Evaluation of Viable Dynamic Treatment Regimes in a Sequentially Randomized Trial of Advanced Prostate Cancer

Lu Wang, Andrea Rotnitzky, Xihong Lin, Randall E. Millikan, and Peter F. Thall *

Abstract

We present new statistical analyses of data arising from a clinical trial designed to compare two-stage dynamic treatment regimes (DTRs) for advanced prostate cancer. The trial protocol mandated that patients were to be initially randomized among four chemotherapies, and that those who responded poorly were to be rerandomized to one of the remaining candidate therapies. The primary aim was to compare the DTRs' overall success rates, with success defined by the occurrence of successful responses in each of two consecutive courses of the patient's therapy. Of the one hundred and fifty study participants, forty seven did not complete their therapy per the algorithm. However, thirty five of them did so for reasons that precluded further chemotherapy; i.e. toxicity and/or progressive disease. Consequently, rather than comparing the overall success rates of the DTRs in the unrealistic event that these patients had remained on their assigned chemotherapies, we conducted an analysis that compared viable switch rules defined by the per-protocol rules but with the additional provision that patients who developed toxicity or progressive disease switch to a non-prespecified therapeutic or palliative strategy. This modification involved consideration of bivariate per-course outcomes encoding both efficacy and toxicity. We used numerical scores elicited from the trial's Principal Investigator to quantify the clinical desirability of each bivariate per-course outcome, and defined one endpoint as their average over all courses of treatment. Two other simpler sets of scores as well as log survival time also were used as endpoints. Estimation of each DTR-specific mean score was conducted using inverse probability weighted methods that assumed that missingness in the twelve remaining drop-outs was informative but explainable in that it only depended on past recorded data. We conducted additional worst-best case analyses to evaluate sensitivity of our findings to extreme departures from the explainable drop-out assumption.

Key Words: Causal inference; Efficiency; Informative dropout; Inverse probability weighting; Marginal structural models; Optimal regime; Simultaneous confidence intervals.

*Lu Wang (email: luwang@umich.edu) is Assistant Professor, Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109; Andrea Rotnitzky (email: andrea@utdt.edu) is Professor, Department of Economics, Di Tella University, Buenos Aires, 1425, Argentina and Adjunct Professor, Department of Biostatistics, Harvard School of Public Health, Boston, MA 02115, Xihong Lin (email: xlin@hsph.harvard.edu) is Professor, Department of Biostatistics, Harvard School of Public Health, Boston, MA 02115, Randall E. Millikan (email: rmillika@mdanderson.org) is Associate Professor, Department of Genitourinary Medical Oncology, M.D. Anderson Cancer Center, Houston, TX 77030, and Peter F. Thall (email: rex@mdanderson.org) is Professor, Department of Biostatistics, M.D. Anderson Cancer Center, Houston, TX 77030. Wang and Lin's research is partially supported by a grant from the National Cancer Institute (R37-CA–76404). Rotnitzky's research is partially supported by grants R01-GM48704 and R01-AI051164 from the National Institutes of Health. Thall's research was partially supported by grant 2RO1 CA083932 from the National Institutes of Health.

1 INTRODUCTION

Therapy of cancer, cardiovascular disease, behavioral disorders, infections and many other diseases typically is conducted in multiple stages. A physician begins a therapeutic process by obtaining baseline information diagnosing a patient's disease and quantifying its severity, as well as covariates that may be related to therapeutic outcomes, and chooses the patient's first treatment on that basis. It is a common medical practice to repeat a treatment that has obtained a favorable response or to switch to an alternative treatment if the current response is unfavorable. The physician's choice of the alternative treatment is often guided by updated data on the patient's disease status and covariates. This decision making process is often repeated until either a response considered to be a definitive therapeutic success is achieved, or the therapy is discontinued. Common reasons for discontinuation include drop-out, the physician's decision that further therapy is futile, or regimen-related adverse events that preclude further therapy.

Multi-stage therapeutic strategies, in which dose or treatment are modified at each stage according to a patient's current history and disease status, have been given a number of different names in the statistical literature, including dynamic treatment regimes (DTRs), treatment policies, adaptive treatment strategies, multi-stage treatment strategies, and individualized treatment rules. In recent years, there has been a great deal of activity in the statistical community in design and analysis of studies aimed at evaluating the effects of DTRs. A number of recent articles discuss design of randomized trials that aim at evaluating DTRs rather than individual treatments (Lavori and Dawson, 2000, 2004; Murphy, 2005; Thall, et al., 2000; Thall, et al., 2002; Oetting, et al., 2007). A vast literature also exists on analytic tools to estimate the effects of DTRs using longitudinal observational data or data from randomized studies. Statistical methods include g-estimation of structural nested models (Robins, 1986, 1989, 1993, 1997), some clever variations of g-estimation for optimal treatment regime estimation (Murphy, 2003; Robins, 2004) and inverse probability weighted estimation of marginal structural models (Murphy, et al.,2001; Robins, et al., 2008; van der Laan, 2006; van der Laan and Petersen, 2007; Orellana, et al., 2010). In observational studies, specific versions of these methods have been developed to control for high dimensional time dependent confounders (i.e. time varying risk factors that affect future treatments) that are themselves predicted by past treatments. In controlled studies, these methods can be used to analyze sequentially randomized designs with randomization probabilities that possibly can depend on past health status and covariates (Lunceford, et al., 2002; Wahed and Tsiatis, 2004, 2006; Bembom and van der Laan 2007, 2008).

The purpose of this paper is to present new statistical analyses of data arising from a clinical trial of advanced prostate cancer conducted at M.D. Anderson Cancer Center from December 1998 to January 2006. The study was a groundbreaking early example of a sequentially multiple randomized assignment

trial (SMART) (Murphy, 2005) specifically designed to evaluate well defined DTRs. Its goal was to compare rules resembling those that oncologists often use when treating cancer patients, namely repeating a previous treatment if it has proved to be favorable and otherwise administering a different treatment ("repeat a winner and switch away from a loser"). The primary goal of the trial was to evaluate and compare 12 different sequential decision rules in which patients could be switched from an initial combination chemotherapy (hereafter, "chemo") chosen from the set $\mathcal{A} = \{\text{CVD,KA/VE,TEC,TEE}\}$ to a second, different chemo from the same set. This goal is different than the conventional goal of evaluating and comparing the four individual chemos given initially. The ultimate goal was to use the results of the trial as a basis for generating hypotheses and planning a future, confirmatory trial.

One hundred and fifty patients were randomized at enrollment to receive one of the four chemos. According to the protocol, depending on the per-course responses, patients could receive two to four courses of chemotherapy, the first at baseline, the second at week 8, and possibly additional courses at weeks 16 and 24. Specifically, the patient's per-protocol treatment assignments would end and he would be switched to a non-randomized therapeutic or palliative option immediately after the occurrence of a second non-favorable course, or two consecutive favorable courses, whichever occurred first. The protocol mandated randomization to a second, different chemo immediately after the patient's first non-favorable course if such an event occurred. Per-course favorable response was defined on the basis of a compound score involving lack of tumor growth and change in prostate specific antigen (PSA) (see Section 3.1). The protocol stipulated the recording of baseline PSA and disease volume, and per-course toxicity level and tumor status while receiving one of the chemos being studied. It also stipulated the recording of PSA values every eight weeks until week 32, regardless of patient discontinuation or not of study chemos. As of March 1, 2011, one patient was still alive, one had been lost to follow-up, and all others had died. Only nine patients died during the course of the trial. All death times but one were recorded and available for data analysis. Additional study design details are given in Thall, et al. (2000) and Thall, et al. (2007).

The seven possible response sequences were $s_a s_a$, $\overline{s}_a s_{a^*} s_{a^*}$, $s_a \overline{s}_a s_{a^*} s_{a^*}$, $\overline{s}_a \overline{s}_{a^*}$, $\overline{s}_a s_{a^*} \overline{s}_{a^*}$, $s_a \overline{s}_a \overline{s}_{a^*}$ and $s_a \overline{s}_a s_{a^*} \overline{s}_{a^*}$, where s and $\overline{s}$ stand for the per-course favorable and non-favorable responses and we subscript the response in each course by the treatment given in that course, a and a^* in the set $\mathcal{A}$ ($a \neq a^*$).The protocol stipulated the primary endpoint to be an overall success/failure outcome with success defined as the occurrence of any of the first three sequences and failure the occurrence of any of the four remaining ones. That is, overall success was defined as two consecutive favorable courses. Figure 1 illustrates the possible multi-stage outcomes per the protocol's treatment assignment algorithm and the number of patients for each outcome history at each course, including those with missing per-course response data.

Analyses of data arising from this trial reported by Thall, et al. (2007) generated some controversy

regarding both statistical methods (Bembom and van der Laan, 2007) and comparison of the regimens used in the trial to a particular combination chemotherapy reported in the medical literature while the trial was still ongoing (Tannock, et al., 2004; Armstrong, et al., 2007; Armstrong and Eisenberger, 2006; Millikan, et al., 2008). Using logistic regression to estimate the per-course probabilities of favorable response, Thall, et al. (2007) concluded that TEC was the best and CVD the worst treatment in course 1, while KA/VE was the best and TEE the worst salvage therapy. Bembom and van der Laan (2007) noted that the estimation strategy of Thall et al. (2007) was useful for identifying the chemotherapy that would give the best success rate in each course, but cannot identify the regime that gives the best overall success rate. These authors used an inverse probability of treatment weighted (IPTW) analysis to estimate the overall success rates of the 12 two-stage strategies, and found that (CVD, KA/VE) was the best, with (TEC, CVD) and (TEE, CVD) ranked second and third but all three estimated mean overall success rates were very similar. Both Bembom and van der Laan, and Thall et al, assumed that drop-outs were non-informative and carried out a complete case analyses, i.e. ignoring all data of the dropout subjects.

In the analyses that we will describe here, unlike Thall et al. (2007) and as recommended in Bembom and van der Laan (2007, 2008), we apply IPTW methods to estimate the endpoint means under different dynamic treatment regimes. Our analyses differ from those in Thall et al. (2007) and Bembom and van der Laan, 2007, in three important ways: (1) following the recommendation of Robins (1986, 2004), van der Laan and Petersen (2007) and Bembom and van der Laan (2008), we modify the definition of the candidate DTRs, (2) we study different endpoints, defined on the basis of data compiled subsequent to these earlier analyses, that identify specific reasons for discontinuing an assigned strategy and (3) based on this new information we define and analyze drop-outs differently. Specifically,

(1) *Viable DTRs.* As shown in Figure 1, 47 (31%) of the 150 patients in the trial who received initial treatment did not complete their therapy per the algorithm. During the process of inspecting the reasons for study chemo discontinuation, we determined that the switch rules prescribed by the protocol were, in fact, not feasible for patients who developed either severe toxicity or progressive disease (PD). Such events ordinarily preclude further chemotherapy, and in the prostate cancer trial they necessarily superseded the protocol's treatment assignment. Consequently, the DTRs that our analysis compares, throughout referred to as viable DTRs, differ from the two-stage switch-rules prescribed by the trial protocol in that they mandate discontinuation of study chemos after the occurrence of severe toxicity or PD. This modification involves the consideration of course-specific responses that encode not only information on efficacy, but also on toxicity and PD, as the DTR now is defined in terms of treatment decision rules that depend on these three domains. Viable DTRs have been discussed in Robins (1986 and 2004). Robins (2004) calls them "feasible" regimes, and van der Laan and Petersen (2007) call them "realistic" regimes.

(2) *Compound endpoints.* Trial investigators adopted the treatment assignment algorithm and primary endpoint defined earlier because, when designing the study, these were intended to reflect how oncologists actually assign treatments and evaluate overall response. The choice of endpoint led the trial investigators to judge unnecessary the collection of tumor status and toxicity data after treatment with the assigned chemo was discontinued. In Section 3.3, following Murphy (2005), we argue that this should not take place in SMARTs designed to evaluate DTRs because the primary endpoint in such trials should quantify the health experience of the patient over a pre-specified fixed period, the same period for all patients, for example, the period spanning the maximum possible duration of treatment, which in the prostate cancer trial was 32 weeks. In our analysis, we exploit the available information on toxicity and tumor status so as to compare the DTRs on the basis of endpoints that we judge are better predictors of the health status of patients over the entire 32 week maximum duration of therapy compared to the overall success/failure endpoint originally defined in the protocol. Specifically, we compare the regime-specific means of a compound score which was constructed by eliciting from the Principal Investigator (PI) of the trial subjective numerical values to quantify the clinical desirability of each per-course efficacy/toxicity/PD response. Bembom and van der Laan (2008) recommended analysis of endpoints based on utility functions that integrate per-course responses, but they did not carry such analysis because at that time the extended data set considered here that includes toxicity and PD was not available. To assess the sensitivity of the analytical conclusions to the chosen scores, we also repeated the analyses using the overall success/failure endpoint score per the trial protocol and another endpoint score that distinguishes therapies that provide transient benefits from those that do not. These alternative scores represent different viewpoints about the clinical desirability of the DTRs, in terms of their ability to diminish disease burden over the duration of therapy, which could last up to eight months. Yet another important dimension is the comparison of the effect of the distinct DTRs on long term survival time. Consequently, we also have estimated the mean log-survival time of the twelve viable DTRs.

(3) *Drop-outs.* Even after re-defining the regimes of interest as the viable DTRs, there still were 12 patients, 8% of the total sample of 150, who did not comply with the re-defined rules. These patients discontinued their assigned chemo neither because it was stipulated by protocol, nor because of severe toxicity or PD. We thus considered these patients to have dropped-out at the course where their therapy was discontinued. The analyses that we report here account for possibly informative, yet explainable, drop-out. That is, we analyze the data under the assumption that drop-out can depend on the history of PSA up to the time of withdrawal but is otherwise independent of the outcomes that would have been measured in the absence of drop-out. In addition, we conduct additional worse-best case analyses to evaluate the sensitivity of our findings to extreme departures from the preceding assumption.

The remainder of the paper is organized as follows. In Section 2, we elaborate on the need to focus on DTRs different from those defined in the protocol's algorithm, in order to account for the clinical decisions routinely made by oncologists when faced with toxicity or PD. In Section 3, we establish formal notation for the more complex outcomes considered in our analysis and define the dynamic treatment regimes that we compared. We define the subjective, PI-specified scoring function used to calculate one of the endpoints of our analysis, and we describe the two additional endpoints that we consider to evaluate effects over the duration of the trial. In Section 4, we discuss the inverse probability of treatment weighted methodology that we applied to estimate the outcome means associated with each of the two-stage strategies. We present the resulting data analyses in Section 5, and close with a brief discussion in Section 6.

2 VIABLE SWITCH RULES

If all patients enrolled in the trial had received treatment as stipulated by the study protocol, the data recorded in the trial would have allowed the assessment of the effects of 12 different two-stage treatment regimes which, for later reference, we call the per-protocol rules. Each such rule prescribes that an initial treatment be given with a specific chemo in $\mathcal{A}$, that treatment decisions be made every eight weeks immediately after the recording of the response to the prior treatment course, and that a switch take place from the initial chemo to either a second specific chemo in $\mathcal{A}$ or to a non-prespecified therapeutic or palliative strategy, with the latter given immediately after two favorable courses of the initial treatment and the former after one non-favorable course with it. The rule also stipulates that, in case of receiving a second chemo in $\mathcal{A}$, a switch to a non-prespecified therapeutic or palliative strategy would be made if the first course with the second chemo was not favorable.

As indicated in the Introduction, not all study participants received treatment per the trial protocol's algorithm. Thirty five patients did not because their treating physicians switched them to non-randomized therapeutic or palliative strategies because they developed either toxicity or PD. Eleven patients did not adhere to their randomized treatments because their physicians decided to remove them from the study for other unknown reasons, and one patient left the study on his own.

Decisions about how to analyze the trial data in the presence of subjects who did not receive treatment as specified in the protocol necessarily depend on the treatment regimes that one wishes to compare. An analysis that would disregard any data collected after the patient left the study protocol, and would use missing data techniques such as multiple imputation, model based-likelihood analysis, or IPW methods, would be aimed at comparing the per-protocol rules in idealized worlds in which all patients would follow the per-protocol rules that they were asked to follow. However, the presence of patients who left the study for reasons that precluded further administration of study therapies raises serious concerns about

the reasonableness and usefulness of such analysis. If a patient cannot continue on a given per-protocol rule due to adverse events that preclude further chemo, then the given rule is unrealistic for that patient, and it makes no sense to pretend that the patient would have followed it. It is an ill-defined task to attempt to compare per-protocol rules on the entire patient population if any such rule is not a viable option for a subset of the patients. An alternative, more reasonable approach, proposed by Robins (1986, 2004) and van der Laan and Petersen (2007) and recommended but not carried out by Bembom and van der Laan (2007), is to change the target of the analysis to viable switch rules that could actually be implemented in the study population. This is how we proceeded in the analyses presented here. The switch rules compared in our analysis, referred to throughout as "viable switch rules", are defined just as the per-protocol rules but have the important additional provision that patients who develop either toxicity or PD are mandated to switch to a non-prespecified therapeutic or palliative strategy, left to the discretion of the physician. The viable rules that we study simply state that the treatment strategy decision, whether palliative or therapeutic, after the development of toxicity or PD is left to the physician. Ideally, we would like to compare more refined rules that specify whether the switch should be to a therapeutic or to a palliative option according to whether the patient develops toxicity or PD. However, we could not evaluate the effects of these more detailed switch rules because the records available for data analysis did not indicate the specific course of action taken after the occurrence of toxicity or PD.

In our analysis, we regard as missing the outcome data subsequent to chemo discontinuation for subjects who went off study for reasons other than toxicity or PD. Our rationale for doing this is that we believe that, for such patients, the protocol treatment remained a viable option and, as such, it was conceivable that these subjects could have followed the viable-switch rule to which they would have been assigned.

In the next section we formally define the data compiled for this analysis, the viable switch rules, and the target parameters used in our analysis as a basis for comparing the distinct viable rules.

3 SOME FORMAL NOTATION

3.1 The data

When no meaning is lost, for simplicity we will suppress the subject index i. The trial recorded baseline covariates and per-course variables measured at the end of each course of chemo and just prior to assignment of the next chemo until just prior to discontinuation of study chemos. The per-course variables included PSA (a positive continuous variable), and a compound binary favorable/unfavorable response indicator defined in terms of PSA and an indicator of advance of disease (AD). AD was defined as any of the four events (1) new spots of bone involvement on bone scans; (2) increase in product of cross-sectional diameters of soft tissue of visceral metastases by 25% or more; (3) increase in cancer related symptoms; or (4) increase

in PSA from baseline by 25% or more confirmed by serial measurements one week apart. As per-protocol, a favorable response in the course that a chemo was first given was defined as a drop in PSA of at least 40% compared to baseline without evidence of AD, and a favorable response in the second consecutive course with the same chemo as per protocol was defined as a drop in PSA of at least 80% compared to baseline without evidence of AD (Thall et al., 2000).

For subjects departing from the study protocol there also were records indicating the reasons for doing so. In particular, it was recorded whether the decision to stop the study therapy was due to the development of severe toxicity or severe PD, or for other reasons. Note that PD was AD considered by the attending physician to be so severe that it precluded further therapy per the protocol algorithm. The extended data set which incorporated new compound per-course variables that recorded the development of toxicity or PD contained, for each patient, entries for the following 19 variables :

$$P_1,\ V_1,\ A_1,\ P_2,\ T_2,\ E_2,\ A_2,\ P_3,\ T_3,\ E_3,\ A_3,\ P_4,\ T_4,\ E_4,\ A_4,\ P_5,\ T_5,\ E_5, X.$$

Variable $A_j, j = 1, ..., 4$, records the chemo in the set $\mathcal{A}$ received at the start of course j if the patient actually received one. If the patient had discontinued the study chemos at or prior to the start of course j, A_j was coded either with OFF or with N/A. It was coded with OFF if the patient was alive at the start of course j and discontinuation was as mandated by protocol (i.e. due to the occurrence of two consecutive per-course favorable responses, or two unfavorable responses, consecutive or not), or due to either PD or severe toxicity. It was coded as N/A if discontinuation was for other reasons, including death. This data-coding convention is needed for our formal definition of viable rules given in the next section. Note that a patient with A_j = OFF in course j would still be adhering to the viable rule during course j, whereas one alive and with A_j = N/A would not.

Variables P_1 and V_1 are measured at baseline, prior to receiving the first chemo, P_1 records PSA, and V_1 is a binary indicator of high (versus low) disease volume, defined as at least 4 areas of presumed pathologic uptake or involvement of the appendicular skeleton as shown by bone scan, or visceral involvement (Thall et al., 2007).

Variables P_j, T_j and $E_j, j = 2, .., 5$ record PSA, toxicity and our compound measure of efficacy, all computed at the end of course $j - 1$ and just prior to A_j, provided the subject received a study chemo in course $j - 1$, otherwise they are coded as N/A. Toxicity T_j was a three level ordinal variable: TOX0 (no toxicity), TOX1 (toxicity occurring at a level of severity that precludes further therapy but allows efficacy to be evaluated) and TOX2 (toxicity so severe that therapy must be stopped and efficacy cannot be evaluated). Efficacy E_j was a four level variable: EFF0 (favorable response to a chemo in course j), EFF1 (non favorable response but no PD), EFF2 (PD) and EFF3 (inevaluable response due to severe toxicity).

Although the protocol stipulated that PSA values should be recorded even after study therapy discontinuation, these values were recorded in a very small number of subjects and, for several of them, only intermittently. We have chosen to disregard the few available post-study therapy PSA values and code them as N/A, since any analysis that used them would need to make untestable assumptions about the mechanism leading to the missing PSA values.

The variable X records the time to death measured in months from the time the first chemo was administered. All but two subjects were known to have died by March 1, 2011 and their death times were all recorded. Of the remaining two, one was last recorded to be alive 28.7 months after study enrollment. The other was still alive as of March 1, 2011. The death times of these two subjects were imputed as the last time they were known to have been alive.

In the sequel, we denote $L_1 = (P_1, V_1)$ and let L_j denote the entries for the covariates (P_j, T_j, E_j) at the end of course $j-1$ and the indicator that the person is alive at the start of course j, i.e. that X is greater than month $2 \times (j-1)$,

$$L_j = \left(P_j, T_j, E_j, I_{(2(j-1),\infty)}(X)\right), \ j = 2, ..., 5.$$

Figure 2 illustrates the possible per-course trajectories for (E_j, T_j), with the numbers observed to have followed each trajectory in parentheses. The figure also displays the courses of action prescribed by the viable DTRs defined in the next section and the number of subjects that dropped-out from the viable DTRs at each course. Comparison of Figures 1 and 2 shows that only 12 of the 47 cases that dropped-out of the per-protocol DTRs remain drop-outs of the viable DTRs.

3.2 The viable switch rules

To define the viable switch rules, we first consider the hypothetical world in which the only reasons for not adhering to the trial protocol are discontinuation of treatment because of PD and/or severe toxicity. In this hypothetical world, A_j will be coded as N/A only if the person is dead at the start of course j. In section 4.3 we will extend our definition to the case in which drop-outs for other reasons are present.

We will use the notational convention $\overline{V}_j = (V_1, ..., V_j)$ to represent the information accumulated on the variable V_l up to course j, and we use an unsubscribed $\overline{V}$ to denote the entire history. For any viable switch rule, described in section 2, the patient initially is treated with chemo $a \in \mathcal{A}$ and, if and when he qualifies for a switch to a second pre-specified chemo, he receives chemo $a^* \in \mathcal{A} - \{a\}$, but otherwise is treated with therapy left to the doctors' discretion. This is defined by four functions, $g_{a,a^*,j}\left(\overline{L}_j\right)$, for $j = 1, 2, 3, 4$. The function $g_{a,a^*,j}\left(\overline{L}_j\right)$ returns the therapy prescribed by the rule for course j when a patient has data $\overline{L}_j$.

To define $g_{a,a^*,j}(\cdot)$ let

$$S_j = I_{\{(\text{TOX0,EFF0})\}}[(T_j, E_j)] \text{ and } F_j = I_{\{(\text{TOX0,EFF1})\}}[(T_j, E_j)]$$

where $I_{\mathcal{B}}[B]$ is the indicator that B is in the set $\mathcal{B}$. Thus, S_j is the indicator of a favorable response without toxicity in course $j-1$ and F_j is the indicator of a non-favorable response without toxicity or PD. The functions $g_{a,a^*,j}$, $j = 1, 2, 3, 4$, are defined as follows:

$$g_{a,a^*,1}(L_1) = a, \qquad g_{a,a^*,2}\left(\overline{L}_2\right) = \begin{cases} a & \text{if } S_2 = 1 \\ a^* & \text{if } F_2 = 1 \\ \text{OFF} & \text{if } S_2 \neq 1, F_2 \neq 1, X > 2 \end{cases}$$

$$g_{a,a^*,3}\left(\overline{L}_3\right) = \begin{cases} a^* & \text{if } S_2F_3 = 1 \text{ or } F_2S_3 = 1 \\ \text{OFF} & \text{if } S_2F_3 \neq 1, \quad F_2S_3 \neq 1 \text{ and } X > 4 \end{cases}$$

$$g_{a,a^*,4}\left(\overline{L}_4\right) = \begin{cases} a^* & \text{if } S_2F_3S_4 = 1 \\ \text{OFF} & \text{if } S_2F_3S_4 \neq 1 \text{ and } X > 6 \end{cases}$$

Although X is not a component of $\overline{L}_j$, the indicator that $X > 2(j-1)$ is. Thus, $g_{a,a^*,j}(\cdot)$ is a well defined function of just the components of $\overline{L}_j$. Recall that an OFF in a course j indicates that the patient is no longer receiving a chemotherapy from the sequence (a, a^*) at the start of course j and has been switched to a therapeutic/palliative action decided by the treating physician. For example, at the start of course 2, a patient who had both $S_2 = 0$ and $F_2 = 0$ must have $T_2 = TOX1$ or $TOX2$ or $E_2 = EFF2$ or $EFF3$, i.e. he must have experienced severe toxicity and/or PD after the first course of chemo. As such, he should be taken off study chemo and switched to a therapeutic/palliative action, so $g_{a,a^*,2}\left(\overline{L}_2\right) = OFF$. Of course, no treatment action at the start of a given course needs to be specified if death has occurred prior to that time.

3.3 Outcome scores

In our analysis, we are interested in comparing DTRs on the basis of their effects on both long-term survival and efficacy in diminishing disease burden over 32 weeks. For the first goal, we analyze $U = \log X$, log survival time. For the second goal, we analyze three endpoints of the form $Y = y\left(\overline{L}\right)$ for specific scoring functions $y(\cdot)$ taking values in the interval $[0, 1]$. The value taken by $y\left(\bar{l}\right)$ is a numerical score that quantifies the clinical desirability of the response trajectory $\bar{l}$. Each choice of $y(\cdot)$ reflects a different viewpoint on what is desirable in a given response trajectory while receiving study chemos. All three scores are composites defined as functions of toxicity and efficacy while on study chemo. The first two scores are functions of the indicators

$$\widetilde{S}_j = I_{\{(\text{TOX0,EFF0}),(\text{TOX1,EFF0})\}}[(T_j, E_j)]$$

of evaluable (with or without toxicity) favorable response at course j.

1. Binary Scores: This scoring system simply assigns the value 1 if there were two consecutive per-course favorable responses and 0 otherwise. That is,

$$Y^{\text{bin}} = y^{\text{bin}}\left(\overline{L}\right) = \begin{cases} 1 \text{ if } \widetilde{S}_j\widetilde{S}_{j+1} = 1 \text{ for } j = 2, 3 \text{ or } 4 \\ 0 \text{ otherwise} \end{cases}.$$

The score Y^{bin} regards therapies that provide transient benefits, in the sense of having a positive probability of either only one successful course or two non-consecutive courses that are successful, to be equally undesirable as therapies that provide no benefits at all. This score is not quite the same as the overall success/failure endpoint stipulated by the trial protocol, since Y^{bin} takes the value 0 for a subject that drops out due to toxicity or PD, whereas the trial endpoint would be missing for such a subject.

2. Ordinal Scores: This scoring function differs from Y^{bin} in that the outcomes of patients for whom therapy achieved one successful course, or two non-consecutive successful courses, were scored as 0.5. Thus, it distinguishes therapies that produce transient efficacy benefits from therapies that don't. Specifically,

$$Y^{\text{ord}} = y^{\text{ord}}\left(\overline{L}\right) = \begin{cases} 1 \text{ if } \widetilde{S}_j\widetilde{S}_{j+1} = 1 \text{ for } j = 2, 3 \text{ or } 4 \\ 0.5 \text{ if } \widetilde{S}_2\left(1-\widetilde{S}_3\right)\left(1-\widetilde{S}_5\right) = 1 \text{ or } \left(1-\widetilde{S}_2\right)\widetilde{S}_3\left(1-\widetilde{S}_4\right) = 1 \\ 0 \text{ otherwise} \end{cases}$$

3. Expert Score: This score reflects the viewpoint of the PI of the trial regarding the relative clinical desirability of each of the possible per-course toxicity and efficacy outcomes while the patient was on study therapies. It thus distinguishes therapies on the basis of their benefits over the entire available trajectory of efficacy and toxicity. To construct this score, we elicited numerical values $C_j = c\left(E_j, T_j\right), j = 2, ..., 5$, between 0 and 1 for each of the possible combinations of values of (E_j, T_j) for every j such that the subject received a study chemo in course $j-1$. The seven possible numerical values of C_j are listed in Table 1. They reflect the clinical viewpoint that a course success, EFF0, is highly desirable, the absence of PD even if a success is not achieved, EFF1, is desirable, and extreme toxicity, TOX2, is highly undesirable. The symbol X in the table indicates that the corresponding combination of (E_j, T_j) is not feasible. The overall outcome score, which we call the "expert score," is defined as the mean of the per-course scores while the patient was on a study chemo, formally,

$$Y^{\text{expert}} = y^{\text{expert}}(\overline{L}) = \frac{\sum_{j=2}^{5}\left\{1 - I_{\{\text{OFF,N/A}\}}\left[A_{j-1}\right]\right\}C_j}{\sum_{j=2}^{5}\left\{1 - I_{\{\text{OFF,N/A}\}}\left[A_{j-1}\right]\right\}}.$$

Note that $1 - I_{\{\text{OFF,N/A}\}}\left[A_{j-1}\right]$ equals 1 if the subject is alive and received a study chemo at the beginning of course $j-1$, and equals 0 otherwise. Recall that (E_j, T_j) denotes the efficacy and toxicity measured at the end of course $j-1$.

The expert score is more informative than the ordinal score, as it distinguishes not only regimes that provide transient efficacy benefits from those that don't, but it also quantifies the clinical desirability of

the different transient benefits. For example, consider two subjects who had a favorable outcome with no toxicity in the first course of chemotherapy ($E_2 = EFF0$, $T_2 = TOX0$) but no more favorable outcomes afterwards. Suppose the first subject experienced PD and no toxicity to the second course of chemo ($E_3 = EFF2$, $T_3 = TOX0$) so his chemotherapy was discontinued, whereas the second subject experienced no PD and no toxicity in the second and third courses of chemo ($E_3 = E_4 = EFF1, T_3 = T_4 = TOX0$). The response trajectory of the second patient, while not an overall success, is still preferable to the response trajectory of the first patient. This is reflected in the expert score but not in the ordinal score; for both patients the ordinal score is 0.5 whereas the expert scores for the first and second patients are 0.55=(1+0.1)/2 and 0.67=(1+0.5+0.5)/3, respectively.

For comparing the benefits of the different DTRs in reducing disease burden over 32 weeks, we use scores computed using only outcome data while the patient was on study chemo. We do so because, by design, data on efficacy and toxicity were not collected subsequent to discontinuation of the study chemos and, as indicated earlier, even though PSA records were obtained for some subjects even after they went off study chemo, these records were very incomplete. The lack of off-study chemo outcome data limits our ability to compare the effects of different viable DTRs on disease burden, while alive, over the fixed period of 32 weeks. Our choice to analyze expert score endpoints is an attempt to remedy this problem insofar as we believe this score is a good predictor of health trajectory over the 32 weeks. The binary and categorical scores can be viewed as alternative, possibly poorer, substitute endpoints. Of course, if data on efficacy and toxicity had been collected over the 32 weeks even after chemotherapy discontinuation, this would have avoided the need for substitute endpoints.

The three scores $Y^{\text{bin}}, Y^{\text{ord}}$ and Y^{expert} are meant to quantify the health trajectory over 32 weeks since the first course of chemo. Yet, because they do not depend on survival, they rank equally two individuals who have the same outcomes while on study chemos, even if one dies soon after chemo discontinuation and the other remains alive at the end of the 32 weeks. A more reasonable utility function would score these two individuals differently, penalizing the former and rewarding the latter. Nevertheless, for simplicity, we have chosen to analyze scores that do not incorporate survival because only nine out of the 150 patients died in the first 32 weeks, all but one did so after study chemo discontinuation, and they were spread evenly among the four initial treatment arms. Comparing treatments on the basis of the log-survival means $E\left(U_{(a,a^*)}\right)$ informs about the long term effects of the different DTRs but not about their immediate effects, while comparisons based on the means of the three scores informs about their more immediate effects.

3.4 Counterfactual outcomes and the target of inference

To compare the different switch rules used in the trial, we apply the counterfactual framework for causal inference as originally developed by Rubin (1978) for time independent treatments and later extended by Robins (1986, 1987) for time dependent treatments in longitudinal studies. Henceforth, we define the vector $\overline{L}_{\overline{a}} = (L_{\overline{a},1}, L_{\overline{a},2}, L_{\overline{a},3}, L_{\overline{a},4},\ L_{\overline{a},5})$ of potential outcomes and the potential survival time $X_{\overline{a}}$ for each possible value $\overline{a} = (a_1, a_2, a_3, a_4)$ that $\overline{A}$ can take. Each $L_{\overline{a},j}$ denotes the value of L_j that would have been recorded at the end of course $j-1$ in a given subject in the hypothetical world in which his $\overline{A}$ would have been equal to $\overline{a}$. Likewise, $X_{\overline{a}}$ denotes the survival time if $\overline{A}$ had been equal to $\overline{a}$. We then define the collection $\mathcal{P} = \left\{\left(\overline{L}_{\overline{a}}, X_{\overline{a}}\right) : \overline{a} \text{ is in the range of } \overline{A}\right\}$ comprised of the potential outcome vectors and survival times under all possible treatment sequences $\overline{a}$. The set $\mathcal{P}$ includes potential outcome vectors $\overline{L}_{\overline{a}}$ corresponding even to values of $\overline{a}$ with some components equal to OFF. For those, the corresponding entries of the vector $\overline{L}_{\overline{a}}$ are set equal to N/A. For example, if $\overline{a}$ = (CVD, TEC, OFF, OFF) then $\overline{L}_{\overline{a}} = (L_{\overline{a},1}, L_{\overline{a},2}, L_{\overline{a},3}, \text{N/A}, \text{N/A})$. We use this convention because we want $L_{\overline{a},j}$ to reflect the value that would have been entered for L_j in the event that the person had $\overline{A}$ equal to $\overline{a}$, and recall that by convention, we code an outcome after discontinuation of study chemos as N/A. Given the complete collection of potential outcomes $\mathcal{P}$, we define for each switch rule g_{a,a^*} the hypothetical outcome vector $\overline{L}_{(a,a^*)}$, the potential survival $X_{(a,a^*)}$, and the potential endpoint $Y_{(a,a^*)} = y\left(\overline{L}_{(a,a^*)}\right)$. These are the values of $\overline{L}$, survival time X, and score Y that would have been recorded on a given patient if he had been randomized, perhaps contrary to fact, to follow the switch rule g_{a,a^*}. Thus, for example, $\overline{L}_{(a,a^*)} = \overline{L}_{\overline{a}}$ where $a_1 = a$, $a_2 = g_{a,a^*,2}\left(\overline{L}_{a_1}\right)$, etc.

In our analysis, we use the mean scores $E\left[Y_{(a,a^*)}\right]$ and mean log-survival times $E\left[U_{(a,a^*)}\right]$ where $U_{(a,a^*)} = \log X_{(a,a^*)}$, with (a, a^*) ranging over all 12 possible pairs, as the target parameters that form the basis for comparing the different switch rules in the trial. In particular, we will estimate each $E\left[Y_{(a,a^*)}\right]$ and $E\left[U_{(a,a^*)}\right]$ and the optimal switch rules $g_{a_{opt},a^*_{opt}}$, where

$$\left(a_{opt}, a^*_{opt}\right) = \arg\max_{(a,a^*)} E\left[Y_{(a,a^*)}\right] \text{ or } \left(a_{opt}, a^*_{opt}\right) = \arg\max_{(a,a^*)} E\left[U_{(a,a^*)}\right]$$

depending on whether our goal is to compare DTRs on the basis of their benefits for transitory diminishing disease burden or for prolonging survival.

SMART trials like the one considered here furnish data that identifies the effects of the DTRs they were designed to compare on the basis of a predetermined endpoint. This is so because at each stage each subject is randomized to one of the treatment options that would be available to him if he were to follow any of the DTRs being compared. One immediate question is whether the prostate cancer trial data could also identify the effects of the viable DTRs that we consider in our analysis. In fact, our modification of the definition of the switch rule does not impede identification. This is because the viable DTRs differ

from the original DTRs only in that they prescribe a switch to a non-prespecified therapy in the event of high toxicity or PD, and this rule was followed by all participating physicians. Intuitively, after a patient develops toxicity or PD, there is only one possible treatment option -the non-prespecified therapy- so identification is possible so long as everybody in the study complies to this added mandate, which indeed happened in the prostate cancer trial.

4 ESTIMATION METHODOLOGY

4.1 The requirements for the validity of the methodology

Our analysis of the trial data relies on estimation techniques described in Murphy, et al. (2001). Following an idea raised by Robins (1993), these authors discussed the use of IPTW methods to estimate the mean of a counterfactual outcome under a given DTR, possibly conditional on baseline covariates. Murphy et al. discussed their methods in the context of analyzing follow-up observational data. However, their methods also apply to analysis of sequentially randomized (SR) trials because they are valid under the following three requirements that, as we indicate next, are satisfied by design in SR trials.

The first requirement is that the collection of potential outcomes for the n study subjects $\mathcal{P}_i$, $i = 1, 2, ..., n$, be independent and identically distributed random vectors. This represents the idealization that the trial participants are a random sample from a large target population. This assumption is made routinely in the analysis of clinical trials and is reasonable for the prostate cancer trial.

The second requirement is unconfoundedness, which stipulates that A_j is independent of the counterfactual data $\mathcal{P}$ given the information $\left(\overline{L}_j, \overline{A}_{j-1}\right)$ recorded until just prior to assigning A_j,

$$\Pr\left(A_j = a_j | \overline{A}_{j-1} = \overline{a}_{j-1}, \overline{L}_j, \mathcal{P}\right) = \Pr\left(A_j = a_j | \overline{A}_{j-1} = \overline{a}_{j-1}, \overline{L}_j\right). \quad (1)$$

This requirement obviously is fulfilled in SR trials like the prostate cancer trial, where the randomization probabilities to the next treatment can depend at most on the information available to the investigator just prior to randomization, which is comprised of prior treatment assignments and recorded outcomes.

In the prostate cancer trial, $\Pr\left(A_j = a_j | \overline{A}_{j-1} = \overline{a}_{j-1}, \overline{L}_j\right) = p\left(a_j | \overline{a}_{j-1}, \overline{L}_j\right)$, $a_j \in \mathcal{A} \cup \{\text{OFF}\}, j =$

$1, .., 4$, where

$$
\begin{aligned}
p\left(a_1|L_1\right) &= 1/4\left\{1 - I_{\{\text{OFF}\}}\left[a_1\right]\right\}, \\
p\left(a_2|a_1, \overline{L}_2\right) &= \begin{cases} I_{\{a_1\}}\left(a_2\right) \text{ if } S_2 = 1 \\ 1/3\left\{1 - I_{\{a_1\}}\left[a_2\right]\right\} \text{ if } F_2 = 1, \\ I_{\{\text{OFF}\}}\left[a_2\right] \text{ if } S_2 \neq 1, F_2 \neq 1, X > 2 \end{cases} \\
p\left(a_3|\overline{a}_2, \overline{L}_3\right) &= \begin{cases} 1/3\left\{1 - I_{\{a_2\}}\left[a_3\right]\right\} \text{ if } S_2F_3 = 1 \\ I_{\{a_2\}}\left[a_3\right] \text{ if } F_2S_3 = 1 \\ I_{\{\text{OFF}\}}\left[a_3\right] \text{ if } S_2F_3 \neq 1, F_2S_3 \neq 1, X > 4 \end{cases}, \\
p\left(a_4|\overline{a}_3, \overline{L}_4\right) &= \begin{cases} I_{\{a_3\}}\left[a_4\right] \text{ if } S_2F_3S_4 = 1 \\ I_{\{\text{OFF}\}}\left[a_4\right] \text{ if } S_2F_3S_4 \neq 1, X > 6 \end{cases}
\end{aligned}
$$

The third requirement, often referred to as positivity, stipulates that any given subject in the study population has a positive probability of following any given DTR in the set of regimes being studied. This assumption obviously holds for the twelve viable switch rules. At the start of the trial any given patient has a positive chance of being assigned to, following any of the twelve rules. This assumption would not have been true if we had instead focused on the per-protocol rules, since subjects who would develop severe toxicity or PD under any given switch rule would have had probability zero of following it.

4.2 The heuristics of the IPTW estimators in our trial.

The IPTW methodology is based on the key observation that, under unconfoundedness and positivity, the means of the potential outcomes $Y_{(a,a^*)}$ and $U_{(a,a^*)}$ under a given viable switch rule g_{a,a^*} are equal to weighted means of the actual outcome values Y and U respectively, among subjects randomized to the switch rule under consideration (Murphy, et al., 2001). Specifically, for subject i let $\Delta_{a,a^*,i} = 1$ if subject i followed the switch rule g_{a,a^*} and $\Delta_{a,a^*,i} = 0$ otherwise. Furthermore, for $j = 1, ..., 4$, let

$$\omega_{j,i} = \Pr\left(A_j = A_{j,i}|\overline{A}_{j-1} = \overline{A}_{j-1,i}, \overline{L}_{j,i}\right)^{-1}$$

and let

$$\omega_i = \omega_{1,i} \times \omega_{2,i} \times \omega_{3,i} \times \omega_{4,i}.$$

Note that $\omega_{j,i}$ is inverse of the probability that a hypothetical patient having the same PSA, efficacy, toxicity, and treatment history up to j as subject i receives, at time j, the same treatment A_j that subject i actually received. It follows from Murphy et al. that, under unconfoundedness and positivity,

$$E\left[Y_{(a,a^*)}\right] = \frac{E\left(\Delta_{a,a^*}\omega Y\right)}{E\left(\Delta_{a,a^*}\omega\right)} = E\left(\Delta_{a,a^*}\omega Y\right). \tag{2}$$

We focus on estimation of $E\left[Y_{(a,a^*)}\right]$ because the arguments are identical for estimation of the mean of $U_{(a,a^*)}$ if survival is uncensored, as is essentially the case in our study. To interpret $E\left(\Delta_{a,a^*}\omega Y\right)$ it is helpful to regard a subject who did not follow the switch rule g_{a,a^*} as being censored at the first course that he departed from the rule. The product form of the weights ω effectively produces a stratified redistribution to the right, wherein those who are censored are redistributed, at the time of censoring, among those who remain uncensored and with the same outcomes and treatments in the past. This redistribution produces the right estimand because, by unconfoundedness, a subject who remained uncensored was chosen fairly from those at risk of being censored with the same past. Consequently, the future experience of a selected uncensored subject is representative of the experience that the censored subject would have had if he had continued to follow the rule g_{a,a^*}.

Table 2 lists the possible values that $\overline{A}$ can take for the subjects in the trial who followed regime g_{a,a^*}, together with the corresponding values of $\omega_j, j = 1,..,4$. To simplify the exposition, we assume that no subject died during the 32 weeks since first receiving chemo. Death induces only slight modifications that we discuss subsequently.

In Table 2, all groups receive an initial weight $\omega_1 = 4$. This is because the probability of initial randomization to chemo a was 1/4. For any given a, three patients are expected to be randomized to a chemo other than a for each patient randomized to a. These three patients have $\Delta_{a,a^*} = 0$. The factor $\omega_1 = 1/(1/4) = 1+3$ effectively makes each subject randomized to a represent three other subjects expected to be randomized to any of the alternative three chemos.

Subjects in groups 1 and 2 of Table 2 ended the study therapy without being randomized to a second treatment option. Those in group 1 ended the study therapies because they experienced toxicity or PD after the first course with chemo a. Those in group 2 had a successful response to the first course with chemo a, so they received a second course with the same chemo, and they were then removed from the study therapies either because they responded successfully to the second course, or because they developed adverse events. From the second course onwards, all patients in both groups followed what the rule g_{a,a^*} stipulated. They receive no more weight from this course and onwards, i.e. for them $\omega_2 = \omega_3 = \omega_4 = 1$, as they have nobody censored to account for.

Next consider subjects in groups 3 and 4. They had a non-successful response to the first course but they qualified for a second randomization at course 2 because they did not experience toxicity or PD, i.e. F_2 was 1. In course 2, these subjects were randomized to receive one of the three remaining chemo combinations in $A - \{a\}$ with probability 1/3 each and ended up being assigned to a^*. For every one of them, there are two patients expected to be assigned to a chemo other than a^* and who will therefore stop following rule g_{a,a^*}, and hence are censored at this course. The factor $\omega_2 = 1/\left(1/3\right) = 1 + 2$ effectively

makes each patient in groups 3 and 4 represent two expected censored patients. After course 2, all patients in groups 3 and 4 followed rule g_{a,a^*} regardless of whether or not they were removed from chemo a^* after course 2. They receive no additional weight, i.e. for them $\omega_3 = \omega_4 = 1$, because they have nobody to account for other than themselves.

Finally, consider subjects in groups 5 and 6. They received a second course of chemo a because they had a successful response to the first course with chemo a. Since this is precisely the action stipulated by rule g_{a,a^*} for such patients, all of them obeyed the rule at this stage. Thus, they receive the weight $\omega_2 = 1$ at this stage, as there is no censored subject they have to account for. However, patients in these groups were randomized to the second chemo at the third course because they had a non-successful response to the second course but they did not experience toxicity or PD, i.e. F_3 was 1. The factor $\omega_3 = 1/(1/3) = 1 + 2$ effectively makes each patient in these groups represent the two expected patients with the same treatment and response as those in courses 1 and 2 who will be censored at course 3 because they will not be randomized to a^*. At course 4, all patients in groups 5 and 6 followed rule g_{a,a^*} regardless of whether or not they were removed from chemo a^*. They receive no additional weight, i.e. for them $\omega_4 = 1$, because they have nobody to account for other than themselves. The last factor $\omega_4 = 1$ in all groups due to the fact that, at the fourth (last possible) course, there is no opportunity for re-randomization.

Suppose now that death could have occurred over the 32 weeks since first receiving chemo. In such a case, $\omega_j = 1$ at every course j in which the subject is dead as it should be, since after dying the dead person has nobody to account for. The equality (2) implies that the weighted sample average of Y among those that followed the switch rule, i.e.

$$\frac{\sum_{i=1}^{n} \Delta_{a,a^*,i}\omega_i Y_i}{\sum_{i=1}^{n} \Delta_{a,a^*,i}\omega_i} \tag{3}$$

is a consistent estimator of $E\left[Y_{(a,a^*)}\right]$.

Subjects in each of the six groups in Table 2 contribute to the sums in (3). For example, subjects who were initially randomized to chemo a and who developed toxicity or PD by the end of the first course are in group 1 and contribute with total weight $\omega = 4$. Note that these subjects contribute to the estimation of $E\left[Y_{(a,a^*)}\right]$ for all three viable DTRs that start with a and switch one $a^* \in \mathcal{A} - \{a\}$.

It is interesting to contrast the weighted average (3) with the unweighted sample average $\left\{\sum_{i=1}^{n} \Delta_{a,a^*,i}Y_i\right\} / \left\{\sum_{i=1}^{n} \Delta_{a,a^*,i}\right\}$ for those who followed regime $g_{a,a*}$. The weight ω_i is equal to 12 for a patient i who complied with the switch rule g_{a,a^*} and was randomized twice, i.e. a patient in groups 3 – 6, and is equal to 4 for a complier to the rule who did not reach the chance of a second randomization, i.e. someone in groups 1 and 2. In contrast, subjects in all six groups are given the same weight in the unweighted sample average. The unweighted average is not a consistent estimator of the counterfactual

mean $E[Y_{(a,a^*)}]$. Intuitively, the unweighted average suffers from bias due to confounding by indication because those failing at a given course are under-represented since, save chance variation, only 1/3 of them are assigned to the chemo a^*.

4.3 Handling drop-outs

In the trial, one subject who qualified for randomization to a second chemo in the second course and eleven subjects who qualified to chemo in the third course did not receive a second chemo for reasons other than toxicity or PD. As explained in section 3, these subjects, whom for ease of reference we call "drop-outs", differ from those who did not adhere to the study protocol because they developed toxicity or PD, in that it is conceivable that they could have continued on the chemo to which they would have been assigned. This consideration leads us to keep as our analytic target comparison of the viable rules defined in section 3.2 on the basis of the potential outcome means in the hypothetical world in which we could prevent drop-out from occurring.

However, addressing properly this analytic target raises new challenges because the twelve drop-outs departed from the viable rules they were being assigned to follow at the time of dropping out and the decision to drop-out was not driven by some exogenous random mechanism. That is, the unconfoundedness assumption on which the IPTW methodology relies is no longer automatically satisfied, essentially because embedded in each arm of the trial there is an observational study with self-selection to drop-out.

Formally, let $R_j = \left\{1 - I_{\{N/A\}}\left(A_j\right)\right\}$, $j = 1, 2, 3, 4$, be the indicator of being neither dead nor a drop-out at the start of course j. The rules we wish to compare are comprised of decisions at each j for two courses of action (R_j, A_j): the first, with regard to drop-out, stipulates that $R_j = 1$ for all subjects alive at the start of course j regardless of their past, i.e. drop-out is not allowed; the second, with regard to therapy, stipulates that a subject with past $\overline{L}_j$ should be assigned to A_j equal to $g_{a,a^*}\left(\overline{L}_j\right)$ where g_{a,a^*} is as defined in section (3.2) . The unconfoundedness requirement for this modified viable DTR demands that the conditional "action" probabilities

$$\begin{aligned} & \Pr\left(R_j = 1, A_j = a_j | R_{j-1} = 1, \overline{A}_{j-1} = \overline{a}_{j-1}, \overline{L}_j, X > 2\left(j-1\right), \mathcal{P}\right) \\ = \; & \Pr\left(R_j = 1 | R_{j-1} = 1, \overline{A}_{j-1} = \overline{a}_{j-1}, \overline{L}_j, X > 2\left(j-1\right), \mathcal{P}\right) \\ & \times \Pr\left(A_j = a_j | R_j = 1, \overline{A}_{j-1} = \overline{a}_{j-1}, \overline{L}_j, X > 2\left(j-1\right), \mathcal{P}\right). \end{aligned} \tag{4}$$

$j = 1, ..., 4$, be independent of the counterfactuals $\mathcal{P}$. Unfortunately this requirement is no longer satisfied by design. Specifically, the second factor on the right hand side is the probability that $A_j = a_j$ for a subject with treatment history $\overline{a}_{j-1}$, response history $\overline{L}_j$ and counterfactual data $\mathcal{P}$ who is alive at the start of course j and who, during course j, is either still on one of the study chemos or has been switched off

the study chemos earlier due to toxicity or PD. This probability equals $p\left(a_j|\overline{a}_{j-1},\overline{L}_j\right)$, defined in section 4.1, and hence it is indeed independent of $\mathcal{P}$. The first probability, on the other hand, is the conditional probability of not dropping-out at the start of course j. We cannot assert that by design this probability is independent of $\mathcal{P}$ because, unlike the study chemotherapies, dropping-out is not an option that has been assigned by randomization. Thus, in our randomized study we cannot guarantee that

$$\begin{aligned} & \Pr\left(R_j=1|R_{j-1}=1,\overline{A}_{j-1}=\overline{a}_{j-1},\overline{L}_j,X>2\left(j-1\right),\mathcal{P}\right) \\ = \; & \Pr\left(R_j=1|R_{j-1}=1,\overline{A}_{j-1}=\overline{a}_{j-1},\overline{L}_jX>2\left(j-1\right)\right). \end{aligned} \tag{5}$$

In our data analysis, we will adopt (5) as an assumption, recognizing that its validity is not ensured by design. However, because this assumption is not empirically verifiable, we will explore the sensitivity of our findings to departures from this assumption. Assumption (5) allows the possibility of informative drop-out as it permits dependence of the probability of drop-out on the (likely) correlates of prognosis $\overline{L}_j$ and $\overline{A}_{j-1}$. However, it stipulates that drop-out is explained by the measured outcome and treatment history, i.e. that $\overline{L}_j$ and $\overline{A}_{j-1}$ are the only correlates of $\mathcal{P}$ (i.e. prognosis) that are associated with stopping one of the study chemos at course j for reasons other than toxicity or PD.

Even after adopting assumption (5), we cannot directly apply the IPTW formula (3) to estimate $E\left[Y_{(a,a^*)}\right]$. This is because, for $j=2$ and 3, the weights

$$\begin{aligned} \omega_{j,i} \; = \; & \Pr\left(A_j=A_{j,i}|R_{j-1,i}=1,\overline{A}_{j-1}=\overline{A}_{j-1,i},\overline{L}_{j,i},X_i>2\left(j-1\right)\right) \\ & \times\Pr\left(R_j=1|R_{j-1}=1,\overline{A}_{j-1}=\overline{A}_{j-1,i},\overline{L}_{j,i},X_i>2\left(j-1\right)\right) \end{aligned}$$

are unknown functions of $\left(\overline{A}_{j-1,i},\overline{L}_{j,i}\right)$ since the non-drop-out probabilities,

$$\pi_{j,i}=\Pr\left(R_j=1|R_{j-1}=1,\overline{A}_{j-1}=\overline{A}_{j-1,i},\overline{L}_{j,i},X_i>2\left(j-1\right)\right)$$

are unknown, whereas for $j=4$ they are equal to 1 because there was no drop-out occurring at that course. Thus, we must estimate the drop-out probabilities, but to do so, we must make some modeling assumptions. This is because with only 150 patients and only 12 dropping out, we could not hope to estimate the unknown drop-out probabilities non-parametrically even if L_j were a finitely valued variable, much less if it includes, as it does, the continuous component PSA. In the analysis reported in section 5, we assume that the drop-out probabilities depend on $\left(\overline{A}_{j-1},\overline{L}_j\right)$ only through the past treatments $\overline{A}_{j-1}$ and on the indicator that PSA dropped over course $j-1$, i.e. that $P_{j-1}-P_j>0$. Our estimators $\widehat{\pi}_{j,i}$ of $\pi_{j,i}$ were computed as the proportion of subjects not dropping out at course j among subjects with treatment history $\overline{A}_{j-1,i}$ as subject i up to course $j-1$ and with PSA change $P_{j-1}-P_j$ of the same sign as that of subject i.

The preceding discussion implies that, if (5) holds and our model for the drop-out probabilities is correct, then

$$\widehat{E}\left[Y_{(a,a^*)}\right] \;=\; \frac{\sum_{i=1}^{n} \Delta_{a,a^*,i}\widehat{\omega}_i Y_i}{\sum_{i=1}^{n} \Delta_{a,a^*,i}\widehat{\omega}_i}$$

is a consistent estimator of $E\left[Y_{(a,a^*)}\right]$, where $\widehat{\omega}_i = \omega_{1,i} \times \widehat{\omega}_{2,i} \times \widehat{\omega}_{3,i}$, $\omega_{1,i} = 1/4$,

$$\widehat{\omega}_{j,i} = p\left(A_{j,i}|\overline{A}_{j-1,i}, \overline{L}_{j,i}\right)^{-1} \widehat{\pi}_{j,i}^{-1} \quad \text{for } j = 2, 3$$

and $p\left(A_{j,i}|\overline{A}_{j-1,i}, \overline{L}_{j,i}\right)$ is the function $p\left(a_j|\ \overline{a}_{j-1}, \overline{l}_j\right)$ defined in section 4.1 evaluated at $a_j = A_{j,i}, \overline{a}_{j-1} = \overline{A}_{j-1,i}$ and $\overline{l}_j = \overline{L}_{j,i}$. In the data analysis in section 5, we do not report the values of $\widehat{E}\left[Y_{(a,a^*)}\right]$, but rather the asymptotically more efficient estimators $\widetilde{E}\left[Y_{(a,a^*)}\right]$ that utilize estimated treatment probabilities, as described in the next section. The preceding discussion applies equally when the endpoint $Y_{(a,a^*)}$ is replaced by the log-survival endpoint $U_{(a,a^*)}$. Although we have records of the death times of the 12 drop-outs, these are not used in the calculation of the estimators of the means $E\left[U_{(a,a^*)}\right]$. The estimator $\widehat{E}\left[U_{(a,a^*)}\right]$ essentially censors each drop-out at the start of the course at which the subject first fails to comply with his assigned regimen and redistributes him among all non drop outs who share the same history of treatments, outcomes and covariates.

4.4 Estimating the known treatment probabilities as a tool for improving efficiency

In IPTW estimation, efficiency can be improved by replacing the known treatment probabilities that form the factors in the weights ω by maximum likelihood estimates under correctly specified models (see, for example, Robins, et al., 1994). While this may seem paradoxical, it can be understood by noticing that this replacement effectively corrects chance imbalances in the covariates in each arm. For example, in the prostate cancer trial 50 subjects and 100 subjects had low and high disease volume at baseline, respectively, so the proportion with low disease volume in the trial was 1/3. However, the respective numbers with low and high disease volume in the patients who initially received KA/VE were 10 and 26, which gives the slightly smaller proportion 10/36 with low disease volume in this group. Suppose that, although we know that $\Pr\left(A_1 = A_{1,i}|L_1 = L_{1,i}\right) = 1/4$ for each subject i, we choose to replace this probability in the computation of $\omega_{1,i}$ by its estimated value $\widehat{\Pr}\left(A_1 = A_{1,i}|L_1 = L_{1,i}\right) = m_{1,i}/n_{1,i}$, the observed proportion of trial participants in the arm of subject i having the same value of disease volume V_1 as him, i.e. with $m_{1,i} = \sum_{i=1}^{n} I_{\{A_{1,i}\}}\left[A_{1,j}\right] I_{\{V_{1,i}\}}\left[V_{1,j}\right]$ and $n_{1,i} = \sum_{i=1}^{n} I_{\{V_{1,i}\}}\left[V_{1,j}\right]$. Thus, in the group of patients who initially received KA/VE, all 10 subjects with low disease volume receive a weight of $1/(10/50) = 5$ and all 26 subjects with high disease volume receive a weight of $1/(26/100) = 100/26 = 3.85$. This effectively forms a pseudo-sample of 150 subjects, 50 with low and 100 with high disease volume, consequently recovering

the disease volume distribution in the entire trial. In contrast, because $\omega_{1,i} = 1/(1/4) = 4$ is the same for all i, this factor is inconsequential in the estimator $\widehat{E}\left[Y_{(a,a^*)}\right]$, i.e., it can be ignored without altering the value in (3). This implies that no correction for chance imbalances on the distribution of the baseline covariate disease volume takes place by the operation of multiplying by $\omega_{1,i}$ if we compute the weights with the known randomization probability.

The preceding argument suggests that it would be advantageous to non-parametrically estimate the treatment probabilities conditional on all recorded past information, i.e. baseline disease volume and PSE and past per-course PSA and treatments. However, in our trial, only 48 subjects qualified for randomization to a second chemo at month 2 (course 2) and only 39 qualified for randomization at month 4 (course 3). With these sample sizes we had to inevitably reduce dimensionality. We thus chose to estimate the treatment probabilities under parsimonious parametric models. Specifically, our estimators of the twelve estimated DTR means were computed as

$$\widetilde{E}\left[Y_{(a,a^*)}\right] = \frac{\sum_{i=1}^{n} \Delta_{a,a^*,i}\widetilde{\omega}_i Y_i}{\sum_{i=1}^{n} \Delta_{a,a^*,i}\widetilde{\omega}_i}$$

where $\widetilde{\omega}_i = \widetilde{\omega}_{1,i} \times \widetilde{\omega}_{2,i} \times \widetilde{\omega}_{3,i}$. The weight $\widetilde{\omega}_{1,i}$ was equal to $\Pr\left(A_1 = A_{1,i}|\overline{L}_{1,i};\widehat{\gamma}\right)$, where $\widehat{\gamma}$ is the maximum likelihood of the parameters $\gamma = \left(\gamma_{j,a}\right)_{j\in\{1,2,3\},a\in\{\text{KA/VE,TEC, TEE}\}}$ in the proportional odds model

$$\log \frac{\Pr\left(A_1 = a|\overline{L}_1;\gamma\right)}{\Pr\left(A_1 = \text{CVD}|\overline{L}_1;\gamma\right)} = \gamma_{1,a} + \gamma_{2,a}V_1 + \gamma_{3,a}\log\left(P_1\right), \tag{6}$$

for $a =$ KA/VE, TEC and TEE.

For $j = 2$ and 3, we computed

$$\widetilde{\omega}_{j,i} = \widehat{\lambda}_{j,i}^{-1} \times \widehat{\pi}_{j,i}^{-1} \tag{7}$$

where $\widehat{\pi}_{j,i}$ are the estimated drop-out probabilities computed as indicated in the preceding section and $\widehat{\lambda}_{j,i}$ are estimators of the treatment probabilities $\Pr\left(A_j = A_{j,i}|\overline{A}_{j-1} = \overline{A}_{j-1,i}, \overline{L}_j = \overline{L}_{j,i}, X_i > 2\left(j-1\right)\right)$ and were computed as follows. We set $\widehat{\lambda}_{j,i}$ to 1 unless subject i qualified for randomization to a second chemo at the start of course j, that is $\widehat{\lambda}_{2,i} = 1$ unless $F_{2,i} = 1$ and $\widehat{\lambda}_{3,i} = 1$ unless $S_{2,i}F_{3,i} = 1$. We computed the remaining values of $\widehat{\lambda}_{j,i}$ as follows. We postulated two models sharing the same parameters, the first for the probability of assignment to a second chemo in course 2 among those that had one course of chemo, A_1, equal to a^* and that qualified to randomization at course 2,

$$\log \frac{\Pr\left(A_2 = a|R_2 = 1, A_1 = a^*, \overline{L}_2, X > 2;\alpha\right)}{\Pr\left(A_2 = a_0\left(a^*\right)|R_2 = 1, A_1 = a^*, X > 2, \overline{L}_2;\alpha\right)} = \alpha_{a,a^*}\log\left(P_2\right),\ a \neq a^* \tag{8}$$

and the second for the probability of assignment to a second chemo in course 3 among those that had two

courses of chemo, A_1 and A_2, equal to a^* and that qualified to randomization at course 3,

$$\log \frac{\Pr\left(A_3 = a | R_3 = 1, A_1 = A_2 = a^*, \overline{L}_3, X > 4; \alpha\right)}{\Pr\left(A_3 = a_0\left(a^*\right) | R_3 = 1, A_1 = A_2 = a^*, \overline{L}_3, X > 4; \alpha\right)} = \alpha_{a,a^*} \log\left(P_3\right), \; a \neq a^* \tag{9}$$

where $a_0\left(\text{CVD}\right) = \text{KA/VE}$, $a_0\left(\text{KA/VE}\right) = \text{CVD}$, $a_0\left(\text{TEC}\right) = \text{CVD}$ and $a_0\left(\text{TEE}\right) = \text{CVD}$. Thus, the two models assume that the probability of assignment to a second chemo, say a, among subjects that received a first chemo, say a^*, is the same function of the last PSA value regardless of whether the assignment is at course 2 or course 3. The functions, however, may be different for subjects that received a different first chemo, as α_{a,a^*} depends on a^*. We computed the maximum likelihood estimator $\widehat{\alpha}_{a,a^*}$ of α_{a,a^*} and for subjects i that qualified for randomization to a second chemo at course j we computed $\widehat{\lambda}_{j,i}$ as $\Pr\left(A_j = A_{j,i} | \overline{A}_{j-1} = \overline{A}_{j-1,i}, \overline{L}_j = \overline{L}_{j,i}, X_i > 2\left(j-1\right); \widehat{\alpha}\right)$. Models (6), (8) and (9) are correctly specified, because the true assignment probabilities are 1/4 for the first randomization and 1/3 for the second randomization, and consequently, the models hold with $\gamma_{j,a} = 0$, for $a = \text{KA/VE}$, TEC and TEE, $j = 1, 2, 3$ and $\alpha_{a,a^*} = 0$ for all a and a^*.

4.5 Inference about the optimal switch rule

Non-parametric bootstrap standard error estimators can be used to construct regime-specific Wald type confidence intervals centered at the regime-specific IPTW estimators $\widetilde{E}\left[Y_{(a,a^*)}\right]$. The bootstrap produces consistent estimators of the asymptotic variance of regular asymptotically linear (RAL) estimators (Gill, 1989) and $\widetilde{E}\left[Y_{(a,a^*)}\right]$ is a RAL estimator since its computation involves solving jointly smooth estimating equations for it and for the parameters of the treatment and drop-out models. However, these regime-specific confidence intervals cannot be used to conduct inference about the optimal regime as they do not account for the multiple comparisons involved in the calculation of the optimal rule. Nevertheless, in the Appendix we show that we can still use the bootstrap to construct simultaneous confidence intervals using a procedure similar to the one described in Bembom and van der Laan (2008). These intervals are computed in such a way that, given a nominal level τ, in at least τ percent of infinitely many hypothetical repetitions of the trial, each of the 12 counterfactual means $E\left[Y_{(a,a^*)}\right]$ would be covered by its corresponding interval. The data analyses in the next section report these simultaneous confidence intervals.

The simultaneous confidence intervals serve for the construction of a confidence set $\mathcal{C}$ for the optimal DTR as follows. We identify the $100 \times \tau$ percent confidence interval corresponding to the regime g_{a,a^*} with the largest estimated mean $E\left[Y_{(a,a^*)}\right]$ and then construct the set $\mathcal{C}$ to be the one comprised by all the DTRs whose confidence intervals overlap with this interval. This random set $\mathcal{C}$ includes the optimal DTR with probability at least τ. In spite of being conservative, the set $\mathcal{C}$ helps narrow down the collection of switch rule candidates for being optimal, in that DTRs that fall outside it are, with high confidence, DTRs that do not yield the largest outcome mean.

In addition to computing the simultaneous confidence intervals we conducted a Wald-type test of the null hypothesis of no overall treatment effect

$$H_0 : E\left(Y_{(a,a^*)}\right) \text{ does not depend on } (a, a^*)\,. \tag{10}$$

The test rejects if $S = (\widetilde{\mu} - \widetilde{\mu}_0)' W_{boot}^{-1} (\widetilde{\mu} - \widetilde{\mu}_0)$ is greater than the 95^{th} percentile of a chi-squared distribution with 11 degrees of freedom. Here, $\widetilde{\mu}$ denotes the 12×1 vector of the 12 estimated means $\widetilde{E}\left[Y_{(a,a^*)}\right]$, $\widetilde{\mu}_0$ is the 12×1 vector with all components equal to the consistent estimator of the common outcome mean under H_0, $\left\{ \sum_{(a,a^*)} \sum_i \Delta_{a,a^*,i} \widetilde{\omega}_i Y_i \right\} / \left\{ \sum_{(a,a^*)} \sum_i \Delta_{a,a^*,i} \widetilde{\omega}_i \right\}$, and W_{boot} is the non-parametric bootstrap estimator of the covariance matrix of $\widetilde{\mu} - \widetilde{\mu}_0$.

5 DATA ANALYSES

Figure 3 displays plots of the twelve estimators $\widetilde{E}\left[Y_{(a,a^*)}\right]$ and their simultaneous 95% confidence intervals for each of the three scores. Table 3 provides the numerical values of these means and confidence intervals. The results reported in this table indicate that the switch rule with the highest estimated mean expert score is the one that starts with TEC and switches to CVD (estimated mean expert score = 0.78). In fact, the other two regimes that start with TEC, i.e. (TEC, KA/VE) and (TEC, TEE), also have high estimated mean expert scores compared to the other regimes, 0.73 and 0.74 respectively. The lowest estimated mean expert score is 0.56, corresponding to regime (CVD, TEE). The uncertainty in the estimated mean scores is, nevertheless, substantial. In fact, the confidence interval for the mean expert score of (TEC, CVD) overlaps with the confidence intervals for the mean expert scores of each of the remaining eleven regimes, thus resulting in a 95% confidence set $\mathcal{C}$ for the optimal regime that does not exclude any of the twelve regimes. Interestingly, the Wald-type test of the overall null hypothesis H_0 defined in (10) rejected at the 95% level ($S = 33.00$). Thus, the data indeed provide evidence that not all DTR mean expert scores are the same, but it is too noisy to allow the detection of the ordering of the outcome means.

The estimated mean ordinal scores gave a similar ranking, with (TEC, CVD) having the highest estimated score (0.67) and (CVD, TEE) the lowest (0.31). Once again, the 95% confidence set for the optimal rule did not exclude any of the twelve regimes, but the test of the overall null H_0 rejected. The increment in mean ordinal score conferred by (TEC, CVD) over (CVD, TEE) was greater than 100% whereas this increment was about 30 % for the mean expert score. An even greater difference between (TEC, CVD) and (CVD, TEE) is obtained when the binary scores are considered: 0.11 for the first and 0.44 for the second. A comparison between the increment conferred by the binary scores and the ordinal scores by (TEC, CVD) over (CVD, TEE) indicates that, whereas the latter produces only few overall successes compared to the former, when transient successes are also considered, the distinction between

the two regimes is less profound. Further comparison using the expert scores indicates that the distinction between these two regimes is even less profound when the nature of the transient successes is also taken into account.

The estimated mean log survival times of the 12 regimes, in the last column of Table 3, are ordered quite differently from the mean score ordering. For instance, (CVD, TEC) and (TEE, CVD) have the largest mean log survival estimate 3.36. However, the 95% confidence set $\mathcal{C}$ for the optimal DTR includes all regimes, and the test at the 95% level that $E\left[U_{(a,a^*)}\right]$ is the same for all (a, a^*) fails to reject ($S = 15.65$), so the results are inconclusive. This is not surprising. Given that most patients survived long after the duration of the studied therapies, the effect on survival of these earlier therapies was likely washed out by treatment decisions made subsequently, and moreover small differences among the mean log survival of the DTRs would not be detectable with the sample size of this study. In any case, even if the analyses would have shed convincing evidence that regimes rank differently on the basis of mean log survival times compared to mean scores, this could have been explained by the fact that some switch rules might be preferable for temporarily diminishing disease burden whereas others might be preferable for prolonging survival.

To assess the sensitivity of our inferences to departures from assumption (5), and in particular, to evaluate to what extent the benefit attributed to the switch rule estimated as optimal based on the expert score and on mean log survival depended on our assumptions about the drop-out process, we conducted four extreme analyses. The first two analyses imputed extreme values for the expert scores of the 12 drop-outs and the last two analyses imputed extreme values for the death times of the 12 drop-outs as follows.

In the first analysis, for patients who were first assigned to CVD (the stage 1 chemo corresponding to the regime with smallest estimated mean expert score) we imputed the highest score 1, whereas for all others we imputed the lowest score 0. This analysis was conducted to examine the robustness of the conclusion that the regime (CVD, TEE) has the lowest estimated mean score to assumptions about the drop-out mechanism. In the second analysis, for patients who were first assigned to TEC (the stage 1 chemo corresponding to the regime with largest estimated mean expert score) we imputed the lowest score 0, whereas for all others we imputed the highest score 1. This analysis was conducted to examine the robustness of the conclusion that the regime (TEC, CVD) has the highest estimated mean score to assumptions about the drop-out mechanism. Results reported in the first two columns of Table 4 indicate that the regimes (CVD, TEE) and (TEC, CVD) remain as the ones with the smallest and largest mean expert scores respectively. However, as in the earlier analysis, the results are inconclusive as the confidence intervals overlap.

In the third and fourth analyses, the survival time imputed for each drop-out whose last course of chemo was $j-1$ and who was alive at the end of that course, was calculated using the survival times of subjects that had the same courses of chemo up to and including $j-1$ and had the same values of E_j and T_j as the given drop-put, throughout referred to as the reference group. In the third analysis, each drop-out who was first assigned to KA/VE (the stage 1 chemo corresponding to the regime with smallest estimated mean log survival time) was imputed the longest survival time of his reference group. For all other drop-outs, each survival time was computed by adding to $2\,(j-1)$, where $j-1$ was the last course of chemo received by the given drop-out, one half of the shortest remaining survival time of his reference group. In the fourth analysis we replicated this last imputation scheme only to impute the survival times of the drop-outs who were first assigned to TEE and CVD (the stage 1 chemos corresponding to the regimes with largest estimated mean log survival) and we imputed the survival times of all remaining drop-outs with the longest survival time in the corresponding reference group. The results reported in the last two columns of Table 4 show that the order of the DTRs does not stay the same as in the earlier analysis that assumed informative but explainable drop-out. In particular, regime (TEE, CVD) is no longer a regime with the largest mean log-survival time and regime (KA/VE, TEE) is no longer the one with the smallest mean. Nevertheless, (CVD, TEC) stays as the regime with the largest mean log-survival time. Once again, these rankings are not firm as any given pair of confidence intervals overlap.

6 DISCUSSION

In this paper, we have presented a new statistical analysis of a novel clinical trial in which prostate cancer patients were initially randomized to one of four chemotherapies, and those who responded poorly to their initial regimen were randomly reassigned to one of the remaining candidate chemos. Such sequentially randomized trials mimic the way that oncologists actually behave when treating cancer patients and thus they allow investigators to study adaptive treatment strategies. Our analysis was motivated by the fact that, as is routine in oncology practice, quite a few (47) patients enrolled in this trial discontinued their assigned therapy due to either severe toxicity or PD. Because many of them (35) did so for reasons that precluded further therapy, we switched the target of analysis to comparison of viable dynamic treatment regimes that additionally stipulate that patients developing toxicity or PD should be removed from study therapy. This was made possible by expanding the data set to include toxicity and PD as additional per-course outcomes, using additional information provided by the principal investigator of the trial. We thus re-defined patient outcome as a more informative compound event combining information on both efficacy and toxicity. The remaining non-compliers (12 patients) were assumed to have followed a possibly informative, but explainable, drop out mechanism given the history of PSA up to the time of withdrawal.

We applied IPTW methods to estimate counterfactual regime-specific means of the compound endpoint, an elicited expert score, under different dynamic treatment regimes. We found that (TEC, CVD) had the highest estimated mean expert score, followed by (TEC, TEE) and (TEC, KA/VE), while (CVD, TEE) had the lowest estimated mean expert score. However, the uncertainty in the estimated mean scores is substantial, as indicated by the 95% simultaneous confidence intervals.

We also applied our proposed methodology to the overall success/failure endpoint score and another ordinal endpoint score that distinguishes therapies providing transient benefits. The former score was used by Bembom and van der Laan (BV, 2007) as well, although their analysis was restricted to complete cases. These authors found that (CVD, KA/VE) has the highest overall success rate. In contrast, (CVD, KA/VE) is no longer the top choice. This is not surprising, because among those patients who followed regime (CVD, KA/VE), 9 patients who developed severe toxicity or severe PD were excluded by Bembom and van der Laan. After re-defining the compound endpoint, only one of these 9 patients is still missing. The other 8 patients were assigned 0 for the overall success/failure endpoint score in our analysis. Therefore, the estimated regime-specific mean score for (CVD, KA/VE) is greatly shrunk to a lower number. Our result, with (TEC, KA/VE) and (TEC,CVD) having the highest estimated overall success rates, is more consistent with Thall, et al. (2007), who used all patients in their first line analysis based on the success/failure endpoint and concluded that the best initial chemo is TEC while the worst initial chemo is CVD. We found the ranking of the 12 regimes in our analyses to be relatively insensitive to the choice of scores.

The estimators of mean log survival times of the 12 regimes were ordered quite differently than the means of the three considered scores. We interpret this distinct ordering as a manifestation that the DTRs might possibly rank differently with regard to their ability to temporarily reduce disease burden compared to prolonging survival.

One limitation of this study is its sample size. The trial was designed to be hypothesis generating (Thall, et al., 2007), hence had 150 patients. The sample size is far too small to draw confirmatory conclusions comparing the 12 treatment pairs. Sample size calculations for sequentially randomized trials are important to provide practical guidance (see Murphy, 2005; Dawson and Lavori 2011; Feng and Wahed 2009). Our proposal to construct simultaneous CIs is conservative and assumes normality of the estimators of the counterfactual mean scores, which is justified only in large samples. Future research is needed to improve finite sample inferences about optimal regimes.

7 Appendix

To compute the simultaneous confidence intervals for the 12 means $E\left[Y_{(a,a^*)}\right]$, following Bembom and van der Laan (2008) we reasoned as follows. If μ and $\widetilde{\mu}$ denote the 12×1 vectors comprised by the 12 means

$E\left[Y_{(a,a^*)}\right]$ and estimated means $\widetilde{E}\left[Y_{(a,a^*)}\right]$ respectively, we know that

$$\sqrt{n}\left(\widetilde{\mu}-\mu\right)\rightarrow N\left(0,\Sigma\right)$$

The asymptotic normality follows, under regularity conditions, after standard Taylor expansion arguments, because computation of $\widetilde{\mu}$ involves solving jointly smooth estimating equations for it and the parameters α and η of the treatment and drop-out model.

Then $\max\left\{\left|\frac{\widetilde{\mu}_j-\mu_j}{\sqrt{\Sigma_{jj}/n}}\right|;j=1,...,12\right\}$ is distributed like $Z_{\max}=\max\left\{\left|Z_j\right|;j=1,...,12\right\}$ where $Z=(Z_1,...,Z_{12})\sim N\left(0,\Omega\right)$ with $\Omega=diag\left(\Sigma\right)^{-1/2}\Sigma diag\left(\Sigma\right)^{-1/2}$. If $z_{\max,.95}$ is the 95th percentile of $Z_{\max}$, and I_j is the interval

$$\left(\widetilde{\mu}_j-z_{\max,.95}\sqrt{\Sigma_{jj}/n},\widetilde{\mu}_j+z_{\max,.95}\sqrt{\Sigma_{jj}/n}\right)$$

then

$$\begin{aligned}\Pr\left\{\mu_j\in I_j\text{ for }j=1,...,12\right\} &= \Pr\left\{\left|\frac{\widetilde{\mu}_j-\mu_j}{\sqrt{\Sigma_{jj}/n}}\right|\leq z_{\max,.95}\text{ for }j=1,...,12\right\}\\ &= \Pr\left\{Z_{\max}\leq z_{\max,.95}\text{ for }j=1,...,12\right\}=0.95\end{aligned}$$

In the construction of our confidence interval, we replaced Σ_{jj}/n with the non-parametric bootstrap estimator $V_{boot,j}$ of the asymptotic variance of $\widetilde{\mu}_j$ using 1000 bootstrap replications. We also used a Monte Carlo procedure to compute an estimate $\widehat{z}_{\max,.95}$ of the unknown value $z_{\max,.95}$. Specifically, we computed the non-parametric bootstrap estimator $\widehat{\Omega}_{boot}$ of the correlation matrix Ω. Next, we generated $\widehat{Z}_k\overset{iid}{\sim}N\left(0,\widehat{\Omega}_{boot}\right),k=1,...,10,000$, and for each k we computed $Z_{\max,k}=\max\left\{\left|Z_{k,j}\right|;j=1,...,12\right\}$. Finally, we computed $\widehat{z}_{\max,.95}$ as the 95th percentile of the empirical distribution of $Z_{\max,k},k=1,...,10,000$ and we calculated the confidence intervals as $\left(\widetilde{\mu}_j-\widehat{z}_{\max,.95}\sqrt{V_{boot,j}},\widetilde{\mu}_j+\widehat{z}_{\max,.95}\sqrt{V_{boot,j}}\right)$.

References

Armstrong, A.J., Garrett-Mayer, E.S., Yang, Y.C., DeWit, R., Tannock, I.F. and Eisenberger, M. (2007a) A contemporary prognostic nomogram for men with hormone-refractory metastatic prostate cancer: a TAX327 study analysis. *Clinical Cancer Research*, 13(21), 6396-403.

Armstrong AJ, Eisenberger MA. Does gonadotropin-releasing-hormone agonist therapy for prostate cancer increase the risk of fracture? Nat Clin Pract Urol. 2006 May; 3 (5):246-7.

Bembom, O. and van der Laan, M. (2007) Commentary: Statistical methods for analyzing sequentially randomized trials. *Journal of the National Cancer Institute*, 99, 1577-1582.

Bembom, O. and van der Laan, M. (2008) Analyzing sequentially randomized trials based on causal effect models for realistic individualized treatment rules. *Statistics in Medicine,* 27 (19), 3689-3716.

Gill, R. D. (1989), "Non- and Semi-Parametric Maximum Likelihood Estimators and the von Mises Method," *Scandinavian Journal of Statistics*, 16, 97–128.

Lavori, P. and Dawson, R. (2000) A design for testing clinical strategies: biased individually tailored within-subject randomization. *Journal of the Royal Statistical Society, A* 163, 29–38.

Lavori, P. and Dawson, R. (2004) Dynamic treatment regimes: practical design considerations. *Clinical Trials* 1, 9–20.

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

Millikan, Randall E., Logothetis, Christopher J., Thall, Peter F. (2008) Response:Re: Adaptive Therapy for Androgen-Independent Prostate Cancer: A Randomized Selection Trial of Four Regimens. Journal of the National Cancer Institute, volume 100, pages 682-683

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

Murphy, S.A., van der Laan, Robins, J.M. and CPPRG. (2001) Marginal mean models for dynamic treatment regimes. *Journal of the American Statistical Association*, 96, 1410-1424.

Murphy S.A. (2003) Optimal Dynamic Treatment Regimes (with discussion). *Journal of the Royal Statistical Society, B*, 65, 331-366

Oetting, A.I., Levy, R.D., Weiss, S.A., Murphy, S.A. (2007). Statistical methodology for a SMART design in the development of adaptive treatment strategies, in *Causality and Psychopathology: Finding the Determinants of Disorders and their Cures.* P.E. Shrout, Editors. Arlington, VA: American Psychiatric Publishing, Inc.

Orellana, L., Rotnitzky, A. and Robins, J. (2010). Dynamic Regime Marginal Structural Mean Models for Estimation of Optimal Dynamic Treatment Regimes, Part I: Main Content. The International Journal of Biostatistics: Vol. 6 : Iss. 2, Article 8. DOI: 10.2202/1557-4679.1200

Robins, J.M. (1986) A new approach to causal inference in mortality studies with sustained exposure periods - application to control of the healthy survivor effect. *Mathematical Modeling*, 7, 1393-1512.

Robins, J.M. (1987). Addendum to "A new approach to causal inference in mortality studies with sustained exposure periods - Application to control of the healthy worker survivor effect." *Mathematical Modeling* 14, 923-945

Robins, J. M. (1989) The analysis of randomized and nonrandomized AIDS treatment trials using a new approach to causal inference in longitudinal studies. In Health Service Research Methodology: a Focus on AIDS (eds L. Sechrest, H. Freeman and A. Mulley), pp. 113–159. US Public Health Service.

Robins, J. M. (1993) Analytic methods for estimating HIV treatment and cofactor effects. In: Methodological issues of AIDS Mental Health Research. Eds.: D. G. Ostrow & R. Kessler. Plenum Publishing, New York, pp. 213–290.

Robins JM, Rotnitzky A and Zhao LP. (1994). Estimation of Regression-Coefficients When Some Regressors Are Not Always Observed. *Journal of the American Statistical Association,* 89(427): 846-866.

Robins, J.M. (1997). Causal Inference from Complex Longitudinal Data Latent Variable Modeling and Applications to Causality. Lecture Notes in Statistics (120), M. Berkane, Editor. NY: Springer Verlag, pp. 69-117.

Robins J. M. (1998). Marginal Structural Models. *Proceedings of the American Statistical Association. section on Bayesian Statistics*, pp. 1-10.

Robins, J. M. (2004). Optimal structural nested models for optimal sequential decisions. In D. Y. Lin and P. Heagerty (Eds.), *Proceedings of the Second Seattle Symposium on Biostatistics*, New York. Springer.

Robins, J., Orellana, L. and Rotnitzky, A. (2008). Estimation and extrapolation of optimal treatment and testing strategies. Statistics in Medicine. 27(23):4678-721.

Rubin, D.B., (1978). Bayesian inference for causal effects: The role of randomization. The Annals of Statistics, 6, 34-58.

Tannock, I.F., DeWit, R., Berry, W.R., Horti, J., Pluzanska, A., Chi, K.N. et al. (2004) Docetaxel plus prednisone or mitoxantrone plus prednisone for advanced prostate cancer. *New England Journal of Medicine*, 351(15), 1502-1512.

Thall, P.F., Millikan, R. and Sung, H-G. (2000) Evaluating multiple treatment courses in clinical trials. *Statistics in Medicine*, 19, 1011-1028.

Thall, P.F., Sung, H-G and Estey, E.H. (2002) Selecting therapeutic strategies based on efficacy and death in multi-course clinical trials. *Journal of the American Statistical Association*, 97, 29-39.

Thall, P.F., Logothetis, C., Pagliaro, L., Wen, S., Brown, M.A., Williams, D. and Millikan, R. (2007) Adaptive therapy for androgen independent prostate cancer: A randomized selection trial including four regimens. *Journal of the National Cancer Institute*, 99, 1613-1622.

van der Laan, MJ. (2006) Causal effects models for intention to treat and realistic individualized treatment rules. http://www.bepress.com/ucbbiostat/paper203

van der Laan, Mark J. and Petersen, Maya L. (2007) "Causal Effect Models for Realistic Individualized Treatment and Intention to Treat Rules," *The International Journal of Biostatistics:* Vol. 3 : Iss. 1, Article 3.

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

Wahed, A.S. and Tsiatis, A.A. (2006) Semiparametric efficient estimation of survival distribution for treatment policies in two-stage randomization designs in clinical trials with censored data. *Biometrika*, 93, 163-177.

Table 1: Expert score for the possible combinations of efficacy and toxicity outcomes.

$C_j = c(E_j, T_j)$		E_j = Efficacy Outcome			
		EFF0	EFF1	EFF2	EFF3
T_j =	TOX0	1.0	0.5	0.1	X
Toxicity	TOX1	0.8	0.3	0	X
Outcome	TOX2	X	X	X	0

Table 2: Inverse probability of treatment weights for each possible treatment sequence.

Group	A_1	A_2	A_3	A_4	ω_1	ω_2	ω_3	ω_4	ω
1	a	OFF	OFF	OFF	4	1	1	1	4
2	a	a	OFF	OFF	4	1	1	1	4
3	a	a^*	OFF	OFF	4	3	1	1	12
4	a	a^*	a^*	OFF	4	3	1	1	12
5	a	a	a^*	OFF	4	1	3	1	12
6	a	a	a^*	a^*	4	1	3	1	12

Table 3: IPTW estimated mean scores and log survival times for the twelve viable DTRs. 95% simultaneous confidence intervals are given in parenthesis.

	Binary score	Ordinal score	Expert score	Log-survival
(CVD, KA/VE)	0.41 (0.14, 0.75)	0.50 (0.19, 0.81)	0.61 (0.46, 0.76)	2.95 (2.60, 3.30)
(CVD, TEC)	0.19 (0.00, 0.94)	0.47 (0.21, 0.73)	0.63 (0.49, 0.78)	3.36 (2.96, 3.77)
(CVD, TEE)	0.11 (0.00, 0.99)	0.31 (0.07, 0.55)	0.56 (0.41, 0.71)	2.95 (2.32, 3.59)
(KA/VE, CVD)	0.21 (0.06, 0.52)	0.43 (0.19, 0.67)	0.66 (0.54, 0.78)	3.10 (2.48, 3.72)
(KA/VE, TEC)	0.19 (0.05, 0.49)	0.55 (0.42, 0.69)	0.71 (0.61, 0.81)	2.93 (2.61, 3.25)
(KA/VE, TEE)	0.24 (0.07, 0.56)	0.37 (0.10, 0.65)	0.63 (0.48, 0.78)	2.87 (2.41, 3.33)
(TEC, CVD)	0.44 (0.15, 0.79)	0.67 (0.45, 0.89)	0.78 (0.66, 0.89)	3.08 (2.73, 3.43)
(TEC, KA/VE)	0.46 (0.21, 0.72)	0.57 (0.31, 0.83)	0.73 (0.56, 0.90)	3.26 (2.70, 3.81)
(TEC, TEE)	0.31 (0.13, 0.58)	0.54 (0.35, 0.73)	0.74 (0.63, 0.84)	3.12 (2.78, 3.45)
(TEE, CVD)	0.44 (0.12, 0.81)	0.52 (0.13, 0.90)	0.69 (0.51, 0.87)	3.36 (2.57, 4.16)
(TEE, KA/VE)	0.25 (0.09, 0.55)	0.41 (0.17, 0.65)	0.65 (0.51, 0.79)	2.94 (2.52, 3.36)
(TEE, TEC)	0.37 (0.11, 0.74)	0.53 (0.24, 0.82)	0.70 (0.55, 0.85)	3.05 (2.58, 3.51)

Table 4: Sensitivity analysis: Estimated mean expert score and log survival time using worse case and best case imputation schemes for drop-outs.

	Expert Score[a]	Expert Score[b]	Log Survival[c]	Log Survival[d]
(CVD, KA/VE)	0.62 (0.47, 0.77)	0.62 (0.47, 0.77)	2.93 (2.59, 3.26)	2.92 (2.58, 3.26)
(CVD, TEC)	0.63 (0.49, 0.77)	0.63 (0.48, 0.78)	3.28 (2.88, 3.67)	3.27 (2.85, 3.68)
(CVD, TEE)	0.57 (0.43, 0.71)	0.57 (0.43, 0.71)	2.93 (2.32, 3.54)	2.92 (2.31, 3.53)
(KA/VE, CVD)	0.65 (0.52, 0.77)	0.67 (0.55, 0.80)	3.20 (2.65, 3.76)	3.20 (2.64, 3.77)
(KA/VE, TEC)	0.70 (0.59, 0.81)	0.73 (0.62, 0.84)	3.05 (2.69, 3.41)	3.05 (2.68, 3.42)
(KA/VE, TEE)	0.62 (0.47, 0.77)	0.65 (0.50, 0.80)	3.00 (2.54, 3.46)	3.00 (2.53, 3.47)
(TEC, CVD)	0.77 (0.65, 0.89)	0.77 (0.65, 0.89)	3.02 (2.68, 3.36)	3.18 (2.89, 3.47)
(TEC, KA/VE)	0.72 (0.56, 0.87)	0.72 (0.56, 0.88)	3.13 (2.60, 3.67)	3.31 (2.80, 3.82)
(TEC, TEE)	0.73 (0.62, 0.83)	0.73 (0.62, 0.83)	3.03 (2.63, 3.42)	3.17 (2.83, 3.50)
(TEE, CVD)	0.65 (0.50, 0.80)	0.68 (0.54, 0.83)	3.06 (2.43, 3.69)	3.02 (2.42, 3.63)
(TEE, KA/VE)	0.63 (0.50, 0.75)	0.66 (0.53, 0.79)	2.83 (2.38, 3.28)	2.79 (2.36, 3.23)
(TEE, TEC)	0.67 (0.53, 0.81)	0.71 (0.57, 0.84)	2.87 (2.42, 3.31)	2.83 (2.39, 3.27)

[a] 1 imputed for the dropouts with CVD in the 1st course, and 0 imputed for all other dropouts.

[b] 0 imputed for the dropouts with TEC in the 1st course, and 1 imputed for all other dropouts.

[c] Maximum of the survival time in reference group imputed for dropouts with KA/VE in the 1st course and 1/2 of the minimum remaining survival time imputed for all other dropouts

[d] 1/2 of the minimum remaining survival time in reference group imputed for dropouts with CVD or TEE in the 1st course and Maximum of the survival time imputed for all other dropouts

Figure 1: The possible courses of action prescribed by the per-protocol DTRs: first a, then a^*. The possible per-course responses are as defined in the original protocol. In the parentheses are the numbers of patients (pooled across all DTRs) observed to have a given per-course response sequence history at each given course, and the numbers of patients that have dropped-out from the per-protocol DTR at each course (pooled across all DTRs). For each course, s stands for per-protocol success, and $\bar{s}$ stands for per-protocol failure.

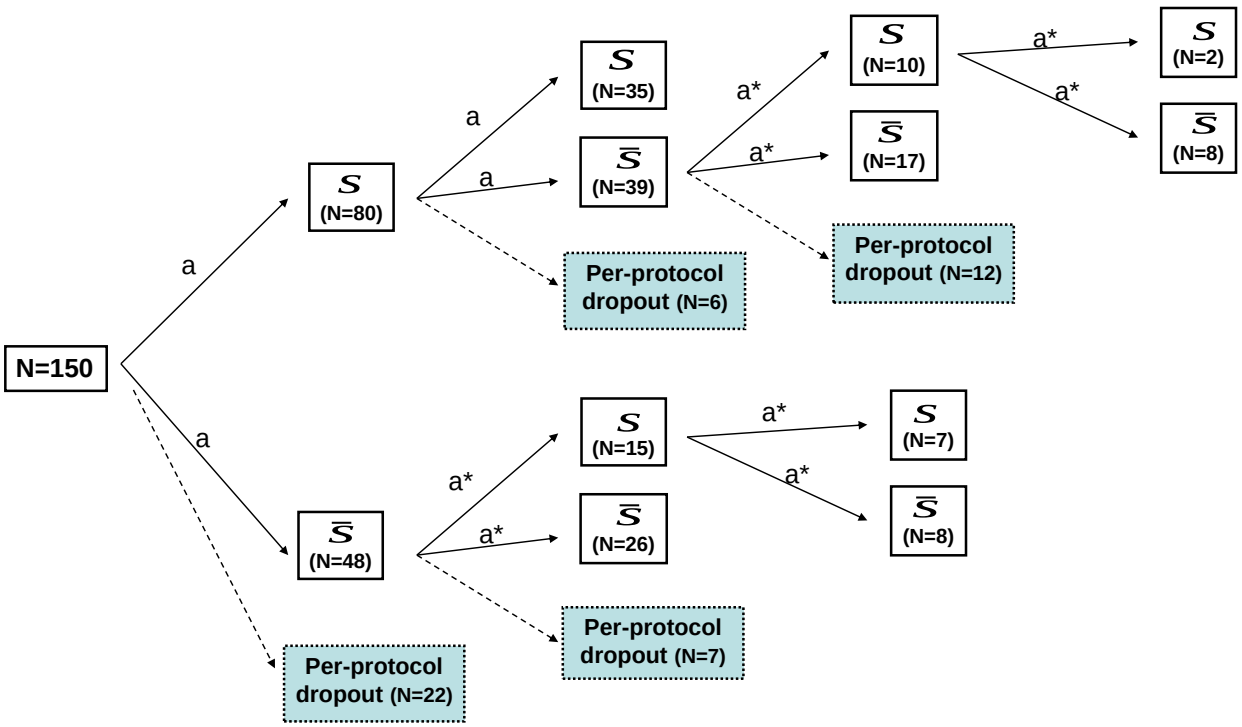

Figure 2: The possible courses of action prescribed by the viable DTRs: first a, then a^*. The possible per-course efficacy and toxicity responses are as defined in section 3. The number of patients (pooled across all DTRs) observed to have a given per-course efficacy and toxicity response sequence history at each given course and the number of patients that have dropped-out from the viable DTR at each course (pooled across all DTRs) are given in parentheses.

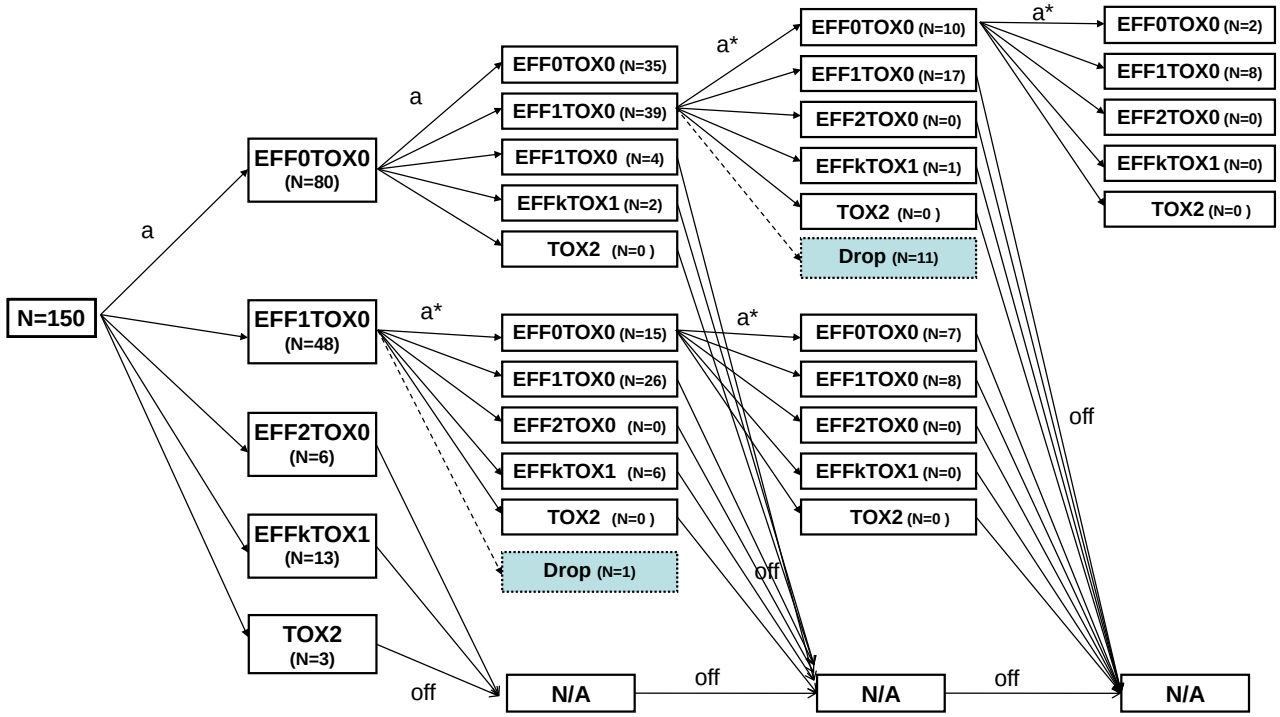

Figure 3: Estimated regime-specific mean scores (binary score, ordinal score, expert score, and modified expert score) for twelve chemotherapy pairs, using the inverse of the estimates of the randomization probabilities and the inverse of the estimates of the drop-out probabilities. The rectangles are the 95% simultaneous confidence intervals, and each middle bar is the estimated counterfactual mean score.

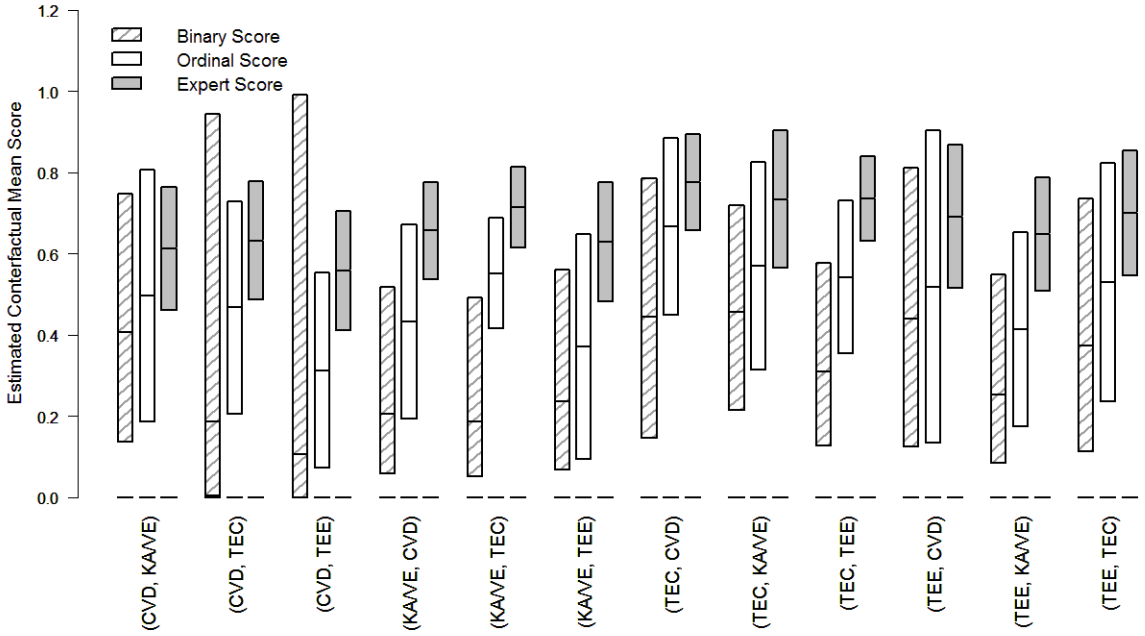

Figure 4: Estimated regime-specific mean log survival time for twelve chemotherapy pairs, using the inverse of the estimates of the randomization probabilities and the inverse of the estimates of the drop-out probabilities. The rectangles are the 95% simultaneous confidence intervals, and each middle bar stands for the estimated counterfactual mean score.