

Propensity score methods for estimating treatment delay effects

Jooyi Jung^a, Hyunjin Song^a, Seungbong Han^{1, a}

^aDepartment of Biostatistics, Korea University College of Medicine, Korea University

Abstract

Oftentimes, the time dependent treatment covariate and the time dependent confounders exist in observation studies. It is an important problem to correctly adjust for the time dependent confounders in the propensity score analysis. Recently, In the survival data, Hade *et al.* (2020) used a propensity score matching method to correctly estimate the treatment delay effect when the time dependent confounder affects time to the treatment time, where the treatment delay effects is defined to the delay in treatment reception. In this paper, we proposed the Cox model based marginal structural model (Cox-MSM) framework to estimate the treatment delay effect and conducted extensive simulation studies to compare our proposed Cox-MSM with the propensity score matching method proposed by Hade *et al.* (2020). Our simulation results showed that the Cox-MSM leads to more exact estimate for the treatment delay effect compared with two sequential matching schemes based on propensity scores. Example from study in treatment discontinuation in conjunction with simulated data illustrates the practical advantages of the proposed Cox-MSM.

Keywords: Cox marginal structural model, propensity score matching, survival analysis, time-dependent confounder, treatment delay effect

1. 서론

관찰 연구에서 종종 환자의 처치가 시간에 의존하는 경우가 존재하는데 임상 의사가 추적관찰 도중 환자의 상태를 보며 처치의 종류를 결정하는 경우가 그때이다. 이때 시간 의존적인 처치에 따른 처치 효과를 올바르게 추정하기 위해서는 처치에 영향을 미치는 다른 교란요인들이 적절하게 보정되어야 한다. 즉 처치의 인과 효과를 올바르게 추정하기 위해서는 처치를 받은 그룹과 받지 않은 그룹 간 다른 공변량의 균형을 적절하게 맞춘 후, 분석하는 것이 중요하다. 일반적으로, 처치의 인과 효과를 추정하기 위해서는 다양한 방법들이 존재하며, 대표적으로 성향 점수(propensity score)를 이용한 방법이 있다 (Austin, 2012; Stuart, 2010; Zhao 등, 2021). 일반적으로 이 방법들은 처치에 영향을 미치는 환자의 특성을 시간에 따라 변하지 않는, 고정된 기저 공변량을 사용하는 방법이다. 하지만, 보다 복잡한 상황인 시간-의존 교란 변수가 시간-의존 처치에 영향을 미치는 경우, 확장된 분석 방법이 요구되는데 시간-의존 교란 변수의 효과를 보정하면서 시간-의존 처치에 대한 인과 효과를 추정하는 것을 목표로 한다. 최근 Hade 등 (2020), Young 등 (2010) 는 시간-의존 교란 변수가 환자의 처치 시점에 영향을 미칠 때, 해당 효과를 보정하여 treatment delay effect를 추정하는 방법을

This work was supported in part by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) [No. 2022R1F1A1063027] and the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI) funded by the Ministry of Health & Welfare, Republic of Korea [No. HI22C045400].

¹Corresponding author: Department of Biostatistics, Korea University College of Medicine, Korea University, 73 Goryeodae-ro, Seongbuk-gu, Seoul 02841, Korea. E-mail: hanseungbong@gmail.com

제시하였다. 여기서 *treatment delay effect*란 환자가 늦게 처치를 받을 경우 제 때 받는 경우에 비해 사건 발생 위험률에 미치는 영향을 의미하며, 환자의 처치 시점은 시간-의존 처치로 작용한다. 한 시점 또는 여러 시점 늦게 처치가 가해질 경우 발생하는 *treatment delay effect*를 추정하기 위해 Hade 등 (2020)는 성향점수에 기반한 순차적 매칭(*sequential matching*) 방법을 제안하였다. 하지만 기본적으로 매칭을 수행하게 되면 적어도 얼마의 표본 수를 잃게 되어 통계적 검정력의 감소를 피할 수 없게 되는 문제점이 존재한다. 본 논문에서는 시간-의존 교란 변수가 존재하는 생존 자료에서 환자의 처치가 지연되었을 때 발생하는 *treatment delay effect*를 추정하기 위해, 적용가능한 Cox-MSM (*marginal structural model*) 프레임워크를 제시하였다. 또한 추가적으로 기존에 Hade 등 (2020)가 제시한 한 시점에서의 *delay effect* 추정법을 확장하여 여러 시점에 대한 *delay*가 존재할 때 적용가능한 추정법을 제시하였다. 본 논문은 다음 형식으로 구성되는데 2장에서는 먼저 순차적 매칭과 새롭게 제시된 Cox-MSM 방법을 설명하고 3장에서는 실제 자료에 대한 적용사례를 통해서 비교해 본다. 4장에서는 대규모 모의실험을 통해서 제시한 방법의 효용성을 설명하고 마지막으로 5장에서는 제시한 모형의 제한점 및 확장 가능성에 대해서 논의하며 앞으로의 연구 방향에 대한 고찰을 제시한다.

2. 연구 방법

2.1. Treatment delay effect 정의

먼저 우리는 *treatment delay effect*를 정의할 필요가 있는데 늦게 처치를 받은 환자와 제 때 처치를 받은 환자 간 *treatment delay effect*는 잠재적 결과 프레임워크(*potential outcomes framework*)를 통해서 정의된다. 어떤 종단 연구에서 N 명의 환자가 반복 측정되었다고 가정할 때 총 관찰 시점의 수를 T 라고 정의한다. 환자 i 의 시점 t 에서의 처치 상태 O_{it} 는 해당 시점에서 처치를 받았을 경우는 1, 받지 않은 경우는 0으로 정의하자. 이 때 처치를 받은 이후에는 처치 효과가 계속 발생한다고 가정하면, 처치를 처음 받은 시점 O_i^{tr} 이후의 시점 O_{it} 는 1의 값을 가지게 된다. 따라서, 환자의 처치 이력 $O_i = \{O_{it}\}$ 은 *monotonically non-decreasing* 벡터이며, $(O_i^{tr} - 1)$ 개의 0과 $(T - O_i^{tr} + 1)$ 개의 1을 원소로 가지는 벡터이다. 만약 연구 종료 시점까지 처치를 받지 않은 환자가 존재하는 경우 그들은 $T + 1$ 시점에서 처치를 받는다고 가정한다. 즉 이를 식으로 나타내면,

$$O_i^{tr} = \begin{cases} \min \{t \mid O_{it} = 1, t = 1, \dots, T\}, & \text{if } O_{it} = 1, \\ T + 1, & \text{otherwise.} \end{cases} \quad (2.1)$$

각 시점 t 에서, 환자 i 는 한 쌍의 잠재적 결과 변수(*potential outcome*) (Y_{it}^0, Y_{it}^1) 를 가지며, 환자의 해당 시점에서의 처치 상태에 따라 관찰된 결과 변수(*observed outcome*) $Y_i(O_i^{tr})$ 가 정의된다 (Hade 등, 2020). 예를 들어, 만약 환자 i 에 대해서 총 네 번의 반복된 측정 시점이 존재 $T = 4$ 하고 두 번째 시점에서 처치를 처음 받았다고 가정할 경우 $O_i^{tr} = 2$, 각 환자에 대한 잠재적 결과 변수는

$$Y_i = \left\{ (Y_{i1}^0, Y_{i1}^1), (Y_{i2}^0, Y_{i2}^1), (Y_{i3}^0, Y_{i3}^1), (Y_{i4}^0, Y_{i4}^1) \right\} \quad (2.2)$$

그리고 관찰된 결과 변수는 $Y_i(2) = \{Y_{i1}^0, Y_{i2}^1, Y_{i3}^1, Y_{i4}^1\}$ 로 정의한다. 이를 바탕으로 환자 i 의 시점 $t = k$ 에서 l 시점만큼 처치 지연이 발생한 경우에 *treatment delay effect* (Ψ_i^l)를 정의하면 다음과 같다.

$$\Psi_i^l = S_{iT}(k) - S_{iT}(k + l), \quad (2.3)$$

여기서 $S_{iT}(k)$ 는 환자 i 가 시점 k 에서 처치를 처음 받은 경우에 시점 T 에서의 잠재적 생존 확률(*potential survival probability*)을 의미한다.

2.2. Sequential matching scheme

여기서는 생존 자료에 대한 확장된 성향 점수 매칭 방법을 소개하고자 한다. 무작위 배정 임상 시험과 달리 관찰연구에서는 환자의 시간-의존 공변량(ex. 병의 진행 속도)에 따른 처치의 지연이 발생할 수 있다. 따라서, treatment delay effect를 올바르게 추정하기 위해서는 서로 다른 시점에서 처치를 받았지만, 공변량이 유사한 환자 간 집합군을 생성해야 하고 이 집합군에서 매칭이 수행되게 된다. 이 때, 처치를 받은 시점을 고려한 순차적 매칭(sequential matching)방법은 각 시점에서 초기 위험자 집합(risk set)내에 존재하는 환자들을 대상으로 매칭하는 방법으로, 각 위험 집합에 대해 시간 순으로 매칭이 이루어진다. 자세한 알고리즘에 대한 설명은 아래와 같다.

1. 먼저 N 명의 환자에 대해서 S 개의 서로 다른 개별 시점 ($0 \leq t_1 < t_2 \dots < t_s$) 이 존재할 때 각 시점에서 처치를 받은 환자의 수를 n_1, n_2, \dots, n_s 로 가정하자.
2. 시점 t 에서의 초기 위험자 집합은 처치를 받을 확률을 위험(risk)으로 가정했을 때, 매우 작은 양수 $\epsilon > 0$ 에 대해 $t - \epsilon$ 시점에서 처치를 받지 않은 환자와 중도 절단되지 않은 환자를 모두 포함하여 정의된다. 각 시점별로 초기 위험자 집합 내에 처치를 받은 환자와 아직 처치를 받지 못한 환자 간 매칭이 수행된다. 해당 시점에서 처치를 받은 환자를 case, 아직 처치를 받지 못한 환자를 control로 정의했을 때, 위험 집합 내에 case가 1명인 경우에는 성향 점수가 가장 유사한 control과 매칭이 수행되고, 대상자가 여럿인 경우 사건 발생 시간(event time)이 가장 빠른 case부터 control과 매칭이 수행된다.
3. 첫 시점인 t_1 에서의 위험 집합 내에는 연구에 참여한 모든 사람 즉, N 명을 포함한다. 이 때, 사전에 정의된 성향 점수에 기반하여, t_1 시점에 처치를 받은 n_1 명의 환자와 $N - n_1$ 명의 아직 처치를 받지 못한 환자 간 1 : 1 매칭을 수행한 후, n_1^m 개의 매칭 쌍(pair)을 생성한다. 해당 시점에서의 매칭이 끝나면, 다음 시점인 t_2 로 넘어가게 되고, t_2 시점의 위험 집합 내에는 t_1 에서 처치를 받은 모든 환자(case)와 이들 환자들과 매칭이 완료된 control이 배제된다. 따라서, 다음 단계의 매칭 대상군은 $N - (n_1 + n_1^m)$ 명의 환자가 존재하게 되며, t_1 시점과 동일한 매칭 과정을 거쳐서, n_2^m 개의 매칭 그룹이 생성된다.
4. 위 알고리즘을 반복하면, 모든 시점 t 에서 시간 순차적으로 각 시점별 매칭 집합을 생성하게 된다. 즉, 임의의 시점 t_k 에서의 위험 집합 내에는 $[N - (n_1 + n_2 + \dots + n_{k-1}) - (n_1^m + n_2^m + \dots + n_{k-1}^m)]$ 명의 매칭 가능한 환자가 존재하며, n_k 명의 처치를 받은 환자와 $[N - (n_1 + n_2 + \dots + n_{k-1}) - (n_1^m + n_2^m + \dots + n_{k-1}^m)] - n_k$ 명의 아직 처치를 받지 않은 환자를 대상으로, n_k 명에 대해 성향 점수가 유사한 control 환자가 매칭되게 된다.

위에서 설명한 알고리즘은 Appendix 내 toy example을 통해 추가적으로 자세히 설명하였다.

2.3. Time-dependent propensity score matching

단면 연구와 다르게 종단 연구에서의 처치 시점은 시간에 대한 함수로 고려해야 하며, 잠재적인 시간-의존 교란 변수를 보정하는 것이 중요하다. 따라서, 콕스 회귀 모형(Cox proportional hazard regression; Cox PH)에 기반한 시간-의존 성향 함수를 추정하여 매칭을 수행해야 한다 (Lu 2005; Austin 2014). 먼저 각 시점에서 환자별 기저 공변량 및 시간-의존 교란 변수를 고려하여 처치를 사건 발생(event)으로 고려한 위험(hazard rate)을 도출한다. 즉, 환자 i 의 시점 t 에서 처치를 받을 위험율은 아래와 같이 정의한다.

$$h_i(t) = h_0(t) \exp \left[\beta^T X_i(t) \right], \quad (2.4)$$

여기서 $h_0(t)$ 는 기저 위험율을 나타내고 X_i 는 기저 공변량 및 시간-의존 교란 변수를 모두 포함하며, 성향점수(propensity score)는 위험 함수(hazard function)에서 위험 점수(risk score)인 $\beta^T X_i(t)$ 로 성향 점수를 정의하게 된다. 따라서, 매칭을 위한 환자 간 성향 점수 거리(distance metric)는 아래와 같이 유클리디안 거리의 제곱을

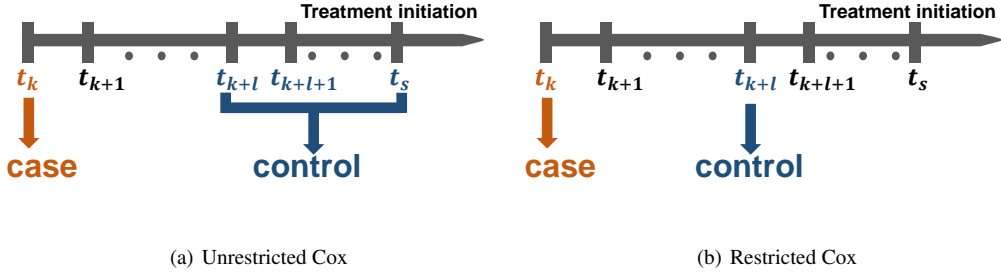


Figure 1: Sequential matching design.

사용한다.

$$\delta_r(m_i, m_j) = [\beta^T X_i(t) - \beta^T X_j(t)]^2. \quad (2.5)$$

위의 성향 점수 거리를 기반으로, 이전 section에서 정의한 순차적 매칭 프레임워크를 이용해 매칭을 수행하며 크게 두 가지 방법으로 매칭이 가능하다 (Hade 등, 2020).

1. Unrestricted Cox matching

Figure 1(a)와 같이, 시점 t_k 에서 처치를 받은 환자와 t_{k+l} 시점 이후 처치를 받은 환자 간 매칭을 수행하는 방법이다. 이 때, 매칭된 집합 내에서 delay 변수는 t_k 시점에 처치를 받은 case 환자는 제 때 처치를 받은 환자로 간주되어 0으로 정의한다. 반면, t_k 시점 이후에 처치를 받은 control 환자는 처치를 받은 시점을 t_m 이라고 가정했을 때, 매칭 시점과의 차이만큼 처치 지연이 발생하였으므로 $(t_m - t_k)$ 로 정의하며 식으로 표현하면 다음과 같다.

$$\text{Delay} = \begin{cases} 0, & \text{if treated at } t_k, \\ t_m - t_k, & \text{if treated at } t_m \text{ where } m \geq k + l. \end{cases} \quad (2.6)$$

2. Restricted Cox matching

두 번째 방법으로 Figure 1(b)와 같이, 처치를 l 단위 시점만큼 늦게 받은 경우 treatment delay effect를 확인하기 위해 시점 t_k 에서 처치를 받은 환자와 정확히 l 시점 뒤인 t_{k+l} 에서 처치를 받은 환자 간 매칭을 수행하는 방법이다. 이 경우 delay 변수는 t_k 시점에 처치를 받은 case환자는 0으로, t_{k+l} 시점에서 처치를 받은 control 환자는 l 로 정의한다.

$$\text{Delay} = \begin{cases} 0, & \text{if treated at } t_k, \\ l, & \text{if treated at } t_{k+l}. \end{cases} \quad (2.7)$$

위 두 가지 방법을 이용하여 성향 점수가 유사한 case와 control 그룹을 매칭함으로써 시간-의존 교란 변수의 효과를 보정할 수 있으며, 올바르게 treatment delay effect를 추정할 수 있다. 모든 시점에서 매칭이 완료되면, 매칭 그룹(matched set)을 층(strata)으로 하는 층화 콕스 회귀 모형(stratified Cox regression model)을 적합하여 treatment delay effect를 추정한다 (Lu 등, 2018). 총 M 개의 매칭 그룹이 생성되었을 때, m 번째 매칭 그룹 내에 존재하는 환자 i 에 대한 위험 함수(hazard function)는 식 (2.8)과 같이 정의하여 모형을 적합한다.

$$\lambda_i(t) = \lambda_{0m} \exp\{\psi \times \text{Delay}_i\}, \quad m = 1, \dots, M. \quad (2.8)$$

2.4. Cox marginal structural model

주변 구조적 모형(marginal structural model; MSM)은 종단 연구에서 교란 변수로 작용하는 시간-의존적 공변량이 존재하는 상황에서 시간-의존 처치의 인과 효과를 추정하기 위해 제안되었다. 역확률 가중치(IPTW)를 이용하여 추정하는 방법으로, 이를 통해 처치와 교란 변수간 연관성을 제거한다. 보다 확장된 모형인 콕스 주변 구조적 모형(Cox marginal structural model; Cox-MSM)은 반복 측정된 생존 자료에서 적용 가능한 MSM 모형으로, 각 측정 시점의 위험 집합 내에서 역확률 가중치를 추정하여 해당 가중치를 환자에게 부여함으로써 생성된 가상의 모집단 내에서 시간-의존 처치의 인과 효과를 추정한다 (Hernan 등, 2000; Robins 등, 2000). 이에 대한 자세한 알고리즘은 다음과 같다.

1. 종단 연구 내 N 명의 환자를 대상으로 얻어지는 생존 자료에서 환자 i 에 대한 정보를 다음과 같이 정의하자. 환자 i 의 사건 발생 시간은 T_i , 시점 t 에서의 환자 i 의 처치 유무에 대한 이항 변수는 $O_i(t)$, 환자의 기저 공변량은 X_i 그리고 시간-의존 교란 변수는 $Z_i(t)$ 로 표기한다. 즉, t 시점에서 처치를 받은 경우에는 $O_i(t) = 1$ 그리고 받지 못한 경우는 0의 값을 가진다. $O_i(t)$ 및 $Z_i(t)$ 에 대해 overbar를 붙여 표기한 변수, 즉 $\bar{O}_i(t)$, $\bar{Z}_i(t)$ 는 각각 시점 t 까지의 처치 및 시간-의존 교란 변수의 모든 과거력(history)을 의미하며, 식으로 표기하면 다음과 같다 (Xiao 등, 2010).

$$\begin{aligned}\bar{O}_i(t) &= \{O_i(u); 0 \leq u < t\}, \\ \bar{Z}_i(t) &= \{Z_i(u); 0 \leq u < t\}.\end{aligned}\quad (2.9)$$

처치 과거력 $\bar{o}(t) = \{o(t); 0 \leq t < \infty\}$ 를 따르는 환자의 가상(counterfactual) 사건 발생 시간을 $T_{\bar{o}(t)}$ 라 하면, Cox-MSM 모형은 식 (2.10)으로 정의할 수 있다.

$$\lambda_{T_{\bar{o}(t)}}(t | X) = \lambda_0(t) \exp(\psi \times f(\bar{o}(t)) + \beta^T X). \quad (2.10)$$

식 (2.9)에서 $\lambda_{T_{\bar{o}(t)}}(t | X)$ 는 기저 공변량 X 가 주어졌을 때, 시점 t 에서의 위험 함수를 나타내고, $\lambda_0(t)$ 는 시점 t 까지 처치를 받지 않은 경우 즉, $\bar{o}(t) = \bar{0}$ 일 때의 기저 위험 함수를 의미한다. $f(\cdot)$ 는 처치 과거력에 대한 함수식으로 사전에 연구자에 의해 정의된다. 본 논문에서는 treatment delay effect를 주 관심 효과로 추정하므로, 환자의 처치 시점이 k 이고, l 시점만큼 delay가 발생했을 때의 Cox-MSM 모형식은 식 (2.11)과 같이 정의한다.

$$\lambda_{T_{\bar{o}(t)}}(t | X) = \lambda_0(t) \exp(\psi \times (k - l) + \beta^T X), \quad (2.11)$$

위의 식에서 $(k - l)$ 은 treatment delay 변수를 의미한다.

2. 이전 단계에서 정의한 Cox-MSM을 적합하기 위해, 가중된 콕스 회귀 모형(weighted Cox regression)을 이용하며, 이를 위해 환자별 가중치 추정이 필요하다. 이 때, 가중치 추정을 위해 역확률 가중치(IPTW) 방법을 이용한다. 시간-의존 교란 변수의 효과를 보정하는 방법으로 가중치 또한 시간-의존적이며 측정 시점에 따라 가중치가 달라진다. 환자별 가중치를 계산하기 위해서는 처치 가중치(treatment weight)와 중도 절단 가중치(censoring weight)의 추정이 각각 필요하다. 먼저, 환자 i 에 대해 $t(1 < t < T)$ 시점에서의 처치 가중치는 식 (2.12)와 같이 정의한다.

$$w_i(t) = \prod_{j=1}^{t^*} \frac{1}{P[O_i(j) = o_i(j) | \bar{O}_i(j-1) = \bar{o}_i(j-1), X = x_i, \bar{Z}_i(j) = \bar{z}_i(j)]}. \quad (2.12)$$

식 (2.11)에서 분모는 시점 j 이전까지의 처치 과거력 및 시간-의존 교란 변수 그리고 기저 공변량이 주어졌을 때의 시점 j 에서의 처치를 받을 확률을 의미한다. 즉 t^* 를 t 보다 같거나 작은 정수로 정의했을 때, 환자

Table 1: Variable description of data on discontinued treatment

Variable	Description
Patient id	Patient registration number
tstart	The left side of the time interval for time-dependent confounder $z(t)$
tstop	The right side of the time interval for time-dependent confounder $z(t)$
$z(t)$	A time-dependent confounder
X	A baseline covariate
Y	A binary indicator of death
TY	Time to death
D	A binary indicator of treatment discontinuation
TD	Time to treatment discontinuation

i 의 시점 t 에서의 가중치는 t 이전 모든 시점 j ($1 \leq j \leq t^*$)에서의 확률에 대한 역수의 곱으로 정의된다. 이 때, 해당 가중치는 1보다 큰 값을 계속 곱해주게 되므로, 측정 시점 수가 늘어날수록 가중치가 크게 증가하게 되며, 이와 더불어 환자가 처치를 받을 확률이 0에 가까운 값을 가질 때, 가중치는 무한대로 발산하는 문제가 발생할 수 있다. 따라서, 이를 보완하기 위해 식 (2.13) 와 같이 stable 가중치(stabilized weight)를 사용한다.

$$sw_i(t) = \prod_{j=1}^{t^*} \frac{P\left[O(j) = o_i(j) \mid \bar{O}(j-1) = \bar{o}_i(j-1), X = x_i\right]}{P\left[O(j) = o_i(j) \mid \bar{O}(j-1) = \bar{o}_i(j-1), X = x_i, \bar{Z}(j) = \bar{z}_i(j)\right]} \quad (2.13)$$

$w_i(t)$ 식을 변형하여, 분자를 이전 시점까지의 처치 과거력과 기저 공변량 만을 조건부로 했을 때의 처치를 받을 확률로 대체한다. 이 stable 가중치는 측정 시점이 증가할수록 분모와 분자의 확률이 동시에 증가하므로 가중치의 변동이 감소하게 된다. 한편, 생존 자료에서 환자가 중도 절단되는 경우, 이를 고려하지 않고 처치 가중치만을 고려하여 환자에게 부여하게 되면 선택 편향(selection bias)이 발생하게 된다. 따라서, 중도 절단 가중치도 동시에 고려되어야 하며, 해당 가중치는 처치 가중치와 동일한 방법으로 정의할 수 있다. 환자 i 에 대한 시점 t 에서의 중도 절단 가중치 식은 아래와 같다.

$$sw_i^*(t) = \prod_{j=1}^{t^*} \frac{P\left[C(j) = 0 \mid \bar{C}(j-1) = 0, \bar{O}(j-1) = \bar{o}_i(j-1), X = x_i\right]}{P\left[C(j) = 0 \mid \bar{C}(j-1) = 0, \bar{O}(j-1) = \bar{o}_i(j-1), X = x_i, \bar{Z}(j) = \bar{z}_i(j)\right]} \quad (2.14)$$

식 (2.13)에서 $C(j)$ 는 시점 j 에서 중도 절단이 된 경우에는 1, 그렇지 않은 경우에는 0으로 정의한다. 기존 연구 모집단에서 중도 절단되었던 환자들은 가상의 모집단 내에서 동일한 공변량 및 처치 과거력을 가지면서 관심 사건이 발생한 환자로 대체된다. 따라서, 가상의 모집단은 중도 절단이 발생하기 이전의 모집단과 동일한 의미를 가진다. 앞서 정의한 처치 가중치와 중도 절단 가중치를 바탕으로, 시점 t 에서의 위험 집합 내의 환자 i 에게 부여되는 최종 가중치는 두 가중치의 곱으로 정의한다(i.e. $sw_i(t) \times sw_i^*(t)$). 각 시점 t 에서 환자별로 도출된 가중치를 부여하여 가상의 모집단을 생성한 뒤, 앞서 정의한 식 (2.11) 을 위험 함수로 가지는 가중된 콕스 회귀 모형을 적합한다. 이 때, 가상의 모집단 내에서는 시간-의존 교란 변수인 $z_i(t)$ 의 효과가 제거되었으므로, Cox-MSM 모형 식에 보정 변수로 포함되지 않는다.

3. 적용 사례

시간-의존 교란 변수가 존재하는 경우 treatment delay effect를 추정하는 3가지 모형 간 비교를 위해, R-package ‘contTimeCausal’ 내 자료를 이용하였다 (Yang과 Holloway, 2021). 1,000명을 대상으로 하는 처치 중단에 대

Table 2: Results of estimating delayed effects of treatment discontinuation by three methods

Method	HR	95% CI of HR		<i>p</i> -value
		LB	UB	
Unrestricted Cox	1.124	0.617	2.049	0.703
Restricted Cox	1.130	0.696	1.833	0.623
Cox-MSM	1.280	0.759	2.160	0.355

한 자료로, 자세한 변수에 대한 설명은 Table 1과 같다. 이 때, 처치 중단에 대한 delay effect를 추정하기 위해 1,000명의 전체 연구 대상자 중 처치 중단이 발생한 432명을 대상으로 분석을 수행하였다. 환자는 5주마다 시간-의존 교란 변수 $z(t)$ 와 처치 중단 발생 유무 D 에 대해 측정되었으며, X 는 첫 방문 시점에만 측정된 기저 공변량이다. 이 때, 사망 여부가 관심 사건으로 추적 기간(follow-up time) 동안 처치 중단이 발생한 환자 432명 중 159명이 사망하였으며, 273명이 중도 절단(right-censoring)되었다. 해당 자료에서 시간-의존 교란 변수인 $z(t)$ 의 효과를 보정한 상태에서 처치 중단이 5주 지연되었을 때의 delay effect를 각 모형별로 추정하였다. 올바른 인과 효과 추정을 위해 두가지의 성향 점수 매칭과 Cox-MSM, 총 3가지의 모형을 고려하였다. 먼저, 각 측정 시점별로 해당 시점에서 처치 중단이 발생한 환자와 아직 처치 중단을 하지 않은 환자 간 매칭을 수행한 후, 매칭 그룹을 층으로 하여 콕스 모형을 적합하는 unrestricted Cox matching 모형을 고려하였다. 매칭 시, 처치 중단이 발생한 환자와 처치 중단 시점이 5주 이상 차이가 나는 환자 간 매칭을 수행하였다. 두 번째로, 각 측정 시점별로 매칭을 수행할 때, 해당 시점에서 처치 중단이 발생한 환자와 정확히 5주 뒤에 처치 중단이 발생한 환자 간 매칭을 수행하는 restricted Cox matching 모형을 고려하였다. 매칭 후에는 unrestricted Cox matching과 동일하게 층화 콕스 모형을 이용하여 적합하였다. 마지막으로, 시간-의존 공변량을 고려하여 도출된 가중치를 통해 생성된 가상 모집단에서 적합한 Cox-MSM 모형을 이용하여 분석하였다. 두 가지의 성향 점수 매칭 방법에서의 성향 점수는 R 프로그램 내 'survival' 패키지의 Coxph를 이용하여 도출하였으며, 매칭은 'optmatch' 내 'pairmatch'를 이용하였다. Cox-MSM에서의 가중치 계산은 'ipw' 패키지 내 'ipwtm'을 이용하여 도출하였다 (van der Wal과 Geskus, 2011). 앞서 설명한 세 가지의 모형 간 시간-의존 처치 중단에 대한 5주 delay effect의 추정값을 비교한 결과, unrestricted Cox와 Cox-MSM의 위험비 추정값은 각각 1.124, 1.130으로 유사한 값이 도출되었으며, 처치 중단이 5주 지연될 경우 사망할 위험이 약 1.1배 높아진다고 해석할 수 있다. 반면, restricted Cox의 위험비 추정값은 1.280으로 앞서 두 모형에 비해 약간 큰 값을 가지며, 처치 중단이 5주 지연될 경우 사망할 위험이 약 1.3배 증가하는 것으로 나타났다 (Table 2). 이와 같은 차이는 restricted Cox의 경우 처치 중단 시점이 정확히 5주만 차이가 나는 환자들 간 매칭을 수행하도록 제한되기 때문에 매칭되는 환자의 수가 감소하게 되면서 발생하는 것으로 판단된다. 실제로 해당 자료를 이용하여 매칭을 수행한 결과를 살펴보면 전체 환자 432명 중 123명의 환자만 매칭이 이루어진 것을 확인할 수 있었다. 결론적으로 위험비의 추정값에서는 차이를 보였으나, *p*-value는 각각 0.703, 0.623 그리고 0.355로 유의 수준 0.05 하에서 처치 중단의 delay effect가 모두 유의하지 않음을 확인하였다. 이는 처치 중단이 5주 늦게 발생하더라도 환자의 사망에는 유의하게 영향을 미치지 않는다고 해석할 수 있다.

4. 모의 실험

4.1. 목적 및 설계

생존 자료로 이루어진 종단 연구에서 시간-의존 교란 변수가 존재하는 경우, 해당 효과를 보정하여 올바르게 treatment delay effect를 추정해야 한다. 따라서, 모의 실험을 통해 성향 점수를 이용한 서로 다른 세 가지의 방법, unrestricted Cox, restricted Cox, Cox-MSM 간 시간-의존 treatment delay effect 추정에 대해 비교하고자 한다. 모의 실험은 R 프로그램(version 4.1.0)을 이용하여 수행하였다. 표본 수(N)는 500, 1,000 그리고 treatment

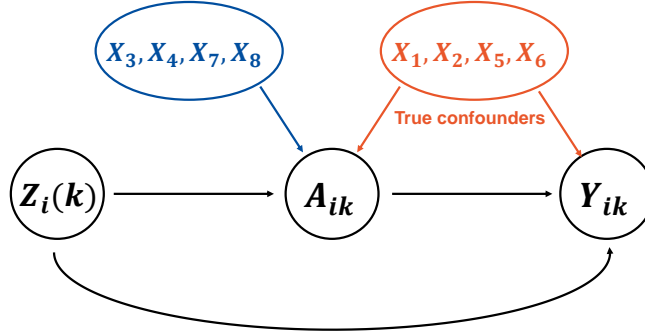


Figure 2: Causal relationships between confounders, time-dependent treatment, and outcome in simulations.

delay effect(ψ)의 크기는 $\log(1)$, $\log(1.2)$ 두 가지를 고려하였다. 측정 시점의 단위는 1주(week)로 가정하여, 처치를 받는 시점이 1주 및 2주 간 지연되는 각각의 경우에 대해 treatment delay effect를 추정하였다. 전체 표본에서 중도 절단된 환자의 비율을 20%, 30% 그리고 50%로 총 3가지 경우를 고려하였다. 공변량의 경우 시간-독립 교란 변수(time-invariant confounder)는 총 8개이며, 아래 식 (4.1), (4.2)를 이용하여 생성하였으며 표준정규 분포를 따르는 4개의 연속형 변수와 4개의 이분형 변수를 고려하였다.

$$X_1, X_2, X_3, X_4 \sim N(0, 1), \quad (4.1)$$

$$X_5, X_6, X_7, X_8 \sim \text{BIN}(1, 0.5). \quad (4.2)$$

변수 중에서 2개의 연속형 변수(X_1, X_2)와 2개의 이분형 변수(X_5, X_6)는 시간-의존 처치와 종속 변수 모두에 대해 영향을 미친다고 가정하여, 해당 변수들은 실제 교란 변수(true confounder)로 작용한다. 나머지 4개의 변수(X_3, X_4, X_7 그리고 X_8)는 시간-의존 처치에만 영향을 미치는 변수로 가정하였다. 추가적으로, 시간-의존 교란 변수(time-dependent confounder) $Z_i(t)$ 는 시간에 비례하여 증가하도록 설정하였으며, 아래식 (4.3)과 같이 정의하였다.

$$Z_i(t) := c_i t, \quad c_i \in \{0.01, 0.2\}, \quad (4.3)$$

이 때, $Z_i(t)$ 는 병의 진행 속도를 나타내는 변수이며, $c_i = 0.01$ 의 값을 가지는 환자는 병의 진행이 느린 환자를 뜻하고, 0.2는 진행이 빠른 환자를 의미한다. 이와 같이 병의 진행 정도를 두 가지로 설정하여, 대상자에 둘 중 하나의 값을 가지도록 랜덤하게 할당하였다. 정의된 교란 변수들을 바탕으로 처치 및 종속 변수와의 관계를 도식화하면 Figure 2와 같이 나타낼 수 있다 (Hernan과 Robins, 2010). 처치를 받는 시간(O_i^m)은 지수 분포(exponential distribution), 고펜르츠 분포(Gompertz distribution) 두 가지 분포를 이용하여 생성하였으며, 위험 함수는 각각 다음과 같다 (Bender 등, 2005).

$$h_{O_i}(t) = \lambda_0 \exp \left(\sum_{j=1}^8 \beta_{O_j} X_{ji} + \beta_{O_Z} Z_i(t) \right), \quad \text{where } \lambda_0 = \frac{1}{6}. \quad (4.4)$$

$$h_{O_i}(t) = \lambda_0 e^{\theta t} \exp \left(\sum_{j=1}^8 \beta_{O_j} X_{ji} + \beta_{O_Z} Z_i(t) \right), \quad \text{where } \lambda_0 = \frac{1}{6}, \theta = \frac{1}{5}. \quad (4.5)$$

이때, 기저위험함수(baseline hazard function)의 경우, 지수 분포는 상수 $\lambda_0 = 1/6$ 으로 설정하였으며, 고펜르츠 분포는 $\lambda_0 e^{\theta t} = (1/6)e^{t/5}$ 로 설정하였다. 시간-독립 교란 변수들의 효과는 $(\beta_{O_1}, \beta_{O_2}, \beta_{O_3}, \beta_{O_4}, \beta_{O_5}, \beta_{O_6}, \beta_{O_7}, \beta_{O_8}) =$

Table 3: Simulation study results of different propensity score methods for comparing treatment delay estimates

Exp(ψ)	Confounder	Model	$\widehat{\psi}$	RMSE	S.D	S.E	95% CP(%)
1	All	Unrestricted Cox	0.997	0.015	0.020	0.021	96.8
		Restricted Cox	1.012	0.092	0.130	0.131	95.8
		Cox-MSM	0.999	0.012	0.016	0.015	94.6
	True	Unrestricted Cox	0.999	0.012	0.018	0.019	96.6
		Restricted Cox	1.011	0.083	0.118	0.118	94.4
		Cox-MSM	1.000	0.011	0.015	0.019	96.0
1.2	All	Unrestricted Cox	1.189	0.025	0.033	0.026	90.6
		Restricted Cox	1.203	0.103	0.145	0.123	96.0
		Cox-MSM	1.209	0.020	0.027	0.031	96.4
	True	Unrestricted Cox	1.190	0.023	0.031	0.024	89.2
		Restricted Cox	1.201	0.092	0.130	0.111	95.8
		Cox-MSM	1.204	0.018	0.025	0.029	97.0

The table provides the average of $\widehat{\psi}$, standard deviation of $\widehat{\psi}$ (SD), average of standard error (SE), 95% coverage probability (CP,%) and root mean squared error (RMSE). The scenario settings are as follows: Censoring rate = 20%, 1 week delay effect, distribution for time to treatment = exponential distribution, N = 1,000.

Table 4: Simulation study results of different propensity score methods for comparing treatment delay estimates

Exp(ψ)	Confounder	Model	$\widehat{\psi}$	RMSE	S.D	S.E	95% CP(%)
1	All	Unrestricted Cox	0.999	0.022	0.031	0.030	95.0
		Restricted Cox	1.016	0.125	0.177	0.178	97.0
		Cox-MSM	1.001	0.019	0.028	0.033	96.2
	True	Unrestricted Cox	1.001	0.020	0.028	0.027	94.8
		Restricted Cox	1.015	0.118	0.166	0.161	95.6
		Cox-MSM	0.999	0.017	0.024	0.028	96.6
1.2	All	Unrestricted Cox	1.194	0.028	0.039	0.031	91.2
		Restricted Cox	1.204	0.130	0.183	0.154	95.8
		Cox-MSM	1.201	0.019	0.027	0.030	96.6
	True	Unrestricted Cox	1.196	0.025	0.035	0.029	93.4
		Restricted Cox	1.215	0.115	0.161	0.139	95.8
		Cox-MSM	1.199	0.016	0.020	0.020	96.0

The table provides the average of $\widehat{\psi}$, standard deviation of $\widehat{\psi}$ (SD), average of standard error (SE), 95% coverage probability (CP,%) and root mean squared error (RMSE). The scenario settings are as follows: Censoring rate = 50%, 1 week delay effect, distribution for time to treatment = exponential distribution, N = 1,000.

$\log((1.05, 1.1, 1.15, 1.2, 1.1, 1.2, 1.3, 1.4))$ 로, 시간-의존 교란 변수의 효과는 $\beta_{0Z} = \log(1.3)$ 으로 정의하였다. 처치를 받은 시간은 생성된 확률 변수 t 를 반올림하여 확률 변수보다 크지 않은 가장 작은 값으로 설정하였다. k 주(week)에 처치를 받은 환자의 기대 생존 시간(potential survival outcome)을 추정하기 위해 아래 위험 함수를 이용하였다. 해당 식에 대해 지수 분포를 따르는 확률 변수 t 를 생성한 후, 기저 상수(baseline constant value) 20주를 더해 반올림한 값으로 기대 생존 시간($Y_{Ti}(k)$)을 설정하였다.

$$h_{Y_i}^k(t) = \lambda_Y \exp(\beta_{Y1}X_{1i} + \beta_{Y2}X_{2i} + \beta_{51}X_{5i} + \beta_{61}X_{6i} + \beta_{Y2}Z_i(t) + (k - l)\psi), \quad \text{where } \lambda_Y = \frac{1}{60}. \quad (4.6)$$

도출된 확률 변수 t 에 기저 상수(baseline constant value)인 20 주를 더한 값을 반올림하여 그 값보다 크지 않은 가장 작은 값으로 기대 생존 시간을 추정하였다. 이때, 8개의 시간-독립 교란 변수 중 종속 변수에 영향을

Table 5: Simulation study results of different propensity score methods for comparing treatment delay estimates

Exp(ψ)	Confounder	Model	$\widehat{\psi}$	RMSE	S.D	S.E	95% CP(%)
1	All	Unrestricted Cox	0.999	0.028	0.040	0.042	96.4
		Restricted Cox	1.018	0.101	0.142	0.138	95.8
		Cox-MSM	0.999	0.026	0.037	0.047	96.4
	True	Unrestricted Cox	1.001	0.027	0.038	0.038	97.0
		Restricted Cox	1.013	0.091	0.128	0.128	95.4
		Cox-MSM	1.004	0.023	0.032	0.039	96.4
1.2	All	Unrestricted Cox	1.190	0.037	0.052	0.044	94.6
		Restricted Cox	1.203	0.112	0.158	0.133	95.0
		Cox-MSM	1.209	0.035	0.049	0.059	96.4
	True	Unrestricted Cox	1.192	0.036	0.050	0.040	94.4
		Restricted Cox	1.199	0.102	0.144	0.123	95.2
		Cox-MSM	1.204	0.030	0.049	0.059	96.2

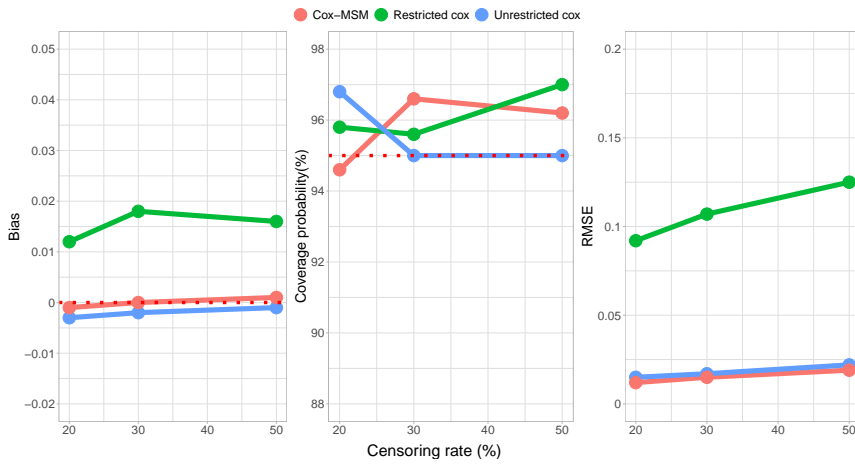
The table provides the average of $\widehat{\psi}$, standard deviation of $\widehat{\psi}$ (SD), average of standard error (SE), 95% coverage probability (CP,%) and root mean squared error (RMSE). The scenario settings are as follows: Censoring rate = 20%, 2 week delay effect, distribution for time to treatment = exponential distribution, N = 1,000.

Table 6: Simulation study results of different propensity score methods for comparing treatment delay estimates

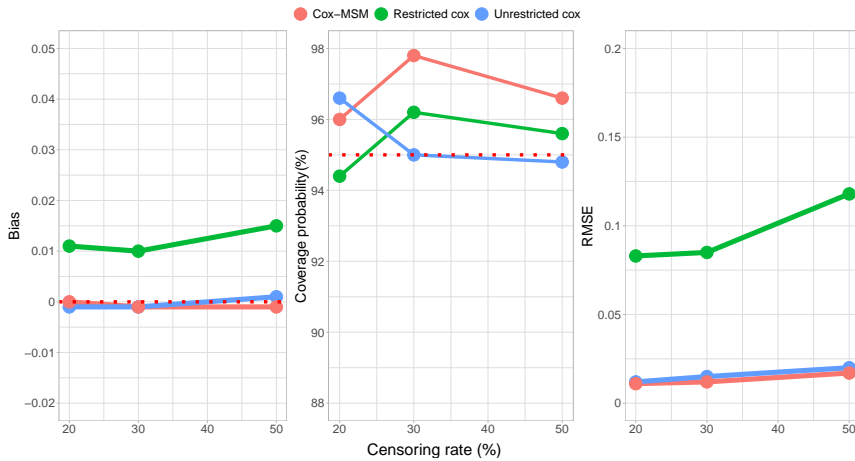
Exp(ψ)	Confounder	Model	$\widehat{\psi}$	RMSE	S.D	S.E	95% CP(%)
1	All	Unrestricted Cox	0.997	0.014	0.023	0.022	94.6
		Restricted Cox	1.013	0.101	0.142	0.133	93.8
		Cox-MSM	0.997	0.014	0.020	0.019	93.8
	True	Unrestricted Cox	0.997	0.014	0.020	0.021	96.2
		Restricted Cox	1.001	0.077	0.110	0.113	96.0
		Cox-MSM	0.997	0.013	0.018	0.017	94.6
1.2	All	Unrestricted Cox	1.188	0.021	0.028	0.028	92.2
		Restricted Cox	1.200	0.095	0.134	0.120	96.6
		Cox-MSM	1.206	0.022	0.030	0.028	96.6
	True	Unrestricted Cox	1.190	0.019	0.025	0.022	94.2
		Restricted Cox	1.199	0.084	0.119	0.102	96.4
		Cox-MSM	1.205	0.019	0.027	0.026	97.2

The table provides the average of $\widehat{\psi}$, standard deviation of $\widehat{\psi}$ (SD), average of standard error (SE), 95% coverage probability (CP,%) and root mean squared error (RMSE). The scenario settings are as follows: Censoring rate = 20%, 1 week delay effect, distribution for time to treatment = Gompertz distribution, N = 1,000.

미치는 변수는 4개로 설정하였으므로, 위험 함수 식에는 X_1, X_2, X_5, X_6 4개의 변수만 포함되는 것을 확인할 수 있으며, 변수들의 효과는 $(\beta_{Y1}, \beta_{Y2}, \beta_{Y5}, \beta_{Y6}) = \log((1.05, 1.1, 1.1, 1.2))$ 로 가정하였다. 시간-의존 교란 변수 $Z_i(t)$ 의 효과는 $\beta_{YZ} = \log(1.2)$ 로, 처치를 받는 시간을 생성할 때 정의한 효과 β_{OZ} 와 다르게 설정하였다. 본 연구의 주된 관심 효과인 treatment delay effect는 $\psi = \log(1.2)$ 로 설정하였으며, 이는 l 주 지연되어 처치를 받을 경우 사건 발생 위험(hazard)이 20% 증가함을 의미한다. 마지막으로, $\lambda_C = \lambda_Y/1.5$ 을 모수로 가지는 지수 분포를 따르는 확률 변수를 생성한 후, 기저 상수 20을 더하여 중도 절단 시간을 생성하였다. 앞서 정의한 기대 생존 시간 $Y_{Ti}(k)$ 과 비교하여, 생존 시간이 작으면 사건 발생으로, 반대의 경우는 중도 절단으로 환자의 사건 발생을 정의하였다. 모든 과정을 1,000번 반복한 뒤, unrestricted Cox matching, restricted Cox matching 그리고 Cox-MSM 모형 각각에서 treatment delay effect를 추정하여 비교 및 평가하였다.



(a) True confounders



(b) All confounders

Figure 3: Results of bias, coverage probability and root mean squared error for comparison according to the censoring rates. (a) Treatment delay effects was estimated by including only 4 confounders (X_1, X_2, X_5, X_6) for each model. (b) Treatment delay effects was estimated by including all confounders (X_1, \dots, X_8) for each model. Scenario settings are as follows. treatment delay effects $\exp(\psi) = 1$, delay unit = 1 week, time to treat = exponential distribution, $N = 1,000$.

4.2. 모의실험 결과

총 48가지의 다양한 시나리오를 가정한 모의 실험을 통해 생존 자료에서 시간-의존 교란 변수가 시간-의존 처치에 영향을 미칠 때, 교란 변수의 효과를 보정하여 treatment delay effect를 추정하였으며, 2가지의 순차적 매칭 방법과 Cox-MSM간 수행력을 비교하였다. 추정하고자 하는 treatment delay effect 크기, 전체 대상자 중 중도 절단된 비율, 처치 지연 기간, 처치 시점을 생성한 분포, 그리고 표본수에 따라 시나리오를 생성하여 비교하였다. 그 중 4개의 시나리오에 대한 결과를 Tables 3-6을 통해 확인하였으며, Figures 3-6은 12가지의

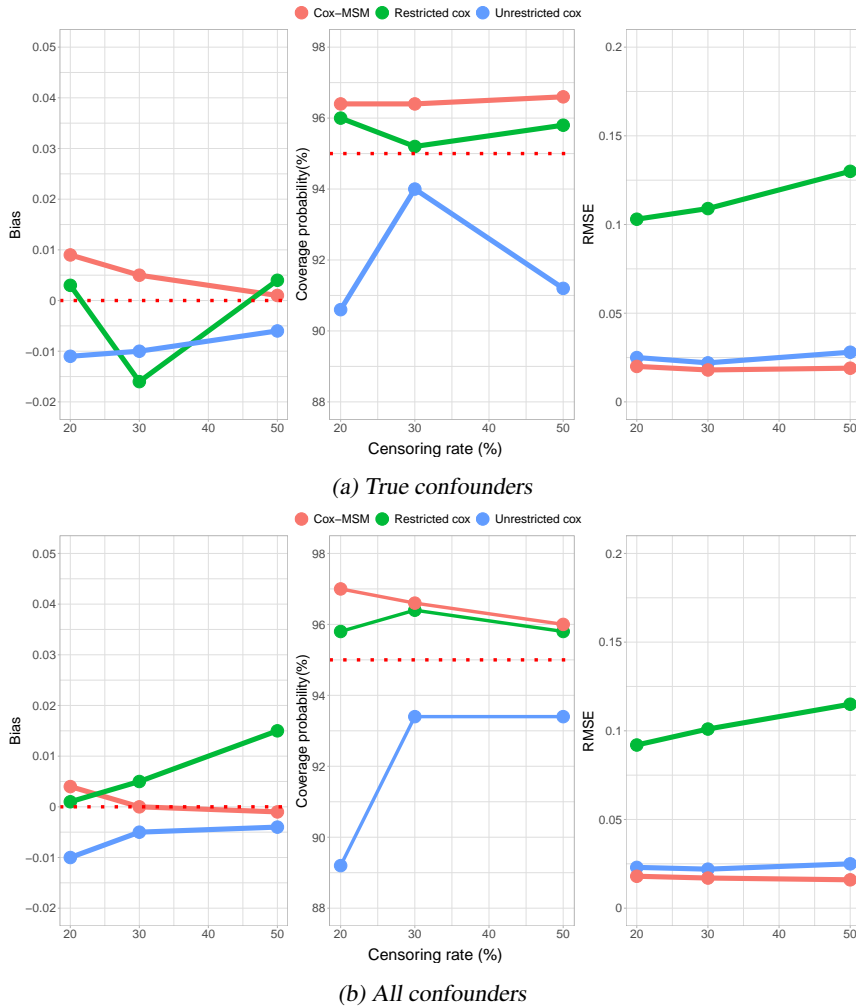


Figure 4: Results of bias, coverage probability and root mean squared error for comparison according to the censoring rates. (a) Treatment delay effects was estimated by including only 4 confounders (X_1, X_2, X_5, X_6) for each model. (b) Treatment delay effects was estimated by including all confounders (X_1, \dots, X_8) for each model. Scnario settings are as follows. treatment delay effects $\exp(\psi) = 1.2$, delay unit = 1 week, time to treat = exponential distribution, $N = 1,000$.

시나리오에 대한 결과를 제공한다. 나머지 36가지 시나리오에 대한 Figure는 Appendix B를 통해 확인할 수 있다. Tables 3–6 에서 treatment delay effect의 추정값($\hat{\psi}$), 제공근평균제곱오차(RMSE), 추정값의 표준편차 (S.D), 표준오차의 평균(S.E) 그리고 95% 포함 확률(95% CP)을 통해 세 가지 모형을 비교하였다.

Table 3은 중도 절단 비율이 20%, 1주의 treatment delay가 발생, 지수 분포에서 처치를 받은 시간 생성 그리고 표본수 $N = 1,000$ 인 시나리오에서의 세 가지 모형의 비교 결과이다. 먼저 treatment delay effect가 없는 경우에는 (i.e. $\exp(\psi) = 1$) 교란 변수의 종류에 관계없이 restricted Cox matching의 추정값이 1.012, 1.011로 실제 모수값과 가장 큰 차이를 보였으며, 제공근평균제곱오차(RMSE) 또한 0.092, 0.083으로 가장 큰 값을

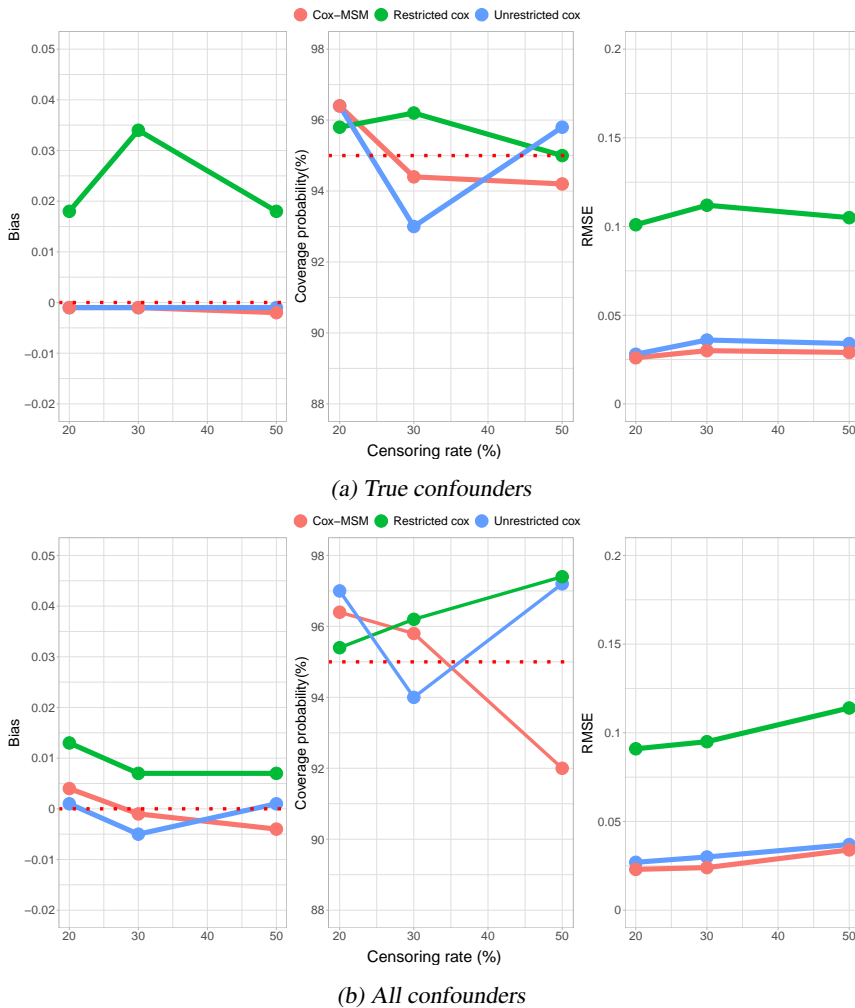


Figure 5: Results of bias, coverage probability and root mean squared error for comparison according to the censoring rates. (a) Treatment delay effects was estimated by including only 4 confounders (X_1, X_2, X_5, X_6) for each model. (b) Treatment delay effects was estimated by including all confounders (X_1, \dots, X_8) for each model. Scenario settings are as follows. treatment delay effects $\exp(\psi) = 1$, delay unit = 2 weeks, time to treat = exponential distribution, $N = 1,000$.

가지는 것을 확인하였다. 이와 반대로 treatment delay effect가 증가하게 되면(i.e. $\exp(\psi) = 1.2$), unrestricted Cox matching의 추정값이 1.189, 1.190으로 다른 두 모형에 비해 실제 모수값과 차이를 보였다. 그러나, 제공 근평균제곱오차는 restricted Cox matching에서 0.103, 0.092 값을 가지며 treatment delay effect가 없는 경우와 동일한 결과를 얻었다. Tables 4-6은 중도 절단 비율, treatment delay 시점, 처치를 받은 시간을 생성한 분포를 바꿔가며 각각 다른 시나리오 내에서 확인한 결과이며, Table 3과 유사한 결과를 제공한다. Treatment delay effect가 없을 때, 즉 $\exp(\psi) = 1$ 인 경우에는 unrestricted Cox matching 과 Cox-MSM 모형은 편향이 $-0.003 \sim -0.001$ 그리고 $-0.001 \sim 0$ 의 값을 가지며 정확도가 높았으나, restricted Cox matching 모형은 두

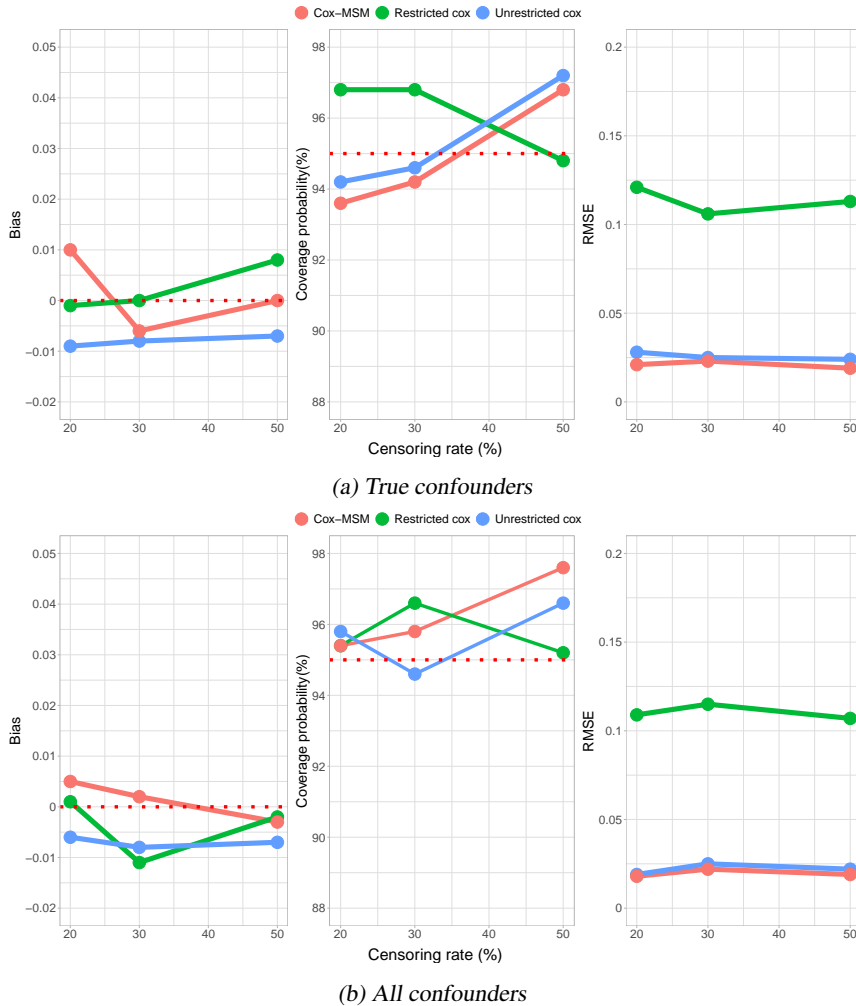


Figure 6: Results of bias, coverage probability and root mean squared error for comparison according to the censoring rates. (a) Treatment delay effects was estimated by including only 4 confounders (X_1, X_2, X_5, X_6) for each model. (b) Treatment delay effects was estimated by including all confounders (X_1, \dots, X_8) for each model. Scenario settings are as follows. treatment delay effects $\exp(\psi) = 1.2$, delay unit = 1 week, time to treat = exponential distribution, $N = 500$.

모형에 비해 편향이 약 0.01의 값을 가지며 상대적으로 과대 추정됨을 확인하였다 (Figure 3). 95% 포함 확률 (95% CP)의 경우 Cox-MSM 과 restricted Cox matching 모형은 중도 절단 비율이 증가함에 따라 포함 확률이 96% 이상이 되어 보수적(conservative)인 경향을 보인 반면, unrestricted Cox matching는 95%에 근사하는 것을 확인하였다. 한편, unrestricted Cox matching와 Cox-MSM의 경우 약 0.01의 제공근평균제곱오차(RMSE)를 가졌으나, restricted Cox matching 모형은 표준 오차의 평균이 두 모형에 비해 약 10배 큰 값을 가짐으로써, 제공근평균제곱오차가 0.08 ~ 0.09의 값으로 비교적 크게 도출되었다. Treatment delay effect가 증가하게 되면(i.e. $\exp(\psi) = 1.2$), restricted Cox matching의 경우 중도 절단 비율에 따라 편향의 정도에 차이를 보였으며,

이를 제외한 두 모형은 중도 절단 비율이 증가함에 따라 실제 모수값과 가까워지는 경향을 보였다 (Figure 4). 그러나, 앞서 treatment delay effect가 없는 경우의 결과와 다르게, 포함 확률이 95% 근방의 값을 가졌으며, unrestricted Cox matching은 포함 확률이 89%에서 94% 사이로 낮은 값을 가지는 것을 확인하였다. Unrestricted Cox matching에서 편향이 커짐에 따라 포함 확률이 낮아지게 되고, 이는 실제 지연 시점과 무관하게 매칭하는 방법이 추정에 영향을 미치는 것으로 해석할 수 있다. 제곱근평균제곱오차의 경우에는 restricted Cox matching 모형이 두 모형에 비해 큰 값을 가지며, treatment delay effect가 없는 경우와 동일한 결과를 얻었다. 두 번째로, 관심있는 delay 기간(delay unit)을 1주(week)가 아닌 2주로 설정했을 때, 표준 오차는 unrestricted Cox matching과 Cox-MSM 모형의 경우 0.03 ~ 0.04의 값을 가지며, 1주인 경우에 비해 증가하는 것을 볼 수 있다. 표준 오차가 증가함에 따라 제곱근평균제곱오차 또한 증가하는 것을 확인할 수 있다 (Figure3, Figure5). 관심있는 delay 기간이 증가하게 되면, unrestricted Cox matching의 경우에는 매칭이 가능한 control 군 내 환자가 줄어들게 되어, 1주에 비해 상대적으로 성향 점수가 덜 유사한 환자 간 매칭이 이루어질 확률이 높으며, 매칭에 성공한 매칭 집합의 수 또한 줄어들게 되므로, 분석에 사용되는 표본 수가 감소되게 되어 표준 오차가 증가하는 것으로 판단된다. 세 번째로, 표본수 $N = 500$ 의 결과는 $N = 1,000$ 에서 확인한 결과와 편향 및 제곱근평균제곱오차 측면에서 유사한 것을 확인할 수 있으나, 포함 확률의 경우 unrestricted Cox matching 모형은 $N = 500$ 에서 더 높은 값을 가졌다 (Figure 4, Figure 6). 마지막으로, 모형에 보정을 위해 포함한 교란 변수의 종류에 따른 결과를 확인했을 때, 대부분의 시나리오에서는 세 모형 간 큰 차이를 보이지 않았다. 그러나, 예외적으로 Figure 5 결과를 살펴보면, restricted Cox matching 모형에서 모든 교란 변수를 포함한 경우에 비해 참 교란 변수(true confounder)를 포함한 경우에 편향이 높은 값을 가지며, 나머지 두 모형은 유사한 결과를 제공한다. 이를 통해, 실제 연구에서 참 교란 변수에 대한 정보를 정확히 알 수 없는 상황에서는 모든 교란 변수를 포함시켜 얻는 추정값이 보다 정확한 결과를 제공할 것으로 판단된다.

요약하면, unrestricted 및 restricted Cox matching, Cox-MSM 세 모형 모두 편향, 포함 확률, 제곱근평균제곱오차가 크게 벗어나지 않은 값을 가지며, 올바르게 추정하는 것을 확인할 수 있다. 하지만, 정확도 및 정밀도 측면에서 차이를 보였으며, 우선 treatment delay effect가 없는 경우에는 restricted Cox matching의 편향이 증가하지만, 반대로 존재하는 경우에는 unrestricted Cox matching의 편향이 증가하는 것을 확인할 수 있었다. Restricted Cox matching 모형의 경우, 제곱근평균제곱오차가 모든 시나리오에서 가장 큰 값을 가지며 나머지 두 모형에 비해 낮은 성능을 가진다. 추가적으로, 중도 절단 비율 그리고 교란 변수의 종류가 변화함에 따라 두 모형에 비해 상대적으로 민감하게 반응하였다. 이를 토대로 종합해보면, 실제 연구에서 treatment delay effect를 올바르게 추정하기 위해서는 unrestricted Cox matching과 Cox-MSM 모형 사용이 적절하다고 사료된다.

5. 결론 및 고찰

치치의 효과를 평가하고자 하는 관찰연구에서 일반적으로 치치는 시간-의존 변수로 작용한다. 또한, 환자의 과거력 및 특성이 상이하기 때문에 단순히 치치를 받은 그룹과 치치를 받지 않은 그룹 간 비교하는 것은 편향을 야기할 가능성이 높다 (Hade 등, 2020). 따라서, 두 그룹 간 공변량의 균형이 이루어진 후에 치치의 효과를 평가해야 한다. 두 그룹 간 공변량을 균형화하는 연구된 다양한 방법들이 있으며, 그 중 성향 점수에 기반한 방법들이 있다 (Lu, 2005). 이 때, 일반적인 성향 점수에 기반한 방법들은 공변량을 시간에 따라 변하지 않는 변수 즉, 기저 공변량으로 고려한다. 하지만, 기저 공변량 뿐 아니라 환자의 병의 진행 속도와 같이 시간에 따라 변하는 환자의 특성이 시간-의존 치치에 영향을 주는 상황 또한 존재한다. 즉, 시간에 따라 변하는 환자의 특성은 시간-의존 교란 변수로 고려해야 하며, 이에 대한 효과를 제거한 후 치치의 효과를 평가해야 한다. 해당 효과를 제거하기 위해서는 일반적인 성향 점수에 기반한 모형 보다 복잡한 모형을 사용해야 하며, 이를 통해 올바른 인과 효과의 추정이 가능하다. Hade 등 (2020)는 시간-의존 교란 변수가 환자의 치치 시점에 영향을

미치는 경우, treatment delay effect 추정 방법을 제안하였다. 즉 환자의 처치가 관심 있는 기간만큼 지연되었을 때, 결과 변수에 미치는 영향으로 treatment delay effect를 정의한 뒤, 해당 효과를 올바르게 추정하는 모형을 소개하였다. 이 때, treatment delay effect를 추정하기 위해 성향 점수 매칭 방법에 기반한 순차적 매칭 모형을 고려하였으며, 매칭 기준에 따라 unrestricted Cox와 restricted Cox 2가지 모형을 제시하였다. 기존 연구에서는 성향 점수를 이용한 방법들 중 매칭만을 이용하여 treatment delay effect를 추정하였으나, 성향 점수에 기반한 다른 방법 또한 적용이 가능한지에 대한 연구는 아직 이루어지지 않았다. 따라서 본 논문에서는 성향 점수를 이용한 또 다른 방법인 Cox-MSM 모형 또한 올바른 treatment delay effect 추정이 가능한지를 모의실험을 통해 평가하였으며, 다양한 시나리오 내에서 기존 두 가지의 순차적 매칭 모형과의 비교를 통해 상황에 따른 적절한 모형을 제시하였다.

모의 실험은 treatment delay effect의 크기, 처치 시간 생성 분포, 중도 절단 비율, 연구 대상자 수 그리고 delay 기간을 각각 변화시키며 조합한 총 48가지의 시나리오로 진행되었다. 모형의 평가를 위해, 추정된 효과의 크기, 추정효과의 표준 편차, 표준 오차의 평균, 95% 포함 확률, 그리고 제곱근평균제곱오차 총 6개의 지표를 사용하였다. 모의 실험 결과 세 모형 모두 여러 평가 지표에서 적절하게 추정하고 있음을 확인하였으나 정확도 및 정밀도의 차이는 존재하였다. 모의 실험에서 고려한 시나리오 중, treatment delay effect의 유무에 따라 모의 실험 결과가 고려한 모형 간 차이를 보였다. Treatment delay effect가 없는 경우에는 편향 및 포함 확률 측면에서 unrestricted Cox와 Cox-MSM이 restricted Cox에 비해 나은 결과를 보였으나, 효과 크기(effect size)가 증가함에 따라, unrestricted Cox의 포함 확률이 떨어짐을 확인하였다. 이는, 효과가 없는 경우에는 매칭이 이루어지는 시점 이후에 처치를 받은 모든 환자들 내에서 control을 선택해도 무관하지만, 효과가 커짐에 따라 정확히 보고자 하는 지연 시점만큼 차이가 나는 환자들 간 매칭을 하는 것이 적절하다고 해석할 수 있다. 제곱근평균제곱오차의 경우에는 모든 시나리오 내에서 restricted Cox가 나머지 두 모형에 비해 큰 값을 얻었다. 추가적으로, 각 모형별로 평가 지표가 민감하게 반응하는 변수가 다름을 확인하였으며, 먼저 unrestricted Cox의 경우 앞서 설명한 바와 같이 treatment delay effect의 유무에 따라 결과의 차이를 보인 반면, restricted Cox의 경우에는 보정하고자 하는 교란 변수에 따라 다른 결과를 얻었다.

따라서 결과를 종합해보면, 모든 시나리오 내에서 3개의 모형 모두 추정이 잘 이루어지고 있음을 다양한 평가 지표를 통해 확인하였고, 정확도 및 정밀도 측면에서 정도의 차이를 보임을 알 수 있었다. 따라서, Cox-MSM 모형 또한 2가지의 기존 매칭 모형과 동일하게 treatment delay effect 추정이 가능하다는 결과를 알 수 있었다. 모든 시나리오 내에서 제곱근평균제곱오차가 restricted Cox 모형이 가장 큰 값을 가지므로, unrestricted Cox와 Cox-MSM을 사용하는 것이 보다 올바른 추정이 가능할 것으로 판단된다.

본 논문의 제한점으로는 첫째, 처치 자체의 효과를 추정하는 것이 아닌 delay effect를 추정하기 때문에 처치를 받은 사람만을 대상으로 분석이 가능하다는 점으로, 즉 연구 종료 시점까지 처치를 받지 않은 환자는 분석에서 제외된다. 추후, 처치를 받지 않은 환자를 제외하는 것이 아닌 분석에 포함할 수 있는 방법에 대한 추가적인 연구가 필요할 것으로 보인다. 또한, 본 논문에서는 성향 점수에 기반한 모형만을 이용하여 treatment delay effect를 추정하였으나, 그 외 G-formula를 이용한 모형에서도 동일한 결과를 얻을 수 있는지에 대한 연구 또한 이루어져야 한다.

Appendix A: Toy example for the sequential matching

Table A.1: Time to treat and time-dependent confounders for 9 patients

Time to treatment	Patient(Covariates)
Week 1	$S_1(1, 3) S_2(2, 3) S_3(3, 3)$
Week 3	$S_4(1, 3, 1) S_5(2, 3, 2) S_6(3, 3, 4) S_7(7, 3, 2)$
Week 5	$S_8(8, 5, 3, 4) S_9(7, 3, 2, 4)$

Table A.2: Results of sequential matching

Time to treatment	Patients in the risk set	Matching set
Week 1	Treated patients : S_1, S_2, S_3	$(S_1, S_4), (S_2, S_5), (S_3, S_6)$
	Untreated patients not yet : $S_4, S_5, S_6, S_7, S_8, S_9$	
Week 3	Treated patients : S_7	(S_7, S_9)
	Untreated patients not yet : S_8, S_9	

만약 하나의 시간-의존 교란변수가 처치 시점에 영향을 미칠 때, 순차적 매칭이 이루어지는 과정은 다음 예제를 통해 보여줄 수 있다. 1주, 3주, 5주 시점에 처치를 받은 9명의 환자가 존재한다고 가정하자. 환자의 병의 진행 정도가 처치 시점에 영향을 미친다고 가정하면, 해당 변수는 시간-의존 교란변수로 작용하여 측정 시점별로 값이 변하게 되고, 환자가 처치를 받는 시점까지만 측정이 된다. 예를 들어, Table A.1과 같이, 1주에 처치를 받은 환자 S_1 는 연구 시작 시점을 포함한 2개의 측정값을 가지는 반면, 3주에 처치를 받은 환자 S_4 는 연구 시작 시점 및 1주 시점을 포함한 3개의 측정값을 가지게 된다.

순차적 매칭을 수행하기 위해서, 먼저 1주에서의 위험 집합을 고려한다. 이 위험 집합 내에는 3명의 처치를 받은 환자 및 6명의 아직 처치를 받지 않은 환자로 이루어진 총 9명의 환자들이 포함된다. 처치를 받은 환자 S_1, S_2, S_3 를 매칭하기 위해, 아직 처치를 받지 않은 환자들 중 가장 유사한 공변량을 가지는 환자와 매칭을 수행한다. S_1 환자의 경우, 가장 유사한 공변량을 가지는 환자 S_4 와 매칭이 이루어진다. S_2, S_3 환자들에 대해서도 동일하게 매칭을 수행하게 되면, 1주 시점에서는 $(S_1, S_4), (S_2, S_5), (S_3, S_6)$ 총 3개의 매칭 집합이 생성된다. 1주 시점과 마찬가지로 3주에서 매칭을 수행하기 위해서는 해당 시점에서의 위험 집합을 고려해야 하는데, 이때, 1주에서 매칭이 이루어진 환자들은 위험 집합에서 제외된다. 따라서, 3주 시점에서의 위험 집합 내에는 1명의 처치를 받은 환자와 2명의 처치를 받지 않은 환자 총 3명의 환자가 포함된다. 환자 S_7 와 유사한 공변량을 가지는 환자는 S_9 이므로, 3주 시점에서는 (S_7, S_9) 로 1개의 매칭 집합이 생성된다. 매칭 결과를 요약하면 Table A.2 와 같다.

Appendix B: Supplementary figure

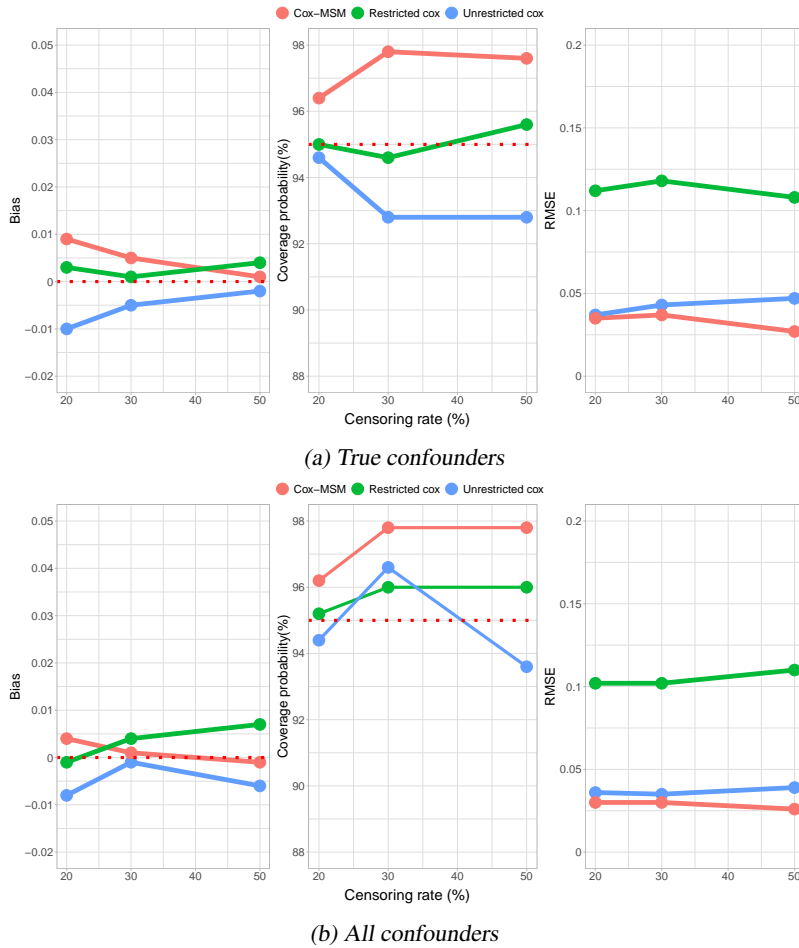


Figure B.1: Results of bias, coverage probability and root mean squared error for comparison according to the censoring rates. (a) Treatment delay effects was estimated by including only 4 confounders (X_1, X_2, X_5, X_6) for each model. (b) Treatment delay effects was estimated by including all confounders (X_1, \dots, X_8) for each model. Scenario settings are as follows. Treatment delay effects $\exp(\psi) = 1.2$, delay unit = 2 weeks, time to treat = exponential distribution, $N = 1,000$.

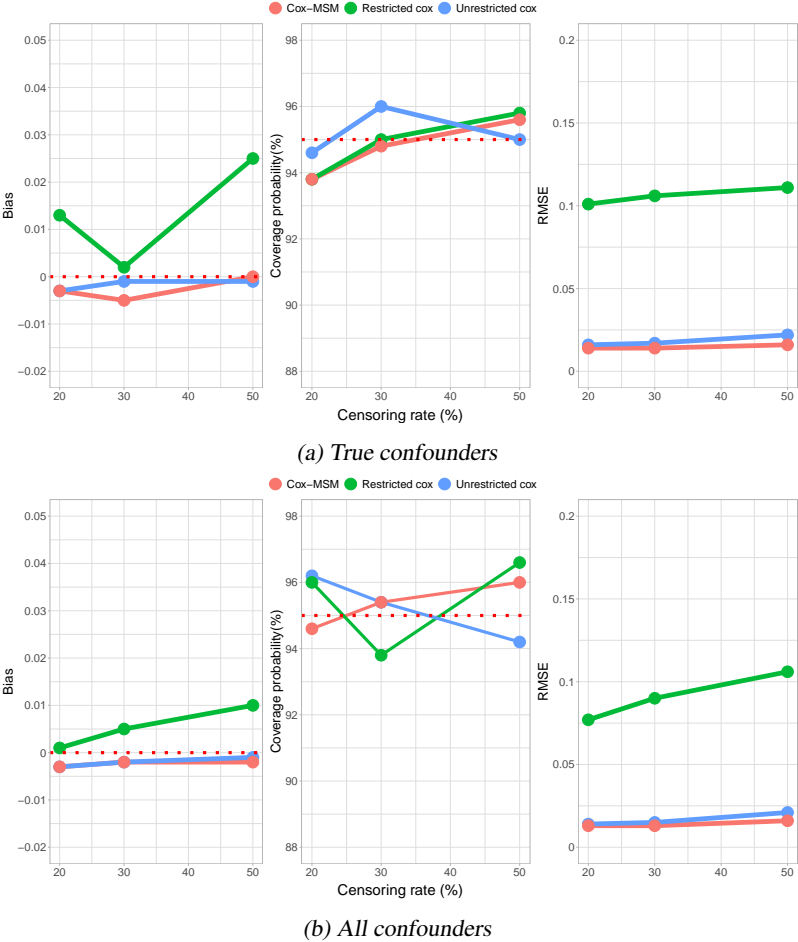


Figure B.2: Results of bias, coverage probability and root mean squared error for comparison according to the censoring rates. (a) Treatment delay effects was estimated by including only 4 confounders (X_1, X_2, X_5, X_6) for each model. (b) Treatment delay effects was estimated by including all confounders (X_1, \dots, X_8) for each model. Scenario settings are as follows. Treatment delay effects $\exp(\psi) = 1$, delay unit = 1 week, time to treat = Gompertz distribution, $N = 1,000$.

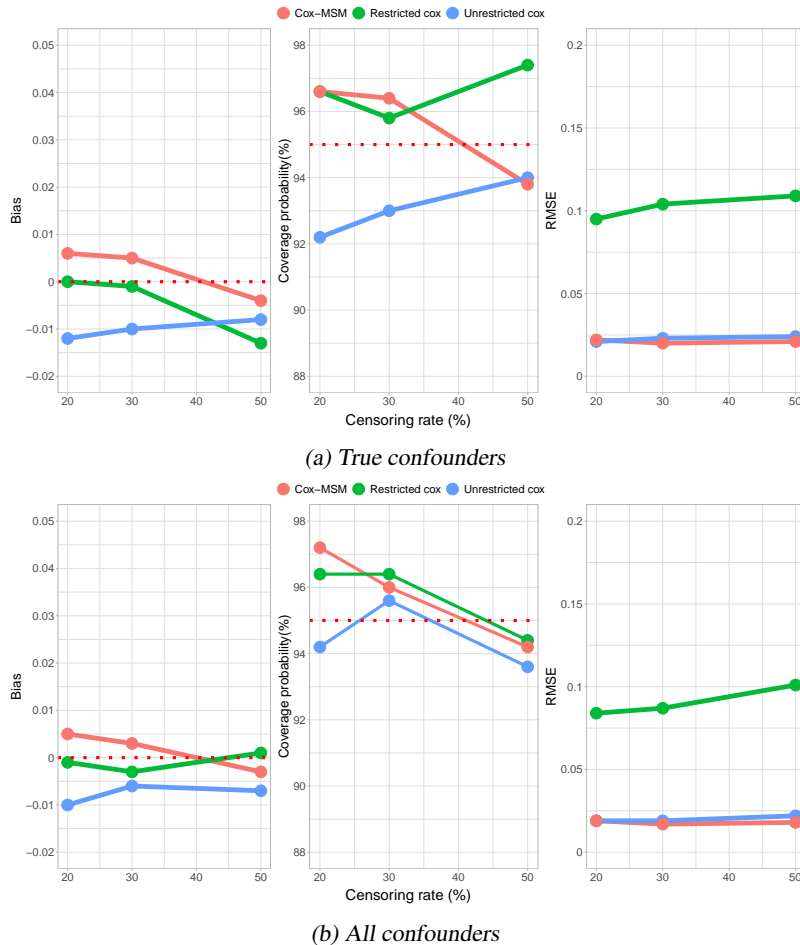


Figure B.3: Results of bias, coverage probability and root mean squared error for comparison according to the censoring rates. (a) Treatment delay effects was estimated by including only 4 confounders (X_1, X_2, X_5, X_6) for each model. (b) Treatment delay effects was estimated by including all confounders (X_1, \dots, X_8) for each model. Scenario settings are as follows. Treatment delay effects $\exp(\psi) = 1.2$, delay unit = 1 week, time to treat = Gompertz distribution, $N = 1,000$.

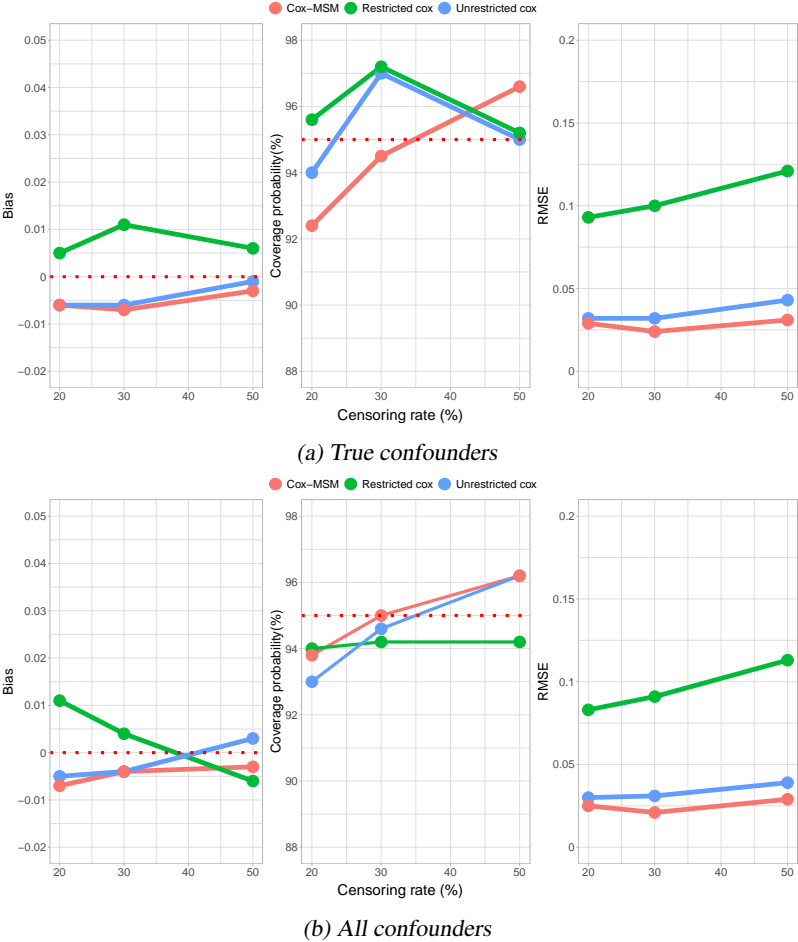


Figure B.4: Results of bias, coverage probability and root mean squared error for comparison according to the censoring rates. (a) Treatment delay effects was estimated by including only 4 confounders (X_1, X_2, X_5, X_6) for each model. (b) Treatment delay effects was estimated by including all confounders (X_1, \dots, X_8) for each model. Scenario settings are as follows. Treatment delay effects $\exp(\psi) = 1$, delay unit = 2 weeks, time to treat = Gompertz distribution, $N = 1,000$.

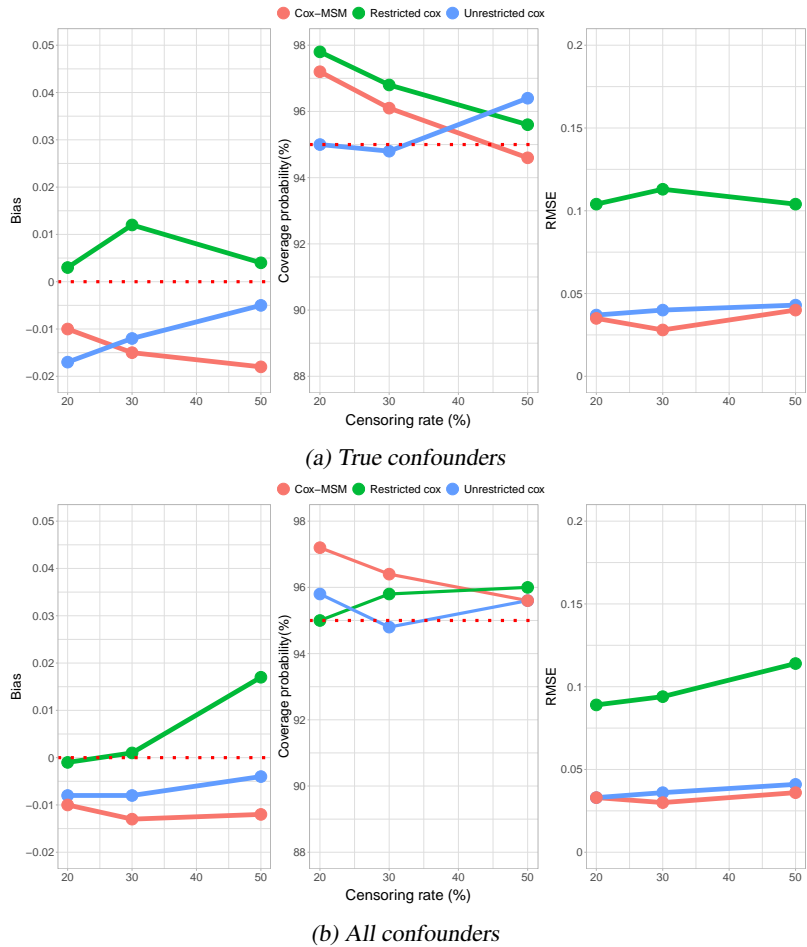


Figure B.5: Results of bias, coverage probability and root mean squared error for comparison according to the censoring rates. (a) Treatment delay effects was estimated by including only 4 confounders (X_1, X_2, X_5, X_6) for each model. (b) Treatment delay effects was estimated by including all confounders (X_1, \dots, X_8) for each model. Scenario settings are as follows. Treatment delay effects $\exp(\psi) = 1.2$, delay unit = 2 weeks, time to treat = Gompertz distribution, $N = 1,000$.

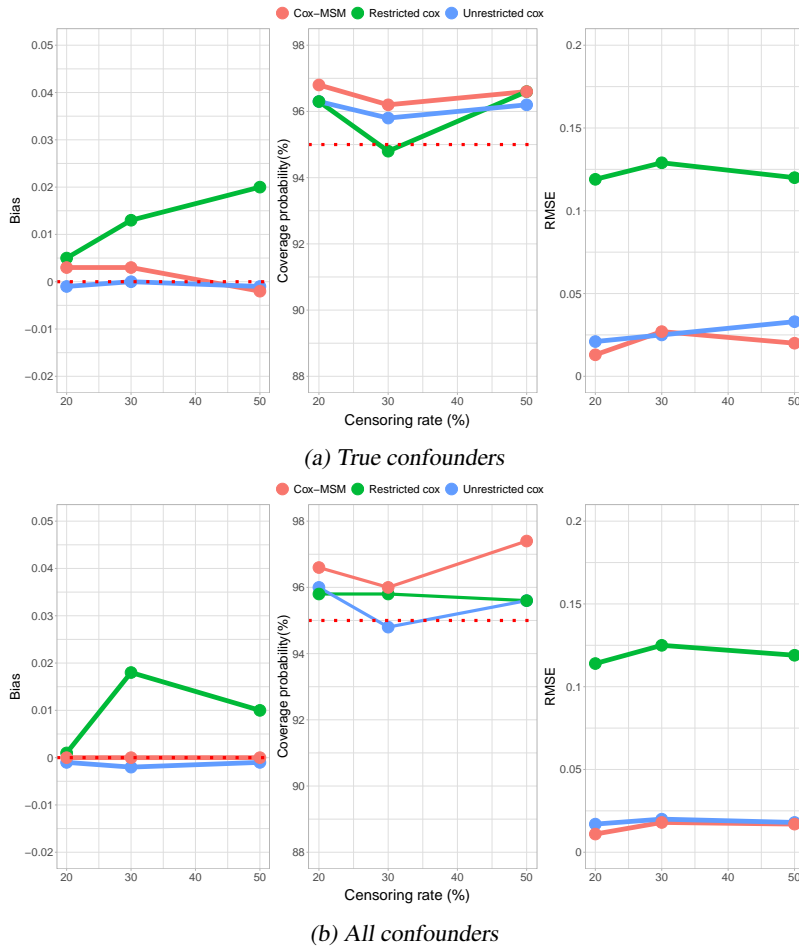
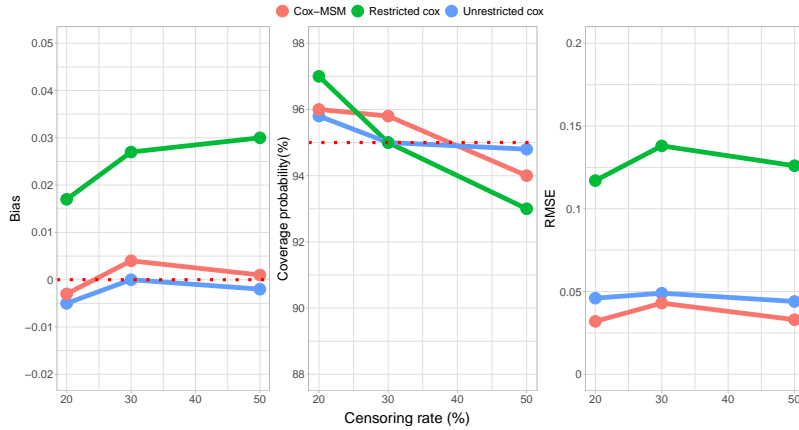
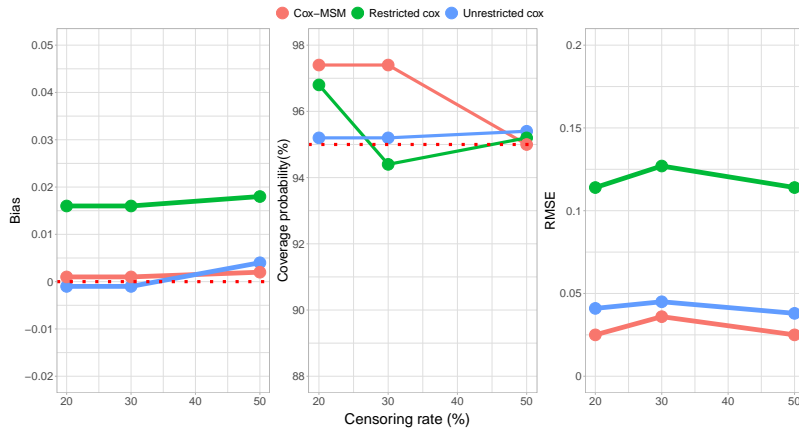


Figure B.6: Results of bias, coverage probability and root mean squared error for comparison according to the censoring rates. (a) Treatment delay effects was estimated by including only 4 confounders (X_1, X_2, X_5, X_6) for each model. (b) Treatment delay effects was estimated by including all confounders (X_1, \dots, X_8) for each model. Scenario settings are as follows. Treatment delay effects $\exp(\psi) = 1$, delay unit = 1 week, time to treat = exponential distribution, $N = 500$.



(a) True confounders



(b) All confounders

Figure B.7: Results of bias, coverage probability and root mean squared error for comparison according to the censoring rates. (a) Treatment delay effects was estimated by including only 4 confounders (X_1, X_2, X_5, X_6) for each model. (b) Treatment delay effects was estimated by including all confounders (X_1, \dots, X_8) for each model. Scenario settings are as follows. Treatment delay effects $\exp(\psi) = 1$, delay unit = 2 weeks, time to treat = exponential distribution, $N = 500$.

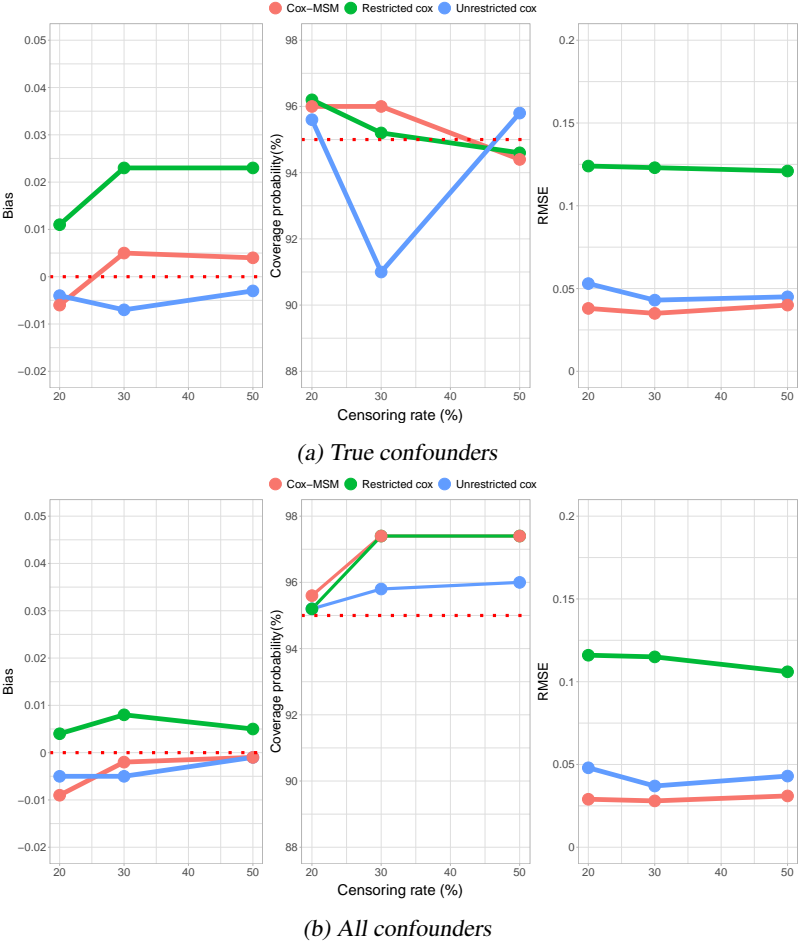


Figure B.8: Results of bias, coverage probability and root mean squared error for comparison according to the censoring rates. (a) Treatment delay effects was estimated by including only 4 confounders (X_1, X_2, X_5, X_6) for each model. (b) Treatment delay effects was estimated by including all confounders (X_1, \dots, X_8) for each model. Scenario settings are as follows. Treatment delay effects $\exp(\psi) = 1.2$, delay unit = 2 weeks, time to treat = exponential distribution, $N = 500$.

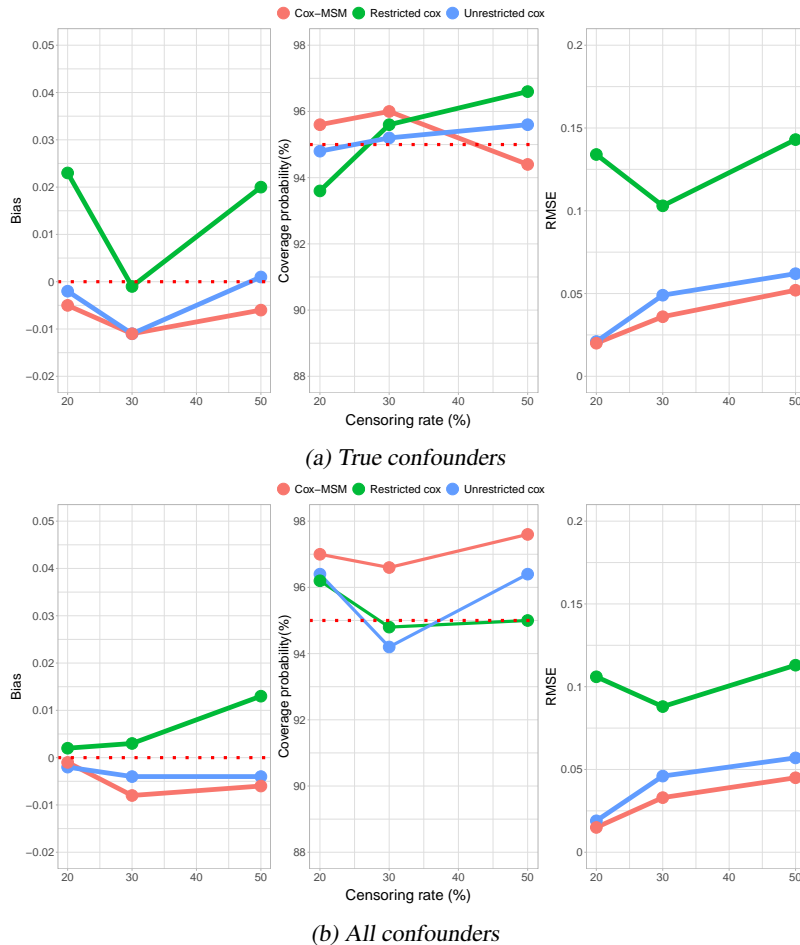


Figure B.9: Results of bias, coverage probability and root mean squared error for comparison according to the censoring rates. (a) Treatment delay effects was estimated by including only 4 confounders (X_1, X_2, X_5, X_6) for each model. (b) Treatment delay effects was estimated by including all confounders (X_1, \dots, X_8) for each model. Scenario settings are as follows. Treatment delay effects $\exp(\psi) = 1$, delay unit = 1 week, time to treat = Gompertz distribution, $N = 500$.

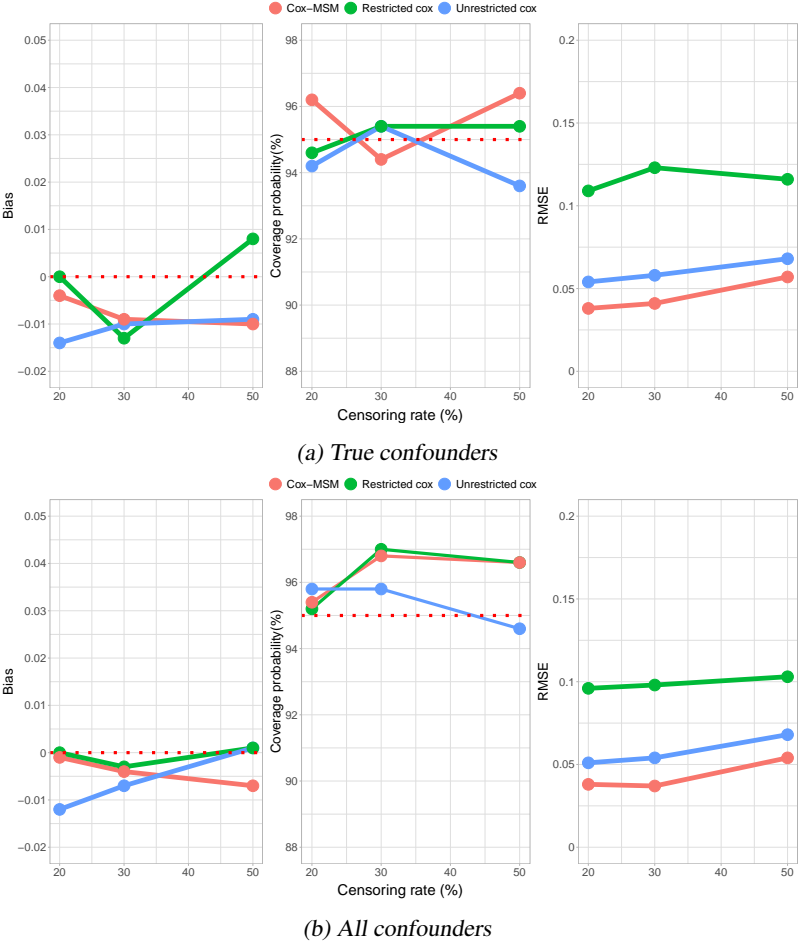


Figure B.10: Results of bias, coverage probability and root mean squared error for comparison according to the censoring rates. (a) Treatment delay effects was estimated by including only 4 confounders (X_1, X_2, X_5, X_6) for each model. (b) Treatment delay effects was estimated by including all confounders (X_1, \dots, X_8) for each model. Scenario settings are as follows. Treatment delay effects $\exp(\psi) = 1.2$, delay unit = 1 week, time to treat = Gompertz distribution, $N = 500$.

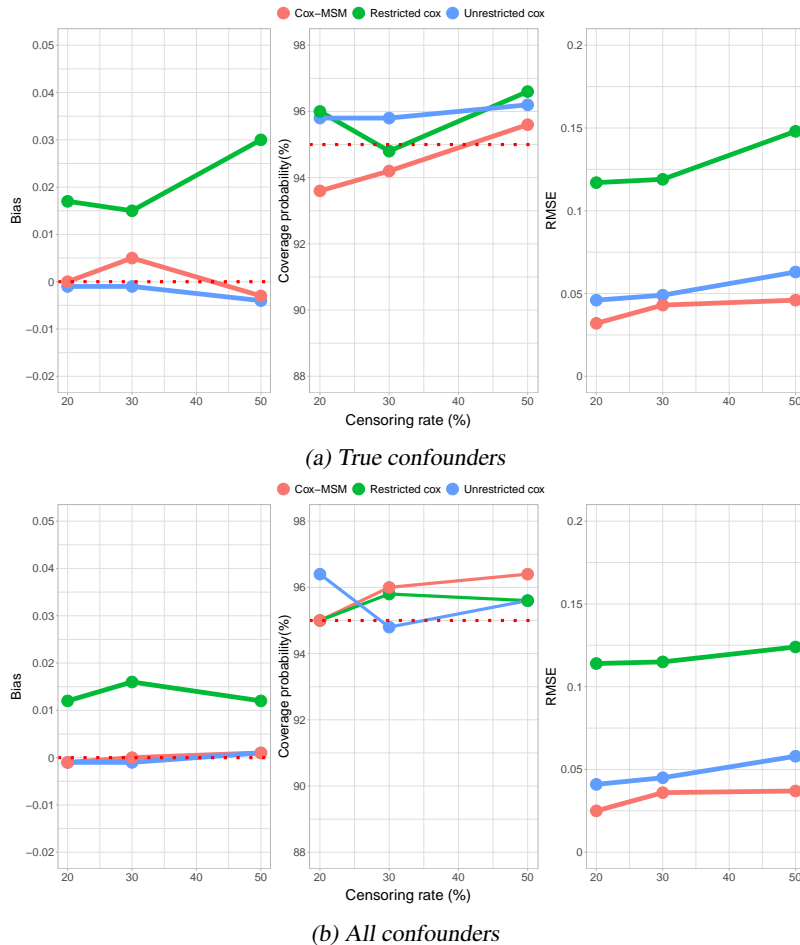


Figure B.11: Results of bias, coverage probability and root mean squared error for comparison according to the censoring rates. (a) Treatment delay effects was estimated by including only 4 confounders (X_1, X_2, X_5, X_6) for each model. (b) Treatment delay effects was estimated by including all confounders (X_1, \dots, X_8) for each model. Scenario settings are as follows. Treatment delay effects $\exp(\psi) = 1$, delay unit = 2 weeks, time to treat = Gompertz distribution, $N = 500$.

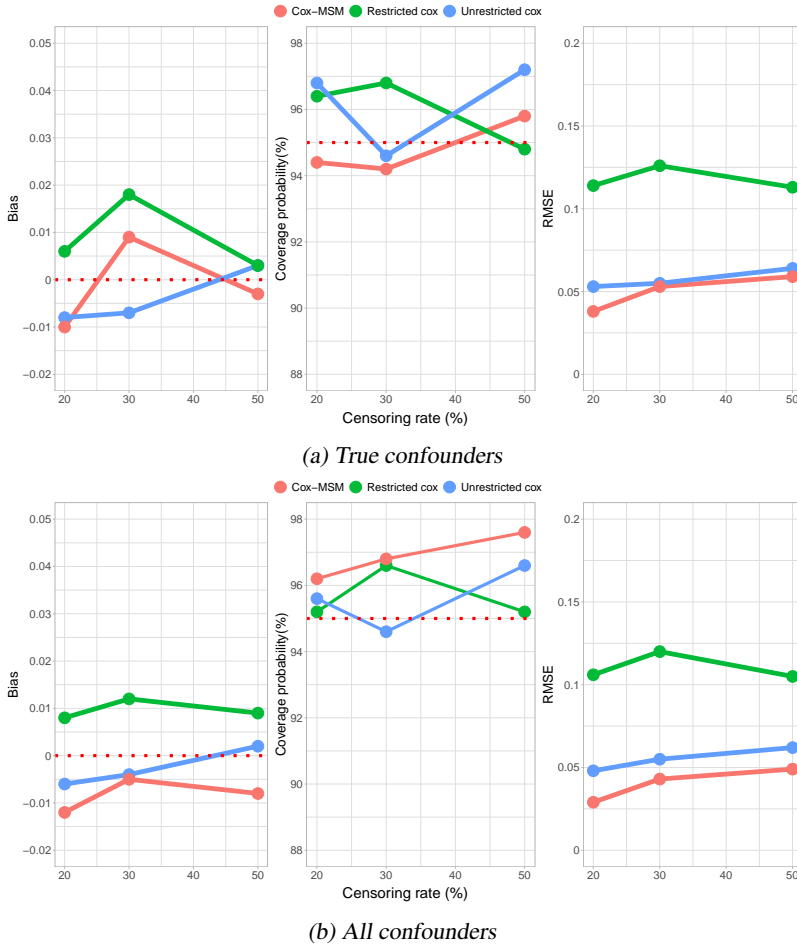


Figure B.12: Results of bias, coverage probability and root mean squared error for comparison according to the censoring rates. (a) Treatment delay effects was estimated by including only 4 confounders (X_1, X_2, X_5, X_6) for each model. (b) Treatment delay effects was estimated by including all confounders (X_1, \dots, X_8) for each model. Scenario settings are as follows. Treatment delay effects $\exp(\psi) = 1.2$, delay unit = 2 weeks, time to treat = Gompertz distribution, $N = 500$.

References

- Austin PC (2012). The performance of different propensity score methods for estimating marginal hazard ratios, *Statistics in Medicine*, **32**, 2837–2849.
- Austin PC (2014). The use of propensity score methods with survival or time-to-event outcomes: Reporting measures of effect similar to those used in randomized experiments, *Statistics in Medicine*, **33**, 1242–1258.
- Bender R, Augustin T, and Blettner M (2005). Generating survival times to simulate Cox proportional hazards models, *Statistics in Medicine*, **24**, 1713–1723.
- Hade EM, Nattino G, Frey HA, and Lu B (2020). Propensity score matching for treatment delay effects with observational survival data, *Statistical Methods in Medical Research*, **29**, 695–708.
- Hernán MÁ, Brumback B, and Robins JM (2000). Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men, *Epidemiology*, **11**, 561–570.
- Hernán MA and Robins JM (2010). *Causal Inference: What If*, CRC Boca Raton, FL.
- Lu B (2005). Propensity score matching with time-dependent covariates, *Biometrics*, **61**, 721–728.
- Lu B, Cai D, and Tong X (2018). Testing causal effects in observational survival data using propensity score matching design, *Statistics in Medicine*, **37**, 1846–1858.
- Robins JM, Hernan MA, and Brumback B (2000). Marginal structural models and causal inference in epidemiology, *Epidemiology*, **11**, 550–560.
- Stuart EA (2010). Matching methods for causal inference: A review and a look forward, *Statistical Science*, **25**, 1–21.
- van der Wal WM and Geskus RB (2011). ipw: An R-package for inverse probability weighting, *Journal of Statistical Software*, **43**, 1–23.
- Xiao Y, Abrahamowicz M, and Moodie EE (2010). Accuracy of conventional and marginal structural Cox model estimators: A simulation study, *The International Journal of Biostatistics*, **6**, 1–28.
- Yang S and Holloway ST (2021). contTimeCausal-Continuous Time Causal Models, Available from: <https://cran.r-project.org/web/packages/contTimeCausal/index.html>
- Young JG, Hernán MA, Picciotto S, and Robins JM (2010). Relation between three classes of structural models for the effect of a time-varying exposure on survival, *Lifetime Data Analysis*, **16**, 71–84.
- Zhao Q-Y, Luo J-C, Su Y, Zhang Y-J, Tu G-W, and Luo Z (2021). Propensity score matching with R: Conventional methods and new features, *Annals of Translational Medicine*, **9**, 1–39.

Received April 4, 2023; Revised May 1, 2023; Accepted May 8, 2023

생존자료분석에서 성향 점수를 이용한 treatment delay effect 추정법에 대한 연구

정주이^a, 송현진^a, 한승봉^{1,a}

^a고려대학교 의학통계학협동과정

요 약

생존 자료에서 Hade 등 (2020) 은 시간-의존 교란 변수가 환자의 처치 시점에 영향을 미칠 때, 해당 효과를 보정하여 treatment delay effect를 올바르게 추정하기 위해 성향 점수 매칭 방법을 이용하였다. 이 때, treatment delay effect란 환자가 관심 있는 지연 시점만큼 늦게 처치를 받는 경우 제 때 받는 경우에 비해 사건 발생 위험에 미치는 영향을 의미한다. 본 연구에서는 또 다른 성향 점수 기반 모형인 Cox-MSM 모형 또한 해당 효과를 올바르게 추정할 수 있는지 모의 실험을 통해 확인 및 기존 매칭 모형과 비교하였다. 모의실험 결과, 세 가지 모형 모두 다양한 시나리오 내에서 treatment delay effect를 올바르게 추정함을 확인하였다. 특히 모든 시나리오 내에서 Cox-MSM의 제곱근평균제곱오차의 값이 가장 낮았으며, restricted Cox matching 모형에서 가장 큰 값을 가지는 것으로 나타났다. 결론적으로, 성향 점수에 기반하나 매칭이 아닌 방법 또한 treatment delay effect 적용이 가능하다는 결과를 제공한다. 추후 G-formula과 같이 성향 점수 기반이 아닌 모형에서도 적용이 가능한지에 대한 상세 연구가 필요하다고 사료된다.

주요용어: 생존 분석, 시간-의존 교란 변수, 성향 점수 매칭, treatment delay effect, Cox-MSM

본 연구는 과학기술정보통신부 재원 한국연구재단 (No. 2022R1F1A1063027)과 보건복지부 재원 한국보건산업진흥원의 보건의료기술연구개발사업 (No. HI22C045400)의 지원을 받아 수행된 연구임.

¹교신저자: (12841) 서울시 성북구 고려대로 73, 고려대학교 의학통계학협동과정. Email: hanseungbong@gmail.com