PENNSTATE

The Methodology Center
advancing methods, improving health

qlaci R Package

for Using Q-Learning to Construct Adaptive Interventions Using Data From a SMART

Ashkan Ertefaie
Kun Deng
Aaron T. Wagner
Susan A. Murphy

Please send questions and comments to *MChelpdesk@psu.edu*.

The development of this R package was supported by the National Institute on Drug Abuse Grant P50-DA10075 to The Center for Prevention and Treatment Methodology.

The suggested citation for this vignette is

Ertefaie, A., Deng, K., Wagner, A. T., & Murphy, S. A. (2014). qlaci R package for using q-learning to construct adaptive interventions using data from a SMART (Version 1.0). University Park: The Methodology Center, Penn State. Available from methodology.psu.edu.

Contents

1 Overview

This document describes how to use the **qlaci** package in R. **qlaci** implements a generalization of Q-learning, a method developed in computer science, to inform the development of adaptive interventions. An adaptive intervention is an individualized sequence of treatments. The individualization occurs via decision rules; the decision rules indicate how to adjust treatment over time in response to an individual's symptoms, side effects or other information. Q-Learning generalizes standard regression methods for use in developing the decision rules underlying an adaptive intervention.

Here we describe how to use **qlaci** with data from a sequential, multiple assignment, randomized trial (SMART). SMARTs are clinical trial designs that generate high-quality data explicitly for the purpose of developing adaptive interventions. **qlaci** can be used to analyze data from a SMART involving two treatment stages.

qlaci requires free, open-source, R statistical software, available from http://www.r-project.org/

1.1 Adaptive Interventions

An adaptive intervention is composed of a sequence of decision rules that specify whether, how, or when to alter the intensity, type, or delivery of treatment at decision stages in the health care process. Adaptive interventions aim to take advantage of heterogeneity in response to treatments in order to maximize individual health outcomes. They do this by adapting and re-adapting treatments to the individual, over time, based on observations made on the individual. Adaptive interventions are also known as dynamic treatment regimes (Robins, 1986; Murphy, Van Der Laan, & Robins, 2001), adaptive treatment strategies (Lavori, Dawson, & Rush, 2000; Murphy, 2005), multi-stage treatment strategies (Thall, Sung, & Estey, 2002), and treatment policies (Lunceford, Davidian, & Tsiatis, 2002; Wahed & Tsiatis, 2004; Wahed & Tsiatis, 2006).

There are three important parts of an adaptive intervention: treatment options, tailoring variables, and decision rules. Tailoring variables (information concerning the individual) serve as inputs to the decision rules; these variables indicate responsiveness to (or can be used to

determine the need for) treatment type or dosage. Decision rules use the tailoring variables to recommend an individualized treatment option.

There are two types of tailoring variables: baseline and time-varying. Baseline tailoring variables may include age, comorbidities, race, gender, treatment history, measures of severity, and contextual risk or protective factors. Baseline tailoring variables are observed prior to treatment and are used to make the first treatment decision. Time-varying tailoring variables are observed during treatment and are used to make subsequent treatment decisions. Because we only consider two-stage adaptive interventions in this package, variables observed up to and at the end of first-stage treatment are potential time-varying tailoring variables. Potential tailoring variables include adherence to first-stage treatment and response to first-stage treatment. An example of a simple adaptive intervention is provided in Figure 1.

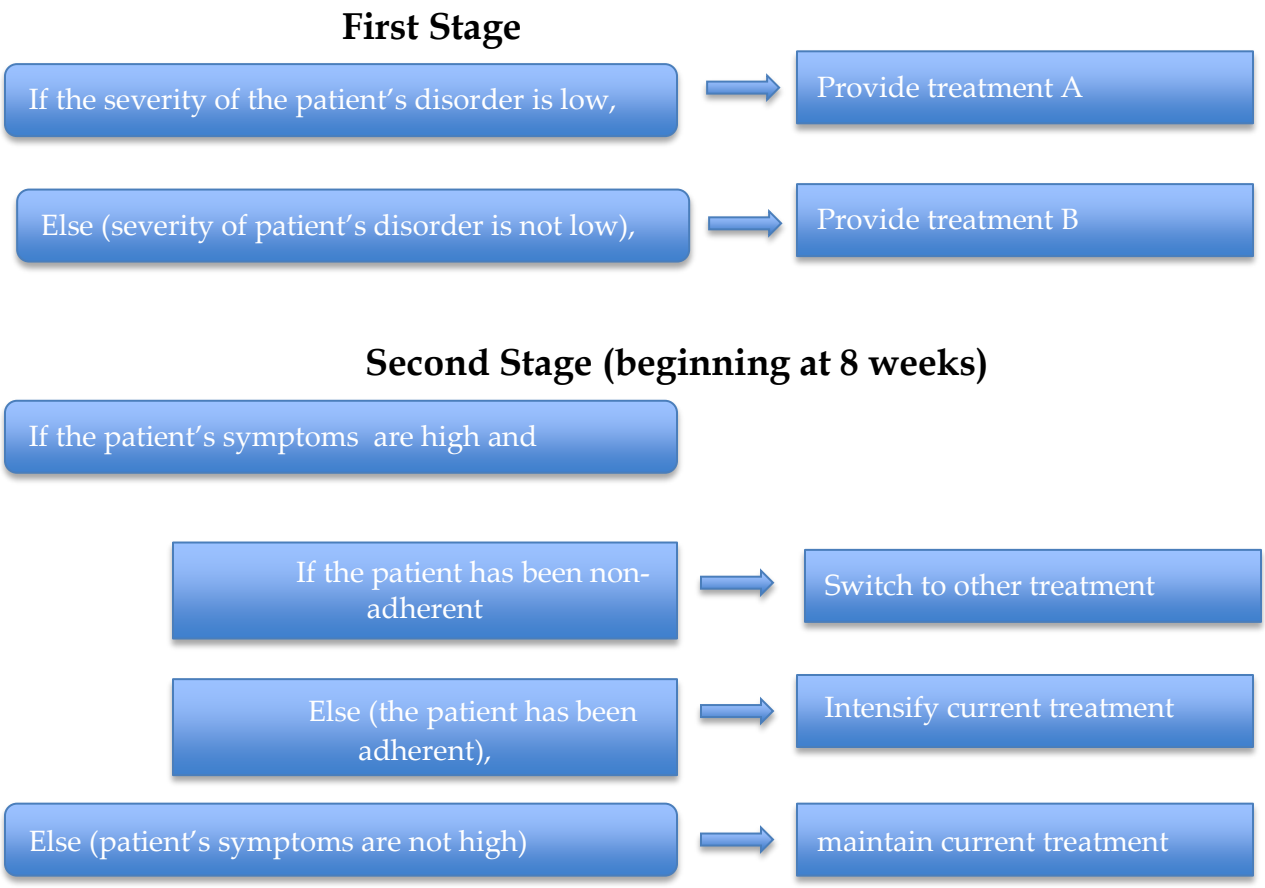

Figure 1: An Example Adaptive Intervention.

In this hypothetical example,

- patient severity is the baseline tailoring variable,
- treatments *A* and *B* are first-stage treatment options,
- symptoms and adherence are the time-varying tailoring variables, and
- intensifying and switching are second-stage treatment options.

1.2 Sequential, Multiple Assignment, Randomized Trial (SMART)

The SMART is a clinical trial design in which each individual proceeds through stages of treatment. At each treatment stage, individuals are randomized to one of the available treatment options at that stage. For example, at the first-stage, all individuals may be randomized to treatment *A* versus treatment *B*. In the second-stage (following the first-stage treatment), responders, non-responders, a subset of responders and/or non-responders, or all individuals may be randomized to second-stage treatment options. Randomizations may be restricted based on ethical concerns, feasibility, availability, or suitability of different sets of treatment options. For more information on SMART, see Murphy (2005), Nahum-Shani et al., (2012a), Lei et al., (2012) and Almirall et al., (2012).

1.3 Q-learning

Q-learning (Watkins, 1989; Watkins & Dayan, 1992) is a generalization of regression to multiple stages of decision making. Q-learning, as implemented in **qlaci**, can be used with data from a SMART to help build an optimal adaptive intervention. The optimal adaptive intervention should lead to the best sum of the stage 1 outcome and stage 2 outcome.

In Q-learning, the second-stage regression is implemented first. The estimated regression coefficients are then used to select the best second-stage treatment. Next, using the selected second-stage treatment, a first-stage regression is implemented (Nahum-Shani et al., 2012b). The estimated regression coefficients from the first-stage regression are then used to select the best first-stage treatment. The backward ordering of regressions used by Q-learning avoids selecting treatment options that appear to be optimal in the short term but may lead to an undesirable or less desirable primary outcomes.

2 Technical Details

2.1 Data Structure

qlaci can be used to analyze longitudinal data arising from a SMART with two decision stages. The observed data on each individual are given by trajectory (O_1, A_1, Y_1, O_2, S, A_2, Y_2). O_i, for i=1,2 is a set of covariates available at the beginning of the ith stage. A_i denotes the treatment options in the ith stage. **qlaci** requires at most two treatment options at each decision stage; this implementation of **qlaci** requires that A_i is coded as -1 and 1 (contrast coding). S is a binary variable coded as 1 if an individual has been rerandomized at stage 2 and 0 otherwise. Y_1 is the stage 1 outcome observed after the stage 1 treatment and Y_2 is the stage 2 outcome, observed after stage 1 treatment. This implementation of qlaci assumes that Y_1 and Y_2 are observed for everyone (if there is no Y_1 in the study then Y_1 should be set to 0) and that both Y_1 and Y_2 are continuous. The vectors O_1 and O_2 are called baseline covariates and intermediate outcomes, respectively. O_1 and O_2 include candidate baseline and time-varying tailoring variables, respectively. Define the history at each stage as $H_1 = O_1$ and $H_2 = (O_1, A_1, Y_1, O_2)$. Note that Y_1 can be a part of O_2 and thus can be a candidate time-varying tailoring variable.

2.2 Estimation procedures

In **qlaci**, the regression coefficients of the fitted models are estimated using the least squares method, and the confidence intervals of the parameters are estimated using a bootstrap technique developed in Laber and Murphy (2011; see also Laber, Qian, Lizotte, & Murphy, 2010).

2.3 Missing Data

Missing covariates and/or missing primary outcomes are not permitted. In particular, the second-stage regression requires that individuals with S=1 do not have any missing values in H_2, A_2 or the stage 2 outcome Y_2. Also, the first-stage regression requires that individuals with S=0 do not have any missing data in H_1, A_1 , Y_1 or Y_2 (if there is no Y_1 in the study then Y_1 should be set to 0). The variable S also needs to be observed for everyone.

3 Q-learning: Statistical Models

As implemented by **qlaci**, Q-learning fits a linear regression model at each decision stage. The regressions are fit sequentially starting from the second-stage, working backward to the first-stage. At each stage, the estimated regression function is used to evaluate usefulness of candidate tailoring variables and to find the treatment option at that stage that maximizes the outcome. We explain how Q-learning is applied in studies with two stages.

To fit the second-stage model, only data from individuals who are re-randomized is used. The stage 2 outcome is regressed on the intermediate outcomes, baseline covariates, and the first and second-stage treatments; this results in the estimated second-stage regression function, Q_2. The second-stage treatment option that maximizes the estimated second-stage regression function is derived (see below). Next, data from all individuals are used to fit the first-stage model by regressing the appropriate outcome (see below) on the baseline covariates and the first-stage treatment; this results in the estimated first-stage regression function, Q_1. Lastly, the first-stage treatment option that maximizes the estimated first-stage regression function is derived (see below).

3.1 Regression Models

qlaci uses linear regression models for Q_1 and Q_2 as follows.

1. The second-stage model (using the data from individuals who are rerandomized at stage 2) is

$$Q_2(H_2, A_2; \beta_{21}, \beta_{22}) = \beta_{21} H_{21} + \beta_{22} H_{22} A_2, \quad (0)$$

where H_{21} and H_{22} are vectors of selected covariates from the individual's history at the second-stage, H_2. Also, β_{21} and β_{22} are vectors of parameters. H_{21} includes a "1" as the first element; this is the intercept for the model. H_{22} also includes "1" as the first element, so that the first parameter in β_{22} represents the main effect of the second-stage treatment. The parameters of this model are estimated by regressing the stage 2 outcome Y_2 on $(H_{21}, H_{22}A_2)$. This results in estimators, $\hat{\beta}_{21}$, $\hat{\beta}_{22}$ and estimated regression function,

$$Q_2(H_2, A_2; \hat{\beta}_{21}, \hat{\beta}_{22}) = \hat{\beta}_{21} H_{21} + \hat{\beta}_{22} H_{22} A_2.$$

The estimated interaction parameters $\hat{\beta}_{22}$ are crucial to identifying the tailoring variables. Since, in **qlaci**, we are using contrast coding (-1,+1) for A_2, the second-stage treatment option that maximizes $Q_2(H_2, A_2; \hat{\beta}_{21}, \hat{\beta}_{22})$ is derived as follows: if $\hat{\beta}_{22} H_{22} > 0$ then the stage 2 treatment coded by 1 is the best; if $\hat{\beta}_{22} H_{22} < 0$ then the stage 2 treatment coded by -1 is the best. *This is the stage 2 decision rule.* Another way to write this is $d_2(H_2) = sign(\hat{\beta}_{22} H_{22})$.

From this regression, the estimated stage 2 outcome if the optimal treatment were taken at stage 2 is

$$\hat{Y}_2^{opt} = \max_{a_2} Q_2(H_2, a_2; \hat{\beta}_{21}, \hat{\beta}_{22}) = \hat{\beta}_{21} H_{21} + \max_{a_2} \{\hat{\beta}_{22} H_{22} a_2\}$$

Since $A_2 \in \{-1, +1\}$, the preceding equation can be written as

$$\hat{Y}_2^{opt} = \max_{a_2} Q_2(H_2, a_2; \hat{\beta}_{21}, \hat{\beta}_{22}) = \hat{\beta}_{21} H_{21} + |\hat{\beta}_{22} H_{22}|. \quad (0)$$

The dependent variable for the first-stage regression, $\tilde{Y}$, is

$$Y_1 + S\hat{Y}_2^{opt} + (1 - S)Y_2. \quad (0)$$

Note: In studies with no stage 1 outcome Y_1, set Y_1=0 for everyone (see the Appendices).

2. The first-stage model (fit using data from all individuals) is

$$Q_1(H_1, A_1; \beta_{11}, \beta_{12}) = \beta_{11} H_{11} + \beta_{12} H_{12} A_1,$$

where H_{11} and H_{12} are vectors of selected covariates from the individual's baseline data, H_1. H_{11} includes a "1" as the first element; this is the intercept for the model. H_{12} also includes "1" as the first element, so that the first parameter in β_{12} represents the main effect of the first-stage treatment. The parameters of this model are estimated by regressing the outcome $\tilde{Y}$ on $(H_{11}, H_{12} A_1)$. This results in estimators $\hat{\beta}_{11}$, $\hat{\beta}_{12}$ and the estimated regression function

$$Q_1(H_1, A_1; \hat{\beta}_{11}, \hat{\beta}_{12}) = \hat{\beta}_{11} H_{11} + \hat{\beta}_{12} H_{12} A_1.$$

Similar to the second-stage, the estimated optimal treatment at the first-stage is as follows: if

$\hat{\beta}_{12}H_{12} > 0$, then the treatment coded 1 is the best; if $\hat{\beta}_{12}H_{12} < 0$, then the treatment coded -1 is the best. *This is the stage 1 decision rule.* Another way to write this is $d_1(H_1) = sign(\hat{\beta}_{12}H_{12})$.

The regression parameters in Q_j, for $j = 1, 2$, are consistently estimated (unbiased in large samples) if the linear models are correctly specified (Laber, Qian, Lizotte, & Murphy, 2010). Confidence intervals for stage 1 and stage 2 parameters are described in section 3.3.

3.2 Contrast Matrix for First- and Second-Stage Regressions

qlaci allows users to specify both a stage 1 and a stage 2 contrast matrix. These matrices are used to estimate a "contrast," that is, a linear combination of either the stage 1 or stage 2 regression parameters. For example, a contrast for stage 1 parameters can be used to estimate the mean optimized outcome (mean of $\tilde{Y}$) for individuals with particular covariate values or can be used to estimate differences in the mean optimized outcome. The number of columns in contrast matrices for stage 1/2 are equal to the number of regression parameters at the stages 1/2, respectively, and the number of rows is the number of contrasts that a user wishes to estimate. Below, we explain how to specify and interpret a contrast matrix for the first-stage regression. The second-stage contrast matrix can be specified similarly.

As an example, suppose the first-stage estimated model is

$$\hat{\beta}_{11}H_{11} + \hat{\beta}_{12}H_{12}A_1,$$

where $H_{11} = (1, O_{11}, O_{12}, O_{13})^T$, $H_{12} = (1, O_{13})^T$ and A_1 is the treatment indicator at the first-stage (coded as -1 and +1). Since we have six parameters in this model, the contrast matrix must have six columns. In the following example, we decide to estimate four contrasts, so our contrast matrix has four rows. Let the 4×6 contrast matrix C_1 be

$$C_1 = \begin{pmatrix} 1 & 0 & 0 & 1 & -1 & -1 \\ 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & -2 & -2 \\ 0 & 0 & 0 & 0 & -2 & 0 \end{pmatrix}$$

The first row of C_1 estimates $\beta_{11}(1,0,0,1)^T + \beta_{12}(-1,-1)^T$ which is the mean response, $\tilde{Y}$, among individuals with $O_{11} = 0$, $O_{12} = 0$, $O_{13} = 1$, and $A_1 = -1$ (therefore, $A_1 O_{13} = -1$). The second row

of C_1 estimates $\beta_{11}(1,0,0,1)^T + \beta_{12}(1,1)^T$, which is the mean response, $\tilde{Y}$, among individuals with $O_{11} = 0$, $O_{12} = 0$, $O_{13} = 1$, and $A_1 = 1$ (therefore, $A_1 O_{13} = 1$). The third contrast estimates the difference between the first two contrasts. More specifically, the third row of the C_1 estimates $[\beta_{11}(1,0,0,1)^T + \beta_{12}(-1,-1)^T] - [\beta_{11}(1,0,0,1)^T + \beta_{12}(1,1)^T] = \beta_{11}(0,0,0,0)^T + \beta_{12}(-2,-2)^T$ which is the difference in mean response between individuals with $A_1 = -1$ and $A_1 = 1$ when $O_{13} = 1$. The last row estimates $[\beta_{12}(1,0,0,0)^T + \beta_{12}(-1,0)^T] - [\beta_{11}(1,0,0,0)^T + \beta_{12}(1,0)^T] = \beta_{11}(0,0,0,0)^T + \beta_{12}(-2,0)^T$, which is the difference in mean response between individuals with $A_1 = -1$ and $A_1 = 1$ when $O_{11} = 0$, $O_{12} = 0$, $O_{13} = 0$.

Comparing the estimated contrast obtained from the first and the second rows of this matrix allows us to estimate which first-stage treatment is best for individuals with $O_{11} = 0$, $O_{12} = 0$, $O_{13} = 1$. More specifically, if the mean response given $O_{11} = 0$, $O_{12} = 0$, $O_{13} = 1$, $A_1 = -1$ is greater than the mean response given $O_{11} = 0$, $O_{12} = 0$, $O_{13} = 1$, $A_1 = +1$, then the group of individuals with characteristics $O_{11} = 0$, $O_{12} = 0$, $O_{13} = 1$ is estimated to benefit more from the treatment coded as $A_1 = -1$ than from the treatment coded as $A_1 = 1$.

The third and fourth contrasts can be used to test whether a group of individuals benefit from one treatment option as compared to the other treatment option. If the confidence interval corresponding to the contrast in the third row contains zero, then there is no evidence that individuals with $O_{11} = 0$, $O_{12} = 0$, and $O_{13} = 1$ respond differently to the two different treatments. However, if the confidence interval lies entirely on the positive side of the real line, then we can conclude that the treatment coded as -1 leads to a better mean response among individuals with $O_{11} = 0$, $O_{12} = 0$, and $O_{13} = 1$, than the treatment coded 1. The last row can be interpreted in a similar manner to the third contrast; however this contrast is for the group of individuals with $O_{11} = 0$, $O_{12} = 0$, and $O_{13} = 0$.

Remark: The intermediate variable O_{13} can be potentially used to tailor the first-stage treatment if the confidence intervals of the two estimated contrasts obtained from the third and last rows lie entirely on the different sides of the real line.

3.3 Confidence Intervals

qlaci uses a bootstrap percentile method to construct confidence intervals for the second-stage parameters (Efron & Tibshirani (1993)). Confidence intervals for the parameters in the first-stage model Q_1 (see (1)), however, require a generalization of the bootstrap. Recall (see (2) and (3)) that the parameters in Q_1 are estimated by regressing $\tilde{Y}$ on the covariates in H_1. Non-differentiability of the absolute value in (2) at zero (when $\beta_{22}H_{22}A_2 = 0$) can result in poor performance of standard bootstrap-based confidence intervals (see Robins, 2004; Moodie and Richardson, 2010; and Chakraborty et. al., 2010).

qlaci uses a bootstrap-based confidence interval called an adaptive confidence interval (ACI) developed by Laber et al. (2010) to construct the confidence intervals. The ACI is formed by partitioning the individuals in sets for which the estimated stage 2 treatment effect (e.g., $\hat{\beta}_{22}H_{22}$) is near zero and for which the estimated stage 2 effect is far from zero and constructing smooth upper and lower bounds using the two partitions. The upper and lower bounds are calculated using a grid search method. The number of bootstraps, grid points and the limits of the grid search for each parameter are specified by the arguments **nb**, **ngrid** and **gridscale**, respectively. Refer to the R help using "?qlaci" command for more information about these arguments. The upper and lower bounds are then bootstrapped to form the confidence interval. As described in section 3.2, two-sided hypothesis tests can also be performed for the parameters of the first-stage model by checking whether the corresponding confidence intervals contain zero.

4 Example: Simulated SMART for Children With Minimal Verbal Ability

4.1 Design and Data Structure

Figure 2 shows an example SMART that might be used to develop an adaptive intervention to improve spoken communication in children who are minimally verbal and have been diagnosed with autism spectrum disorder (ASD). A data set was simulated of n=200 children following this SMART design. The simulated data set is called ASDdat.

To understand the structure of the simulated data, consider the following study design features.

- At stage 1, all children are randomized to either the joint attention/joint engagement (JAE) treatment combined with the enhanced milieu teaching (EMT) treatment (coded as A_1=+1) or the joint attention/joint engagement (JAE) treatment combined with the augmentative and alternative communication (AAC) treatment (coded as A_1=-1).
- After 3 months, all children are classified as responders or non-responders to the stage 1 treatment based on whether they meet a pre-specified criterion for improvement in spoken communication.
- At stage 2, responders continue with their initial treatment for an extra 3 months. Children who did not respond to JAE+AAC continue with intensified JAE+AAC. Non-responders to JAE+EMT are randomized to either JAE+AAC (coded as A_2=-1) or intensified JAE+EMT (coded as A_2=+1).
- After 6 months, the number of different spontaneous words is assessed (the stage 2 outcome, Y_2).

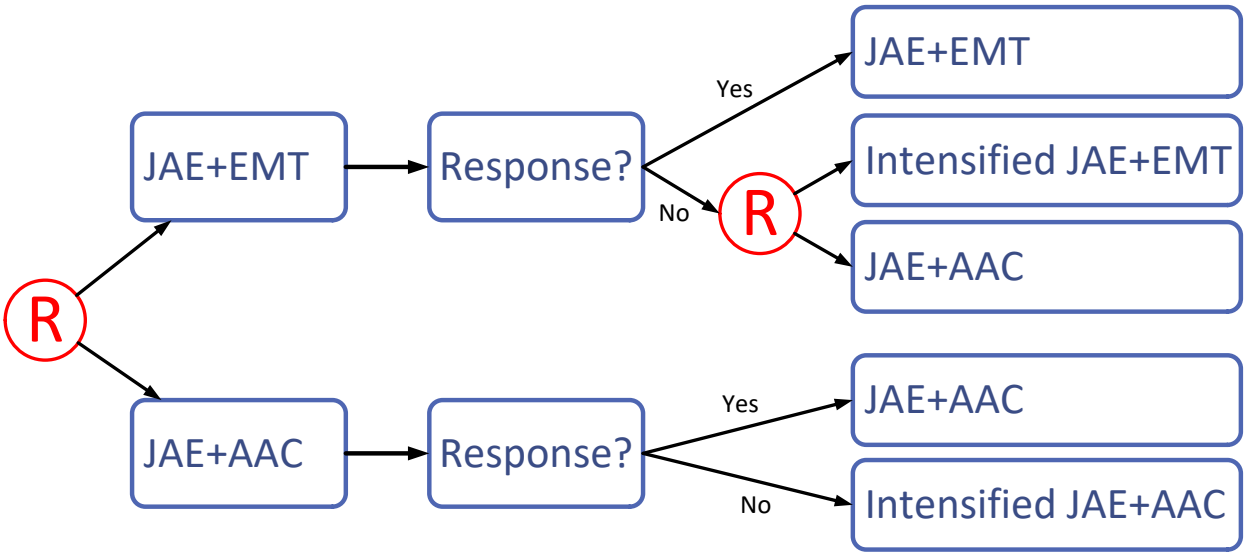

Figure 2: SMART design of the ASD study. AAC = augmentative and alternative communication; EMT = enhanced milieu teaching; JAE = joint attention/joint engagement. The "R" with a circle around it denotes randomization.

The simulated data set, called ASDdat, contains a number of measures that can serve as baseline tailoring variables for making a stage 1 decision between JAE+EMT vs. JAE+AAC. The candidate baseline tailoring variables include

- O_{11}: Number of different spontaneous words. High values are preferred. This is the same measure as the stage 2 outcome Y_2, but is measured at baseline.
- O_{12}: Number of unintelligible utterances by the child. Lower values are preferred.

The simulated data set also contains candidate time-varying tailoring variables from stage 1. These variables can be used, along with the candidate baseline tailoring variables, to make a decision between intensifying vs. augmenting treatment among those who are labeled as non-responders to initial JAE+EMT. The candidate time-varying tailoring variables include

- Y_1: Number of different spontaneous words is assessed during stage 1 treatment. Higher values are preferred. This is the stage 1 outcome.
- O_{21}: Number of communicative functions by the child (e.g., to request, to initiate a social interaction, to direct another's attention to an object). Higher values are preferred.

Two variables Y_1 and O_{21} are used to assess the response/non-response status. The simulated data set includes one additional variable, *S*, which is the rerandomization indicator. Since only non-responders to the first-stage treatment with JAE+EMT are rerandomized at the second-stage, only these children's data can be used in the stage 2 regression model. As a result *S* is coded 1 if the child is non-responder to JAE+EMT and is coded 0 otherwise.

4.2 Models, Contrasts and Codes

Suppose we decide to fit the following regression models (for clarity, the matrix terms in the regression models are written out).

- Stage 1: $\beta_{10} + \beta_{11} O_{11} + \beta_{12} A_1 + \beta_{13} A_1 O_{11}$
- Stage 2: $\beta_{20} + \beta_{21} O_{11} + \beta_{22} O_{21} + \beta_{23} A_2 + \beta_{24} A_2 O_{21}$

The model at stage 1 has only one interaction term involving first-stage treatment (A_1) (between baseline number of different spontaneous words (O_{11}) and A_1); thus, in this illustrative analysis O_{11} is the only candidate baseline tailoring variable. In this example, our goal is to find the best stage 1 treatment option for those with low or high number of different spontaneous words (where $O_{11} \leq 2$ is low and $O_{11} \geq 4$ is high). We use the following contrast matrix for the stage 1 regression model.

```
# Constructing the contrast matrices for stage 1
row1<-c(1,2,1,2) # mean Ỹ under JAE+EMT for children w/O11 = 2 = low;
row2<-c(1,2,-1,-2) #mean Ỹ under JAE+AAC for children w/O11 = 2 = low;
row3<-c(0,0,2,4) #mean difference in Ỹ for JAE+EMT - JAE+AAC for children
w/O11 = 2 = low;
row4<-c(0,0,2,8) #mean difference in Ỹ for JAE+EMT - JAE+AAC for children
w/O11 = 4 = hi;
c1<-rbind(row1, row2, row3, row4)
```

The contrast matrix, c1, has four columns corresponding to the fact that there are four parameters in the stage 1 regression and four rows corresponding to the fact that in four contrasts are considered. Although omitted below, a stage 2 contrast matrix, c2, can also be

formed in order to have qlaci provide confidence intervals for contrasts at stage 2.

The following code presents the **qlaci** syntax to fit the above stage 1 and 2 regression models.

```
data(ASDdat)
attach(ASDdat);
set.seed(300);
H10<- cbind(1, O11);
colnames(H10)<-c("int","O11");
H11<- cbind(1,O11); #O11 is a candidate tailoring variable for stage 1
colnames(H11)<-c("A1","A1O11");
#qlaci automatically multiplies H11 by A1 thus the column names of H11
#include A1
H20<- cbind(1, O11,O21);
colnames(H20)<-c("int","O11","O21");
H21<- cbind(1, O21); #O21 is a candidate tailoring variable for stage 2
colnames(H21)<-c("A2","A2O21");
#qlaci automatically multiplies H21 by A2 thus the column names of H21
#include A2
S<- as.logical(S);
result1<-qlaci(H10, H11, A1, Y1, H20, H21, A2, Y2, S,c1=t(c1)
,nb=1000);
print(result1);
```

4.3 qlaci Output

Table 1 shows the output obtained by running the example **qlaci** code above using the simulated data. The first contrast estimates the mean outcome under JAE+EMT (A_1=+1) for children when O_{11}=2 (e.g., baseline number of spontaneous words, O_{11}, is low). The second contrast estimates the mean $\tilde{Y}$ under JAE+AAC (A_1=-1) for children when O_{11}=2. The third contrast estimates the effect of JAE+EMT (A_1=+1) vs. JAE+AAC (A_1=-1) on the mean outcome, $\tilde{Y}$, when the baseline number of spontaneous words, O_{11}, is low. The last contrast estimates the effect of JAE+EMT (A_1=+1) vs. JAE+AAC (A_1=-1) on the mean $\tilde{Y}$ when the baseline number of spontaneous words, O_{11}, is high.

The estimated mean using contrast 1 & 2 suggests that individuals with a low baseline number of spontaneous words, O_{11}=2, benefit from starting off on JAE+EMT (56.35-54.00>0) and contrast 3 shows that this difference (2.35) is significant since the 95% confidence interval does not include zero. The last contrast shows that JAE+AAC significantly improves the final outcome for individuals with high baseline number of spontaneous words, O_{11}=4. Since the suggested first-stage treatment depends on the value of the variable O_{11} (low vs. high), this analysis indicates that this variable is likely to be a significant baseline tailoring variable.

Table 1. ASD example **qlaci** output.

```
$stg1coeff
      int       O11        A1     A1O11
45.880596  4.648178  6.555096 -2.689879

$stg2coeff
        int         O11         O21          A2       A2O21
43.40979346 -3.72296950  0.05873156  1.68595430  0.04902778

$ci1
           est       low       upp
row1 56.352290 55.630908 57.093178
row2 54.001614 52.879513 55.146183
row3  2.350676  1.077563  3.638551
row4 -8.408840 -9.441824 -7.379712

$ci2
NULL
```

Note that one can construct the individual confidence intervals for the first- and second-stage coefficients by setting the contrast matrices as $c1 = diag(4)$ and $c2 = diag(5)$, respectively.

5 Appendices

5.1 Simulated SMART in which all individuals are randomized at each stages

The simulated data set is called dat2. This simulated SMART involves 200 individuals and has the following features (Figure 3).

- At stage 1, all the individuals are randomized to one of the stage 1 treatment options (coded as $A_1 = \pm 1$).

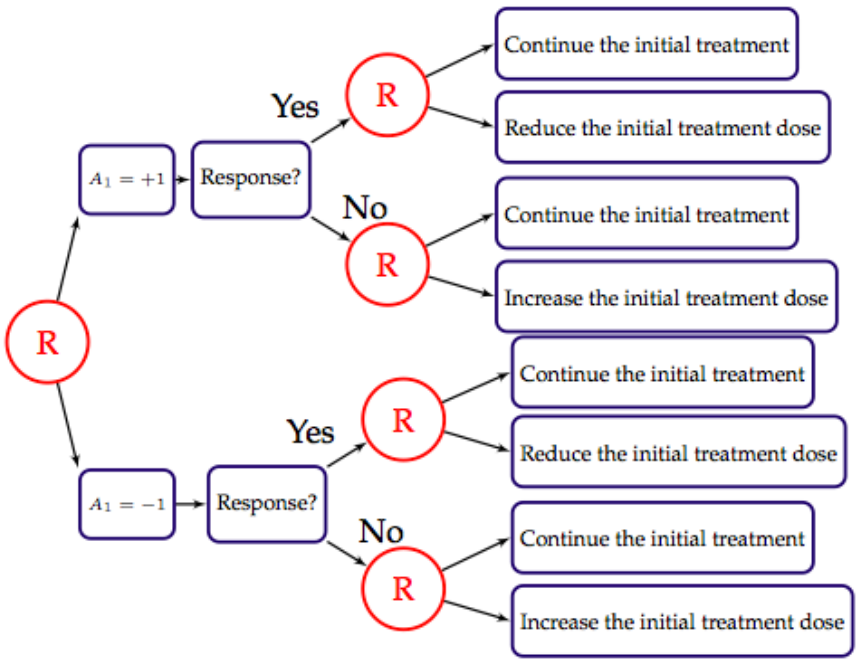

Figure 3: SMART design in which all individuals are randomized at each stage. The "R" with a circle around it denotes randomization.

- All individuals are classified as responders or non-responders to the stage 1 treatment.
- At stage 2, responders are randomized to either continue with their initial treatment or reduce the dose. Non-responders are randomized to either continue with their initial treatment or intensify.
- After stage 2, the final outcome is assessed (Y_2).

In this example, only the final outcome Y_2 is assessed and the objective is to learn an

adaptive intervention that comes close to maximizing the expected value of the final outcome ($E[Y_2]$). In this study there is no Y_1; we set Y_1=0 for everyone.

We fit the following regression models

- Stage 1: $\beta_{10} + \beta_{11} O_{11} + \beta_{12} A_1 + \beta_{13} A_1 O_{11}$
- Stage 2: $\beta_{20} + \beta_{21} O_{11} + \beta_{22} O_{21} + R(\beta_{23} A_2 + \beta_{24} A_2 O_{21}) + (1-R)(\beta_{25} A_2 + \beta_{26} A_2 O_{21})$

where R is a binary variable coded as 1 if responder and 0 otherwise. Since the stage 2 treatment options depend on whether or not each individual is labeled a responder, the stage 2 model is nested within responders and non-responders. Note that since everyone is randomized at stage 2, S is equal to one for everyone and the second model is fit among all the individuals.

The following code presents the **qlaci** syntax to fit the above stage 1 and 2 regression models.

```
data(dat2)
attach(dat2);
head(dat2);
## construct covariates used in the first-stage and the second-stage
## regression
 H10<- cbind(1, O11);
colnames(H10)<-c("int","O11");
H11<- cbind(1,O11);
colnames(H11)<-c("A1","A1O11");
Y1<- rep(0,200); # there is no Y1 in this simulated data
H20<- cbind(1,O11,O21);
colnames(H20)<-c("int","O11","O21");
H21<- cbind(R,R*O21,1-R,(1-R)*O21);  # R=1=responders
colnames(H21)<-c("A2R","A2RO21","A2(1-R)","A2(1-R)O21");
Y2<- Y;
S<- rep(1,200); # everyone is randomized at stage 2
## Construct contrast matrices
c1<-diag(4);  #number of rows must be equal to the number of parameters
              #in the stage 1 model
c2<-diag(7); #number of rows must be equal to the number of parameters
             #in the stage 2 model
```

```
      ## Run qlaci function to get estimates and confidence intervals for
      ##the contrasts
      set.seed(300);
      result2<-qlaci(H10, H11, A1, Y1, H20, H21, A2, Y2, S,c1=t(c1),c2=t(c2)
,nb=1000);
      # here we used 1000 bootstraps.
      print(result2);
```

The following table summarizes the output corresponding to this example.

Table 2. dat2 example **qlaci** output.

```
$stg1coeff
       int        O11         A1      A1O11
 9.9736728  1.0321597 -1.6113811 -0.3419619

$stg2coeff
       int        O11        O21        A2R     A2RO21    A2(1-R) A2(1-R)O21
 6.6386713  0.5972870  1.6897144  1.0626422  1.5863444  3.3274834  0.1768726


$ci1
            est        low         upp
[1,]  9.9736728  9.5478612 10.43201237
[2,]  1.0321597  0.5810393  1.46765503
[3,] -1.6113811 -2.0382856 -1.18664203
[4,] -0.3419619 -0.6909388  0.01724626

$ci2
           est        low       upp
[1,] 6.6386713  6.3038058 6.9640510
[2,] 0.5972870  0.3044086 0.8920901
[3,] 1.6897144  1.4719698 1.9270822
[4,] 1.0626422  0.6982854 1.4800429
[5,] 1.5863444  1.1850816 2.0064770
[6,] 3.3274834  2.6981402 3.8890823
[7,] 0.1768726 -0.1608220 0.5169883
```

5.2 Simulated SMART with no embedded tailoring variable

The simulated data set is called dat3. This simulated SMART involves 200 individuals and has the following features (Figure 4).

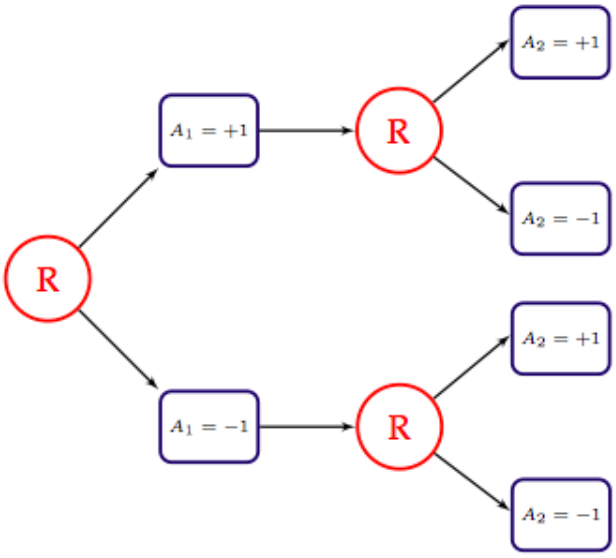

Figure 4: SMART design with no embedded tailoring variable. The "R" with a circle around it denotes randomization.

- At stage 1, all the individuals are randomized to one of the stage 1 treatment options (coded as $A_1 = \pm 1$).
- At stage 2, all the individuals are randomized to one of the stage 2 treatment options (coded as $A_2 = \pm 1$).
- After stage 2, the outcome is assessed (Y_2).

There is no embedded tailoring variable in this simulated data. Similar to the simulated SMART in section 4.2, just the final outcome Y_2 is assessed and the objective is to learn an adaptive intervention which comes close to maximizing the expected value of the final outcome ($E[Y_2]$). Since there is no Y_1 in this study, we set Y_1=0 for everyone.

We fit the following regression models:

- Stage 1: $\beta_{10} + \beta_{11} O_{11} + \beta_{12} A_1 + \beta_{13} A_1 O_{11}$
- Stage 2: $\beta_{20} + \beta_{21} O_{11} + \beta_{22} O_{21} + \beta_{23} O_{22} + \beta_{24} A_2 + \beta_{25} A_2 O_{21}$

Similar to the example in section 5.1, the second model is fit among all the individuals.

The following code presents the **qlaci** syntax to fit the above stage 1 and 2 regression models.

```
data(dat3)
attach(dat3);
head(dat3);
## construct covariates used in the first-stage and the second-stage
##regression
H10<- cbind(1, O11);
colnames(H10)<-c("int","O11");
H11<- cbind(1,O11); #O11 is a candidate tailoring variable for stage 1
colnames(H11)<-c("A1","A1O11");
Y1<- rep(0,200); # there is no Y1 in this simulated data
H20<- cbind(1,O11,O21,O22);
colnames(H20)<-c("int","O11","O21","O22");
H21<- cbind(1,O21);  #O21 is a candidate tailoring variable for stage 2
colnames(H21)<-c("A2","A2O21");
Y2<- Y;
S<- rep(1,200); # everyone is randomized at stage 2
## Construct contrast matrices
c1<-diag(4);  #number of rows must be equal to the number of parameters
              ##in the stage 1 model
c2<-diag(6); #number of rows must be equal to the number of parameters
             ##in the stage 2 model
## Run qlaci function to get estimates and confidence intervals for
##the contrasts
set.seed(300);
result3<-qlaci(H10, H11, A1, Y1, H20, H21, A2, Y2, S,c1=t(c1),c2=t(c2)
```

```
,nb=1000);
      print(result3);
```

The following Table summarizes the output corresponding to this example.

Table 3. dat3 example **qlaci** output.

```
$stg1coeff
       int        O11         A1      A1O11
 9.5527575  1.2420832 -1.6283943 -0.4366322

$stg2coeff
        int         O11         O21         O22          A2       A2O21
 7.01327606  0.61942202  1.69015898 -0.06039361  0.99006407  1.70521812


$ci1
            est        low          upp
[1,]  9.5527575  9.0791442 10.042931301
[2,]  1.2420832  0.7052559  1.740402724
[3,] -1.6283943 -2.0730089 -1.160290640
[4,] -0.4366322 -0.8294057 -0.000831816

$ci2
             est        low       upp
[1,]  7.01327606  5.6773712 8.2842926
[2,]  0.61942202  0.3521772 0.8982917
[3,]  1.69015898  1.4466815 1.9409173
[4,] -0.06039361 -0.2429888 0.1462929
[5,]  0.99006407  0.6761274 1.2687471
[6,]  1.70521812  1.4935963 1.9178780
```

6 References

Almirall, D., Compton, S. N., Gunlicks-Stoessel, M., Duan, N., Murphy, S.A., (2012). Designing a Pilot Sequential Multiple Assignment Randomized Trial for Developing an Adaptive Treatment Strategy.*Statistics in Medicine* 31(17):1887-1902.

Bickel, P. & Freedman, D. (1981). Some asymptotic theory for the bootstrap. *The Annals of Statistics, 9*(6), 1196–1217.

Chakraborty, B., Murphy, S. A., & Strecher, V. (2010). Inference for non-regular parameters in optimal dynamic treatment regimes. *Statistical Methods in Medical Research, 19*(3), 317–343.

Efron, B., & Tibshirani, R. (1993). An introduction to the bootstrap. Chapman and Hall.

Ertefaie, A., Almirall, D., Huang, L., Dziak, J. J., Wagner, A. T., & Murphy, S. A. (2012). *SAS PROC QLREAN users' guide* (version 1.0.0). P. S. University Park: The Methodology Center. Retrieved from http://methodology.psu.edu

Kasari, C. (2009). Developmental and augmented intervention for facilitating expressive language, Available at http://clinicaltrials.gov/ct2/show/NCT01013545?term =kasari&rank=5.

Kasari, C., Freeman, S., & Paparella, T. (2006). Joint attention and symbolic play in young children with autism: A randomized controlled intervention study. *Journal of Child Psychology and Psychiatry, 47*(6): 611–620.

Kasari, C., Paparella, T., Freeman, S., & Jahromi, L. B. (2008). Language outcome in autism: Randomized comparison of joint attention and play interventions. *Journal of Consulting and Clinical Psychology, 76*(1):125.

Laber, E., & Murphy, S. (2011). Adaptive confidence intervals for the test error in classification. *Journal of the American Statistical Association.* Advance online publication. doi: 10.1198/jasa.2010.tm10053

Laber, E., Qian, M., Lizotte, D., & Murphy, S. (2010). Statistical inference in dynamic treatment regimes. *Arxiv preprint arXiv:*1006.5831.

Lavori, P., Dawson, R., & Rush, A. (2000). Flexible treatment strategies in chronic disease: Clinical and research implications. *Biological Psychiatry 48*(6), 605–614.

Lei, H., Nahum-Shani, I., Lynch, K., Oslin, D., & Murphy, S. (2012). A "SMART" design for building individualized treatment sequences. *Annual Review of Clinical Psychology, Review in Advance first posted online on December 12, 2011* Vol. 8: 21-48. PMC22224838

Lunceford, J., Davidian, M., & Tsiatis, A. (2002). Estimation of survival distributions of treatment policies in two-stage randomization designs in clinical trials. *Biometrics 58*(1), 48–57.

Moodie, E, & Richardson, T. (2010) Estimating optimal dynamic regimes: Correcting bias under the null. *Scandinavian Journal of Statistics* 37.1: 126-146.

Murphy, S. (2003). Optimal dynamic treatment regimes. *Journal of the Royal Statistical Society, Series B, 65*(2), 331–355.

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

Murphy, S., Van Der Laan, M., & Robins, J. (2001). Marginal mean models for dynamic regimes. *Journal of the American Statistical Association, 96*(456), 1410–1423.

Nahum-Shani, I., Qian, M., Almirall, D., Pelham, W., Gnagy, B., Fabiano, G., Waxmonsky, J., Yu, J., & Murphy, S. A. (2012a). Experimental design and primary data analysis methods for comparing adaptive interventions. *Psychological Methods*, Advance online publication. doi: 10.1037/a0029372

Nahum-Shani, I., Qian, M., Almirall, D., Pelham, W., Gnagy, B., Fabiano, G., Waxmonsky, J., Yu, J., &. Murphy, S. (2012b). Q-Learning: A data analysis method for constructing adaptive interventions. *Psychological Methods*, Advance online publication. doi: 10.1037/a0029373

Pelham, W., & Fabiano, G. (2008). Evidence-based psychosocial treatments for attention-deficit/hyperactivity disorder. *Journal of Clinical Child and Adolescent Psychology, 37*(1), 184–214.

Pelham, W., Hoza, B., Pillow, D. R., Gnagy, E.M., Kipp, H. L., Greiner, A. R. , Waschbusch, D. A., Trane, S. T., Greenhouse, J., Wolfson, L. (2002). Effects of methyphenidate and expectancy on children with ADHD: Behavior, academic performance, and attributions in a summer treatment program and regular classroom settings. *Journal of Consulting and Clinical Psychology, 70*(2): 320.

Robins, J. (1986). A new approach to causal inference in mortality studies with a sustained exposure period–application to control of the healthy worker survivor effect. *Mathematical Modelling, 7*(9-12), 1393–1512.

Robins, J. (2004). Optimal structural nested models for optimal sequential decisions. In *Proceedings of the Second Seattle Symposium on Biostatistics* (pp. 189–326). New York, NY: Springer.

Thall, P., Sung, H., & Estey, E. (2002). Selecting therapeutic strategies based on efficacy and death in multicourse clinical trials. *Journal of the American Statistical Association, 97*(457), 29–39.

Thall, P. & Wathen, J. (2005). Covariate-adjusted adaptive randomization in a sarcoma trial with multi-stage treatments. *Statistics in Medicine, 24*(13), 1947–1964.

Wahed, A., & Tsiatis, A. (2004). Optimal estimator for the survival distribution and related quantities for treatment policies in two-stage randomization designs in clinical trials. *Biometrics, 60*(1), 124–133.

Wahed, A., & Tsiatis, A. (2006). Semiparametric efficient estimation of survival distributions in

two-stage randomisation designs in clinical trials with censored data. *Biometrika, 93*(1), 163–177.

Watkins, C. J. C. H. (1989). *Learning from delayed rewards* (Doctoral thesis, University of London, King's College. Available from ProQuest Dissertations and Theses Database. (ProQuest Document ID No. 303776498.

Watkins, C. J. C. H. & Dayan, P. (1992). Q-learning. *Machine Learning, 8*(3), 279–292.