Received: 20 February 2023 | Revised: 14 September 2023 | Accepted: 23 October 2023

DOI: 10.1002/sim.9952

RESEARCH ARTICLE

Statistics in Medicine WILEY

Statistical inference for time-to-event data in non-randomized cohorts with selective attrition

Tuo Wang[1] | **Lu Mao**[1] | **Aldo Cocco**[2] | **KyungMann Kim**[1]

[1]Department of Biostatistics and Medical Informatics, School of Medicine and Public Health, University of Wisconsin-Madison, Madison, Wisconsin, USA

[2] Indigo.ai, Milan, Italy

Correspondence
KyungMann Kim, Department of Biostatistics and Medical Informatics, School of Medicine and Public Health, University of Wisconsin-Madison, Madison, WI 53726, USA.
Email: kyungmann.kim@wisc.edu

Funding information
National Institutes of Health, Grant/Award Numbers: R01 HL149875, U01 HL130204

In multi-season clinical trials with a randomize-once strategy, patients enrolled from previous seasons who stay alive and remain in the study will be treated according to the initial randomization in subsequent seasons. To address the potentially selective attrition from earlier seasons for the non-randomized cohorts, we develop an inverse probability of treatment weighting method using season-specific propensity scores to produce unbiased estimates of survival functions or hazard ratios. Bootstrap variance estimators are used to account for the randomness in the estimated weights and the potential correlations in repeated events within each patient from season to season. Simulation studies show that the weighting procedure and bootstrap variance estimator provide unbiased estimates and valid inferences in Kaplan-Meier estimates and Cox proportional hazard models. Finally, data from the INVESTED trial are analyzed to illustrate the proposed method.

KEYWORDS
bootstrap variance estimation, composite endpoints, inverse probability of treatment weighting, propensity score

1 | INTRODUCTION

The Influenza Vaccine to Effectively Stop cardio Thoracic Events and Decompensated heart failure (INVESTED) trial (ClinicalTrials.gov Identifier: NCT02787044) was a randomized, double-blind, active-controlled, two-arm study in which two different influenza vaccines were compared for its effectiveness in reducing all-cause death or cardiopulmonary (CP) hospitalization in high-risk cardiovascular patients.[1] Patients were randomized to receive a standard dose quadrivalent inactivated influenza vaccine (SD-QIV) or a high-dose trivalent inactivated influenza vaccine (HD-TIV) for up to three influenza seasons with vaccination according to the randomize-once strategy.

As a multi-season study, patients randomized in previous seasons, if alive and not dropped out, were re-vaccinated according to the initial randomization for up to three seasons. The primary endpoint of the study is the composite of all-cause death or cardiopulmonary hospitalization in each season as the elapsed time from vaccination to the death or cardiopulmonary hospitalization, whichever occurs first. The study protocol planned to use a modified intention-to-treat (mITT) analysis in which patients' clocks would be reset two weeks after vaccination in each season, and the primary endpoint would be counted until July 31 of that season.

The unique design, as shown in Figure 1, and the mITT analysis pose two statistical challenges. First, because of the randomize-once strategy in the first year of enrollment, patients who continue enrollment in subsequent seasons are no longer comparable between the two treatment arms due to potentially differential survivorship and dropout. For

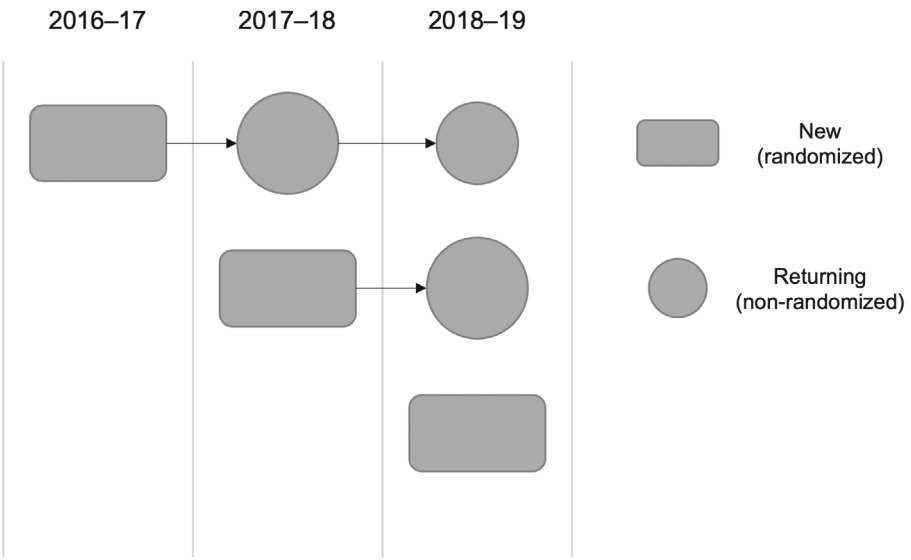

FIGURE 1 Newly enrolled and returning cohorts for each season.

example, suppose that one treatment tends to cause higher mortality or greater discomfort in sicker patients so that only healthier patients return to the current season. Then the naive analysis of the returning cohorts in the current season will be biased in favor of that treatment. Second, each patient can contribute to CP hospitalization in more than one influenza season, leading to correlated events within each patient. Therefore, the assumption of independent samples required by many standard inference procedures is violated. Traditional methods such as Kaplan-Meier estimates and Cox proportional hazard models do not account for potential bias and within-subject correlation and, therefore, may produce biased statistical inference for the difference, and underestimate of its variance, in effectiveness of the two vaccines.

To adjust for selection bias in observational studies, propensity scores have been widely used.[2-7] The propensity score is the probability of receiving a particular treatment given a vector of observed confounders. These scores can be used as the basis for matching or inverse probability of treatment weighting (IPTW) to offset confounding and thereby produce unbiased estimates of treatment effect.[8,9] Moreover, existing methods to generate propensity scores mostly rely on parametric models such as logistic regression against baseline covariates. These methods are inappropriate for the INVESTED trial because the propensity to enroll in future seasons depends crucially on the time from randomization to the start of the current season. This time dependency is hard to model correctly by using parametric models. Meanwhile, since the attrition and efficacy endpoints overlap (eg, death), their correlations are likely to be strong and must be properly accounted for to ensure valid inference.

To address these issues, we develop an IPTW method based on propensity scores estimated by survival models on the attrition endpoints and apply the weights to Kaplan-Meier estimates and Cox proportional hazard models for the efficacy endpoints. Here "treatment weighting" deals with selection bias in continuing the same treatment rather than confounding in treatment (re)assignment in each new season. For robust inference, we use bootstrap variance estimators to account for both the randomness in the estimated weights and the within-subject correlations.[10-14]

The rest of the paper is organized as follows. In Section 2, we lay out the details about how to estimate the propensity score from the observed data and how to conduct bootstrap on the multi-season data to obtain an unbiased variance estimator. In Section 3, we conduct simulations to evaluate the performance of the proposed method in different scenarios. In Section 4, the data from the INVESTED trial is used to illustrate the proposed method. Section 5 concludes with some practical considerations and discussions.

2 | INVERSE PROBABILITY OF TREATMENT WEIGHTING AND BOOTSTRAP VARIANCE ESTIMATION

2.1 | Full data

We first introduce the notation for the data. We focus on the cohort recruited in Year 1 (2016–2017); the situation with those enrolled in later years is analogous. A subset of the Year 1 cohort returned for Year 2 (2017–2018), and a further

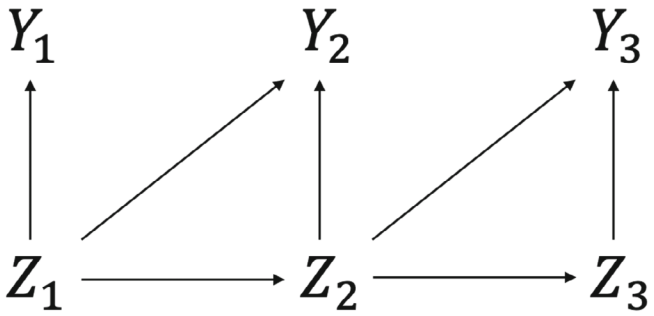

FIGURE 2 Directed acyclic graph (DAG) for patient data across multiple seasons.

subset returned for Year 3 (2018–2019). For Year k ($k = 1, 2, 3$), write $\boldsymbol{Y}_k = (T_k, C_k, \widetilde{T}_k, \widetilde{C}_k)$, where T_k is the time to death or CP hospitalization (primary efficacy endpoint), C_k is the censoring time for the primary endpoint (ie, administrative censoring or loss to follow-up), $\widetilde{T}_k$ is the time to death or dropout (attrition endpoint) and $\widetilde{C}_k$ is the administrative censoring time, with (re-)vaccination for Year k as the starting point of all event times defined above. By study design, T_2 or C_2 can be observed only if the patient returns for Year 2, that is, $\widetilde{T}_1 > \widetilde{C}_1$. Similarly, T_3 or C_3 can be observed only if $\widetilde{T}_1 > \widetilde{C}_1$ and $\widetilde{T}_2 > \widetilde{C}_2$.

2.2 | Estimation of propensity score

When using the returning sub-cohorts in the analysis of the primary endpoint, their non-random nature must be accounted for. To adjust for the potential selection bias, we define a set of covariates $\boldsymbol{Z}_k = (\boldsymbol{Z}, \boldsymbol{Z}_k^*)$, where $\boldsymbol{Z}$ contains baseline characteristics such as randomization status (time-constant by design), age, sex, race, region and clinical characteristics, and $\boldsymbol{Z}_k^*$ ($\boldsymbol{Z}_1^* \equiv 0$) contains season-specific variables such as the number of adverse events or of all-cause hospitalizations in the previous season. We estimate the "propensity score" as defined by the probability of returning to receive the (initially randomized) treatment given the covariates in previous seasons. To do so, we assume that the association between the outcome of the current season and attrition from previous seasons is accounted for by the covariates. Specifically, we make the following assumptions.

(C1) The relationship among the season-specific data follows the directed acyclic graph (DAG) depicted in Figure 2.[15]

(C2) Time to efficacy endpoint and the corresponding censoring are conditionally independent given the randomization status A, that is,

$$(T_k \perp\!\!\!\perp C_k)|A.$$

(C3) Time to attrition and the corresponding censoring are conditionally independent given the covariates, that is,

$$(\widetilde{T}_k \perp\!\!\!\perp \widetilde{C}_k)|\boldsymbol{Z}_k.$$

Remark 1. Conditions (C2) and (C3) are standard independent censoring assumptions in survival analysis to ensure that the distributions of both the efficacy and attrition endpoints are estimable from the observed data. Condition (C1) is a season-specific version of the "no unmeasured confounders" assumption[16]—that the confounders influencing both continued enrollment and the outcome in the next season are fully captured in the covariates for the current season. Some of (C1)'s most important implications are given below.

Let $\boldsymbol{\mathcal{H}}_k = \{(\boldsymbol{Y}_j, \boldsymbol{Z}_j) : j = 1, \ldots, k\}$.

(C1.1) $(\boldsymbol{Y}_{k+1} \perp\!\!\!\perp \boldsymbol{\mathcal{H}}_k)|\boldsymbol{Z}_k$;
(C1.2) $(\boldsymbol{Y}_{k+1} \perp\!\!\!\perp \boldsymbol{Y}_k)|(\boldsymbol{Z}_k, \boldsymbol{Z}_{k+1})$;
(C1.3) $(\boldsymbol{Y}_k \perp\!\!\!\perp \boldsymbol{Z}_{k+1})|\boldsymbol{Z}_k$.

In particular, (C1.1) means that outcomes in the $(k+1)$th season depend on patient life history only through covariates in the previous season.

Write $S_1(t|\boldsymbol{Z}_1) = pr(\widetilde{T}_1 > t|\boldsymbol{Z}_1)$ and $S_2(t|\boldsymbol{Z}_1, \boldsymbol{Z}_2) = pr(\widetilde{T}_2 > t|\boldsymbol{Z}_1, \boldsymbol{Z}_2)$. It can be shown that, under conditions (C1) – (C3), the selection biases in the returning sub-cohorts for season 2 (ie, those with $\widetilde{T}_1 > \widetilde{C}_1$) and for season 3 (ie, those with $\widetilde{T}_1 > \widetilde{C}_1$ and $\widetilde{T}_2 > \widetilde{C}_2$) can be adjusted by weighing the subjects by the inverse of $S_1(t|\boldsymbol{Z}_1)$ and $S_1(t|\boldsymbol{Z}_1)S_2(t|\boldsymbol{Z}_1, \boldsymbol{Z}_2)$, respectively.

Proposition 1. *Under conditions (C1) and (C3), for an arbitrary integrable function $f(\cdot)$, we have that*

$$E\{w_1 f(T_2, C_2)\} = E\{f(T_2, C_2)\} \tag{1}$$

and

$$E\{w_1 w_2 f(T_3, C_3)\} = E\{f(T_3, C_3)\}, \tag{2}$$

where

$$w_1 = \frac{I(\widetilde{T}_1 > \widetilde{C}_1)}{S_1(\widetilde{C}_1|\boldsymbol{Z}_1)} \text{ and } w_2 = \frac{I(\widetilde{T}_2 > \widetilde{C}_2)}{S_2(\widetilde{C}_2|\boldsymbol{Z}_1, Z_2)}.$$

The proof of Proposition 1 can be found in the Appendix. Let n denote the number of patients enrolled in season 1 and use subscript i to denote individual-level data. Then $E\{f(T_2, C_2)\}$ can be estimated by $\frac{1}{n}\sum_{i=1}^{n} \frac{I(\widetilde{T}_{1i} > \widetilde{C}_{1i})}{S_1(\widetilde{C}_{1i}|\boldsymbol{Z}_{1i})} f(T_{2i}, C_{2i})$ and $E\{f(T_3, C_3)\}$ can be estimated by $\frac{1}{n}\sum_{i=1}^{n} \frac{I(\widetilde{T}_{1i} > \widetilde{C}_{1i})}{S_1(\widetilde{C}_{1i}|\boldsymbol{Z}_{1i})} \frac{I(\widetilde{T}_{2i} > \widetilde{C}_{2i})}{S_2(\widetilde{C}_{2i}|\boldsymbol{Z}_{1i}, \boldsymbol{Z}_{2i})} f(T_{3i}, C_{3i})$, which only contains the data of returning patients. We can estimate $S_1(t|\boldsymbol{Z}_1)$ and $S_2(t|\boldsymbol{Z}_1, Z_2)$ using Cox proportional hazards models for attrition endpoints and obtain the estimated weights $\widehat{w}_1$ and $\widehat{w}_2$ accordingly. Then, by (C2), we can fit weighted Kaplan–Meier curves[3] or weighted Cox models[7] for valid efficacy analysis using standard statistical packages such as R-functions `survfit()` and `coxph()` with the `weights` option in the `survival` R-package.[17] In the INVESTED trial, for example, patients in season 1 can contribute to the primary endpoint up to three times. Therefore, the patients who enrolled in the study in season 1 and returned in season 2 and season 3 will be weighted by $\widehat{w}_1$ and $\widehat{w}_1\widehat{w}_2$, respectively.

Remark 2. If we consider attrition as patients being "censored" from the study cohort, then Proposition 1 suggests weights similar in form to those used in inverse probability of censoring weighting (IPCW) to adjust for dependent censoring,[18] with the survival functions of the censoring time in the denominator. The cumulative product across multiple seasons is also reminiscent of the weighting scheme in marginal structural models for time-varying treatments.[19] A major difference, however, is that our weight functions, that is, the survival functions of attrition endpoints, must be re-constructed for each season (instead of the same function re-evaluated at different times) because of the multiple season-specific outcomes for each patient.

2.3 | Bootstrap variance estimation

Once the efficacy estimates are computed, whether from Kaplan-Meier curves or Cox models, it is necessary to compute the variance of the estimates and construct confidence intervals for statistical inference. Although the IPTW produces unbiased estimates under large samples, the naive or robust sandwich-type variance estimators produced by standard software are invalid due to within-subject correlation across multiple seasons and additional randomness in the estimated propensity scores.[20-22] Austin conducted comprehensive simulations to show that naive and robust variance estimators give conservative variance estimators, whereas the bootstrap variance estimator maintains the correct coverage probability.[11] Hajage et al[13] and Shu et al[14] developed two different closed-form variance estimators for IPTW Cox models to incorporate the randomness in propensity scores. These variance estimators are not directly applicable to our case, as propensity scores are estimated using survival models, and each patient can contribute to the primary endpoint for multiple seasons. Consequently, we resort to the bootstrap method for variance estimation. Specifically, suppose there are n_k new patients enrolled in season k ($k = 1, 2, 3$). We sample with replacement n_k patients in each season k for B times (eg, $B = 1{,}000$). When a patient is sampled at enrollment, his/her data for all subsequent seasons are sampled along, including the information on whether he/she returns for the second season, and (if yes) his/her next season data. Note that for each bootstrap sample, the number of newly enrolled subjects is the same as the original data. The number of returning patients, however, is usually different from the original data.

Denote the parameter of interest θ (survival function at a time point or log-hazard ratio). Write $\widehat{\theta}$ as the estimate of θ calculated from the original data, whether it is the Kaplan-Meier estimate or the regression coefficient in a Cox model. For each bootstrap sample $j = 1, \ldots, B$, the entire estimation procedure in Section 2.2 is repeated, including the estimation of the propensity scores and weights, to obtain the estimator $\widehat{\theta}_j$. Then the bootstrap variance estimator is obtained by

$$\widehat{\mathrm{Var}}(\widehat{\theta}) = \frac{1}{B-1}\sum_{j=1}^{B}\left(\widehat{\theta}_j - \frac{1}{B}\sum_{j=1}^{B}\widehat{\theta}_j\right)^2.$$

Finally, the $100(1-\alpha)\%$ confidence interval for θ is constructed by $\left[\widehat{\theta} - z_{1-\alpha/2}\mathrm{se}(\widehat{\theta}), \widehat{\theta} + z_{1-\alpha/2}\mathrm{se}(\widehat{\theta})\right]$ where $z_{1-\alpha/2}$ is the $100(1-\alpha/2)\%$ quantile of the standard normal distribution, $\widehat{\theta}$ is the estimated parameter of interest and $\mathrm{se}(\widehat{\theta})$ is the square root of $\widehat{\mathrm{Var}}(\widehat{\theta})$. For each bootstrap sample, the propensity score model is re-estimated based on the attrition endpoint in season 1 and season 2. Therefore, the bootstrap variance estimator can capture the randomness in the estimation of propensity scores.

3 | SIMULATION STUDIES

We conducted simulation studies to assess the performance of the proposed methodology. For simplicity, we focused only on one cohort randomized in Year 1 with sub-cohorts returning for Years 2 and 3 (ie, the first line of Figure 1). To incorporate new randomized cohorts in later seasons, one only needs to set their weights to 1 under Proposition 1. For the kth season, four latent event times are of interest, namely, time to death T_{kD}, time to first CP hospitalization T_{kH}, time to dropout (loss to follow-up) C_{kL}, and the administrative censoring time C_{kM}. Using the notation of Section 2.2, we have that $T_k = T_{kD} \wedge T_{kH}$, $C_k = C_{kL} \wedge C_{kM}$, $\widetilde{T}_k = T_{kD} \wedge C_{kL}$, and $\widetilde{C}_k = C_{kM}$, where $a \wedge b = \min(a, b)$. For all the simulation studies, let treatment $A \sim$ Bernoulli(0.5).

3.1 | IPTW Kaplan-Meier estimates

In the first set of simulations, we assessed the performance of the IPTW Kaplan-Meier estimates. We considered two scenarios for potential selection bias. Scenario 1 only considers time-independent covariates at randomization, while Scenario 2 also includes season-specific covariates.

Scenario 1 assumes the selection bias depends only on the baseline covariate $Z \sim \mathrm{Uniform}[0, 1]$. Let $T_{kH} \sim \mathrm{Expn}(\eta \exp(\mu A))$ and $T_{kD} \sim \mathrm{Expn}(\gamma Z \exp(\mu A))$. Under this set-up, we can calculate the explicit formula of the true survival function for time-to-first event (TFE) $T_k = \min(T_{kH}, T_{kD})$ for treatment A. Specifically, we have

$$P(T_k > t|A, Z) = e^{-\gamma Z \exp(\mu A)t} e^{-\eta \exp(\mu A)t}.$$

Thus, the survival function for T_k for treatment A can be calculated by integration, which is

$$P(T_k > t|A) = e^{-\eta \exp(\mu A)t}\left(\frac{-1}{\gamma \exp(\mu A)t}\right)\left(e^{-\gamma \exp(\mu A)t} - 1\right).$$

In the INVESTED trial, the administrative censoring time is fixed for each participant as time from vaccination to July 31. Therefore, we emulated the design in the INVESTED trial and let $C_{kM} \sim \mathrm{Uniform}[1, 6]$. For simplicity, assume there is no loss to follow-up, that is, $C_{kL} = \infty$, and fix $\gamma = 0.2$, $\eta = 0.05$ and $\mu = 0.2$. Under this setup, around 70% and 52% subjects return to seasons 2 and 3, respectively. Next, we used the Cox model on the event $\left\{\widetilde{T}_k \wedge \widetilde{C}_k, I(\widetilde{T}_k \leq \widetilde{C}_k)\right\}$ to calculate the propensity scores for the season $k = 2, 3$. We considered sample size $n = 1{,}000$ in season 1 and drew 50 bootstrap samples for each simulated dataset to obtain the bootstrap variance estimator. In our explorations, 50 bootstrap samples were sufficient to guarantee accurate results, though some authors generally recommend 200 samples.[10,11] We replicated the simulations 2,000 times. For sample size $n = 500$ replicated 2,000 times, each with 50 bootstrap samples, it took approximately 8 h on our Departmental computing server (32 CPUs) to run. Graphical results for the standard (unweighted) and IPTW Kaplan-Meier estimator are shown in Figure 3. Graphical results for the naive and bootstrap

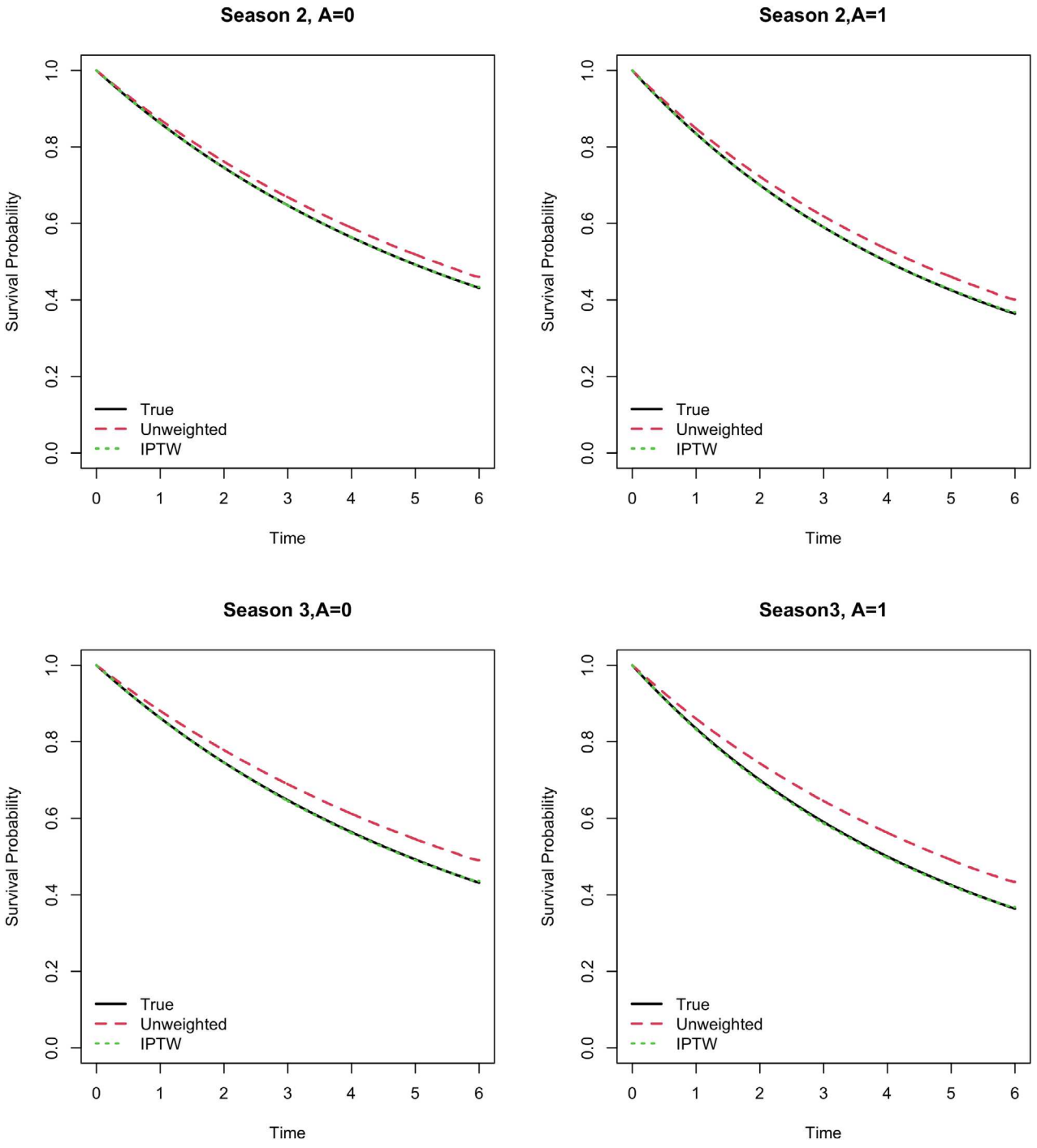

FIGURE 3 Unweighted and IPTW Kaplan-Meier estimator simulation results under Scenario 1 with $\gamma = 0.2$, $\eta = 0.05$, $\mu = 0.2$, $n = 1{,}000$, $B = 50$.

variance estimator for the IPTW Kaplan-Meier estimator are shown in Figure 4. The results of the survival functions at some selected time points are presented in Table 1.

Scenario 2 assumes the selection bias depends on both the time-independent covariate Z and season-specific covariate Z_k^*. Let $Z \sim \text{Uniform}[0,1]$ and, for simplicity, $Z_k^* \sim \text{Poisson}(2+3A)$ for $k = 2, 3$ and $\boldsymbol{Z}_1^* \equiv 0$. Let $T_{kH} \sim \text{Expn}(\eta \exp(\mu A))$ and $T_{kD} \sim \text{Expn}\big((\gamma Z + \delta Z_k^*)\exp(\mu A)\big)$. Under this setup, the true survival function for TFE in season k for treatment A is

$$P(T_k > t|A) = e^{-\eta \exp(\mu A)t}\left(\int_0^1 e^{-\gamma z \exp(\mu A)t} \mathrm{d}z\right) e^{(2+3A)(\exp(-\delta \exp(\mu A)t)-1)}.$$

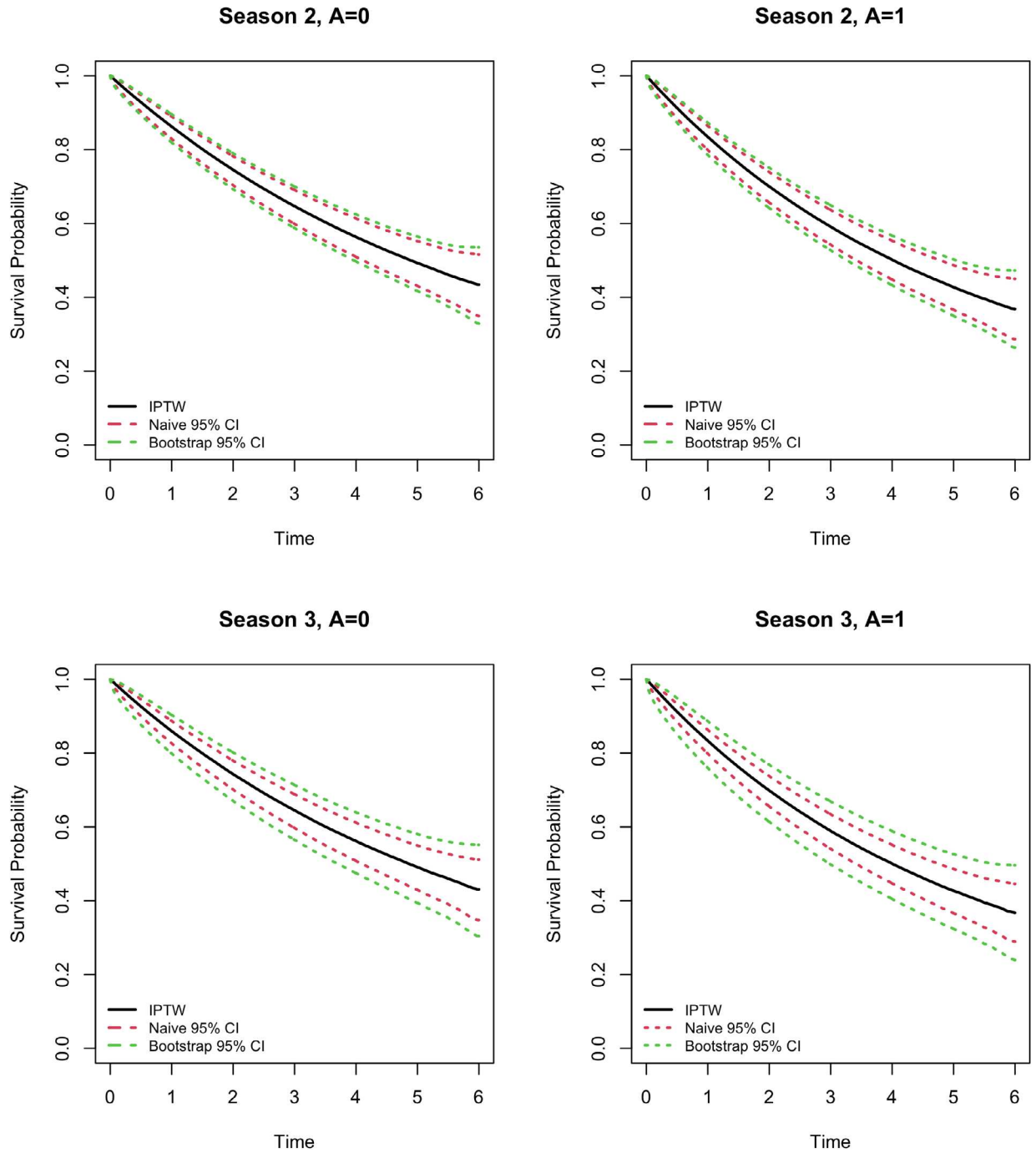

FIGURE 4 95% confidence intervals based on naive variance estimator and bootstrapped variance estimator under Scenario 1 with $\gamma = 0.2$, $\eta = 0.05$, $\mu = 0.2$, $n = 1{,}000$, $B = 50$.

Let $C_{kM} \sim \text{Uniform}[1, 6]$ and $C_{kL} = \infty$. Fix $\gamma = 0.2$, $\eta = 0.05$, $\delta = 0.01$, and $\mu = 0.2$. Under this setup, around 70% and 45% subjects return to season 2 and 3, respectively. We used the Cox model on the event $\{\widetilde{T}_k \wedge \widetilde{C}_k, I(\widetilde{T}_k \leq \widetilde{C}_k)\}$ to calculate the propensity scores. We considered sample size $n = 1{,}000$ in season 1 and drew 50 bootstrap samples. We replicated the simulation 2,000 times. Graphic results for the unweighted and IPTW Kaplan-Meier estimator are shown in Figure 5. Graphic results for the 95% confidence intervals based on the naive and bootstrapped standard error estimator for the IPTW Kaplan-Meier estimator are shown in Figure 6. The results of the survival functions at some selected time points are presented in Table 2.

For both scenarios 1 and 2, the unweighted Kaplan-Meier curves are biased, with season 3 more so than season 2 as patients go through another season of selective attrition. By contrast, the IPTW Kaplan-Meier curves correct the biases for

TABLE 1 Simulation results for selected time points under Scenario 1.

				Unweighted				IPTW					
										Naive		Bootstrap	
Season	Treatment	t	$S(t)$	Bias	SE	SEE	CP	Bias	SE	SEE	CP	SEE	CP
2	0	0.5	0.928	0.006	0.013	0.013	0.957	0.000	0.014	0.011	0.892	0.015	0.950
		1.0	0.862	0.010	0.017	0.017	0.925	0.000	0.019	0.015	0.883	0.019	0.950
		1.5	0.802	0.013	0.021	0.021	0.906	0.000	0.023	0.018	0.882	0.022	0.949
		2.0	0.746	0.016	0.023	0.023	0.904	0.000	0.025	0.020	0.885	0.025	0.950
		3.0	0.647	0.021	0.027	0.027	0.894	0.000	0.029	0.023	0.896	0.029	0.947
		4.0	0.564	0.025	0.031	0.031	0.884	0.000	0.033	0.027	0.890	0.033	0.946
		5.0	0.492	0.027	0.036	0.036	0.888	0.001	0.038	0.031	0.886	0.038	0.941
	1	0.5	0.913	0.007	0.014	0.015	0.941	0.000	0.017	0.013	0.853	0.017	0.946
		1.0	0.835	0.014	0.019	0.019	0.911	0.000	0.022	0.017	0.862	0.022	0.952
		1.5	0.764	0.019	0.022	0.022	0.887	0.000	0.025	0.019	0.868	0.025	0.952
		2.0	0.700	0.023	0.024	0.025	0.870	0.000	0.027	0.021	0.874	0.028	0.954
		3.0	0.590	0.030	0.029	0.029	0.834	0.001	0.031	0.024	0.868	0.031	0.948
		4.0	0.500	0.034	0.033	0.032	0.824	0.002	0.036	0.027	0.867	0.034	0.936
		5.0	0.425	0.035	0.039	0.038	0.841	0.002	0.040	0.031	0.864	0.039	0.939
3	0	0.5	0.928	0.009	0.014	0.015	0.935	0.002	0.020	0.012	0.744	0.020	0.936
		1.0	0.862	0.018	0.020	0.020	0.886	0.003	0.027	0.015	0.729	0.026	0.938
		1.5	0.802	0.026	0.023	0.023	0.833	0.002	0.031	0.018	0.748	0.030	0.940
		2.0	0.746	0.032	0.027	0.026	0.795	0.003	0.035	0.020	0.736	0.033	0.931
		3.0	0.647	0.042	0.031	0.031	0.754	0.002	0.038	0.023	0.764	0.038	0.942
		4.0	0.564	0.048	0.036	0.035	0.735	0.002	0.043	0.026	0.766	0.042	0.943
		5.0	0.492	0.053	0.043	0.042	0.757	0.001	0.050	0.031	0.777	0.048	0.940
	1	0.5	0.913	0.015	0.016	0.016	0.898	0.001	0.026	0.012	0.680	0.024	0.927
		1.0	0.835	0.027	0.022	0.022	0.825	0.001	0.033	0.016	0.683	0.032	0.936
		1.5	0.764	0.037	0.026	0.026	0.749	0.002	0.037	0.019	0.684	0.036	0.931
		2.0	0.700	0.045	0.029	0.029	0.700	0.001	0.039	0.021	0.701	0.039	0.939
		3.0	0.590	0.057	0.033	0.033	0.649	0.001	0.044	0.024	0.724	0.044	0.942
		4.0	0.500	0.065	0.037	0.038	0.630	0.001	0.048	0.027	0.733	0.047	0.940
		5.0	0.425	0.068	0.044	0.044	0.666	0.002	0.054	0.031	0.734	0.052	0.942

Note: Bias is the absolute difference between the true survival function and its estimator parameter. SE is standard error of the survival function estimator; SEE is the mean of the standard error estimator; CP is the coverage probability of the 95% confidence interval.

both treatment groups in season 2 and 3. The bootstrap standard error estimators are very close to the empirical standard error, and the corresponding 95% confidence intervals cover the true values at approximately the nominal level. On the other hand, the naive standard error estimators underestimate the true variations, leading to under-coverage of the true values (again with season 3 more severe than season 2 due to the accumulation of seasonal biases).

3.2 | IPTW Cox proportional hazards models

Next, we conducted simulations to evaluate the performance of the IPTW Cox model. For simplicity, we let $T_{kH} = \infty$, and considered only time-independent covariates. Let the baseline covariate $Z \sim$ Bernoulli(0.5). We simulated T_{kD} from the following method:

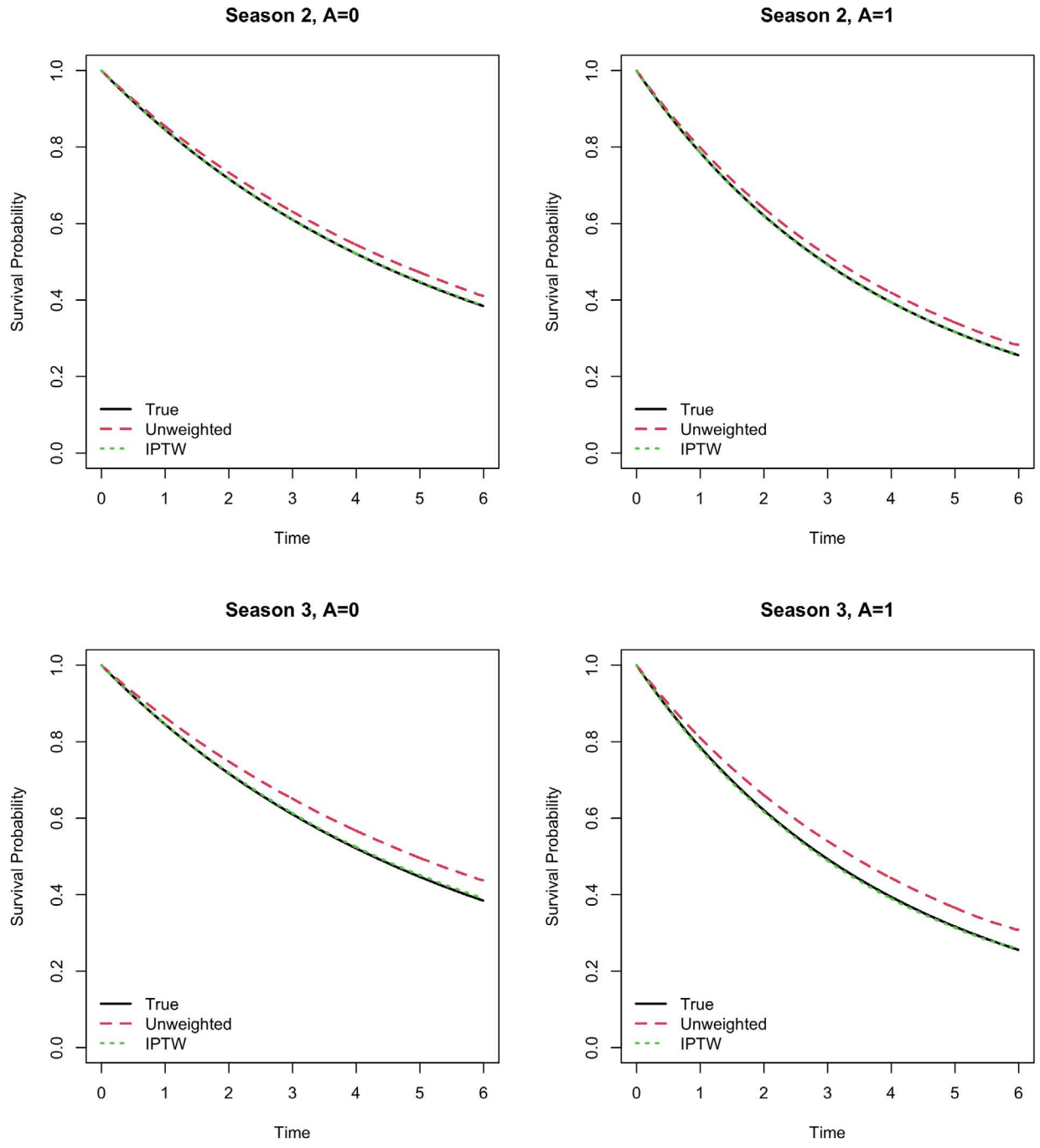

FIGURE 5 Unweighted and IPTW Kaplan-Meier estimator simulation results under Scenario 2 with $\gamma = 0.2$, $\eta = 0.05$, $\delta = 0.01$, $\mu = 0.2$, $n = 1{,}000$, $B = 50$.

- $T_{kD}|A = 0 \sim \text{Expn}(\lambda)$
- $T_{kD}|Z = 0, A = 1 \sim \text{Expn}(\eta\lambda \exp(\mu))$
- For $T_{kD}|Z = 1, A = 1$, we used numerical methods to simulate data from

$$S(t) = P(T_{kD} > t|Z = 1, A = 1) = 2\exp(-\eta\lambda \exp(\mu)t) - \exp(-\eta\lambda \exp(\mu)t).$$

Under this setup, we have that

$$P(T_{kD} > t|A) = \exp(-\lambda \exp(\mu A)t),$$

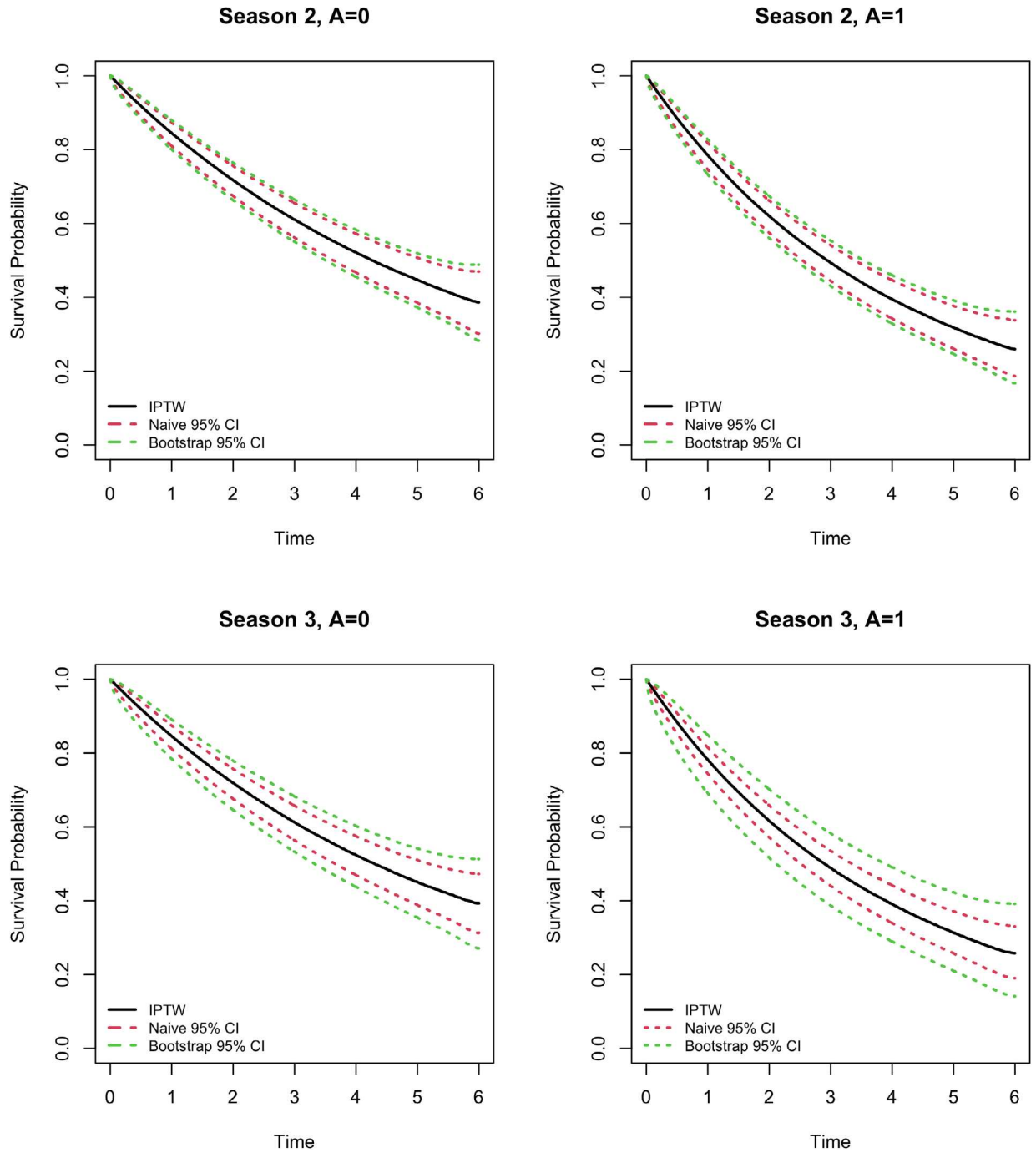

FIGURE 6 95% confidence intervals based on naive variance estimator and bootstrapped variance estimator under Scenario 2 with $\gamma = 0.2$, $\eta = 0.05$, $\delta = 0.01$, $\mu = 0.2$, $n = 1{,}000$, $B = 50$.

which satisfies the proportional hazards assumption with a log-hazard ratio of μ. Fix $\lambda = 0.09$, $\eta = 1.5$. Let $C_{kM} \sim$ Uniform$[0, 6]$ and assume no loss to follow-up. Consider the log-hazard ratios $\mu = 0.0, 0.2, 0.5$. We calculated the weights based on the method in Section 2.2. For each scenario, we considered sample size $n = 200{,}300, 500, 1{,}000, 2{,}000$, replicated the simulation 2,000 times, and drew 50 bootstrap samples. Under this setup, about 75% and 57% patients would return to seasons 2 and 3, respectively.

The results for unweighted-Cox and IPTW-Cox models with naive and bootstrap standard error estimator are summarized in Table 3. While the unweighted estimator is obvious biased, the estimates from the IPTW-Cox models are close to true log-hazard ratios. The bootstrapped standard error estimator agrees well with the empirical standard deviation

TABLE 2 Simulation results for selected time points under Scenario 2.

				Unweighted				IPTW					
										Naive		Bootstrap	
Season	Treatment	t	$S(t)$	Bias	SE	SEE	CP	Bias	SE	SEE	CP	SEE	CP
2	0	0.5	0.919	0.005	0.013	0.014	0.961	0.000	0.015	0.012	0.897	0.015	0.946
		1.0	0.845	0.009	0.018	0.018	0.933	0.000	0.020	0.016	0.878	0.020	0.941
		1.5	0.778	0.013	0.022	0.021	0.913	0.000	0.024	0.019	0.870	0.023	0.932
		2.0	0.717	0.017	0.024	0.024	0.905	0.001	0.026	0.021	0.883	0.026	0.949
		3.0	0.610	0.020	0.028	0.028	0.896	0.000	0.030	0.024	0.877	0.029	0.936
		4.0	0.521	0.023	0.032	0.031	0.881	0.000	0.033	0.027	0.884	0.033	0.937
		5.0	0.447	0.025	0.037	0.037	0.894	0.001	0.038	0.031	0.883	0.038	0.944
	1	0.5	0.886	0.007	0.016	0.017	0.953	0.001	0.019	0.014	0.879	0.019	0.948
		1.0	0.786	0.012	0.022	0.022	0.919	0.001	0.025	0.018	0.852	0.024	0.942
		1.5	0.698	0.016	0.025	0.025	0.904	0.001	0.028	0.021	0.846	0.027	0.939
		2.0	0.621	0.020	0.027	0.027	0.886	0.000	0.030	0.022	0.857	0.029	0.939
		3.0	0.493	0.024	0.030	0.030	0.874	0.001	0.032	0.025	0.862	0.032	0.945
		4.0	0.394	0.026	0.033	0.032	0.876	0.001	0.034	0.027	0.883	0.034	0.943
		5.0	0.316	0.026	0.037	0.037	0.884	0.002	0.038	0.030	0.875	0.037	0.948
3	0	0.5	0.919	0.010	0.016	0.016	0.939	0.000	0.021	0.012	0.768	0.020	0.941
		1.0	0.845	0.018	0.021	0.021	0.899	0.001	0.027	0.016	0.762	0.027	0.943
		1.5	0.778	0.025	0.025	0.025	0.864	0.001	0.031	0.019	0.771	0.031	0.944
		2.0	0.717	0.031	0.028	0.028	0.836	0.002	0.034	0.021	0.762	0.034	0.947
		3.0	0.610	0.039	0.033	0.033	0.808	0.002	0.039	0.024	0.767	0.038	0.941
		4.0	0.521	0.044	0.038	0.037	0.786	0.003	0.044	0.027	0.765	0.042	0.941
		5.0	0.447	0.048	0.045	0.043	0.788	0.004	0.051	0.031	0.762	0.048	0.931
	1	0.5	0.886	0.013	0.021	0.021	0.918	0.002	0.033	0.014	0.603	0.031	0.929
		1.0	0.786	0.024	0.027	0.028	0.893	0.003	0.042	0.018	0.605	0.040	0.932
		1.5	0.698	0.033	0.032	0.031	0.848	0.003	0.047	0.020	0.612	0.045	0.930
		2.0	0.621	0.039	0.035	0.034	0.819	0.004	0.050	0.022	0.605	0.047	0.936
		3.0	0.493	0.047	0.039	0.039	0.780	0.004	0.054	0.024	0.635	0.050	0.930
		4.0	0.394	0.050	0.043	0.042	0.794	0.003	0.056	0.026	0.662	0.052	0.919
		5.0	0.316	0.049	0.048	0.048	0.812	0.003	0.060	0.029	0.680	0.055	0.931

Note: Bias is the absolute difference between the true survival function and its estimator parameter. SE is standard error of the survival function estimator; SEE is the mean of the standard error estimator; CP is the coverage probability of the 95% confidence interval.

of the IPTW estimator. The 95% confidence interval based on bootstrapped standard error estimator maintains correct empirical coverage probability.

4 | EXAMPLE: INVESTED TRIAL

In the INVESTED trial, the primary efficacy analysis was performed on the time to the first occurrence of all-cause death or CP hospitalization within each enrollment season according to the mITT analysis described in Section 1. Based on the unadjusted Cox proportional hazard model stratified by enrollment season, the original article reported a hazard ratio of 1.06 (95% CI: [0.97 − 1.17]; p-value = 0.21).[1]

TABLE 3 Simulation results for the IPTW-Cox model.

			Unweighted				IPTW					
									Naive		Bootstrap	
log(HR)	*n*	season	Bias	SE	SEE	CP	Bias	SE	SEE	CP	SEE	CP
0.0	200	2	0.054	0.357	0.350	0.953	0.003	0.365	0.305	0.908	0.377	0.961
		3	0.106	0.411	0.405	0.950	0.007	0.431	0.309	0.859	0.494	0.971
	300	2	0.048	0.283	0.283	0.953	0.005	0.289	0.247	0.913	0.299	0.959
		3	0.105	0.335	0.327	0.941	0.002	0.348	0.249	0.849	0.363	0.962
	500	2	0.054	0.219	0.218	0.943	0.001	0.222	0.190	0.908	0.226	0.951
		3	0.110	0.250	0.251	0.933	0.006	0.260	0.190	0.855	0.271	0.959
	1000	2	0.059	0.153	0.153	0.936	0.006	0.154	0.133	0.913	0.157	0.951
		3	0.109	0.177	0.176	0.903	0.000	0.184	0.134	0.843	0.186	0.949
	2000	2	0.054	0.108	0.108	0.926	0.001	0.109	0.094	0.913	0.111	0.952
		3	0.107	0.124	0.124	0.861	0.000	0.129	0.094	0.846	0.130	0.950
0.2	200	2	0.065	0.338	0.341	0.956	0.006	0.346	0.293	0.911	0.370	0.968
		3	0.117	0.408	0.400	0.942	0.007	0.437	0.298	0.828	0.492	0.970
	300	2	0.067	0.276	0.275	0.948	0.005	0.280	0.236	0.907	0.293	0.960
		3	0.120	0.329	0.323	0.936	0.001	0.347	0.239	0.835	0.366	0.963
	500	2	0.061	0.212	0.212	0.946	0.001	0.217	0.182	0.910	0.222	0.953
		3	0.120	0.253	0.248	0.922	0.001	0.269	0.183	0.824	0.274	0.954
	1000	2	0.063	0.149	0.149	0.933	0.000	0.151	0.128	0.911	0.154	0.951
		3	0.126	0.178	0.174	0.883	0.002	0.186	0.129	0.829	0.188	0.950
	2000	2	0.061	0.106	0.105	0.914	0.001	0.107	0.090	0.905	0.108	0.949
		3	0.121	0.122	0.123	0.828	0.002	0.129	0.090	0.831	0.132	0.952
0.5	200	2	0.063	0.334	0.331	0.946	0.006	0.345	0.279	0.893	0.366	0.962
		3	0.134	0.404	0.398	0.938	0.012	0.444	0.285	0.805	0.514	0.973
	300	2	0.070	0.275	0.268	0.939	0.002	0.282	0.225	0.884	0.291	0.954
		3	0.148	0.330	0.320	0.924	0.011	0.357	0.228	0.798	0.386	0.963
	500	2	0.076	0.208	0.206	0.932	0.002	0.213	0.173	0.892	0.220	0.955
		3	0.146	0.245	0.246	0.911	0.005	0.268	0.174	0.802	0.289	0.963
	1000	2	0.076	0.147	0.145	0.918	0.000	0.151	0.122	0.887	0.153	0.950
		3	0.143	0.177	0.173	0.868	0.006	0.192	0.122	0.793	0.200	0.956
	2000	2	0.074	0.102	0.102	0.886	0.001	0.105	0.086	0.891	0.107	0.950
		3	0.150	0.119	0.122	0.770	0.001	0.128	0.086	0.813	0.139	0.961

Note: Bias is the absolute difference between the true log-hazard ratio (log(HR)) and its estimator; SE standard error of the estimator; SEE is the mean of the standard error estimator; CP is the coverage probability of the 95% confidence interval.

To illustrate the proposed method, we re-analyzed the data using the proposed propensity score approach. Based on eTable 1 in the supplementary material of Vardeny et al,[1] ejection fraction, qualifying event, history of renal impairment, history of ischemic stroke, and prior myocardial infarction (MI) showed a significant difference between patients who returned and those who did not return. Therefore, we selected the qualifying event, history of renal impairment, history of ischemic stroke, and prior MI as baseline covariates in the estimation of the propensity score. The ejection fraction was not used due to an excessive amount of missing data. In addition, a season-specific CP hospitalization event rate was also included when estimating the propensity score. In each season, the CP hospitalization event rate is defined as the number of total CP hospitalizations divided by the total duration of event-specific follow-up. We excluded one subject with missing values in the covariates in the following analyses.

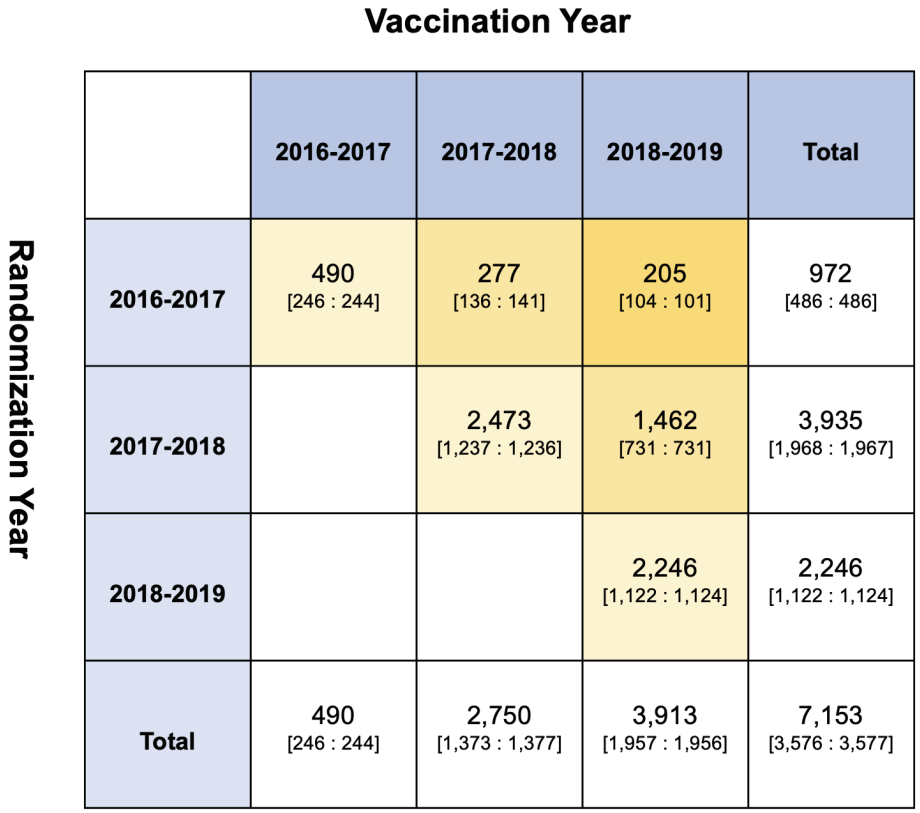

FIGURE 7 Number of subjects by randomized vaccination groups ([HD-TIV: SD-QIV]) randomized and returned in the INVESTED trial.

To compare with the original analysis in Vardeny et al,[1] we conducted the IPTW analysis on all cohorts with all newly enrolled and returning patients. After excluding subjects who didn't get vaccinations and had missing values in covariates, a total of 5,209 subjects and 7,153 subject-seasons over three influenza seasons were included for the IPTW analysis (see Figure 7). Among them, 2,605 subjects received HD-TIV and 2,604 subjects received SD-QIV. The Kaplan-Meier estimator or the log-hazard ratio in a season-stratified Cox model was calculated using the propensity score method described in Section 2.2. The variance estimators were obtained by using bootstrap method as described in Section 2.3. Specifically, we sampled with replacement 490 patients enrolled in Year 1, 2,473 enrolled in Year 2, and 2,264 enrolled in Year 3. For each bootstrapped sample, we computed the estimate of the Kaplan-Meier estimator or the log-hazard ratio in an unadjusted season-stratified Cox model (with treatment as the only covariate), using the proposed weighting procedures. We drew 1,000 bootstrap samples and re-calculated the weights for each bootstrap sample. It took 2 min on an iMac (24-inch, M1, 2021) to complete the entire analysis. Finally, we obtained the bootstrap variance estimator by the sample variance of the 1,000 bootstrap samples.

Figure 8 shows the unweighted, IPTW Kaplan-Meier estimator, and the 95% confidence intervals based on the naive and bootstrap standard error estimator for the IPTW Kaplan-Meier estimator. The IPTW Kaplan-Meier estimates are slightly smaller than the unweighted ones for both treatment groups. For the IPTW Kaplan-Meier estimator, the bootstrap confidence intervals are wider than the naive ones, which is in line with the simulation results in Section 3. In addition, an unadjusted Cox model stratified by enrolling season was used to estimate the treatment effect. The unweighted Cox model gives a hazard ratio of 1.06 (95% CI: [0.97 − 1.16]; p-value = 0.18) (slightly different than reported in the original analysis[1] due to exclusion of a few patients with missing covariates) and the IPTW Cox model gives a hazard ratio of 1.07. The naive and bootstrap standard error estimator for the IPTW Cox model are 0.04 and 0.08, yielding p-values of 0.11 and 0.44, respectively. The bootstrap standard error is bigger than the naive standard error leading to a wider confidence interval (naive 95% CI: [0.99 − 1.15]; bootstrap 95% CI: [0.90 − 1.25]). The difference between the IPTW and unweighted results appears minor because the two vaccines did not differ meaningfully in attrition. As illustrated in the simulations, however, when the treatment has a significant effect on attrition, the unweighted analyses tend to be biased,

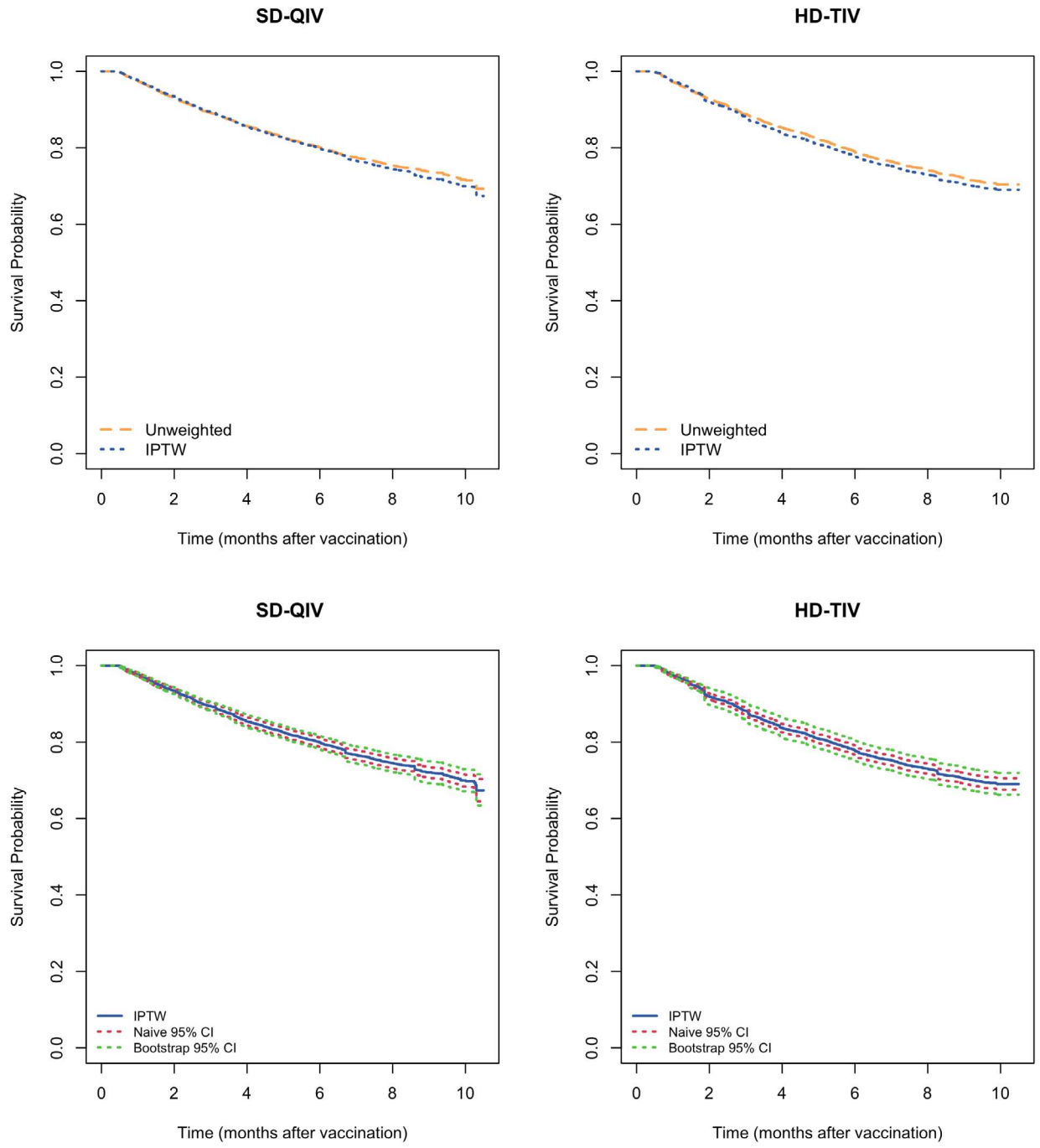

FIGURE 8 Unweighted and IPTW Kaplan-Meier estimators for INVESTED trial data with 95% confidence intervals for the IPTW estimator based on the naive variance estimator and bootstrap variance estimator.

while the IPTW analyses correct the biases. To help investigators correctly analyze other trials with a similar design, we have included the R-programs implementing the IPTW and bootstrap methodology in the Supporting Information, along with a mock data example of the INVESTED trial as an illustration.

5 | DISCUSSION

The proposed method is motivated by the special design for the INVESTED trial in which the primary endpoint is counted anew in each influenza season over multiple seasons following a randomize-once strategy. To address the non-random

cohorts in returning seasons, we have developed an IPTW inference scheme that improves upon standard procedures in two aspects. First, the conventional approach to the estimation of the propensity score typically relies on simplistic models against baseline covariates. By contrast, we use survival models for the event of discontinuation (death/dropout) to capture the temporal nature of re-enrollment in the trial. The proposed IPTW method relies on the assumption that the association between the primary and attrition endpoint of the current season and of the previous season is accounted for by the observed confounders. Second, we develop a robust variance estimation routine using the bootstrap to simultaneously account for the randomness in the estimated weights and correlations among attrition/efficacy endpoints across multiple seasons.

The simulation results show that the proposed IPTW method produces unbiased estimation of both Kaplan–Meier estimates and the (log-)hazard ratio in Cox models. The 95% confidence interval based on the bootstrap variance is much more accurate than the naive one treating weights as fixed and events as independent, which underestimates the variance. This underestimation may largely arise from ignoring the correlations between the multi-season outcomes. Even in the absence of with-patient clustering, however, Austin[11] has found that using estimated weights can add variability to the weighted estimator for some time-to-event outcomes, despite standard theory for the *vice versa*.[23] The exact reason for such apparent exceptions is unclear. In any case, we recommend the routine use of the bootstrap approach to quantify the variations in the IPTW estimators.

In the INVESTED trial the between-season attrition rates have been moderate (< 45%). If an overwhelming majority of patients died or dropped out, we might need to contend with near-zero survival functions of attrition and thus excessively large inverse weights. Methods to deal with extreme weights have been well established in the literature, including stabilization (normalize the weight by its average across similar patients) and truncation/trimming.[9,24-27] Such techniques can be readily applied to our IPTW estimators. Meanwhile, statistical efficiency can be added to the estimators by augmentation with (working) models for the season-specific outcome (instead of attrition) against covariates in the current and previous seasons.[23] This augmentation, however, needs additional derivation and programming.

Data from the INVESTED trial were used for illustration. There are a number of other trials that adopt randomize-once designs. For example, a randomized, double-blind, placebo-controlled clinical trial of the M-001 universal influenza vaccine enrolled over 12,000 seniors (ClinicalTrials.gov Identifier: NCT03450915). Subjects were vaccinated twice on days 0 and 21 according to their initial randomization for up to two influenza seasons during 2018–2020. The proposed methods can be used in such trials to address the potential selection bias in returning cohorts.

A key assumption in our methodology is that drivers of attrition are all captured as covariates–a variation of the standard "no unmeasured confounding" condition in causal inference.[16] The assumption fails, for example, if patients in the INVESTED trial missed re-vaccination for reasons other than those considered in Section 4 (eg, qualifying event, medical histories, risk of hospitalization, etc.). In the presence of such unknown factors, alternative approaches, like using an extraneous "instrumental variable",[28] could be taken to either point-estimate the treatment effect or bound its bias as sensitivity analysis.[29,30] To adapt such methods from standard settings to multi-season trials requires additional work.

ACKNOWLEDGEMENTS

This research was supported in part by grants U01 HL130204 (The INVESTED Trial Data Coordinating Center) and R01 HL149875 from the National Institutes of Health and by the University of Wisconsin-Madison Office of the Vice Chancellor for Research and Graduate Education with funding from the Wisconsin Alumni Research Foundation. The authors would like to acknowledge Scott D. Solomon, MD, of Brigham and Women's Hospital/Harvard Medical School and Orly Vardeny, PharmD, MS, of the University of Minnesota, who were Multiple Principal Investigators of the INVESTED Trial Clinical Coordinating Center. We also thank the Associate Editor and two anonymous referees for helpful comments.

DATA AVAILABILITY STATEMENT

The INVESTED data that support the findings in this paper are available on request from the corresponding author. The INVESTED data will be made available at the Biologic Specimen and Data Repository Information Coordinating Center (BioLINCC) of the National Heart, Lung, and Blood Institute.

ORCID

Tuo Wang https://orcid.org/0000-0003-1055-355X
Lu Mao https://orcid.org/0000-0002-8626-9822

REFERENCES

1. Vardeny O, Kim K, Udell JA, et al. Effect of high-dose trivalent vs standard-dose quadrivalent influenza vaccine on mortality or cardiopulmonary hospitalization in patients with high-risk cardiovascular disease: a randomized clinical trial. *Jama*. 2021;325(1):39-49.
2. Hernán MÁ, Brumback B, Robins JM. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology*. 2000;11(5):561-570.
3. Xie J, Liu C. Adjusted Kaplan–Meier estimator and log-rank test with inverse probability of treatment weighting for survival data. *Stat Med*. 2005;24(20):3089-3110.
4. Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivar Behav Res*. 2011;46(3):399-424.
5. Austin PC. A tutorial and case study in propensity score analysis: an application to estimating the effect of in-hospital smoking cessation counseling on mortality. *Multivar Behav Res*. 2011;46(1):119-151.
6. Austin PC. The performance of different propensity score methods for estimating marginal hazard ratios. *Stat Med*. 2013;32(16):2837-2849.
7. Austin PC. The use of propensity score methods with survival or time-to-event outcomes: reporting measures of effect similar to those used in randomized experiments. *Stat Med*. 2014;33(7):1242-1258.
8. Rosenbaum PR. Modern algorithms for matching in observational studies. *Ann Rev Stat Appl*. 2020;7:143-176.
9. Hernán MA, Robins JM. *Causal Inference: What If*. Boca Raton: Chapman & Hall/CRC; 2020.
10. Efron B, Tibshirani RJ. *An Introduction to the Bootstrap*. New York: Chapman & Hall/CRC; 1994.
11. Austin PC. Variance estimation when using inverse probability of treatment weighting (IPTW) with survival analysis. *Stat Med*. 2016;35(30):5642-5655.
12. Austin PC, Small DS. The use of bootstrapping when using propensity-score matching without replacement: a simulation study. *Stat Med*. 2014;33(24):4306-4319.
13. Hajage D, Chauvet G, Belin L, Lafourcade A, Tubach F, De Rycke Y. Closed-form variance estimator for weighted propensity score estimators with survival outcome. *Biom J*. 2018;60(6):1151-1163.
14. Shu D, Young JG, Toh S, Wang R. Variance estimation in inverse probability weighted cox models. *Biometrics*. 2021;77(3):1101-1117.
15. Causality PJ. *Causality*. Cambridge, UK: Cambridge University Press; 2009.
16. Rosenbaum P, Rubin D. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983;70(1):41-55.
17. Therneau TM. A Package for Survival Analysis in R. 2022 R package version 3.4-0.
18. Robins JM, Finkelstein DM. Correcting for noncompliance and dependent censoring in an AIDS clinical trial with inverse probability of censoring weighted (IPCW) log-rank tests. *Biometrics*. 2000;56(3):779-788.
19. Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology*. 2000;11(5):550-560.
20. Andersen PK, Gill RD. Cox's regression model for counting processes: a large sample study. *Ann Stat*. 1982;10(4):1100-1120.
21. Lin DY, Wei LJ. The robust inference for the cox proportional hazards model. *J Am Stat Assoc*. 1989;84(408):1074-1078.
22. Binder DA. Fitting Cox's proportional hazards models from survey data. *Biometrika*. 1992;79(1):139-147.
23. Tsiatis AA. *Semiparametric Theory and Missing Data*. New York: Springer; 2006.
24. Lee BK, Lessler J, Stuart EA. Weight trimming and propensity score weighting. *PloS One*. 2011;6(3):e18174.
25. Xiao Y, Moodie EE, Abrahamowicz M. Comparison of approaches to weight truncation for marginal structural cox models. *Epidemiol Methods*. 2013;2(1):1-20.
26. Li F, Thomas LE, Li F. Addressing extreme propensity scores via the overlap weights. *Am J Epidemiol*. 2019;188(1):250-257.
27. Chesnaye NC, Stel VS, Tripepi G, et al. An introduction to inverse probability of treatment weighting in observational research. *Clin Kidney J*. 2022;15(1):14-20.
28. Wang L, Tchetgen Tchetgen E. Bounded, efficient and multiply robust estimation of average treatment effects using instrumental variables. *J R Stat Soc Series B Stat Methodology*. 2018;80(3):531-550.
29. Van der Weele TJ, Arah OA. Bias formulas for sensitivity analysis of unmeasured confounding for general outcomes, treatments, and confounders. *Epidemiology*. 2011;22(1):42-52.
30. Mao L. Identification of the outcome distribution and sensitivity analysis under weak confounder-instrument interaction. *Stat Probab Lett*. 2022;189:109590.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Wang T, Mao L, Cocco A, Kim KM. Statistical inference for time-to-event data in non-randomized cohorts with selective attrition. *Statistics in Medicine*. 2024;43(2):216-232. doi: 10.1002/sim.9952

APPENDIX

Proof of Proposition 1. Let $f(\cdot)$ be an arbitrary integrable function of $\boldsymbol{Y}_2$ or $\boldsymbol{Y}_3$. For Equation (1), we have that

$$\begin{aligned} E\{w_1 f(\boldsymbol{Y}_2)\} &= E\big[E\{w_1 f(\boldsymbol{Y}_2)|\boldsymbol{Z}_1\}\big] \\ &= E\big[E(w_1|\boldsymbol{Z}_1)E\{f(\boldsymbol{Y}_2)|\boldsymbol{Z}_1\}\big] && \text{(by C1.1)} \\ &= E\big[E\{f(\boldsymbol{Y}_2)|\boldsymbol{Z}_1\}\big] && \text{(by C3)} \\ &= E\{f(\boldsymbol{Y}_2)\}. \end{aligned}$$

For Equation (2), we have that

$$\begin{aligned} E\{w_1 w_2 f(\boldsymbol{Y}_3)\} &= E\big[E\{w_1 w_2 f(\boldsymbol{Y}_3)|\boldsymbol{Z}_2\}\big] \\ &= E\big[E(w_1 w_2|\boldsymbol{Z}_2)E\{f(\boldsymbol{Y}_3)|\boldsymbol{Z}_2\}\big] && \text{(by C1.1),} \end{aligned}$$

where

$$\begin{aligned} E(w_1 w_2|\boldsymbol{Z}_2) &= E\{E(w_1 w_2|\boldsymbol{Z}_1, \boldsymbol{Z}_2)|\boldsymbol{Z}_2\} \\ &= E\{E(w_1|\boldsymbol{Z}_1, \boldsymbol{Z}_2)E(w_2|\boldsymbol{Z}_1, \boldsymbol{Z}_2)|\boldsymbol{Z}_2\} && \text{(by C1.2)} \\ &= E\{E(w_1|\boldsymbol{Z}_1, \boldsymbol{Z}_2)|\boldsymbol{Z}_2\} && \text{(by C3)} \\ &= E\{E(w_1|\boldsymbol{Z}_1)|\boldsymbol{Z}_2\} && \text{(by C1.3)} \\ &= 1. \end{aligned}$$

This completes the proof. ■