

## Propensity score matching method의 소개

고려대학교 의과대학 구로병원 마취통증의학과

이동규

### An introduction to propensity score matching methods

Dong Kyu Lee

Department of Anesthesiology and Pain Medicine, Korea University Guro Hospital, Seoul, Korea

Propensity score matching method (PSM) is widely used in observational study to reduce selection bias. Observational study lacks randomization, hence, statistical inferences without bias adjustments usually include observed or unobserved effects of covariates. If a subject with specific characteristics has a higher chance to be selected for a specific treatment, the characteristics have a possible effect on statistical results. PSM is the method for controlling covariate imbalance that produces the selection bias. In this paper, we introduce the basic concepts of PSM and simplified methods of PSM process. However, PSM is a rapidly developing statistical area with many limitations and some disadvantages. These points are described in the concluding section to emphasize the importance of considering the various features of PSM in the study design. (*Anesth Pain Med* 2016; 11: 130-148)

**Key Words:** Observational study, Propensity score matching method, Selection bias.

### 서 론

전향적 연구(prospective study)는 특정 처치(treatment)를 받는 처치군(treatment group)과 대조군(control group)에 해당되

Received: January 18, 2016.

Revised: February 2, 2016.

Accepted: February 22, 2016.

Corresponding author: Dong Kyu Lee, M.D., Ph.D., Department of Anesthesiology and Pain Medicine, Korea University Guro Hospital, 148, Gurodong-ro, Guro-gu, Seoul 08308, Korea. Tel: 82-2-2626-1437, Fax: 82-2-2626-1438, E-mail: entopic@naver.com

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

는 연구대상을 선정하는 것에서 출발한다. 특정 처치가 나타내는 결과를 추정(causal inference) 하기 위하여 연구대상을 선정할 때는 연구 설계 단계에서 무작위 배정(random assignment)을 사용하여 결과에 영향을 주는 특성의 차이가 없도록 해야 한다. 그러나 임상에서 환자에게 위험(hazard)이 예상되는 경우 혹은 이미 특정 치료방법이 시행되고 있는 경우에는 무작위 배정이 윤리적 문제를 발생시킬 수 있다. 예를 들어 수술 전 간기능의 손상이 예상되는 환자를 대상으로 간기능에 영향을 미치는 마취약으로 전신마취를 시행하여 간기능의 손상이 진행 되거나, 폐암 치료방법으로 수술이 권장되는 경우 다른 방법과 비교를 하여 환자가 수술에 의해 치료를 받을 수 있는 기회를 박탈하는 경우 등이 있다.

준실험(quasi-experiment)은 무작위 배정의 과정을 거치지 않고 행하는 실험 방법이며, 대표적으로 관측연구(observational study)가 해당된다[1]. 관측연구는 무작위 배정없이 특정 집단을 대상으로 연구를 진행하기 때문에, 조금 더 현실적이고 일반적인 장점이 있다. 또한 추정된 치료효과를 임상에 적용할 수 있도록 조정이 가능하며, 특정 혼란변수에 대한 통계적 통제방법을 적용할 수 있다. 그러나 관측연구는 연구대상의 선정에서 선택편향(selection bias)을 피할 수 없고, 무작위 배정에 의하지 않기 때문에 어떤 현상의 원인을 추론하는 것이 근본적으로 불가능하다. 선택편향이란 특정 공변량(covariate)에 의하여 연구대상이 특정 처치를 받을 확률이 증가하는 것을 말하며, 이 공변량은 처치군과 대조군 사이의 결과 차이를 규명하는데 혼란변수로 작용하게 된다.

관측연구에서 선택편향을 감소시키는 방법으로 고전적으로 사용하는 방법은 짝짓기(matching)이다[1]. 짝짓기는 결과에 영향을 미치는 공변량의 수준을 맞춘 연구대상들을 처치군과 대조군에 고르게 포함시키는 방법이다. 그러나 발생률이 낮은 현상을 연구 대상으로 한다면, 짝짓기 방법으로 통계적 검정력(statistical power)을 가지는 수만명 표본을 수집하는 것은 매우 어려우며, 공변량이 여러 개라면 두 군을 완전히 맞추는 것이 불가능하고, 오히려 여러 개의 공변량을 모두 맞추면 편향되고 왜곡된 통계 결과를 도출하기도 한다. 관측연구에서 고전적인 짝짓기 방법이 가지는 단점을

극복하고 선택편향을 최소화 하기 위해 propensity score를 이용한 짝짓기 방법이 사용되고 있다. Propensity score matching (PSM)은 관측연구 뿐 아니라 무작위 배정의 적용이 어려운 후향적 연구 등과 같이 준실험에 해당되는 연구방법에서 선택편향을 감소시키는 방법으로 사용된다[2].

이 글은 PSM의 개념, 방법과 PSM이 가지는 장점과 단점 그리고 실제 사용에 있어서 주의 사항을 소개하는 데 목적이 있다.

### Propensity Score의 정의와 계산방법

Propensity score란 연구 대상이 특정 공변량에 의해 대조군이 아닌 처치군에 포함될 확률이다[3]. 예를 들어 연구대상이 대형병원에 가까이 거주할수록, 대형병원에서 치료를 받을 확률이 높아진다. Propensity score는 확률이기 때문에 0에서 1사이의 값을 가진다. 만약 연구대상이 처치군과 대조군에 무작위 배정되었다고 가정한다면, 이때의 propensity score는 0.5가 된다. 그러나 실제 준실험 설계에서 무작위 배정은 불가능하므로 각 공변량에 대한 propensity score를 추정하여 이 점수(확률)를 이용한 조정(adjustment)을 시행하게 된다. Propensity score는 공변량과 특정 처치를 받아 처치군에 포함되는 것과의 연관성에 따라 결정된다. 다른 공변량의 효과가 없다고 가정한다면, 특정 공변량에 대해 같은 propensity score를 가지는 연구대상들이 처치군과 대조군에 같은 수로 포함된다면 이 두 집단을 대상으로 한 통계적 추론의 결과는 같은 propensity score를 가지는 공변량에 의한 효과보다는 특정 처치를 받아 처치군에 포함되었기 때문에 발생한 차이에서 기인한 것이라고 할 수 있다.

Propensity score는 두가지 방법으로 계산한다[1]. 한가지 방법은 2003년에 Lemon 등이 처음 소개한 non-parametric decision tree method로 'classification and regression tree analysis (CART)'라는 방법이다. 그러나 이 방법은 너무 복잡하고 특정 통계적 상황에서만 적용이 가능하기 때문에 널리 사용되지 않는다. 다른 방법은 로지스틱 회귀분석을 사용하는 방법이다. 이 방법은 처치군에 포함되는 경우를 1, 대조군에 포함되는 경우를 0으로 하는 이항반응(binary response) 형태로 종속변수를 설정하고, 보정하려는 공변량을 독립변수로 지정하여 로지스틱 회귀분석을 시행한다. 이 경우 로지스틱 회귀분석은 propensity score model을 추정하며, 이 모형에 의해 각 대상들의 추정된 확률(각 대상이 주어진 공변량에 의해 처치군에 포함될 확률)이 propensity score에 해당된다. 이후 대조군과 처치군에 포함된 모든 연구대상에 대하여 propensity score가 같은 혹은 유사한 대상끼리 짝을 맞추어 자료를 선정하게 되며 짝을 이루지 못한 것들은 통계분석에서 제외한다.

### Propensity Score를 이용한 짝짓기 방법들

PSM이 등장한 이후로 다양한 matching method가 소개되었으나, 최근에는 Greedy matching method와 optimal matching method를 많이 사용한다. 이 두가지 방법은 다른 PSM의 장점들을 사용하여 이루어지며, 방대한 양의 자료를 다루기에 적합하다. PSM은 다음에 설명되는 방법들 중 단독으로 혹은 몇가지 방법을 함께 사용하여 왔으며, 마지막에는 최근 많이 사용하는 Greedy와 optimal matching method에 대해서 기술하였다[4,5].

#### 층화 짝짓기(stratified matching)

Propensity score의 범위에 따라 몇 개의 간격으로 연구에 포함된 개체를 분류하는 방법이다. 다섯 개 정도의 층(strata)으로 자료를 분류하면 공변량에 의한 bias를 90-95% 가량 제거할 수 있다. 또한 통계분석 결과가 관찰되지 않은 공변량(unobserved covariate)에 의해 영향을 받는 것이 의심되는 경우, 층화 짝짓기 방법이 매우 유용하다. 비슷한 propensity score를 가지는 대상들을 같은 계층에 모아 분석을 하면, 관찰되지 않은 공변량에 의한 효과가 의미 있게 감소하는 것으로 알려져 있다[5].

#### 최 근접 짝짓기(nearest neighbor matching)

대조군과 처치군에 포함된 모든 연구 대상들의 추정된 propensity score 차이의 절대값이 가장 작은 순서대로 짝짓기를 하는 방법이다.

#### N : N matching

대조군과 처치군에 포함된 연구 대상을 무작위로 나열한 뒤, 처치군의 첫번째 개체부터 N개의 대조군 중에서 propensity score가 가장 차이가 적은 개체와 짝짓기를 하는 방법이다. 짝짓기의 방법은 대조군과 처치군의 각각 한 개씩 짝을 이루게 하면 1 : 1, 대조군 하나와 여러 개의 처치군을 짝을 이루게 하면 1 : N, 그 반대의 경우에는 N : 1 matching이 되며 상황에 따라 선택하여 사용한다. 1 : 1 matching은 대조군과 처치군의 propensity score가 비슷하지 않으면 많은 수의 처치군의 자료가 탈락되는 단점이 있다. 따라서 처치군과 대조군의 표본수가 같은 경우에는 1 : 2 matching이 1 : 1 matching보다 좋은 검정력을 가지나, 1 : 5 이상으로 matching을 하면 검정력의 증가는 미미한 것으로 알려져 있다. 또한, 대조군의 표본수가 작은 경우에는 한번 짝짓기에 사용된 대조군의 연구대상을 반복해서 사용하는 'matching with replacement'방법이 효과적이다. 이 방법을 사용하는 경우에는 대조군의 자료 분포를 설명할 때 반복되어 사용된 개체가 있음을 명시해야 한다. 만약 대조군의 표

본수가 충분히 많다면 대조군과 처치군을 1 : 2로 짝짓기를 하는 것이 적절하다.

### Radius matching

처치군의 propensity score로부터 미리 설정한 간격 이내의 대조군을 짝짓기 하는 방법이다. 모든 처치군의 연구대상이 짝을 이루게 되나, 이들의 propensity score 차이는 반드시 설정한 간격 이내이어야 한다.

### Kernel matching

처치군과 대조군의 propensity score의 차이에 반비례하는 값에 의하여 가중치(weight)를 결정하고 대조군의 가중평균(weighted average)에 의거하여 처치군의 개체들과 짝을 이루는 방법이다.

### Caliper matching

추정된 propensity score의 표준오차 (standard error)의 1/4에 해당되는 값을 범위로 지정하여 짝짓기에 사용한다. 대조군과 처치군의 짝을 이룰 대상들 사이의 propensity score 차이가 이 범위 내에 해당되는 경우에만 짝을 이루어 분석에 포함시키고 제외되는 모든 개체는 분석에서 제외한다.

### Mahalanobis metric matching

처치군의 연구대상과 Mahalanobis distance가 가장 작은 대조군이 짝을 이루는 방법이다. 짝짓기는 모든 처치군 대상이 짝을 이룰 때까지 시행되며, 남은 대조군 개체들은 분석에서 제외한다.

Mahalanobis distance는 다음과 같이 정의된다(등식 1).

$$D_{ij} = \sqrt{(x_i - y_j)^T S^{-1} (x_i - y_j)} \quad \text{등식 1}$$

$S^{-1}$ : 짝을 이루는 변수들의 공분산행렬의 역행렬

$x_i, y_j$ : propensity score에 따라 짝을 이룬 변수들

$(x_i - y_j)^T$ : 공분산행렬의 전달행렬(transfer matrix)

관찰연구에서 Mahalanobis matching method는 block randomization design과 같은 것으로 간주한다. 또한 대조군의 표본수가 많지만 비대칭 분포를 하는 경우에는 Radius나 Kernel 혹은 Mahalanobis matching이 사용된다.

### Greedy matching method

Greedy matching method는 다양한 post-matching analysis이 가능하기 때문에 유용하다[6]. 또한 연구 대상이 정규분포를 하지 않거나, 연속변수가 아닌 경우에는 Caliper matching에서 정의된 범위 내로 최 근접 짝짓기를 이용한 Greedy matching method가 적절하며, 표본수가 작은 경우에는 Ma-

halanobis metric matching을 사용한 Greedy matching이 유용하다. 여러 장점에도 불구하고 Greedy matching method는 몇 가지 제한점을 가지고 있기 때문에 널리 사용되지는 않는다. 첫째, 이 방법은 가능한 정확한 matching을 하기 때문에 많은 수의 연구대상이 배제되게 된다. 둘째, Greedy matching은 대조군과 처치군의 추정된 propensity score들의 분포가 서로 겹쳐야 가능하며 대푯값(예: 평균 혹은 중앙값)이 유사하여야 사용이 가능하다. 이외에도 다양한 논문에서 제한점들이 제기되고 있다[6,7].

### Optimal matching method

PSM의 방법 중 가장 최근에 소개된 Optimal propensity score matching (OPMS)은 network flow theory에 기초하여 짝짓기를 최적화하며, 관찰연구에서 bias를 통제하는데 가장 많이 사용된다[7,8]. 특히 2002년에 Hansen 등이 통계 프로그램인 R에서 사용할 수 있는 optmatch package를 공개하면서부터 사용빈도가 급증하였다. OPMS는 유사한 propensity score를 가진 대조군과 처치군의 연구대상들이 하나의 계층(stratum)으로 분류하여, 자료 전반에 걸쳐 층화(stratification)를 시행한다. 여러 개의 계층으로 나누어지는 과정 중에 짝짓기가 이루어지며, 각 계층 내에서 처치군과 대조군 표본수의 비율에 따라 matching process가 결정된다. Matching process는 1:1 matching (pair matching), 1:N matching (variable ratio 혹은 variable matching, 하나의 처치군에 여러 개의 대조군을 짝짓기 하는 방법), N:N matching (full matching, 여러 개의 처치군과 여러 개의 대조군을 짝짓기 하는 방법)중 선택이 가능하다. 이 방법으로 전체 표본에 대한 propensity score의 통계적 거리(statistical distance)를 최소화하는 계층을 만들어 통계분석을 시행한다.

### Validating PMS

Propensity score가 같은 대상들만 짝짓기를 했다면 짝짓기가 완료된 자료는 완벽하게 균형을 이루는 처치군과 대조군을 가지게 된다. 그러나 propensity score는 propensity score model에 포함된 모든 공변량에 대한 확률이고, PSM은 propensity score가 가까운 것들끼리 짝짓기를 이루는 과정이다. 따라서 PSM후에는 짝짓기 과정의 적절성에 대해 검토가 필요하다.

Balance diagnostics는 공변량이 연속변수 이거나 혹은 이분변수(binary variable)인 경우에 평균과 표준편차 혹은 빈도를 이용하여 두 군의 공변량의 차이를 알아보는 가장 간단한 방법이다. 만약 공변량이 범주형 변수라면 가변수(dummy variable)를 생성하여 분석이 가능하다. 이 방법에 의한 표준화된 차이(standardized difference; d, 등식 2, 3)가 0.1보다 작다면 두 군의 공변량에 의한 차이는 무시할 정도

라고 결정한다[7].

$$d = \frac{(\bar{x}_{\text{처치군}} - \bar{x}_{\text{대조군}})}{\sqrt{\frac{s_{\text{처치군}}^2 + s_{\text{대조군}}^2}{2}}} \quad \text{등식 2}$$

$\bar{x}$ : 처치군과 대조군의 공변량의 평균

$s^2$ : 처치군과 대조군의 공변량의 분산

$$d = \frac{(\hat{p}_{\text{처치군}} - \hat{p}_{\text{대조군}})}{\sqrt{\frac{\hat{p}_{\text{처치군}}(1 - \hat{p}_{\text{처치군}}) + \hat{p}_{\text{대조군}}(1 - \hat{p}_{\text{대조군}})}{2}}} \quad \text{등식 3}$$

$\hat{p}$ : 처치군과 대조군의 공변량의 빈도

PSM을 이용한 자료는 짝짓기 전후의 balance diagnostics를 함께 표시하여 짝짓기 결과의 적절함을 반드시 밝혀야 하며, 가능하다면 paired, variable 그리고 full matching의 타당성 검사 결과를 모두 제시하여야 한다[9].

### Matching Data의 분석

PSM에 의해 생성된 자료의 분석은 일반적인 이분변수를 분석하는 방법과 같다. 결과 변수가 중도 절단된 형태의 자료(censored data)라면, 생존분석을 이용하는 방법도 가능하다. 흔하게는 stratified log-rank test나 stratified Cox regression을 사용한다[10].

Optimal matching method를 사용한 경우 회귀분석이 유용하다. 1:1 optimal matching method를 예를 들면, 짝을 이룬 두 군의 개체들의 결과변수 차이를 종속변수로 지정하고, 짝을 이룬 개체들의 공변량의 차이에 대해 regression을 시행하면 절편(intercept)이 평균적인 처치의 효과(average treatment effect)를 나타내게 된다(difference score regression). Variable ratio나 full matching의 경우에는 Hodges-Lehmann aligned rank test를 시행한다. 몇몇 통계 프로그램에서 이 방법이 package나 내장된 함수 혹은 매크로로 제공이 된다.

### 결론

준실험 형태의 연구에서 PSM을 이용하여 선택편향을 감소시키는 방법은 위양성 결과의 발현을 억제한다. 그러나 현존하는 PSM 방법은 많은 제한점을 가지고 있다. PSM 과정에는 선택되지 않고(분석에 포함되지 않고) 버려지는 표본의 수가 상당히 존재하므로, 발생율이 낮거나 표본수가 작은 경우에는 사용이 어렵다. 또한 상당히 보수적인 방법(conservative method)임에도 불구하고 선택편향을 완전히 제거할 수 없다. 이외의 제한점들로는 관찰되지 않은 공변량

에 대해서는 조정이 불가능하고, 많은 표본수를 요구하거나, 처치와 관련된 공변량을 대상으로 조정하나 그 공변량이 결과변수에 전혀 관련이 없거나 혹은 아주 높은 상관관계를 가지게 되는 경우에는 분석의 결과에 의도하지 않은 영향을 미치게 된다[11].

PSM은 최근 빠르게 발전하고 있는 분야로 여러 가지 방법들이 여전히 시험과정을 거치고 있다는 것은, 이 글에 소개한 방법들도 많은 제한점과 약점을 내재하고 있다는 것이다[11]. 따라서 PSM을 이용할 때는 각 방법이 기본적으로 내포하고 있는 가정(assumption)을 반드시 만족해야만 한다. PSM은 두가지 중요한 기본 가정을 필요로 한다. 첫번째는 conditional independence assumption이다. 즉, 한 연구대상이 처치군에 속하게 되는 것은 관찰된 특징들에 의해 결정되어야 한다. 만약 관찰되지 않은 요인이 존재한다면 matching estimator가 편향되게 된다. 두번째 가정은 common support condition 혹은 overlap condition이다. 즉, 비교를 시행하는 두 군의 propensity score들은 반드시 어느정도 중복되는 분포를 가지고 있어야 한다. 만약 중복되지 않는다면 PSM은 불가능하다.

PSM을 사용할 때는 위에 언급한 두가지 가정을 면밀하게 살펴보는 것이 반드시 필요하며, 많은 의학논문에서 잘못된 PSM에 의한 왜곡된 결과를 도출하고 있는 점도 반드시 고려해야 한다[12,13].

### PSM을 위한 Statistical Packages

통계 프로그램으로 R (The R Foundation for Statistical Computing)을 사용한다면 쉽게 matching package를 구할 수 있다[14-16]. R과 PSM을 위한 package를 사용하는 방법이 가장 사용자가 원하는 대로, 쉽게 결과를 얻을 수 있기 때문에 이 방법을 가장 추천한다. Stata (StataCorp LP, Texas, USA)의 경우에도 PSM을 위해 내장된 함수 기능이 있기 때문에 비교적 용이하게 가능하다[17-19]. SAS (SAS Institute Inc. NC, USA)나 SPSS (IBM® SPSS® Statistics, IBM Corporation, NY, USA)의 경우에는 macro 기능을 이용하여 PSM을 사용할 수 있다[20,21]. SPSS에서는 이전버전에서 사용자 정의 대화상자 패키지 기능을 이용하여 국내에서 개발된 package가 소개된 적이 있으나 최근 버전에서는 작동하지 않는 아쉬움이 있다.

### 감사의 글

I appreciate to professor Heezoo Kim in Korea univ. Guro hospital who helped me to write and proofread this article.

## REFERENCES

1. Luellen JK, Shadish WR, Clark MH. Propensity scores: an introduction and experimental test. *Eval Rev* 2005; 29: 530-58.
2. Day AG. Why the propensity for propensity scores? *Crit Care Med* 2015; 43: 2024-6.
3. Caliendo M, Kopeinig S. Some practical guidance for the implementation of propensity score matching. *J Econ Surv* 2008; 22: 31-72.
4. Heinrich C, Maffioli A, Vazquez G. A primer for applying propensity score matching. Inter-American Development Bank, 2010 [cited 2016 Jan 15]. Available from <http://idbdocs.iadb.org/wsdocs/getdocument.aspx?docnum=35320229>.
5. Hullsiek KH, Louis TA. Propensity score modeling strategies for the causal analysis of observational data. *Biostatistics* 2002; 3: 179-93.
6. Rosenbaum PR. Constructing matched sets and strata. In: *Observational studies*. 2nd ed. New York, Springer-Verlag. 2002, pp 295-331.
7. NCSS statistical Software. Chapter 123 Data Matching - Optimal and Greedy [Internet]. Kaysville: NCSS LLC. [cited 2016 Jan 15]. Available from [http://www.ncss.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Data\\_Matching-Optimal\\_and\\_Greedy.pdf](http://www.ncss.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Data_Matching-Optimal_and_Greedy.pdf).
8. Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behav Res* 2011; 46: 399-424.
9. Austin PC, Mamdani MM. A comparison of propensity score methods: a case-study estimating the effectiveness of post-AMI statin use. *Stat Med* 2006; 25: 2084-106.
10. Lu B. Propensity score matching with time-dependent covariates. *Biometrics* 2005; 61: 721-8.
11. King G, Nielsen R. Why propensity scores should not be used for matching [Internet]. 2016 [cited 2016 Jan 15]. Available from <http://gking.harvard.edu/files/gking/files/psnot.pdf>.
12. Shah BR, Laupacis A, Hux JE, Austin PC. Propensity score methods gave similar results to traditional regression modeling in observational studies: a systematic review. *J Clin Epidemiol* 2005; 58: 550-9.
13. Weitzen S, Lapane KL, Toledano AY, Hume AL, Mor V. Principles for modeling propensity scores in medical research: a systematic literature review. *Pharmacoepidemiol Drug Saf* 2004; 13: 841-53.
14. Ho DE, Imai K, King G, Stuart EA. MatchIt: nonparametric preprocessing for parametric causal inference [Internet]. *J Stat Soft* 2011 Jun; 42. Available from <http://www.jstatsoft.org/article/view/v042i08>
15. Sekhon JS. Multivariate and propensity score matching software with automated balance optimization: The matching package for R [Internet]. *J Stat Soft* 2011 May; 42. Available from <http://www.jstatsoft.org/article/view/v042i07>
16. Hansen B, Klopfer SO. Optimal full matching and related designs via network flows. *J Comput Graph Stat* 2006; 15: 609-27.
17. Leuven E, Sianesi B. PSMATCH2. Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. Statistical Software Components S432001. Chestnut Hill, Massachusetts: Boston College Department of Economics; 2003 [cited 2016 Jan 15]. Available from <https://ideas.repec.org/c/boc/bocode/s432001.html>.
18. Abadie A, Drukker D, Herr JL, Imbens GW. Implementing matching estimators for average treatment effects in Stata. *Stata J* 2004; 4: 290-311.
19. Becker SO, Ichino A. Estimation of average treatment effects based on propensity scores. *Stata J* 2002; 2: 358-77.
20. Parsons LS. Reducing bias in a propensity score matched pair sample using greedy matching techniques [Internet]. Cary NC: SAS Institute Inc., Proceedings of the Twenty-Sixth Annual SAS Users Group International Conference; 2001. 214-26 Available from <http://www2.sas.com/proceedings/sugi26/p214-26.pdf>.
21. Levesque R, Balabanov A. Raynald's SPSS Tools. Match cases on basis of propensity scores [Internet]. 2001 Dec 8 [cited 2016 Jan 15]. Available from <http://spsstools.net/en/syntax/375/>

## Appendix

## Optimal propensity score matching method using R system

## 서론

R 시스템은 공개 컴퓨터 언어로 사용의 제약이 없고 무료로 제공이 되며, 통계분석에 유용하게 사용할 수 있는 여러가지 패키지가 제공되기 때문에 쉽게 propensity score matching (PSM)을 적용할 수 있다. 이 글의 목적은 독자가 공개 컴퓨터 언어인 R 시스템과 PSM의 예제를 통하여 PSM을 이해하고 응용할 수 있도록 하는데 있다. 이 글은 예제에 사용된 통계 방법의 자세한 설명은 제외하고 PSM의 설명에 중점을 두고 있기 때문에, 통계 방법에 대한 설명은 여러 문헌을 통하여 참고할 것을 당부한다.

## R 시스템을 이용한 PSM 방법

## R 시스템의 설치

R 시스템의 최신 정보 확인과 R 시스템의 최신판을 다운로드 하는 방법은 R 홈페이지 <http://www.r-project.org>를 통해서 가능하다. 홈페이지에서 CRAN (Comprehensive R Archive Network) mirror를 통하여 최신판의 R 시스템을 설치할 수 있다.

## PSM에 필요한 패키지 설치

- R 시스템을 구동하면 상단에 패키지 메뉴를 확인할 수 있다. 이 메뉴를 통해 PSM에 필요한 몇 가지 패키지를 설치한다.
- **MatchIt**: 통계적인 인과관계 추론을 위한 짝짓기 소프트웨어이다. MatchIt은 2004년 Ho 등에[1] 의해 처음 소개되었으며 [1], 비모수적 짝짓기 방법으로 구성된 자료로부터 개선된 모수적 통계 모형을 제공한다. MatchIt에 대해 자세한 내용은 R 시스템을 구동한 뒤, help ("MatchIt") 명령을 통해 확인할 수 있다.
  - **Optmatch**: Optimal full matching을 위해 반드시 필요한 패키지이다. 2002년에 Hansen 등에[2] 의하여 처음 소개 되었으며, 관측연구에서 처치군과 대조군 사이에 공변량의 불균형을 줄이기 위해 본문에 소개된 여러가지 distance에 근거하여 쌍을 이루는 짝짓기 방법을 제공한다.

예제 파일 소개: <http://www.anesth-pain-med.org/submission/File/psm.example.apm.csv>

이 자료는 R 시스템에서 무작위수 생성의 기능을 이용하여 PSM을 시현하기 위해 만든 자료로서, 위의 링크에서 다운로드 가능하다.

일부 공변량의 불균형을 만들기 위하여 수학적 조작을 시행 하였으며, 7개의 변수로 이루어져 있고 표본수는 3000이다. 첨부된 예제 파일을 C:\temp\ 폴더에 복사한 뒤 R 시스템에서 사용하기 위하여 다음과 같은 방법을 이용한다. R 코드는 이탤릭체, 출력결과는 일반글꼴로 표시되어 있다. 앞서 기술한 바와 같이 예제 파일은 환자 일련번호(X)와 공변량(AST, ALT, HbA1c, Aspirin 복용여부)과 특정 처치 여부인 treatment 그리고 결과 변수인 bleeding으로 이루어져 있다.

```
> # Load Original dataset
> original = read.csv("c:/temp/psm.example.apm.csv")
> head(original)
```

	X	AST	ALT	HbA1c	Aspirin	treatment	bleeding
1	1	360.7339	435.8191	10.014484	0	0	0
2	2	324.7640	282.6584	9.167625	0	0	0
3	3	476.4515	344.7153	8.729882	0	0	0
4	4	484.5870	373.1589	8.300526	1	1	0
5	5	351.8709	328.7666	9.040898	0	0	0
6	6	529.5012	529.7834	6.458388	1	0	0

## 예제 파일 분석

우선 연속변수들에 대해 모수통계를 적용하기 위하여 정규성 검정과 등분산성 검정을 시행한다.

```
> ##### Initial analysis #####
> # Normality test
> shapiro.test (resid(lm(AST ~ treatment, data = original)))

        Shapiro-Wilk normality test

data: resid(lm(AST ~ treatment, data = original))
W = 0.99956, p-value = 0.7684

> shapiro.test (resid(lm(ALT ~ treatment, data = original)))

        Shapiro-Wilk normality test

data: resid(lm(ALT ~ treatment, data = original))
W = 0.99958, p-value = 0.8002

> shapiro.test (resid(lm(HbA1c ~ treatment, data = original)))

        Shapiro-Wilk normality test

data: resid(lm(HbA1c ~ treatment, data = original))
W = 0.99923, p-value = 0.2353

> # Equal variance test
> var.test (AST ~ treatment, data = original)

        F test to compare two variances

data: AST by treatment
F = 1.1074, num df = 2517, denom df = 481, p-value = 0.1567
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.9614096 1.2671484
sample estimates:
ratio of variances
 1.107368

> var.test (ALT ~ treatment, data = original)

        F test to compare two variances

data: ALT by treatment
```

```
F = 1.0678, num df = 2517, denom df = 481, p-value = 0.3631
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.927078 1.221899
sample estimates:
ratio of variances
 1.067824
```

```
> var.test (HbA1c ~ treatment, data = original)
```

```
      F test to compare two variances
```

```
data: HbA1c by treatment
```

```
F = 1.001, num df = 2517, denom df = 481, p-value = 0.999
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.8690916 1.1454722
sample estimates:
ratio of variances
 1.001034
```

처치군과 대조군의 연속변수인 AST, ALT, HbA1c는 정규성과 등분산성을 만족하고 있음을 알 수 있다.

이제 자료의 분석을 시행한다. 공변량들에 대해서는 two sample t-test와 chi-square test, test of population proportion을 시행한다.

```
> # Comparing original data
> t.test (AST ~ treatment, var.equal = T, data = original)
```

```
      Two Sample t-test
```

```
data: AST by treatment
```

```
t = 0.1618, df = 2998, p-value = 0.8715
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -9.044961 10.671970
sample estimates:
mean in group 0 mean in group 1
 399.5292      398.7157
```

```
> t.test (ALT ~ treatment, var.equal = T, data = original)
```

```
      Two Sample t-test
```

```
data: ALT by treatment
```

```
t = 0.71153, df = 2998, p-value = 0.4768
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
```



```

-6.327051 13.534593
sample estimates:
mean in group 0 mean in group 1
    409.0967      405.4929

> t.test (HbA1c ~ treatment, var.equal = T, data = original)

      Two Sample t-test

data: HbA1c by treatment
t = -0.86409, df = 2998, p-value = 0.3876
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-0.13684743 0.05312717
sample estimates:
mean in group 0 mean in group 1
    9.006312      9.048172

>
> library (MASS) # For chi-square analysis
> tbl.aspirin = table (original$Aspirin, original$treatment)
> rownames(tbl.aspirin) = c("(-)Aspirin", "(+)Aspirin")
> colnames(tbl.aspirin) = c("(-)treatment", "(+)treatment")
> tbl.aspirin

      (-)treatment  (+)treatment
(-)Aspirin      1561          278
(+)Aspirin       957          204
> chisq.test(tbl.aspirin)

      Pearson's Chi-squared test with Yates' continuity correction

data: tbl.aspirin
X-squared = 2.9992, df = 1, p-value = 0.08331

>
> tbl.bleeding = table (original$bleeding, original$treatment)
> rownames(tbl.bleeding) = c("(-)bleeding", "(+)bleeding")
> colnames(tbl.bleeding) = c("(-)treatment", "(+)treatment")
> tbl.bleeding

      (-)treatment  (+)treatment
(-)bleeding      1317          227
(+)bleeding      1201          255

> prop.test (table (original$treatment, original$bleeding), correct = F)

```

2-sample test for equality of proportions without continuity correction

```
data: table(original$treatment, original$bleeding)
X-squared = 4.3929, df = 1, p-value = 0.03609
alternative hypothesis: two.sided
95 percent confidence interval:
 0.003434916 0.100724679
sample estimates:
 prop 1 prop 2
0.5230342 0.4709544
>
```

Matching전의 자료 분석 결과를 정리하면 다음과 같다(Table 1).

**Table 1.** Statistical Analysis Results before Propensity Score Matching

Covariates				
Variables		Observed values		P value
		Control group	Treated group	
AST		399.5 ± 101.9	398.7 ± 96.9	0.872
ALT		409.1 ± 102.4	405.5 ± 99.1	0.477
HbA1c		9.01 ± 0.97	9.05 ± 0.97	0.388
Aspirin	Not used	1561 (84.9)	278 (15.1)	0.083
	Used	957 (82.4)	204 (17.6)	
Result				
Bleeding	Not occurred	1317 (52.3)	227 (47.1)	0.036*
	Occurred	1201 (47.7)	255 (52.9)	

모든 공변량에서 두 군의 차이는 없었으며, 출혈의 발생은 두 군 사이에 유의한 차이가 있다.

#### PSM: Standardized difference와 MatchIt 패키지 사용방법

본문에 언급한 standardized difference를 이용하여 짝짓기 전후의 balanced diagnostics를 비교하여 공변량의 차이를 확인하는 과정이 PSM에는 반드시 필요하다. 우선은 PSM전에 각 공변량에 대한 standardized difference를 추정한다.

```
> # Standard difference of original dataset
> treated = original$treatment == 1
> cov = original[,2:4] # Covariates, continuous data
> cov [1:10,]

      AST      ALT      HbA1c
1  360.7339  435.8191 10.014484
2  324.7640  282.6584  9.167625
3  476.4515  344.7153  8.729882
4  484.5870  373.1589  8.300526
5  351.8709  328.7666  9.040898
6  529.5012  529.7834  6.458388
```

```

7 392.6652 342.4583 7.973046
8 299.9401 522.1397 9.316708
9 288.0765 364.0826 8.105721
10 431.9856 586.1401 7.471799
>
> std.diff = apply(cov, 2, function(x) 100*(mean(x[treated]) - mean(x[!treated]))/(.5*(var(x[treated]) + var(x[!treated]))))
> std.diff
      AST      ALT      HbA1c
-0.008229608 -0.035499310  4.410494451
>
> p.tx0 = tbl.aspirin[2,1]/(tbl.aspirin[1,1]+tbl.aspirin[2,1])
> p.tx1 = tbl.aspirin[2,2]/(tbl.aspirin[1,2]+tbl.aspirin[2,2])
> std.diff.p = (p.tx1 - p.tx0)/(sqrt((p.tx1*(1-p.tx1)+p.tx0*(1-p.tx0))/2))
> std.diff.p
[1] 0.08815193

```

계산된 standardized difference를 정리하면 Table 2와 같다.

**Table 2.** Statistical Analysis Results and Standard Differences before Propensity Score Matching

Covariates					
Variables		Observed values		P value	Standardized difference
		Control group	Treated group		
AST		399.5 ± 101.9	398.7 ± 96.9	0.872	-0.00823
ALT		409.1 ± 102.4	405.5 ± 99.1	0.477	-0.0355
HbA1c		9.01 ± 0.97	9.05 ± 0.97	0.388	4.410494
Aspirin	Not used	1561 (84.9)	278 (15.1)	0.083	0.088152
	Used	957 (82.4)	204 (17.6)		
Result					
Bleeding	Not occurred	1317 (52.3)	227 (47.1)	0.036*	
	Occurred	1201 (47.7)	255 (52.9)		

일반적으로 standardized difference가 0.1 이하이면 공변량의 차이를 무시할 수 있다. 이 예제에서는 HbA1c가 상당한 차이를 나타내고 있으며 다른 공변량들은 의미있는 차이를 보이지 않고 있다. 이 결과는 PSM 시행 이후와 비교하여 보도록 한다.

PSM의 시행은 AST, ALT, HbA1c, Aspirin을 대상으로 시행하였으며, optimal full matching method로 시행하였다.

```

> ##### Propensity Score Matching: Optimal 1:1 method #####
> library (MatchIt)
필요한 패키지를 로딩중입니다: MASS
>
> original1 = as.data.frame(na.omit(original))
> opt.psm.out = matchit(treatment ~ AST + ALT + HbA1c + Aspirin, method = "optimal", data = original1)
Loading required package: optmatch
Loading required package: survival
The optmatch package has an academic license. Enter relaxinfo() for more information.
Warning message:
In fullmatch(d, min.controls = ratio, max.controls = ratio, omit.fraction = (n0 - :

```

Without 'data' argument the order of the match is not guaranteed to be the same as your original data.

```
> summary (opt.psm.out)
```

Call:

```
matchit(formula = treatment ~ AST + ALT + +HbA1c + Aspirin, data = original1,
method = "optimal")
```

Summary of balance for all data:

	Means Treated	Means Control	SD Control	Mean Diff	eQQ Med	eQQ Mean	eQQ Max
distance	0.1619	0.1604	0.0140	0.0015	0.0013	0.0015	0.0080
AST	398.7157	399.5292	101.9251	-0.8135	3.6490	5.2687	69.4231
ALT	405.4929	409.0967	102.3945	-3.6038	3.8694	5.2492	75.8239
HbA1c	9.0482	9.0063	0.9745	0.0419	0.0554	0.0573	0.3797
Aspirin	0.4232	0.3801	0.4855	0.0432	0.0000	0.0436	1.0000

Summary of balance for matched data:

	Means Treated	Means Control	SD Control	Mean Diff	eQQ Med	eQQ Mean	eQQ Max
distance	0.1619	0.1619	0.0146	0.0000	0.0002	0.0002	0.0010
AST	398.7157	397.9533	101.7633	0.7624	5.6879	6.7025	88.6984
ALT	405.4929	405.5981	100.4716	-0.1052	4.4827	5.3914	33.9137
HbA1c	9.0482	9.0274	0.9479	0.0208	0.0364	0.0532	0.9915
Aspirin	0.4232	0.4274	0.4952	-0.0041	0.0000	0.0041	1.0000

Percent Balance Improvement:

	Mean Diff.	eQQ Med	eQQ Mean	eQQ Max
distance	98.5859	85.2916	85.1643	88.1270
AST	6.2836	-55.8756	-27.2127	-27.7649
ALT	97.0804	-15.8492	-2.7090	55.2731
HbA1c	50.2591	34.2689	7.2095	-161.1719
Aspirin	90.3889	0.0000	90.4762	0.0000

Sample sizes:

	Control	Treated
All	2518	482
Matched	482	482
Unmatched	2036	0
Discarded	0	0

결과는 mean difference의 변화를 보여주며, 마지막에 짝을 이룬 자료가 482개로 PSM에 의해 생성된 자료의 표본수는 964개 이다. PSM에 의해 짝짓기가 이루어진 값들을 자료로 저장하고 그래프를 통하여 변화를 확인하는 과정도 필요하다.

```
> opt.data = match.data(opt.psm.out)
```

```
> head(opt.data)
```

	X	AST	ALT	HbA1c	Aspirin	treatment	bleeding	distance	weights	subclass
4	4	484.5870	373.1589	8.300526	1	1	0	0.1718407	1	321
8	8	299.9401	522.1397	9.316708	0	1	0	0.1487687	1	383
9	9	288.0765	364.0826	8.105721	1	1	0	0.1725665	1	399

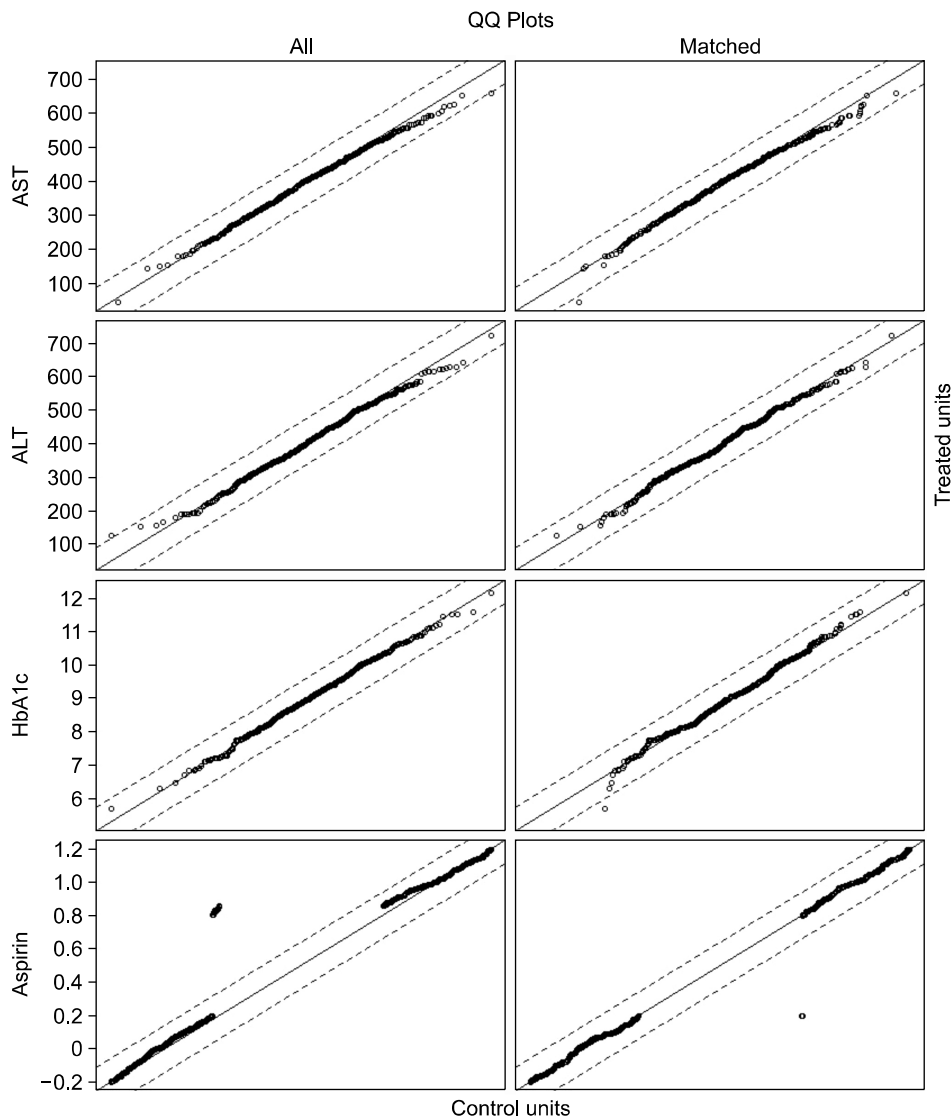
```

10 10 431.9856 586.1401 7.471799 1 1 0 0.1574736 1 1
15 15 447.1536 574.9611 9.320016 0 1 0 0.1455818 1 6
17 17 280.3349 375.2379 8.798687 0 1 0 0.1521404 1 54
>
> plot(opt.psm.out)
Hit <Return> to see next plot:
> plot(opt.psm.out, type = "jitter")
[1] "To identify the units, use first mouse button; to stop, use second."
integer(0)
> plot(opt.psm.out, type = "hist")
>

```

Opt.psm.out은 기존의 7개 변수에 distance, weight, subclass라는 세 개의 변수가 더 추가되었다. Distance는 각 propensity score의 차이에 의한 거리, weight는 weight option을 사용했을 때 사용된 값(이 자료는 weight option을 사용하지 않음), subclass는 짝짓기가 이루어진 자료들의 번호이다. 짝짓기가 된 자료는 같은 subclass번호를 부여 받는다.

Plot 명령들에 의해 생성된 그래프는 다음과 같다(Figs. 1-3).



**Fig. 1.** The graph created using the R command 'plot'. Using this QQ plots, visual check analysis for post matched data compared to original data is possible according to each covariate. In case of covariate 'aspirin', all outliers existed in original data were disappeared after propensity score matching except one.

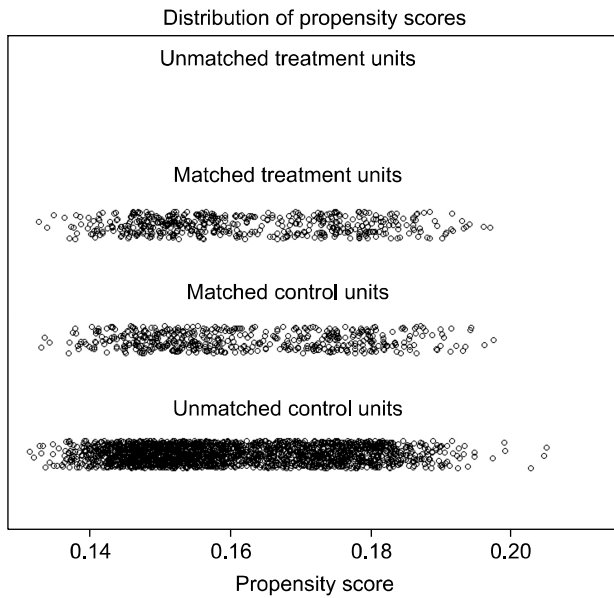


Fig. 2. The graph created using option 'Type=jitter'. The propensity scores of matched treatment and matched control units are considerably overlapped.

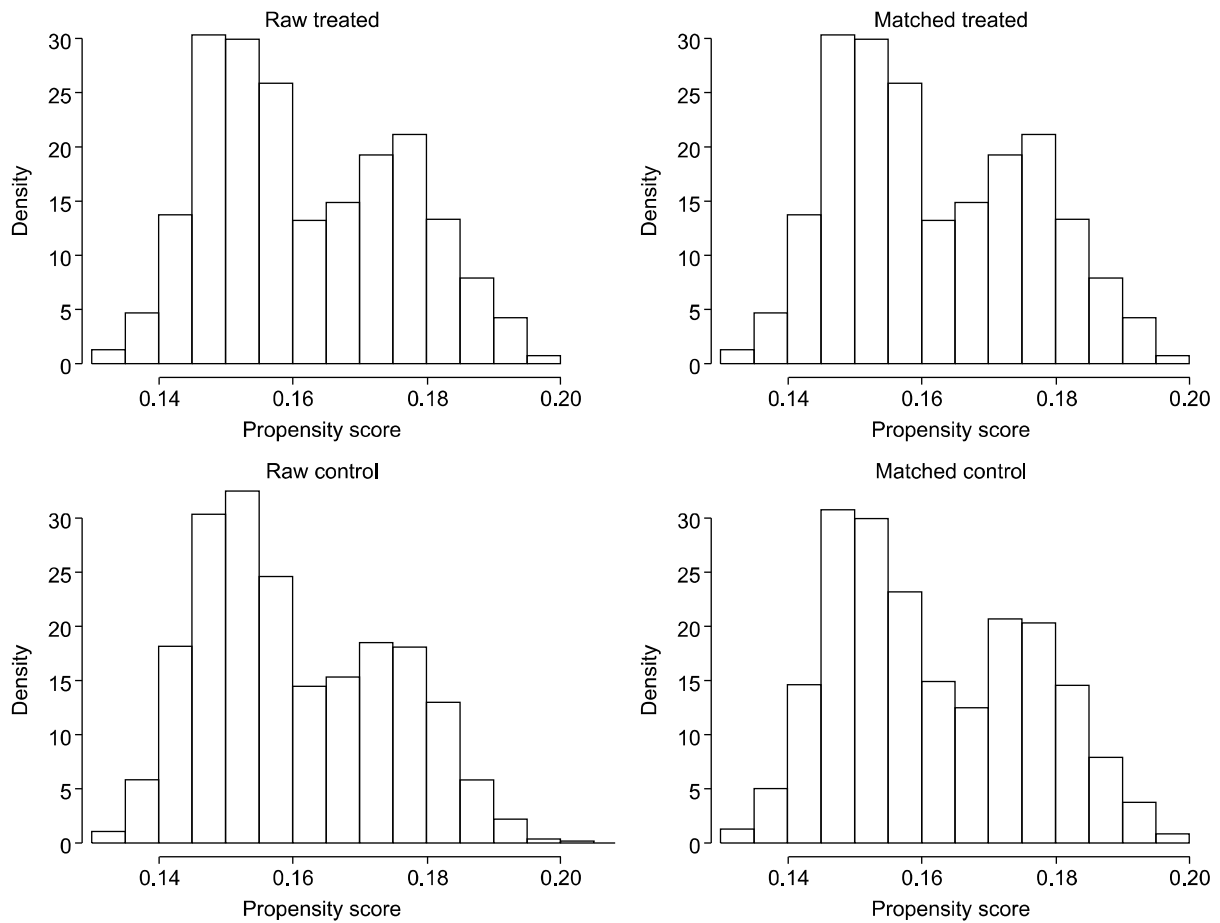


Fig. 3. Propensity score distributions of before and after propensity score matching. When propensity score matching process is adequate, the distributions of matched treated and matched control groups are actually similar.

## Post matching analysis

Matching된 자료는 일반적인 이분변수를 분석하는 방법을 이용한다. 이 예제에서는 편의상 짝짓기 이전의 자료와 같은 방법으로 통계처리를 하였다. 통계처리를 한 뒤에는 결과변수의 차이를 확인할 수 있으며, standardized difference를 계산하여 기존 자료에서 공변량의 불균형이 얼마나 해소 되었는지 확인할 수 있다.

```
> ###Post matching analysis###
> #normality test
>
> shapiro.test (resid(lm(AST ~ treatment, data = opt.data)))

        Shapiro-Wilk normality test

data: resid(lm(AST ~ treatment, data = opt.data))
W = 0.99882, p-value = 0.792

> shapiro.test (resid(lm(ALT ~ treatment, data = opt.data)))

        Shapiro-Wilk normality test

data: resid(lm(ALT ~ treatment, data = opt.data))
W = 0.99863, p-value = 0.6704

> shapiro.test (resid(lm(HbA1c ~ treatment, data = opt.data)))

        Shapiro-Wilk normality test

data: resid(lm(HbA1c ~ treatment, data = opt.data))
W = 0.99759, p-value = 0.1709

>
> #equal variance test
> var.test (AST ~ treatment, data = opt.data)

        F test to compare two variances

data: AST by treatment
F = 1.1093, num df = 481, denom df = 481, p-value = 0.2558
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.927515 1.326633
sample estimates:
ratio of variances
 1.109267

> var.test (ALT ~ treatment, data = opt.data)
```

F test to compare two variances

data: ALT by treatment

F = 1.0027, num df = 481, denom df = 481, p-value = 0.9764

alternative hypothesis: true ratio of variances is not equal to 1

95 percent confidence interval:

0.8384097 1.1991851

sample estimates:

ratio of variances

1.002701

```
> var.test (HbA1c ~ treatment, data = opt.data)
```

F test to compare two variances

data: HbA1c by treatment

F = 0.93467, num df = 481, denom df = 481, p-value = 0.459

alternative hypothesis: true ratio of variances is not equal to 1

95 percent confidence interval:

0.7815236 1.1178205

sample estimates:

ratio of variances

0.9346674

>

```
> # comparing matched data
```

```
> t.test (AST ~ treatment, var.equal = T, data = opt.data)
```

Two Sample t-test

data: AST by treatment

t = -0.11777, df = 962, p-value = 0.9063

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-13.32855 11.81938

sample estimates:

mean in group 0 mean in group 1

397.9611 398.7157

```
> t.test (ALT ~ treatment, var.equal = T, data = opt.data)
```

Two Sample t-test

data: ALT by treatment



```

t = 0.33367, df = 962, p-value = 0.7387
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -10.40327 14.66569
sample estimates:
mean in group 0 mean in group 1
 407.6241 405.4929

> t.test (HbA1c ~ treatment, var.equal = T, data = opt.data)

      Two Sample t-test

data: HbA1c by treatment
t = -1.8946, df = 962, p-value = 0.05845
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.2380002 0.0041857
sample estimates:
mean in group 0 mean in group 1
 8.931265 9.048172

>
> library (MASS)
> opt.tbl.aspirin = table (opt.data$Aspirin, opt.data$treatment)
> opt.tbl.aspirin

   0    1
0 277 278
1 205 204
> chisq.test(opt.tbl.aspirin)

      Pearson's Chi-squared test with Yates' continuity correction

data: opt.tbl.aspirin
X-squared = 0, df = 1, p-value = 1

>
> opt.tbl.bleeding = table (opt.data$bleeding, opt.data$treatment)
> opt.tbl.bleeding

   0    1
0   0 227
1 482 255
>
> prop.test (table (opt.data$treatment, opt.data$bleeding), correct = F)

```

2-sample test for equality of proportions without continuity correction

```
data: table(opt.data$treatment, opt.data$bleeding)
X-squared = 296.92, df = 1, p-value < 2.2e-16
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.5155159 -0.4263928
sample estimates:
 prop 1 prop 2
0.0000000 0.4709544
>
> opt.treated = opt.data$treatment == 1
> opt.cov = opt.data[,2:4]
> opt.cov [1:10,]
      AST      ALT  HbA1c
4 484.5870 373.1589 8.300526
8 299.9401 522.1397 9.316708
9 288.0765 364.0826 8.105721
10 431.9856 586.1401 7.471799
15 447.1536 574.9611 9.320016
17 280.3349 375.2379 8.798687
26 461.5947 623.6744 8.424403
30 473.8413 190.3081 9.279516
36 321.5720 363.5726 10.179795
38 518.5353 391.3981 6.892140
>
> std.opt.diff = apply(opt.cov, 2, function(x) 100*(mean(x[opt.treated]) -
mean(x[!opt.treated]))/(.5*(var(x[opt.treated])+var(x[!opt.treated]))))
> std.opt.diff
      AST      ALT      HbA1c
0.007725369 -0.001056726 2.254509818
>
> p.opt.tx0 = opt.tbl.aspirin[2,1]/(opt.tbl.aspirin[1,1]+opt.tbl.aspirin[2,1])
> p.opt.tx1 = opt.tbl.aspirin[2,2]/(opt.tbl.aspirin[1,2]+opt.tbl.aspirin[2,2])
> std.diff.p.opt = (p.opt.tx1 - p.opt.tx0)/(sqrt((p.opt.tx1*(1-p.opt.tx1)+p.opt.tx0*(1-p.opt.tx0))/2))
> std.diff.p.opt
[1] -0.008392996
```

## PSM 전후의 결과 비교

짜짓기 전후의 결과 비교는 Table 3과 같다. PSM을 시행한 후에는 기존의 자료와 짜짓기를 한 자료의 standardized difference를 반드시 표기해 주어야 한다.

**Table 3.** Statistical Analysis Results and Standardized Differences before and after Propensity Score Matching

Variables		Before propensity score matching				After propensity score matching			
		Observed values		P value	Standardized difference	Observed values		P value	Standardized difference
Covariates		Control group	Treated group			Control group	Treated group		
AST		399.5 ± 101.9	398.7 ± 96.9	0.872	−0.00823	398 ± 101.8	398.7 ± 96.9	0.906	0.007725
ALT		409.1 ± 102.4	405.5 ± 99.1	0.477	−0.0355	405.6 ± 100.5	405.5 ± 99.1	0.739	−0.00106
HbA1c		9.01 ± 0.97	9.05 ± 0.97	0.388	4.410494	9.03 ± 0.95	9.05 ± 0.97	0.058	2.25451
Aspirin	Not used	1561 (84.9%)	278 (15.1%)	0.083	0.088152	276 (49.8%)	278 (50.2%)	1.000	−0.00839
	Used	957 (82.4%)	204 (17.6%)			206 (50.2%)	204 (49.8%)		
Result									
Bleeding	Not occurred	1317 (52.3%)	227 (47.1%)	0.036		1 (0.2%)	227 (47.1%)	<0.001	
	Occurred	1201 (47.7%)	255 (52.9%)			481 (99.8%)	255 (52.9%)		

Optimal full matching을 시행하였음에도 불구하고 HbA1c의 경우에는 standardized difference가 2를 넘는 값이 계산되었다. 이런 경우에는 optimal full matching 뿐 아니라 다른 방법들도 반드시 고려해 봐야 하며, 1 : 1 뿐만 아니라 1 : 2, 1 : 3 등 비율을 바꾸어 가면서 짝짓기를 해 봐야 한다. 또한 propensity score 계산에 사용된 공변량의 종류도 재고할 필요가 있다. 본 예제에 서는 optimal full matching 방법만 다루었지만, MatchIt은 기본적으로 exact matching, subclassification, nearest neighbor, genetic matching, coarsened exact matching 등 여러 가지 짝짓기 방법을 제공하며 가중치를 부여하거나, 추정된 propensity score들의 짝을 이루는 간격 등을 지정할 수 있다[3].

## References

1. Ho D, Imai K, King G, Stuart E. Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference. *Political Analysis* 2007; 15: 199-236.
2. Hansen BB, Fredricson M, Buckner J, Errickson J, Solenberger P. Package 'optmatch'. Available from <https://cran.r-project.org/web/packages/optmatch/optmatch.pdf>
3. Randolph JJ, Falbe K, Manuel AK, Balloun JL. A Step-by-Step Guide to Propensity Score Matching in R. *Pract Assess Res Eval* 2014; 19: No 18. Available from <http://pareonline.net/getvn.asp?v=19&n=18>