with an additional treatment. Patients were asked whether they would prefer to switch or augment and then randomized to one of some number of treatments consistent with their preference. This design, in many ways, mimics actual clinical practice: Patients are monitored at regular intervals, and, based on their current health status, recent treatment, and personal preferences, treatment may be changed.

Sequenced Treatment Alternatives to Relieve Depression is an example of a SMART: At each of a sequence of stages, patients are randomized to one of the available treatment options, potentially dependent on patient preference or response to prior treatment. First introduced under the name *biased coin adaptive within-subject* design,[18,19] the general framework was proposed by Murphy,[20] with such designs aiming to reflect standard clinical practice in chronic disorder management within a randomized setting. Other examples of SMARTs include the multi-stage Clinical Antipsychotic Trials of Intervention Effectiveness for the treatment of Alzheimer's disease[21] and schizophrenia[22–24] along with numerous two-stage trials investigating cancer[25,26] and smoking cessation.[27,28]

A hypothetical example

We illustrate SMART designs with a hypothetical simplification of STAR*D. Consider a 6-week study where patients with MDD are monitored at clinic visits at weeks 0 (study entry), 3 and 6 (study end). Three treatments are considered: citalopram (a standard first-line therapy for depression), lithium (a mood stabilizer) and cognitive behavioral therapy (CBT, a counselling-based approach). The outcome of interest is QIDS score measured at 6 weeks (which we wish to minimize), with treatment for weeks 0–3 and 3–6 assigned as follows (Figure 1):

- Week 0: Patients are randomized to citalopram or CBT.
- Week 3: Patients are classified as *responders* if their QIDS score has improved or *non-responders* otherwise. Patients are randomized to one of two treatments as follows:
  Responders: Patients are maintained on the current therapy, or the treatment is augmented with citalopram or CBT (as appropriate).
  Non-responders: Patients switch to one of the other two therapies.

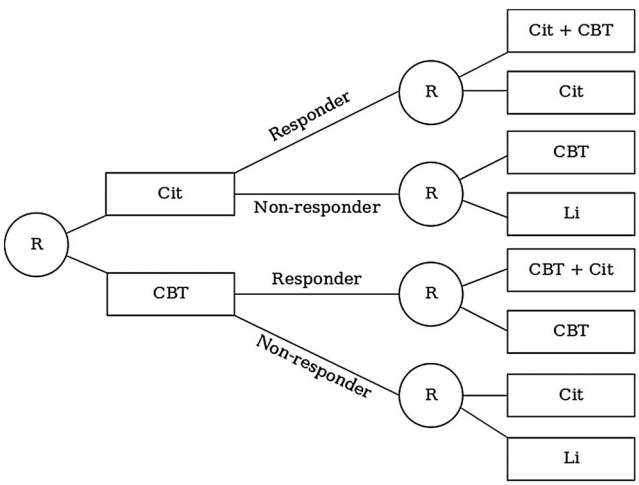

Figure 1. Hypothetical two-stage SMART design based on STAR*D trial. Circled Rs represent randomization to boxed treatments: Cit = citalopram, CBT = cognitive behavioral therapy, Li = lithium

DOI: 10.1002/pds

This example, featuring two points at which a treatment decision is made, is referred to as a *two-stage* design. It could, of course, be extended to more stages or incorporate more treatment options, if desired.

Using a SMART design, researchers are able to collect high-quality data so as to identify treatment *strategies* that will optimize (minimize) a patient's QIDS score. For example, such a trial may ultimately lead to the strategy

> *'At study entry, prescribe citalopram if patient's initial QIDS score is 16 or higher; prescribe CBT otherwise. At three weeks maintain those with an improved QIDS score on their original treatment; switch others to lithium.'*

or some other sequence of treatment decision rules.

Notation

We introduce some notation to accompany the example above. The outcome, Y, can be either an end-of-study variable or a summary of intermediate variables. Y is typically chosen so that larger values are preferred; in situations where we wish to minimize the 'natural' outcome variable, a simple transformation will usually suffice. In our depression example, we may be interested in a patient's QIDS score at 6 weeks and so could take Y to be a patient's QIDS score subtracted from the maximum possible 27. In a longer study, a summary statistic such as the number of deviations from a target therapeutic range over the course of the trial could be used.[29]

Observed patient characteristics (covariates) at time t before treatment is assigned, such as QIDS score or whether the patient responded to the previous treatment, are denoted by X_t. Because we work in stages, 'time' in SMARTs is necessarily discrete, and so, t will typically be the stage number at which the measurement was taken. For example, X_1 would be patient information at stage one, such as QIDS score at study entry. Similarly, A_t indicates the treatment received (or action taken) at stage t; a patient receiving citalopram at stage two would be recorded in A_2.

Total patient information at each stage is referred to as their *history* and denoted as H_t. This includes all the patient information known *prior* to treatment assignment at that stage ($H_1 = (X_1)$ and $H_2 = (X_1, A_1, X_2)$, for example) and can be viewed as the information available to inform that treatment decision. An entire ATS is denoted by d. Such a strategy is a sequence of individual decision rules: functions of a patient's history that return a recommended treatment decision. A stage-specific strategy may be identified by the standard subscripting. These notations, along with some frequently seen alternatives, are summarized in Table 1.

Advantages: why SMARTs are smart

A natural consideration is whether an alternative (cheaper) study design could be used to address the same research question as effectively as a SMART. For example, a more 'obvious' approach might be to conduct two single-stage RCTs. The first of these would simply compare citalopram and CBT, perhaps finding that over a 3-week follow-up, patients responded best to the former. The second study would then begin by prescribing citalopram to all patients for 3 weeks before randomizing responders to maintenance or augmentation and non-responders to one of the two switching options. The approach of conducting a series of single-stage trials is, however, inferior to the SMART. First, it cannot be used to identify delayed treatment effects: Suppose that while citalopram is more effective than CBT in the short term, patients who had previously received CBT then go on to respond better to subsequent treatments (perhaps because CBT improves later compliance). A related issue is that observing a patient receive a specific treatment may in itself provide useful information for future management decisions, information that would be lost if that treatment was not recommended in the short term. For example, one might find that in the process of receiving CBT, patients provide information that could be used to better tailor follow-up treatment (for instance, whether or not they would struggle to comply with lithium treatment more than citalopram). In both cases—delayed effects and the gain of diagnostic information—the single-stage trial, unlike the SMART, would not identify these benefits of initial CBT to overall outcome and instead simply recommend citalopram over CBT as a first-line treatment.

Finally, SMARTs may be less restricted in their recruitment criteria and so more representative of the study population. Unlike a traditional RCT, where recruitment criteria may be highly specific to reduce variability in treatment effects, SMARTs seek to identify heterogeneity in response and use that variability to better tailor treatments.[22,30] Further, patients who do not respond to treatment in a single-stage study often have little choice but to continue with an ostensibly ineffective therapy or drop out.[31] In a SMART,

Table 1. Notation (and some common alternatives)

Object	Notation	Alternative notations
Outcome of interest	Y	
Patient information at stage t	X_t	O_t, L_t, S_t
Treatment assigned at stage t	A_t	
History at stage t	$H_t = (\overline{X}_t, \overline{A}_{t-1})$*	
Treatment strategy	d	π

*Overline notation can be used to denote the past, e.g. the sequence of treatment decisions $(A_1, A_2, \ldots, A_j)$ can instead be represented by $\overline{A}_j$.

DOI: 10.1002/pds

treatment changes are built into the trial protocol and such patients need not be lost or viewed as off study protocol. Thus, results from a SMART may be more generalizable than those from a single-stage trial.

Design considerations

When designing a SMART, a variety of hypotheses may be investigated. These might concern identifying the best initial treatment, the best follow-up treatment, or comparing two strategies in their entirety. Because of its randomized, multi-treatment, multi-stage design, a single SMART can provide unbiased answers to more than one such question. For the purposes of power and sample size calculations, however, a single *primary* research question is typically specified (although this is not essential). While other secondary hypotheses may also be investigated, the study may be insufficiently powered for them.

Sample size calculation for a SMART will vary depending on the primary research question. For example, if the researcher is concerned with identifying the best initial treatment, sample size is calculated in a similar fashion to a standard two-group RCT. Alternatively, if we modified our SMART example such that responders to CBT were randomized to either CBT and citalopram or citalopram alone (i.e. the same follow-up options as those who respond to an initial citalopram treatment), then we may wish to target the second-line therapy for inference and ask which of these two treatments is preferred among those who respond to their initial therapy. In this instance, sample size calculations are similarly straightforward but also require estimation of the overall response rate to initial treatment.

Alternatively, two entire treatment strategies may be compared. This involves comparing the two subgroups of patients whose treatment allocations are consistent with the strategies in question and estimating the difference in expected outcome between them. Calculations are most straightforward when the subgroups are distinct; to ensure this, Murphy[20] suggests that the strategies should begin with different initial treatments. A more general procedure, meanwhile, may seek to identify the *best* strategy (rather than merely comparing two specific examples) by estimating long-term outcomes.

While test statistics and sample size formulae or algorithms have been derived for comparing entire treatment strategies,[20,32] such questions can nevertheless become difficult in even relatively simple settings. When designing a SMART, it is therefore advisable to limit both the treatment options as well as the decision rules that may be pursued to minimize the risk of a high-dimensional (and potentially intractable) problem and unrealistic sample size requirements. A pilot study may also be required to provide reasonable estimates of the parameters required for the sample size calculations.[33]

ANALYSIS BY Q-LEARNING

Our final consideration is how one can use data collected in a SMART to identify an *optimal* ATS. Formally, the optimal strategy we wish to identify is defined as the sequence of decision rules that maximizes the expected outcome: $d^{opt}(H) = arg\,max_{A_2} E[Y|d(H)]$. In a single-stage setting, this is potentially straightforward. For example, suppose that a linear relationship exists between our outcome, the treatment and a single patient covariate. Using our above notation (but dropping the subscripts for convenience), we might propose the model

$$E[Y|X,A] = \beta_0 + \beta_1 X + A(\psi_0 + \psi_1 X)$$

where X is a patient's initial QIDS score and A is an indicator variable that takes the value 1 if a patient is prescribed citalopram and 0 if they are prescribed CBT. The first 'half' of this model (parameterized by $\boldsymbol{\beta}$) will be familiar as part of a standard linear regression model relating the outcome Y to initial QIDS score X. The second 'half' (parameterized by $\boldsymbol{\psi}$) represents the *interaction* between patient history and treatment A. Our treatment strategy is concerned with choosing between citalopram ($A=1$) and CBT ($A=0$) so as to maximize $E[Y|X,A]$. Consequently, only the sign of $\psi_0+\psi_1 X$ matters: if it is positive, then $E[Y|X,A]$ is maximized by $A=1$; otherwise, we would choose $A=0$. This suggests the following procedure for identifying the optimal treatment strategy:

(1) Use ordinary least squares to obtain estimates $\hat{\psi}_0, \hat{\psi}_1$ (and $\hat{\beta}_0, \hat{\beta}_1$).
(2) Use these estimates to define a decision rule:
    (a) Prescribe CBT ($A=0$) if $\hat{\psi}_0 + \hat{\psi}_1 X \leq 0$.
    (b) Prescribe citalopram ($A=1$) otherwise.

Matters are inevitably more complex with multiple stages of treatment. A detailed tutorial of ATS analytical methods is beyond the scope of this article; however, we provide details of one such approach owing to its relative intuitiveness and simplicity in implementation.

Q-learning has its roots in computer science, where, as a *reinforcement learning*[34,35] technique, it is used to solve optimization problems. We demonstrate Q-learning in the two-stage setting but note that the approach extends naturally to any (finite) number of stages. We begin by identifying a decision rule for second-stage treatments and then use this rule to identify the first-stage treatment

DOI: 10.1002/pds

that maximizes a patient's outcome *assuming that they then go on to receive optimal treatment*.

Persisting with the previous notation, we assume that we observe patient outcome Y with history (X_1,A_1,X_2,A_2). The Q-learning process is then formally defined in terms of *quality* or *Q-functions*:

$$Q_2(H_2,A_2) = E[Y|H_2,A_2], \text{ and}$$
$$Q_1(H_1,A_1) = E[\max_{A_2} Q_2(H_2,A_2)|H_1,A_1].$$

The second-stage Q-function is reassuringly familiar: a conditional mean that can be modelled in the usual ways. Suppose that we assume a linear model such that $Q_2(H_2,A_2;\beta_2,\psi_2) = \beta_2^T H_2 + \psi_2^T H_2 A_2$; we begin by regressing the outcome on patient history to obtain estimates of the stage-two regression parameters $\hat{\beta}_2$ and $\hat{\psi}_2$. These estimates can then be used to construct a decision rule for the second-stage of treatment: 'Prescribe $A_2 = 0$ if $\hat{\psi}_2^T H_2 \leq 0$; prescribe $A_2 = 1$ otherwise'.

Having fit our model for Q_2, we next use our second-stage parameter estimates to generate a *pseudo-outcome* $Y' = \max_{A_2} Q_2\left(H_2, A_2; \hat{\beta}_2, \hat{\psi}_2\right)$. This represents the expected outcome for a patient with the given history if they go on to receive optimal second-stage treatment. 'Plugging' this into our first-stage Q-function yields a more familiar form: $Q_1(H_1,A_1) = E[Y' \mid H_1,A_1]$. If we model Q_1 by, say, $E\left[Y'|H_1,A_1;\beta_1,\psi_1\right] = \beta_1^T H_1 + \psi_1^T H_1 A_1$, then linear regression of the pseudo-outcome Y' on stage-one history yields estimates $\left(\hat{\beta}_1, \hat{\psi}_1\right)$ that can be used to derive a decision rule for the first stage of treatment. Combining this with the second-stage decision rule already identified gives a full two-stage ATS.

It is important to note that Q-learning requires the models for the Q-functions to be correctly specified; alternative approaches such as g-estimation can be employed more generally.[29]

DISCUSSION

The growing interest in personalized medicine has expedited the need for novel statistical methods that provide an empirical alternative to the ad hoc approaches often adopted in practice. Furthermore, an ageing population has increased the prevalence of chronic diseases and comorbidities often requiring numerous treatment decisions over time. ATSs provide a formalized, evidence-based framework wherein optimal management plans can be derived and implemented.

In this paper, we have provided something of a 'crash course' in ATSs, demonstrating why personalized medicine can lead to better results, the importance of SMART study design, and a brief introduction to how we can use the resulting data to identify optimal management plans. The interested reader is referred to other recent surveys of the ATS literature in general and SMARTs and Q-learning in particular.[10,36–40] We have, however, deliberately avoided spending much time on describing specific analytical methods. While, owing to its simplicity, we have mentioned Q-learning, there are of course numerous alternative methods available, each with their own advantages and drawbacks. Many can be seen as variations on Q-learning or as following the same recursive approach.

A final consideration is the data types themselves we wish to work with. In this paper, we have considered continuous outcome variables with binary treatment decisions, and while commonly encountered in practice (perhaps even by deliberate construction), we will often wish to work beyond these limitations. Recent publications have begun to consider discrete outcome variables,[25,41] continuous treatments,[20,29,42–44] and survival outcomes.[45,46]

Despite these issues, ATSs remain an attractive, important, and expanding area of health research. While at their most powerful when working with specifically collected data (such as those from SMARTs), they can be used with observational data as well, provided that a number of assumptions are satisfied.[47,48] With new methods and applications continually in development, it is a field rich in possibilities—both theoretical and practical—and one that has the potential to dramatically impact health research and delivery.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

KEY POINTS

- Adaptive treatment strategies, which allow for the personalization of medicine by identifying treatment strategies that adapt to evolving patient characteristics, can thereby improve outcomes.
- Sequential multiple assignment randomized trials (SMARTs) are the most valid means of learning about and identifying treatment sequences and how they should be adapted over time.
- Q-learning is a relatively simple method by which to identify optimal treatment strategies.
- Research into adaptive treatment strategies is expanding; new methods and applications are in development that have the potential to significantly impact health research and delivery.

DOI: 10.1002/pds

ETHICS STATEMENT

The authors state that no ethical approval was needed.

ACKNOWLEDGEMENTS

This work is funded by a Natural Sciences and Engineering Research Council (NSERC) of Canada Discovery grant. Dr Moodie is sponsored by a Chercheur-boursier career award from the Fonds de recherche du Québec-Santé (FRSQ).

REFERENCES

1. Gail M, Simon R. Testing for qualitative interactions between treatment effects and patient subsets. *Biometrics* 1985; **41**: 361–372.
2. *World population prospects: The 2012 revision*. United Nations: New York, 2012.
3. Conditions C. *Making the Case for Ongoing Care; September 2004 Update*. Partnership for Solutions, Johns Hopkins University: Baltimore, MD, 2004.
4. World Health Organisation: Chronic diseases. Available at: http://www.who.int/topics/chronic_diseases/en/ [24 January 2014].
5. Topol E. *The Creative Destruction of Medicine*. Basic Books, 2012.
6. Grove W, Zald D, Lebow B, *et al*. Clinical Versus Mechanical Prediction: A Meta-Analysis. *Psychol Assess* 2000; **12**(1): 19–30.
7. Thall P, Millikan R, Sung H. Evaluating multiple treatment courses in clinical trials. *Stat Med* 2000; **19**: 1011–1028.
8. Lenzini P, Wadelius M, Kimmel S, *et al*. Integration of genetic, clinical, and INR data to refine warfarin dosing. *Clin Pharmacol Ther* 2010; **87**(5): 572–578.
9. Nielsen P, Lundbye-Christensen S, Larsen T, *et al*. Monitoring of anticoagulant therapy applying a dynamic statistical model. *Comput Methods Programs Biomed* 2013; **110**: 380–388.
10. Chakraborty B, Murphy S. Dynamic treatment regimes. *Annual Review of Statistics and its Application* 2014; **1**: 447–464.
11. Chakraborty B. *Moodie E*. Statistical Methods for Dynamic Treatment Regimes: Springer, 2013.
12. Lunceford J, Davidian M, Tsiatis A. Estimation of survival distributions of treatment policies in two-stage randomization designs in clinical trials. *Biometrics* 2002; **58**: 48–57.
13. Lizotte D, Laber E, Murphy S. Assessing confidence in policies learned from sequential randomized trials. Technical Report 481, Department of Statistics, University of Michigan, 2009.
14. Dawson R, Lavori P. Efficient design and inference for multistage randomized trials of individualized treatment policies. *Biostatistics* 2012; **13**: 142–152.
15. Rush A, Fava M, Wisniewski S, *et al*. Sequenced treatment alternatives to relieve depression (STAR*D): Rationale and design. *Control Clin Trials* 2004; **25**: 119–142.
16. Fava M, Rush A, Trivedi M, *et al*. Background and rationale for the Sequenced Treatment Alternatives to Reve Depression (STAR*D) study. *Psychiatr Clin North Am* 2003; **26**: 457–494.
17. Rush A, Trivedi M, Ibrahim H, *et al*. The 16-item Quick Inventory of Depressive Symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. *Biol Psychiatry* 2003; **54**(5): 573–583.
18. Lavori P, Dawson R. A design for testing clinical strategies: Biased adaptive within-subject randomization. *J R Stat Soc Ser A* 2000; **163**: 29–38.
19. Lavori P, Dawson R. Dynamic treatment regimes: Practical design considerations. *Clin Trials* 2004; **1**: 9–20.
20. Murphy S. Optimal dynamic treatment regimes (with discussions). *J R Stat Soc Ser B* 2003; **65**: 331–366.
21. Schneider L, Tariot P, Lyketsos C, *et al*. National Institute of Mental Health Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE): Alzheimer disease trial methodology. *Am J Geriatr Psychiatry* 2001; **9**(4): 346–360.
22. Stroup TS, McEvoy JP, Swartz MS, *et al*. The National Institute of Mental Health clinical antipschotic trials of intervention effectiveness (CATIE) project: schizophrenia trial design and protocol deveplopment. *Schizophr Bull* 2003; **29**: 15–31.
23. Swartz M, Perkins D, Stroup T, *et al*. Assessing clinical and functional outcomes in the Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) schizophrenia trial. *Schizophr Bull* 2003; **29**(1): 33–43.
24. Shortreed S, Moodie E. Estimating the optimal dynamic antipsychotic treatment regime: evidence from the sequential multiple-assignment randomized Clinical Antipsychotic Trials of Intervention and Effectiveness schizophrenia study. *J R Stat Soc Ser C Appl Stat* 2012; **61**(4): 577–599.
25. Thall P, Millikan R, Sung H. Evaluating multiple treatment courses in clinical trials. *Stat Med* 2000; **19**(8): 1011–1128.
26. Wahed A, Tsiatis A. Optimal estimator for the survival distribution and related quantities for treatment policies in two-stage randomized designs in clinical trials. *Biometrics* 2004; **60**(1): 124–133.
27. Chakraborty B, Murphy S, Strecher V. Inference for non-regular parameters in optimal dynamic treatment regimes. *Stat Methods Med Res* 2010; **19**(3): 317–343.
28. Strecher V, McClure J, Alexander G, *et al*. Web-based smoking cessation programs: Results of a randomized trial. *Am J Prev Med* 2008; **34**(5): 373–381.
29. Rich B, Moodie E, Stephens D. Simulating sequential multiple assignment randomized trials to generate optimal personalized warfarin dosing strategies. *Clin Trials* 2013, in press. doi: 10.1177/1740774513517063.
30. Sinyor M, Schaffer A, Levitt A. The Sequenced Treatment Alternatives to Relieve Depression (STAR*D) Trial: A Review. *Can J Psychiatry* 2010; **55**(3): 126–135.
31. Murphy S, Lynch K, Oslin D, *et al*. Developing adaptive treatment strategies in substance abuse research. *Drug Alcohol Depend* 2007; **88**: s24–s30.
32. Oetting A, Levy J, Weiss R, *et al*. 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*, Shrout P, Keyes K, Ornstein K (eds). American Psychiatric Publishing: Arlington, 2011; 179–205.
33. Almirall D, Compton S, Gunlicks-Stoessel M, *et al*. Preparing for a Sequential Multiple Assignment Randomized Trial for Developing an Adaptive Treatment Strategy: Designing a SMART Pilot Study. *Stat Med* 2012; **31**(17): 1887–1902.
34. Watkins C. Learning from Delayed Rewards. PhD dissertation, Cambridge University Press, Cambridge, England, 1989.
35. Sutton R, Andrew G. *Reinforcement learning: An introduction*. MIT press: Cambridge, 1998.
36. Zhao Y, Zeng D. Recent development on statistical methods for personalized medicine discovery. *Front Med* 2013; **7**(1): 102–110.
37. Lei H, Nahum-Shani I, Lynch K, *et al*. A "SMART" design for building individualized treatment sequences. *Annu Rev Clin Psychol* 2012; **8**: 21–48.
38. Nahum-Shani I, Qian M, Almirall D, *et al*. Q-learning: A data analysis method for constructing adaptive interventions. *Psychol Methods* 2012; **17**(4): 478.
39. Schulte PJ, Tsiatis AA, Laber EB, Davidian M. Q- and A-learning methods for estimating optimal dynamic treatment regimes. 2012. arXiv, 1202.4177v2.
40. Qian M, Nahum-Shani I, Murphy S. Dynamic treatment regimes. In *Modern Clinical Trial Analysis*, Tang W, Tu X (eds). Springer, 2013; 127–148.
41. Moodie E, Dean N, Sun Y. Q-learning: Flexible learning about useful utilities. *Stat Biosci* 2013, in press. DOI: 10.1007/s12561-013-9103-z.
42. Joffe M, Yang W, Feldman H. G-estimation and Artificial Censoring: Problems, Challenges, and Applications. *Biometrics* 2012; **68**(1): 275–286.
43. Cotton C, Heagerty P. A Data Augmentation Method for Estimating the Causal Effect of Adherence to Treatment Regimens Targeting Control of an Intermediate Measure. *Stat Biosci* 2011; **3**(1): 28–44.
44. Rosthøj S, Fullwood C, Henderson R, *et al*. Estimation of Optimal Dynamic Anticoagulation Regimes from Observational Data: A Regret-Based Approach. *Stat Med* 2006; **25**: 4197–4215.
45. Zhao Y, Zeng D, Socinski M, *et al*. Reinforcement learning strategies for clinical trials in nonsmall cell lung cancer. *Biometrics* 2011; **67**: 1422–1433.
46. Thall P, Wooten L, Logothetis C, *et al*. Bayesian and frequentist two-stage treatment strategies based on sequential failure times subject to interval censoring. *Stat Med* 2007; **26**: 4687–4702.
47. Rubin D. Discussion of randomized analysis of experimental data: The Fisher randomization test by D. Basu. *J Am Stat Assoc* 1980; **75**: 591–593.
48. Robins J, Wasserman L. Estimation of effects of sequential treatments by reparameterizing directed acyclic graphs. In *Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence*, Geiger D, Shenoy P (eds). Morgan Kaufmann: San Francisco, 1997; 409–430.