Impact of near-positivity violations on IPTW-estimated marginal structural survival models with time-dependent confounding

Spreafico, M.

Biometrical Journal

WILEY-VCH

Biometrical Journal

RESEARCH ARTICLE OPEN ACCESS

Impact of Near-Positivity Violations on IPTW-Estimated Marginal Structural Survival Models With Time-Dependent Confounding

Marta Spreafico[1,2]

[1]Mathematical Institute, Leiden University, Leiden, The Netherlands | [2]Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, The Netherlands

Correspondence: Marta Spreafico (m.spreafico@math.leidenuniv.nl)

Received: 10 January 2025 | **Revised:** 27 June 2025 | **Accepted:** 14 August 2025

Keywords: inverse probability of treatment weighting | marginal structural models | positivity assumption | simulation studies | survival outcomes

ABSTRACT

In longitudinal observational studies, marginal structural models (MSMs) are used to analyze the causal effect of an exposure on the (time-to-event) outcome of interest, while accounting for exposure-affected time-dependent confounding. In the applied literature, inverse probability of treatment weighting (IPTW) has been widely adopted to estimate MSMs. An essential assumption for IPTW-based MSMs is *positivity*, which requires that, for any combination of measured confounders among individuals, there is a nonzero probability of receiving each treatment strategy. Positivity is crucial for valid causal inference through IPTW-based MSMs, but is often overlooked compared to confounding bias. Near-positivity violations, where certain treatments are theoretically possible but rarely observed due to randomness, are common in practical applications, particularly when the sample size is small, and they pose significant challenges for causal inference. This study investigates the impact of near-positivity violations on estimates from IPTW-based MSMs in survival analysis. Two algorithms are proposed for simulating longitudinal data from hazard-MSMs, accommodating near-positivity violations, a time-varying binary exposure, and a time-to-event outcome. Cases of near-positivity violations, where remaining unexposed is rare within certain confounder levels, are analyzed across various scenarios and weight truncation (WT) strategies. Through comprehensive simulations, this study shows that even minor near-positivity violations in longitudinal survival analyses can substantially destabilize IPTW-based estimators, inflating variance and bias, especially under aggressive WT. This work aims to serve as a critical warning against overlooking the positivity assumption or naively applying WT in causal studies using longitudinal observational data and IPTW.

1 | Introduction

In longitudinal observational studies, exposure-affected time-varying confounding represents a major challenge for estimating the effect of a treatment on the (time-to-event) outcome of interest, as standard analyses fail to give consistent estimators (Clare et al. 2019; Daniel et al. 2013). Over the past decades, considerable progress has been made in developing causal inference methods tailored to such complex data. Among these, marginal structural models (MSMs) offer a rigorous modeling approach for estimating and summarizing longitudinal causal effects in the presence of time-dependent confounding (Daniel et al. 2013; Hernán et al. 2000; Hernán and Robins 2020; Robins et al. 2000; Williamson and Ravani 2017). MSMs are particularly useful when multiple time-varying treatment regimens are possible, as they help synthesize complex treatment histories into interpretable summaries, making them highly valuable for applied work. MSMs are models for the *potential* or *counterfactual*

outcome that individuals would have experienced if they had received a particular treatment or exposure value. This study focuses on counterfactual time-to-event outcomes by considering marginal structural hazard models (hazard-MSM) or a discrete-time analogue. The parameters of an MSM can be consistently estimated through various methods, including Inverse Probability of Treatment Weighting (IPTW) estimators, G-computation, or doubly robust methods (Clare et al. 2019; Daniel et al. 2013; Hernán and Robins 2020; Gabriel et al. 2024; Robins et al. 2000; van der Laan and Gruber 2016). Despite being less robust, IPTW-based MSMs have largely been adopted in the applied literature, especially in epidemiology and medicine, due to their simplicity in both implementation and interpretation (Clare et al. 2019). IPTW-based MSMs require the correct specification of the exposure model conditional on confounders (i.e., the *weighting model*) and special attention to the identifiability assumptions of consistency, no unmeasured confounding, and positivity (Cole and Frangakis 2009; Cole and Hernán 2008; Hernán and Robins 2020; Williamson and Ravani 2017). This work focuses on the latter, which is often overlooked compared to confounding bias.

Positivity holds if, for any combination of the measured confounders occurring among individuals in the population, there is a nonzero (i.e., positive) probability of receiving every level of the exposure possible under the target treatment strategies to be compared (Cole and Hernán 2008; Hernán and Robins 2020). While less well-recognized than bias due to incomplete control of confounding, violations of the positivity assumption can increase both the variance and bias of causal effect estimates (Léger et al. 2022; Petersen et al. 2012). Positivity violations can occur in two situations (Y. Zhu et al. 2021). *Strict* (or *theoretical*) violations occur when certain treatment levels are impossible for specific subgroups of subjects. For example, if a certain treatment a is never given to individuals with severe comorbidities, then the causal effect of a cannot be estimated for that subgroup; the analysis should therefore focus only on individuals without severe comorbidities. Even in the absence of structural zeros, *random* zeros may occur by chance due to small sample sizes or highly stratified data by numerous confounders. *Near* (or *practical*) violations refer to situations where the assignment to a specific treatment is always theoretically possible but is not (or rarely) observed in the data due to randomness. Sampling variability may indeed result in subjects having a near-zero probability of being exposed (or unexposed) for certain combinations of covariate values. These situations are common in practical applications, particularly when the sample size is small, and they pose significant challenges for causal inference. While treatment remains technically possible within a subgroup, its rarity makes reliable estimation difficult, particularly when using IPTW. Empirical studies across various clinical areas have reported evidence of near-positivity violations, often indicated by extremely large inverse probability weights resulting from very small treatment probabilities in some covariate strata. Examples include studies investigating the effect of methotrexate on mortality in rheumatoid arthritis patients (Fewell et al. 2004), the impact of initiating highly active antiretroviral therapy on changes in HIV-1 RNA viral load in HIV-infected individuals (Cole and Hernán 2008), the effect of the anticoagulant warfarin on the risk of gastrointestinal bleeding (Platt et al. 2012), and the impact of metformin on colon cancer recurrence among diabetic patients (Y. Zhu et al. 2021). Such extreme weights can destabilize the analysis and inflate variance. To address these challenges and stabilize the variance of estimates, weight truncation (WT) is often applied in IPTW to down-weight observations in regions where near-violations occur (Cole and Hernán 2008; Xiao et al. 2013; Y. Zhu et al. 2021). However, if applied inappropriately, this technique could result in excessive truncation, which may introduce bias into the estimates.

In the literature, research studies on positivity violations have been previously carried out in a pedagogical manner by using real data to illustrate how incorrect inference occurs in estimating MSMs when positivity is violated (Bembom and van der Laan 2007; Cole and Hernán 2008; Mortimer et al. 2005; Rudolph et al. 2022; Y. Zhu et al. 2021; A. Zhu et al. 2023). Findings across different studies agreed that positivity violations have a more severe impact on the IPTW estimator than other causal estimators. However, when using real data, disentangling the effect of positivity violations from other sources of bias is typically not possible: important confounders may be undetected or unmeasured and the fulfillment of the remaining assumptions underlying the IPTW estimator is generally difficult to ascertain. Moreover, real data do not allow us to design scenarios that could be of interest, such as studying performance under different sample sizes. To overcome these limitations, other investigations were conducted more systematically by setting up simulation studies (Léger et al. 2022; Neugebauer and van der Laan 2005; Naimi et al. 2011; Petersen et al. 2012; Wang et al. 2006). Results confirmed that under positivity violations IPTW estimator performs worse than other methods, becoming very unstable and exhibiting high variability. However, these studies were limited to assessing the causal effect of a treatment assigned either at a single time point or twice.

This study investigates the impact of near-positivity violations on the performance of IPTW-estimated MSMs in longitudinal survival contexts with time-varying confounding using a simulation-based approach. No systematic simulation studies currently exist in this framework, largely due to the challenges of simulating longitudinal survival data under conditions of both exposure-affected time-varying confounding and near-positivity violations. To address this gap, two algorithms are proposed for simulating longitudinal data from hazard-MSMs, accommodating near-positivity violations, a time-varying binary exposure, and a survival outcome. These methods build on the works of Havercroft and Didelez (2012) and Keogh et al. (2021). Two simulation studies analyzing cases of near-positivity violations, where remaining unexposed is rare within certain levels of confounders, through various scenarios and WT strategies are performed.

This study aims to highlight the critical importance of carefully considering the positivity assumption in causal studies that utilize longitudinal observational data and IPTW. The final purpose is to warn against the risks of underestimating the assumption's significance or uncritically applying WT methods. This is fundamental given that the past several decades have seen an exponential growth in causal inference approaches and their applications to observational data (Hammerton and Munafò 2021; Mitra et al. 2022; Olier et al. 2023), including emerging areas such as target trial emulations (Hernán and Robins 2016; Hernán 2021), prediction modeling under hypothetical interventions (Keogh and van

Geloven 2024; Lin et al. 2021; van Geloven et al. 2020), or causal machine learning (Feuerriegel et al. 2024; Moccia et al. 2024).

This study is organized as follows: Section 2 briefly recalls the notation, MSMs for survival outcomes, and IPTW. Section 3 explains the proposed mechanism to enforce positivity violations in algorithms to simulate longitudinal data from MSMs. Sections 4 and 5 present the two simulation studies. Sections 6 and 7 finally provide a set of practical recommendations and discuss the findings. Statistical analyses were performed in the R software environment (R Core Team 2023). Source code is available at https://github.com/mspreafico/PosViolMSM. A vignette illustrating how to use the developed code and algorithms in practice is provided in the Supporting Information.

2 | MSMs for Potential Survival Outcomes

2.1 | Notation

Let us consider a set of $i = 1, \dots, n$ subjects and a set of regular visits $k = 0, 1, \dots, K$ performed at times $q_0 < q_1 < \dots < q_K$ (assumed to be the same for everybody). Each subject undergoes each visit up until event time $T_i^* = \min(C_i, T_i)$, that is, the earlier of the time of the actual event of interest T_i and the administrative censoring time C_i. At each visit k, if $T_i \geq q_k$, a binary treatment status $A_{i,k} \in \{0, 1\}$ (unexposed vs. exposed; control vs. treatment) and a set of time-dependent covariates $L_{i,k}$ are observed. A bar over a time-dependent variable indicates the history, that is, $\bar{\boldsymbol{A}}_{i,k} = \left(A_{i,0}, \dots, A_{i,k}\right)$ and $\bar{\boldsymbol{L}}_{i,k} = \left(L_{i,0}, \dots, L_{i,k}\right)$. Finally, the binary failure indicator process is denoted by $Y_{i,k+1}$, where $Y_{i,k+1} = 1$ if subject i has failed (e.g., died) in period $(q_k; q_{k+1}]$, that is, $q_k < T_i \leq q_{k+1}$, or $Y_{i,k+1} = 0$ otherwise.

2.2 | MSMs for Counterfactual Hazard Rates

Marginal structural hazards models (hazard-MSMs) are a class of causal models that focus on *counterfactual* time-to-event variables (Hernán and Robins 2020; Hernán et al. 2000; Robins et al. 2000). These variables represent the time at which an event would have been observed had a patient been administered a specific exposure strategy $\bar{\boldsymbol{a}} = (a_0, a_1, \dots, a_K)$ with $a_k \in \{0, 1\}$ for all k. Vector $\bar{\boldsymbol{a}}$ might differ from the actual treatment received $\bar{\boldsymbol{A}}_i = \bar{\boldsymbol{A}}_{i,K} = \left(A_{i,0}, \dots, A_{i,K}\right)$. The *counterfactual event time* that would be observed in a subject under complete exposure history $\bar{\boldsymbol{a}}$ is denoted by $T^{\bar{\boldsymbol{a}}}$. Hazard-MSMs hence model the counterfactual hazard rate:

$$\lambda^{\bar{\boldsymbol{a}}}(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T^{\bar{\boldsymbol{a}}} < t + \Delta t \mid T^{\bar{\boldsymbol{a}}} \geq t)}{\Delta t}.$$

In case of discrete-time hazard of failure, a *marginal structural logistic regression model* (*logit-MSM*) can be assumed to model the counterfactual probability of failure in a single interval $(q_k, q_{k+1}]$, given survival up to q_k. The *logit-MSM* for the counterfactual hazard at visit k is defined as follows:

$$\lambda_k^{\bar{\boldsymbol{a}}} = \Pr\left(Y_{k+1}^{\bar{\boldsymbol{a}}} = 1 \mid Y_k^{\bar{\boldsymbol{a}}} = 0\right) = \text{logit}^{-1}[\tilde{\gamma}_0 + g(\tilde{\boldsymbol{\gamma}}_A; \bar{\boldsymbol{a}}_k)], \quad (1)$$

where $\bar{\boldsymbol{a}}$ is the complete treatment strategy, $Y_k^{\bar{\boldsymbol{a}}}$ is the counterfactual event indicator at visit k, $g(\cdot)$ is a function (to be specified) of the treatment strategy history up to visit k (denoted by $\bar{\boldsymbol{a}}_k$), and $(\tilde{\gamma}_0, \tilde{\boldsymbol{\gamma}}_A)$ is the vector of log odds ratios, with $\tilde{\gamma}_0$ as the intercept. Depending on the desired information provided in $g(\cdot)$, the hazard at visit k can thus assume different forms (see Appendix A.1).

In the context of continuous-time hazard, a *marginal structural Aalen's additive hazard model* (*Aalen-MSM*) can be assumed to model the counterfactual hazard at time t given treatment history $\bar{\boldsymbol{a}}$:

$$\lambda^{\bar{\boldsymbol{a}}}(t) = \tilde{\alpha}_0(t) + g\left(\tilde{\boldsymbol{\alpha}}_A(t); \bar{\boldsymbol{a}}_{\lfloor t \rfloor}\right) \quad (2)$$

where $\tilde{\alpha}_0(t)$ is the baseline hazard at time t, $\bar{\boldsymbol{a}}_{\lfloor t \rfloor}$ denotes treatment pattern up to the most recent visit prior to time t (i.e., $\lfloor t \rfloor = \max_{k \leq t} k$), $g(\cdot)$ is a function (to be specified) of treatment pattern $\bar{\boldsymbol{a}}_{\lfloor t \rfloor}$, and $\tilde{\boldsymbol{\alpha}}_A(t)$ is the vector of coefficients at time t. Depending on the desired information provided in $g(\cdot)$, the hazard at time t can thus assume different forms (see Appendix A.1).

2.2.1 | Hazard-Based Estimands and Marginal Survival Probabilities

The logit-MSM (1) estimates log odds ratios $\tilde{\boldsymbol{\gamma}}_A$ and the Aalen-MSM (2) the cumulative regression coefficients $\int_0^t \tilde{\boldsymbol{\alpha}}_A(s)ds$. Since hazard-based estimands may not have a straightforward interpretation, estimates from the MSMs are typically transformed into estimates for an interpretable causal estimand (Didelez and Stensrud 2022; Hernán 2010; Keogh et al. 2021; Martinussen et al. 2020). One example is comparing the marginal survival probabilities at time t, that is, $S^{\bar{\boldsymbol{a}}}(t) = \Pr(T^{\bar{\boldsymbol{a}}} > t)$, for *always treated* $\bar{\boldsymbol{a}} = \mathbf{1} = (1, \dots, 1)$ (i.e., sustained use of the treatment) versus *never treated* $\bar{\boldsymbol{a}} = \mathbf{0} = (0, \dots, 0)$ (i.e., sustained nonuse of the treatment), or evaluating the marginal risk difference between them. The marginal survival probability at time t under treatment history $\bar{\boldsymbol{a}}$ can be computed based on the different hazard forms:

i. for the logit-MSMs in (1) is given by

$$S^{\bar{\boldsymbol{a}}}(t) = \prod_{k \leq t} \left(1 - \lambda_k^{\bar{\boldsymbol{a}}}\right) = \prod_{k \leq t} \left(1 - \text{logit}^{-1}[\tilde{\gamma}_0 + g(\tilde{\boldsymbol{\gamma}}_A; \bar{\boldsymbol{a}}_k)]\right); \quad (3)$$

ii. for the Aalen-MSM in (2) is given by

$$S^{\bar{\boldsymbol{a}}}(t) = \exp\left(- \int_0^t \tilde{\alpha}_0(s)\, ds - \int_0^1 g(\tilde{\boldsymbol{\alpha}}_A(s); a_0)\, ds - \int_1^2 g(\tilde{\boldsymbol{\alpha}}_A(s); \bar{\boldsymbol{a}}_1)\, ds - \dots - \int_{\lfloor t \rfloor}^t g(\tilde{\boldsymbol{\alpha}}_A(s); \bar{\boldsymbol{a}}_{\lfloor t \rfloor})\, ds \right). \quad (4)$$

2.3 | Inverse Probability of Treatment Weighting

In the presence of confounders, and assuming there are no unmeasured confounders, MSMs can be estimated from the observed data by applying a technique called IPTW (Hernán and

Robins 2020). IPTW involves weighting the contribution of each subject i by the inverse of the probability of receiving their actual exposure level given their confounding covariates. This process creates a pseudo-population where the effects of time-dependent confounding are balanced, so association in hazard regression models is causation (Hernán and Robins 2020). To optimize the variance estimation, stabilized (or standardized) weights are usually preferred (Hernán and Robins 2020; Hernán et al. 2000; Léger et al. 2022; Robins et al. 2000). The stabilized weight for subject i at time t is defined as

$$sw_i(t) = \prod_{k=0}^{\lfloor t \rfloor} \frac{\Pr\left(A_{i,k} \mid \bar{\boldsymbol{A}}_{i,k-1}, T_i \geq q_k\right)}{\Pr\left(A_{i,k} \mid \bar{\boldsymbol{A}}_{i,k-1}, \bar{\boldsymbol{L}}_{i,k}, T_i \geq q_k\right)}, \tag{5}$$

where $\lfloor t \rfloor = \max_{k \leq t} k$ is the largest visit-time prior to t, and A_{-1} is defined to be 0. These weights are well-defined only when the denominator probabilities are strictly greater than zero, that is, for each visit k, if $\Pr\left(\bar{\boldsymbol{A}}_{i,k-1} = \bar{\boldsymbol{a}}_{k-1}, \bar{\boldsymbol{L}}_{i,k} = \bar{\boldsymbol{l}}_k, T_i \geq q_k\right) \neq 0$, then

$$\Pr\left(A_{i,k} = a \mid \bar{\boldsymbol{L}}_{i,k} = \bar{\boldsymbol{l}}_k, \bar{\boldsymbol{A}}_{i,k-1} = \bar{\boldsymbol{a}}_{k-1}, T_i \geq q_k\right) > 0$$

$$\text{for all } a \in \{0, 1\}.$$

This means that, at each visit k, there is a nonzero (i.e., positive) probability of receiving every level of exposure $A_{i,k}$ for every combination of values of exposure and covariate histories $\bar{\boldsymbol{A}}_{i,k-1}$ and $\bar{\boldsymbol{L}}_{i,k}$ that occur among at-risk individuals $(T_i \geq q_k)$ in the population, which is what the positivity assumption guarantees (Cole and Hernán 2008; Hernán and Robins 2020).

Even when standardized, weights $sw_i(t)$ can largely inflate for a subject i concerned by near-positivity violation: when the denominator probabilities are very close to zero, weights become extremely large. In such cases, the common approach is to consider truncated stabilized weights $s\tilde{w}_i(t)$ obtained by truncating the lowest and the highest estimations by the first and 99th (1–99) percentiles, or alternatively by narrower truncations, such as the 2.5–97.5, 5–95, or 10–90 percentiles (Cole and Hernán 2008; Xiao et al. 2013; Y. Zhu et al. 2021).

2.4 | Simulating Longitudinal Survival Data From Marginal Structural Hazard Models

Even when positivity holds, simulating longitudinal data when the model of interest is a model for potential outcomes, as for MSMs, is generally not straightforward (Evans and Didelez 2024). The main challenges consist of (i) replicating the complex dynamics of time-varying confounding; (ii) generating data in such a way that the model of interest is correctly specified; and (iii) in case of survival or other non-collapsible models (Didelez and Stensrud 2022; Robinson and Jewell 1991), reconciling the MSM with the conditional model used in Monte Carlo studies to generate the data. For these reasons, only a few methods for simulating data from hazard-MSMs have been published in the literature (Evans and Didelez 2024; Havercroft and Didelez 2012; Keogh et al. 2021; Seaman and Keogh 2024; Young et al. 2010; Young and Tchetgen Tchetgen 2014; Xiao et al. 2010). These methods impose restrictions on the data-generating mechanisms to address the issues mentioned, allowing for accurate simulation of longitudinal data from prespecified hazard-MSMs.

Among these, the approaches by Havercroft and Didelez (2012) and Keogh et al. (2021) present similar structures of the temporal causal relationships between variables. The directed acyclic graphs (DAGs) in the top panels of Figure 1 display for both cases the assumed data structure and inform which variables, measured at which time points, are confounders of the association between treatment at a given time point and the outcome. Both DAGs are illustrated in discrete time for a follow-up with $k = 0, \ldots, K$ visits. Variables are assumed to be constant between visits. By imagining splitting the time intervals between successive visits into smaller and smaller intervals, both structures approach the continuous-time setting. For the current study, these two approaches are used as "truth" benchmarks for cases where longitudinal data generation from the desired hazard-MSM has already been demonstrated and the positivity assumption is valid.

3 | Simulating Longitudinal Survival Data With Random Positivity Violations

To impose near-positivity violations within a data-generating mechanism, certain treatment levels (e.g., exposure or non-exposure in binary treatments) may become unobservable (though theoretically possible) for specific subgroups defined by confounders, due to randomness. Suppose the interest is in the subgroup of subjects presenting a poor health condition. Near-positivity violations occur when the probability of remaining unexposed (or being exposed), given a poor health condition, is very close to zero (or approaches one). This happens when remaining unexposed to treatment at visit k is rarely observable for subjects in the poor health subgroup.

Let us define the subgroup of subjects presenting a poor health condition at visit k as determined by a range of values $\mathcal{I}_\tau$ of the confounder $L_{i,k}$. Violations occurring by chance can be introduced in a data-generating mechanism by considering (i) a latent individual propensity P_i to exposure given a poor health condition, and (ii) an exposure cutoff $\pi \in [0, 1]$. For each subject i, a random uniform variable P_i is generated on the interval [0,1] and treatment assignment may be modified according to the exposure cutoff π. At each visit k, subjects in poor health with $P_i < \pi$ have a positive probability of either receiving exposure or remaining unexposed; in contrast, subjects in poor health with $P_i \geq \pi$ are deterministically assigned to exposure. This results in

$$\Pr\left(A_{i,k} = 1 \mid L_{i,k} \in \mathcal{I}_\tau, \bar{\boldsymbol{L}}_{i,k-1}, \bar{\boldsymbol{A}}_{i,k-1}, T_i \geq q_k\right)$$

$$= \begin{cases} 1 & \text{if } P_i \geq \pi \\ p^A_{i,k} \in (0,1) & \text{if } P_i < \pi, \end{cases} \tag{6}$$

when $\Pr\left(\bar{\boldsymbol{A}}_{i,k-1}, \bar{\boldsymbol{L}}_{i,k-1}, L_{i,k} \in \mathcal{I}_\tau, T_i \geq q_k\right) \neq 0$. This leads to near-violations of the positivity assumption because, within the poor health subgroup defined by individuals i with $L_{i,k} \in \mathcal{I}_\tau$, both exposure and unexposure are theoretically possible. However, non-exposure may be rarely observed due to the randomness in drawing the individual propensities for exposure, creating situations where noncompliance with positivity occurs more frequently than expected, especially for low values of π.

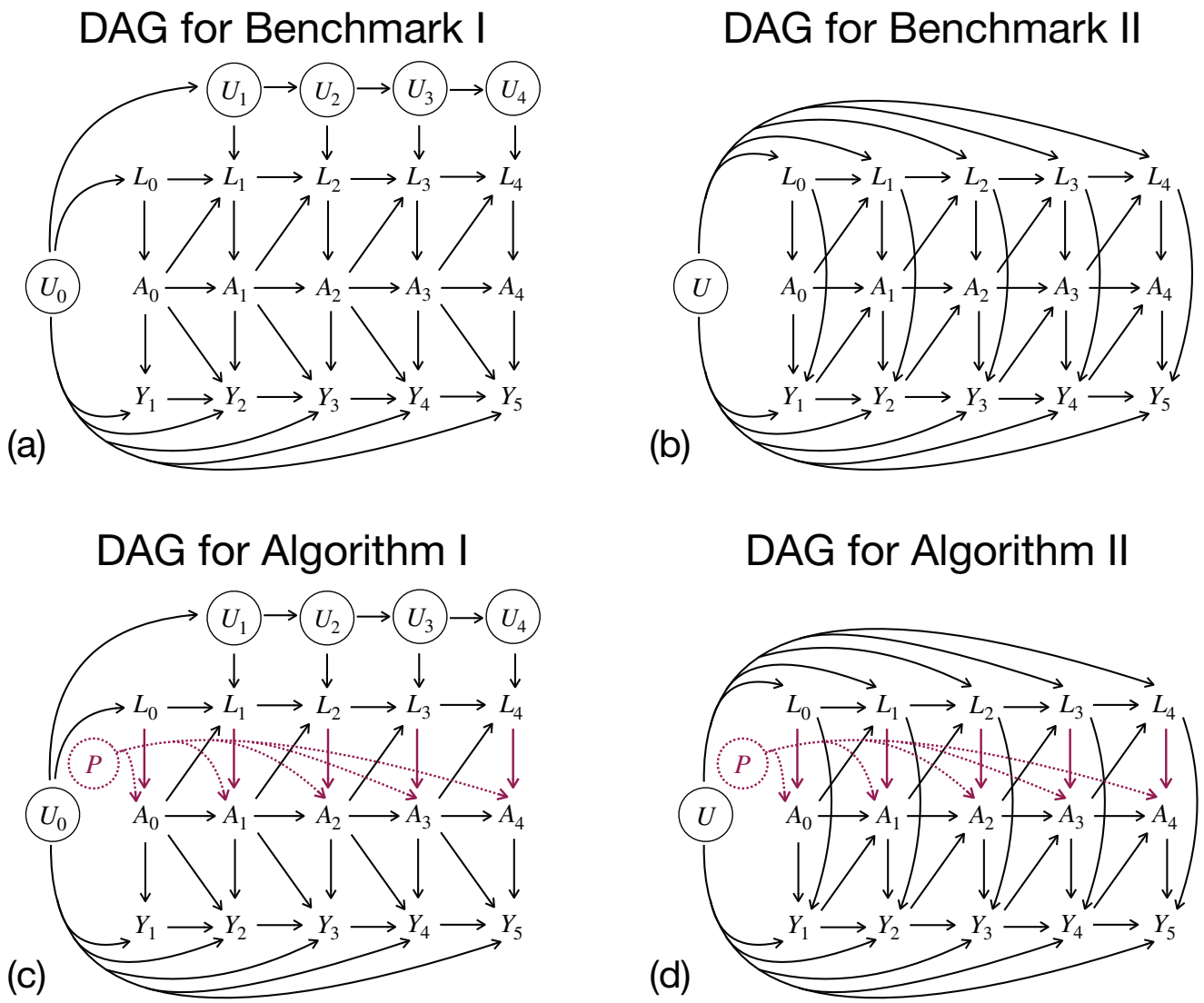

FIGURE 1 | Top: Directed acyclic graphs (DAGs) illustrating the temporal causal relationships between variables in the data-generating mechanisms proposed by (a) Havercroft and Didelez (2012) and (b) Keogh et al. (2021), that is, Benchmarks I and II. Bottom: DAGs illustrating the temporal causal relationships between variables in the proposed data-generating mechanisms, that is, (c) Algorithm 2 and (d) Algorithm 3.

The parameter π represents the expected proportion of subjects for whom both exposure and non-exposure are observable within each subgroup defined by the measured confounder, reflecting the *expected positivity support proportion*. Specifically,

- when $\pi = 1$, all subjects have a nondeterministic probability $p^A_{i,k}$ of receiving treatment, ensuring that the positivity assumption is satisfied;
- when $\pi = 0$, all subjects in poor health are deterministically assigned to receive treatment, representing a strict violation of the positivity assumption within the poor health group;
- when $0 < \pi < 1$, approximately $(1 - \pi) \times 100\%$ of subjects are expected to be deterministically assigned to exposure when they are in poor health. If unexposure is rarely observed within the remaining π proportion of subjects, the positivity assumption may be nearly violated within the poor health group due to limited support for the unexposed condition.

In other words, $(1 - \pi)$ can be interpreted as the expected proportion of subjects who deterministically contribute to violations of the positivity assumption within the poor health subgroup, while the remaining π proportion may contribute to near-violations if the unexposed condition is rarely observed among them. Therefore, for a fixed $\mathcal{I}_\tau$, the higher the cutoff π, the less severe the violation.

Given an algorithm to simulate longitudinal survival data from MSMs in the presence of time-varying confounding, near-positivity violations can be incorporated by using the pseudocode structure in panel Algorithm 1. The main advantage of imposing near-positivity violations in an existing approach, where longitudinal data generation from the desired hazard-MSM has already been confirmed, is the ability to directly examine the impact on IPTW estimators solely attributable to positivity violations, rather than other sources of bias. In this way, the original data-generating mechanism can be considered as the "truth" or benchmark case where the positivity assumption holds.

ALGORITHM 1 | General pseudocode for random positivity violation.

Initialize parameters: $(\mathcal{I}_\tau, \pi, \dots)$
for $i = 1, \dots, n$ **do**
  …
  $P_i \sim \mathcal{U}(0,1)$ ▷ Draw the individual propensity
  for $k = 0, \dots, K$ **do**
    $L_{i,k}$ is assigned based on the generating algorithm
    if $P_i \geq \pi$ and $L_{i,k} \in \mathcal{I}_\tau$ **then**
      exposure is assigned deterministically: $A_{i,k} = 1$
    else
      exposure $A_{i,k}$ is assigned stochastically with $p^A_{i,k} \in (0,1)$ based on the generating algorithm
    end if
    …
  end for
end for

4 | Simulation Study I

4.1 | Data Generation

The first algorithm proposed in this work is based on the data-generating mechanism introduced by Havercroft and Didelez (2012) to simulate from a discrete-time logit-MSM. This mechanism is now briefly introduced and then extended by imposing positivity violations.

4.1.1 | Benchmark I in a Nutshell

Building upon the DAG in Figure 1a, Havercroft and Didelez (2012) proposed an algorithm to emulate longitudinal data from the Swiss HIV Cohort Study (Sterne et al. 2005). The authors considered a discrete-time setting where visit times correspond to visit numbers, that is, $q_k = k$ for all $k = 0, \dots, K$. The time-dependent binary treatment process $A_{i,k}$ represents exposure to the highly active antiretroviral therapy (HAART) versus no treatment (unexposure). Once HAART has started for a subject i, it continues until failure or the end of the follow-up period. The only measured time-dependent confounder $L_{i,k}$ is the nonnegative CD4 cell count, measured in cells/μL. Variable $U_{i,k}$ represents the individual general latent health process, indicating a poor individual health status at visit k for values close to 0, or good health conditions for values close to 1. The latent process $U_{i,k}$ and the survival process $Y_{i,k+1} \in \{0,1\}$ are updated at each time point k, whereas CD4 cell count $L_{i,k} \geq 0$ and HAART exposure $A_{i,k} \in \{0,1\}$ are updated every κth time point, for a chosen κ, named checkup visits. Specifically, $U_{i,k}$, $L_{i,k}$, and $A_{i,k}$ are generated based on the latent general health process at visit $k = 0$, $U_{i,0}$, which is also transformed to obtain $Y_{i,k+1}$, using the desired MSM survival function, once the actual treatment history is known.

Despite there being no direct arrow from $L_{i,k}$ to $Y_{i,k+1}$, the DAG (Figure 1a) exhibits time-dependent confounding due to $U_{i,0}$ being a common ancestor of $\bar{\boldsymbol{A}}_i$ via $\bar{\boldsymbol{L}}_i$ and $\bar{\boldsymbol{Y}}_i$. Moreover, $A_{i,k}$ is independent from $\bar{\boldsymbol{U}}_{i,k}$ given $\left(\bar{\boldsymbol{L}}_{i,k}, \bar{\boldsymbol{A}}_{i,k-1}\right)$ and the vector $\bar{\boldsymbol{L}}_i$ is sufficient to adjust for confounding. Based on this mechanism, the authors proposed an algorithm to correctly simulate data from the following discrete-time logit-MSM:

$$\begin{aligned}\lambda_k^{\bar{a}} &= \text{logit}^{-1}\left[\tilde{\gamma}_0 + \tilde{\gamma}_{A1} \cdot \{(1-a_k)k + a_k k^*\}\right. \\ &\quad \left. + \tilde{\gamma}_{A2} \cdot a_k + \tilde{\gamma}_{A3} \cdot a_k(k-k^*)\right] \\ &= \text{logit}^{-1}[\tilde{\gamma}_0 + \tilde{\gamma}_{A1} \cdot d_{1k} + \tilde{\gamma}_{A2} \cdot a_k + \tilde{\gamma}_{A3} \cdot d_{3k}], \end{aligned} \tag{7}$$

where a_k is the binary treatment strategy at time k, k^* is the treatment initiation time, $d_{1k} = \min\{k, k^*\}$ and $d_{3k} = \max\{k-k^*, 0\}$ represent the time elapsed before and after treatment initiation, respectively. Note that $g(\tilde{\boldsymbol{\gamma}}_A; \bar{\boldsymbol{a}}_k) = \tilde{\gamma}_{A1} \cdot d_{1k} + \tilde{\gamma}_{A2} \cdot a_k + \tilde{\gamma}_{A3} \cdot d_{3k}$ in (7), reflecting that the hazard-MSM depends on a summary of the treatment history rather than only on the current treatment. In particular, Havercroft and Didelez proved that the parameters $(\tilde{\gamma}_0, \tilde{\gamma}_{A1}, \tilde{\gamma}_{A2}, \tilde{\gamma}_{A3})$ in the desired logit-MSM (7) are collapsible with the conditional distribution parameters $(\gamma_0, \gamma_{A1}, \gamma_{A2}, \gamma_{A3})$ in the following conditional logit model:

$$\begin{aligned}\lambda_{i,k} &= \text{logit}^{-1}\left[\gamma_0 + \gamma_{A1} \cdot \left\{(1-A_{i,k})k + A_{i,k}K_i^*\right\}\right. \\ &\quad \left. + \gamma_{A2} \cdot A_{i,k} + \gamma_{A3} \cdot A_{i,k}(k-K_i^*)\right], \end{aligned} \tag{8}$$

where K_i^* is the individual treatment initiation time and $\lambda_{i,k}$ represents the individual probability of failure in the interval $k < t \leq k+1$, conditional on survival up to visit k.

4.1.2 | Algorithm 2: Imposing Random Positivity Violations in Benchmark I

As illustrated in the DAG in Figure 1c, the first proposed algorithm to account for potential near-positivity violations builds upon Benchmark I (Figure 1a) by incorporating two additional components.

i. First, a poor health subgroup identified by $\mathcal{I}_\tau$ and acting on the purple path $L_{i,k} \to A_{i,k}$ must be defined. Since a CD4

count below 500 cells/μL indicates the patients' immune system may be weakened, making them susceptible to developing serious infections from viruses, bacteria, or fungi that typically do not cause problems in healthy individuals, it is reasonable to assume that subjects in a poor health condition at visit k are identified by $L_{i,k} < \tau$, where $\tau \in [0; 500]$ cells/μL. This is equivalent to a nonnegative CD4 range of the form $\mathcal{I}_\tau = [0, \tau)$, where the upper threshold τ has to be defined according to the simulation scenario. The higher the upper threshold τ, the wider $\mathcal{I}_\tau$ and the more severe the violations.

ii. Then, the latent individual propensity for exposure $P_i \sim \mathcal{U}(0,1)$ directly acts on $A_{i,k}$. Subjects in poor health condition with propensity P_i above the *exposure cutoff* π (to be defined according to the simulation scenario) are forced to start the treatment.

The procedure proposed below extends the algorithm by Havercroft and Didelez (2012) by incorporating the possibility of near-positivity violations. For details regarding the chosen parameter values, please refer to their primary work.

Procedure For each subject $i = 1, \ldots, n$, the simulation procedure with K discrete time points and checkups every κth visit is as follows:

1. Generate the individual propensity to exposure: $P_i \sim \mathcal{U}(0,1)$.
2. Generate the general latent health status at baseline: $U_{i,0} \sim \mathcal{U}(0,1)$.
3. Generate the baseline CD4 as a transformation of $U_{i,0}$ by the inverse cumulative distribution function of $\Gamma(3, 154)$ distribution plus an error $\epsilon_{i,0} \sim \mathcal{N}(0, 20)$: $L_{i,0} = F^{-1}_{\Gamma(3,154)}(U_{i,0}) + \epsilon_{i,0}$.
4. If $P_i \geq \pi$ and $L_{i,0} < \tau$, the subject starts HAART and $A_{i,0} = 1$. Otherwise, draw treatment decision $A_{i,0} \sim Be\left(p^A_{i,0}\right)$ where $p^A_{i,0} = \text{logit}^{-1}\left[-0.405 - 0.00405 \cdot (L_{i,0} - 500)\right]$. If $A_{i,0} = 1$, set the treatment initiation time K^*_i to 0.
5. Compute the conditional individual hazard $\lambda_{i,0}$ for $k = 0$ using (8). If $\lambda_{i,0} \geq U_{i,0}$, death has occurred in the interval (0,1] and set $Y_{i,1} = 1$. Otherwise, the subject survived and set $Y_{i,1} = 0$.

For $k = 1, \ldots, K$, if the individual is still at risk:

6. Draw $U_{i,k} = \min\{1, \max\{0, U_{i,k-1} + \mathcal{N}(0, 0.05)\}\}$ as a perturbation of $U_{i,k-1}$ restricted to [0,1].
7. If k is not a checkup visit, CD4 cell counts are not updated and $L_{i,k} = L_{i,k-1}$. Otherwise, update the count as $L_{i,k} = \max\left\{0, L_{i,k-1} + 150 \cdot A_{i,k-1} + \epsilon_{i,k}\right\}$, where the addition of 150 indicates the positive effect of exposure to HAART on CD4 count, and $\epsilon_{i,k} \sim \mathcal{N}(100(U_{i,k} - 2), 50)$ is a Gaussian drift term implying that the worse is the general health condition $U_{i,k}$ (i.e., value closer to 0), the stronger the negative drift in CD4.
8. If k is not a checkup visit, treatment is not updated and $A_{i,k} = A_{i,k-1}$. Otherwise, assign exposure
   a. *deterministically*: if $P_i \geq \pi$ and $L_{i,k} < \tau$ or if treatment has started at previous checkup ($A_{i,k-\kappa} = 1$), patient i is exposed to HAART and $A_{i,k} = 1$;
   b. *stochastically*: otherwise, draw treatment decision $A_{i,k} \sim Be\left(p^A_{i,k}\right)$, where

$$p^A_{i,k} = \text{logit}^{-1}\left[-0.405 + 0.0205 \cdot k - 0.00405 \cdot (L_{i,k} - 500)\right].$$

   As in Benchmark I, the conditional distribution parameters have been set to calibrate the logistic function such that $\Pr(A_{\bullet,0} = 1 \mid L_{\bullet,0} = 500) = 0.4$, $\Pr(A_{\bullet,0} = 1 \mid L_{\bullet,0} = 400) = 0.5$, and $\Pr(A_{\bullet,10} = 1 \mid L_{\bullet,10} = 500) = 0.45$, where "$\bullet$" represents a placeholder indicating that the statements apply to all possible values of the first subscript.

   If the subject starts the treatment at visit/time k, set the treatment initiation time K^*_i equal to k.
9. Compute $\lambda_{i,k}$, that is, the individual probability of failure in the interval $(k; k+1]$ conditional on survival up to visit k, using (8). If $S_i(t) = \prod_{j=0}^{k}(1 - \lambda_{i,j}) \leq 1 - U_{i,0}$, the death has occurred in the interval $(k; k+1]$ and $Y_{i,k+1} = 1$. Otherwise, the subject remains at risk and $Y_{i,k+1} = 0$.

The related pseudocode is provided in Appendix A.2. Note that when $\pi = 1$ the positivity assumption always holds and this procedure corresponds to the data-generating mechanism of Benchmark I. An example of a dataset simulated using Algorithm 2 is available in the vignette provided as Supporting Information.

4.2 | Simulation Study Using Algorithm 2

4.2.1 | Methods and Estimands

Investigations are performed in several scenarios by considering different sample sizes (n = 50, 100, 250, 500, 1000), exposure cutoff values ($\pi = 0.05, 0.1, 0.3, 0.5, 0.8, 1$), WT strategies (NoWT, 1–99, 5–95, 10–90), and poor health subgroups $\mathcal{I}_\tau = [0; \tau)$ with varying upper thresholds ($\tau = 0, 100, 200, 300, 400, 500$ measured in cells/μL). The other parameters are set to be identical to those used by Havercroft and Didelez (2012) to consider their results as a benchmark for this analysis. Specifically, $K = 40$ time points with checkups every ($\kappa = 5$)th visit are considered, and the desired conditional distribution parameters in Equation (8) are $(\gamma_0, \gamma_{A1}, \gamma_{A2}, \gamma_{A3}) = (-3, 0.05, -1.5, 0.1)$. In this way, the true values of the parameters in logit-MSM (7) are $(\tilde{\gamma}^*_0, \tilde{\gamma}^*_{A1}, \tilde{\gamma}^*_{A2}, \tilde{\gamma}^*_{A3}) = (-3, 0.05, -1.5, 0.1)$.

For each scenario, $B = 1000$ simulated datasets are generated. The logit-MSM (7) is fitted to each simulated dataset through IPTW estimation using (truncated) stabilized weights. Weight components at each checkup visit ($k = 0, \kappa, 2\kappa, \ldots$) are estimated by logistic regression models for the probability of treatment initiation, with numerator and denominator in (5) defined, respectively, as

$$\Pr\left(A_{i,k}=1 \mid \bar{\boldsymbol{A}}_{i,k-1}=\bar{\mathbf{0}}, Y_{i,k-1}=0\right)=\text{logit}^{-1}[\theta_0+\theta_1\cdot k] \text{ and}$$

$$\Pr\left(A_{i,k}=1 \mid \bar{\boldsymbol{A}}_{i,k-1}=\bar{\mathbf{0}}, \bar{\boldsymbol{L}}_{i,k}, Y_{i,k-1}=0\right)$$
$$=\text{logit}^{-1}\left[\theta_0+\theta_1\cdot k+\theta_2\cdot(L_{i,k}-500)\right].$$

In this way, since $\pi \neq 0$, the denominator model is correctly specified according to the data-generating mechanism (Bryan 2004; Havercroft and Didelez 2012).

The estimands of interest are the regression coefficients $(\tilde{\gamma}_0, \tilde{\boldsymbol{\gamma}}_A)$ and the marginal survival probabilities in Equation (3) for the *always treated* versus *never treated* regimens, where $g(\tilde{\boldsymbol{\gamma}}_A; \bar{\boldsymbol{a}}_k) = \tilde{\gamma}_{A1}\cdot d_{1k}+\tilde{\gamma}_{A2}\cdot a_k+\tilde{\gamma}_{A3}\cdot d_{3k}$.

Section 4.2.2 presents the results across all scenarios. For each regression coefficient, estimated bias, empirical standard error (empSE), and root mean squared error (RMSE) (Morris et al. 2019) are considered as performance measures. Marginal survival curves are presented graphically by showing the mean estimated curves across the $B = 1000$ repetitions. Note that simulation settings with $\pi = 1$ and NoWT are equivalent to Benchmark I, regardless of τ (positivity always holds when $\pi = 1$). In such cases, the analyses are based on correctly specified logit-MSMs and correctly specified models for the weights, so the resulting estimates are expected to be approximately unbiased.

Section 4.2.3 presents the results of three specific scenarios with $(n, \pi, \tau) = (500, 0.1, 400)$ and different WT strategies (NoWT; WT 1–99; WT 5–95). For each scenario, the log-transformed within-dataset summary measures (i.e., mean, maximum, and minimum) of the estimated standardized IPTW weights over repetitions are shown along with the corresponding estimation errors for the regression coefficients and relative performance measures. Estimated marginal survival curves for each simulated dataset are presented graphically, along with the mean estimated curve across repetitions.

4.2.2 | Results Across all Scenarios

Figure 2 shows the mean marginal survival curves, along with the true ones (in orange), in each simulated scenario without WT. Each line refers to a different τ value; the darker the line color, the more severe the violation (i.e., the bigger τ). Each row refers to a different sample size ($n = 50, 100, 250, 500, 1000$), and each column refers to a different exposure cutoff ($\pi = 0.05, 0.1, 0.3, 0.5, 0.8, 1$). Similar behaviors are observed across different sample sizes as π varies: major deviations from true curves occur for small values of π, especially for the *never treated* group (dashed lines), while all scenarios eventually converged to the true values when no positivity violations are present ($\pi = 1$). Contrary to expectations based on the true survival curves, for more severe violations ($\pi = 0.05, 0.1; \tau = 300, 400, 500$), the *never treated* group (dashed lines) exhibits notably better estimated marginal survival curves than the *always treated* group (solid lines), particularly in scenarios with small sample sizes.

This finding is supported by the magnitude and direction of the bias in the estimated regression coefficients $\left(\hat{\tilde{\gamma}}_0, \hat{\tilde{\gamma}}_{A1}, \hat{\tilde{\gamma}}_{A2}, \hat{\tilde{\gamma}}_{A3}\right)$, whose performance metrics (bias, empSE, RMSE) are reported in Supporting Information S1. Smaller sample sizes result in poorer performance across all regression coefficients. While increasing the sample size reduces bias and variability caused by finite sample limitations, estimation errors resulting from violations still persist, particularly for low values of π. Across all scenarios, as the severity of violations increases (i.e., the bigger τ and the lower π), the absolute bias, empSE, and RMSE grow substantially, particularly for $\tilde{\gamma}_0$ and $\tilde{\gamma}_{A2}$ (i.e., the intercept and the parameter most directly related to the effect of exposure). This outcome reflects how large weights resulting from near-violations increase variability and reduce precision in IPTW-based estimates (see Section 4.2.3 for further insights on this aspect). Consequently, since curves for *never treated* depend on $(\hat{\tilde{\gamma}}_0, \hat{\tilde{\gamma}}_{A1})$, they generally show a greater deviation compared to those of *always treated*, which depend on $(\hat{\tilde{\gamma}}_0, \hat{\tilde{\gamma}}_{A2}, \hat{\tilde{\gamma}}_{A3})$—see Equation (1). Indeed, in the *always treated* the negative (positive) bias of $\hat{\tilde{\gamma}}_0$ can be balanced by the positive (negative) one of $\hat{\tilde{\gamma}}_{A2}$. This is not true for the *never treated*, so the deviation notably increases as near-positivity violations become more concrete (i.e., as π decreases).

The paradox deviation of better survival for *never* treated becomes even more pronounced when any WT strategy is applied, as illustrated in Figure 3, where each row corresponds to a different WT strategy with a sample size of $n = 1000$. The estimated mean curves are very close to the true ones for an expected positivity support proportion of 80% under NoWT, or even for 50% under 1–99 WT. However, under a more aggressive WT (5–95 or 10–90), the estimated curves deviate from the true ones, even at high exposure cutoffs. This pattern is again supported by the performance metrics of the estimated regression coefficients (see Supporting Information S1). Adopting a WT strategy does reduce variability and slightly decreases bias by truncating extreme weights, particularly for larger values of τ. However, further narrowing the truncation range (e.g., from WT 1–99 to WT 5–95 or 10–90) does not improve performance and may, in fact, degrade it. The following section offers a more in-depth example illustrating this aspect.

4.2.3 | Focused Examination of WT in Selected Scenarios

To more closely examine the impact of applying or not applying WT in the presence of near-violations, the results of three specific scenarios are presented below. Each scenario is defined by a sample size of $n = 500$, an exposure cutoff of $\pi = 0.1$, the poor health subgroup $\mathcal{I}_{400} = [0; 400)$, and one of three WT strategies: NoWT, WT 1–99, or WT 5–95.

Figure 4 shows the boxplots of the logarithm of the within-dataset mean (left panel), maximum (center panel), and minimum (right panel) of the estimated standardized IPTW weights, $\widehat{sw}_i^b(t)$, computed across individuals ($i = 1, \ldots, 500$) in each simulated dataset ($b = 1, \ldots, 1000$). Each color refers to a different WT strategy (magenta: NoWT; yellow: WT 1–99; blue: WT 5–95). Under NoWT, several patterns indicate potential issues with weight stability. Deviations of log(mean) from 0 suggest shifts in the distribution of estimated stabilized weights across simulated datasets, driven by sampling variability, model sensitivity, or changes in covariate distributions over time, particularly as subjects die and exit the risk set. An observable increase in the range of log(mean) over

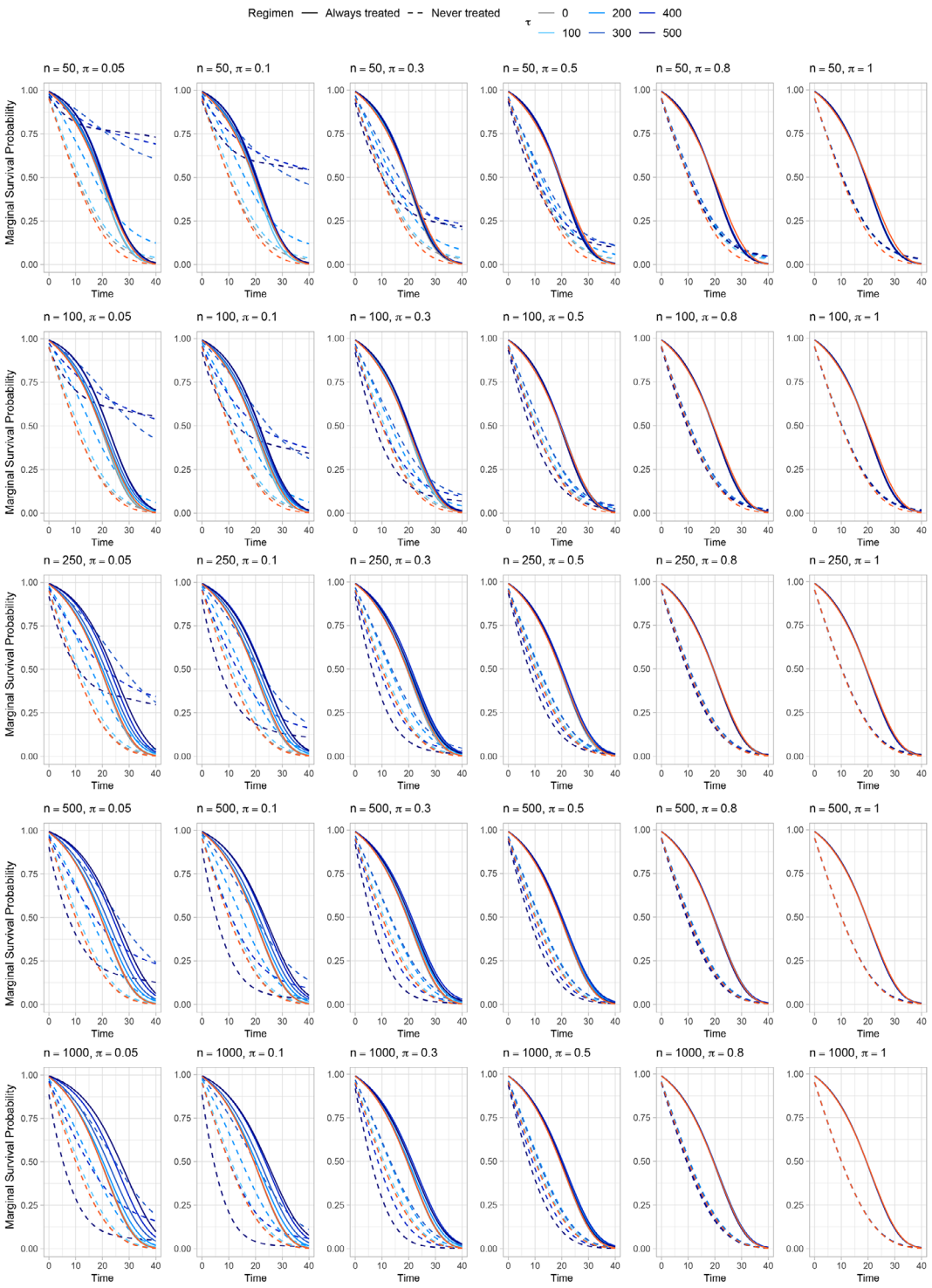

FIGURE 2 | Marginal survival probability curves averaged across all the $B = 1000$ repetitions for different settings without weight truncation (NoWT) of simulation study I. Each row refers to a different sample size $n = 50, 100, 250, 500, 1000$. Each column refers to a different exposure cutoff $\pi = 0.05, 0.1, 0.3, 0.5, 0.8, 1$. Dashed lines refer to the *never treated* regimen, while solid ones refer to the *always treated* regimen. Curves are colored according to different values of rule-threshold τ. True marginal survival curves are shown in orange.

time further indicates growing instability, likely due to deteriorating model performance in later periods. Values of log(max) greater than 3 reflect the presence of very large weights (e.g., ≥ 20), indicating limited covariate overlap and violations of the positivity assumption. The range of log(max) also increases over time, reflecting the greater influence of extreme weights as fewer subjects remain under observation. Although log(min) remains mostly above −5, suggesting that excessively small weights (e.g., <0.01) are rare, some variability is still observed, especially at later time points. This signals sensitivity to sparse covariate patterns or near-deterministic treatment assignments among a shrinking subset of the population. These issues are most pronounced in

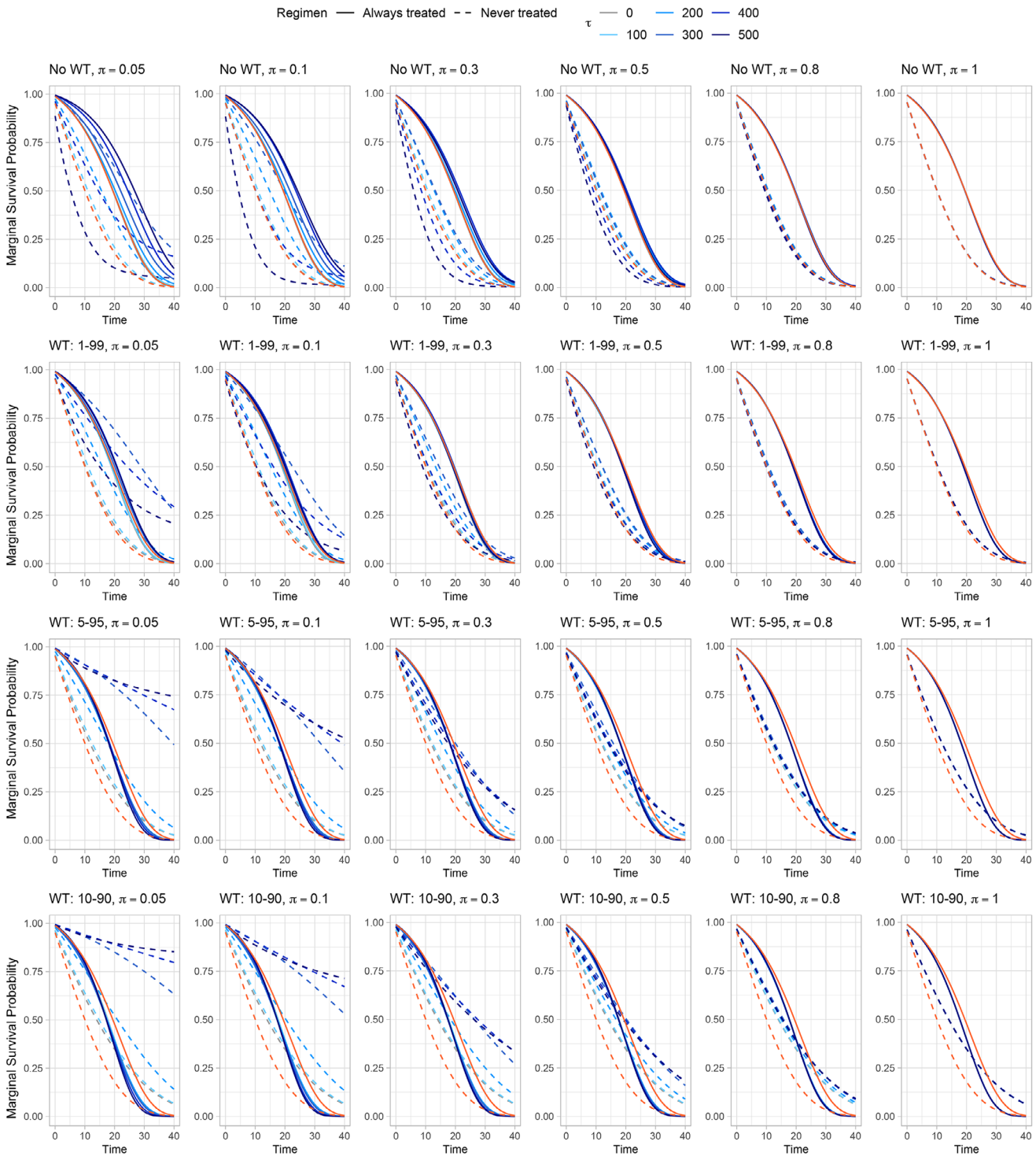

FIGURE 3 | Marginal survival probability curves averaged across all the $B = 1000$ repetitions for different settings with sample size $n = 1000$ of simulation study I. Each row refers to a different weight truncation (WT) strategy: NoWT, 1–99, 5–95, 10-99. Each column refers to a different exposure cutoff $\pi = 0.05, 0.1, 0.3, 0.5, 0.8, 1$. Dashed lines refer to the *never treated* regimen, while solid ones refer to the *always treated* regimen. Curves are colored according to different values of rule-threshold τ. True marginal survival curves are shown in orange.

the NoWT scenario (magenta), where extreme weights are not controlled. In contrast, truncation strategies such as WT 1–99 (yellow) or WT 5–95 (blue) effectively reduce the presence of extreme outliers, improving stability over time. However, high log(max) values under WT 1–99 still suggest potential positivity violations, while increasingly negative log(mean) under WT 5–95 indicates that some subgroups of the population have very low probabilities of receiving certain treatments, even after truncation.

Boxplots of the estimation errors of the regression coefficients across the simulated datasets for each scenario are shown in Figure 5. Mean estimated coefficients and relative performance

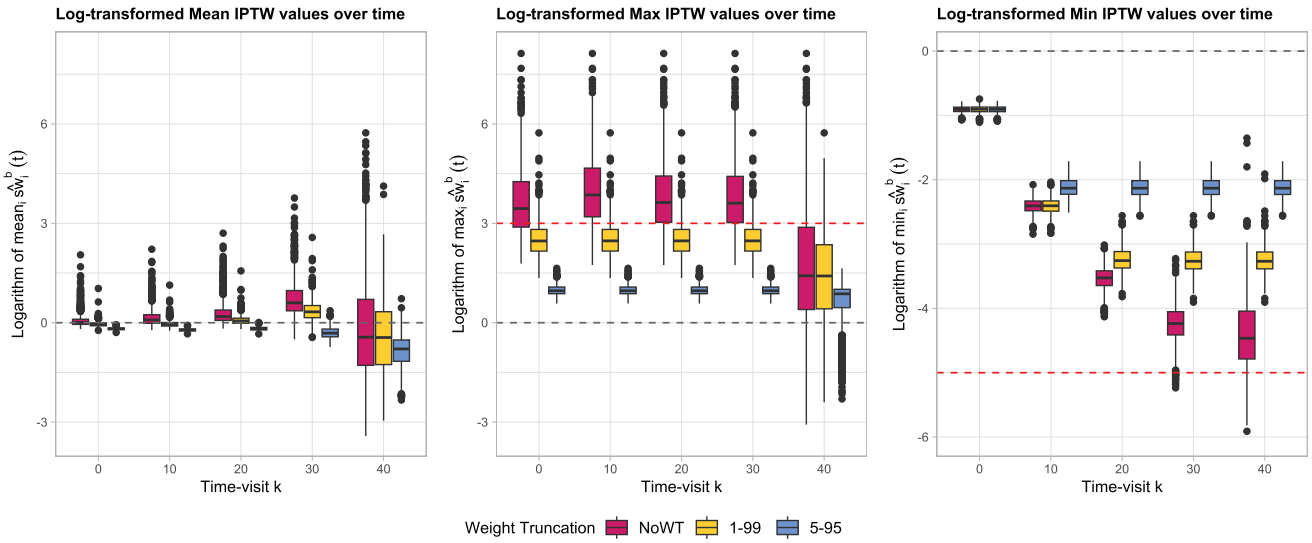

FIGURE 4 | Boxplots of the logarithm of the within-dataset mean (left panel), maximum (center panel), and minimum (right panel) of the estimated standardized IPTW weights over time, $\widehat{sw}_i^b(t)$, computed across individuals in each dataset ($b = 1, \ldots, 1000$) simulated using Algorithm 2 with a sample size of $n = 500$, an exposure cutoff of $\pi = 0.1$, the poor health subgroup $\mathcal{I}_{400} = [0; 400)$, and different WT strategies (magenta: NoWT; yellow: WT 1–99; blue: WT 5–95).

TABLE 1 | Mean, bias, empirical standard error (empSE), and root mean squared error (RMSE) of the coefficient estimates for three different settings of simulation study I with sample size $n = 500$, exposure cutoff of $\pi = 0.1$, poor health subgroup $\mathcal{I}_{400} = [0; 400)$, and WT strategies $\in$ {NoWT, 1–99, 5–95}.

True coefficient	Weight Truncation	Mean	Bias	empSE	RMSE
$\tilde{\gamma}_0^* = -3$	NoWT	−3.216	−0.216	0.881	0.907
	1–99	−3.520	−0.520	0.580	0.778
	5–95	−4.327	−1.327	0.463	1.405
$\tilde{\gamma}_{A1}^* = 0.05$	NoWT	0.050	−0.000	0.076	0.076
	1–99	0.028	−0.022	0.034	0.040
	5–95	0.011	−0.039	0.018	0.043
$\tilde{\gamma}_{A2}^* = -1.5$	NoWT	−1.817	−0.317	0.919	0.972
	1–99	−1.235	0.265	0.581	0.638
	5–95	−0.269	1.231	0.466	1.316
$\tilde{\gamma}_{A3}^* = 0.1$	NoWT	0.106	0.006	0.019	0.020
	1–99	0.107	0.007	0.010	0.012
	5–95	0.118	0.018	0.008	0.020

measures in terms of bias, empSE, and RMSE are shown in Table 1. Results reflect how large weights resulting from near-violations increase variability and reduce precision in IPTW-based estimates. The error magnitude is notably higher for the intercept $\tilde{\gamma}_0$ and the coefficient most directly related to the effect of exposure $\tilde{\gamma}_{A2}$—see Equation (7)—compared to other coefficients, indicating that these parameters are more challenging to estimate accurately. The WT 1–99 strategy (yellow) effectively reduces the variability of the estimates, suggesting that this approach may stabilize the estimation process and improve precision compared to NoWT (magenta). In contrast, further narrowing to WT 5–95 (blue) introduces a clear bias in the estimates. This suggests that more aggressive WT strategies systematically shift the estimates away from the true values, compromising accuracy despite any gains in stability.

This behavior is reflected in the estimates of marginal survival probabilities. Figure 6 shows the estimated marginal survival curves for the *never treated* (top panels) and *always treated* (bottom panels) groups across 1000 simulated datasets. Each panel displays the estimated curves for each dataset (in gray), their mean (in the same color as the WT strategy), and the true survival curve (in orange), under the three truncation strategies (left: NoWT; center: WT 1–99; right: WT 5–95). The plots highlight how highly variable IPTW weights, resulting from near-violations of the positivity assumption, lead to substantial variability in the estimated survival curves, particularly for the *never treated* group. WT can reduce this variability but introduces bias, as it systematically excludes observations in the tails of the weight distribution. These observations often correspond to individuals in regions of the covariate space where the unexposed condi-

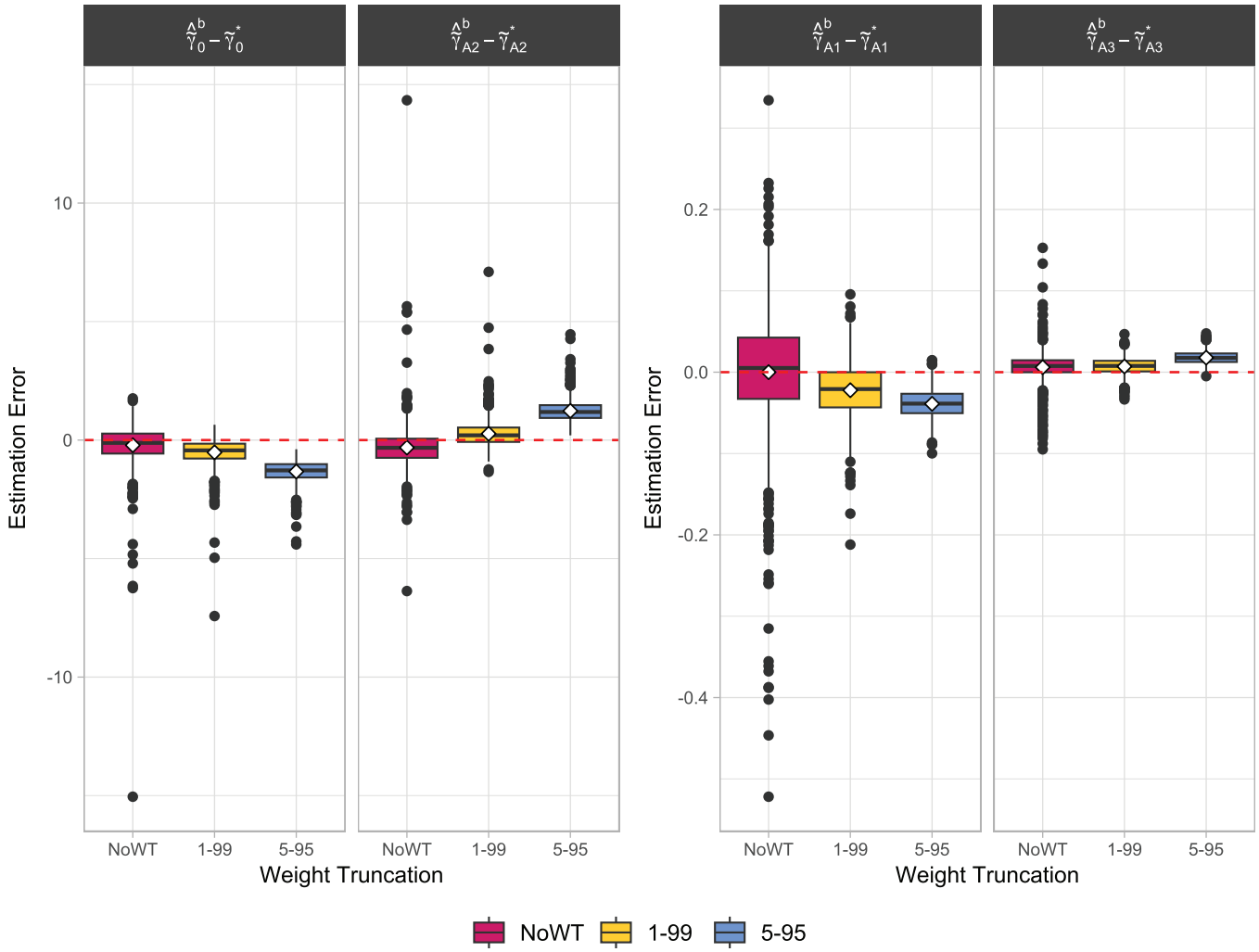

FIGURE 5 | Boxplots of the estimation errors of the regression coefficients across the datasets simulated using Algorithm 2 with $(n, \pi, \tau) = (500, 0.1, 400)$ and three WT strategies (magenta: NoWT; yellow: WT 1–99; blue: WT 5–95). Each column refers to a different coefficient ($\tilde{\gamma}_0$: first column; $\tilde{\gamma}_{A1}$: third column; $\tilde{\gamma}_{A2}$: second column; $\tilde{\gamma}_{A3}$: fourth column). The white diamonds represent the bias across repetitions. Note that the ranges of the y-axes differ between panels.

tion is rare but informative. Truncating these extreme weights disproportionately downweights such individuals, diminishing the representativeness of the weighted pseudo-population. This creates a trade-off: truncation improves stability but may compromise validity by altering the target estimand. As observed in Section 4.2.2, this issue becomes particularly concerning in settings with severe near-violations of positivity, as extreme weights are more prevalent and the impact of excluding informative observations is amplified.

5 | Simulation Study II

5.1 | Data Generation

The second algorithm proposed in this work is based on the data-generating mechanism introduced by Keogh et al. (2021) to simulate from an Aalen-MSM. This mechanism is now briefly introduced and then extended by imposing positivity violations.

5.1.1 | Benchmark II in a Nutshell

Keogh et al. (2021) proposed a setting where, at each visit $k = 0, \ldots, K$, a binary treatment process $A_{i,k} \in \{0, 1\}$ (control vs. treatment) and a time-dependent biomarker $L_{i,k} \in \mathbb{R}$ are observed for each subject i. The assumed data structure is illustrated in Figure 1b using a discrete-time DAG setting where visit times correspond to visit numbers (i.e., $q_k = k\ \forall k$) and $Y_{i,k+1} = I(k < T_i \leq k + 1)$ is an indicator of whether the event T_i occurs between visits k and $k + 1$. As the time intervals become very small, the algorithm approaches the continuous time setting. The DAG also includes a baseline latent variable U_i, representing a subject-specific unmeasured individual frailty, which has a direct effect on $L_{i,k}$ and $Y_{i,k+1}$, but not on $A_{i,k}$.

The DAG in Figure 1b exhibits time-dependent confounding due to $L_{i,k}$ that predicts subsequent treatment use $A_{i,k}$, is affected by earlier treatment $A_{i,k-1}$, and affects the outcome $Y_{i,k+1}$ through pathways that are not just through subsequent treatment. Moreover, because U_i is not a confounder of the association between

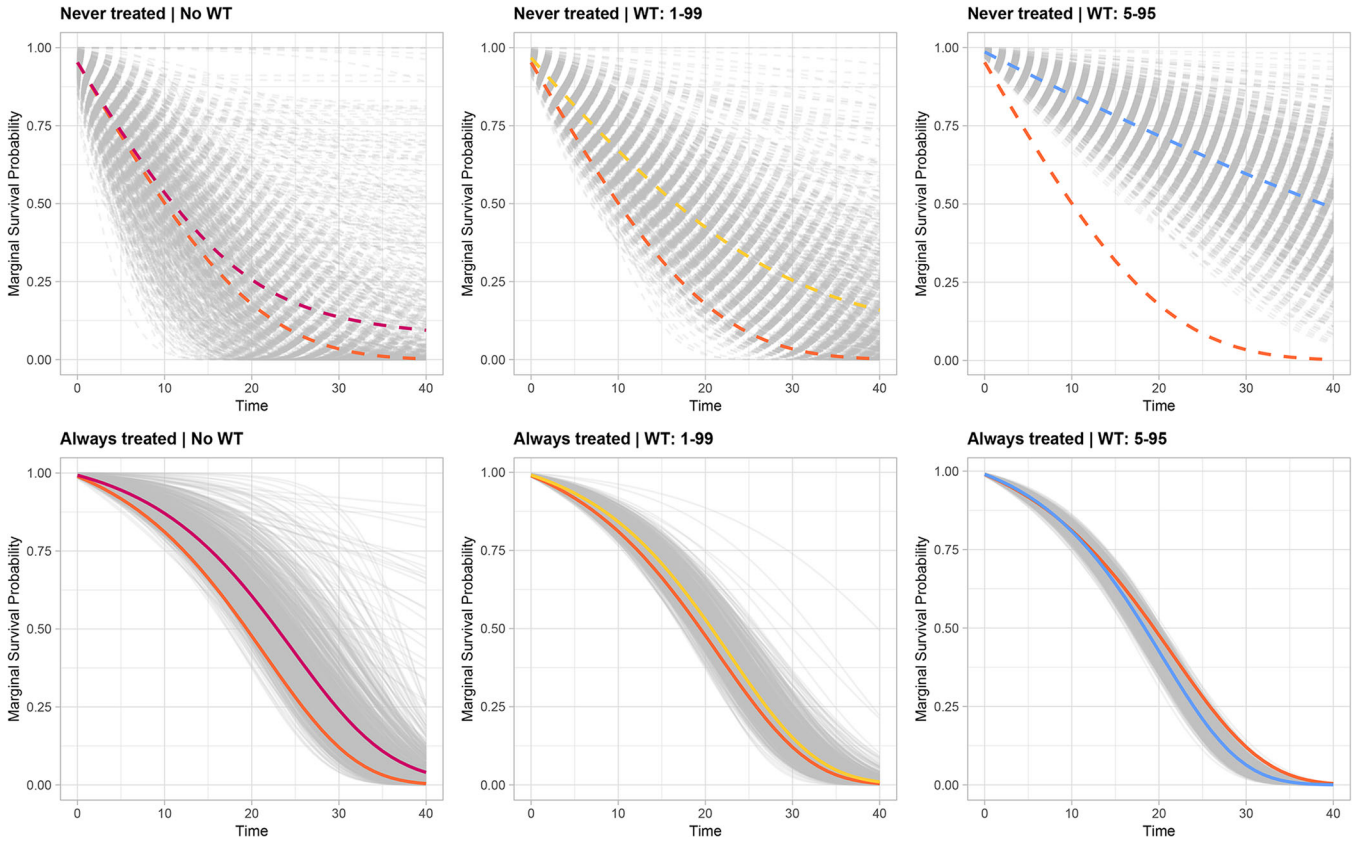

FIGURE 6 | Estimated marginal survival curves for the *never treated* (top panels) and *always treated* (bottom panels) groups across $B = 1000$ datasets simulated using Algorithm 2 with $(n, \pi, \tau) = (500, 0.1, 400)$ and different WT strategies (left panels: NoWT; middle panels: WT 1–99; right panels: WT 5–95). Each panel displays the true survival curve (in orange), the estimated curves for each dataset (in gray), and their mean (magenta: NoWT; yellow: WT 1–99; blue: WT 5–95).

the treatment and the outcome, the fact that it is unmeasured does not affect the ability to estimate causal effects of treatments. The authors demonstrated that using a conditional additive hazard of the form

$$\lambda_i\left(t \mid \bar{\boldsymbol{A}}_{i,\lfloor t \rfloor}, \bar{\boldsymbol{L}}_{i,\lfloor t \rfloor}, U_i\right) = \alpha_0 + \alpha_A \cdot A_{i,\lfloor t \rfloor} + \alpha_L \cdot L_{i,\lfloor t \rfloor} + \alpha_U \cdot U_i, \tag{9}$$

their data-generating mechanism correctly simulates data from the additive Aalen-MSM of the form:

$$\lambda^{\bar{a}}(t) = \tilde{\alpha}_0(t) + \sum_{j=0}^{\lfloor t \rfloor} \tilde{\alpha}_{Aj}(t) \cdot a_{\lfloor t \rfloor - j}, \tag{10}$$

that is an Aalen-MSM including as treatment-pattern $g\left(\tilde{\boldsymbol{\alpha}}_A(t); \bar{\boldsymbol{a}}_{\lfloor t \rfloor}\right)$ the main effect terms at each visit (see Table A1). Researchers using this approach can only specify the parameters $(\alpha_0, \alpha_A, \alpha_L, \alpha_U)$ of the conditional model (9). The true values of the cumulative regression coefficients $C_0(t) = \int_0^t \tilde{\alpha}_0(s)ds$ and $C_{Aj}(t) = \int_0^t \tilde{\alpha}_{Aj}(s)ds$ $(j = 0, \ldots, 4)$ of the Aalen-MSM (10) must be computed using a simulation-based approach, as detailed in Keogh et al. (2021).

Thanks to the collapsibility property of the Aalen's additive hazard model, the generating mechanism of Benchmark II includes the direct arrow from $L_{i,k}$ to $Y_{i,k+1}$, making it more realistic in practice compared to Benchmark I. Unlike Benchmark I, which is restricted to generating data closely matching the Swiss HIV Cohort Study (Sterne et al. 2005), Benchmark II can hence be applicable in more general contexts. However, its parameter values need to be carefully selected to ensure that the chance of obtaining a negative hazard, a common drawback in Aalen's model, is negligible.

5.1.2 | Algorithm 3: Imposing Random Positivity Violations in Benchmark II

Analogously to Algorithm 2, the second algorithm extends Benchmark II (Section 5.1.1) by imposing positivity violations randomly. As illustrated in the DAG in Figure 1d, compared to Benchmark II's structure in Figure 1b, two components are introduced.

i. The latent individual propensity for exposure $P_i \sim \mathcal{U}(0, 1)$ directly acts on $A_{i,k}$: subjects in poor health condition with propensity P_i above the *exposure cutoff* π (to be defined according to the simulation scenario) are always exposed.

ii. The poor health subgroup identified by $\mathcal{I}_\tau$ acts on the purple path $L_{i,k} \to A_{i,k}$. In the framework of Keogh et al.'s procedure, the confounder $L_{i,k}$ represents a general biomarker with no direct real-world interpretation, yet its higher values at time k correspond to an increased likelihood of exposure and

an increased hazard. It is hence reasonable to assume that subjects with a poor health condition at time k are identified by $L_{i,k} > \tau$. This is equivalent to a range $\mathcal{I}_\tau = (\tau, \infty)$, where the lower threshold τ has to be defined according to the simulation scenario. Here, the lower the threshold τ, the wider the interval $\mathcal{I}_\tau$ and the more severe the violation.

The proposed procedure extends Keogh et al.'s algorithm by incorporating the possibility for near-positivity violations. For details regarding the chosen parameter values, please refer to their primary work.

Procedure For each subject $i = 1, \ldots, n$, the simulation procedure with $K+1$ as administrative censoring time is as follows:

1. Generate the individual propensity to exposure: $P_i \sim \mathcal{U}(0,1)$.
2. Generate the individual frailty term: $U_i \sim \mathcal{N}(0, 0.1)$.
3. Generate the baseline biomarker as a transformation of U_i: $L_{i,0} \sim \mathcal{N}(U_i, 1)$.
4. If $P_i \geq \pi$ and $L_{i,0} > \tau$, the subject is exposed to treatment and $A_{i,0} = 1$. Otherwise, draw treatment decision $A_{i,0} \sim Be\left(p^A_{i,0}\right)$, where $p^A_{i,0} = \text{logit}^{-1}\left[-2 + 0.5 \cdot L_{i,0}\right]$.
5. Event times in the period $0 < t < 1$ are generated by calculating $\Delta_i = -\log(v_{i,0})/\lambda_i\left(t \mid A_{i,0}, L_{i,0}, U_i\right)$, where at numerator $v_{i,0} \sim \mathcal{U}(0,1)$ and the denominator is the individual conditional hazard in (9) with $\lfloor t \rfloor = 0$ and desired parameters. If $\Delta_i < 1$, death occurred in the interval $t \in (0,1)$: the event time is set to be $T_i = \Delta_i$, and the failure process is $Y_{i,1} = 1$. Otherwise, subjects with $\Delta_i \geq 1$ remain at risk at time $t = 1$ and set $Y_{i,1} = 0$.
   For $k = 1, \ldots, K$, if the individual is still at risk:
6. Update the biomarker value as: $L_{i,k} \sim \mathcal{N}(0.8 \cdot L_{i,k-1} - A_{i,k-1} + 0.1 \cdot k + U_i, 1)$.
7. Assign exposure
   a. *deterministically*: if $P_i \geq \pi$ and $L_{i,k} > \tau$, subject i is exposed to treatment and $A_{i,k} = 1$;
   b. *stochastically*: otherwise, draw treatment decision $A_{i,k} \sim Be\left(p^A_{i,k}\right)$ where

$$p^A_{i,k} = \text{logit}^{-1}\left[-2 + 0.5 \cdot L_{i,k} + A_{i,k}\right].$$

8. Event times in the period $k \leq t < k+1$ are generated by calculating $\Delta_i = -\frac{\log(v_{i,k})}{\lambda_i\left(t \mid \bar{A}_{i,k}, \bar{L}_{i,k}, U_i\right)}$, where $v_{i,k} \sim \mathcal{U}(0,1)$ and the denominator is the individual conditional hazard in (9) with $\lfloor t \rfloor = k$ and desired parameters. If $\Delta_i < 1$, death occurred in the interval $[k, k+1)$: the event time is set to be $T_i = k + \Delta_i$ and the failure process is $Y_{i,k+1} = 1$. Otherwise, subjects with $\Delta_i \geq 1$ remain at risk at time $k+1$, that is, $Y_{i,k+1} = 0$.

Subjects who do not have an event time generated in the period $0 < t < K+1$ are administratively censored at time $K+1$.

The related pseudocode is provided in Appendix A.3. Note that when $\pi = 1$ the positivity assumption always holds and this procedure corresponds to the data-generating mechanism of Benchmark II. An example of a dataset simulated using Algorithm 3 is available in the vignette provided as Supporting Information.

5.2 | Simulation Study Using Algorithm 3

5.2.1 | Methods and Estimands

Investigations are performed in several scenarios by considering different sample sizes (n = 50, 100, 250, 500, 1000), exposure cutoff values (π = 0, 0.05, 0.1, 0.3, 0.5, 0.8, 1), WT strategies (NoWT, 1–99, 5–95, 10–90), and poor health subgroups identified by intervals $\mathcal{I}_\tau = (\tau; \infty)$ with varying lower threshold τ. Since $L_{i,k}$ represents a general biomarker with no direct real-world interpretation, the choice of possible values for τ relies on the distribution of the complete history of biomarker values generated using Benchmark II with 100,000 subjects. Specifically, the rounded values closest to the 80th, 90th, 95th, 99th, and 100th percentiles (i.e., $\tau = 1, 1.5, 2, 3, 7$) plus an extreme value outside the observed range (i.e, $\tau = 10$) are considered as possible lower thresholds. The other parameters are set to be identical to those considered by Keogh et al. (2021) in order to (i) have the same true values of the estimands of interest for the Aalen-MSM (10) (see Tables 1 and 2 in Keogh et al. (2021)), (ii) use their results as a benchmark for this analysis, and (iii) ensure that the probability of obtaining a negative hazard is negligible for Benchmark II. Specifically, $K = 4$ time points with administrative censoring at $K+1$ are considered, and the conditional distribution parameters in Equation (9) were $(\alpha_0, \alpha_A, \alpha_L, \alpha_U) = (0.7, -0.2, 0.05, 0.05)$.

For each scenario, $B = 1000$ simulated datasets are generated. The Aalen-MSM (10) is fitted to each simulated dataset through IPTW estimation using (truncated) stabilized weights. Weight components at time k are estimated by logistic regression models for the probability of being exposed at time k, with the numerator and denominator in (5) defined respectively as

$$\Pr\left(A_{i,k} = 1 \mid \bar{A}_{i,k-1}, T_i \geq k\right) = \text{logit}^{-1}\left[\theta_0 + \theta_1 \cdot A_{i,k-1}\right] \quad \text{and}$$

$$\Pr\left(A_{i,k} = 1 \mid \bar{A}_{i,k-1}, \bar{L}_{i,k}, T_i \geq k\right) = \text{logit}^{-1}\left[\theta_0 + \theta_1 \cdot A_{i,k-1} + \theta_2 \cdot L_{i,k}\right].$$

In this way, since $\pi \neq 0$, the denominator model is correctly specified according to the data generation mechanism (Keogh et al. 2021).

The estimands of interest are the cumulative regression coefficients $C_0(t) = \int_0^t \tilde{\alpha}_0(s)ds$ and $C_{Aj}(t) = \int_0^t \tilde{\alpha}_{Aj}(s)ds$ ($j = 0, \ldots, 4$), and the marginal survival probabilities in Equation (4) for the *always treated* and *never treated* regimens, where $g\left(\tilde{\boldsymbol{\alpha}}_A(t); \bar{\boldsymbol{a}}_{\lfloor t \rfloor}\right) = \sum_{j=0}^{\lfloor t \rfloor} \tilde{\alpha}_{Aj}(t) \cdot a_{\lfloor t \rfloor - j}$.

Section 5.2.2 presents the results across all scenarios. For the cumulative regression coefficients, results are presented graphically by showing the performance (i.e., bias, empSE, and RMSE) measured at times $t = 1, 2, 3, 4, 5$. For the marginal survival curves, the mean value of the estimates across repetitions is presented graphically across time points $t = 1, 2, 3, 4, 5$. Note that simulation settings with $\pi = 1$ and NoWT are equivalent to

TABLE 2 | Mean, bias, empirical standard error (empSE), and root mean squared error (RMSE) of the cumulative coefficient estimates for three different settings of simulation study II with sample size $n = 500$, exposure cutoff of $\pi = 0.05$, poor health subgroup $\mathcal{I}_1 = (1; \infty)$, and WT strategies $\in$ {NoWT, 1–99, 5–95}.

Cum. Coefficient	Time	True	Weight Truncation	Mean	Bias	empSE	RMSE
$C_0(t) = \int_0^t \tilde{\alpha}_0(s)ds$	1	0.700	NoWT	0.693	−0.007	0.055	0.056
			1–99	0.694	−0.006	0.056	0.057
			5–95	0.688	−0.012	0.052	0.053
	2	1.408	NoWT	1.377	−0.031	0.176	0.179
			1–99	1.378	−0.030	0.124	0.127
			5–95	1.371	−0.037	0.094	0.101
	3	2.128	NoWT	2.069	−0.059	0.344	0.348
			1–99	2.059	−0.069	0.240	0.250
			5–95	2.050	−0.078	0.171	0.188
	4	2.863	NoWT	2.768	−0.095	0.478	0.487
			1–99	2.725	−0.138	0.388	0.412
			5–95	2.733	−0.130	0.285	0.314
	5	3.623	NoWT	3.457	−0.166	0.631	0.652
			1–99	3.424	−0.199	0.568	0.602
			5–95	3.403	−0.220	0.452	0.502
$C_{A0}(t) = \int_0^t \tilde{\alpha}_{A0}(s)ds$	1	−0.198	NoWT	−0.222	−0.024	0.192	0.194
			1–99	−0.213	−0.015	0.160	0.160
			5–95	−0.182	0.016	0.111	0.113
	2	−0.396	NoWT	−0.434	−0.038	0.371	0.372
			1–99	−0.417	−0.021	0.283	0.283
			5–95	−0.356	0.040	0.198	0.201
	3	−0.594	NoWT	−0.633	−0.039	0.555	0.556
			1–99	−0.613	−0.019	0.417	0.417
			5–95	−0.520	0.074	0.311	0.320
	4	−0.790	NoWT	−0.821	−0.031	0.747	0.747
			1–99	−0.767	0.023	0.610	0.610
			5–95	−0.676	0.114	0.464	0.477
	5	−0.987	NoWT	−0.974	0.013	0.969	0.969
			1–99	−0.934	0.053	0.832	0.834
			5–95	−0.819	0.168	0.660	0.681

Benchmark II, regardless of τ (positivity always holds). In such cases, the analyses are based on correctly specified Aalen-MSMs and correctly specified models for the weights, so the resulting estimates are expected to be approximately unbiased.

Section 5.2.3 presents the results of three specific scenarios with $(n, \pi, \tau) = (500, 0.1, 1)$ and different WT strategies (NoWT; WT 1–99; WT 5–95). For each scenario, the log-transformed within-dataset summary measures (i.e., mean, maximum, and minimum) of the estimated standardized IPTW weights over repetitions are shown along with the corresponding estimation errors for the cumulative regression coefficients and relative performance measures. Estimated marginal survival curves for each simulated dataset are presented graphically, along with the mean estimated curve across repetitions.

5.2.2 | Results Across All Scenarios

Figure 7 shows the mean marginal survival curves over times $t = 1, 2, 3, 4, 5$ across repetitions estimated in the various scenarios without WT, along with the true ones (in green). Each line refers to a different τ value; the darker the line color, the more severe the violation (i.e., the smaller τ). Each row refers to a different sample size ($n = 50, 100, 250, 500, 1000$), and each column to a different exposure cutoff ($\pi = 0.05, 0.1, 0.3, 0.5, 0.8, 1$). Even at high exposure cutoffs, scenarios with small sample sizes ($n = 50, 100, 250$) heavily suffer from the main drawback of the additive hazard model, which does not restrict the hazard to be nonnegative. This determines survival probabilities for the *always treated* (solid lines) that wrongly increase over time. This issue is mitigated with bigger sample sizes ($n = 500, 1000$), where

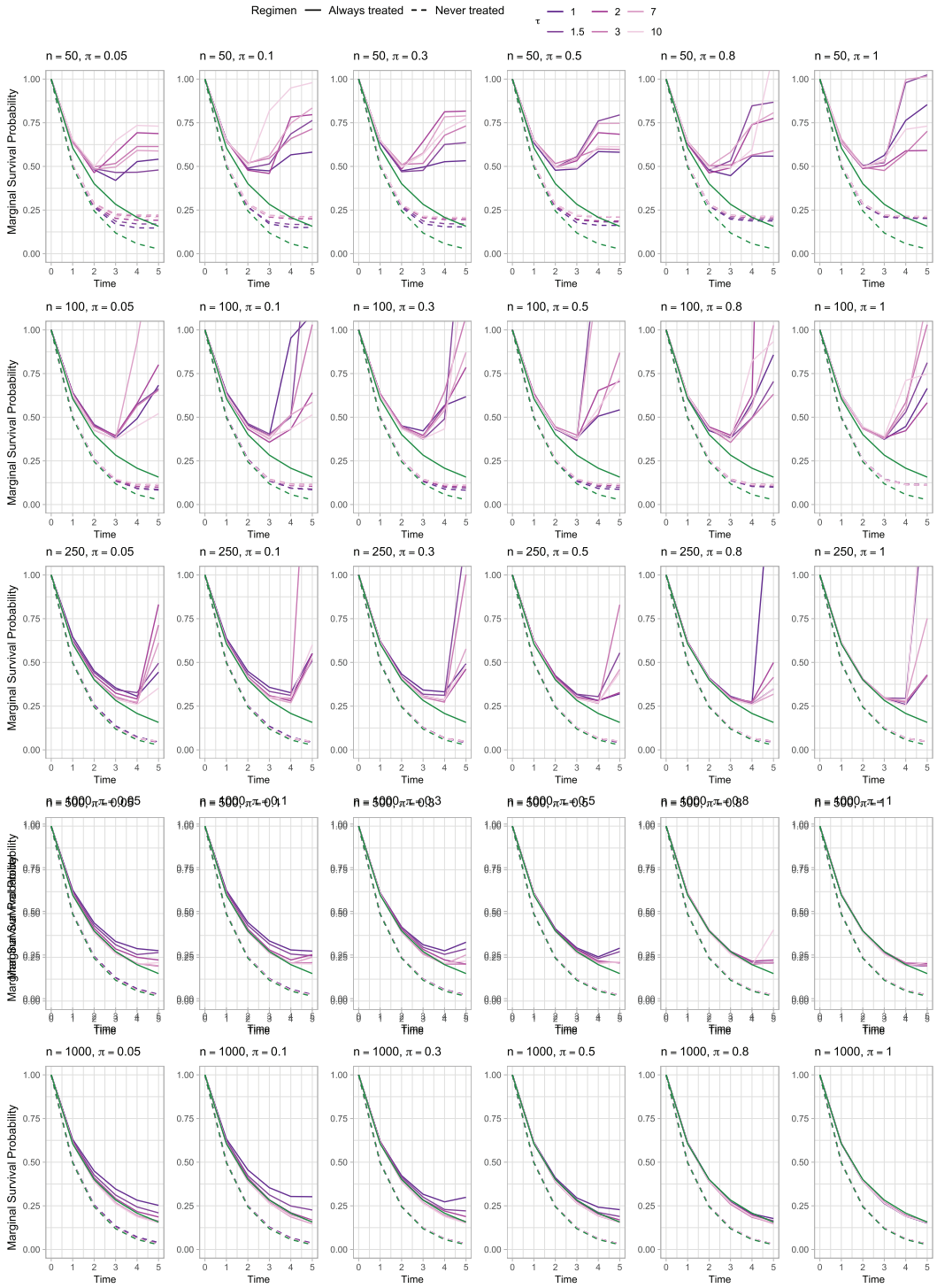

FIGURE 7 | Marginal survival probability curves averaged across all the $B = 1000$ repetitions for different settings without weight truncation (NoWT) of simulation study II. Each row refers to a different sample size $n = 50, 100, 250, 500, 1000$. Each column refers to a different exposure cutoff $\pi = 0.05, 0.1, 0.3, 0.5, 0.8, 1$. Dashed lines refer to the *never treated* regimen, while solid ones refer to the *always treated* regimen. Curves are colored according to different values of rule-threshold τ. True marginal survival curves are shown in green.

increasing the exposure cutoff π leads to estimated mean curves that closely align with the true ones, particularly when the expected positivity support proportion is 80% or higher. Adopting 1–99 WT (see Figure 8) improves performance compared to NoWT, with the estimated mean curves aligning with the true ones even at expected positivity support proportion as low as 30%. Nonetheless, further narrowing the WT range results in increased bias over time compared to 1–99 WT, suggesting that more aggressive truncation excludes important tail observations, reduces the representativeness of the pseudo-population, and may introduce bias that outweighs the benefits of reduced variance.

These findings are supported by the performance metrics for the estimated cumulative coefficients over time ($t = 1, 2, 3, 4, 5$)

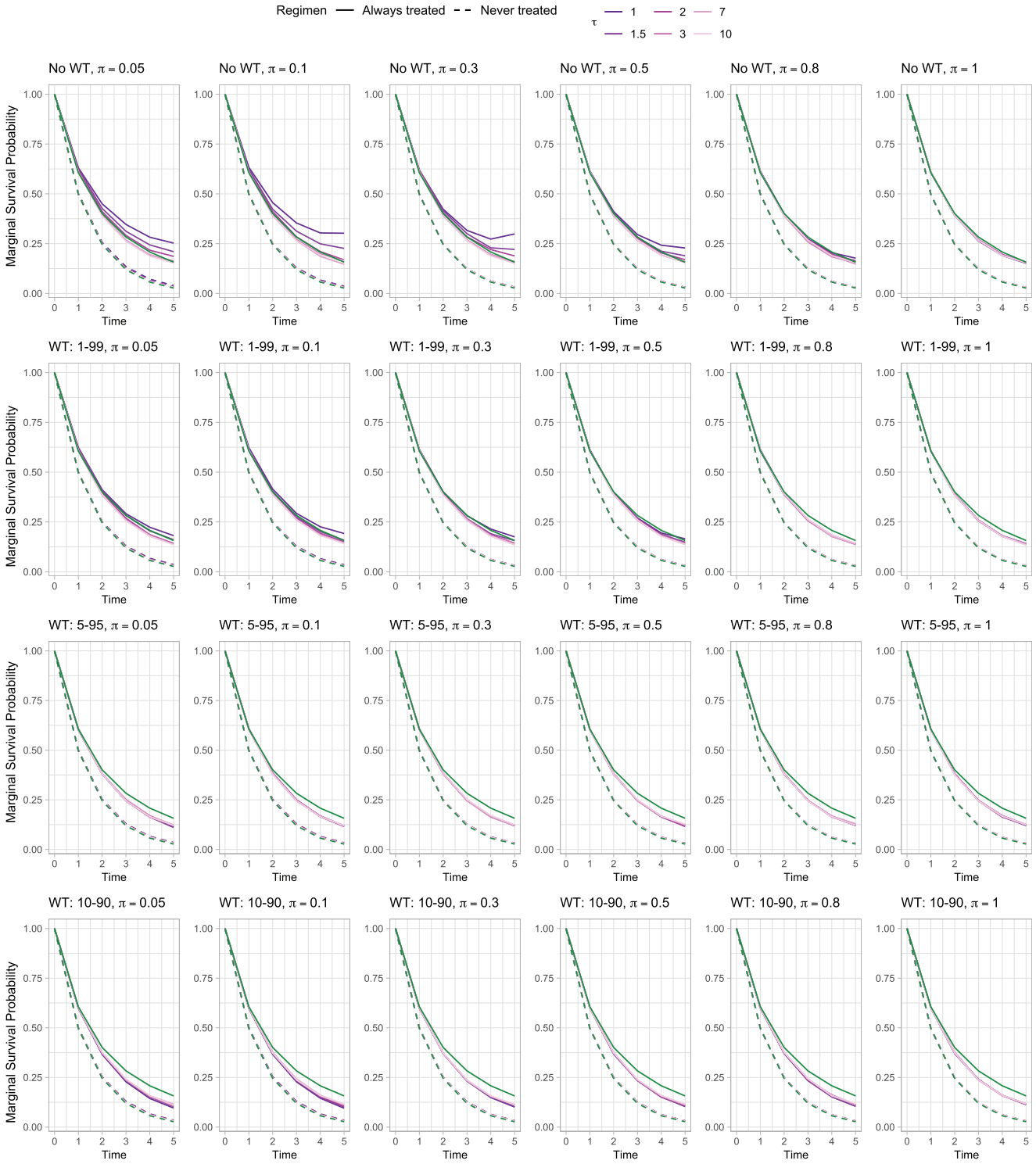

FIGURE 8 | Marginal survival probability curves averaged across all the $B = 1000$ repetitions for different settings with sample size $n = 1000$ of simulation study II. Each row refers to a different weight truncation (WT) strategy: NoWT, 1–99, 5–95, 10–99. Each column refers to a different exposure cutoff $\pi = 0.05, 0.1, 0.3, 0.5, 0.8, 1$. Dashed lines refer to the *never treated* regimen, while solid ones refer to the *always treated* regimen. Curves are colored according to different values of rule-threshold τ. True marginal survival curves are shown in green.

in each simulated scenario, as reported in Supporting Information S2. Indeed, the curves are derived from the cumulative coefficients—as in Equation (4)—whose estimates determine how closely the estimated mean curves match the true (green) ones. At each time point $t = 1, 2, 3, 4, 5$, the curves for *never treated* (dashed lines) depend solely on $\hat{C}_0(t)$ (the cumulative intercept), while all cumulative coefficients contribute to estimating the curves for *always treated* (solid lines). The estimated bias for $\hat{C}_0(t)$ is negative and decreases over time, with minimal differences across π values for large sample sizes ($n = 500, 1000$).

As a result, the estimated mean curves for *never treated* align with the true (green) curves for big sample sizes, exhibiting very low bias across time points. Conversely, summing the contributions of each cumulative coefficient results in higher bias for the *always treated*, particularly at later time points. Indeed, the results for the cumulative coefficients lead to similar conclusions: (i) performance worsens with time, (ii) adopting a WT strategy reduces the variability, and (iii) the estimated performance eventually converges to that observed under Benchmark II. Specifically, for the cumulative coefficient $\hat{C}_{A0}(t)$ related to the current main effect terms, smaller sample sizes exhibit worse performance, as increasing the sample size only mitigates the bias induced by finite sample issues. Across scenarios, bias, empSE, and RMSE increase with time. In general, the more severe the violation (i.e., low π, low τ), the higher empSE and RMSE. Adopting a WT strategy decreases empSE and RMSE, as extreme weights are truncated, especially for more severe violations. However, compared to WT 1–99, narrowing down the WT resulted in worse bias over time. Results eventually converge to be unbiased under NoWT, but a small bias still persists under 1–99 WT for $t = 4, 5$. In terms of variability, the empSE are comparable to the one estimated in Benchmark II, even when the expected positivity support proportion is 80%. The other treatment-related cumulative coefficients, $\hat{C}_{Aj}(t)$ for $j = 1, 2, 3, 4$, exhibit similar patterns, with higher empSE and RMSE; however, unlike $\hat{C}_{A0}(t)$, no discernible relationship is found between the values of τ and the resulting bias.

5.2.3 | Focused Examination of WT in Selected Scenarios

Three specific scenarios are presented below to more closely examine the impact of WT on the estimation of the IPTW weights and the target estimands. Each scenario is generated $B = 1000$ times and is defined by a sample size of $n = 500$, an exposure cutoff of $\pi = 0.05$, the poor health subgroup $I_1 = (1; \infty)$, and one of three WT strategies: NoWT, WT 1–99, or WT 5–95.

Figure 9 shows the boxplots of the logarithm of the within-dataset mean (left panel), maximum (center panel), and minimum (right panel) of the estimated standardized IPTW weights, $\widehat{sw}_i^b(t)$, computed across individuals ($i = 1, \ldots, 500$) in each simulated dataset ($b = 1, \ldots, 1000$). Several patterns indicate potential issues with weight stability. Under NoWT (magenta), deviations of log(mean) from 0 with an observable increase in the range over time suggest shifts in the weight distribution and growing instability. Values of log(max) greater than 3 indicate the presence of very large weights (e.g., ≥ 20), indicating limited covariate overlap and violations of the positivity assumption. The range of log(min) also increases over time, reflecting the greater influence of extreme weights as fewer subjects remain under observation. As expected, WT 1–99 (yellow) substantially decreases the range of log(mean) over time; however, high log(max) values greater than 3 are still observed, indicating possible violations of the positivity assumption. WT 5–95 (blue) more effectively limits extreme outliers; however, the increasingly negative log(mean) indicates that the mean of the standardized weights is drifting further from 1, suggesting potential positivity violations and raising concerns about bias.

Figure 10 displays the boxplots of the estimation errors over time points $t = 1, \ldots, 5$ of the cumulative regression coefficients $C_0(t) = \int_0^t \tilde{\alpha}_0(s)ds$ and $C_{Aj}(t) = \int_0^t \tilde{\alpha}_{Aj}(s)ds$ ($j = 0, \ldots, 4$) across the simulated datasets for each scenario (magenta: NoWT; yellow: WT 1–99; blue: WT 5–95). Mean estimated coefficients and relative performance measures in terms of bias, empSE, and RMSE over time are shown in Tables 2 and 3. Results reflect how extreme weights resulting from near-violations increase variability and reduce precision in IPTW-based estimates, especially at later time points. By trimming extreme observations, WT strategies improve stability (lower empSE and RMSE) at the cost of increased bias, reflecting a trade-off with estimate accuracy.

This pattern is evident in the estimated marginal survival probabilities across the simulated datasets presented in Figure 11 (gray lines). The plots highlight how extreme IPTW weights, resulting from near-violations of the positivity assumption, lead to substantial variability in the estimated survival curves, particularly for the *always treated* group (bottom panels), which is more affected than the *never treated* group (top panels) by the inability of the additive hazards model to constrain the hazard function to be nonnegative. Compared to the NoWT strategy (left panels), WT 1–99 (middle panels) and WT 5–95 (right panels) help mitigate this variability, resulting in mean curves (colored lines) that more closely approximate the true survival curves (in green). However, the issue of individual non-monotonic survival curves still persists.

6 | Practical Tips for Applied Statisticians

The simulation studies discussed above provide the basis for deriving a set of practical recommendations for statisticians conducting real-world studies in the presence of near-positivity violations.

First, it is essential to conduct early diagnostic checks to identify potential positivity issues. This involves inspecting the distribution of IPTW weights, where extreme or highly variable weights, or stabilized weights with a mean far from one, may indicate limited treatment variation or model misspecification (Cole and Frangakis 2009; Cole and Hernán 2008). Routine use of summary statistics and diagnostic plots can help detect such problems. Additionally, quantitative measures—such as weighted standardized differences comparing covariate means or prevalences—and qualitative graphical methods should be employed to assess covariate balance between treatment groups in the weighted sample (Austin and Stuart 2015; Desai and Franklin 2019). Since the weights are time-varying in longitudinal settings, these diagnostics should be performed separately at each time point, as the risk set can change over time.

Second, WT must be applied with caution. Truncating weights at high percentiles (e.g., 1–99) can substantially reduce estimator variance and improve stability. However, as shown in the simulation studies, more aggressive truncation usually leads to detrimental effects in terms of bias. Sensitivity analyses using a range of truncation thresholds are recommended to assess the robustness of the findings.

Third, analysts should consider alternative estimation strategies when near-violations are suspected. Approaches like tar-

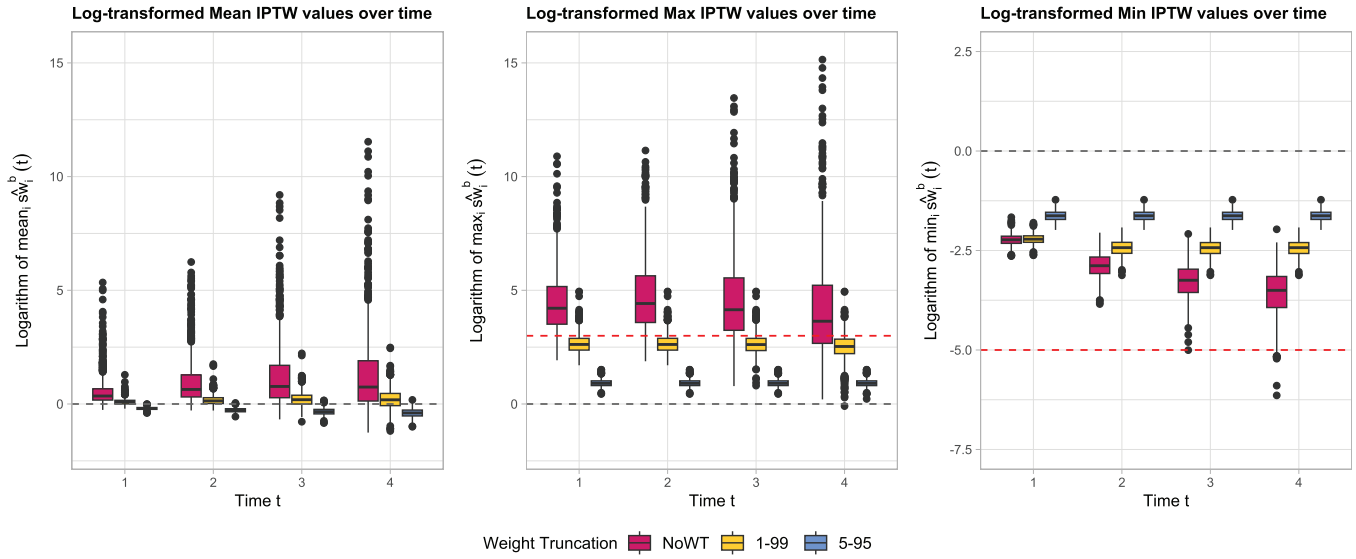

FIGURE 9 | Boxplots of the logarithm of the within-dataset mean (left panel), maximum (center panel), and minimum (right panel) of the estimated standardized IPTW weights over time, $\widehat{sw}_i^b(t)$, computed across individuals in each dataset ($b = 1, \ldots, 1000$) simulated using Algorithm 3 with a sample size of $n = 500$, an exposure cutoff of $\pi = 0.05$, the poor health subgroup $\mathcal{I}_1 = (1; \infty)$, and different WT strategies (magenta: NoWT; yellow: WT 1–99; blue: WT 5–95).

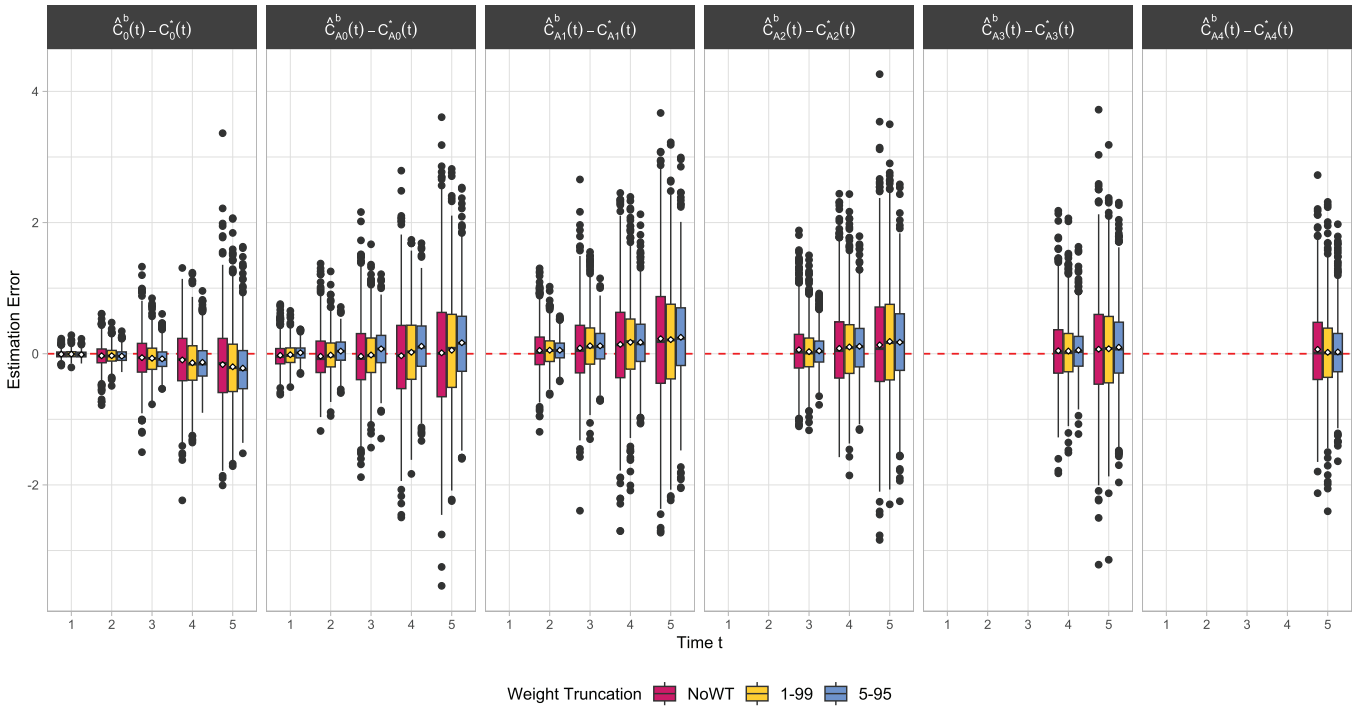

FIGURE 10 | Boxplots of the estimation errors over time of the cumulative regression coefficients across the datasets simulated using Algorithm 3 with $(n, \pi, \tau) = (500, 0.05, 1)$ and three WT strategies (magenta: NoWT; yellow: WT 1–99; blue: WT 5–95). Each column refers to a different cumulative coefficient (first column: $C_0(t) = \int_0^t \tilde{\alpha}_0(s)ds$; columns from two to six: $C_{Aj}(t) = \int_0^t \tilde{\alpha}_{Aj}(s)ds$ with $j = 0, \ldots, 4$, respectively). The white diamonds represent the bias across repetitions over time.

geted maximum likelihood estimation, g-computation, or doubly robust estimators can offer improved performance in these settings (Clare et al. 2019; Daniel et al. 2013; Léger et al. 2022; Petersen et al. 2012; Robins et al. 2000; van der Laan and Gruber 2016).

Fourth, when extreme sparsity in treatment–confounder combinations is observed, it may be appropriate to revise the causal estimand. Focusing on a subpopulation where sufficient support exists for both treatment arms can help preserve identifiability and yield more reliable estimates.

TABLE 3 | Mean, bias, empirical standard error (empSE), and root mean squared error (RMSE) of the cumulative coefficient estimates for three different settings of simulation study II with sample size $n = 500$, exposure cutoff of $\pi = 0.05$, poor health subgroup $\mathcal{I}_1 = (1; \infty)$, and WT strategies $\in$ {NoWT, 1–99, 5–95}.

Coefficient	Time	True	Weight Truncation	Mean	Bias	empSE	RMSE
$C_{A1}(t) = \int_0^t \tilde{\alpha}_{A1}(s)ds$	2	−0.098	NoWT	−0.048	0.050	0.344	0.348
			1–99	−0.044	0.054	0.241	0.247
			5–95	−0.045	0.053	0.163	0.172
	3	−0.195	NoWT	−0.109	0.086	0.571	0.577
			1–99	−0.074	0.121	0.421	0.438
			5–95	−0.075	0.120	0.285	0.310
	4	−0.291	NoWT	−0.149	0.142	0.795	0.807
			1–99	−0.117	0.174	0.618	0.642
			5–95	−0.117	0.174	0.468	0.499
	5	−0.386	NoWT	−0.156	0.230	1.013	1.038
			1–99	−0.171	0.215	0.856	0.883
			5–95	−0.135	0.251	0.696	0.740
$C_{A2}(t) = \int_0^t \tilde{\alpha}_{A2}(s)ds$	3	−0.077	NoWT	−0.016	0.061	0.444	0.448
			1–99	−0.045	0.032	0.355	0.356
			5–95	−0.033	0.044	0.246	0.250
	4	−0.153	NoWT	−0.071	0.082	0.644	0.649
			1–99	−0.050	0.103	0.590	0.599
			5–95	−0.040	0.113	0.439	0.453
	5	−0.228	NoWT	−0.093	0.135	0.921	0.930
			1–99	−0.041	0.187	0.875	0.894
			5–95	−0.053	0.175	0.669	0.691
$C_{A3}(t) = \int_0^t \tilde{\alpha}_{A3}(s)ds$	4	−0.060	NoWT	−0.015	0.045	0.525	0.526
			1–99	−0.019	0.041	0.482	0.483
			5–95	−0.005	0.055	0.357	0.361
	5	−0.121	NoWT	−0.052	0.069	0.812	0.814
			1–99	−0.045	0.076	0.770	0.773
			5–95	−0.022	0.099	0.620	0.628
$C_{A4}(t) = \int_0^t \tilde{\alpha}_{A4}(s)ds$	5	−0.047	NoWT	0.019	0.066	0.674	0.677
			1–99	−0.025	0.022	0.607	0.607
			5–95	−0.020	0.027	0.475	0.476

Note: For each $j = 1, 2, 3, 4$, the cumulative coefficient $C_{Aj}(t)$ is equal to 0 for $t \leq j$.

Finally, transparency in reporting is crucial. Researchers should clearly document any evidence of near-positivity violations, the diagnostic tools employed, and the strategies adopted to address the issue, whether through WT strategies, alternative methods, or modified target populations. Such reporting not only strengthens the credibility of the analysis but also aids interpretability and reproducibility.

7 | Discussion

Simulation studies play a key role in evaluating robustness to assumption violations, enabling the examination of various properties (Friedrich and Friede 2023; Morris et al. 2019). While existing literature on positivity violations in MSMs has largely focused on incorrect inferences using real data or simulations with exposure assigned at a single or two time points, this study fills the gap by presenting two simulation studies in realistic survival contexts involving a time-varying binary treatment and a continuous time-dependent confounder. Two distinct algorithms were proposed to simulate data from hazard-MSMs and to account for potential near-positivity violations, where remaining unexposed is rare within certain confounder levels. Systematic simulations were conducted to evaluate the impact of near-positivity violations on the performance of target estimands obtained via IPTW under various scenarios and WT strategies.

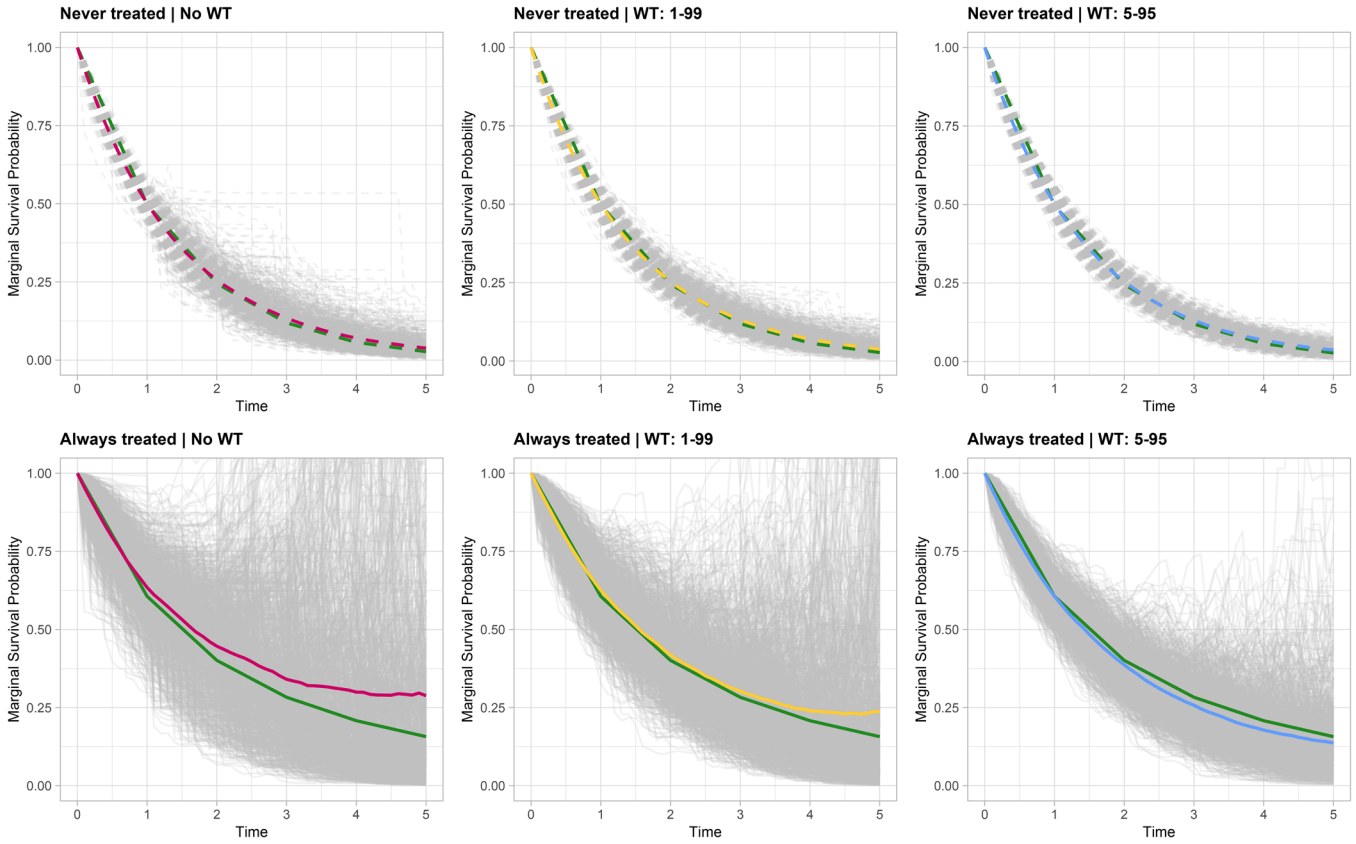

FIGURE 11 | Estimated marginal survival curves for the *never treated* (top panels) and *always treated* (bottom panels) groups across $B = 1000$ datasets simulated using Algorithm 3 with $(n, \pi, \tau) = (500, 0.05, 1)$ and different WT strategies (left panels: NoWT; middle panels: WT 1–99; right panels: WT 5–95). Each panel displays the true survival curve (in green), the estimated curves for each dataset (in gray), and their mean (magenta: NoWT; yellow: WT 1–99; blue: WT 5–95).

Findings from both studies revealed a consistent trend: as the violation becomes more severe (i.e., low π), performance deteriorates. Increasing the sample size mitigates bias and variability due to finite sample size, but incorrect inference resulting from positivity violations persists. Even when $\mathcal{I}_\tau$ is small, performance may still be poor due to the presence of extreme weights. Under NoWT, the higher the expected positivity support proportion π, the better the performance aligns with Benchmarks I and II. Adopting a WT strategy always reduces variability by truncating extreme weights, especially for wider $\mathcal{I}_\tau$. WT 1–99 generally offers a better balance between reducing variability and maintaining accuracy than more aggressive truncations that may not improve bias. This suggests that bias becomes the more dominant factor when the positivity assumption is violated. The decision to adopt the 1–99 WT strategy in cases of near-violations should carefully consider the bias-variance trade-off. For intermediate positivity support proportions ($\pi = 0.3, 0.5$), the 1–99 WT strategy generally outperformed NoWT. In contrast, for high values ($\pi = 0.8, 1$), NoWT was more effective.

Algorithms 2 and II (in the Supporting Information) proposed in this work were built on prior algorithms by Havercroft and Didelez (2012) and Keogh et al. (2021), respectively. The advantage of extending existing algorithms was threefold. First, the issue of non-collapsibility (Didelez and Stensrud 2022; Robinson and Jewell 1991) between conditional and marginal models and the replication of complex confounding dynamics has already been overcome in the original studies. Second, by controlling the exposure-confounder path and avoiding misspecification of the weighting model, the effect due to the imposed positivity violations was separated from other sources of bias. Third, the original Benchmarks I and II were used as the references for the expected true estimates when positivity is valid (i.e., for $\pi = 1$).

While a direct comparison is not feasible as they pertain to different data-generating mechanisms, Algorithm 2 generally exhibited poorer performance compared to Algorithm 3. This difference may stem from their distinct treatment decision mechanisms. Algorithm 2 requires continuous exposure until failure or censoring once treatment begins, whereas Algorithm 3 does not have such a requirement. This constraint limits the possible combinations of treatment-covariate history in Algorithm 2, with a significant impact on the estimated coefficients, even though very few combinations are missing. Consequently, this influences the estimated mean survival curves, leading to incorrect survival probabilities for the *never treated* group. On the other hand, Algorithm 3 suffers, especially with small sample sizes, from the linear form of the Aalen-MSM, which does not restrict the hazard to be nonnegative, resulting in unrealistic survival estimates.

This work has its limitations, which also open up intriguing possibilities for future research. Both studies focused on instances where violations occur within a single interval of the confounder variable and examined only a single continuous confounding

variable. However, in real-world scenarios, violations may span varied intervals, and multiple continuous/categorical confounding factors are typically present. This highlights interesting directions for extending the proposed algorithms, though adapting them to new contexts will require meticulous adjustments. Nonetheless, in their current form, Algorithms I and II developed in this study represent a valuable contribution to the literature. They could serve as data-generating tools for systematic analyses, enabling (i) the comparison of different techniques for estimating causal effects from observational data under near-positivity violations and (ii) the evaluation of potential new methods designed to address near-positivity violations in a longitudinal-treatment framework. Since current methods for detecting and addressing positivity violations are primarily tailored to point-treatment settings (Danelian et al. 2023; Karavani et al. 2019; Traskin and Small 2011; A. Zhu et al. 2023; Zivich et al. 2024), developing methodologies specifically suited to a longitudinal framework presents a challenging direction for future research.

In summary, this study emphasizes the importance of carefully assessing positivity compliance to ensure robust and reliable causal inference in survival studies, while also highlighting the risks of underestimating it. By demonstrating the substantial impact of near-positivity violations, it underscores the need for rigor in causal inference, particularly given the exponential growth of causal inference approaches and their applications to observational data. In practical analyses, researchers are strongly encouraged to examine group-wise descriptives for the original and weighted populations, utilize bootstrap to quantify uncertainty in weights, and conduct sensitivity analysis for further insights (Austin and Stuart 2015; Cole and Hernán 2008; Desai and Franklin 2019). The causal effect of interest must be defined with consideration of positivity violations. While adopting a WT strategy may reduce variability, it should be approached with caution due to the potential risk of increased bias. Although IPTW-based MSMs are widely used in applied studies for their simplicity in implementation and interpretation, analysts must remain vigilant about blindly accepting the positivity assumption, as doing so can lead to detrimental consequences. Finally, the two algorithms developed in this study also serve as valuable tools for generating data in future systematic analyses of novel causal inference methodologies.

Acknowledgments

The author gratefully acknowledges Prof. Dr. Marta Fiocco (Mathematical Institute, Leiden University) for her valuable initial inputs and discussions on this research.

Funding

The author received no specific funding for this work. However, the author's position has been supported by KWF Kankerbestrijding (grant number 2023-3 DEV / 15461).

Conflicts of Interest

The author declares no conflicts of interest.

Data Availability Statement

The data supporting the findings of this study can be generated using the code provided in the Supporting Information.

Open Research Badges

This article has earned an Open Data badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available in the Supporting Information section.

This article has earned an open data badge "**Reproducible Research**" for making publicly available the code necessary to reproduce the reported results. The results reported in this article could fully be reproduced.

References

Austin, P. C., and E. A. Stuart. 2015. "Moving Towards Best Practice When Using Inverse Probability of Treatment Weighting (IPTW) Using the Propensity Score to Estimate Causal Treatment Effects in Observational Studies." *Statistics in Medicine* 34, no. 28: 3661–3679.

Bembom, O., and M. J. van der Laan. 2007. "A Practical Illustration of the Importance of Realistic Individualized Treatment Rules in Causal Inference." *Electronic Journal of Statistics* 1: 574–596.

Bryan, J. 2004. "Analysis of Longitudinal Marginal Structural Models." *Biostatistics* 5, no. 3: 361–380.

Clare, P. J., T. A. Dobbins, and R. P. Mattick. 2019. "Causal Models Adjusting for Time-Varying Confounding—A Systematic Review of the Literature." *International Journal of Epidemiology* 48, no. 1: 254–265.

Cole, S. R., and C. E. Frangakis. 2009. "The Consistency Statement in Causal Inference." *Epidemiology* 20, no. 1: 3–5.

Cole, S. R., and M. A. Hernán. 2008. "Constructing Inverse Probability Weights for Marginal Structural Models." *American Journal of Epidemiology* 168, no. 6: 656–664.

Danelian, G., Y. Foucher, M. Léger, F. Le Borgne, and A. Chatton. 2023. "Identification of In-Sample Positivity Violations Using Regression Trees: The PoRT Algorithm." *Journal of Causal Inference* 11, no. 1:20220032.

Daniel, R. M., S. N. Cousens, B. L. De Stavola, M. G. Kenward, and J. A. Sterne. 2013. "Methods for Dealing With Time-Dependent Confounding." *Statistics in Medicine* 32, no. 9: 1584–1618.

Desai, R. J., and J. M. Franklin. 2019. "Alternative Approaches for Confounding Adjustment in Observational Studies Using Weighting Based on the Propensity Score: A Primer for Practitioners." *BMJ* 367: l5657.

Didelez, V., and M. J. Stensrud. 2022. "On the Logic of Collapsibility for Causal Effect Measures." *Biometrical Journal* 64, no. 2: 235–242.

Evans, R. J., and V. Didelez. 2024. "Parameterizing and Simulating From Causal Models." *Journal of the Royal Statistical Society, Series B (Statistical Methodology)* 86, no. 3: 535–568.

Feuerriegel, S., D. Frauen, V. Melnychuk. et al. 2024. "Causal Machine Learning For Predicting Treatment Outcomes." *Nature Medicine* 30, no. 4: 958–968.

Fewell, Z., M. A. Hernán, F. Wolfe, K. Tilling, H. Choi, and J. A. C. Sterne. 2004. "Controlling for Time-Dependent Confounding Using Marginal Structural Models." *The Stata Journal* 4, no. 4: 402–420.

Friedrich, S., and T. Friede. 2023. "On the Role of Benchmarking Data Sets and Simulations in Method Comparison Studies." *Biometrical Journal* 66, no. 1: 2200212.

Gabriel, E. E., M. C. Sachs, I. Waernbaum. et al. 2024. "Propensity Weighting Plus Adjustment in Proportional Hazards Model Is Not Doubly Robust." *Biometrics* 80, no. 3:ujae069.

Hammerton, G., and M. R. Munafò. 2021. "Causal inference with observational data: the need for triangulation of evidence." *Psychological Medicine* 51, no. 4: 563–578.

Havercroft, W. G., and V. Didelez. 2012. "Simulating From Marginal Structural Models With Time-Dependent Confounding." *Statistics in Medicine* 31, no. 30: 4190–4206.

Hernán, M. A. 2010. "The Hazards of Hazard Ratios." *Epidemiology* 21, no. 1: 13–15.

Hernán, M. A. 2021. "Methods of Public Health Research—Strengthening Causal Inference From Observational Data." *New England Journal of Medicine* 385, no. 15: 1345–1348.

Hernán, M. A., B. Brumback, and J. M. Robins. 2000. "Marginal Structural Models to Estimate the Causal Effect of Zidovudine on the Survival of HIV-Positive Men." *Epidemiology* 11, no. 5: 561–570.

Hernán, M. A., and J. M. Robins. 2016. "Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available." *American Journal of Epidemiology* 183, no. 8: 758–764.

Hernán, M., and J. Robins. 2020. *Causal Inference: What If*. Chapman & Hall/CRC.

Karavani, E., P. Bak, and Y. Shimoni. 2019. "A Discriminative Approach for Finding and Characterizing Positivity Violations Using Decision Trees." Preprint, arXiv, July 18. https://doi.org/10.48550/arXiv.1907.08127.

Keogh, R. H., S. R. Seaman, J. M. Gran, and S. Vansteelandt. 2021. "Simulating Longitudinal Data From Marginal Structural Models Using the Additive Hazard Model." *Biometrical Journal* 63, no. 7: 1526–1541.

Keogh, R. H., and N. van Geloven. 2024. "Prediction Under Interventions: Evaluation of Counterfactual Performance Using Longitudinal Observational Data." *Epidemiology* 35, no. 3: 329–339.

Léger, M., A. Chatton, F. Le Borgne, R. Pirracchio, S. Lasocki, and Y. Foucher. 2022. "Causal Inference in Case of Near-Violation of Positivity: Comparison of Methods." *Biometrical Journal* 64, no. 8: 1389–1403.

Lin, L., M. Sperrin, D. A. Jenkins, G. P. Martin, and N. Peek. 2021. "A Scoping Review of Causal Methods Enabling Predictions Under Hypothetical Interventions." *Diagnostic and Prognostic Research* 5, no. 1: 3.

Martinussen, T., S. Vansteelandt, and P. K. Andersen. 2020. "Subtleties in the Interpretation of Hazard Contrasts." *Lifetime Data Analysis* 26, no. 4: 833–855.

Mitra, N., J. Roy, and D. Small. 2022. "The Future of Causal Inference." *American Journal of Epidemiology* 191, no. 10: 1671–1676.

Moccia, C., G. Moirano, M. Popovic. et al. 2024. "Machine Learning in Causal Inference for Epidemiology." *European Journal of Epidemiology* 39, no. 10: 1097–1108.

Morris, T. P., I. R. White, and M. J. Crowther. 2019. "Using Simulation Studies to Evaluate Statistical Methods." *Statistics in Medicine* 38, no. 11: 2074–2102.

Mortimer, K. M., R. Neugebauer, M. van der Laan, and I. B. Tager. 2005. "An Application of Model-Fitting Procedures for Marginal Structural Models." *American Journal of Epidemiology* 162, no. 4: 382–388.

Naimi, A. I., S. R. Cole, D. J. Westreich, and D. B. Richardson. 2011. "A Comparison of Methods to Estimate the Hazard Ratio Under Conditions of Time-Varying Confounding and Nonpositivity." *Epidemiology* 22, no. 5: 718–723.

Neugebauer, R., and M. J. van der Laan. 2005. "Why Prefer Double Robust Estimators in Causal Inference?" *Journal of Statistical Planning and Inference* 129, no. 1–2: 405–426.

Olier, I., Y. Zhan, X. Liang, and V. Volovici. 2023. "Causal Inference and Observational Data." *BMC Medical Research Methodology* 23: 227.

Petersen, M. L., K. E. Porter, S. Gruber, Y. Wang, and M. J. van der Laan. 2012. "Diagnosing and Responding to Violations in the Positivity Assumption." *Statistical Methods in Medical Research* 21, no. 1: 31–54.

Platt, R. W., J. A. C. Delaney, and S. Suissa. 2012. "The Positivity Assumption and Marginal Structural Models: The Example of Warfarin Use and Risk of Bleeding." *European Journal of Epidemiology* 27, no. 2: 77–83.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing.

Robins, J. M., M. A. Hernán, and B. Brumback. 2000. "Marginal Structural Models and Causal Inference in Epidemiology." *Epidemiology* 11, no. 5: 550–560.

Robinson, L. D., and N. P. Jewell. 1991. "Some Surprising Results About Covariate Adjustment in Logistic Regression Models." *International Statistical Review* 59, no. 2: 227.

Rudolph, J. E., D. Benkeser, E. H. Kennedy, E. F. Schisterman, and A. I. Naimi. 2022. "Estimation of the Average Causal Effect in Longitudinal Data With Time-Varying Exposures: The Challenge of Nonpositivity and the Impact of Model Flexibility." *American Journal of Epidemiology* 191, no. 11: 1962–1969.

Seaman, S. R., and R. H. Keogh. 2024. "Simulating Data From Marginal Structural Models for a Survival Time Outcome." *Biometrical Journal* 66, no. 8: e70010.

Sterne, J. A. C., M. A. Hernán, B. Ledergerber. et al. 2005. "Long-Term Effectiveness of Potent Antiretroviral Therapy in Preventing AIDS and Death: A Prospective Cohort Study." *Lancet* 366, no. 9483: 378–384.

Traskin, M., and D. S. Small. 2011. "Defining the Study Population for an Observational Study to Ensure Sufficient Overlap: A Tree Approach." *Statistics in Biosciences* 3, no. 1: 94–118.

van der Laan, M., and S. Gruber. 2016. "One-Step Targeted Minimum Loss-Based Estimation Based on Universal Least Favorable One-Dimensional Submodels." *International Journal of Biostatistics* 12, no. 1: 351–378.

van Geloven, N., S. A. Swanson, C. L. Ramspek. et al. 2020. "Prediction Meets Causal Inference: The Role of Treatment in Clinical Prediction Models." *European Journal of Epidemiology* 35, no. 7: 619–630.

Wang, Y., M. Petersen, D. Bangsberg, and M. J. van der Laan. 2006. "Diagnosing Bias in the Inverse Probability of Treatment Weighted Estimator Resulting From Violation of Experimental Treatment Assignment." *U.C. Berkeley Division of Biostatistics Working Paper Series* Working Paper 211.

Williamson, T., and P. Ravani. 2017. "Marginal Structural Models in Clinical Research: When and How to Use Them?" *Nephrology, Dialysis, Transplantation* 32, no. S2: ii84–ii90.

Xiao, Y., M. Abrahamowicz, and E. E. M. Moodie. 2010. "Accuracy of Conventional and Marginal Structural Cox Model Estimators: A Simulation Study." *International Journal of Biostatistics* 6, no. 2: 13.

Xiao, Y., E. E. M. Moodie, and M. Abrahamowicz. 2013. "Comparison of Approaches to Weight Truncation for Marginal Structural Cox Models." *Epidemiology Method* 2, no. 1: 1–20.

Young, J. G., M. A. Hernán, S. Picciotto, and J. M. Robins. 2010. "Relation Between three Classes of Structural Models for the Effect of a Time-Varying Exposure on Survival." *Lifetime Data Analysis* 16, no. 1: 71–84.

Young, J. G., and E. J. Tchetgen Tchetgen. 2014. "Simulation From a Known Cox MSM Using Standard Parametric Models for the g-Formula." *Statistics in Medicine* 33, no. 6: 1001–1014.

Zhu, A. Y., N. Mitra, and J. Roy. 2023. "Addressing Positivity Violations in Causal Effect Estimation Using Gaussian Process Priors." *Statistics in Medicine* 42, no. 1: 33–51.

Zhu, Y., R. A. Hubbard, J. Chubak, J. Roy, and N. Mitra. 2021. "Core Concepts in Pharmacoepidemiology: Violations of the Positivity Assumption in the Causal Analysis of Observational Data: Consequences and Statistical Approaches." *Pharmacoepidemiology and Drug Safety* 30, no. 11: 1471–1485.

Zivich, P. N., J. K. Edwards, E. T. Lofgren, S. R. Cole, B. E. Shook-Sa, and J. Lessler. 2024. "Transportability Without Positivity: A Synthesis of Statistical and Simulation Modeling." *Epidemiology* 35, no. 1: 23–31.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.

Supporting File 1: bimj70093-sup-0001-Datacode.zip. **Supporting File 2:** bimj70093-sup-0002-SuppMat.pdf. **Supporting File 3:** bimj70093-sup-0003-SuppMat.pdf.

APPENDIX

A.1 | Treatment-Pattern Forms for Marginal Structural Hazard Models

As mentioned in Section 2.2, marginal structural hazards models (hazard-MSMs) are a class of causal models for the *counterfactual event time* $T^{\bar{a}}$ that would be observed in a subject under complete exposure history $\bar{a}$ (Hernán and Robins 2020; Hernán et al. 2000; Robins et al. 2000). In the case of discrete-time hazard of failure, the *logit-MSM* form in (1) can be assumed to model the counterfactual probability of failure in a single interval $(q_k, q_{k+1}]$, given survival up to q_k. In the context of continuous-time hazard, the *Aalen-MSM* form in (2) can be assumed to model the counterfactual hazard at time t given treatment history $\bar{a}$. An alternative often used is the *marginal structural Cox proportional hazard model* (*Cox-MSM*) form, defined as

$$\lambda^{\bar{a}}(t) = \lambda_0(t) \exp\left\{g\left(\tilde{\boldsymbol{\beta}}_A; \bar{\boldsymbol{a}}_{\lfloor t \rfloor}\right)\right\}, \tag{A1}$$

where $\lambda_0(t)$ is the baseline hazard function, $\bar{\boldsymbol{a}}_{\lfloor t \rfloor}$ denotes treatment pattern up to the most recent visit prior to time t, $g(\cdot)$ is a function (to be specified) of treatment pattern $\bar{\boldsymbol{a}}_{\lfloor t \rfloor}$, and $\tilde{\boldsymbol{\beta}}_A$ is a vector of log hazard ratios.

TABLE A1 | Examples of treatment-pattern forms for function $g(\cdot)$ in Equations (1), (2), and (A1).

Form of treatment pattern	**Logit-MSM**	**Aalen-MSM**	**Cox-MSM**
Function $g(\cdot)$	$g\left(\tilde{\gamma}_A; \bar{a}_k\right)$ **in (1)**	$g\left(\tilde{\alpha}_A(t); \bar{a}_{\lfloor t \rfloor}\right)$ **in (2)**	$g\left(\tilde{\beta}_A; \bar{a}_{\lfloor t \rfloor}\right)$ **in (A1)**
Current level of treatment	$\tilde{\gamma}_A \cdot a_k$	$\tilde{\alpha}_A(t) \cdot a_{\lfloor t \rfloor}$	$\tilde{\beta}_A \cdot a_{\lfloor t \rfloor}$
Duration of treatment	$\tilde{\gamma}_A \cdot \sum_{j=0}^{k} a_{k-j}$	$\tilde{\alpha}_A(t) \cdot \sum_{j=0}^{\lfloor t \rfloor} a_{\lfloor t \rfloor - j}$	$\tilde{\beta}_A \cdot \sum_{j=0}^{\lfloor t \rfloor} a_{\lfloor t \rfloor - j}$
Main effect terms at each visit	$\sum_{j=0}^{k} \tilde{\gamma}_{Aj} \cdot a_{k-j}$	$\sum_{j=0}^{\lfloor t \rfloor} \tilde{\alpha}_{Aj}(t) \cdot a_{\lfloor t \rfloor - j}$	$\sum_{j=0}^{\lfloor t \rfloor} \tilde{\beta}_{Aj} \cdot a_{\lfloor t \rfloor - j}$

In the hazard-MSMs (1), (2), and (A1), the function $g(\cdot)$ combines information from the treatment pattern up to the most recent visit k prior to time t. Depending on the desired information provided in $g(\cdot)$, the hazard at time t (or visit k) can thus assume different forms. Table A1 shows examples of three forms of the treatment pattern function $g(\cdot)$, specifically: (i) the current level of treatment, (ii) the duration of treatment, or (iii) the history of treatment up to time t through the main effect terms for treatment at each visit. Any other desired form can be alternatively specified.

A.2 | Pseudocode of Algorithm 2

ALGORITHM 2 | Pseudocode of Algorithm 2 introduced in Section 4.1.2.

Input parameters:

n = sample size, τ = upper threshold of $\mathcal{I}_\tau$, π = exposure cutoff

K = number of visits/time points, κ = checkup times,

$(\gamma_0, \gamma_{A1}, \gamma_{A2}, \gamma_{A3})$ = conditional distribution parameters in Equation (8)

Algorithm:

for each subject $i = 1, \ldots, n$ **do**

  $P_i \sim \mathcal{U}(0,1)$ ▷ Individual propensity

  $U_{i,0} \sim \mathcal{U}(0,1)$

  $L_{i,0} = F^{-1}_{\Gamma(3,154)}(U_{i,0}) + \epsilon_{i,0}$ where $\epsilon_{i,0} \sim \mathcal{N}(0, 20)$

  if $P_i \geq \pi$ and $L_{i,0} < \tau$ **then**

    $A_{i,0} = 1$ ▷ Deterministic exposure assignment

  else

    $p^A_{i,0} = \text{logit}^{-1}\left[-0.405 - 0.00405 \cdot (L_{i,0} - 500)\right]$ ▷ Stochastic exposure assignment

    $A_{i,0} \sim Be\left(p^A_{i,0}\right)$

  end if

  if $A_{i,0} = 1$ **then** $K^*_i = 0$ **end if**

  $\lambda_{i,0} = \text{logit}^{-1}\left[\gamma_{A0} + \gamma_{A2} \cdot A_{i,0}\right]$ ▷ Conditional hazard (8)

  if $\lambda_{i,0} \geq U_{i,0}$ **then** $Y_{i,1} = 1$ **else** $Y_{i,1} = 0$ **end if**

  $k = 1$

  while $Y_{i,k} = 0$ and $k \leq K$ **do**

    $U_{i,k} = \min\{1, \max\{0, U_{i,k-1} + \varepsilon_{i,k}\}\}$ where $\varepsilon_{i,k} \sim \mathcal{N}(0, 0.05)$

(Continues)

if $k \bmod \kappa \neq 0$ **then**

$L_{i,k} = L_{i,k-1}$

$A_{i,k} = A_{i,k-1}$

else

$L_{i,k} = \max\left\{0, L_{i,k-1} + 150 \cdot A_{i,k-1} + \epsilon_{i,k}\right\}$ where $\epsilon_{i,k} \sim \mathcal{N}(100(U_{i,k} - 2), 50)$

if $(P_i \geq \pi$ and $L_{i,k} < \tau)$ or $A_{i,k-\kappa} = 1$ **then**

$A_{i,k} = 1$ ▷ Deterministic exposure assignment

else

$p^A_{i,k} = \text{logit}^{-1}\left[-0.405 + 0.0205 \cdot k - 0.00405 \cdot (L_{i,k} - 500)\right]$

▷ Stochastic exposure assignment

$A_{i,k} \sim Be\left(p^A_{i,k}\right)$

end if

if $A_{i,k} = 1$ and $A_{i,k-1} = 0$ **then** $K^*_i = k$ **end if**

end if

$\lambda_{i,k} = \text{logit}^{-1}\left[\gamma_0 + \gamma_{A1} \cdot \left\{(1 - A_{i,k})k + A_{i,k}K^*_i\right\} + \gamma_{A2} \cdot A_{i,k} + \gamma_{A3} \cdot A_{i,k}(k - K^*_i)\right]$

▷ Conditional hazard (8)

if $\prod_{j=0}^{k}(1 - \lambda_{i,j}) \leq 1 - U_{i,0}$ **then** $Y_{i,k+1} = 1$ **else** $Y_{i,k+1} = 0$ **end if**

$k = k + 1$

end while

end for

A.3 | Pseudocode of Algorithm 3

ALGORITHM 3 | Pseudocode of Algorithm 3 introduced in Section 5.1.2.

Input parameters:

n = sample size, τ = lower threshold of $\mathcal{I}_\tau$, π = exposure cutoff

K = number of visits/time points

$(\alpha_0, \alpha_A, \alpha_L, \alpha_U)$ = desired true parameters in Equation (9)

Algorithm:

for each subject $i = 1, \ldots, n$ **do**

$P_i \sim \mathcal{U}(0, 1)$ ▷ Individual propensity

$U_{i,0} \sim \mathcal{N}(0, 1)$

$L_{i,0} \sim \mathcal{N}(U_i, 1)$

if $P_i \geq \pi$ and $L_{i,0} > \tau$ **then**

$A_{i,0} = 1$ ▷ Deterministic exposure assignment

else

$p^A_{i,0} = \text{logit}^{-1}\left[-2 + 0.5 \cdot L_{i,0}\right]$ ▷ Stochastic exposure assignment

$A_{i,0} \sim Be\left(p^A_{i,0}\right)$

end if

$\lambda_i\left(t \mid A_{i,0}, L_{i,0}, U_i\right) = \alpha_0 + \alpha_A \cdot A_{i,0} + \alpha_L \cdot L_{i,0} + \alpha_U \cdot U_i$ ▷ Conditional hazard (9)

$\Delta_i = -\log(\upsilon_{i,0})/\lambda_i(t \mid A_{i,0}, L_{i,0}, U_i)$ where $\upsilon_{i,0} \sim \mathcal{U}(0, 1)$

if $\Delta_i < 1$ **then** $T_i = \Delta_i$ and $Y_{i,1} = 1$ **else** $Y_{i,1} = 0$ **end if**

$k = 1$

(Continues)

while $Y_{i,k} = 0$ and $k \leq K$ **do**

$L_{i,k} \sim \mathcal{N}(0.8 \cdot L_{i,k-1} - A_{i,k-1} + 0.1 \cdot k + U_i, 1)$

if $P_i \geq \pi$ and $L_{i,k} > \tau$ **then**

$A_{i,k} = 1$ ▷ Deterministic exposure assignment

else

$p^A_{i,k} = \text{logit}^{-1}\left[-2 + 0.5 \cdot L_{i,k} + A_{i,k}\right]$ ▷ Stochastic exposure assignment

$A_{i,k} \sim Be\left(p^A_{i,k}\right)$

end if

$\lambda_i\left(t \mid \bar{\boldsymbol{A}}_{i,k}, \bar{\boldsymbol{L}}_{i,k}, U_i\right) = \alpha_0 + \alpha_A \cdot A_{i,k} + \alpha_L \cdot L_{i,k} + \alpha_U \cdot U_i$ ▷ Conditional hazard (9)

$\Delta_i = -\log(\upsilon_{i,k})/\lambda_i(t \mid A_{i,0}, L_{i,0}, U_i)$ where $\upsilon_{i,k} \sim \mathcal{U}(0,1)$

if $\Delta_i < 1$ **then** $T_i = k + \Delta_i$ and $Y_{i,k+1} = 1$ **else** $Y_{i,k+1} = 0$ **end if**

$k = k + 1$

end while

if is.null(T_i) **then** $T_i = K + 1$ **end if** ▷ Administrative censoring at $K + 1$

end for