

Pharmacological Treatment of Adult Attention-Deficit/Hyperactivity Disorder (ADHD) in a Longitudinal Observational Study: Estimated Treatment Effect Strengthened by Improved Covariate Balance

Ole Klungsoyr^{1*}, Mats Fredriksen²

¹Oslo Centre for Biostatistics and Epidemiology, Section for Treatment Research, Department for Research and Education, Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway

²Division of Mental Health and Addiction, Vestfold Hospital Trust, Tønsberg, Norway

Email: *ole.klungsoyr@medisin.uio.no

How to cite this paper: Klungsoyr, O. and Fredriksen, M. (2017) Pharmacological Treatment of Adult Attention-Deficit/Hyperactivity Disorder (ADHD) in a Longitudinal Observational Study: Estimated Treatment Effect Strengthened by Improved Covariate Balance. *Open Journal of Statistics*, 7, 988-1012.

<https://doi.org/10.4236/ojs.2017.76070>

Received: October 27, 2017

Accepted: December 10, 2017

Published: December 13, 2017

Copyright © 2017 by authors and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

An improved method for estimation of causal effects from observational data is demonstrated. Applications in medicine have been few, and the purpose of the present study is to contribute new clinical insight by means of this new and more sophisticated analysis. Long term effect of medication for adult ADHD patients is not resolved. A model with causal parameters to represent effect of medication was formulated, which accounts for time-varying confounding and selection-bias from loss to follow-up. The popular marginal structural model (MSM) for causal inference, of Robins *et al.*, adjusts for time-varying confounding, but suffers from lack of robustness for misspecification in the weights. Recent work by Imai and Ratkovic [1] [2] achieves robustness in the MSM, through improved covariate balance (CBMSM). The CBMSM (freely available software) was compared with a standard fit of a MSM and a naive regression model, to give a robust estimate of the true treatment effect in 250 previously non-medicated adults, treated for one year, in a specialized ADHD outpatient clinic in Norway. Covariate balance was greatly improved, resulting in a stronger treatment effect than without this improvement. In terms of treatment effect per week, early stages seemed to have the strongest influence. An estimated average reduction of 4 units on the symptom scale assessed at 12 weeks, for hypothetical medication in the 9 - 12 weeks period compared to no medication in this period, was found. The treatment effect persisted throughout the whole year, with an estimated aver-

age reduction of 0.7 units per week on symptoms assessed at one year, for hypothetical medication in the last 13 weeks of the year, compared to no medication in this period. The present findings support a strong and causal direct and indirect effect of pharmacological treatment of adults with ADHD on improvement in symptoms, and with a stronger treatment effect than has been reported.

Keywords

Covariate Balance, Propensity Score, Marginal Structural Model, Causal Treatment Effect, ADHD

1. Introduction

Estimation of causal treatment effects from observational studies has obvious limitations and challenges. However, the randomized controlled trial is often no alternative, due to practical or ethical reasons. If a scientific question of interest really is a causal one, the analysis should target the specific question, even if the price is strong assumptions. In most cases, a scientist would more easily relate to subject specific assumptions, than an association with few assumptions and no causal statement, which can be misleading (in both magnitude and direction). Causal methods are more explicit in the assumptions for a causal interpretation, and often more robust for certain types of bias, even though they are just as susceptible to e.g. unmeasured confounding as more traditional methods.

Short term effect of medication for treatment of ADHD is well documented [3] [4] [5] [6]. However, questions about long-term effects are less resolved, and more long-term prospective studies on treatment of adult ADHD patients with no prior medication, in a practical clinical setting are warranted [7]. Selection bias, high drop-out rate and limited reports on side-effects characterizes the few that have been conducted [8] [9].

To estimate the causal effect of medication, a successful study design would compare two arms in a trial, an active on medication and a control arm without medication. However, as for ADHD, present knowledge makes inclusion to the control-arm questionable. Also, if all eligible patients are offered treatment, a high number of patients are more likely to participate. A research design where all patients are offered medication and follow-up assessments, resembles clinical practice, with findings that apply to the population of interest. Causal effects can still be estimated under additional assumptions.

If all patients start off with medication, and those that experience intolerance, or a high number of side-effects, terminate the treatment, it means that there are patients on and off medication. A direct comparison of these groups yields a biased estimate of the causal treatment effect, due to feedback between the treatment assignment process and the outcome, e.g. symptoms. If treatment decreases subsequent symptoms on the one hand, but on the other hand increases side –

effects and intolerance which again leads to termination of treatment, one might expect that the direct comparison between those on and off medication is an overestimate of the true treatment effect. Bias in the opposite direction can be expected due to time-varying confounding. Prior medication likely improves subsequent symptoms which represents a confounder for the association between continued medication and final symptoms. The positive association between prior symptoms and both continued medication and final symptoms will weaken the negative association between continued medication and final symptoms, and result in underestimation of the true treatment effect.

A marginal structural model (MSM), with inverse probability weighting (IPW) can account for such feedback, and make an unbiased estimate of the treatment effect, under specific assumptions [10]. In spite of its popularity, in a wide range of fields, the MSM also has limitations. For example, it is sensitive to miss-specification of the treatment assignment model in the weights. Fitting a MSM can give an overly low estimated probability for treatment for certain covariate combinations, which yields a high weight in the analysis (inverse of the probability). This high weight will propagate through time since the weight in a later time period is the cumulative product of the weights at earlier periods. Imai and Ratkovic (I&R) have successfully improved the robustness of such weighting methods as in the MSM, by introducing the covariate balancing propensity score (CBPS) methodology, first for cross-sectional, and later for longitudinal data [1] [2].

In this study, a re-analysis of longitudinal follow-up data on medication history and symptoms over one year for adult ADHD patients is presented. Ordinary longitudinal analysis of the same data, has been previously published, with limited focus on causal treatment effect, and little attempt to account for time-varying confounding mentioned above [11]. The present analysis fits an ordinary MSM, and compares it with the refinement of improved covariate balance (CBMSM), to estimate the causal effect of treatment on symptoms at one year follow-up.

The study sample, measures used, and the formulation of the MSM and CBMSM, including necessary assumptions for causal interpretation, will be described in the Method section. In Results, the findings from the CBMSM method is presented and contrasted to the ordinary fitted MSM. The Discussion section, relates the results to the previously published analysis, lists some strengths and limitations in the methods and interprets the results.

2. Methods

2.1. Sample and Treatment Schedule

Patients were included at a specialized outpatient clinic in Vestfold, Norway, between May 2009 and December 2010. Referred patients were aged 18 - 60, had to fulfill DSM-5 criteria for ADHD [12], and without experience with any ADHD medication in adulthood. A total of 250 patients were included. Relevant baseline characteristics were: 129 were female (52%), 140 were unemployed or

work disabled (56%), 188 had at least one comorbid disorder (75%), 156 used concomitant non-ADHD medication (69%), mean age was 32.5 years (SD = 9.8), mean body weight was 77.0 kg (SD = 16.8), mean number of years of education was 11 (SD = 2.3). By DSM-IV subgroups of adult ADHD, 97 patients were categorized as ADHD-inattentive, 113 as ADHD-combined, 17 as ADHD-hyperactive, and 23 as ADHD-residual. A more complete description of the sample can be found in a recent publication [11].

All patients received methylphenidate as first-line medication and psychosocial treatment according to the national treatment guidelines (Norwegian Directorate of Health, 2005). Patients were assessed for symptoms, functioning and side-effects at scheduled follow-up visits at baseline, 6 weeks, 12 weeks, 26 weeks and 52 weeks. Standard-titration with immediate-release methylphenidate (MPH-IR) was prescribed for the first six weeks; 5 mg three times a day, and if tolerated stepwise increase until maximum 60 mg/day. Thereafter a flexible dose titration was applied to optimize efficacy (maximum 120 mg/day). Shift into extended-release methylphenidate (MPH-ER) was offered at the three-month visit if patients reported difficulties with compliance, annoying fluctuations in effect or otherwise wanted to try an easier administration form. If MPH was not tolerated or was ineffective, second-line medications were short-acting dextroamphetamine (dAMP) or atomoxetine (ATX). The dose of dAMP was escalated until a maximum 50 mg/day, and dose of ATX to a maximum of 120 mg/day [11].

2.2. Measures

2.2.1. Predictive Baseline Characteristics

Baseline characteristics of the sample have been thoroughly described previously [11] [13], and constitute demographics, measures of functioning, mental distress, ADHD symptom-level, and comorbid anxiety, depression, bi-polar disorder, and drug and alcohol use disorders. To be diagnosed with ADHD, patients must retrospectively have endorsed at least 6 out of 9 DSM-IV symptoms of inattention and/or hyperactivity/impulsivity in childhood, and currently been assessed by two board certified psychiatrists using the Norwegian version of the Diagnostic Interview for ADHD in Adults, second edition (DIVA 2.0) [14] with at least 5 out of 9 DSM-IV symptoms of inattention and/or hyperactivity/impulsivity for the last 6 months (according to diagnostic threshold for ADHD in adults in DSM-5 [12], and with their ADHD symptoms related to significant impairment in social, academic, or occupational functioning. Comorbid psychiatric disorders were assessed with the MINI International Neuropsychiatric Interview Plus (M.I.N.I.-Plus) [15] [16] [17]. The time-fixed covariates selected in the regression model for treatment assignment (time-varying) were baseline side-effect level (predefined) measured by mean score of the Canadian Attention Deficit Hyperactivity Disorder Resource Alliance (CADDRA) patient and ADHD medication form (8.36) for indicating side effects [18], alcohol use disorder (dichotomous), any anxiety disorder (dichotomous), baseline psychosocial function by the Global Assessment of Functioning

(GAF)—symptom level (GAF-S) [19], and age at baseline.

2.2.2. Time-Varying Covariates and Outcome

The primary outcome measure was current ADHD symptoms on the 18-item Adult ADHD Self-Report Scale version 1.1 (ASRS), Norwegian version [20]. A continuous scoring method was used, and frequency of ADHD symptoms present since last visit, self-rated on a 5-point scale (0 - 4) on each item, with a sum score from 0 to 72 points. Cronbach's alpha was 0.86.

Two psychiatrists, not involved in the treatment, assessed overall psychosocial functioning (last two weeks) by the Global Assessment of Functioning (GAF) Scale [19], the split version of symptom (GAF-S) and function (GAF-F) to improve reliability [21]. The intra-class correlation coefficient between the raters was 0.83 for GAF-S and 0.79 for GAF-F in the pilot [13].

Level of mental distress over the last week was self-rated on 90 items on a 5 point scale, and the mean of all items is referred to as Global-Severity-Index (GSI) [22] [23].

Side-effects (Mean Side Effects—MSE) was quantified by a measure of tolerability Patient and ADHD medication form. The questionnaire lists symptoms frequently associated with stimulant treatment, and each item is scored by frequency (score 0 - 3) [18].

Dose is a time-varying covariate and expresses daily dose in mg (or dose equivalent) per day.

Medication (MED) is a time-varying dichotomous indicator (1/0) of whether on medication or not, at the time of assessment.

2.3. Estimation of Causal Effects

Standard notation in the causal inference literature, makes use of “counterfactual outcomes”. A counterfactual, or potential outcome, is an outcome for a hypothetical treatment regime. There are general (non-parametric) assumptions necessary to estimate causal effects from observational data [24]. *Consistency* means that a person's counterfactual outcome under his/her observed treatment history is precisely his/her observed outcome. *Exchangeability* means that the treatment groups are representative for each other. In a randomized controlled trial (RCT) exchangeability is guaranteed by design, and implies that if the treatment group had been untreated, they would have responded similar to the control group. In observational data, exchangeability is unrealistic, so instead one hopes for conditional exchangeability (exchangeability within levels of confounding covariates), often called “no unmeasured confounding”. *Positivity* means a positive probability for all levels of treatment, in all levels of measured covariates. In other words, with treatment = yes/no, positivity requires that there are both treated and untreated persons for all combinations of the confounding covariates. *Correct model specification* is necessary for all models, also in causal inference, both for the outcome model and models for the weights in the present analysis.

These assumptions are helpful to interpret results, but not testable from data,

although some indication of non-positivity or misspecification is given by the weight distribution in weighting methods like the MSM [24]. A mean of the “stabilized weights” far from 1 indicates non-positivity or model misspecification. If conditional exchangeability is satisfied, an average causal effect in the population can be estimated by combining stratified treatment effects within levels of confounding covariates, or by regression models. However, this becomes unfeasible with a high number of covariates, some of which may be continuous. With many strata and limited sample size (or severe differences between treatment groups), some strata are bound to lack a treatment group. This non-positivity is not easy detectable, and leads to extrapolation in a regression model. In light of this, propensity score (PS) methods have become popular [25]. PS is the probability of being treated, conditional on all confounding covariates. It is a balancing score, a univariate measure that summarizes all confounding covariates. To adjust for confounding, it suffices to stratify/condition on the PS, instead of all covariates, which avoids extrapolation, and allows for diagnosing differences in the treatment groups [26] [27]. If the treatment groups have little overlap in the PS distribution, conditional exchangeability is questionable and causal inference might be flawed. The weights in the MSM are functions of the PS.

Exchangeability between treatment groups implies covariate balance, that for any covariate, the treatment groups have equal distributions, e.g. equality of weighted means. Conditional exchangeability implies that all confounding covariates are measured and balanced. In an observational study, without knowing whether or not conditional exchangeability is satisfied, to check covariate balance in measured covariates is informative and recommended [28], even though measured covariate balance does not guaranty balance in unmeasured confounding covariates. Balance diagnostics in measured covariates is a minimum for unbiased estimation, and routinely reported in PS studies, but often ignored in the MSM literature, maybe due to the more complex time-varying weights [29]. The MSM has been found to be highly sensitive to model misspecification in the treatment assignment model [2]. Rare covariate combinations with low number of treated, results in large weights that dominate the analysis. Recent work of Imai and Ratkovic has focused on automatic improvement in covariate balance for the MSM in a longitudinal setting, to make it more robust. Their method, here denoted covariate balanced MSM (CBMSM), generalizes their covariate balance propensity score (CBPS) from 2014 [1], to the longitudinal setting, and is available in open-source software as an R package (CBPS) at CRAN [30] [31]. In the present analysis, the causal effect of ADHD medication on symptoms at one year follow-up is estimated, with adjustment for time-varying confounding (MSM), with and without improved covariate balance (CBMSM).

2.4. MSM for ADHD-Symptoms

A standard longitudinal analysis (linear mixed model) of the ASRS symptoms in the present study-sample, has previously been published (Figures 1-3) [11].

Large individual variation in symptoms over time was found, but rapid average improvement (during first 6 weeks), followed by persistent low level (Figure 1). Different medication histories reflected acute intolerance, continuous medication, treatment termination after some time, and off and on trajectories (Figure 2).

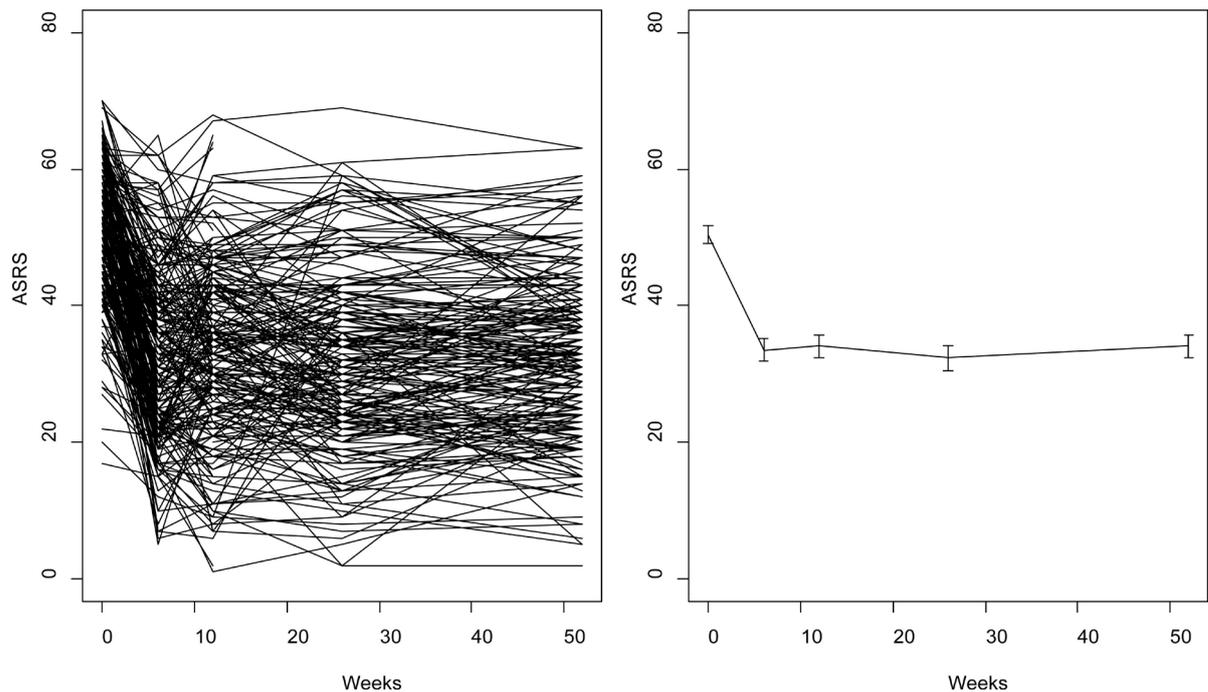


Figure 1. Symptom level (ASRS) at different assessments during one year treatment, in a Norwegian sample of 250 patients with ADHD diagnosed in adulthood.

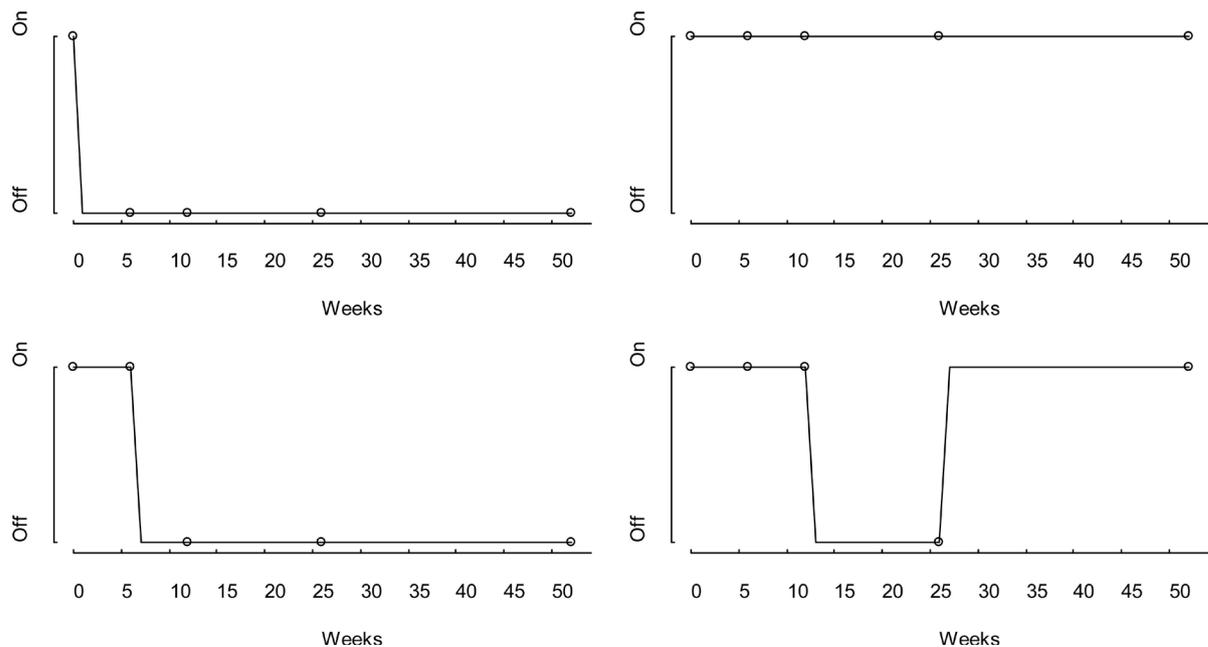


Figure 2. Examples of different medication histories, intolerance (top left), continuous (top right), short term (bottom left), on/off (bottom right), in a Norwegian sample of 250 patients with ADHD diagnosed in adulthood.

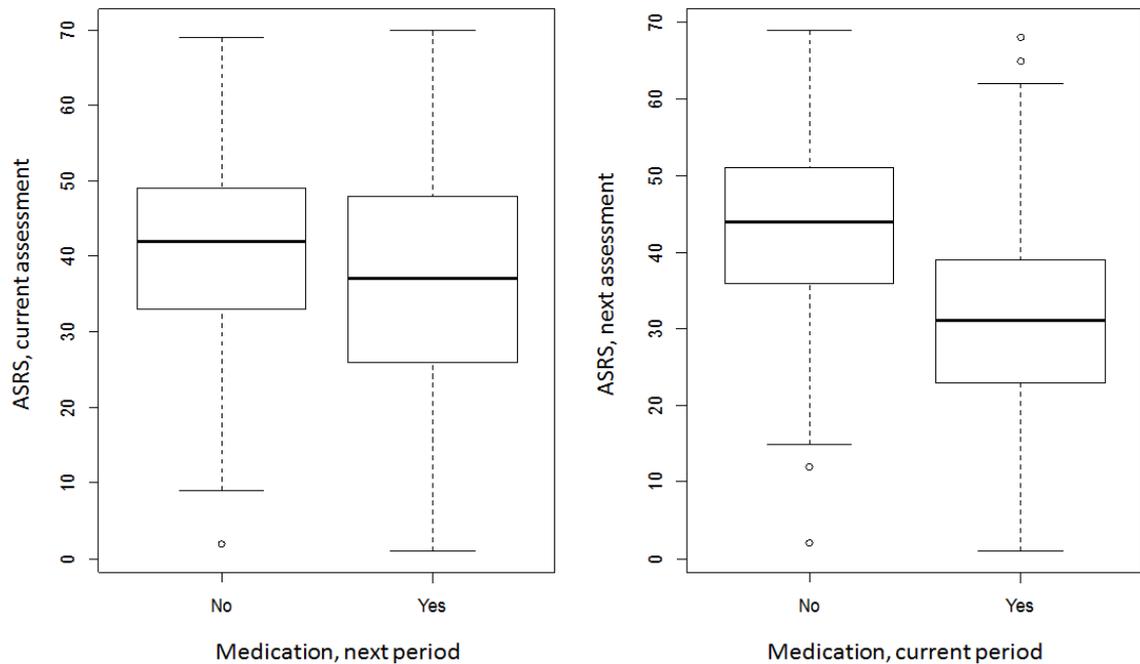


Figure 3. Feedback between treatment (on/off medication each period) and symptoms (ASRS). Boxplot of ASRS distribution for those on/off medication in next period (left panel), and at the end of current period (right panel), in a Norwegian sample of 250 patients with ADHD diagnosed in adulthood.

Feedback between medication and symptoms is illustrated in **Figure 3**. Receiving medication in time-period j , was followed by subsequent average lower symptom level than no-medication (**Figure 3**, right panel). On the other hand, the off-medication group at time-period $j + 1$, had a history of higher average symptom level at the prior time-period than the on-medication group (**Figure 3**, left panel). These box-plots suggest that the data contains information about the true treatment effect, and calls for methods with the ability to account for different biases, and target the effect of interest. A MSM (with IPW) can adjust for time-varying confounding, and selection bias from differential loss to follow-up, in settings where standard regression models (univariate or longitudinal) fail [10].

The MSM is a model for a counterfactual outcome, univariate or repeated measure (longitudinal). In the present application, the univariate model is sufficient and easy to interpret. The MSM for the mean counterfactual end-of-study univariate ASRS level, denoted by Y (ADHD symptoms at one year follow-up), can be formulated as:

$$E\left[Y\left(\overline{med}_{52}\right)\right] = \beta_0 + \beta_1 med_6 + \beta_2 med_{12} + \beta_3 med_{26} + \beta_4 med_{52} \quad (1)$$

where med_6 , med_{12} , med_{26} , and med_{52} are dichotomous (1/0) indicators of on/off medication at baseline, 6 weeks, 12 weeks, 26 weeks and 52 weeks, respectively. \overline{med}_{52} which denotes medication history at 52 weeks, was limited to one of five different regimes (**Table 1**), for reasons that will become clear later. $Y\left(\overline{med}_{52}\right)$ is the counter-factual final outcome under a hypothetical one year

Table 1. Different medication regimes (\overline{med}) considered in this study, of a Norwegian sample of 250 patients with ADHD diagnosed in adulthood.

Medication at	Baseline	6 weeks	12 weeks	26 weeks	52 weeks
Average of 3 weeks of medication ($N=4$)	Yes	No	No	No	No
Average of 9 weeks of medication ($N=8$)	Yes	Yes	No	No	No
Average of 19 weeks of medication ($N=13$)	Yes	Yes	Yes	No	No
Average of 39 weeks of medication ($N=14$)	Yes	Yes	Yes	Yes	No
One year of medication ($N=131$)	Yes	Yes	Yes	Yes	Yes

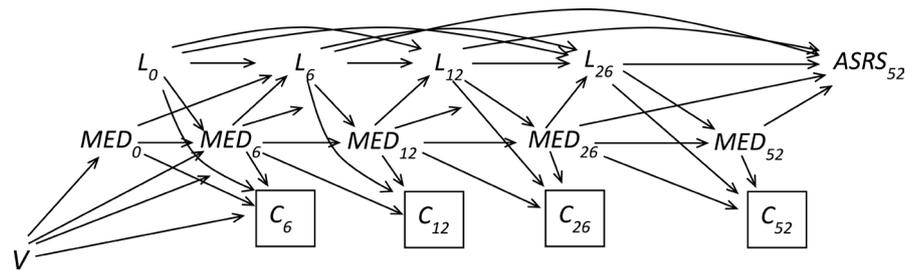


Figure 4. Causal diagram (DAG) of the longitudinal structure of the data, a Norwegian sample of 250 patients with ADHD diagnosed in adulthood.

medication history. β_0 is the marginal mean of Y with medication at baseline only, β_1 corresponds to average change in Y under medication regime (1, 1, 0, 0, 0) relative to (1, 0, 0, 0, 0), β_2 corresponds to average change in Y under medication regime (1, 1, 1, 0, 0) relative to (1, 1, 0, 0, 0), and so on.

The time-varying covariates GSI, GAF-S and GAF-F, are viewed as closest in time to the actual follow-up assessment. MSE (side-effects) and ASRS (symptom-level) contain information from the whole preceding period, but with most weight close to the follow-up assessment. MED (indicator for on/off medication) describes the treatment status at the time of assessment. $MED = 0$ means that medication was terminated some time during the preceding period. The parameters in Equation (1) therefore each represents the effect of half the previous and half the following period of extra medication (on average), with uniformly distributed termination times. With respect to the direction of effects, the influence is allowed from medication to symptom-level in the same period, and from symptom-level to medication in the following period (Figure 4).

To estimate the MSM in (1), a weighted univariate linear regression model (associational), conditional on medication history only, is fitted [10]:

$$E\left[Y|\overline{MED}_{52}\right] = \beta_0 + \beta_1 MED_6 + \beta_2 MED_{12} + \beta_3 MED_{26} + \beta_4 MED_{52} \quad (2)$$

with weights (stabilized) for each person at time j , specified by:

$$SW_j = \prod_{s=1}^j \frac{Pr\left(MED_s|\overline{MED}_{s-1}, V\right)}{Pr\left(MED_s|\overline{MED}_{s-1}, \overline{L}_{s-1}, V\right)} \quad (3)$$

The weights are estimated by a series of logistic regressions. V represents

baseline confounders and \bar{L}_{s-1} the time-varying confounders (only in denominator). Robust standard errors are used, e.g. by the “sandwich” software package in R [32] to account for the fact that the weights are unknown and have to be estimated [10]. The parameters in (2) have causal interpretations, and are unbiased estimates of the theoretical parameters in (1), under the assumptions in the previous section, no selection bias from loss to follow-up, and no measurement error [10]. All confounder adjustment is achieved through the weights. Weights can be calculated using the “ipw” software package in R [33]. With no time-varying confounding, the numerator and denominator in (3) are equal and the weights equals one. If a combination of the confounders in the denominator gives a small estimated probability for treatment, the weight can become very large, and will propagate through time, due to the cumulative product. This is automatically alleviated with the CBMSM algorithm.

2.5. Improved Covariate Balance with CBMSM

Estimating the weights in (3) for the MSM, is usually done with a series of logistic regressions by maximum likelihood (ML), in a generalized linear model (GLM). With misspecification in these PS-models, maximizing the likelihood might not balance the covariates [1]. I&R resolve this by explicitly imposing moment conditions (covariate balance equations), e.g. the weighted mean of a covariate in the treated group is set equal to the weighted mean in the untreated group (moment condition equal to zero). Parameter estimation and moment conditions are solved simultaneously. With only one dichotomous treatment variable (yes/no) in a cross-sectional setting, total number of balance equations equals the number of parameters, and CBPS in the simplest form is just-identified (**Appendix**) [1]. In a longitudinal setting, the potential number of different treatment histories is large, and balance is needed across all of them, which results in over-identification, *i.e.* more balance equations than parameters. This over-identified set of equations does not have one unique solution, but is solved to minimize a quadratic function of the moment conditions (as close as possible to zero) to give “minimum imbalance” by the generalized method of moments (GMM) (**Appendix**) [34].

With medication = yes/no in four consecutive time periods (everyone started off on medication), the number of potential treatment histories are 16. To avoid loss of precision in parameter estimates from patterns with few patients, only monotone treatment regimens were allowed (**Table 1**). This means that once a patient temporarily terminated medication, a restart was excluded from the analysis, instead the data for this patient was censored from the point of restart and forward. In this way, e.g. β_3 in model (1) refers to average change in Y among those with an average of 39 weeks of medication, relative to those with an average of 19 weeks (**Table 1**).

To assess differences between treatment groups, overlap in the PS distributions for the five groups, was examined. Here, the PS represented probability for

termination of medication sooner or later, as a function of baseline covariates.

2.5.1. Covariate (im) Balance

To assess balance, the standardized mean difference (*SMD*) for each covariate (difference in weighted means between two treatment groups, divided by the population standard deviation), is calculated. More precisely, the *SMD* express imbalance, which is smaller for better balance. As a rule of thumb, a value less than 0.25 is commonly considered to be acceptable [26] [29] in a cross-sectional setting. No such acceptance level has been suggested for longitudinal data. Balance is more challenging in a longitudinal setting, because of the over-identified set of equations. For each time period, every covariate is balanced on all possible current and future treatment patterns conditional on the past treatment history. With $J = 4$ as the number of time periods (as in the present data), each covariate enters $2^J - 2^{j-1}$ moment conditions at time period j [2], for example in the first time period ($j = 1$), covariates enter 15 moment conditions of which each represent the same number of equations as number of unknown parameters and number of covariates. Obviously, for a covariate to satisfy a higher number of equations, the “best” solution is expected to be further away from an exact solution, which implies more imbalance. The GMM estimation algorithm minimizes imbalance across all moment conditions and time periods (**Appendix**) [2].

2.5.2. Estimating the Causal Effect of ADHD Medication

Model selection for the weights in Equation (3) was based on combinations of covariates that were significant predictors for continued medication at several time-points, and with resulting imbalance as low as possible. The following logistic model was chosen:

$$\begin{aligned} & \text{logit}\left(\Pr\left(MED_j = 1 \mid \dots\right)\right) \\ &= \alpha_0 + \alpha_1 anx + \alpha_2 mse_0 + \alpha_3 age + \alpha_4 age^2 + \alpha_5 gafs_0 + \alpha_6 alc + \alpha_7 alc \times mse_0 \quad (4) \\ & \quad + \alpha_8 alc \times gafs_0 + \alpha_9 dose_{j-1} + \alpha_{10} asrs_{j-1} + \alpha_{11} gsi_{j-1} + \alpha_{12} gaff_{j-1}^2 \end{aligned}$$

with the abbreviations *anx*—indicator for any anxiety disorder at baseline, *mse₀*—mean side-effects at baseline, *age*—age at baseline, *gafs₀*—baseline psychosocial function, symptom part, *alc*—indicator for baseline alcohol use disorder, *dose_{j-1}*—dose in previous period, *asrs_{j-1}*—ADHD symptoms at previous assessment, *gsi_{j-1}*—distress at previous assessment, *gaff_{j-1}²*—squared psychosocial function, function part at previous assessment.

A similar model for censoring in the censoring weights was constructed. Censoring was used for deviation from monotone treatment (all patients that deviated from the five different regimes in **Table 1**), and for those with missing values in the ASRS outcome, at some point.

2.6. Missing Data

In these data, there were missing values, both in the outcome and in covariates. The 26 week assessment had the most missing, *gafs₂₆*, *gaff₂₆*—66 missing,

mse_{26} —63 missing, gsi_{26} —49 missing, $asrs_{26}$ —46 missing, and 27 variables had no missing values. 132 observations had complete cases with no missing values at any assessment. With substantial missingness for some covariates (more than 20%), possibly caused by side-effects, multiple imputation (MI) was considered a suitable method to reduce the impact of loss of data. The missing covariates were multiple imputed, with chained equations [35] under the assumption of missing at random (MAR), and performed with the MICE package in R [36]. The ASRS symptom outcome, was also imputed when serving as a covariate, but not as outcome. Censoring weights were chosen to correct for potential bias from missing outcomes (loss to follow-up), a straight forward extension to the inverse probability of treatment weights to correct for time-varying confounding. For the missing covariates, the data were imputed 10 times, and all analysis (measures of balance and parameter estimates) was repeated for each dataset, and combined using Rubin's rules [37]. After imputation, the dataset consisted of 170 observations with no missing outcome and covariates, and with monotone treatment history.

3. Results

3.1. Overlap of Different Treatment Groups

The five different groups (Table 1) were comparable, with good overlap in the whole range of predicted probability for termination of treatment (PS), although some variation in the right tail (Figure 5). The two groups with termination fol-

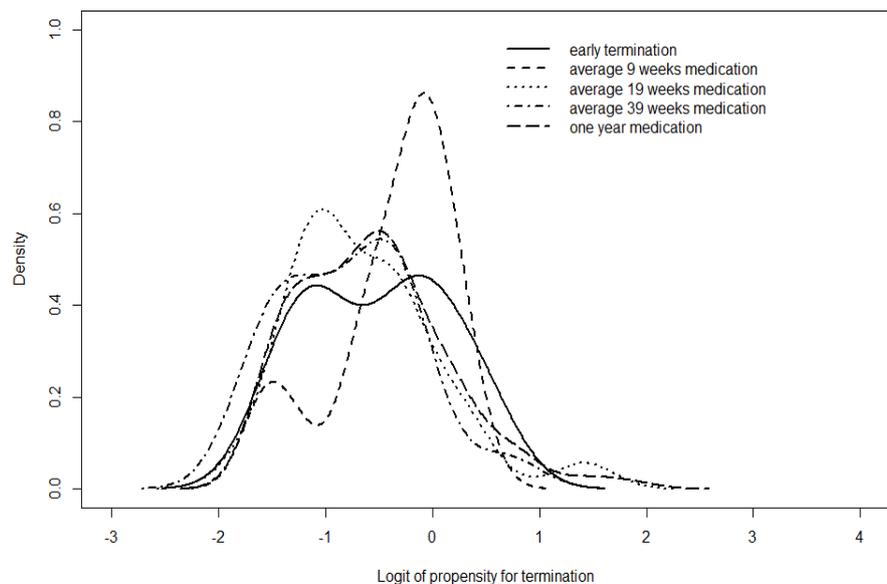


Figure 5. Overlap in the five different treatment groups, with respect to predicted probability for termination of medication (PS), by baseline covariates: $gaff_0$ —psychosocial functioning at baseline, anx —indicator for baseline anxiety disorder, alc —indicator for baseline alcohol use disorder, $asrs_0$ —baseline ADHD symptoms, gsi_0 —baseline measure of distress, $asrs_0 \times gsi_0$ —interaction, in a Norwegian sample of 250 patients with ADHD diagnosed in adulthood.

lowing baseline and 6 weeks had clearly right-shifted probability mass (higher probability for termination) compared to the other groups, which is intuitively reasonable, being closest in time to the explanatory covariates. No signs of serious violations of conditional exchangeability (among measured baseline covariates) or positivity were found (Figure 5 and Figure 9).

3.2. Covariate (im) Balance

Figures 6-8 shows graphically the *SMDs* for each covariate in the model from Equation (4), in estimation of the parameters in the MSM, and Table 2 summarizes these findings, with a comparison between the results from the CBMSM weights and from an ordinary fit of the MSM, with GLM weights (“ipw” package in R) [33]. All results are MI-estimates based on 10 multiple imputed datasets. In estimation of both β_3 and β_4 , the imbalance is clearly reduced in the estimation from the CBMSM weights compared to unweighted estimation (Figure 6, Figure 7) and compared with ordinary fitted MSM with GLM weights (Figure 8). The recommended acceptance-level of ± 0.25 in the cross-sectional setting, is indicated with dotted vertical lines. Three covariates exceeded this limit, however those considered important time-varying confounders, like previous

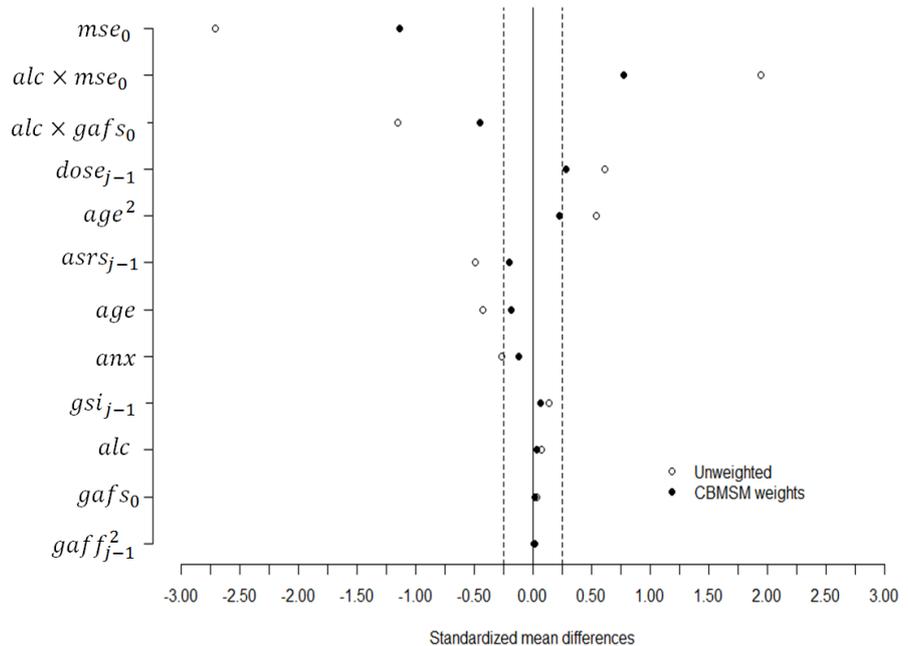


Figure 6. Improved balance in covariates for estimation of β_3 , by standardized mean differences (*SMD*), with covariates from Equation (4) and CBMSM weights (sorted by unweighted imbalance), in a Norwegian sample of 250 patients with ADHD diagnosed in adulthood. Results are based on multiple imputation to limit influence from missing covariates (10 imputed datasets, $N = 170$). mse_0 : baseline mean side effects, alc : baseline alcohol use disorder, $gafs_0$: baseline psychosocial function—symptom part, $dose_{j-1}$: dose in previous period, age : age at baseline, $asrs_{j-1}$: ADHD symptoms at previous assessment, anx : baseline anxiety disorder, gsi_{j-1} : distress at previous assessment, $gaff_{j-1}$: psychosocial function—function part at previous assessment.

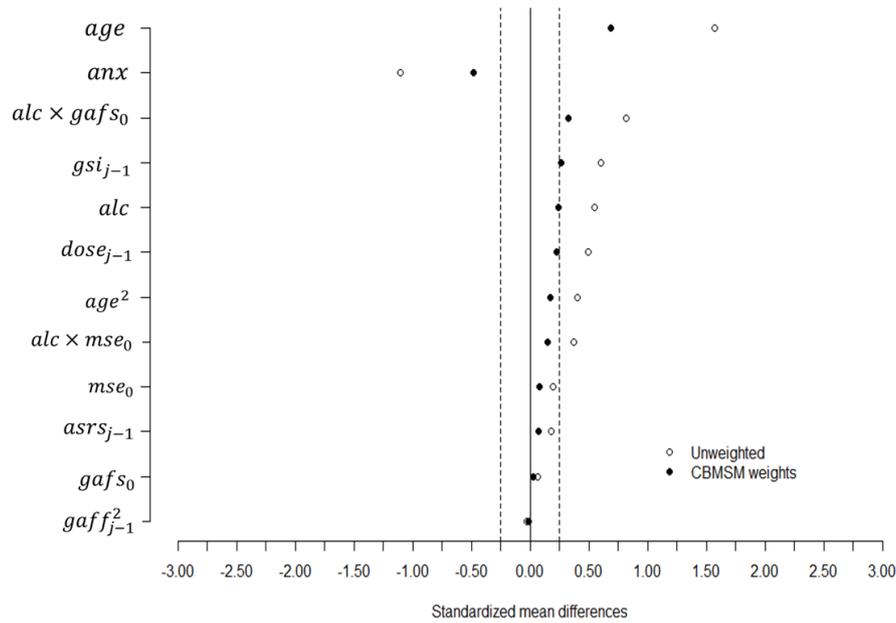


Figure 7. Improved balance in covariates for estimation of β_4 , by standardized mean differences (SMD), with covariates from Equation (4) and CBMSM weights (sorted by unweighted imbalance), in a Norwegian sample of 250 patients with ADHD diagnosed in adulthood. Results are based on multiple imputation to limit influence from missing covariates (10 imputed datasets, $N=170$). mse_0 : baseline mean side effects, alc : baseline alcohol use disorder, $gafs_0$: baseline psychosocial function—symptom part, $dose_{j-1}$: dose in previous period, age : age at baseline, $asrs_{j-1}$: ADHD symptoms at previous assessment, anx : baseline anxiety disorder, gsi_{j-1} : distress at previous assessment, $gaff_{j-1}$: psychosocial function—function part at previous assessment.

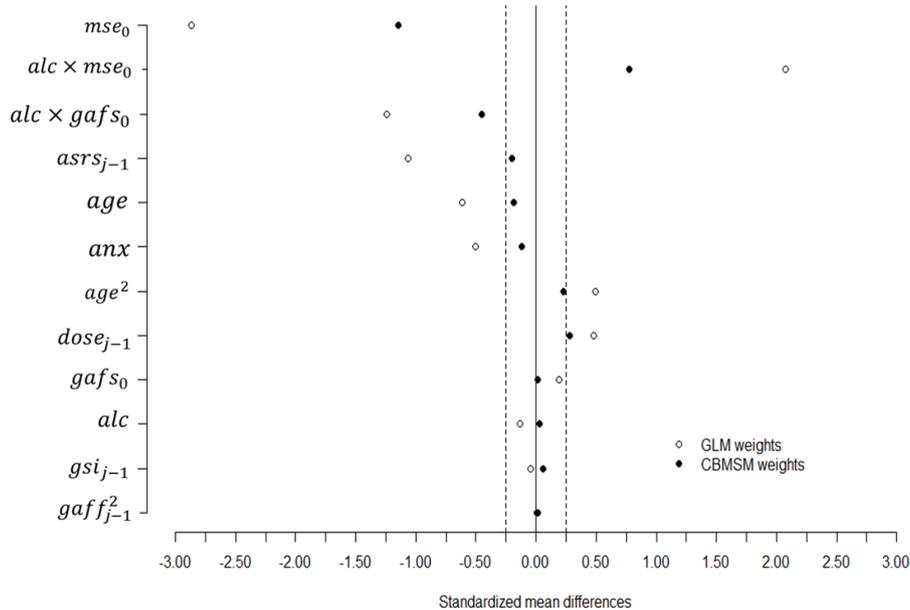


Figure 8. Comparison of balance in covariates for estimation of β_3 , by standardized mean differences (SMD), covariates from Equation (4) and with GLM versus CBMSM weights (sorted by GLM weight imbalance), in a Norwegian sample of 250 patients with ADHD diagnosed in adulthood. Results are based on multiple imputation to limit influence from missing covariates (10 imputed datasets, $N=170$). mse_0 : baseline mean side effects, alc : baseline alcohol use disorder, $gafs_0$: baseline psychosocial function—symptom part, $dose_{j-1}$: dose in previous period, age : age at baseline, $asrs_{j-1}$: ADHD symptoms at previous assessment, anx : baseline anxiety disorder, gsi_{j-1} : distress at previous assessment, $gaff_{j-1}$: psychosocial function—function part at previous assessment.

Table 2. Average and spread in absolute imbalance (*SMD*) among covariates in treatment assignment model, for estimation of parameters in the MSM. Results are based on multiple imputation to limit influence from missing covariates (10 imputed datasets, $N = 170$).

	CBMSM		GLM	
	mean ($ SMD $)	sd ($ SMD $)	mean ($ SMD $)	sd ($ SMD $)
β_1	0.87	0.65	1.82	1.43
β_2	0.95	0.89	2.3	2.25
β_3	0.29	0.35	0.81	0.88
β_4	0.23	0.2	0.51	0.38

symptom-level, psychosocial functioning and distress were within the limit (Figures 6-8). Slightly less imbalance in estimation of β_4 , compared to β_3 is also indicated in the figures. This is confirmed in Table 2, where average absolute imbalance over all covariates are presented and compared with GLM weights. Estimates of early effects of treatment (β_1 and β_2) are clearly based on more imbalance than later effects (β_3 and β_4). This is due to more moment conditions for each covariate early in the observation period, and the fact that groups with early termination were more different from the completers than those with late termination. CBMSM weights are superior to GLM weights for estimation of all effects with respect to imbalance (Table 2). Finally, the average imbalance for estimation of both β_3 and β_4 are close to the cross-sectional acceptance level of 0.25 (Table 2).

3.3. Estimation of Causal Effects of ADHD Medication

In Table 3(a) different estimates (MI-estimates) of the causal parameters in the MSM from Equation (1) are presented, and estimates for the CBMSM method across imputations are presented in Table 3(b). Each parameter has the interpretation of an effect of hypothetical medication in a period, in addition to being on medication up until that period. CBMSM estimates (Table 3(a), left-most column) are believed to be closest to the true unknown causal parameters, and show strong effects from medication in both the 19 - 39 weeks period, and the 39 - 52 weeks period. Hypothetical medication relative to no medication in the 19 - 39 weeks period (being on medication until 19 weeks), would be expected to reduce ASRS symptoms significantly at one year with 12.06 units ($\beta_3 = -12.06$, 95% CI: $-20.3, -3.88$, $p = 0.004$), and hypothetical medication relative to no medication in the 39 - 52 weeks period (being on medication until 39 weeks) would result in a significant reduction in symptoms of 8.72 units ($\beta_4 = -8.72$, 95% CI: $-15.6, -1.88$, $p = 0.013$). The β_3 coefficient represents the longest period of medication (20 weeks on average) and had the strongest effect. It represents a “direct effect”, not mediated through the last period. Effects of hypothetical medication in the first two periods (3 - 9 weeks, and 9 - 19 weeks) were non-significant, indicating possible mediation through subsequent periods.

Compared to an ordinary fitted MSM with GLM weights (**Table 3(a)**, column 4 - 6) the differences in magnitude are modest, when the sizable improvement in average imbalance is taken into account (**Table 2**). The effect from the last period is reduced by 25% relative to the CBMSM, but still strong and significant. To indicate the magnitude of selection bias from non-monotone treatment pattern or loss to follow-up, CBMSM estimates without censoring weights are presented in columns 10 - 12. Compared to the fully adjusted CBMSM estimates, the selection bias was not negligible and seemed to result in negative bias (underestimation) mostly in the 19 - 39 weeks effect with a 15% reduction in the estimate. Without any adjustment, the effects would be greatly underestimated, with a relative bias of 11% (19 - 39 weeks period) and 47% (39 - 52 weeks period), respectively (**Table 3(a)**, column 13 - 15). For reference, an ordinary linear regression, with adjustment for time-varying covariates is included (**Table 3(a)**, column 7 - 9). The results show serious underestimation with this approach, with 37% and 53% relative bias for the 19 - 39 weeks, and 39 - 52 weeks periods, respectively. The effect of the 39 - 52 weeks period is no longer significant, and the bias is increased, also compared to the no adjustment (unweighted) case. This clearly demonstrates failure of standard regression to account for time-varying confounding, together with direction and magnitude of this bias. In **Table 3(b)**, parameter estimates for the CBMSM method are shown across imputations. In each imputation, the missing covariates are predicted from all available data (e.g. including side-effects), and little between-imputation variability was found.

Separate models (CBMSM) to assess influence of different periods of medication on ASRS symptoms at 26 weeks, and at 12 weeks were fitted (MI-estimates) to examine the course of treatment effects and symptoms (**Table 4**). With $ASRS_{26}$ as outcome, the coefficient for the most recent period of medication represents the effect of hypothetical medication relative to no medication in the 19 - 26 weeks period (when being on medication prior to 19 weeks). This seven-week period of medication had a strong and significant effect, and would be expected to reduce symptoms with 20 units ($\beta_3 = -19.98$, 95% CI: $-23.6, -16.37$, $p < 0.001$), a magnitude of the same size as the total treatment effect from two periods in the $ASRS_{52}$ -model. Estimated effects of earlier periods were nonsignificant, in spite of β_2 representing the longest period of medication (10 weeks), possibly resulting from mediation through the last period. With $ASRS_{12}$ as outcome, the coefficient for the most recent period of medication represents the effect of hypothetical medication relative to no medication in the 9 - 12 weeks period (when being on medication prior to 9 weeks). This three-week period of medication had a strong and significant effect of 13 units expected reduction in symptoms ($\beta_2 = -13.1$, 95% CI: $-18.54, -7.65$, $p < 0.001$). Estimated effect of the earlier period, 3 - 9 weeks, was nonsignificant, in spite of longer duration, possibly mediated through the last period.

In summary, medication seemed to have strong and positive effects on subsequent symptoms, across the whole year, both for early and late periods. Symp-

Table 3. (a) Estimated average causal effects of medication in different time periods on final symptoms (ASRS) at one year follow-up, in a Norwegian sample of 250 patients with ADHD diagnosed in adulthood. Results are based on multiple imputation to limit influence from missing covariates (10 imputed datasets, $N = 170$). CBMSM: results of improved covariate balance from Kosuke & Ratkovic, GLM: ordinary IPW estimation by logistic regression in a MSM, LM: ordinary least squares regression with adjustment for time-varying covariates, CBMSMcens: CBMSM algorithm without adjustment for censoring (loss to follow-up or nonmonotone treatment regimes), Unweighted: without any adjustment (for confounding or censoring); (b) Parameter estimates for the CBMSM method, in each imputed dataset (10 imputed datasets, $N = 170$).

(a)															
	CBMSM			GLM			LM			CBMSMcens			Unweighted		
	beta	95% CI	p-value	beta	95% CI	p-value	beta	95% CI	p-value	beta	95% CI	p-value	beta	95% CI	p-value
<i>Int</i>	44.67	38.17, 51.16	<0.001	44.67	38.64, 50.71	<0.001	32.17	24.67, 39.67	<0.001	45.4	38.19, 52.61	<0.001	45.5	37.06, 53.94	<0.001
<i>med₆</i>	-3.22	-16.13, 9.68	0.62	1.15	-13.71, 16.01	0.88	-0.63	-10.56, 9.3	0.9	-4.03	-15.92, 7.85	0.51	-3.38	-16.21, 9.46	0.61
<i>med₁₂</i>	7.46	-4.77, 19.7	0.23	1.85	-12.64, 16.34	0.8	0.43	-8.58, 9.44	0.92	5.18	-5.44, 15.8	0.34	2.72	-8.07, 13.52	0.62
<i>med₂₆</i>	-12.06	-20.3, -3.88	0.004	-12.24	-19.69, -4.79	0.001	-7.56	-14.09, -1.03	0.02	-10.25	-18.65, -1.84	0.02	-10.77	-18.26, -3.29	0.005
<i>med₅₂</i>	-8.72	-15.6, -1.88	0.013	-6.51	-12.31, -0.71	0.03	-4.08	-9.35, 1.2	0.13	-8.10	-15.31, -0.9	0.03	-4.65	-10.7, 1.4	0.13

(b)					
Imputation	CBMSM				
	β_0	β_1	β_2	β_3	β_4
1	45.08	-3.4	7.69	-12.17	-9.28
2	44.05	-2.53	7.20	-12.09	-8.54
3	45.10	-3.36	7.19	-11.84	-9.05
4	44.76	-3.53	7.73	-12.0	-8.73
5	45.07	-4.0	7.21	-11.44	-8.61
6	44.29	-3.83	8.51	-12.21	-8.73
7	43.6	-2.7	8.21	-12.83	-8.11
8	44.47	-3.04	7.29	-11.98	-8.61
9	44.96	-2.88	6.98	-12.11	-8.64
10	45.31	-2.96	6.64	11.96	-8.89

Column means are MI-estimates in first column, **Table 3(a)**.

Table 4. Estimated average causal effects (CBMSM) of medication in different time periods on symptoms (ASRS) at one year, 26 weeks, and 12 weeks of follow-up, in a Norwegian sample of 250 patients with ADHD diagnosed in adulthood. Results are based on multiple imputation to limit influence from missing covariates (10 imputed datasets, $N = 170$).

	ASRS ₅₂			ASRS ₂₆			ASRS ₁₂		
	beta	95% CI	p-value	beta	95% CI	p-value	beta	95% CI	p-value
<i>Int</i>	44.67	38.17, 51.16	<0.001	49.29	42.88, 55.69	<0.001	34.19	13.76, 54.63	0.001
<i>med₆</i>	-3.22	-16.13, 9.68	0.62	-7.12	-16.84, 2.59	0.15	10.64	-10.4, 31.71	0.32
<i>med₁₂</i>	7.46	-4.77, 19.7	0.23	5.67	-2.23, 13.58	0.16	-13.1	-18.54, -7.65	<0.001
<i>med₂₆</i>	-12.06	-20.3, -3.88	0.004	-19.98	-23.6, -16.37	<0.001			
<i>med₅₂</i>	-8.72	-15.6, -1.88	0.013						

toms, measured at early stages by $ASRS_{12}$ and $ASRS_{26}$ would be expected to be largely reduced by medication immediately prior to symptom assessment, with no direct effects from earlier medication. Symptoms at one year follow-up, $ASRS_{52}$, would be expected to be influenced by more medication history. Hypothetical medication in the mid-period, 19 - 39 weeks (with no medication in the last period) seemed to be most influential, with a direct effect in addition to an indirect effect through medication in the 39 - 52 week period. This is in line with a dose-response effect (longer duration of treatment corresponds to larger effect), but in contrast to symptoms at early stages, with nonsignificant effects from earlier periods of treatment with longer duration. Treatment effect per week seemed strongest in the early stages, but persisted across the whole year.

4. Discussion

In the present application of the CBMSM method, strong causal effects of medication on ADHD symptoms were found, under standard assumptions for causal inference. A MSM (GLM weights) was fitted, to account for time-varying confounding and selection bias, but with covariate balance that was greatly improved with the CBMSM. The magnitude of the treatment effects from the CBMSM, are the ones believed to be closest to the unknown true levels. The alternative estimates were smaller (25% relative difference compared to the standard MSM), which indicates that for these data, standard analysis yields underestimated treatment effects.

With a maximum of one year medication in four consecutive time periods, all periods seemed to represent improvement in ADHD symptoms, over the course of the study. In terms of treatment effect per week, early stages seemed to have the strongest influence, with an average reduction of approximately 4 units per week of symptoms at 12 weeks ($ASRS_{12}$) ($\beta_2 = -13.1$, 95% $CI: -18.54, -7.65$, $p < 0.001$) for hypothetical medication in the 9 - 12 week period. However, the treatment effect persisted over the whole year, with an expected reduction of 0.7 units per week of symptoms at the final assessment ($ASRS_{52}$) with hypothetical continued medication over the last 13 weeks.

The CBMSM model revealed causal information, not accessible in a standard longitudinal analysis. The average symptom level during the course of the observation period is characterized by a rapid drop in the first 6 weeks, followed by an enduring constant low level for the rest of the year (**Figure 1**). The improvement in symptoms could falsely be attributed to medication in the first 6 weeks alone. After 6 weeks of medication, either maximum improvement was obtained or medication could be terminated, or would be necessary to maintain a low symptom level. Adjustment for time-varying side-effects, distress, psychosocial functioning and symptoms would not result in unbiased effect of treatment. The reduction in average symptoms seen in **Figure 1** underestimates the causal effects of treatment, better predicted from the sum of effects from different periods in the CBMSM model. In addition to persistent and decreasing treatment

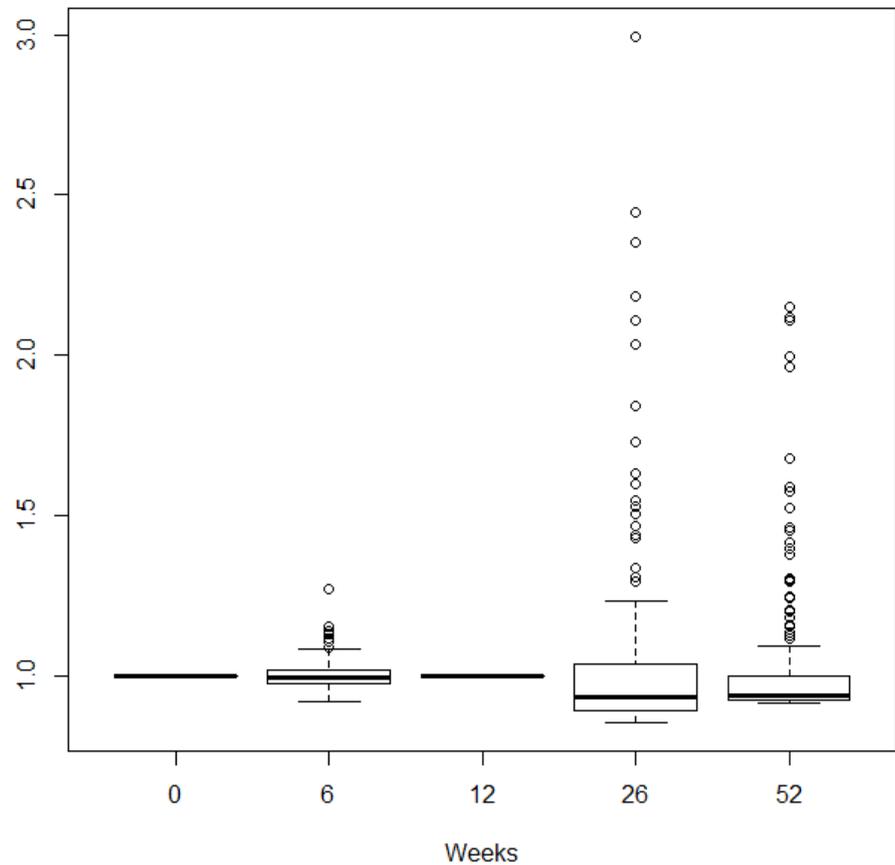


Figure 9. Stabilized censoring weight distributions (boxplots) for each time period, for censoring from missing in outcome ($ASRS_{52}$) or non-monotone treatment patterns, in a Norwegian sample of 250 patients with ADHD diagnosed in adulthood.

effect across the whole period, another notable change in the treatment effect was revealed. The different models for symptoms at 12 weeks ($ASRS_{12}$) and 26 weeks ($ASRS_{26}$), both showed significant causal effects of hypothetical medication in the most recent treatment period, with no significant effects from earlier periods, in spite of their length. As for symptoms at end of follow-up ($ASRS_{52}$), significant causal effects were found for hypothetical medication in the two most recent periods. The causal direct effect from hypothetical medication in the 19 - 39 week period on symptoms at one year, may represent other pathways than through continued treatment in the 39 - 52 week period (pharmacological pathway). Clinical insight has suggested that a patient's social environment needs time to adapt and trust the improvement, when the patient is treated. This adaption process is slower than the pharmacological effect, but can be important for further improvement, through motivation, support, and positive feedback. A four and a half months adaption process is one possible explanation for the direct effect from the 19 - 39 week period. In this case, treatment effect in the early periods would be mostly pharmacological. Alternatively, the direct effect from the 19 - 39 week period corresponds to a dose-response effect. In that case, other explanations for the lack of such dose-response effects on symptoms at 12 and

26 weeks are needed.

If the models for ASRS₁₂, ASRS₂₆, and ASRS₅₂ had been similar with respect to causal effects of medication, they could have been combined in a repeated measures MSM (longitudinal model), for possible gain in efficiency. In the present application, however, the differences in the separate univariate models were informative.

5. Conclusions

In conclusion, the CBMSM improved covariate balance substantially in these data, compared to the standard fitted MSM, and should therefore represent estimates closer to the causal effect of medication one would find in a successful RCT.

The improved covariate balance, strengthened the treatment effect, compared to the MSM, and even more so compared to ordinary longitudinal analysis with naive adjustment for time-varying confounding, reported in the literature [11]. The causal model also provided possible new clinical insight with respect to the dynamics of the effect of pharmacological treatment on adults with ADHD. A persistent and strong effect across the whole year on improvement in symptoms, was supported, and with both direct and indirect causal pathways.

References

- [1] Imai, K. and Ratkovic, M. (2014) Covariate Balancing Propensity Score. *Journal of the Royal Statistical Society B*, **76**, 243-263. <https://doi.org/10.1111/rssb.12027>
- [2] Imai, K. and Ratkovic, M. (2015) Robust Estimation of Inverse Probability Weights for Marginal Structural Models. *Journal of the American Statistical Association*, **110**, 1013-1023. <https://doi.org/10.1080/01621459.2014.956872>
- [3] Faraone, S.V. and Buitelaar, J. (2010) Comparing the Efficacy of Stimulants for ADHD in Children and Adolescents using Meta-Analysis. *European Child & Adolescent Psychiatry*, **19**, 353-364. <https://doi.org/10.1007/s00787-009-0054-3>
- [4] Faraone, S.V. and Glatt, S.J. (2010) A Comparison of the Efficacy of Medications for Adult Attention-Deficit/Hyperactivity Disorder Using Meta-Analysis of Effect Sizes. *Journal of Clinical Psychiatry*, **71**, 754-763. <https://doi.org/10.4088/JCP.08m04902pur>
- [5] Castells, X., *et al.* (2011) Amphetamines for Attention Deficit Hyperactivity Disorder (ADHD) in Adults. The Cochrane Library. <https://doi.org/10.1002/14651858.CD007813.pub2>
- [6] Castells, X., *et al.* (2011) Efficacy of Methylphenidate for Adults with Attention-Deficit Hyperactivity Disorder: A Meta-Regression Analysis. *CNS Drugs*, **25**, 157-169. <https://doi.org/10.2165/11539440-000000000-00000>
- [7] Fredriksen, M., *et al.* (2013) Long-Term Efficacy and Safety of Treatment with Stimulants and Atomoxetine in Adult ADHD: A Review of Controlled and Naturalistic Studies. *European Neuropsychopharmacology*, **23**, 508-527. <https://doi.org/10.1016/j.euroneuro.2012.07.016>
- [8] Torgersen, T., Gjervan, B. and Rasmussen, K. (2008) Treatment of Adult ADHD: Is Current Knowledge Useful to Clinicians? *Neuropsychiatric Disease and Treatment*, **2008**, 177-186. <https://doi.org/10.2147/NDT.S1223>

- [9] Fredriksen, M. and Peleikis, D.E. (2016) Long-Term Pharmacotherapy of Adults with Attention Deficit Hyperactivity Disorder: A Literature Review and Clinical Study. *Basic & Clinical Pharmacology & Toxicology*, **118**, 23-31. <https://doi.org/10.1111/bcpt.12477>
- [10] Robins, J.M., Hernan, M.A. and Brumback, B. (2000) Marginal Structural Models and Causal Inference in Epidemiology. *Epidemiology*, **11**, 550-560. <https://doi.org/10.1097/00001648-200009000-00011>
- [11] Fredriksen, M., *et al.* (2014) Effectiveness of One-Year Pharmacological Treatment of Adult-Deficit/Hyperactivity Disorder (ADHD): An Open-Label Prospective Study of Time in Treatment, Dose, Side-Effects and Comorbidity. *European Neuropsychopharmacology*, **24**, 1873-1884. <https://doi.org/10.1016/j.euroneuro.2014.09.013>
- [12] Association, A.P. (2013) Diagnostic and Statistical Manual of Mental Disorders. 5th Edition. <https://doi.org/10.1176/appi.books.9780890425596>
- [13] Fredriksen, M., *et al.* (2014) Childhood and Persistent ADHD Symptoms Associated with Educational Failure and Long-Term Occupational Disability in Adult ADHD. *ADHD Attention Deficit and Hyperactivity Disorders*, **6**, 87-99. <https://doi.org/10.1007/s12402-014-0126-1>
- [14] Kooij, J.J.S. and Francken, M.H. (2010) DIVA 2.0. Diagnostic Interview Voor ADHD in Adults bij volwassenen [DIVA 20 Diagnostic Interview ADHD in Adults]. DIVA Foundation.
- [15] Sheehan, D.V., *et al.* (1998) The Mini-International Neuropsychiatric Interview (MINI): The Development and Validation of a Structured Diagnostic Psychiatric Interview for DSM-IV and ICD-10. *Journal of Clinical Psychiatry*, **59**, 22-33.
- [16] Sheehan, D., *et al.* (2002) MINI PLUS. Mini International Neuropsychiatric Interview. Norwegian Translation Version 5.0.0, University of South Florida Institute for Research in Psychiatry, Tampa.
- [17] Leiknes, K., Leganger, S. and Malt, E. (2009) MINI International Neuropsychiatric Interview (1992-2009). [MINI Internasjonalt neuropsykiatrisk intervju, Norwegian Translation Version 6.0.0] Norway. Norway Mapi Research Institute, Lexington.
- [18] CADDRA (2011) Canadian ADHD Practice Guidelines, 3rd Edition, CADDRA, Toronto.
- [19] Aas, I.H. (2011) Guidelines for Rating Global Assessment of Functioning (GAF). *Annals of General Psychiatry*, **10**, 2. <https://doi.org/10.1186/1744-859X-10-2>
- [20] Kessler, R.C., *et al.* (2005) The World Health Organization Adult ADHD Self-Report Scale (ASRS): A Short Screening Scale for Use in the General Population. *Psychological Medicine*, **35**, 245-256. <https://doi.org/10.1017/S0033291704002892>
- [21] Pedersen, G., Hagtvet, K.A. and Karterud, S. (2007) Generalizability Studies of the Global Assessment of Functioning-Split Version. *Comprehensive Psychiatry*, **48**, 88-94. <https://doi.org/10.1016/j.comppsy.2006.03.008>
- [22] Derogatis, L.R. and Savitz, K.L. (1999) The SCL-90-R and Brief Symptom Inventory, and Matching Clinical Rating Scales. In: Maruish, M.E., Ed., *The Use of Psychological Testing for Treatment Planning and Outcomes Assessment*, 2nd Edition, Lawrence Erlbaum Associates Publishers, 1507.
- [23] Vassend, O. and Skrondal, A. (1999) The Problem of Structural Indeterminacy in Multidimensional Symptom Report Instruments. The Case of SCL-90-R. *Behavioural Research Therapy*, **37**, 685-701.
- [24] Cole, S.R. and Hernan, M.A. (2008) Constructing Inverse Probability Weights for

- Marginal Structural Models. *American Journal of Epidemiology*, **168**, 656-664. <https://doi.org/10.1093/aje/kwn164>
- [25] Rosenbaum, P.R. and Rubin, D.B. (1983) The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika*, **70**, 41-55. <https://doi.org/10.1093/biomet/70.1.41>
- [26] Ho, D.E., *et al.* (2007) Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference. *Political Analysis*, **15**, 199-236. <https://doi.org/10.1093/pan/mpl013>
- [27] Schafer, J.L. and Kang, J. (2008) Average Causal Effects from Nonrandomized Studies: A Practical Guide and Simulated Example. *Psychological Methods*, **13**, 279-313. <https://doi.org/10.1037/a0014268>
- [28] Harder, V.S. and Stuart, E.A. (2010) Propensity Score Techniques and the Assessment of Measured Covariate Balance to Test Causal Associations in Psychological Research. *Psychological Methods*, **15**, 234-249. <https://doi.org/10.1037/a0019623>
- [29] Vandecandelaere, M., *et al.* (2016) Time-Varying Treatments in Observational Studies: Marginal Structural Models of the Effects of Early Grade Retention on Math Achievement. *Multivariate Behavioural Research*, **51**, 843-864. <https://doi.org/10.1080/00273171.2016.1155146>
- [30] R Core Team (2014) A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna.
- [31] Fong, C.C., *et al.* (2016) CBPS: Covariate Balancing Propensity Score. R Package Version 0.13.
- [32] Zeileis, A.A. (2004) Econometric Computing with HC and HAC Covariance Matrix Estimators. *Journal of Statistical Software*, **11**, 1-17. <https://doi.org/10.18637/jss.v011.i10>
- [33] Wal, W.M.V.D. and Geskus, R.B. (2011) Ipw: An R Package for Inverse Probability Weighting. *Journal of Statistical Software*, **43**, 1-23. <https://doi.org/10.18637/jss.v043.i13>
- [34] Hansen, L.P. (1982) Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica*, **50**, 1029-1054. <https://doi.org/10.2307/1912775>
- [35] Azur, M.J., *et al.* (2011) Multiple Imputation by Chained Equations: What Is It and How Does It Work? *International Journal of Methods in Psychiatric Research*, **20**, 40-49. <https://doi.org/10.1002/mpr.329>
- [36] Buuren, S.V. and Groothuis-Oudshoorn, K. (2011) Mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, **45**, 1-67. <https://doi.org/10.18637/jss.v045.i03>
- [37] Rubin, D.B. (1987) Multiple Imputation for Nonresponse in Surveys. Wiley, New York. <https://doi.org/10.1002/9780470316696>

Appendix: Covariate Balance in Cross-Sectional and Longitudinal Data

First, the cross-sectional situation is considered; a sample of size n from a population, individuals indexed by $i=1, \dots, n$, one single dichotomous treatment indicator $T_i = t, t \in \{0, 1\}$ ($T_i = 1$ for treatment and 0 for no treatment), a K -dimensional vector of confounders X_i with influence on treatment assignment and outcome, $Y_i(t)$ is the counterfactual univariate outcome for hypothetical treatment assignment $T_i = t$ and observed outcome Y_i . The propensity score is (PS) assumed to satisfy $0 < Pr(T_i = 1 | X_i) = \pi(1, X_i) < 1$, for all X_i [25]. They showed that with conditional independence between counterfactual outcomes and observed treatment, conditional on covariates (no unmeasured confounding/ignorable treatment assignment assumption), this implies the same conditional independence by conditioning on the true PS

$$\{Y_i(1), Y_i(0)\} \perp\!\!\!\perp T_i | \pi(1, X_i) \tag{A1}$$

which means that unbiased estimation of the treatment effect is possible by conditioning on the PS alone. It also implies covariate balance between treatment groups, for example equal weighted mean in the treated and untreated. (A1) represents a dimension reduction and has led to development of propensity methods, like weighting. Inverse probability of treatment weighting [10] can consistently estimate the marginal mean of the counterfactual outcome for any treatment, and the method of which the marginal structural model is based. Each observation is weighted by the inverse of the probability of his/her observed treatment assignment,

$$w_i(t) = \pi(t, X_i)^{-1}, \quad t \in \{0, 1\} \tag{A2}$$

In observational studies, the PS has to be estimated, commonly by e.g. logistic regression and maximum likelihood, for example parameterized by

$$\pi_\alpha(T_i, X_i) = Pr(T_i | X_i) = \text{expit}((2T_i - 1)\alpha^T X_i) \tag{A3}$$

where $\text{expit}(z) = [1 + \exp(-z)]^{-1}$. If the model in (A3) is misspecified, the covariates are possibly unbalanced. To make estimation more robust for misspecification, I&R proposed to estimate the PS under an additional condition of covariate balance, formulated as

$$E \left\{ \frac{T_i X_i}{\pi_\alpha(1, X_i)} - \frac{(1 - T_i) X_i}{\pi_\alpha(0, X_i)} \right\} = 0 \tag{A4}$$

By iterated expectation it is easily seen that both terms in (A4) equals $E(X_i)$, the population mean, and also a weighted conditional mean of the treated/untreated, respectively. With (A3) as the parametric model for the PS (1-PS), the number of equations in (A4) (one for each of the K covariates), equals the number of unknown parameters, which is the just-identified case [1]. I & R suggested to solve these moment conditions, either by generalized method of moments (GMM) or empirical likelihood, because these methods easily generalize to the over-identified

set of equations, including the longitudinal case.

In the present application, the data is longitudinal with four time periods, $J = 4$. The time varying covariate at a given time period j depends possibly on the past treatment history until the previous time period ($j - 1$), written \bar{t}_{j-1} . It needs to be balanced on all current and future treatment trajectories, written $t_j = \{t_j, t_{j+1}, \dots, t_J\}$, and is conditional on the past history. The covariate balancing conditions are written

$$E\{X_{ij}(\bar{t}_{j-1})\} = E\{1(\bar{T}_{i,j-1} = \bar{t}_{j-1}, T_{ij} = t_j)w_i(\bar{t}_j, X_{ij}(\bar{t}_{j-1}))X_{ij}(\bar{t}_{j-1})\} \quad (A5)$$

and can be represented in an orthogonal way, in the following manner: Let the time varying covariates be combined in a $4K \times 1$ dimensional column vector $\tilde{X}_i = (w_i X_{i1}^T, w_i X_{i2}^T, w_i X_{i3}^T, w_i X_{i4}^T)^T$. To determine the sign of each term in the moment conditions, the following $2^J - 1$ dimensional column vector M_i is needed:

$$M_i^T = (-1^{T_{i1}}, -1^{T_{i2}}, -1^{T_{i1}+T_{i2}}, -1^{T_{i3}}, -1^{T_{i1}+T_{i3}}, -1^{T_{i2}+T_{i3}}, -1^{T_{i1}+T_{i3}+T_{i4}}, -1^{T_{i4}}, -1^{T_{i1}+T_{i4}}, -1^{T_{i2}+T_{i4}}, -1^{T_{i1}+T_{i2}+T_{i4}}, -1^{T_{i3}+T_{i4}}, -1^{T_{i1}+T_{i3}+T_{i4}}, -1^{T_{i2}+T_{i3}+T_{i4}}, -1^{T_{i1}+T_{i2}+T_{i3}+T_{i4}})$$

As described above, the moment conditions balance covariates measured at time j across all possible current and future treatments, but not past treatments and their interactions. Therefore, moment conditions on past treatments and their interactions are not binding, and can be zero'd out. As time progresses, the number of not binding moment conditions increase. The four different "selection matrix-es" to identify which conditions to zero out, are given by:

$$R_1 = I_{15}, R_2 = \begin{bmatrix} 0 & 0 \\ 0 & I_{14} \end{bmatrix}, R_3 = \begin{bmatrix} 0_{3 \times 3} & 0_{3 \times 12} \\ 0_{12 \times 3} & I_{12} \end{bmatrix}, R_4 = \begin{bmatrix} 0_{7 \times 7} & 0_{7 \times 8} \\ 0_{8 \times 7} & I_8 \end{bmatrix}$$

where I_d is the identity matrix of dimension $d \times d$.

The sample moment conditions for time-period j can then be written [2]:

$$G_j = \frac{1}{n} \sum_i [M_i^T R_j \otimes \tilde{X}_i] \quad (A6)$$

where \otimes is the Kronecker product (matrix on right-hand side is multiplied by each element in matrix on left-hand side), and with

$$\begin{aligned} G_1 &= \frac{1}{n} \sum_i (-1^{T_{i1}} \quad \dots \quad -1^{T_{i1}+T_{i2}+T_{i3}+T_{i4}}) I_{15} \otimes w_i X_{i1} \\ &= \frac{1}{n} \sum_i (-1^{T_{i1}} w_i X_{i1} \quad \dots \quad -1^{T_{i1}+T_{i2}+T_{i3}+T_{i4}} w_i X_{i1}) \\ G_2 &= \frac{1}{n} \sum_i (-1^{T_{i1}} \quad \dots \quad -1^{T_{i1}+T_{i2}+T_{i3}+T_{i4}}) \begin{bmatrix} 0 & 0 \\ 0 & I_{14} \end{bmatrix} \otimes w_i X_{i2} \\ &= \frac{1}{n} \sum_i (0 \quad -1^{T_{i2}} w_i X_{i2} \quad \dots \quad -1^{T_{i1}+T_{i2}+T_{i3}+T_{i4}} w_i X_{i2}) \\ G_3 &= \dots = \frac{1}{n} \sum_i (0 \quad 0 \quad 0 \quad -1^{T_{i3}} w_i X_{i3} \quad \dots \quad -1^{T_{i1}+T_{i2}+T_{i3}+T_{i4}} w_i X_{i3}) \end{aligned}$$

$$G_4 = \dots = \frac{1}{n} \sum_i \left(0 \quad -1^{T_{i4}} w_i X_{i4} \quad \dots \quad -1^{T_{i1}+T_{i2}+T_{i3}+T_{i4}} w_i X_{i4} \right)$$

and combining for all time-periods yields the matrix G with dimension $4K \times 15$

$$G = \begin{bmatrix} G_1 \\ G_2 \\ G_3 \\ G_4 \end{bmatrix} \tag{A7}$$

with corresponding covariance matrix (dimension $60K \times 60K$)

$$W = \frac{1}{n} \sum_i E \left(M_i M_i^T \otimes \tilde{X}_i \tilde{X}_i^T \mid X_i \right) \tag{A8}$$

where

$$\tilde{X}_i \tilde{X}_i^T = \begin{bmatrix} w_i^2 X_{i1} X_{i1}^T & w_i^2 X_{i1} X_{i2}^T & w_i^2 X_{i1} X_{i3}^T & w_i^2 X_{i1} X_{i4}^T \\ w_i^2 X_{i2} X_{i1}^T & w_i^2 X_{i2} X_{i2}^T & w_i^2 X_{i2} X_{i3}^T & w_i^2 X_{i2} X_{i4}^T \\ w_i^2 X_{i3} X_{i1}^T & w_i^2 X_{i3} X_{i2}^T & w_i^2 X_{i3} X_{i3}^T & w_i^2 X_{i3} X_{i4}^T \\ w_i^2 X_{i4} X_{i1}^T & w_i^2 X_{i4} X_{i2}^T & w_i^2 X_{i4} X_{i3}^T & w_i^2 X_{i4} X_{i4}^T \end{bmatrix}$$

Since each moment condition set equal to zero, like in (A4) (for time period 1 there are 15 conditions), has the same number of equations as the number of unknown parameters, the set of equations is over-identified, and there is no unique solution. Instead a quadratic function of the moment conditions is minimized to come as close as possible to zero, to “minimize imbalance” and this is achieved by the GMM estimator [34]. The optimal GMM estimator for α is given by [2]

$$\hat{\alpha} = \arg \min_{\alpha} \text{vec}(G)^T W^{-1} \text{vec}(G) \tag{A9}$$

where G is from (A7) and the covariance from (A8), where the expectation can be calculated analytically in the logistic regression case [1]. These estimates are used in the construction of the weights for the MSM to give estimates of the causal effects from a marginal weighted outcome regression model (Equation (2)).