



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

치료군이 3 개인 경우 성향점수를
이용한 방법의 비교 : 모의실험 연구

연세대학교 대학원
의학전산통계학협동과정
의학통계학전공
신 정 수

치료군이 3 개인 경우 성향점수를
이용한 방법의 비교 : 모의실험 연구


지도 남 정 모 교수


이 논문을 석사 학위논문으로 제출함


2018 년 12 월 일

연세대학교 대학원
의학전산통계학협동과정
의학통계학전공
신 정 수

신정수의 석사 학위논문을 인준함

심사위원 남 정 모 

심사위원 정 인 경 

심사위원 송 기 준 

연세대학교 대학원

2018년 12월 일

차 례

그림 차례	iii
표 차례	v
국문요약	vi
제 1 장 서론	1
1.1 연구 배경 및 목적	1
1.2 연구 내용 및 방법	2
1.3 논문의 구성	3
제 2 장 이론적 배경	4
2.1 다범주 치료에서의 성향점수	4
2.2 성향점수 역확률 가중치 방법	6
2.3 성향점수 매칭 가중치 방법	7
2.4 Three-way 매칭 방법	8
2.4.1 삼각형의 둘레를 이용한 매칭	9
2.4.2 중심점에서의 거리를 이용한 매칭	10
2.4.3 원의 반지름을 이용한 매칭	10

2.5 모형 평가 지표	12
제 3 장 모의실험	13
3.1 모의실험 설계	13
3.2 모의실험 자료 분석 결과	16
제 4 장 결론 및 고찰	29
참고문헌	31
부록	33
영문요약	42

그림 차례

그림 1. Three-way 매칭 방법	11
그림 2. 상황 1 에서 샘플수	19
그림 3-1. 상황 1 에서 좋은 공변량 중복인 경우 (치료 할당 확률 분포의 공통영역이 큰) 표준화 평균 차이	20
그림 3-2. 상황 1 에서 좋지 않은 공변량 중복인 경우 (치료 할당 확률 분포의 공통영역이 작은) 표준화 평균 차이	21
그림 4. 상황 2 에서 샘플수	24
그림 5-1. 상황 2 에서 좋은 공변량 중복인 경우 (치료 할당 확률 분포의 공통영역이 큰) 표준화 평균 차이	25
그림 5-2. 상황 2 에서 좋지 않은 공변량 중복인 경우 (치료 할당 확률 분포의 공통영역이 작은) 표준화 평균 차이	26
부록 그림 6-1. 상황 1 에서 좋은 공변량 중복인 경우 (치료 할당 확률 분포의 공통영역이 큰) 편향	34
부록 그림 6-2. 상황 1 에서 좋지 않은 공변량 중복인 경우 (치료 할당 확률 분포의 공통영역이 작은) 편향	35

부록 그림 7-1. 상황 1 에서 좋은 공변량 중복인 경우	
(치료 할당 확률 분포의 공통영역이 큰) 평균 제곱 오차	36
부록 그림 7-2. 상황 1 에서 좋지 않은 공변량 중복인 경우	
(치료 할당 확률 분포의 공통영역이 작은) 평균 제곱 오차	37
부록 그림 8-1. 상황 2 에서 좋은 공변량 중복인 경우	
(치료 할당 확률 분포의 공통영역이 큰) 편향	38
부록 그림 8-2. 상황 2 에서 좋지 않은 공변량 중복인 경우	
(치료 할당 확률 분포의 공통영역이 작은) 편향	39
부록 그림 9-1. 상황 2 에서 좋은 공변량 중복인 경우	
(치료 할당 확률 분포의 공통영역이 큰) 평균 제곱 오차	40
부록 그림 9-2. 상황 2 에서 좋지 않은 공변량 중복인 경우	
(치료 할당 확률 분포의 공통영역이 작은) 평균 제곱 오차	41

표 차례

표 1-1. 상황 1 에서 좋은 공변량 중복인 경우

(치료 할당 확률 분포의 공통영역이 큰) 치료 효과 비교 22

표 1-2. 상황 1 에서 좋지 않은 공변량 중복인 경우

(치료 할당 확률 분포의 공통영역이 작은) 치료 효과 비교 23

표 2-1. 상황 2 에서 좋은 공변량 중복인 경우

(치료 할당 확률 분포의 공통영역이 큰) 치료 효과 비교 27

표 2-2. 상황 2 에서 좋지 않은 공변량 중복인 경우

(치료 할당 확률 분포의 공통영역이 작은) 치료 효과 비교 28

국 문 요 약

치료군이 3 개인 경우 성향점수를 이용한 방법의 비교 : 모의실험 연구

무작위 배정 임상시험은 비용, 윤리적인 문제로 시행이 어려운 경우가 많다. 이러한 경우에는 관찰연구를 통한 연구 설계를 할 수 있지만 치료군의 효과에는 편향이 발생한다. 3 개 이상의 치료군의 효과를 편향 없이 추정하도록 일반화 성향점수를 이용한 여러 방법들이 제안되었다.

본 연구는 Yoshida(2017)의 연구를 확장하여 성향점수를 이용한 방법의 성능을 비교하고자 한다. 성향점수를 이용한 가중치 방법으로는 안정된 역확률 가중치, 매칭 가중치 방법을 사용하고, 매칭 방법으로는 three-way 매칭을 사용한다. Three-way 매칭은 성향점수를 이용하여 2 차원 공간에서 삼각형의 둘레를 이용한 방법과 원의 반지름을 이용한 방법, 중심점에서의 거리를 이용한 방법이다.

각 방법들의 성능비교를 위해 모의실험을 시행하였다. 독립적으로 생성한 5 개의 공변량과 이 변수들의 교호작용과 이차항 등으로 생성한 5 개의 공변량을 이용하여 3 개의 범주를 갖는 치료군과 2 개의 결과변수를 가정한 모형에 따라 자료를 생성하였다. 치료군이 3 개인 경우 성향점수를 추정하기 위해 다항 로짓 회귀모형을

이용하였다. 치료 할당 확률 분포의 공통영역이 큰 경우를 좋은 공변량 중복이라고 하고, 작은 경우를 좋지 않은 공변량 중복이라고 한다. 상황 1 에서는 10 개의 공변량을 이용하여 성향점수를 추정하고, 상황 2 에서는 독립적으로 생성한 5 개의 공변량을 이용하여 성향점수를 추정한다. 10 개의 공변량 각각의 표준화 평균 차이는 3 가지 치료군을 세 쌍으로 하여 평균을 계산하였다. 편향과 평균 제곱 오차를 이용하여 치료 효과를 비교하였다.

상황 1 에서 three-way 매칭 방법은 비슷한 성능을 나타냈다. 매칭 가중치 방법은 표준화 평균 차이와 편향에서 좋은 성능을 보였지만 좋지 않은 공변량 중복에서 가중치를 사용한 표본수가 작고, 평균 제곱 오차가 컸다. 안정된 역확률 가중치 방법은 원본 자료와 동일한 표본수를 갖지만 좋지 않은 공변량 중복에서 성능이 크게 떨어졌다.

상황 2 에서 three-way 매칭 방법은 비슷한 성능을 나타냈다. 매칭 가중치 방법은 표준화 평균 차이가 치료 분율에 따른 차이가 컸고, 좋지 않은 공변량 중복에서 가중치를 사용한 표본수가 작고, 평균 제곱 오차가 컸다. 안정된 역확률 가중치 방법은 원본 자료와 표본수가 동일 하지만 좋지 않은 공변량 중복에서 성능이 크게 떨어졌다.

현재 연구에서는 three-way 매칭에서 사용하는 캘리퍼의 알려진 기준이 없고, 치료군이 4 개 이상인 경우에는 계산량이 기하급수적으로 증가하는 한계가 있다.

핵심 되는 말 : 성향점수 매칭, three-way 매칭, 역확률 가중치, 매칭 가중치

제 1 장 서론

1.1 연구 배경 및 목적

무작위 배정 임상시험 (randomized controlled clinical trials : RCT)은 시험군과 대조군의 효과의 차이를 편향 없이 추정하는 실험설계이다. 하지만 무작위 배정 임상시험은 윤리적인 고려사항, 비용, 환자의 선호 등의 문제로 시행이 어려운 경우가 많다. 예를 들어, 수술적 치료와 약물치료를 비교하는 연구에서는 윤리적인 이유로 무작위 배정이 불가능하다.

이러한 경우에는 관찰연구 (observational study)를 통하여 시험군과 대조군의 효과를 비교하는 연구 설계를 할 수 있다. 그러나 관찰연구에서는 연구자가 치료를 중재할 수 없고, 연구설계 단계에서 교란변수의 영향을 통제할 수 없다. 그래서 관찰연구에서는 시험군과 대조군 간의 효과를 비교하기 위하여 여러 가지 방법을 이용하여 교란변수를 보정하고 편향을 제거해야 한다. 관찰연구에서 발생하는 편향을 줄이기 위하여 성향점수 (propensity score)를 이용한 방법이 연구되어 왔다.

1.2 연구 내용 및 방법

성향점수 방법론은 대부분 시험군과 대조군으로 나뉜 경우와 같이 이분형 치료군에서 사용되었다. 그러나 최근 들어 치료 수준이 세 가지 이상인 경우에도 성향점수를 적용할 수 있도록 연구가 이루어지고 있다. 예를 들어, 질병이 복잡해짐에 따라 합병증이 발생하고 효과적인 치료를 위하여 세 가지 이상의 치료가 이루어 지는 경우가 있다. 의사는 환자에게 수술치료, 약물치료, 물리치료 등의 세 가지 이상의 치료방법을 취할 수 있고, 또는 치료의 세 가지 이상의 다른 기간이나 약물의 세 가지 이상의 다른 용량 등으로 적용할 수 있다. 이와 같이 치료군이 3 개 이상인 경우에는 이분형 치료군의 성향점수 방법론을 적용하기 어렵다.

치료군이 3 개 이상인 경우, 성향점수 방법을 확장시킨 일반화 성향점수 (generalized propensity score)가 Imbens(2000)에 의하여 제안되었다. 그러나 일반화 성향점수의 적용은 제한적이고 한계점이 있다(Onur, 2008).

본 논문은 Yoshida(2017)의 연구를 확장하여 치료군이 3 개인 경우 성향점수 방법을 이용한 안정된 역확률 가중치, 매칭 가중치, three-way 매칭을 적용하여 각 방법의 성능을 비교하고자 한다.

1.3 논문의 구성

제 1 장에서는 연구의 배경 및 목적과 연구 내용 및 방법을 소개하였고, 제 2 장에서는 성향점수 가중치 방법과 성향점수 매칭 방법의 개념과 모형을 평가하는 지표를 설명하였다. 제 3 장에서는 제 2 장에서 소개한 방법들을 비교하기 위하여 설정한 각 시나리오에서의 모의실험을 정의하며 설계 방법과 각 시나리오를 비교한 결과를 제시하였다. 마지막으로 제 4 장에서는 모의실험의 결과를 바탕으로 본 논문에 대한 결론 및 고찰에 대하여 논의하였다.

제 2 장 이론적 배경

2.1 다범주 치료에서의 성향점수

Rosenbaum 과 Rubin (1983)이 제안한 성향점수는 시험군과 대조군으로 나누어진 경우에 적용할 수 있는 한계점이 있다. 이러한 한계를 극복하기 위하여 Imbens (2000)는 치료군이 다범주인 경우에 대한 성향점수를 정의하였다.

우리가 관심 있는 치료를 T 라 하고, 다범주 치료군을 m 개라 가정한다. 즉, $T \in \{1, 2, \dots, m\}$. $Y_i(t)$ 는 i 번째 환자가 치료 t 에 할당되었을 때 잠재 결과변수 (potential outcome)이고, Y_{it} 는 i 번째 환자의 관찰된 결과 (observed response) 라고 한다. $I(t)$ 는 치료 t 를 받았는지 여부를 나타내는 지시변수이다.

$$I(t) = \begin{cases} 1 & \text{if } T = t \\ 0 & \text{otherwise} \end{cases}$$

T 는 환자가 받을 치료를 나타내는 임의의 변수이고, 각 환자는 m 개의 잠재 결과변수를 가지고 있다. T_i 는 i 번째 대상자가 실제로 받은 치료를 나타낸다. 그러나 실제로는 할당된 하나의 관찰된 결과만 관찰할 수 있다. 우리가 관찰할 수 없는 잠재 결과변수를 반사실적 관계 (counterfactuals) 라고 한다 (Greenland, 2004).

공변량 X 가 주어졌을 때, 치료 t 를 받을 조건부 확률을 일반화 성향점수 e_{ti} 라고 하며, Imbens(2000)는 다음과 같이 정의하였다.

$$e_{ti} = \text{pr}(T_i = t, |X_i = x_i) = E\{I(t)|X_i = x_i\}$$

일반화 성향점수 e_{ti} 가 주어졌을 때, 치료 t 를 받을 환자의 기댓값을 $\beta(t, e_{ti})$ 로 다음과 같이 정의하며 모든 t 에 대하여 공변량 X 가 주어졌을 때 결과는 아래와 같다.

$$\beta(t, e_{ti}) = E\{Y_i(t)|e_{ti}\} = E\{Y_i|T_i = t, e_{ti}\}$$

따라서, 치료 t 하에서 잠재 결과변수의 기댓값은 다음과 같다.

$$E\{Y_i(t)\} = E\{\beta(t, e_{ti})\}$$

본 연구에서는 치료군이 다범주 이므로 다항 로짓 회귀모형 (multinomial logit regression)을 적용하여 일반화 성향점수를 추정한다.

2.2 성향점수 역확률 가중치 방법

치료군과 대조군이 있는 경우 공변량이 주어졌을 때 치료를 받을 조건부 확률은 성향점수 $e_{1i} = pr(T_i = 1, |X_i = x_i)$ 로 나타낸다. 이때, 성향점수 역확률 가중치 (inverse probability of treatment weighting : IPTW)는 다음과 같이 정의한다.

$$IPTW = \frac{1}{I(Z_i = 1)e_{1i} + I(Z_i = 0)e_{0i}}$$

역확률 가중치를 이용한 방법은 표본의 크기가 비정상적으로 증가하기 때문에 귀무가설을 너무 자주 기각하는 경향이 있다. 이를 개선하기 위하여 안정된 역확률 가중치 (stabilized IPTW)가 제안되었다 (Xu and Ross, 2010). 낮은 성향점수를 가지며 치료를 받은 환자나 높은 성향점수를 가지며 치료를 받지 않은 환자에 대하여 가중치를 줄이는 것이 목적이다.

전체 환자가 N 명일 때, 치료를 받은 환자와 받지 않은 환자를 각각 n_1 명, n_0 명이라 하면 $N = n_1 + n_0$ 이다. 공변량을 고려하지 않았을 경우 치료를 받을 확률은 $p = n_1/N$ 이고, 치료를 받지 않을 확률은 $1 - p$ 이다. 치료를 받은 환자에서의 안정된 역확률 가중치는 p/e_{1i} 이고, 치료를 받지 않은 환자에서의 안정된 역확률 가중치는 $(1 - p)/(1 - e_{1i})$ 로 정의된다.

치료군이 다범주인 경우로 확장시키면 $T_i = t$ 인 i 번째 대상자의 안정된 역확률 가중치는

$$\text{stabilized IPTW} = \frac{p_t}{e_{ti}} \quad (1)$$

이며, 이때 $p_t = n_t/N$ 이고 n_t 는 t 라는 치료를 받은 환자의 수 이다.

안정된 역확률 가중치는 성향점수 매칭과는 다르게 샘플수를 줄이지 않고, 모든 환자를 분석에 사용할 수 있는 장점이 있다.

2.3 성향점수 매칭 가중치 방법

성향점수 역확률 가중치 방법은 낮은 성향점수를 가지는 치료를 받은 환자나 높은 성향점수를 가지는 치료를 받지 않은 환자에게 비정상적으로 큰 가중치를 주는 문제가 발생한다. 이를 해결하기 위하여 매칭 가중치 (matching weight)가 다음과 같이 제안되었다 (Li and Greene, 2013).

$$\text{Matching weight} = \frac{\min(e_{1i}, \dots, e_{mi})}{\sum_{k=1}^m I(Z_i = k)e_{ki}} \quad (2)$$

매칭 가중치는 역확률 가중치의 변형이며 분자에 1 대신 $\min(e_{1i}, \dots, e_{mi})$ 이 놓인다. 역확률 가중치와 비교하여 매칭 가중치는 성향점수 분포의 균형을 맞추기 위하여 과도하게 부여되는 가중치는 낮추며 수치적 안정성을 부여한다. 따라서 역확률 가중치는 $(1, \infty)$ 값을 가지는 반면에 매칭 가중치의 범위는 $(0, 1)$ 이다. 매칭 가중치 방법은 비정상적으로 큰 가중치가 부여되는 문제를 해결할 수 있기 때문에 보다 효율적이다.

2.4 Three-way 매칭 방법

성향점수 매칭 방법은 치료가 이분형인 경우에 대해 강한 무관성 (strongly ignorable treatment assignment assumption)의 가정에 기초한다. 이 가정은 두 가지 가정을 만족시킬 때 충족된다.

가정 1) 조건부 독립성의 가정 (conditional independence assumption)

$$(Y_0, Y_1) \perp T \mid X$$

공변량 X 가 주어졌을 때, 치료 할당은 반응변수 (Y_i)와 독립이다.

가정 2) 공통 영역의 가정 (common support assumption)

$$0 < P(T = 1 \mid X) < 1$$

시험군과 대조군 개체들의 치료 할당 확률 분포는 공통 영역에 존재한다.

위 가정이 모두 만족되면 시험군에 대한 특성을 측정할 수 있는 변수들이 충분히 존재할 때, 이들을 통제하는 것만으로 선택 편향이 없는 치료효과 추정치를 산출할 수 있다.

Three-way 매칭 방법은 치료군이 3 가지인 경우, 치료 1에 대한 성향점수와 치료 2에 대한 성향점수를 이용하며 그와 동시에 치료 3에 대한 성향점수가 결정된다. Three-way 매칭 방법은 2 차원 공간에서 최적화된 매칭 방법이다. 치료

1 을 받은 환자와 치료 2 를 받은 환자, 치료 3 을 받은 환자의 2 차원 상에서의 거리가 최소가 되도록 하는 세 명의 환자를 찾는 것이 목적이다 Rassen(2013).

2.4.1 삼각형의 둘레를 이용한 매칭

치료 1 을 받은 대상자와 치료 2 를 받은 대상자, 치료 3 을 받은 대상자를 잇는 삼각형의 둘레가 최소가 되도록 하는 방법이다. 이때 2 차원 공간은 치료 1 과 치료 2 에 대한 성향점수로 이루어져 있다. 캘리퍼 (caliper)의 폭은 각 치료군의 대상자들을 잇는 삼각형의 둘레를 기준으로 설정하였다. 삼각형의 둘레의 최대값은 다음과 같다.

$$0.6 \times \sqrt{\frac{\tau_1^2 + \tau_2^2 + \tau_3^2}{3}}$$

위의 식에서 τ_k^2 은 다음과 같이 정의한다.

$$\tau_k^2 = \frac{Var(e_{1i}|T = k) + Var(e_{2i}|T = k)}{2}, \quad k = 1, 2, 3.$$

2.4.2 중심점에서의 거리를 이용한 매칭

치료 1 을 받은 대상자와 치료 2 를 받은 대상자, 치료 3 을 받은 대상자의 2 차원 공간에서 세 대상자의 중심을 찾고, 그 중심으로부터 각 대상자까지의 거리의 합이 최소가 되도록 하는 방법이다. 캘리퍼의 폭은 중심점에서 각 대상자까지의 거리의 합을 기준으로 설정하였고, 삼각형의 둘레를 이용한 매칭 방법에서의 캘리퍼에 상수를 곱하여 정의하였다. 이 방법에서 캘리퍼에 대한 알려진 기준이 없기 때문에 삼각형의 둘레를 이용한 매칭 방법의 결과와 비슷한 비율의 표본수로 매칭되도록 경험적으로 상수를 설정하였다. 중심점에서 각 대상자를 잇는 거리의 합의 최대값은 다음과 같다.

$$\frac{\sqrt{3}}{2} \times 0.6 \times \sqrt{\frac{\tau_1^2 + \tau_2^2 + \tau_3^2}{3}}$$

2.4.3 원의 반지름을 이용한 매칭

치료 1 과 치료 2 에 대한 성향점수로 이루어진 2 차원 공간 상에서 치료 1 을 받은 대상자와 치료 2 를 받은 대상자, 치료 3 을 받은 대상자를 지나는 원은 하나로 정의 된다. 이때 원의 반지름이 최소가 되도록 하는 세 명의 대상자를 찾는 방법이다. 이 방법에서의 캘리퍼는 원의 반지름을 기준으로 설정하였고, 삼각형의 둘레를 이용한 매칭 방법에서의 캘리퍼에 상수를 곱하여 정의하였다. 이 방법에서 캘리퍼에 대한 알려진 기준이 없기 때문에 삼각형의 둘레를 이용한 매칭 방법의 결과와 비슷한

비율의 표본수로 매칭되도록 경험적으로 상수를 설정하였다. 원의 반지름의 최대값은 다음과 같다.

$$\frac{1}{3} \times 0.6 \times \sqrt{\frac{\tau_1^2 + \tau_2^2 + \tau_3^2}{3}}$$

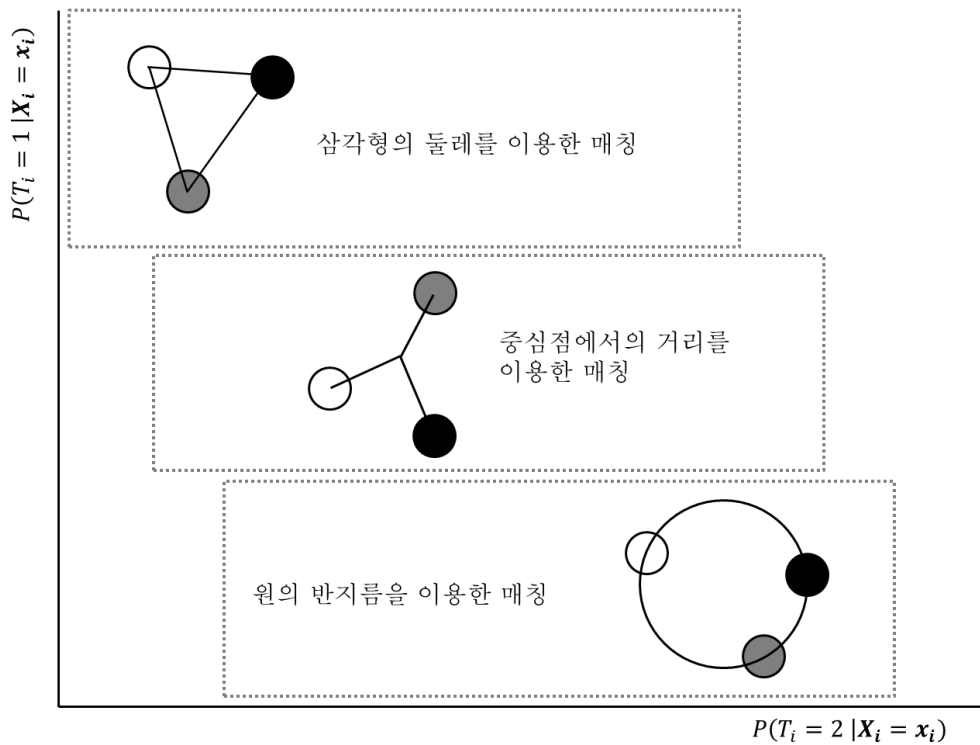


그림 1. Three-way 매칭 방법

그림 1. 은 삼각형의 둘레를 이용한 매칭 방법, 중심점에서의 거리를 이용한 매칭 방법, 원의 반지름을 이용한 매칭 방법을 2 차원 공간에서 표현하였다.

2.5 모형 평가 지표

본 연구에서는 여러가지 모형의 평가지표로서 분석에 사용되는 표본수, 각 공변량의 표준화 평균 차이, 편향 그리고 평균 제곱 오차를 사용한다.

표본수를 비교하기 위하여 가중치 방법에서는 가중치를 사용한 표본수를, 매칭 방법에서는 매칭 후 표본수를 이용한다. 가중치를 사용한 표본수는 $\sum_{i=1}^n W_i$ 로 나타내며, W_i 는 식(1)과 식(2)를 이용한다.

표준화 평균 차이는 3 가지 치료군을 세 쌍으로 한 표준화 된 평균 차이를 계산하고 각 공변량에 대하여 평균을 계산하였다. 평균 차이를 pooled within-group 분산의 제곱근으로 나누어 표준화 하였고 다음과 같다.

$$d = \frac{|\bar{x}_i - \bar{x}_j|}{\sqrt{\frac{s_i^2 + s_j^2}{2}}}, \quad i, j = 1, 2, 3, \quad i \neq j$$

치료 효과는 편향(bias)과 평균 제곱 오차(mean squared error)를 이용하여 성능을 비교하며, 다음과 같다.

$$bias = \frac{1}{M} \sum_{s=1}^M (\hat{\beta}_{Tks} - \beta_{Tks}), \quad k = 1, 2, 3$$

$$mean \ squared \ error = \frac{1}{M} \sum_{s=1}^M (\hat{\beta}_{Tks} - \beta_{Tks})^2, \quad k = 1, 2, 3$$

여기서 M 은 모의실험 반복 횟수를 의미한다.

제 3 장 모의실험

3.1 모의실험 설계

성향점수를 이용하여 총 5 가지의 가중치 방법과 매칭 방법을 12 가지의 시나리오 하에서 비교하고자 한다. 자료의 생성은 Franklin(2014) 의 설정을 그대로 이용하였고, 결과변수는 이분형이며 치료군은 세 가지로 구성되어 다음과 같이 생성하였다.

$$\log\left(\frac{P(T_i = 2 | \mathbf{X}_i = \mathbf{x}_i)}{P(T_i = 1 | \mathbf{X}_i = \mathbf{x}_i)}\right) = \alpha_{20} + \boldsymbol{\alpha}_{2X}^T \mathbf{x}_i$$

$$\log\left(\frac{P(T_i = 3 | \mathbf{X}_i = \mathbf{x}_i)}{P(T_i = 1 | \mathbf{X}_i = \mathbf{x}_i)}\right) = \alpha_{30} + \boldsymbol{\alpha}_{3X}^T \mathbf{x}_i$$

치료 할당은 실제 성향점수에 기초하여 다항 확률 변수로 생성되었다. 이때, α_{20} 와 α_{30} 는 치료 분율을 결정하고, $\boldsymbol{\alpha}_{2X}$ 와 $\boldsymbol{\alpha}_{3X}$ 는 공변량의 중복 수준을 결정한다. 치료의 분율은 $\{(33:33:33), (10:45:45), (10:10:80)\}$ 로 설정하였다. 치료 할당 확률 분포의 공통영역(common support)이 큰 경우를 좋은 공변량 중복(Good covariate overlap)이라 하고, 작은 경우를 좋지 않은 공변량 중복(Poor covariate overlap)이라 한다.

좋은 공변량 중복은 α_{2X} 를 $1/16 * [0.8, 0.06, 0.06, 1.5, 1.5, 0.4, 0.8, 0.04, 0.08, 1.0]$ 로 설정하고, α_{3X} 를 $1/8 * [0.8, 0.06, 0.06, 1.5, 1.5, 0.4, 0.8, 0.04, 0.08, 1.0]$ 로 설정한다. 좋지 않은 공변량 중복은 α_{2X} 를 $1 * [0.8, 0.06, 0.06, 1.5, 1.5, 0.4, 0.8, 0.04, 0.08, 1.0]$ 로 설정하고, α_{3X} 를 $2 * [0.8, 0.06, 0.06, 1.5, 1.5, 0.4, 0.8, 0.04, 0.08, 1.0]$ 로 설정한다. 실제 성향점수는 다음과 같다.

$$e_{1i} = P(T_i = 1 | \mathbf{X}_i = \mathbf{x}_i) = \frac{1}{1 + \exp(\alpha_{20} + \alpha_{2X}^T \mathbf{x}_i) + \exp(\alpha_{30} + \alpha_{3X}^T \mathbf{x}_i)}$$

$$e_{2i} = P(T_i = 2 | \mathbf{X}_i = \mathbf{x}_i) = \frac{\exp(\alpha_{20} + \alpha_{2X}^T \mathbf{x}_i)}{1 + \exp(\alpha_{20} + \alpha_{2X}^T \mathbf{x}_i) + \exp(\alpha_{30} + \alpha_{3X}^T \mathbf{x}_i)}$$

$$e_{3i} = P(T_i = 3 | \mathbf{X}_i = \mathbf{x}_i) = \frac{\exp(\alpha_{30} + \alpha_{3X}^T \mathbf{x}_i)}{1 + \exp(\alpha_{20} + \alpha_{2X}^T \mathbf{x}_i) + \exp(\alpha_{30} + \alpha_{3X}^T \mathbf{x}_i)}$$

범주형과 연속형으로 구성된 10 개의 공변량이 있으며, 공변량 중복의 수준과 치료의 분율 및 치료 효과가 다양하게 존재한다. log-probability 모델을 사용하였고, 다음과 같다.

$$\log(P(Y_i = 1 | T_i = t_i, \mathbf{X}_i = \mathbf{x}_i)) = \beta_0 + \mathbf{x}_i^T \boldsymbol{\beta}_X + \beta_{T2} I(t_i = 2) + \beta_{T3} I(t_i = 3)$$

$\boldsymbol{\beta}_X$ 는 공변량의 주효과를 나타내며, β_{T2} 과 β_{T3} 는 치료효과 (treatment effect)를 나타내고 치료효과가 있는 경우와 없는 경우로 설정하였다. $\boldsymbol{\beta}_X = 1/5 * [0.8, 0.06, 0.06, 1.5, 1.5, 0.4, 0.8, 0.04, 0.08, 1.0]$ 로 설정하였다. 치료효과가 없는 경우는 β_{T2} 과 β_{T3} 를 각각 0 으로 설정하였고, 치료효과가 있는 경우는 β_{T2} 를 $\log(0.9)$, β_{T3} 를 $\log(0.6)$ 으로 설정하였다. β_0 는 baseline 에서의 질병 위험으로 5%로 설정하였다. 3 가지의 치료 분율, 2 가지의 공변량 중복 수준과 2 가지의 치료효과 수준을

이용하여 이 모의실험의 경우를 조합하여 총 12 개의 시나리오를 구성하였다. 자료마다 600 명의 환자를 생성하였고, 각 시나리오는 100 번씩 반복 하였다. 즉, $M=100$.

변수	생성 과정
X_{1i}	$Normal(0, 1^2)$
X_{2i}	$Log - Normal(0, 0.5^2)$
X_{3i}	$Normal(0, 10^2)$
X_{4i}	$Bernoulli(p_i = e^{2X_{1i}} / (1 + e^{2X_{1i}}))$ where $E[p_i] = 0.5$
X_{5i}	$Bernoulli(p = 0.2)$
X_{6i}	$Multinomial(p = (0.5, 0.3, 0.1, 0.05, 0.05)^T)$
X_{7i}	$\sin(X_{1i})$
X_{8i}	X_{2i}^2
X_{9i}	$X_{3i} \times X_{4i}$
X_{10i}	$X_{4i} \times X_{5i}$

본 논문에서는 두 가지 상황을 설정하고 각 상황에서 성향점수를 이용한 방법의 성능을 비교한다.

상황 1. 10 개의 공변량을 이용하여 성향점수를 계산한다.

상황 2. 독립적으로 생성한 5 개의 공변량(X_1, X_2, X_3, X_5, X_6)을 이용하여 성향점수를 계산한다. 성향점수를 추정할 때 추정 모형이 잘못 설정된 경우를 의미한다.

3.2 모의실험 자료 분석 결과

상황 1 의 결과, 그림 2 는 좋은 공변량 중복과 좋지 않은 공변량 중복에서의 표본수를 나타낸다. 좋은 공변량 중복에서 원본 자료와 안정된 역확률 가중치 방법에서의 표본수는 동일하게 나타났고, 삼각형의 둘레를 이용한 매칭 방법, 원의 반지름을 이용한 매칭, 중심점에서의 거리를 이용한 매칭, 매칭 가중치 방법에서의 표본수는 원본 자료보다 작은 것으로 나타났다. 또한 치료의 분율에 따라서 표본수가 달라지는 것을 확인 할 수 있었고, $\{(33:33:33), (10:45:45), (10:10:80)\}$ 순으로 표본수가 크게 나타났다. 좋지 않은 공변량 중복에서 안정된 역확률 가중치 방법에서의 표본수는 원본 자료와 동일하게 나타났고, three-way 매칭 방법과 매칭 가중치 방법에서의 표본수는 원본 자료보다 작게 나타났다. 좋은 공변량 중복에서와 달리 치료의 분율에 따른 차이가 크지 않았고, 특히 매칭 가중치 방법에서의 표본수가 크게 감소 하였다.

그림 3-1. 은 좋은 공변량 중복에서 공변량 각각의 표준화 평균 차이를 나타내고, 그림 3-2. 는 좋지 않은 공변량 중복에서 공변량 각각의 표준화 평균 차이를 나타낸다. 좋은 공변량 중복에서는 three-way 매칭 방법은 비슷한 결과를

나타냈고, 모두 0.2 미만이었다. 매칭 가중치, 안정된 역확률 가중치 방법은 0.1 미만이었다. 좋지 않은 공변량 중복에서는 원본 자료에 비하여 three-way 매칭 방법과 매칭 가중치 방법에서 표준화 평균 차이가 대체로 감소하는 양상을 보였고, 안정된 역확률 가중치 방법에서는 원본 자료보다는 감소하였지만 다른 방법들에 비하여 크게 증가하였다.

표 1-1. 은 좋은 공변량 중복에서 치료 효과를 비교하였다. 치료효과가 없는 경우 실제 치료효과는 0이며, 치료효과가 있는 경우 실제 치료효과는 $\beta_{T2} = \log(0.9)$, $\beta_{T3} = \log(0.6)$ 이다. 편향은 모든 방법이 0에 가깝다. 평균 제곱 오차는 모든 방법에서 비슷한 값을 나타냈다. 표 1-2. 는 좋지 않은 공변량 중복에서 치료 효과를 비교하였다. 편향은 three-way 매칭과 매칭 가중치 방법에서 0에 가깝게 나타났지만, 안정된 역확률 가중치 방법에서는 0.7이상까지 크게 증가했다.

상황 2 의 결과, 그림 4. 는 좋은 공변량 중복과 좋지 않은 공변량 중복에서의 표본수를 나타낸다. 원본 자료와 안정된 역확률 가중치 방법에서의 표본수는 동일하고, three-way 매칭 방법과 매칭 가중치 방법에서의 표본수는 원본 자료보다 작게 나타났다. 또한 치료의 분율이 $\{(33:33:33), (10:45:45), (10:10:80)\}$ 순으로 표본수가 컸다. 좋지 않은 공변량 중복에서는 치료 분율에 따른 차이가 크지 않았고, 특히 매칭 가중치 방법에서의 표본수가 크게 감소하였다.

그림 5-1. 은 좋은 공변량 중복에서 공변량 각각의 표준화 평균 차이를 나타내고, 그림 5-2. 는 좋지 않은 공변량 중복에서 공변량 각각의 표준화 평균 차이를 나타낸다. 좋은 공변량 중복에서 성향점수를 계산하기 위해 포함된 5개의

공변량은 가중치 방법이 0.1 미만으로 작았고, three-way 매칭 방법은 0.2 미만으로 대부분 비슷했지만 중심점을 이용한 매칭 방법에서 조금 증가하였다. 성향점수를 계산하기 위해 포함되지 않은 5개의 공변량에서는 특히 X4 변수에서 three-way 매칭 방법이 크게 증가하였고, 치료 분율에 따른 차이도 컸다. 좋지 않은 공변량 중복에서 성향점수를 계산하기 위해 포함된 5개의 공변량은 three-way 매칭 방법과 매칭 가중치 방법에서 0.2미만이었고, 안정된 역확률 가중치 방법에서는 0.2 이상으로 크게 증가하였다. 성향점수를 계산하기 위해 포함되지 않은 5개의 공변량에서는 X4 변수에서 0.3 이상으로 크게 증가하였고, three-way 매칭 방법은 비슷한 성능을 보였지만 매칭 가중치 방법은 치료 분율에 따른 차이가 컸다. X9, X10 변수에서 가중치 방법에 비하여 three-way 매칭 방법이 더 좋은 성능을 나타냈고 매칭 가중치 방법은 치료 분율에 따른 차이가 컸다.

표 2-1. 은 좋은 공변량 중복에서 치료 효과를 비교하였다. 치료효과가 없는 경우 실제 치료효과는 0이며, 치료효과가 있는 경우 실제 치료효과는 $\beta_{T_2} = \log(0.9)$, $\beta_{T_3} = \log(0.6)$ 이다. 편향은 모든 방법에서 0에 가까웠고, 평균 제곱 오차는 모든 방법이 비슷하게 나타났다. 표 2-2. 는 좋지 않은 공변량 중복에서 치료 효과를 비교하였다. 편향은 three-way 매칭과 매칭 가중치 방법에서 0에 가까웠지만, 안정된 역확률 가중치 방법에서는 0.8이상까지 크게 증가했다.

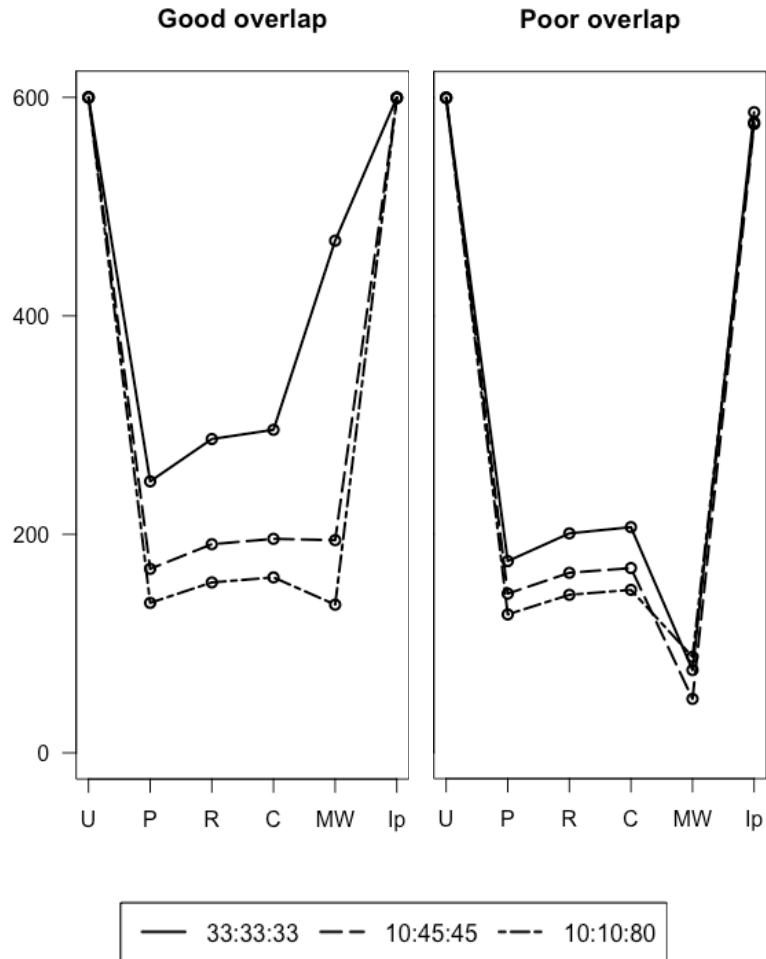


그림 2. 상황1에서 표본수

약어: U : 원본 자료, P : 삼각형 둘레를 이용한 매칭 방법, R : 원의 반지름을 이용한 매칭 방법, C : 중심점에서의 거리를 이용한 매칭 방법, MW : 매칭 가중치 방법, Ip : 안정된 역확률 가중치 방법.

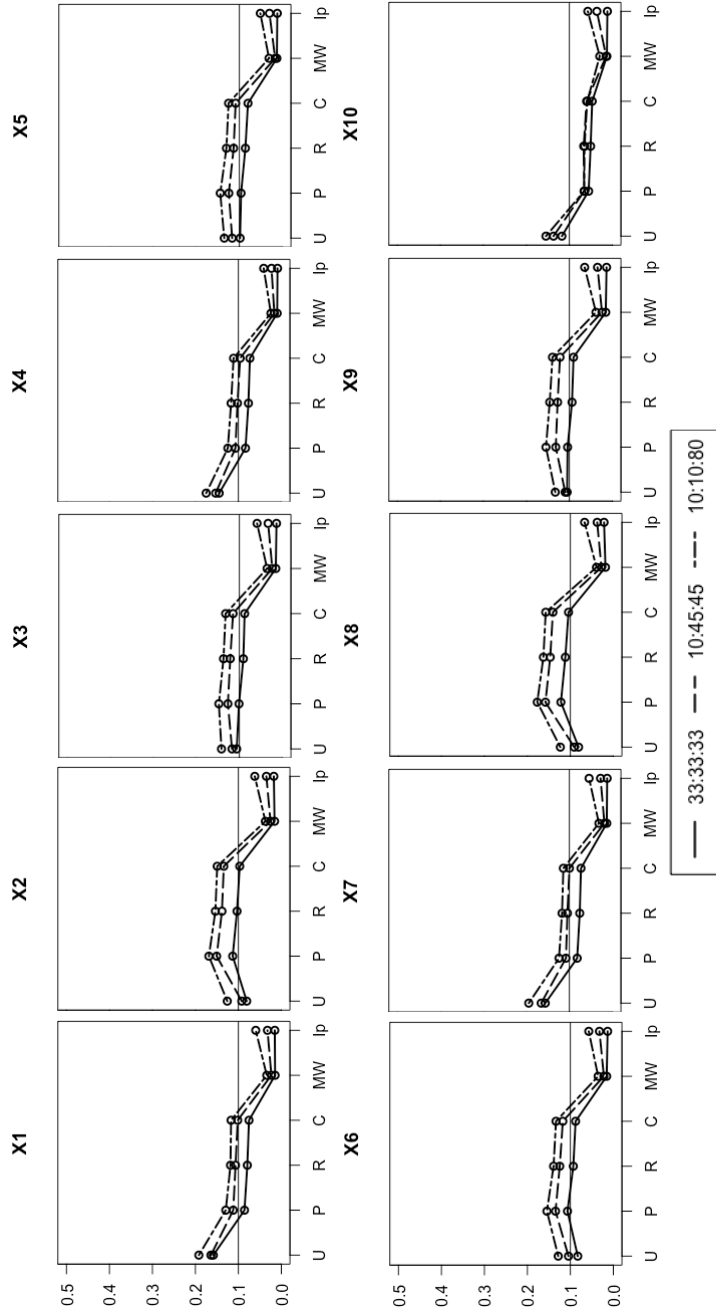


그림 3-1. 상황1에서 좋은 공변량 중복인 경우(치료 할당 확률 분포의 공통영역이 큰) 표준화 평균 차이

약어: U : 원본 자료, P : 삼각형 물레를 이용한 매칭 방법, R : 원의 반지름을 이용한 매칭 방법, C : 중심점에서의 거리를 이용한 매칭 방법, MW : 매칭 가중치 방법, Ip : 안정된 역확률 가중치 방법.

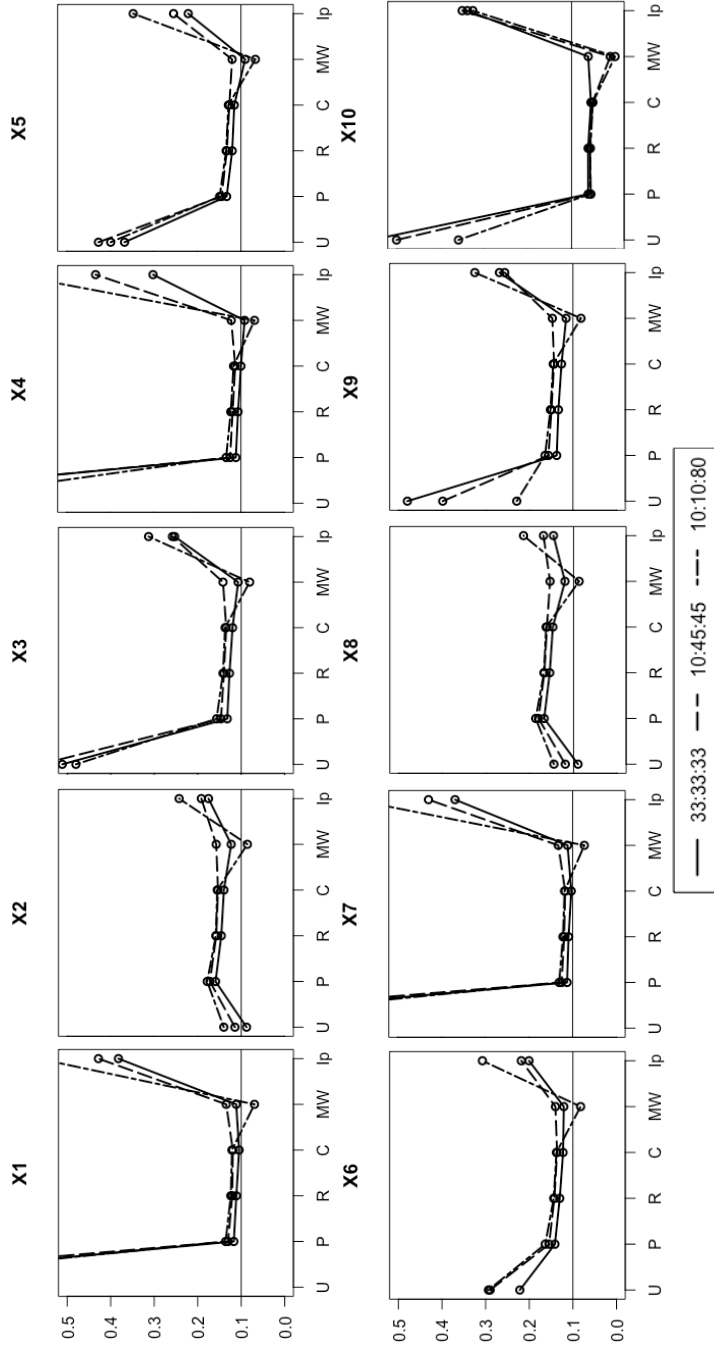


그림 3-2. 상황1에서 좋지 않은 공변량 중복인 경우(치료 할당 확률 분포의 공통영역이 작음) 표준화 평균 차이

약어: U : 원본 자료, P : 삼각형 틀레를 이용한 매칭 방법, R : 원의 반지름을 이용한 매칭 방법, C : 중심점에서의 거리를 이용한 매칭 방법, MW : 매칭 가중치 방법, Ip : 안정된 역확률 가중치 방법.

표1-1. 상황1에서 좋은 공변량 중복인 경우(치료 할당 확률 분포의 공통영역이 큰) 치료 효과 비교

prevalence ¹	method ²	bias			MSE		
		1vs2	1vs3	2vs3	1vs2	1vs3	2vs3
33:33:33	U	0.059	0.202	0.143	0.130	0.169	0.154
	P	-0.011	0.033	0.039	0.268	0.280	0.280
	R	-0.017	0.031	0.048	0.200	0.222	0.227
	C	-0.025	0.041	0.066	0.226	0.208	0.218
	MW	-0.020	0.044	0.064	0.127	0.133	0.141
	Ip	-0.014	0.051	0.065	0.131	0.134	0.141
10:45:45	U	0.157	0.228	0.069	0.308	0.313	0.081
	P	-0.014	0.120	0.098	0.460	0.557	0.572
	R	0.003	0.091	0.088	0.479	0.566	0.518
	C	0.000	0.092	0.075	0.476	0.500	0.600
	MW	0.063	0.058	-0.006	0.303	0.278	0.104
	Ip	0.134	0.126	-0.008	0.397	0.334	0.076
10:10:80	U	0.050	0.272	0.218	0.445	0.366	0.325
	P	-0.030	0.120	0.106	0.502	0.589	0.617
	R	-0.029	0.104	0.198	0.547	0.496	0.643
	C	-0.064	0.047	0.073	0.548	0.565	0.587
	MW	-0.037	0.125	0.163	0.484	0.342	0.334
	Ip	-0.021	0.171	0.189	0.547	0.416	0.345

¹ prevalence : 치료할당 분율, ² method (U: 원본 자료, P : 삼각형 들레를 이용한 매칭 방법, R : 원의 반지름을 이용한 매칭 방법, C : 중심점에서의 거리를 이용한 매칭 방법, MW : 매칭 가중치 방법, Ip : 안정된 역확률 가중치 방법)

표1-2. 상황1에서 좋지 않은 공변량 중복인 경우 (치료 할당 확률 분포의 공통영역이 작은) 치료 효과 비교

prevalence ¹	method ²	bias			MSE		
		1vs2	1vs3	2vs3	1vs2	1vs3	2vs3
33:33:33	U	0.445	0.994	0.549	0.384	1.169	0.421
	P	0.056	0.162	0.080	0.471	0.505	0.434
	R	0.030	0.088	0.066	0.475	0.480	0.358
	C	0.027	0.141	0.101	0.520	0.526	0.488
	MW	0.308	0.030	-0.294	0.838	1.333	0.872
	Ip	0.399	0.474	0.060	0.795	0.924	0.460
10:45:45	U	0.371	1.064	0.686	0.471	1.445	0.576
	P	0.083	0.303	0.171	0.207	0.278	0.205
	R	0.129	0.226	0.131	0.234	0.276	0.289
	C	0.078	0.304	0.220	0.232	0.327	0.228
	MW	0.184	0.012	-0.107	1.250	1.854	1.102
	Ip	0.722	0.834	0.118	1.771	1.954	0.345
10:10:80	U	0.086	0.823	0.735	0.611	0.998	0.934
	P	0.004	0.183	0.180	0.305	0.275	0.218
	R	-0.041	0.144	0.205	0.312	0.422	0.372
	C	0.001	0.168	0.170	0.310	0.366	0.375
	MW	-0.042	-0.206	-0.134	1.069	1.295	1.210
	Ip	0.317	0.932	0.674	2.021	2.156	1.620

¹ prevalence : 치료할당 분율, ² method (U: 원본 자료, P : 삼각형 들레를 이용한 매칭 방법, R : 원의 반지름을 이용한 매칭 방법, C : 중심점에서의 거리를 이용한 매칭 방법, MW : 매칭 가중치 방법, Ip : 안정된 역확률 가중치 방법)

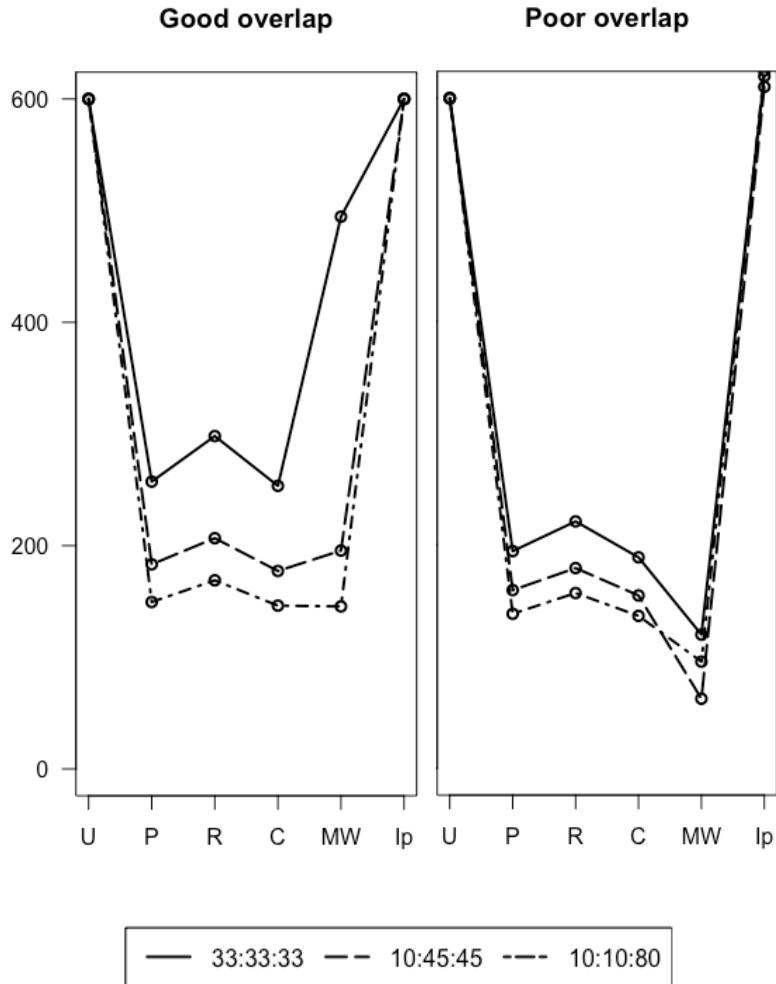


그림 4. 상황 2에서 표본수

약어: U : 원본 자료, P : 삼각형 둘레를 이용한 매칭 방법, R : 원의 반지름을 이용한 매칭 방법, C : 중심점에서의 거리를 이용한 매칭 방법, MW : 매칭 가중치 방법, Ip : 안정된 역확률 가중치 방법.

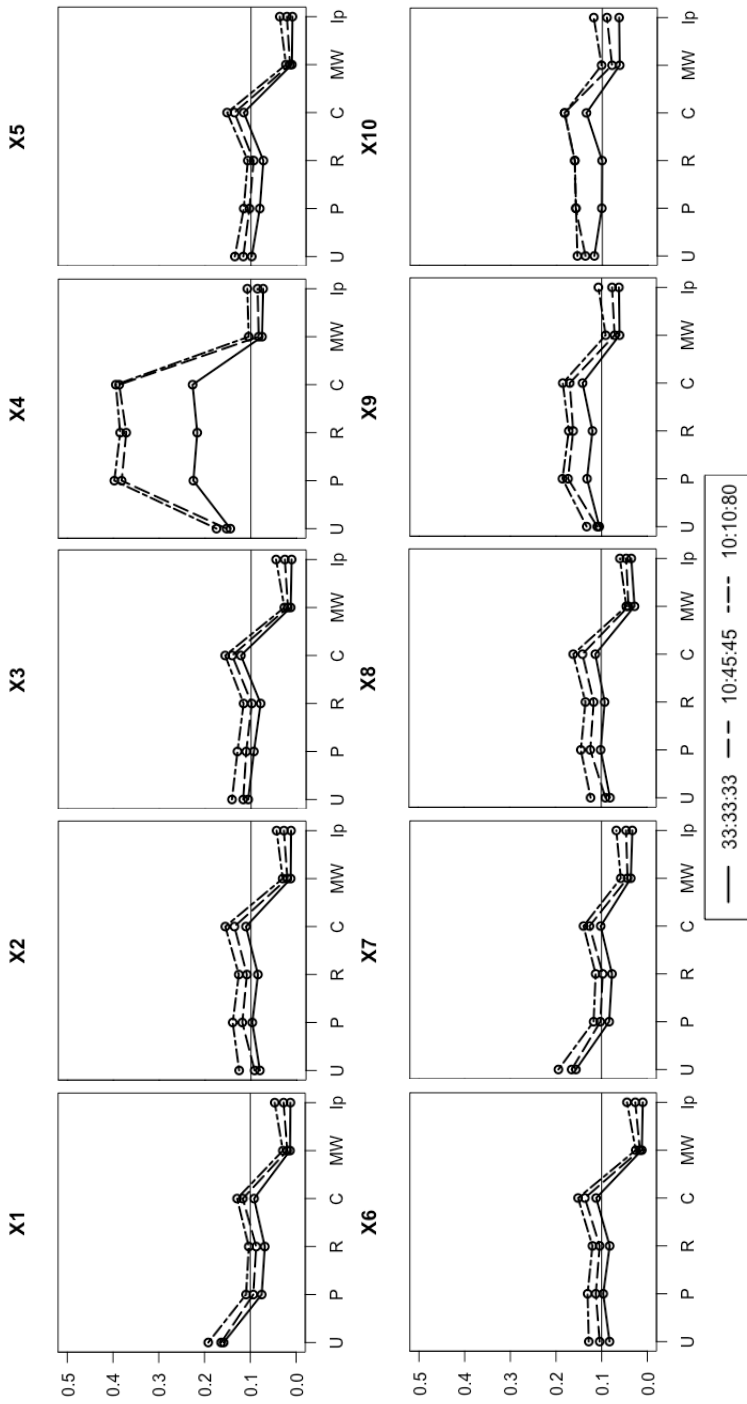


그림 5-1. 상황 2에서 좋은 공변량 중복인 경우(치료 할당 확률 분포의 공통영역이 큰) 표준화 평균 차이

약어: U : 원본 자료, P : 삼각형 돌레를 이용한 매칭 방법, R : 원의 반지름을 이용한 매칭 방법, C : 중심점에서 거리
 를 이용한 매칭 방법, MW : 매칭 가중치 방법, Ip : 안정된 역확률 가중치 방법.

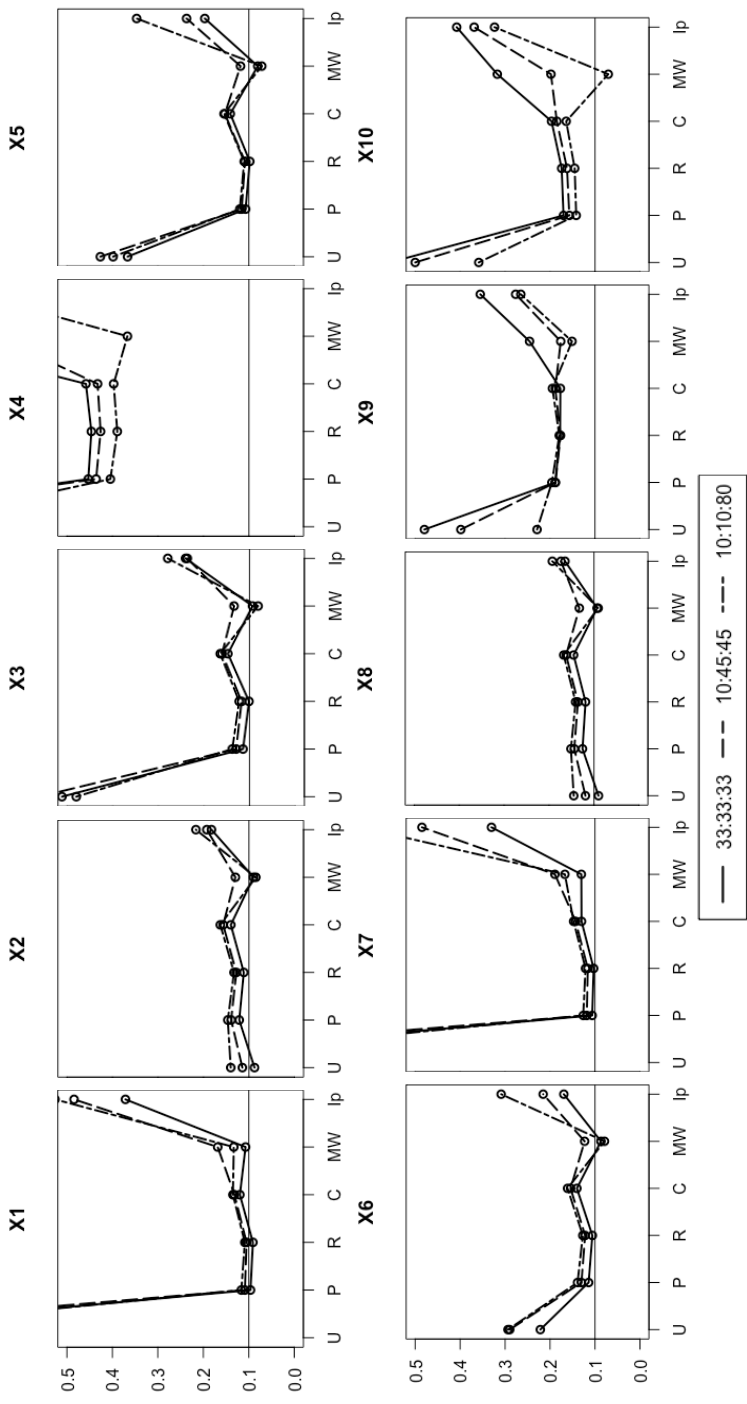


그림 5-2. 상황 2에서 좋지 않은 공변량 증복인 경우(치료 할당 확률 분포의 공통영역이 작은) 표준화 평균 차이

약어: U : 원본 자료, P : 삼각형 들레를 이용한 매칭 방법, R : 원의 반지름을 이용한 매칭 방법, C : 중심점에서의 거리를 이용한 매칭 방법, MW : 매칭 가중치 방법, Ip : 안정된 역화를 가중치 방법.

표2-1. 상황2에서 좋은 공변량 중복인 경우(치료 할당 확률 분포의 공통영역이 큰) 치료 효과 비교

prevalence ¹	method ²	bias			MSE		
		1vs2	1vs3	2vs3	1vs2	1vs3	2vs3
33:33:33	U	0.059	0.202	0.144	0.130	0.169	0.154
	P	0.004	0.019	0.016	0.237	0.243	0.297
	R	0.010	0.036	0.027	0.186	0.234	0.235
	C	-0.033	0.061	0.091	0.265	0.212	0.343
	MW	-0.001	0.066	0.067	0.124	0.135	0.147
	Ip	0.001	0.069	0.068	0.125	0.135	0.138
10:45:45	U	0.157	0.228	0.070	0.308	0.313	0.081
	P	0.018	0.097	0.057	0.517	0.469	0.499
	R	0.006	0.001	-0.010	0.498	0.477	0.459
	C	0.071	0.083	0.016	0.534	0.551	0.522
	MW	0.085	0.092	0.006	0.294	0.268	0.082
	Ip	0.133	0.136	0.002	0.364	0.314	0.076
10:10:80	U	0.050	0.272	0.218	0.445	0.366	0.325
	P	-0.070	0.150	0.192	0.538	0.493	0.601
	R	-0.137	0.138	0.181	0.574	0.579	0.545
	C	0.044	0.174	0.102	0.519	0.592	0.494
	MW	-0.018	0.149	0.164	0.477	0.322	0.327
	Ip	0.001	0.178	0.174	0.522	0.387	0.337

¹ prevalence : 치료할당 분율, ² method (U: 원본 자료, P : 삼각형 들레를 이용한 매칭 방법, R : 원의 반지름을 이용한 매칭 방법, C : 중심점에서의 거리를 이용한 매칭 방법, MW : 매칭 가중치 방법, Ip : 안정된 역확률 가중치 방법)

표2-2. 상황2에서 좋지 않은 공변량 중복인 경우(치료 할당 확률 분포의 공통영역이 작은) 치료 효과 비교

prevalence ¹	method ²	bias			MSE		
		1vs2	1vs3	2vs3	1vs2	1vs3	2vs3
33:33:33	U	0.445	0.994	0.549	0.384	1.169	0.421
	P	0.220	0.323	0.090	0.598	0.631	0.555
	R	0.184	0.349	0.139	0.573	0.717	0.676
	C	0.240	0.364	0.108	0.596	0.761	0.504
	MW	0.357	0.322	-0.009	0.664	0.980	0.447
	Ip	0.437	0.649	0.212	0.795	1.186	0.360
10:45:45	U	0.371	1.064	0.686	0.471	1.445	0.576
	P	0.009	0.232	0.131	0.240	0.384	0.263
	R	0.048	0.225	0.186	0.261	0.431	0.362
	C	0.109	0.384	0.179	0.259	0.387	0.275
	MW	0.200	0.199	0.003	1.038	1.445	0.697
	Ip	0.589	0.839	0.218	1.583	1.911	0.350
10:10:80	U	0.086	0.823	0.735	0.611	0.998	0.934
	P	0.031	0.219	0.200	0.242	0.330	0.298
	R	0.038	0.232	0.228	0.390	0.445	0.448
	C	0.065	0.249	0.224	0.291	0.379	0.343
	MW	-0.095	-0.056	-0.024	1.211	1.085	1.087
	Ip	0.215	0.866	0.684	1.872	1.951	1.514

¹ prevalence : 치료할당 분율, ² method (U: 원본 자료, P : 삼각형 돌레를 이용한 매칭 방법, R : 원의 반지름을 이용한 매칭 방법, C : 중심점에서서의 거리를 이용한 매칭 방법, MW : 매칭 가중치 방법, Ip : 안정된 역확률 가중치 방법)

제 4 장 결론 및 고찰

본 연구에서는 치료군이 3 개인 경우 관찰연구에서 발생하는 편향을 줄이기 위하여 성향점수를 이용한 5 가지 방법을 2 가지 가정상황에서 비교하였다.

상황 1 에서 three-way 매칭 방법은 전체적으로 비슷한 성능을 나타냈다. matched sample size 는 치료할당 분율에 따라 차이를 보였고, 표준화 평균 차이, 편향, 평균 제곱 오차는 공변량 중복 수준과 치료의 분율에 관계없이 비슷한 성능을 보였다. 매칭 가중치 방법은 표준화 평균 차이, 편향에서 가장 좋은 성능을 보였지만 좋지 않은 공변량 중복에서 가중치를 사용한 표본수가 가장 작고, 평균 제곱 오차가 크게 나타났다. 안정된 역확률 가중치 방법은 원본 자료와 동일한 표본수를 갖지만 좋지 않은 공변량 중복에서 성능이 크게 떨어졌다.

상황 2 에서 three-way 매칭 방법은 전체적으로 비슷한 성능을 보였다. 매칭 가중치 방법은 좋은 공변량 중복에서는 가장 좋은 성능을 나타냈지만 좋지 않은 공변량 중복에서는 성능이 떨어졌다. 안정된 역확률 가중치 방법은 좋지 않은 공변량 중복에서 여전히 성능이 크게 떨어졌다. 측정되지 않은 변수가 있는 경우를 가정하였을 때, 측정되지 않은 변수에서의 표준화 평균 차이는 가중치 방법보다 three-way 매칭 방법의 성능이 더 좋았다. 좋은 공변량 중복에서는 매칭 가중치 방법이 가장 성능이 좋았지만, 좋지 않은 공변량 중복에서는 가중치 방법은 성능이 떨어졌다. 좋지 않은 공변량 중복에서 매칭 가중치 방법은 표본수가 감소하므로 평균

제곱 오차가 커지는 것을 알 수 있다. three-way 매칭 방법과 매칭 가중치 방법은 치료의 분율과 공변량 중복 수준에 영향을 받는 것으로 나타났고, 이것은 추정치가 자료의 특성에 영향을 받는다는 것을 의미한다. 안정된 역확률 가중치 방법은 치료의 분율과 공변량 중복에 관계없이 원본 자료와 동일한 표본수를 나타내므로 이것은 three-way 매칭, 매칭 가중치 방법과 달리 전체 집단에서의 영향을 추정한다는 것을 의미한다. 좋지 않은 공변량 중복에서 안정된 역확률 가중치 방법의 성능이 좋지 않은 것은 일부 대상자가 공통 영역 외부에 존재할 때 안정된 역확률 가중치가 전체 집단을 추정할 수 없을 가능성이 높다는 것을 의미한다. 반면 three-way 매칭 방법과 매칭 가중치 방법은 공통 영역의 효과가 강조되었기 때문에 대부분의 결과가 합리적인 것으로 나타났다고 할 수 있다.

현재 연구에서는 three-way 매칭 방법에서 사용하는 캘리퍼의 알려진 기준이 없으므로 three-way 매칭 방법의 표본수를 비교하는 것은 의미가 없고, 각 방법에서 캘리퍼의 폭을 정의하는 기준에 대한 연구가 필요하다. 또한 치료군이 4 개 이상인 경우에는 거리를 정의하기 어렵고 계산량이 기하급수적으로 증가하는 한계가 있다. 따라서 치료군이 4 개 이상인 경우 계산량을 줄이면서 성능을 향상시킬 수 있는 방법에 대한 연구와, 통계량의 안정성을 위하여 붓스트랩을 이용한 추후연구가 필요하다.

참고 문헌

Rosenbaum P.R and Rubin D.B. (1983a). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70: 41–55.

Rosenbaum P.R and Rubin D.B. (1983b). Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *Journal of the Royal Statistical Society. Series B(Statistical Methodology)*, 45:212–218.

Rosenbaum P.R and Rubin D.B. (1984). Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American statistical Association*, 79: 516–524.

Rosenbaum P.R and Rubin D.B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, 39: 33–38.

Imbens G.W. (2000). The role of the propensity score in estimating dose–response functions. *Biometrika*, 87(3): 706–710.

Kazuki Yoshida. (2017). Matching weights to simultaneously compare three treatment groups: Comparison to three–way matching. *Epidemiology*, 28(3):387–395.

Li L, Greene T. (2013). A weighting analogue to pair matching in propensity score analysis. *Int J Biostat*, 9(2):215–234.

Xu S, Ross C, Raebel MA, Shetterly S, Blanchette C, Smith D. (2010). Use or stabilized inverse propensity scores as weights to directly estimate relative risk and its confidence intervals. *Value Health*, 13(2):273–277.

임보라미. (2012). 처치가 다범주인 경우 성향점수 보정방법의 비교. 석사학위논문. 고려대학교, 서울.

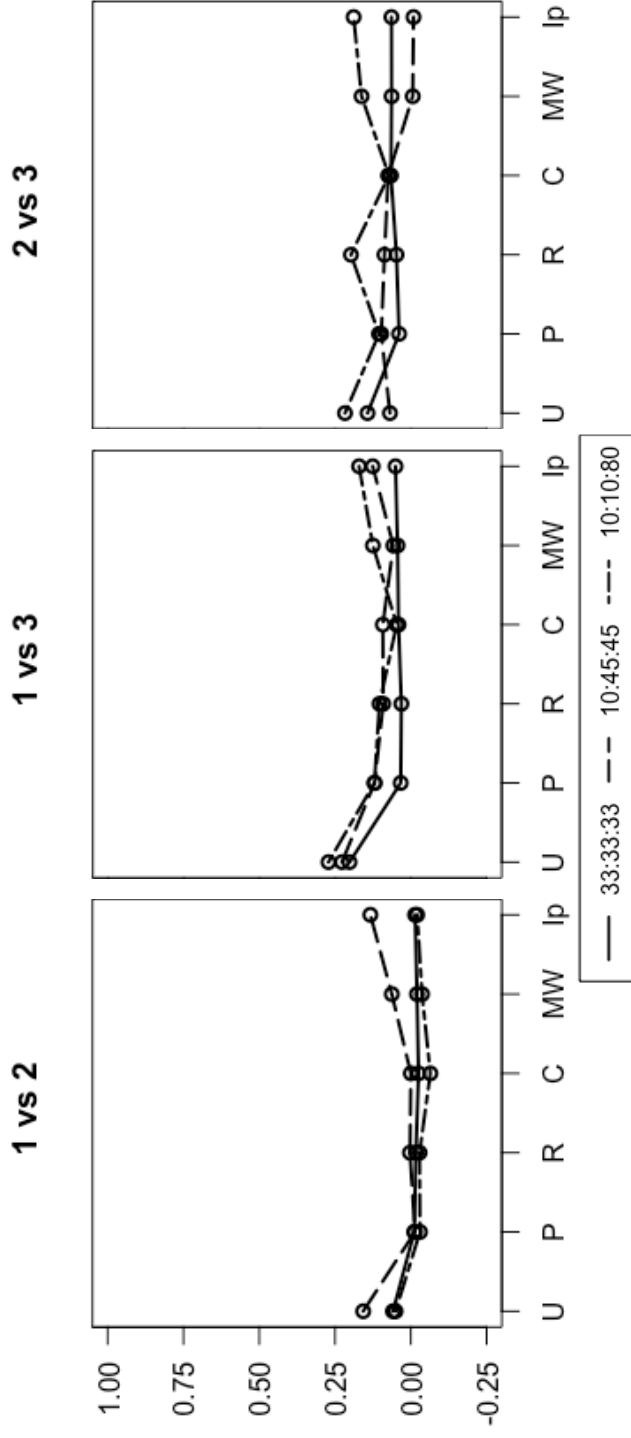
Jeremy A. Rassen. (2013). Matching by propensity score in cohort studies with three treatment groups. *Epidemiology*, 24(3):401–409.

Jessica M. Franklin. (2012). Metrics for covariate balance in cohort studies of causal effects. *Statistics in Medicine*, 33:1685–1699.

Greenland S. (2004). An overview of methods for inference from observational studies. *Applied Bayesian modeling and causal inference from incomplete–data perspectives*. Wiley New York 2004; ISBN: 0–470–09043–X.

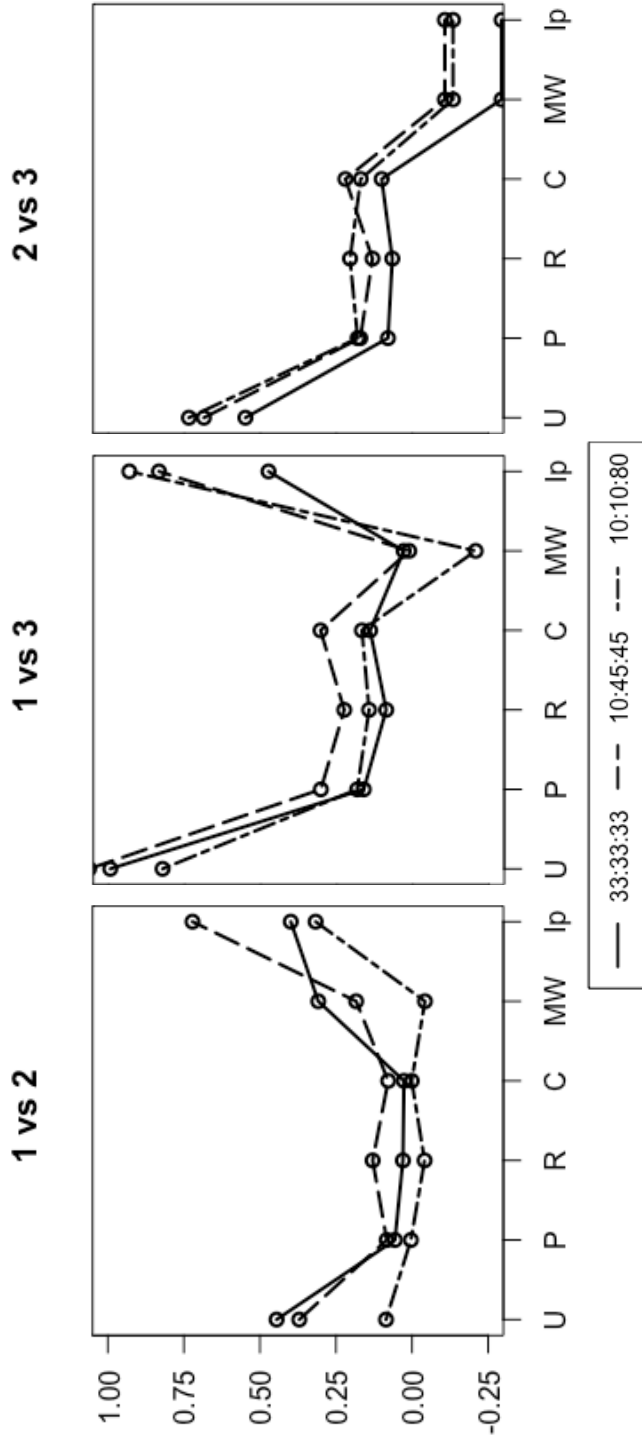
Onur Baser. (2008). Propensity score matching with multi–level categories: an application. *International society for pharmacoconomics and outcomes research*.

목
차



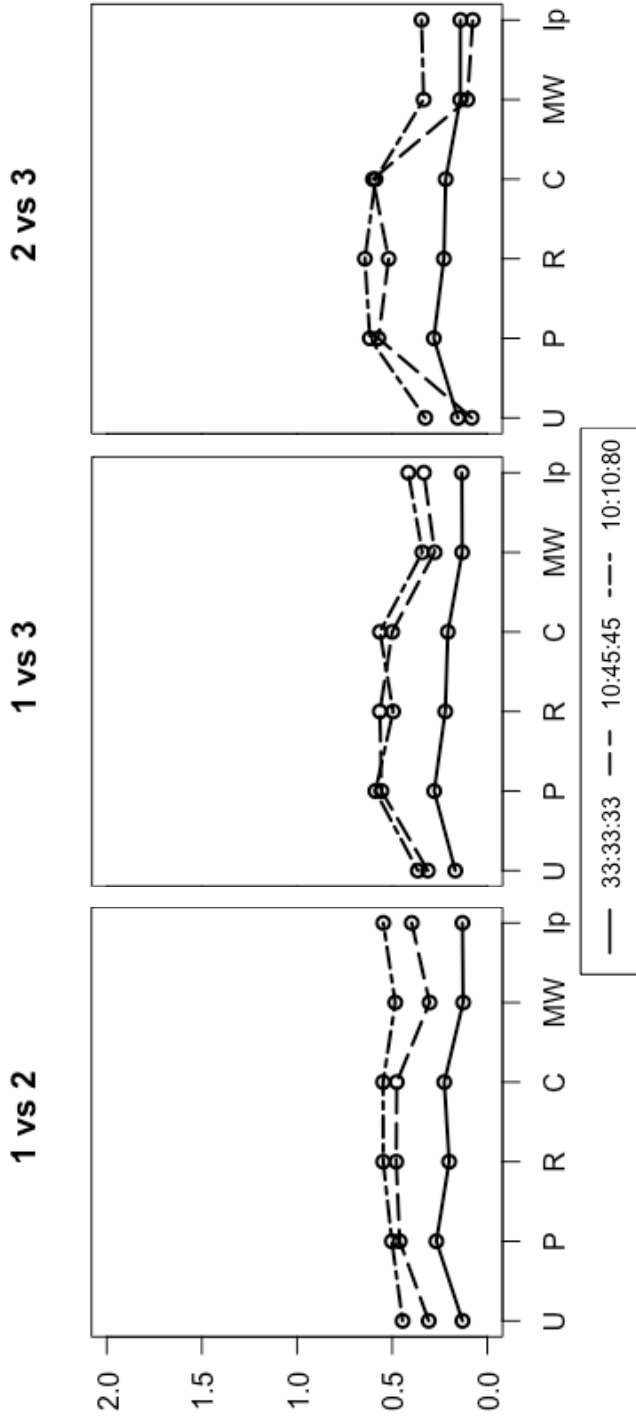
부록 그림 6-1. 상황1에서 좋은 공변량 중복인 경우(치료 할당 확률 분포의 공통영역이 큰) 편향

약어: U : 원본 자료, P : 삼각형 돌레를 이용한 매칭 방법, R : 원의 반지름을 이용한 매칭 방법, C : 중심점에서
 거리를 이용한 매칭 방법, MW : 매칭 가중치 방법, Ip : 안정된 역확률 가중치 방법.



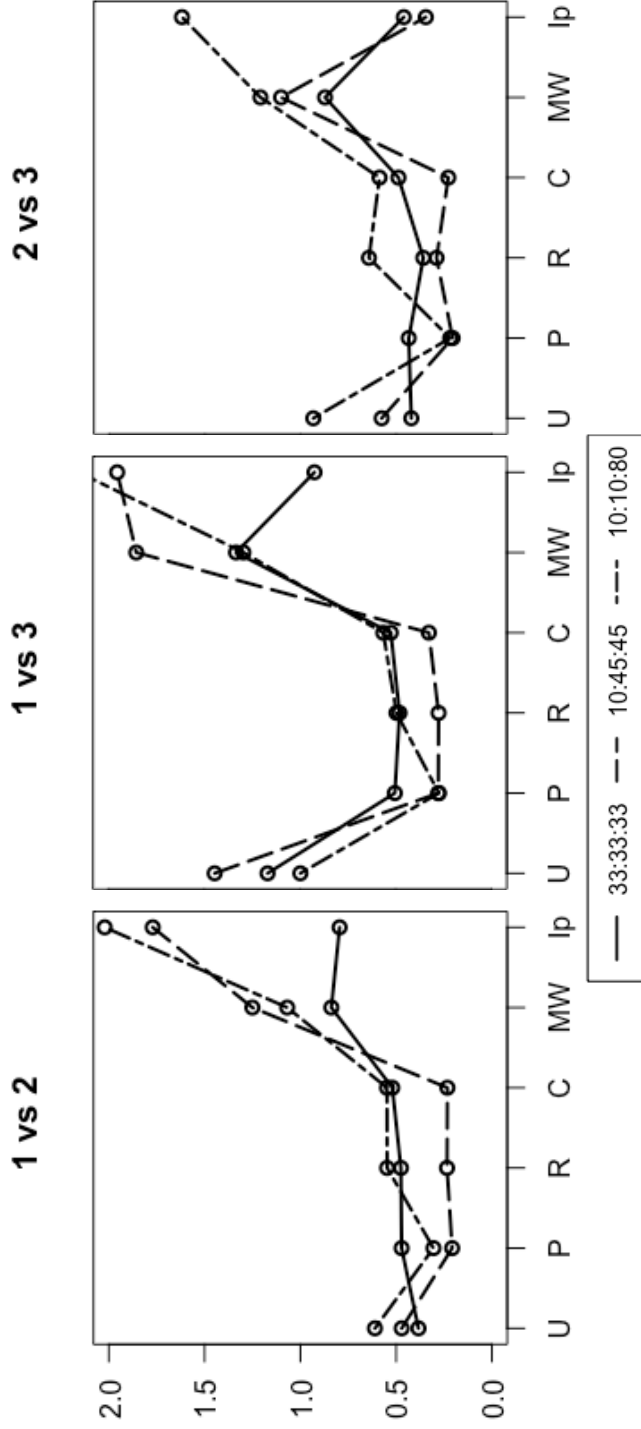
부록 그림 6-2. 상황1에서 좋지 않은 공변량 중복인 경우 (치료 할당 확률 분포의 공통영역이 작은) 편향

약어: U : 원본 자료, P : 삼각형 틀레를 이용한 매칭 방법, R : 원의 반지름을 이용한 매칭 방법, C : 중심점에서의 거리를 이용한 매칭 방법, MW : 매칭 가중치 방법, Ip : 안정된 역확률 가중치 방법.



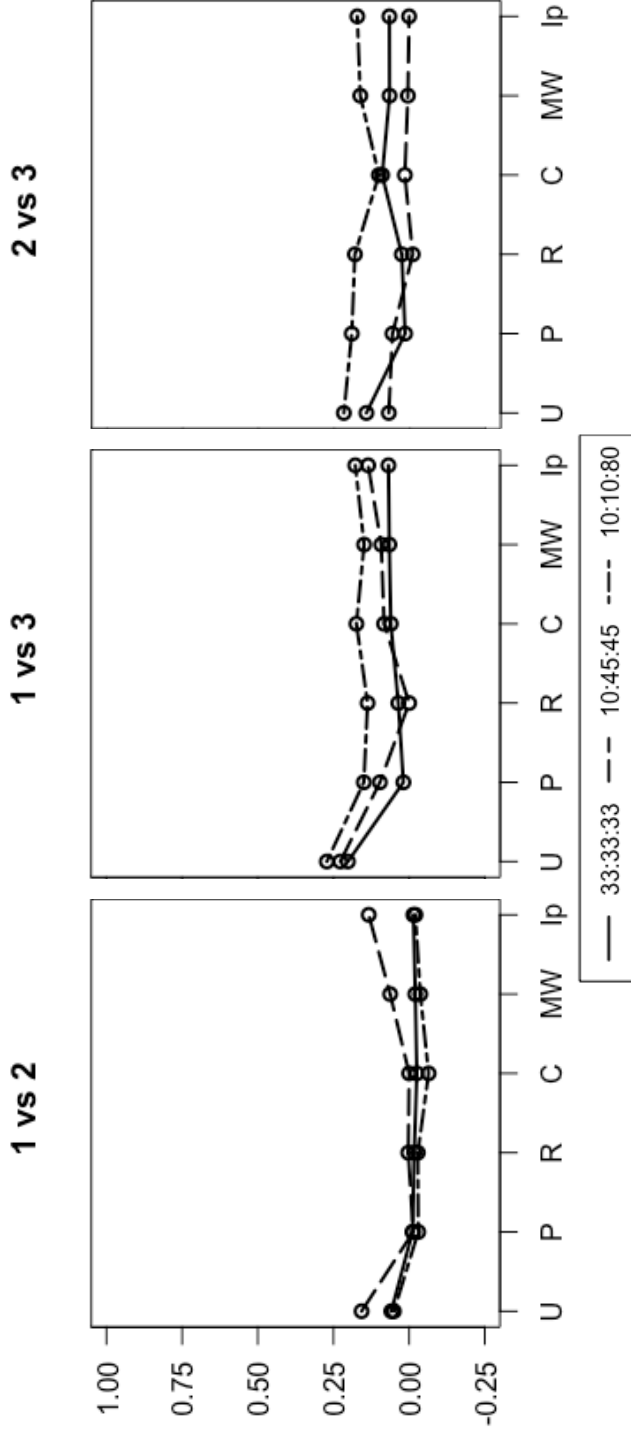
부록 그림 7-1. 상황1에서 좋은 공변량 중복인 경우(치료 할당 확률 분포의 공통영역이 큰) 평균 제곱 오차

약어: U : 원본 자료, P : 삼각형 들레를 이용한 매칭 방법, R : 원의 반지름을 이용한 매칭 방법, C : 중심점에서의 거리를 이용한 매칭 방법, MW : 매칭 가중치 방법, Ip : 안정된 역확률 가중치 방법.



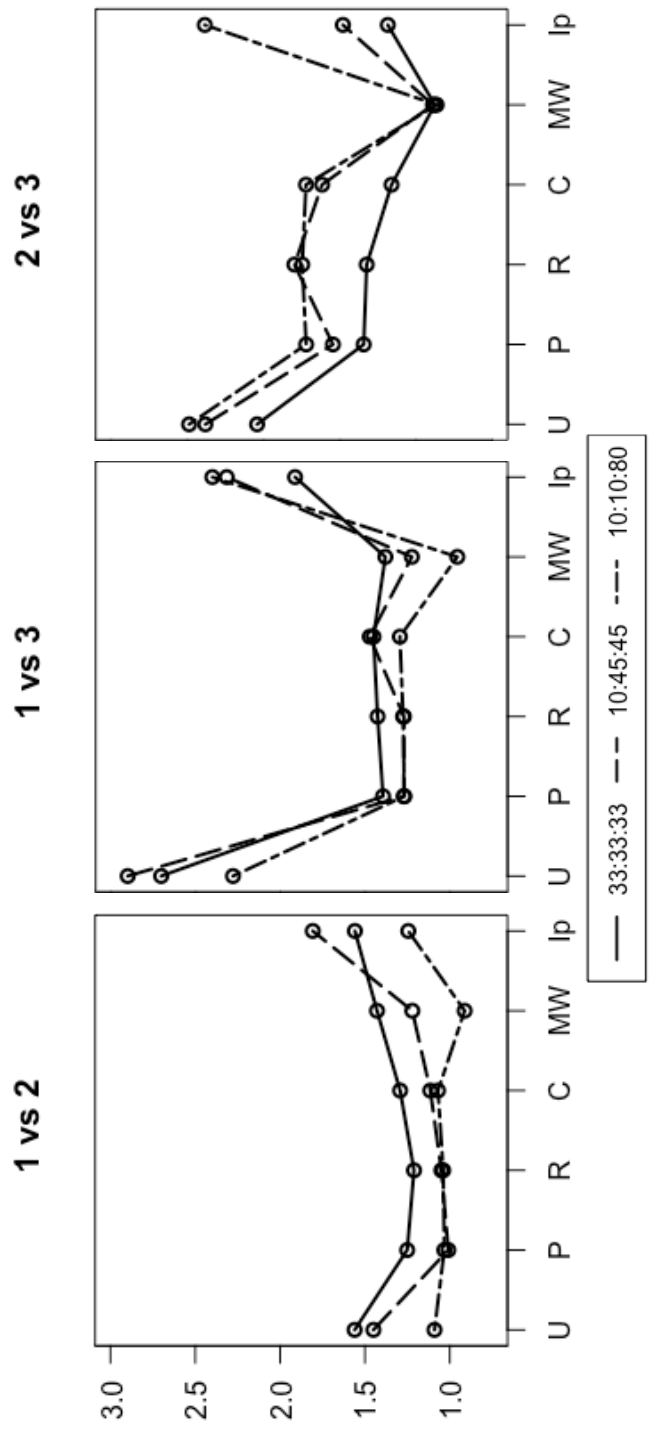
부록 그림 7-2. 상황1에서 좋지 않은 공변량 중복인 경우 (치료 할당 확률 분포의 공통영역이 작은) 평균 제곱 오차

약어: U : 원본 자료, P : 삼각형 틀레를 이용한 매칭 방법, R : 원의 반지름을 이용한 매칭 방법, C : 중심점에서의 거리를 이용한 매칭 방법, MW : 매칭 가중치 방법, Ip : 안정된 역확률 가중치 방법.



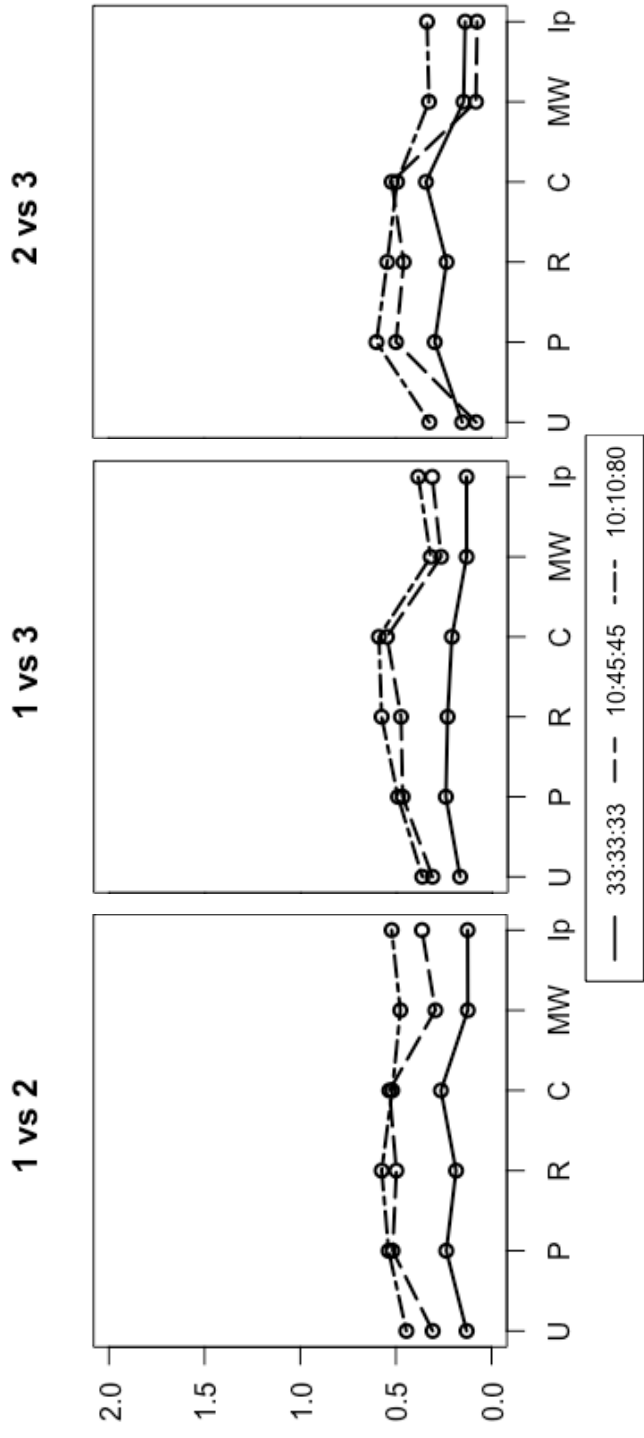
부록 그림 8-1. 상황 2에서 좋은 공변량 중복인 경우(치료 할당 확률 분포의 공통영역이 큰) 편향

약어: U : 원본 자료, P : 삼각형 돌레를 이용한 매칭 방법, R : 원의 반지름을 이용한 매칭 방법, C : 중심점에서의 거리를 이용한 매칭 방법, MW : 매칭 가중치 방법, Ip : 안정된 역확률 가중치 방법.



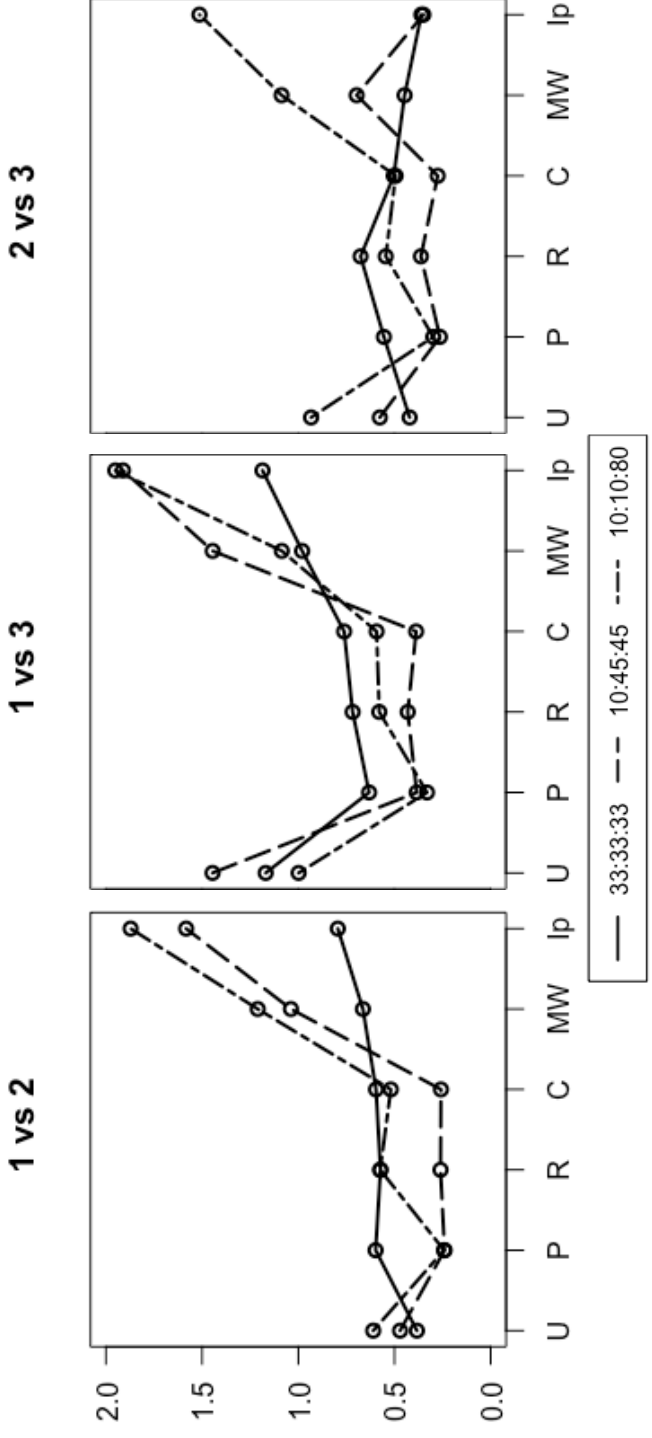
부록 그림 8-2. 상황 2에서 좋지 않은 공변량 중복인 경우 (지료 할당 확률 분포의 공통영역이 작은) 편향

약어: U : 원본 자료, P : 삼각형 들레를 이용한 매칭 방법, R : 원의 반지름을 이용한 매칭 방법, C : 중심점에서의 거리를 이용한 매칭 방법, MW : 매칭 가중치 방법, Ip : 안정된 역확률 가중치 방법.



부록 그림 9-1. 상황 2에서 좋은 공변량 중복인 경우(치료 할당 확률 분포의 공통영역이 큰) 평균 제곱 오차

약어: U : 원본 자료, P : 삼각형 물레를 이용한 매칭 방법, R : 원의 반지름을 이용한 매칭 방법, C : 중심점에서의 거리를 이용한 매칭 방법, MW : 매칭 가중치 방법, Ip : 안정된 역확률 가중치 방법.



부록 그림 9-2. 상황 2에서 좋지 않은 공변량 중복인 경우(치료 할당 확률 분포의 공통영역이 작은) 평균 제곱 오차

약어: U : 원본 자료, P : 삼각형 둘레를 이용한 매칭 방법, R : 원의 반지름을 이용한 매칭 방법, C : 중심점에서의 거리를 이용한 매칭 방법, MW : 매칭 가중치 방법, Ip : 안정된 역확률 가중치 방법.

ABSTRACT

Comparison of methods using propensity score for three treatment groups : A simulation study.

Shin, Jeong Soo

Dept. of Biostatistics and Computing

The Graduate School

Yonsei University

Randomized controlled clinical trials are often difficult to implement due to cost and ethical issues. In such cases, we can design observational studies. However, observational studies are included bias in the effectiveness of treatment group. Several methods using generalized propensity scores have been proposed to estimate the effects of multiple treatment groups without bias.

This study extends Kabuki Yoshida' s study and compares the performance of methods using propensity score. There are stabilized IPTW, matching weight as weighting method using propensity score and three-way matching as matching method using propensity score. The three-way matching is a method using the perimeter of the triangle, the method of using the radius of the circle,

and the distance from the center point in the two-dimensional space using the propensity score.

We compared matching weight, stabilized IPTW, and the three-way matching method in simulated datasets. There are five covariates generated independently, and five covariates generated by the interactions and secondary term of these variables. We generate data set with three treatment groups and binary outcome. The propensity score model was fit by multinomial logit regression. When the common support is large, it is called good covariate overlap. When the common support is small, it is called poor covariate overlap.

In situation 1, ten covariates were used to estimate propensity score. In situation 2, five covariates generated independently were used to estimate propensity score. We calculated standardized mean differences for three pairwise contrasts and averaged them for each covariate. Bias and MSE are used to compare treatment effects.

In situation 1, three-way method performed similar. Matching weight method performed well in standardized mean differences and bias but weighed sample size is small and MSE is large in poor overlap. Stabilized IPTW method sample size is equal to the sample size of original data but performance is reduced in poor overlap.

In situation 2, three-way method performed similar. In the matching weight method, the difference in standardized mean difference is large according to the treatment prevalence. Matching weight sample size (weighted sample size) is small and MSE is large in poor overlap. In the stabilized IPTW method, weighted sample size is equal to the sample size of original data but performance is reduced in poor overlap.

In the current study, there is no known standard of caliper width for three-way matching method. If there are more than four treatment groups, there is a limit to the exponential increase of the calculation amount.

Key words : propensity score matching, three-way matching, inverse probability of treatment weighting, matching weight