

Propensity Score Matching 방법을 이용한 간호중재 효과 평가

이숙정¹⁾ · 유지수²⁾ · 신미경³⁾ · 박창기⁴⁾ · 이현철⁵⁾ · 최은진⁶⁾

서 론

연구의 필요성

실험연구의 목적은 중재를 준 집단(실험군)과 중재를 주지 않은(대조군) 두 집단간에 중재에 의한 명확한 차이를 확인하는데 있으며, 중재 전 두 집단의 동질성 여부를 확인하는 것은 중재의 효과를 판단하기 앞서 반드시 거쳐야 할 조건이다. 선택편중(selection bias)이란 중재집단에 참여를 결정하는 요인과 중재의 효과가 서로 상관관계를 가지는 경우를 가리킨다. 무작위 할당 방법이 중재 전 두 집단의 구성을 위해 사용되지 않았거나 불완전한 무작위할당에 의한 선택편중이 발생했을 때, 중재집단의 변화된 결과가 오직 중재에 의한 결과로 해석되는 오류를 범하기 쉽다. 이러한 실험연구에서의 선택편중을 효과적으로 제거할 수 있는 방법으로는 무작위(random) 표본 추출 방법, 유사 실험연구(quasi-experimental designs), 그리고 짝짓기(matching) 방법 등이 이용되고 있다.

선택편중을 제거하는 방법으로 무작위 추출 방법은 확률 이론에 의하여 두 집단(실험군과 대조군)을 분류하기 때문에 두 집단간에 관찰 가능한 특성들 이외에 관찰 불가능한 특성들까지도 평균적으로 동일한 기대치를 갖게 함으로서(Lee, 2003), 중재에 의한 진정한 결과의 효과를 추정할 수 있는 평가 방법으로 여겨지고 있다(Bloom et al., 2002). 하지만 환자나 지역사회 주민을 대상으로 하는 간호연구의 경우 현실적으로 환자를 대상으로 한 무작위 할당은 어려울 뿐 아니라

비용 효과적이지 못하며, 무작위 할당을 하는 경우라 하더라도 대상자의 부족으로 실험군과 유사한 특성의 대조군 집단을 구성하는 데 어려움이 있을 때가 많다. 국내 간호 중재 연구에 있어서 무작위 할당을 실시한 연구는 거의 없으며, 대한간호학회지에 1998년 이후 당뇨병환자를 대상으로 실시한 중재 연구 6개 중에서 5개(83%)가 비동등성 대조군 실험설계로 실시되었다.

비동등성 대조군 실험설계는 유사 실험설계 중 하나로, 중재 전 두 집단을 무작위 할당하지 못했기 때문에 사전조사를 통하여 두 집단의 특성 중 종속변수에 영향하는 변수들의 차이를 탐색하여 통제하는 방법이다. 즉 실험군과 대조군이 외생변수에 있어 동질하지 않은 경우 연구결과의 분석 시 외생변수를 통제하게 된다. 하지만 비동등성 대조군 실험 설계는 두 집단간에 선택 편중이 발생할 수 있는 가능성 때문에 중재의 인과적인 관계를 검증하는데 있어 연구중재의 결과과소 또는 과대하게 평가될 수 있는 단점을 가지게 된다.

비동등성 대조군 실험 설계 외에, 무작위 할당 하지 않은 연구에서 최대한 선택편중을 줄일 수 있는 방법으로 짝짓기(matching) 방법이 종종 사용되고 있는데(D'Agostino, 1998), 이 방법은 실험군과 대조군의 실험 참가에 줄 수 있는 영향을 통제하기 위하여 외생변수인 연령, 성별 등 결과에 영향을 줄 수 있는 변수에 의해 실험군에 맞는 대조군을 짝짓는 방법이다. 하지만 할당할 변수의 수가 많아지면 대조군 선정에 어려움이 생기게 되고, 따라서 통제가 필요한 많은 변수를 통제할 수 없음이 한계점으로 지적되고 있다(Dehejia &

주요어 : 경향점수, 중재 평가

1) 적십자 간호대학 전임 강사, 2) 연세대학교 간호대학 간호학과 교수, 3) University of Illinois at Chicago 연구원

4) University of Illinois at Chicago 연구원, 5) 연세대학교 의과대학 의학과 교수, 6) 연세대학교 간호대학 간호정책 연구소 연구원

투고일: 2006년 6월 14일 심사완료일: 2007년 4월 11일

Wahba, 1998). 변수의 수와 상관없이 선택편종을 줄이는 방법으로 1980년대 초반에 소개되어 최근에 사회과학과 의학 및 역학 분야에서 다양하게 활용되고 있는 방법이 Propensity Score Matching(PSM) 방법이다(Rosenbaum & Rubin, 1983).

Propensity Score라는 용어는 Rosenbaum과 Rubin(1983)이 그들의 논문에서 관찰 자료의 인과관계를 보다 명확히 계산하기 위한 방법으로 처음 소개되었다. PSM 개념은 counterfactual(대응) framework을 기초로 한다(Rubin, 1978). 고전적인 실험 연구는 실험군의 평균 결과 값을 대조군의 평균 결과 값과 비교하는 것이다. 이 경우 대조군은 실험군의 counterfactual이다. 실험군의 counterfactual이란 실험군이 실험에 참여하지 않았을 경우를 가정한 것이다. 연구자들의 실제적인 관심은 비교군의 상황에서 나타나는 비교군 결과 값이 아니고, 실험군이 실험에 참여하지 않았을 때 나타날 결과 값이다. 그런데 실험군이 실험에 참여하지 않을 상황은 관찰할 수 없기 때문에 대응으로 대조군을 두게 된다. 전통적인 실험 연구는 실험군과 대조군이 선택 편종이 없고, 두 그룹이 같다는 가정을 한다. 그런데 Heckman과 Smith(1995)는 선택 편종 때문에 실험군이 실험하지 않았을 상황에서 나타나는 평균 결과값은 proxy로 사용된 비교군의 평균 결과 값과 정확히 같지 않음을 보고하였다. 이들은 randomization을 통해 selection bias가 생기지 않으려면, 누가 프로그램에 참여하는가, 어떤 참여자가 그 프로그램을 수용하는가, 누가 프로그램에 계속 남아 있는가 등의 내용이 가정에서 포함되어야 한다고 명시하였다.

이러한 framework 내에서 실험군과 대조군과의 차이의 평균값을 구할 때 관찰가능한 대조군의 결과는 실험군의 counterfactual의 결과로 간주된다. 이 때 선택 편종의 문제가 있다면 전통적인 대조군의 결과값은 실험군의 counterfactual의 결과값으로 대치하기 어렵다. 실제로 많은 변수가 대조군의 결과값과 실험군의 counterfactual의 결과값의 차이에 영향을 미치게 된다. Rosenbaum과 Rubin(1983)은 실제 상황에서 관찰되는 공변량이 같다면, 실험군에게 만약 중재하지 않았다면 나타나는 잠재적 결과나 대조군에서 만약 중재했다면 나타나는 잠재적 결과는 같다고 보고한다. 위의 기본가정이 만족되지 못할 때는 연구자는 공변량이 같지 않음으로 인해 생기는 편종을 제거해야 한다. PSM은 실험군과 대조군의 공변량의 영향력을 반영하여 이들이 중재에 참여할 조건부 확률값인 Propensity Score에 따라 최대한 같은 공변량을 가진 두 그룹으로 짝짓기 함으로써 공변량이 같아야 한다는 가정을 만족시키려는 방법이다. PSM방법은 모든 대상자의 공변량을 통제된 상황에서 이들이 중재에 참여할 조건부 확률값인 Propensity Score(PS)점수를 구하고, 이 PS를 이용하여 유사한 대상자들끼리 짝짓기 하는 방법으로서 선택 편종을 줄인 상

태에서 실험군과 대조군을 재구성함으로써 이에 따라 중재의 효과를 보다 정확하게 비교할 수 있는 방법이다(Rosenbaum & Rubin, 1983).

간호연구는 대부분 끊임없이 환경과 상호작용하고 변화하는 인간을 대상으로 하기 때문에 그 모든 상황을 통제하는 무작위 할당 방법을 효과적으로 진행하는데 어려움을 겪고 있다. 본 연구는 간호학 분야에서 선택편종의 제한점을 극복할 수 있는 한 방법으로 PSM을 이용한 분석을 소개함으로써 간호연구에서 보다 정확한 평가를 가능하게 할 것이다.

연구 목적

본 연구는 두 그룹간의 비교 연구에서 선택편종을 최대한 줄이면서 중재의 효과를 비교할 수 있는 PSM 방법을 소개하는 것이 목적이다. 이를 위해서 당뇨병환자를 대상으로 생활습관 개선 프로그램을 실시한 후 그 효과를 평가한 선행 연구의 결과분석(Yoo et al., 2005)을 사례로, 선택편종을 줄일 수 있는 방법인 Propensity Score Matching을 이용한 분석 결과를 소개하고, 그 결과를 비교 하고자 한다.

용어의 정의

● 공변량

실험에서 치료요인이 아니면서 반응변수에 영향을 주는 요인을 말하며 confounding factor 또는 covariate라고 한다.

● Propensity Scores

Propensity Score(PS)는 대상자들의 관찰 가능한 특성들이 주어졌을 때 이들이 실험군이 될 조건부 확률을 나타내는 값으로, 대상자의 서로 다른 공변량의 차이를 근거하여 하나의 점수(0에서 1)로 계산된다. 이는 Rosenbaum과 Rubin에 의해 처음으로 소개된 이후(Rosenbaum & Rubin, 1983), 공변량 x_i 가 주어졌을 때 실험($Z_i=1$)을 받을 조건부 확률값으로 다음과 같이 계산될 수 있다; $\text{Propensity Score} = e(x_i) = \text{pr}(Z_i=1 | X_i=x_i)$.

● Propensity Score Matching(PSM)

실험군과 대조군 선정시 나타나는 선택 편종을 줄이기 위해 Propensity Score를 기준으로 짝짓기 하는 방법이다. 실험군과 대조군이 각각 실험군이 될 확률값(PS)이 추정되면 실험군을 구성하는 모든 관측치들마다 가장 가까운 PS 값을 가지는 대조군의 관측치로 짝짓기 하게 된다.

먼저 X명의 실험군과 Y명의 대조군으로 중재실험이 이루어졌다고 생각해보자($X < Y$). 기존의 일대일 짝짓기 방법으로는

X명의 실험군과 X명의 짝지어진 대조군이 요구되나, 만약 통제 변수들의 수가 많을 경우에는 짝짓기 해야 하는 조합이 상당히 많아지게 되고, 따라서 실험군과 같은 특성을 가진 대조군을 짝짓기 할 수 없는 경우도 발생할 확률이 커지기 때문에 결과적으로 기존의 X명에 비해 매우 작은(<X) 짝짓기 집단이 생성되거나 짝짓기가 잘 되지 않을 가능성이 생긴다. PSM은 이러한 문제를 해결하기 위해 대상자 (실험군 과 대조군)의 관찰 가능한 특성들(공변량)을 한 가지 점수(실험에 참여할 조건부 확률 값)인 Propensity Score를 이용하여 짝짓기 하는 방법으로, 비슷한 PS 점수에 의해 짝짓기로 구성된 두 집단을 통해 최대한 선택편향을 줄이고, 무작위 할당을 실시한 연구와 같이 중재에 대한 보다 정확한 효과를 계산할 수 있다.

Propensity Score Matching 방법을 정리하면 다음과 같다.

- 각 대상자(실험군과 대조군)의 PS점수를 구한다. 본 연구에서는 성별, 직업유무, 당뇨 약물 복용여부, 연령, 당뇨 유병기간, 사전 당화혈색소 수치, 체질량지수를 이용해서 logistic regression을 통해 PS를 계산하였다.
- PS 점수에 의해 대조군과 실험군을 짝짓는다. 각 환자의 실험군이 될 확률값(PS)이 하나의 수치로 추정되면 실험군을 구성하는 대상자의 PS 점수와 가장 가까운 PS점수를 가지는 대조군의 대상자와 짝짓기하게 된다. 본 연구에서는 실험군의 PS와 가장 가까운 대조군의 PS를 짝짓는 방법으로 최소 거리법(minimum distance method)을 사용하였으며, 이 최소 거리법을 통해 1명의 실험군 대상자가 1명의 대조군 대상자와 짝짓는 방법(1:1 짝짓기)을 사용하였다.
- 짝짓기 한 결과를 이용하여 두 그룹간의 결과의 차이를 비교분석한다.

연구 방법

연구설계

본 연구는 제 2형 당뇨병 환자에게 간호 중재를 실시한 연구 결과를 바탕으로 기존의 방법과 새로운 방법인 PSM 방법을 이용하여 비교 분석한 후향적 비교 연구(retrospective, comparative study)이다. PS를 통해 짝짓는 방법으로 일대일 짝짓기한 결과를 비교하였다.

연구과정

연구 결과의 이해를 위해 사례가 된 중재연구를 설명하면 다음과 같다. 당뇨병자를 위한 중재 연구인 포괄적인 생활개선 프로그램은 2003년 10월부터 2005년 9월까지 일 종합병원

에 통원치료를 받고 있는 당뇨병자를 위한 모집단으로 하여 무작위 할당으로 초기 계획되었던 연구이다(Yoo et al., 2005). 연구 과정에서 초기 실험군 30명과 대조군 30명은 무작위 할당 되었으나, 실험이 진행되는 과정에서 대조군은 그대로 유지하고 실험군은 더 많은 환자를 참여시키게 되었으며, 그 과정은 임의 표출 방법을 사용하게 되어서 결국 비동등성 대조군 그룹이 형성되게 되었다. 결과적으로 연구과정에서 탈락자 및 누락된 데이터를 가진 대상자를 제외하고 실험군 총 45명, 대조군 22명의 결과가 분석에 사용되었다. 프로그램은 제 2형 당뇨병 환자들에게 필요한 운동 및 식이 교육과 상담 및 토론으로 총 12주간 이루어졌으며(Yoo et al., 2005), 그 후 4주간은 스트레스 관리를 위한 점진적 근육이완 운동 및 심신의 안정을 위한 상담 등이 포함되어 총 16주 간의 중재가 이루어 졌다. 실험군에게는 4개월간 주 1회 1시간의 교육과 그 후 1년까지 월 1회의 추후 상담을 실시했고, 대조군에게는 당뇨의 일상적인 관리와 실험 초 1회의 영양교육을 실시했다. 실험군과 대조군의 동질성 분석을 위해 성별, 연령, 직업유무, 당뇨 약물 복용 유무, 당뇨 유병 기간, 실험 전 당화혈색소 수치, 실험 전 BMI 수치가 비교되었고 프로그램 12개월 후의 당화혈색소의 수치가 중재결과를 확인하기 위한 종속변수로 분석되었다.

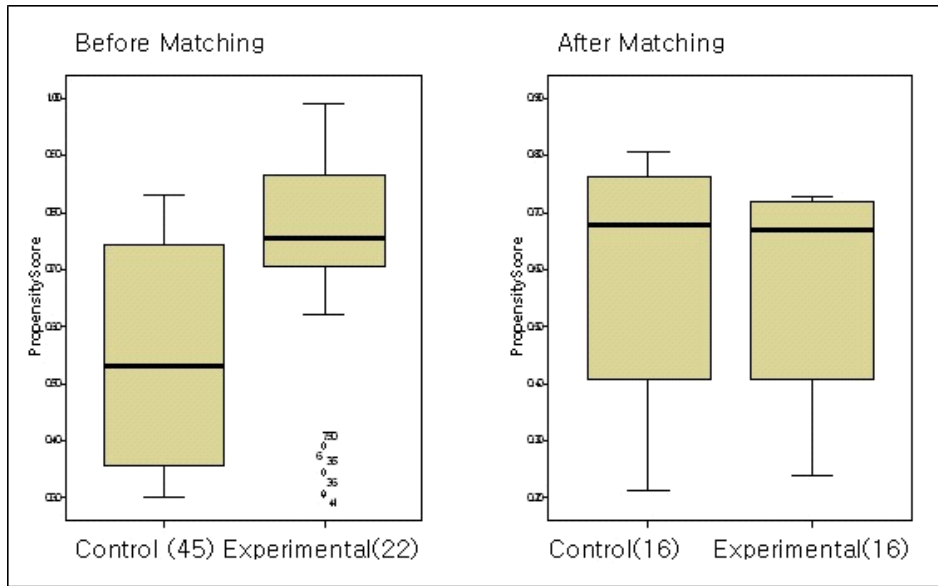
자료분석방법

기존의 연구에서 실험군(45명)과 대조군(22명)의 일반적 사항과 당화혈색소의 차이는 백분율, t-test, χ^2 -test로 분석되었다. PSM을 위해 기존연구에 참여했던 모든 실험군과 대조군의 Propensity Score는 Stata 8.0을 이용하여 계산되었고, 이 Propensity Score에 의해 일대일 짝짓기 한 실험군(16명)과 대조군(16명)의 당화혈색소의 차이는 다시 t-test, χ^2 -test로 비교되었다. 이 결과를 기존 data에서 공변량을 통제한 후에 나타난 결과와 비교하기 위하여 짝짓지 않은 12개월의 당화혈색소 결과를 종속 변수로 다중 회귀분석이 실시되었다.

연구 결과

짝짓기 전후 Propensity Score의 분포

성별, 직업유무, 당뇨 약물 복용여부, 연령, 당뇨 유병기간, 사전 당화혈색소 수치와 체질량지수를 공변량으로 실험군에 들어갈 확률을 Propensity Score로 구한 후 짝짓기 전과 후의 대상자들의 PS 분포를 그림1에 제시하였다(Figure 1). 짝짓기 전에는 실험군의 PS 평균이 0.7에서 0.9 사이에 몰려 있으며 0.4 이하인 대상자가 몇 명 나타났고, 대조군은 PS 평균이



<Figure1> Comparison of propensity score

0.4에서 0.7로 넓게 퍼져 있었다. 이 두 그룹은 유사한 부분이 0.7에서 0.75 사이로 적은 부분 겹쳐져 있어, 두 그룹 간 유사한 특징을 가진 사람의 수가 적음을 알 수 있다. PS 점수에 의한 일대일 짝짓기 한 경우는 실험군과 대조군 각각 16명씩 짝짓기 되었으며, 두 그룹 간 PS평균이 약 0.7로 같고 분포는 두 그룹이 유사한 양상을 보였다.

짝짓기 전 후 각 변수의 차이 비교

짝짓기 전 실험군은 45명, 대조군은 22명이었다. 이 연구에서는 성별, 연령, 직업유무, 당뇨 약물 복용여부, 당뇨 유병기간, 사전 당화혈색소 수치, 체질량지수가 공변량으로 비교되었다. 이 중 직업의 유무, 당뇨약물복용 여부는 통계적으로 유의하게 실험군과 대조군 사이 차이를 보였다. 실험군

(31.1%)은 대조군(59.1%)에 비해 직업이 있는 대상자가 더 적었고, 당뇨약물 복용에 있어서는 실험군(57.8%)이 대조군(86.4%)보다 약물 복용을 적게 하였다($p<0.05$). 연령은 평균 연령이 대조군은 53.7세, 실험군은 52.9세로 유사했으며, 당뇨 유병기간은 실험군은 6.4년, 대조군은 8.2년으로 대조군 대상자가 더 길었으나 이들 두 변수의 통계적인 차이는 없었다. 사전 당화 혈색소는 실험군과 대조군 각각 8.2%와 8.3%로 두 군간 유사한 수치였으며, 체질량지수도 실험군과 대조군 각각 25.5kg/m^2 와 26.2kg/m^2 로 유사한 수치를 나타내었다. 실험12개월 후 당화 혈색소는 실험군이 7.5%, 대조군이 8.7%로 실험군이 대조군에 비해 1.2% 통계적으로 유의하게 낮은 수치를 보였다.

PS점수에 의해 일대일 짝짓기를 했을 때 실험군과 대조군은 각각 16명씩 짝짓기 되었다. 이 때 성별, 연령, 직업유무,

<Table 1> Comparison of differences of the variables between the two groups before and after propensity score matching

| | Before matching | | | | | | One to one matching | | | | | |
|-------------------------|-----------------|----------------|--------------|----------------|----------------------------|-------|---------------------|----------------|--------------|----------------|----------------------------|------|
| | Exp Mean | n=45 (%)/SD | Cont Mean | n=22 (%)/SD | Comparisons χ^2 /t | Diff | Exp Mean | n=16 (%)/SD | Cont Mean | n=16 (%)/SD | Comparisons χ^2 /t | Diff |
| Male:y(%) | 11 | (24.4) | 7 | (31.8) | 0.41 | -7.4 | 5 | (31.2) | 6 | (37.5) | 0.1 | -6.3 |
| Age | 53.7 | 7.96 | 52.9 | 8.51 | 0.38 | 0.8 | 55.9 | 8.93 | 55.4 | 6.33 | 0.16 | 0.5 |
| Occupation:y(%) | 14 | (31.1) | 13 | (59.1) | 4.81* | -28.0 | 7 | (43.8) | 8 | (50.0) | 0.13 | -6.2 |
| DM drug:y(%) | 26 | (57.8) | 19 | (86.4) | 5.47* | -28.6 | 15 | (93.8) | 14 | (87.5) | 0.37 | 6.3 |
| DM Duration | 6.4 | 6.24 | 8.2 | 6.16 | -1.05 | -1.8 | 8.9 | 6.41 | 8.8 | 6.15 | 0.03 | 0.26 |
| HbA1c(%) 0m | 8.2 | 1.41 | 8.3 | 1.34 | -0.37 | -0.1 | 8.4 | 1.25 | 8.7 | 1.27 | -0.72 | -0.3 |
| BMI(Kg/m ²) | 25.5 | 3.70 | 26.2 | 3.99 | -7.70 | -0.7 | 26.9 | 4.18 | 26.0 | 4.45 | 0.63 | 0.9 |
| HbA1c(%) 12m | 7.5 | 1.21 | 8.7 | 1.62 | -2.74* | -1.2 | 7.6 | 1.48 | 8.8 | 1.37 | -2.36* | -1.2 |

* $p<.05$, Comparisons on gender, occupation, and DM drug taking are difference in % between experimental and control group.

Comparisons on age, DM duration, HbA1c(0m), BMI and HbA1c(12m) are difference in mean between experimental and control group.

Exp:experimental group, Cont: control group, Diff: difference in % or mean between experimental and control group, y:yes

<Table 2> Predictors of HbA1c from the multiple regression

| Model | B | Std. Error | β | t | p | R ² | F | p |
|-------------|-------|------------|---------|-------|------|----------------|-------|-------|
| (Constant) | 7.03 | 2.06 | | 3.41 | 0.00 | | | |
| Group | -0.88 | 0.32 | -0.29 | -2.73 | 0.01 | | | |
| Sex | 0.54 | 0.42 | 0.17 | 1.28 | 0.21 | | | |
| Job | -0.28 | 0.42 | -0.10 | -0.67 | 0.51 | | | |
| DM drug | 0.13 | 0.39 | 0.04 | 0.33 | 0.74 | 0.477 | 6.378 | 0.001 |
| Age | -0.04 | 0.02 | -0.23 | -1.88 | 0.07 | | | |
| DM duration | 0.06 | 0.03 | 0.28 | 2.05 | 0.05 | | | |
| HbA1c(0m) | 0.46 | 0.11 | 0.44 | 4.04 | 0.00 | | | |
| BMI 0month | -0.03 | 0.04 | -0.08 | -0.71 | 0.48 | | | |

당뇨 약물 복용, 당뇨 유병기간은 각각 통계적으로 유의한 차이를 보이지 않았다<Table 1>. 이 때 12개월 후 당화혈색소의 차이는 실험군이 1.2% 통계적으로 유의하게 낮아서, 짹짹하기 전 기존연구 결과와 같은 결과를 나타내었다. 그러나 짹짹하기 전 실험군은 실험 전 당화혈색소 수치 8.2%에서 12개월 후 0.7% 감소하여 7.7%가 되었고, 대조군은 8.3%에서 0.4% 증가하여 8.7%가 되었으나, 일대일 짹짹하기 후에는 실험군이 8.4%에서 12개월 후 0.8%로 감소하여 7.6%가 되었고, 대조군은 8.7%에서 0.1% 증가하여 8.8%를 보였다<Table 1>.

기존 실험군(45명)과 대조군(22명)의 자료에서 위의 모든 공변량을 통제한 상태에서 12개월 후 당화혈색소를 종속변수로 다중회귀 분석 한 결과 실험군의 12개월 후 당화혈색소가 대조군에 비하여 통계적으로 유의하게 0.88% 낮게 나타났다.

논 의

본 연구는 PS 점수에 의한 일대일 짹짹하기 방법을 통해 선택 편증을 줄인 상태에서 보다 유사한 그룹을 형성한 후에 그 결과를 비교할 수 있는 PSM 방법을 소개하기 위해, 이미 기존에 실험 연구로서 그 효과를 검증한 연구를 사례로 재분석 하였다. 본 연구의 사례가 되었던 기존의 중재 프로그램은 본 연구자들이 직접 중재에 참여 했으며 기존연구의 제한점은 다음과 같다. 기존 중재연구는 실험 초기에는 무작위 할당의 방법으로 실험군과 대조군을 구성하였지만, 중재가 진행되면서 중도 탈락자로 인해 실험군과 대조군을 더 모집하게 되었고, 이 과정에서 대상자의 참여 의지가 반영됨으로써 두 그룹 간의 일반적 사항이 동질하지 않은 특성이 나타나게 되었다(직업의 유무와 당뇨 약물 복용 여부). 실험군 대상자들의 특징 중 직업이 없는 사람이 통계적으로 더 많았는데, 중재를 진행하면서 직업 유무는 단순히 직업 유무의 차이를 넘어서 당뇨 관리 이행에 매우 깊은 관계가 있음을 발견할 수 있었다. 직업이 있는 경우 외식이 잦고, 시간적 여유가 없어서 프로그램에 충실하게 참석할 수 없었고, 직업이 없는 경우는 식사와 운동 등의 기록 등 자가 모니터를 충실히 하는 경향이

있었다. 또한 실험군은 대조군에 비해 통계적으로 유의하게 약물 복용하는 사람 수가 적었는데, 이것은 당뇨의 정도가 대조군에 비해 경함을 간접적으로 의미하는 것이며, 실제로 만성적으로 오래된 대상자 보다 당뇨병을 진단 받고 얼마 되지 않았거나 아직 약물 사용을 시작하지 않은 대상자가 당뇨병을 오래 앓고 있던 대상자 보다 본 프로그램 실험군 으로서 지속적으로 출석하는 경향을 보였다.

직업 유무와 당뇨 약물 복용의 정도로 보아 시간적 여유가 있고, 당뇨병의 정도가 경한 대상자들이 실험군에 통계적으로 유의하게 많이 포함되어 있었기 때문에 이러한 특징이 당화혈색소의 변화에 긍정적으로 작용할 수 있다. 이들이 중재에 참여하지 않았다면 1년 후에 당화혈색소 수치가 현재 대조군에게 나타난 것과 같은 정도의 변화를 보일 것인가는 의문이다. 반대로 대조군이 현재 실험군에게 실시한 중재를 실시했을 때 현재 실험군에게 나타난 정도의 변화를 보일 것인가 또한 의문이다. 이러한 가정 속에서 시작된 짹짹하기 방법이 PSM 방법이다.

PSM 방법은 결과변수에 영향을 줄 수 있는 공변량을 통제한 상태에서 실험군에 포함될 조건부 확률값을 하나의 수치로 산출하여 짹짹하기 때문에 그 수치에 의해 유사한 특성의 대상자들로 짹짹할 수 있고, 두 군의 특징을 반영하는 변수들이 많을 때에도 충분히 짹짹하기 할 수 있는 장점이 있다.

중재 연구를 할 때 사전에 무작위 할당으로 대상자 선정을 계획했다면 초기 변수들에 대한 선택 편증은 별로 문제 거리가 되지 않는다. 그것은 무작위 할당이 대상자의 특징뿐만 아니라 그 잠재적인 특징 까지도 확률적으로 선택한다는 가정을 만족시키기 때문이다. 그러나 무작위 할당을 하지 않았거나, 어떠한 이유로 선택 편증이 발생한 경우에 있어서는 그 중재의 순수한 효과 검증을 위해서 선택편증을 줄일 수 있는 방법으로 분석을 시도해야 하겠다. 기존의 두 그룹 비교 방법인 t-test와 χ^2 -test의 경우는 이 선택 편증을 조절할 수 없다. 다중회귀 분석의 경우는 공변량을 통제할 수 있기 때문에 t-test와 χ^2 -test와는 달리 선택 편증을 어느 정도 조절할 수 있다. 그러나 다중회귀 분석은 공변량과 종속 변수간에 선형함

수 관계를 가정하고 있기 때문에 비모수 추정방법인 PSM에 비하여 그 결과의 신뢰도가 낮아질 가능성이 크다. 또한 실험군인지 대조군인지를 나타내는 dummy변수와 다른 특성 변수 간의 상관관계로 인하여 추정되는 중재효과가 부정확하게 추정될 수 있다. 본 연구에서도 짝짓기 전 다중회귀 분석 한 결과 실험군의 12개월 후 당화혈색소가 통계적으로 유의하게 0.88% 낮게 나타나 짝짓기 한 후 나타난 1.2%의 차이보다 낮게 나타났는데, 이 결과는 보편적으로 많이 사용되고 있는 다중 회귀 분석 방법이 선택편중의 불완전한 통제에 의하여 프로그램 효과가 실제로보다 과소평가 될 수 있음을 보여주고 있다. Rubin과 Thomas(2000)은 전체자료를 사용하는 다중회귀 분석 보다 짝지어진 그룹에 따르는 PSM 방법이 추정오류의 확률이 일반적으로 낮다고 보고하였다.

PS점수를 이용하여 실험군의 PS와 최단 거리에 있는 대조군 PS를 가진 사람끼리 일대일 짝짓기를 했을 때에 기존에 실험군과 대조군 사이에 통계적으로 차이가 있던 직업유무와 약물 복용여부가 유사한 16명으로 구성된 두 집단이 형성되었다. 기존연구와 PSM 방법 모두에서 12개월 후의 실험군과 대조군의 당화 혈색소의 차이는 모두 1.2% 였으나 그 내부적인 차이를 비교해 보면 기존의 연구에서는 실험군의 당화혈색소가 8.2%에서 0.7% 감소하였고, 대조군은 8.3%에서 0.4% 증가하였다. 반면 PS점수에 의해 일대일 짝짓기 했을 때는 실험군의 당화혈색소가 8.4%에서 0.8% 감소하였고, 대조군은 8.7%에서 0.1% 증가를 보였다. PSM 후의 결과를 다시 해석해 보면 실험군이 중재에 참여하지 않았다면 1년 후에 당화혈색소가 중재전에 비해 0.1% 증가했을 것이라고 해석할 수 있다. 짝짓기 전 대조군이 12개월 후 당화혈색소가 0.4% 증가하였는데, 이것은 대조군이 당뇨 조절이 잘 안되는 대상자가 많이 포함되어 있음을 보여 주는 것이고, 이것이 원인이 되어 실험군과의 당화혈색소의 차이가 심해지고 중재의 효과가 과장될 수 있음을 추론 할 수 있겠다.

이상에서 보았듯이 PSM을 이용하는 것이 기존의 방법에서 벗어나서 선택 편중을 줄일 수 있기 때문에 실제로 가까운 평가를 할 수 있는 방법이라 할 수 있다. 하지만 본 사례가 되었던 연구가 PSM 적용의 바람직한 예로 쓰이기에는 몇 가지 제한점을 갖는다. 본 사례 연구에서는 실험군이 45명이었고 대조군이 22명이기 때문에 실험군을 기준으로 짝짓기 하는데, 짝짓기할 대조군이 오히려 적기 때문에, 실험대상자 모두가 짝짓기 할 수 없다는 문제를 가지고 있다. 기존에 실험연구로 그 효과를 검증하고, 그 실험군 자료를 다시 다른 집단의 대상자들과 짝짓기 하여 비교함으로써 그 효과를 재검증한 연구(Baskett et al., 2005)의 경우 1896명의 관상동맥 bypass 수술시 balloon pump를 실시한 대상자와 balloon pump를 실시하지 않은 4332명의 대조군 대상자를 일대일 짝짓기 하여 1896명

씩 짝지어 졌고, 그 두 그룹에게 나타나는 결과를 비교하였다. 본 사례 연구가 만약 대조군 대상자가 많아서 대조군을 실험군 45명 모두와 PSM 방법으로 짝지울 수 있었다면, 이것은 16명씩 일대일 짝짓기 된 결과보다 중재의 효과를 보다 정확하게 측정 할 수 있는 결과가 도출 될 수 있을 것이다.

선택 편중을 줄일 수 있는 PSM 방법을 활용하는 것은 간호 연구에 있어서 보다 정확한 분석을 하게 한다. 현재 사회과학, 의학 및 역학 분야에서는 연구가 시작되고 있으나 아직 간호연구에는 활용된 연구는 거의 없는 실정이다. 간호연구에서는 그동안 무작위 할당하지 못해서 생긴 편중을 그대로 분석하거나 부분적인 통제 방법을 사용하여 많은 제한점들을 보였었는데, 기존의 자료나 연구 결과를 후향적 조사 연구로 대단위의 대조군 코호트대상으로 PSM 방법으로 그룹을 형성한 후, 그 효과를 분석해 보면 보다 순수한 효과를 검증 할 수 있겠다. PSM 방법의 출현은 실험연구 분석에서 시작된 것이 아니고 조사 연구에서부터 시작 된 것으로 관찰 집단과 비교 집단 두 그룹의 비교에 있어서 유사한 특징의 그룹으로 짝짓기 한 후에 비교하는 방법으로부터 활용되어 왔다. Ozminkowski 등(2006)은 류마티스 관절염을 앓고 있는 직장인들의 의료비용을 류마티스 관절염이 없는 대 집단 코호트를 대조군으로 PSM방법을 이용하여 짝짓기 해서 비교하였다. 그 동안 간호 행위는 여러 분야에서 다양하게 진행되었고, 지속적으로 발전하고 있음에도 불구하고, 기록의 미비나 연구 결과의 부족 등으로 간호행위에 쏟은 노력만큼 그 의미가 제대로 평가되지 못한 경우가 많이 있었다. PSM의 적용은 임상에서 어떠한 형태의 간호행위를 받은 자와 그렇지 않은 자의 비교 또는 어떠한 상태에 처해 있는 자와 그렇지 않은 자의 비교 등 다양한 형태의 비교 연구에서 보다 실제에 가까운 결과를 도출할 수 있기 때문에 실험연구를 직접 실행하지 않고도 유사 무작위 할당과 같은 효과를 가져 올 수 있어서 간호 연구에 매우 유용할 것이라 생각된다. 또한 보다 널리 알려지고 타당도와 신뢰도가 검증된 측정도구를 사용하여 국제적인 자료를 분석한다면 국가간의 편중을 통제한 상태에서 객관화된 분석 결과를 얻을 수 있을 것이다.

결론

본 연구는 선택 편중을 통제할 수 있는 PSM 방법을 이용하여 이미 실험 연구로서 그 효과를 검증한 당뇨병환자를 대상으로 실시한 중재 연구를 사례로 분석하고 그 결과를 기존의 결과와 비교하였다. PSM은 대상자 선정시 나타나는 선택 편의를 줄이기 위해 Propensity Score를 기준으로 matching하는 방법으로, Propensity Score(PS)는 대상자들의 관찰 가능한 특성들이 주어 졌을 때 이들이 실험군이 될 조건부 확률값이다.

분석은 기존의 실험군 45명과 대조군 22명의 중재결과를 PSM 방법을 사용하여 일대일 짝짓기 하고, 선정된 실험군과 대조군 각각 16명의 중재결과와 비교 분석하였다. 분석 결과 PSM 방법으로 짝짓기 한 후 12개월 후 당화혈색소 수치는 실험군과 대조군간 유의한 차이를 보임으로써, 기존연구가 가진 선택편중의 제한점을 보완한 후에도 간호중재가 당화혈색소를 저하시킨다는 인과적인 평가를 가능하게 하였다. 본 연구는 선택 편중을 최대한 줄여서 실제에 가까운 두 그룹을 비교하는 PSM 방법의 활용이 간호 연구에 있어서 무작위 할당의 어려움을 해결할 수 있는 방법임을 제시하였다.

두 그룹 비교 결과의 신뢰도 재고를 위해 다음을 제언한다. 관찰 연구에 있어서 다양한 집단으로 구성된 대조군 집단을 형성하거나, 국제적인 자료 또는 국가적인 통계자료를 대조군 집단으로 구성하여 관찰 대상자들과 PSM을 통해 보다 객관적으로 비교하는 연구를 제언한다. 이를 위해서는 타당도와 신뢰도가 높고, 세계적으로 널리 이용되고 있는 측정도구를 사용하는 것이 바람직하겠다. 실험 연구에 있어서 두 그룹 비교 연구를 재검증하는 의미로 새로운 대조군 집단과 PSM을 통해 기존의 실험군과의 결과를 비교함으로써 유사 무작위 할당과 같은 상태에서 실제에 가까운 효과를 검증하는 연구를 제언한다.

References

- Baskett, R. J., O'Connor, G. T., Hirsh, G. M., Ghali, W. A., Sabadosa, K. A., Morton, J. R., Ross, C. S., Hernandez, F., Nugent, W. C., Lahey, S. J., Sisto, D., Dacey, L. J., Klemperer, J. D., Helm, R. E., & Maitland, A. (2005). The preoperative intra-aortic balloon pump in coronary bypass surgery: A lack of evidence of effectiveness. *Am Heart J*, 150(6), 1122-1127.
- Bloom, H. S., Charles M., Caryoly, J. H., & Ying Lei. (2002). Can non-experimental comparison group methods the findings from a random assignment evaluation of mandatory welfare to work programs? *MDRS working Papers on Research Methodology*. New York, NY: Manpower Demonstration Research Corporation.
- D'agostino, R. B. (1998). Tutorial in biostatistics propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Statist Med*, 17, 2265-2281.
- Dehejia, R. H., & Sadek, W. (1999). Causal effects in nonexperimental studies; Reevaluating the evaluation of training programs. *J Am Stat Assoc*, 94(448), 1053-1062.
- Heckman, J., & Smith, J. (1995). Assessing the case for social experiments. *J Econ Perspect*, 9, 85-110.
- Lee, S. W. (2003). Evaluating the effectiveness of vocational training programs in Korea using propensity score matching. *Korean Public Adm Rev*, 37(3), 181-199.
- Ozminkowski, R. J., Burton, W. N., Goetzel, R. Z., Maclean, R., & Wang, S. (2006). The impact of rheumatoid arthritis on medical expenditures, absenteeism, and short-term disability benefits. *JOEM*, 48(2), 135-147.
- Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41-55.
- Rubin, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *Ann Stat*, 6(1), 34-58
- Rubin, D. B., & Thomas, N. (2000). Combining propensity score matching with assertional adjustment for prognostic covariates. *J Am Stat Assoc*, 95, 573-585.
- Yoo, J. S, Lee, S. J, Lee, H. C., Kim, S. H., Kang, E. S., Park, E. J. (2005). The effects of short term comprehensive life style modification program on glycemic metabolism, lipid metabolism and body composition in type 2 diabetes mellitus. *J Korean Acad Nurs*, 34(7), 1277-1287.

The Use of Propensity Score Matching for Evaluation of the Effects of Nursing Interventions

Lee, Suk Jeong¹⁾ · Yoo, Ji Soo²⁾ · Shin, Mikyung³⁾ · Park, Chang Gi⁴⁾ · Lee, Hyun Chul⁵⁾ · Choi, Eun Jin⁶⁾

1) Full Time Lecturer, Red Cross College of Nursing, 2) Professor, College of Nursing, Yonsei University

3) Doctor of Public Health, University of Illinois at Chicago, 4) Researcher, University of Illinois at Chicago

5) Professor, College of Medicine, Yonsei University, 6) Researcher, Nursing Policy Research Institute, College of Nursing, Yonsei University

Background: Nursing intervention studies often suffer from a selection bias introduced by failure of random assignment. Evaluation with selection bias could under or over-estimate any intervention's effects. PS matching (PSM) can reduce a selection bias through matching similar Propensity Scores (PS). PS is defined as the conditional probability of being treated given the individual's covariates and it can be reused to balance the covariates of two groups. **Purpose:** This study was done to assess the significance of PSM as an alternative evaluation method of nursing interventions. **Method:** An intervention study for patients with some baseline individual characteristic differences between two groups was used for this demonstration. The result of a t-test with PSM was compared with a t-test without matching. **Results:** The level of HbA1c at 12 months after baseline was different between the two groups in terms of matching or not. **Conclusion:** This study demonstrated the effects of a quasi-random assignment. Evaluation using PSM can reduce a selection bias impact that affects the result of the nursing intervention. Analyzing nursing research more objectively to reduce selection bias using PSM is needed.

Key words : Propensity score matching, Nursing intervention, Evaluation

• Address reprint requests to : Lee, Suk Jeong

Red Cross College of Nursing,

98 Saemoonangil Jongno-Gu, Seoul 110-102, Korea

Tel: 82-2-2129-1155 Fax: 82-2-2129-1122 E-mail: LSJ1109@redcross.ac.kr.