Biometrics 1, 1–??

DOI: 000

October 2014

Model assessment in dynamic treatment regimen estimation via double robustness

Michael P. Wallace[1,*], Erica E. M. Moodie[1], and David A. Stephens[2]

McGill University,

[1]Department of Epidemiology, Biostatistics and Occupational Health,

1020 Pine Avenue West, Montreal, QC H3A 1A2,

[2]Department of Mathematics and Statistics, Montreal, QC H3A 2K6.

email: michael.wallace@mcgill.ca

Summary: Dynamic treatment regimens (DTRs) recommend treatments based on evolving subject-level data. The optimal DTR is that which maximizes expected patient outcome and as such its identification is of primary interest in the personalized medicine setting. When analyzing data from observational studies using semi-parametric approaches there are two primary components which can be modeled: the expected level of treatment and the expected outcome for a patient given their other covariates. In an effort to offer greater flexibility, so-called doubly-robust methods have been developed which offer consistent parameter estimators as long as at least one of these two models is correctly specified. However, in practice it can be difficult to be confident if this is the case. Using G-estimation as our example method, we demonstrate how the property of double robustness itself can be used to provide evidence that a specified model is or is not correct. This approach is illustrated through simulation studies as well as data from the Multicenter AIDS Cohort Study.

Key words: Adaptive treatment strategies; double robustness; dynamic treatment regimens; G-estimation; model selection; personalized medicine.

1. Introduction

Personalized medicine is an expanding area of interest in health research wherein patient management is driven by subject-level, rather than diagnosis-level, data. Within this framework theory concerning dynamic treatment regimens, or DTRs (Chakraborty and Moodie, 2013; Chakraborty and Murphy, 2014; Zhao and Laber, 2014), has developed. Its focus lies in estimating the optimal sequence of decision rules - taking patient information as input and recommending treatment as output - that maximize long-term patient outcome. Identifying the optimal DTR is therefore of considerable interest, and numerous methods have been developed to tackle this particular problem (for a comprehensive review, see Chakraborty and Moodie 2013).

A common theme in the DTR literature is the trade-off between methodological complexity and robustness to model mis-specification. This idea is manifest in the dichotomy between *single* and *double robustness*; while methods of the former type require a specific model to be correctly specified to ensure consistent estimators, the latter requires correct specification of *at least one* of two models. These doubly-robust methods therefore offer an attractive property that typically comes at the cost of more difficult implementation. For example, the relatively well-established method of Q-learning (Watkins, 1989; Sutton and Andrew, 1998) relies largely on standard regression techniques, but also on the correct specification of the outcome mean model. An alternative approach of G-estimation (Robins, 2004), meanwhile, introduces some additional methodological complexity but in return offers consistent estimators if either the outcome model or the model of expected treatment level given patient information is correct. Other doubly-robust approaches to DTR estimation include dynamic weighted ordinary least squares (Wallace and Moodie, 2015), A-learning (Murphy, 2003; Robins, 2004; Blatt et al., 2004; Schulte et al., 2014), targeted maximum likelihood (van der Laan and Rubin, 2006; Neugebauer et al., 2010), and augmented inverse

probability of treatment weighting (Zhang et al., 2012). In this paper, we will focus on G-estimation, however our proposed diagnostic approach applies directly to other doubly-robust methods of estimation.

It is usually the case that the analyst does not know with certainty whether either of the two models required for consistency is correct, and as such doubly-robust methods cannot be viewed as a 'statistical panacea' to be used with impunity. Indeed, some authors have counselled caution when both models are even 'slightly' mis-specified (see Kang and Schafer 2007 and associated Comments), and with DTRs often encompassing multiple stages of modeling the combined impact of repeated mis-specification could be considerable. Thus, there is a real need to develop methods of model assessment, with a view to ensuring at least one model is indeed correctly specified.

In this paper we present how the property of double robustness can be exploited to either provide evidence that neither model has been correctly specified, or reassurance that at least one has. The basic principle we shall use has been briefly noted by Robins and Rotnitzky (2001) when considering comparing doubly-robust estimates with two different singly-robust ones, while Bang and Robins (2005) also touch on the issue in the context of a missing data problem. Pursuing these ideas in greater depth, we provide a detailed discussion of earlier work, consider the difficulties with formal testing, present some novel approaches within this framework and demonstrate how we can proceed using only the desired doubly-robust method (rather than requiring additional singly-robust approaches to be considered). These ideas are illustrated in the specific case of G-estimation within the DTR environment, wherein some additional challenges lie.

2. DTRs and G-estimation

We first introduce the general DTR estimation problem as well as the G-estimation process by which we shall illustrate the model checking approach. For simplicity of exposition and

notation we shall first limit ourselves to the *single-stage* setting - where one treatment decision is to be made for each patient - but note everything presented extends to the multi-stage setting (which will be illustrated through simulated and real datasets).

For a given dataset we aim to identify the *optimal* DTR: the decision rule recommending the treatment a given patient information (or history) x that maximizes expected patient outcome $\mathbb{E}[Y|A = a, X = x]$ (with Y chosen so that larger values are preferred). A typical model for Y could be $\mathbb{E}[Y|A = a, X = x; \beta, \psi] = f(x^\beta; \beta) + \gamma(a, x^\psi; \psi)$ where x^β and x^ψ are two (potentially identical) subsets of the full patient information x. This formulation sees the mean outcome split into two components: the *treatment-free* (sometimes *expected counterfactual*) $f(x^\beta; \beta)$ and the *blip* $\gamma(a, x^\psi; \psi)$. The blip function is the expected mean difference between patients who receive treatment a and those who receive some baseline (or control) treatment. As such, the treatment only affects outcome through the blip function, and so the optimal treatment for a patient with information x^ψ is that which maximizes $\gamma(a, x^\psi; \psi)$. In the common case of a binary treatment $a \in \{0, 1\}$ (where $a = 0$ corresponds to no treatment or a control) and a linear blip $\gamma(a, x^\psi; \psi) = ax^\psi\psi$ the optimal regimen is particularly straightforward: "treat ($a = 1$) if $x^\psi\psi > 0$ and do not treat ($a = 0$) otherwise". If we obtain estimates of the blip parameters then the estimated optimal regimen follows.

While this appears a straightforward regression problem it is complicated by the treatment-free component of the outcome model which is seldom known or reliably estimable. G-estimation is one approach that has been developed to circumvent this. In addition to specifying models for the blip function (which must be done correctly), and the treatment-free model, one may also specify a *treatment model* $\mathbb{E}[A|x^\alpha; \alpha]$ (which is the propensity score when treatment is binary). G-estimation then produces consistent estimates of ψ as long as at least one of the treatment or treatment-free models is correctly specified. Thus G-estimation is said to be doubly-robust.

The G-estimation approach, like some other DTR estimation methods such as Q-learning and dWOLS, proceeds in a recursive manner. In the multi-stage setting, the analyst begins by estimating the optimal final-stage treatment and then, using this information, moves backwards through stages estimating each previous optimal decision rule in turn. This recursive approach allows each subsequent decision to be based on the assumption that all future treatments are optimal. The specific details of G-estimation are included in the Web Appendix. An important aspect of G-estimation within the DTR framework is that its recursive nature means blip parameter estimates from later stages are used for estimation at earlier ones. As such, any mis-specified models at later treatment stages could 'feed' into this process, potentially adversely affecting results at earlier stages even if the proposed models at those stages are correct. Additional care must therefore be taken with model selection and validation in the DTR setting.

3. Exploiting double robustness

G-estimation offers estimators that are robust to mis-specification of one of two component models, but in practice it may not be possible to say with much certainty whether at least one model is correctly specified. Nevertheless we can still investigate this in a relatively simple manner. Suppose one model is correctly specified. By the definition of double robustness, estimators will remain consistent if analyses are repeated using the same correct model while changing the other. Furthermore, it is known that variance of the estimators attains a minimum when both models are correctly specified (see Web Appendix). For example, suppose our treatment model is correctly specified but not our treatment-free model. If we repeat the analysis but with a *different* treatment-free model then the resulting parameter estimates should be similar (if n is large) to those obtained from our first analysis. If, however, changing our treatment-free model resulted in a dramatic shift in our parameter estimates,

this would provide evidence our treatment model was mis-specified. The same argument applies to a specific treatment-free model with varying treatment models.

Robins and Rotnitzky (2001) mention this property of doubly-robust estimators in passing and comment that comparison of estimates resulting from doubly-robust analyses with those from the appropriate singly-robust ones provides a "useful informal goodness of fit test". However, we note a singly-robust method can be viewed as a special case of a doubly-robust approach where the second model is mis-specified. We further observe that this logic extends to comparing doubly-robust estimates where one model is kept fixed while the other is varied.

Taking advantage of this property to inform analysis is, however, non-trivial. Deriving formal statistical tests presents a particular challenge as both the interdependence of our model choices and the resulting parameter estimates needs to be considered. There are, of course, standard tests available which may appear to serve our requirements. For example, in their setting of estimating the mean of a partially observed outcome Bang and Robins (2005) explicitly outline a typical Z-test type approach. Using estimates $\widehat{\mu}^D$ and $\widehat{\mu}^S$ resulting from doubly- and singly-robust analyses respectively, they observe that if $\widehat{\tau}^2$ is the empirical variance of $\widehat{\mu}^D - \widehat{\mu}^S$ (estimated via bootstrap, perhaps) then a test with rejection region $\left|(\widehat{\mu}^D - \widehat{\mu}^S)/(\widehat{\tau})\right| > 1.96$ is a valid - though inconsistent - large sample 0.05 level test of the null hypothesis that the model used in the singly-robust method is correctly specified.

The lack of consistency relates to a general problem when using doubly-robust methods in this way: two wrong models may still lead to the same wrong estimate. A further complication here is that Bang and Robins' example concerns a single parameter estimate, whereas in practice we will typically deal with vectors of blip parameters. As such a multivariate extension (such as Hotelling's T-squared test) may prove useful in identifying whether our proposed models have been correctly specified or, as we shall demonstrate, by considering statistics which 'collapse' our vectors of parameter estimates into univariate statistics. Again,

however, incautious use of such tests may prove unwise given the potential for deceptive results.

Within the multi-stage DTR framework, methods such as G-estimation provide an additional challenge. A recursive approach uses results from each stage of an analysis in the estimation process for previous stages. Incorrect modeling at the final stage, for example, could then render any subsequent work useless. We observe, however, that in practical situations we implicitly assume that our model fitting at every stage is correct – or at least 'correct enough'. We also note that if we happened to know the correct stage 1 treatment model (such as via knowledge of the study design) we may be able to assess our stage 2 model specification. If analysis using the correct stage 1 treatment model and a variety of stage 1 treatment-free models resulted in widely varying stage 1 parameter estimates, for example, this could be suggestive of an error in our stage 2 model fitting. This latter example is, however, something of a special case, and as such we shall not pursue it here.

Thus, rather than pursuing a formal test, we suggest instead a variety of quick and simple investigations following analysis with a number of different models. This gives rise to the following informal 'step-by-step' procedure where, at each stage of our analysis we perform the following steps:

(1) Propose a set of candidate treatment models and treatment-free models.

(2) Using bootstrap replicated datasets, perform G-estimation using every possible combination of treatment and treatment-free models.

(3) Within each bootstrap sample, examine the variation in estimates obtained for each given treatment model under every different treatment-free model. The correct treatment model should be identifiable by being that which results in the lowest such variation most often. The treatment-free models may be assessed similarly.

(4) If all models lead to widely varying estimates (identifiable through, for example, boxplots

of bootstrapped point estimates), consider the possibility of all models being wrong and investigate new candidate models.

(5) If there are further stages to analyze, use the above to choose the best combination of treatment and treatment-free model and use these when analyzing the earlier stages.

More precise details of this approach are demonstrated in the examples that follow.

3.1 Simulated example: a simple illustration

While our natural interest lies in the multi-stage setting, we begin with a simple hypothetical example in the single-stage setting so as to most clearly illustrate our basic principles. As per our previous notation we focus on a single treatment decision a based on a single piece of patient information x and wish to maximize an expected outcome. Our data (sample size $n = 1,000$) are generated as $X \sim U(0,1)$, $A \in \{0,1\}$ with $\mathbb{P}(A = 1|X = x) = x$, and $Y \sim N(1 + x + e^x + A(\psi_0 + \psi_1 x), 0.5)$ with $\psi_0 = \psi_1 = 1$. The treatment-free model in this setup is thus $f(x) = 1 + x + e^x$, the blip model is $\gamma(x, a; \psi) = a(\psi_0 + \psi_1 x)$ and the treatment model is $\mathbb{P}(A = 1|X = x) = x$.

We conduct analyses using three treatment models and three treatment-free models, with only the first of each three correctly specified. Our treatment models are $T_1 : \mathbb{P}(A = 1|X = x) = x$, $T_2 : \mathbb{P}(A = 1|X = x) = x^2$, and $T_3 : \mathbb{P}(A = 1|X = x) = 0.5$, while our treatment-free models are $F_1 : f(x;\beta) = \beta_0 + \beta_1 x + \beta_2 e^x$, $F_2 : f(x;\beta) = \beta_0 + \beta_1 x + \beta_2 \log(x)$, and $F_3 : f(x;\beta) = \beta_0 + \beta_1 x$. Performing our analyses nine times - once for each treatment and treatment-free model pair - the double robustness property of G-estimation should lead to biased or unbiased results depending on whether at least one is correctly specified.

We carry out such analyses on $10,000$ simulated datasets. In Table 1 we first present the results of analysis on just one of these datasets to illustrate our general principle: when either the treatment or treatment-free model is correct (T_1 or F_1) estimates exhibit little bias, whereas when both are mis-specified bias is evident. Moreover, if we compare the standard deviation in parameter estimates across rows (keeping the treatment model the same), or

down columns (keeping the treatment-free model the same) the lowest values occur when one model is correctly specified. This pattern manifests across the entire simulation run (Figure 1), and can also be applied to a single dataset via bootstrapping, which yields near-identical results to those observed across simulation runs.

[Table 1 and Figure 1 about here.]

Another graphical assessment method is possible by comparing point estimates from various models on each individual bootstrap resample: when one model is correct we should find that each analysis yields more similar estimates than if both models are wrong. An illustration is presented in the Web Appendix where, from analysis of bootstrap samples of a single dataset, estimates resulting from analyses using different treatment-free models are visibly more correlated when the treatment model is correctly specified.

While our primary interest lies in these blip parameter estimates, comparisons such as those above become impractical once we move into more complex (higher dimensional) settings. Instead, we can consider the linear combination of our blip parameters and covariates: $\psi_0 + \psi_1 x$ (i.e. our blip function without the preceding a term) which we regard as our 'treatment effect': the change in expected outcome from receiving treatment. This combines our parameter estimates into a single numeric value, and while this necessarily comes at the cost of a loss of some information, it yields a more straightforward means to assess our models. For example, from a model selection perspective we might choose the treatment model that results in the lowest standard deviation in mean treatment effect estimates for the dataset across the three treatment-free models, and vice-versa. Across our 10,000 simulation runs this selects the correct treatment-free model 99% of the time, but only selects the correct treatment model some 60% of the time, selecting the 'null' model ($\mathbb{P}(A = 1) = 0.5$) the remaining 40%. Alternatively, working with an individual dataset we could identify the model which results in the lowest standard deviation when analyzing each bootstrap sample,

and then choose that which yields the lowest standard deviation most often. Across 1,000 simulated datasets, with 200 bootstrap samples of each, this selected the correct treatment-free model for over 99% of the datasets and the correct treatment model in 75%.

The apparent poor performance across multiple simulation runs for selecting the correct treatment model is a consequence of the simplifying step of collapsing our parameter estimates into a single function. In contrast to when we compared blip parameter estimates directly (where the correct treatment *and* treatment-free models are selected in over 98% of simulations), this simplification led to a situation where despite the 'null' model resulting in markedly biased estimates, they combined with the covariate x to return very similar treatment effect estimates across the three treatment-free models. This further highlights how a 'quick fix' for model selection in this manner may be optimistic, and that it is prudent to look at individual parameter estimates.

3.2 A two-stage example

The preceding example demonstrated some fundamental principles behind model assessment and selection using doubly-robust approaches. Interested as we are in the DTR framework, however, we now pursue a more complex, two-stage example. For these simulations we introduce the stage j *regret* $\mu_j(h_j, a_j)$, defined as the expected decrease in our outcome from using treatment a_j at stage j instead of the optimal treatment, assuming optimal treatment thereafter. In the binary treatment setting the regret relates to the blip via $\mu_j(h_j, a_j) = \gamma_j(h_j, a_j^{opt}) - \gamma_j(h_j, a_j)$, where a_j^{opt} denotes the stage-j optimal treatment.

Our simulations are intended to reflect a common real-world scenario where the same covariates are recorded at each stage, and where stage 2 covariates may be affected by the treatment received at stage 1. Data are generated as follows:

- Stage 1: $X_{11} \sim N(0,1)$, $X_{12} \sim N(0,1)$, $X_{13} \sim N(0,1)$

$$\mathbb{P}(a_1 = 1 | x_{11}, x_{12}, x_{13}) = \text{expit}(x_{11} + x_{12} + x_{13})$$

$$\mu_1(h_1; \boldsymbol{\psi}_1) = (a_1^{opt} - a_1)(\psi_{10} + \psi_{11} x_{11});$$

- Stage 2: $X_{21} \sim N(x_{11} + a_1, 1)$, $X_{21} \sim N(x_{12} + a_1, 1)$, $X_{23} \sim N(x_{13} + a_1, 1)$

$$\mathbb{P}(a_2 = 1 | x_{21}, x_{22}, x_{23}) = \text{expit}(x_{21} + x_{22} + x_{23})$$

$$\mu_2(h_2; \boldsymbol{\psi}_2) = (a_2^{opt} - a_2)(\psi_{20} + \psi_{21} x_{21});$$

- Outcome: $Y \sim N(x_{11} + x_{12} + x_{13} - \mu_1(x_1, a_1; \psi_1) - \mu_2(x_2, a_2; \psi_2, 1);$

where $\text{expit}(x) = 1/(1+e^{-x})$ is the logistic function, $\psi_{10} = \psi_{20} = -\frac{1}{2}$ and $\psi_{11} = \psi_{21} = 1$. Our regret functions allow an alternative formulation of our expected outcome, which we can view as an optimal outcome $Y^{opt} = x_{11} + x_{12} + x_{13}$ from which the regret functions are subtracted. The regret functions correspond to blips of the form $\gamma_j(x_j, a_j; \psi_j) = a_j(\psi_{j0} + \psi_{j1} x_j)$, and so $a_j^{opt} = 1$ if $\psi_{j0} + \psi_{j1} x_j > 0$ and 0 otherwise. This leads to true treatment-free models complicated by the presence of indicator functions:

- $f_1(\boldsymbol{x}; \boldsymbol{\beta}_1) = x_{11} + x_{12} + x_{13} - 1_{\{\psi_{10} + \psi_{11} x_{11} > 0\}}(\psi_{10} + \psi_{11} x_{11})$; and
- $f_2(\boldsymbol{x}; \boldsymbol{\beta}_2) = x_{11} + x_{12} + x_{13} - (1_{\{\psi_{10} + \psi_{11} x_{11} > 0\}} - a_1)(\psi_{10} + \psi_{11} x_1) - 1_{\{\psi_{20} + \psi_{21} x_{21} > 0\}}(\psi_{20} + \psi_{21} x_{21})$.

The precise form of these models depends on the very parameters we are trying to estimate, and so in reality their identification can be difficult. However, one ad hoc approach in this setting is to specify the general form of our treatment-free models based on our blips and an estimate of the treatment decision rule which governs the indicator functions. We could obtain such an estimate through a preliminary analysis, a prior belief, or some other means. Here, for the sake of illustration, we propose reforming the indicator function used at stage 1 as "treat ($a_1 = 1$) if a subject's x_{11} value is greater than the population mean of x_{11}", and similarly for stage 2 based on x_{21}. As with our single-stage analysis, we consider 3 treatment models and 3 treatment-free models, each of which corresponds to including 1, 2, or all 3 of the available covariates. This mimics real-world model assessment problems where inclusion or exclusion of a particular variable is often of interest.

At each stage we assume a logistic regression model for treatment, with model T_{jp} denoting

the treatment model for stage j that contains p covariates, with $p = 1, 2, 3$ corresponding to (x_{j1}), (x_{j1}, x_{j2}) and (x_{j1}, x_{j2}, x_{j3}), respectively. At stage 1 our treatment-free models are linear, containing covariates $(\widetilde{a}_1^{opt}, \widetilde{a}_1^{opt} x_{11})$ where $\widetilde{a}_1^{opt}$ takes the value 1 for a particular subject if their x_{11} value is larger than the overall mean of x_{11}, and 0 otherwise (in other words, our 'working estimate' of the optimal stage 1 treatment). These covariates represent the component of our treatment-free model corresponding to the stage 1 regret when $a_1 = 0$. Treatment-free model F_{1p} then contains these terms along with p covariates, analogous to the definition of our treatment models. For example, the covariates in treatment-free model F_{12} are $(\widetilde{a}_1^{opt}, \widetilde{a}_1^{opt} x_{11}, x_{11}, x_{12})$. In similar fashion, our stage 2 treatment-free models contain covariates $(a_1, a_1 x_{11}, \widetilde{a}_1^{opt}, \widetilde{a}_1^{opt} x_{11}, \widetilde{a}_2^{opt}, \widetilde{a}_2^{opt} x_{21})$, where $\widetilde{a}_2^{opt}$ takes the value 1 if x_{21} is larger than the overall mean of x_{21}, and 0 otherwise. These terms correspond to the stage 1 regret and the stage 2 regret with $a_2 = 0$. Treatment-free model F_{2p} then contains these terms along with p covariates. For example, the covariates in treatment-free model F_{23} are $(a_1, a_1 x_{11}, \widetilde{a}_1^{opt}, \widetilde{a}_1^{opt} x_{11}, \widetilde{a}_2^{opt}, \widetilde{a}_2^{opt} x_{21}, x_{11}, x_{12}, x_{13})$.

While our treatment-free models may seem somewhat complex, they merely reflect what we might propose in this context based on our assumed form of the blip. In particular, we note that by specifying the form of our blip, this in turn defines the form of the regret, and thence some of the terms of our treatment-free model. While this requires us to estimate the optimal treatment despite this being the objective of our analysis, we have found that in practice including some sort of indicator function term can dramatically improve estimation even when that estimate is not particularly accurate. Our focus here, meanwhile, is on identifying which covariates to include in our treatment and treatment-free models.

Each of 1,000 simulation runs involved taking 100 bootstrap samples of a size $n = 1,000$ dataset generated as per the above procedure, fitting every combination of stage 1 and stage 2 models. When investigating our stage 1 estimates we separate our results between those

where at least one of the stage 2 models was correct and those where both stage 2 models were incorrect, to investigate whether mis-specification had a marked effect on the subsequent stage 1 analysis. Focusing on the mean treatment effect estimates we find the correct models, as expected, tend to result in lower variation across the different treatment models, with the correct models being identified in every simulation run. Furthermore, mis-specification of stage 2 models had little impact on model selection at stage 1.

In Figure 2 we illustrate visual diagnostics for assessing model validity by looking at the bootstrap blip parameter estimates for a single simulation run. In particular, we note that when analysis is repeated using the same mis-specified treatment (or treatment-free) model and each of the different treatment-free (or treatment) models, then considerable variation in blip parameter estimates is evident. In contrast, when analysis is repeated using a fixed correct model there is little variation. Plots for the overall treatment effect (rather than a specific parameter estimate) exhibited a similar pattern.

[Figure 2 about here.]

Standard methods for assessing the treatment model do, of course, exist, and in these simulations the use of the Bayesian Information Criterion in fitting our logistic regression models resulted in the same success rate in identifying the correct treatment model. In contrast, not only are there no established methods for identifying the correct treatment-free model in such scenarios, none have, to our knowledge, even been employed. Regardless, a key advantage of this approach of exploiting double robustness is that it offers the potential for model *validation*, and not just model assessment.

Here, for instance, we have simulated an example where there are three covariates that influence our models. However, it is entirely plausible to encounter a situation such as this where one of those covariates is not considered, either through omission at the modeling stage, or even not collected at all. In this scenario standard model selection techniques would be

limited in their ability to detect that a covariate was missing: they can tell us which model is *best*, but are less able to tell us that model is still mis-specified. In contrast, were we to conduct an analysis based on examining variation in blip parameter (or treatment effect) estimates through boxplots such as Figure 2 we may have cause for concern. As a simple illustration, suppose in the preceding analysis we had not considered models containing x_{13} or x_{23}, in effect simulating the real-world scenario of omitting a covariate altogether. Ignoring those boxes corresponding to these models in Figure 2 we would have considerable concerns about either of our two remaining models being correct given the clear variation in parameter estimates. (On the other hand, when we do include the correct model, this approach offers reassurance that we have selected the *correct* choice, and not merely the best one.)

3.3 The Multicenter AIDS Cohort Study

We now apply our approach to data from the Multicenter AIDS Cohort Study (MACS, Kaslow et al. 1987). Beginning in 1984, and accumulating information from HIV-1 infected homosexual and bisexual men in four US cities, MACS is a longitudinal observational study designed to examine the natural history of AIDS. Participants are seen every three to six months where they complete a questionnaire, undergo physical examination, and provide blood. To illustrate our approach we present a relatively simplistic analysis of these data.

Following Moodie et al. (2007) we seek to identify the optimal time to implement zidovudine (AZT, used in the treatment of HIV/AIDS) with a view to maximizing CD4 count at 12 months after the baseline visit. We consider two treatment intervals: baseline to six months (stage 1), and six to 12 months (stage 2), during which AZT could be initiated. Our analysis is restricted to a dataset of 1,879 HIV-positive, AIDS-free men recruited after March 1986 when AZT became available, and we use last observation carried forward when observations were unobserved (Hernán et al., 2000). As per Moodie et al. (2007) we use the previous interval's CD4 count as the sole tailoring variable in our blip models. Treatment during the

first six months was assumed dependent on year of study entry, presence of AIDS symptoms at baseline, and baseline CD4 count. Treatment during the second six months was modeled on presence of symptoms at six months and six month CD4 count.

We denote stage j CD4 count by c_j and presence of symptoms s_j (binary), where $j = 1$ corresponds to baseline and $j = 2$ to six months. Year of study entry is time-invariant and denoted by e, while our outcome of 12 month CD4 count is denoted by y. AZT initiation during the first and second six months is denoted by $a_1 = 1$ and $a_2 = 1$, respectively, with $a_j = 0$ otherwise. Note that once treatment is initiated it is not discontinued, and so from a DTR perspective all patients who initiated in the first six months are receiving the 'best' (i.e. only) treatment possible in stage 2 given their first stage treatment. As such, we exclude these patients when estimating stage 2 treatment effects (but they are still used for stage 1 modeling). Of the 1,879 participants, 166 initiated AZT in stage 1 and so while all 1,879 are used in our estimation of stage 1 parameters, only $1,713$ are used for the stage 2 analysis.

As noted above, the models chosen at stage 2 will influence the estimates obtained at stage 1. Here, we therefore first compare treatment and treatment-free models in terms of their stage 2 blip parameter estimates before selecting one choice for each and proceeding to compare models at stage 1. At both stages we assume simple linear blip functions of the form $\gamma(a_j, x_j^{\psi}; \psi_j) = a_j c_j^{\psi} \psi_j$. Our candidate stage 2 treatment models are:

- T_{21}: $\mathbb{P}(A_2 = 1 | C_2 = c_2, S_2 = s_2) = \text{expit}(\alpha(1, c_2, s_2)^T)$;
- T_{22}: $\mathbb{P}(A_2 = 1 | S_2 = s_2) = \text{expit}(\alpha(1, s_2)^T)$; and
- T_{23}: $\mathbb{P}(A_2 = 1) = 0.5$

where parameters α are estimated via logistic regression of stage 2 treatment a_2 on the corresponding covariates. As such, our 'primary' treatment model T_{21} is a logistic regression of treatments on stage 2 CD4 count and stage 2 presence of symptoms, T_{22} is a logistic regression of treatments on stage 2 presence of symptoms alone, while T_{23} can be considered the 'null' model assuming no relationship between treatment and patient history. We consider

four treatment-free models of the form $f(h_2;\beta) = h_2\beta$ where for F_{21}, F_{22}, F_{23}, and F_{24}, $h_2 = (1, c_2)$, $(1, c_2, \log(c_2))$, $(1, c_2, c_1)$ and $(1, c_2, \log(c_2), c_1, \log(c_1))$, respectively. This represents two pairs of dichotomies: whether to include c_1 as well as c_2, and whether to include log-transforms of the CD4 counts. We took 200 bootstrap samples and carried out G-estimation for each of these model combinations.

In Figure 3 we summarize the standard deviations of blip parameter and treatment effect estimates for each fixed model, noting that while the blip parameters seem to suggest the first treatment model, things are rather less clear with the treatment effect results. Meanwhile, all three metrics suggest that it is advisable to include log-transformations of CD4 counts in our treatment-free models, along with some evidence that including baseline CD4 counts (c_1) as well as those at six months is worthwhile. If we inspect which models resulted in the lowest variation in estimates for each bootstrap sample (Table 2), we find treatment model T_{21} is overwhelmingly preferred based on blip parameter estimates, and is still the preferred model based on treatment effect estimates, albeit less clearly. The treatment-free models F_{22} and F_{24} were clearly preferred by both metrics with little to distinguish them. Meanwhile, the resulting point estimates of the blip parameters (Table 3) exhibit considerably more variation under treatment models T_{22} and T_{23}, and treatment-free models F_{21} and F_{23}. Based on these results, we proceed to stage 1 of the analysis using models T_{21} and F_{24} (the latter chosen over F_{22} given its marginally better performance).

[Figure 3 and Table 3 about here.]

With our stage 2 models determined we now consider possible stage 1 models. Again we consider three treatment models analogous to those investigated at the second stage with the addition of year of entry (e):

- T_{11}: $\mathbb{P}(A_1 = 1|C_1 = c_1, S_1 = s_1, E = e) = \text{expit}(\alpha(c_1, s_1, e)^T)$;
- T_{12}: $\mathbb{P}(A_1 = 1|S_1 = s_1, E = e) = \text{expit}(\alpha(s_1, e)^T)$; and
- T_{13}: $\mathbb{P}(A_1 = 1) = 0.5$

with α estimated from fitting logistic regression models of stage 1 treatment a_1 on the corresponding covariates. We consider just two treatment-free models of the form $f(h_1;\beta) = h_1\beta$ where for F_{11} and F_{12}, $h_1 = (1, c_1)$ and $(1, c_1, \log(c_1))$, respectively.

G-estimation proceeds using the recommended models for the second stage of treatment followed by all combinations of these stage 1 treatment and treatment-free model pairs. Analogous statistics based on standard deviations are also summarized in Table 2. We observe in particular that while the parameter-based metrics again strongly favour model T_{11} this apparently clear recommendation is largely disguised when considering the treatment effect statistic instead. Meanwhile, model F_{12} is - perhaps unsurprisingly given our stage 2 results - convincingly preferred. The resulting parameter estimates are also summarized in Table 3, where again our selected models seem to result in considerably less variable parameter estimates. At both stages, however, these estimates are not significantly different from zero.

4. Discussion

The potential for model mis-specification exists in virtually all data analyses, and the estimation of dynamic treatment regimens is no exception. Indeed, in this setting recursive methods and covariate-dependent treatments present additional challenges which can make correct model specification both more important and more challenging. Doubly-robust methods offer a convenient, intuitive, and novel route to diagnosing model inadequacies. While doubly-robust methods are not universally preferred (Kang and Schafer, 2007), the methods proposed in this paper can shed light on their reliability using straightforward, easy-to-apply checks. By exploiting the simple fact that estimators are consistent as long as one model is correct we are able to investigate the validity - or otherwise - of other models under consideration.

This fundamental principle has been noted before (Robins and Rotnitzky, 2001; Bang and Robins, 2005; Ogburn et al., 2014), though not in great depth. Further, the use of multiple

singly-robust estimates to check a doubly-robust estimate seems unnecessarily complex. In this paper, we have presented how the consistency of doubly-robust estimators can be exploited for the purpose of model assessment, and provided a concrete and intuitive means to do so. Along with simple inspection of the resulting point estimates, visual diagnostics offer tangible and intuitive insights into how one's models compare. We would also argue that, just as the study of residual plots is essential in a standard linear regression, the diagnostic plots we propose should form a routine part of any analysis conducted using doubly-robust methods.

A key property of this approach is that rather than merely an avenue towards model selection, double robustness may be used for model *validation*. In particular, while standard methods for selecting a treatment model may be used within the DTR framework, they will often only be able to identify the *best* model, without necessarily providing insights into whether that model is correct. Moreover, assessment of the treatment-free model remains much harder, with Rich et al. (2010) proposing the only notable method to date.

A valuable direction for future work would be an investigation of more formal testing procedures. We are also interested in the idea of the degree to which models may be misspecified (informally, the idea of one model being 'worse' than another on some scale). We note that our approach can potentially identify when two or more models are 'more correct' than others (as we saw in the MACS example), and that in this instance choosing between such models should have little impact as the resulting parameter estimates are necessarily close to one another. Additional simulations that introduce this concept (as well as others demonstrating the methods discussed in this paper) are included in the Web Appendix.

Related work has been conducted in the general structural mean model (SMM) environment for a single interval in the context of non-compliance in randomized trials. Recently Taguri et al. (2014), building on the work of Pan (2001), introduced a model selec-

tion approach based on *quasi-likelihood* and an extension of Akaike's Information Criterion (Akaike, 1973). While their work could be translated to the DTR setting, the focus of their approach is solely on selection of the general analog of our blip model, whereas we have been primarily concerned with the two 'nuisance' models besides the blip. That is, Taguri et al. (2014) assume fixed nuisance models so that consistency of the blip parameters is assured provided at least one of the two is correct. In contrast, we do not assume that model specification of the nuisance models is correct but rather aim to 'diagnose' whether this might be the case. Like Taguri et al. (2014), Fischer et al. (2011) also consider SMMs in the presence of non-compliance with treatment and propose a 'partial' goodness of fit test. We have investigated application of this approach in the G-estimation setting and found that it does not distinguish between mis-specification of treatment and treatment-free (or indeed blip) models. Also, as Taguri et al. (2014) note, this test will tend to select models which are too large. Nevertheless this presents a possible avenue for further investigation of the general model selection problem in doubly-robust settings.

An alternative perspective has been illustrated in the missing data setting by Han (2014). In contrast to our general approach of comparing several estimators from multiple models, Han combines the working models so as to derive a single final estimator that is consistent as long as at least one of the working models is correct. It may be possible to extend this idea to the DTR framework. Until such a 'multiply-robust' model is proposed, the techniques that we have proposed, which fully exploit the properties of double robustness, are important new tools for model assessment and selection.

Supplementary Materials

Details of G-estimation's implementation (referenced in Section 2), some additional theory concerning the variance of its estimators (referenced in Section 3), results of some further simulation studies (referenced in Sections 3.1 and the Discussion), and code used for gener-

ating the simulations reported in Sections 3.1 and 3.2 are available with this paper at the *Biometrics* website on Wiley Online Library.

Acknowledgments

This work is funded by Discovery Grants from the Natural Sciences and Engineering Research Council (NSERC) of Canada. Dr. Moodie is sponsored by a Chercheur-boursier career award from the Fonds de recherche du Québec-Santé (FRQS). Data were collected by the Multicenter AIDS Cohort Study with centers (Principal Investigators) at: Johns Hopkins University Bloomberg School of Public Health (Joseph Margolick), U01-AI35042; Northwestern University (Steven Wolinsky), U01-AI35039; University of California, Los Angeles (Roger Detels), U01-AI35040; University of Pittsburgh (Charles Rinaldo), U01-AI35041; the Center for Analysis and Management of MACS, Johns Hopkins University Bloomberg School of Public Health (Lisa Jacobson), UM1-AI35043. The MACS is funded primarily by the National Institute of Allergy and Infectious Diseases (NIAID) with additional co-funding from the National Cancer Institute (NCI). Targeted supplemental funding for specific projects was also provided by the National Heart, Lung, and Blood Institute (NHLBI) and the National Institute on Deafness and Communication Disorders (NIDCD). MACS data collection is also supported by UL1-TR000424 (JHU CTSA). Website: http://www.statepi.jhsph.edu/macs/macs.html.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In Petov, B. and Csḱi, F., editors, Second International Symposium on Information Theory, pages 267–281. Budapest, Hungary: Akadémiai Kiadó.

Bang, H. and Robins, J. M. (2005). Doubly-robust estimation in missing data and causal inference models. Biometrics **61,** 962 – 972.

Blatt, D., Murphy, S. A., and Zhu, J. (2004). A-learning for appropriate planning. Technical Report 04-63, The Methodology Center, Pennsylvania State University.

Chakraborty, B. and Moodie, E. E. M. (2013). *Statistical Methods for Dynamic Treatment Regimes*. New York: Springer.

Chakraborty, B. and Murphy, S. A. (2014). Dynamic treatment regimes. Annual Review of Statistics and Its Application **1,** 447–464.

Fischer, K., Goetghebeur, E., Vrijens, B., and White, I. R. (2011). A structural mean model to allow for non-compliance in a randomized trial comparing two active treatments. Biostatistics **12,** 247–57.

Han, P. (2014). Multiply robust estimation in regression analysis with missing data. Journal of the American Statistical Association **109,** 1159–1173.

Hernán, M. A., Brumback, B., and Robins, J. M. (2000). Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. American Journal of Epidemiology **11,** 561–570.

Kang, J. D. Y. and Schafer, J. L. (2007). Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. Statistical Science **22(4),** 523–539.

Kaslow, R. A., Ostrow, D. G., Detels, R., Phair, J. P., Polk, B. F., and Rinaldo, C. R. (1987). The multicenter aids cohort study: rationale, organization, and selected characteristics of the participants. American Journal of Epidemiology **126,** 310–318.

Moodie, E. E. M., Richardson, T. S., and Stephens, D. A. (2007). Demystifying optimal dynamic treatment regimes. Biometrics **63(2),** 447–455.

Murphy, S. (2003). Optimal Dynamic Treatment Regimes (with Discussions). Journal of the Royal Statistical Society, Series B **65,** 331–366.

Neugebauer, R., Silverberg, M. J., and Van der Laan, M. J. (2010). Observational study and individualized antiretroviral therapy initiation rules for reducing cancer incidence in hiv-infected patients. Technical report, U. C. Berkeley Division of Biostatistics Working

Working Paper Series.

Ogburn, E. G., Rotnitzky, A., and Robins, J. M. (2014). Doubly robust estimation of the local average treatment effect curve. Journal of the Royal Statistical Society Series B: Statistical Methodology (In press).

Pan, W. (2001). Akaike's information criterion in generalized estimating equations. Biometrics **57,** 120–125.

Rich, B., Moodie, E. E. M., Stephens, D. A., and Platt, R. W. (2010). Model checking with residuals for G-estimation of optimal dynamic treatment regimes. International Journal of Biostatistics **6(2),** Article 12.

Robins, J. M. (2004). Optimal Structural Nested Models for Optimal Sequential Decisions. In D, L. and P, H., editors, Proceedings of the second Seattle symposium on biostatistics, pages 189–326. New York: Springer.

Robins, J. M. and Rotnitzky, A. (2001). Comment on the Bickel and Kwon article "on double robustness". Statistica Sinica **11,** 920–936.

Schulte, P., Tsiatis, A. A., Laber, E. B., and Davidian, M. (2014). Q- and A-learning methods for estimating optimal dynamic treatment regimes. Statistical Science. In press.

Sutton, R. S. and Andrew, G. B. (1998). *Reinforcement Learning: An Introduction.* MIT press: Cambridge.

Taguri, M., Matsuyama, Y., and Ohashi, Y. (2014). Model selection criterion for causal parameters in structural mean models based on a quasi-likelihood. Biometrics **70,** 724–733.

van der Laan, M. J. and Rubin, D. B. (2006). Targeted maximum likelihood learning. International Journal of Biostatistics **2(1),** 1–40.

Wallace, M. P. and Moodie, E. E. M. (2015). Doubly-Robust Dynamic Treatment Regimen Estimation Via Weighted Least Squares. Biometrics doi:10.1111/biom.12306.

Watkins, C. J. C. H. (1989). *Learning from Delayed Rewards*. [PhD dissertation]. Cambridge University Press: Cambridge, England.

Zhang, B., Tsiatis, A. A., Laber, E. B., and Davidian, M. (2012). A robust method for estimating optimal treatment regimes. Biometrics **68,** 1010–1018.

Zhao, Y. and Laber, E. B. (2014). Estimation of Optimal Dynamic Treatment Regimes. Clinical Trials **11,** 400–407.

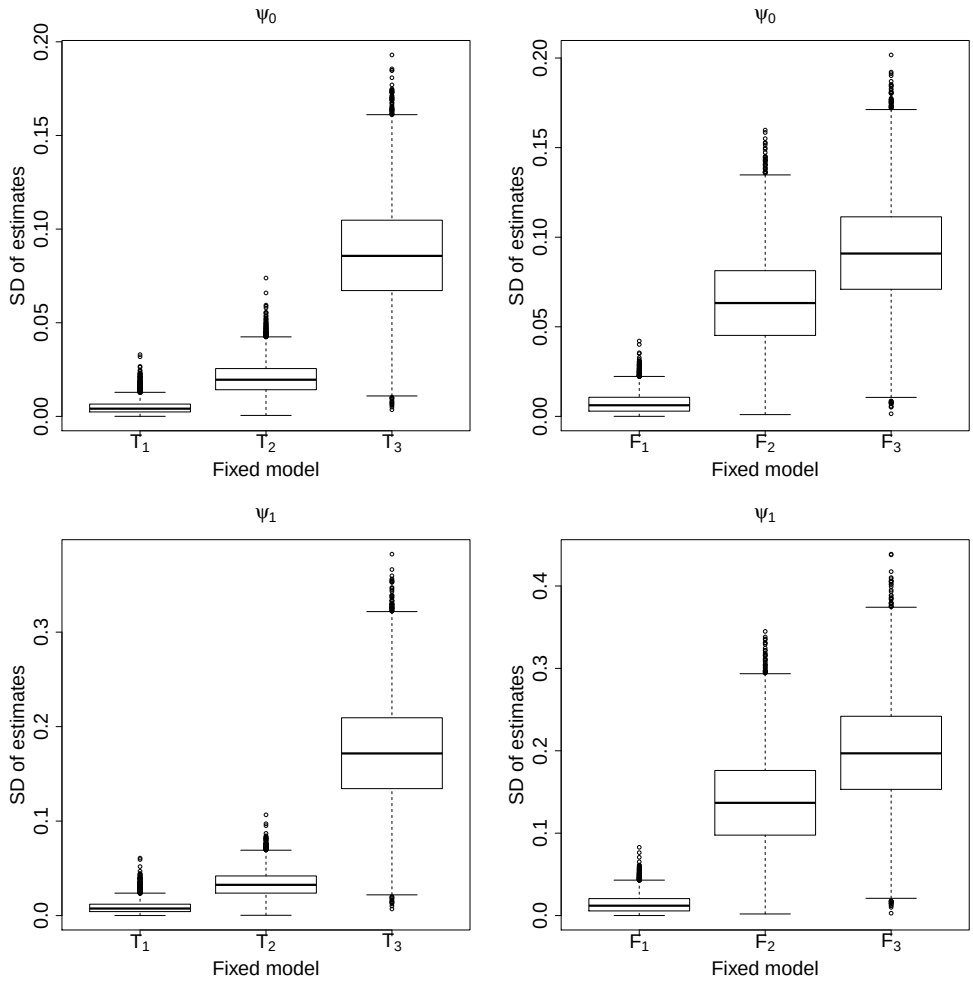

Figure 1: Standard deviations of sets of parameter estimates for each of $10,000$ simulated datasets resulting from applying G-estimation with one model fixed and the other varying. For example the leftmost box in the top-left figure summarizes the distribution of standard deviations of the three parameter estimates of ψ_0 obtained by applying G-estimation with treatment model 1 and treatment-free models 1, 2 and 3.

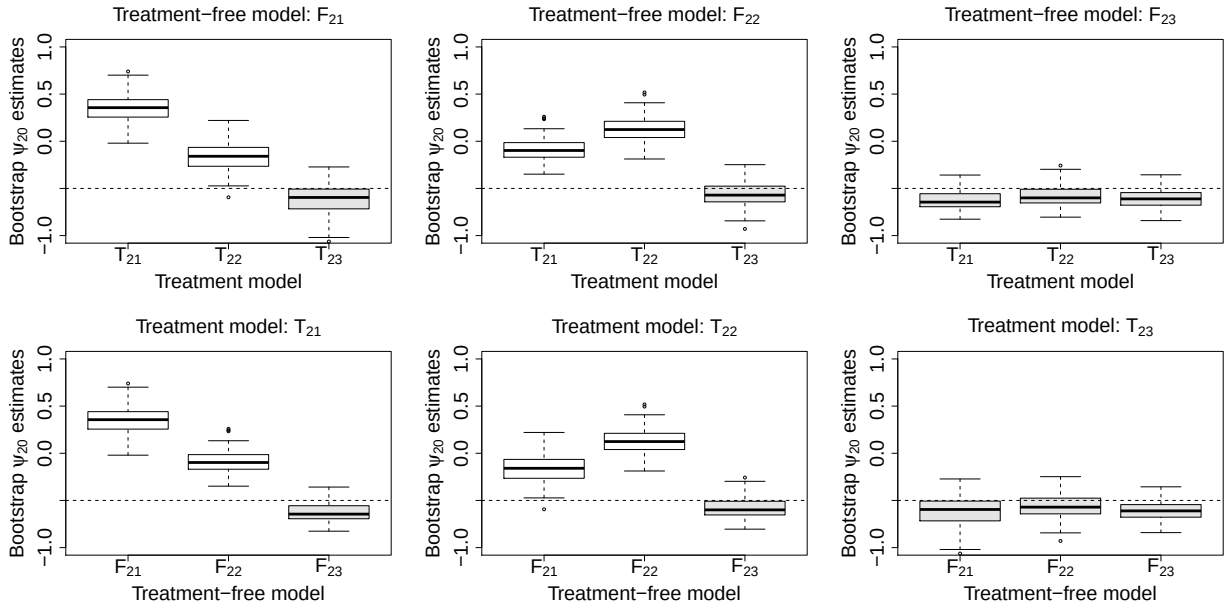

Figure 2: Estimates of ψ_{20} from 100 bootstrap samples of a single dataset generated as per the procedure in Section 3.2. Each plot corresponds to three sets of analyses combining one fixed treatment (treatment-free) model with each of the candidate treatment-free (treatment) models. Dashed line indicates true $\psi_{20} = -0.5$, grayed boxes indicate analyses where at least one model is correctly specified.

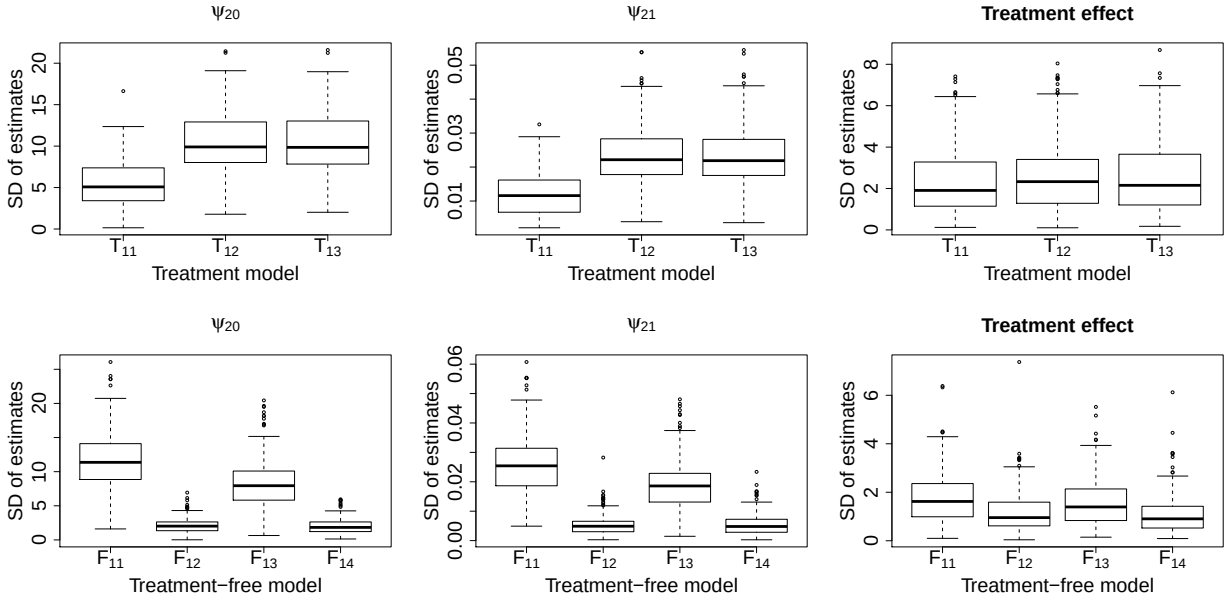

Figure 3: Standard deviations of sets of bootstrapped parameter and mean treatment effect estimates via G-estimation of second stage blip parameter estimates of MACS data where one model is fixed and the other varies. The top row corresponds to the variation in estimates when each corresponding treatment model was held fixed (while the treatment-free model varied), while the bottom corresponds to variation when each treatment-free model was held fixed.

Table 1: Results of G-estimation analysis of a single simulated dataset with different combinations of treatment (T) and treatment-free (F) models.

	Treatment-free							
	ψ_0				ψ_1			
Treatment	F_1	F_2	F_3	*SD*	F_1	F_2	F_3	*SD*
T_1	1.030	1.034	1.035	*0.0026*	1.008	1.001	0.999	*0.0046*
T_2	1.049	1.054	1.034	*0.0130*	0.972	0.930	0.953	*0.0207*
T_3	1.036	0.920	0.899	*0.0738*	0.996	1.226	1.266	*0.1459*
SD	*0.0100*	*0.0726*	*0.0781*		*0.0185*	*0.1544*	*0.1689*	

Table 2: Analysis of results of G-estimation on MACS data. Columns indicate how many times each model was selected during 200 bootstrap samples based on minimal standard deviation of the corresponding estimate (blip parameters or T.E.$_j$ = stage j treatment effect)

	ψ_{j0}	ψ_{j1}	T.E.$_j$
T_{11}	199	199	61
T_{12}	0	0	91
T_{13}	1	1	48
F_{11}	22	18	31
T_{12}	178	182	169
T_{21}	187	185	88
T_{22}	2	1	62
T_{23}	11	14	50
F_{21}	1	1	7
F_{22}	79	99	89
F_{23}	1	1	32
F_{24}	119	99	72

Table 3: Results of G-estimation analysis of MACS data with different combinations of treatment (T_{2i}) and treatment-free (F_{2i}) models. Bold indicates estimates from recommended models.

Stage 1	Treatment-free			
	ψ_{10}		ψ_{11}	
Treatment	F_{11}	F_{12}	F_{11}	F_{12}
T_{11}	-1.233	**0.751**	-0.029	**-0.030**
T_{12}	-16.152	-1.165	0.003	-0.024
T_{13}	-16.167	-0.860	0.005	-0.024

Stage 2	Treatment-free							
	ψ_{20}				ψ_{21}			
Treatment	F_{21}	F_{22}	F_{23}	F_{24}	F_{21}	F_{22}	F_{23}	F_{24}
T_{21}	27.332	27.803	18.660	**20.564**	-0.095	-0.094	-0.079	**-0.080**
T_{22}	8.800	26.286	5.515	18.145	-0.054	-0.092	-0.048	-0.074
T_{23}	7.108	24.868	4.988	17.922	-0.050	-0.090	-0.049	-0.075

Web-based Supplementary Materials for "Model assessment in dynamic treatment regimen estimation via double robustness" by Michael P. Wallace, Erica E. M. Moodie and David A. Stephens.

1 G-estimation: theoretical details

We first present the step-by-step approach to implementation of the G-estimation method, where we introduce a subscript j notation to indicate the j^{th} (of K) treatment stage, and h_j denotes all information available immediately prior to the j^{th} treatment decision.

1. Propose a model for the blip function $\gamma_j(h_j, a_j; \psi_j)$, and define a vector-valued function $S_j(a_j)$ (with dimension equal to that of ψ_j) chosen to contain variables thought to interact with treatment to effect a difference in expected outcome. The optimal form of $S_j(a_j)$, though known (Robins, 2004), requires additional assumptions and so typically $\frac{\partial}{\partial \psi}\gamma_j$ is used.

2. Define the function $G_j(\psi) = y - \gamma_j(h_j, a_j; \psi_j) + \sum_{k=j+1}^{K} \gamma(h_k, a_k^{opt}; \psi_k) - \gamma(h_k, a_k; \psi_k)$, where a_k^{opt} is the optimal stage k treatment decision.

3. Propose a model for $\mathbb{E}[G_j(\psi)|H_j = h_j; \beta_j]$ and use the data to estimate its parameters in terms of ψ_j. This is our *treatment-free outcome model*. For example $f_j(h_j; \beta_j)$ may be a posited linear model for $\mathbb{E}[G_j(\psi)|H_j = h_j; \beta_j]$ from which ordinary least squares estimates $\widehat{\beta}_j(\psi_j)$ may be determined up to a parameter.

4. Propose a *treatment model* $\mathbb{E}[A_j|H_j = h_j; \alpha]$; use the data to obtain estimates $\widehat{\alpha}_j$.

5. Construct the function

   $$U_j(\psi_j, \beta_j, \alpha_j) = (G_j(\psi_j) - \mathbb{E}[G_j(\psi_j)|H_j = h_j; \beta_j])(S_j(a_j) - \mathbb{E}[S_j(a_j)|H_j = h_j; \alpha_j])$$

   and by using parameter estimates from the previous steps estimate the blip parameters ψ_j by solving the system of estimating equations $0 = \sum_{i=1}^{n} U_j(\psi_j, \widehat{\beta}_j(\psi_j), \widehat{\alpha}_j)$.

1.1 A more friendly presentation

While G-estimation is a powerful tool for DTR identification, its typical presentation can prove intimidating, particularly to those without extensive statistical and computational experience. Although simpler methods such as Q-learning require little more than familiarity with standard regression techniques, G-estimation introduces considerable pre-computation,

requires estimating the β parameters in terms of other unknowns, and familiarity with estimating equation techniques. We take a moment aside from the primary focus of this paper to offer a more accessible presentation of G-estimation implementation when the blip function is linear. Our approach extends to the multi-stage setting, but for simplicity we introduce in the single-stage case. In this context we assume our outcome Y can be modeled as

$$\mathbb{E}[Y|X;\beta,\psi] = X^\beta\beta + AX^\psi\psi$$

where the treatment-free model $f(X^\beta\beta)$ is now assumed linear in β. Here, our blip is $AX^\psi\psi$, and so for G-estimation as above we define $G = Y - AX^\psi\psi$.

If treatment A depends on $X^\alpha \in X$ then we can write the final estimating equation step of G-estimation as

$$0 = \sum_{i=1}^{n} X_i^\psi(A_i - \mathbb{E}[A_i|X_i^\alpha;\hat{\alpha}])(Y_i - A_iX_i^\psi\psi - X_i^\beta\hat{\beta}(\psi))$$

where estimates $\hat{\alpha}$ are obtained from the treatment model (such as via logistic regression) and estimates $\hat{\beta}(\psi)$ are typically obtained via ordinary least squares. That is, we obtain $\hat{\beta}(\psi)$ by solving

$$0 = \sum_{i=1}^{n} X_i^\beta(Y_i - A_iX_i^\psi\psi - X_i^\beta\beta)$$

for β in terms of ψ. We can, however, make an aesthetic tweak by instead writing these two steps as a single set of equations

$$0 = \sum_{i=1}^{n} \begin{pmatrix} X_i^\beta \\ X_i^\psi(A_i - \mathbb{E}[A_i|X_i^\alpha;\hat{\alpha}]) \end{pmatrix} (Y_i - A_iX_i^\psi\psi - X_i^\beta\beta) \tag{1}$$

and if we convert this to matrix form by defining $X^\delta = (X^\beta, AX^\psi)$, $X^w = (X^\beta, X^\psi(A - \mathbb{E}[A|X^\alpha;\hat{\alpha}])$ and $\delta = (\beta,\psi)$ then we can write (1) as

$$0 = (X^w)^T(Y - X^\delta\delta)$$

from which estimates of δ can be obtained via

$$\hat{\delta} = ((X^w)^TX^\delta)^{-1}(X^w)^TY.$$

The required $\hat{\psi}$ terms can then be simply 'read off' from the vector $\hat{\delta}$ and used for regimen estimation in the usual way.

This approach therefore allows G-estimation to be carried out by standard matrix equations rather than via the more complex implications of its standard presentation. As such, the need for specialized software is greatly reduced, with G-estimation possible with only a few lines of programming code and little specialized knowledge. Furthermore, this approach can be extended to the continuous treatment (as per Rich et al. 2014, for example) and multi-stage environments in a straightforward manner.

1.2 Variance for recursive G-estimation

Moodie (2009) showed the variance of the optimal decision rule parameters $\hat{\psi}$ adjusting for the plug-in estimates of nuisance parameters in the estimating function $U_j\{\psi_j,\hat{\beta}_j,\hat{\alpha}_j\}$ is given by

$$\Sigma^\psi = E\left[\left\{\left(E\left[\frac{\partial}{\partial\psi}U_{\text{adj}}\{\psi,\beta^*,\alpha^*\}\right]\right)^{-1}U_{\text{adj}}\{\psi,\beta^*,\alpha^*\}\right\}^{\otimes 2}\right],$$

where α^* and β^* are the limiting values of the nuisance parameters for the treatment model and treatment-free model, respectively, and

$$\begin{aligned} U_{\text{adj}}(\psi) &= U\{\psi, \beta^*, \alpha^*\} + E\left[\frac{\partial}{\partial\beta}U\{\psi, \beta^*, \alpha^*\}\right]\left(\hat{\beta} - \beta^*\right) \\ &\quad + E\left[\frac{\partial}{\partial\alpha}U\{\psi, \beta^*, \alpha^*\}\right](\hat{\alpha} - \alpha^*) + o_p(1). \end{aligned}$$

Note that this variance calculation does not assume correct specification of either of the two nuisance models, but Robins (2004) has shown that the G-estimator is semiparametric efficient when both are correctly specified, and so this variance must attain its minimum under correct specification of the nuisance models.

Let $\alpha^\sharp$, $\beta^\sharp$ denote the ("true") limiting values of the nuisance models *when they are correctly specified*. We can then re-write $U_{\text{adj}}(\psi)$ as follows (ignoring the $o_p(1)$ term):

$$\begin{aligned} U_{\text{adj}}(\psi) &= U\{\psi, \beta^*, \alpha^*\} + E\left[\frac{\partial}{\partial\beta}U\{\psi, \beta^*, \alpha^*\}\right]\left(\hat{\beta} - \beta^*\right) \\ &\quad + E\left[\frac{\partial}{\partial\alpha}U\{\psi, \beta^*, \alpha^*\}\right](\hat{\alpha} - \alpha^*) \\ &= U\{\psi, \beta^*, \alpha^*\} + E\left[\frac{\partial}{\partial\beta}U\{\psi, \beta^*, \alpha^*\}\right]\left(\hat{\beta} \pm \beta^\sharp - \beta^*\right) \\ &\quad + E\left[\frac{\partial}{\partial\alpha}U\{\psi, \beta^*, \alpha^*\}\right](\hat{\alpha} \pm \alpha^\sharp - \alpha^*) \\ &= U\{\psi, \beta^*, \alpha^*\} + E\left[\frac{\partial}{\partial\beta}U\{\psi, \beta^*, \alpha^*\}\right](\hat{\beta} - \beta^\sharp) \\ &\quad + E\left[\frac{\partial}{\partial\alpha}U\{\psi, \beta^*, \alpha^*\}\right](\hat{\alpha} - \alpha^\sharp) \\ &\quad + E\left[\frac{\partial}{\partial\beta}U\{\psi, \beta^*, \alpha^*\}\right](\beta^\sharp - \beta^*) \\ &\quad + E\left[\frac{\partial}{\partial\alpha}U\{\psi, \beta^*, \alpha^*\}\right](\alpha^\sharp - \alpha^*). \end{aligned}$$

It can then be seen that the variance inflation due to incorrect specification of the treatment model is a function of $E\left[\frac{\partial}{\partial\alpha}U\{\psi, \beta^*, \alpha^*\}\right](\alpha^\sharp - \alpha^*)$ and similarly, variance inflation due to incorrect specification of the treatment-free model is a function of $E\left[\frac{\partial}{\partial\beta}U\{\psi, \beta^*, \alpha^*\}\right](\beta^\sharp - \beta^*)$. Of course for consistency of the DTR parameters, at least one model must be correctly specified so that either $\alpha^\sharp = \alpha^*$ or $\beta^\sharp = \beta^*$ and the corresponding variance-increasing term is eliminated from $U_{\text{adj}}(\psi)$.

2 Supplementary figure and tables

We also include some additional results mentioned in the associated paper. In Figure 1 we see that when the correct treatment model is specified the blip parameter estimates obtained by using two different treatment-free models are visibly more correlated than when the treatment model is mis-specified. In Tables 1 and 2, meanwhile, we present some analogous results to those discussed in Section 3.1 of the paper for some smaller sample sizes.

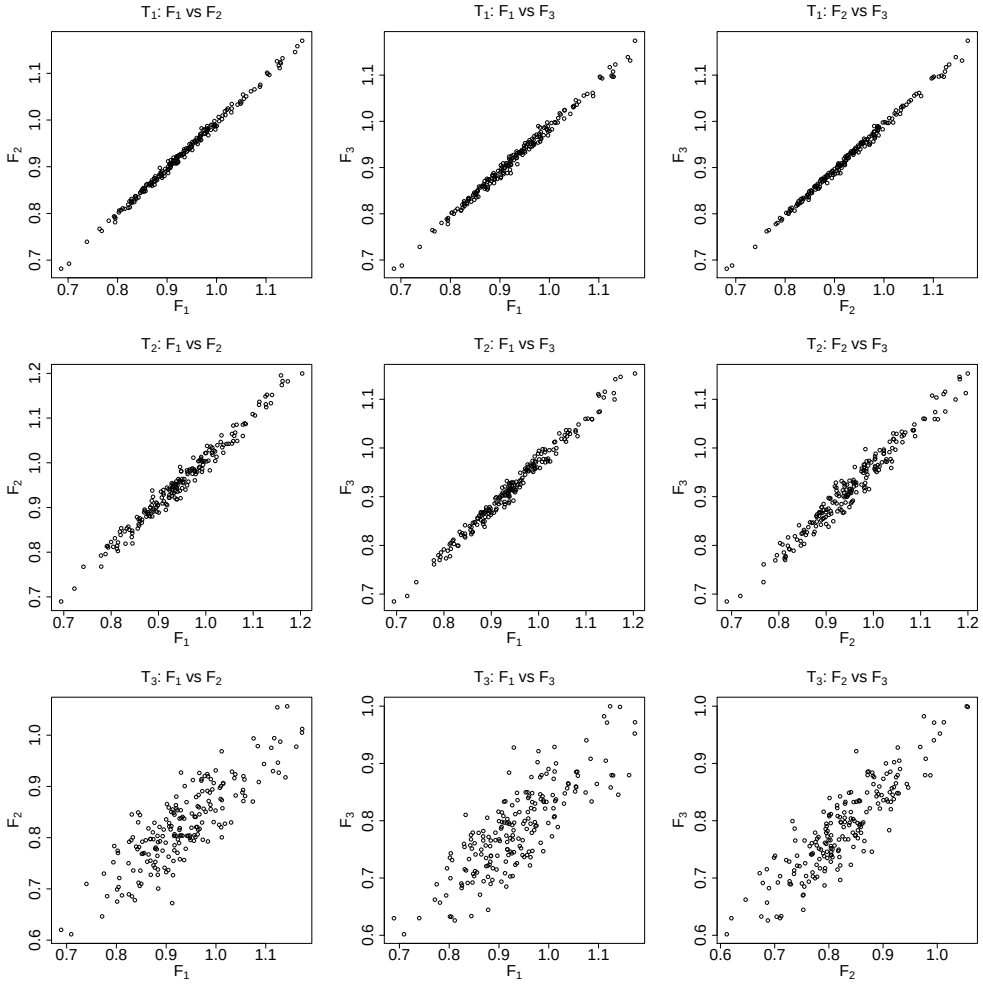

Figure 1: Contrasting bootstrapped G-estimation parameter estimates of ψ_0 for each pair of treatment-free models (F) while keeping each treatment model fixed (T) as discussed in Section 3.1 of the associated paper.

Table 1: Simulation results as per the setup in Section 3.1 with varying sample sizes. Numbers indicate the proportion of simulation runs where the correct model was selected based on which resulted in the smallest standard deviation in resulting estimates of the treatment effect (TE) and blip parameters (ψ_0 and ψ_1) when the given model was held fixed. First six rows correspond to results from 10,000 simulated datasets, bottom two rows (TE (BS)) correspond to results from analysis of 200 bootstrap samples across 1,000 simulated datasets.

		Sample size		
Method	Model	100	500	1,000
TE	Treatment	0.560	0.590	0.588
TE	Treatment-free	0.874	0.972	0.995
ψ_0	Treatment	0.821	0.943	0.989
ψ_0	Treatment-free	0.835	0.940	0.983
ψ_1	Treatment	0.810	0.954	0.990
ψ_1	Treatment-free	0.856	0.946	0.986
TE (BS)	Treatment	0.752	0.755	0.748
TE (BS)	Treatment-free	0.977	0.999	0.999

Table 2: Simulation results as per the setup in Section 3.1 with varying sample sizes. Numbers indicate the proportion of 1,000 simulation runs where a one-way analysis of variance on bootstrapped treatment effect estimates (while keeping the corresponding model fixed) smaller than 0.05.

	Sample size		
Model	100	500	1,000
T_1	0.147	0.055	0.045
T_2	0.487	0.877	0.981
T_3	0.220	0.128	0.104
F_1	0.006	0.000	0.000
F_2	0.264	0.554	0.817
F_3	0.568	0.889	0.991

3 Simulated example: degrees of model mis-specification

An interesting issue in the context of model selection is whether some candidate models can be viewed as 'more mis-specified' than others. Some basic intuition would, of course, suggest so. If we suppose a true treatment model $\mathbb{P}(A = 1|X) = X$, for example, then it would seem reasonable to consider the model $\mathbb{P}(A = 1|X) = X^{1.1}$ as being 'more correct' (or closer to the truth) than, say, the model $\mathbb{P}(A = 1|X) = X^2$. In general, however, things are far less clear-cut. We take this opportunity to present some related observations in the context of doubly-robust model assessment.

We simulate datasets (size $n = 1,000$) as follows:

- $X \sim U(0, 1)$;
- $A \in \{0, 1\}$ with $\mathbb{P}(A = 1|X) = x$;
- $Y \sim N(1 + x + x^3 + a(\psi_0 + \psi_1 x), 0.5)$ with $\psi_0 = \psi_1 = 1$;

and conduct G-estimation analyses using five different treatment and treatment-free models:

- Treatment models: $\mathbb{P}(A = 1|X = x) = x^t$ for $t = 0, 1, 2, 3, 4$; and
- Treatment-free models: $f(x; \boldsymbol{\beta}) = \beta_0 + \beta_1 x + 1_{f\neq 1}\beta_2 x^f$ for $f = 1, 2, 3, 4, 5$,

where $1_{f\neq 1}$ is the indicator function taking value 0 when $f = 1$ and 1 otherwise. As such, the treatment model is correct when $t = 1$ and the treatment-free model is correct when $f = 3$. Intuitively we might expect the treatment models to perform worse as t moves further from 1, and for the treatment-free models to perform worse as f moves further from 3. The analyses were carried out across $10,000$ simulated datasets, and within-model standard deviations of parameter estimates are summarized (in analogous fashion to Figure 1 in the main article) in Figure 2.

For both parameters, and for both the treatment and treatment-free models, we see our intuition is reflected in the results. When the treatment model is correctly specified the parameter estimates resulting from the five different treatment-free model analyses exhibit the lowest standard deviation, and vice-versa for the correct treatment-free model. We note, however, that while the treatment model results give a fairly clear indication of the 'best' model, the treatment-free model results exhibit considerable overlap. This is reflected in how often the correct model is identified if we look, for each simulation, at which models return the lowest standard deviation of estimates. The correct treatment model is estimated for both stages over 97% of the time, while the correct treatment-free model is identified for both stages in around 40% of simulations. This latter result, however, should be set in the context of the other possibilities: the next 'most likely' treatment-free model ($f(X) = \beta_0 + \beta_1 X + \beta_2 X^2$) was indicated only 27% of the time and so the correct choice is still clear.

How, then, do these observations translate to the results of the analyses themselves? While in practice (and thus the primary purpose of this paper) we do not know what the true parameter values are, we can in this example assess model performance. In Table 3 we summarize the mean parameter estimates for each pair of treatment and treatment-free model analyses. While the least biased estimates arise when at least one of the models is correctly specified, there is some association between how 'wrong' our models are (both intuitively and in terms of the standard deviation results presented in Figure 2) and the bias in the resulting estimates.

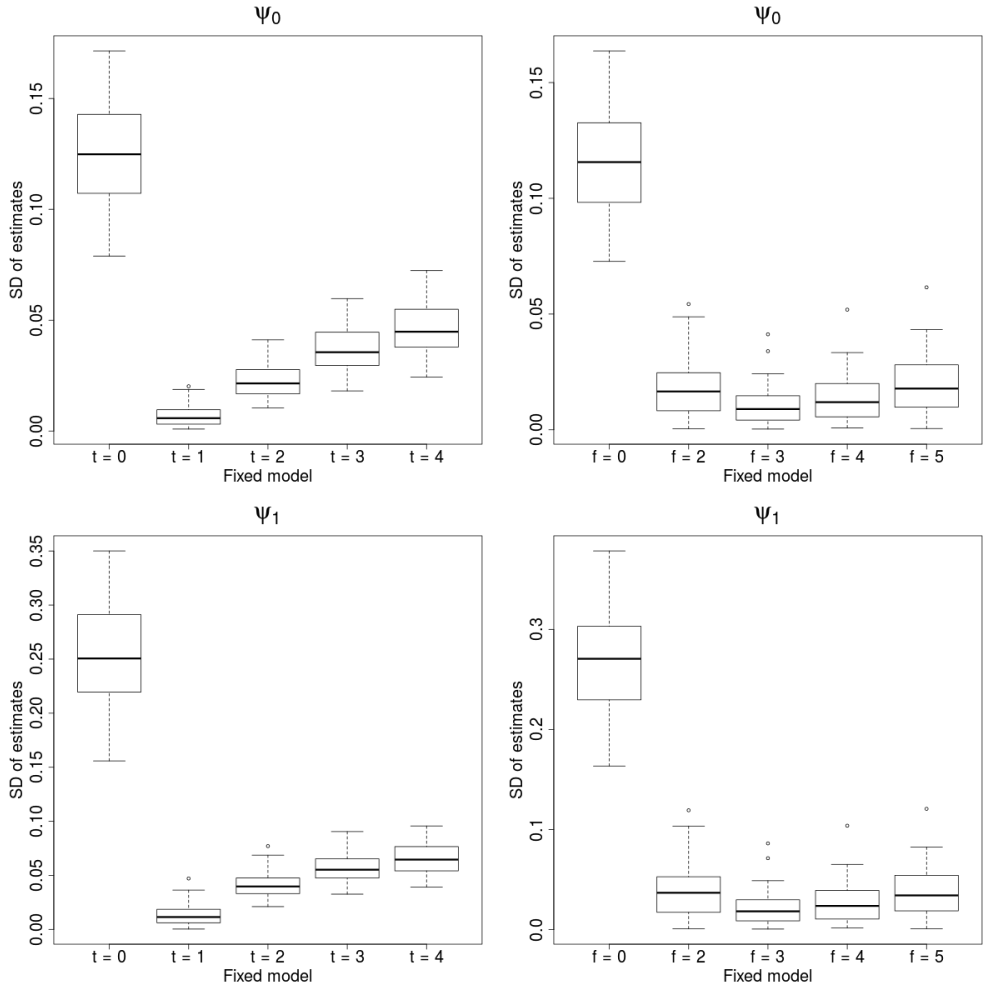

Figure 2: Standard deviations of sets of bootstrapped parameter estimates via G-estimation for each of $10,000$ simulated datasets where one model is fixed and the other varies. Treatment models: $\mathbb{P}(A = 1|X = x) = x^t$. Treatment-free models: $f(X;\boldsymbol{\beta}) = \beta_0 + \beta_1 X + 1_{f\neq 1}\beta_2 X^f$.

Table 3: Mean blip parameter estimates from G-estimation analysis of 10,000 simulated datasets with different combinations of treatment ($\mathbb{P}(A = 1|X = x) = x^t$) and treatment-free ($f(X) = \beta_0 + \beta_1 X + 1_{f\neq 1}\beta_2 X^f$) models. Bold rows/columns indicate correct models.

	Treatment-free									
			ψ_0					ψ_1		
Treatment	$f=1$	$f=2$	$\boldsymbol{f=3}$	$f=4$	$f=5$	$f=1$	$f=2$	$\boldsymbol{f=3}$	$f=4$	$f=5$
$t=0$	0.695	0.997	**0.997**	0.980	0.959	1.607	1.003	**1.003**	1.004	1.004
$\mathbf{t=1}$	**0.996**	**0.997**	**0.997**	**0.997**	**0.997**	**1.005**	**1.003**	**1.003**	**1.004**	**1.004**
$t=2$	0.967	1.019	**0.998**	0.981	0.970	0.961	0.959	**1.002**	1.031	1.050
$t=3$	0.932	1.028	**0.997**	0.974	0.956	0.975	0.937	**1.003**	1.048	1.078
$t=4$	0.908	1.031	**0.997**	0.970	0.949	0.999	0.926	**1.003**	1.058	1.097

References

Moodie, E. (2009). A note on the variance of doubly-robust g-estimates. Biometrika **96,** 998–1004.

Rich, B., Moodie, E. E. M., and Stephens, D. A. (2014). Simulating sequential multiple assignment randomized trials to generate optimal personalized warfarin dosing strategies. Clinical Trials (London, England) **11(4),** 435–444.

Robins, J. M. (2004). Optimal Structural Nested Models for Optimal Sequential Decisions. In D, L. and P, H., editors, Proceedings of the second Seattle symposium on biostatistics, pages 189–326. New York: Springer.