STATISTICAL CONSIDERATIONS FOR CLINICAL TRIALS AIMING TO IDENTIFY INDIVIDUALIZED TREATMENT RULES

A DISSERTATION SUBMITTED TO THE FACULTY OF THE
UNIVERSITY OF MINNESOTA
BY

Charles H Cain

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

ADVISORS:

David Vock Thomas Murray

COMMITTEE MEMBERS:

David Vock Thomas Murray Kyle Rudser Alex Rothman

JULY, 2021

Abstract

Clinical trials have traditionally been designed to identify the best treatment on average or a treatment that results in benefit for a majority of a population. However, treatment that works well for a majority may not work at all for the minority, which motivates the need for personalized treatment rules in modern clinical trials. For a single stage randomized control trial, identification of an individualized treatment rule (ITR) is often set as an aim, particularly in behavioral intervention trials. ITRs assign treatment as a function of patients' clinical information which contrasts with a static treatment rule that assign everyone the same treatment. Much of the focus on ITRs revolves around identifying rules that are close to a theoretical optimal rule, which could lead to identifying rules that perform worse than the optimal static rule particularly in the absence of substantial effect heterogeneity. This limitation motivates new methods that reliably recommend the estimated optimal static rule when evidence of effect heterogeneity is lacking, and considerations for sample size regarding reliable identification of a beneficial ITR, which is an ITR that performs better than the optimal static rule. To address these limitations, we introduce a Monte Carlo integration based calculation of the probability to identify a beneficial ITR which requires specification of a data generating model. We also introduce an approach to selecting the penalty parameter in a LASSO model such that the static rule is identified with high probability in the absence of treatment effect heterogeneity which mitigates the risk of identifying a harmful ITR. A Dynamic Treatment Regime (DTR) is a clinical tool to guide the treatment decisions of clinicians which assigns treatment at each decision point over time based on patient characteristics including prior response to treatment. A Sequential Multiple Assignment Randomized Trial (SMART) aims to

identify optimal DTR through randomization at multiple time points. We introduce beneficial DTRs as DTRs that performs better than the estimated optimal embedded DTR. This definition implies that in the absence of treatment effect heterogeneity, an identified more deeply tailored DTR would be harmful in a population. To address this, we introduce a permutation test method to select the penalty parameter in a LASSO model such that no treatment interaction coefficients are selected for regression using Q-learning in the absence of treatment effect heterogeneity with specified probability at each stage of treatment assignment. The use of Q-learning, however, presents challenges in that the stage one model is frequently incorrectly specified. IQ-learning avoids this by not directly modeling the Q-function at the first stage of treatment. However, variable selection methods have not been considered when using IQ-learning and we apply a group LASSO method where the penalty parameter is selected through the same permutation-based methods. We apply all of our methods to two separate SMARTs, the Program for LUng Cancer Screening and TObacco Cessation (PLUTO) which aims to identify DTRs to assist with smoking cessation and the M-Bridge study which aimed to estimate an optimal DTR to prevent binge drinking in college freshman.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Randomized controlled trials (RCTs) often are designed to identify the treatment that maximizes an outcome on average in a given population. The result in these trials is an optimal static rule that assigns the same treatment to every person in a given population based on treatment main effect, which may lead to some people receiving a treatment that provides them with no benefit, or even causes them harm. Because of this problem, many trials set out to personalize treatment by identifying Individualized Treatment Rules (ITRs) that assign treatment based on patient characteristics. They do so, however, with little consideration given to the statistical framework needed to successfully identify personalized rules. This dissertation defines a statistical framework for personalization of treatment in both single stage RCTs and Sequential Multiple Assignment Randomized Trial (SMART). We define the success of a personalized rule as a beneficial rule (or at least a non-harmful rule), meaning a rule that results in a greater (or at least as large) value as the optimal treatment assignment already embedded into the study.

In Chapter 2 we introduce a beneficial/harmful ITR as one that performs better/worse than the optimal static rule. We also introduce a Monte Carlo integration to calculate the probability to identify a beneficial ITR as no closed form expression exists. In order to perform the Monte Carlo integration, certain nuisance parameters must be specified. We provide examples of obtainable values such as the proportion of variance explained in the outcome by the predictors, and the proportion of the population that would benefit from an ITR which can be used to determine these nuisance parameters. We also introduce a model selection technique that is designed specifically to identify a non-harmful ITR in the absence of treatment effect heterogeneity.

A Dynamic Treatment Regime (DTR) is a sequence of decision rules that specify treatment at different time points based on response to preceding treatment. A SMART is a multi-stage clinical trial designed to analyze DTRs by assigning treatment at various decision points. Within a SMART are multiple embedded DTRs which are the treatment sequences that participants are randomized. Similar to single stage RCTs, many SMARTs aim to identify more deeply tailored DTRs which assign treatment based on patient characteristics in addition to response to preceding treatment assignment. However, the definition of a successfully identified more deeply tailored DTR is unclear and few model selection methods have been considered previously for identifying DTRs.

The M-Bridge study is a SMART aimed at identifying treatment sequences to prevent binge drinking in college freshmen. One of the secondary aims of this study is to identify subgroups which experience varied outcomes from the rest of the population and in turn identify more deeply tailored DTRs. In Chapter 3 we extend the

methods from Chapter 2 to a two-stage SMART design. We define beneficial DTR as one that results in an expected outcome that is larger than that of the optimal embedded DTR, which is the embedded DTR that performs better than all other embedded DTRs. We define the λ_q-LASSO model selection in the context of a two stage SMART study where modified Q-learning is used to identify DTRs. These methods identify the embedded DTR at a given stage with user specified probability in the absence of treatment effect heterogeneity.

Q-learning is a backwards induction method commonly applied to identify the optimal DTR from SMARTs. In a two-stage study, the DTR is identified by first estimating the optimal treatment assignment at the second stage of treatment. The optimal treatment at the first stage is then estimated assuming that the optimal assignment will be given at the second stage. To do this, the predicted outcome from the stage two model is used to estimate the model at stage 1. The predicted outcome used as the outcome for the stage one model, however, will almost certainly be non-linear and non-smooth due to interaction terms in the second stage model. A linear model is still frequently used, however, for the stage one model in Q-learning.

Interactive Q-learning, or IQ-learning, is an adaptation to Q-learning that estimates the optimal treatment at stage one using a combination of two models that can be linearly associated with the stage one predictors. In Chapter 4 we discuss applying IQ-learning to SMART studies with our definition of beneficial DTRs. Existing model selection methods are not naturally applied to IQ-learning due to it's use of two models to estimate the stage one Q-learning models. We introduce the λ_q-GLASSO to identify the estimated optimal embedded rule at the first stage of treatment with user specified probability in the absence of treatment effect heterogeneity. We discuss in

Chapter 4 some of the difficulties we continue to handle when comparing IQ-learning to Modified Q-learning.

Chapter 2

Design Considerations and Analytical Framework for Reliably Identifying a Beneficial Individualized Treatment Rule

2.1 Introduction

The primary aim of many randomized trials is to evaluate the effect of an intervention. Assuming the intervention maximizes a key clinical endpoint in a target population, the result is to recommend a static treatment rule, i.e., a rule which assigns one treatment for everyone in a given population. Since certain subgroups may experience higher or lower outcomes than the rest of the population, many trials aim

to identify an individualized treatment rule (ITR), i.e., a rule which assigns treatment based on measured patient characteristics. The *value* of an ITR is defined as the expected outcome in the population when each individual receives the treatment indicated by the corresponding rule. An ITR is considered optimal if its value is greater than or equal to the value of any other ITR. Identifying an (optimal) ITR is frequently a secondary aim or exploratory aim of many randomized trials, especially in behavioral intervention trials. Frequently, in our experience, investigators of moderate-sized government-funded studies promise some form of personalization to make their proposals appear more cutting-edge and fundable.

One trial with personalization of treatment as a secondary aim is the Program for Lung Cancer Screening and Tobacco Cessation (PLUTO) trial [1]. PLUTO is a Sequential Multiple Assignment Randomized Trial (SMART) aimed at identifying sequences of treatment to assist with smoking cessation. The primary outcome of interest is long-term abstinence from smoking with secondary outcomes such as change in number of cigarettes smoked. One secondary aim focuses on identifying ITRs, either for a single stage or multiple stages of treatment.

There has been substantial research to develop methods for estimating ITRs and the related problem of identifying subgroups that benefit from a particular treatment relative to control or standard of care. Since personalizing treatment is often a pre-specified secondary aim, the definition and likelihood of success when developing an ITR should be established *a priori* as we routinely do for other aims. However, there are a number of challenges which preclude using existing methodology to calculate such a probability. First, there is no generally accepted criterion by which

to determine if an ITR is successful, and existing approaches have significant limitations. Second, calculating the *a priori* probability of estimating a beneficial or non-harmful ITR requires eliciting a targeted benefit from tailoring treatment which needs to be incorporated into a data generative model. Additionally, ITRs often involve non-smooth functionals of estimated regression parameters; therefore, formulae for calculating the *a priori* probability of a beneficial ITR do not exist in closed form. Third, under the null scenario of no treatment effect heterogeneity, no existing methods can guarantee that they will return a static rule with sufficiently high probability.

This study addresses each of these limitations by developing a conceptual and analytical framework for reliably estimating a beneficial ITR. Specifically, we first formalize the problem and argue that a successful or beneficial ITR is one whose value is larger than the value of the static rule rather than only being "close" to the true optimal. Additionally, we show how the pre-trial probability of estimating a beneficial ITR can be computed using a simulation-based calculator. We then cover how we can elicit the data generating model including nuisance parameters in terms of interpretable quantities, such as the proportion of people that benefit from an ITR and the proportion of variance explained in the outcome by the predictors. Then, we summarize some of the existing methods to identify ITRs as well as introducing an approach for selecting the penalty parameter for the LASSO model such that the static rule is recommended with high probability in the absence of treatment effect heterogeneity. We also simulate examples of how the probability is affected by certain values as well as examples of how often identifying ITRs could potentially cause harm in a population. Lastly, we apply the proposed methods to the PLUTO Study [1].

2.2 Probability of Estimating a Beneficial ITR

We consider two-arm randomized clinical trials with outcome, Y, assuming without loss of generality that a larger value is better, $A = 1$ or -1 for treatment or control arm respectively, and p dimensional vector of features, $X \in \mathcal{X}$, derived from baseline covariates. The purpose of an ITR is to identify the treatment that given baseline characteristics, X, on average will result in the higher expected outcome, Y. We define an ITR as a function, $d : \mathcal{X} \to \{-1, 1\}$, that maps baseline characteristics of a patient to either control or treatment. We define the potential outcome, $Y(a)$, as the outcome value under treatment a, possibly contrary to fact[2]. Given a participant with characteristics $X = x$, the potential outcome under treatment rule d is denoted as

$$Y(d) = \mathbb{1}\{d(x) = 1\}Y(1) + \mathbb{1}\{d(x) = -1\}Y(-1),$$

i.e. $Y(d) = Y(1)$ when $d(x) = 1$ and $Y(-1)$ otherwise.

We assume the observed data are $(Y_i, X_i, A_i) \overset{iid}{\sim} F,\ i = 1, \ldots, n$. We are interested in the *value* of an ITR, d, defined as the expected potential outcome assuming everyone follows d,

$$V(d) = E\{Y\,(d)\}.$$

An ITR, d^{opt} is said to be optimal if $V(d^{opt}) \geq V(d)$ for any other ITR d. We define the optimal static rule as $sign[E\{Y(1) - Y(-1\}]$. Assuming that there are no unmeasured confounders, which is guaranteed under randomization, that is,

$$A \perp\!\!\!\perp \{Y(-1), Y(1)\}|X,$$

then $E_{Y|X,A}\{Y|X, A = d(x)\} = E_{Y|X}\{Y(d)|X\}$.

For scalar functions $g, h : \mathcal{X} \to \mathbb{R}$, we write

$$E\{Y|X = x, A = d(x)\} = g(x) + d(x)h(x) \tag{2.1}$$

so that the true optimal treatment rule arises as $d^{opt}(x) = sign\{h(x)\}$. Note that if we have estimated $h(x)$, say $\widehat{h}(x)$, then an estimator for d^{opt} is $\widehat{d}^{opt} = sign\{\widehat{h}(x)\}$. So,

$$\begin{aligned} V(d) &= E_X\left[E_{Y|X}\{Y(d)|X\}\right] \\ &= E_X\{g(X) + d(X)h(X)\}. \end{aligned} \tag{2.2}$$

Many approaches for evaluating the likely quality of an expected ITR[3, 4, 5, 6, 7, 8, 9, 10, 11, 12] compare $V(\widehat{d}^{opt})$ to $V(d^{opt})$. For example, Laber et. al.[3] proposed a sample size estimate using pilot data to satisfy

$$P\{|\widehat{V}(\widehat{d}^{opt}) - V(d^{opt})| < \epsilon\} > 1 - \gamma, \tag{2.3}$$

where $\widehat{V}(\widehat{d}^{opt})$ is defined the same as in (2.2) with $\widehat{g}$ and $\widehat{h}$ replacing g and h and $1 - \gamma$ represents the power. The sample size found to satisfy (2.3) is dependent on the choice of ϵ, however, an appropriate choice for ϵ is not always clear. This presents a problem if ϵ is chosen to be too large. For example, if ϵ is chosen to be larger than the difference in the values of the optimal rule, d^{opt}, and the estimated optimal static rule, $\widehat{w}^{opt}$, i.e. $\epsilon > V(d^{opt}) - V(\widehat{w}^{opt})$, then we could potentially identify an ITR which performs worse than $\widehat{w}^{opt}$. The estimated optimal static rule is the static rule which results in the highest value among possible static rules (note that in a two-arm

trial, only two static rules are possible). Furthermore, when there is no treatment effect heterogeneity, $d^{opt} = w^{opt}$, any ITR that recommends different treatments to certain subpopulations will perform worse than $\widehat{w}^{opt}$ because $\widehat{w}^{opt} = sign[E_X\{\widehat{h}(x)\}]$ and $w^{opt} = sign[E_X\{h(x)\}]$ so $\widehat{w}^{opt} = w^{opt}$ in most reasonable examples. Also, since identifying w^{opt} by estimating $\widehat{w}^{opt}$ is already well-established, easiest to implement, and the primary aim in most clinical trials, we argue any ITR should perform at least as well as $\widehat{w}^{opt}$.

To express this we call $\widehat{d}^{opt}$ a *Beneficial ITR* when

$$V(\widehat{d}^{opt}) > V(\widehat{w}^{opt}) \tag{2.4}$$

and thus, the probability of identifying a beneficial ITR, arises as

$$P_B = P\{V(\widehat{d}^{opt}) - V(\widehat{w}^{opt}) > 0\}.$$

Conversely, the probability of an ITR causing harm in a population is

$$P_H = P\{V(\widehat{d}^{opt}) - V(\widehat{w}^{opt}) < 0\}.$$

We are often also interested in the probability of not causing harm which is equivalent to $1 - P_H$. When model selection is incorporated to identify an ITR, it is possible to have $V(\widehat{d}^{opt}) - V(\widehat{w}^{opt}) = 0$, so $P_B \neq 1 - P_H$.

To calculate either P_B or P_H, we assume that we will use a two-step process to identify the beneficial ITR. That is, $h(x)$ will be estimated first and then the estimated

optimal rule is the sign of $\widehat{h}(x)$. Then $V(\widehat{d}^{opt})$ and $V(\widehat{w}^{opt})$ can be written as

$$\begin{aligned} V(\widehat{d}^{opt}) &= E_X\{g(X) + \widehat{d}^{opt}(X)h(X)\} \\ &= \int [g(x) + sign\{\widehat{h}(x)\}h(x)]dF_X(x), \\ V(\widehat{w}^{opt}) &= E_X\{g(X) + \widehat{w}^{opt}(X)h(X)\} \\ &= \int \{g(x) + \widehat{w}^{opt}h(x)\}dF_X(x), \end{aligned}$$

and

$$V(\widehat{d}^{opt}) - V(\widehat{w}^{opt}) = \int [sign\{\widehat{h}(x)\} - \widehat{w}^{opt}]h(x)dF_X(x),$$

where $F_X(x)$ is the distribution function of X. Note that the above expressions for $V(\widehat{d}^{opt})$ are conditioned on the function $\widehat{d}^{opt}$ which is itself a function of the sample data $\{Y_i, X_i, A_i\}_{i=1,\ldots,n}$. To emphasize this we write

$$\widehat{d}^{opt}(x) = \widehat{d}^{opt}(x; \{Y_i, X_i, A_i\}_{i=1,\ldots,n}).$$

Thus the probability of a beneficial rule is determined using

$$\begin{aligned} &E_{\{Y_i,X_i,A_i\}_{i=1,\ldots,n}} \left(E_X \left[\{\widehat{d}^{opt}(x; \{Y_i, X_i, A_i\}_{i=1,\ldots,n}) - \widehat{w}^{opt}\}h(X) \right] \right) \\ &= \int \cdots \int [sign\{\widehat{h}(x)\} - \widehat{w}^{opt}]h(x)dF_X(x) \prod_{i=1}^{n} dF_{y_i,x_i,a_i}(x_i, y_i, a_i) \end{aligned}$$

The above integral is analytically intractable due to the use of the *sign* function on $\widehat{h}(x)$, so we evaluate it using Monte Carlo integration. Specifically, we either generate N_1 samples of X from F_X or simply use baseline characteristics of the study participants that will be used for tailoring. We label the resulting data set as the

population set. We then either generate N_2 samples $(y_i, x_i, a_i)_{i=1,\ldots,n}$ from F or, if study data is used as the population set, sample x_i from the population set N_2 times with replacement and generate y_i and a_i parametrically according to the assumed distributions $F_{Y|X,A}$ and F_A. P_B is estimated as

$$\hat{P}_B = \frac{1}{N_1 N_2} \sum_{r=1}^{N_1} \sum_{s=1}^{N_2} \mathbb{1}\left[\{\widehat{d}^{opt}(x_r; \{y_{is}, x_{is}, a_{is}\}_{i=1,\ldots,n}) - \widehat{w}^{opt}\} h(x_r) > 0\right]. \tag{2.5}$$

Functions in R for the methods described in this paper are available at https://github.com/charlescain/BeneficialITR.

2.3 Eliciting Data Generative Model

For this paper, we primarily focus on when outcome Y is continuous and conditionally normal, and g and h from (2.1) are linear, i.e.

$$g(X) = \beta_0 + X'\beta_1 \quad \text{and} \quad h(X) = \beta_2 + X'\beta_3$$

where X is p-dimensional and standardized, so the data generative model is

$$E[Y_i|X_i, A_i] = \beta_0 + X_i'\beta_1 + A_i\beta_2 + A_i X_i'\beta_3. \tag{2.6}$$

Specifying these coefficients, $\beta = (\beta_0', \beta_1', \beta_2', \beta_3')'$ as well as the residual variance σ^2, is not always intuitive and may be difficult to elicit from a collaborator(s) or existing literature. The coefficients could instead be determined by eliciting the following more intuitive (or readily available) values:

(i) Δ, the expected main effect of the treatment,

(ii) ν, the proportion of people expected to benefit from control,

(iii) R_T^2 or R_C^2, the proportion of variance explained in the outcome by X in the treatment or control group, and

(iv) V_y, the variance of the outcome in the control group.

Additionally, F_X would need to be specified but we always consider F_X to be multivariate normal with independent correlation matrix Σ_p except when otherwise specified. We outline how the coefficients may be determined from (i)-(iv) in three key scenarios: (a) only one predictor main effect and interaction is associated with the outcome, (b) all predictor main effects and interactions are equally associated with the outcome, and (c) predictor main effects and interactions associations' are diminishing.

Assuming $E[X] = 0$ which is always feasible by centering, we determine β_2 from Δ as,

$$\beta_2 = \Delta/2.$$

Without loss of generality, we assume $\Delta > 0$, we determine β_3 from

$$\nu = P\left(\beta_2 + X'\beta_3 < 0\right),$$

i.e.

$$\nu = \Phi\left(\frac{-\beta_2}{\sqrt{\beta_3'\Sigma_p\beta_3}}\right). \tag{2.7}$$

ν could colloquially be referred to as the proportion that would benefit from the

optimal ITR. If Δ is assumed to be 0, (2.7) cannot be used to determine β_3. However, few trials assume the main effect of treatment will be truly 0 *a priori* so using (2.7) is almost always viable.

Lastly, we determine β_1 and σ^2 from R_C^2 and V_y as follows

$$R_C^2 = \frac{(\beta_1 - \beta_3)'\Sigma_p(\beta_1 - \beta_3)}{\sigma^2 + (\beta_1 - \beta_3)'\Sigma_p(\beta_1 - \beta_3)} \quad V_y = \sigma^2 + (\beta_3 - \beta_1)'\Sigma_p(\beta_3 - \beta_1). \tag{2.8}$$

By solving the system equations from (2.7) and (2.8) we obtain values for β_1, β_3, and σ^2. R_C^2 can be easily replaced by $R_T^2 = \frac{(\beta_1+\beta_3)'\Sigma_p(\beta_1+\beta_3)}{\sigma^2+(\beta_1+\beta_3)'\Sigma_p(\beta_1+\beta_3)}$ to solve the system of equations depending on the availability of information. Under scenario (a), the equations simplify quite nicely but when more than one predictor/interaction effect is non-zero, the equations are not always solvable which is why we place restrictions on the coefficients in scenarios (b) and (c). In particular, in scenario (b), we assume that $\beta_{11} = \ldots = \beta_{1p}$ and $\beta_{31} = \ldots = \beta_{3p}$. In scenario (c), we assume that $\beta_{1j} = \beta_1/j$ for $j = 1, \ldots, p$ and similarly for β_3. Using these restrictions we are able to solve the system of equations from (2.7) and (2.8) for β using Newton-Raphson.

While specifying values such as variance of the outcome and the main effect of treatment is commonplace for planning trials, specifying values such as the proportion of people that benefit from tailoring treatment or variance explained by X in the treatment group may be less intuitive. R_C^2 and ν, may be found from previous data if it is available and we provide examples that illustrate the effect of these parameters on the overall probability to identify a beneficial ITR.

2.4 Methods to Identify an ITR

2.4.1 Existing Methods

When estimating $h(X)$ in equation (2.1) (which is used to estimate the optimal ITR, $\widehat{d}^{opt}$), many standard methods incorporate the use of model selection, e.g., forward selection[13], LASSO[5, 14, 8, 7, 15], Elastic Net[16] or Random Forest[6, 4, 9, 10]. We briefly review these techniques and the implementation details that we used.

To perform forward selection, we begin with a model that includes only a main effect for treatment and then for features $X_j, j = 1, \ldots, p$, we consider adding main effects or an interaction with treatment one at a time based on which addition results in the greatest decrease in AIC. Forward selection provides a straight-forward method for model selection but becomes computationally intensive for datasets with a large amount of predictors.

LASSO model selection uses L_1 penalized regression to select parameters in a model with non-zero coefficients. Some form of cross-validation is typically used to select a value for the penalty parameter. For this paper, we consider 10-fold cross validation[17]. For each candidate value of λ, we evaluate the cross-validation error of the resulting model and pick the value, λ_{min}, that results in the lowest cross-validation error. Another common choice is to select λ_{1se} which results in a cross validation error within one standard error of the minimum. For a given λ, we estimate

$$\hat{\beta} = \underset{\beta \in \mathbb{R}}{\operatorname{argmin}} \sum_{i=1}^{n} (Y_i - X_i\beta) + \sum_{j=1}^{p} \widetilde{\lambda}_j |\beta_j|$$

where $\widetilde{\lambda}$ is the vector of the penalty parameters such that $\widetilde{\lambda}_j = 0$ for the intercept

and the main effect coefficient and $\widetilde{\lambda}_j = \lambda$ otherwise. LASSO provides a flexible and computationally fast way to perform model selection by identifying terms that are non-zero after applying the penalty parameter. However, due to the inherent bias of penalized coefficient estimators, we estimate $\widehat{d}^{opt}$ by fitting a linear regression model using least-squares with the non-zero terms from the LASSO model.

LASSO performs poorly when $p \gg n$ which motivates consideration of the Elastic Net model[16]. Using an Elastic Net model for variable selection is similar to using LASSO, but we estimate

$$\widehat{\beta} = \underset{\beta \in \mathbb{R}}{\operatorname{argmin}} \sum_{i=1}^{n} (Y_i - X_i\beta) + \alpha \sum_{j=1}^{p} \widetilde{\lambda}_j |\beta_j| + (1-\alpha) \sum_{j=1}^{p} \widetilde{\lambda}_j (\beta_j)^2$$

where α is a mixing parameter that is selected through cross validation simultaneously with λ.

Additionally we consider the use of Random Forest. To do this, we estimate a random forest of 500 trees separately for the treatment and control conditions. Note that,

$$\hat{h}(x_i) = \tfrac{1}{2}\{\hat{E}\{Y_i | A_i = 1, X = x_i\} - \hat{E}\{Y_i | A_i = -1, X = x_i\}\},$$

where $\hat{E}\{Y_i | A_i = 1, X = x_i\}$ is from the tree estimated within the treatment condition and similarly for $\hat{E}\{Y_i | A_i = -1, X = x_i\}$. Trees are generated by testing $\frac{p}{3}$ variables at each split.

2.4.2 Modified selection of LASSO penalty parameter

While the above methods are able to identify beneficial ITRs, they do not always avoid causing harm. In particular, in the case when the effect of treatment is not

heterogeneous (i.e., $sign\{h(x)\} = sign\{h(E_X[x])\}$ for all x), the model selection approaches detailed above are not guaranteed to result in a static rule with sufficiently high probability. In the scenario without treatment effect heterogeneity, any non-static rule will perform worse than the static rule which results in harm overall in a population. Selecting the LASSO penalty parameter to minimize the cross validation error can often result in a penalty parameter that is not conservative enough, leading to the selection of erroneous interaction effects. The LASSO with λ_{1se} has been mostly used ad-hoc and can result in model selection that is too conservative, i.e., the variable selection often fails to select interaction terms. We propose a permutation-based method to select the LASSO penalty parameter which aims to bound the probability of identifying a harmful ITRs. The resulting method recommends $\widehat{w}^{opt}$ with a greater than user-specified level when no treatment effect heterogeneity exists while using model selection of the LASSO model when there is heterogeneity.

The permutation based method takes data (Y_i, X_i, A_i) and a set of values $\mathbf{\Lambda}$ for the LASSO penalty parameter. For each $\lambda \in \mathbf{\Lambda}$, we first subtract the main effect of treatment, Δ, from Y to obtain pseudo-outcome Z,

$$Z_i = \begin{cases} Y_i - \Delta, & \text{if } A_i = 1 \\ Y_i, & \text{if } A_i = -1 \end{cases} \quad i = 1, \ldots, n.$$

If Δ is unknown, we use $\widehat{\Delta}$ instead. We then permute the treatment assignment to obtain data $\widetilde{D} = (Z_i, X_i, \widetilde{A}_i)_{i=1,\ldots,n}$, where $\widetilde{A}_i$ denotes the permuted treatment assignment for the i^{th} observation. Fitting a LASSO model with $\widetilde{D}$, we evaluate if all

interactions $\hat{\beta}_3^{(\lambda)}$ were zero within the model,

$$I_\lambda = \mathbb{1}\left(\hat{\beta}_3^{(\lambda)} = \mathbf{0}\right),$$

where $\mathbf{0}$ is a vector of zeros and $\hat{\beta}_3^{(\lambda)}$ is the interaction terms from the LASSO regression model using the penalty parameter λ. Repeating this process B times, the proportion of models that result in one or more non-zero interaction term given λ is estimated using

$$P\left(I_\lambda = 1\right) \approx \frac{1}{B}\sum_{b=1}^{B} \mathbb{1}\left(\hat{\beta}_{3b}^{(\lambda)} = \mathbf{0}\right).$$

We then select the modified LASSO penalty parameter, λ_q, as the smallest value in $\mathbf{\Lambda}$ such that $P\left(I_\lambda = 1\right) > q$.

By subtracting Δ from the outcome of those with $A = 1$, we remove any treatment main effect. By permuting the treatment assignment we remove any semblance of treatment interaction effects. As such, the selected λ_q will limit the false discovery of an interaction effect at the specified q in the absence of treatment effect heterogeneity. In other words, using λ_q as a penalty parameter will result in $\widehat{w}^{opt}$ at least $q \times 100\%$ of the time in the absence of treatment effect heterogeneity.

2.5 Examples

We consider a variety of plausible scenarios in which investigators may wish to calculate the probability of a beneficial and harmful ITR. We also investigate the sensitivity of these probabilities to inputs in the data generating model, the number of parameters considered, and the model selection procedure. The data generating models

presented in the following examples are not as sophisticated as those presented with other methods, however, calculating the probability to identify a beneficial ITR can be likened to power calculations where simplifying assumptions are almost always made. To perform forward selection based on the AIC values, we used the MASS package[18] in R with the function `stepAIC`. Both the LASSO and the Elastic Net model selection were performed using the glmnet package[19] with the `glmnet` function. The random forests were generated using the randomForest package[20] using function `randomForest`.

2.5.1 Example 1: Effect of ν

We consider a scenario in which investigators select a sample size for a study to ensure a sufficiently low probability of identifying a harmful ITR. We investigate how the probability of a beneficial and harmful ITR varies across different values of the proportion which benefit from the ITR, the number of tailoring variables considered, and the model selection method used. More specifically, we determined the sample size to identify an ITR that causes harm only $P_H = 5\%$ of the time where we assumed 16% of the population would benefit from control over treatment. We additionally defined the generative model assuming $V_y = 1$, $\Delta = 0.3$, $R^2_C = 0.2$, and $p = 1$ resulting in a sample size of $n = 266$ assuming no model selection is used.

We examined the effect that p and ν have on the probability to identify a beneficial ITR. Across all scenarios, the data generating mechanism has only one predictor with a non-zero interaction effect and has $V_y = 1$, $\Delta = 0.3$, and $R^2_C = 0.2$. We varied ν among $0, 0.02, 0.16, 0.3$ to represent when there was not treatment effect heterogeneity and when only people whose value of X is two, one, and one-half standard deviations

below the mean, respectively, would benefit from control over treatment. Even though the sample size was determined assuming that the investigator would only consider a single covariate to interact with treatment, we also varied the number of predictors considered, $p = 1, 5, 20$, during model selection when identifying an ITR to mimic a likely situation in which an investigator later identifies more covariates for possible inclusion in the model.

	Treatment rule identification method					
ν	None	Forward	λ_{min}-LASSO	$\lambda_{0.8}$-LASSO	Elastic Net	Random Forest
One Potential Tailoring Variable ($p = 1$)						
0	0.01/0.51	0.00/0.16	0.01/0.45	0.01/0.17	0.00/0.49	0.01/0.75
0.02	0.37/0.40	0.16/0.34	0.33/0.38	0.17/0.35	0.36/0.39	0.08/0.74
0.16	0.96/0.02	0.90/0.02	0.95/0.02	0.91/0.02	0.96/0.02	0.64/0.29
0.3	1.00/0.00	1.00/0.00	1.00/0.00	1.00/0.00	1.00/0.00	0.99/0.01
Five Potential Tailoring Variables ($p = 5$)						
0	0.01/0.97	0.00/0.25	0.01/0.64	0.01/0.16	0.01/0.77	0.01/0.99
0.02	0.03/0.96	0.15/0.40	0.06/0.72	0.03/0.31	0.05/0.83	0.01/0.99
0.16	0.68/0.32	0.88/0.05	0.71/0.25	0.69/0.09	0.69/0.29	0.23/0.77
0.3	1.00/0.00	1.00/0.00	1.00/0.00	1.00/0.00	1.00/0.00	0.99/0.01
Twenty Potential Tailoring Variables ($p = 20$)						
0	0.01/0.99	0.01/0.50	0.01/0.74	0.00/0.15	0.01/0.86	0.01/0.99
0.02	0.01/0.99	0.09/0.60	0.02/0.79	0.01/0.21	0.02/0.88	0.01/0.99
0.16	0.03/0.97	0.78/0.17	0.31/0.63	0.48/0.12	0.25/0.73	0.46/0.54
0.3	0.99/0.01	1.00/0.00	1.00/0.00	0.99/0.00	1.00/0.00	1.00/0.00

Table 2.1: Simulated values for P_B/P_H after varying values of ν, the number of predictors, and the method used to identify the treatment rule. $\Delta = 0.3$, $V_y = 1$, $R^2_C = 0.2$. None refers to using no model selection and λ-LASSO refers to when model selection was performed with a LASSO regression model using the λ penalty parameter λ_{min} or $\lambda_{0.8}$.

Table 2.1 shows the values of P_B and P_H from the simulations under varied values of ν, and, p. Not surprisingly, increasing ν resulted in a larger probability of estimating a beneficial ITR and lower probability of estimating a harmful one. The number of predictors considered for inclusion in the model had a dramatic effect

on the probability of identifying a beneficial ITR. When $p = 1$, all model selection methods, with the exception of random forests, led to a probability of identifying a beneficial ITR between 90 an 95% when $\nu = 0.16$. However, when $p = 5$ and $p = 20$ these probabilities decrease to between 68 and 88% and 3 and 78%, respectively. The effect was particularly dramatic for those methods (no model selection and elastic net) which do not perform variable selection. Random forest had consistently smaller values of P_B. Random forests are the only model in which the data generating mechanism is not contained in the model space which may contribute to lower values of P_B. Importantly, random forests are an extremely flexibly class of models and adding additional parameters to a model (either through additional covariates or flexible basis expansions and interactions of the covariates) can lead to lower values of P_B. Only $\lambda_{0.8}$-LASSO was able to control the probability of identifying a harmful ITR at less than 20% when there was no treatment effect heterogeneity. However, there were relatively low values of P_B for $\lambda_{0.8}$-LASSO when compared to other methods like Forward selection, λ_{min}-LASSO, and Elastic Net.

Additionally, we examined the values for $V(\widehat{d}^{opt}) - V(\widehat{w}^{opt})$ in Table A1.1 in the supplement. When $R_C^2 \geq 0.2$ and $\nu \geq 0.16$, the mean difference in values was positive. The mean difference was negative when $\nu < 0.16$ even for the $\lambda_{0.8}$-LASSO which identified a beneficial ITR over 80% of the time. This implied that the $\lambda_{0.8}$-LASSO was likely selecting either $\widehat{w}^{opt}$ or a rule that performs worse than $\widehat{w}^{opt}$.

2.5.2 Example 2: Post hoc identification of an ITR

Estimation of an ITR is frequently done as a post-hoc analysis. In this example we consider a trial which is powered to detect a main effect of treatment. In many trials,

estimation of an optimal ITR is done to "salvage" a negative trial, so we examine how the probability of identifying a beneficial and harmful ITR changes as the main effect of the treatment is attenuated from the effect used to power the trial.

Specifically, we considered a trial with sample size $n = 468$, which is the sample size needed to detect average treatment effect with a Cohen's D of 0.3 at 90% power. We evaluated three scenarios: The true average treatment effect was zero, the true average treatment effect was half of the effect used to power the trial, and the targeted average treatment effect was equal to the effect used to power the trial. We set $V_y = 1$ so that these three scenarios correspond to $\Delta = 0.3, 0.15$, and 0 and $\beta_2 = 0.15, 0.075$, and 0, respectively. In all scenarios we set $R^2_C = 0.1$ and set one interaction term in the generative model to be non-zero. Specifically, we set $\beta_3 = 0.15$, which corresponded to $\nu = 16\%$ of people benefiting from control over treatment when the true main effect was equal to the targeted main effect, $\Delta = 0.3$. Across the three data generating scenarios, we kept β_3 constant (i.e., a constant interaction effect); this implies $\nu = 0.16, 0.3, 0.5$ when $\Delta = 0.3, 0.15, 0$, respectively.

We used each method outlined in this paper to identify an ITR. Additionally we included additional predictors in the model whose main and interaction effects were zero by varying the dimension of X, p. In the scenario where there was no main effect, P_B was consistently high and did not substantially attenuate as the number of predictors used increased. The LASSO Model using the $\lambda_{0.8}$ penalty parameter had the lowest value of P_B but still identified a beneficial ITR around 80% of the time when there were 20 predictors compared to no model selection which performed the best by identifying a beneficial ITR 99.9% of the time. Performing no model selection resulted in the highest probability since it is the only method that guarantees that

interaction terms are included whereas the $\lambda_{0.8}$-LASSO and other methods that use model selection may have been too conservative in this context. In the context of no treatment main effect and a moderate interactive effect, we could add additional covariates to the model without significant consequence, however, as is demonstrated in other examples, adding in additional parameters when there is no interaction effect can have a deleterious effect. When the main effect was half of what was used to power the trial, each method performed relatively well when $p \leq 5$. As the number of predictors increased, no model selection failed to identify a beneficial ITR whereas forward selection identified a beneficial ITR 97.1% of the time. There was not a large difference between the model selection methods in this scenario. In the scenario where the main effect was equal to the effect used to power the trial almost every ITR identification method except for Forward Selection and the $\lambda_{0.8}$-LASSO performed poorly as the number of predictors increased when attempting to identify a beneficial ITR. Forward selection performed the best by identifying a beneficial ITR at least 90% of the time whereas the $\lambda_{0.8}$-LASSO identified a beneficial ITR at least 70% of the time. We also show the values of P_H and see that both forward selection and $\lambda_{0.8}$-LASSO have low probability of harm even when 20 predictors were tested and $\Delta = 0.3$.

Overall, using the $\lambda_{0.8}$-LASSO for model selection performed well across all of the scenarios. While it did perform poorly in the first scenario when compared to other model selection methods, it still performed well by identifying a beneficial ITR often and avoided causing harm in scenarios where other methods could not. Forward Selection performed exceptionally well in each scenario, and we expect that this was due to the simplicity of the simulation with only one predictor having a true interaction

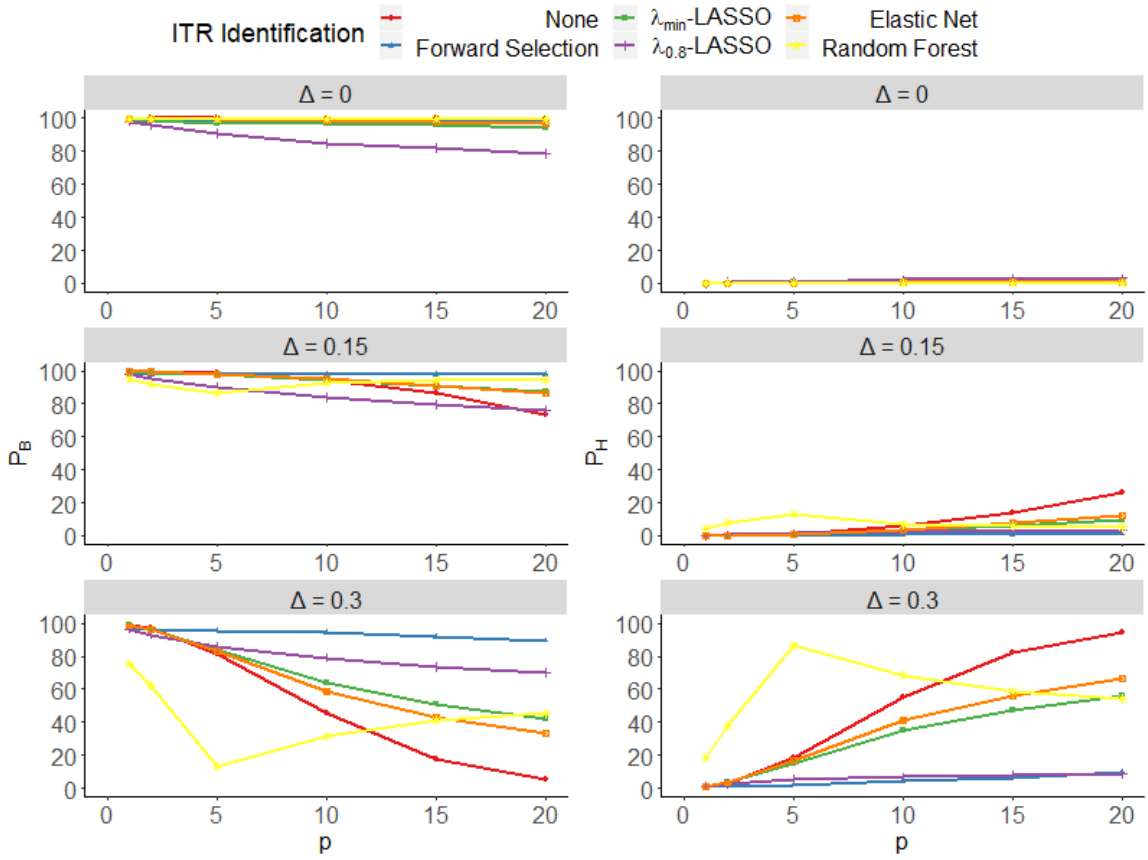

Figure 2.1: Simulated values for P_B (left) and P_H (right) where n=468. The first scenario represents when there was no true main effect ($\Delta = 0$), the second scenario represents when the main effect was half of what was used to power the trial ($\Delta = 0.15$), and the third scenario represents when the main effect was equal to what was used to power the trial ($\Delta = 0.3$).

effect with treatment.

The empirical CDFs of $V(\widehat{d}^{opt}) - V(w^{opt})$ for $p = 2, 10$ are given in Figure A1.1 in the supplement. The plots are bounded on the right by $V(d^{opt}) - V(w^{opt})$. Note that even when $V(\widehat{d}^{opt})$ is near $V(d^{opt})$, it is still possible to have $V(\widehat{d}^{opt}) < V(w^{opt})$ resulting in a harmful ITR. We notice this the most in the scenario where the main effect is equal to the effect used to power the study and as p increases.

2.5.3 Example 3: Multiple interaction effects

To evaluate more complicated data generating mechanisms, we considered the effects of more than one non-zero predictor main effect and treatment interaction. Similar to example 1 we considered simulated trial data with a sample size ($n = 196$) determined to identify a treatment rule that does not cause harm 95% of the time in a population for which 16% would benefit from control over treatment.

We defined the generative model using $V_y = 1$, $\Delta = 0.3$, and $R^2_C = 0.3$. Two cases of non-zero interactions were considered. The first case considered all the interaction and main effect coefficients for the predictors to be equal across all predictors, i.e. , $\beta_{11} = \ldots \beta_{1p}$ and $\beta_{31} = \beta_{3p}$. The second case considered a diminishing effect in the predictor main effect and interaction coefficients, i.e. $\beta_{1j} = \frac{\beta_1}{j}$ and $\beta_{3j} = \frac{\beta_3}{j}$ for $j = 1, \ldots, p$. To examine the impact of correlation in the predictors we used an AR(1) type structure for the correlation matrix of the features, X, i.e. $Cor(X_j, X_{j'}) = \rho^{|j-j'|}$ for $j \neq j' = 1, \ldots, p$. We compared the simulation when we have correlated predictors to the simulation where the predictors are independent, i.e. $\rho = 0$ or 0.8. Once again, we varied the number of predictors, $p = 1, 5, 20$, used for model selection. Table 2.2 shows the simulated values for P_B and P_H using each ITR identification method.

Across all scenarios there was a positive association between a positive correlation structure in the predictors and P_B when using any of the methods to identify an ITR. Even if the selection method selected the wrong predictors for an ITR, the selected predictors would still provide information towards an ITR better than $\widehat{w}^{opt}$. There was almost no effect of the correlation structure and the probability to identify a beneficial ITR when no model selection is performed.

		Treatment rule identification method					
ρ	p	None	Forward	λ_{min}-LASSO	$\lambda_{0.8}$-LASSO	Elastic Net	Random Forest
Single predictor with non-zero coefficient							
0	1	0.93/0.03	0.84/0.03	0.93/0.03	0.84/0.03	0.93/0.03	0.56/0.36
0	5	0.59/0.41	0.81/0.07	0.62/0.31	0.57/0.11	0.62/0.35	0.28/0.72
0	20	0.04/0.96	0.68/0.24	0.26/0.64	0.37/0.13	0.20/0.75	0.44/0.56
0.8	5	0.59/0.41	0.82/0.06	0.72/0.24	0.66/0.12	0.68/0.30	0.37/0.63
0.8	20	0.03/0.97	0.75/0.15	0.39/0.54	0.43/0.18	0.32/0.66	0.50/0.50
All predictor with non-zero and equal coefficients							
0	5	0.55/0.45	0.27/0.64	0.55/0.45	0.06/0.46	0.56/0.44	0.22/0.78
0	20	0.03/0.97	0.03/0.96	0.03/0.97	0.02/0.25	0.03/0.97	0.08/0.92
0.8	5	0.58/0.41	0.76/0.14	0.77/0.21	0.71/0.10	0.75/0.24	0.50/0.50
0.8	20	0.04/0.96	0.47/0.47	0.44/0.55	0.26/0.41	0.38/0.62	0.61/0.39
All predictor with non-zero and diminishing coefficients							
0	5	0.57/0.43	0.53/0.39	0.56/0.43	0.31/0.26	0.57/0.42	0.25/0.75
0	20	0.03/0.97	0.26/0.69	0.07/0.92	0.11/0.24	0.05/0.94	0.27/0.73
0.8	5	0.59/0.41	0.78/0.13	0.76/0.21	0.70/0.11	0.75/0.24	0.49/0.51
0.8	20	0.03/0.97	0.59/0.34	0.42/0.56	0.39/0.28	0.36/0.64	0.62/0.38

Table 2.2: P_B/P_H after varying values of the correlation between the predictors, the form of the coefficients for the predictors, the number of predictors, and the method used to identify the treatment rule. $\Delta = 0.3$, $V_y = 1$, $R_C^2 = 0.3$, $\nu = 0.16$. We also vary the number of coefficients associated with the outcome with only one predictor has a non-zero coefficient, the effect of the predictors evenly dispersed among the predictors, one predictor strongly associated with the outcome and treatment and a diminished effect in the others.

When more than one predictor was non-zero, the probability to identify a beneficial ITR was diminished compared to scenarios in which there was a single interaction. This effect was less noticeable in the context where the predictors were correlated or when any model selection was used. When using Forward Selection, which had performed well in the previous simulations, P_H was high and P_B was low when the effect of the predictors was evenly distributed across predictors. When the interaction effects between the predictors and treatment are all equally associated with the outcome, Forward selection struggled to identify the best single predictors at a time to add to the model that decrease AIC and it does not default back to $\widehat{w}^{opt}$. While

the $\lambda_{0.8}$-LASSO did not always show the highest probability of identifying a beneficial ITR, it did result in the lowest probability to cause harm in the population. When using the $\lambda_{0.8}$-LASSO model selection, we were able to default to $\widehat{w}^{opt}$ when the heterogeneity was difficult to identify.

2.5.4 Example 4: Comparison of LASSO Penalty Parameters

To Evaluate the LASSO model selection using the $\lambda_{0.8}$ penalty, we compared it to the LASSO models with penalty parameters λ_{min} and λ_{1se}. We used the same simulation described in example 2 where we attempted to identify an ITR after the trial was powered to identify a main effect. We considered the same three scenarios —the main effect was equal to the effect used to power the trial, the main effect was half of the effect used to power the trial, and there was no main effect—as well as an additional scenario where the true main effect is equal to the effect used to power the trial but there is no true treatment effect heterogeneity, $\nu = 0$. We compare the values of P_B as well as P_H in each scenario when using $\lambda_{0.8}$ to that of when using λ_{min} or λ_{1se}.

In the absence of treatment effect heterogeneity, $\nu = 0$, P_B is likely to be zero, assuming $\widehat{w}^{opt}$ estimates w^{opt} accurately, since $V(d^{opt}) = V(w^{opt})$ so it is almost impossible to have a beneficial ITR. In this scenario, the penalty parameter λ_{min} for the LASSO model selection identified a harmful ITR, i.e. failed to identify $\widehat{w}^{opt}$, with very high probability compared to when the other penalty parameters were used. Using λ_{1se} resulted in low probability to identify a harmful ITR when the main effect was non-zero including when there was no treatment effect heterogeneity but resulted in low values of P_B when the main effect for treatment was zero. This shows that λ_{1se} was too conservative of a penalty parameter in this example since an interaction term

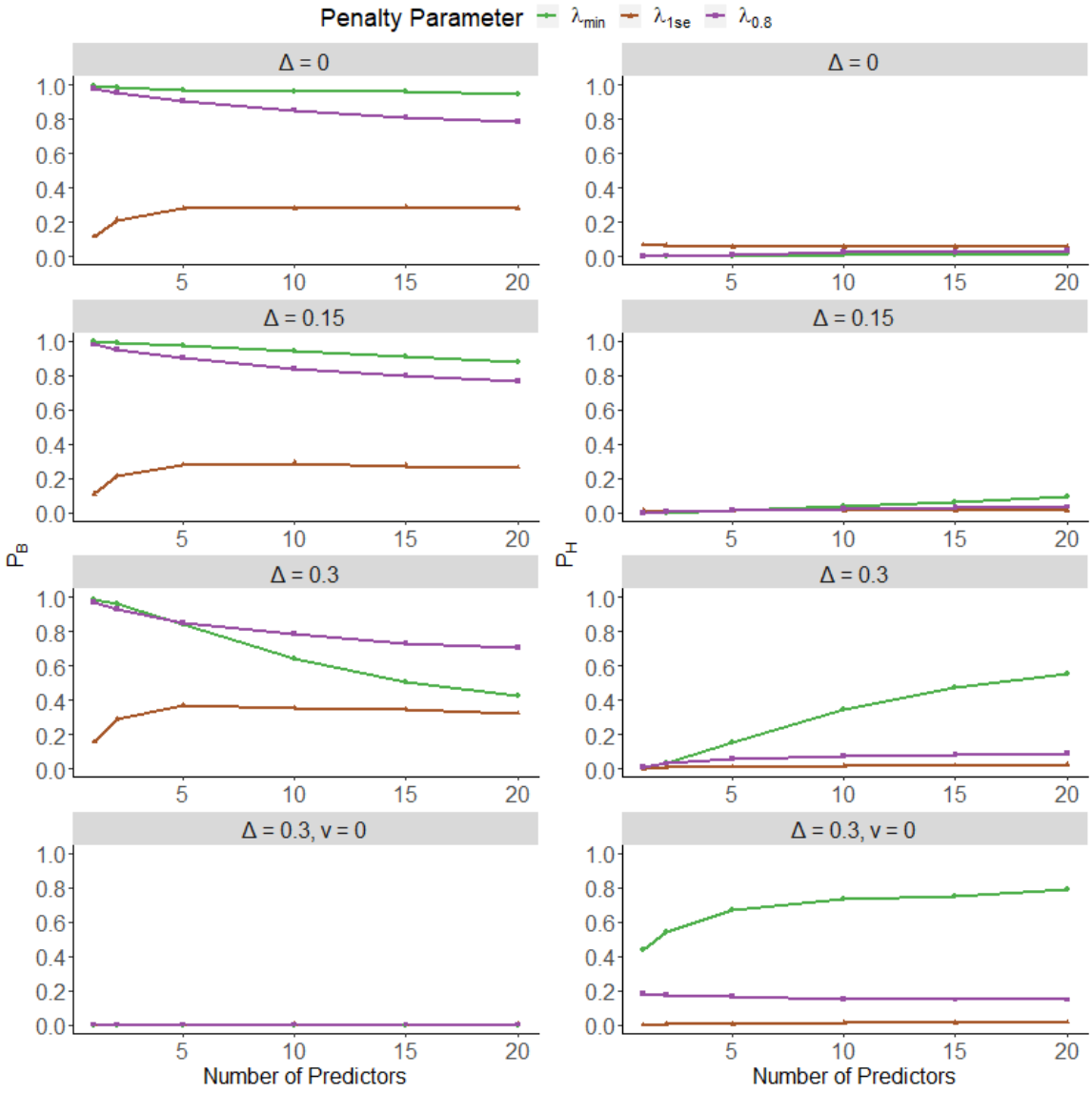

Figure 2.2: Values of P_B (left) and P_H (right) for three LASSO models with different penalty parameter under the same scenarios in Figure 2.1 plus an additional scenario where $\Delta = 0.3$, $\nu = 0$.

was hardly used in any of the models using λ_{1se}. The permutation based method to select the penalty parameter value, $\lambda_{0.8}$, resulted in the most consistent results across the different scenarios with high values of P_B and values of P_H which were

controlled. Using $\lambda_{0.8}$ resulted in a less aggressive model selection than using λ_{min} while not being as conservative as when using λ_{1se}.

2.6 Application to the PLUTO Study

The PLUTO Study is an ongoing two-stage SMART aimed at identifying sequences of treatment to assist with smoking cessation and reduction in number of cigarettes smoked per day. One secondary aim focuses on identifying personalized treatment rules but the likelihood of estimating a beneficial ITR has not been determined. Here we demonstrate how the proposed methods can be used in a trial in which baseline data have been collected but follow-up is ongoing. Usually treatment rules that are identified in SMARTs, known as dynamic treatment rules (DTR) assign treatment at multiple time points but for the purpose or this paper we focused on identifying beneficial ITRs at the first time point.

A total of 643 current smokers eligible for lung cancer screening were enrolled in the study and have completed baseline questionnaires. The study population is 64.4% male and 88.8% white, 92.7% have completed high school or above, 45.2% are married, and 90.4% have an income of $100,000 or less. Additional demographic information is given in Table 2.3. Participants were randomized with equal probability to four or eight weeks of Tobacco Longitudinal Care (TLC) which includes intensive telephone counselling along with nicotine replacement therapy. After the first stage of treatment, all participants are randomized to a different set of treatments, depending on how they respond to initial treatment. We focus solely on the development of an ITR for the number of cigarettes smoked per day 18 months after initial randomization.

Characteristic	N (%)	Characteristic	N (%)
N	**643**		
Age; Mean (SD)	64.4 (5.78)	Marital Status (%)	
Gender		Never Married	66 (10.3%)
Female	229 (35.6%)	Married	291 (45.3%)
Male	414 (64.4%)	Widowed	58 (9%)
Hispanic/Latino (%)	6 (0.9%)	Separated	16 (2.5%)
Missing	2 (0.3%)	Divorced	210 (32.7%)
Race (%)		Missing	2 (0.3%)
White	571 (88.8%)	Income (%)	
Black	32 (5%)	< $8K	15 (2.3%)
Other	10 (1.6%)	$8K-$15K	67 (10.4%)
More than one race	25 (3.9%)	$15K-$25K	61 (9.5%)
Missing	5 (0.8%)	$25K-$35K	78 (12.1%)
Education (%)		$35K-$50K	86 (13.4%)
≤ 8th grade	5 (0.8%)	$50K-$65K	88 (13.7%)
Some HS	41 (6.4%)	$65K-$80K	71 (11%)
HS Grad	176 (27.4%)	$80K-$100K	55 (8.6%)
Post HS Training	69 (10.7%)	>$100K	68 (10.6%)
Some College	211 (32.8%)	Missing	54 (8.4%)
BA/BS Degree	93 (14.5%)		
Grad/Professional Degree	47 (7.3%)		
Missing	1 (0.2%)		

Table 2.3: Summary of demographics from the 643 participants enrolled in the PLUTO study.

In order to calculate the probability to identify a beneficial ITR, we use the baseline data from the first stage of the PLUTO study. This data is then used as the population set for the simulation, i.e. $N_1 = 643$. We examine a set of 55 variables (table A1.3) consisting of demographic information, past 30 day cigarettes per day, age of first cigarette, number of quit attempts, certain medical history questions, readiness to quit, and various scored measures such as nicotine and alcohol dependence, depression, anxiety, etc.

We consider both the probability of benefit and the probability of harm, P_B and P_H, and simulated $N_2 = 1000$ data sets of size $n = 643$ by sampling with replacement from the study data. The coefficients for the relationship between X and Y were determined using the methods described in this paper. We assumed that only 5 variables had diminishing main and interaction effects and the rest had not association with Y. As an illustration, the 5 variables associated with Y were, in order of assumed importance: cigarettes smoked per day over the previous 30 days, number of past quit attempts, readiness to quit ladder, calculated pack years, and age. We assumed the variance of Y, which we have not observed yet, to be 1 since the data can always be standardized to have variance 1. We assumed the average treatment main effect to be 0.3. We ran two different analyses where we vary $R^2_C \in \{0.1, 0.9\}$ in one analysis and then vary $\nu \in \{0, 0.4\}$ in another analysis. In both analyses, we also examined how P_B and P_H are affected if we were *a priori* to narrow down the number of variables used in variable selection from 65 down to only the 5 actually associated with outcome.

For the first analysis, we varied the value of R^2_C and fixed $\nu = 0.16$. In figure 2.3 both P_B was substantially higher and P_H was lower overall when we narrowed down the number of predictors used in variable selection to five. According to plot (a), in order to reliably identify a beneficial or non-harmful ITR, a high value of R^2_C, 0.6 or above, is necessary. When $p = 65$ all model selection techniques resulted in values of P_B below 0.8 for almost all values of R^2_C. This demonstrates the importance of hypothesizing which predictors may be used to develop an ITR ahead of time rather than post-hoc. In plot (a), the $\lambda_{0.8}$-LASSO performed the best by identifying a harmful ITR the least but from plot (b), which shows values for P_B, we observe that this was largely due to the $\lambda_{0.8}$-LASSO identifying w^{opt} more often than other

model selection methods when and ITR cannot be successfully identified. This is seen most clearly when $p = 65$, P_H is much lower for the $\lambda_{0.8}$-LASSO than any other model selection but when looking at P_B it performed similar to other model selection methods.

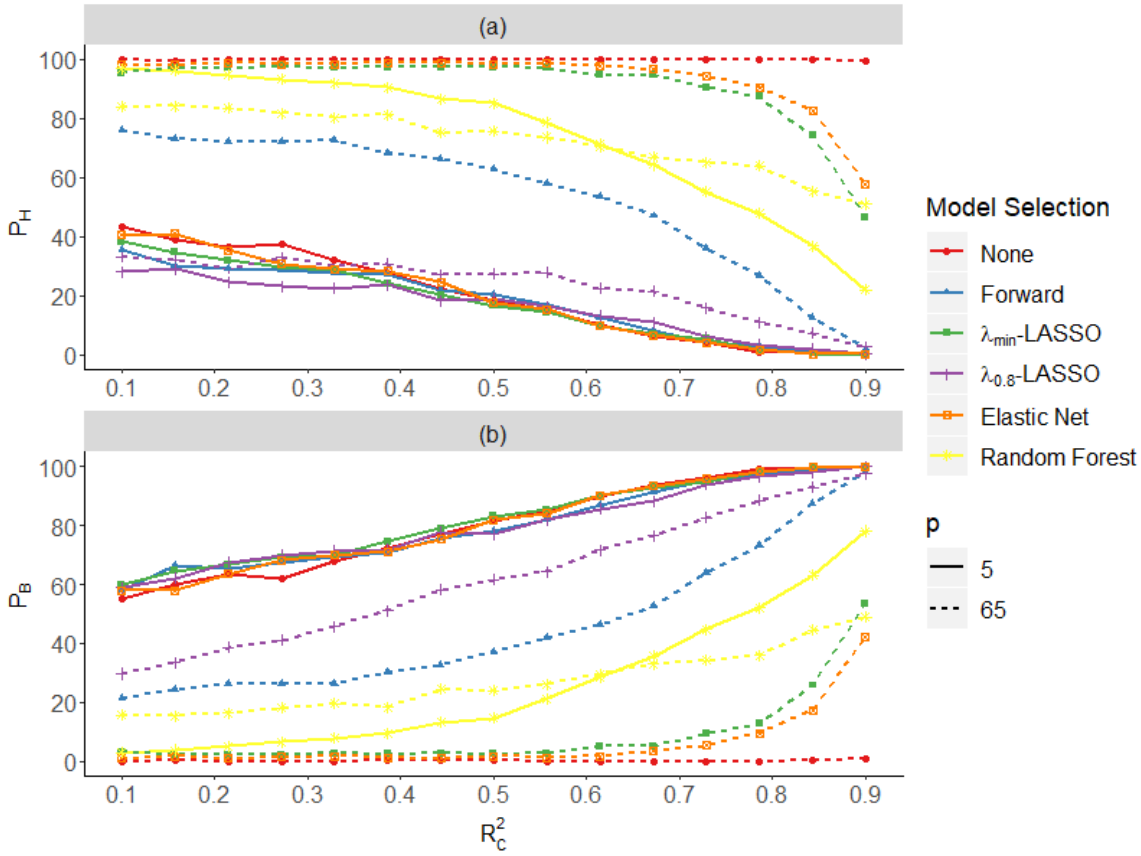

Figure 2.3: Comparison of (a) the probability of harm, P_H, and (b) the probability of benefit, P_B, for values of R_C^2 is between 0.1 and 0.9 in the PLUTO Study.

In the second analysis we fixed $R_C^2 = 0.3$ and looked into how P_B and P_H changed after varying the proportion of people that would benefit from control over treatment. In most cases, P_H in figure 2.4(a) is lower when $p = 5$ over when $p = 65$. When no treatment effect heterogeneity exists, the $\lambda_{0.8}$-LASSO identifies a harmful ITR

only 10% of the time, i.e. identifies $\widehat{w}^{opt}$, around 90% of the time. This is higher than expected since we should only identify $\widehat{w}^{opt}$ 80% of the time but we likely are observing an additional 10% due to interaction effects included but are so small that the identified ITR does not assign anyone to control in this population. When $p = 65$ the $\lambda_{0.8}$-LASSO identifies a harmful ITR less often than when $p = 5$ and when ν is small but greater than 0. Using figure 2.4(b), which shows the values of P_B, we can deduce that the $\lambda_{0.8}$-LASSO outperforms at $p = 65$ by selecting $\widehat{w}^{opt}$ more often than when $p = 5$.

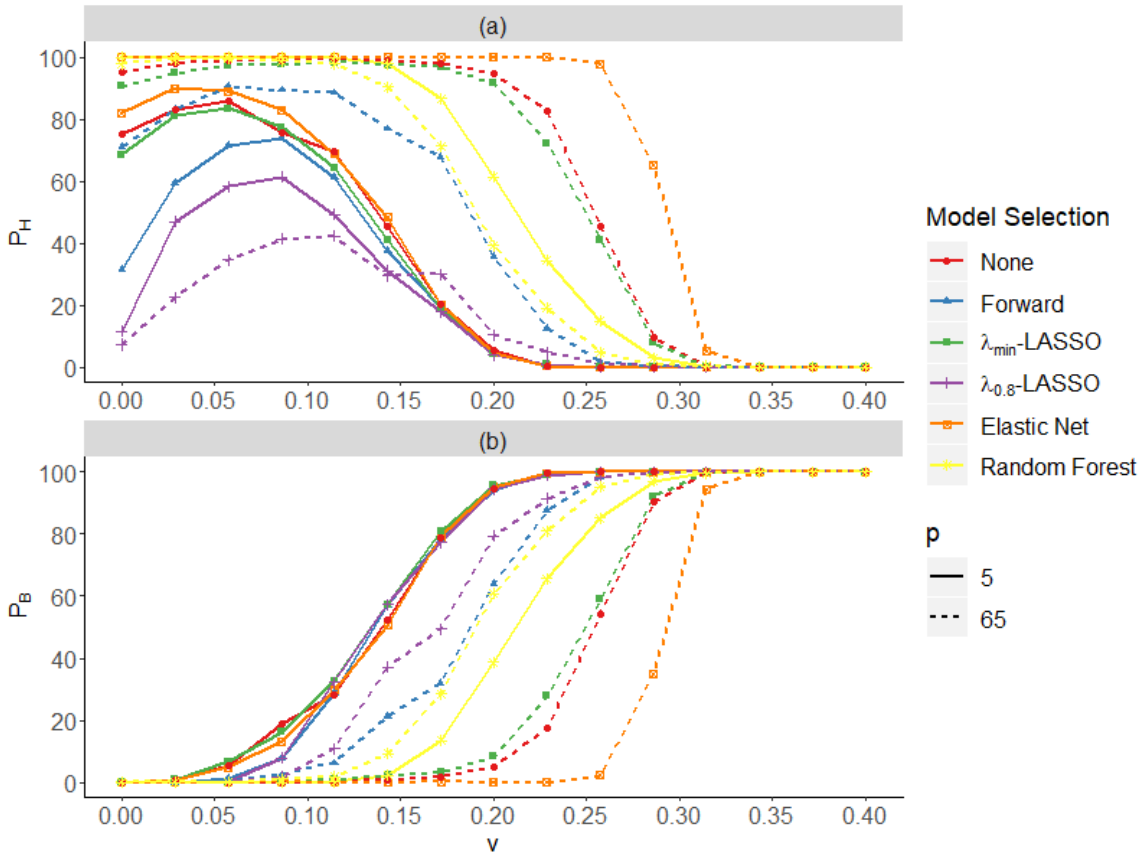

Figure 2.4: Comparison of (a) the probability of harm, P_H, and (b) the probability of benefit, P_B for values of ν between 0 and 0.4 in the PLUTO study.

In figure 2.4(a) there was a strange "peak" as ν increased. Specifically, P_H increased rapidly when $p = 5$ with the $\lambda_{0.8}$-LASSO once ν is greater than 0. When ν is small the ability to identify an ITR that has a small interaction effect is difficult as both the sign and magnitude of the interaction must be estimated accurately. This proves difficult when the interaction is quite small. For example when $\nu = 0.05$, only 5% of the population would benefit from the ITR and need to be assigned to control. Since there is true treatment effect heterogeneity, the $\lambda_{0.8}$-LASSO would not guarantee 80% probability to find a rule that is at least as good as $\widehat{w}^{opt}$ and it would instead identify an ITR more than 20% of the time. An identified ITR may assign 11% of the population to control which could be considered a success since only 6% of the population would be treated sub optimally. However, if w^{opt} were used, only 5% of the population would be treated sub optimally as only 5% would have benefited from control. Therefore, in this example the ITR causes harm since it did not perform at least as well as $\widehat{w}^{opt}$, assuming $\widehat{w}^{opt}$ estimated w^{opt} accurately.

2.7 Discussion

We introduced the idea of a beneficial ITR, which is a treatment rule that will lead to an average outcome in the population that is larger than if all individuals followed the estimated optimal static rule. To evaluate the probability to identify such a treatment rule a Monte Carlo numerical integration was required, and consequently we provided values to use to specify a data generative model. As well as evaluating the probability to identify a beneficial ITR, we also introduced the λ_q-LASSO method to select the LASSO penalty parameter to constrain the probability of identifying harmful ITRs

by returning a static rule with the specified probability in the absence of treatment effect heterogeneity. We did observe a slight increase in the probability of harm when identifying an ITR using the $\lambda_{0.8}$-LASSO when the treatment effect heterogeneity was small but non-zero. A possible fix for this increase in probability could be identifying λ_q across other values of ν to be even more conservative when the heterogeneity is small.

Many other methods have been presented to identify ITRs, they could allow for identification of a beneficial ITR but their ability to do so is yet to be determined. Zhang[12] introduced a robust method to estimate an ITR via augmented inverse probability weighting. This method, while robust to misspecified models still performed best under a correctly specified model. Many methods to identify ITRs have incorporated data mining methods such as LASSO regression or Random Forests. Imai and Ratkovic[15] incorporated LASSO constraints with Support Vector Machines in order to identify heterogeneous treatment effects. Qian and Murphy[8] also used LASSO regression to identify an ITR such that its value converges to the value of the true optimal treatment rule as sample size increases. Gunter et. al.[7] used a LASSO model as a model selection tool to identify qualitative interactions. Variables are added based on the Adjusted Gain in Value which compares the value of an ITR to that of the static rule. Ballarini et. al.[5] also used LASSO regression as a model selection tool to identify variables that may interact with the treatment's effect on the outcome and determined subgroups such that the CATE was larger than a specified value. Classification and Regression Trees (CART) and Random Forests[21] have also seen extensive use in the identification of ITRs due to their ability to detect qualitative treatment-predictor interactions. Ruberg et. al.[9] discussed how Random Forests

can be used to identify heterogeneous treatment effects and Su et. al.[10] introduced regression trees to identify ITRs where the variable for each split is determined such that the square of a T-test statistic for the interaction term of said variable with treatment in a linear model is maximized. Foster et. al.[6] introduced the use of random forests to obtain the CATE and determine an ITR such that the CATE is maximized. Athey and Imbens[4] introduced the idea of causal trees which modified the conventional CART specifically to estimate CATEs and provide an application to data mining heterogeneous treatment effects when they are not hypothesized *a priori*.

One limitation is that the data generating models and analysis methods considered were not as sophisticated as other methods. Similarly, we estimate the probabilities using the true value of an estimated ITR rather than the estimated value which assumes that the trial sample is actually representative of the population. However, our probability calculation can be likened to a power calculation and simplifying assumptions are made in almost all power calculations in order to make them possible but in practice there may be different components of variability that are not considered with power. Another limitation for the data generative model is that values such as R_C^2 and ν may not be readily available and no prior data may exist so it may be necessary to collect pilot data in order to reliably calculate the power to identify a beneficial ITR in this scenario, however, this is not required. Like many clinical trials, the measurement of outcome or predictors could be inaccurate based on availability of measurement tools and can also affect our ability to identifying beneficial ITRs.

Using data from the PLUTO study, we demonstrated difficulty with identification of a beneficial or non-harmful ITR in real world data. We require a high amount of

variance explained, a large proportion of the population that would benefit from control over treatment, or both in order to reliably identify an beneficial ITR. The LASSO model where the penalty parameter is specifically chosen to minimize error when no treatment effect heterogeneity exists performed well and was able to overcome many of the difficulties of personalization of treatment although we saw significant drops in the probability of benefit when the proportion of people that benefit from an ITR over a static rule was small but non-zero. Additionally we found that reducing the number of predictors used when attempting to identify a beneficial ITR resulted in a higher probability showing the importance of testing for an ITR with variables with prior evidence of heterogeneous treatment effect whenever possible.

Clinical trials that aim to identify an ITR often do not consider the risks of identifying an ITR that fails to improve on a static rule. Some trials provide possible personalization of treatment post-hoc when a null or less than expected main effect is found, and the probability to identify a beneficial ITR is hardly considered in these contexts as well. While we did not consider it in any of our examples, cost or toxicity of treatment could also be used in addition to the outcome to adjust the definition of what is considered a beneficial ITR. Whether personalization of treatment is set as a secondary aim or primary aim, the probability to identify a beneficial ITR (as well as a non-harmful ITR) should be evaluated *a priori* to provide guidance on whether identification of ITRs should be performed at all. This ultimately will avoid implementing ITRs that perform worse than the optimal static rule.

Chapter 3

Identification of Non-Harmful DTRs Using LASSO With Permutation-Based Selection of the Penalty Parameter

3.1 Introduction

When evaluating patients with chronic conditions (e.g. hypertension, cancer, drug and alcohol abuse), clinicians are often faced with a variety of decisions over time. A common approach is to assign one intervention and then evaluate the effectiveness of that treatment after some time. This is done to avoid an excess of treatment, monitor side effect severity, manage intervention risk, and account for non-adherence. Depending on how well the participant/patient responds to the treatment a new treatment may

be assigned, the same treatment continued as is, or the same treatment continued augmented with a different treatment or with a modified dose or frequency. A Dynamic Treatment Regime (DTR) is a clinical tool to guide the treatment decisions of clinicians which assigns treatment at each decision point over time based on patient characteristics including prior response to treatment. Most randomized controlled trials do not test a sequence of treatments and instead examine the assignment of a single treatment and compare the resulting clinical outcome to that of a control. Sequential Multiple Assignment Randomized Trials [24, 25] (SMARTs) were designed to assess the effectiveness of DTR by randomizing participants to treatment options at multiple decision points. Several DTRs are often *embedded* in the trial design, (i.e., embedded DTRs) depending on the number of decision points considered and the number of treatment options tested at each decision point.

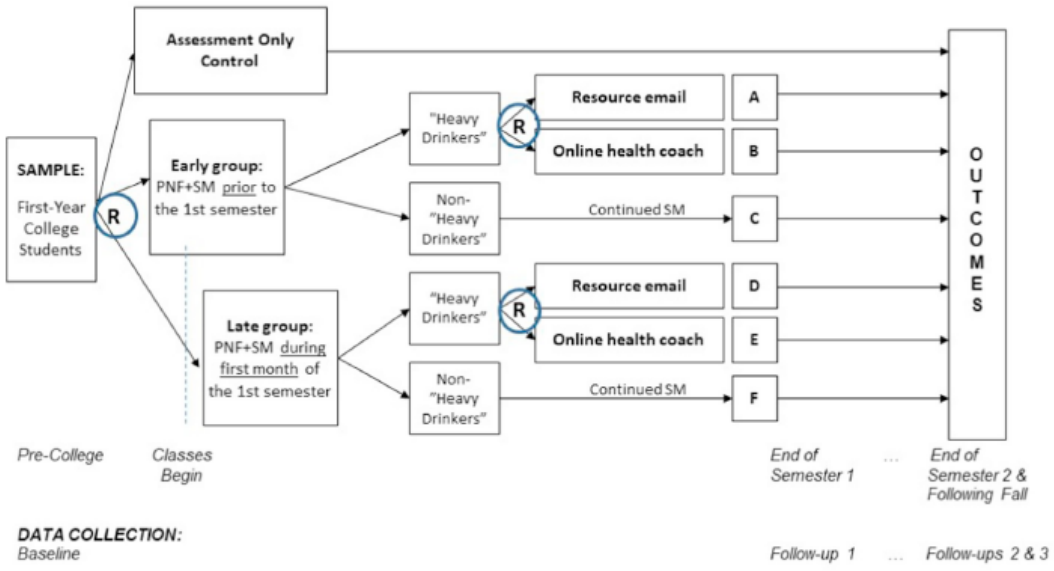

Figure 3.1: M-Bridge Study design overview.

One example of a SMART is the M-bridge trial [26] which was aimed at assessing

DTRs to prevent binge drinking in college freshman. 891 full-time, first-year college students aged 18-21 years at the University of Minnesota Twin Cities were enrolled in the trial. Participants were initially assigned with 2:1 randomization to DTRs which begin with personalized normative feedback (PNF) and self-monitoring (SM) or an assessment only control (no PNF or SM). Of those assigned to DTRs, half were randomly assigned to the early group in which participants received PNF 2 weeks before classes started with SM for 8 weeks afterward, and the other half were randomly assigned to the late group in which participants received PNF 2 weeks after classes started with SM for 8 weeks afterward. Each SM survey assessed alcohol use which determined response to the initial intervention. Non-responders, i.e. participants assessed to be "heavy drinkers" (2 or more binge drinking episodes in the previous 2 weeks or at least 1 high intensity drinking episode), were re-randomized to receive access to an online health coach or a resource email (Figure 3.1). Responders, i.e. participants never assessed to be "heavy drinkers," were not re-randomized and continued to receive SM surveys. In this trial, assessment of whether a student was a "heavy-drinker" was an embedded tailoring variable, i.e., a variable which determines the set of treatments to which a person can be randomized. There were four DTRs within the M-bridge study that only depend on the embedded tailoring variable, shown in Table 3.1. Each embedded DTR incorporated one of the initial treatments then a follow up treatment for any non-responders, e.g., DTR 2 provided everyone with PNF+SM 2 weeks before start of classes and then non-responders were given an invitation to online health coaching. The primary outcome of interest was number of binge drinking episodes in the previous month assessed at the end of the fall semester; secondary outcomes included health services utilization and maximum

number of drinks consumed in a given day.

	Treatment Assignment			
DTR	Initial	Responders	Non-responders	Figure 3.1 paths
1	Early PNF+SM	Continue SM	Resource email	A & C
2	Early PNF+SM	Continue SM	Online health coach invitation	B & C
3	Late PNF+SM	Continue SM	Resource email	D & F
4	Late PNF+SM	Continue SM	Online health coach invitation	E & F

Table 3.1: Embedded DTRs withing the M-Bridge study

One secondary aim of the M-Bridge study focuses on identifying more deeply tailored DTRs which assign treatment based on participant characteristics in addition to the embedded tailoring variables. Many methods to identify DTRs involve some form of backward induction or dynamic programming algorithms [27]. Backward induction identifies the best treatment at the final decision point in the sequence as the treatment that maximizes a prespecified value such as the expected outcome conditioned on clinical history. The best penultimate treatment is selected to maximize the same value assuming the optimal treatment will be assigned at the final decision point. This process continues until a treatment rule is decided at each decision point. Q-learning [28], A-learning [24, 29], and Backwards Outcome Weighted Learning [30, 31] all use backwards induction and select the treatment rules at each decision point to maximize different values. Q-learning maximizes the estimated Q-functions in order to estimate the optimal treatment rule at each decision point. The Q-functions are the expected outcome given clinical history assuming future/subsequent treatments are assigned accordingly, usually the estimated optimal DTR.

As Q-learning involves specifying regression models, some form of variable selection is needed to improve performance and derive parsimonious decision rules. Existing variable selection methods used in Q-learning such as S-Scores [7, 32], which

focus on selecting qualitative interactions, and forward selection [33] do not specifically aim to identify DTRs which improve upon the DTRs embedded in the trial design (which we term a beneficial DTR) with specified probability. In the absence of treatment effect heterogeneity for each treatment, a beneficial DTR is not possible (i.e., a more deeply tailored rule cannot improve upon the optimal embedded DTR), and no existing variable selection methods for Q-learning guarantee the identification of the estimated optimal embedded DTR in that scenario with a specified a priori probability. This study introduces a permutation-based method to select the LASSO penalty parameter such that non-harmful DTRs are identified by selecting the estimated optimal embedded DTR when no treatment effect heterogeneity exists. These methods often result in a more conservative approach to identify more deeply tailored DTRs and can result in more interpretable DTRs.

We first formalize the framework surrounding beneficial DTRs and Q-learning in section 3.2. Then we introduce the variable selection method to identify non-harmful DTRs using Q-learning in the context of the M-Bridge study in section 3.3 and review some existing methods to identify a more deeply tailored DTR in section 3.4. We also simulate several possible scenarios in section 3.5, which reflect a variety of possible combinations of treatment main effect and heterogeneity, and compare our proposed approach to existing methods. In section 3.6 we apply these methods to identify a more deeply tailored DTR in the M-Bridge study. We conclude in section 3.7

3.2 Notation

We consider SMARTs with two decision points $t = 1, 2$ although the proposed approach could generalize to more than two decision points. Let Y be the final outcome, X_t be the p_t dimensional vector of features collected before the t^{th} decision point which includes R_{t-1}, the response to treatment at the previous stage and let $A_t \in \{-1, 1\}$ be the treatment assignment at the t^{th} stage. Note that A_t in principle may depend on R_{t-1} (i.e. responders and non-responders can be re-randomized to different sets of treatments). If a participant is not re-randomized at the t^{th} stage then we set $A_t = 0$. The observed data is a time ordered quintuple $(X_{1i}, A_{1i}, X_{2i}, A_{2i}, Y_i)_{i=1,\ldots,n} \sim F$ and we denote $H_t,\ t = 1, 2$ as the cumulative information available when treatment A_t is assigned, i.e. $H_2 = (X_1, A_1, X_2)$ and $H_1 = X_1$.

One of the goals in SMARTs is to identify a DTR, d that maximize a given outcome in a population, if everyone were to follow the regime. A DTR, $d = (d_1, d_2)$, is comprised of decision rules at each time point, d_t which map the domain of H_t, $\mathcal{H}_t$, to the set of treatments, $d_t : \mathcal{H}_t \to \{-1, 1\}$. We define the potential outcome of the response if the DTR d is followed as $Y(d)$ and the value of d is defined as $V(d) = E\{Y(d)\}$, the expected potential outcome. The optimal DTR, d^{opt}, is defined as the DTR with largest value. The features X_2 are dependent on X_1 and A_1 so the potential outcome of X_2 under DTR, d is written as $X_2(d_1)$ as it is only dependent on the first decision rule in d.

Q-learning [28] is a regression based algorithm to identify DTRs. The algorithm relies on the Q-functions:

$$Q_2(h_2, a_2) = E\{Y(a_2)|H_2 = h_2\} = E\{Y|H_2 = h_2, A_2 = a_2\}, \tag{3.1}$$

$$Q_1(h_1, a_1) = E\{Y(a_1, d_2^{opt})|H_1\} = E\{\max_{a_2 \in \mathcal{A}(h_2)} Q_2(h_2, a_2)|H_1 = h_1, A_1 = a_1\}, \tag{3.2}$$

where $\mathcal{A}(H_2)$ is the set of available treatment options at the second stage given history H_2 (e.g., $\{-1, 1\}$ for non-responders and 0 for responders in the M-Bridge study) and where the second equality follows from the standard causal identifying assumptions [2] including no unmeasured confounding which is guaranteed when using randomization.

The optimal treatment rules are determined by maximizing each Q-function, $d_t^{opt}(h_t) = \text{argmax}_{a_t \in \mathcal{A}_t(h_t)} Q_t(h_t, a_t)$. The Q-functions, $Q_t, t = 1, 2$ can be expressed with real valued functions, $b_t, c_t : H_t \to \mathbb{R}, t = 1, 2$, which may depend on a set of finite dimensional parameters, α and β.

$$Q_2(h_2, a_2; \alpha_2, \beta_2) = b_2(h_2; \alpha_2) + c_2(h_2; \beta_2)a_2 \tag{3.3}$$

$$Q_1(h_1, a_1; \alpha_1, \beta_1) = b_1(h_1; \alpha_1) + c_1(h_1; \beta_1)a_1. \tag{3.4}$$

For this study we only consider modified Q-learning [34] and the algorithm to determine the optimal DTR proceeds by first estimating b_2 and c_2 using standard regression-based techniques with Y as the dependent variable. If $\mathcal{A}_t(H_2) = \{-1, 1\}$ then the estimated optimal treatment rule at the second stage can be written as $\widehat{d}_2^{opt}(h_2) = sign\{c_2(h_2; \widehat{\beta}_2)\}$. Then, we predict the outcome if the participant were to follow the estimated optimal regime at the second stage as the outcome plus the

"regret" [24] from not assigning the optimal treatment assignment,

$$\tilde{Y}_i = Y_i + [Q_2(H_{2i}, \widehat{d}_{2i}(h_{2i}); \widehat{\alpha}_2, \widehat{\beta}_2) - Q_2(H_{2i}, a_{2i}; \widehat{\alpha}_2, \widehat{\beta}_2)]. \tag{3.5}$$

Note that when the estimated optimal treatment is assigned to participant i, i.e. $a_{2i} = \widehat{d}_{2i}(h_{2i})$, the regret is equal to 0 so $\tilde{Y} = Y$. Additionally, $\tilde{Y} = Y$ for those that are not re-randomized, e.g. the responders in the M-Bridge study. We use standard regression-based methods with $\tilde{Y}$ as the outcome in order to estimate b_1 and c_1 in equation (3.4). In general, the estimated optimal treatment rules taken from Q-learning are defined as

$$\widehat{d}_t^{opt}(H_t) = \text{argmax}_{a_t \in \mathcal{A}_t(h_t)} Q_t(h_t, a_t; \widehat{\alpha}_t, \widehat{\beta}_t) = sign\{c_t(H_t; \widehat{\beta}_t)\}. \tag{3.6}$$

The value of a DTR, d, identified through Q-learning is

$$\begin{aligned}
V(d) &= E_Y\{Y(d)\} = E_{X_1, X_2(d)}[E_{Y|X_1, X_2(d_1)}\{Y(d)|X_2(d_1), X_1\}] \\
&= E_{X_1, X_2(d_1)}[E_{Y|X_1, X_2(d_1)}\{Y|X_2(d_1), X_1, A_2 = d_2(H_2), A_1 = d_1(H_1)\}] \\
&= E_{X_1}(E_{X_2|X_1}[E_{Y|X_1, X_2}\{Y|X_2, X_1, A_2 = d_2(H_2), A_1 = d_1(H_1)\}|X_1, A_1 = d_1(H_1)]) \\
&= E_{X_1}(E_{X_2|X_1}[b_2(H_2) + c_2(H_2)d_2(H_2)|X_1, A_1 = d_1(H_1)]) \\
&= \iint b_2\{(x_1, d_1(h_1), x_2)\} + c_2\{(x_1, d_1(h_1), x_2)\}d_2(h_s)dF_{X_2|X_1, d_1(H_1)}(x_2)dF_{X_1}(x_1)
\end{aligned}$$

Many statistical methods for identifying DTRs involve estimating the optimal DTR, d^{opt}. Doing so often results in a more deeply tailored DTR that depends on multiple patient characteristics and how they respond to treatment. These more deeply tailored DTRs can be time and cost intensive for both clinicians and patients;

whereas, the optimal embedded DTR may perform just as well or even better. We define an embedded DTR, w, as a DTR within a SMART that depends on only the embedded tailoring variables. The estimated optimal embedded DTR $\widehat{w}^{opt}$ is defined such that $V(w^{opt}) \geq V(w)$ for all embedded DTRs, w. When considering a potentially complex and more deeply tailored DTR, we believe that a more deeply tailored DTR must perform as well or better (i.e., result in a higher value) than the simpler estimated optimal embedded DTR. We define a beneficial DTR, d such that

$$V(d) > V(\widehat{w}^{opt}). \tag{3.7}$$

Conversely, a harmful DTR is defined such that $V(d) < V(\widehat{w}^{opt})$; a non-harmful DTR replaces the strict inequality in equation (3.7) with a non-strict inequality.

3.3 λ_q-LASSO method to identify a non-harmful DTR

Typically, the Q-functions in Q-learning are estimated using standard regression techniques. There is no shortage of model selection techniques for regression problems which could be implemented in this context. However, we develop a variable selection approach to ensure that if a more deeply tailored DTR is selected, it is less likely to result in harm compared to other available variable selection methods. This is especially crucial in the scenario where there is no treatment effect heterogeneity where our method identifies a non-harmful DTR (i.e. the estimated optimal embedded DTR) with high a priori probability.

LASSO variable selection [14] uses L_1 penalized regression to select parameters in a model with non-zero coefficients. LASSO model coefficients, β, with outcome, Y, and predictors, X, are estimated as

$$\hat{\beta} = \underset{\beta \in \mathbb{R}}{\operatorname{argmin}} \sum_{i=1}^{n} (Y_i - X_i\beta) + \sum_{j=1}^{p} \widetilde{\lambda}_j |\beta_j|, \tag{3.8}$$

where $\widetilde{\lambda}$ is the vector of the penalty parameters such that $\widetilde{\lambda}_j = 0$ would imply no penalty is given to predictor j. Typically $\lambda_j = \lambda s_j$ where s_j is an indicator if predictor j is to be penalized. Some form of cross-validation [17] to minimize mean square prediction error or other measure of predictive performance is typically used to select a value for the penalty parameter, but we consider a permutation-based algorithm for this study.

When estimating c_1 and c_2 during Q-learning, it is often assumed that model selection is done ahead of time and many methods do not consider performance under the "null" scenario where there is no treatment effect heterogeneity. When treatment effect heterogeneity is absent no method can estimate a (strictly) beneficial DTR since the optimal treatment rule would be the optimal embedded DTR. A non-harmful DTR can be observed but no current methods guarantee its identification with specified probability in the absence of treatment effect heterogeneity.

We consider linear models for the Q-functions i.e.,

$$\begin{aligned} b_1(h_1) &= \alpha_{10} + h_1'\boldsymbol{\alpha}_{11}, \quad c_1(h_1) = \beta_{10} + h_1'\boldsymbol{\beta}_{11}, \\ b_2(h_2) &= \alpha_{20} + h_2'\boldsymbol{\alpha}_{21} \quad c_2(h_2) = \beta_{20} + h_2'\boldsymbol{\beta}_{21}. \end{aligned}$$

To induce variable selection, we estimate α_{20}, $\boldsymbol{\alpha}_{21}$, β_{20} and $\boldsymbol{\beta}_{21}$ by solving equation

(3.8) with outcome Y and covariates (h_2, a_2, a_2h_2). Similarly, we estimate α_{10}, α_{11}, β_{10}, and β_{11} by solving equation (3.8) with outcome $\tilde{Y}$ and covariates (h_1, a_1, a_1h_1). However, rather than choose the penalty parameter using cross-validation, we propose the following algorithm which we refer to as the λ_q-LASSO. This algorithm permutes the treatment assignment and subtracts the main effect of treatment from the outcome to simulate data where treatment and outcome are not associated. This breaks any treatment by covariate interactions and their association with the outcome if present and imitates data without treatment effect heterogeneity. Then using a LASSO model to estimate the coefficients we evaluate if any interactions are present over multiple iterations. We then select the penalty parameter, λ, such that no interactions are selected with user specified probability under the scenario where association between treatment and outcome have been removed from the data.

We define the λ_q-LASSO with data (H_2, A_2, Y) and a set of penalty parameter values $\mathbf{\Lambda} = (\mathbf{\Lambda}_1, \mathbf{\Lambda}_2)$ for $t = 1$ and 2 respectively. We initially create a pseudo outcome Z_2:

$$Z_2 = \begin{cases} Y - \Delta_{2a}, & \text{if } A_1 = a, A_2 = 1 \\ Y, & A_2 = -1 \text{ or } 0, \end{cases}$$

where Δ_{2a} represents the main effect of the second treatment in initial treatment group, $A_1 = a$. Since Δ_{2a} is often unknown, we instead use the estimated treatment effect, which can be found by taking the difference between the sample means of both second stage treatment groups. For each iteration b, we permute A_2 to create $\tilde{D}_2^{(b)} = (H_2, \tilde{A}_2^{(b)}, Z_2)$. Then, for each $\lambda_2 \in \mathbf{\Lambda}_2$ we fit a LASSO model for Q_2 regressing Z_2 on H_2 and $\tilde{A}_2^{(b)}$ using $\tilde{D}_2^{(b)}$ with penalty parameter λ_2. We evaluate if interaction

terms in the model are zero,

$$I_{\lambda_2}^{(b)} = \mathbb{1}(\widehat{\boldsymbol{\beta}}_{21}^{\lambda_2(b)} = \mathbf{0})$$

where $\mathbf{0}$ represents a vector of zeros and $\widehat{\beta}_{21}^{\lambda_2(b)}$ represents the estimated interaction coefficients in the LASSO model under penalty parameter λ_2 for iteration b. Repeating this process B (typically chosen to be 1000) times, the proportion of LASSO models that result in no interaction is estimated using

$$\hat{P}(I_{\lambda_2} = 1) = \frac{1}{B}\sum_{b=1}^{B} I_{\lambda_2}^{(b)},$$

where $I_{\lambda_2}^{(b)}$ represents the evaluation of I_{λ_2} for iteration b. We then select λ_{2q} as the smallest value in $\mathbf{\Lambda_2}$ such that $\hat{P}(I_{\lambda_2} = 1) > q$ where q is the user specified probability that the embedded DTR should be selected in the absence of treatment effect heterogeneity at the current treatment stage.

Using the covariates with non-zero coefficients selected using the LASSO model, we estimate Q_2 using least-squares to determine $\widehat{d}_2^{opt}(h_2) = sign\{\widehat{\beta}_{20} + h_2'\widehat{\boldsymbol{\beta}}_{21}\}$ and predict values $\tilde{Y}$ using equation 3.5. We then perform a similar LASSO variable selection with Q_1 where the pseudo outcome is the final outcome assuming the estimated best second treatment is assigned with no average effect from the first treatment,

$$Z_1 = \begin{cases} \tilde{Y} - \Delta_1, & \text{if } A_1 = 1 \\ \tilde{Y}, & A_1 = -1, \end{cases}$$

where Δ_1 is estimated, $\widehat{\Delta}_1 = \widehat{E}\{\tilde{Y}|A_1 = 1\} - \widehat{E}\{\tilde{Y}|A_1 = -1\}$.

3.4 Existing methods to identify a more deeply tailored DTR

Some methods available focus on estimating Individualized Treatment Rules (ITRs) at a single time point but can easily be translated to two-stage SMART designs. We consider two existing methods to identify treatment rules: forward selection based on QIC [33] and S-Scores [7].

3.4.1 Forward Selection

The methods presented by Wallace [33] use Forward Selection to identify DTRs based on QIC values for G-estimation [35]. To perform Forward Selection at stage t based on QIC in Q-learning we first consider generalized estimating equations with independent covariance matrix structure and only one observation per participant at the time point t. Using generalized estimating equations in this way results in estimated coefficients equal to the coefficient estimates obtained by least squares but allows us to easily calculate QIC. We proceed with traditional forward selection by adding in feature main effect or interactions with treatment one at a time based one whichever results in the lowest QIC [36]. We continue to add variables until adding any additional variables results in a non-decreased QIC value. The models identified are used within Q-learning as defined in section 3.2.

3.4.2 S-Scores

S-Scores [7], allow for the identification of ITRs by means of selecting qualitative interactions at a single stage. To use S-scores in Q-learning at stage t of treatment

assignment, we first consider LASSO models estimating Q_t with all predictor main effects and interactions using a sequence of penalty parameters Λ. We start by estimating coefficients for LASSO models for each $\lambda \in \Lambda$ and sort the model predictors in the order that they become zero with the coefficient that becomes zero due to the highest value λ first and the coefficient that becomes zero due to the lowest value λ last. We then put the coefficients into nested subsets indexed $k = 0, \ldots, K$ such that the 0^{th} subset contains only A_t and any previous treatment assignment and the k^{th} subset contains the predictors included in the $k^{th} - 1$ subset and the k^{th} predictor in the sorted list of predictors. If the predictor interaction with treatment is included in the subset, we also force the predictor main effect to be included in that subset as well per Gunter [7]. For each subset of predictors we fit a linear model Q_t^k using those predictors, estimate treatment rule d_t^k (which may be a static rule that assigns the same treatment to all) and obtain predicted values as the empirical average $\widehat{Q}_t^k = Q_t(H_t, A_t = d_t(h_t); \widehat{\boldsymbol{\alpha}}_t, \widehat{\boldsymbol{\beta}}_t)$. Using those predicted values, we can compute the S-Score for subset k,

$$S_t(k) = \frac{\widehat{Q}_t^k - \widehat{Q}_t^0}{\widehat{Q}_t^m - \widehat{Q}_t^k} \frac{m}{k},$$

where $Q_t^m = \max_k Q_t^k$. We select the variable set t that results in the highest $S(k)$ and estimate Q_t using those variables and determine $d_t(h_2)$.

3.5 Simulation Study

For all simulations we considered a continuous outcome Y with baseline covariates $X_1 \sim N_p(0, I_p)$ where I_p was an independent correlation matrix of size p and initial treatment, $A_1 \in \{-1, 1\}$ was assigned with 1:1 randomization and response to initial

treatment. Response was simulated $R_1|A_1 = a_1, X_1 \sim Bernoulli(P_{a_1})$ where P_{a_1} represents the probability of response to treatment a_1 and treatment, $A_2 \in \{-1, 1\}$ assigned with 1:1 randomization to the non-responders with the outcome of interest, Y simulated such that

$$E[Y|X_1, R_1, A_1, A_2] = X_1'\beta_1 + A_1(\beta_2 + X_1'\beta_3) + R_1\beta_4 + A_2(1-R_1)(\beta_5 + A_1\beta_6 + X_1'\beta_7)$$

and $Var(Y|X_1, R_1, A_1, A_2) = \sigma^2$ is fixed to be 0.9.

Due to the interactions in the model many of the parameters are difficult to interpret individually. However, these parameters can be related to more interpretable values such as proportion of people to benefit from an individualized rule at the second stage in treatment group $A_1 = a$, $\nu_{2a} = \Phi(\frac{-(\beta_5 + a\beta_6)}{\beta_7 \Sigma_X})$, main effect of second treatment assignment in treatment group $A_1 = a, \Delta_{2a} = 2(\beta_5 + a\beta_6)$, main effect of responding to $A_1 = a$, $\Delta_R = \beta_4 - E_{X_1}[E_{A_2}\{A_2\beta_5 + aA_2\beta_6 + A_2X_1'\beta_7\}] = \beta_4$. Additionally, we can define proportion of variance explained by X in the treatment group $A_1 = A_2 = 1, R_{11}^2 = \frac{(\beta_1+\beta_3+\beta_7)'\Sigma_X(\beta_1+\beta_3+\beta_7)}{\sigma^2+(\beta_1+\beta_3+\beta_7)'\Sigma_X(\beta_1+\beta_3+\beta_7)}$, the main effect of the first treatment, $\Delta_1 = \beta_2 + E_{X_1,R_1}\{(1 - R_1)(|\beta_5 + \beta_6 + X_1'\beta_7| - |\beta_5 - \beta_6 + X_1'\beta_7|\}$, and the proportion of people that benefit from a heterogeneous treatment for $t = 1$, $\nu_1 = P(\beta_2 + X_1'\beta_3 + (1-R_1)\beta_6 < 0)$. In describing the simulation scenarios, we report these summary measures as they are more meaningful quantities to a domain-are expert.

Data were simulated by generating $n = 500$ observations simulated from the data generating model specified in table A2.1. We simulated $M = 2500$ datasets and performed modified Q-learning to estimate an optimal DTR using one of four variable

selection methods: No variable selection, forward selection using QIC, S-Scores, and $\lambda_{0.8}$-LASSO. We simulated data under three possible scenarios at each decision point: Main treatment effect (i.e., non-zero main effect of treatment) with heterogeneous effect, main treatment effect with no heterogeneous effect, and no main or heterogeneous treatment effect. We simulated each possible combination of scenarios at the two decision points resulting in 9 total scenarios. Within each scenario, we varied the number of predictors and interaction effects with treatment considered in the data generating model, p_{cons}, as well as the number of interaction effects that are truly non-zero, p_{true}. Specifically, we considered $(p_{cons}, p_{true}) = (1,1), (5,1), (5,5), (20,5), (20,20)$. In all scenarios we defined coefficients such that R^2_{11} was approximately 0.4 when there were heterogeneous effects at both $t = 1, 2$ and the main effect of responding to treatment at $t = 1$, $\Delta_R = 0.3$. In scenarios where there was a main treatment effect at $t = 2$, we set the average main treatment to be 0.3, i.e. $\frac{\Delta_{20}+\Delta_{21}}{2} = 0.3$. Likewise, when there was a main treatment effect at $t = 1$, the average treatment effect was set to be 0.3 with some variation due to in the scenarios where the first stage treatment had a heterogeneous effect with second stage treatment. When there was a heterogeneous treatment effect at $t = 1$, we defined coefficients such that the proportion of people that would benefit from a more deeply tailored rule at $t = 1$ was approximately 0.2. At $t = 2$, we defined the treatment-predictor interactions such that ν_{21} and ν_{20} averaged to be approximately 0.2 as well as at $t = 1$.

For each scenario we calculated the probability that an embedded DTR was identified as the estimated optimal DTR, i.e. no interactions were included via variable selection. We also summarized the probability to identify a beneficial/harmful DTR across each scenario.

Simulation Characteristics				Variable Selection			
$t=1$	$t=2$	p_{cons}	p_{true}	None	$\lambda_{0.8}$-LASSO	S-Scores	Forward
m,nh	m,nh	1	1	0.0/0.0 (0.0)	80.4/79.2 (64.2)	96.9/79.4 (76.8)	86.4/72.8 (62.9)
		5	1	0.0/0.0 (0.0)	79.4/79.2 (62.9)	81.5/57.2 (46.4)	47.0/40.0 (18.8)
		5	5	0.0/0.0 (0.0)	79.5/80.2 (64.0)	81.9/58.3 (47.4)	46.4/39.4 (18.2)
		20	5	0.0/0.0 (0.0)	80.1/79.7 (64.0)	81.4/60.8 (49.8)	4.4/ 4.0 (0.1)
		20	20	0.0/0.0 (0.0)	80.0/80.4 (64.1)	80.4/60.4 (49.4)	3.8/ 4.4 (0.2)
	m,h	1	1	0.0/0.0 (0.0)	75.9/ 3.4 (2.7)	96.9/14.9 (14.2)	82.8/ 2.2 (1.8)
		5	1	0.0/0.0 (0.0)	79.6/12.9 (10.8)	83.2/14.4 (11.6)	45.6/ 1.5 (0.5)
		5	5	0.0/0.0 (0.0)	79.9/25.0 (20.6)	83.9/18.6 (15.3)	44.8/ 3.6 (1.7)
		20	5	0.0/0.0 (0.0)	80.2/46.2 (37.1)	80.0/28.7 (23.2)	4.7/ 0.5 (0.0)
		20	20	0.0/0.0 (0.0)	80.8/52.1 (42.1)	80.1/37.5 (29.7)	4.0/ 0.4 (0.0)
m,h	m,nh	1	1	0.0/0.0 (0.0)	0.1/80.4 (0.1)	10.3/82.1 (8.0)	0.1/73.6 (0.1)
		5	1	0.0/0.0 (0.0)	0.7/80.5 (0.6)	9.3/60.4 (5.0)	0.0/39.1 (0.0)
		5	5	0.0/0.0 (0.0)	4.9/80.8 (4.0)	20.2/58.6 (12.4)	0.4/38.5 (0.2)
		20	5	0.0/0.0 (0.0)	21.0/79.6 (17.4)	28.4/62.6 (17.7)	0.1/ 3.9 (0.0)
		20	20	0.0/0.0 (0.0)	35.0/81.1 (28.4)	58.8/58.9 (35.3)	0.0/ 3.6 (0.0)
	m,h	1	1	0.0/0.0 (0.0)	0.2/ 5.1 (0.0)	12.3/17.2 (1.5)	0.2/ 3.0 (0.0)
		5	1	0.0/0.0 (0.0)	1.0/12.7 (0.1)	11.1/16.0 (1.4)	0.0/ 1.3 (0.0)
		5	5	0.0/0.0 (0.0)	6.0/27.3 (1.1)	24.0/17.9 (3.7)	0.4/ 3.6 (0.0)
		20	5	0.0/0.0 (0.0)	23.0/50.2 (11.6)	28.0/31.5 (8.8)	0.0/ 0.5 (0.0)
		20	20	0.0/0.0 (0.0)	36.9/54.8 (21.3)	61.1/36.8 (21.6)	0.0/ 0.4 (0.0)

Table 3.2: Estimates for the probability of identifying an embedded DTR as the estimated optimal DTR when performing variable selection methods. Values are the probability of identifying an embedded DTR at $t = 1/t = 2$ and (both $t = 1$ and $t = 2$). At each time point the main we considered one of two scenarios: treatment main and heterogeneous effect (m,h) or treatment main non-heterogeneous effect (m,nh). Additionally we varied the number of true predictor interactions ($p_{true} = 1, 5, 20$) and the variables considered for interaction with treatment ($p_{cons} = 1, 5, 20$).

Table 3.2 shows the probability of identifying an embedded rule at each decision point, $t = 1, 2$ as well as the probability an embedded DTR is identified. For the results presented here we focused on four main scenarios with the full results presented in the supplemental material. In the scenarios where there was no heterogeneity with one or both treatment assignments, the $\lambda_{0.8}$-LASSO resulted in the highest likelihood of identifying an embedded DTR as the estimated optimal DTR across most values of (p_{cons}, p_{true}). Additionally, we see that at individual stages the $\lambda_{0.8}$-LASSO identifies

the embedded rule at that stage with the specified probability thus preserving the nominal rate of selecting a not more deeply tailored DTR. The overall probability of selecting an embedded DTR using the $\lambda_{0.8}$-LASSO in the absence of treatment effect heterogeneity is approximately 0.8^2 suggesting that identifying the embedded rule at one decision point is uncorrelated with identifying the embedded rule at other decision points. In the scenarios where there was heterogeneity at both decision points, both the $\lambda_{0.8}$-LASSO and S-Scores resulted in higher probabilities to identify an embedded DTR as the estimated optimal DTR compared to no variable selection or forward selection.

We additionally considered the probability to identify a beneficial/harmful DTR in table 3.3. In the scenario where there was no heterogeneity the probability for benefit was almost zero for every method of variable selection. This was expected since, without heterogeneity the true optimal embedded DTR is the optimal DTR and a beneficial rule is identified only when the estimated optimal embedded DTR is not the same as the true optimal embedded DTR, an unlikely occurrence. In the absence of treatment effect heterogeneity at either stage of treatment assignment the $\lambda_{0.8}$-LASSO resulted in the lowest probability of harm compared to the other variable selection methods. The $\lambda_{0.8}$-LASSO often resulted in lower probability of benefit when compared to other variable selection methods but this is due to it identifying an embedded DTR (neither benefit or harm) more often. When there was treatment effect heterogeneity at both treatment stages the $\lambda_{0.8}$-LASSO, S-Scores, and Forward selection resulted in similar probability of benefit or harm across multiple values of (p_{cons}, p_{true}).

Simulation Characteristics				Variable Selection			
$t=1$	$t=2$	p_{cons}	p_{true}	None	$\lambda_{0.8}$-LASSO	S-Scores	Forward
m,nh	m,nh	1	1	1.0/ 81.1	0.6/30.6	0.9/20.1	0.6/30.1
		5	1	1.2/ 98.8	0.7/35.8	1.1/52.0	1.0/78.0
		5	5	1.1/ 98.8	0.6/34.8	1.0/51.1	0.9/78.9
		20	5	0.7/ 99.3	0.6/35.6	1.0/49.2	1.0/98.8
		20	20	0.8/ 99.2	0.5/35.6	0.7/49.8	0.9/98.8
	m,h	1	1	94.1/ 5.6	92.4/ 4.2	83.1/ 2.6	93.7/ 3.8
		5	1	49.0/ 51.0	76.0/12.8	72.2/16.2	75.0/24.2
		5	5	46.6/ 53.4	30.2/47.8	40.1/44.6	47.4/50.6
		20	5	1.4/ 98.6	10.4/51.9	17.5/59.1	4.8/95.2
		20	20	2.0/ 98.0	4.0/53.7	7.6/62.7	4.0/96.0
m,h	m,nh	1	1	93.2/ 6.7	95.6/ 4.4	85.2/ 6.5	94.7/ 5.2
		5	1	73.5/ 26.5	95.5/ 3.9	82.6/12.3	88.6/11.4
		5	5	63.4/ 36.6	76.0/19.9	50.9/36.6	71.5/28.2
		20	5	1.8/ 98.2	45.0/37.7	34.2/48.0	14.8/85.2
		20	20	1.8/ 98.2	2.2/69.7	5.6/59.0	3.1/96.9
	m,h	1	1	100.0/ 0.0	99.4/ 0.6	97.9/ 0.5	99.6/ 0.4
		5	1	99.2/ 0.8	98.8/ 1.1	97.4/ 1.2	99.1/ 0.9
		5	5	98.2/ 1.8	77.9/20.8	73.9/22.2	93.5/ 6.5
		20	5	24.7/ 75.3	44.7/43.8	43.8/47.4	41.2/58.8
		20	20	33.9/ 66.1	2.8/76.1	10.5/67.7	10.0/90.0

Table 3.3: Estimates for the probability of identifying a beneficial DTR and the probability of identifying a harmful DTR, P_b/P_h, under various scenarios with varying DGM characteristics. At each time point the main we consider one of two scenarios: treatment main and heterogeneous effect (m,h) and treatment main non-heterogeneous effect (m,nh). Additionally we varied the number of true predictor interactions ($p_{true} = 1, 5, 20$) and the variables considered for interaction with treatment ($p_{cons} = 1, 5, 20$).

Increasing the number of parameters considered when performing variable selection resulted in a negative impact on the probability of benefit for all variable selection methods even when the number true interactions with treatment remained the same. For example, the probability of benefit when there was heterogeneity with both treatment assignments and $p_{true} = p_{cons} = 5$ is 77.9% when using the $\lambda_{0.8}$-LASSO

but drops to 44.7% when p_{cons} increased to 20 even though p_{true} remained equal to 5. This emphasizes that even when the variable selection method performed well or as expected, considering fewer variables as candidates for interaction to determine a more deeply tailored DTR resulted in higher probability of benefit.

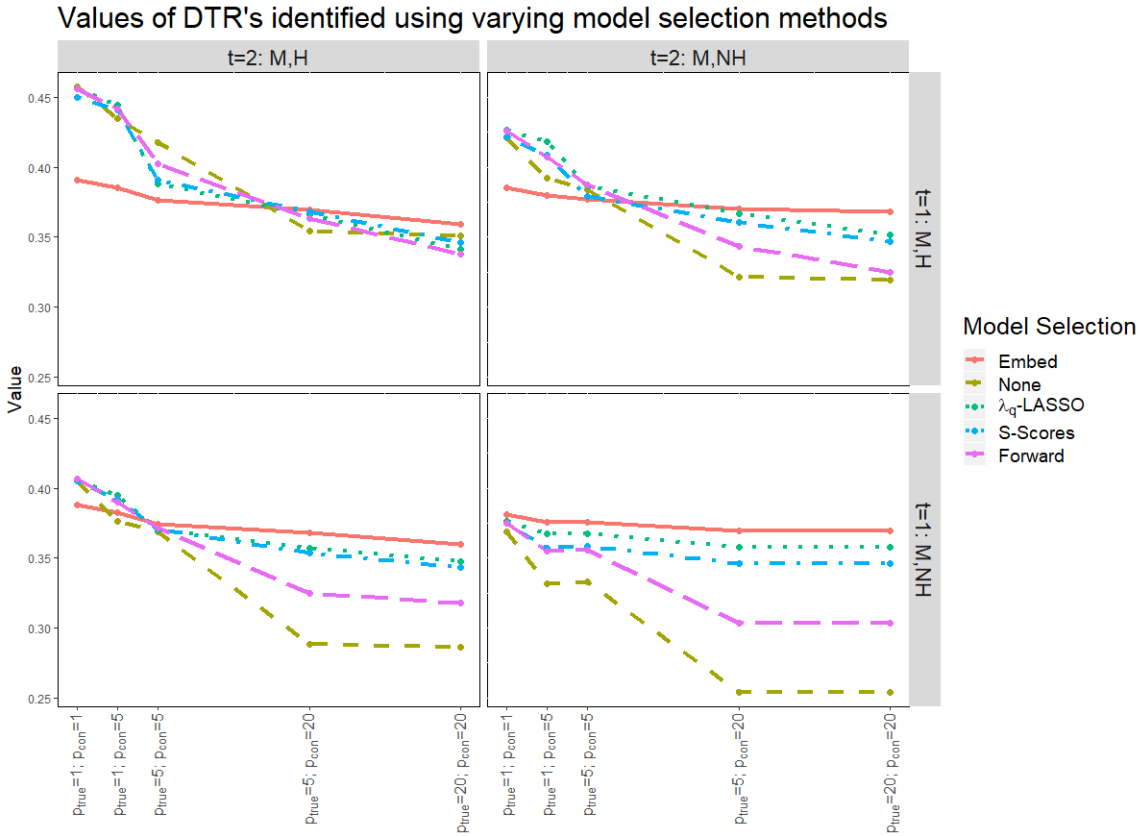

Figure 3.2: Values of estimated DTRs resulting from the embedded DTR, no variable selection, $\lambda_{0.8}$-LASSO, S-Scores, and Forward Selection. At each time point the main we consider one of two scenarios: treatment main and heterogeneous effect (m,h) and treatment main non-heterogeneous effect (m,nh). Additionally we varied the number of true predictor interactions ($p_{true} = 1, 5, 20$) and the variables considered for interaction with treatment ($p_{cons} = 1, 5, 20$).

Figure 3.2 shows the average value resulting from using each variable selection

method as well as from the estimated optimal embedded DTR. Both Forward Selection and no variable selection produced high values when there was heterogeneity with both treatment assignments and both p_{cons} and p_{true} are both low whereas, the λ_q-LASSO was the most consistent with values near embedded rules with no heterogeneity and still some values above the estimated optimal embedded DTR when heterogeneity existed and the number of predictors considered was low.

3.6 M-Bridge Study

The M-Bridge Study [26] was a two stage SMART aimed at preventing binge drinking among college freshman. One of the secondary aims look to identify modifiers of treatment effect at $t = 1$ or 2. We satisfied this aim by attempting to identify more deeply tailored DTRs using each variable selection method presented.

A total of 891 participants were enrolled with 300 randomized to an "assessment-only control" group leaving 591 participants randomized to receive one of two possible initial treatments for the DTRs considered, early ($A_1 = 1$) or late ($A_1 = -1$) intervention. A total of 158 participants were flagged as "Heavy Drinkers", i.e. non-responders, through self-monitoring and were re-randomized to one of two possible stage 2 treatment assignments, online health coaching ($A_2 = 1$) or resource email ($A_2 = -1$). Complete final outcome data on the number of instances of binge drinking in those assigned to receive an DTR was obtained in the first follow-up survey completed by $N = 500$ people, 142 non-responders and 358 responders. The study population was largely female (64%) and white (75.8%) with only 11.4% intending to pledge greek life at baseline.

Ten variables were set *a priori* to be tested as potential modifiers for treatment effect at both stage 1 or 2: sex, race/ethnicity, pre-college drinking norms (perceived percentage of students that drink/binge drink and perceived typical number/max number of drinks consumed by college students), and pre college intentions to drink (number of drinks per month, number of drinks in a typical sitting, number of times getting drunk per month), with two additional variables considered at stage 2 only: reported frequency of binge drinking and high intensity drinking in the most recent self-monitoring assessment. We multiplied the outcome, number of binge drinking instances, by -1 so that a higher value was better and centered and scaled all predictors at 0 before considering them in variable selection. We performed variable selection on models that consider these variables as possible interactions.

Since the simulation study demonstrated the benefit of considering fewer variables for a more deeply tailored DTR, we also considered a reduced set of variables to identify more deeply tailored DTRs. We considered a single propensity score [37] variable which represents baseline probability of binge drinking as a potential modifier at each stage of treatment calculated as the predicted probability of binge drinking even once based on the potential modifiers for both stages or a propensity score. To estimate the propensity score model, data from the assessment-only control group was used. We considered the variables collected at baseline and considered potential treatment effect modifiers such as sex, race, intention to pledge Greek, perceived drinking norms, and pre-college intentions to drink. Using a LASSO model with the penalty parameter selected using the "one standard error" rule we selected variables to be used in a generalized linear model with a logit link with binary outcome with 1=binge drinking

at first follow-up. Intention to pledge Greek, drinking norms: percent drink and percent binge, and drinking intentions: frequency/month and typical number of drinks were all selected as part of the propensity score model and positively associated with binge drinking.

	$t = 1$			$t = 2$		
Predictor	$\lambda_{0.8}$-LASSO	S-Scores	Forward	$\lambda_{0.8}$-LASSO	S-Scores	Forward
A2				-0.147	1.51	1.482
A1	0.008	0.008	0.008			0.208
Female						0.454
Non-white					0.908	0.885
Intention to pledge greek						-0.425
Drinking Norms: Percent drink						-0.012
Drinking Norms: Num drinks typical						-0.053
Drinking Norms: Most drinks typical					-0.104	-0.138
Drinking Norms: Percent binge						0.022
Drinking Intentions: Freq/month						0.167
Drinking Intentions: Num drinks						
Drinking Intentions: Drunk/month						
Self Monitoring: Binge drinking	-	-	-		-0.392	-0.503
Self Monitoring: High-intensity drinking	-	-	-			

Table 3.4: Treatment main effect and interaction coefficients for estimated DTRs identified using one of three variable selection methods

Table 3.4 shows the identified more deeply tailored DTRs discovered using $\lambda_{0.8}$-LASSO, S-Scores, and Forward selection. None of the variable selection methods resulted in a more deeply tailored rule at the first stage of treatment. At the second stage a more deeply tailored rule was identified by S-Scores and Forward selection although they were difficult to interpret. The $\lambda_{0.8}$-LASSO did not identify a more deeply tailored rule although Figure 3.3, which shows values of the interaction coefficients estimated for the LASSO model using λ, suggests that the most recent self-monitoring binge drinking might have some non-zero interaction but was just barely excluded when using $\lambda_{0.8}$.

	$t=1$			$t=2$		
Predictor	$\lambda_{0.8}$-LASSO	S-Scores	Forward	$\lambda_{0.8}$-LASSO	S-Scores	Forward
A2				0.534	0.524	0.272
A1	0.08	0.08	0.08		0.205	0.212
Propensity Score						0.554
Self Monitoring: Binge drinking	-	-	-	-0.278	-0.273	-0.245
Self Monitoring: High-intensity drinking	-	-	-			

Table 3.5: Treatment main effect and interaction coefficients for estimated DTRs identified using one of three variable selection methods where fewer variables are consider by combining several baseline covariates into a propensity score.

We also considered identifying a more deeply tailored rule where only the propensity score and self-monitoring variables were considered in table 3.5. The $\lambda_{0.8}$-LASSO resulted in a more deeply tailored DTR dependent only on the most recent self-monitoring binge drinking reported. Assigning treatment based on the identified treatment rule would assign 5% of this population of non-responders to an online health coach and the rest to a resource email. According to this rule a resource email is more beneficial for those with higher reported binge drinking. More specifically, non-responders with ≥ 2 binge drinking instances should be assigned to a resource email and non-responders with < 2 binge drinking instances should be assigned to online health coach.

To consider a more continuous outcome we considered an additional analysis with an alternative outcome: Max number of drinks in a 24 hour period in the past 30 days. We first performed variable selection on a model where all predictor-treatment interactions were considered as well as a model where the baseline variables were combined into a propensity score estimating the probability to binge drink. When all predictor-treatment interactions were considered, a more deeply tailored DTR was identified where those whose baseline intention to drink: number of drinks in a sitting was less than 3 would benefit from the online health coaching and those

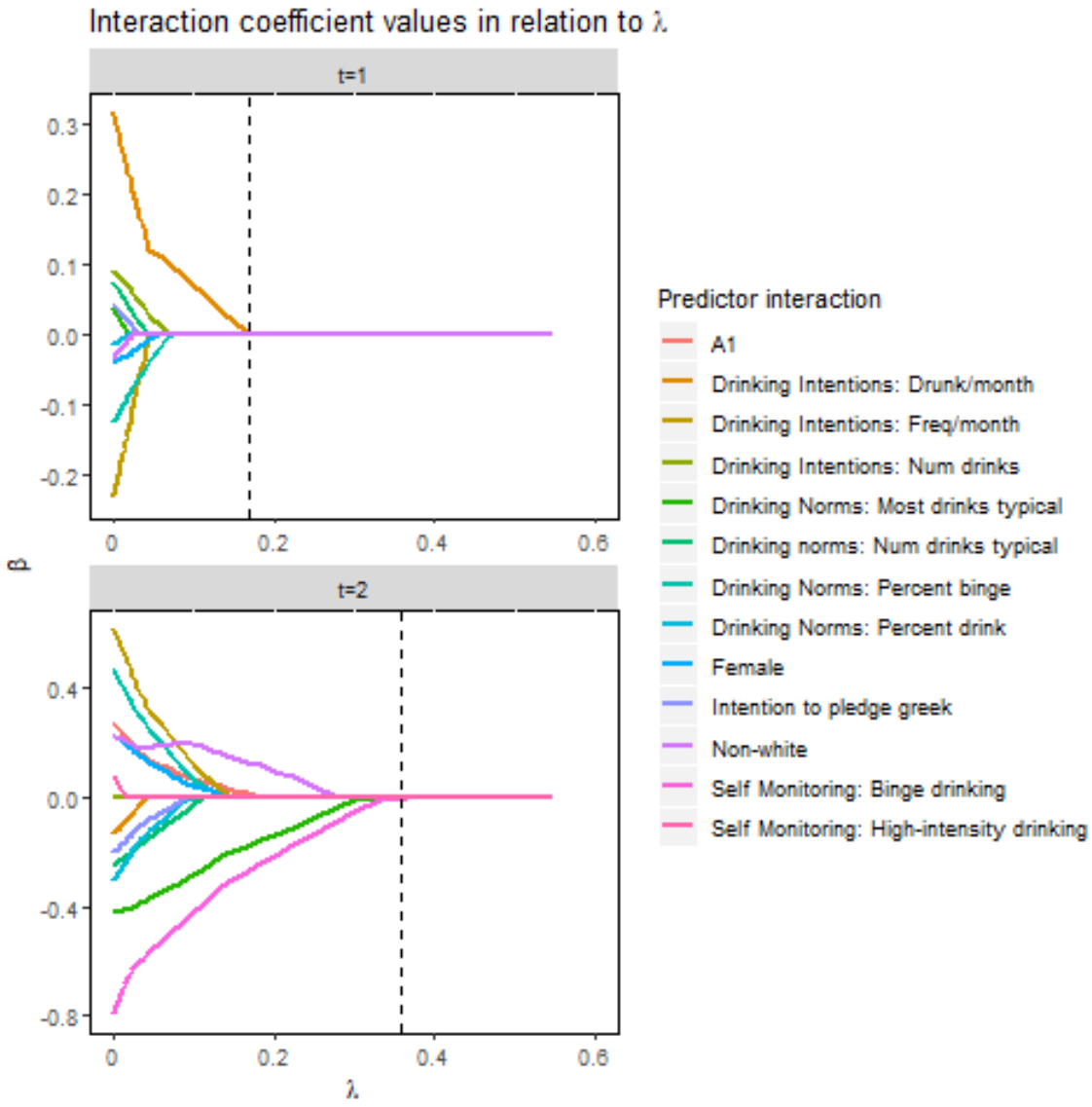

Figure 3.3: Interaction coefficient values from the LASSO model estimated using λ. The dashed line represents the value of $\lambda_{0.8}$

who intended to drink 3 or more drinks per drinking period would benefit more from the resource email. By reducing the number of predictors tested to be the baseline propensity score for binge drinking along with the self-monitoring predictors

	$t=1$			$t=2$		
Predictor	$\lambda_{0.8}$-LASSO	S-Scores	Forward	$\lambda_{0.8}$-LASSO	S-Scores	Forward
A2				1.594	1.594	2.36
A1	0.053	0.053	0.053			0.285
Female						
Non-white						-1.775
Intention to pledge greek						-0.899
Drinking Norms: Percent drink						-0.014
Drinking Norms: Num drinks typical						
Drinking Norms: Most drinks typical						
Drinking Norms: Percent binge						0.019
Drinking Intentions: Freq/month						0.128
Drinking Intentions: Num drinks				-0.545	-0.545	-0.699
Drinking Intentions: Drunk/month						-0.092
Self Monitoring: Binge drinking	-	-	-			
Self Monitoring: High-intensity drinking	-	-	-			0.079
Baseline variables combined into propensity score						
A2				1.53	1.505	1.467
A1	0.088	-0.034	0.088	0.581		0.552
Propensity Score	-0.175		-0.175	-2.938	-2.927	-2.694
Self Monitoring: Binge drinking	-	-	-			
Self Monitoring: High-intensity drinking	-	-	-			-0.173

Table 3.6: Treatment main effect and interaction coefficients for estimated DTRs identified using one of three variable selection methods using an alternative outcome: Max drinks in 24 hour period in past 30 days. All predictor-treatment interactions as well as propensity score-treatment interaction were considered.

for the second stage we identify a slightly different estimated more deeply tailored DTR. The $\lambda_{0.8}$-LASSO identifies a rule to minimize the max drinks in a 24 hour period that assigns early initial treatment to those whose baseline propensity to binge drink is less than 0.503 and assigns late initial treatment to those whose baseline propensity to binge drink is greater than or equal 0.503. At stage 2 heavy-drinkers who received early initial treatment would benefit from the resource email if their baseline propensity to binge drink was greater than 0.719 where as heavy drinkers who received late initial treatment would benefit form the resource email if their baseline propensity to binge drink was greater than 0.347. S-Scores identified a more deeply tailored DTR that assigns treatment based on propensity score at baseline but does not have differing cut-off points at stage 2 based on initial treatment assignment.

3.7 Discussion

We define a beneficial/harmful DTR as a DTR that results in a higher/lower value than the optimal estimated embedded DTR. To identify non-harmful DTRs we introduced the use of a permutation-based method to determine the penalty parameter of a LASSO model such that when no treatment effect heterogeneity exists, the optimal embedded DTR is identified. Our simulation study showed that the λ_q-LASSO identified the estimated optimal embedded DTR with specified probability in the absence of treatment effect heterogeneity and avoided harm more often than other variable selection methods. We also observed that the probability of benefit was lower and the probability of harm was higher when more variables were considered even if the number of variables with a true heterogeneous effect with treatment remained the same.

Using data from the M-Bridge study we demonstrated the usefulness of the λ_q-LASSO by identifying an easily interpretable more deeply tailored DTR. The rule suggests that among non-responders to the initial treatment, people who reported more binge drinking would benefit more from the resource email than online health coaching. Much like the simulations, rules were easier to identify when less variables were considered.

When personalization of treatment by way of more deeply tailored DTRs is considered, one should always consider not only the potential benefit in the population but also the potential harm. Estimated more deeply tailored DTRs can often result in a lower average value than the easier to identify optimal embedded DTR especially in instances where there is no treatment effect heterogeneity or the number of predictors considered is large. Therefore, variable selection methods that identify the

estimated optimal embedded DTR in the absence of treatment effect heterogeneity should always be considered to avoid causing harm in the population.

Chapter 4

Variable Selection to Identify a Non-Harmful DTR when using IQ-learning

4.1 Introduction

For many chronic conditions including smoking, substance use disorder, depression, obesity, ADHD, autism, or schizophrenia, patient response to available treatments is heterogeneous, both between patients and within a single patient over time. Frequently, clinicians make an initial treatment decision, monitor for side effects, non-adherence, and changes in symptoms, and then make adjustments to treatment based on a patient's changing course. A dynamic treatment regime (DTR) formalizes this process; a DTR is a set of decision rules, one for each decision point, which recommends a treatment, intervention, or action to take based on patient characteristics,

including prior response to treatment. A DTR embedded in a Sequential Multiple Assignment Randomized Trial (SMART) assigns treatment sequentially often based on response but can also assign treatment based on participant characteristics in the case of a more deeply tailored DTR. In chapter 3, beneficial/harmful more deeply tailored DTRs were defined as DTRs that result in a higher/lower value than the estimated optimal embedded DTR. A harmful DTR most likely occurs in the absence of treatment effect heterogeneity so methods were introduced to identify a non-harmful DTR with user specified probability in this scenario.

A SMART can be used to assess the efficacy of DTRs by randomizing participants at multiple time points often based on response to previous treatment. We are motivated by the Program for LUng Cancer Screening and TObacco Cessation (PLUTO) SMART. PLUTO is aimed at identifying DTRs to assist with smoking cessation among those eligible for lung cancer screening. The primary outcome of interest is long-term abstinence from smoking with secondary outcomes such as change in number of cigarettes smoked. One secondary aim focuses on identifying more deeply tailored DTRs to guide treatment to improve long-term abstinence from smoking or reduce number of cigarettes smoked.

Identification of an optimal DTR is considered in most SMARTs, including the PLUTO study, and a variety of methods have been developed to accomplish this goal. One such method is Q-learning [28], a regression-based backwards induction method. In Q-learning the optimal treatment rule at the final stage of treatment assignment is identified by fitting a regression model for the outcome given treatment at the last stage and prior covariate and treatment history. The optimal rule at the preceding timepoint is determined by fitting a regression model with covariates and treatment

from that timpoint and the expected response assuming that the optimal treatment assignment is followed at the final stage as the outcome. This process is repeated for each stage of treatment assignment in which the optimal treatment assignment is determined assuming the optimal is assigned at future stages. In practice, the assumed regression models are typically linear models (i.e., linear in the covariates and treatment history). When there is treatment effect heterogeneity at a given stage of treatment assignment, the predicted outcome assuming the optimal treatment is used thereafter will have a non-linear and non-smooth relationship with the predictors used at the preceding stage under most plausible data generating mechanisms. Thus, linear models at the preceding stages of treatment assignment are usually incorrectly specified. As a solution to this problem, Interactive Q-learning or IQ-learning [39] was introduced. In the context of a two-stage treatment decision, the regression model at the final stage proceeds as in standard Q-learning. However, to learn the optimal decision at stage one, two separate models are fit: the contrast and the main effect at the second stage are each the outcomes and baseline covariates are the predictors. Under many plausible data generating scenarios, the main effect and contrast at the second stage can be linearly associated with baseline covariates. In particular, both of these models are used to derive the estimated optimal treatment strategy at the first stage.

IQ-learning uses conditional mean and variance modeling to identify easy to interpret treatment rules while overcoming the non-linear and non-smooth relationship between the predicted outcome in non-terminal models and covariates. However, there is scant literature on how to incorporate variable selection into IQ-learning. Variable selection is needed to improve model performance and obtain treatment

regimes which are more parsimonious. However, care must be taken as there are two separate models at the first stage which contribute to the estimated optimal treatment rule. In particular, if a baseline covariate by first-stage treatment interaction is selected in either model, then the covariate will be included in the estimated optimal decision rule (at the first stage). This study presents a group LASSO method for variable selection for both models considered at non-terminal stages in IQ-learning where the penalty parameter is selected through a permutation-based method similar to the methods presented in Chapter 3.

In section 4.2 we introduce both Q-learning and IQ-learning and their applications to SMARTs. Then we introduce our group LASSO method of variable selection to be used with IQ-learning in section 4.3. We compare our variable selection method for IQ-learning against similar variable selection methods with different forms of Q-learning such as modified Q-learning [34] using simulation in section 4.4. In section 4.5, we apply these methods using Q-learning, Modified Q-learning, and IQ-learning to the Program for Lung Cancer Screening and Tobacco Cessation (PLUTO) trial [1]. We close the chapter in section 4.6.

4.2 Q-learning and IQ-Learning

We consider SMARTs with two decision points $t = 1, 2$ with final outcome Y, treatment assignments $A_t \in \{-1, 1\}$, and p_t dimensional vectors of features $X_t \in \mathcal{X}_t$. We denote the cumulative information available when A_t is assigned as $H_t,\ t = 1, 2$ i.e. $H_2 = (X_1, A_1, X_2)$ and $H_1 = X_1$. Although we focus on two-stage SMARTs, as in the PLUTO study, the proposed methods easily generalize to multiple stages.

We define a DTR, d, as a series of decision rules at each time point, d_t, which map the domain of H_t to the set of treatments, $d_t : \mathcal{H}_t \to \{-1, 1\}$. We define the value of d as the expected potential outcome, $V(d) = E\{Y(d)\}$, where the potential outcome [2] is the outcome, assuming d is followed, denoted as $Y(d)$. Many statistical methods for identifying DTRs involve estimating the optimal DTR, d^{opt}, i.e. $V(d^{opt}) \geq V(d)$ for all DTRs d. Doing so often results in a more deeply tailored DTR that depends on multiple patient characteristics including response to treatment. These more deeply tailored DTRs can be time and cost intensive for both clinicians and patients whereas the optimal embedded DTR may perform just as well or even better. An embedded DTR, w, is a DTR within a SMART that depends on only on the embedded tailoring variables. The estimated optimal embedded DTR $\widehat{w}^{opt}$, is the embedded DTR that results in the highest estimated value out of all embedded DTRs. When considering a potentially complex more deeply tailored DTRs we believe that it should perform as well or better than the simpler estimated optimal embedded DTR. We define a beneficial DTR, d such that

$$V(d) > V(\widehat{w}^{opt}). \tag{4.1}$$

Conversely, a harmful DTR is defined such that $V(d) < V(\widehat{w}^{opt})$, so a non-harmful DTR replaces the strict inequality in equation (4.1) with a non-strict inequality.

4.2.1 Q-Learning

Q-learning [28] is a regression-based algorithm to identify DTRs. The algorithm relies on the Q-functions,

$$Q_2(h_2, a_2) = E(Y|H_2 = h_2, A_2 = a_2), \tag{4.2}$$

$$Q_1(h_1, a_1) = E\{\max_{a_2} Q_2(h_2, a_2)|H_1 = h_1, A_1 = a_1\}. \tag{4.3}$$

The optimal treatment rules are determined by maximizing each Q-function, $d_t^{opt}(h_t) = \text{argmax}_{a_t} Q_t(h_t, a_t)$. The Q-functions, $Q_t, t = 1, 2$ can be expressed with real valued functions, $b_t, c_t : H_t \to \mathbb{R}, t = 1, 2$, which may depend on a set of finite dimensional parameters α_t and β_t.

$$Q_2(h_2, a_2) = b_2(h_2; \alpha_2) + c_2(h_2; \beta_2)a_2 \tag{4.4}$$

$$Q_1(h_1, a_1) = b_1(h_1; \alpha_1) + c_1(h_1; \beta_1)a_1. \tag{4.5}$$

To determine the optimal DTR we first estimate α_2 and β_2 in Q_2 using standard regression-based methods with Y as the dependent variable. Then, we predict the outcome $\tilde{Y}$ assuming the optimal treatment rule is used at the second stage,

$$\tilde{Y}_i = \max_{a_{2i}} Q_2(H_{2i}, a_{2i}; \widehat{\alpha}_2, \widehat{\beta}_2) = b_2(h_2; \widehat{\alpha}_2) + |c_2(h_2; \widehat{\beta}_2)|.$$

Assuming the b_1 and c_1 are linear in (some possibly transformed) H_1, we regress $\tilde{Y}$ on H_1 and A_1 for all participants in order to estimate α_1 and β_1 in equation (4.5). The estimated optimal treatment rules taken from Q-learning are defined as

$$\widehat{d}_t^{opt}(H_t) = \text{argmax}_{a_t} Q_t(h_t, a_t; \widehat{\alpha}_t, \widehat{\beta}_t) = sign\{c_t(H_t; \widehat{\beta}_t)\} \tag{4.6}$$

Modified Q-learning [34] is identical to Q-learning except for the estimation of $\tilde{Y}$. In modified Q-learning $\tilde{Y}$ is defined as the outcome plus the "regret" [24] from not assigning the optimal treatment assignment,

$$\tilde{Y}_i = Y_i + [Q_2(H_{2i}, \widehat{d}_{2i}(h_{2i}); \widehat{\alpha}_2, \widehat{\beta}_2) - Q_2(H_{2i}, a_{2i}; \widehat{\alpha}_2, \widehat{\beta}_2)]. \tag{4.7}$$

Note that when the estimated optimal treatment is assigned to participant i, i.e. $a_{2i} = \widehat{d}_{2i}(h_{2i})$, the regret is equal to 0 so $\tilde{Y} = Y$.

The value of a DTR, d, identified through Q-learning is written as,

$$V(d) = \iint b_2\{(x_1, d_1(h_1), x_2)\} + c_2\{(x_1, d_1(h_1), x_2)\} d_2(h_s) dF_{X_2|X_1, d_1(H_1)}(x_2) dF_{X_1}(x_1).$$

4.2.2 IQ-learning

One criticism of Q-learning is that $\tilde{Y}$ is often predicted using a model with interactions between A_2 and predictors correlated with H_1. This leads to an inherently non-linear and non-smooth relationship between H_1 and $\tilde{Y}$. Yet, Q_1 is frequently estimated with a linear model. IQ-learning [39] instead models Q_1 using the main effect, $\mu(H_2)$, and contrast, $\Delta(H_2)$, functions of Q_2,

$$\widehat{\mu}(H_2) = \frac{1}{2}\{\widehat{Q}_2(H_2, 1) + \widehat{Q}_2(H_2, -1)\},\ \widehat{\Delta}(H_2) = \frac{1}{2}\{\widehat{Q}_2(H_2, 1) - \widehat{Q}_2(H_2, -1)\}.$$

IQ-Learning estimates Q_2 the same as in Q-Learning and then estimates Q_1 using

$$Q_1 = E\{\mu(H_2)|H_1 = h_1, A_1 = a_1\} + \int |z| g_{h_1, a_1}(z) dz, \tag{4.8}$$

where g_{h_1,a_1} denotes the conditional distribution of $\Delta(H_2)$ given $H_1 = h_1$ and $A_1 = a_1$. The estimate of Q_1 has the form

$$\widehat{Q}_1(h_1, a_1) = \widehat{L}(h_1, a_1) + \int |z| \widehat{g}_{h_1,a_1}(z) dz,$$

where $\widehat{L}(h_1, a_1)$ is the estimated outcome (i.e., estimated conditional mean) from the linear regression of $\widehat{\mu}(H_2)$ on H_1 and A_1 and $\widehat{g}_{h_1,a_1}$ is obtained by way of a location-scale model described in Laber et. al. [39] where the mean, $\widehat{m}(h_1, a_1)$ is estimated by regressing $\widehat{\Delta}(H_2)$ on H_1 and A_1. The variance of the location-scale model can also be estimated through regression on H_1 but here we consider only constant variance for the location-scale model. Then, $\int |z| \widehat{g}_{h_1,a_1}$ is estimated empirically and the estimated optimal treatment rule at stage t is still

$$\widehat{d}_t^{opt}(H_t) = \text{argmax}_{a_t} \widehat{Q}_t(h_t, a_t).$$

In particular, assuming that the conditional mean for $\mu(H_2)$ and $\Delta(H_2)$ given H_1 and A_1 is linear in the covariates, the estimated optimal treatment rule at stage 1 will depend on a particular covariate if that covariate interacts with stage 1 treatment in either the model for $\mu(H_2)$ or $\Delta(H_2)$.

4.3 λ_q-GLASSO

IQ-learning accounts for the inherent non-linear and non-smooth relationship between H_1 and the predicted value assuming a heterogeneous rule at $t = 2$ by avoiding directly modeling Q_1. However, because two separate models are fit to estimate Q_1,

special care must be taken in variable selection. In particular, a covariate by stage 1 treatment interaction must be selected out of both the models for $\mu(H_2)$ and $\Delta(H_2)$ for that covariate to not be a part of the estimated optimal rule at stage 1. Excluding covariates from the stage 1 treatment regime is important not only for parsimony but including irrelevant covariates in the decision rule can degrade performance. In particular, in the case in which there is no heterogeneity at $t = 1$, any heterogeneous rule will result in a value of that rule which is less than the non-heterogeneous one. However, no current variable selection methods for IQ-learning can identify a non-heterogeneous rule at $t = 1$ in the absence of heterogeneity with specified a priori probability. The λ_q-LASSO presented in chapter 3 selects the embedded DTR with user specified probability in two stage SMART studies and is appropriate to use at $t = 2$ in IQ-learning since the algorithm is the same as in both Q-learning and Modified Q-learning. At $t = 1$, however, λ_q-LASSO as defined in chapter 3 does not handle the multiple models used to estimate Q_1. To perform variable selection across multiple models we consider a group LASSO [40, 41] model regressing both $\mu(H_2)$ and $\Delta(H_2)$ on H_1.

The group LASSO model extends the LASSO model [14] by way of shrinking groups of coefficients (some shrunk to zero) rather than single coefficients done in the LASSO model. For a general regression problem, given grouped matrices of characteristics, $X_1, \ldots, X_K$ with corresponding vectors of coefficients $\beta_1, \ldots, \beta_K$ for the linear model regressing Y on $(X_1', \ldots, X_K')'$, the group LASSO model under penalty parameter, λ, is defined as minimizing

$$\sum_{i=1}^{n} \left(Y_i - \sum_{j=1}^{K} X_{ij}' \beta_j \right)^2 + \lambda \sum_{j=1}^{K} \|\beta_j\|_{\zeta_j},$$

where $\|\eta\|_\zeta = (\eta'\zeta\eta)^{1/2}$. The matrix, ζ_j serves as a weight matrix for the coefficients included in grouping j. For the context of this study we consider only two cases for ζ_j where it is either the identity matrix when the penalty parameter is applied to the j^{th} group of coefficients or a matrix of zeros when the penalty is not applied. Group LASSO models shrink entire groups of parameters to zero to allow for variable selection of coefficients on a group level by selecting only the groups that are non-zero to then be used in a separate regression model.

Traditionally, when using a group LASSO model for variable selection, the purpose is to perform variable selection on grouped variables within a single model. For variable selection on Q_1 of IQ-learning we aim to perform variable selection on variables across two models. This is accomplished by combining the main effect and contrast outcomes as well as each covariate coefficient into 2 dimensional vectors and creating diagonal block matrices for each predictor with the predictor repeated on the diagonal to represent groupings. Specifically, to apply the group LASSO with models, $L(h_1, a_1)$ and $m(h_1, a_1)$ from IQ-learning we write

$$L(h_1, a_1; \gamma) = \gamma_0 + h_1'\gamma_1 + a_1(\gamma_2 + h_1'\gamma_3), \quad m(h_1, a_1; \theta) = \theta_0 + h_1'\theta_1 + a_1(\theta_2 + h_1'\theta_3),$$

where $\gamma_1, \gamma_3, \theta_1$, and θ_3 are all p dimensional vectors and p is the number of predictors in h_1 and $\boldsymbol{\gamma}$ and $\boldsymbol{\theta}$ represent the $2p + 2$ dimensional vectors of model coefficients for L and m respectively. Let the identical $2p + 2$ dimensional vector of predictors and treatment-predictor interactions for the ith individual in both models be written as, $\boldsymbol{H_i}$. The model parameters are grouped pairwise as $\tau_j = \{\boldsymbol{\gamma}_j, \boldsymbol{\theta}_j\}$, $j = 0, \ldots 2p + 1$, which reflect the coefficient of a single model predictor, e.g. $\tau_0 = \{\gamma_0, \theta_0\}$, $\tau_1 =$

$\{\gamma_{11}, \theta_{11}\}$, etc. For individual i, we define a model outcome $(\widehat{\mu}(h_{2i}), \widehat{\Delta}(h_{2i}))'$ and 2×2 matrices

$$\boldsymbol{H}_{ij}^{*} = \begin{pmatrix} \boldsymbol{H}_{ij} & 0 \\ 0 & \boldsymbol{H}_{ij} \end{pmatrix},$$

where $\boldsymbol{H}_{ij}$ denotes the $(j+1)^{th}$ element from $\boldsymbol{H}_{i}$. So the Group LASSO estimates of grouped coefficients, β_j are defined as

$$\text{argmin}_{\tau} \sum_{i=1}^{n} \left[\left(\widehat{\mu}(h_{2i}) - \sum_{j=0}^{2p+1} \boldsymbol{H}_{ij} \boldsymbol{\gamma}_j \right)^2 + \left(\widehat{\Delta}(h_{2i}) - \sum_{j=0}^{2p+1} \boldsymbol{H}_{ij} \boldsymbol{\theta}_j \right)^2 \right] + \lambda \sum_{j=0}^{2p+1} \|\tau_j\|_{\zeta_j}, \tag{4.9}$$

where ζ_j is a 2×2 identity matrix for all parameters except for those for the main effect of treatments of treatment, $(\gamma_2, \theta_2)'$, or the model intercepts where ζ_j is a 2×2 matrix of zeros.

Using a group LASSO model we are able to perform variable selection on both $L(h_1, a_1)$ and $m(h_1, a_1)$ simultaneously by simultaneously selecting parameters to include in the model. In the absence of treatment effect heterogeneity at $t = 1$, the optimal rule is to follow the embedded DTR so we want to select our model such that no interaction terms are selected for either model. To control this we select the penalty parameter λ through a permutation-based method similar to that of the λ_q-LASSO seen in chapter 3.

We define the λ_q-GLASSO with data $\{H_1, A_1, \widehat{\mu}(H_2), \widehat{\Delta}(H_2)\}$ and a set of penalty parameters $\boldsymbol{\Lambda}$. We create two psuedo outcomes Z_μ and Z_Δ defined as,

$$Z_\mu = \begin{cases} \widehat{\mu}(H_2) - \delta_\mu, & \text{if } A_1 = 1, \\ \widehat{\mu}(H_2), & \text{if } A_1 = -1, \end{cases} \qquad Z_\Delta = \begin{cases} \widehat{\Delta}(H_2) - \delta_\Delta, & \text{if } A_1 = 1, \\ \widehat{\Delta}(H_2), & \text{if } A_1 = -1, \end{cases}$$

where δ_μ and δ_Δ are the estimated main effect of treatment on $\widehat{\mu}(H_2)$ and $\widehat{\Delta}(H_2)$ respectively. For each iteration b, we permute treatment assignment A_1 to create $\tilde{D}^{(b)} = (H_1, \tilde{A_1}^{(b)}, Z_\mu, Z_\Delta)$. Then, for each $\lambda \in \Lambda$, we fit the Group LASSO model in equation (4.9) regressing both Z_μ and Z_Δ on H_1 and $A_1^{(b)}$ using $\tilde{D}^{(b)}$ with penalty parameter λ. We evaluate the presence of interaction terms in either model under λ,

$$I_\lambda^{(b)} = \mathbb{1}(\widehat{\gamma}_3^{\lambda(b)} = \mathbf{0} \wedge \widehat{\theta}_3^{\lambda(b)} = \mathbf{0})$$

where $\mathbf{0}$ represents a vector of zeros and $\widehat{\gamma}_3^{\lambda(b)}$ and $\widehat{\theta}_3^{\lambda(b)}$ are the predictor-treatment interaction terms in the group lasso model under λ for iteration b. The proportion of the group LASSO models that result in no interactions is estimated by repeating B (typically chosen to be 1000) times and calculating

$$\hat{P}(I_\lambda = 1) = \frac{1}{B}\sum_{b=1}^{B} I_\lambda^{(b)}.$$

We select λ_q, as the smallest value in $\mathbf{\Lambda}$ such that $\hat{P}(I_\lambda = 1) > q$. Using λ_q for the penalty parameter for variable selection via LASSO selects a non-heterogeneous rule at $t = 1$ in the absence of treatment effect heterogeneity $q \times 100\%$ of the time.

When there is estimated treatment effect heterogeneity at $t = 2$, $\tilde{Y}$ will have a non-linear and non-smooth relationship with H_1 and IQ-learning solves this issue. Without treatment effect heterogeneity at $t = 2$, however, a variable selection method such as λ_q-LASSO can be applied and a non-heterogeneous treatment rule will likely be identified. In this case $\tilde{Y}$ will not have a non-linear relationship with H_1 thus eliminating the benefit of IQ-learning. In fact, when a non-heterogeneous rule is identified at $t = 2$, $\widehat{\Delta}(H_2)$ will not be continuous and could even be constant in some

scenarios. Since, we want to apply variable selection at both stages of treatment the steps to estimate a DTR using these methods are,

1. Perform variable selection using the λ_q-LASSO and estimate b_2 and c_2 to evaluate d_2.

If a heterogeneous rule is identified, i.e. $sign\{\widehat{c}_2(h_{2i})\} \neq sign\{\widehat{c}_2(h_{2j})\}$ for some i, j then,

2. Predict $\widehat{\mu}(H_2)$ and $\widehat{\Delta}(H_2)$ using $\widehat{Q}_2$,

3. Perform variable selection on $L(h_1, a_1)$ and $m(h_1, a_1)$ using λ_q-GLASSO,

4. Continue with IQ-learning with the selected models to evaluate d_1.

If a heterogeneous rule is not identified, i.e. $sign\{\widehat{c}_2(h_i)\} = sign\{\widehat{c}_2(h_j)\}$ for all i, j then,

2. Predict $\tilde{Y}$ using Q_2 according to modified Q-learning,

3. Perform variable selection on Q_1 using λ_q-LASSO at $t = 1$ described in chapter 3

4. Continue with Q-learning with the selected model to evaluate d_1

4.4 Simulation Study

We considered simulations of a SMART where all participants are re-randomized regardless of response to initial treatment. The outcome, Y, is continuous with baseline predictors, $X_1 \sim N_p(0, \Sigma_{AR(1)}(0.5))$, where $\Sigma_{AR(1)}(0.5)$ denoted a covariance matrix

with AR(1) correlation structure with $\rho = 0.5$ and both A_1 and A_2 are independently assigned 1:1 to -1 or 1. Treatment assignment was assigned at 1:1 randomization, for $t = 1, 2$. We assumed all predictors were remeasured at $t = 2$, thus $X_2 = 1.5X_1 + \xi$, where $\xi \sim N(0, I_p)$ and I_p was an independent correlation matrix of size p. We simulated the outcome Y such that

$$E[Y|X_2, A_1, A_2] = X_2'\beta_1 + A_1(\beta_2 + X_2'\beta_3) + A_2(\beta_4 + A_1\beta_5 + X_2'\beta_6)$$

and $Var(Y|X_2, A_1, A_2) = \sigma^2$ was selected such that the R^2 of the second stage model was fixed at 0.6.

We consider two scenarios of treatment effect heterogeneity or no treatment effect heterogeneity at $t = 1, 2$ resulting in 4 total scenarios denoted as "nh" or "h" to denote non-heterogeneity and heterogeneity respectively at each stage. We vary the number of predictors considered, p_{cons} and the number of predictors that have a true heterogeneous treatment effect, p_{true}. The main effect of the predictors was set as, $\beta_1 = \mathbb{1}_{p_{cons}}/p_{cons}$ where $\mathbb{1}_p$ is a vector of ones with length p, and the main effect coefficient of each treatment was set to be $\beta_2 = \beta_4 = 0.15$ for A_1 and A_2 respectively. When there was heterogeneity at a given stage, the interaction coefficients were selected such that the proportion of the population that would benefit from the heterogeneous rule was 0.2, i.e. $P(\beta_2 + X_2'\beta_3 + A_2\beta_5 < 0) = 0.2$ and $P(\beta_4 + A_1\beta_5 + X_2'\beta_6 < 0) = 0.2$ for $t = 1$ or $t = 2$ respectively where β_5 was set to be 0.05 when there was treatment effect heterogeneity at both $t = 1$ and $t = 2$.

Each simulation was created $M = 1000$ times for each scenario. The probability of identifying the estimated optimal DTR was evaluated as well as both the probability

of benefit and probability of harm. We compare Q-learning, Modified Q-learning, and IQ-learning with no variable selection to using the λ_q-LASSO with both Q-learning and Modified Q-learning or λ_q-GLASSO with IQ-learning. The Q-functions prior to variable selection were defined as,

$$Q_2(x_2, a_1, a_2) = \beta_{20} + x_2'\beta_{21} + a_1(\beta_{22} + x_2'\beta_{23}) + a_2(\beta_{24} + a_1\beta_{25} + x_2'\beta_{26}),$$

$$Q_1(x_1, a_2) = \beta_{10} + x_1'\beta_{11} + a_1(\beta_{12} + x_1'\beta_{13})$$

Simulation Characteristics				variable selection					
$t=1$	$t=2$	p_{cons}	p_{true}	None (Q)	None (Mod-Q)	None (IQ)	$\lambda_{0.8}$-LASSO (Q)	$\lambda_{0.8}$-LASSO (Mod-Q)	$\lambda_{0.8}$-GLASSO (IQ)
m,nh	m,nh	1.00	1.00	0.0/0.0 (0.0)	0.0/0.0 (0.0)	0.0/0.0 (0.0)	72.9/80.2 (58.2)	79.5/80.2 (63.8)	78.7/80.2 (64.1)
		5.00	1.00	0.0/0.0 (0.0)	0.0/0.0 (0.0)	0.0/0.0 (0.0)	66.7/79.8 (53.3)	77.3/79.8 (61.8)	74.3/79.8 (62.0)
		5.00	5.00	0.0/0.0 (0.0)	0.0/0.0 (0.0)	0.0/0.0 (0.0)	65.9/80.3 (52.8)	76.9/80.3 (61.3)	74.2/80.3 (61.3)
		20.00	5.00	0.0/0.0 (0.0)	0.0/0.0 (0.0)	0.0/0.0 (0.0)	66.6/80.8 (52.8)	79.2/80.8 (63.3)	77.2/80.8 (63.5)
		20.00	20.00	0.0/0.0 (0.0)	0.0/0.0 (0.0)	0.0/0.0 (0.0)	66.7/80.2 (53.9)	80.2/80.2 (63.5)	76.6/80.2 (63.2)
	m,h	1.00	1.00	0.0/0.0 (0.0)	0.0/0.0 (0.0)	0.0/0.0 (0.0)	69.8/ 2.9 (2.4)	79.7/ 2.9 (2.3)	72.8/ 2.9 (2.3)
		5.00	1.00	0.0/0.0 (0.0)	0.0/0.0 (0.0)	0.0/0.0 (0.0)	68.3/14.1 (9.9)	80.7/14.1 (11.8)	69.8/14.1 (11.8)
		5.00	5.00	0.0/0.0 (0.0)	0.0/0.0 (0.0)	0.0/0.0 (0.0)	68.2/10.7 (7.1)	80.4/10.7 (8.4)	70.4/10.7 (8.4)
		20.00	5.00	0.0/0.0 (0.0)	0.0/0.0 (0.0)	0.0/0.0 (0.0)	68.4/ 4.3 (3.3)	79.5/ 4.3 (3.5)	68.7/ 4.3 (3.4)
		20.00	20.00	0.0/0.0 (0.0)	0.0/0.0 (0.0)	0.0/0.0 (0.0)	69.1/ 5.9 (4.2)	81.2/ 5.9 (4.4)	69.9/ 5.9 (4.5)
m,h	m,nh	1.00	1.00	0.0/0.0 (0.0)	0.0/0.0 (0.0)	0.0/0.0 (0.0)	1.2/81.0 (1.2)	7.2/81.0 (5.7)	5.9/81.0 (5.9)
		5.00	1.00	0.0/0.0 (0.0)	0.0/0.0 (0.0)	0.0/0.0 (0.0)	3.8/78.2 (3.2)	6.2/78.2 (4.6)	5.0/78.2 (4.4)
		5.00	5.00	0.0/0.0 (0.0)	0.0/0.0 (0.0)	0.0/0.0 (0.0)	3.6/79.1 (2.5)	7.3/79.1 (5.4)	6.6/79.1 (5.5)
		20.00	5.00	0.0/0.0 (0.0)	0.0/0.0 (0.0)	0.0/0.0 (0.0)	0.6/79.0 (0.3)	2.3/79.0 (1.7)	2.0/79.0 (1.7)
		20.00	20.00	0.0/0.0 (0.0)	0.0/0.0 (0.0)	0.0/0.0 (0.0)	2.8/80.7 (2.3)	14.6/80.7 (11.6)	12.3/80.7 (11.8)
	m,h	1.00	1.00	0.0/0.0 (0.0)	0.0/0.0 (0.0)	0.0/0.0 (0.0)	0.1/ 5.1 (0.0)	2.8/ 5.1 (0.0)	0.1/ 5.1 (0.0)
		5.00	1.00	0.0/0.0 (0.0)	0.0/0.0 (0.0)	0.0/0.0 (0.0)	4.4/22.0 (0.9)	4.7/22.0 (0.8)	4.3/22.0 (0.7)
		5.00	5.00	0.0/0.0 (0.0)	0.0/0.0 (0.0)	0.0/0.0 (0.0)	2.0/16.0 (0.7)	4.1/16.0 (0.9)	2.6/16.0 (1.0)
		20.00	5.00	0.0/0.0 (0.0)	0.0/0.0 (0.0)	0.0/0.0 (0.0)	0.6/ 8.1 (0.0)	0.9/ 8.1 (0.2)	0.9/ 8.1 (0.2)
		20.00	20.00	0.0/0.0 (0.0)	0.0/0.0 (0.0)	0.0/0.0 (0.0)	0.6/ 8.5 (0.0)	4.4/ 8.5 (0.6)	1.3/ 8.5 (0.6)

Table 4.1: Probability of identifying embedded treatment rule under various simulated scenarios using different forms of Q-learning or IQ-learning with or without variable selection to identify a possibly more deeply tailored DTR.

Table 4.1 shows the probability of identifying the embedded DTR at either stage or overall for each scenario. Using no variable selection clearly did not result in an embedded treatment rule at either stage. In the absence of treatment effect heterogeneity Modified Q-learning with $\lambda_{0.8}$-LASSO identified the embedded rule at $t = 1$ and $t = 2$ with specified probability. IQ-learning with the $\lambda_{0.8}$-GLASSO also resulted

in high probability to identify an embedded rule in the absence of treatment effect heterogeneity. However, both Q-learning with $\lambda_{0.8}$-LASSO and IQ-learning with $\lambda_{0.8}$-GLASSO did not identify the embedded treatment rule at $t = 1$ at the specified probability of 0.8. This was caused by a small proportion of scenarios where $\beta_{23} \neq \mathbf{0}$ resulting in a heterogeneous relationship between A_1 and X_2 with both $\tilde{Y}$ and $\widehat{\mu}(H_2)$. Since X_1 and X_2 were simulated as highly correlated, this resulted in a heterogeneous treatment effect for $\tilde{Y}$ and $\widehat{\mu}(H_2)$ with X_1 and A_1. As a result, λ_q-LASSO and the λ_q-GLASSO would not necessarily identify the embedded rule at $t = 1$ since there is heterogeneity in the model being estimated at $t = 1$. Modified Q-learning is not affected by this because it uses the actual outcome Y or Y plus the regret as the outcome to estimate Q_1 which is not completely dependent on the variable selection at $t = 2$.

One possible consideration to solve the decreased probability to identify the embedded DTR at $t = 1$ from λ_q-GLASSO would be to not include β_{23} in Q_2 before performing variable selection. This would, however, lead to intentional misspecification of Q_2 when there is true treatment effect heterogeneity at $t = 1$ which could be missed when estimating Q_1. Another consideration could be to adapt the methods used in modified Q-learning to IQ-learning to use the actual outcome or outcome plus regret for the estimation of $\mu(H_2)$ and $\Delta(H_2)$. Lastly, we could modify the algorithm of the λ_q-GLASSO so that when selecting λ_q each permutation evaluates interactions at $t = 2$ but also estimates the probability of interaction at $t = 1$. This would lead to exponentially increased computation to the algorithm.

Table 4.2 shows the probability of benefit and probability of harm in identifying a more deeply tailored DTR. Using both the $\lambda_{0.8}$-LASSO and the $\lambda_{0.8}$-GLASSO resulted

Simulation Characteristics				variable selection					
$t=1$	$t=2$	p_{cons}	p_{true}	None (Q)	None (Mod-Q)	None (IQ)	$\lambda_{0.8}$-LASSO (Q)	$\lambda_{0.8}$-LASSO (Mod-Q)	$\lambda_{0.8}$-GLASSO (IQ)
m,nh	m,nh	1.00	1.00	3.2/ 79.6	3.3/ 81.1	3.2/ 79.6	3.3/29.7	3.3/27.6	3.3/27.2
		5.00	1.00	0.0/100.0	0.0/100.0	0.0/100.0	0.0/38.0	0.0/37.8	0.0/37.6
		5.00	5.00	3.5/ 96.5	3.5/ 96.5	3.5/ 96.5	3.5/40.3	3.5/35.9	3.5/36.0
		20.00	5.00	0.0/100.0	0.0/100.0	0.0/100.0	0.0/37.5	0.0/36.3	0.0/36.2
		20.00	20.00	2.6/ 97.4	2.6/ 97.4	2.6/ 97.4	2.6/38.8	2.6/33.2	2.6/34.0
	m,h	1.00	1.00	85.9/ 14.1	83.6/ 16.4	87.1/ 12.9	89.8/ 7.4	86.9/10.6	90.3/ 7.2
		5.00	1.00	0.3/ 99.7	0.3/ 99.7	0.3/ 99.7	0.4/88.8	0.4/87.7	0.4/87.7
		5.00	5.00	31.5/ 68.5	24.0/ 76.0	30.5/ 69.5	39.2/53.2	39.7/51.9	39.0/52.6
		20.00	5.00	0.4/ 99.6	0.2/ 99.8	0.3/ 99.7	62.3/34.0	59.0/37.5	62.7/33.9
		20.00	20.00	36.4/ 63.6	27.8/ 72.2	36.5/ 63.5	83.0/12.8	80.0/15.8	83.5/12.2
m,h	m,nh	1.00	1.00	62.1/ 37.0	62.0/ 37.0	62.1/ 37.0	84.4/14.0	79.6/14.1	79.9/14.2
		5.00	1.00	0.0/100.0	0.0/100.0	0.0/100.0	43.1/53.0	25.0/70.4	25.5/70.0
		5.00	5.00	10.5/ 89.5	8.6/ 91.4	10.5/ 89.5	73.4/23.8	66.3/28.5	66.9/27.8
		20.00	5.00	0.1/ 99.9	0.1/ 99.9	0.1/ 99.9	80.2/19.3	75.7/22.4	76.6/21.7
		20.00	20.00	6.8/ 93.2	5.4/ 94.6	6.8/ 93.2	76.1/21.2	64.1/24.1	64.2/23.9
	m,h	1.00	1.00	99.9/ 0.1	99.9/ 0.1	99.9/ 0.1	99.3/ 0.7	99.3/ 0.7	99.3/ 0.7
		5.00	1.00	79.6/ 20.4	65.8/ 34.2	81.0/ 19.0	27.6/71.3	24.5/74.5	26.4/72.7
		5.00	5.00	97.8/ 2.2	96.8/ 3.2	98.0/ 2.0	92.8/ 6.6	89.8/ 9.4	92.4/ 6.7
		20.00	5.00	89.2/ 10.8	78.9/ 21.1	90.6/ 9.4	94.4/ 5.5	94.3/ 5.5	94.5/ 5.3
		20.00	20.00	96.8/ 3.2	95.0/ 5.0	96.7/ 3.3	96.8/ 3.2	95.4/ 4.0	96.1/ 3.3

Table 4.2: Simulated values of the probability of benefit/harm under various scenarios using different forms of Q-learning or IQ-learning with or without variable selection to identify a possibly more deeply tailored DTR.

in a lower probability of harm in the absence of treatment effect heterogeneity. Both methods also maintained a high probability of benefit when there was treatment effect heterogeneity at both stages. The probability benefit from the $\lambda_{0.8}$-GLASSO seemed to be slightly higher when there was heterogeneity at $t = 2$ and no heterogeneity at $t = 1$ but the effect is quite small. Overall, both tables 4.1 and 4.2 show encouraging results for using λ_q-GLASSO for IQ-learning over using λ_q-LASSO with Q-learning but not necessarily over modified Q-learning.

4.5 Application to the PLUTO Study

The PLUTO Study is an ongoing two-stage SMART aimed at identifying an optimal DTR for assisting with smoking cessation and reduction. A secondary aim of the study is to identify more deeply tailored DTRs. A total of 643 current smokers

eligible for lung cancer screening were enrolled in the study and have completed baseline questionnaires. Participants were randomized with equal probability to four or eight weeks of Tobacco Longitudinal Care (TLC) which includes intensive telephone counselling along with nicotine replacement therapy and then re-randomized to either continue TLC at the same or reduced rate for responders or either continue TLC at the same rate or TLC plus medication therapy management (MTM) for non-responders. Non-response is classified as any smoking (even a puff) in the last 7 days.

We plan on use the methods developed in this paper to identify more deeply tailored DTRs to lead to a larger decrease in smoking than any of the embedded treatment rules if there is true treatment effect heterogeneity. We plan to identify any DTRs using both modified Q-learning with λ_q-LASSO and IQ-learning with λ_q-GLASSO. Variables to use when identifying a more deeply tailored DTRs have not been specified in advance so we will likely have to specify them for this context here. In chapter 2 we considered five variables to be used for an individualized treatment rule: cigarettes smoked per day over the previous 30 days, number of past quit attempts, readiness to quit ladder, calculated pack years, and age which we plan to use here as well.

4.6 Discussion

IQ-learning resolves an issue in Q-learning where the model for the first stage of treatment is incorrect when there is treatment effect heterogeneity at the second stage of treatment. We introduced variable selection methods that can be applied to IQ-learning that result in an embedded DTR at the first stage with user specified

probability in the absence of treatment effect heterogeneity. When there is not treatment effect heterogeneity at the second stage of treatment, the benefits of IQ-learning are diminished thus the λ_q-GLASSO is only applied when there is treatment effect heterogeneity at the second stage and the λ_q-LASSO with modified Q-learning is used otherwise.

Our simulation study showed that IQ-learning with the λ_q-GLASSO performed at least as well as the λ_q-LASSO with modified Q-learning. There was some possible evidence suggesting that the λ_q-GLASSO leads to higher probability of benefit when there is heterogeneity at the second treatment stage even though the embedded rule was selected less often than expected. When an interaction with the first stage treatment was included in the second stage model, the resulting predicted value would sometimes prevent the identification of the embedded rule at the first stage of treatment even in the absence of treatment effect heterogeneity at the first stage of treatment. We suggested some possible solutions to this problem but evaluation of these solutions is ongoing. Additionally we outlined our plans to apply these methods to the PLUTO study.

Variable selection has been shown to reduce the probability of harm when identifying a more deeply tailored DTR especially in the absence of treatment effect heterogeneity at either stage. With IQ-learning, variable selection at the first stage presents difficulties due to having two separate models to estimate. The group LASSO presented in this paper provide an efficient method to perform variable selection that avoids harm with user specified probability when there is no treatment effect heterogeneity at the first stage of treatment.

Chapter 5

Conclusion

In Chapter 2 we introduced the concept of beneficial personalized treatment rules in the context of a single stage RCT. We also discussed the need for the prior calculation of the probability to identify a beneficial ITR and introduced methods to do so. Few methods to identify ITRs consider a scenario where there is no treatment effect heterogeneity and the optimal treatment rule is truly the static treatment rule. To fill this gap we developed the λ_q-LASSO for a single stage to identify the static rule in the absence of treatment effect heterogeneity.

To extend the methods from Chapter 2 to a two stage SMART we apply the λ_q-LASSO to modified Q-learning. The λ_q-LASSO performs well compared to S-Scores and Forward selection in simulation studies identifying the embedded rule at each stage with user specified probability in the absence of treatment effect heterogeneity. Each model selection method was also applied to data from the M-Bridge study identifying in easily interpretable more deeply tailored rule using λ_q-LASSO and S-Scores.

We did not solve the issue of inherent non-linearity non-smoothness in the stage 1 model between the predicted outcome and predictors in Chapter 3 so we introduce the λ_q-GLASSO applied to IQ-learning in Chapter 4. The group LASSO methods presented should result in identification of the embedded rule at the first stage of treatment but we noticed through simulation that inclusion of an interaction between stage one treatment and a predictor can lead to heterogeneity between the predicted outcome and the predictors in the stage one model. When this occurs, the λ_q-GLASSO fails to select the embedded rule with specified probability because heterogeneity was introduced into the stage one model. We did not notice these issues when using modified Q-learning so we are currently developing solutions to this problem when using IQ-learning.

Another future problem to tackle when considering beneficial or harmful personalized treatment rules is to incorporate them into adaptive randomization (AR) methods. Cheung et. al. [43], adapted some of the adaptive randomization methods from Thall and Wathen [44] to be used in SMARTs based around the Q-functions called SMART-AR. Additional steps in adaptive randomization could also be modified so that participants are only assigned treatment according to beneficial DTRs during the trial rather determining the beneficial DTRs after the trial has completed.

References

[1] Steven S. Fu, Alexander J. Rothman, David M. Vock, Bruce Lindgren, Daniel Almirall, Abbie Begnaud, Anne Melzer, Kelsey Schertz, Susan Glaeser, Patrick Hammet, and Anne M. Joseph. Program for Lung Cancer Screening and Tobacco Cessation: Study Protocol of a Sequential, Multiple Assignment, Randomized Trial. *Contemporary Clinical Trials*, 60(9):86–95, 2017.

[2] Rubin D. B. Estimating causal effects of treatment in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688–701, 1974.

[3] Eric B. Laber, Ying Qi Zhao, Todd Regh, Marie Davidian, Anastasios Tsiatis, Joseph B. Stanford, Donglin Zeng, Rui Song, and Michael R. Kosorok. Using pilot data to size a two-arm randomized trial to find a nearly optimal personalized treatment strategy. *Statistics in Medicine*, 35(8):1245–1256, apr 2016.

[4] Susan Athey and Guido Imbens. Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences of the United States of America*, 113(27):7353–60, 2016.

[5] Nicolás M. Ballarini, Gerd K. Rosenkranz, Thomas Jaki, Franz König, and Martin Posch. Subgroup identification in clinical trials via the predicted individual

treatment effect. *PLoS ONE*, 13(10), oct 2018.

[6] Jared C Foster, Jeremy M G Taylor, and Stephen J Ruberg. Subgroup identification from randomized clinical trial data. *Stat Med.*, 30(24), 2011.

[7] Lacey Gunter, Ji Zhu, and Susan Murphy. Variable Selection for Qualitative Interactions in Personalized Medicine While Controlling the Family-Wise Error rate. *Journal of biopharmaceutical statistics*, 21(6):1063–1078, 2012.

[8] Min Qian and Susan A. Murphy. Performance guarantees for individualized treatment rules. *The Annals of Statistics*, 39(2):1180–1210, 2011.

[9] Stephen J. Ruberg, Lei Chen, and Yanping Wang. The mean does not mean as much anymore: Finding sub-groups for tailored therapeutics. In *Clinical Trials*, volume 7, pages 574–583, oct 2010.

[10] Xiaogang Su, Chih Ling Tsai, Hansheng Wang, David M. Nickerson, and Bogong Li. Subgroup analysis via recursive partitioning. *Journal of Machine Learning Research*, 10:141–158, 2009.

[11] Yaoyao Xu, Menggang Yu, Ying Qi Zhao, Quefeng Li, Sijian Wang, and Jun Shao. Regularized Outcome Weighted Subgroup Identification for Differential Treatment Effects. *Biometrics*, 71(3):645–653, 2015.

[12] Baqun Zhang, Anastasios A. Tsiatis, Eric B. Laber, and Marie Davidian. A Robust Method for Estimating Optimal Treatment Regimes. *Biometrics*, 68(4):1010–1018, dec 2012.

[13] Michael H. Kutner, Christopher J. Nachtsheim, and John Neter. *Applied Linear Regression Models*. McGraw-Hill Irwin, fourth edition, 2004.

[14] Robert Tibshirani. Regression Shrinkage and Selection via the LASSO. *Journal of the Royal Statistical Society*, 58(1):267–288, 1996.

[15] Kosuke Imai and Marc Ratkovic. Estimating treatment effect heterogeneity in randomized program evaluation. *Annals of Applied Statistics*, 7(1):443–470, 2013.

[16] Hui Zou and Trevor Hastie. Regression and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2):301–320, 2005.

[17] M. Stone. Cross-Validatory Choice and Assessment of Statistical Predictions. *Journal of the Royal Statistical Society*, 36(2):111–147, 1974.

[18] Brian Ripley, Bill Venables, Douglas M Bates, Kurt Hornik, Albrecht Gebhardt, and David Firth. Support Functions and Datasets for Venables and Ripley's MASS, Version 7.3-51.4. pages 1–169, 2019.

[19] Author Jerome, Trevor Hastie, Rob Tibshirani, and Noah Simon. Package ' glmnet '. 2019.

[20] Leo Breiman, Adele Cutler, Andy Liaw, and Matthew Wiener. Package 'randomForest'. 2018.

[21] Leo Breiman. Random Forests. *Machine Learning*, (45):5–32, 2001.

[22] Bibhas Chakraborty and Erica E.M. Moodie. *Statistical Methods for Dynamic Treatment Regimes*. Statistics for Biology and Health. Springer New York, New York, NY, 2013.

[23] Ilya Lipkovich, Alex Dmitrienko, and Ralph B. D'Agostino. Tutorial in biostatistics: data-driven subgroup identification and analysis in clinical trials. *Statistics in Medicine*, 36(1):136–196, jan 2017.

[24] S. A. Murphy, Elja Arjas, C. Jennison, A. P. Dawid, D. R. Cox, Stephen Senn, Robert G. Cowell, V. Didelez, Richard D. Gill, J. B. Kadane, and James M. Robins. Optimal dynamic treatment regimes. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 65(2):331–366, 2003.

[25] David M. Vock and Daniel Almirall. Sequential Multiple Assignment Randomized Trial (SMART). In *Wiley StatsRef: Statistics Reference Online*, pages 1–11. John Wiley and Sons, Ltd, aug 2018.

[26] Megan E. Patrick, Jeffrey A. Boatman, Nicole Morrell, Anna C. Wagner, Grace R. Lyden, Inbal Nahum-Shani, Cheryl A. King, Erin E. Bonar, Christine M. Lee, Mary E. Larimer, David M. Vock, and Daniel Almirall. A sequential multiple assignment randomized trial (SMART) protocol for empirically developing an adaptive preventive intervention for college student drinking reduction. *Contemporary Clinical Trials*, 96(July), 2020.

[27] Richard Bellman. *Dynamic Programming*. Princeton University Press, 1957.

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

[29] D Blatt, S a Murphy, and J Zhu. A-learning for approximate planning. *Technical Report 04-63, The Methodology Center, Pennsylvania State University*, 2004.

[30] Yingqi Zhao, Donglin Zeng, John A. Rush, and Michael R. Kosorok. Estimating Individualized Treatment Rules Using Outcome Weighted Learning. *Journal of the American Statistical Association*, 107(449):1106–1118, 2012.

[31] Ying Qi Zhao, Donglin Zeng, Eric B. Laber, and Michael R. Kosorok. New Statistical Learning Methods for Estimating Optimal Dynamic Treatment Regimes. *Journal of the American Statistical Association*, 110(510):583–598, 2015.

[32] Peter Biernot and Erica E.M. Moodie. A comparison of variable selection approaches for dynamic treatment regimes. *International Journal of Biostatistics*, 6(1), 2010.

[33] Michael P. Wallace, Erica E.M. Moodie, and David A. Stephens. Model selection for G-estimation of dynamic treatment regimes. *Biometrics*, 75(4):1205–1215, 2019.

[34] Xuelin Huang, Sangbum Choi, Lu Wang, and Peter F. Thall. Optimization of Multi-Stage Dynamic Treatment Regimes Utilizing Accumulated Data. *Stat Med.*, 34(26):3424–3443, 2015.

[35] James M. Robins. Optimal Structural Nested Models for Optimal Sequential Decisions. *Proceedings of the Second Seattle Symposium in Biostatistics*, pages 189–326, 2003.

[36] Wei Pan. Akaike ' s Information Criterion in Generalized. *Biometrics*, 57(1):120–125, 2001.

[37] Paul R. Rosenbaum and Donald B. Rubin. The central role of the propensity score in observational studies for causal effects. *Matched Sampling for Causal Effects*, 70(1):41–55, 1983.

[38] Phillip J. Schulte, Anastasios A. Tsiatis, Eric B. Laber, and Marie Davidian. Q- and A-learning Methods for Estimating Optimal Dynamic Treatment Regimes. *Statistical Science*, 29(4):640–661, 2014.

[39] Eric B. Laber, Kristin A. Linn, and Leonard A. Stefanski. Interactive model building for Q-learning. *Biometrika*, 101(4):831–847, 2014.

[40] Ming Yuan and Yi Lin. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 68(1):49–67, 2006.

[41] Yi Yang and Hui Zou. A fast unified algorithm for solving group-lasso penalize learning problems. *Statistics and Computing*, 25(6):1129–1141, 2015.

[42] Brandon Koch, David M. Vock, and Julian Wolfson. Covariate selection with group lasso and doubly robust estimation of causal effects. *Biometrics*, 74(1):8–17, 2018.

[43] Ying Kuen Cheung, Bibhas Chakraborty, and Karina W. Davidson. Sequential multiple assignment randomized trial (SMART) with adaptive randomization for quality improvement in depression treatment program. *Biometrics*, 71(2):450–459, jun 2015.

[44] Peter F. Thall and J. Kyle Wathen. Practical Bayesian Adaptive Randomization in Clinical Trials. *European Journal of Cancer*, 43(5):859–866, 2007.

Appendix 1-Supplemental materials for Design Considerations and Analytical Framework for Reliably Identifying a Beneficial Individualized Treatment Rule

	Treatment rule identification method					
ν	None	Forward	LASSO	Modified	Elastic	Random Forest
One Potential Tailoring Variable ($p = 1$)						
0	-0.006	-0.003	-0.005	-0.003	-0.006	-0.018
0.02	-0.003	-0.003	-0.003	-0.003	-0.003	-0.014
0.16	0.021	0.021	0.022	0.021	0.021	0.005
0.3	0.111	0.111	0.112	0.111	0.112	0.083
Five Potential Tailoring Variables ($p = 5$)						
0	-0.034	-0.006	-0.021	-0.006	-0.026	-0.042
0.02	-0.025	-0.005	-0.018	-0.007	-0.022	-0.035
0.16	0.005	0.019	0.009	0.015	0.007	-0.010
0.3	0.098	0.109	0.100	0.109	0.100	0.075
Twenty Potential Tailoring Variables ($p = 20$)						
0	-0.082	-0.016	-0.042	-0.007	-0.053	-0.022
0.02	-0.070	-0.014	-0.038	-0.007	-0.047	-0.018
0.16	-0.036	0.013	-0.012	0.008	-0.017	0.000
0.3	0.057	0.105	0.081	0.107	0.077	0.076

Table A1.1: Average value of $V(\widehat{d}^{opt}) - V(\widehat{w}^{opt})$ after varying values of ν and p, and the method used to identify the treatment rule. $\Delta = 0.3$, $V_y = 1$.

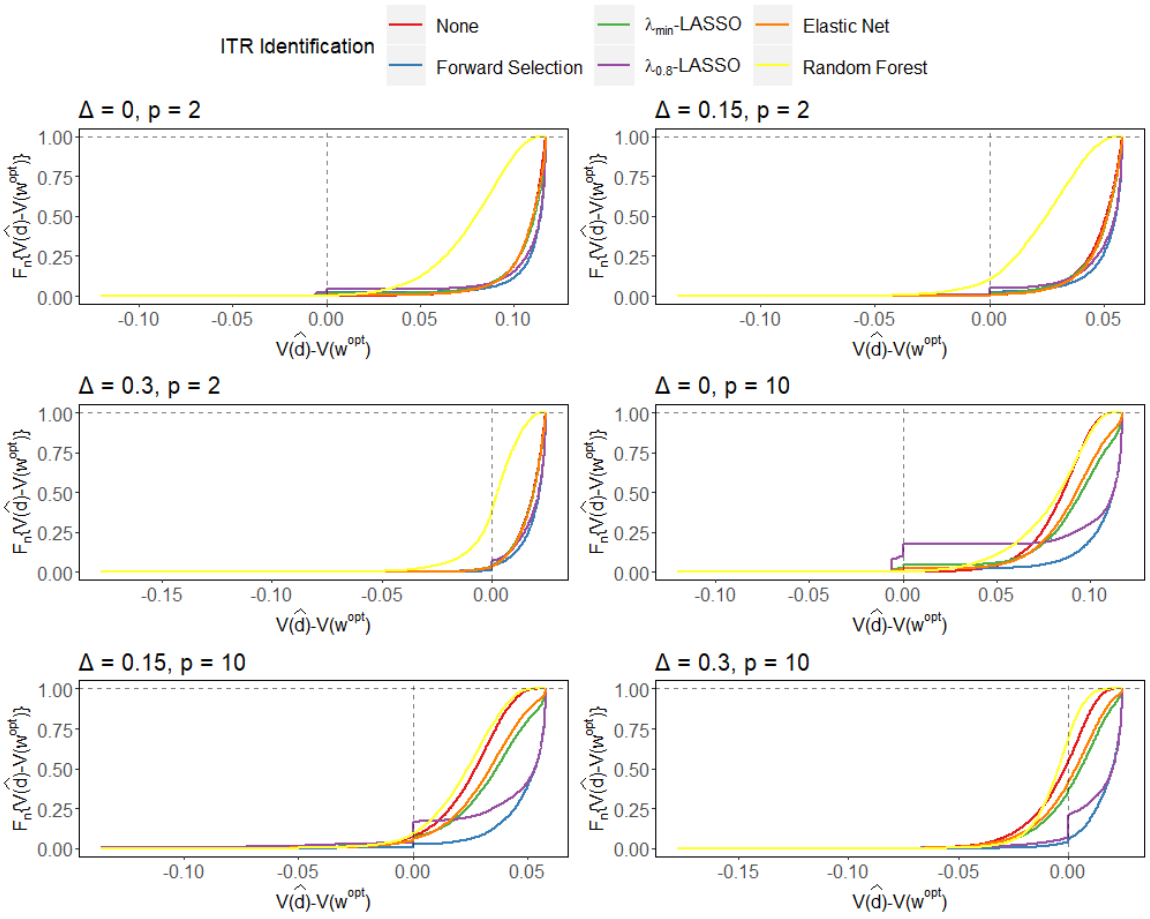

Figure A1.1: Empirical CDF's of $V(\widehat{d^{opt}}) - V(w^{opt})$ when there are two or ten predictors after using treatment rule identification methods. No main effect refers to when $\beta_2 = 0$, Over-estimate main effect refers to when $\beta_2 = 0.075$, Correctly assumed main effect refers to when $\beta_2 = 0.15$.

		Treatment rule identification method					
ρ	p	None	Forward	λ_{min}-LASSO	$\lambda_{0.8}$-LASSO	Elastic	Random Forest
Single predictor with non-zero coefficient							
0.00	1	0.03	0.02	0.02	0.02	0.02	0.01
0.00	5	0.01	0.02	0.01	0.02	0.01	-0.00
0.00	20	-0.04	0.01	-0.01	0.01	-0.02	0.00
0.80	5	0.01	0.02	0.01	0.02	0.01	-0.00
0.80	20	-0.04	0.02	-0.00	0.01	-0.01	0.00
All predictors with non-zero and equal coefficients							
0.00	5	0.00	-0.00	0.00	-0.00	0.00	-0.01
0.00	20	-0.04	-0.04	-0.04	-0.01	-0.04	-0.01
0.80	5	0.00	0.01	0.02	0.02	0.01	0.00
0.80	20	-0.04	0.00	-0.00	0.00	-0.01	0.00
All predictors with non-zero and diminishing coefficients							
0.00	5	0.00	0.00	0.00	0.00	0.01	-0.01
0.00	20	-0.04	-0.01	-0.03	-0.00	-0.04	-0.00
0.80	5	0.01	0.02	0.02	0.02	0.01	0.00
0.80	20	-0.04	0.01	-0.00	0.01	-0.01	0.01

Table A1.2: Average value of $V(\widehat{d}^{opt}) - V(\widehat{w}^{opt})$ after varying values of the correlation between the predictors, the form of the coefficients for the predictors, the number of predictors, and the method used to identify the treatment rule. $\Delta = 0.3$, $V_y = 1$, $R^2_C = 0.3$, $\nu = 0.16$. The β-Form column refers to the form of the coefficients, where "Single" refers to only one predictor has a non-zero coefficient, "Even" refers to the effect of the predictors evenly dispersed among the predictors, and "Diminishing" refers to one predictor strongly associated with the outcome and treatment and a diminished effect in the others.

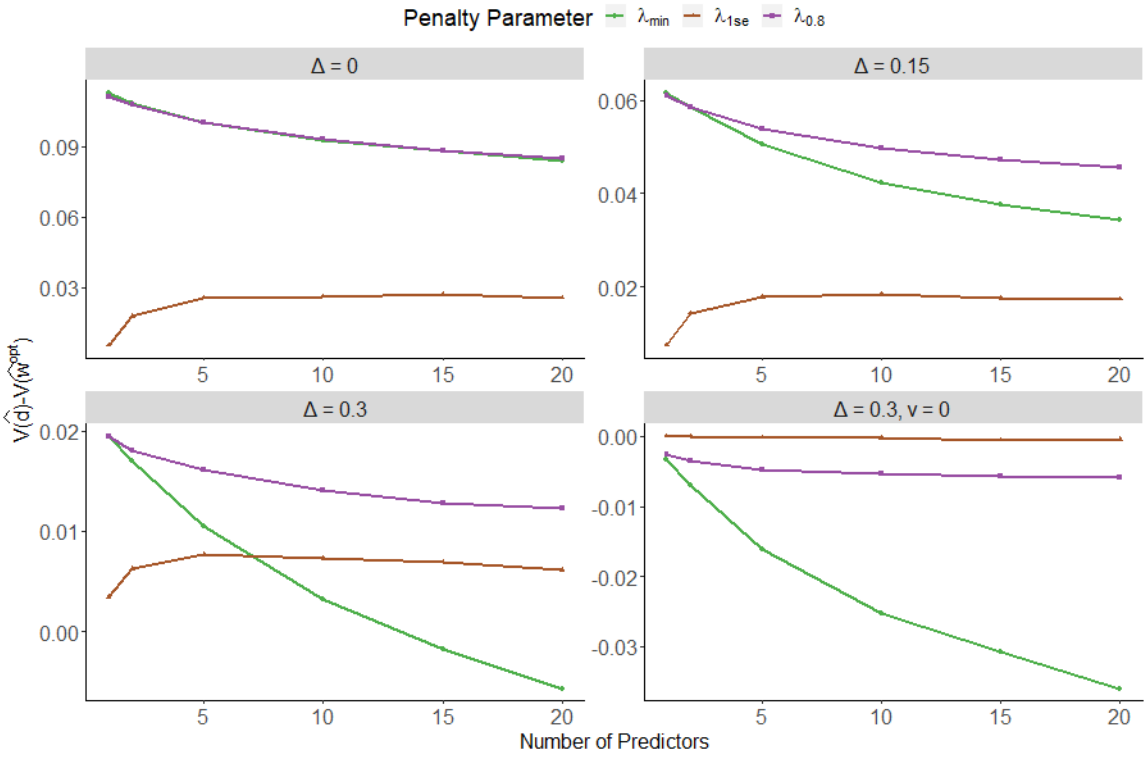

Figure A1.2: $V(\widehat{d}^{opt}) - V(\widehat{w}^{opt})$ using three LASSO models for model selection with varied penalty parameters.

Variable	Variable.1
Age at baseline	Readiness to Quit Ladder
Gender identity	Heart attack
Ethnicity	Heart failure
Race	Peripheral vascular disease
Highest grade completed	Cerebrovascular disease
Current marital status	Asthma
Past year pre-tax income	Emphysema, chronic bronchitis or COPD
Past 30 day average CPD	Peptic ulcer disease
Age at 1st cigarette	Diabetes
# past quit attempts	Kidney disease
Told by medical provider to quit within 12 months	Lupus, polymyalgia rheumatic, rheumatoid arthritis
History of use of in-person individual counseling	Cigarette Dependence Scale (3 separate questions)
History of use of group counseling	Self-efficacy for smoking cessation (8 separate questions)
History of use of quit line or telephone help line	MN Nicotine Withdrawal Scale
History of use of internet/web-based quit app	Questionnaire on Smoking Urges
History of use of NRT: patch	Regulatory focus/personality (Prevention, Promotion)
History of use of NRT: gum	PEG (Pain)
History of use of NRT: inhaler	AUDIT-C (Alcohol)
History of use of NRT: nasal spray	PHQ-8 (Depression)
History of use of NRT: lozenge	GAD-7 (Anxiety)
History of use of varenicline	Charlson Index (Self report medical hx)
History of use of bupropion	Calculated field: pack year history
History of use of other prescription cessation med	PIQ: Current social support

Table A1.3: List of the 55 variables included PLUTO analysis

Appendix 2-Supplemental materials for Identification of non-harmful DTRs using LASSO with permutation-based selection of the penalty parameter

Simulation Characteristics				DGM								
$t=1$	$t=2$	p_{cons}	p_{true}	P_1	P_{-1}	R_{11}	Δ_{21}	Δ_{2-1}	Δ_1	ν_{21}	ν_{2-1}	ν_{11}
		1.00	1.00	0.60	0.40	0.15	0.00	0.00	0.00	0.00	0.00	0.00
		5.00	1.00	0.60	0.40	0.13	0.00	0.00	0.00	0.00	0.00	0.00
	nm,nh	5.00	5.00	0.60	0.40	0.13	0.00	0.00	0.00	0.00	0.00	0.00
		20.00	5.00	0.60	0.40	0.18	0.00	0.00	0.00	0.00	0.00	0.00
		20.00	20.00	0.60	0.40	0.18	0.00	0.00	0.00	0.00	0.00	0.00
		1.00	1.00	0.60	0.40	0.15	0.30	0.30	0.00	0.00	0.00	0.00
		5.00	1.00	0.60	0.40	0.13	0.30	0.30	0.00	0.00	0.00	0.00
nm,nh	m,nh	5.00	5.00	0.60	0.40	0.13	0.30	0.30	0.00	0.00	0.00	0.00
		20.00	5.00	0.60	0.40	0.18	0.30	0.30	0.00	0.00	0.00	0.00
		20.00	20.00	0.60	0.40	0.18	0.30	0.30	0.00	0.00	0.00	0.00
		1.00	1.00	0.60	0.40	0.29	0.30	0.30	0.00	0.22	0.22	0.00
		5.00	1.00	0.60	0.40	0.22	0.30	0.30	0.00	0.22	0.22	0.00
	m,h	5.00	5.00	0.60	0.40	0.26	0.30	0.30	0.00	0.21	0.21	0.00
		20.00	5.00	0.60	0.40	0.18	0.30	0.30	0.00	0.21	0.21	0.00
		20.00	20.00	0.60	0.40	0.29	0.30	0.30	0.00	0.22	0.22	0.00
		1.00	1.00	0.60	0.40	0.15	0.00	0.00	0.30	0.00	0.00	0.00
		5.00	1.00	0.60	0.40	0.13	0.00	0.00	0.30	0.00	0.00	0.00
	nm,nh	5.00	5.00	0.60	0.40	0.13	0.00	0.00	0.30	0.00	0.00	0.00
		20.00	5.00	0.60	0.40	0.18	0.00	0.00	0.30	0.00	0.00	0.00
		20.00	20.00	0.60	0.40	0.18	0.00	0.00	0.30	0.00	0.00	0.00
		1.00	1.00	0.60	0.40	0.15	0.30	0.30	0.30	0.00	0.00	0.00
		5.00	1.00	0.60	0.40	0.13	0.30	0.30	0.30	0.00	0.00	0.00
m,nh	m,nh	5.00	5.00	0.60	0.40	0.13	0.30	0.30	0.30	0.00	0.00	0.00
		20.00	5.00	0.60	0.40	0.18	0.30	0.30	0.30	0.00	0.00	0.00
		20.00	20.00	0.60	0.40	0.18	0.30	0.30	0.30	0.00	0.00	0.00
		1.00	1.00	0.60	0.40	0.29	0.40	0.20	0.33	0.15	0.30	0.00
		5.00	1.00	0.60	0.40	0.22	0.40	0.20	0.33	0.15	0.30	0.00
	m,h	5.00	5.00	0.60	0.40	0.26	0.40	0.20	0.33	0.14	0.29	0.00
		20.00	5.00	0.60	0.40	0.18	0.40	0.20	0.33	0.14	0.29	0.00
		20.00	20.00	0.60	0.40	0.29	0.40	0.20	0.33	0.15	0.31	0.00

Table A2.1: Simulation characteristics of the 9 scenarios considered. At each time point the we consider one of two scenarios: treatment main and heterogeneous effect (m,h) and treatment main non-heterogeneous effect (m,nh). Additionally we vary the number of variables considered for interaction with treatment $p = 1, 5, 20, 20$ and the number of variables that truly have a heterogeneous effect with treatment $p = 1, 5, 20, 5$

Simulation Characteristics				DGM								
$t=1$	$t=2$	p_{cons}	p_{true}	P_1	P_{-1}	R_{11}	Δ_{21}	Δ_{2-1}	Δ_1	ν_{21}	ν_{2-1}	ν_{11}
		1.00	1.00	0.60	0.40	0.29	0.00	0.00	0.30	0.00	0.00	0.22
		5.00	1.00	0.60	0.40	0.22	0.00	0.00	0.30	0.00	0.00	0.22
	nm,nh	5.00	5.00	0.60	0.40	0.26	0.00	0.00	0.30	0.00	0.00	0.21
		20.00	5.00	0.60	0.40	0.18	0.00	0.00	0.30	0.00	0.00	0.21
		20.00	20.00	0.60	0.40	0.29	0.00	0.00	0.30	0.00	0.00	0.21
		1.00	1.00	0.60	0.40	0.29	0.30	0.30	0.30	0.00	0.00	0.22
		5.00	1.00	0.60	0.40	0.22	0.30	0.30	0.30	0.00	0.00	0.22
m,h	m,nh	5.00	5.00	0.60	0.40	0.26	0.30	0.30	0.30	0.00	0.00	0.21
		20.00	5.00	0.60	0.40	0.18	0.30	0.30	0.30	0.00	0.00	0.21
		20.00	20.00	0.60	0.40	0.29	0.30	0.30	0.30	0.00	0.00	0.21
		1.00	1.00	0.60	0.40	0.42	0.40	0.20	0.33	0.15	0.30	0.19
		5.00	1.00	0.60	0.40	0.34	0.40	0.20	0.33	0.15	0.30	0.19
	m,h	5.00	5.00	0.60	0.40	0.38	0.40	0.20	0.33	0.14	0.29	0.18
		20.00	5.00	0.60	0.40	0.24	0.40	0.20	0.33	0.14	0.29	0.18
		20.00	20.00	0.60	0.40	0.40	0.40	0.20	0.33	0.15	0.31	0.18

Table A2.1 continued:

Simulation Characteristics				Variable Selection			
$t=1$	$t=2$	p_{cons}	p_{true}	None	$\lambda_{0.8}$-LASSO	S-Scores	Forward
		1.00	1.00	0.0/0.0 (0.0)	79.6/79.7 (63.8)	61.1/39.2 (23.8)	86.0/73.5 (63.6)
		5.00	1.00	0.0/0.0 (0.0)	79.0/78.7 (62.2)	26.2/20.4 (5.9)	45.7/38.3 (17.0)
	nm,nh	5.00	5.00	0.0/0.0 (0.0)	81.6/79.2 (64.2)	26.0/20.9 (6.3)	46.4/38.3 (18.2)
		20.00	5.00	0.0/0.0 (0.0)	78.7/79.0 (62.1)	38.4/34.4 (14.3)	4.1/ 4.0 (0.2)
		20.00	20.00	0.0/0.0 (0.0)	79.8/80.0 (63.7)	40.8/35.3 (15.0)	4.2/ 4.0 (0.2)
		1.00	1.00	0.0/0.0 (0.0)	78.6/80.0 (62.8)	58.5/79.8 (46.0)	84.6/73.8 (62.6)
		5.00	1.00	0.0/0.0 (0.0)	79.6/78.8 (62.9)	26.0/56.6 (15.2)	46.4/39.6 (18.5)
nm,nh	m,nh	5.00	5.00	0.0/0.0 (0.0)	79.6/79.9 (63.6)	25.4/58.9 (14.7)	44.9/39.8 (18.0)
		20.00	5.00	0.0/0.0 (0.0)	80.3/79.5 (63.9)	38.7/59.8 (23.7)	4.5/ 4.2 (0.4)
		20.00	20.00	0.0/0.0 (0.0)	80.0/80.4 (64.4)	39.5/60.5 (24.6)	4.8/ 4.1 (0.2)
		1.00	1.00	0.0/0.0 (0.0)	78.9/ 4.9 (4.1)	60.2/18.5 (11.4)	84.2/ 3.5 (3.1)
		5.00	1.00	0.0/0.0 (0.0)	78.6/13.7 (10.8)	25.2/17.3 (5.0)	42.6/ 2.3 (1.2)
	m,h	5.00	5.00	0.0/0.0 (0.0)	79.1/26.8 (20.9)	27.0/22.6 (5.7)	44.5/ 4.2 (2.0)
		20.00	5.00	0.0/0.0 (0.0)	80.2/48.4 (39.5)	39.6/30.6 (12.3)	4.1/ 0.4 (0.0)
		20.00	20.00	0.0/0.0 (0.0)	80.3/54.0 (43.6)	39.0/39.0 (15.0)	4.2/ 0.5 (0.0)
		1.00	1.00	0.0/0.0 (0.0)	81.5/79.8 (65.3)	97.0/38.6 (37.4)	86.9/74.0 (64.4)
		5.00	1.00	0.0/0.0 (0.0)	79.3/80.9 (63.9)	83.0/21.2 (17.8)	44.2/40.9 (18.4)
m,nh	nm,nh	5.00	5.00	0.0/0.0 (0.0)	78.6/80.1 (63.2)	83.2/21.2 (18.0)	46.2/39.6 (18.8)
		20.00	5.00	0.0/0.0 (0.0)	80.3/79.8 (64.4)	81.7/35.1 (28.9)	4.0/ 4.0 (0.2)
		20.00	20.00	0.0/0.0 (0.0)	79.6/80.7 (63.9)	82.2/34.7 (28.5)	4.5/ 5.1 (0.2)
		1.00	1.00	0.0/0.0 (0.0)	0.1/81.4 (0.1)	10.5/44.2 (4.8)	0.1/74.3 (0.1)
		5.00	1.00	0.0/0.0 (0.0)	0.5/81.9 (0.4)	8.8/25.5 (2.7)	0.0/41.3 (0.0)
m,h	nm,nh	5.00	5.00	0.0/0.0 (0.0)	5.6/80.4 (4.6)	22.6/21.5 (4.8)	0.4/40.0 (0.2)
		20.00	5.00	0.0/0.0 (0.0)	19.3/81.2 (15.8)	27.5/36.8 (10.2)	0.0/ 4.1 (0.0)
		20.00	20.00	0.0/0.0 (0.0)	35.6/78.2 (27.9)	64.0/32.4 (20.9)	0.0/ 3.8 (0.0)

Table A2.2: Estimates for the probability of identifying an embedded DTR as the estimated optimal DTR when performing variable selection methods in additional scenarios. Values are the probability of identifying an embedded DTR at $t = 1/t = 2$ and (both $t = 1$ and $t = 2$). At each time point the main we considered one of three scenarios: treatment main and heterogeneous effect (m,h), treatment main non-heterogeneous effect (m,nh) no treatment main or heterogeneous effect (nm, nh). Some of these scenarios were already presented in table 3.2. Additionally we varied the number of true predictor interactions ($p_{true} = 1, 5, 20$) and the variables considered for interaction with treatment ($p_{cons} = 1, 5, 20$).

Simulation Characteristics				Variable Selection			
$t=1$	$t=2$	p_{cons}	p_{true}	None	$\lambda_{0.8}$-LASSO	S-Scores	Forward
		1.00	1.00	0.0/ 0.0	0.0/ 0.0	0.0/ 0.0	0.0/ 0.0
		5.00	1.00	0.0/ 0.0	0.0/ 0.0	0.0/ 0.0	0.0/ 0.0
	nm,nh	5.00	5.00	0.0/ 0.0	0.0/ 0.0	0.0/ 0.0	0.0/ 0.0
		20.00	5.00	0.0/ 0.0	0.0/ 0.0	0.0/ 0.0	0.0/ 0.0
		20.00	20.00	0.0/ 0.0	0.0/ 0.0	0.0/ 0.0	0.0/ 0.0
		1.00	1.00	1.2/ 68.0	0.9/13.2	1.1/16.8	1.0/17.2
		5.00	1.00	0.8/ 98.4	0.5/19.6	0.8/42.4	0.7/56.0
nm,nh	m,nh	5.00	5.00	0.8/ 98.7	0.5/18.4	0.7/40.0	0.6/55.3
		20.00	5.00	1.1/ 98.9	0.6/20.1	0.9/39.2	1.1/94.2
		20.00	20.00	1.4/ 98.6	0.6/19.4	1.1/38.3	1.4/94.0
		1.00	1.00	97.5/ 2.2	93.1/ 1.4	79.9/ 1.4	94.3/ 1.7
		5.00	1.00	80.9/ 19.1	81.0/ 5.1	74.7/ 7.9	86.6/10.6
	m,h	5.00	5.00	74.4/ 25.6	31.2/41.7	37.2/40.2	56.6/39.0
		20.00	5.00	6.0/ 94.0	11.5/40.0	16.0/53.5	15.8/83.7
		20.00	20.00	15.6/ 84.4	3.2/43.0	6.5/54.6	9.1/90.3
		1.00	1.00	0.0/ 39.8	0.0/18.0	0.0/ 3.0	0.0/12.9
		5.00	1.00	0.0/ 95.3	0.0/20.7	0.0/16.5	0.0/54.6
m,nh	nm,nh	5.00	5.00	0.0/ 95.2	0.0/21.3	0.0/16.3	0.0/52.9
		20.00	5.00	0.0/100.0	0.0/19.7	0.0/18.1	0.0/95.3
		20.00	20.00	0.0/100.0	0.0/20.4	0.0/17.6	0.0/94.4
		1.00	1.00	100.0/ 0.0	99.9/ 0.0	89.5/ 0.0	99.9/ 0.0
		5.00	1.00	99.7/ 0.3	99.4/ 0.2	91.1/ 0.0	99.8/ 0.2
m,h	nm,nh	5.00	5.00	99.2/ 0.8	85.8/ 8.6	68.8/ 8.6	97.2/ 2.4
		20.00	5.00	49.4/ 50.6	54.9/25.8	51.6/21.0	69.2/30.8
		20.00	20.00	51.0/ 49.0	1.2/63.2	6.8/29.2	27.2/72.8

Table A2.3: Estimates for the probability of identifying a beneficial DTR and the probability of identifying a harmful DTR, P_b/P_h, under in additional scenarios scenarios with varying DGM characteristics. At each time point the main we considered one of three scenarios: treatment main and heterogeneous effect (m,h), treatment main non-heterogeneous effect (m,nh) no treatment main or heterogeneous effect (nm, nh). Some of these scenarios were already presented in table 3.3. Additionally we varied the number of true predictor interactions ($p_{true} = 1, 5, 20$) and the variables considered for interaction with treatment ($p_{cons} = 1, 5, 20$).

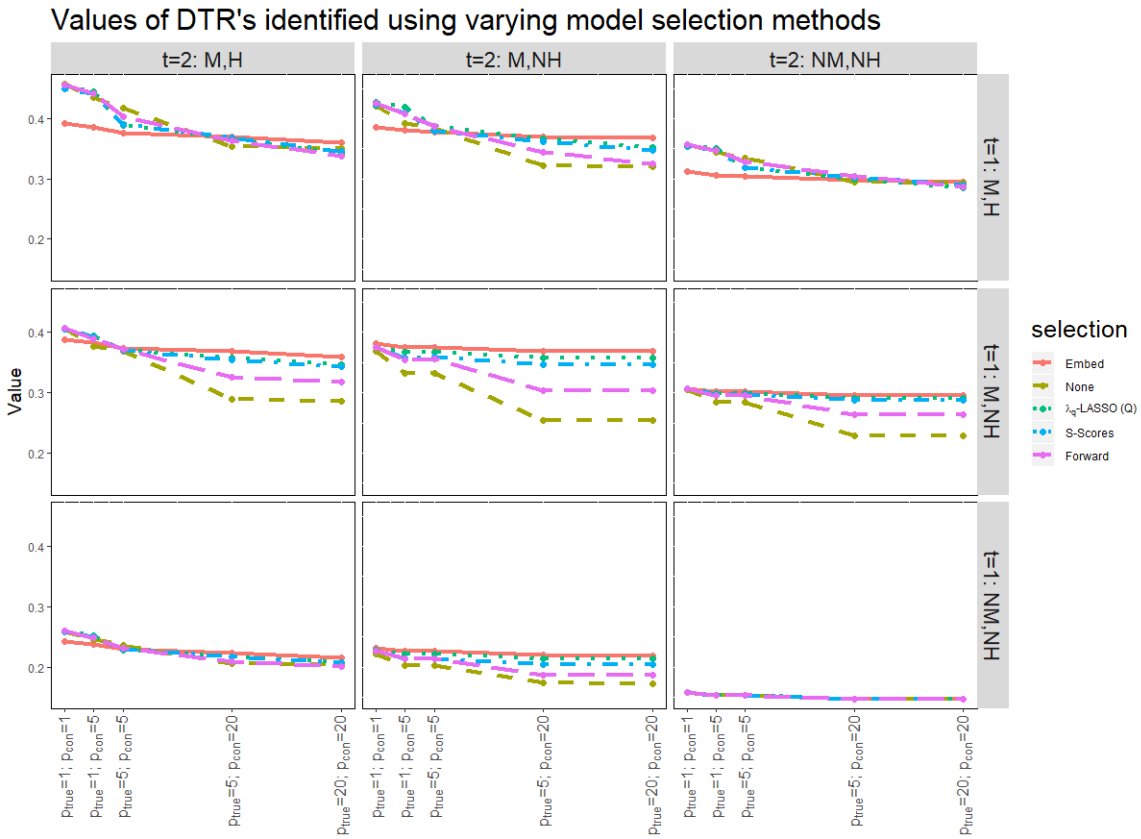

Figure A2.1: Values of estimated DTRs resulting from the embedded DTR, no variable selection, $\lambda_{0.8}$-LASSO, S-Scores, and Forward Selection. At each time point the main we considered one of three scenarios: treatment main and heterogeneous effect (m,h), treatment main non-heterogeneous effect (m,nh) no treatment main or heterogeneous effect (nm, nh). Some of these scenarios were already presented in figure 3.2. Additionally we varied the number of true predictor interactions ($p_{true} = 1, 5, 20$) and the variables considered for interaction with treatment ($p_{cons} = 1, 5, 20$).

	Overall	Non-Heavy Drinkers	Heavy Drinkers
Sex (%)			
Male	180 (36.0%)	130 (36.3%)	50 (35.2%)
Female	320 (64.0%)	228 (63.7%)	92 (64.8%)
Race/ethnicity (%)			
White	379 (75.8%)	253 (70.7%)	126 (88.7%)
Asian	50 (10.0%)	45 (12.6%)	5 (3.5%)
Black	15 (3.0%)	15 (4.2%)	0 (0.0%)
Hispanic/Latinx	26 (5.2%)	22 (6.1%)	4 (2.8%)
Other/Multi	30 (6.0%)	23 (6.4%)	7 (4.9%)
Intention to pledge greek (%)			
No	312 (62.4%)	242 (67.6%)	70 (49.3%)
Yes	57 (11.4%)	29 (8.1%)	28 (19.7%)
Undecided	131 (26.2%)	87 (24.3%)	44 (31.0%)
Drinking Norms: Mean percent drink (SD)	52.3 (20.1)	50.4 (20.1)	56.9 (19.2)
Drinking Norms: Mean num drinks typical (SD)	5.2 (6.9)	5 (7.2)	5.8 (6.1)
Drinking Norms: Mean most drinks typical (SD)	5.5 (3.5)	5.2 (3.4)	6.2 (3.6)
Drinking Norms: Mean percent binge (SD)	22.4 (17.2)	20 (15.8)	28.4 (19)
Drinking Intentions: Mean freq/month (SD)	2.3 (2.8)	1.5 (2.1)	4.4 (3.2)
Drinking Intentions: Mean num drinks (SD)	2 (1.8)	1.4 (1.5)	3.5 (1.7)
Drinking Intentions: Mean drunk/month (SD)	1.4 (2.2)	0.8 (1.6)	3 (2.8)
Self Monitoring: Binge drinking (SD)	1 (1.5)	0.1 (0.3)	2.9 (1.5)
Self Monitoring: High-intensity drinking (SD)	0.2 (0.8)	0 (0)	0.7 (1.4)

Table A2.4: Summary statistics of predictors in M-Bridge.