Research Article

Statistics in Medicine

Received XXXX

(www.interscience.wiley.com) DOI: 10.1002/sim.0000

A cautionary note concerning the use of stabilized weights in marginal structural models

Denis Talbot[a,b], Juli Atherton[a], Amanda M. Rossi[c,d], Simon L. Bacon[c,d], Geneviève Lefebvre[a] *

Marginal structural models (MSMs) are commonly used to estimate the causal effect of a time-varying treatment in presence of time-dependent confounding. When fitting a MSM to data, the analyst must specify both the structural model for the outcome and the treatment models for the inverse-probability-of-treatment weights. The use of stabilized weights is recommended since they are generally less variable than the standard weights. In this paper, we are concerned with the use of the common stabilized weights when the structural model is specified to only consider partial treatment history, such as the current or most recent treatments. We present various examples of settings where these stabilized weights yield biased inferences while the standard weights do not. These issues are first investigated on the basis of simulated data and subsequently exemplified using data from the Honolulu Heart Program. Unlike common stabilized weights, we find that basic stabilized weights offer some protection against bias in structural models designed to estimate current or most recent treatment effects. Copyright © 2010 John Wiley & Sons, Ltd.

Keywords: Time-dependent confounding, marginal structural models, inverse-probability-weighting, repeated measures, stabilized weights

1. Introduction

Marginal structural models (MSMs) [1–4] are nowadays a common longitudinal data analytical approach for estimating the effects of time-varying treatments in presence of time-dependent confounding [5–11]. When fitting a MSM to data, an analyst faces two important decisions: 1) the specification of the structural model for the outcome, done in accordance

[a] *Université du Québec à Montréal, Département de mathématiques*
[b] *Université Laval, Département de médecine sociale et préventive*
[c] *Concordia University, Department of Exercise Science*
[d] *Montreal Behavioural Medicine Centre, Hôpital du Sacré-Coeur de Montréal*
**Correspondence to: Université du Québec à Montréal, Département de mathématiques, 201 Président-Kennedy, Montréal, Québec, H2X 3Y7. E-mail: lefebvre.gen@uqam.ca*

Contract/grant sponsor: D. Talbot is supported by a Fonds de recherche du Québec: Nature et Technologies (FRQNT) doctoral scholarship. Dr. Atherton's research is supported by the Natural Sciences and Engineering Research Council of Canada (NSERC). A. M. Rossi is supported by a Canadian Institutes of Health Research (CIHR) Vanier Graduate Scholarship and a Fond de recherche Santé - Québec (FRQS) doctoral scholarship. Dr. S. L. Bacon is supported by the CIHR New Investigator Award and a FRQS Chercheur-Boursier. Dr. Lefebvre's is a FRQS Chercheur-Boursier and is also supported by NSERC.

with the causal contrast of interest; 2) the specification of the treatment models which are used to calculate the inverse-probability-of-treatment received at each time point, i.e. the weights [5]. For the structural model, a single measure is commonly used to summarize treatment history, such as the treatment received at the last time point, a cumulative measure of the treatment or an indicator of "ever started treatment" [5, 12]. The covariates included in the treatment models are typically the baseline covariates and the histories of time-varying covariates and prior treatments. Platt et al. [12] outline strategies for marginal structural model specifications and introduce a quasi-likelihood information criterion to help with the selection of the structural model on the basis of data.

Stabilized weights are recommended to be used in MSMs in place of the standard weights since they are generally less variable than the latter [3]. The stabilized weights are similar to the standard weights but are commonly defined so that the numerator is the marginal probability of observed treatment history predicted using prior treatments only while a numerator equal to 1 is instead used for the standard weights [3–5]. The denominator is the same for both types of weights. In MSMs, it has been shown that, when saturated structural models are specified, the treatment effect estimates that result from the use of stabilized or standard weights are the same [4]. In correctly specified unsaturated structural models however, the estimates differ but this difference is only due to sampling variability [4].

This note is concerned with the impact of using the common stabilized weights under different and frequently used specifications of the structural model in ordinary MSMs. As such, we focus on the estimation of the causal effect of a static treatment regime, that is, the estimation of the causal effect that a pre-specified treatment regime would have. In contrast, inferences about a dynamic treatment would consist in estimating the causal effect of a treatment regime where the treatment a subject receives at a given time point is decided according to a pre-specified rule, which might involve time-varying covariates and prior treatments. It has already been recommended not to use stabilized weights for estimating the causal effect of dynamic treatments [13]. In the sequel, we present various settings where the common stabilized weights lead to biased structural model parameter estimates while the standard weights do not. This curious (and perhaps unexpected) phenomenon is observed when the structural model targets the effect of the current treatment or the most recent treatments. This result concerns both classical MSMs and MSMs with repeated measures, although MSMs with repeated measures are arguably more susceptible to this type of structural model specification.

The paper is organized as follows. In Section 2, we introduce the notation and review the MSMs. Section 3 focuses on a very simple example that captures the problem presented in this work. In Section 4, we present the description of a simulation study devised to illustrate the potential problems of using the common stabilized weights in MSMs. The results of the simulation study are presented in Section 5. In Section 6, we investigate these issues using data from the Honolulu Heart Program. In particular, we find that the estimated effect of the current level of physical activity on blood pressure differs depending on whether standard or stabilized weights are used. We conclude with a short discussion in Section 7.

2. Notation and MSM implementations

In the following, we distinguish between two types of implementations of MSMs: classical and repeated measures.

2.1. Classical marginal structural model

Based on Robins et al. [3], we briefly review the classical MSM. In the sequel, we use capital letters to represent random variables and lower-case letters to represent possible realizations (values) of random variables.

Consider a follow-up study consisting of n sampled subjects from a population, along with covariates measured at $K+1$ time points (visits). Let $A_{k,i}$ be subject i's ($i = 1, \ldots, n$) treatment level at the kth visit from the start of the follow-up ($k = 0, \ldots, K$) and let Y_i be his outcome measured at end of follow-up, i.e. $Y_i = Y_{K+1,i}$. For the sake of simplicity, we consider continuous outcome and binary treatment variables (with $A_{k,i} = 1$ if subject i receives treatment at time k and $A_{k,i} = 0$ otherwise). For subject i, $L_{k,i}$ consists of the outcome at time k, $Y_{k,i}$, and the vector of all other measured risk factors for Y_i at time k, $V_{k,i}$, i.e. $L_{k,i} = (V_{k,i}, Y_{k,i})$. We suppose that $L_{k,i}$ temporally precedes $A_{k,i}$ for all i and k. Let $\bar{A}_{k,i} = (A_{0,i}, A_{1,i}, ..., A_{k,i})$ be subject i's treatment history through time k and let $\bar{A}_i = \bar{A}_{K,i}$. We define $\bar{L}_{k,i}$ and $\bar{L}_i$ similarly. Finally, $Y_{\bar{a}k,i}$ is subject i's counterfactual outcome at visit k, that is the outcome that would have been observed if, possibly contrary to the fact, subject i had received treatment regime $\bar{a}$ instead of his own treatment regime $\bar{a}_i$. Note that $Y_{\bar{a}k,i} = Y_{k,i}\ \forall\ k$ if $\bar{a} = \bar{a}_i$. As in Hernán et al. [4], we assume that every subject's data are independently drawn from a common distribution; therefore we drop subscript i unless it is required for clarity.

The classical marginal structural model is defined as a model for the population's mean of the counterfactual outcome at visit $K+1$ under treatment history $\bar{a}$:

$$E[Y_{\bar{a}}] = g(\bar{a};\gamma), \tag{1}$$

where g is a user defined function. Possible g functions are $g(\bar{a};\gamma) = \gamma_0 + \gamma_1 a_K + \ldots + \gamma_{K+1} a_0$, $g(\bar{a};\gamma) = \gamma_0 + \gamma_1 a_K$, $g(\bar{a};\gamma) = \gamma_0 + \gamma_1 \text{cum}(\bar{a})$ where $\text{cum}(\bar{a}) = \sum_{k=0}^{K} a_k$, or $g(\bar{a};\gamma) = \gamma_0 + \gamma_1 I_{\{\text{cum}(\bar{a})\geq 1\}}$. The parameters γ of model (1) encode the causal effect of the treatment history on the last outcome. For example, when selecting $g(\bar{a};\gamma) = \gamma_0 + \gamma_1 \text{cum}(\bar{a})$, it is hypothesized that the effect of treatment history on the mean outcome increases linearly as a function of the cumulative treatment. Thus for two treatment regimes $\bar{a}$ and $\bar{a}'$ being compared, $\gamma_1(\text{cum}(\bar{a}) - \text{cum}(\bar{a}'))$ can be interpreted as the mean difference in outcome Y, i.e. $E[Y_{\bar{a}} - Y_{\bar{a}'}]$. In particular, if $\bar{a} = \{1, 1, \ldots, 1\}$ and $\bar{a}' = \{0, 0, \ldots, 0\}$ - corresponding to the always and never treated regimes, respectively - then the expected difference in outcome is $\gamma_1(K+1)$. Similarly, if $g(\bar{a};\gamma) = \gamma_0 + \gamma_1 a_K$ is selected, then it is hypothesized that the effect of treatment history on the mean outcome only depends on the last treatment. In this case, γ_1 corresponds to the expected difference in outcome when $\bar{a} = \{\cdot, \ldots, \cdot, 1\}$ and $\bar{a}' = \{\cdot, \ldots, \cdot, 0\}$, where symbol $\cdot$ is used to represent either of the two possible levels for treatment. The issues we are concerned with in this paper stem from using structural model specifications such as this one.

The parameters γ of structural model (1) can be consistently estimated using a weighted linear regression model for $E[Y|\bar{A}]$, where each subject is weighted by the inverse probability of his observed treatment history conditional on covariates and prior treatments. Specifically, the standard weight for subject i is

$$w_i = \left\{ \prod_{k=0}^{K} \frac{1}{P(A_k = a_{k,i}|\bar{A}_{k-1} = \bar{a}_{k-1,i}, \bar{L}_k = \bar{l}_{k,i})} \right\}, \qquad i = 1, \ldots, n, \tag{2}$$

where $\bar{A}_{k-1}$ is ignored in the conditioning when $k = 0$. The standard weights w are often highly variable; therefore it is usually advised to instead use stabilized weights sw, where

$$sw_i = \left\{ \prod_{k=0}^{K} \frac{P(A_k = a_{k,i}|\bar{A}_{k-1} = \bar{a}_{k-1,i})}{P(A_k = a_{k,i}|\bar{A}_{k-1} = \bar{a}_{k-1,i}, \bar{L}_k = \bar{l}_{k,i})} \right\}. \tag{3}$$

In both (2) and (3) the $\bar{L}$ covariates are selected to ensure that the sequential (conditional) randomized assumption holds [2], that is

$$Y_{\bar{a}} \perp\!\!\!\perp A_k|\bar{A}_{k-1}, \bar{L}_k \quad \forall\ \bar{a} \text{ and } k, \tag{4}$$

where $\perp\!\!\!\perp$ symbolizes statistical independence. Perhaps underrealized is that conditioning on $\bar{A}_{k-1}$ in (4) implies that, in addition to $\bar{L}_k$, the previous treatment variables should also be regarded as potential confounding variables. This last remark is crucial for understanding the possible introduction of bias when using stabilized weights sw in MSMs.

2.2. Marginal structural model with repeated measures

Instead of modelling the mean counterfactual outcome at the end of follow-up, a MSM with repeated measures [4] aims to model the mean counterfactual outcome at each time $k+1$ ($k = 0, \dots, K$) as a function of treatment history up to time k, that is

$$E\left[Y_{\bar{a}(k+1)}\right] = g(\bar{a}_k; \gamma). \tag{5}$$

Popular choices of g function for this type of MSM implementation are $g(\bar{a}_k;\gamma) = \gamma_0 + \gamma_1 a_k + \gamma_2 k$, $g(\bar{a}_k;\gamma) = \gamma_0 + \gamma_1 a_k + \gamma_2 a_{k-1} + \gamma_3 k$, $g(\bar{a}_k;\gamma) = \gamma_0 + \gamma_1 \text{cum}(\bar{a}_k) + \gamma_2 k$, where $\text{cum}(\bar{a}_k) = \sum_{t=0}^{k} a_t$ or $g(\bar{a}_k;\gamma) = \gamma_0 + \gamma_1 I_{\{\text{cum}(\bar{a}_k)\geq 1\}} + \gamma_2 k$. Model (5) is then fitted using a weighted linear generalized estimating equation (GEE) regression for $E[Y_{k+1}|\bar{A}_k]$, where person-visit $(i, k+1)$ is weighted by its standard or stabilized weight

$$w_{k,i} = \left\{\prod_{t=0}^{k} \frac{1}{P(A_t = a_{t,i}|\bar{A}_{t-1} = \bar{a}_{t-1,i}, \bar{L}_t = \bar{l}_{t,i})}\right\} \quad \text{or} \quad sw_{k,i} = \left\{\prod_{t=0}^{k} \frac{P(A_t = a_{t,i}|\bar{A}_{t-1} = \bar{a}_{t-1,i})}{P(A_t = a_{t,i}|\bar{A}_{t-1} = \bar{a}_{t-1,i}, \bar{L}_t = \bar{l}_{t,i})}\right\}, \tag{6}$$

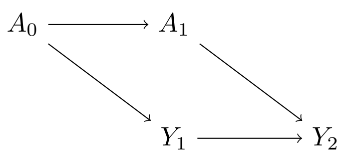
respectively. The choice of covariates $\bar{L}$ to include in these weights must also be dictated by the sequential randomized assumption [4, 14]:

$$Y_{\bar{a}(k+1)} \perp\!\!\!\perp A_t|\bar{A}_{t-1}, \bar{L}_t \quad \forall\, \bar{a} \quad \text{and} \quad k \geq t. \tag{7}$$

3. A striking example

The issues raised in this paper are best first illustrated with the simple directed acyclic graph depicted in Figure 1 (DAG1). In DAG1, Y_1 depends on A_0, A_1 depends on A_0 and Y_2 depends on both A_1 and Y_1. Here $L_0 = \emptyset$ and $L_1 = \{Y_1\}$: no L covariates other than the outcome at time 1 are considered since they are irrelevant to illustrate our problem. Covariates denoted by V are however later incorporated in our simulation scenarios presented in Section 4.1.

Figure 1. DAG1



Consider the implementation of a classical MSM based on data compatible with DAG1. While a first logical step would be the specification of the structural model, we momentarily delay this step and examine the definition of weights w and sw with regard to the sequential randomization assumption (4). Because of the presence of the open back-door path $A_1 \leftarrow A_0 \rightarrow Y_1 \rightarrow Y_2$ (★) from A_1 to Y_2 in DAG1, it follows that $Y_{\bar{a}2} \not\!\perp\!\!\!\perp A_1$ and therefore the (unconditional) randomization assumption (4) does not hold [15]. This path can be closed by A_0, which leads to $Y_{\bar{a}2} \perp\!\!\!\perp A_1|A_0$. The sequential randomization assumption is achieved conditional on treatment history since for all $\bar{a}$ and $k = 0, 1$, $Y_{\bar{a}} \equiv Y_{\bar{a}2} \perp\!\!\!\perp A_k|\bar{A}_{k-1}$ (we already have $Y_{\bar{a}2} \perp\!\!\!\perp A_0$). In principle, a MSM can thus be validly implemented with the following standard and stabilized weight definitions for subject i:

$$w_i = \frac{1}{P(A_0 = a_{0,i})} \times \frac{1}{P(A_1 = a_{1,i}|A_0 = a_{0,i})}, \tag{8}$$

and

$$sw_i = \frac{P(A_0 = a_{0,i})}{P(A_0 = a_{0,i})} \times \frac{P(A_1 = a_{1,i}|A_0 = a_{0,i})}{P(A_1 = a_{1,i}|A_0 = a_{0,i})} = 1. \quad (9)$$

Note that the second denominators in (8) and (9) could have been set to $P(A_1 = a_{1,i}|A_0 = a_{0,i}, Y_1 = y_{1,i})$ to follow the generic notation (2) and (3) for the specification of the weights. However DAG1 implies that $P(A_1 = a_{1,i}|A_0 = a_{0,i}, Y_1 = y_{1,i}) = P(A_1 = a_{1,i}|A_0 = a_{0,i})$, and thus it suffices to condition on A_0 only.

The simplification of the stabilized weight sw_i to the value 1 in (9) indicates that, in the setting represented by DAG1, *the implementation of a classical MSM with weights sw is equivalent to the implementation of an unweighted (crude) MSM*. This leads to biased or unbiased parameter estimators depending on the form of the structural model selected.

Suppose the structural model $E[Y_{\bar{a}}] = \gamma_0 + \gamma_1 a_1 + \gamma_2 a_0$ is chosen, where parameters γ_1 and γ_2 encode the causal effect of A_1 on $Y_2 \equiv Y$ and of A_0 on $Y_2 \equiv Y$, respectively. Using stabilized weights sw with this structural model yields an unbiased estimator for both γ_1 and γ_2. The parameter γ_1 is of particular interest in this case since, recall, $Y_{\bar{a}2} \not\perp\!\!\!\perp A_1$ due to the open back-door path (★). Although the confounding introduced by this back-door path is not handled by the weights (because $sw_i = 1 \, \forall \, i$), it is nonetheless accounted for by the inclusion of the treatment covariate A_0 in the regression model $E[Y|\bar{A}] = \beta_0 + \beta_1 a_1 + \beta_2 a_0$. This implies that the associational parameter β_1 coincides with the structural parameter γ_1, that is $\beta_1 = \gamma_1$, as desired.

Suppose we now consider the structural model $E[Y_{\bar{a}}] = \gamma_0 + \gamma_1 a_1$ and its associated regression model $E[Y|\bar{A}] = \beta_0 + \beta_1 a_1$. Although this reduced structural model is misspecified since A_0 has an effect on $Y_{\bar{a}2}$, it is much relevant to be able to obtain unbiased estimation for the effect this structural model is capable of identifying, namely, the effect of the most recent exposure effect (A_1) on $Y_{\bar{a}2}$. If the stabilized weights sw are used, then β_1 and γ_1 do not coincide anymore as the confounding is neither accounted for in the weights nor the regression model. With this structural model, unbiased γ_1 estimation can however be obtained by using the standard weights w since these weights do account for the confounding caused by A_0.

This example is simple and admittedly a bit artificial since a traditional regression-based approach could have correctly identified the causal effect targeted by the structural model $E[Y_{\bar{a}}] = \gamma_0 + \gamma_1 a_1$ [5]. However, it unravels a potential problem with the use of stabilized weights sw along with structural models that only include partial treatment history (e.g., current treatment or current treatment with lag 1 treatment). Indeed, a consequence of such a stabilization of the weights may be that the unconfounding achieved by the denominator is cancelled out (at least partially) by the numerator. This phenomenon is empirically demonstrated in Section 5. Also seen in Section 5 is that similar problems occur when using stabilized weights sw in MSMs with repeated measures.

4. Description of the simulation study

In this section, we present the four simulation scenarios investigated as well as the definitions of the standard and stabilized weights used in the classical implementation of the MSMs (the weights for the repeated mesures implementation are defined in a similar manner). We conclude the section with a description of the analyses we performed.

4.1. Simulation scenarios

Scenario 1. Our first simulation scenario is compatible with DAG1 (recall Figure 1). The causal relationships between the variables are as follows:

$$P(A_0 = 1) = 0.5$$
$$Y_1 = A_0 + \varepsilon_{Y_1}$$
$$P(A_1 = 1) = expit(A_0)$$
$$Y_2 = A_1 + Y_1 + \varepsilon_{Y_2},$$

where $expit(z) = e^z/(e^z + 1)$ and ε_{Y_1} and ε_{Y_2} are independent $N(0, 1)$ random variables. The standard and stabilized weights used in the classical MSM implementation are defined in (8) and (9).

Scenario 2. The second simulation scenario is only slightly more complex than the first scenario (see Figure 2):

$$P(A_0 = 1) = 0.5$$
$$Y_1 = A_0 + \varepsilon_{Y_1}$$
$$P(A_1 = 1) = expit(0.5A_0 + 0.5Y_1)$$
$$Y_2 = A_1 + Y_1 + \varepsilon_{Y_2},$$

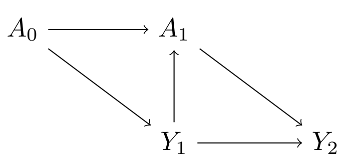
where ε_{Y_1} and ε_{Y_2} are independent $N(0, 1)$ random variables. In this scenario, the presence of the causal link between Y_1 and A_1 makes the adjustment for Y_1 in the denominator of the weights necessary to achieve (4); the standard and stabilized weights are thus defined as

$$w_i = \frac{1}{P(A_0 = a_{0,i})} \times \frac{1}{P(A_1 = a_{1,i}|A_0 = a_{0,i}, Y_1 = y_{1,i})},$$

and

$$sw_i = \frac{P(A_0 = a_{0,i})}{P(A_0 = a_{0,i})} \times \frac{P(A_1 = a_{1,i}|A_0 = a_{0,i})}{P(A_1 = a_{1,i}|A_0 = a_{0,i}, Y_1 = y_{1,i})}.$$

Figure 2



Scenario 3. The third scenario is a typical MSM representation and includes a time-dependent confounder V that is affected by previous treatment (see Figure 3):

$$V_0 = \varepsilon_{V_0}$$
$$P(A_0 = 1) = expit(0.5V_0)$$
$$Y_1 = A_0 + V_0 + \varepsilon_{Y_1}$$
$$V_1 = 0.5A_0 + \varepsilon_{V_1}$$
$$P(A_1 = 1) = expit(0.5A_0 + 0.5Y_1 + 0.5V_1)$$
$$Y_2 = A_1 + 0.5Y_1 + V_1 + \varepsilon_{Y_2},$$

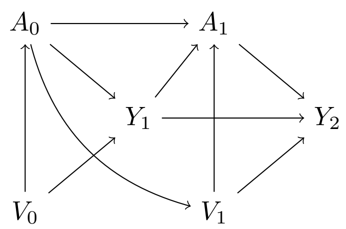
where $\varepsilon_{V_0}, \varepsilon_{Y_1}, \varepsilon_{V_1}$, and ε_{Y_2} are independent $N(0,1)$ random variables. For this scenario, we adopt the naïve strategy of including all possible covariates for the specification of the weights, that is

$$w_i = \frac{1}{P(A_0 = a_{0,i}|V_0 = v_{0,i})} \times \frac{1}{P(A_1 = a_{1,i}|A_0 = a_{0,i}, Y_1 = y_{1,i}, V_1 = v_{1,i}, V_0 = v_{0,i})},$$

and

$$sw_i = \frac{P(A_0 = a_{0,i})}{P(A_0 = a_{0,i}|V_0 = v_{0,i})} \times \frac{P(A_1 = a_{1,i}|A_0 = a_{0,i})}{P(A_1 = a_{1,i}|A_0 = a_{0,i}, Y_1 = y_{1,i}, V_1 = v_{1,i}, V_0 = v_{0,i})}.$$

Figure 3



Scenario 4. The fourth scenario is similar to the previous scenario but generates data for an additional follow-up visit (see Figure 4):

$$\begin{aligned}
&V_0 = \varepsilon_{V_0}\\
&P(A_0 = 1) = expit(0.5V_0)\\
&Y_1 = A_0 + V_0 + \varepsilon_{Y_1}\\
&V_1 = 0.25A_0 + \varepsilon_{V_1}\\
&P(A_1 = 1) = expit(0.5A_0 + 0.5Y_1 + 0.5V_1)\\
&Y_2 = A_1 + 0.25A_0 + 0.5Y_1 + V_1 + \varepsilon_{Y_2}\\
&V_2 = 0.25A_1 + \varepsilon_{V_2}\\
&P(A_2 = 1) = expit(0.5A_1 + 0.3A_0 + 0.5Y_2 + 0.5V_2)\\
&Y_3 = A_2 + 0.25A_1 + 0.5Y_2 + 0.5Y_1 + V_2 + \varepsilon_{Y_3},
\end{aligned}$$

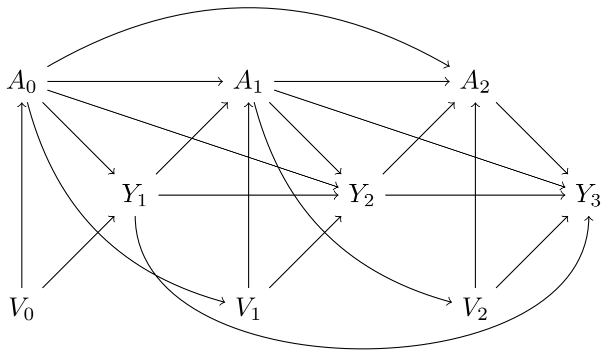
where $\varepsilon_{V_0}, \varepsilon_{Y_1}, \varepsilon_{V_1}, \varepsilon_{Y_2}, \varepsilon_{V_2}$ and ε_{Y_3} are independent $N(0,1)$ random variables. For this scenario, we also include all possible covariates for the specification of the weights, that is

$$\begin{aligned}
w_i &= \frac{1}{P(A_0 = a_{0,i}|V_0 = v_{0,i})} \times \frac{1}{P(A_1 = a_{1,i}|A_0 = a_{0,i}, Y_1 = y_{1,i}, V_1 = v_{1,i}, V_0 = v_{0,i})}\\
&\times \frac{1}{P(A_2 = a_{2,i}|A_1 = a_{1,i}, A_0 = a_{0,i}, Y_2 = y_{2,i}, Y_1 = y_{1,i}, V_2 = v_{2,i}, V_1 = v_{1,i}, V_0 = v_{0,i})},
\end{aligned}$$

and

$$\begin{aligned}
sw_i &= \frac{P(A_0 = a_{0,i})}{P(A_0 = a_{0,i}|V_0 = v_{0,i})} \times \frac{P(A_1 = a_{1,i}|A_0 = a_{0,i})}{P(A_1 = a_{1,i}|A_0 = a_{0,i}, Y_1 = y_{1,i}, V_1 = v_{1,i}, V_0 = v_{0,i})}\\
&\times \frac{P(A_2 = a_{2,i}|A_1 = a_{1,i}, A_0 = a_{0,i})}{P(A_2 = a_{2,i}|A_1 = a_{1,i}, A_0 = a_{0,i}, Y_2 = y_{2,i}, Y_1 = y_{1,i}, V_2 = v_{2,i}, V_1 = v_{1,i}, V_0 = v_{0,i})}.
\end{aligned}$$

Figure 4



4.2. Description of analyses

We generated 10 000 datasets of size $n = 1000$ for each of the four scenarios described in Section 4.1. A series of MSM analyses was performed on each dataset. The set of structural models we considered include a variety of models that have been seen in recent classical and repeated measures MSM implementations [6–11]. For the classical version of the MSMs (*cMSM*), we considered the following three structural models:

- *Full*: $E[Y_{\bar{a}}] = \gamma_0 + \gamma_1 a_K + \gamma_2 a_{K-1} + \ldots + \gamma_{K+1} a_0$;
- *Current*: $E[Y_{\bar{a}}] = \gamma_0 + \gamma_1 a_K$;
- *Cumulative*: $E[Y_{\bar{a}}] = \gamma_0 + \gamma_1 cum(\bar{a})$.

We also considered three structural models for the repeated measures implementation of the MSMs (*rmMSM*):

- *Current*: $E[Y_{\bar{a}(k+1)}] = \gamma_0 + \gamma_1 a_k + \gamma_2 k$;
- *Current+Lag1*: $E[Y_{\bar{a}(k+1)}] = \gamma_0 + \gamma_1 a_k + \gamma_2 a_{k-1} + \gamma_3 k$;
- *Cumulative*: $E[Y_{\bar{a}(k+1)}] = \gamma_0 + \gamma_1 cum(\bar{a}_k) + \gamma_2 k$.

For Scenarios 1-3, the *Full*, *Cumulative* (*cMSM* and *rmMSM*) and *Current+Lag1* structural models are correctly specified. For Scenario 4, only the *Full* and *Cumulative* (*cMSM* and *rmMSM*) structural models are correctly specified. For every scenario and structural model (both *cMSM* and *rmMSM* implementations), the data generating equations presented in Section 4.1 imply that $\gamma_1 = 1$. Recall however that γ_1 has different interpretations across structural models (see Section 2.1).

We obtained the unweighted results (which is equivalent to setting weights equal to 1) as well as the results using the standard and stabilized weights w and sw for each scenario, implementation and structural model. Specifically, for every combination of implementation/structural model/weight, we estimated the mean and standard deviation of $\hat{\gamma}_1$ based on the 10 000 datasets generated from each scenario. As recommended, we used an independence working correlation structure for the estimation of the GEEs [16, 17]. The analyses were performed using the function `geeglm` from the `R` [18] package `geepack` [19–21].

To comply with `geeglm`'s requirements, for every scenario we fitted the *Current+Lag1* structural model by deleting all the data pertaining to the first visit since the Lag1 treatment (i.e., a_{k-1}) is structurally missing when $k = 0$ [10, 11]. As a by-product of this deletion, the *rmMSM* implementation with the *Current+Lag1* structural model ends up being equivalent to the *cMSM* implementation with the *Full* structural model in the simpler scenarios (Scenarios 1-3).

5. Simulation results

The results of the simulation study are presented in Table 1.

Table 1. Results for Scenarios 1-4 by structural model and MSM implementation. The mean and the standard deviation (in parenthesis) of the estimates of γ_1 for each weight definition are provided (calculated from 10 000 datasets of size 1000).

	Classical MSM (*cMSM*)			Repeated measures MSM (*rmMSM*)		
Weight (by scenario):	*Full* ($\gamma_1 = 1$)	*Current* ($\gamma_1 = 1$)	*Cumulative* ($\gamma_1 = 1$)	*Current* ($\gamma_1 = 1$)	*Current+Lag1* ($\gamma_1 = 1$)	*Cumulative* ($\gamma_1 = 1$)
$S_1 : 1$	0.999 (0.094)	1.243 (0.096)	1.000 (0.057)	1.118 (0.061)	0.999 (0.094)	1.000 (0.053)
$S_1 : w$	0.999 (0.095)	0.999 (0.095)	1.000 (0.058)	0.999 (0.066)	0.999 (0.095)	1.000 (0.054)
$S_1 : sw$	0.999 (0.094)	1.243 (0.096)	1.000 (0.057)	1.118 (0.061)	0.999 (0.094)	1.000 (0.053)
$S_2 : 1$	1.474 (0.092)	1.681 (0.094)	1.179 (0.057)	1.332 (0.061)	1.474 (0.092)	1.126 (0.053)
$S_2 : w$	1.001 (0.073)	1.000 (0.074)	1.000 (0.054)	1.000 (0.053)	1.001 (0.073)	1.000 (0.051)
$S_2 : sw$	1.001 (0.071)	1.232 (0.078)	1.000 (0.054)	1.113 (0.056)	1.001 (0.071)	1.000 (0.051)
$S_3 : 1$	1.861 (0.103)	2.168 (0.103)	1.413 (0.061)	1.810 (0.074)	1.861 (0.103)	1.430 (0.056)
$S_3 : w$	1.003 (0.098)	1.002 (0.101)	1.001 (0.072)	1.002 (0.071)	1.003 (0.098)	1.001 (0.062)
$S_3 : sw$	1.003 (0.095)	1.314 (0.102)	1.001 (0.071)	1.153 (0.064)	1.003 (0.095)	1.001 (0.058)
$S_4 : 1$	2.112 (0.130)	2.900 (0.136)	1.527 (0.054)	2.133 (0.075)	2.061 (0.082)	1.486 (0.048)
$S_4 : w$	1.019 (0.185)	1.011 (0.195)	1.006 (0.110)	1.006 (0.116)	1.011 (0.138)	1.004 (0.085)
$S_4 : sw$	1.013 (0.175)	1.612 (0.193)	1.004 (0.091)	1.282 (0.079)	1.084 (0.097)	1.002 (0.065)

LEGEND. "1": unweighted; w: standard weights; sw: stabilized weights.

We first discuss the results for the classical MSM implementation. As expected, the use of either weights w or sw with the full structural model (*cMSM Full*) yields unbiased estimates for the true current effect of the treatment on the outcome ($\gamma_1 = 1$) in every scenario. Note that the slight bias of about 1% seen under the more complex Scenario 4 disappears when samples of size 5000 are considered (results not shown). The results for the cumulative structural model (*cMSM Cumulative*) are also unbiased under both types of weights. In Scenarios 1-4, when only the current treatment covariate is included in the structural model (*cMSM Current*), the standard weights w yield unbiased γ_1 estimates whereas the stabilized weights sw do not.

Now examining the results for the repeated measures MSM implementation, we observe that, as with the classical MSM implementation, the cumulative structural model (*rmMSM Cumulative*) yields unbiased γ_1 estimates under both weights w and sw. Moreover, the repeated measures MSM with only the current treatment covariate in the model (*rmMSM* Current) similarly yields biased estimates of γ_1 when using stabilized weights sw. The repeated measures structural model with current and previous treatments (*rmMSM Current + Lag1*) produces unbiased results for weights w and sw in Scenarios 1-3 but biased results for weights sw in Scenario 4. Unlike results for *cMSM Full*, this bias does not vanish as sample size is increased (the bias remains at 8% when $n = 5000$). This last set of results does not come as a surprise given that Scenario 4 involves three post-baseline visits ($K + 1 = 3$) whereas only two visits ($K + 1 = 2$) are considered in Scenarios 1-3. More precisely, recall that the *Current + Lag1* structural model is not misspecified in Scenarios 1-3, as opposed to Scenario 4.

The biased results for weights sw under implementation/structural model *cMSM Current*, *rmMSM Current* and *rmMSM Current + Lag1* can be explained using arguments similar to those in Section 3. First, conditioning on the past treatment(s) in the numerators of the stabilized weights sw neutralizes some deconfounding acting through the denominators of the weights, and, second, the remaining confounding is not handled by the structural model.

It is also worthwhile to mention that, while our analyses focus on parameter γ_1 for simplicity, other parameters of the structural models considered are prone to be estimated with bias when using stabilized weights sw. For instance, in Scenario 4, $\hat{\gamma}_2$ is also biased in the implementation/structural model *rmMSM Current + Lag1* when using weights sw. Indeed, for this scenario, the mean and standard deviation (in parenthesis) of the 10 000 estimates of $\gamma_2 = 1$ under the three different weighting strategies are 1) unweighted: 1.380 (0.078); 2) standard weights w: 1.007 (0.155); 3) stabilized weights sw: 1.118 (0.107). The same reasoning as the one put forward for γ_1 explains the bias found when using weights sw to estimate γ_2.

To conclude, we observed, from our simulations, that when the structural models were correctly specified, unbiased estimators were obtained when using either stabilized weights sw or standard weights w. In this case, and as expected, a reduction in variance was also seen for the structural parameter estimators resulting from the use of weights sw, as opposed to weights w. However, when the structural models were misspecified, only standard weights w led to unbiased estimation of the structural parameters. Given that selecting an appropriate structural model is a challenging issue, robustness of the weights to misspecification of this model is believed to be desirable. We feel this is particularly relevant for repeated measures implementations of MSMs, for which simplified structural model specifications could also be preferred to better take advantage of available data (e.g., see [10]). For instance, in our results, remark there is a decrease in variability for the current treatment effect estimator ($\hat{\gamma}_1$) in the *rmMSM Current* implementation/structural model as opposed to the same estimator in the *rmMSM Current + Lag1* implementation/structural model (as a result, in all scenarios, from the use of many more data points for the estimation of this effect in the former structural model).

In the next section, we investigate if other types of stabilized weights would consistently provide unbiased parameter estimates under differentially specified structural models.

5.1. *Additional analyses*

Although weights sw follow the typical definition for stabilized weights found in the MSM literature, other stabilization strategies could be employed. For a classical MSM for instance, basic stabilized weights which avoid conditioning on the past treatments in the numerators are

$$swb_i = \left\{ \prod_{k=0}^{K} \frac{P(A_k = a_{k,i})}{P(A_k = a_{k,i} | \bar{A}_{k-1} = \bar{a}_{k-1,i}, \bar{L}_k = \bar{l}_{k,i})} \right\}. \tag{10}$$

For both the classical and repeated measures implementations, we therefore also fitted the MSMs with weights swb to verify the impact of such a stabilization strategy on the distribution of $\hat{\gamma}_1$ (see Table 2). From these results, we observe that all estimates are unbiased and that notable variance reduction can be obtained by using the basic stabilized weights swb as opposed to the standard weights w (see the results for the repeated measures MSM implementation in particular).

Table 2. Results from Scenarios 1-4 by structural model and MSM implementation using basic stabilized weights swb. The mean and the standard deviation (in parenthesis) of the estimates of γ_1 are provided (calculated from 10 000 datasets of size 1000).

	Classical MSM (*cMSM*)			Repeated measures MSM (*rmMSM*)		
Scenario	*Full* ($\gamma_1 = 1$)	*Current* ($\gamma_1 = 1$)	*Cumulative* ($\gamma_1 = 1$)	*Current* ($\gamma_1 = 1$)	*Current+Lag1* ($\gamma_1 = 1$)	*Cumulative* ($\gamma_1 = 1$)
S_1	0.999 (0.094)	0.999 (0.094)	1.000 (0.058)	1.000 (0.056)	0.999 (0.094)	1.000 (0.053)
S_2	1.001 (0.073)	1.000 (0.074)	1.000 (0.054)	1.000 (0.048)	1.001 (0.073)	1.000 (0.051)
S_3	1.003 (0.098)	1.002 (0.101)	1.001 (0.071)	1.001 (0.060)	1.003 (0.098)	1.001 (0.057)
S_4	1.012 (0.179)	1.008 (0.189)	1.003 (0.092)	1.002 (0.071)	1.006 (0.103)	1.001 (0.062)

6. The Honolulu Heart Program results

In this section, data from the Honolulu Heart Program (HHP) are used to illustrate how the choice of weights can influence the exposure effect estimates in non-simulated MSM analyses.

The HHP is a study of Japanese-American men living in Oahu, Hawaii, which examined 8006 participants. Participants were born between 1900 and 1919 (aged 45-68 years old at study entry) and were recruited from the selective service registry. They were evaluated at multiple time points beginning in 1965 and followed until 1994 for deaths and morbid events. Information regarding physical activity participation was collected by questionnaire at Exam 1 (1965-68), Exam 2 (1968-1971) and Exam 4 (1991-1993). Blood pressure (BP) was measured manually (in mmHg) by a trained professional during each exam. More details about HHP can be found elsewhere [22].

Repeated measures MSMs were used to estimate the causal effect of physical activity on systolic blood pressure (SBP) and diastolic blood pressure (DBP). Since physical activity was not measured at Exam 3, and since there was a long delay between Exam 2 and Exam 4, we chose to only use data from the first two exams. Our belief is that the effect of current and prior physical activity history on current BP is primarily a function of current physical activity. Our structural model for each type of BP thus has the following form:

$$E[Y_{\bar{a}k}] = \gamma_0 + \gamma_1 a_k + \gamma_2 k, \tag{11}$$

where unlike equation (5), which has a delayed treatment effect, the treatment effect in (11) is immediate. In our structural models, $Y_{\bar{a}k}$ is the counterfactual outcome (either SBP or DBP) at Exam k ($k = 1, 2$) and a_k is the physical activity level (active or inactive) reported at Exam k.

For both MSM analyses, the covariates used to calculate the visit specific weights at the first time point (Exam 1) were: age (in years) at Exam 1 and employment at Exam 1 (employed or unemployed). For the second time point (Exam 2), the weights were calculated using: employment at Exam 1, physical activity level at Exam 1, hypertension medication usage at Exam 1 (yes or no), BMI at Exam 1 (in kg/m^2), age at Exam 2 and employment at Exam 2. Note that hypertension medication usage at Exam 1 and BMI at Exam 1 were not considered in the calculation of the weights at the first time point because these variables are believed to be effects of the physical activity level at Exam 1. Subjects with missing data at a given time point were removed from the analyses (about 1% for Exam 1 and about 3% for Exam 2).

We estimated the effect of current level of physical activity on current SBP and DBP using repeated measures MSMs and the same four weights that were investigated in the simulation studies ("1", w, sw and swb). For the estimation of the GEEs, a robust variance estimator was used along with an independence working correlation structure. The results are summarized in Table 3.

Table 3. Estimated effect of current physical activity level on current systolic (SBP) and diastolic (DBP) blood pressure.

Weights	Estimate for SBP (95% CI)	Estimate for DBP (95% CI)
1	-2.29 (-3.35, -1.22)	-0.82 (-1.40, -0.24)
w	-1.85 (-2.94, -0.75)	-0.43 (-1.04, 0.17)
sw	-1.94 (-3.59, -0.29)	-1.29 (-2.18, -0.39)
swb	-1.56 (-2.56, -0.55)	-0.29 (-0.84, 0.26)

LEGEND. "1": unweighted; w: standard weights; sw: (common) stabilized weights; swb: basic stabilized weights.

Upon the examination of Table 3, we remark that the estimates of the effect of current physical activity on current SBP are relatively robust to the choice of weights. However, the choice of weights has a notable impact on the estimates of the effect of physical activity on DBP. In this case, the estimates obtained using an unweighted MSM or a MSM with common stabilized weights sw exhibit a significant decrease of DBP with physical activity at level $\alpha = 0.05$, whereas a non signifiant decrease is obtained from the MSMs with standard weights w and basic stabilized weights swb. These last results are in accordance with the *rmMSM Current* results from the simulation study where the unweighted and common stabilized weights sw estimates departed from those obtained with standard weights w and basic stabilized weights swb. Because there is believed to be time-dependent confounding, the unweighted repeated measures MSM is considered to be inappropriate for estimating the causal effect of current physical activity on current DBP. We also note that the confidence intervals obtained with the basic stabilized weights swb are slightly narrower than those obtained with the standard weights w.

7. Discussion

Although it is widely known that the weighting scheme affects the variance of MSM estimators, it is less well known that it can also affect their bias. Using a series of simulated examples, we showed that the utilization of the most common stabilized weights (weights sw) may lead to biased parameter estimates when structural models feature only partial information on treatment history, such as the current or most recent treatments. The diffusion of this result is critical since such structural model specifications are often seen in repeated measures MSMs, a type of MSMs which is increasingly used to perform causal inferences [6–11].

The phenomenon documented in this paper adds to the number of subtle issues arising in the implementation of MSMs [5]. Indeed, our results suggest that the choice of weights needs to be done according to the structural model that is specified. Particularly, we advise analysts to avoid using the common stabilized weights when the analyses target the estimation of the current or most recent treatment causal effects. In this context, the analysts could adopt the basic stabilized weights swb put forward herein, simple weights which have been found to yield unbiased results under all scenarios and structural models investigated.

References

[1] Robins JM. Marginal structural models. Proceedings of the Section on Bayesian Statistical Science. American Statistical Association: Alexandria, VA. 1998; 1-10.

[2] Robins JM. Marginal structural models versus structural nested models as tools for causal inference. In Statistical models in epidemiology, the environment, and clinical trials, 95–133. Springer 2000.

[3] Robins JM, Hernán MA, Brumback BA. Marginal structural models and causal inference in epidemiology. *Epidemiology*. 2000; **11(5)**: 550–560.

[4] Hernán MA, Brumback BA, Robins JM. Estimating the causal effect of zidovudine on cd4 count with a marginal structural model for repeated measures. *Statistics in medicine*. 2002; **21(12)**: 1689–1709.

[5] Yang W, Joffe MM. Subtle issues in model specification and estimation of marginal structural models. *Pharmacoepidemiology and Drug Safety*. 2012; **21(3)**: 241–245.

[6] Fairall LR, Bachmann MO, Louwagie G, van Vuuren C, Chikobvu P, Steyn D, Staniland GH, Timmerman V, Msimanga M, Seebregts CJ, Boulle A, Nhiwatiwa R, Bateman ED, Zwarenstein MF, Chapman RD. Effectiveness of antiretroviral treatment in a south african program: a cohort study. *Archives of Internal Medicine*. 2008; **168(1)**: 86–93.

[7] Patel K, Hernán MA, Williams PL, Seeger JD, McIntosh K, Dyke RB, Seage GR 3rd, Pediatric AIDS Clinical Trials Group 219/219C Study Team. (2008). Long-term effects of highly active antiretroviral therapy on cd4+ cell evolution among children and adolescents infected with hiv: 5 years and counting. *Clinical Infectious Diseases*. 2008; **46(11)**: 1751–1760.

[8] Sampson RJ, Laub JH, Wimer C. Does marriage reduce crime? A counterfactual approach to within-individual causal effects. *Criminology*. 2006; **44(3)**: 465–508.

[9] Schildcrout JS, Haneuse S, Peterson JF, Denny JC, Matheny ME, Waitman LR, Miller RA. Analyses of longitudinal, hospital clinical laboratory data with application to blood glucose concentrations. *Statistics in medicine*. 2011; **30(27)**: 3208–3220.

[10] VanderWeele TJ, Hawkley LC, Thisted RA, Cacioppo JT. A marginal structural model analysis for loneliness: implications for intervention trials and clinical practice. *Journal of consulting and clinical psychology*. 2011; **79(2)**: 225–235.

[11] VanderWeele TJ, Hawkley LC, Cacioppo JT. On the reciprocal association between loneliness and subjective well-being. *American journal of epidemiology*. 2012; **176(9)**: 777–784.

[12] Platt RW, Brookhart MA, Cole SR, Westreich D, Schisterman EF. An information criterion for marginal structural models. *Statistics in Medicine*. 2013; **32(8)**:1383–93.

[13] Robins JM, Hernán MA. Estimation of the causal effects of time-varying exposure. In: Fitzmaurice GM, Davidian M, Verbeke G, Molenberghs G. ed. *Longitudinal data analysis*; Boca Raton: CRC Press; 2009: 553–599.

[14] Brumback BA, Hernán MA, Haneuse SJ, Robins JM. Sensitivity analyses for unmeasured confounding assuming a marginal structural model for repeated measures. *Statistics in medicine*. 2004; **23(5)**: 749–67.

[15] Pearl J. Reflection, elaborations, and discussions with readers. In: Pearl J, ed. *Causality: Models, reasoning, and inference*; 2nd ed. New York: Cambridge University Press; 2009: 341–344.

[16] Tchetgen Tchetgen EJ, Glymour MM, Weuve J, Robins J. A cautionary note on specification of the correlation structure in inverse-probability-weighted estimation for repeated measures. *Harvard University Biostatistics Working Paper Series*. 2012; Working Paper 140.

[17] Tchetgen Tchetgen EJ, Glymour MM, Weuve J, Robins JM. Specifying the correlation structure in inverse-probability-weighting estimation for repeated measures. *Epidemiology*. 2012; **23(4)**:644–646.

[18] R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria 2013; `http://www.R-project.org/`.

[19] Højsgaard S, Halekoh U, Yan J. The R package geepack for generalized estimating equations. *Journal of Statistical Software*. 2006; **15(2)**:1–11.

[20] Yan J, Fine JP. Estimating equations for association structures *Statistics in Medicine*. 2004; **23(6)**: 859–880.

[21] Yan J. geepack: yet another package for generalized estimating equations. *R-News*. 2002; **2/3**: 12–14.

[22] Kagan A, Harris BR, Winkelstein Jr W, Johnson KG, Kato H, Syme SL, Rhoads GG, Gay ML, Nichaman MZ, Hamilton HB, et al. 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)**:345–364.

