An Experimental Design for the Development of Adaptive Treatment Strategies

S.A. Murphy   SAMURPHY@UMICH.EDU

Department of Statistics & Institute for Social Research
University of Michigan
Ann Arbor, MI USA

Abstract

In adaptive treatment strategies, the treatment level and type is repeatedly adjusted according to ongoing individual response. Since past treatment may have delayed effects, the development of these treatment strategies is challenging. This paper advocates the use of sequential multiple assignment randomized trials in the development of adaptive treatment strategies. Both a simple ad hoc method for ascertaining sample sizes and simple analysis methods are provided.

Keywords: Therapeutic strategies. Treatment algorithms. Sequential multiple assignment randomized trial. Multi-stage decision problems. Dynamic treatment regimes.

1. Introduction

Adaptive treatment strategies are emerging as a new paradigm for the treatment and long term management of chronic, relapsing disorders such as alcoholism, smoking cessation, cocaine abuse, depression and hypertension (Glasgow et al., 1989; Kreuter and Strecher, 1996; Breslin et al, 1999; Unützer et al., 2001; Brooner and Kidorf, 2002; Lavori, Dawson and Roth, 2000; Murphy and McKay, 2003). In adaptive treatment strategies the treatment level and type is repeatedly adjusted according to the individual's need. The rationale is that the response is optimized by varying treatment type and dosage as a function of measures of time-dependent information such as response to past treatment. An adaptive treatment strategy is characterized by a sequence of decision rules, one per treatment decision. Adaptive treatment strategies are frequently called dynamic treatment regimes (Robins, 1986, 1989, 1993, 1997, Murphy et al., 2001, Lavori and Dawson, 2003).

Currently scientists use a combination of clinical experience, trial and error, behavioral, psychosocial and biological theories, results from observational studies and randomized experimental studies conducted for other purposes, to formulate the decision rules composing adaptive treatment strategies. In this paper we propose the use of experimental trials for the development of the decision rules. Our long term goal is to optimize the adaptive treatment strategies, that is produce a treatment strategy that yields the best mean value of the outcome. We concentrate on a major challenge, that of delayed effects, to designing randomized trials in this setting. We propose the use of sequential multiple assignment randomized (SMAR) trials for developing decision rules as these trials do not make apriori assumptions on the existence or form of the delayed effects. In a SMAR trial, each individual may be randomized multiple times; the multiple randomizations occur sequentially through time.

A number of SMAR trials have been or are being conducted. These include the CATIE trial for antipsychotic medications in patients with Alzheimer's (Schneider et al., 2001), the CATIE trial for antipsychotic medications in patients with Schizophrenia (see http://www.catie.unc.edu/schizophrenia/synopsis.html), STAR*D for treatment of depression(Rush et al., 2003; Lavori et al., 2001) and phase II trials at MD Anderson for treating cancer (Thall et al., 2000). In most cases the goals of these trials are to develop adaptive treatment strategies (sequences of treatments for an individual) and are not confirmatory. That is, the goal does not include confirming that one adaptive treatment strategy is better than control or standard treatment. Confirmatory trials comparing an optimized adaptive treatment strategy with the appropriate control or standard treatment would follow such studies.

To fix ideas we use the following simplified example throughout.

> <u>EXAMPLE</u>: Consider an addiction management study for alcohol dependent subjects. In this simple example there are only two decisions, choosing the initial treatment and the secondary treatment. Following an intensive outpatient program, alcohol dependent subjects may be prescribed either an opiate antagonist (med with low level of counselling) or intensive cognitive behavioral therapy (cbt). Then in the next two months clinicians record heavy drinking, adherence, self-management skill and other intermediate outcomes. Individuals are classified as responders or nonresponders based on their level of heavy drinking. Note that the terms responder/nonresponder refer to the level of heavy drinking while the individual is on the initial treatment. If an individual is a nonresponder to med then we must decide whether to either switch to cbt or assign an enhanced motivational program with a step up($em + cbt + med$). If an individual is a nonresponder to cbt then we must decide whether to switch to med or provide the step-up in treatment, $em + cbt + med$. Responders to the initial treatment should be assigned either to telephone monitoring (tm) or telephone counselling and monitoring (tmc). The goal in the decision making is to formulate an adaptive treatment strategy that results in the highest percent of days abstinent over a one year period. We seek two decision rules that will maximize the mean percent days abstinent over the defined period. The first decision rule can use pretreatment information such as level of addiction to choose the initial treatment. The second decision rule can utilize intermediate outcomes such as adherence to initial treatment, present level of self-management skill and number of heavy drinking days while on initial treatment to choose the secondary treatment.

In the next section we define adaptive treatment strategies and the optimal adaptive treatment strategy. In section 3 we use the formulae of the optimal adaptive treatment strategy as a vehicle to discuss delayed effects. It is the presence of delayed effects that motivates our experimental design proposal of SMAR trials in section 4. Section 4 also contains a discussion of other experimental design principles. In section 5 we propose a particular primary analysis; associated with this analysis is a simple approach for ascertaining the sample size. Lastly we provide an simulation based evaluation of our sample size formula and in the discussion section we discuss methods and areas that need further scientific development.

Note that the term "adaptive" as used here refers to a time-varying therapy for managing a chronic, relapsing illness rather than an experimental design. This should not be confused with "adaptive" experimental designs such as designs in which treatment allocation probabilities for the present patient depend on the responses of *past patients* (see Berry, 2001 or Rosenberger et al., 2001 for examples) nor should this be confused with two stage adaptive experimental designs in which design parameters of an experimental trial are altered mid-trial (see Fisher, 1998 or Jennison and Turnbull, 2003 for examples).

2. Adaptive Treatment Strategies

Suppose we will need to make at most k decisions, $a_1, a_2, \ldots, a_k$ per individual. S_1 denotes pretreatment information whereas S_j, $j \leq k$ denotes the intermediate outcome (ongoing information) available after decision a_{j-1} and prior to decision a_j. Thus the time order is $S_1, a_1, S_2, a_2, \ldots, S_k, a_k, S_{k+1}$. We use an overbar to denote past and present information (e.g. $\bar{S}_j = \{S_1, \ldots S_j\}$). The primary outcome is $Y = u(\bar{S}_{k+1}, \bar{a}_k)$ where u is a known summary function. In the addiction management study, Y is the mean percent days abstinent so u counts the number of days abstinent and divides by study length in days. An *adaptive treatment strategy* is a sequence of decision rules, one per decision. Thus we denote an adaptive treatment strategy by the decision rules $\{d_1, d_2, \ldots, d_k\}$ where the decision rule d_j takes the information available at time j ($\bar{S}_j = \{S_1, \ldots, S_j\}$ and past treatment $\bar{a}_{j-1} = \{a_1, \ldots, a_{j-1}\}$) and outputs a treatment type/level, a_j. For example in the addiction management study, $k = 2$ and the possible treatments, values for a_1, at the first decision time point are *med* and *cbt*. The information available for making the second decision includes pretreatment information denoted by S_1, the first treatment, and the intermediate outcomes denoted by S_2. A simple example of a decision rule, d_2, is: if the individual responds to initial treatment assign telephone monitoring with counselling or if the individual does not respond to initial treatment assign $em + med + cbt$.

As discussed in the introduction, the goal is to develop an adaptive treatment strategy that leads to the highest mean of Y (as compared to other possible adaptive treatment strategies). If we knew the multivariate distribution of $\bar{S}_{k+1}$ for each array of treatments $\bar{a}_k$, then this goal would be easily achieved. In this case, the ascertainment of the optimal or best adaptive treatment strategy is an optimization problem treated at great length by many authors (Bellman, 1957; Sutton and Barto, 1998; Cowell, Dawid, Lauritzen and Spiegelhalter (1999)). Classical algorithms use dynamic programming which is based on a backwards induction argument. We denote the optimal adaptive treatment strategy by $\{d_1^*(s_1), d_2^*(\bar{s}_2, a_1), \ldots, d_k^*(\bar{s}_k, \bar{a}_{k-1})\}$. Below we review the formulae for the optimal adaptive treatment strategy when $k = 2$; the formula for general k is similar (Cowell, Dawid, Lauritzen and Spiegelhalter (1999)).

The optimal decision at time 2 maximizes $E_{\bar{a}_2}[Y|\bar{S}_2]$ over a_2. Note that $E_{\bar{a}_2}[Y|\bar{S}_2]$ is the conditional mean of Y given $\bar{S}_2$ when the treatment decisions are set at $\bar{a}_2$. The optimal decision is given by

$$d_2^*(\bar{s}_2, a_1) = \arg\max_{a_2} E_{\bar{a}_2}[Y|\bar{S}_2 = \bar{s}_2]. \tag{1}$$

Define

$$V_2(\bar{s}_2; a_1) = \max_{a_2} E_{\bar{a}_2}[Y|\bar{S}_2 = \bar{s}_2].$$

Intuitively V_2 is the highest mean of Y achievable given that we are at time 2, have made the previous decision, a_1 and have recorded information $\bar{s}_2$. V_2 is called the optimal benefit-to-go at time 2. The optimal decision at time 1 is

$$d_1^*(s_1) = \arg\max_{a_1} E_{a_1}[V_2(\bar{S}_2; a_1)|S_1 = s_1]. \tag{2}$$

That is we integrate the optimal benefit to go at time 2 over the conditional distribution of S_2 given S_1 for treatment a_1 and then maximize. The optimal benefit-to-go at time 1 is given by

$$V_1(s_1) = \max_{a_1} E_{\bar{a}_1}[V_2(\bar{S}_2; a_1)|S_1 = s_1].$$

The mean of Y when the optimal rules are used to assign treatment (set $a_1 = d_1^*(S_1), a_2 = d_2^*(\bar{S}_2, a_1)$) is given by the expected optimal benefit-to-go at time 1:

$$E[V_1(S_1)] = E\left[\max_{a_1} E_{a_1}[\max_{a_2} E_{\bar{a}_2}[Y|\bar{S}_2]|S_1]\right]. \tag{3}$$

The *optimal adaptive treatment strategy* is (d_1^*, d_2^*). Observe that to obtain the optimal decision at time 2, we use the previously obtained optimal decision at time 1.

The above formulae are imprecise because we did not indicate the set of possible treatments over which each $\max_{a_j}$ is formed. In general the set of treatments is not the same for each time point and in addition past S_j's may restrict the set of possible treatments. In the addiction management study, the possible treatments at time 1 are *med* and *cbt*. But at time 2 the possible treatments depend on response to initial treatment.

In this paper the multivariate distribution of $\bar{S}_{k+1}$ indexed by the treatments $\bar{a}_k$ is unknown; thus we can not directly use the arguments given above to develop an optimal adaptive treatment strategy. Our goal is to propose experimental trials that will yield data for the formulation of an optimal adaptive treatment strategy. First we discuss the issue of delayed effects; this will motivate the design of the experimental trial.

3. Delayed Effects

In this section, we discuss why we want to observe intermediate and long term outcomes following each possible pattern of treatments across time; this motivates our design proposal of randomizing each individual multiple times. For clarity, we temporarily ignore sample variability and focus on correctly estimating which initial and subsequent treatments are best. Thus we use expectations rather than sample averages below. Once we have justified why it is important to observe intermediate and long term outcomes following each possible pattern of treatments we consider sampling issues (see the following sections).

For simplicity consider $k = 2$. Recall that the optimal rule for assigning treatment at time two is the output of $\arg\max_{a_2} E_{\bar{a}_2}[Y|\bar{S}_2 = \bar{s}_2]$. Certainly if this rule yields different secondary treatments depending on the initial treatment then in order to estimate the best secondary treatment we need to observe outcomes following each pattern of initial and secondary treatments. Also estimating the best initial treatment may require outcomes following each pattern of initial and secondary treatments. To see this consider the criterion function for ascertaining the best decision at time 1 (see (2)) and note that this criterion

function depends on the optimal decision at time 2. To ascertain the first decision we must maximize

$$E_{a_1}[\max_{a_2} E_{\bar{a}_2}[Y|\bar{S}_2]|S_1 = s_1] = \int_{s_2} \max_{a_2} E_{\bar{a}_2}[Y|\bar{S}_2 = \bar{s}_2] f_{a_1}(s_2|s_1) \, ds_2 \qquad (4)$$

over a_1 (note $f_{a_1}(s_2|s_1)$ is the conditional density of S_2 given S_1 indexed by a_1). Thus a_1 plays a role in the criterion function (4), both through the product of the density $f_{a_1}(s_2|s_1)$ and the maximized conditional mean $\max_{a_2} E_{\bar{a}_2}[Y|\bar{S}_2 = \bar{s}_2]$. Thus if $\max_{a_2} E_{\bar{a}_2}[Y|\bar{S}_2 = \bar{s}_2]$ depends on a_1 then in general to estimate the best initial treatment we need to observe intermediate and long term outcomes following each possible pattern of $\{a_1, a_2\}$. Certainly a treatment a_1 that produces very good values of S_2 (e.g. optimizes the mean of the density $f_{a_1}(s_2|s_1)$) may not be the best initial treatment in an adaptive treatment strategy.

We have argued that if $\max_{a_2} E_{\bar{a}_2}[Y|\bar{S}_2 = \bar{s}_2]$ depends on a_1 then in general to estimate the best initial treatment we need to observe intermediate and long term outcomes following each possible pattern of $\{a_1, a_2\}$. One way in which the maximized conditional mean, $\max_{a_2} E_{\bar{a}_2}[Y|\bar{S}_2 = \bar{s}_2]$, may depend on a_1 is if the effect of treatment at time 2 varies by the initial treatment (an interaction between initial and secondary treatments on the primary outcome, Y). There are at least two ways in which the effect of treatment at time 2 varies by the initial treatment. We discuss these two ways in the context of our simple addiction management study. First a particular initial treatment may set the stage for an enhanced effect by a particular secondary treatment. For example, suppose counselling is more effective than monitoring among responders to *cbt*; that is the individual learns to use counselling during *cbt* and thus is able to take advantage of the counselling offered to responders. Individuals who received *med* during the initial treatment will not have learned to use counselling and thus among responders to *med* the addition of counselling to the monitoring does not improve abstinence (Y) relative to monitoring alone. If an individual is a responder to *cbt* it is best to offer *tmc* as the secondary treatment, but if the individual is a responder to *med* it is best to offer the less expensive *tm* as the secondary treatment. In summary even if *cbt* and *med* result in the same proportion of responders, *cbt* may be the best initial treatment as part of the adaptive treatment strategy. This would be due to the enhanced effect of *tmc* when preceded by *cbt*.

A second way in which the effect of secondary treatment varies by the initial treatment is a consequence of Berkson's fallacy (Gail & Benichou, pg. 73, 2000). Suppose that there is an unobserved factor U that affects one's ability to respond to treatment. Here we conceptualize this factor as level of structure in one's life; $U = H$ if highly structured environment and $U = L$ if chaotic environment. We can expect that U is positively related to the intermediate outcome of responder/nonresponder status and also U is positively related to Y, percent days abstinent. Suppose also that the initial treatments have differing effects on responder/nonresponder status. Berkson's fallacy is the conditional correlation between U and initial treatment assignment given nonresponse to initial treatment. Intuitively a nonresponder who received the better initial treatment is more likely to live in a chaotic environment.

The following numerical example illustrates how Berkson's fallacy results in an interaction of secondary and initial treatments on outcome Y. Suppose the following relationship holds among nonresponders:

Mean of Y among nonresponders by secondary treatment and given structure of environment, $E_{a_1,a_2}[Y|S_2 = NR, U]$

	$U = H$ (highly structured)	$U = L$ (chaotic)
$a_2 = switch$	20	15
$a_2 = step - up$	30	10

Recall Y is percent days abstinent and thus large values are good. Thus nonresponders who live in a chaotic environment do not benefit from the step-up in treatment (possibly due to nonadherence), yet if a nonresponder lives in highly structured environment, the step-up in treatment is more beneficial than switching treatment. The conditional means in the table are the same regardless of the value of the initial treatment a_1. This table reflects the influence of the unobserved background variable U on Y. The level of structure in one's life U is also related to the intermediate outcome of responder/nonresponder as follows:

Probability of response by initial treatment and given structure of environment, $P_{a_1}[S_2 = NR|U]$

	$U = H$ (highly structured)	$U = L$ (chaotic)
$a_1 = med$	.3	.3
$a_1 = cbt$	.1	.5

Lastly suppose that $1/2$ of addicted individuals live in a chaotic environment. Using the formula,

$$E_{a_1,a_2}[Y|S_2 = NR] = \frac{\sum_{u\in\{H,L\}} E_{a_1,a_2}[Y|S_2 = NR, U = u]P_{a_1}[S_2 = NR|U = u]P[U = u]}{\sum_{u\in\{H,L\}} P_{a_1}[S_2 = NR|U = u]P[U = u]}$$

we obtain the table

Mean of Y among nonresponders by secondary and initial treatments, $E_{a_1,a_2}[Y|S_2 = NR]$

	$a_2 = switch$	$a_2 = step - up$
$a_1 = med$	17.5	20
$a_1 = cbt$	15.8	13.3

Thus if an individual is a nonresponder to med it is best to offer the step-up in treatment (to $em+med+cbt$) whereas if the individual is a nonresponder to cbt it is best to offer a switch in treatment (to med). So we see that there is an interaction between initial and secondary treatments; this interaction is due to Berkson's fallacy (there is no interaction conditional

on unobserved U). Since many chronic disorders (in particular, substance abuse and mental illness) appear to have psychosocial, biological and genetic causes, one can expect that U is high dimensional and that Berkson's fallacy may occur; $\max_{a_2} E_{\bar{a}_2}[Y|\bar{S}_2 = \bar{s}_2]$ may depend on a_1. Thus in order to estimate the best initial treatment we should observe the primary outcome following each pattern of initial and secondary treatments (recall (4)).

In many settings prior randomized trials of initial treatments (followed by standard or no treatment or treatment as available in the community) may indicate that a particular treatment rule, say, $d'_1(S_1)$ is best in terms of maximizing long term outcome (Y). Unfortunately the best initial treatment when the secondary treatment is set to a standard treatment (say b_2) is not necessarily the best initial treatment in an adaptive treatment strategy. That is, $d'_1(S_1)$ an estimator of

$$\arg \max_{a_1} E_{a_1}[\ E_{a_1,b_2}[Y|\bar{S}_2]|S_1 = s_1]$$

is not necessarily an estimator of

$$\arg \max_{a_1} E_{a_1}[\max_{a_2} E_{a_1,a_2}[Y|\bar{S}_2]|S_1 = s_1].$$

Thus the resulting rule may not be optimal in the context of an adaptive treatment strategy (the resulting rule is optimal if the only available secondary treatment is the standard treatment).

An interesting open problem would be to document the kinds of assumptions that make it possible to estimate the best adaptive treatment strategy from multiple experiments, one per decision. In the field of medical decision making (see for example, Cowell, Dawid, Lauritzen and Speigelhalter, 1999 or Parmigiani, ch. 5, 2002 or the applied example in Tengs, et al., 1998), Markovian assumptions along with the fact that earlier decisions are used only to reveal information (early decisions are diagnostic tests and only the last decision is a treatment) permit one to combine information across multiple experiments.

4. Sequential Multiple Assignment Randomized Trials

We propose that *SMAR trials* be used to develop adaptive treatment strategies. This type of trial has been proposed by Lavori and Dawson, (2000, 2003). In a SMAR trial each individual is randomized multiple times; that is, at each decision time. Thus individuals following each of the possible treatment patterns are observed. The randomization probabilities may depend on past and present observations. Since the treatments are randomized we use A_j instead of lower case a_j as in previous sections. Thus the data for one individual in a SMAR trial is given by $(\bar{S}_{k+1}, \bar{A}_k)$. Data from SMAR trials satisfies $P[S_{j+1} \leq s_{j+1}|\bar{S}_j = \bar{s}_j, \bar{A}_j = \bar{a}_j] = P_{\bar{a}_j}[S_{j+1} \leq s_{j+1}|\bar{S}_j = \bar{s}_j]$ for each j; that is the conditional distributions required to form optimal adaptive treatment strategies (for example, the optimal decisions in (2) and (1)) are identified by the corresponding empirical distributions in the SMAR trial. See the appendix for a proof of this equality.

Adaptive treatment strategies can be viewed as multi-component treatments with many components (e.g. which primary treatment, when to alter treatment, which treatment to provide nonresponders, form of monitoring or secondary treatment to provide responders to the primary treatment, what information to use to decide which treatment is best,

etc.). Consider the addiction management study. It is natural to expect that pretreatment information such as the presence of co-occurring disorders such as mental illness and homelessness might discriminate between individuals for whom it is best to receive *med* as the initial treatment in the strategy versus individuals for whom it is best to receive *cbt* initially. Furthermore, on-initial-treatment information such as the level of adherence to initial treatment and evolving level of self management skill in addition to responder status may be useful in discriminating who should get which secondary treatment. Thus the development of an adaptive treatment strategy not only requires consideration of possible treatments but also consideration of what information should be used in the decision rules. Viewed in this light the development of an adaptive treatment strategy is indeed a high dimensional problem. We propose several design principles that have the goal of reducing the dimensionality and we propose randomization probabilities that equalize sample size among different adaptive treatment strategies.

Principle I: When designing the trial restrict the class of treatment options at each decision point only by ethical, scientific or feasibility considerations. We conceptualize the S_j's as containing only information that is be used to constrain treatment options due to ethical, scientific or feasibility considerations. In practice S_j would be a known function of a larger amount of data, X_j that is collected prior to the jth decision time. The X_j's contain auxiliary information that may be useful in optimizing treatment but except via the information in S_j will not constrain treatment options in the trial. Thus the randomization probability for each A_j will depend on at most $(\bar{S}_j, \bar{A}_{j-1})$. We expect that in most practical settings S_j will be of low dimension (e.g. binary or trinary). Note however that the formulae given in the previous sections hold with X in place of S and with $Y = u(\bar{X}_{k+1}, \bar{A}_k)$. An additional rationale for allowing the randomization probabilities to only depend on the S's (the ethical, scientific or feasibility constraints) is to prevent undue restriction of the class of adaptive treatment strategies. When the experimental design for a SMAR trial specifies that the class of alternative treatments is to be restricted by past history, we are in effect restricting the class of adaptive treatment strategies to be considered. This point surfaces in the next section in the form of assumption (8).

As an example of this principle consider the addiction management study. Here the X's might include measures of an individual's experiences with the past treatments such as acceptability, adherence, burden, level of side effects, motivation level and also might include ongoing developments such as the level of craving, stress levels and other potential indicators of an impending relapse to heavy drinking. In general, information that is predictive of how one might respond differently to different kinds of treatments should be included in the X's. Indeed it would be tempting to specify different secondary treatment options for initial-treatment-nonadhering-nonresponders and initial-treatment-adhering-nonresponders. However this would mean that we would only consider strategies that treat these two groups of nonresponders differently. Unless this is known to be necessary, this constitutes an unnecessary restriction on the class of possible adaptive treatment strategies. Of course observing degree of adherence to initial treatment makes possible secondary analyses in which we test if adherence to initial treatment should be used to discriminate between secondary treatments.

Principle II: View this trial as one trial in a series of developmental, randomized trials (Box, Hunter and Hunter, 1978, see ch. 9, 15) prior to a confirmatory trial. This viewpoint

reduces the dimensionality problems since we do not expect this one trial to produce an optimized adaptive treatment strategy. Note this viewpoint is consistent with the present ongoing SMAR trials discussed in the introduction(CATIE and STAR*D); these trials are not designed to confirm a best optimal adaptive treatment strategy. Each developmental trial would build on the results of prior trials. A full development of Box's approach in this setting is an open problem (see the discussion for simple examples of how such developmental trials might proceed). A confirmatory randomized control trial of the developed adaptive treatment strategy versus the appropriate comparison (control or standard care) would follow.

Randomization Probabilities Proposal: We provide randomization probabilities that create equal sample sizes across strategies. In other words we set the randomization probabilities so that the proportion of subjects whose treatment assignments are consistent with an adaptive treatment strategy is constant across adaptive treatment strategies; equivalently, $P[\bar{A}_k = \bar{d}_k]$ will not depend on $\bar{d}_k$. Achieving this is more complicated than at first appears as the treatments for any subject are consistent with more than one adaptive treatment strategy. One can view this approach as being consistent with the working assumption that the variance of Y under different strategies is constant (Recall the classical result that in a large sample comparison of two means, assuming equal variances, equal sample sizes in the two groups maximizes power.)

Actually in setting the randomization probabilities we go one step further. We equalize sample sizes at each decision j among the possible treatment strategies beginning at that decision. That is, if we consider the adaptive treatment strategies beginning at decision j with "pretreatment" information equal to $(\bar{S}_j, \bar{A}_{j-1})$, then the sample size for each of these strategies should be the same; for each j, $P[\bar{A}_k = \bar{d}_k | \bar{S}_j, \bar{A}_{j-1} = \bar{d}_{j-1}]$ will not depend on $(d_j, \ldots, d_k)$. This approach of balancing sample sizes across different strategies is ad hoc; the best choice of randomization probabilities is an open question.

Denote the number of treatment options for the history, $(\bar{s}_j, \bar{a}_{j-1})$ by $n_j(\bar{s}_j, \bar{a}_{j-1})$. Note the possible dependence both on prior treatment and on the S's. First we provide simple formula for the randomization probabilities when each $n_j(\bar{s}_j, \bar{a}_{j-1}) = n_j(\bar{a}_{j-1})$, that is the number of treatment options does not depend on the S's and for $k = 2$. See the appendix for arbitrary k. In this case formulae for the randomization probabilities do not depend on the S's and are as follows: Set $N_2(a_1) = n_2(a_1)$. Denote the randomization probability of assigning treatment a_2 at time 2 for an individual with past history, $(\bar{S}_2, A_1) = (\bar{s}_2, a_1)$ by $p_2(a_2 | \bar{s}_2, \bar{a}_1)$ and similarly for initial treatment. Then

$$p_2(a_2 | \bar{s}_2, a_1) = N_2(a_1)^{-1} \tag{5}$$

Note the time 2 randomization probabilities are uniform. Next

$$N_1 = \sum_{b=1}^{n_1} N_2(b)$$

and

$$p_1(a_1 | s_1) = \frac{N_2(a_1)}{N_1}. \tag{6}$$

9

See the appendix for a proof that these randomization probabilities result in strategies with equal expected sample sizes.

The above formulae simplify further when the number of treatment possibilities do not depend on past treatment; in this case $N_j = \prod_{\ell=j}^{k} n_\ell$ and each $p_j(a_j) = 1/n_j$. The randomization probabilities are uniform over the options at each time. Consider the addiction management study; there are two treatment options for each responder class at decision 2 and two treatment options at time 1, thus all randomization probabilities are equal to .5. Suppose instead that the only secondary treatments following medication (*med*) are telephone monitoring for responders, and a step-up for nonresponders. Secondary treatments remain the same for individuals initially assigned cognitive behavioral therapy (*cbt*). In this case *med* responders are assigned telephone monitoring with probability one whereas *cbt* responders are assigned one of the two options with probability .5. Similarly for *med* nonresponders and *cbt* nonresponders. Next, $N_2(med) = 1$ and $N_2(cbt) = 2$, so $N_1 = 3$ and we assign *cbt* therapy with probability 2/3 ($N_2(cbt)/N_1$) and medication with probability 1/3 ($N_2(med)/N_1$).

When the number of treatment options depend on the S's, randomization probabilities that equalize sample size across strategies become more complicated and more importantly depend on the distribution of each S_j given $(\bar{S}_{j-1}, \bar{A}_{j-1})$. We provide the general formulae in the appendix along with a proof. When $k = 2$, these formulae yield, $p_2(a_2|\bar{S}_2, A_1) = n_2(\bar{S}_2, A_1)^{-1}$ and

$$
p_1(a_1|S_1) = \frac{E\left[n_2(\bar{S}_2, a_1)^{-1}|S_1, A_1 = a_1\right]^{-1}}{\sum_{b=1}^{n_1(S_1)} E[n_2(\bar{S}_2, b)^{-1}|S_1, A_1 = b]^{-1}}.
$$

To use these formulae, we need to make a working assumption concerning the probability distribution of S_2 given (S_1, A_1). In the simulations we make the extreme working assumption that the distribution of each S_j given $(\bar{S}_{j-1}, \bar{A}_{j-1})$ assigns mass 1 to the value of S_j with the largest number of subsequent treatment strategies. For $k = 2$ this implies that $p_2(a_2|\bar{S}_2, A_1) = n_2(\bar{S}_2, A_1)^{-1}$ and

$$
p_1(a_1|S_1) = \frac{\max_{s_2} n_2(s_2, S_1, a_1)}{\sum_{b=1}^{n_1(S_1)} \max_{s_2} n_2(s_2, S_1, b)}.
$$

See the appendix for the formula for general k. In both of the CATIE studies, there are multiple treatment possibilities for nonresponders but only one possible treatment for responders. Consider the addiction management study and suppose responders are provided one particular maintenance treatment and nonresponders are randomized between the two treatment alternatives. If we assume that the responder rates for *cbt* and *med* are the same, then the randomization probabilities for nonresponders are 1/2 and the initial randomization probabilities are also 1/2.

5. Test Statistic and Sample Size Formula

In order to power the SMAR trial a primary analysis must be specified. For example in the addiction management study, a natural and simple primary analysis would be to test if *cbt* differs from *med* in terms of the intermediate outcome, responder status. But we saw in

section three that absent additional scientific assumptions, the results of such an analysis are not useful in ascertaining the initial treatment in an optimal adaptive treatment strategy. Of course an advantage of this primary analysis is that it is of low dimension! We advocate that the primary analysis test if two adaptive treatment strategies each beginning with a different initial treatment yield different mean outcomes. In this primary analysis we lower the dimension by considering only very simple adaptive treatment strategies.

Principle III: The primary analysis should consider adaptive treatment strategies that depend only on S's (the ethical, scientific or feasibility constraints). Thus, for example, the primary analysis for the addiction management study would not consider rules depending on adherence to initial treatment. Note however that secondary analyses may investigate decision rules depending on the larger data including the X's.

Power Proposal: Power the developmental SMAR trial to discriminate between two adaptive treatment strategies, each depending only on the S's and each beginning with a different initial treatment. That is, compare pairs of adaptive treatment strategies in which the initial treatment varies by member of the pair. The rationale is that we need to decide which of the possible initial treatments are worthy of further investigation. The second and further randomizations will create comparable groups receiving different subsequent treatments, so even though we have not powered the study for discriminating between subsequent treatments, the randomization will permit an valid causal comparison between subsequent treatments. Subsequent developmental trials would then consider only the initial treatments that performed well in this trial; a simple, ideal result would be that a particular initial treatment A produces a much better adaptive treatment strategy (in terms of mean Y) than alternative initial treatments, then further experimental trials would use treatment A and be powered to contrast adaptive treatment strategies with different secondary treatments.

Note that one could power the trial so as to compare initial treatments, ignoring subsequent treatments. However this comparison will not generally assist us in formulating an adaptive treatment strategy because the mean of Y ignoring subsequent treatments will depend on the randomization probabilities. That is, the estimand will change depending on how one randomizes to subsequent treatments. Algebraically the mean response to the initial treatment a_1 ignoring subsequent treatment is given by (k=2)
$E_{a_1}\left[\left(\sum_{a_2} E_{\bar{a}_2}[Y|\bar{S}_2]p_2(a_2|s_1,a_1)\right)\Big|S_1 = s_1\right]$. The argument of the maximum of this quantity need not be the same estimand as (2).

The fact that SMAR trials are relatively new, with little practical experience by scientists, motivates us to consider sample size formulae in a simple setting. Thus in the following we focus attention on the setting in which there are a few major decisions (hence randomizations, say $k = 2$-4 randomizations per individual). Interpreting S_1 as containing only strong scientific, ethical or feasibility constraints on the class of initial treatments, there will be no S_1 if there is no need to place restrictions on the class of initial treatments. So the formulae below have no S_1 with the understanding that if restrictions on the class of initial treatments are necessary then we consider the subpopulations defined by values of S_1 separately. A consequence of this is that the initial decision rule, d_1, is simply a constant function specifying the same initial treatment for all within a subpopulation. Note this does not preclude consideration of d_1 a function of X_1 in secondary analyses.

The test statistic is based on the following estimator of the mean of Y under treatment strategy $\bar{d}_k$ (denote by $\mu_{\bar{d}_k}$). Define

$$U(\bar{S}_k, \bar{A}_k; \bar{d}_k, \mu_{\bar{d}_k}) = \left(\prod_{j=1}^{k} \frac{I\{A_j = d_j(\bar{S}_j, \bar{A}_{j-1})\}}{p_j(d_j|\bar{S}_j, \bar{A}_{j-1})} \right) (Y - \mu_{\bar{d}_k}).$$

where the indicator, $I\{B\}$ is 1 if the event B occurs and 0 otherwise. Solve

$$0 = \mathbb{P}_n U(\bar{S}_k, \bar{A}_k; \bar{d}_k, \mu_{\bar{d}_k})$$

for $\mu_{\bar{d}_k}$ and call the solution $\hat{\mu}_{\bar{d}_k}$. $\hat{\mu}_{\bar{d}_k}$ is our estimator of $\mu_{\bar{d}_k}$. In this equation, $\mathbb{P}_n f(X)$ is defined as the average over the sample, $1/n \sum_{i=1}^{n} f(X_i)$. In this case

$$\hat{\mu}_{\bar{d}_k} = \frac{\mathbb{P}_n \left[\prod_{j=1}^{k} \frac{I\{A_j = d_j(\bar{S}_j, \bar{A}_{j-1})\}}{p_j(d_j|\bar{S}_j, \bar{A}_{j-1})} Y \right]}{\mathbb{P}_n \left[\prod_{j=1}^{k} \frac{I\{A_j = d_j(\bar{S}_j, \bar{A}_{j-1})\}}{p_j(d_j|\bar{S}_j, \bar{A}_{j-1})} \right]} \tag{7}$$

See Murphy, van der Laan and Robins (2001) for technical details. The estimating formula can be generalized to be doubly robust (Robins, 2000); see the appendix for details. A crucial assumption underlying the consistency of this estimator is that

$$P_{\bar{d}_k} \left[\prod_{j=1}^{k} p_j(d_j|\bar{S}_j, \bar{d}_{j-1})) > 0 \right] = 1 \tag{8}$$

where subscript, $\bar{d}_k$ on the probability symbol indicates that this probability is calculated under the distribution of $\bar{S}_{k+1}$ when all treatments are assigned according to the strategy $\bar{d}_k$. That is we assume that at any decision time, and for any given history, the treatment specified by the adaptive strategy has a positive probability of being assigned. Since we will be comparing strategies, intuitively this assumption is that all feasible treatments have a nonzero probability of being assigned in the study.

Given data from a SMAR trial, we can perform two-way comparisons of adaptive treatment strategies with different initial treatments, say $\bar{d}_k$ and $\bar{d}'_k$, by approximating the distribution of the difference in estimators of mean response by a normal distribution. Since the primary purpose of the trial is to develop adaptive treatment strategies (as compared to attempting to confirm the superiority of one strategy over the others), we are more concerned about missing significant differences between means rather than accidently concluding that two means are equal; thus we have not provided an adjustment for multiple comparisons. A consistent estimator of the variance of $\sqrt{n}(\hat{\mu}_{\bar{d}_k} - \hat{\mu}_{\bar{d}'_k})$ is $\mathbb{P}_n(U^2(\bar{S}_k, \bar{A}_k; \bar{d}_k, \hat{\mu}_{\bar{d}_k}) + U^2(\bar{S}_k, \bar{A}_k; \bar{d}'_k, \hat{\mu}_{\bar{d}'_k}))$. Since one individual's treatment assignments can not be consistent with multiple adaptive treatment strategies each beginning with different initial treatments, there is no cross product in the above formula for the variance estimator. The statistic

$$\frac{\sqrt{n}(\hat{\mu}_{\bar{d}_k} - \hat{\mu}_{\bar{d}'_k})}{\sqrt{\mathbb{P}_n(U^2(\bar{S}_k, \bar{A}_k; \bar{d}_k, \hat{\mu}_{\bar{d}_k}) + U^2(\bar{S}_k, \bar{A}_k; \bar{d}'_k, \hat{\mu}_{\bar{d}'_k}))}} \tag{9}$$

has, in large samples, a standard normal distribution under the null hypothesis that $E_{\bar{d}_k}[Y] = E_{\bar{d}'_k}[Y]$. See Murphy et al.(2001) for a list of technical assumptions. The primary assumption is (8).

Sample Size Formula: To derive a sample size formula, we re-express the variance of $\sqrt{n}\hat{\mu}_{\bar{d}_k}$ in terms terms of the distribution of $\bar{S}_{k+1}$ when all treatments are assigned according to $\bar{d}_k$ as

$$E_{\bar{d}_k}\left[\prod_{\ell=1}^{k}\frac{1}{p_\ell(d_\ell|\bar{S}_\ell,\bar{d}_{\ell-1}))}\left(Y-\mu_{\bar{d}_k}\right)^2\right]. \tag{10}$$

The subscript, $\bar{d}_k$ on the expectation symbol is meant to act as a reminder that this expectation is calculated under the distribution of $\bar{S}_{k+1}$ when all treatments are assigned according to $\bar{d}_k$. This formula for the variance is proved in the appendix.

We provide sample size formula utilizing simple upper bounds on the variance (10). We consider two cases. First suppose the randomization probabilities are uniform on the possible treatment alternatives; it is immediate that an upper bound on the asymptotic variance of $\sqrt{n}\hat{\mu}_{\bar{d}_k}$ is $\max_{\bar{s}_k,\bar{a}_{k-1}}\prod_{j=1}^{k}n_j(\bar{s}_j,\bar{a}_{j-1})\sigma^2$ where σ^2 is the variance in response between individuals following the same adaptive treatment strategy ($\sigma^2 = Var_{\bar{d}_k}[Y]$).

Alternately suppose we use randomization probabilities that produce equal sample sizes across the strategies as in (5,6). If the number of treatment options depend at most on the past treatment ($n_j(\bar{s}_j,\bar{a}_{j-1})$ depends at most on $\bar{a}_{j-1}$ for each j), then (10) is equal to $N_1\sigma^2$. This equality is derived in the appendix. Indeed if the number of treatment options does not depend on either $\bar{s}_j$ or on $\bar{a}_{j-1}$ for each j then (5,6) are uniform randomization probabilities and $N_1 = \prod_{j=1}^{k}n_j$. If number of treatment options depends on the S's, we make the extreme working assumption that the distribution of each S_j given $(\bar{S}_{j-1},\bar{A}_{j-1})$ assigns mass 1 to the value of S_j with the largest number of subsequent treatment strategies ($j \leq k$). In this case, $N_1\sigma^2$ is again an upper bound on (10) where

$$N_1 = \sum_{a_1=1}^{n_1}\max_{s_2}\sum_{a_2=1}^{n_2(s_2,a_1)}\max_{s_3}\ldots\sum_{a_{k-1}=1}^{n_{k-1}(\bar{s}_{k-1},\bar{a}_{k-2})}\max_{s_k}n_k(\bar{s}_k,\bar{a}_{k-1}).$$

See the appendix for a proof.

Recall that we set the sample size so that we can compare two adaptive treatment strategies each beginning with a different initial treatment. Let $z_{\alpha/2}$ be the standard normal $(1-\alpha/2)$ percentile. Then the total sample size should be,

$$n = 2(z_{\alpha/2}+z_\beta)^2\tilde{N}\frac{\sigma^2}{(\Delta\mu)^2}$$

where $\tilde{N}$ is the appropriate upper bound from the previous paragraph, α is the Type I error of the two-sided test, and $1-\beta$ is the power of the test to detect a difference in the mean response of $\Delta\mu$. Since these trials are meant to be developmental, α might be set at a value larger than .05 and β might be small (say .05, .1).

6. Secondary Analyses

The multiple randomizations permit a variety of experimental secondary analyses. All of these analyses use randomization to justify causal inferences. For example, since the last decision optimizes a conditional mean (1) we can use regression analyses to test for differences between the final treatments. However note that only the last decision maximizes a conditional mean of Y (compare (1) with (2)) so use of a simple regression is limited to the last decision. See both Murphy (2003) and Robins (2003) for comments about other regression methods. Alternately we can compare two strategies with the same initial treatment using the statistic:

$$\frac{\sqrt{n}(\hat{\mu}_{\bar{d}_k} - \hat{\mu}_{\bar{d}'_k})}{\sqrt{\mathbb{P}_n(U(\bar{S}_k, \bar{A}_k; \bar{d}_k, \hat{\mu}_{\bar{d}_k}) - U(\bar{S}_k, \bar{A}_k; \bar{d}'_k, \hat{\mu}_{\bar{d}'_k}))^2}}.$$

An asymptotic justification for this statistic can be found in Murphy et al.(2001). However if X is of moderate or high dimensions, then this type of statistic will not be useful for consideration of adaptive treatment strategies depending on the X's (each d_j is a function from $(\bar{X}_j, \bar{a}_{j-1})$ to a class of treatments) because very few or no individuals in the data will have treatment patterns consistent with any given treatment strategy based on the X's. In this case one could estimate each conditional distribution, $P[S_{j+1} \le s_{j+1} | \bar{X}_j = \bar{x}_j, \bar{A}_j = \bar{a}_j]$ and then use Robin's "G-computation formula" to estimate the mean response to a given strategy. This method has not yet received much attention.

Other secondary analyses might estimate optimal decision rules; that is rules, $\bar{d}_k$, leading to the highest mean response (assume large values of Y are preferred). Such rules might very well be a function of the larger data, the X_j's. These methods use the sequential multiple assignment randomization to eliminate bias; see Murphy (2003) and Robins (2003) for methods that can be used to estimate such rules.

7. Simulation Results

The purpose of the simulations is to evaluate the utility of the sample size formula and corresponding test statistic. We consider settings in which there are two decisions ($k = 2$). There are three scenarios. Each has no S_1, two initial treatments and a binary S_2. In Scenario I, there are two secondary treatments for each of the values of S_2. In Scenario II, there are two possible secondary treatments for each of the values of S_2 as long as the initial treatment is treatment 1. If the initial treatment is treatment 2, then there is only one possible secondary treatment for each value of S_2. In Scenario III, there are two secondary treatments if the value of $S_2 = 0$ and only one possible secondary treatment if the value of $S_2 = 1$. To recap, in scenario I, the number of secondary treatments, n_2 does not depend on s_2, a_1; in scenario II, n_2 depends on a_1 but not on s_2 and in scenario III, n_2 depends on s_2.

In the simulations, S_2 is binary with the probability of a 1 equal to f_1 or f_2 depending on whether the initial treatment, $A_1 = 1$ or 2. The possible secondary treatments are $t_{s_2,a_1,1}$ or $t_{s_2,a_1,2}$ following treatment a_1 and outcome $S_2 = s_2$. We generated the response Y using Murphy's (2003) model for adaptive treatment strategies. This model is described in the appendix. In all of the scenarios, the optimal strategy, $\bar{d}_2^*$, is to assign treatment

14

1 initially then if $S_2 = 0$ assign the secondary treatment $t_{0,1,1}$ and if $S_2 = 1$ assign the secondary treatment $t_{1,1,1}$. In each of the three scenarios, we simulated the data so that the difference between the mean response to the optimal strategy $(E_{\bar{d}_2^*}Y)$ and the mean response to a strategy with initial treatment 2 (say $E_{\bar{d}_2}Y$) is proportional to a common standard deviation, $\sqrt{Var_{\bar{d}_2^*}Y} = \sqrt{Var_{\bar{d}_2}Y}$. This proportionality constant or signal-to-noise ratio is $\Delta\mu/\sigma$. For each scenario we simulated data for a variety of signal-to-noise ratios and values of f_1, f_2. We also ensured that there were at least one pair of adaptive treatment strategies with different initial treatments but with the same mean response. In this way we are able to check that the proposed procedure has the correct type one error rate.

In all cases the estimator is (7); the test statistic is given by (9). We set $\alpha = .1$ and $\beta = .1$ for the two-sided tests. Each simulation is of size 1000 data sets.

Consider first scenario I. We use randomization probabilities that balance the sample size among strategies; in this case these randomization probabilities are uniform (.5) because there always two treatment options at each decision. Given a signal-to-noise ratio, we set the total sample size equal to the smallest integer greater or equal to

$$2(z_{\alpha/2} + z_\beta)^2(2 \times 2)\frac{\sigma^2}{(\Delta\mu)^2}.$$

Table 1: Scenario I

$\Delta\mu/\sigma$(sample size)	$P[S_2 = 1\mid A_1 = 1]$	$P[S_2 = 1\mid A_1 = 2]$	Type 1 Error	Power
.25(1097)	.5	.5	.093	.900
.25(1097)	.3	.7	.109	.902
.5(275)	.5	.5	.097	.901
.5(275)	.3	.7	.093	.902
.75(122)	.5	.5	.117	.891
.75(122)	.3	.7	.109	.905

Clearly in this simple scenario the sample size formula and associated statistic perform well (Note that the binomial standard deviation for 1000 trials with $p = .1$ is approximately .01). Recall for any data set, we can estimate the mean outcome to multiple adaptive treatment strategies with differing initial treatments. The entries under the column labelled "Type 1 Error" are due to comparisons of adaptive treatment strategies with different initial treatments but the same mean ($\Delta\mu = 0$), whereas the entries under the column labelled "Power" are due to comparisons of adaptive treatment strategies with different initial treatments but means differ by $\Delta\mu/\sigma$.

In Scenario II we also use randomization probabilities that balance the sample size among strategies. Recall that there are two possible secondary treatments for each of the values of S_2 as long as the initial treatment is treatment 1. If the initial treatment is treatment 2, then there is only one possible secondary treatment for each value of S_2. So the $p_2(a_2|s_2, 1) = .5$ for $a_2 = t_{s_2,1,1}$ or $t_{s_2,1,2}$ whereas $p_2(t_{s_2,2,1}|s_2, 2) = 1$. The randomization probability at time 1 is $p_1(1) = 2/3$. Given a signal-to-noise ratio, we set the total sample size equal to the smallest integer greater or equal to

$$2(z_{\alpha/2} + z_\beta)^2(2 + 1)\frac{\sigma^2}{(\Delta\mu)^2}.$$

Table 2: Scenario II

$\Delta\mu/\sigma$(sample size)	$P[S_2 = 1|A_1 = 1]$	$P[S_2 = 1|A_1 = 2]$	Type 1 Error	Power
.25(823)	.5	.5	.107	.860
.25(823)	.3	.7	.115	.905
.5(206)	.5	.5	.113	.900
.5(206)	.3	.7	.122	.895
.75(92)	.5	.5	.120	.882
.75(92)	.3	.7	.127	.905

As in scenario I, the sample size formula and associated statistic perform well (Note that the binomial standard deviation for 1000 trials with $p = .1$ is approximately .01).

Recall that in Scenario III, the number of treatment options at decision 2 depends on S_2; there are two secondary treatments if the value of $S_2 = 0$ and only one possible secondary treatment if the value of $S_2 = 1$. If we make the extreme working assumption that $P[S_2 = 0|A_1] = 1$, then the $p_2(a_2|0, a_1) = .5$ for $a_2 = t_{0,a_1,1}$ or $t_{0,a_1,2}$ whereas $p_2(t_{1,a_1,1}|1, a_1) = 1$. The randomization probability at time 1 is $p_1(1) = 2/4$. Given a signal-to-noise ratio, we set the total sample size equal to the smallest integer greater or equal to

$$2(z_{\alpha/2} + z_\beta)^2(2+2)\frac{\sigma^2}{(\Delta\mu)^2}.$$

As expected the sample size formula is (see Table 3) conservative since not all individuals will respond with $S_2 = 0$ and thus be eligible for 2 rather than 1 secondary treatments.

Table 3: Scenario III

$\Delta\mu/\sigma$(sample size)	$P[S_2 = 1|A_1 = 1]$	$P[S_2 = 1|A_1 = 2]$	Type 1 Error	Power
.25(1097)	.5	.5	.088	.965
.25(1097)	.3	.7	.095	.968
.5(275)	.5	.5	.095	.964
.5(275)	.3	.7	.101	.953
.5(275)	.2	.2	.115	.909
.75(122)	.5	.5	.105	.958
.75(122)	.3	.7	.108	.954

However note the additional fifth row in which the probability of an individual responding with $S_2 = 1$ is .2. In this setting most of the individuals will be eligible for two secondary treatments rather than one and the working assumption that $P[S_2 = 0|A_1] = 1$ more closely approximates reality. In this case the sample size formula is accurate (.909 is close to .9).

8. Discussion

As stated above we view the SMAR trial as one trial in a series of developmental, randomized trials (Box, Hunter and Hunter, 1978, see ch. 9, 15) prior to a confirmatory trial. The generalization of Box's approach is an open problem. However to provide a indication of how such a sequence of trials might proceed consider the development of an adaptive treatment strategy for managing alcohol addiction. Suppose the treatment options are as

in our example. To set the sample size, we begin with $\alpha = \beta = .1$ and a given effect size (i.e., signal to noise ratio, $\Delta\mu/\sigma$). Then we use the formula in section 5 with $\tilde{N} = 4$ to calculate the sample size. After conducting the trial we find that a strategy beginning with *med* produces a higher percent abstinent days than the strategies beginning with *cbt*. In secondary analyses we use regression to discover an unanticipated result that nonadhering-nonresponders to the initial treatment do better when assigned the step up in treatment ($em + cbt + med$) whereas adhering-nonresponders to the initial treatment do better when switched to *cbt*. In secondary analyses, using the methods of Murphy (2003) or Robins (2003) we do not find any pretreatment covariates that would produce subpopulations responding best (in terms of Y) to adaptive treatment strategies beginning with *cbt*. Now we are ready for a second randomized trial. This trial would seek to replicate the unexpected results of the secondary analysis and further develop possible secondary treatments; this trial would involve only one randomization per individual as follows. We power the trial using the responder proportions from the prior trial so as to ensure sufficient nonresponders and responders. All individuals would be initially assigned *med*; responders would then be randomly assigned either to *tm* or *tmc*; nonresponders might be assigned to one of *cbt*, $em + med$, or $em + cbt + med$. Thus in addition to attempting to replicate the results associated with adherence level, we also provide the opportunity to check if the improved Y of nonresponders is due to the addition of *em* or to the combination of $em + cbt$. Note that analyses for this second trial are standard regression analyses. Suppose that we see no difference in the mean of Y for individuals receiving *tm* or *tmc*; since counselling is expensive we then decide to use *tm* as our secondary treatment for responders to *med*. Furthermore suppose that in the regression analysis of the nonresponders to *med* the adherence level does not appear to make a difference in mean Y and $em + med + cbt$ continues to yield better mean Y than other secondary treatments. Now we are ready to proceed to a randomized-controlled confirmatory trial. This would be a two group trial, the first group assigned to the strategy: *med* if responder then *tm*, if nonresponder then $em + med + cbt$, and the second group assigned to treatment as usual.

Here is an alternate possible sequence of trials. Suppose as before a strategy beginning with *med* produces a higher percent abstinent days than the strategies beginning with *cbt*. However using the methods of Murphy (2003) or Robins (2003), secondary analyses of the data from the SMAR trial lead to the unexpected finding that if an individual has a low pretreatment addiction level then adaptive treatment strategies beginning with *cbt* produce (scientifically) significant higher percent days abstinent than treatment strategies beginning with *med*. In this case a second SMAR trial with the same design would be run in an effort to replicate this result. If this result replicates then separate trials would need to be conducted for the individuals with different pretreatment levels of addiction.

Even the simplistic scenarios presented above illustrate how the development of adaptive treatment strategies presents a difficult challenge to experimental design due to the fact that these treatment strategies are high dimensional multi-component strategies evolving through time. Ideally the formulation of a powerful adaptive treatment strategy involves many decisions such as, when to start treatment, which treatment to start, when to step up treatment, which step-up treatment is best, when to step down treatment to maintenance/monitoring, which maintenance/monitoring treatment is best and what information to use to make each of the above decisions. This high dimensional problem appears insur-

mountable, but is crucially important as adaptive treatment strategies provide a paradigm whereby empirical evidence and statistical methods can improve clinical practice in which most treatment is adaptive, particularly in the management of chronic, relapsing disorders.

We envision that Box's approach (Box, Hunter and Hunter, 1978) of multiple experimental trials (that take advantage of working assumptions based on clinical experience, scientific theories and past studies) will provide a paradigm for reducing the dimensionality in a principled fashion. How best to formulate the paradigm in this setting remains an open problem. However before Box's approach can be utilized, we must understand the array of potential working assumptions, including working assumptions concerning delayed effects. This is an open area of research. The work presented here makes steps in this direction however it ignores an important issue that must be considered in order to fully assess the importance of delayed effects. This is the issue of how to provide a principled method for trading off immediate outcomes against longer term outcomes. The rich area of reinforcement learning may offer some guidance (see Sutton and Barto, 1998).

An alternate to Box's approach is the adaptive experimental design as developed by D. Berry and colleagues (Berry and Fristedt, 1985; Berry and Eick, 1995; Berry, 2001; Berry, 2003, and Thall, Sung and Estey, 2002). Here the word "adaptive" refers to the experimental design rather than the treatment strategy as in this paper. That is, adaptive refers to the practice of using outcomes from prior individuals to specify the probability of assigning a given treatment to the present individual. Thall, et al. (2002) considers a combination of two stage experimental design (drop poorly performing strategies halfway through trial) with a SMAR trial for developing strategies in the treatment of acute myelogenous leukemia. This combination along with the Bayesian analysis represents an attractive alternative to our approach. It is unclear how the adaptive experimental design would deal with the high dimensionality present in the development of adaptive treatment strategies.

Other open problems include the development and evaluation of sample size formulae in settings in which the number of treatment alternatives depends on the intermediate outcomes (S's) and one can make good working assumptions concerning the conditional distributions of the intermediate outcomes. In this paper we made the extreme working assumption that with probability one, S takes the value leading to the largest number of strategies. Clearly in many settings this assumption will lead to excessively large sample sizes. Another open problem is the best choice of randomization probabilities. We chose to set these randomization probabilities so that the sample size is balanced at each decision among subsequent strategies. It is unclear when, if ever, this is optimal. Indeed since the variance of the estimator of the mean response depends on the randomization probabilities it may be that in some settings using randomization probabilities depending on the larger X will reduce variance.

Acknowledgments

9. References

Bellman R. (1957) <u>Dynamic Programming</u>. Princeton, N.J.: Princeton University Press.

Berry, DA, Fristedt, B (1985). <u>Bandit Problems: Sequential Allocation of Experiments</u> Chapman-Hall, London.

Berry, DA, Eick SG (1995). Adaptive assignment versus balanced randomization in clinical trials: A decision analysis. *Statistics in Medicine* **14**:231-246.

Berry, DA (2001). Adaptive Trials and Bayesian Statistics in Drug Development. *Biopharmaceutical Report* **9**:1-7.

Berry, DA (2003). General Keynote: Clinical Trial Design. *Gynecologic Oncology* **88**:S114-S116.

Box, GP, Hunter WG and JS Hunter (1978) <u>Statistics for Experimenters</u> Wiley & Sons, New York.

Brooner, RK and M. Kidorf (2002). Using behavioral reinforcement to improve methadone treatment participation. *Science and Practice Perspectives* **1**:38-48.

Breslin, FC Sobell MB Sobell LC Cunningham JA Sdao-Jarvie K. and D Borsoi, (1999). Problem drinkers: evaluation of a stepped-care approach. *Journal of Substance Abuse* **10(3)**:217-232.

Cowell RG, Dawid A.P., Lauritzen S.L. and Spiegelhalter, D.J. (1999). <u>Probabilistic Networks and Expert Systems</u>. New York: Springer-Verlag New York, Inc.

Fisher LD. Self-designing clinical trials. <u>Statistics in Medicine</u> **17**:1551-1562.

Gail MH and Benichou J. (editors) (2000). <u>Encyclopedia of epidemiologic methods</u>. Chichester;New York: Wiley.

Glasgow, MS Engel BT and BC D'Lugoff (1989). A controlled study of a standardized behavioral stepped treatment for hypertension. *Psychosomatic Medicine* **51**:10-26.

Jennison C, Turnbull BW (2003) Mid-course sample size modification in clinical trials based on the observed treatment effect. *Statistics in Medicine* **22**:971-993.

Kreuter, MW and VJ Strecher (1996) Do tailored behavior change messages enhance the effectiveness of health risk appraisal? Results from a randomized trial. *Health Education Research* **11(1)**:97-105.

Lavori, PW, R Dawson, & AJ Roth, (2000). Flexible treatment strategies in chronic disease: Clinical and research implications. *Biological Psychiatry*, **48**: 605-614.

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

Lavori PW and R Dawson (2003). Dynamic treatment regimes: practical design considerations. *Clinical Trials*, **1**:9-20.

Lavori PW AJ Rush SR Wisniewski, J Alpert M. Fava, DJ Kupfer A. Nierenberg, FM Quitkin, HA Sacheim, ME Thase and M. Trivedi (2001) *Biological Psychiatry* **50**:792-801.

Murphy, SA (2003) Optimal Dynamic Treatment Regimes *Journal of the Royal Statistical Society, Series B*, **65(2)**, 331-366.

Murphy, SA and McKay, JR (2003) Adaptive treatment strategies: an emerging approach for improving treatment effectiveness. *to appear in the newsletter of the Society for a Science of Clinical Psychology*

Murphy, SA van der Laan MJ, Robins JM and CPPRG (2001), marginal mean models for dynamic regimes *JASA*, **96** 1410-1423.

Parmigiani, G. (2002). Modeling in Medical Decision Making: A Bayesian Approach. Wiley & Sons, Ltd: England.

Prochaska JO, Velicer WF, Guadagnoli E, Rossi JS and CC DiClemente (1991). Patterns of change: dynamic typology applied to smoking cessation *Multivariate Behavioral Research* **26(1)**: 83-107.

Prochaska JO, Velicer WF, Fava JL, Rossi JS and JY Tsoh (2001). Evaluating a population-based recruitment approach and a stage-based expert system intervention for smoking cessation *Addictive Behaviors* **26**: 583-602.

Robins, J.M. (1986). A new approach to causal inference in mortality studies with sustained exposure periods–application to control of the healthy worker survivor effect. *Computers and Mathematics with Applications* **14**, 1393–1512.

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

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

Robins, J.M. (1993). Information recovery and bias adjustment in proportional hazards regression analysis of randomized trials using surrogate markers. *Proceedings of the Biopharmaceutical Section, American Statistical Association*, 24–33.

Robins, J.M. (1997). Causal Inference from complex longitudinal data. *Latent Variable Modeling and Applications to Causality. Lecture Notes in Statistics (120)*, 69–117, (eds: M. Berkane). Springer-Verlag, Inc, New York.

Robins JM (2003) Optimal structural nested models. *unpublished manuscript.*

Robins, J.M. and Greenland S. (1994). Adjusting for differential rates of PCP prophylaxis in high- versus low-dose AZT treatment arms in an AIDS randomized trial. *Journal of the American Statistical Association*, **89**, 737-749.

Rosenberger WF, Stallard N, Ivanova A, Harper CN, Ricks ML (2001). Optimal adaptive designs for binary response trials. *Biometrics* **57**:909-913.

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

Rubin, D.B., (1986). Comment: Which ifs have causal answers. *Journal of the American Statistical Association*, **81**, 961-962.

Rush, AJ, M. Trivedi, M. Fava (2003). Depression,IV *American Journal of Psychiatry* **160(2)**:237.

Schneider LS, PN Tariot, CG Lyketsos, KS Dagerman, KL Davis, S. Davis, JK Hsiao, DV Jeste IR Katz, JT Olin, BG Pollock, PV Rabins, RA Rosenheck, GW Small, B Lebowitz, JA Lieberman (2001). National Institute of Mental Health clinical antipsychotic trials of intervention effectiveness (CATIE) Alzheimer disease trial methodology, *American Journal of Geriatric Psychiatry* **9(4)**: 346-360.

Sutton RS and AG Barto (1998). Reinforcement Learning : An Introduction MIT Press, Cambridge, Mass.

Tengs TO, Winer EP, Paddock S, Aguilar-Chavez and DA Berry (1998). Testing for the BRCA1 and BRCA2 breast-ovarian cancer susceptibility genes: A decision analysis. *Medical Decision Making* **18**:365-375.

Thall PF, Millikan RE and HG Sung (2000) Evaluating multiple treatment courses in clinical trials. *Statistics and Medicine* **19**:1011-1028.

Thall PF, Sung HG and EH Estey (2002). Selecting therapeutic strategies based on efficacy and death in multicourse clinical trials. *Journal of the American Statistical Association* **97**:29-39.

Untzer J, Katon W, Williams JW, Callahan CM, Harpole L, Hunkeler EM, Hoffing M, Arean P, Hegel MT, Schoenbaum M, Oishi SM, and CA Langston (2001) Improving primary care for depression in late life: the design of a multicenter randomized trial. it Medical Care **39(8)**, 785-799.

10. Appendix

<u>Sequential multiple assignment randomized (SMAR) trial:</u> In order to prove that SMAR trial produces data that permits the development of optimal adaptive treatment strategies, we must prove that

$$P[S_{j+1} \leq s_{j+1} | \bar{S}_j = \bar{s}_j, \bar{A}_j = \bar{a}_j] = P_{\bar{a}_j}[S_{j+1} \leq s_{j+1} | \bar{S}_j = \bar{s}_j]$$

holds for all values of $\bar{s}_{j+1}$, $\bar{a}_j$ and for each j. Note that data from a SMAR trial provides information about the conditional probabilities on the left hand side; whereas the optimal adaptive treatment strategy is defined in terms of probabilities on the right hand side. The difference between these two probabilities is subtle. The former probability refers to a multivariate distribution that permits variation in treatments; possibly because treatments are chosen by individuals or are chosen for them or treatments are randomized. $P[S_{j+1} \leq s_{j+1} | \bar{S}_j = \bar{s}_j, \bar{A}_j = \bar{a}_j]$ is the distribution of S_{j+1} among the subpopulation of individuals with values $(\bar{S}_j = \bar{s}_j, \bar{A}_j = \bar{a}_j)$. The latter probability refers to a multivariate distribution corresponding to all individuals having the same treatments $\bar{a}_k$. There is no variation in treatment in this multivariate distribution; all have treatments $\bar{a}_k$. According to this multivariate distribution $P_{\bar{a}_j}[S_{j+1} \leq s_{j+1} | \bar{S}_j = \bar{s}_j]$ is the distribution of S_{j+1} among subpopulation of individuals with values $(\bar{S}_j = \bar{s}_j)$. Note the differing multivariate distributions and differing definitions of the subpopulations. In general these two probabilities need not be equal; however they are equal under sequential multiple assignment randomization. To prove this equality we use the potential outcomes model (Rubin, (1978, 1986); Robins (1986, 1987)).

Instead of considering a different distribution of the vector $\bar{S}_j$ for each possible value of the treatments, $\bar{a}_{j-1}$, we consider a different vector of random variables corresponding to each value of the treatment; these different random variables are the potential outcomes. Thus corresponding to each fixed value of the treatment vector, $\bar{a}_{j-1}$ we conceptualize a potential outcome denoted by $S_j(\bar{a}_{j-1})$ where $S_j(\bar{a}_{j-1})$ is the response at the end of the jth interval that a subject would have if he/she were assigned the treatments, $\bar{a}_{j-1}$. For each $j = 1, \ldots, k$, the multivariate distribution of $(S_1, S_2(a_1), \ldots, S_j(\bar{a}_{j-1})$ is the section 2 distribution of $\bar{S}_j$ for the treatment vector $\bar{a}_j$, e.g. for a measurable set B, $P_{\bar{a}_{j-1}}[S_j \in B | \bar{S}_{j-1} = \bar{s}_{j-1}]$ is the conditional probability that the potential outcome, $S_j(\bar{a}_{j-1}) \in B$ given the potential outcomes, $(S_1, S_2(a_1), \ldots, S_{j-1}(\bar{a}_{j-2})) = \bar{s}_{j-1}$. Similarly $E[Y(\bar{a}_k)]$ is the same as $E_{\bar{a}_k}[Y]$.

Let $\mathcal{A}_k$ be the collection of all possible k-vectors of treatments decisions. The set of all potential outcomes is $\{(S_1, S_2(a_1), S_3(\bar{a}_2), \ldots, S_{k+1}(\bar{a}_{k+1})) : \bar{a}_k$ varying in $\mathcal{A}_k\}$ (S_1 does not vary by treatments but is included for completeness). We can rewrite the section 2 formulae for the optimal rules in terms of potential outcomes. For simplicity let $k = 2$. Define $Y(\bar{a}_2)$ as $u(S_1, S_2(a_1), S_3(\bar{a}_2), \bar{a}_2)$. Then the optimal decision rules are given by

$$d_2^*(\bar{s}_2, a_1) = \arg \max_{a_2} E[Y(a_1, a_2) | S_1 = s_1, S_2(a_1) = s_2]$$

and

$$d_1^*(s_1) = \arg \max_{a_1} E[Y(a_1, d_2^*) | S_1 = s_1].$$

And the optimal mean response is given by,

$$E\left[Y(d_1^*, d_2^*)\right].$$

We use the potential outcomes to connect the formulae in section 2 to the distribution for the observable data $(\bar{S}_{k+1}, \bar{A}_k)$ from a SMAR trial. First we make the consistency assumption (Robins, 1997) that S_{k+1} from the observable data is the same as the potential outcome $S_{k+1}(\bar{a}_k)$ evaluated at $\bar{a}_k = \bar{A}_k$, and so on, including $S_2 = S_2(A_1)$. Robins' consistency assumption is the assumption that an individual's response to treatment (the S's) in the study is affected only by his or her treatment not by the treatment of others and that we have represented the treatment adequately by the a_j's (i.e. if there are 5 values of a_j there are only 5 treatments). Because the treatments are randomized with randomization probabilities depending only on past and present information, we have that for each $j = 1, ..., k$, A_j is independent of $\{(S_1, S_2(a_1), S_3(\bar{a}_2), \ldots, S_{k+1}(\bar{a}_{k+1})) : \bar{a}_k$ varying in $\mathcal{A}_k\}$ given $(\bar{S}_j, \bar{A}_{j-1})$. Thus

$$
\begin{aligned}
P[S_{j+1} \leq s_{j+1}|\bar{S}_j = \bar{s}_j, \bar{A}_j = \bar{a}_j] &= P[S_{j+1}(\bar{a}_j) \leq s_{j+1}|\bar{S}_j = \bar{s}_j, \bar{A}_j = \bar{a}_j] \\
&\qquad \text{(consistency assumption)} \\
&= P[S_{j+1}(\bar{a}_j) \leq s_{j+1}|\bar{S}_j = \bar{s}_j, \bar{A}_{j-1} = \bar{a}_{j-1}] \\
&\qquad \text{(randomization)} \\
&= P[S_{j+1}(\bar{a}_j) \leq s_{j+1}|\bar{S}_{j-1} = \bar{s}_{j-1}, S_j(\bar{a}_{j-1}) = s_j, \bar{A}_{j-1} = \bar{a}_{j-1}] \\
&\qquad \text{(consistency assumption)} \\
&= P[S_{j+1}(\bar{a}_j) \leq s_{j+1}|\bar{S}_{j-1} = \bar{s}_{j-1}, S_j(\bar{a}_{j-1}) = s_j, \bar{A}_{j-2} = \bar{a}_{j-2}] \\
&\qquad \text{(use randomization} \\
&\qquad \text{and consistency repeatedly....)} \\
&= P[S_{j+1}(\bar{a}_j) \leq s_{j+1}|\bar{S}_j(\bar{a}_{j-1}) = \bar{s}_j] \\
&= P_{\bar{a}_j}[S_{j+1} \leq s_{j+1}|\bar{S}_j = \bar{s}_j] \\
&\qquad \text{(by definition).}
\end{aligned}
$$

So the consistency assumption in addition to the proof above allow us to connect the distribution of the data, denoted by P or E with no subscript to the section 2 multivariate distributions of $\bar{S}_{k+1}$ indexed by $\bar{a}_K$, denoted by $P_{\bar{a}_k}$ or $E_{\bar{a}_k}$. Now we see that the formulae for optimal decisions in section 2 can be written as functions of conditional distributions for the observable data; thus optimal decisions are identifiable from a SMAR trial.

Randomization Probabilities: Set $N_k(\bar{s}_k, \bar{a}_{k-1}) = n_k(\bar{s}_k, \bar{a}_{k-1})$ and then

$$p_k(a_k|\bar{s}_k, \bar{a}_{k-1}) = N_k(\bar{s}_k, \bar{a}_{k-1})^{-1} \tag{11}$$

Note the time k randomization probabilities are uniform. Then for each j set

$$N_{j-1}(\bar{s}_{j-1}, \bar{a}_{j-2}) = \sum_{a_{j-1}=1}^{n_{j-1}(\bar{s}_{j-1}, \bar{A}_{j-2})} E[N_j(\bar{S}_j, \bar{a}_{j-1})^{-1}|\bar{S}_{j-1} = \bar{s}_{j-1}, \bar{A}_{j-1} = \bar{a}_{j-1}]^{-1} \tag{12}$$

and

$$p_{j-1}(a_{j-1}|\bar{s}_{j-1}, \bar{a}_{j-2}) = \frac{E[N_j(\bar{S}_j, \bar{A}_{j-1})^{-1}|\bar{S}_{j-1} = \bar{s}_{j-1}, \bar{A}_{j-1} = \bar{a}_{j-1}]^{-1}}{N_{j-1}(\bar{s}_{j-1}, \bar{a}_{j-2})}. \tag{13}$$

Lastly set

$$N_1(s_1) = \sum_{a_1=1}^{n_1(s_1)} E[N_2(\bar{S}_2, A_1)^{-1}|S_1 = s_1, A_1 = b]^{-1}$$

and

$$p_1(a_1|s_1) = \frac{E[N_2(\bar{S}_2, A_1)^{-1}|S_1 = s_1, A_1 = a_1]^{-1}}{N_1(s_1)}$$

Proceeding backward through time we will see that the randomization probabilities defined above balance the sample size among subsequent strategies at each decision. First at the last decision, note that $P[A_k = d_k|\bar{S}_k = \bar{s}_k, \bar{A}_{k-1} = \bar{d}_{k-1}] = p_k(d_k|\bar{s}_k, \bar{d}_{k-1})$ which does not depend on the rule d_k (see (11)). Next

$$P[A_k = d_k, A_{k-1} = d_{k-1}|\bar{S}_{k-1} = \bar{s}_{k-1}, \bar{A}_{k-2} = \bar{d}_{k-2}] =$$
$$p_{k-1}(d_{k-1}|\bar{s}_{k-1}, \bar{d}_{k-2})E[N_k(\bar{S}_k, \bar{d}_{k-1})^{-1}|\bar{S}_{k-1} = \bar{s}_{k-1}, \bar{A}_{k-1} = \bar{d}_{k-1}].$$

By definition of p_{k-1}, this probability does not depend on the rule d_{k-1} (see (13). Continuing backwards we see that for each j $P[\bar{A}_k = \bar{d}_k|\bar{S}_j = \bar{s}_j, \bar{A}_{j-1} = \bar{d}_{j-1}]$ does not depend on $(d_j, \ldots, d_k)$. Thus if these randomization probabilities are used then we have that

$$P[\bar{A}_k = \bar{d}_k|\bar{S}_j, \bar{A}_{j-1} = \bar{d}_{j-1}] = E\left[\prod_{\ell=j}^{k} p_\ell(d_\ell|\bar{S}_\ell, \bar{d}_{\ell-2})|\bar{S}_j, \bar{A}_{j-1} = \bar{d}_{j-1}\right]$$
$$= N_j(\bar{d}_{j-1})^{-1}.$$

Note this probability is the same for all rules $(d_j, \ldots, d_k)$ and also note that this probability implicitly depends on $\bar{S}_{j-1}$ and past treatment through the decision rules $\bar{d}_{j-1}$.

If the number of treatment options at each time depends only on past treatment (i.e. $n_j(\bar{s}_j, \bar{a}_{j-1})$ can be written as $n_j(\bar{a}_{j-1})$ for each j) then N_k depends only on past treatment and furthermore from (12) we see that if any N_j is a function only of past treatment, N_{j-1} must also be a function of only past treatment.

Consider the situation in which the number of treatment options at each time can depend not only on past treatment but also on past S's. Suppose we make the working assumption that the distribution of S_j given $(\bar{S}_{j-1}, \bar{A}_{j-1})$ is a point mass at a value of S_j with a maximal number of subsequent strategies. Thus $P[S_k = s_0|\bar{S}_{k-1} = \bar{s}_{k-1}, \bar{A}_{k-1} = \bar{a}_{k-1}] = 1$ if $s_0 = \arg\max_{s_k} n_k(\bar{s}_k, \bar{a}_{k-1})$. And for any j $P[S_{j-1} = s_0|\bar{S}_{j-2} = \bar{s}_{j-2}, \bar{A}_{j-2} = \bar{a}_{j-2}] = 1$ if $s_0 = \arg\max_{s_{j-1}}(n_{j-1}(\bar{s}_{j-1}, \bar{a}_{j-2})\max_{s_j, \ldots, s_k}\prod_{i=j}^{k} n_k(\bar{s}_j, \bar{a}_{j-1}))$. Using this working assumption (12) becomes

$$N_{j-1}(\bar{s}_{j-1}, \bar{a}_{j-2}) = \sum_{a_{j-1}=1}^{n_{j-1}(\bar{s}_{j-1}, \bar{a}_{j-2})} \max_{s_j} N_j(\bar{s}_j, \bar{a}_{j-1}).$$

The formulae for the randomization probabilities follow immediately.

<u>Test Statistic:</u> In the following the potential outcomes notation from the beginning of this appendix is used. The following estimating formula is a doubly robust generalization of (7). The goal is to estimate the mean response to strategy $\bar{d}_k$ ($EY(\bar{d}_k)$ or $\mu_{\bar{d}_k}$). For each $j = 2, \ldots, k$, denote $E[Y(\bar{d}_k)|\bar{S}_j = \bar{s}_j, \bar{A}_{j-1} = \bar{d}_{j-1}]$ by $\mu_j(\bar{s}_j, \bar{d}_{j-1})$. Define

$$U(\bar{S}_k, \bar{A}_k; \bar{d}_k, \bar{\mu}_k, \mu_{\bar{d}_k}) = \left(\prod_{j=1}^{k} \frac{I\{A_j = d_j(\bar{S}_j, \bar{A}_{j-1})\}}{p_j(d_j | \bar{S}_j, \bar{A}_{j-1})} \right) (Y - \mu_k(\bar{S}_k, \bar{d}_{k-1}))$$

$$+ \sum_{\ell=2}^{k-1} \left(\prod_{j=1}^{\ell} \frac{I\{A_j = d_j(\bar{S}_j, \bar{A}_{j-1})\}}{p_j(d_j | \bar{S}_j, \bar{A}_{j-1})} \right) \left(\mu_{\ell+1}(\bar{S}_{\ell+1}, \bar{d}_\ell) - \mu_\ell(\bar{S}_\ell, \bar{d}_{\ell-1}) \right)$$

$$+ \left(\frac{I\{A_1 = d_1\}}{p_1(d_j)} \right) \left(\mu_2(S_2, d_1) - \mu_{\bar{d}_k} \right).$$

where the indicator, $I\{B\}$ is 1 if the event B occurs and 0 otherwise. Given approximations for each μ_j, (say $\hat{\mu}_j$) solve

$$0 = \mathbb{P}_n U(\bar{S}_k, \bar{A}_k; \bar{d}_k, \hat{\bar{\mu}}_k, \mu_{\bar{d}_k}) \tag{14}$$

for $\mu_{\bar{d}_k}$ and call the solution $\hat{\mu}_{\bar{d}_k}$. $\hat{\mu}_{\bar{d}_k}$ is our estimator of $EY(\bar{d}_k)$. In this equation, $\mathbb{P}_n f(X)$ is defined as the average over the sample, $1/n \sum_{i=1}^{n} f(X_i)$. One possible estimator for each $\hat{\mu}_j$ is

$$\hat{\mu}_\ell(\bar{s}_\ell, \bar{a}_{\ell-1}) = \frac{\mathbb{P}_n \left(\prod_{j=\ell}^{k} \frac{I\{A_j = d_j(\bar{S}_j, \bar{A}_{j-1})\}}{p_j(d_j | \bar{S}_j, \bar{A}_{j-1})} Y I\{(\bar{S}_\ell, \bar{A}_{\ell-1}) = (\bar{s}_\ell, \bar{a}_{\ell-1})\} \right)}{\mathbb{P}_n I\{(\bar{S}_\ell, \bar{A}_{\ell-1}) = (\bar{s}_\ell, \bar{a}_{\ell-1})\}}. \tag{15}$$

$\hat{\mu}_\ell(\bar{s}_\ell, \bar{d}_{\ell-1})$ is a consistent nonparametric estimator of $E[Y(\bar{d}_k)|\bar{S}_\ell = \bar{s}_\ell, \bar{A}_{\ell-1} = \bar{d}_{\ell-1}]$. The estimating formula, U can be greatly simplified because it is doubly robust (Robins, 2000) in that even if the approximations $\hat{\mu}_j$ are inconsistent, as long as the $\hat{\mu}_j$'s are bounded in probability the estimator $\hat{\mu}_{\bar{d}_k}$ of $EY(\bar{d}_k)$ remains consistent. Thus, for example, the estimator (7) for $EY(\bar{d}_k)$ results if we set each $\hat{\mu}_j$ to the constant mean $\mu_{\bar{d}_k}$ and solve for $\mu_{\bar{d}_k}$ in (14).

<u>Asymptotic Variance of $\sqrt{n}\hat{\mu}_{\bar{d}_k}$:</u> Again the potential outcomes notation from the beginning of this appendix is used. We derive the asymptotic variance for the doubly robust estimator; the estimator (7) follows if one replaces each μ_j by $\mu_{\bar{d}_k}$. Classical Taylor series argument along with technical assumptions (e.g. Serfling, 1980) imply that $\sqrt{n}(\hat{\mu}_{\bar{d}_k} - \mu_{\bar{d}_k})$ is, in large samples, equal to

$$\left(-E\left[\frac{\partial}{\partial \mu_{\bar{d}_k}} U(\bar{S}_k, \bar{A}_k; \bar{d}_k, \bar{\mu}_k, \mu_{\bar{d}_k}) \right] \right)^{-1} \left(\sqrt{n}\mathbb{P}_n U(\bar{S}_k, \bar{A}_k; \bar{d}_k, \bar{\mu}_k, \mu_{\bar{d}_k}) \right.$$

$$\left. + \sum_{j=1}^{k} E\left[\frac{\partial}{\partial \mu_j} U(\bar{S}_k, \bar{A}_k; \bar{d}_k, \bar{\mu}_k, \mu_{\bar{d}_k}) \right] \sqrt{n}(\hat{\mu}_j - \mu_j) \right)$$

plus terms converging in probability to zero. Note however that $E\left[\frac{\partial}{\partial \mu_j} U(\bar{S}_k, \bar{A}_k; \bar{d}_k, \bar{\mu}_k, \mu_{\bar{d}_k}) \right] = 0$; this is the Robins' double robustness property. Additionally it is easy to see that

25

$E\left[\frac{\partial}{\partial\mu_{\bar{d}_k}}U(\bar{S}_k,\bar{A}_k;\bar{d}_k,\bar{\mu}_k,\mu_{\bar{d}_k})\right]=1$. Thus the asymptotic variance of $\sqrt{n}(\hat{\mu}_{\bar{d}_k}-\mu_{\bar{d}_k})$ is equal to the asymptotic variance of $\sqrt{n}\mathbb{P}_n U(\bar{S}_k,\bar{A}_k;\bar{d}_k,\bar{\mu}_k,\mu_{\bar{d}_k})$. Next express all variables in terms of potential outcomes as in the beginning of this appendix and use the fact that randomization implies that for $j=1,...,K$, A_j is independent of $\{(S_1,S_2(a_1),S_3(\bar{a}_2),...,S_{k+1}(\bar{a}_{k+1})):$ $\bar{a}_k$ varying in $\mathcal{A}_k\}$ given $(\bar{S}_j,\bar{A}_{j-1})$. To be complete yet conserve space we provide a proof for $k=2$ (the general proof is similar but requires more space).

$$Var(U)=Var\left[\frac{I\{A_2=d_2(S_2(d_1),d_1)\}I\{A_1=d_1\}}{p_2(d_2(S_2(d_1),d_1)|S_2(d_1),d_1)p_1(d_1)}\Big(Y(d_1,d_2(S_2(d_1),d_1))-\mu_2(S_2(d_1),\bar{d}_1)\Big)\right.$$
$$\left.+\frac{I\{A_1=d_1\}}{p_1(d_1)}\Big(\mu_2(S_2(d_1),d_1)-\mu_{\bar{d}_2}\Big)\right].$$

Using the independence property in addition to the fact that for any random variables, U, Z, $Var(U)=Var(E(U|Z))+E(Var(U|Z))$ (put $Z=(Y(d_1,d_2(S_2(d_1),d_1)),S_2(d_1)))$ we obtain,

$$Var(U)=Var\left[\frac{I\{A_1=d_1\}}{p_1(d_1)}\Big(Y(d_1,d_2(S_2(d_1),d_1))-\mu_2(S_2(d_1),\bar{d}_1)\Big)\right.$$
$$+\frac{I\{A_1=d_1\}}{p_1(d_1)}\Big(\mu_2(S_2(d_1),d_1)-\mu_{\bar{d}_2}\Big)\right]$$
$$+E\left[\frac{(1-p_2(d_2(S_2(d_1),d_1)|S_2(d_1),d_1))\,I\{A_1=d_1\}}{p_2(d_2(S_2(d_1),d_1)|S_2(d_1),d_1)p_1(d_1)^2}\Big(\mu_2(S_2(d_1),d_1)-\mu_{\bar{d}_2}\Big)^2\right]$$

Simplify the first term and use the independence property on the second to obtain,

$$Var(U)=Var\left[\frac{I\{A_1=d_1\}}{p_1(d_1)}\Big(Y(d_1,d_2(S_2(d_1),d_1))-\mu_{\bar{d}_2}\Big)\right]$$
$$+E\left[\frac{(1-p_2(d_2(S_2(d_1),d_1)|S_2(d_1),d_1))}{p_2(d_2(S_2(d_1),d_1)|S_2(d_1),d_1)}\frac{1}{p_1(d_1)}\Big(\mu_2(S_2(d_1),d_1)-\mu_{\bar{d}_2}\Big)^2\right].$$

Once again applying similar arguments to the first variance term yields

$$Var(U)=Var\left[Y(d_1,d_2(S_2(d_1),d_1))-\mu_{\bar{d}_2}\right]$$
$$+E\left[\frac{1-p_1(d_1)}{p_1(d_1)}\Big(Y(d_1,d_2(S_2(d_1),d_1))-\mu_{\bar{d}_2}\Big)^2\right]$$
$$+E\left[\frac{(1-p_2(d_2(S_2(d_1),d_1)|S_2(d_1),d_1))}{p_2(d_2(S_2(d_1),d_1)|S_2(d_1),d_1)}\frac{1}{p_1(d_1)}\Big(\mu_2(S_2(d_1),d_1)-\mu_{\bar{d}_2}\Big)^2\right].$$

Now if all treatments are assigned according to $\bar{d}_k$ then each observed S_j is equal to the potential outcome, $S_j(\bar{d}_{j-1})$ and the observed Y is equal to the potential outcome $Y(d_1,d_2(S_2(d_1),d_1))$, hence we can re-express the above (for general k) as

$$Var_{\bar{d}_k}(Y)+$$
$$\sum_{j=2}^{k}E_{\bar{d}_k}\left[\frac{1-p_j(d_j|\bar{S}_j,\bar{d}_{j-1})}{p_j(d_j|\bar{S}_j,\bar{d}_{j-1})}\prod_{\ell=1}^{j-1}\frac{1}{p_\ell(d_\ell|\bar{S}_\ell,\bar{d}_{\ell-1})}\Big(Y-\mu_j(\bar{S}_j,\bar{d}_{j-1})\Big)^2\right]$$
$$+E_{\bar{d}_k}\left[\frac{1-p_1(d_1)}{p_1(d_1)}\Big(Y-\mu_{\bar{d}_k}\Big)^2\right]$$

where $\mu_j(\bar{s}_j, \bar{a}_{j-1})$ is the limit, in probability, of $\hat{\mu}_j(\bar{s}_j, \bar{a}_{j-1})$.

The variance in (10) follows immediately by setting each $\mu_j = \mu_{\bar{d}_k}$ in the above variance formula. Murphy et al. (2001) prove that U is the efficient influence function for estimation of $\mu_{\bar{d}_k}$; thus the variance of U is no larger than the variance of any regular estimator of $\mu_{\bar{d}_k}$. This means that (10) which is the variance of (7) must be no smaller than the above variance formula.

Alternate Representation of the Variance in (10): Suppose we use the randomization probabilities as in (5,6). Then

$$\prod_{j=1}^{k} p_j(A_j|\bar{S}_j, \bar{A}_{j-1})^{-1} = \left(\prod_{j=2}^{k} N_j(\bar{S}_j, \bar{A}_{j-1}) E\left[N_j(\bar{S}_j, \bar{A}_{j-1})^{-1} | \bar{S}_{j-1}, \bar{A}_{j-1} \right] \right) N_1(S_1).$$

Now if the number of treatment options at each time depends only on past treatment (i.e. $n_j(\bar{s}_j, \bar{a}_{j-1})$ can be written as $n_j(\bar{a}_{j-1})$ for each j) then as seen above each $N_j(\bar{S}_j, \bar{A}_{j-1})$ does not depend on $\bar{S}_j$. This means that the term in parentheses is one and the above simplifies to $\prod_{j=1}^{k} p_j(A_j|\bar{S}_j, \bar{A}_{j-1})^{-1} = N_1$. In this case the variance in (10) is equal to $N_1 \sigma^2$.

Upper Bound on Variance Under Extreme Working Assumption: We make the extreme working assumption that the distribution of each S_j given $(\bar{S}_{j-1}, \bar{A}_{j-1})$ assigns mass 1 to the value of S_j with the largest number of subsequent treatment strategies $(j \leq k)$. From the beginning of this appendix we have that $p_k(a_k|\bar{s}_k, \bar{a}_{k-1}) = N_k(\bar{s}_k, \bar{a}_{k-1})^{-1}$ and for each $j < k$,

$$p_j(a_j|\bar{s}_j, \bar{a}_{j-1}) = \frac{\max_{s_j} N_j(\bar{s}_j, \bar{a}_{j-1})}{N_{j-1}(\bar{s}_{j-1}, \bar{a}_{j-2})}$$

where

$$N_{j-1}(\bar{s}_{j-1}, \bar{a}_{j-2}) = \sum_{a_{j-1}=1}^{n_{j-1}(\bar{s}_{j-1}, \bar{a}_{j-2})} \max_{s_j} N_j(\bar{s}_j, \bar{a}_{j-1}).$$

Then

$$\prod_{j=1}^{k} p_j(a_j|\bar{s}_j, \bar{a}_{j-1})^{-1} = \prod_{j=1}^{k} \frac{N_j(\bar{s}_j, \bar{a}_{j-1})}{\max_{s_j} N_j(\bar{s}_j, \bar{a}_{j-1})} N_1.$$

Each fraction in the product is bounded above by 1, leaving N_1 as the upper bound.

Simulation Design:

The model is

$$Y(\bar{a}_2) = \beta_0 - \beta_1(a_1) + \phi(S_2, a_1) - \beta_2(S_2, \bar{a}_2) + \epsilon$$

where the effect of treatment relative to the optimal treatment at decision one and two is encoded in the nonnegative β_1 and β_2 respectively. We call β_1 and β_2 the regrets as they

are the regret one has for not assigning the optimal treatment. Recall the optimal adaptive treatment strategy is $d_1^* = 1$, $d_2^*(s_2, 1) = t_{s_2,1,1}$. Thus $\beta_1(1) = 0$ as we have no regret for assigning the initial treatment 1 and $\beta_2(0, 1, t_{0,1,1}) = \beta_2(1, 1, t_{1,1,1}) = 0$ as $t_{0,1,1}$ and $t_{1,1,1}$ are the optimal treatments following $a_1 = 1$ for each value of S_2. The term $\phi(S_2, a_1)$ has conditional mean zero $(f_{a_1}\phi(1, a_1) + (1 - f_{a_1})\phi(0, a_1) = 0$ for $a_1 = 1$ or 2). The error term ϵ, has conditional mean zero (given $S_2(a_1)$) for each value of a_2; otherwise the distribution of the error term can depend arbitrarily on $(S_2(a_1), \bar{a}_2)$. See Murphy (2003) for a detailed discussion of this model.

In all simulations, ϵ is a normal random variable with mean zero and variance depending on the value of $(S_2, \bar{a}_2)$, (say $\Sigma(S_2, \bar{a}_2)$). Also $\beta_0 = 0$.

In scenario 1, we set all the values of β_1, β_2 to zero except $\beta_1(2) = 1$, $\beta_2(1, 1, t_{1,1,2}) = \beta_2(1, 2, t_{1,2,2}) = 1$, indicating that if the initial treatment is 2 and then $S_2 = 1$ then it is best to choose treatment $t_{1,2,1}$ over $t_{1,2,2}$. Also $\beta_2(0, 1, t_{0,1,2}) = \beta_2(0, 2, t_{0,2,2}) = 1$ as well. The nuisance parameters are $\phi(0, 1) = f_1^2(1 - f_1)$ and $\phi(0, 2) = f_2^2(1 - f_2)$ (note since $\phi(S_2, a_1)$ has conditional mean zero this defines $\phi(1, 1), \phi(1, 2)$ as well). The values of f_1, f_2 are specified in Table 1. This is enough to determine the difference in mean response to the optimal strategy as compared to the least desirable strategy ($\Delta\mu$). In Table 1, we specify the signal-to-noise ratio, $\Delta\mu/\sigma$. This ratio divided by $\Delta\mu$ specifies the desired standard deviation $\sigma = Y(\bar{d}_2)$ for each strategy. We set the values of $\Sigma(S_2, \bar{a}_2)$ so as to achieve this standard deviation. In this scenario the poorest strategy is $(d_1 = 2, d_2(s_2, 2) = t_{s_2,2,2})$. Two strategies that performed equally well (so as to permit an assessment of the realized type I error rate) are $(d_1 = 1, d_2(s_2, 1) = t_{s_2,1,2})$ and $(d_1 = 2, d_2(s_2, 2) = t_{s_2,2,1})$.

In scenario 2, we set all the values of β_1, β_2 to zero except $\beta_1(2) = 1$, $\beta_2(1, 1, t_{1,1,2}) = 1$ and $\beta_2(0, 1, t_{0,1,2}) = 1$. Recall there is only one possible secondary treatment for $a_1 = 2$ and both values of S_2. The nuisance parameters are as before, $\phi(0, 1) = f_1^2(1 - f_1)$ and $\phi(0, 2) = f_2^2(1 - f_2)$. In this scenario the poorest strategy is $(d_1 = 2, d_2(s_2, 1) = t_{s_2,2,1})$. Two strategies that performed equally well are $(d_1 = 1, d_2(s_2, 1) = t_{s_2,1,2})$ and $(d_1 = 2, d_2(s_2, 2) = t_{s_2,2,1})$.

In scenario 3, we set $\beta_1(2) = 1 - f_2$, $\beta_1(1) = 0$ and all the values of β_2 to zero except $\beta_2(0, a_1, t_{0,a_1,2}) = 1$ for both $a_1 = 1$ and 2. Recall there is only one possible secondary treatment following $S_2 = 1$. Setting $\beta_1(2) = 1 - f_2$ was one way to ensure that there were two strategies with equal mean response. The nuisance parameters are as before, $\phi(0, 1) = f(_1^2(1 - f_1)$ and $\phi(0, 2) = f_2^2(1 - f_2)$. The values of f_1, f_2 are specified in Table 3. As in scenarios 1 and 2, we set the values of $\Sigma(S_2, \bar{a}_2)$ so achieve the standard deviation σ. In this scenario the poorest strategy is $(d_1 = 2, d_2(1, 1) = t_{1,2,1}, d_2(0, 1) = t_{0,2,2})$. Two strategies that performed equally well are $(d_1 = 1, d_2(0, 1) = t_{0,1,2}, d_2(1, 1) = t_{1,1,1})$ and $(d_1 = 2, d_2(s_2, 2) = t_{s_2,2,1})$.