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

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 be initially randomized among four chemotherapies, and that those who responded poorly be re-randomized 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 150 study participants, 47 did not complete their therapy as per the algorithm. However, 35 of them did so for reasons that precluded further chemotherapy, that is, 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 were also used as endpoints. Estimation of each DTR-specific mean score was conducted using inverse probability weighted methods that assumed that missingness in the 12 remaining dropouts was informative but explainable in that it only depended on past recorded data. We conducted additional worst- and best-case analyses to evaluate sensitivity of our findings to extreme departures from the explainable dropout assumption.

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

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 dropout, 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 is 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; Thall, Millikan, and Sung 2000; Thall, Sung, and Estey 2002; Murphy 2005; Oetting et al. 2011). 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, van der Laan, Robins, and CPPRG 2001; van der Laan 2006; van der Laan and Petersen 2007; Robins, Orellana, and Rotnitzky 2008; Orellana, Rotnitzky, and Robin 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

Lu Wang is Assistant Professor, Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109 (E-mail: *luwang@umich.edu*). Andrea Rotnitzky is Professor, Department of Economics, Di Tella University, Buenos Aires, 1425, Argentina (E-mail: *andrea@utdt.edu*) and Adjunct Professor, Department of Biostatistics, Harvard School of Public Health, Boston, MA 02115. Xihong Lin is Professor, Department of Biostatistics, Harvard School of Public Health, Boston, MA 02115 (E-mail: *xlin@hsph.harvard.edu*). Randall E. Millikan is Associate Professor, Department of Genitourinary Medical Oncology, M.D. Anderson Cancer Center, Houston, TX 77030 (E-mail: *rmillika@mdanderson.org*). Peter F. Thall is Professor, Department of Biostatistics, M.D. Anderson Cancer Center, Houston, TX 77030 (E-mail: *rex@mdanderson.org*). 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.

Journal of the American Statistical Association
June 2012, Vol. 107, No. 498, Applications and Case Studies
DOI: 10.1080/01621459.2011.641416

sequentially randomized (SR) designs with randomization probabilities that possibly can depend on past health status and covariates (Lunceford, Davidian, and Tsiatis 2002; Wahed and Tsiatis 2004, 2006; Bembom and van der Laan 2007, 2008).

The purpose of this article 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 sequential multiple assignment randomized 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 or 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 2–4 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 nonrandomized therapeutic or palliative option immediately after the occurrence of a second nonfavorable course, or two consecutive favorable courses, whichever occurred first. The protocol mandated randomization to a second, different chemo immediately after the patient's first nonfavorable 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 are 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 $s_a s_a$, $\bar{s}_a s_{a^*} s_{a^*}$, $s_a \bar{s}_a s_{a^*} s_{a^*}$, $\bar{s}_a \bar{s}_{a^*}$, $\bar{s}_a s_{a^*} \bar{s}_{a^*}$, $s_a \bar{s}_a \bar{s}_{a^*}$ and $s_a \bar{s}_a s_{a^*} \bar{s}_{a^*}$, where s and $\bar{s}$ stand for the per-course favorable and nonfavorable responses, respectively, 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 as 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 as 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 and Eisenberger 2006; Armstrong et al. 2007; Millikan, Logothetis, and Thall 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 (2007) and Thall et al. (2007) assumed that dropouts were noninformative and carried out a complete case analysis, that is, 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 DTRs. 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 dropouts 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 as 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 information not only 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 were discussed in Robins

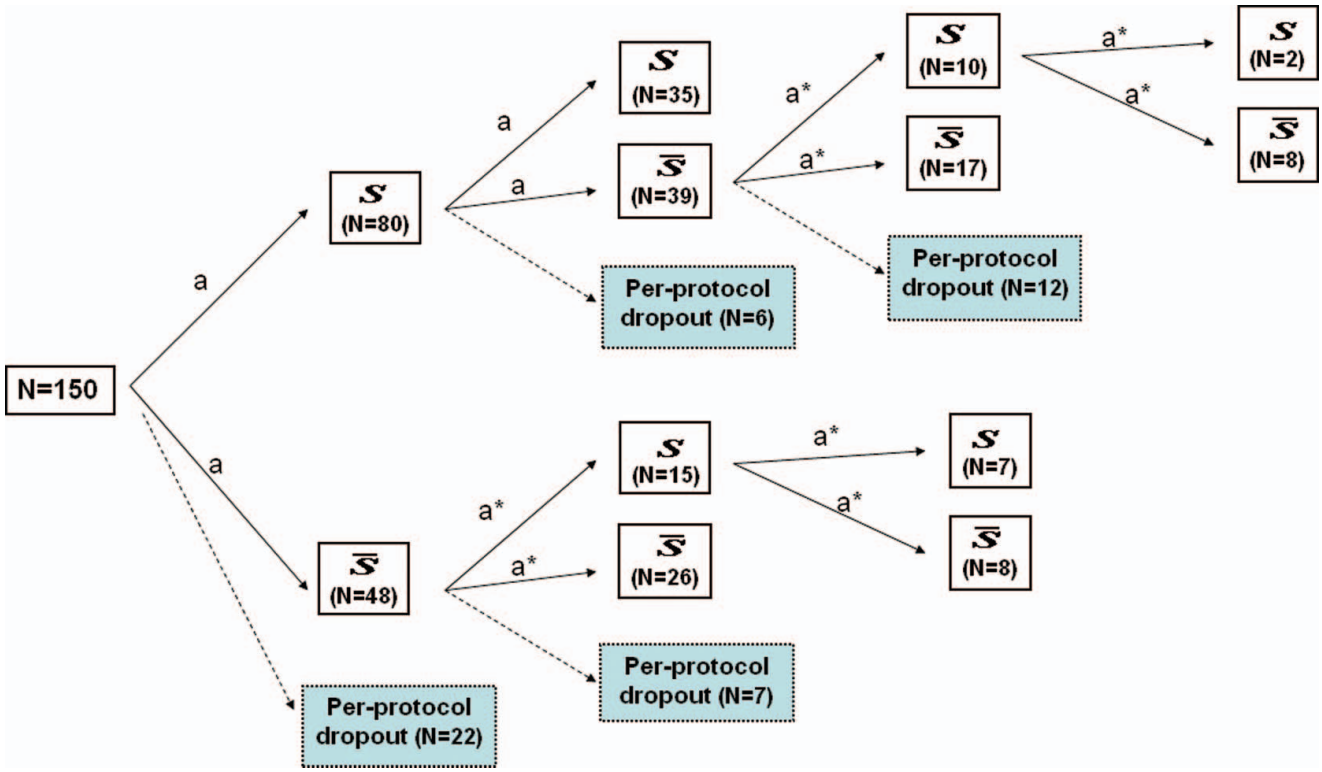

Figure 1. The possible courses of action prescribed by the per-protocol DTRs: first a, then a^*. The possible per-course responses are as defined in the original protocol. In the parentheses are the numbers of patients (pooled across all DTRs) observed to have a given per-course response sequence history 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. The online version of this figure is in color.

(1986, 2004). Robins (2004) called them "feasible" regimes, and van der Laan and Petersen (2007) called 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 prespecified 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 out such analysis because at that time, the extended dataset 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 as 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 12 viable DTRs.

(3) *Dropouts.* Even after redefining 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 redefined 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, dropout. That is, we analyze the data under the assumption that dropout 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 dropout. In addition, we conduct additional worst- and best-case analyses to evaluate the sensitivity of our findings to extreme departures from the preceding assumption.

The remainder of the article 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 DTRs 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 nonfavorable course with it. The rule also stipulates that in the 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 as 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 inverse probability weighted (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 following 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 as per the protocol algorithm. The extended dataset 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, $j = 1, \ldots, 4$, 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 four 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, $j = 2, \ldots, 5$, 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, that is, that X is greater than month $2 \times (j - 1)$,

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

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 who 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 dropouts 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 dropouts for other reasons are present.

We will use the notational convention $\overline{V}_j = (V_1, \ldots, V_j)$ 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 prespecified 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, $g_{a,a^*,j}(\overline{L}_j)$, for $j = 1, 2, 3, 4$. The function $g_{a,a^*,j}(\overline{L}_j)$ 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:

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

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 nonfavorable response without toxicity or PD. The functions $g_{a,a^*,j}$, $j = 1, 2, 3, 4$ are defined as follows:

$$g_{a,a^*,1}(L_1) = a,$$

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

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

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

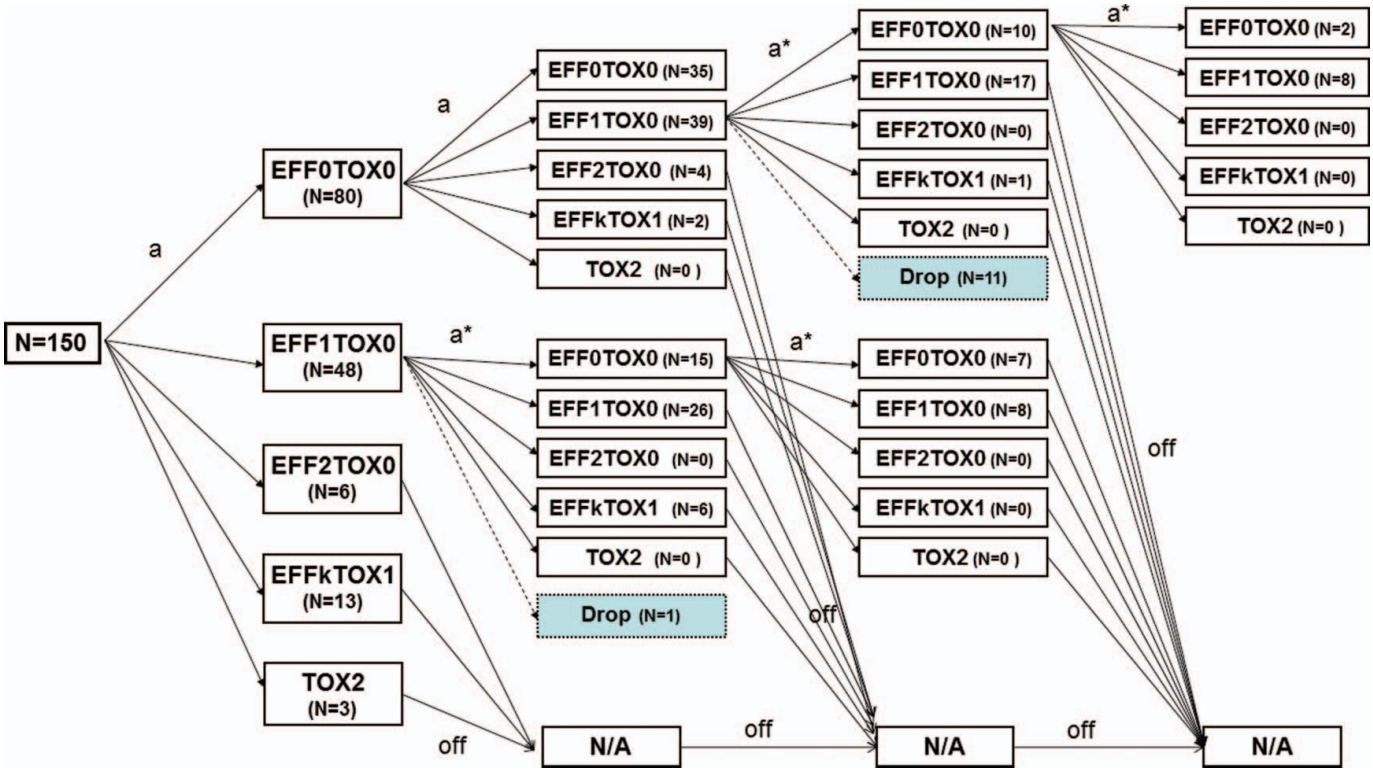

Figure 2. The possible courses of action prescribed by the viable DTRs: first a, then a^*. The possible per-course efficacy and toxicity responses are as defined in Section 3. The number of patients (pooled across all DTRs) observed to have a given per-course efficacy and toxicity response sequence history at each given course and the number of patients who have dropped out from the viable DTR at each course (pooled across all DTRs) are given in parentheses. "EFFkTOX1" stands for toxicity at level 1 and efficacy at any level. The online version of this figure is in color.

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, that is, 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 $g_{a,a^*,2}(\overline{L}_2) =$ 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 $Y = y(\overline{L})$ for specific scoring functions $y(\cdot)$ taking values in the interval [0, 1]. The value taken by $y(\bar{l})$ 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 or 0 otherwise. That is,

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

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 nonconsecutive 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 who 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 nonconsecutive successful courses, were scored as 0.5. Thus, it distinguishes therapies that produce transient efficacy benefits from therapies that do not.

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

Specifically:

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

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 $C_j = c(E_j, T_j)$, $j = 2, \ldots, 5$, 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:

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

Note that $1 - I_{\{\text{OFF,N/A}\}}[A_{j-1}]$ equals 1 if the subject is alive and received a study chemo at the beginning of course $j-1$, or 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 not only distinguishes regimes that provide transient efficacy benefits from those that do not, but 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 afterward. 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^{bin}, Y^{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 $E(U_{(a,a^*)})$ 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 $\mathcal{P} = \{(\overline{L}_{\overline{a}}, X_{\overline{a}}) : \overline{a} \text{ is in the range of } \overline{A}\}$ 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 $\overline{a} = (\text{CVD}, \text{TEC}, \text{OFF}, \text{OFF})$, 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 $Y_{(a,a^*)} = y(\overline{L}_{(a,a^*)})$. 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 $a_1 = a$, $a_2 = g_{a,a^*,2}(\overline{L}_{a_1})$, etc.

In our analysis, we use the mean scores $E[Y_{(a,a^*)}]$ and mean log survival times $E[U_{(a,a^*)}]$ 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 $E[Y_{(a,a^*)}]$ and $E[U_{(a,a^*)}]$ and the optimal switch rules $g_{a_{\text{opt}},a^*_{\text{opt}}}$, where

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

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. (2001) discussed their methods in the context of analyzing follow-up observational data. However, their methods also apply to analysis of 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 $\mathcal{P}_i$, $i = 1, 2, \ldots, n$, 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 $(\overline{L}_j, \overline{A}_{j-1})$ 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 such as 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 composed of prior treatment assignments and recorded outcomes.

In the prostate cancer trial, $\Pr(A_j = a_j | \overline{A}_{j-1} = \overline{a}_{j-1}, \overline{L}_j) = p(a_j | \overline{a}_{j-1}, \overline{L}_j)$, $a_j \in \mathcal{A} \cup \{\text{OFF}\}$, $j = 1, \ldots, 4$, where

$$p(a_1|L_1) = 1/4\{1 - I_{\{\text{OFF}\}}[a_1]\},$$

$$p(a_2|a_1, \overline{L}_2) = \begin{cases} I_{\{a_1\}}(a_2) & \text{if } S_2 = 1 \\ 1/3\{1 - I_{\{a_1\}}[a_2]\} & \text{if } F_2 = 1, \\ I_{\{\text{OFF}\}}[a_2] & \text{if } S_2 \neq 1, F_2 \neq 1, X > 2 \end{cases}$$

$$p(a_3|\overline{a}_2, \overline{L}_3) = \begin{cases} 1/3\{1 - I_{\{a_2\}}[a_3]\} & \text{if } S_2F_3 = 1 \\ I_{\{a_2\}}[a_3] & \text{if } F_2S_3 = 1, \\ I_{\{\text{OFF}\}}[a_3] & \text{if } S_2F_3 \neq 1, F_2S_3 \neq 1, X > 4 \end{cases}$$

$$p(a_4|\overline{a}_3, \overline{L}_4) = \begin{cases} I_{\{a_3\}}[a_4] & \text{if } S_2F_3S_4 = 1 \\ I_{\{\text{OFF}\}}[a_4] & \text{if } S_2F_3S_4 \neq 1, X > 6. \end{cases}$$

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 12 viable switch rules. At the start of the trial any given patient has a positive chance of being assigned to, following any of the 12 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^*} or $\Delta_{a,a^*,i} = 0$ otherwise. Furthermore, for

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

$j = 1, \ldots, 4$, let

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

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. (2001) that under unconfoundedness and positivity,

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

We focus on estimation of $E[Y_{(a,a^*)}]$ 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 $E(\Delta_{a,a^*}\omega Y)$, 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 ω_j, $j = 1, \ldots, 4$. 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 onward, all patients in both groups followed what the rule g_{a,a^*} stipulated. They receive no more weight from this course and onward, that is, 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 nonsuccessful response to the first course but they qualified for a second randomization at course 2 because they did not experience toxicity or PD, that is, 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 $\omega_2 = 1/(1/3) = 1 + 2$ 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, that is, 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 nonsuccessful response to the second course but they did not experience toxicity or PD, that is, 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, that is, 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 who followed the switch rule, that is,

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

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

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 $E[Y_{(a,a^*)}]$ 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 $\{\sum_{i=1}^{n} \Delta_{a,a^*,i} Y_i\}/\{\sum_{i=1}^{n} \Delta_{a,a^*,i}\}$ 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, that is, 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, that is, 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 one-third of them are assigned to the chemo a^*.

4.3 Handling Dropouts

In the trial, one subject who qualified for randomization to a second chemo in the second course and 11 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 "dropouts," 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 dropout from occurring.

However, addressing properly this analytic target raises new challenges because the 12 dropouts departed from the viable rules they were being assigned to follow at the time of dropping out and the decision to dropout 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 dropout.

Formally, let $R_j = \{1 - I_{\{N/A\}}(A_j)\}$, $j = 1, 2, 3, 4$ be the indicator of being neither dead nor a dropout at the start of course j. The rules we wish to compare are composed of decisions at each j for two courses of action (R_j, A_j): the first, with regard to dropout, stipulates that $R_j = 1$ for all subjects alive at the start of course j regardless of their past, that is, dropout 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 $g_{a,a^*}(\overline{L}_j)$ 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

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

$j = 1, \ldots, 4$, 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, either is still on one of the study chemos or has been switched off the study chemos earlier due to toxicity or PD. This probability equals $p(a_j | \overline{a}_{j-1}, \overline{L}_j)$, 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

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

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 dropout as it permits dependence of the probability of dropout on the (likely) correlates of prognosis $\overline{L}_j$ and $\overline{A}_{j-1}$. However, it stipulates that dropout is explained by the measured outcome and treatment history, that is, 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 $E[Y_{(a,a^*)}]$. This is because, for $j = 2$ and 3, the weights

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

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

$$\begin{aligned}\pi_{j,i} = \Pr(R_{j,i} = 1 | R_{j-1,i} = 1, \overline{A}_{j-1} = \overline{A}_{j-1,i}, \overline{L}_{j,i}, \\ X_i > 2(j-1))\end{aligned}$$

are unknown, whereas for $j = 4$, they are equal to 1 because there was no dropout occurring at that course. Thus, we must estimate the dropout 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 dropout probabilities nonparametrically 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 dropout probabilities depend on $(\overline{A}_{j-1}, \overline{L}_j)$ only through the past treatments $\overline{A}_{j-1}$ and on the indicator that PSA dropped over course $j-1$, that is, 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 dropout probabilities is correct, then

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

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

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

where $p(A_{j,i}|\overline{A}_{j-1,i}, \overline{L}_{j,i})$ is the function $p(a_j|\,\overline{a}_{j-1}, \overline{l}_j)$ defined in Section 4.1 and 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 $\widehat{E}[Y_{(a,a^*)}]$, but rather the asymptotically more efficient estimators $\widetilde{E}[Y_{(a,a^*)}]$ that use 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 dropouts, these are not used in the calculation of the estimators of the means $E[U_{(a,a^*)}]$. The estimator $\widehat{E}[U_{(a,a^*)}]$ essentially censors each dropout at the start of the course at which the subject first fails to comply with his assigned regimen and redistributes him among all non-dropouts 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 (e.g., Robins, Rotnitzky, and Zhao 1994). While this may seem paradoxical, it can be understood by noting 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 $\Pr(A_1 = A_{1,i}|L_1 = L_{1,i}) = 1/4$ for each subject i, we choose to replace this probability in the computation of $\omega_{1,i}$ by its estimated value $\widehat{\Pr}(A_1 = A_{1,i}|L_1 = L_{1,i}) = m_{1,i}/n_{1,i}$, the observed proportion of trial participants in the arm of subject i having the same value of disease volume V_1 as him, that is, with $m_{1,i} = \sum_{i=1}^n I_{\{A_{1,i}\}}[A_{1,j}]I_{\{V_{1,i}\}}[V_{1,j}]$ and $n_{1,i} = \sum_{i=1}^n I_{\{V_{1,i}\}}[V_{1,j}]$. 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 $\widehat{E}[Y_{(a,a^*)}]$, that is, 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 nonparametrically estimate the treatment probabilities conditional on all recorded past information, that is, baseline disease volume and PSA 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 12 estimated DTR means were computed as:

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

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 $\Pr(A_1 = A_{1,i}|\overline{L}_{1,i}; \widehat{\gamma})$, where $\widehat{\gamma}$ is the maximum likelihood of the parameters $\gamma = (\gamma_{j,a})_{j\in\{1,2,3\},a\in\{\text{KA/VE,TEC, TEE}\}}$ in the proportional odds model

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

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 dropout probabilities computed as indicated in the preceding section and $\widehat{\lambda}_{j,i}$ are estimators of the treatment probabilities $\Pr(A_j = A_{j,i}|\overline{A}_{j-1} = \overline{A}_{j-1,i}, \overline{L}_j = \overline{L}_{j,i}, X_i > 2(j-1))$ 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 who had one course of chemo, A_1, equal to a^* and who qualified to randomization at course 2,

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

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

A_1 and A_2, equal to a^* and who qualified to randomization at course 3,

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

where a_0(CVD) = KA/VE, a_0(KA/VE) = CVD, a_0(TEC) = CVD, and a_0(TEE) = CVD. Thus, the two models assume that the probability of assignment to a second chemo, say a, among subjects who 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 who 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 who qualified for randomization to a second chemo at course j, we computed $\widehat{\lambda}_{j,i}$ as $\Pr(A_j = A_{j,i} | \overline{A}_{j-1} = \overline{A}_{j-1,i}, \overline{L}_j = \overline{L}_{j,i}, X_i > 2(j-1); \widehat{\alpha})$. 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 = 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

Nonparametric bootstrap standard error estimators can be used to construct regime-specific Wald-type confidence intervals centered at the regime-specific IPTW estimators $\widetilde{E}[Y_{(a,a^*)}]$. The bootstrap produces consistent estimators of the asymptotic variance of regular asymptotically linear (RAL) estimators (Gill 1989) and $\widetilde{E}[Y_{(a,a^*)}]$ is an RAL estimator since its computation involves solving jointly smooth estimating equations for it and for the parameters of the treatment and dropout 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 % of infinitely many hypothetical repetitions of the trial, each of the 12 counterfactual means $E[Y_{(a,a^*)}]$ 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 $100 \times \tau$ % confidence interval corresponding to the regime g_{a,a^*} with the largest estimated mean $E[Y_{(a,a^*)}]$ 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:

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

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

5. DATA ANALYSES

Figure 3 displays plots of the 12 estimators $\widetilde{E}[Y_{(a,a^*)}]$ 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, that is, (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 11 regimes, thus resulting in a 95% confidence set $\mathcal{C}$ for the optimal regime that does not exclude any of the 12 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 12 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.

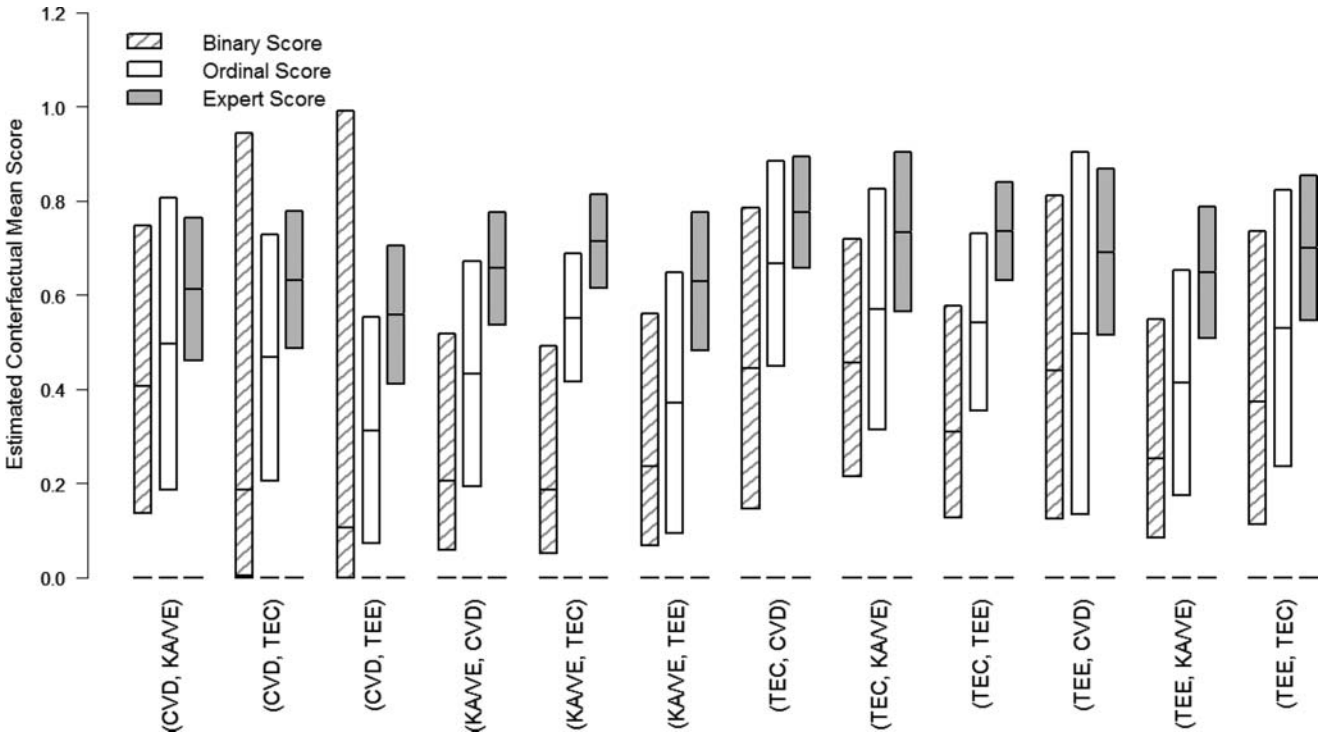

Figure 3. Estimated regime-specific mean scores (binary score, ordinal score, expert score, and modified expert score) for 12 chemotherapy pairs, using the inverse of the estimates of the randomization probabilities and the inverse of the estimates of the dropout probabilities. The rectangles are the 95% simultaneous confidence intervals, and each middle bar is the estimated counterfactual mean score.

However, the 95% confidence set $\mathcal{C}$ for the optimal DTR includes all regimes, and the test at the 95% level that $E[U_{(a,a^*)}]$ 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 dropout process, we conducted four extreme analyses. The first two analyses imputed extreme values for the expert scores of the 12 dropouts and the last two analyses imputed extreme values for the death times of the 12 dropouts as follows.

In the first analysis, for patients who were first assigned to CVD (the stage 1 chemo corresponding to the regime with

Table 3. IPTW estimated mean scores and log survival times for the 12 viable DTRs. 95% simultaneous confidence intervals are given in parentheses

	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)

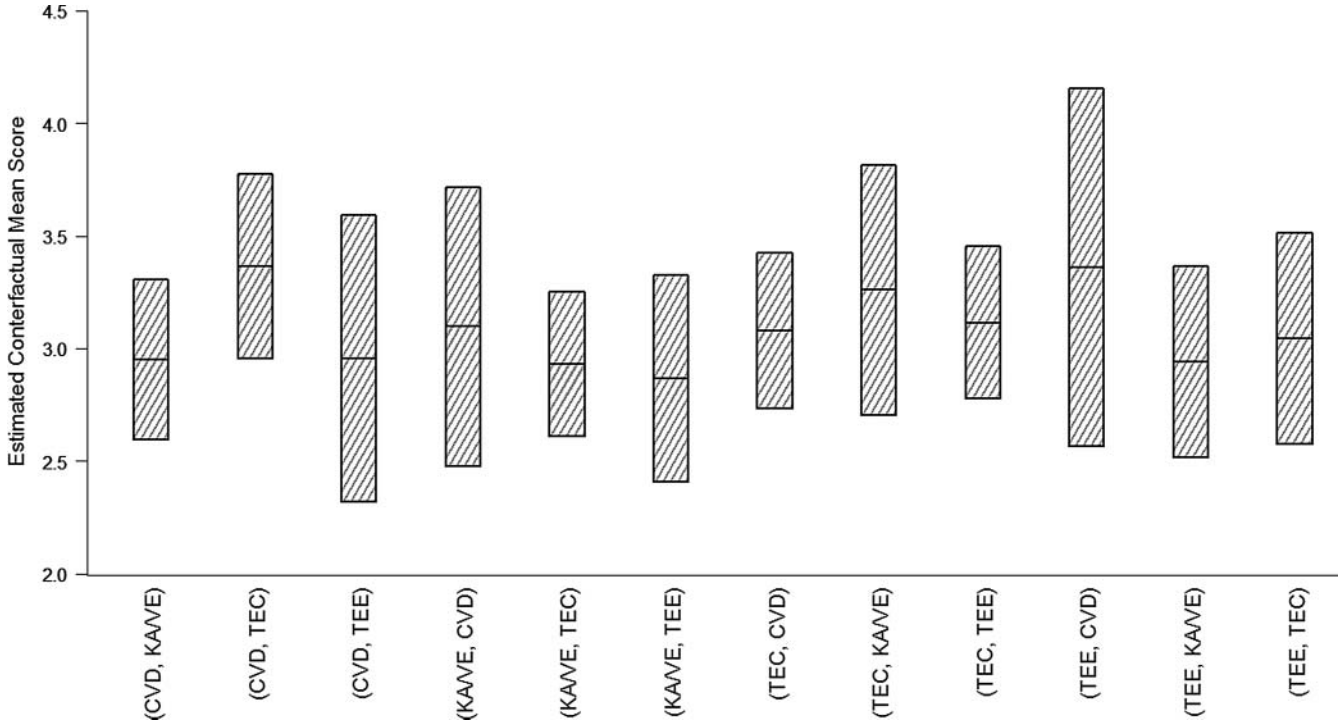

Figure 4. Estimated regime-specific mean log survival time for 12 chemotherapy pairs, using the inverse of the estimates of the randomization probabilities and the inverse of the estimates of the dropout probabilities. The rectangles are the 95% simultaneous confidence intervals, and each middle bar stands for the estimated counterfactual mean score.

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 dropout 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 dropout 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 dropout whose last course of chemo was $j-1$ and who was alive at the end of that course was calculated using

Table 4. Sensitivity analysis: Estimated mean expert score and log survival time using worst-case and best-case imputation schemes for dropouts. 95% simultaneous confidence intervals are given in parentheses

	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)

NOTE: [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 0.5 of the minimum remaining survival time imputed for all other dropouts. [d]Half 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.

the survival times of subjects who 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 dropout, throughout referred to as the reference group. In the third analysis, each dropout 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 dropouts, each survival time was computed by adding to $2(j-1)$, where $j-1$ was the last course of chemo received by the given dropout, 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 dropouts who were first assigned to TEE and CVD (the stage 1 chemos corresponding to the regimes with largest estimated mean log survival) and imputed the survival times of all remaining dropouts 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 dropout. 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 article, 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 SR trials mimic the way in which 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 DTRs 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 PI of the trial. We thus redefined patient outcome as a more informative compound event combining information on both efficacy and toxicity. The remaining noncompliers (12 patients) were assumed to have followed a possibly informative, but explainable, dropout 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 DTRs. 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 (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), nine patients who developed severe toxicity or severe PD were excluded by Bembom and van der Laan (2007). After redefining the compound endpoint, only one of these nine patients is still missing. The other eight 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, it had 150 patients. The sample size is far too small to draw confirmatory conclusions comparing the 12 treatment pairs. Sample size calculations for SR trials are important to provide practical guidance (see Murphy 2005; Feng and Wahed 2009; Dawson and Lavori 2012). Our proposal to construct simultaneous confidence intervals 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.

APPENDIX

To compute the simultaneous confidence intervals for the 12 means $E[Y_{(a,a^*)}]$, 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 $E[Y_{(a,a^*)}]$ and estimated means $\widetilde{E}[Y_{(a,a^*)}]$ respectively, we know that

$$\sqrt{n}\,(\widetilde{\mu} - \mu) \to N\,(0, \Sigma)\,.$$

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 dropout model.

Then, $\max\{|\frac{\widetilde{\mu}_j - \mu_j}{\sqrt{\Sigma_{jj}/n}}|;\, j = 1, \ldots, 12\}$ is distributed like $Z_{\max} = \max\{|Z_j|;\, j = 1, \ldots, 12\}$ where $Z = (Z_1, \ldots, Z_{12}) \sim N(0, \Omega)$ with $\Omega = \mathrm{diag}(\Sigma)^{-1/2}\Sigma\mathrm{diag}(\Sigma)^{-1/2}$. 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

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

In the construction of our confidence interval, we replaced Σ_{jj}/n with the nonparametric bootstrap estimator $V_{\text{boot},j}$ of the asymptotic variance of $\widetilde{\mu}_j$ using 1000 bootstrap replications. We also used a Monte Carlo procedure to compute an estimate $\widehat{z}_{\max,0.95}$ of the unknown value $z_{\max,0.95}$. Specifically, we computed the nonparametric bootstrap estimator $\widehat{\Omega}_{boot}$ of the correlation matrix Ω. Next, we generated $\widehat{Z}_k \overset{\text{iid}}{\sim} N(0, \widehat{\Omega}_{\text{boot}}), k = 1, \ldots, 10,000$, and for each k, we computed $Z_{\max,k} = \max\{|Z_{k,j}|; j = 1, \ldots, 12\}$. Finally, we computed $\widehat{z}_{\max,0.95}$ as the 95th percentile of the empirical distribution of $Z_{\max,k}, k = 1, \ldots, 10,000$ and calculated the confidence intervals as $(\widetilde{\mu}_j - \widehat{z}_{\max,0.95}\sqrt{V_{\text{boot},j}}, \widetilde{\mu}_j + \widehat{z}_{\max,0.95}\sqrt{V_{\text{boot},j}})$.

[Received March 2010. Revised October 2011.]

REFERENCES

Armstrong, A. J., Garrett-Mayer, E. S., Yang, Y. C., DeWit, R., Tannock, I. F., and Eisenberger, M. (2007), "A Contemporary Prognostic Nomogram for Men with Hormone-Refractory Metastatic Prostate Cancer: A TAX327 Study Analysis," *Clinical Cancer Research*, 13(21), 6396–403. [494]

Armstrong, A. J., and Eisenberger, M. A. (2006), "Does Gonadotropin-Releasing-Hormone Agonist Therapy for Prostate Cancer Increase the Risk of Fracture?," *Natural Clinical Practice Urology*, 3(5), 246–7. [494]

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. [493,494,496,507]

——— (2008), "Analyzing Sequentially Randomized Trials Based on Causal Effect Models for Realistic Individualized Treatment Rules," *Statistics in Medicine*, 27(19), 3689–3716. [493,494,495,496,504,507]

Dawson, R., and Lavori, P. W. (2012), "Efficient Design and Inference for Multistage Randomized Trials of Individualized Treatment Policies," *Biostatistics*, 13, 142–152. [507]

Gill, R. D. (1989), "Non- and Semi-Parametric Maximum Likelihood Estimators and the von Mises Method," *Scandinavian Journal of Statistics*, 16, 97–128. [504]

Lavori, P., and Dawson, R. (2000), "A Design for Testing Clinical Strategies: Biased Individually Tailored Within-Subject Randomization," *Journal of the Royal Statistical Society*, Series A, 163, 29–38. [493]

——— (2004), "Dynamic Treatment Regimes: Practical Design Considerations," *Clinical Trials*, 1, 9–20. [493]

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. [493]

Millikan, R. E., Logothetis, C. J., and Thall, P. F. (2008), "Response: Adaptive Therapy for Androgen-Independent Prostate Cancer: A Randomized Selection Trial of Four Regimens," *Journal of the National Cancer Institute*, 100, 682–683. [494]

Murphy, S. A. (2003), "Optimal Dynamic Treatment Regimes (with discussion)," *Journal of the Royal Statistical Society*, Series B, 65, 331–366. [493,494,495,507]

——— (2005), "An Experimental Design for The Development of Adaptive Treatment Strategies," *Statistics in Medicine*, 24, 1455–1481. [493,500]

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. [493]

Oetting, A. I., Levy, R. D., Weiss, S. A., and Murphy, S. A. (2011), "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*, ed. P. E. Shrout, Arlington, VA: American Psychiatric Publishing, pp. 175–205. [493]

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," *International Journal of Biostatistics*, 6(2), Article 8. DOI: 10.2202/1557-4679.1200 [493]

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. [493,494,496,499]

——— (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. [499]

——— (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, Rockville, MD: U.S. Public Health Service, pp. 113–159. [493]

——— (1993), "Analytic Methods for Estimating HIV Treatment and Cofactor Effects," in *Methodological issues of AIDS Mental Health Research.* eds. D. G. Ostrow, and R. Kessler, New York: Plenum Publishing, pp. 213–290. [493,500]

——— (1997), "Causal Inference from Complex Longitudinal Data," in *Latent Variable Modeling and Applications to Causality.* ed. M. Berkane, New York: Springer Verlag, pp. 69–117. [493]

——— (2004), "Optimal Structural Nested Models for Optimal Sequential Decisions," in *Proceedings of the Second Seattle Symposium on Biostatistics*, eds. D. Y. Lin, and P. Heagerty, New York: Springer, pp. 189–326. [493]

Robins, J., Orellana, L., and Rotnitzky, A. (2008), "Estimation and Extrapolation of Optimal Treatment and Testing Strategies," *Statistics in Medicine*, 27(23), 4678–721. [494]

Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1994), "Estimation of Regression-Coefficients When Some Regressors Are Not Always Observed," *Journal of the American Statistical Association*, 89(427), 846–866. [503]

Rubin, D. B. (1978), "Bayesian Inference for Causal Effects: The Role of Randomization," *Annals of Statistics*, 6, 34–58. [499]

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. [494]

Thall, P. F., Millikan, R., and Sung, H. -G. (2000), "Evaluating Multiple Treatment Courses in Clinical Trials," *Statistics in Medicine*, 19, 1011–1028. [493,494,496]

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. [493]

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. [494,497,507]

van der Laan, M. J. (2006), "Causal Effects Models for Intention to Treat and Realistic Individualized Treatment Rules," *http://www.bepress.com/ucbbiostat/paper203* [493]

van der Laan, M. J., and Petersen, Maya L. (2007), "Causal Effect Models for Realistic Individualized Treatment and Intention to Treat Rules," *The International Journal of Biostatistics*, 3 (1), Article 3. DOI: 10.2202/1557-4679.1022. [493,494,496]

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. [493]

——— (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. [493]

Comment

Daniel ALMIRALL, Daniel J. LIZOTTE, and Susan A. MURPHY

1. INTRODUCTION

Sequential treatments, in which treatments are adapted over time based on the changing clinical status of the patient, are often necessary because treatment effects are heterogeneous across patients: not all patients will respond (similarly) to treatment, calling for changes in treatment to achieve an acute response or to place all patients on a positive health trajectory. Further, a treatment that is effective now for one patient may not work as well in the future for the same patient, again necessitating a sequence of treatments. Moreover, it is often necessary to balance benefits (e.g., symptom reduction) with burden (e.g., toxicity), a trade-off that may unfold over time. As a result, in clinical practice clinicians often find themselves implicitly or explicitly using a sequence of treatments with the goal of optimizing both short- and long-term outcomes, or, as may be the case in cancer treatment, to prevent death. Dynamic treatment regimes (DTRs) operationalize such sequential decision making. A DTR individualizes treatment over time via decision rules that specify whether, how, or when to alter the intensity, type, or delivery of treatment at critical clinical decision points. Sequential multiple assignment randomized trials (SMARTs) or equivalently, sequentially randomized trials, have been developed explicitly for the purpose of constructing proposals for high-quality DTRs.

In the article "Evaluation of Viable Dynamic Treatment Regimes in a Sequentially Randomized Trial of Advanced Prostate Cancer," Wang et al. (2012, hereinafter WRLMT) provide an excellent and lucid re-analysis of data from a SMART study and both motivate and encourage a discussion about design and analysis issues around SMARTs. In our comment, we focus on two important ideas raised by WRLMT: (1) the *design* of SMARTs (as opposed to the *analysis* of SMARTs), and (2) the analysis of, and presentation of results based on, multiple outcomes.

2. DESIGNING SMART STUDIES

2.1 Ensuring Viable Embedded DTRs in the Design of a SMART

It is critically important to ensure *prior* to the conduct of a SMART that the DTRs embedded within it are indeed viable. A first step to ensure that the embedded DTRs are viable is a clear operationalization of the embedded tailoring variable used to restrict subsequent treatment options within the SMART. Often, the embedded tailoring variable is a well-operationalized notion of early (or "in treatment") response and nonresponse. WRLMT use the phrase "course-specific success or failure." In this regard, the efforts of the prostate SMART study designers are commendable: As noted in WRLMT and in more detail in Thall et al. (2007), after an eight-week course of initial treatment, success (versus failure) was defined as a decline of at least 40% in prostate-specific antigen (PSA), no regression of any magnitude on any measurable disease dimension, no symptom increase in pain, anorexia, asthenia, or cachexia, and no new lesions or new cancer-related symptoms. Further, criteria for scoring success after being offered the second course of the same treatment were also clearly operationalized.

A second important step to ensure that the embedded DTRs are viable is a clear operationalization of how to treat patients in the event that additional common contingencies (e.g., beyond what is typically thought of as course-specific success or failure) arise during treatment. Such contingencies may include intolerable side effects (such as toxicity in the treatment of cancer, or weight gain in the treatment of schizophrenia), excessive treatment burden (such as is possible with preventive and behavioral interventions), and treatment drop-out or refusal to receive subsequent treatment (such as may happen with any intervention). Indeed, in our experience, the clinical trial protocol (including the materials provided to either the Data Safety Monitoring Board and/or Institutional Review Boards) will detail a plan for what will happen in the event that any of these common contingencies arise. Often, the plan may be a transfer of the patient to "treatment as usual by patient's clinician," or in some settings, the plan may include a behavioral therapy aimed at re-engagement of the patient in treatment. The embedded DTRs are only viable if they incorporate the trial protocol plans for these commonly occurring contingencies.

In their re-analysis of the prostate SMART, WRLMT report that among the 47 participants who did not complete their therapy according to the 12 originally conceived DTRs embedded within the SMART, 35 of them did so due to severe toxicity or progressive disease (PD). WRLMT note that in actual oncology practice, severe toxicity or PD preclude further chemotherapy for patients with advanced prostate cancer and, instead, indicate a therapeutic or palliative treatment of some sort. They further note that this is precisely what was discovered to have happened during the conduct of the prostate SMART.

Given the relatively large proportion ($35/150 = 23\%$) of participants who were affected by PD or toxicity so severe as to preclude chemotherapy, we suspect that the trial protocol likely detailed a plan (i.e., the provision of therapeutic or palliative care for those with severe toxicity or PD) for these contingencies.

In other SMART studies, a common contingency is that a patient misses the clinic visit during which course-specific

Daniel Almirall is Faculty Research Fellow, Institute for Social Research, University of Michigan, Ann Arbor, MI 48106-1248 (E-mail: *dalmiral@umich.edu*). Daniel J. Lizotte is Assistant Professor, David R. Cheriton School of Computer Science, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1 (E-mail: *dlizotte@uwaterloo.ca*). Susan A. Murphy is Professor of Statistics and Psychiatry, and Research Scientist at the Institute for Social Research, University of Michigan, Ann Arbor, MI 48106-1248 (E-mail: *samurphy@umich.edu*). Funding for this work was provided by the grants R01-MH-080015 (Murphy) and P50-DA-010075 (Murphy).

Journal of the American Statistical Association
June 2012, Vol. 107, No. 498, Applications and Case Studies
DOI: 10.1080/01621459.2012.665615

success or failure is assessed. If the patient returns to the clinic for treatment beyond the window of time during which subsequent (originally planned) treatments are appropriate, a clear alternative treatment plan is necessary to make the embedded DTRs viable. This issue is particularly important in the design of SMARTs because the trial design depends on course-specific success or failure.

To summarize, common contingencies that may arise during treatment require a prespecified treatment plan and the absence of such a plan may lead to the consideration of nonviable embedded DTRs. Of course, we do not mean to imply that the embedded DTRs need to be tailored to any and all contingencies that may arise.

The primary consequence of having nonviable DTRs in the design of a SMART is that it becomes unclear what DTRs the SMART is designed to make inferences about. As a result, the effects of the proposed DTRs resulting from the study will be less replicable as it will be unclear to future investigators how these contingencies were handled in the trial. Since the primary motivation to conduct a SMART in the first place is to inform decision making concerning the sequencing of treatments, this is a consequential omission in the design of SMARTs. Thus, the consideration of additional contingencies as part of the embedded DTRs is not merely a quality control measure or an ethical concern. Rather, the treatment plan following one of these contingencies is part of the *definition* of the embedded DTRs. Without accounting for these additional contingencies, as we learned from WRLMT, we do not have realistic DTRs that are applicable to the population of interest and, therefore, useful in informing sequential decision making.

Our message that it is critical to ensure that embedded DTRs are viable during the design phase of a SMART is not new. Indeed, as emphasized in the recent National Research Council's study report on "The Prevention and Treatment of Missing Data in Clinical Trials, 2010," the consideration of viable interventions *should* be a basic tenet of all RCT designs. However, unlike standard RCTs, SMARTs force us to confront this issue.

Since the primary focus of standard RCTs is often on initial treatment offerings, this means that even when there exist (or should exist) treatment plans for common contingencies that may arise, they are not always explicitly considered to be part of the definition of treatment. Instead, these common contingencies are often considered treatment outcomes. We note that while the rate at which these contingencies occur is a treatment outcome, the plan for how to treat participants in these situations is part of the definition of treatment, a subtle but important distinction.

In contrast, since SMARTs are explicit in their aim to develop DTRs, the issue of what to do next given common contingencies that may arise (even beyond what is considered course-specific success or failure) is less easily "swept under the rug."

2.2 Matching the Statistical Analysis to the Rationale for a SMART

In our experience in designing SMART studies, the overarching goal of the study is to construct one or more proposals for high-quality DTRs. These proposals would then be combined with the results of other studies and emerging science to produce DTR(s) that would form one or more of the intervention arms in a future randomized confirmatory trial. Thus, the goal of a SMART is often quite different from the more "confirmatory" goal of most standard RCTs.

This appears to be the case in the prostate SMART considered here, as WRLMT write: "The ultimate goal was to use the results of the trial as a basis for generating hypotheses and planning a future, confirmatory trial."

This goal is quite similar to the goal of randomized factorial designs used in engineering (Box, Hunter, and Hunter 1978) and its emerging use in the development of behaviorial interventions (Collins et al. 2005; Collins, Murphy, and Strecher 2007; Collins et al. 2009, 2011; Strecher et al. 2008; Chakraborty and Murphy 2009). Indeed, SMARTs can be viewed as sequentially randomized factorial designs (Murphy and Bingham 2009).

Similar to the use of factorial designs in engineering, SMARTs are intended to aid in the construction of a multicomponent intervention (namely, a DTR) as opposed to confirm a best DTR. Accordingly, the statistical analysis of a SMART need not have a confirmatory flavor. For example, in these factorial designs, investigators might not conduct hypothesis tests. Instead, investigators might rank order the treatment/intervention factors in terms of estimated effect sizes and keep the x most highly ranked factors; similar "ranking and selection" ideas have been proposed in the clinical trial literature (Simon, Wittes, and Ellenberg 1985; Sargent and Goldberg 2001) as well. If hypothesis tests are used, the focus is on reducing the Type II error as opposed to controlling the Type I error rate.

For example, scientists might test a small number of prespecified hypothesis, each at a specified, marginal, significance level, and then control the overall error rate of the remaining hypothesis (Collins, Murphy, and Strecher 2007; Chakraborty and Murphy 2009).

Despite the fact that the prostate SMART study discussed by WRLMT appears to be focused on constructing proposals for high quality DTRs, WRLMT control the experiment-wise error rate (i.e., they construct simultaneous confidence intervals). It is thus easy to misinterpret the use of simultaneous confidence intervals as implying that the trial was intended to be confirmatory. We maintain that regime-specific confidence intervals should and can be used to conduct inference in building a high quality DTR. Of course, these nonsimultaneous confidence intervals will not confirm that one regime is best. In fact, much of the current work on sample size planning for SMARTs (Feng and Wahed 2008, 2009; Oetting et al. 2010; Li and Murphy 2011) does not focus on devising sample size formulas that control the experiment-wise error rate. Rather, the focus has been on formulas that control the Type I error of a prespecified primary aim that aides in building an effective DTR. We acknowledge that SMARTs can be designed to confirm which of the embedded DTRs is best; indeed in some settings, such as in the development of internet-based interventions where large sample sizes are inexpensively obtained, this approach may be desirable and is likely feasible.

3. COMPOSITE ENDPOINTS

Our second comment describes a new approach that can be used in addition to endpoints such as those considered by

WRLMT to further quantify the trade-off between toxicity and efficacy.

In Section 3.3, WRLMT discuss three composite endpoints (Y^{binary}, Y^{ordinal}, and Y^{expert}) of treatment efficacy and toxicity. These three composite endpoints, as well as log-survival time, were used in the re-analysis by WRLMT. The three endpoints exploit newly available toxicity and efficacy data (not available during the primary analysis of the trial) that the authors show serve as better surrogates for the overall health status of patients over the entire 32 week treatment period compared to the endpoints stipulated by the original study protocol.

The three new endpoints differ in terms of how they trade-off toxicity and efficacy during chemotherapy. We suspect this is important in oncology because the most efficacious chemotherapies are likely the treatments associated with high levels of toxicity and vice-versa, that is, some chemotherapies that are less effective may also be the ones that are also less toxic. This question, of how to trade-off opposing outcomes, also arises in other areas of clinical research such as in the treatment of schizophrenia where the trade-off is between symptom relief and weight gain.

As implied by the authors, the choice of the endpoint Y necessarily influences the conclusions drawn from the study, but the "correct" choice of the endpoint is often not obvious or even well-defined.

The authors note that Y^{ordinal} is in a sense finer-grained than Y^{binary} because it "distinguishes therapies that produce transient efficacy benefits from therapies that don't." They further note that Y^{expert} "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."

In examining the results in WRLMT, we can see that using all-or-nothing "success" (i.e., Y^{binary}) as the desired outcome produces a different estimated optimal DTR than the result obtained when considering (possibly transient) efficacy benefits: the results show that the estimated optimal DTR according to Y^{binary} is (TEC, KA/VE), while the estimated optimal DTR according to Y^{ordinal} or Y^{expert} is (TEC, CVD).

As a possible adjunct to the high-quality analysis presented in the article by WRLMT, we can conduct a sensitivity analysis with respect to the endpoint definition that provides further insight into how the results would change depending on the relative utilities of different joint outcomes. For example, this can be done by considering outcomes $Y^*_{(a,a^*)}(\delta) = (1-\delta)\cdot Y^{\text{bin}}_{(a,a^*)} + \delta \cdot Y^{\text{expert}}_{(a,a^*)}$, where $\delta \in [0,1]$ is used to interpolate between the binary outcome score and the expert outcome score. Note that for $\delta = 0$, we recover the binary outcome score, whereas for $\delta = 1$ we recover the expert outcome score, and for intermediate values of δ, we define an outcome that is a combination of both. The different outcomes indexed by δ under this framework represent not just different levels of granularity, but also different sets of *preferences* for $Y^{\text{expert}}_{(a,a^*)}$ vs $Y^{\text{bin}}_{(a,a^*)}$.

For this example, since the authors' estimate of the DTR means, $\widetilde{E}[Y_{(a,a^*)}]$, is linear in the observed outcome scores Y_i, we can use the estimates provided in the paper to compute DTR means for $Y^*_{(a,a^*)}(\delta)$. Figure 1 illustrates how the estimated DTR means change as a function of δ.

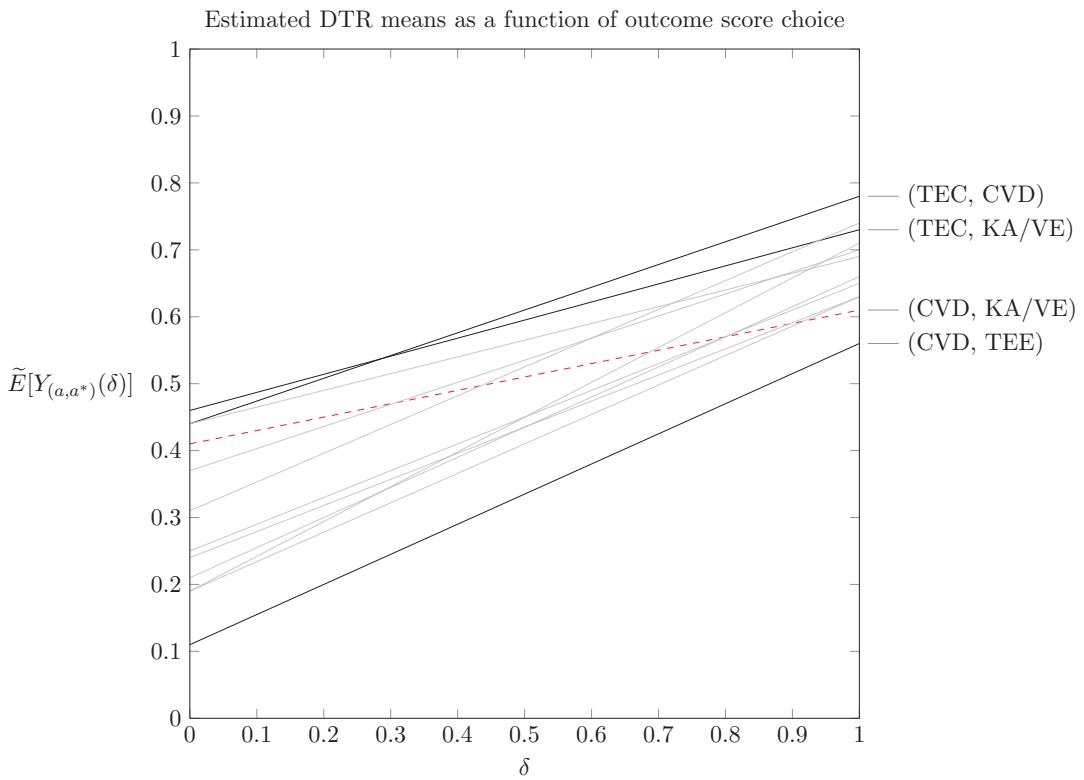

Figure 1. Estimated DTR means as a function of δ, where $Y_{(a,a^*)}(\delta) = (1-\delta)\cdot Y^{\text{bin}}_{(a,a^*)} + \delta \cdot Y^{\text{expert}}_{(a,a^*)}$. The online version of this figure is in color.

As expected, we see that (TEC, KA/VE) is preferable for Y^{bin} ($\delta = 0$), whereas (TEC, CVD) is preferable for Y^{expert} ($\delta = 1$). The cross-over point, at $\delta \approx 0.28$, helps us understand at what preference point the results will differ. Such an analysis is useful since different patients or clinicians may have different preferences in how they trade-off Y^{expert} versus Y^{bin}. Figure 1 also shows that (CVD, TEE) appears the worse no matter what the trade-off. Further, we see that some DTRs change ranking substantially for different outcomes: for example, (CVD, KA/VE) goes from 4th for Y^{bin} to second-last for Y^{expert}.

One could also imagine doing a similar analysis that trades off two different expert scores, for example $Y^{\text{expert}-1}_{(a,a^*)}$ versus $Y^{\text{expert}-2}_{(a,a^*)}$, which may represent two opposing views (operationalized by different choices for C_j) on how to trade-off efficacy versus toxicity. Or one could imagine a similar analysis that trades off two continuous, direct measures of toxicity versus efficacy, such as $Y_{(a,a^*)}(\delta) = (1-\delta)\cdot T(a,a^*) + \delta \cdot E(a,a^*)$. The latter may only be possible if toxicity and efficacy can be placed on "similar footing" so that a linear convex trade-off of this sort is clinically meaningful. This could be done by first "calibrating" or "scaling" the measures of T and E to ensure the linear combination is meaningful so that, for example, $\delta = 0.5$ represents a moderate or typical clinical preference for one outcome over the other. This last idea may not be possible in oncology research since, as noted by WRLMT, severe toxicity may be so highly undesirable that no level of efficacy (no matter how high) could trade-off with it.

This approach to trading off two or more opposing outcomes is being developed further for use in data analyses that build optimal DTRs (say, using Q-Learning) from data arising from SMARTs (Lizotte, Bowling, and Murphy 2010).

[Received January 2012. Revised January 2012.]

REFERENCES

Box, G., Hunter, W., and Hunter, J. (1978), *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building*, New York: Wiley. [510]

Chakraborty, B., Collins, L., Strecher, V., and Murphy, S. (2009), "Developing Multicomponent Interventions Using Fractional Factorial Designs," *Statistics in Medicine*, 28(21), 2687–2708. [510]

Collins, L., Baker, T., Mermelstein, R., Piper, M., Jorenby, D., Smith, S., Christiansen, B., Schlam, T., Cook, J., and Fiore, M. (2011), "The Multiphase Optimization Strategy for Engineering Effective Tobacco Use Interventions," *Annals of Behavioral Medicine*, 41(2), 208–226. [510]

Collins, L., Chakraborty, B., Murphy, S., and Strecher, V. (2009), "Comparison of a Phased Experimental Approach and a Single Randomized Clinical Trial for Developing Multicomponent Behavioral Interventions," *Clinical Trials*, 6, 5–15. [510]

Collins, L., Murphy, S., Nair, V., and Strecher, V. (2005), "A Strategy for Optimizing and Evaluating Behavioral Interventions," *Annals of Behavioral Medicine*, 30, 65–73. [510]

Collins, L., Murphy, S., and Strecher, V. (2007), "The Multiphase Optimization Strategy (MOST) and the Sequential Multiple Assignment Randomized Trial (SMART): New Methods for More Potent e-health Interventions," *American Journal of Preventive Medicine*, 32, S112–S118. [510]

Feng, W., and Wahed, A. (2008), "Supremum Weighted Log-Rank Test and Sample Size for Comparing Two-Stage Adaptive Treatment Strategies," *Biometrika*, 95(3), 695–707. [510]

——— (2009), "Sample Size for Two-Stage Studies with Maintenance Therapy," *Statistics in Medicine*, 28, 2028–2041. [510]

Li, Z., and Murphy, S. (2011), "Sample Size Formulas for Two-Stage Randomized Trials with Survival Outcomes," *Biometrika*, 98(3), 503–518. [510]

Lizotte, D., Bowling, M., and Murphy, S. (2010), "Efficient Reinforcement Learning with Multiple Reward Functions for Randomized Clinical Trial Analysis," *Proceedings of the Twenty-Seventh International Conference on Machine Learning (ICML)*, pp. 695–702. [512]

Murphy, S., and Bingham, D. (2009), "Screening Experiments for Developing Dynamic Treatment Regimes," *Journal of the American Statistical Association*, 184, 391–408. [510]

National Research Council. (2010), *The Prevention and Treatment of Missing Data in Clinical Trials*, Washington, DC: The National Academies Press. [510]

Oetting, A., Levy, J., Weiss, R., and Murphy, S. (2010), "Statistical Methodology for a SMART Design in the Development of Adaptive Treatment Strategies," in *Causality and Psychopathology*, eds. P. Shrout, K. Keyes, and K. Ornstein, New York: Oxford University Press, pp. 179–205. [510]

Sargent, D., and Goldberg, R. (2001), "A Flexible Design for Multiple Armed Screening Trials," *Statistics in Medicine*, 20(7), 1051–1060. [510]

Simon, R., Wittes, R., and Ellenberg, S. (1985), "Randomized Phase II Clinical Trials," *Cancer Treatment Reports*, 69(12), 1375–1381. [510]

Strecher, V., McClure, J., Alexander, G., Chakraborty, B., Nair, V., Konkel, J., Greene, S., Collins, L., Carlier, C., Wiese, C., Little, R., Pomerleau, C., and Pomerleau, O. (2008), "Web-based Smoking Cessation Programs: Results of a Randomized Trial," *American Journal of Preventive Medicine*, 34, 373–381. [510]

Thall, P., Logothetis, C., Pagliaro, L., Wen, S., Brown, M., Williams, D., and Millikan, R. (2007), "Adaptive Therapy for Androgen-Independent Prostate Cancer: A Randomized Selection Trial of Four Regimens," *Journal of the National Cancer Institute*, 99(21), 1613–1622. [509]

Wang, L., Rotnitzky, A., Lin, X., Millikan, R. E., and Thall, P. F. (2012), "Evaluation of Viable Dynamic Treatment Regimes in a Sequentially Randomized Trial of Advanced Prostate Cancer," *Journal of the American Statistical Association*, 107(498). [509]

Comment

Paul CHAFFEE and Mark VAN DER LAAN

1. INTRODUCTION

We would like to congratulate the authors of this article on a thoughtful and thorough design and data analysis. Their analysis represents a good example of how complex longitudinal data from observational studies or randomized trials should be analyzed, and reflects care and thought. The authors target questions of interest using sophisticated causal inference tools instead of allowing available software and common practice to dictate the analysis.

1.1 Background

The authors analyzed data from a sequentially randomized controlled trial, in which the longitudinal data structure on one patient can be coded as $O = (L(1),\ A(1), L(2),\ A(2), L(3),\ A(3),\ L(4),\ A(4), Y = L(5))$, where $L(1)$ denotes baseline covariates, $L(j),\ j = 2, 3, 4$, encodes both efficacy and toxicity measurements, $A(j) = (A_1(j), A_2(j))$, $A_1(j)$ indicates a treatment category among specified options, $A_2(j)$ is a dropout indicator, and Y is a particular final outcome such as a treatment efficacy score measured at 32 weeks. Let P_0 denote the true probability distribution of O. If the person drops out, then the subsequent L-process, including Y, is defined deterministically in some arbitrary manner. Treatment decisions $A_1(j)$ were assigned either deterministically in response to the observed history of the patient or randomly to three categories, with probability 1/3 in cases where this type of experimentation made sense from a medical point of view. Certain patients developed a history that resulted in a medically sensible deterministic switch of $A_1(j)$ to an alternative therapy (coded as OFF) prescribed by the treating physician, even though this deterministic switch was not a priori planned or foreseen in the protocol of the trial. The authors dealt with this by defining this switch as another natural treatment decision, as if planned by design, that can be incorporated in the targeted treatment rules of interest, instead of viewing it as right-censoring and thereby targeting parameters that cannot even be identified.

The authors defined the quantity of interest as the mean of outcome Y (measured at 32 weeks after randomization) under a dynamic intervention d applied to the "intervention" nodes $A(j)$, denoted by E_0Y_d. This intervention is chosen so that it fully respects the designed deterministic treatment switches as well as the unplanned deterministic switch to OFF, and the intervention enforces no dropout till the end (i.e., $A_2(j) = 0$ for $j = 1, 2, 3, 4$). The dynamic intervention d is indexed by an assigned first-line therapy for $A_1(1)$ among four possible choices, and a second-line therapy, among the remaining three, to which the patient will be switched if the first-line therapy fails *and* switching to this second-line therapy does not violate the deterministic switches in the data-generating experiment. If the patient experiences two consecutive successful responses on the same therapy, then the intervention stops. The intervention is also stopped if the patient experiences an event that enforces the treatment to be switched to OFF. Twelve such dynamic rules are considered, indexed by four choices of the first-line therapy and three choices for the second-line therapy. The authors estimated the mean E_0Y_d for each of these 12 rules d and constructed simultaneous confidence bands.

The dynamic rules are chosen so that this quantity of interest, E_0Y_d, can actually be identified from the data-generating experiment: that is, the positivity assumption $P_0(A(j) = d_j(Pa(A(j))) \mid \bar{L}(j), \bar{A}(j-1)) > 0$ almost everywhere (a.e.) holds, where we denote the history of all variables right before $A(j)$—the "parents" of $A(j)$—with $Pa(A(j))$, and we use the notation $\bar{L}(j) = (L(1), L, \ldots, L(j))$. A natural byproduct of selecting such realistic rules is that these dynamic treatments are of practical and scientific interest—a point effectively argued by the authors. Therefore, we can represent E_0Y_d as $\Psi(P_0)$ for a specified mapping $\Psi : \mathcal{M} \to \mathbb{R}$ that maps a probability distribution of O in the statistical model $\mathcal{M}$ into a real number. We remind the reader that a statistical model is an assumed collection of possible probability distributions of the data structure O and thus represents all statistical knowledge about the experiment. The pure statistical estimation problem is now defined as estimation of $\Psi(P_0)$ based on n iid copies of $O \sim P_0$, with the knowledge that $P_0 \in \mathcal{M}$.

The authors estimated E_0Y_d with a stabilized inverse probability of censoring-weighted (IPCW) estimator, defined as the inverse probability of treatment- and censoring-weighted empirical mean of the outcome over all subjects who followed the rule d. Specifically, the stabilized IPCW estimator of $\psi_0 = \Psi(P_0)$ is defined as

$$\psi_n^{\mathrm{IPCW}} \equiv \frac{\sum_i Y_i I(A_i = \bar{d}(L_i))/g_0(A_i \mid L_i)}{\sum_i I(A_i = \bar{d}(L_i))/g_0(A_i \mid L_i)},$$

where $A_i = (A_i(1), \ldots, A_i(4))$, $L_i = (L_i(1), \ldots, L_i(5))$, and $g_0(A_i \mid L_i) \equiv \prod_j P(A_i(j) \mid \bar{L}_i(j), \bar{A}_i(j-1))$, and for notational convenience, we use $A = \bar{d}(L)$ to denote the set of relations $(A(1) = d_1(L(1)), A(2) = d_2(Pa(A(2))), \ldots, A(4) = d_4(Pa(A(4))))$. Thus, g_0 denotes the conditional probability, given covariates, of following a specified treatment through the entire trial and not being censored. We use subscript n to denote sample-based estimators.

Since, ignoring dropout for now, the treatment mechanism is known in a sequentially randomized trial, such estimators are guaranteed to be consistent and asymptotically linear (and thereby asymptotically normally distributed), and they are also

Paul Chaffee is a PhD candidate in Biostatistics, in the Division of Biostatistics, University of California Berkeley, Berkeley, CA 94720 (E-mail: *pchaffee@berkeley.edu*). Mark van der Laan is Jiann-Ping Hsu/Karl E. Peace Professor in Biostatistics, in the Division of Biostatistics, University of California Berkeley, Berkeley, CA 94720 (E-mail: *laan@berkeley.edu*).

Journal of the American Statistical Association
June 2012, Vol. 107, No. 498, Applications and Case Studies
DOI: 10.1080/01621459.2012.665197

quite stable since the positivity assumption is practically satisfied. By estimating the treatment mechanism and incorporating covariates that are predictive of the outcome, thus ignoring the knowledge that the treatment mechanism is known, such an IPCW estimator can be made highly efficient, or even fully efficient.

The authors indeed used an estimator of g_0 for this reason. If dropout would have been quite severe and informative, then a typical IPCW estimator may be quite inefficient, biased, and unstable (see, e.g., Stitelman, Gruttola, and van der Laan 2011; van der Laan and Rose 2011). Given that dropout was quite minimal in this particular trial, the selected estimation procedure, which is easy to understand and implement, appears to represent a reasonable choice.

In our previous discussion of the first analysis of this trial (Bembom and van der Laan 2007), we were given some room to provide constructive suggestions and demonstrated an analysis of realistic dynamic treatment rules with the available data, but we find the current analysis thorough. At that time, some of the data compiled for analysis in the current article were not available. The new, augmented dataset led the authors to redefine the set of outcomes and to propose target parameters indexed by a set of treatment rules different from those based on the earlier data.

1.2 Substitution Estimators

We define the parameter of interest as a mapping $\Psi : \mathcal{M} \to \mathbb{R}$ from the statistical model $\mathcal{M}$ to the parameter space, where for simplicity, we consider scalar parameters. A substitution estimator of Ψ is a plug-in estimator $\Psi(\mathbf{P}_n)$, with $\mathbf{P}_n \in \mathcal{M}$ being an estimator of the true P_0. In our research over the last decade, we have focused on the development of double robust locally efficient *substitution* estimators based on the targeted minimum loss-based learning framework proposed in van der Laan and Rubin (2006), building on a rich literature in semiparametric models. As substitution estimators, these estimators are not only asymptotically optimized but also fully respect the global constraints implied by the statistical model and the target parameter, which makes them robust under sparsity. We refer to van der Laan and Rose (2011) for a comprehensive book on targeted learning, including contributions of many authors, documenting the motivation, developments, and many of the applications of targeted minimum-loss-based estimation (TMLE) to complex data structures. We would like to use the opportunity of this discussion to present a double robust, locally efficient substitution estimator for causal effects of multiple time point interventions, recently presented in van der Laan and Gruber (2011), which represents a particular TMLE inspired by innovative advances in general double robust estimation in Bang and Robins (2005). TMLE provides an alternative to IPCW estimation that is generally more efficient, less biased (TMLE is double robust), more finite-sample-robust under sparsity, and inherits the virtues of substitution estimators mentioned above. Without a reanalysis of the dataset, it is difficult to judge the potential benefit of this procedure for this particular study, but certainly if bias due to informative dropout had been a serious concern, it would have been worthwhile to apply this procedure.

In this discussion, we will also demonstrate a particular TMLE, proposed in van der Laan and Rubin (2006), chosen to guarantee to outperform the IPCW estimator, even under severe misspecification of the time-dependent covariate and outcome regressions. Though not discussed here, we refer to recent original work by Rotnitzky et al. (forthcoming), who go beyond this by presenting a general method for constructing a double robust, locally efficient estimator that is guaranteed to beat any user-supplied estimator (not just this IPCW estimator). This has nice applications in the sequentially randomized trial analyzed by the authors here. Subsequently, we applied Rotnitzky et al.'s idea to the TMLE framework to construct TMLEs that are guaranteed to outperform any user-supplied estimator (Gruber and van der Laan 2012). We refrain from discussing these important relevant advances in this discussion, but wish to highlight this important work by Rotnitzky and coauthors, and the resulting enhanced property of TMLE. In this discussion, we will also present a small-scale simulation aiming to imitate a longitudinal data structure of a two-time-point, sequentially randomized trial without right-censoring, and we conclude with a few remarks. We will start with some background that motivates the development of TMLE.

2. WHY NOT NPMLE?

The density P_0 of the data-generating distribution factorizes as $P_0 = Q_0 g_0$, where $g_0 = \prod_j g_0(A(j) \mid Pa(A(j)))$ represents the known treatment and unknown right-censoring mechanism, while $Q_0 = \prod_{j=1}^{5} P_0(L(j) \mid Pa(L(j)))$ represents the unspecified relevant factor of the likelihood. Note that $g_0 = g(P_0)$ and $Q_0 = Q(P_0)$ are themselves parameters of the data-generating distribution. The statistical model $\mathcal{M}$ for P_0 consists of all such distributions of O. As a result of sequential randomization, and assuming that right-censoring also satisfies the sequential randomization and positivity assumptions, the distribution of the counterfactual $L_d = (L(1), L_d(2), L_d(3), L_d(4), L_d(5) = Y_d)$ under dynamic intervention d is given by the so-called G-computation formula:

$$Q_{0,d}(l) = \prod_j P_0(L(j) = l(j) \mid \bar{L}(j-1) = \bar{l}(j-1), \bar{A}(j-1) = d(\bar{L}(j-1))).$$

In particular, $E_0 Y_d$ is defined as the mean of the marginal distribution under $Q_{0,d}$ of component Y_d. This defines now EY_d as a statistical target parameter $\Psi(P)$ of $P \in \mathcal{M}$, which depends on P only through $Q = Q(P)$. If $L(j)$ are discrete-valued, then we can estimate each conditional distribution of $L(j)$, given its parents $Pa(L(j))$, with empirical proportions, resulting in a pure empirical estimator, Q_n, of the true $Q_0 = Q(P_0)$. The corresponding plug-in estimator $\Psi(Q_n)$ would be the nonparametric maximum likelihood estimator (NPMLE). The problem with this estimator is that it is often not defined due to empty strata in finite samples, and even when defined, it is often too variable to be useful in finite samples.

3. WHY NOT SMOOTHED NPMLE?

In general, one will need to apply smoothing to estimate Q_0. For example, one could use a kth nearest-neighbor regression estimator for each conditional distribution of a binary variable coding $L(j)$, given its parents. However, to guarantee that these nearest-neighbor regression estimators are consistent, one needs the number of neighbors in the neighborhood over which one smoothes to go to infinity. In a covariate space of dimension k, this means that nh^k has to go to infinity, and thereby, that the smoothing width h converges to zero slower than $n^{-1/k}$. If one averages outcomes over a neighborhood of width h, then the bias will be $O(h)$. As a consequence, if $k \geq 2$, this smoothed NPMLE will have a bias $O(n^{-1/k})$ that cannot be neglected for statistical inference: confidence intervals ignoring this bias will have asymptotically zero coverage, and p-values cannot be trusted either. Using parametric regression models represents an extreme form of smoothing that is guaranteed to result in bias that will not even converge to zero as sample size goes to infinity, making confidence intervals particularly meaningless.

4. SUPER LEARNING: BETTER, BUT NOT ENOUGH

The typical bias of an estimator of Q_0 that converges at a particular rate to Q_0 will be of the same order as that rate. By constructing an adaptive estimator of Q_0 that is able to adapt to underlying smoothness or structure of Q_0, one can construct an estimator that might achieve a better rate of convergence than the typical minimax rate of convergence implied by the statistical model $\{Q(P) : P \in \mathcal{M}\}$ for Q_0. Such an estimator will therefore also have less bias. For that purpose, one can use super learning (van der Laan and Dudoit 2003; van der Laan, Polley, and Hubbard 2007; van der Laan and Rose 2011), which involves using cross-validation to select among a family of candidate estimators of Q_0 that are indexed by different fine-tuning parameters and approximation strategies. For example, the library of the super learner could include kernel regression estimators indexed by different bandwidths and different degrees of orthogonality of the kernels so that the best choice among all these kernel estimators will achieve the minimax rate for the model which assumes that the true underlying smoothness of Q_0 is known. Due to the oracle inequality for the cross-validation selector, the super learner will also achieve this adaptive minimax rate of convergence. So, super learning results in less-biased estimators of $\psi_0 \equiv \Psi(Q_0)$ relative to using a nearest-neighbor type of estimator or standard kernel regression estimator. However, the best rate of convergence one can realistically hope for is still worse than $1/\sqrt{n}$ (i.e., the rate achieved for a correctly specified parametric model), so the bias for ψ_0 implied by the super learner fit will still be nonnegligible and result in a plug-in estimator that is *not* asymptotically linear.

Additional bias reduction is thus essential, but this bias reduction should be fully targeted toward the target parameter. That is, the estimator Q_n does not need to solve all score equations as does the NPMLE, but it needs to solve certain target parameter-specific score equations, such as the efficient influence curve/efficient score for the target parameter. For that purpose, we combine super learning with TMLE, resulting in the targeted learning approach (van der Laan and Rubin 2006; van der Laan and Rose 2011). This step will need to involve an estimator of the true treatment and censoring mechanism g_0, and if g_0 is unknown, super learning is again recommended to obtain maximal bias reduction for this step. For this purpose, it is useful to note that g_0 can be factorized as a product of treatment probabilities and censoring probabilities. If treatment $A_1(j)$ at time j has more than two possible values, then one can code $A_1(j)$ in terms of a few binaries and still estimate the conditional probability of $A(j)$ with corresponding logistic regressions, and corresponding adaptive regression estimators for binary outcomes.

5. TARGETED MINIMUM-LOSS-BASED ESTIMATION

5.1 Efficient Influence Curve

We remind the reader that an efficient influence curve $D^*(P)$ at P is the canonical gradient of the pathwise derivative of $\Psi : \mathcal{M} \to \mathbb{R}$ along a richly selected class of parametric submodels through P (e.g., see Bickel et al. 1997; van der Vaart 1998; van der Laan and Robins 2003; van der Laan and Rose 2011). A regular asymptotically linear estimator has an influence curve that equals one of the gradients of the pathwise derivative, which explains why the canonical gradient is also called the efficient influence curve. In particular, the asymptotic variance of a regular asymptotically linear estimator is larger than or equal to the variance of the efficient influence curve. An estimator ψ_n of ψ_0 is asymptotically efficient at P_0 among the class of regular estimators if and only if it is asymptotically linear at P_0, with the influence curve equal to the efficient influence curve: that is, $\psi_n - \psi_0 = 1/n \sum_{i=1}^n D^*(P_0)(O_i) + o_P(1/\sqrt{n})$. A minimal condition for a substitution estimator $\Psi(\mathbf{P}_n)$ to be asymptotically efficient is that it solves $0 = \sum_{i=1}^n D^*(\mathbf{P}_n)(O_i)$ (or at least up till an $o_P(1/\sqrt{n})$-term). This equation for $\mathbf{P}_n$ does not define $\mathbf{P}_n$ (it is just one equation for an infinite-dimensional $\mathbf{P}_n$), but provides an important and necessary characteristic of $\mathbf{P}_n$ and makes the bias of $\Psi(\mathbf{P}_n)$ second order. The efficient influence curve at P depends on a relevant part of P, $Q(P)$, and a purely nuisance part, $g(P)$, and so, we also use the notation $D^*(P) = D^*(Q(P), g(P))$.

5.2 TMLE Algorithm

The first step in TMLE is to represent $\Psi(P)$ as a mapping of a part of P, say $\Psi(Q)$, with $Q = Q(P)$. One then obtains an initial estimate Q_n of Q_0 and g_n of g_0, and subsequently, one constructs a targeted estimator Q_n^*, g_n^*, obtained by iteratively minimizing an empirical risk w.r.t an appropriate loss function along a cleverly chosen submodel through Q_n and g_n so that the final update (Q_n^*, g_n^*) of the initial estimator, (Q_n, g_n), solves the efficient influence curve score equation $P_n D^*(Q_n^*, g_n^*) = 0$. We use superscript $*$ to indicate that these estimators are targeted. Here, we used the notation $P_n f = 1/n \sum_{i=1}^n f(O_i)$. The iterative updating of g_n is not necessary (in which case $g_n^* = g_n$), but can provide additional gains in the TMLE, as discussed below. The corresponding TMLE of ψ_0 is simply the plug-in estimator $\psi_n^* = \Psi(Q_n^*)$. Under regularity conditions, ψ_n^* is a double robust, locally efficient asymptotically linear estimator: that is, $\Psi(Q_n^*)$ is consistent and asymptotically linear if either g_n^* or Q_n^* is consistent, and it is efficient if both are consistent.

One could construct a targeted estimator of the relevant factor of the likelihood as in van der Laan (2010), Stitelman et al. (2011), and van der Laan and Rose (2011), involving coding $L(j)$ in terms of binaries, fitting the conditional distributions of these binaries with data-adaptive logistic regression, and targeting this fit with a subsequent targeted maximum likelihood update step involving univariate logistic regression with a clever covariate, using the initial fit as offset.

Alternatively, we can note that

$$\begin{aligned}\bar{Q}_4^d(\bar{l}(4)) &= E(Y \mid \bar{A}(4) = d(\bar{l}(4)), \bar{L}(4) = \bar{l}(4)) \\ &= E(Y_d \mid \bar{L}_d(4) = \bar{l}(4)) \\ \bar{Q}_3^d(\bar{l}(3)) &= E(\bar{Q}_4^d(\bar{L}(4)) \mid \bar{A}(3) = d(\bar{l}(3)), \bar{L}(3) = \bar{l}(3)) \\ &= E(Y_d \mid \bar{L}_d(3) = \bar{l}(3)),\end{aligned}$$

and so on so that

$$\bar{Q}_0^d = E\left(\bar{Q}_1^d(L(1))\right) = EY_d.$$

In this manner, it follows that EY_d is a function Ψ of a vector of iteratively defined conditional means $\bar{Q}^d = (\bar{Q}_4^d, \ldots, \bar{Q}_0^d)$. Suppose that Y is binary. The initial regression $\bar{Q}_4^d$ can be estimated with logistic regression. The targeting step involves adding a clever covariate $I(\bar{A} = d(\bar{L}))/g$ whose coefficient can be fitted with univariate logistic regression, using the initial estimator as offset. This targeted estimator $\bar{Q}_{4,n}^{d,*}$ of $\bar{Q}_4^d$ is now used as an outcome in (0,1) in the next *logistic* regression of $\bar{Q}_4^d$ onto $\bar{A}(3) = d(\bar{L}(3))$, $\bar{L}(3)$, and again, the resulting fit is targeted by fluctuating the fit with a univariate logistic regression model using a clever covariate $I(\bar{A}(3) = d(\bar{L}(3))/\prod_{j=1}^3 g_j$. This sequential regression procedure finally ends by taking the empirical mean of the estimator of $\bar{Q}_1^d$ over the baseline covariates $L(1)$, which is the final targeted estimate of EY_d. The same estimator applies to continuous $Y \in (0, 1)$, and thereby, for bounded continuous outcomes.

For details about this TMLE $\psi_n^* = \Psi(\bar{Q}_n^{d,*})$, obtained by plugging in a targeted estimator $\bar{Q}_n^{d,*}$ of these iteratively defined conditional means $\bar{Q}_0^d$, we refer to van der Laan and Gruber (2011). In the latter article, we also present some simulations demonstrating its practical performance.

6. STATISTICAL PROPERTIES

If g_0 is estimated consistently, then under appropriate regularity conditions, the TMLE of ψ_0 described above will be consistent and asymptotically linear even if all the conditional regressions for $\bar{Q}_j^d$, $j = 4, 3, 2, 1$ are inconsistent. In addition, if $\bar{Q}^d$ is estimated consistently, then this TMLE is asymptotically efficient. That is, as with all TMLEs, this TMLE is a double robust, locally efficient substitution estimator. It respects the global constraints of the statistical model and target parameter by using logistic regressions (see Gruber and van der Laan 2010; van der Laan and Gruber 2011). The asymptotic variance of ψ_n^* can be estimated with $1/n^2 \sum_{i=1}^n \{D^*(\bar{Q}_n^*, g_n)\}^2(O_i)$, and statistical inference proceeds accordingly.

7. TMLE THAT IS GUARANTEED TO BEAT IPCW

The efficient influence curve can also be represented as $D^*(P) = D_{\text{IPCW}}(P) - D_{\text{CAR}}(P)$, where $D_{\text{IPCW}}(P) = I(A = d(L))(Y - \psi)/g$ is the estimating function solved by the stabilized IPCW estimator, while D_{CAR} is a score of g (e.g., see Robins and Rotnitzky 1992; van der Laan and Robins 2003). As proposed in van der Laan and Rubin (2006), given a current estimator $\bar{Q}_n^k, g_n^k$, we can construct a submodel through g_n^k with score at zero-fluctuation equal to $D_{\text{CAR}}(\bar{Q}_n^k, g_n^k)$. By coding $A(j)$ in terms of binaries, this submodel can be chosen to correspond with fluctuating each of these binary conditional distributions with a univariate logistic regression in a cleverly chosen covariate, using the current estimator as offset. By also iteratively updating g_n^k in the TMLE algorithm described above along this clever submodel, we obtain a TMLE Q_n^*, g_n^* that solves both the efficient influence curve equation $P_n D^*(Q_n^*, g_n^*) = 0$ and the score $P_n D_{\text{CAR}}(Q_n^*, g_n^*) = 0$ for the censoring mechanism. As a consequence, it also solves $P_n D_{\text{IPCW}}(Q_n^*, g_n^*) = 0$ so that ψ_n^* is now also a stabilized IPCW estimator, and is guaranteed to be at least as efficient as the stabilized IPCW estimator used in the study.

Table 1. Comparison of TMLE and IPCW in estimation of a single treatment history-specific parameter $EY_{\bar{a}}$ at $n = 100$ in simulated data under two different data-generating mechanisms, P_0 and $P_0^\dagger$

P_0 ($n = 100$)	Bias	Var.	Rel. MSE
TMLE	–0.0068	0.0018	0.25
IPCW (g_n)	*	0.0040	0.52
IPCW (g_0)	*	0.0077	1.00
$P_0^\dagger$ ($n = 100$)	Bias	Var.	Rel. MSE
TMLE	*	0.0036	0.6
IPCW (g_n)	*	0.0064	1.0
IPCW (g_0)	*	0.0061	1.0

NOTE: Results based on 2000 simulations. IPCW (g_0) refers to the IPCW estimator using the true g_0 for the treatment mechanism; IPCW (g_n) is the IPCW estimator with g_0 estimated from the data.
*Indicates $< 10^{-3}$.

8. A SIMULATED EXAMPLE

We simulated data that mimic a simplified two-time-point, sequential randomized controlled trial with structure for a single observation $O = (L(1), A(1), L(2), A(2), L(3) = Y)$. Each $A(j)$ was binary and completely randomized, and each $L(j)$ (including $L(3)$) was a single, continuous random variable in the interval [0, 1]. We did not incorporate dropout in this data-generating experiment.

We chose as parameters of interest for these data the four possible treatment-history-specific counterfactual parameters $EY_{\bar{a}}$, where $\bar{a} \in \{(0, 0), (0, 1), (1, 0), (1, 1)\}$. In words, these are the mean population outcomes under the specified sequential interventions on the variables $(A(1), A(2))$. We report results for one of them ($EY_{0,0}$) under two different data-generating distributions. In the first distribution, $L(j)$ is a particular function of $\bar{A}(j-1)$ and $\bar{L}(j-1)$. The second incorporates similar relationships, but the influence of $Pa(L(j))$ on $L(j)$ for each j is very weak (see the Appendix).

The true values of these parameters of the simulated data distribution are, of course, not of interest, since they are entirely fabricated. The results of interest are the relative performances of the estimators in estimating these values.

The sequential regressions required for the TMLE were estimated with main-terms logistic regressions, and thus possibly misspecified for $\bar{Q}_j$, $j < 3$. $P(A(j) = a|Pa(A(j)))$, for the use in the IPCW estimator was also estimated using main-terms logistic regression, and included all elements of $Pa(A(j))$, which is a consistent estimator of the true conditional treatment probabilities. We found the TMLE to be quite insensitive to the choice of g incorporated (i.e., g_n vs. g_0), and we thus only include results using g_0 (Table 1).

We simulated data at three sample sizes, 50, 100, and 500, but include results only for $n = 100$ since the relative performances of the three estimators were quantitatively similar at all samples sizes. We report estimated bias, variance, and relative mean square error (MSE), the latter defined as $\text{MSE}(\cdot)/\text{MSE}(\psi^{\text{IPCW}}_{n,g_0})$.

A more thorough simulation study would first evaluate the relative variance of the efficient influence curve and the influence curve of the stabilized IPCW estimator under a variety of settings so that one knows how much information has been neglected by the IPCW estimator. One can then also investigate the relative gain of an estimator such as TMLE, which takes full advantage of the asymptotic efficiency bound.

9. CONCLUDING REMARKS

Though randomized trials make the IPCW estimator a valid estimator that is relatively efficient and unbiased in the absence of censoring, knowledge of the treatment mechanism can be utilized by constructing locally efficient substitution estimators of the target parameter that are guaranteed asymptotically linear and more efficient than, for example, the standard IPCW estimator, even under severe misspecification of the relevant factor of the likelihood. The potential gain in efficiency and bias of these estimators relative to current practice in actual clinical trials that are subject to informative censoring remains an important area of research. In addition, the utilization of machine (super) learning to maximize efficiency and minimize bias due to informative dropout represents another potential improvement in the analysis of randomized controlled trials. Obviously, such estimators still need to be specified a priori. Asymptotic theory teaches us that adaptive estimation is essential for obtaining an asymptotically linear and thereby consistent and normally distributed estimator of the target parameter in semiparametric models where we lack knowledge of the parametric models involved (see the main theorem in Zheng and van der Laan 2010 and chap. 27 in van der Laan and Rose 2011 by the same authors). There is an evident need for adaptive estimation in typical randomized controlled trials that are subject to dropout or other forms of missingness that are only partially understood.

APPENDIX: DATA-GENERATING FUNCTIONS FOR SIMULATIONS

As mentioned, we ran simulations under two different data-generating functions. The influence of $Pa(L(j))$ on $L(j)$ for $j = 2, 3$ was much stronger under the first distribution.

- Data generation under "P_0":
  $L(1) = \text{expit}(3 * Z_1)$, where $Z_1 \sim N(0, 1)$
  $A(1) \sim \text{Ber}(0.5)$
  $L(2) = \text{expit}(\text{logit}(L(1)) + 2 * A(1) + Z_2)$, where $Z_2 \sim N(0, 0.16)$
  $A(2) \sim \text{Ber}(0.5)$
  $Y = \text{expit}(\text{logit}(L(1)) + 1.5 * A(1) + \text{logit}(L(2)) - 0.5 * A(2) + Z_3)$, with $Z_3 \sim N(0, 0.16)$
- Data generation under "$P_0^\dagger$":
  $L(1) = \text{expit}(3 * Z_4)$, where $Z_4 \sim N(0, 1)$
  $A(1) \sim \text{Ber}(0.5)$
  $L(2) = \text{expit}(0.01 * \text{logit}(L(1)) + 0.05 * A(1) + Z_5)$, where $Z_5 \sim N(0, 16)$
  $A(2) \sim \text{Ber}(0.5)$
  $Y = \text{expit}(0.01 * \text{logit}(L(1)) + 0.1 * A(1) + 0.01 * \text{logit}(L(2)) - 0.05 * A(2) + Z_6)$, with $Z_6 \sim N(0, 12.25)$
  $Z_i \perp Z_j$ for all $i \neq j$.

[Received January 2012. Revised January 2012.]

REFERENCES

Bang, H., and Robins, J. (2005), "Doubly Robust Estimation in Missing Data and Causal Inference Models," *Biometrics*, 61, 962–972. [514]

Bembom, O., and van der Laan, M. (2007), "Comment: Statistical Methods for Analyzing Sequentially Randomized Trials," *Journal of the National Cancer Institute*, 99(21), 1577–1582. [514]

Bickel, P., Klaassen, C., Ritov, Y., and Wellner, J. (1997), *Efficient and Adaptive Estimation for Semiparametric Models*, New York: Springer-Verlag. [515]

Gruber, S., and van der Laan, M. (2010), "A Targeted Maximum Likelihood Estimator of a Causal Effect on a Bounded Continuous Outcome," *International Journal of Biostatistics*, 6, Article 26, Available at *http://www.degruyter.com/view/j/ijb.2010.6.1/ijb.2010.6.1.1260/ijb.2010.6.1.1260.xml?format=INT*. [516]

Gruber, S., and van der Laan, M. (2012), "Targeted Minimum Loss Based Estimator that Outperforms a Given Estimator," *International Journal of Biostatistics*, 8(1), Article 11. [514]

Robins, J., and Rotnitzky, A. (1992), "Recovery of Information and Adjustment for Dependent Censoring Using Surrogate Markers," in *AIDS Epidemiology – Methodological Issues*, eds. N. Jewell, K. Dietz, and V. Farewell, Basel: Birkhäuser, pp. 24–33. [516]

Rotnitzky, A., Lei, Q., Sued, M., and Robins, J. (forthcoming), "Enhanced Double-Robust Locally Efficient Estimation in Regression Models for Longitudinal Studies," *Biometrika*. [514]

Stitelman, O., Gruttola, V. D., and van der Laan, M. (2011), "A General Implementation of TMLE for Longitudinal Data Applied to Causal Inference in Survival Analysis," Technical Report No. 281, Division of Biostatistics, University of California, Berkeley. Available at *http://www.bepress.com/ucbbiostat/article281*. [514,515]

van der Laan, M. (2010), "Targeted Maximum Likelihood Based Causal Inference: Part I," *International Journal of Biostatistics*, 6(2), Article 2. [515]

van der Laan, M., and Dudoit, S. (2003), "Unified Cross-Validation Methodology for Selection Among Estimators and a General Cross-Validated Adaptive Epsilon-Net Estimator: Finite Sample Oracle Inequalities and Examples," Technical Report No. 130, Division of Biostatistics, University of California, Berkeley. [515]

van der Laan, M., and Gruber, S. (2011), "Targeted Minimum Loss Based Estimation of an Intervention Specific Mean Outcome," Technical Report No. 290, Division of Biostatistics, University of California, Berkeley. [514,516]

van der Laan, M., Polley, E., and Hubbard, A. (2007), "Super Learner," *Statistical Applications in Genetics and Molecular Biology*, 6(1), Article 25. [515]

van der Laan, M., and Robins, J. (2003), *Unified Methods for Censored Longitudinal Data and Causality*, Berlin: Springer. [515,516]

van der Laan, M., and Rose, S. (2011), *Targeted Learning: Causal Inference for Observational and Experimental Data*, Berlin: Springer. [514,515,517]

van der Laan, M., and Rubin, D. B. (2006), "Targeted Maximum Likelihood Learning," *International Journal of Biostatistics*, 2(1), Article 11. [514,515,516]

van der Vaart, A. (1998), *Asymptotic Statistics*, New York: Cambridge University Press. [515]

Zheng, W., and van der Laan, M. (2010), "Asymptotic Theory for Cross-Validated Targeted Maximum Likelihood Estimation," Technical Report No. 273, Division of Biostatistics, University of California, Berkeley. [517]

Rejoinder

Lu WANG, Andrea ROTNITZKY, Xihong LIN, Randall E. MILLIKAN, and Peter F. THALL

1. A BRIEF HISTORY OF THE SCRAMBLE

Because the analyses reported in this article are the culmination of a process that began in 1998 when the trial was conceived by Millikan and Thall at MD Anderson Cancer Center (MDACC), it is worthwhile to first provide a brief history, and also some remarks on the relationship between theory and application.

The prostate cancer trial was motivated, in part, by the desire to better reflect actual clinical practice by evaluating multistage treatment strategies. This was a radical idea for oncologists in 1998 and, sadly, it remains so at this writing. This trial was dubbed "The Scramble" by members of the Genitourinary Medical Oncology Department at MDACC, due to its unique and apparently complex structure. When we conceived the prostate trial scramble design in 1998, we were not aware that a literature on dynamic treatment regimes (DTRs) existed. Since then, the scramble has served as a prototype for later oncology trials, including an ongoing trial at MDACC of six two-stage DTRs of targeted agents for advanced kidney cancer (Thall et al. 2007).

The immense research activity in DTRs among statisticians over the past decade seems to have been motivated in large part by their interesting mathematical structure and the challenging technical problems that they present. The original impetus for this area of research was analysis of complex observational data, pioneered by Robins (1986, 1987, 1989). Because DTRs are essentially mathematical formalisms of routine medical practice, clinical trials based on DTRs provide great advantages compared to conventional trials that reduce variables and focus on one stage of therapy. Evidently, as statisticians, we have not done an effective job of communicating these facts to the oncology and larger medical research communities. In fact, actual trials of DTRs are few even as theoretical research activity has exploded. The good news is that many oncologists and research physicians in other areas very much like the idea of designing trials to evaluate DTRs, and moreover implementing such trials is not much harder than conventional trials. Since the ultimate aim of all this research activity is to develop improved medical therapies, it is high time for a more proactive approach to communicating these new statistical ideas to the medical research communities, and undertaking the hard work of actually implementing trials to evaluate DTRs.

2. SMARTS

We thank Almirall, Lizotte, and Murphy (ALM) for their kind words regarding our reanalysis of the data from the prostate cancer trial. ALM provide a nice overview of sequentially multiple assignment randomized trials (SMARTs), and focus on design and analysis as separate issues. ALM argue that it is important to design trials to evaluate viable DTRs by giving "a clear operationalization of how to treat patients," for all foreseen outcomes, that is, codifying the actions to be taken in response to all foreseen outcome histories. We could not agree more. However, a viable DTR is one that actually can and will be implemented in the targeted population by the treating physicians. In most medical settings, there are a myriad of possible outcomes and accompanying actions. Inevitably, this complexity must be simplified to obtain a feasible design, including viable regimes that actually can and will be implemented in practice during actual trial conduct.

Stopping chemotherapy in patients with advanced disease when progressive disease (PD) or severe toxicity (TOX) occurs is so routine in oncology practice that, when we conceived the prostate trial scramble design in 1998, it did not seem worth mentioning in the design. In fact, because all patients who developed PD or TOX were removed from the intervention, the removal rule was effectively operationalized in this trial. Years later, for this third data analysis, our inclusion of PD and TOX as outcomes and numerically scoring the resulting bivariate per-stage outcome were motivated, initially, by what we naively considered to be 47 dropouts. Formally and correctly calling PD and TOX outcomes led to 35 of these 47 patients being fully evaluated, and it enabled us to define viable DTRs that elaborated the simpler DTRs in the original design.

This trial raises three important points. The first point is that one cannot optimally design an experiment until after it has been carried out. Only actually conducting a trial can reveal the many complexities that are relevant to the study goals. That is, science is a learning process. ALM provide a very useful explanation of this issue as it relates to SMARTs in their section 2.2. The second point is that working harder to explain and define actual clinical outcomes led us to define DTRs that were viable, and hence that actually could be carried out in practice by treating physicians. The third point is that treatments have many effects, and accounting for these inevitably leads to both more complex statistical models and the need, as ALM point out, to account for trade-offs between efficacy and toxicity.

Although this all may appear to be quite challenging, ALM point out more good news, namely, that clinical trial protocols almost invariably include detailed explanations of the actual viable DTRs that will be used. After all, the first purpose of a clinical trial is to treat patients, and any reasonable protocol must spell out how this will be done. To simplify things somewhat, ideally, a medical statistician's task is to incorporate this sort of explanation formally into the trial design by defining the viable DTRs that will be evaluated. To motivate this, ALM provide a useful account of the consequences of having nonviable DTRs

We thank the editor, Hal Stern, for inviting this discussion. We also are grateful to the two sets of discussants both for providing many insights into the issues arising in our article, and for pointing the way to new and important methodological research in dynamic treatment regimes.

Journal of the American Statistical Association
June 2012, Vol. 107, No. 498, Applications and Case Studies
DOI: 10.1080/01621459.2012.665198

in a SMART trial. Their discussion of this issue could be used, in its simplest form, as an explanation of why conventional trial designs often fail to provide results that are useful to practicing physicians.

3. COMPOSITE ENDPOINTS

In their discussion of composite endpoints, ALM provide a simple and useful way to combine two of our scoring methods, by taking a weighted average. Because careful consideration of clinical events almost invariably leads to multivariate outcomes, inevitably, there arises the need to somehow reduce this to one dimension, in terms of either observations or parameters. ALM's repeated use of the words "trade-off" and "preference" in their discussion is encouraging, as we believe that this sort of thinking is essential when evaluating multiple, often competing, clinical outcomes. For example, methods based on scores (Bekele and Thall 2004), efficacy–toxicity trade-offs (Thall and Cook 2004), and utilities (Houede et al. 2010) have been proposed for dose-finding trials, while quality-adjusted survival time (Glasziou, Simes, and Gelber 1990; Zhao and Tsiatis 1996) has been used for many years as a more honest and ethical tool than using survival time alone. These and similar methods should be portable for use in designing SMARTs, and we feel that they will play a central role when evaluating DTRs in the future.

4. NEW METHODS

We thank Chaffee and van der Laan (CVDL) for their kind remarks and their concise explanation of the formalism underlying our analyses. CVDL indicate that because in a SMART trial, the treatment probabilities are known by design and bounded away from zero, inverse probability weighting (IPW) estimation of the DTR-specific counterfactual means offers an easy-to-implement procedure that yields estimators that are well behaved, that is, nearly unbiased and with an approximately normal distribution. They reiterate the point we made in Section 4.4, noting that IPW estimators can be made nearly efficient by replacing the known treatment probabilities by probabilities estimated under models for the treatment process that incorporate covariates that are predictive of the outcome. They go on to indicate that when, as in our trial, there are few dropouts, IPW estimators that further weight by the inverse of the conditional probabilities of dropout given past data offer an attractive practical analytic choice for SMART trials. We agree with all of these points. In fact, this is precisely the rationale that led us to analyze the data using the proposed IPW methodology.

Revisiting arguments made in the past by Robins and Ritov (1997), CVDL argue that, in the presence of severe dropout, even if one is prepared to assume that dropout is ignorable, that is, the dropout probabilities are conditionally independent of the outcome given past covariates and treatments, due to the curse of dimensionality, one cannot estimate the counterfactual DTR means without making some dimension reduction assumption about either (a) the dropout probabilities or (b) the conditional law of the endpoint of interest given the data collected up to each study cycle. CVDL offer the targeted maximum likelihood (TML) methodology as an attractive analytical tool for SMART studies with severe ignorable dropout. They indicate that TML is a general strategy for estimation of parameters of nonparametric or semiparametric models. In the context of SMART clinical trials with ignorable dropouts, certain variants of TML produce estimators of the DTR counterfactual means having the following attractive features:

(1) they respect natural constraints on the counterfactual means;
(2) they are double robust, that is, they are consistent and asymptotically normal provided either a working model for (a) or a working model for (b) is correct, but not necessarily both;
(3) they achieve the smallest possible variance for estimators that are consistent and asymptotically normal under the semiparametric model that assumes the working model for (a) to be correct, if indeed the data generating law also satisfies the working model for (b), a property that CVDL refer to as local efficiency; and
(4) they are guaranteed to be at least as efficient as any given IPW estimator that uses known or consistently estimated treatment and dropout probabilities.

CVDL present a simulation study of a moderately sized SMART trial with two time points, no dropout, and four possible static treatment regimes. In their simulation, TML performs better than an IPW estimator that uses treatment probabilities estimated under a logistic regression model with linear terms in the variables recorded up to each time point. Given that, in a clinical trial with no dropouts, the treatment probabilities are known by design, double robustness is not an issue. For such settings, one could also use a simple, easy-to-implement variant of the IPW estimator, which also satisfies (1), (3), and (4) (see, e.g., section 6 of Rotnitzky and Robins 1995). This is an IPW estimator in which the treatment probabilities are estimated under a logistic regression model that, in addition to the linear terms in the past covariates, includes specifically tailored preprocessed covariates computed at a first stage. It would have been interesting to include this estimator in CVDL's simulation study.

Several proposals other than TML exist that yield estimators satisfying some or all of (1)–(4). For example, Robins (2000) and Bang and Robins (2005) satisfied (2) and (3); Tan (2006, 2008) and Cao, Tsiatis, and Davidian (2009) satisfied (2)–(4); and Rotnitzky et al. (2012) and Tan (2010) satisfied (1)–(4). Although (2)–(4) are attractive theoretical large sample properties, it is important to bear in mind that many cancer studies are moderately sized, thus these properties are relevant only so long as they are good proxies for finite sample performance. We thus agree with the statement made by CVDL in their concluding remarks that more research is needed to understand the performance in terms of bias and efficiency of TML and any of the alternative semiparametric approaches in realistic SMART clinical trial settings.

5. CONCLUSIONS

Although it is well known that DTRs reflect more closely what happens in clinical practice, their use in clinical studies is still in its infancy, and there are many practical challenges. While increasing statistical methodological research has been done in recent years on design and analysis of DTRs, there

are still many open methodological issues. Some of these were encountered in our analysis of the prostate cancer trial, and were nicely addressed by the discussants. In the future, promotion of DTRs in clinical studies will require close collaborations between statisticians and clinicians to overcome practical barriers and develop consensus guidelines for design and analysis.

ADDITIONAL REFERENCES

Bang, H., and Robins, J. (2005), "Doubly Robust Estimation in Missing Data and Causal Inference Models," *Biometrics*, 61, 692–972. [519]

Bekele, B. N., and Thall, P. F. (2004), "Dose-Finding Based on Multiple Toxicities in a Soft Tissue Sarcoma Trial," *Journal of the American Statistical Association*, 99, 26–35. [519]

Cao, W., Tsiatis, A., and Davidian, M. (2009), "Improving Efficiency and Robustness of the Doubly Robust Estimator for a Population Mean With Incomplete Data," *Biometrika*, 96, 723–734. [519]

Glasziou, P. P., Simes, R. J., and Gelber, R. D. (1990), "Quality-Adjusted Survival Analysis," *Statistics in Medicine*, 9, 1259–1276. [519]

Houede, N., Thall, P. F., Nguyen, H., Paoletti, X., and Kramar, A. (2010), "Utility-Based Optimization of Combination Therapy Using Ordinal Toxicity and Efficacy in Phase I/II Trials," *Biometrics*, 66, 532–540. [519]

Robins, J. M. (2000), "Robust Estimation in Sequentially Ignorable Missing Data and Causal Inference Models," in *Proceedings of the American Statistical Association Section on Bayesian Statistical Science*, pp. 6–10. [519]

——— (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. [518]

——— (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. [518]

——— (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. [518]

Robins, J. M., and Ritov, Y. (1997), "Toward a Curse of Dimensionality Appropriate (CODA) Asymptotic Theory for Semi-Parametric Models," *Statistics in Medicine*, 16, 285–319. [519]

Rotnitzky, A., Lei, Q., Sued, M., and Robins, J. (2012), "Improved Double-Robust Estimation in Missing Data and Causal Inference Models," *Biometrika*, 99(2), 439–456. [519]

Rotnitzky, A., and Robins, J. M. (1995), "Semiparametric Regression Estimation in the Presence of Dependent Censoring," *Biometrika*, 82, 805–820. [519]

Tan, Z. (2006), "A Distributional Approach for Causal Inference Using Propensity Scores," *Journal of the American Statistical Association*, 101, 1619–1637. [519]

——— (2008), "Comment: Improved Local Efficiency and Double Robustness," *International Journal of Biostatistics*, 4, Article 10. [519]

——— (2010), "Bounded, Efficient and Doubly Robust Estimation With Inverse Weighting," *Biometrika*, 97, 661–682. [519]

Thall, P. F., and Cook, J. D. (2004), "Dose-Finding Based on Efficacy-Toxicity Trade-Offs," *Biometrics*, 60, 684–693. [519]

Thall, P. F., Wooten, L. H., Logothetis, C. J., Millikan, R., and Tannir, N. M. (2007), "Bayesian and Frequentist Two-Stage Treatment Strategies Based on Sequential Failure Times Subject to Interval Censoring," *Statistics in Medicine*, 26, 4687–4702. [518]

Zhao, H., and Tsiatis, A. A. (1996), "A Consistent Estimator for the Distribution of Quality Adjusted Survival Time," *Biometrika*, 84, 339–348. [519]