Web Appendix for Effect Estimation using Structural Nested Models and G-estimation

Anonymized authors.

Introductory concepts and notation

First, we provide some additional details on the general data framework within which G-estimation may be applied. An important concept is that we work in *stages*: treatment decisions are taken at fixed points in time, with such decisions potentially depending on a patient's current covariates as well as past information including treatment history. While the theory can be applied to a situation where a single treatment decision is made (the *single-stage* or *cross-sectional* setting) we are primarily interested in the general, multi-stage case.

We use the following notation:

- y: Patient outcome;
- a_j: The stage j treatment decision (with $j = 1, ..., J$), with a_j^0 denoting "no treatment" (such as a control or standard care);
- $\boldsymbol{x}_j$: Non-treatment information (e.g. age, disease severity, response to previous treatments) measured immediately prior to the j^{th} treatment decision;

and at each stage, specify models of the relationship between patient history up until that point and three variables:

- *Treatment model*: the observed treatment a_j.
- *Treatment-free model*: the expected outcome assuming no treatment is received at or after stage j.
- *Blip model*: the expected change in outcome given a_j is received and no treatment is received after stage j. (This is the structural nested mean model referred to in the accompanying letter.)

The stage j treatment, treatment-free and blip models are parameterized by α_j, β_j and ψ_j, respectively, with model covariates denoted by $\boldsymbol{h}_j^\alpha$, $\boldsymbol{h}_j^\beta$ and $\boldsymbol{h}_j^\psi$.

The blip model characterizes our primary interest: the effect of treatment at any given stage on the outcome y, including any interactions between treatment and covariates. A useful step is to view the relationship between our outcome and covariates/treatments as

$$y = G_1(\boldsymbol{h}_1^\beta; \beta_1) + \sum_{j=1}^{J} \gamma_j(a_j, \boldsymbol{h}_j^\psi; \psi_j).$$

Viewed this way, we can see that the stage 1 treatment-free model is simply $G_1(\boldsymbol{h}_1^\beta;\beta_1)$ (setting all treatments to 0), the stage 2 treatment-free model is $G_1(\boldsymbol{h}_1^\beta;\beta_1)+\gamma_1(a_1,\boldsymbol{h}_1^\psi;\psi_1) = G_2(\boldsymbol{h}_2^\beta;\beta_2)$ (setting all treatments from stage 2 onwards to 0 and reparameterizing), and so on. The effect of treatment at each stage, meanwhile, is characterized by the blip $\gamma_j(a_j\boldsymbol{h}_j^\psi;\psi_j)$. Of particular use is that treatment effects may be estimated solely through estimation of the blip parameters ψ_j: we do not need to correctly estimate the parameters of our treatment-free models β_j. In practice, we assume linear models for the treatment-free and blip models: $G_j(\boldsymbol{h}_j^\beta;\beta_j) = \boldsymbol{h}_1^\beta\beta_1$ and $\gamma_j(a_j,\boldsymbol{h}_j^\psi;\psi_j) = a_j\boldsymbol{h}_j^\psi\psi_j$ (but note that, as with any standard regression setup, functions of covariates such as logarithms or squares can be used for greater flexibility).

Finally, an essential component of G-estimation are what we term the stage-specific *pseudo-outcomes*, defined in terms of our blip parameters ψ as $\widetilde{y}_j = y - \sum_{k=j+1}^{J}\gamma_k(\boldsymbol{h}_k,a_k;\psi_k)$. Defined in this way, $\widetilde{y}_j$ may be thought of as the expected outcome assuming a patient receives no treatment *subsequent* to stage j - this is slightly different to the stage j treatment-free outcome, which is expected outcome assuming no treatment from stage j onwards. We may also write this as $\widetilde{y}_j = \widetilde{y}_{j+1} - \gamma_{j+1}(\boldsymbol{h}_j,a_j;\widehat{\psi}_j)$: the stage j pseudo-outcome being equal to the stage $j+1$ pseudo-outcome minus the effect of observed treatment at stage $j+1$.

We note, however, that with the exception of $\widetilde{y}_J$, which is equal to the observed outcome y, each of these pseudo-outcomes depend on the blip parameters at *future* stages. As such, G-estimation proceeds *recursively*: the final stage blip parameters ψ_J are estimated first, before working backwards until every stage has been analyzed. By working recursively we are able to estimate each pseudo-outcome by 'plugging in' all future blip parameters.

G-estimation

While typically presented in more complex theoretical terms, G-estimation may in fact be conducted as a relatively straightforward series of matrix equations. In addition, effectively the same procedure is conducted at each stage, and so for convenience we can suppress our stage-specific notation and write $\boldsymbol{h}_\alpha$, $\boldsymbol{h}_\beta$ and $\boldsymbol{h}_\psi$ instead of $\boldsymbol{h}_j^\alpha$, $\boldsymbol{h}_j^\beta$ and $\boldsymbol{h}_j^\psi$, respectively.

At each stage G-estimation consists of three steps:

1. Specify each of the treatment, treatment-free, and blip models, and form the pseudo-outcome $\tilde{y}$ for that stage.

2. Estimate the parameters of the treatment model, such as by logistic or linear regression of treatment on covariates $\boldsymbol{h}_\alpha$. Use the results to estimate $\hat{a} = E[A|\boldsymbol{h}_\alpha;\hat{\alpha}]$ and construct $\boldsymbol{d}$: the diagonal $n\times n$ matrix with $(i,i)^{th}$ entry equal to $a-\hat{a}$ for the i^{th} subject.

3. Construct the matrix $\boldsymbol{w} = \boldsymbol{d}^\top - \boldsymbol{d}^\top\boldsymbol{h}_\beta\left(\boldsymbol{h}_\beta^\top\boldsymbol{h}_\beta\right)^{-1}\boldsymbol{h}_\beta^\top$ and estimate ψ by solving

$$\hat{\psi} = \left[\boldsymbol{h}_\psi^T\boldsymbol{w}a\boldsymbol{h}_\psi\right]^{-1}\boldsymbol{h}_\psi^\top\boldsymbol{w}\tilde{y}. \tag{1}$$

While the formulation of the matrix $\boldsymbol{w}$ appears slightly complex, the resulting form of the estimator (1) for $\hat{\psi}$ is fairly simple (and, in fact, is almost identical to a standard weighted ordinary least squares-type estimator). As such, analysis at each stage may proceed by forming the pseudo-outcome $\widetilde{y}$, constructing the above matrices, then directly calculating the blip parameter estimates by matrix calculations; implementable in most software environments.

Inference

A further issue is that of inference about our blip parameters ψ. This is complicated by the process estimating nuisance parameters β and α, although theory for estimating a 'corrected' standard error that takes this into account has been developed (1, 2). However, in our experience a standard robust (or sandwich) variance estimator that ignores the nuisance parameter estimation typically performs as well as the bootstrap or the nuisance parameter corrected standard error (3).

While our derivation above is more simplistic, it is grounded in the fundamental idea that we obtain estimates of ψ by considering the estimating equations

$$U(\psi; \hat{\beta}; \hat{\alpha}) = \boldsymbol{h}_{\psi}^{\top} \boldsymbol{w}(\tilde{y} - a\boldsymbol{h}_{\psi}\psi)$$

and finding ψ that solve $\mathbb{E}_n[U(\psi; \hat{\beta}, \hat{\alpha})] = 0$, where $\mathbb{E}_n$ denotes the empirical average function. Standard estimating equation theory leads to the sandwich estimator via

$$\sqrt{n}(\hat{\theta} - \theta) \underset{d}{\to} N_p(0, \boldsymbol{B}^{-1}\boldsymbol{F}(\boldsymbol{B}^{\top})^{-1}) \tag{2}$$

where $\boldsymbol{B} = -\mathbb{E}\left[\frac{\partial U}{\partial \psi}\right] = \mathbb{E}[\boldsymbol{h}_{\psi}^{T}\boldsymbol{w}a\boldsymbol{h}_{\psi}]$ (the 'bread') and $\boldsymbol{F} = Var[U] = \mathbb{E}[UU^{\top}]$ (the filling), both of which can be estimated from the matrices we use to construct U.

By comparison with (1, 2) we note that the formulation of G-estimation presented here means that the above sandwich estimator already accounts for the estimation of the nuisance parameters β (but not α). Correcting for the estimation of α relies on a number of Taylor expansions that can make implementation challenging but, in our experience, make little difference to inference. The results of the Honolulu Heart Program analysis presented in the letter contain three different approaches to inference: the 'standard' sandwich above, the 'corrected' (or adjusted) sandwich (available in our R routine), and the bootstrap.

R code for simulated datasets

Along with our R package (discussed below), we also present how G-estimation may be conducted via matrix equations using R. The associated R code is available as a file to download as Web Material. We present the code here with comprehensive comments for reference. We note that the code presented could be far more concise, but it is deliberately thorough to ease comprehension.

We begin by generating a dataset as per our setup. First, our stage 1 data are generated as $X_1 \sim N(0, 1)$ and $\mathbb{P}(A_1 = 1|X_1 = x_1) = \text{expit}(x_1)$:

```
# Sample size
n <- 1000
# Stage 1 data:
X1 <- rnorm(n)
A1 <- rbinom(n,1,1/(1+exp(-X1)))
```

Next, our stage 2 data are generated as $X_2 \sim N(a_1, 1)$ and $A_2 \sim N(x_2, 1)$:

```
# Stage 2 data:
X2 <- A1 + rnorm(n)
A2 <- X2 + rnorm(n)
```

Our stage 3 data are generated as $X_3 \sim N(a_2, 1)$ and $\mathbb{P}(A_3 = 1|X_3 = x_3) = \text{expit}(x_3)$:

```
# Stage 3 data:
X3 <- A2 + rnorm(n)
A3 <- rbinom(n,1,1/(1+exp(-X3)))
```

Finally, our outcome $Y \sim N(1 + x_1 + a_1(1 + x_1) + a_2(1 + x_2) + a_3(1 + x_3), 1)$

```
# Outcome Y = 'baseline' outcome + treatment effects (blips)
Y <- 1 + X1 + A1*(1+X1) + A2*(1+X2) + A3*(1+X3) + rnorm(n)
```

Analysis proceeds on a stage-by-stage basis as follows:

Stage 3 (treatment model correctly specified, treatment-free model mis-specified)

1. (a) Estimate α_3 by logistic regression of a_3 on $\boldsymbol{h}_3^\alpha$, write $\hat{a}_3 = \mathbb{E}[A_3|\boldsymbol{h}_3^\alpha; \hat{\alpha}_3]$.

```
# treatment model correct:
H3.alpha <- cbind(rep(1,n),X3)
# estimate treatment model
A3hat <- fitted(glm(A3~0+H3.alpha,binomial))
```

1. (b) Using the estimates $\hat{a}_3$ from (a), write

   $\tilde{y}_3 = y$;

   $\boldsymbol{d}_3$ for the diagonal matrix with entries $(a_3 - \hat{a}_3)$[1]

   $\boldsymbol{w}_3 = \boldsymbol{d}_3^\top - \boldsymbol{d}_3^\top \boldsymbol{h}_3^\beta \left[(\boldsymbol{h}_3^\beta)^\top \boldsymbol{h}_3^\beta\right]^{-1} (\boldsymbol{h}_3^\beta)^\top$

```
# pseudo-outcome
Y3 <- Y
# treatment-free model incorrect
H3.beta <- cbind(rep(1,n))
# blip model
H3.psi <- cbind(rep(1,n),X3)
# w-matrix for analysis
W3 <- diag(A3-A3hat) - (A3-A3hat)*H3.beta
        %*% solve(t(H3.beta) %*% H3.beta) %*% t(H3.beta)
```

1. (c) Estimate ψ_3 by solving $\hat{\psi}_3 = \left[(\boldsymbol{h}_3^\psi)^T \boldsymbol{w}_3 a_3 \boldsymbol{h}_3^\psi\right]^{-1} (\boldsymbol{h}_3^\psi)^\top \boldsymbol{w}_3 \tilde{y}_3$.

[1]Note that in our R code we do not form the matrix $\boldsymbol{d}_j$ explicitly, but instead include it in our definition of $\boldsymbol{w}_j$ as this gains us some computational efficiency.

```
# estimate parameters
psi3 <- solve(t(H3.psi) %*% W3 %*% (A3*H3.psi))
          %*% t(H3.psi) %*% W3 %*% Y3
```

Stage 2 (treatment model mis-specified, treatment-free model correctly specified)

2. (a) Estimate α_2 by linear regression of a_2 on $\boldsymbol{h}_2^\alpha$, write $\hat{a}_2 = \mathbb{E}[A_2|\boldsymbol{h}_2^\alpha; \hat{\alpha}_2]$.

```
# treatment model incorrect
H2.alpha <- rep(1,n)
# estimate treatment model
A2hat <- fitted(lm(A2~0+H2.alpha))
```

2. (b) Using the estimates $\hat{\psi}_3$ from Step 1 and $\hat{a}_2$ from (a), write

$\tilde{y}_2 = y - \gamma_3(\boldsymbol{h}_3, a_3; \hat{\psi}_3) = y - a_3(\hat{\psi}_{30} + \hat{\psi}_{31} x_3)$;

$\boldsymbol{d}_2$ for the diagonal matrix with entries $(a_2 - \hat{a}_2)$

$\boldsymbol{w}_2 = \boldsymbol{d}_2^\top - \boldsymbol{d}_2^\top \boldsymbol{h}_2^\beta \left[(\boldsymbol{h}_2^\beta)^\top \boldsymbol{h}_2^\beta \right]^{-1} (\boldsymbol{h}_2^\beta)^\top$

```
# pseudo-outcome
Y2 <- Y - A3*(H3.psi %*% psi3)
# treatment-free model correct
H2.beta <- cbind(rep(1,n),X1,A1,A1*X1)
# blip model
H2.psi <- cbind(rep(1,n),X2)
# w-matrix for analysis
W2 <- diag(A2-A2hat) - (A2-A2hat)*H2.beta
       %*% solve(t(H2.beta) %*% H2.beta) %*% t(H2.beta)
```

2. (c) Estimate ψ_2 by solving $\hat{\psi}_2 = \left[(\boldsymbol{h}_2^\psi)^T \boldsymbol{w}_2 a_2 \boldsymbol{h}_2^\psi \right]^{-1} (\boldsymbol{h}_2^\psi)^\top \boldsymbol{w}_2 \tilde{y}_2$.

```
# estimate parameters
psi2 <- solve(t(H2.psi) %*% W2 %*% (A2*H2.psi))
          %*% t(H2.psi) %*% W2 %*% Y2
```

Stage 1 (both models mis-specified)

3. (a) Estimate α_1 by logistic regression of a_1 on $\boldsymbol{h}_1^\alpha$, write $\hat{a}_1 = \mathbb{E}[A_1|\boldsymbol{h}_1^\alpha; \hat{\alpha}_1]$.

```
# treatment model incorrect
H1.alpha <- rep(1,n)
# estimate treatment model
A1hat <- fitted(glm(A1~0+H1.alpha,binomial))
```

3. (b) Using the estimates $\hat{\psi}_2$ from Step 2 and $\hat{a}_1$ from (a), write, write

$\tilde{y}_1 = \tilde{y}_2 - \gamma_2(\boldsymbol{h}_2, a_2; \hat{\psi}_2) = \tilde{y}_2 - a_2(\hat{\psi}_{20} + \hat{\psi}_{21} x_2)$;

$\boldsymbol{d}_1$ for the diagonal matrix with entries $(a_1 - \hat{a}_1)$

$\boldsymbol{w}_1 = \boldsymbol{d}_1^\top - \boldsymbol{d}_1^\top \boldsymbol{h}_1^\beta \left[(\boldsymbol{h}_1^\beta)^\top \boldsymbol{h}_1^\beta \right]^{-1} (\boldsymbol{h}_1^\beta)^\top$

```
# pseudo-outcome
Y1 <- Y2 - A2*(H2.psi %*% psi2)
# treatment-free incorrect
H1.beta <- cbind(rep(1,n))
# blip model
H1.psi <- cbind(rep(1,n),X1)
# weight-like matrix for analysis
W1 <- diag(A1-A1hat) - (A1-A1hat)*H1.beta
        %*% solve(t(H1.beta) %*% H1.beta) %*% t(H1.beta)
```

3. (c) Estimate ψ_1 by solving $\hat{\psi}_1 = \left[(\boldsymbol{h}_1^\psi)^T \boldsymbol{w}_1 a_1 \boldsymbol{h}_1^\psi \right]^{-1} (\boldsymbol{h}_1^\psi)^\top \boldsymbol{w}_1 \tilde{y}_1$.

```
# estimate parameters
psi1 <- solve(t(H1.psi) %*% W1 %*% (A1*H1.psi))
          %*% t(H1.psi) %*% W1 %*% Y1
```

R package `DTRreg`

Our R package `DTRreg` was originally designed to be used within the dynamic treatment regimen environment, but includes the option to apply it within the setting focused on here as well. It accepts inputs for the blip, treatment, and treatment free models as lists of formula objects. Such objects, most familiar within commands such as `lm` for linear regression, take the form of an outcome variable, followed by a tilde, followed by an equation detailing the relationship between the expected outcome and the covariates. For example, in our three-stage example we specify our treatment models by a list of three formulas, one for each stage, where the j^{th} formula specifies the stage j treatment as the outcome and the variables thought to influence it as covariates:

```
# treatment model:
treat.mod <- list(A1~1,A2~1,A3~X3)
```

The first two entries tell the command that `A1` and `A2` are, respectively, the stage 1 and stage 2 treatment variables, and that both treatment models are intercept-only. The third entry indicates that the stage 3 treatment model includes the variable `X3`. As with other regression-like commands, when other covariates are specified the intercept term is assumed implied and does not require explicit inclusion. The command infers whether treatments are binary or continuous based on whether they take on more than two values, and applies logistic or linear regression to estimate these models accordingly.

The blip and treatment-free models are even more straightforward, as in these cases we do not specify an outcome variable. Instead we simply need to include the terms in the corresponding models:

```
# blip model:
blip.mod <- list(~X1,~X2,~X3)
# treatment-free model:
tf.mod <- list(~1,~X1+A1+A1:X1,~1)
```

The most illuminating example is the stage 2 treatment-free model, which through this input is specified to take the form $\beta_{20} + \beta_{21}x_1 + \beta_{22}a_1 + \beta_{23}a_1x_1$. Finally, we put all this together into the command itself, whose basic syntax takes the form

```
mod <- DTRreg(Y,blip.mod,treat.mod,tf.mod,
              var.est="sandwich",type="alt")
```

where the option `var.est = "sandwich"` specifies that covariate variance estimation should be performed by the corrected (or adjusted) sandwich approach detailed above, and `type="alt"` indicates we are not working within the dynamic treatment regimen environment. From this command, `summary(mod)` and `coef(mod)` return summaries and blip parameter estimates in a similar fashion to more familiar commands such as `lm` (for linear regression) and `glm` (for generalized linear models). In addition to the basic specification of outcome, blip, treatment and treatment-free models, the command offers a variety of more complex options, including variance estimation via the bootstrap instead of the adjusted sandwich. This is specified by a combination of `var.est = "boot"` and `B = n` for the number of bootstrap replications. In addition, when treatment is binary and the true treatment model is known (as may be the case, for example, in a randomized trial) then this may be specified by setting `treat.mod.man` as a list of vectors which gives the probability of receiving treatment for each subject at each stage. Finally, `DTRreg` will automatically ignore any subjects with missing data (thereby carrying out a complete-cases analysis), but if the option `missing = "ipcw"` is specified, then inverse probability of censored weights is used with the probability of censoring estimated via logistic regression on the full covariate history up to that point.

The Honolulu Heart Program

We now illustrate how G-estimation may be applied to a real dataset. The Honolulu Heart Program (HHP) is a prospective study of causes of cardiovascular disease among Japanese Americans living in Hawaii (4). Run by the National Heart, Lung, and Blood Institute (NHLBI) it began in 1965 and consists of a series of examinations, the last of which was conducted during 1991-1993. Similar to Talbot et al. (5), we focus on the effect of physical activity level on blood pressure, using data from questionnaires administered at Exam 1 (1965-1968) and Exam 2 (1968-1971). We therefore pursue a 2-stage setup, with stage 1 corresponding to Exam 1, and stage 2 to Exam 2.

Within this dataset, activity level is defined in multiple ways. At both examinations data were recorded on activity level at work and at home separately, coded as either "Mostly sitting", "Moderate", or "Much", while at Exam 1 an 'activity score' was calculated based on how long participants spent at a variety of activity levels during a 24 hour period. Typically, we may wish to dichotomize such variables (perhaps as 'active' and 'inactive'), but in doing so we also lose information; especially so with the activity score which can reasonably be regarded a continuous variable. As G-estimation does not require treatment variables be binary, we are able to combine both binary and continuous variables at each of our treatment stages. As such, we shall dichotomize our treatment variable at stage 2, while using the

continuous activity score at stage 1. At Exam 2 we follow Talbot et al. and define participants as active if they reported their activity level as 'Moderate' or 'Much' at home *or* at work.

Next, we must specify our blip, treatment-free, and treatment models. To allow easier comparison with the results of Talbot et al. our first analysis is restricted to blip models which are solely functions of the treatment, and do not include any interaction terms with non-treatment covariates as in our simulated example. Our treatment models, meanwhile, are inspired by Talbot et al. Activity at Exam 1 is modeled via a linear regression on age and whether the participant was employed at the time. At Exam 2, meanwhile, a logistic regression model was fit with the following covariates: Exam 1 activity level, Exam 1 body mass index and whether the participant was receiving treatment for high blood pressure at Exam 1, and age and whether the participant was employed at Exam 2. Finally, our treatment-free models consist of the covariates contained within the corresponding treatment models. Like Talbot et al., body mass index and hypertension medication use at Exam 1 are not included in either the treatment or treatment-free models at stage 1 as these are considered to be effects of the stage 1 treatment, although SNMs can incorporate such covariates.

The complete dataset contains information on 8,006 participants. In the interests of simplicity we shall conduct a complete-cases analysis, and only include those who have data available on all of the covariates of interest at both stages. (Although we note that, if desired, missing data may be accommodated by standard methods such as multiple imputation or weighting.) This reduced our dataset to 7,374 participants, with summary statistics of our variables of interest presented in Table 1 (where we note that activity level score at stage 1 has been divided by 100 from the original scale).

We conduct our analysis in a similar manner to the first two stages of the simulated example of the previous section. Specifically, the matrices $\boldsymbol{h}_1^\alpha$ and $\boldsymbol{h}_1^\beta$ are the same as each other, containing an intercept term, Exam 1 age and Exam 1 employment, while $\boldsymbol{h}_2^\alpha$ and $\boldsymbol{h}_2^\beta$ contain an intercept term, Exam 1 activity level, Exam 1 body mass index, Exam 1 blood pressure treatment, Exam 2 age and Exam 2 employment. Both blip matrices $\boldsymbol{h}_1^\psi$ and $\boldsymbol{h}_2^\psi$ simply contain an intercept term, as both blip models are of the form $a_j\psi_j$ with ψ_j the effect estimate of stage j activity level. The results are summarized in Table 2, where we estimated 95% confidence intervals using the unadjusted sandwich, the 'adjusted' sandwich (accommodating the estimation of nuisance parameters as per (1, 2)), and the standard bootstrap (with 1,000 resamples).

As we might expect, at both stages there is evidence that activity level is associated with a decrease in both systolic and diastolic blood pressure. These results seem largely consistent with those presented in Talbot et al., although we observe that their setup was somewhat different, using a binary measure of activity level at both stages and assuming the treatment effect was the same at both stages. Furthermore, there is very little difference between our three different inferential approaches.

While the preceding analysis was conducted without any terms interacting with treatment to ease comparison with previous results, we note that one of the key properties of SNMs over MSMs is their capacity to handle such interaction terms. To illustrate how they may be interpreted, we conducted a secondary analysis of the data in a manner almost identical to the above, with the exception that our blip models are now of the form $\gamma_j = a_j(\psi_{j0} + \psi_{j1}\text{Age}_{\text{B}j})$, where $\text{Age}_{\text{B}j}$ is a participant's age at stage j, dichotomized so that it takes the value 1 if

aged 60 years and over, and 0 otherwise.

This analysis returned blip parameter estimates of $(\hat{\psi}_{10}, \hat{\psi}_{10}) = (-1.048, -0.367)$ at stage 1, and $(\hat{\psi}_{20}, \hat{\psi}_{20}) = (-1.194, -3.488)$ at stage 2 when the outcome was taken as systolic blood pressure at Exam 2 (with similar results for diastolic blood pressure) and may be interpreted just like those in more standard regression analyses. For example, at stage 2 we might observe that an active (rather than inactive) lifestyle would be associated with a decrease in blood pressure of -1.194 mmHg among those under 60, and a decrease of $-1.194 - 3.488 = -4.682$ mmHg among those over 60. In other words, this might suggest an interaction between activity and age, with older participants experiencing a greater 'benefit' from exercise. We note, however, that the confidence intervals associated with these estimates were large and covered zero, so these results come with considerable uncertainty (and reinforce our initial assumption of no interaction between treatment and time-varying covariates).

References

[1] Robins JM. Optimal Structural Nested Models for Optimal Sequential Decisions. In *Proceedings of the second Seattle symposium on biostatistics*, (eds.) Lin D and Heagerty P. New York: Springer-Verlag, 2004; (pages 189–326).

[2] Moodie E. A note on the variance of doubly-robust g-estimates. *Biometrika* 2009; **96**: 998–1004.

[3] Saarela O, Stephens DA, Moodie EEM, and Klein MB. On Bayesian estimation of marginal structural models. *Biometrics* 2015; **71**: 279–301.

[4] Kagan A, Harris BR, Winkelstein JW, Johnson KG, Kato H, Syme SL, Rhoads GG, Gay ML, Nichaman MZ, Hamilton HB, and Tillotson J. Epidemiologic studies of coronary heart disease and stroke in Japanese men living in Japan, Hawaii and California: demographic, physical, dietary and biochemical characteristics. *Journal of Chronic Diseases* 1974; **27(7-8)**: 345–364.

[5] Talbot D, Atherton J, Rossi AM, Bacon SL, and Lefebvre G. A cautionary note concerning the use of stabilized weights in marginal structural models. *Statistics in Medicine* 2015; **34**: 812–823.

Figure 1. Step-by-step implementation of G-estimation for our simulated example using matrix equations (underlying theory detailed in eAppendix).

Our step-by-step analysis is as follows:

1. (a) Estimate α_3 by logistic regression of a_3 on $\boldsymbol{h}_3^\alpha$; write $\hat{a}_3 = \mathbb{E}[A_3|\boldsymbol{h}_3^\alpha; \hat{\alpha}_3]$.
   (b) Using the estimates $\hat{a}_3$ from (a), write

   $\tilde{y}_3 = y$;

   $\boldsymbol{d}_3$ for the diagonal matrix with $(i,i)^{th}$ entry $(a_3 - \hat{a}_3)$ for subject i; and

   $$\boldsymbol{w}_3 = \boldsymbol{d}_3^\top \left\{ \boldsymbol{I}_n - \boldsymbol{h}_3^\beta \left[(\boldsymbol{h}_3^\beta)^\top \boldsymbol{h}_3^\beta \right]^{-1} (\boldsymbol{h}_3^\beta)^\top \right\}.$$

   (c) Estimate ψ_3 by solving $\hat{\psi}_3 = \left[(\boldsymbol{h}_3^\psi)^T \boldsymbol{w}_3 a_3 \boldsymbol{h}_3^\psi \right]^{-1} (\boldsymbol{h}_3^\psi)^\top \boldsymbol{w}_3 \tilde{y}_3$.

2. (a) Estimate α_2 by linear regression of a_2 on $\boldsymbol{h}_2^\alpha$; write $\hat{a}_2 = \mathbb{E}[A_2|\boldsymbol{h}_2^\alpha; \hat{\alpha}_2]$.
   (b) Using the estimates $\hat{\psi}_3$ from Step 1 and $\hat{a}_2$ from (a), write

   $\tilde{y}_2 = y - \gamma_3(\boldsymbol{h}_3, a_3; \hat{\psi}_3) = y - a_3(\hat{\psi}_{30} + \hat{\psi}_{31} x_3)$;

   $\boldsymbol{d}_2$ for the diagonal matrix with $(i,i)^{th}$ entry $(a_2 - \hat{a}_2)$ for subject i; and

   $$\boldsymbol{w}_2 = \boldsymbol{d}_2^\top \left\{ \boldsymbol{I}_n - \boldsymbol{h}_2^\beta \left[(\boldsymbol{h}_2^\beta)^\top \boldsymbol{h}_2^\beta \right]^{-1} (\boldsymbol{h}_2^\beta)^\top \right\}.$$

   (c) Estimate ψ_2 by solving $\hat{\psi}_2 = \left[(\boldsymbol{h}_2^\psi)^T \boldsymbol{w}_2 a_2 \boldsymbol{h}_2^\psi \right]^{-1} (\boldsymbol{h}_2^\psi)^\top \boldsymbol{w}_2 \tilde{y}_2$.

3. (a) Estimate α_1 by logistic regression of a_1 on $\boldsymbol{h}_1^\alpha$; write $\hat{a}_1 = \mathbb{E}[A_1|\boldsymbol{h}_1^\alpha; \hat{\alpha}_1]$.
   (b) Using the estimates $\hat{\psi}_2$ from Step 2 and $\hat{a}_1$ from (a), write

   $\tilde{y}_1 = \tilde{y}_2 - \gamma_2(\boldsymbol{h}_2, a_2; \hat{\psi}_2) = \tilde{y}_2 - a_2(\hat{\psi}_{20} + \hat{\psi}_{21} x_2)$;

   $\boldsymbol{d}_1$ for the diagonal matrix with $(i,i)^{th}$ entry $(a_1 - \hat{a}_1)$ for subject i; and

   $$\boldsymbol{w}_1 = \boldsymbol{d}_1^\top \left\{ \boldsymbol{I}_n - \boldsymbol{h}_1^\beta \left[(\boldsymbol{h}_1^\beta)^\top \boldsymbol{h}_1^\beta \right]^{-1} (\boldsymbol{h}_1^\beta)^\top \right\}.$$

   (c) Estimate ψ_1 by solving $\hat{\psi}_1 = \left[(\boldsymbol{h}_1^\psi)^T \boldsymbol{w}_1 a_1 \boldsymbol{h}_1^\psi \right]^{-1} (\boldsymbol{h}_1^\psi)^\top \boldsymbol{w}_1 \tilde{y}_1$.

Table 1. Covariates Summary for Honolulu Heart Program Participants, 1965-1971.

	Variable	Type	Summary
Exam 1	Age (years)	Categorical	<50: 1,741
			50-54: 2,591
			55-59: 1.463
			60-64: 1,185
			$\geq$ 65: 394
	Employed	Binary	Employed: 6,820
	BMI	Continuous	Mean (sd): 23.85 (3.05)
	BP treatment	Binary	Receiving treatment: 726
	Activity level	Continuous	Mean (sd): 327.7 (45.2)
Exam 2	Age (years)	Categorical	48-54: 3,337
			55-59: 1,898
			60-64: 1,285
			65-70: 854
	Employed	Binary	Employed: 6,513
	Activity level	Binary	Active: 6,664
	SBP (mmHg)	Continuous	Mean (sd): 134.1 (20.8)
	DBP (mmHg)	Continuous	Mean (sd): 84.1 (11.3)

Table 2. Effect Estimates of Activity Level on Systolic and Diastolic Blood Pressure in Honolulu Heart Program Participants, 1965-1971.

			95% confidence interval		
Blood pressure	Parameter	Estimate	Sandwich	Adj. Sandwich	Bootstrap
Systolic	ψ_1	-1.105	-2.120,-0.091	-2.120,-0.091	-2.123,-0.088
	ψ_2	-1.959	-3.535,-0.201	-3.531,-0.206	-3.494,-0.243
Diastolic	ψ_1	-1.307	-1.874,-0.736	-1.874,-0.736	-1.848,-0.762
	ψ_2	-0.901	-1.831,0.028	-1.831,0.028	-1.816,0.014