

# A SMART data analysis method for constructing adaptive treatment strategies for substance use disorders

Inbal Nahum-Shani<sup>1</sup>, Ashkan Ertefaie<sup>2</sup>, Xi (Lucy) Lu<sup>3</sup>, Kevin G. Lynch<sup>4</sup>, James R. McKay<sup>5</sup>, David W. Oslin<sup>6</sup> & Daniel Almirall<sup>1</sup>

Institute for Social Research, University of Michigan, Ann Arbor, Michigan, USA,<sup>1</sup> Department of Biostatistics and Computational Biology, University of Rochester, Rochester, New York, USA,<sup>2</sup> Department of Statistics, University of Michigan, Ann Arbor, Michigan, USA,<sup>3</sup> Treatment Research Center and Center for Studies of Addictions, Department of Psychiatry, University of Pennsylvania, Philadelphia, Pennsylvania, USA,<sup>4</sup> Center on the Continuum of Care in the Addictions, Department of Psychiatry and Philadelphia Veterans Administration Medical Center, University of Pennsylvania, Philadelphia, Pennsylvania, USA<sup>5</sup> and VISN 4 MIRECC, Cpl Michael J Crescenzo VA Medical Center and the Department of Psychiatry, University of Pennsylvania, Philadelphia, PA, USA<sup>6</sup>

## ABSTRACT

**Aims** To demonstrate how Q-learning, a novel data analysis method, can be used with data from a sequential, multiple assignment, randomized trial (SMART) to construct empirically an adaptive treatment strategy (ATS) that is more tailored than the ATSs already embedded in a SMART. **Method** We use Q-learning with data from the Extending Treatment Effectiveness of Naltrexone (ExTENd) SMART ( $N = 250$ ) to construct empirically an ATS employing naltrexone, behavioral intervention, and telephone disease management to reduce alcohol consumption over 24 weeks in alcohol dependent individuals. **Results** Q-learning helped to identify a subset of individuals who, despite showing early signs of response to naltrexone, require additional treatment to maintain progress. **Conclusions** Q-learning can inform the development of more cost-effective, adaptive treatment strategies for treating substance use disorders.

**Keywords** Adaptive interventions, adaptive treatment strategies, alcohol dependence, Q-learning, Sequential Multiple Assignment Randomized Trial (SMART), stepped-care.

Correspondence to: Inbal Nahum-Shani, Institute for Social Research, University of Michigan, 426 Thompson St., Ann Arbor, MI 48106, USA.

E-mail: inbal@umich.edu

Submitted 5 January 2016; initial review completed 3 June 2016; final version accepted 19 December 2016

## INTRODUCTION

The cyclical and heterogeneous nature of many substance use disorders highlights the need to adapt the type or the dose of treatment to accommodate the specific and changing needs of individuals [1–6]. Adaptive treatment strategies (ATSs) are suited for guiding this type of sequential and tailored treatment decision making [1,7–10]. An ATS is a *treatment design* in which treatment options are tailored based not only on baseline characteristics (e.g., patient demographics), but also on information that is likely to change over time in the course of treatment (e.g., early signs of non-response). This is similar to clinical decision making whereby care providers tailor the type/dose of treatment repeatedly in the course of clinical care based on ongoing information regarding patient progress in treatment. In an ATS, such tailoring is operationalized (formalized) via decision rules that specify what type/dose of treatment should be offered, for whom and when, so as

to enable replicability and evaluation. The sequential, multiple assignment, randomized trial (SMART) was developed specifically for constructing empirically-supported ATSs [11,12]. A SMART is an *experimental design*, which involves multiple stages of randomization. Each randomization stage provides an opportunity to inform how best to tailor the treatment at a specific stage of an ATS.

Several sequences of treatments of scientific interest are embedded in a SMART by design; these often use a single tailoring variable, such as the individual's early response status—offering different subsequent treatments to individuals who show early signs of non-response to initial treatments than to those who respond well. However, investigators are often interested in using data from a SMART to construct ATSs that are more tailored. That is, investigators often collect additional information concerning baseline (e.g., baseline severity) and time-varying status of individuals (e.g., adherence to treatment) and plan to use this information to investigate whether and how

**Table 1** Key terms and definitions.

<i>Key term</i>	<i>Definition</i>
Tailoring	Individualization, namely the use of information from the individual to select when and how to offer treatment.
Tailoring variables	Information concerning the individual that is used for individualization (i.e., to decide when and/or how to offer treatment).
Adaptive Treatment Strategy (ATS)	A treatment design in which treatment options are tailored not only based on baseline characteristics, but also based on time-varying information about the individual, namely information that is likely to change over time in the course of treatment. An ATS involves a sequence of decision rules. The decision rules link the treatment options and tailoring variables in a systematic way. ATSs are also known as Adaptive Interventions.
The Sequential, Multiple Assignment, Randomized Trial (SMART)	A multi-stage randomized trial. Participants progress through the stages and are potentially randomly assigned to one of several treatment options at each stage. Each stage of randomization is designed to address scientific questions concerning the type, dose, mode of delivery, or tailoring of treatments at a specific stage of an ATS. While most clinical trials are designed to evaluate or compare two or more treatments, SMART aims to provide data to construct and optimize an ATS.
Embedded Adaptive Treatment Strategies	Tailored sequences of treatments of scientific interest that are embedded in a SMART by design. These ATSs are often relatively simple, in that they use a single tailoring variable—the individual's early response status.
Q-learning Regression	A data analysis method drawn from computer science that can be used with SMART data to investigate whether and how certain covariates are useful for developing an ATS or improving an existing one. In other words, this method can be used to identify new tailoring variables beyond those used in a SMART by design. The 'Q' in Q-learning indicates that this method is used to assess the relative quality of different treatment options in a sequence of tailored treatments.

treatment could be further tailored according to these variables. Drawn from computer science, Q-learning [13,14] is a novel methodology that can be used for this purpose.

The current manuscript provides an overview of Q-learning to investigators in the area of substance use disorders (SUDs). Despite the growing use of SMART studies in the area of SUDs [15], there are no published applications of Q-learning in this area. Further, existing illustrations of Q-learning are mainly geared towards statisticians (e.g. [16,17]). To close this gap, we use data from the Extending Treatment Effectiveness of Naltrexone (ExTEND) trial—a 24-week study employing a SMART to inform the development of an ATS for supporting naltrexone medication in the treatment of alcohol dependence ( $N = 250$ ; D. Oslin, PI [18–20]). Previous analyses of data from ExTEND compared the relatively simple ATSs embedded in this SMART [18,21]. Here, these data are used for the first time with Q-learning to construct a more tailored ATS. The goal of this application is to demonstrate the scientific yield gained by applying Q-learning to inform the construction of ATSs for SUDs. Key terms and definitions are provided in Table 1.

## ADAPTIVE TREATMENT STRATEGIES

Consider the development of an ATS to treat alcohol dependence using oral naltrexone (NTX)—an opioid receptor

antagonist that blocks the pleasurable effects resulting from endogenous opioid neurotransmitters released by alcohol consumption in some people [22,23]. While NTX is efficacious for treating alcohol dependence, clinical use of NTX has been limited [24], in part because of substantial heterogeneity in treatment response [25]; this heterogeneity is attributed to multiple factors, such as poor adherence, biological response to alcohol and the medication, poor coping skills, and poor social support. Hence, a natural ATS might include treatment components aimed to address these multiple factors, such as the Combined Behavioral Intervention (CBI), an in-person intervention targeting adherence to pharmacotherapy, motivation for change, and coping skills; and telephone disease management (TDM), targeting similar factors via basic (minimal) telephone-delivered clinical support [26,27].

The following is an example ATS in this setting: At the first stage, alcohol dependent individuals are provided NTX, and their drinking behaviors are monitored weekly for eight weeks. At the second stage, the type of treatment is adapted based on the number of heavy drinking days (HDDs)<sup>1</sup> in the past week. Specifically, individuals who experience five or more HDDs during weeks two to eight are considered to be non-responding; as soon as an individual is non-responding, s/he enters the second stage and is offered a rescue intervention: switching to CBI. Individuals

<sup>1</sup>A heavy drinking day (HDD) is defined as four or more drinks per day for women and five or more for men.

who never experience five or more HDDs up to and including week eight (i.e., responders) are offered a maintenance intervention: adding TDM at the end of week eight. An ATS involves a sequence of decision rules; this ATS uses the decision rules in ATS #1 (Table 2).

The decision rules for this example ATS involve a single *tailoring variable*—the individual’s response status. Here, different second-stage treatments are offered to responders than to non-responders. The first-stage treatment and the criterion for non-response are not tailored; they are the same for all individuals.

Traditionally, the sequence of decision rules underlying ATSs used in practice are constructed based on clinical experience, empirical evidence and literature reviews. However, in many cases, there are open questions concerning the best treatment option at specific stages of an ATS, which tailoring variables to use, and how to best use them. For example, in the context of the example ATS above, there may be insufficient evidence to inform (a) the amount of drinking behavior that reflects non-response to NTX, (b) the type of rescue tactic that would be most useful for non-responders, and (c) the type of maintenance tactic that would be most useful in reducing the chance of relapse among responders. The SMART is a clinical trial design that can be used to efficiently obtain data to address scientific questions such as these.

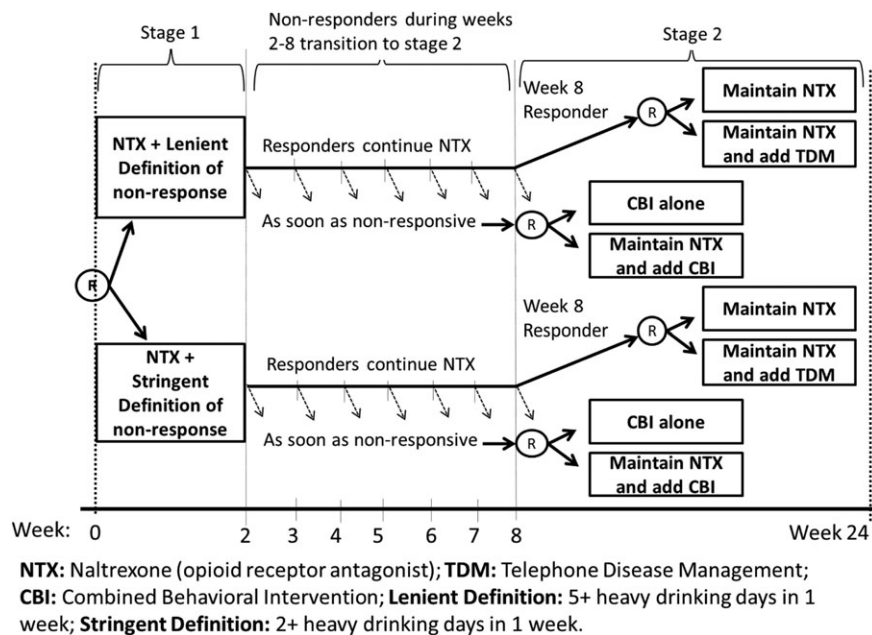
**THE SEQUENTIAL, MULTIPLE ASSIGNMENT, RANDOMIZED TRIAL (SMART)**

Consider the following simplified version of the ExTEND SMART (Fig. 1). In this 24-week trial, NTX was offered to all individuals. The first-stage randomization was to one of two criteria for early non-response: (1) a stringent criterion, in which an individual was classified as a non-responder as soon as s/he reported having two or more HDDs during the first eight weeks of NTX treatment; or (2) a lenient criterion, in which an individual was classified as a non-responder as soon as s/he reported having five or more HDDs during the first eight weeks of NTX treatment. Individuals were assessed weekly for drinking behavior. Starting at week two, as soon as the individual met his/her assigned criterion for non-response, s/he was immediately re-randomized to one of the two rescue tactics: (1) adding CBI (NTX + CBI) or (2) CBI alone (CBI). Individuals who did not meet their assigned non-response criterion by the end of week eight (i.e., responders), were re-randomized at that point (i.e., at week 8) to one of two maintenance tactics: (1) adding TDM (NTX + TDM) or (2) NTX alone (NTX). The primary outcomes were based on weekly assessments of the number of drinking days.

Eight ATSs are embedded in ExTEND (see [18]); one is described above (ATS#1). Each embedded ATS utilizes one tailoring variable—the individual’s early response

Table 2 Adaptive treatment strategies.

ATS #1	ATS #2: A more tailored ATS
<p>At entry into the program</p> <p>First-stage treatment = [NTX]</p> <p>At the end of every week from week 3 to 8 of the initial NTX treatment</p> <p>If response status = non-response (HDD ≥ 5)</p> <p>Then, second-stage treatment = [Switch to CBI immediately]</p> <p>Else if response status = response (HDD &lt; 5)</p> <p>Then, continue first-stage treatment, re-assess non-response in the following week, and at week 8 move to second-stage treatment = [add TDM]</p>	<p>At entry into the program</p> <p>Stage 1 treatment = [NTX]</p> <p>By the end of week 3 of initial NTX treatment</p> <p>Choose between a stringent or lenient criterion for weekly response/non-response</p> <p>At the end of every week from week 3 to 8 of the initial NTX treatment</p> <p>If response status = non-response</p> <p>Then, second-stage treatment = [offer CBI or NTX + CBI]</p> <p>Else if response status = response</p> <p>Then, continue stage 1 treatment and re-assess non-response in the following week, and, at week 8</p> <p>If the proportion of non-abstinence days during first-stage &gt; 10 %</p> <p>Then, move to second-stage treatment = [NTX + TDM]</p> <p>Else if the proportion of non-abstinence days during first-stage ≤ 10 %</p> <p>Then, move to second-stage treatment = [NTX + TDM or NTX alone]</p>
<p>Estimated percentage of abstinence days over the entire study duration among ExTEND participants following this ATS: 78 %</p>	<p>Estimated percentage of abstinence days over the entire study duration among ExTEND participants following this ATS: 77 %</p>



**Figure 1** ExTEND SMART study

status. This is because, by design, different second-stage treatments were offered to responders than to non-responders. Various methods can be employed to compare and select the best ATSs among the eight that are embedded in ExTEND (e.g. [28–30]). In previous analyses [18], ATS#1 was found to be the best among the eight embedded ATSs in terms of the probability of drinking during the second stage of treatment.

However, investigators often wish to explore whether other variables, beyond response status, could be tailoring variables. For example, in ExTEND, it would be useful to explore whether the non-response criterion should be tailored to the individual's baseline years of alcohol consumption. This is driven by empirical evidence suggesting that individuals with more severe histories of alcohol use problems are prone to faster relapse, requiring a more stringent definition of non-response [31,32]. Additionally, it would be useful to explore whether the maintenance tactic for responders should be based on the proportion of non-abstinence (i.e., any use) days during the initial NTX treatment. This is based on the idea that even in those categorized as responders, failure to achieve complete abstinence places the individual at greater risk for poor long-term outcomes, hence requiring additional support in order to maintain long-term improvement [33–35]. In the following section we demonstrate how Q-learning can be used to conduct these analyses.

## Q-LEARNING

Q-learning [13,14] is a multi-stage regression approach that can be used with data from a SMART to investigate

whether and how certain covariates are useful for developing an ATS or improving an existing one. Investigators first select a set of covariates at each stage that are hypothesized to be useful tailoring variables for the randomized treatment options at that stage. Such *candidate tailoring variables* may include any collection of baseline and time-varying variables measured prior to the randomization at each stage. In Q-learning, a regression is used at each stage to investigate whether and how the average treatment effect (i.e., the difference between treatment options) at that stage varies as a function of the candidate tailoring variables, while appropriately controlling for the effects of optimal *future* tailored treatments. Q-learning resembles moderated regression analyses [14], making it familiar and, therefore, easy to understand and implement. However, standard moderated regression analyses typically cannot be used to examine time-varying covariates as candidate tailoring variables for the purpose of empirically developing an ATS. For example, Nahum-Shani *et al.* [14] demonstrate how, compared to Q-learning, using a single standard moderated regression analysis to investigate time-varying candidate tailoring variables in a sequential treatments setting can lead to bias and, therefore, misleading conclusions.

Here we describe the application of Q-learning to data from the ExTEND SMART study. As discussed above, we examine the following two candidate tailoring variables: (a) the individual's baseline years of alcohol consumption, denoted  $O_{11}$ ; and (b) the proportion of non-abstinence days during the first stage, denoted  $O_{21}$ . Notice that  $O_{21}$  is an outcome of the first-stage treatment, rather than a baseline measure.



The outcome,  $Y$ , is the proportion of abstinence days over 24 weeks (high values are desirable). Let  $A_1$  denote the randomized non-response criteria at the first randomization, coded  $-1$  for the stringent criterion and  $1$  for lenient;  $A_{2R}$  denotes the randomized maintenance tactics for responders at the second randomization, coded  $-1$  for NTX alone and  $1$  for NTX + TDM; and  $A_{2NR}$  denotes the randomized rescue tactics for non-responders at the second randomization, coded  $-1$  for CBI alone and  $1$  for NTX + CBI. To apply Q-learning in this context, we use two regressions, one for each randomization stage. Technical details concerning the regression models are provided in Appendix S1; here, we provide a more accessible presentation.

The first regression of  $Y$  on terms involving the predictors ( $O_{11}$ ,  $A_1$ ,  $O_{21}$ ,  $A_{2R}$ ,  $A_{2NR}$ ) focuses on the effects of the second-stage randomized treatment options for responders ( $A_{2R}$ ), and non-responders ( $A_{2NR}$ ). This regression includes not only treatment effects for the maintenance ( $A_{2R}$ ) and rescue ( $A_{2NR}$ ) tactics, but also an interaction between  $A_{2R}$  and the proportion of non-abstinence days during the first stage ( $O_{21}$ ). This is because our goal is to assess the usefulness of  $O_{21}$  in tailoring the best maintenance tactic for responders ( $A_{2R}$ ). No candidate tailoring variables are considered for the rescue tactics ( $A_{2NR}$ ).

If the coefficient for the interaction between  $A_{2R}$  and  $O_{21}$  is different from zero, then the average effect of  $A_{2R}$  differs by levels of  $O_{21}$ . To further investigate how this effect varies, we attend to the conditional effect [36] of  $A_{2R}$ , namely the expected difference between responders offered NTX + TDM ( $A_{2R} = 1$ ) and those offered NTX alone ( $A_{2R} = -1$ ), for varying levels of  $O_{21}$ . This is a simple linear combination of the regression coefficients for  $A_{2R}$  and the interaction between  $A_{2R}$  and  $O_{21}$ . For a given level of  $O_{21}$ , if this conditional effect is positive, it means that the expected outcome is higher for responders offered NTX + TDM, compared to NTX alone; if it is negative, it means that the expected outcome is higher for responders offered NTX alone compared to NTX + TDM. Recall that higher values of  $Y$  (proportion of abstinence days over 24 weeks) are more desirable. Hence, for a given level of  $O_{21}$ , if the conditional effect of  $A_{2R}$  is positive, an ATS based on this model should recommend NTX + TDM for responders with this particular level of the tailoring variable; if the conditional effect is negative, the ATS should recommend NTX alone for responders with this particular level of the tailoring variable.

For non-responders, because no candidate tailoring variables are considered for the rescue tactics, we attend to the regression coefficient for  $A_{2NR}$  to select the best rescue tactic. If this coefficient is positive, it means that the expected outcome is higher for non-responders offered NTX + CBI ( $A_{2NR} = 1$ ), compared to CBI alone ( $A_{2NR} = -1$ ); hence, an ATS based on this model should

recommend NTX + CBI for non-responders. If the coefficient for  $A_{2NR}$  is negative, it means that the expected outcome is higher for non-responders offered CBI alone, compared to NTX + CBI; hence, the ATS should recommend CBI alone to non-responders.

After we have estimated the regression coefficients and assessed the evidence regarding the second-stage randomized treatment options and candidate tailoring variables, we move to a second regression that focuses on the effects of the first-stage randomized treatment options ( $A_1$ ). Here, we examine the effects of the non-response criteria ( $A_1$ ), presuming that in the future we would employ the optimal second-stage treatment options for responders and non-responders. This is done by using an adjusted  $Y$  that takes into account the optimal estimated second-stage tactics based on the first regression. This adjustment is straightforward; see Appendix S1.

The second regression of the adjusted  $Y$  on terms involving the predictors ( $O_{11}$ ,  $A_1$ ), includes not only the effect of the non-response criteria ( $A_1$ ), but also an interaction between  $A_1$  and the individual's baseline years of alcohol consumption ( $O_{11}$ ). This is because our goal is to assess the usefulness of  $O_{11}$  in tailoring the best non-response criteria ( $A_1$ ). If the coefficient for the interaction between  $A_1$  and  $O_{11}$  is different from zero, it means that the average effect of  $A_1$  differs by levels of  $O_{11}$ . To further understand how this effect varies, we attend to the conditional effect of  $A_1$ , namely the difference between the lenient ( $A_1 = 1$ ) and the stringent ( $A_1 = -1$ ) criterion, for varying levels of  $O_{11}$ . This is a simple linear combination of the regression coefficients for  $A_1$  and the interaction between  $A_1$  and  $O_{11}$ . For a given level  $O_{11}$ , if this conditional effect is positive, it means that the expected outcome under the optimal second-stage treatment option is higher for individuals receiving the lenient, rather than the stringent criterion; if the conditional effect is negative, it means that the expected outcome under the optimal second-stage treatment option is higher for individuals receiving the stringent, rather than the lenient criterion. Hence, for a given level of the tailoring variable, if the conditional effect of  $A_1$  is positive, it means that an ATS based on this regression model should recommend the lenient criterion for individuals with this particular level of the tailoring variable; if the conditional effect is negative, the ATS should recommend the stringent criterion for these individuals.

## ILLUSTRATIVE ANALYSIS OF THE EXTEND DATA

The procedure described above was implemented to analyze data from the ExtEND study, using the 'qllearning' package in R [37]. Information from a total of 250 study participants was used in this analysis. In both the first-

and second-stage regressions we included an indicator of gender (female =1) as a covariate. Appendix S2 provides information concerning the baseline characteristics of study participants. Appendix S3 provides information concerning the rates and patterns of missing values in the data. A multiple imputation method specifically adapted for SMART [38] was used to generate ten imputed datasets. Each imputed dataset was analyzed in the same way; results across the imputed datasets were summarized using standard formulae (see [39] for a more detailed discussion of multiple imputation methodologies).

### Measures

Weekly time-line follow-back (TLFB) assessments of the number of standard drinks recorded per day were used to calculate (a) the primary outcome ( $Y$ )—the proportion of abstinence days over the study duration—by dividing the number of non-drinking days by the duration of the study; and (b) the proportion of non-abstinence days during the initial NTX treatment ( $O_{21}$ ), by dividing the number of drinking days by the total number of days the participant was provided the initial NTX treatment (i.e., the total number of days in the first stage prior to re-randomization). Baseline years of alcohol consumption ( $O_{11}$ ) was self-reported by participants prior to initial randomization. Individuals were asked to indicate the number of years they consumed alcohol (i.e., any use) prior to entering the study. Appendix S2 includes sample distributions of variables.

### Results

Table 3 includes the results for the first regression, which focuses on the effects of the second-stage randomized treatment options for responders ( $A_{2R}$ ), and non-responders ( $A_{2NR}$ ). The interaction between  $A_{2R}$  and  $O_{21}$  (the candidate tailoring variable) was significantly different from zero ( $Estimate = .14$ ;  $CI = [.06, .24]$ ), indicating that the effect of maintenance tactics for responders varies depending on the proportion of non-abstinence days during initial NTX treatment ( $O_{21}$ ). Estimates of the conditional effect of  $A_{2R}$  for various levels of  $O_{21}$  indicate that for responders who consumed alcohol during 10 % or fewer of days during the initial NTX treatment, the conditional effect of  $A_{2R}$  is not significantly different from zero (e.g., for 10 % drinking days,  $Estimate = .04$ ;  $CI = [-.002, .06]$ ). However, for responders who consumed alcohol during more than 10 % of days during the initial NTX treatment, the conditional effect of  $A_{2R}$  was positive and significantly different from zero (e.g., for 20 % drinking days,  $Estimate = 0.06$ ;  $CI = [.02, .12]$ ). For these responders (33 % of all study responders), NTX + TDM leads to at least a 6 % increase in the percentage of abstinence days over the entire study duration (i.e., 10 days on average), relative to NTX alone. The coefficient of  $A_{2NR}$  was not significantly different from zero ( $Estimate = -.02$ ;  $CI = [-.06, .04]$ ), indicating inconclusive evidence with respect to the difference between the two rescue tactics for non-responders.

Table 4 includes the results for the second regression, which focuses on the effects of the first-stage randomized treatment options ( $A_1$ ). The results show that neither the effect of  $A_1$ , or the interaction between  $A_1$  and  $O_{11}$

**Table 3** Results for the first regression, which focuses on the effects of the second-stage randomized treatment options for responders ( $A_{2R}$ ), and non-responders ( $A_{2NR}$ ); and estimated conditional effects of the second-stage randomized treatment options for responders.

Parameter	Estimate	90 % CI <sup>a</sup>	
Intercept	-0.06	-0.15	-0.03
Gender	-0.03	-0.08	0.02
$O_{11}$ : Baseline years of alcohol consumption	-0.003	-0.02	0.02
$A_1$ : Non-response criterion	0.01	-0.01	0.03
$O_{21}$ : Proportion of drinking days during stage 1	-1.07	-1.20	-0.96
$A_{2R}$ : Maintenance tactic for responders	0.003	-0.02	0.02
$A_{2NR}$ : Rescue tactic for non-responders	-0.02	-0.06	0.04
$A_{2R} \times O_{21}$ : Maintenance tactic for responders x Proportion of drinking days during stage 1	0.14	0.06	0.24
<b>Estimated Conditional Effects of Maintenance Tactics</b>			
Percent drinking days during stage 1 = 0 (24 % of responders had $O_{21} = 0$ )	0.006	-0.04	0.04
Percent drinking days during stage 1 = 10 % (43 % of responders had $0 < O_{21} \leq 0.1$ )	0.04	-0.002	0.06
Percent drinking days during stage 1 = 20 % (14 % of responders had $0.1 < O_{21} \leq 0.2$ )	0.06	0.02	0.12
Percent drinking days during stage 1 = 30 % (10 % of responders had $0.2 < O_{21} \leq 0.3$ ; and 9 % had $0.3 < O_{21}$ )	0.10	0.04	0.16

<sup>a</sup>The estimated regression coefficients and associated lower and upper limit of the 90 % confidence intervals (CI) are summarized across ten imputed datasets. We set the Type I error rate to 0.10, rather than 0.05, given the illustrative nature of this analysis. Moreover, the aim of the analysis is to generate hypotheses about useful tailoring variables. Hence, from a clinical standpoint, it is sensible to tolerate a greater probability of detecting a false effect in order to improve the ability to detect true effects (see Collins *et al.* [50]; Dziak *et al.* [51]; McKay *et al.* [52]).

(the candidate tailoring variable) were significantly different from zero. Hence, evidence is inconclusive with respect to the difference between the two non-response criteria, as well as with respect to the usefulness of baseline years of alcohol consumption for tailoring the best non-response criterion.

The ATS proposed based on the results above is presented in Table 2 (ATS #2). The percentage of abstinence days over the entire study duration among individuals following ATS #2 was estimated to be 77 %. In additional analyses we found ATS #1 (Table 2) to be the best of the 8 ATSs embedded in ExTEND in terms of the proportion of abstinence days over the entire study duration; that is, consistent with previous studies [18], ATS #1 was the best among the 8 ATSs in which response status was used as the sole tailoring variable. The percentage of abstinence days over the entire study duration among participants following ATS #1 was estimated to be 78 %-- similar to ATS #2 which is more tailored.

**DISCUSSION**

The results of the Q-learning analysis suggests that ATS #2 is advantageous; this strategy recommends adding TDM only to a subset of responders, namely those for whom the percentage of non-abstinence days during the initial NTX treatment was larger than 10 %. Hence, while ATS #1 recommends that TDM should be added to *all* responders, the more tailored ATS (ATS #2) recommends TDM to only 33 % of responders. Although TDM is a cost-effective and potentially cost-saving strategy for treating SUDs compared to face-to-face alternatives [40], costs per session are estimated at \$30.24 for the client and \$30.55 for the health system [41]. Hence, employing the more tailored ATS (ATS #2) may result in substantially

lower cost of treatment, while achieving outcomes similar to ATS #1.

This has potential implications on both the scalability of the ATS as well as on treatment adherence. First, offering more costly treatments only to those who need it most should lead to greater cost-effectiveness, enhancing the scalability of substance abuse treatments [6]. Second, providing no more intervention than needed will reduce treatment burden, hence improve treatment adherence [42]. This demonstrates how ATSs generated by Q-learning have the potential to improve the treatment of SUDs beyond ATSs that are embedded in a SMART study, informing the development of more tailored treatment protocols that optimize outcomes while reducing cost and treatment burden [3]. As with standard linear regression, a potential limitation of Q-learning is that it requires that the linear models be correctly specified [16].

As with any study, the sample size for ExTEND was based on having adequate power to address its primary objective [43], which was to test the effect of CBI alone versus NTX + CBI among individuals who are non-responsive to NTX. Specifically, the total sample size ( $N = 300$ ) was selected to detect (at least) a moderate standardized difference of 0.55 [44] in change in the number of abstinent days following the second randomization between non-responders who were offered CBI alone vs. those offered NTX + CBI, with at least 80 % power (calculations were based on a Type-I error rate of 5 %, an exchangeable correlation structure assuming a within-person correlation of 0.5, 33 % non-response rate, and 15 % attrition rate by the end of the study). Sample size calculators exist for other primary objectives that are typical for SMART studies, such as the comparison of first-stage treatments or the comparison of ATSs that are embedded within the trial (e.g. [30]).

**Table 4** Results for the second regression, which focuses on the effects of the first-stage randomized treatment options ( $A_1$ ); and estimated conditional effects of the first-stage randomized treatment options.

Parameter	Estimate	90 % CI <sup>a</sup>	
Intercept	-0.21	-0.37	-0.19
Gender	-0.06	-0.15	0.02
$O_{11}$ : Baseline years of alcohol consumption	-0.0001	-0.02	0.03
$A_1$ : non-response criteria	0.01	-0.02	0.03
$A_1 \times O_{11}$ : Non-response criteria X Baseline years of alcohol consumption	-0.01	-0.02	0.01
<b>Estimated Conditional Effects of Non-Response Criterion</b>			
For patients with low number of baseline years of alcohol consumption (i.e., $O_{11} = -1$ , namely 1 SD below sample mean)	0.04	-0.02	0.10
For patients with high number of baseline years of alcohol consumption (i.e., $O_{11} = 1$ , namely 1 SD above sample mean)	0.004	-0.06	0.06

<sup>a</sup>The estimated regression coefficients and associated lower and upper limit of the 90 % confidence intervals (CI) are summarized across ten imputed datasets. We set the Type I error rate to 0.10, rather than 0.05, given the illustrative nature of this analysis. Moreover, the aim of the analysis is to generate hypotheses about useful tailoring variables. Hence, from a clinical standpoint, it is sensible to tolerate a greater probability of detecting a false effect in order to improve the ability to detect true effects (see Collins *et al.* [50]; Dziak *et al.* [51]; McKay *et al.* [52]).

As with standard randomized trials, investigating additional ways to tailor treatment (such as the investigation we have conducted here with Q-learning regression) is often a secondary objective in a SMART. Such an objective is often of great interest to investigators designing a SMART [45]. Yet given the exploratory (or hypothesis-generating) nature of this objective, it is still rare for the sample size in a SMART, or in any standard randomized trial, to be planned based on this objective. Recent methodological work in the single-stage setting using data from standard randomized trials has begun to address this gap [46].

## CONCLUSION

Our application of Q-learning demonstrates that even responders to SUD treatment can be heterogeneous, with some exhibiting less progress than others during treatment. This information can be useful in identifying those individuals who respond well initially, but require more support to maintain progress. These insights can contribute to the development of more cost-effective, stepped-care strategies for treating SUDs. Currently, there is growing interest in developing methods that go beyond Q-learning, as implemented here, to identify new ways to tailor treatments (e.g. [47–49]). The straightforward application of Q-learning provided here represents a promising first-step towards the implementation of other novel methodologies to empirically develop more tailored ATSS in SUDs.

## Declaration of interests

Drs. Nahum-Shani, Ertefaie, Almirall, Lynch, KcKay, and Lu reported no biomedical financial interests or potential conflicts of interest. Dr. Oslin has provided consultation to Otsuka Pharmaceuticals in the last year and is a paid consultant to the Hazelden Betty Ford Foundation.

## Acknowledgements

We thank Susan Murphy for her helpful feedback and advice. This study was supported by the following awards from the National Institutes of Health: R01-MH-080015, R03-MH-097954, R01-AA-014851, P01-AA016821, RC1-AA-019092, R01-DA-039901, P50-DA-039838, P50-DA-010075, R01-HD-073975, and U54-EB020404.

## References

- McKay J. R. *Treating substance use disorders with adaptive continuing care*. Washington, DC: American Psychological Association; 2009.
- Marlowe D. B., Festinger D. S., Dugosh K. L., Benasutti K. M., Fox G., Croft J. R. Adaptive programming improves outcomes in drug court: An experimental trial. *Crim Justice Behav* 2012; **39**: 514–32.
- Kranzler H. R., McKay J. R. Personalized treatment of alcohol dependence. *Curr Psychiatry Rep* 2012; **14**: 486–93.
- Black J. J., Chung T. Mechanisms of change in adolescent substance use treatment: How does treatment work? *Subst Abuse* 2014; **35**: 344–51.
- Witkiewitz K., Finney J. W., Harris A. H., Kivlahan D. R., Kranzler H. R. Recommendations for the design and analysis of treatment trials for alcohol use disorders. *Alcohol Clin Exp Res* 2015; **39**: 1557–70.
- Klostermann K., Kelley M. L., Mignone T., Pusateri L., Wills K. Behavioral couples therapy for substance abusers: Where do we go from here? *Subst Use Misuse* 2011; **46**: 1502–9.
- Almirall D., Nahum-Shani I., Sherwood N. E., Murphy S. A. Introduction to SMART designs for the development of adaptive interventions: with application to weight loss research. *Transl Behav Med* 2014; **4**: 260–74.
- Collins L. M., Murphy S. A., Bierman K. L. A conceptual framework for adaptive preventive interventions. *Prev Sci* 2004; **5**: 185–96.
- Lavori P. W., Dawson R. Adaptive treatment strategies in chronic disease. *Annu Rev Med* 2008; **59**: 443.
- Lavori P. W., Dawson R., Rush A. J. Flexible treatment strategies in chronic disease: clinical and research implications. *Biol Psychiatry* 2000; **48**: 605–14.
- Murphy S. A. An experimental design for the development of adaptive treatment strategies. *Stat Med* 2005; **24**: 1455–81.
- Lavori P. W., Dawson R. A design for testing clinical strategies: biased adaptive within-subject randomization. *J R Stat Soc A Stat Soc* 2000; **163**: 29–38.
- Watkins C. J. C. H. *Learning from delayed rewards*. Cambridge, UK: University of Cambridge; 1989.
- Nahum-Shani I., Qian M., Almirall D., Pelham W. E., Gnagy B., Fabiano G. A. et al. Q-learning: A data analysis method for constructing adaptive interventions. *Psychol Methods* 2012; **17**: 478.
- Projects Using SMARTs. The Methodology Center: The Pennsylvania State University; 2016. Available at: <https://methodology.psu.edu/ra/adap-inter/projects> (accessed 21 August 2016) (Archived at <http://www.webcitation.org/6k3FCHmin>).
- Schulte P. J., Tsiatis A. A., Laber E. B., Davidian M. Q- and A-learning methods for estimating optimal dynamic treatment regimes. *Stat Sci: a review journal of the Institute of Mathematical Statistics* 2014; **29**: 640.
- Moodie E. E., Dean N., Sun Y. R. Q-learning: Flexible learning about useful utilities. *Stat Biosci* 2014; **6**: 223–43.
- Lei H., Nahum-Shani I., Lynch K., Oslin D., Murphy S. A. “SMART” design for building individualized treatment sequences. *Annu Rev Clin Psychol* 2012; **8**.
- Qian M., Nahum-Shani I., Murphy S. A. Dynamic treatment regimes. In: *Modern Clinical Trial Analysis*. New York: Springer; 2013, pp. 127–48.
- Murphy S. A., Lynch K. G., Oslin D., McKay J. R., TenHave T. Developing adaptive treatment strategies in substance abuse research. *Drug Alcohol Depend* 2007; **88**: S24–30.
- Lu X., Nahum-Shani I., Kasari C., Lynch K. G., Oslin D. W., Pelham W. E. et al. Comparing dynamic treatment regimes using repeated-measures outcomes: modeling considerations in SMART studies. *Stat Med* 2016; **35**: 1595–615.
- Davidson D., Pallfai T., Bird C., Swift R. Effects of naltrexone on alcohol self-administration in heavy drinkers. *Alcohol Clin Exp Res* 1999; **23**: 195–203.



23. McCaul M. E., Wand G. S., Stauffer R., Lee S. M., Rohde C. A. Naltrexone dampens ethanol-induced cardiovascular and hypothalamic–pituitary–adrenal axis activation. *Neuropsychopharmacology* 2001; **25**: 537–47.
24. Yoon G., Kim S. W., Thuras P., Westermeyer J. Safety, tolerability, and feasibility of high-dose naltrexone in alcohol dependence: an open-label study. *Human Psychopharmacol Clin Exp* 2011; **26**: 125–32.
25. Pettinati H. M., Volpicelli J. R., Pierce J. D. Jr., O'Brien C. P. Improving naltrexone response: an intervention for medical practitioners to enhance medication compliance in alcohol dependent patients. *J Addict Dis* 2000; **19**: 71–83.
26. Longabaugh R., Zweben A., LoCastro J. S., Miller W. R. Origins, issues and options in the development of the combined behavioral intervention. *J Stud Alcohol Drugs* 2005; **15**: 179.
27. Miller W. R., Moyers T. B., Arciniega L. T., DiClemente C., LoCastro J., Longabaugh R. *et al.* A combined behavioral intervention for treating alcohol dependence. *Alcohol Clin Exp Res* 2003; **27**: 113A.
28. Ertefaie A., Wu T., Lynch K. G., Nahum-Shani I. Identifying a set that contains the best dynamic treatment regimes. *Biostatistics* 2015; **17**: 135–48.
29. Nahum-Shani I., Qian M., Almirall D., Pelham W. E., Gnagy B., Fabiano G. A. *et al.* Experimental design and primary data analysis for developing adaptive interventions. *Psychol Methods* 2012; **17**: 457–77.
30. Oetting A., Levy J., Weiss R., Murphy S. Statistical methodology for a SMART design in the development of adaptive treatment strategies. In: Shrout P., editor. *Causality and psychopathology: finding the determinants of disorders and their cures*. Arlington VA: American Psychiatric Publishing; 2007, pp. 179–205.
31. Rando K., Hong K.-I., Bhagwagar Z., Li C. S., Bergquist K., Guarnaccia J. *et al.* Association of Frontal and Posterior Cortical Gray Matter Volume With Time to Alcohol Relapse: A Prospective Study. *Am J Psychiatry* 2011; **168**: 183–92.
32. Heilig M., Egli M. Pharmacological treatment of alcohol dependence: target symptoms and target mechanisms. *Pharmacol Ther* 2006; **111**: 855–76.
33. McKay J. R., Van Horn D. H., Lynch K. G., Ivey M., Cary M. S., Drapkin M. L. *et al.* An adaptive approach for identifying cocaine dependent patients who benefit from extended continuing care. *J Consult Clin Psychol* 2013; **81**: 1063.
34. Cable N., Sacker A. Typologies of alcohol consumption in adolescence: predictors and adult outcomes. *Alcohol Alcohol* 2008; **43**: 81–90.
35. McKay J. R., Lynch K. G., Shepard D. S., Pettinati H. M. The effectiveness of telephone-based continuing care for alcohol and cocaine dependence: 24-month outcomes. *Arch Gen Psychiatry* 2005; **62**: 199–207.
36. Jaccard J., Turrissi R. *Interaction effects in multiple regression*, 2nd edn. Thousand Oaks, CA: Sage; 2003.
37. Qian M., Nahum-Shani I., Kaur A., Ertefaie A., Almirall D., Murphy S. A. R Code for Using Q-Learning to Construct Adaptive Interventions Using Data from a SMART: Instructions for Using the Q-learning Package in R. The Methodology Center, Penn State. 2013. Available at: <https://methodology.psu.edu/ra/adap-treat-strat/qlearning> (accessed 3 February 2017) (Archived at <http://www.webcitation.org/6o0QfTqik>).
38. Shortreed S. M., Laber E., Lizotte D. J., Stroup T. S., Pineau J., Murphy S. A. Informing sequential clinical decision-making through reinforcement learning: an empirical study. *Mach Learn* 2011; **84**: 109–36.
39. Sterne J. A., White I. R., Carlin J. B., Spratt M., Royston P., Kenward M. G. *et al.* Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ* 2009; **338**: b2393.
40. Shepard D. S., Daley M. C., Neuman M. J., Blakman A. P., McKay J. R. Telephone-based continuing care counseling in substance abuse treatment: economic analysis of a randomized trial. *Drug Alcohol Depend* 2016; **159**: 109–16.
41. McCollister K., Yang X., McKay J. R. Cost-effectiveness analysis of a continuing care intervention for cocaine-dependent adults. *Drug Alcohol Depend* 2016; **158**: 38–44.
42. Heckman B. W., Mathew A. R., Carpenter M. J. Treatment burden and treatment fatigue as barriers to health. *Curr Opin Psychol* 2015; **5**: 31–6.
43. Piantadosi S. *Clinical trials: a methodologic perspective*. New York, NY: John Wiley & Sons; 2013.
44. Cohen J. *Statistical power for the behavioral sciences*. Hillsdale, NJ: Erlbaum; 1988.
45. Wallace M. P., Moodie E. E., Stephens D. A. SMART thinking: a review of recent developments in sequential multiple assignment randomized trials. *Curr Epidemiol Rep* 2016; **3**: 225–32.
46. Laber E. B., Zhao Y. Q., Regh T., Davidian M., Tsiatis A., Stanford J. B. *et al.* Using pilot data to size a two-arm randomized trial to find a nearly optimal personalized treatment strategy. *Stat Med* 2016; **35**: 1245–56.
47. Laber E., Zhao Y. Tree-based methods for individualized treatment regimes. *Biometrika* 2015; **102**: 501–14.
48. Zhao Y.-Q., Zeng D., Laber E. B., Song R., Yuan M., Kosorok M. R. Doubly robust learning for estimating individualized treatment with censored data. *Biometrika* 2015; **102**: 151–68.
49. Qian M., Murphy S. A. Performance guarantees for individualized treatment rules. *Ann Stat* 2011; **39**: 1180.
50. Collins L. M., Dziaik J. J., Li R. Design of experiments with multiple independent variables: A resource management perspective on complete and reduced factorial designs. *Psychol Methods* 2009; **14**: 202.
51. Dziaik J. J., Nahum-Shani I., Collins L. M. Multilevel factorial experiments for developing behavioral interventions: Power, sample size, and resource considerations. *Psychol Methods* 2012; **17**: 153–75.
52. McKay J. R., Lynch K. G., Shepard D. S., Morgenstern J., Forman R. E., Pettinati H. M. Do patient characteristics and initial progress in treatment moderate the effectiveness of telephone-based continuing care for substance use disorders? *Addiction* 2005; **100**: 216–26.

## Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

**Appendix S1** Details concerning Q-learning.

**Appendix S2** Sample characteristics.

**Appendix S3** Dropout versus non-dropout missing rates in ExTEND.