Journal of Statistical Research *https://doi.org/10.47302/jsr.2021550109*
2021, Vol. 55, No. 1, pp. 127-145 *ISSN 0256 - 422 X*

OPTIMIZING DYNAMIC TREATMENT REGIMES VIA THRESHOLD UTILITY ANALYSIS ON QUALITY ADJUSTED SURVIVAL

GEOFFREY S JOHNSON
Center for Observational and Real-world Evidence, Merck & Co. Inc., West Point, PA 19426, U.S.A.
Email: geoffrey.johnson@merck.com

ANDREW S. TOPP
AbbVie Inc., Chicago, Illinois 60064, U.S.A.
Email: andrew.topp@abbvie.com

ABDUS S. WAHED*
Department of Biostatistics, University of Pittsburgh, Pittsburgh, PA 15261, U.S.A.
Email: wahed@pitt.edu

SUMMARY

The analysis of quality adjusted lifetime adds an interesting wrinkle to the field of dynamic treatment regimes (DTRs), in that the optimal regime will not only depend on patient information (including treatments taken, intermediate outcomes, and other patient covariates), but it will also depend on information on the treatments themselves, e.g. monetary cost or toxicity. The focus of this paper is to investigate a form of Q-learning using estimating equations for the quality adjusted survival outcome. We use m-out-of-n bootstrap for inference and threshold utility analysis to show how the patient-specific optimal regime varies according to the treatment characteristics (e.g. cost, side effects). Methodologies developed are investigated through a simulation study and are demonstrated to construct optimal treatment regimes for the treatment of children's neuroblastoma.

Keywords and phrases: Counterfactuals; Dynamic treatment regime; Inverse probability weighting; m-out-of-n bootstrap; Potential outcomes; Quality adjusted lifetime; Q-learning; Survival analysis; Threshold utility analysis.

1 Introduction

Cost of treatment is often prohibitive for patients, especially those in developing countries, preventing them from following the optimal treatment regime for treating a condition. Additionally, in many cases adverse effects of the treatments in the optimal regime may prevent a patient from following the regime. This leads to a very important question for researchers to answer - should the cost of treatment and/or adverse effect of a treatment be considered while constructing optimal treatment regimes for a given condition? In what follows we develop a frame work for factoring the cost and

* Corresponding author

adverse event profiles into the analysis of dynamic treatment regimes in the form of quality adjustment. We define a quality function that maps the patient's health status and treatment characteristics (e.g. cost, adverse events) to an index ranging between zero and one, with zero indicating the worst health state (unaffordable treatment, life-threatening adverse effects, or death) and one indicating the best (normal health, affordable treatment, no adverse effects).

Clinical trials for cancer often measure a primary outcome and several secondary outcomes. The secondary outcomes may include, among others, measures of toxicity and adherence. Taken separately, these measures may sometimes lead to different optimal treatments. While one treatment may have the largest expected primary outcome, a second one may be less toxic, and a third might have the best adherence. To address this Gelber et al. (1989), Glasziou et al. (1990), Goldhirsch et al. (1989) and Korn (1993) considered quality adjusted lifetime to adjust the length of life based on its quality. In its simplest form, quality adjusted life assigns a utility weight, ranging from 0 to 1, to separate states of health. If there are k health states, then $U_i = \sum_{j=1}^{k} q_j s_{ji}$ is the quality adjusted lifetime (QAL) for the i^{th} patient, where $s_{1i}, .., s_{ki}$ are the times spent in each state, and $q_1, \ldots, q_k$ are the utility coefficients assigned to each of the health states. Note that the quality adjusted lifetime U_i is simply a fraction of total lifetime for patient i. More recently, Zhao and Tsiatis (1997, 1999, 2000, 2001) have provided consistent and efficient estimators, and provided hypothesis tests for distributional features of quality adjusted lifetime in the presence of right censoring. Wang and Zhao (2007) extended this work to the regression setting, using inverse weighting techniques to form consistent estimating equations for regression parameters.

Assuming larger outcomes are better, it is natural to search for the optimal regime, the one with the largest expected outcome. Murphy (2003), Robins (2004), and others pioneered the use of backwards induction in statistics via Q-learning and g-estimation to identify such optimal regimes. These algorithms work backwards in time by identifying at each stage which treatment has the largest expected outcome, and creating pseudo data for each subject by replacing his/her observed outcomes with the estimated optimal expected outcome at each stage, given prior observed outcomes and covariate information. The optimal treatment at each stage is the one with the largest expected value of this pseudo data. Huang et al. (2014) go on to discuss a version of Q-learning that emphasizes the observed data rather than deterministic modeled outcomes. It preserves the randomness of the data and better satisfies the consistency assumption frequently employed in the analysis of DTRs.

The goal of this manuscript is to develop an optimal dynamic treatment regime to maximize quality adjusted lifetime by using a Q-learning-type approach discussed in Huang et al. (2014). This method will be operationalized using the estimating equations of Wang and Zhao (2007), and a threshold utility analysis will be used to show how the subject-specific optimal DTR not only depends on patient history and intermediate outcomes, but also on quality of life, monetary cost, and other factors during each treatment. Though we optimize quality adjusted lifetime, we provide suggestions on how the quality adjustment can be used for any continuous outcome via Q-learning. We use a simulation study to evaluate these methods, and then apply them to COG study A3891 concerning 379 children receiving treatment for high-risk neuroblastoma (Matthay et al., 1999).

2 Setup

2.1 Quality-adjusted lifetime

Describe the health history for the i^{th} patient with a continuous time stochastic process $\{V_i(t), t \geq 0\}$. $V_i(t)$ maps to the space of health states $S = \{0, 1, 2, \ldots, m\}$, where the state '0' corresponds to the absorbing state of death. Denote the health history up to time t by $V_i^H(t) = \{V_i(s) : s \leq t\}$. Let $V_i(s) = 0$ imply that $V_i(t) = 0$ for $t \geq s$. Let T_i denote the survival time for patient i. Naturally, $V_i(t) = 0$ for $t \geq T_i$. Then we see that $T_i = \inf\{t : V_i(t) = 0\}$. Let $q()$ be a quality of life function mapping $V_i(t)$ to $[0, 1]$, with $q(0) \equiv 0$. The quality adjusted lifetime for the i^{th} patient is defined as $Q(T_i) = \int_0^{T_i} q\{V_i(t)\}dt$.

In the presence of non-informative right censoring, one might consider the restricted survival time where total follow-up time is limited to L, where L is some value less than the maximum survival time for all patients. Therefore, the survival time for all patients will be truncated at L, $T^L = \min(T, L)$. For ease of notation, we will drop the superscript and simply use T. We will denote the i^{th} patient's censoring time by C_i, and the survival distribution of C by $K(t) = P(C > t)$. Define $U_i = \min(T_i, C_i)$ and $\Delta_i = I(T_i \leq C_i)$, respectively, to be the observed time to event (death or censoring), and the death indicator. Then $Q(U_i) = \int_0^{U_i} q\{V_i(t)\}dt$ represents the quality adjusted time to event for the i^{th} patient.

In this construction the quality function q is not patient specific (does not have a subscript i), and was assumed known. One view is that q exists at the population level. This means that every patient in the analysis, and all of the patients they represent, experience the same quality of life when in a particular health state. This allows for a threshold utility analysis, described in detail in Section 3.2, where quality adjusted lifetime (or a function of it) is considered over the entire range of possible values of q, to examine how the value of q affects the estimation of quality adjusted lifetime. As a convention we will take $Q(s, t)$ to refer to $\int_s^t q\{V(u)\}du$ and $Q(t)$ to refer to $\int_0^t q\{V(u)\}du$.

For example, consider a discrete-state health history process $V_i(t)$ with three states: treatment, response (well-being), and death. Suppose each of these states are mapped to [0,1] as $q\{V_i(t)\} = q_a I\{t \leq T_i^R\} + 1I\{T_i^R < t < T_i\} + 0I\{t > T_i\}$. Such a mapping may be reasonable as the quality is the least (zero) after death, one when healthy, and a constant, q_a, between zero and one when being treated due to toxicity related complications and/or monetary cost from receiving treatment $A = a$. Here, time from beginning of treatment to response is denoted by T_i^R. Under this scenario, $Q(T_i) = \int_0^{T_i^R} q_a dt + \int_{T_i^R}^{T_i} 1dt = T_i - (1 - q_a)T_i^R$. If the patient undergoes a maintenance treatment immediately after responding, and remains on maintenance treatment $B = b$ until death, $Q(T_i)$ could be written as $Q(T_i) = \int_0^{T_i^R} q_a dt + \int_{T_i^R}^{T_i} q_b dt = q_b T_i - (q_b - q_a)T_i^R$, where the constant q_b reflects the utility weight of treatment $B = b$ for toxicity, monetary cost, and other factors.

Since quality adjusted lifetime is the area under $q\{V_i(t)\}$ over the health states from 0 to T, for any function $q\{V_i(t)\}$ there exists a constant function in each health state that results in the same area, and produces the same quality adjusted lifetime. Not coincidentally, the example above has the health states of each patient correspond to the sequence of treatments received. When estimating mean quality adjusted lifetime in such settings, the utility weights q_a and q_b factor out, producing

$E[Q(T_i)] = q_a E[T_i^R] + q_b E[T_i - T_i^R]$. When viewed in this way, not only can the utility weights be seen as population constants, they can alternatively be seen as adjustments to the expected utility of each treatment for the prospective patient, depending on his or her aversion to each treatment, with each prospective patient potentially having different values of the utility weights. Such an interpretation of the utility weights offers even more motivation for a threshold utility analysis.

For drawing inference on quality adjusted lifetime, the survival function of quality adjusted lifetime may be used the same way as the survival function of overall survival. In the presence of non-informative censoring one might naturally turn to the Kaplan-Meier estimator, to estimate $S(t) = P\big(Q(T_i) > t\big)$, but Gelber et al. (1989) and Pradhan & Dewanji (2009) showed that this can result in biased estimation because the quality adjustment induces a dependence between the survival times and censoring times. Zhao & Tsiatis (1997) offer an inverse-probability weighted estimator, similar to that proposed by Robins & Rotnitzky (1992) and Robins et al. (1994) , $\hat{S}(t)^{cen} = \frac{1}{n}\sum_{i=1}^{n} \frac{\Delta_i}{\hat{K}(U_i)} I[Q(U_i) > t]$, where $\hat{K}(U_i)$ is the Kaplan-Meier estimator for the censoring random variable evaluated at U_i, and Δ_i and $\hat{K}(U_i)$ can depend on t to improve efficiency. Zhao & Tsiatis (1999) improve the efficiency of their estimator by incorporating each patient's health history. In Zhao & Tsiatis (2000) they used the same principles to estimate the mean quality adjusted lifetime.

Wang & Zhao (2007) extended this work to the regression setting by constructing consistent estimating equations for mean quality adjusted lifetime in the presence of censoring, yielding $\mathcal{U}_n(\beta) = \sum_{i=1}^{n} \frac{\Delta_i}{\hat{K}(U_i)} h(X_i)\{Q(U_i) - g(\beta, X_i)\} = 0$, where X_i denotes a $(p+1) \times 1$ vector of covariates associated with patient i, with the first covariate being the constant 1, $h(X_i)$ is a $(p+1) \times 1$ vector of functions of X_i, β is a $(p+1) \times 1$ vector of parameters, and $g(\beta, X_i) = E[Q(T_i)|X_i]$. The estimator for β solving $\mathcal{U}_n(\beta)$ will be used to operationalize our search for the optimal dynamic treatment regime, described in Section 3.

2.2 Dynamic treatment regimes and corresponding terminology

Consider a two-stage sequential multiple assignment randomized trial (SMART) design where patients are randomized to one of two induction therapies, $\mathcal{A} = \{a_1, a_2\}$. Patients may be resistant to their initial treatment, or they may respond. For each of the induction therapies, if treatment response is observed, patients are further randomized to one of two maintenance treatments, $\mathcal{B} = \{b_1, b_2\}$. This design allows for inference on four DTRs that might be carried out in clinical practice, namely, $d(A_i = a_j; B_i = b_k),\ j, k = 1, 2$, where $d(A_i; B_i)$ stands for "Treat with A_i, if the patient responds, treat with B_i." Our goal is to find the optimal treatment regime among these that maximizes expected quality adjusted lifetime.

Let $G_i^H(t)$ denote all information collected on patient i prior to time t. Some or all of the information in $G_i^H(t)$, for example serum biomarker levels, responses to questionnaires, or tumor size, is used to define $V_i^H(t)$, which then defines R_i and T_i^R, the observed response indicator and the observed time to response given $R_i = 1$, respectively. $G_i^H(t)$ may include additional patient information not used to define $V_i^H(t)$. Then, introducing further indicators for first and second

stage treatment, the observed data for the i^{th} patient in the presence of censoring is written as

$$D_i^{\delta} = \Big(Z_{1i}^{(A)},\ Z_{2i}^{(A)},\ R_i,\ R_i T_i^R,\ R_i Z_{1i}^{(B)},\ R_i Z_{2i}^{(B)},\ U_i,\ \Delta_i,\ V_i^H(U_i),\ G_i^H(U_i) \Big),$$

where $Z_{ji}^{(A)}$=1 if patient i received the j^{th} induction therapy, $Z_{ji}^{(A)}$=0 otherwise, and $Z_{ki}^{(B)}$ denotes the b_k treatment assignment indicator $I\{B = b_k\}$, defined only if R_i=1. Note that $Z_{2i}^{(A)} = 1 - Z_{1i}^{(A)}$ and $Z_{2i}^{(B)} = 1 - Z_{1i}^{(B)}$, but we explicitly define them to facilitate the use of summation.

By design, treatments are assigned independently of prognosis or any observed data measured prior to the second stage. This condition is often referred to as no unmeasured confounders or sequential randomization assumption. This 'no unmeasured confounders' condition holds even if the second-stage randomization probabilities depend on the first-stage treatment assignments.

3 Optimization of Dynamic Treatment Regimes on Quality Adjusted Survival

3.1 Optimization

Following the work of Murphy (2003), Robins (2004), and Huang et al. (2014), we describe a backward induction method to identify the optimal dynamic treatment regime, using mean quality adjusted survival time as the criterion of optimality. From the reinforcement learning literature in the field of DTRs, the typical $\mathcal{Q}$-functions for two stages of our SMART design, assuming no unmeasured confounders, would be

$$\mathcal{Q}_B\Big(A_i = a_j, G_i^H(T_i^R), B_i = b_k \Big) = E\Big[Q(T_i^R, T_i) \Big| A_i = a_j, R_i = 1, G_i^H(T_i^R), B_i = b_k \Big]$$
$$\mathcal{Q}_A\Big(G_i^H(0), A_i = a_j \Big) = E\Big[H_i^{(A)} \Big| G_i^H(0), A_i = a_j \Big],$$

where

$$H_i^{(A)} = \begin{cases} Q(T_i^R) + \underset{b_k}{max}\ \mathcal{Q}_B\Big(A_i, G_i^H(T_i^R), B_i = b_k \Big), & \text{if } R_i = 1 \\ Q(T_i), & \text{if } R_i = 0. \end{cases}$$

Then, the optimal stage 1 treatment given baseline information is

$$A_i^{opt} = \underset{a_k}{argmax}\ E\Big[H_i^{(A)} \Big| G_i^H(0), A_i = a_j \Big],$$

and the optimal stage 2 treatment given stage 1 treatment assignment and information up to stage 2 is

$$B_i^{opt} = \underset{b_k}{argmax}\ E\Big[Q(T_i^R, T_i) \Big| A_i = a_j, R_i = 1, G_i^H(T_i^R), B_i = b_k \Big].$$

Below we walk through the backwards induction used to estimate the optimal treatment at each stage, with a different $H_i^{(A)}$ shown in Huang et al. (2014) that we use in our simulation and application.

We start with the second stage (include only those patients who responded $R_i = 1$). Under assumptions described in Section 2.2, the quality adjusted time from maintenance therapy to death for those patients who responded is $Q(T_i^R, T_i) = \int_{T_i^R}^{T_i} q\{V_i(t)\}dt$, so that

$$\gamma_B \equiv E\Big[\, Q(T_i^R, T_i) \,\Big| A_i = a_j, B_i = b_1, R_i = 1, G_i^H(T_i^R)\Big] - E\Big[\, Q(T_i^R, T_i) \,\Big| A_i = a_j, B_i = b_2, R_i = 1, G_i^H(T_i^R)\Big]$$

is the difference in expected stage 2 outcomes, given prior information. We assume the following linear model for $\mathcal{Q}_B\big(A_i, B_i, R_i = 1, \bar{X}_{Bi}, \boldsymbol{\beta}_B, \boldsymbol{\alpha}_B\big)$

$$E\Big[Q(T_i^R, T_i)\Big| A_i, B_i, R_i = 1, \bar{X}_{Bi}, \boldsymbol{\beta}_B, \boldsymbol{\alpha}_B\Big] = \bar{X}_{Bi}^{'}\boldsymbol{\beta}_B + Z_{1i}^{(B)}\bar{X}_{Bi}^{'}\boldsymbol{\alpha}_B, \tag{3.1}$$

where $\bar{X}_{Bi}$ are the first stage treatment assignment indicators and covariates from $G_i^H(T_i^R)$, and includes an element equal to 1 corresponding to an intercept term, which implies that $\gamma_B = \bar{X}_{Bi}^{'}\boldsymbol{\alpha}_B$, and the estimated optimal stage two treatment given stage 1 treatment assignment and patient information up to stage 2 is

$$\hat{B}^{opt}(\bar{X}_{Bi}) = \underset{b_k}{\textit{argmax}}\ \hat{E}\Big[\, Q(T_i^R, T_i) \,\Big| A_i = a_j, R_i = 1, B_i = b_k, \bar{X}_{Bi}, \boldsymbol{\beta}_B, \boldsymbol{\alpha}_B\Big].$$

If γ_B is positive then b_1 is the optimal stage 2 treatment, otherwise, b_2 is optimal. Using fitted models corresponding to equation (3.1) we can estimate the optimal quality adjusted time from maintenance therapy to death as

$$H_i^{(B)}(\hat{\boldsymbol{\alpha}}_B) \equiv \begin{cases} Q(T_i^R, T_i) + \big|\bar{X}_{Bi}^{'}\hat{\boldsymbol{\alpha}}_B\big|, & \text{if } B_i = b_k,\ \hat{B}_i^{opt} \neq b_k \\ Q(T_i^R, T_i), & \text{if } B_i = b_k,\ \hat{B}_i^{opt} = b_k. \end{cases}$$

Moving to the first stage, under assumptions described in Section 2.2 the quality adjusted survival time with observed stage one treatment and the estimated optimal stage two treatment can be written as

$$H_i^{(A)}(\hat{\boldsymbol{\alpha}}_B) = \begin{cases} Q(T_i^R) + H_i^{(B)}(\hat{\boldsymbol{\alpha}}_B), & \text{if } R_i = 1 \\ Q(T_i), & \text{if } R_i = 0. \end{cases}$$

Let X_{Ai} denote important covariates in $G_i^H(0)$ predictive of residual survival. We assume the following linear model for $\mathcal{Q}_A\big(A_i, X_{Ai}, \boldsymbol{\beta}_A, \boldsymbol{\alpha}_A\big)$

$$E\Big[H_i^{(A)}(\hat{\boldsymbol{\alpha}}_B)\Big| A_i, X_{Ai}, \boldsymbol{\beta}_A, \boldsymbol{\alpha}_A\Big] = X_{Ai}^{'}\boldsymbol{\beta}_A + Z_{1i}^{(A)}X_{Ai}^{'}\boldsymbol{\alpha}_A,$$

where X_{Ai} includes an element equal to 1 corresponding to an intercept term, which implies that

$$\gamma_A \equiv E\Big[H_i^{(A)}(\hat{\boldsymbol{\alpha}}_B)\Big| A_i = a_1, G_i^H(0)\Big] - E\Big[H_i^{(A)}(\hat{\boldsymbol{\alpha}}_B)\Big| A_i = a_2, G_i^H(0)\Big] = X_{Ai}^{'}\boldsymbol{\alpha}_A$$

is the difference in expected outcomes at stage 1, given that each patient received his estimated optimal stage 2 treatment. The estimated optimal stage one treatment is

$$\hat{A}^{opt}(X_{Ai}) = \underset{a_j}{\textit{argmax}}\ \hat{E}\Big[H_i^{(A)}(\hat{\boldsymbol{\alpha}}_B)\Big|A_i = a_j, X_{Ai}, \boldsymbol{\beta}_A, \boldsymbol{\alpha}_A\Big].$$

If γ_A is positive then a_1 is the optimal stage 1 treatment, otherwise, a_2 is optimal. Thus if one could estimate the quantities γ_A or γ_B, or equivalently, the parameters α_B and α_A, the optimal treatment regime could be constructed given the q function for each specific stage.

To estimate these parameters, the simple weighted regression models described in Section 2.1 by Wang & Zhao (2007) can be used. Explicitly, for stage 2 we solve the estimating equation

$$\mathcal{U}_n^{(B)}(\boldsymbol{\beta}_B, \boldsymbol{\alpha}_B) = \sum_{i=1}^{n} \frac{\Delta_i}{\hat{K}(U_i)} R_i \begin{bmatrix} \bar{X}_{Bi} \\ Z_{1i}^{(B)} \bar{X}_{Bi} \end{bmatrix} \Big\{ Q(T_i^R, U_i) - \bar{X}_{Bi}^{'} \boldsymbol{\beta}_B - Z_{1i}^{(B)} \bar{X}_{Bi}^{'} \boldsymbol{\alpha}_B \Big\} = 0,$$

for $\boldsymbol{\beta}_B$ and $\boldsymbol{\alpha}_B$. Similarly, for stage 1 we solve

$$\mathcal{U}_n^{(A)}(\boldsymbol{\beta}_A, \boldsymbol{\alpha}_A) = \sum_{i=1}^{n} \frac{\Delta_i}{\hat{K}(U_i)} \begin{bmatrix} X_{Ai} \\ Z_{1i}^{(A)} X_{Ai} \end{bmatrix} \Big\{ H_i^{(A)}(\hat{\boldsymbol{\alpha}}_B) - X_{Ai}^{'} \boldsymbol{\beta}_A - Z_{1i}^{(A)} X_{Ai}^{'} \boldsymbol{\alpha}_A \Big\} = 0$$

to obtain estimates of $\boldsymbol{\beta}_A$ and $\boldsymbol{\alpha}_A$.

3.2 Threshold utility analysis

Glasziou et al. (1990) perform a threshold utility analysis when studying the effects of adjuvant chemotherapy on quality adjusted lifetime in patients with early breast cancer. Each patient's survival time is quality adjusted based on periods of toxicity of treatment and relapse of disease. These quality weights, ranging from 0 to 1, are plotted against each other and the regions where each treatment is favored are identified via lines (planes) of indifference. This results in a type of sensitivity analysis, allowing one to see all possible treatment decisions drawn depending on the quality weights. In our DTR setting, a patient's course of treatment often depends on his/her state of health, be it response to treatment or relapse of the disease, so that his/her health states correspond to the stages of the DTR. In our approach, each patient's survival time will be weighted according to treatment received, allowing a threshold utility analysis among treatments, and ultimately among regimes.

Optimal decision rules for first and second stage treatments developed in Section 3 are not only a function of the observed data (patient level information), but also of the quality of life function q. In our development in the previous section, we assumed that this q function was known, and we offered two interpretations of its meaning. Rather than performing a single analysis with one q function, a sensitivity analysis can be performed using a variety of reasonable q functions to determine for which functions of q, if any, the choice of optimal regime changes. In the special case of constant q functions, q can be varied from 0 to 1, and a threshold utility plane can be plotted. This

is of importance, since depending on the values of the q function, there may be different optimal treatment regimes.

To be explicit, consider quality adjusting each patient's survival time as

$$Q(T_i) = \begin{cases} T_i q_{a_j}, & \text{if } A_i = a_j,\ R_i = 0, \\ T_i^R q_{a_j} + (T_i - T_i^R) q_{b_k}, & \text{if } A_i = a_j,\ R_i = 1,\ B_i = b_k, \end{cases}$$

for $j = 1, 2$, $k = 1, 2$ where $q_{a_j}, q_{b_k} \in [0, 1]$. For those who responded ($R_i = 1$) and received maintenance treatment, the quality weights $q_{b_1}, q_{b_2} \in [0, 1]$ can be plotted against each other on the x and y axes, with

$$\begin{aligned} \hat{\gamma}_B &= \hat{E}\Big[(T_i - T_i^R) q_{b_1} \Big| A_i = a_j, B_i = b_1, R_i = 1, \bar{X}_{Bi}, \boldsymbol{\beta}_B, \boldsymbol{\alpha}_B \Big] \\ &\quad - \hat{E}\Big[(T_i - T_i^R) q_{b_2} \Big| A_i = a_j, B_i = b_2, R_i = 1, \bar{X}_{Bi}, \boldsymbol{\beta}_B, \boldsymbol{\alpha}_B \Big] \\ &= \bar{X}_{Bi}^{'} \hat{\boldsymbol{\alpha}}_B \end{aligned}$$

from Section 3 plotted on the z axis. This forms a two-dimensional plane in a three-dimensional space. When quality adjusting in this way, the utility weights q_{b_1} and q_{b_2} factor out of the expectations and can be viewed as adjustments to the expected utility of each stage two treatment for the prospective patient, depending on his or her aversion to each treatment. The line where $\hat{\gamma}_B = 0$ is the estimated threshold at which the expected utility of b_1 and b_2 are equal, where the prospective patient is indifferent when choosing between stage two treatments.

Similarly, for those who received an induction treatment, the quality weights $q_{a_1}, q_{a_2} \in [0, 1]$ can be plotted against each other on the x and y axes, with

$$\hat{\gamma}_A = \hat{E}\Big[H_i^{(A)}(\hat{\boldsymbol{\alpha}}_B) \Big| A_i = a_1, X_{A_i}, \boldsymbol{\beta}_A, \boldsymbol{\alpha}_A \Big] - \hat{E}\Big[H_i^{(A)}(\hat{\boldsymbol{\alpha}}_B) \Big| A_i = a_2, X_{A_i}, \boldsymbol{\beta}_A, \boldsymbol{\alpha}_A \Big] = X_{Ai}^{'} \hat{\boldsymbol{\alpha}}_A$$

from Section 3 plotted on the z axis, where

$$H_i^{(A)}(\hat{\boldsymbol{\alpha}}_B) = \begin{cases} T_i^R q_{a_j} + H_i^{(B)}(\hat{\boldsymbol{\alpha}}_B), & \text{if } A_i = a_j,\ R_i = 1 \\ T_i q_{a_j}, & \text{if } A_i = a_j,\ R_i = 0 \end{cases}$$

$$H_i^{(B)}(\hat{\boldsymbol{\alpha}}_B) = \begin{cases} (T_i - T_i^R) q_{b_k} + \Big| \bar{X}_{Bi}^{'} \hat{\boldsymbol{\alpha}}_B \Big|, & \text{if } B_i = b_k,\ \hat{B}_i^{opt} \neq b_k \\ (T_i - T_i^R) q_{b_k}, & \text{if } B_i = b_k,\ \hat{B}_i^{opt} = b_k. \end{cases}$$

The line where $\hat{\gamma}_A = 0$ is the estimated threshold at which the prospective patient is indifferent when choosing between a_1 and a_2.

3.3 Inference

Robins (2004), Chakraborty et al. (2009), and Laber et al. (2014) are quick to point out that the estimators derived from Q-learning have non-regular limiting distributions, because the estimated

stage 1 pseudo data (and hence the estimated stage 1 model parameters) are a non-smooth (non-differentiable at $\bar{X}'_{Bi}\hat{\boldsymbol{\alpha}}_B$=0) function of $\hat{\boldsymbol{\alpha}}_B$. This motivated Chakraborty et al. (2013) to discuss the m-out-of-n bootstrap in the context of DTRs, in place of standard large-sample inference methods. The m-out-of-n bootstrap technique essentially smooths the empirical distribution function, with more smoothing corresponding to smaller values of m, the resample size, by allowing the empirical distribution function to tend to its limiting distribution at a faster rate than the bootstrap empirical distribution tends to the empirical distribution. We use this technique to create confidence regions in the threshold utility analysis, identifying regions of indifference and strong acceptance when choosing between stage 1 and stage 2 treatments. While Chakraborty et al. (2013) provide several data driven methods for determining the smaller resample size m, we find a suitable m through simulation and apply this same m in the analysis of real data.

4 Simulation Study

In this section we conduct a simulation experiment to evaluate the optimization of dynamic treatment regimes for quality adjusted lifetime described in Section 3. Similar to the COG study A3891 that will be presented later in Section 5, we consider a 2-stage SMART design.

We generated 5,000 simulations with sample size n=1000. Patients are randomized to one of two induction therapies with probability one-half, and the probability of non-response for each induction therapy is the same, 0.55. Those who respond to induction therapy are further re-randomized with probability one-half to one of two maintenance therapies. Sojourn times to response and/or death were generated from various exponential distributions.

Table 1 shows the coverage probabilities over the 5,000 simulations for 90% point-wise bootstrap confidence intervals for the estimated difference in mean quality adjusted lifetime between Stage 1 and Stage 2 treatments (γ_A and γ_B, respectively) when searching for the optimal treatment regime using the simple weighted estimating equations from Section 2.1. The 5^{th} and 95^{th} percentiles of the bootstrapped sampling distributions are used to create the confidence intervals. Stage 1 coverage probabilities are estimated at q_{b_1}=0.8 and q_{b_2}=0.6. A variety of re-sample sizes were considered for the stage 1 m-out-of-n bootstrap, and m=850 produced confidence intervals maintaining the nominal coverage probability. The coverage probabilities for the 90% confidence intervals are close to the nominal level for utility weights that are away from zero. This makes sense, as a value of q close to zero greatly reduces the variability in the data, making it difficult to estimate the respective quantities. For some combinations of q_{b_1} and q_{b_2} the estimated stage 2 coverage probabilities are below the nominal level. Although no irregularity issues exist for the stage 2 estimates, the m-out-of-n bootstrap was still employed to improve the coverage probabilities, with m=800. Using the m-out-of-n bootstrap, the stage 1 coverage probabilities for the difference in mean quality adjusted lifetime are well maintained. Similar simulations were performed with a sample size of n=300 and survival times close to that of the COG study A3891. This gives us an idea of an appropriate choice of m. We found that m=240 and m=255 worked well for maintaining the nominal coverage probabilities for stage 2 and stage 1, respectively. Similar results are shown in Table 2 to determine m for a sample size of $n = 200$.

Table 1: Coverage probabilities of 90% point-wise bootstrap confidence intervals (500 bootstrap samples), from simulated data with 5000 replicates of n=1000, stage 2 m=800, stage 1 m=850.

$A = a_1 : B = b_1$ vs $B = b_2$

$q_{b_1} \backslash q_{b_2}$	0.0	0.2	0.4	0.6	0.8	1.0
0.00		0.822	0.822	0.822	0.822	0.822
0.20	0.904	0.916	0.831	0.817	0.812	0.813
0.40	0.904	0.927	0.916	0.872	0.831	0.819
0.60	0.904	0.919	0.930	0.916	0.886	0.855
0.80	0.904	0.915	0.927	0.928	0.916	0.894
1.00	0.904	0.913	0.920	0.928	0.927	0.916

$A = a_2 : B = b_1$ vs $B = b_2$

$q_{b_1} \backslash q_{b_2}$	0.0	0.2	0.4	0.6	0.8	1.0
0.00		0.929	0.929	0.929	0.929	0.929
0.20	0.919	0.901	0.919	0.928	0.929	0.930
0.40	0.919	0.909	0.901	0.912	0.919	0.923
0.60	0.919	0.913	0.903	0.901	0.904	0.916
0.80	0.919	0.914	0.909	0.902	0.901	0.903
1.00	0.919	0.916	0.910	0.907	0.901	0.901

$A = a_1$ vs $A = a_2$

$q_{a_1} \backslash q_{a_2}$	0.0	0.2	0.4	0.6	0.8	1.0
0.00	0.913	0.913	0.915	915	0.916	0.911
0.20	0.912	0.912	0.913	0.913	0.916	0.916
0.40	0.910	0.911	0.912	0.914	0.913	0.917
0.60	0.907	0.909	0.912	0.914	0.917	0.917
0.80	0.909	0.909	0.912	0.914	0.915	0.918
1.00	0.907	0.907	0.911	0.913	0.916	0.918

Table 2: Coverage probabilities of 90% point-wise bootstrap confidence intervals (500 bootstrap samples), from simulated data with 5000 replicates of n=2000, stage 2 m=1600, stage 1 m=1700.

$A = a_1 : B = b_1$ vs $B = b_2$

$q_{b_1} \backslash q_{b_2}$	0.0	0.2	0.4	0.6	0.8	1.0
0.00		0.702	0.702	0.702	0.702	0.702
0.20	0.859	0.924	0.734	0.690	0.687	0.687
0.40	0.859	0.914	0.924	0.813	0.734	0.702
0.60	0.859	0.897	0.927	0.924	0.857	0.780
0.80	0.859	0.884	0.914	0.932	0.924	0.878
1.00	0.859	0.878	0.904	0.923	0.934	0.924

$A = a_2 : B = b_1$ vs $B = b_2$

$q_{b_1} \backslash q_{b_2}$	0.0	0.2	0.4	0.6	0.8	1.0
0.00		0.933	0.933	0.933	0.933	0.933
0.20	0.914	0.877	0.915	0.927	0.930	0.933
0.40	0.914	0.896	0.877	0.901	0.915	0.923
0.60	0.914	0.901	0.890	0.877	0.891	0.910
0.80	0.914	0.907	0.896	0.885	0.877	0.886
1.00	0.914	0.907	0.889	0.892	0.884	0.877

$A = a_1$ vs $A = a_2$

$q_{a_1} \backslash q_{a_2}$	0.0	0.2	0.4	0.6	0.8	1.0
0.00	0.915	0.918	0.918	0.915	0.912	0.908
0.20	0.910	0.913	0.916	0.917	0.915	0.911
0.40	0.905	0.910	0.911	0.914	0.914	0.915
0.60	0.905	0.908	0.911	0.911	0.916	0.914
0.80	0.903	0.904	0.908	0.912	0.913	0.915
1.00	0.900	0.901	0.905	0.910	0.912	0.914

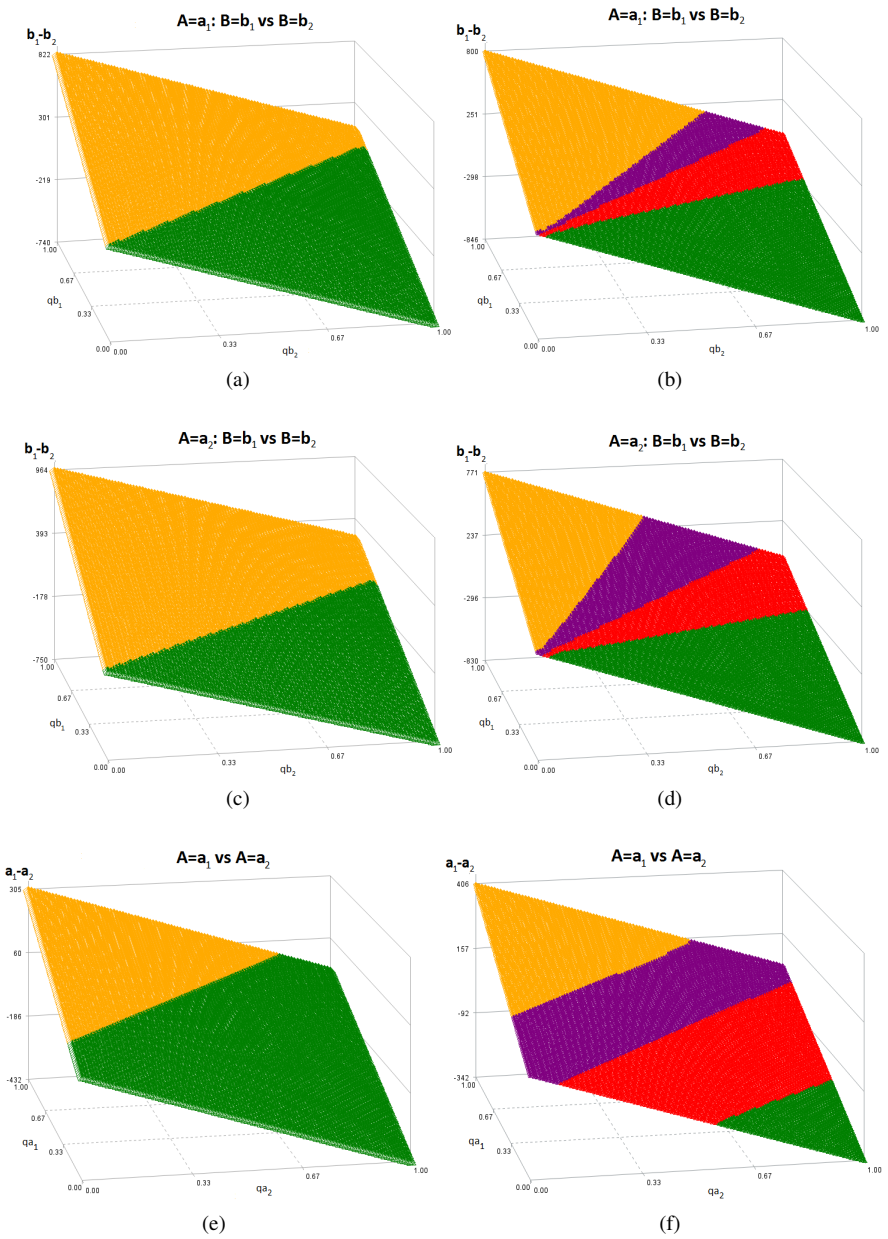

Figure 1: True (left column) and estimated (right column) threshold utility planes for the simulated scenario. For each combination of q_{b_1} and q_{b_2}, or q_{a_1} and q_{a_2}, the estimated difference in mean quality adjusted lifetime is plotted. The yellow and green represent the region of strong acceptance for b_1 and b_2, or a_1 and a_2, respectively. The purple and red near the center of the plane have 90% point-wise bootstrap confidence intervals that cover zero and represent the region of indifference when choosing between b_1 and b_2, or a_1 and a_2.

Figure 1 shows the true (left column) and estimated (right column) threshold utility planes for the simulated scenario with n=300. The estimated threshold utility planes are for a single simulated data set. For each combination of q_{b_1} and q_{b_2}, or q_{a_1} and q_{a_2}, the estimated difference in mean quality adjusted lifetime is plotted. The yellow and green represent the region of strong acceptance for choosing between b_1 and b_2, or a_1 and a_2, respectively. The purple and red near the center of the plane have 90% point-wise bootstrap confidence intervals that cover zero and represent the region of indifference when choosing between b_1 and b_2, or a_1 and a_2. We see that for the estimated threshold utility planes, the estimated line of indifference does not correspond exactly with the true line of indifference, yet the 90% confidence region does contain the true line. These threshold utility planes allow us to visualize how the optimal regime changes depending on the values of q_{b_1}, q_{b_2}, q_{a_1}, and q_{a_2}. For example, assume that the threshold utility planes presented on the right panel of Figure 1 are the planes computed from the observed data. If for these treatments q_{a_1}=0.8, q_{a_2}=0.5, q_{b_1}=0.7, and q_{b_2}=0.5, then the estimated optimal regime is $d(A = a_1; B = b_1)$. However, if q_{a_1}=0.3, q_{a_2}=0.8, q_{b_1}=0.4, and q_{b_2}=0.6, the estimated optimal regime would be $d(A = a_2; B = b_2)$.

5 Application with Threshold Utility Analysis

In this section we apply the optimization methods discussed previously to the COG study A3891 concerning 379 children ages 6-months to 17 years old receiving treatment for high-risk neuroblastoma. All 379 patients were to receive five cycles of chemotherapy before beginning their induction treatment. Of these, 189 patients were randomized to receive continued chemotherapy (three additional cycles), $A = a_1$, and the remaining 190 were randomized to receive bone marrow transplantation, $A = a_2$. After completing the induction therapy, 203 patients were deemed responders (those for whom the disease did not progress) and consented to further randomization to receive six cycles of 13-cis-retinoic acid (160 mg per square meter per day for 14 consecutive days), $B = b_1$, or no further therapy, $B = b_2$. Survival time was truncated to 2452 days, since this was the largest observed death time in the study.

In what follows we assume the role of the prospective patient, considering only quality of life as affected by toxicity of treatment when choosing between treatments. Each of the therapies in this study comes with its own side effects. Following a cohort of lung cancer patients undergoing chemotherapy, Winter et al. (2013) measured quality of life using the EORTC QLQ-C30 questionaire Aaronson et al. (1993) as the patients completed multiple courses of chemotherapy. In the analysis by Winter et al. (2013) , the highest average global quality of life measure (ranging 0 to 100) over multiple courses of chemotherapy was 57. We rescaled these scores between 0 and 1 to have the quality of life weight of those undergoing chemotherapy vary between 0.5 to 0.6. In the case of bone marrow transplant, Felder et al. (2006) analyze the health related quality of life of 68 pediatric patients aged 4 to 18 years old receiving allogeneic bone marrow or stem cell transplantation in a 5-year prospective study using The Pediatric Quality of Life Inventory(PedsQL) and The Health Utilities Index Mark2 + 3(HUI2/3). It is reasonable to interpret these scores as quality weights, indicating that those undergoing bone marrow transplantation have a quality of life near 0.7. Hong et al. (1986) studied the use of 13-cis-retinoic acid in 44 patients with oral leukoplakia,

and found that cheilitis, erythema, and dry skin were most common. Based on the symptoms, mean survival time for patients on 13-cis-retinoic acid could reasonably be quality adjusted by 0.9.

Figure 2 (top row) shows the estimated stage 2 threshold utility planes - the estimated mean survival time for those on 13-cis-retinoic acid minus the estimated mean survival time for those on no further treatment. The yellow and green represent the region of strong acceptance for 13-cis-retinoic acid and no further therapy, respectively. The purple and red near the center of the plane give point estimates that favor 13-cis-retinoic acid and no further therapy, respectively, but the 90% point-wise bootstrap confidence intervals cover zero and represent the region of indifference when choosing between 13-cis-retinoic acid and no further therapy.

When the survival times for stage 2 treatments are both given a weight of 1 (no quality adjustment), those who received no further therapy had larger survival times than those who received 13-cis-retinoic acid, following continued chemotherapy; following bone marrow transplant, those who received 13-cis-retinoic acid had, on average, larger survival times than those who received no further therapy. It should be noted, though, that both of these point estimates fall within the m-out-of-n bootstrap indifference regions (the red and purple shaded areas), suggesting there is no statistically significant difference between the stage 2 treatments following either stage 1 treatment.

As one would begin to lower either q_{b_1} or q_{b_2} towards 0, while holding the other fixed, we see that the estimated difference in mean quality adjusted survival time falls in the region of statistical significance, where one stage 2 treatment truly out performs the other, given the stage 1 treatment. For q_{b_1}=0.9 and q_{b_2}=1, the stage 2 quality of life weights considered earlier for this study, the point estimate for the optimal stage 2 treatment falls in the same region as that for q_{b_1}=1 and q_{b_2}=1 described above and yields 13-cis-retinoic acid for those following bone marrow transplantation, and no further therapy for those following continued chemotherapy. If q_{b_1} is lower than 0.9, the optimal stage 2 treatment would be no further therapy for both induction therapies.

Figure 2 (bottom row) also shows the estimated stage 1 threshold utility plane - the estimated mean survival time for those on continuation chemotherapy minus the estimated mean survival time for those who received a bone marrow transplant. This figure is generated using pseudo data where responders at stage 1 are assumed to take their optimal stage 2 treatment, and their remaining survival time is estimated using the methods from Section 3, with q_{b_1}=0.9 for 13-cis-retinoic acid and q_{b_2}=1 for no further treatment. At q_{a_1}=0.5 and q_{a_2}=0.7 the optimal stage 1 treatment is bone marrow transplant, and the point estimate falls within the strong acceptance region, meaning the 90% point-wise bootstrap confidence interval for the difference in mean survival time between continued chemotherapy and bone marrow transplant does not contain zero. Therefore, with q_{a_1}=0.5, q_{a_2}=0.7, q_{b_1}=0.9, and q_{b_2}=1, the optimal regime is to first treat with bone marrow transplantation and, if a response is observed, treat with 13-cis-retinoic acid.

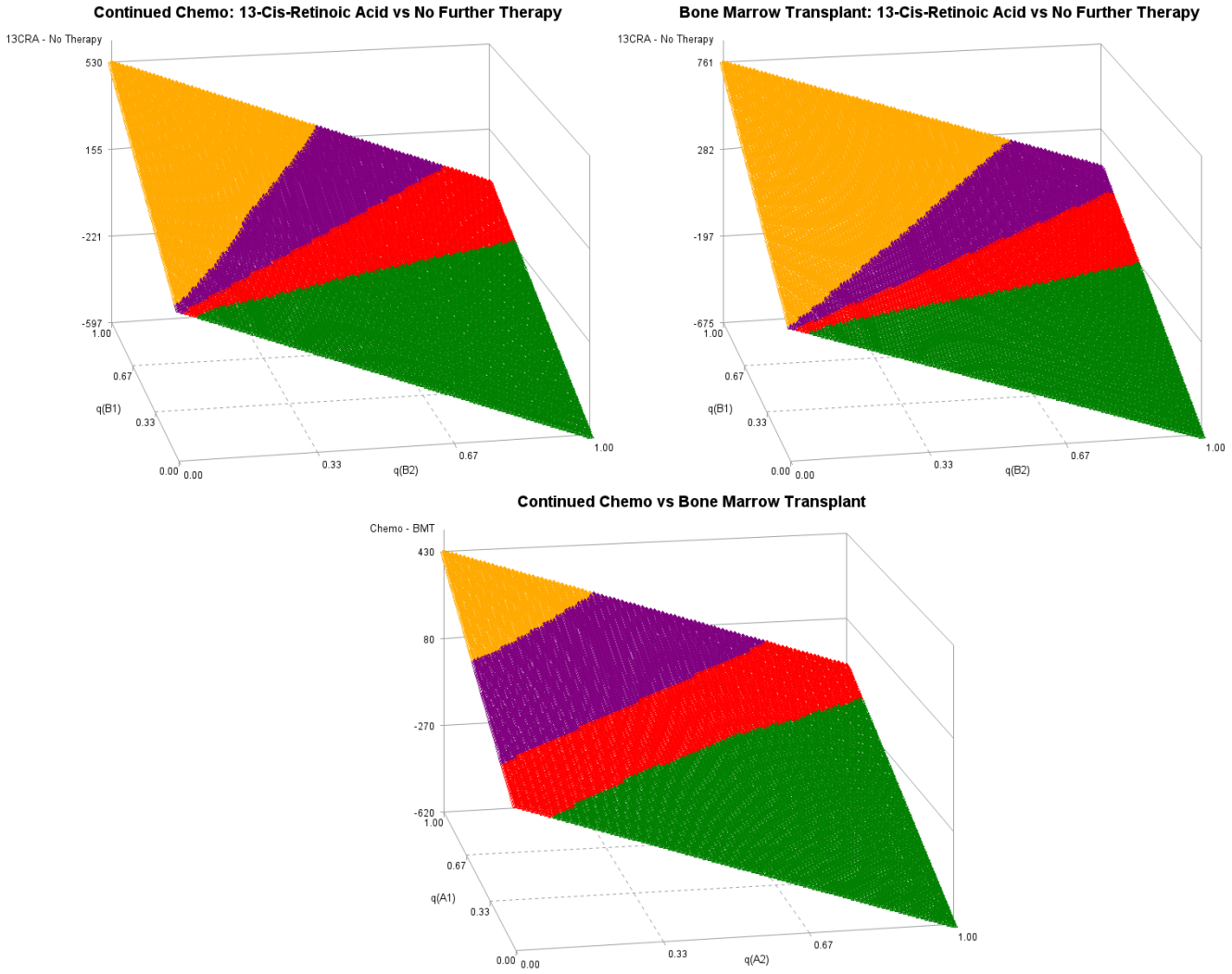

Figure 2: Estimated stage 2 (top row) and stage 1 (bottom row) threshold utility planes for COG study A3891. For each combination of q_{b_1} and q_{b_2}, or q_{a_1} and q_{a_2}, the estimated difference in mean quality adjusted lifetime is plotted. The yellow and green represent the region of strong acceptance for 13-cis-retinoic acid and no further therapy, or continued chemotherapy and bone marrow transplant, respectively. The purple and red near the center of the plane have 90% point-wise bootstrap confidence intervals that cover zero and represent the region of indifference when choosing between treatments.

6 Generalization to Other Outcomes

Our exploration of Q-learning to optimize a dynamic treatment regime on quality adjusted lifetime leads one to consider $\mathcal{Q}$-functions that weight the expected utility at each stage for any continuous outcome, not just survival time. For a 2-stage SMART design depicted earlier with a primary outcome Y at the end of the second stage, one can use the $\mathcal{Q}$-functions

$$\mathcal{Q}_B\Big(A_i = a_j, \bar{X}_{Bi}, B_i = b_k\Big) = q_{b_k} E\Big[Y_i^{(B)}\Big| A_i = a_j, \bar{X}_{Bi}, B_i = b_k\Big], \tag{6.1}$$

$$\mathcal{Q}_A\Big(X_{Ai}, A_i = a_j\Big) = E\Big[H_i^{(A)}\Big| X_{Ai}, A_i = a_j\Big], \tag{6.2}$$

where

$$H_i^{(A)} = \begin{cases} Y_i^{(A)} q_{a_j} + \underset{b_k}{max}\, \mathcal{Q}_B\Big(A_i = a_j, \bar{X}_{Bi}, B_i = b_k,\Big), & \text{if } A_i = a_j,\ R_i = 1 \\ Y_i^{(A)} q_{a_j}, & \text{if } A_i = a_j,\ R_i = 0, \end{cases} \tag{6.3}$$

and where $Y_i^{(A)}$ and $Y_i^{(B)}$ are the outcomes at the first and second stages, respectively, with $Y_i^{(A)} + Y_i^{(B)} = Y_i$. The law of total expectation can be used to improve computational efficiency when performing a threshold utility analysis. Most authors fit a single regression model for $E\big[H_i^{(A)}\big| X_{Ai}, A_i = a_j\big]$; however, a Q-learning model for stage 1 could be built using

$$\begin{aligned} E\Big[H_i^{(A)}\Big| X_{Ai}, A_i = a_j\Big] &= P(R_i = 1 | X_{A_i}, A_i = a_j) \\ &\times \Big\{ q_{a_j} E\Big[Y_i^{(A)}\Big| X_{Ai}, A_i = a_j\Big] + E\Big[\underset{b_k}{max}\, \mathcal{Q}_B\Big(A_i = a_j, \bar{X}_{Bi}, B_i = b_k\Big)\Big| X_{Ai}, A_i = a_j\Big]\Big\} \\ &+ P(R_i = 0 | X_{A_i}, A_i = a_j)\Big\{ q_{a_j} E\Big[Y_i^{(A)}\Big| X_{Ai}, A_i = a_j,\Big]\Big\}. \end{aligned} \tag{6.4}$$

Written this way, it is clear how the utility weights factor out of the expectations and create what we call quality adjusted Q-learning, for any continuous outcome. This could easily be generalized to SMARTs with an arbitrary number of stages. Modeling $E\big[H_i^{(A)}\big| X_{Ai}, A_i = a_j\big]$ in this way improves computational efficiency since each of the component models only needs to be fit once before varying the utility weights and producing a threshold utility analysis. Other authors, including Song et al. (2011), consider $\mathcal{Q}$-functions that have a single utility weight q, regardless of stage or treatment, that is multiplied to every nested expectation (except the first), creating an effect similar to the autoregressive working correlation structure from generalized linear models. Most authors interpret this single q as a utility weight that, when compounded over the nested expectations, diminishes the expected utility of each subsequent stage. The idea being that the prospective patient might not complete every stage of the DTR, and the optimal regime should give more importance to earlier treatments. However, even with this approach, most authors ignore the utility weight by setting it equal to 1. As we showed above, we propose assigning a separate utility weight to each treatment of each stage, representing the prospective patient's aversion to each treatment based on discomfort, side effects, monetary cost, ethical and/or religious beliefs, ability to complete the treatment schedule, and a host of other unmeasurable factors that might vary from one prospective patient to another.

This allows a threshold utility analysis as described in Section 3.2 for any continuous outcome Y, and shows us the decision making process of the prospective patient.

7 Concluding Remarks

Quality adjusted lifetime is a natural endpoint for deciding among treatments that prolong survival time, permitting one to factor toxicity and financial burden of treatment, among other things, into the decision. This is particularly useful in the realm of DTRs, allowing the optimal regime to depend not only on patient level characteristics, but also on treatment characteristics. We have demonstrated how threshold utility analysis can be combined with the standard optimization algorithm to produce optimal regimes accounting for patient and treatment level information. For simplicity, our methods did not include any covariate information other than response status, but additional patient characteristics such as age, race, or sex could be included in the optimization algorithm, producing a separate set of utility planes for, say, males and females, or young children and older children. Patient information could also be used to improve efficiency by using the semiparametric estimating equations in Wang & Zhao (2007) .

References

Aaronson, N. K., Ahmedzai, S., Bergman, B., Bullinger, M., Cull, A., Duez, N. J., Filiberti, A., Flechtner, H., Fleishman, S. B., de Haes, J. C., et al. (1993), "The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology," *Journal of the national cancer institute*, 85, 365–376.

Chakraborty, B., Laber, E. B., and Zhao, Y. (2013), "Inference for Optimal Dynamic Treatment Regimes Using an Adaptive m-Out-of-n Bootstrap Scheme," *Biometrics*, 69, 714–723.

Chakraborty, B., Murphy, S., and Strecher, V. (2009), "Inference for non-regular parameters in optimal dynamic treatment regimes," *Statistical methods in medical research*.

Felder-Puig, R., Di Gallo, A., Waldenmair, M., Norden, P., Winter, A., Gadner, H., and Topf, R. (2006), "Health-related quality of life of pediatric patients receiving allogeneic stem cell or bone marrow transplantation: results of a longitudinal, multi-center study," *Bone marrow transplantation*, 38, 119–126.

Gelber, R. D., Gelman, R. S., and Goldhirsch, A. (1989), "A Quality-of-Life-Oriented Endpoint for Comparing Therapies," *Biometrics*, 45, 781–795.

Glasziou, P., Simes, R., and Gelber, R. (1990), "Quality Adjusted Survival Analysis," *Statistics in Medicine*, 9, 1259–1276.

Goldhirsch, A., Gelber, R. D., Simes, R., Glasziou, P., and Coates, A. (1989), "Costs and benefits of adjuvant therapy in breast cancer: a quality-adjusted survival analysis," *Journal of Clinical Oncology*, 7, 36–44.

Hong, W. K., Endicott, J., Itri, L. M., Doos, W., Batsakis, J. G., Bell, R., Fofonoff, S., Byers, R., Atkinson, E. N., Vaughan, C., et al. (1986), "13-cis-retinoic acid in the treatment of oral leukoplakia," *New England Journal of Medicine*, 315, 1501–1505.

Huang, X., Ning, J., and Wahed, A. S. (2014), "Optimization of individualized dynamic treatment regimes for recurrent diseases," *Statistics in medicine*, 33, 2363–2378.

Korn, E. L. (1993), "On estimating the distribution function for quality of life in cancer clinical trials," *Biometrika*, 80, 535–542.

Laber, E. B., Lizotte, D. J., Qian, M., Pelham, W. E., and Murphy, S. A. (2014), "Dynamic treatment regimes: Technical challenges and applications," *Electronic journal of statistics*, 8, 1225.

Matthay, K. K., Villablanca, J. G., Seeger, R. C., Stram, D. O., Harris, R. E., Ramsay, N. K., Swift, P., Shimada, H., Black, C. T., Brodeur, G. M., et al. (1999), "Treatment of high-risk neuroblastoma with intensive chemotherapy, radiotherapy, autologous bone marrow transplantation, and 13-cis-retinoic acid," *New England Journal of Medicine*, 341, 1165–1173.

Murphy, S. (2003), "Optimal dynamic treatment regimes," *Journal of the Royal Statistical Society B*, 65, 331–366.

Pradhan, B. and Dewanji, A. (2009), "On induced dependent censoring for quality adjusted lifetime (QAL) data in a simple illness–death model," *Statistics & Probability Letters*, 79, 2170–2176.

Robins, J. M. (2004), "Optimal structural nested models for optimal sequential decisions," in *Proceedings of the second seattle Symposium in Biostatistics*, Springer, pp. 189–326.

Robins, J. M. and Rotnitzky, A. (1992), "Recovery of information and adjustment for dependent censoring using surrogate markers," in *AIDS Epidemiology*, Springer, pp. 297–331.

Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1994), "Estimation of regression coefficients when some regressors are not always observed," *Journal of the American Statistical Association*, 89, 846–866.

Song, R., Wang, W., Zeng, D., and Kosorok, M. R. (2011), "Penalized Q-learning for dynamic treatment regimes," *arXiv preprint arXiv:1108.5338*.

Wang, H. and Zhao, H. (2007), "Regression analysis of mean quality-adjusted lifetime with censored data," *Biostatistics*, 8, 368–382.

Wintner, L., Giesinger, J., Zabernigg, A., Sztankay, M., Meraner, V., Pall, G., Hilbe, W., and Holzner, B. (2013), "Quality of life during chemotherapy in lung cancer patients: results across different treatment lines," *British journal of cancer*, 109, 2301–2308.

Zhao, H. and Tsiatis, A. A. (1997), "A consistent estimator for the distribution of quality adjusted survival time," *Biometrika*, 84, 339–348.

— (1999), "Efficient Estimation of the Distribution of Quality-Adjusted Survival Time," *Biometrics*, 55, 1101–1107.

— (2000), "Estimating Mean Quality Adjusted Lifetime with Censored Data," *Sankhya: The Indian Journal of Statistics*, 62, 175–188.

— (2001), "Testing Equality of Survival Functions of Quality-Adjusted Lifetime," *Biometrics*, 57, 861–867.

Received: March 7, 2021

Accepted: April 10, 2021