

비실험자료를 이용한 연구에서 인과적 추론의 강화: 성향점수와 도구변수 방법의 적용

김명희, 도영경¹⁾

울지외과대학교 예방의학교실, 노스캐롤라이나대 보건대학원 보건정책관리학과¹⁾

Strengthening Causal Inference in Studies using Non-experimental Data: An Application of Propensity Score and Instrumental Variable Methods

Myoung-Hee Kim, Young Kyung Do¹⁾

Department of Preventive Medicine, Eulji University College of Medicine,
Department of Health Policy and Administration, University of North Carolina at Chapel Hill School of Public Health¹⁾

Objectives : This study attempts to show how studies using non-experimental data can strengthen causal inferences by applying propensity score and instrumental variable methods based on the counterfactual framework. For illustrative purposes, we examine the effect of having private health insurance on the probability of experiencing at least one hospital admission in the previous year.

Methods : Using data from the 4th wave of the Korea Labor and Income Panel Study, we compared the results obtained using propensity score and instrumental variable methods with those from conventional logistic and linear regression models, respectively.

Results : While conventional multiple regression

analyses fail to identify the effect, the results estimated using propensity score and instrumental variable methods suggest that having private health insurance has positive and statistically significant effects on hospital admission.

Conclusions : This study demonstrates that propensity score and instrumental variable methods provide potentially useful alternatives to conventional regression approaches in making causal inferences using non-experimental data.

J Prev Med Public Health 2007;40(6):495-504

Key words : Causality, Selection bias, Research design, Health insurance, Korea

서론

보건학 분야에서 무작위통제실험(randomized controlled trial: RCT)은 인과성을 밝히는 최선의 연구설계로 여겨지지만, 이것이 항상 가능한 것은 아니다. 연구대상 할당에 윤리적 문제가 수반될 수 있고, 시간이나 비용 상의 이유로 현실적 수행이 불가능한 경우도 있으며 [1,2], 또한 연구 주제에 따라서는 RCT가 최선의 연구설계가 아닐 수도 있다 [1]. 건강의 사회적 결정요인을 다루는 사회역학 분야나 제도적 요인들을 다루는 의료관리 연구 분야에서는 이러한 특징이 더욱 두드러진다. 이를테면 소득 불평등이 인구집단의 평균 수명에 악영향을 미치는지 확인하려고, 확

률 할당을 통해 한 집단에는 소득 수준적인 조세 정책을, 다른 한 집단에는 역진적인 조세 정책을 시행할 수 있을까? 이는 불가능할 뿐 아니라 비윤리적이기도 하다. 또한 실제로 이러한 중재를 시행하였다고 해도, 소득 불평등 이외에 평균 수명에 영향을 주는 다른 사회적 요인들을 효과적으로 통제하기 어려울 뿐 아니라, 소득 불평등의 건강 효과는 수개월에서 수십 년이 걸릴 수도 있기 때문에 적절한 추적기간을 설정하는 것조차 어렵다. 그래서, 이들 분야에서는 주기적으로 수행되는 전국 표본조사 자료나 건강보험 청구자료 같은 대규모 행정관리자료 등의 비실험자료(non-experimental data), 혹은 관찰자료를 이용한 연구들이 주를 이루고 있다.

그러나 이렇게 비실험 자료를 활용하는 경우, RCT에서와 같은 '무작위할당'이 불가능하기 때문에 통계적 보정을 통해 설명변수의 '독립적인' 효과를 추정해야 하며, 이를 위해서는 다양한 공변수들을 포함한 다중회귀모형이 가장 흔히 쓰인다. 여기에서 통제변수를 '보정'한다는 것은, 관심있는 독립변수를 제외하고는 '다른 모든 조건이 동일'한 조건을 통계적으로 재구성함을 뜻한다. 이는 인과성 추론의 핵심적 개념인 반사실적 조건(counterfactual condition)의 개념과 관계 있다 [3]. 사회학 사전의 정의에 의하면, 인과적 관계란 (1) 두 사건 A와 B 사이에 공간적, 시간적 인접성이 존재하고, (2) 한 사건 A가 다른 사건 B에 선행하며, (3) 전자A가 일어나지 않았더라면 후자 B는 발생할 것 같지

않았을 경우를 말한다 [4]. 이 중 세 번째 조건은, 실제로는 A와 B가 일어났지만 만일 A가 발생하지 않았더라면 B가 일어나지 않았을 것이라는, 즉 실제 일어난 사실 (factual)에 반하는 반 사실적 조건을 가정한다. 이는 ‘만일 첫번째 객체가 없었더라면 두 번째 객체가 결코 존재하지 않았을 때, 선행하는 요인을 원인이라고 정의한다’는 철학자 David Hume의 인과관계 개념을 따르고 있다(재인용) [5].

이를테면, 비정규 고용이 노동자의 건강에 악영향을 초래할 것이라는 가설을 상정해보자. 이를 검증하는 가장 좋은 연구 방법은 노동자 A가 동일한 기간 동안 비정규 노동자이면서 동시에 정규직 노동자가 되어 ‘다른 모든 상태가 똑같은 때 비정규 고용의 효과’를 평가하는 것이다. 동일한 인구집단이 각기 다른 쪽로 상태를 경험하였다면 관찰되었을 ‘잠재적 결과 (potential outcome)’를 비교함으로써 인과적 관계를 확인할 수 있다 [6]. 그러나 다른 모든 조건이 동일한 가운데 의심되는 원인 x의 유무에 따른 두 상황을 동시에 관찰하는 것은 현실에서 불가능하다. 그래서 우리는 비정규직 노동자 집단과 정규직 노동자 집단을 비교하되, 이 두 집단이 고용형태를 제외한다면 모든 면에서 동일한 속성을 가졌다고 가정한다. 즉, 잠재적 결과를 가져올 관찰되지 않은 집단(unobserved counterfactual)과 실제 관찰된 대체 집단(observed counterfactual substitute) 사이의 교환성(exchangeability) 보장을 전제하는 것이다. RCT는 확률할당을 통해 폭로(처치) 변수 외에는 다른 모든 조건이 동일한 실험군과 대조군을 구축함으로써 반사실적 조건을 ‘인위적으로 창조’하는 반면, 비실험연구는 회귀모형에 혼란요인들을 포함시키거나 층화 분석을 통해 이러한 교환성을 보장하고자 한다. 즉, 비정규 고용 이외에 연령, 성별, 학력 등 건강에 영향을 미칠 수 있는 혼란 요인들을 보정함으로써 이러한 혼란요인들이 모두 동일하다고 가정된 상황에서 비정규고용의 ‘독립적인’ 효과를 추정하려는 것이다. 그런데 만일, 비정규 노동자들이 모두 나이가 많고 서비스/생산직에 종사하는 저학력 여성들이고, 정규직 노동자들은 이와 반대

로 젊은 연령의, 사무직에 종사하는 고학력 남성들이라면, 이들을 보정한 상태에서 비정규 고용의 독립적 효과를 추정한다는 것이 논리적으로 타당할까? 이는 일종의 구조적 혼란효과(structural confounding)로서, 보다 나은 표본추출 과정이나 표본크기의 증대를 통해 극복될 수 없는 구조적인 차이가 집단간에 존재함을 의미한다. 이러한 문제를 도외시한 채 통상적인 통계적 보정을 실시하는 것은 일종의 외삽(extrapolation)으로서, 지지 범위를 벗어난 추론(off-support inference)이라 할 수 있다 [7].

한편 의료관리 연구는 또다른 측면에서 주의를 기울일 필요가 있다. 이를테면, 허혈성 심질환에 이환된 환자나 이를 진료하는 의사는 예상 결과에 근거하여 심도자술의 시술 여부를 결정한다. 이 때 상대적으로 젊고 건강상태가 양호한 환자들의 경우에 약물치료보다는 적극적으로 심도자술을 시도할 가능성이 높기 때문에, 심도자술의 효과는 과대추정될 수 있다. 의무기록이나 보험청구자료를 이용하는 경우, 회귀모형을 통해 다양한 혼란요인들을 보정하려고 노력하겠지만 이러한 변수들을 모두 확인하는 것은 불가능하며, 그 결과 통제변수로는 포착되지 않는 중요한 차이가 치료군과 대조군 사이에서 남아 있을 수 있다 [8]. 이러한 관찰되지 않은 이질성(unobserved heterogeneity)의 문제 [9]는 진료량-결과(volume-outcome) 연구 [10]나 의료제도 및 보건사업 효과 평가 [11]에서도 나타날 수 있다.

이 논문에서는 이러한 방법론적 한계를 보완할 수 있는 접근법으로 제시되고 있는 방법 중 성향점수(propensity score)와 도구변수(instrumental variable) 방법을 다루고자 한다. 이 두 가지 방법은 모두 반사실적 조건 개념과 밀접하게 연관되어 있다.

성향점수는 이미 20여년 전, Rosenbaum과 Rubin에 의해 보건학 분야에 소개되었으나 [12], 그 이후로도 상당기간, 성향점수를 활용한 실증적 연구결과가 발표되기 보다는 ‘중설’이나 ‘강좌’를 통해 개념과 방법론의 소개에 머무른 경향이 있었다. 이에 비해 최근 성향점수를 활용한 실증 연구가 늘어난 것은, 그 내용이나 방법론

상에서 특별한 변화가 있어서라기보다 건강의 사회적 결정요인, 혹은 이차 자료 활용에 대한 관심이 높아지면서 실질적인 필요성이 늘어났기 때문이라고 할 수 있다 [13,14].

한편, 도구변수 방법은 경제학에서는 비교적 널리 활용되던 것으로, 1990년대 중반부터 의료관리 연구에 활용되면서 보건학 분야에 본격적으로 도입되었다고 할 수 있다 [8,15,16]. 역학 분야에서도 개념과 활용에 대한 소개가 이루어진 바 있으며 [17-20], 특히 사회역학에서 그 잠재적 가치를 주목받고 있다 [21]. 최근에는, 임상적 치료, 중재의 효과 평가 [22-26]나 정책 및 보건사업 효과 평가 [11] 등으로 그 활용이 확대되고 있다.

방법론의 동기 면에서 성향점수와 도구변수 방법은 유사한 부분이 많고, 따라서 이 둘은 개념적 이해를 돕기 위해서나 [27], 실제 적용 면에서나 [24,28-30] 함께 비교, 서술되는 경우가 흔하다. 이 논문에서도 동일 자료에 대한 두 가지 방법의 분석 예시를 통해 비실험자료에 기반한 인과성 추론의 강화 방법을 소개하고자 한다. 예시 사례로는 현재 한국 사회에서 중요한 사회적, 정책적 문제로 부상하고 있는 사적의료보험(이하, 사보험) [31-33] 문제를 다루고자 하였다. 앞서 설명한 반사실적 조건의 개념들에 따라, 다른 모든 조건이 동일하고 사보험 가입(구매) 여부만이 다를 때 입원의료 이용 확률에 차이가 있는지 검정하려는 것이다. 보건경제학적 관점에서 보자면 보험 가입은 자가선택(self-selection)의 가능성이 높은 전형적 변수라 할 수 있다. 향후 의료이용 확률이 큰 사람일수록 보험을 구매할 가능성이 크며 (‘역선택’), 따라서 통상적인 회귀모형에서는 사보험 가입의 의료이용 효과가 과대추정될 수 있다. 다른 한편으로 보험자에게는 위험도가 낮은 사람을 선택적으로 가입시키려는 경향이 존재하기 때문에 (‘위험 선택’), 사보험의 효과는 과소추정될 수도 있다. 이 논문에서는, 두 가지 방법론을 소개하는 것이 본래의 목적인 바, 사보험 자체에 대한 내용적 논의는 방법론 적용과 해석에 필요한 정도로만 최소화하여 다룰 것임을 미리 밝힌다.

대상 및 방법

1. 연구 자료

이 연구에서는 한국노동연구원의 ‘한국노동패널’(이하 노동패널) 제4차년도(2001년) 자료를 이용하였다. 노동패널은 국내 도시지역에 거주하는 약 5,000가구와 이에 속한 15세 이상 가구원을 대상으로 가구 특성, 개인의 사회인구학적 특성과 경제활동 등에 대해 매년 반복적인 면접 조사를 실시하고 있다. 또한 시기별로 특정 주제에 대한 부가조사를 병행하는데, 2001년도에는 ‘건강과 은퇴’라는 주제로 부가조사가 실시되었다 [34].

연구에 사용된 주요 변수들은 다음과 같다.

먼저 독립변수는 사보험 가입(구매) 여부로서, ‘건강과 은퇴’ 부가조사에 포함된 ‘국민건강보험 이외에 추가적으로 질병이나 사고 등에 대비한 건강 또는 상해보험에 가입하고 계십니까?’라는 질문에 대한 개인의 답변(‘가입하고 있다’ 또는 ‘가입하고 있지 않다’)을 이분변수로 이용하였다. 종속변수는 지난 1년간의 입원의료 이용 여부로서, ‘~님께서 지난 1년 중 입원하신 적이 있습니까?’라는 질문에 대한 답변(‘있다’ 혹은 ‘없다’)을 활용하였다. 그 밖의 혼란 변수들과 도구변수는 개인용, 가구용 자료로부터 재구성되었으며, 그 속성은 Table 1에 기술되어 있다.

2. 분석 방법

1) 성향점수 방법

Rosenbaum과 Rubin은 성향점수를 ‘관찰된 공변수 x 의 집합이 주어졌을 때 각 개체가 처치를 받게 될 조건부 확률(the conditional probability of receiving the treatment given the observed covariates x)’로 정의하였다 [12]. 달리 표현하자면, 성향점수란 폭로(처치) 여부에 영향을 미치는 일군의 혼란변수들을 단일 함수로 만든, 대상자들의 특성을 요약하는 공변수들의 단일 복합지표라 할 수 있다 [35]. 성향점수는 처치군과 대조군의 비교를 좀더 의미있게 만드는 일반적인 균형점수(balancing score)의 하나로서, 임의할당이 존재하지 않는 관찰연구에서도 처치군과 대조군 사이의

구조적 차이를 극복하게 함으로써 비뚤림 없는 추정치(unbiased estimate)를 구할 수 있도록, 즉 통계적 추론의 내적 타당도를 향상시킬 수 있도록 한다 [12].

이 연구에서 ‘사보험 가입 여부(D)가 입원의료 이용(Y)에 영향을 미친다’는 가설을 검정하고자 할 때, 폭로요인은 사보험 가입(D=1)과 비가입(D=0)으로 구분되며, 성향점수는 ‘주어진 사회경제적 속성과 건강 특성(x) 하에서 대상자가 사보험에 가입할 기대 확률’로 정의할 수 있다. 즉, 사보험 가입과 관련된 공변수들의 집합을 X로 정의할 때, 성향점수는 확률 $p(D=1|Set X)$ 가 되며, 노동패널 자료의 가용한 변수들을 이용해 식 (1)의 로지스틱 모형으로부터 추정되었다. 이 때 로지스틱 모형은 parsimonious model 을 적합시키는 것보다는, 타당하다면 가급적 많은 공변수와 교차항, 교호항들을 포함시키는 것이 바람직하다 [7] (결과는 Appendix Table 1 왼쪽 열에 제시).

$$\text{logit}(p) = \sum \beta X \dots\dots (1)$$

이러한 수식을 통해 각 관찰값마다 고유한 성향점수 값을 추정한다면 다음에는, 이를 토대로 점수가 비슷한 집단끼리 층화하거나, 가입군과 비가입군의 개별 짝짓기를 하여 폭로 효과를 추정할 수 있으며, 성향점수를 하나의 공변수로 포함한 회귀분석을 실시할 수도 있다 [12,13]. 짝짓기의 경우, 허용 폭(caliper)이 넓을수록 짝을 찾을 가능성이 높아져 표본 수는 늘어나는 반면 교환성의 훼손 가능성은 커지고, 반대로 이를 좁게 만들 경우 비교 가능성은 높아지지만 표본 수가 줄어드는 문제점이 있다. 따라서 기준 폭의 크기는 상황에 따라 저울질을 해야 하며, 그 영향을 평가하기 위해 사후 민감도 분석을 하는 것이 바람직하다 [7]. 이 논문에서는 성향점수에 따라 5군으로 층화분석한 결과와 짝짓기를 통한 분석결과를 함께 제시하였다. 짝짓기의 경우, greedy algorithm [36]을 사용하여, 사보험 가입자에 대한 비가입자의 성향점수 차이가 10^{-8} 부터 시작해서 10^{-1} 에 이를 때까지 비교를 반복하여 더 이상 적합한 짝이 없을 때까지 1:1, 1:2, 1:3 짝짓기

를 하도록 하였다. 이렇게 층화, 혹은 짝지어진 표본은 사보험 가입과 비가입이라는 점만 제외한다면 여타 공변수들의 분포가 같고 (이론적으로) RCT의 확률 할당에서와 같은 교환 가능한 표본이 되며, 따라서 두 집단의 입원의료 이용 차이는 사보험 가입에 의한 차이일 가능성이 높아진다.

한편, 본격적인 효과 추정에 앞서 성향점수를 이용한 층화나 짝짓기 과정을 통해 실제로 공변수들이 균형을 이루었는지, 즉 비교 가능한 표본이 되었는지 점검하는 것이 필요하다. 표준화 차이(standardized difference)로 평가하였을 때 폭로군과 비폭로군의 차이가 10%보다 크면 공변수의 균형이 제대로 이루어졌다고 할 수 없다. 또한 층화와 짝짓기를 통한 비뚤림의 감소 정도를 평가하기 위해 % bias reduction을 산출할 수도 있다 [13].

이러한 과정을 거쳐 짝짓기가 적절하게 이루어졌다는 것이 확인되면, 이제 효과 추정에 들어가게 된다. 여기에서는, 층화 자료에 대해서는 로지스틱 회귀 분석을, 짝지은 자료에 대해서는 조건부 로지스틱 회귀 모형을 사용하여 사보험 가입이 입원의료 이용에 미치는 효과를 추정하였다.

2) 도구변수 방법

도구변수 방법의 핵심은 도구변수를 통해 RCT의 확률할당(randomization)을 모방하여 반사실적 조건을 만들어주는 것이다 [1]. 즉, 다른 모든 조건이 동일하고 사보험 가입의 잠재적 확률만이 다르게 나타나도록 하는 자연적 실험(natural experiment) 상황을 찾는 것이다. 이를테면 규모나 부가급여 등 여러 측면에서 비슷하고, 고용된 직원들도 비슷한 사회경제적 위치에 속해 있으며 건강수준도 다르지 않은 A기업과 B기업을 가정해보자. 그런데 A기업은 직원들에게 사보험 가입 보조금을 지급한 반면 B기업은 비슷한 금액을 문화생활비로 지급하였고, 그 결과, A 직원들의 50%, B 직원들의 30%가 사보험에 가입하였다면 이를 어떻게 평가할 수 있을까? 우리는 이 20%point의 사보험 가입 차이를 보조금 지급의 효과라고 볼 수 있으며, 만일 A 직원들의 의료이용이 더 많은 것으로 나타난다면, 이미 앞서 두 집단의 유사성을 가정하였기 때문에, 이 차이를 사보험 가입

Table 1. Sociodemographic characteristics by purchase of private health insurance

Variable	Total (N=10,766)	With PHI (n=4,113)	Without PHI (n=6,653)
Having private health insurance (PHI)	4,113 (38.2)	-	-
Having at least one admission in the past year	812 (7.5)	321 (7.8)	491 (7.4)
Demographic			
Age (year)	41.43 ± 16.81	39.31 ± 10.94	42.75 ± 19.47
Female	5,584 (51.9)	2,049 (49.8)	3,535 (53.1)
Marital status			
Married currently	6,730 (62.5)	3,306 (80.4)	3,424 (51.5)
Married in the past	1,047 (9.7)	184 (4.5)	863 (13.0)
Never married	2,989 (27.8)	623 (15.2)	2,366 (35.6)
Residence in metropolitan area	6,003 (55.8)	2,233 (54.3)	3,770 (56.7)
Socioeconomic position			
Education			
<Middle school grad	2,314 (21.5)	420 (10.2)	1,894 (28.5)
<High school graduation	2,046 (19.0)	638 (15.5)	1,408 (21.2)
High school graduation	3,403 (31.6)	1,763 (42.9)	1,640 (24.7)
Some college or more	3,003 (27.9)	1,292 (31.4)	1,711 (25.7)
Job classification			
Administrative/Professional	487 (4.5)	278 (6.8)	209 (3.1)
White collar	1,282 (11.9)	789 (19.2)	493 (7.4)
Pink collar	1,381 (12.8)	726 (17.7)	655 (9.9)
Blue collar	2,480 (23.0)	1,062 (25.8)	1,418 (21.3)
No job	5,136 (47.7)	1,258 (30.6)	3,878 (58.3)
Ln(annual individual income in 10K KRW)	2.15 ± 2.37	3.04 ± 2.39	1.59 ± 2.18
Having any disability	364 (3.4)	76 (1.9)	288 (4.3)
Having any chronic disease	2,033 (18.9)	485 (11.8)	1,548 (23.3)
Household characteristics			
Ln(equalized annual income in 10K KRW)	6.60 ± 1.39	6.93 ± 1.13	6.40 ± 1.49
Ln(financial assets in 10K KRW)	4.38 ± 3.46	5.34 ± 3.18	3.78 ± 3.49
Ln(real estate and house price in 10K KRW)	5.96 ± 4.39	5.91 ± 4.43	6.00 ± 4.37
Number of household members	3.91 ± 1.32	3.95 ± 1.16	3.90 ± 1.41
Basic Livelihood Support	207 (1.9)	29 (0.7)	178 (2.7)
Instrumental variables			
Employment providing subsidy for PHI	832 (7.7)	507 (12.3)	325 (4.9)
Household having saving insurance	1,314 (12.2)	792 (19.3)	522 (7.8)

Note: Sample weights were not applied. Continuous variables are presented in mean ± standard deviation, and binary variables in number (%).

Source: Korea Labor Income and Panel Study, 4th wave.

에서 비롯된 것이라고 추론할 수 있다. 이는 일종의 자연적 실험으로서, 개인들의 건강상태나 요구에 관계없이 기업의 보조금 지급여부에 따라 직원의 사보험 가입 확률에 임의적 변이(random variation)가 초래됨으로써, 사보험 가입의 의료이용 효과를 추정할 수 있게 된 것이다.

이 사례는 도구변수 방법의 적용이 가능하기 위한 두 가지 조건을 보여준다. 첫째, 보조금은 보험가입의 충분한 유인으로 작용하여, 실제로 직원들의 사보험 가입률을 증가시켜야 한다. 즉, 사보험 가입 확률 모형에서 도구변수는 충분한 설명력을 가져야 한다. 둘째, 이 보조금은 사보험 가입을 증가시킴으로써 의료이용에 영향을 미치는 것 이외에 제 3의 경로를 통해 의료이용에 영향을 미쳐서는 안 된다. 즉, 도구변수는 최종 결과변수(입원의료이용)와 관계가 없어야 한다.

이러한 맥락에서 이 연구는 직장의 ‘개

인 의료, 상해보험료 지원’ 제공 여부를 일차 도구변수로 사용하였고, 가구의 ‘저축성보험(재테크보험, 교육보험 등)’ 가입 여부를 추가적인 도구변수로 포함하였다. 이는 저축성 보험에 가입한 가구는 위험회피 성향이 높고, 따라서 사보험 가입 가능성도 높을 것이라는 가정에 기초하고 있다.

이 두 가지 도구변수를 이용하여 두 단계 최소자승모형(instrumental variable two-stage least square estimation: IV-2SLS)을 추정하였다. 우선 1단계에서는 선형확률모형(linear probability model)에 근거하여 도구변수를 포함한 최소 자승모형으로 사보험 가입 확률을 예측하였다. 이렇게 예측된 확률을 2단계 주(主) 모형의 사보험 가입 변수에 대체하였는데, 이 계수는 도구변수를 통해 사보험 가입이 입원의료 이용에 미치는 효과를 나타낸다. 이 과정은 널리 쓰이는 통계분석 프로그램 명령문

(예: Stata의 ‘ivregress’ 나 SAS의 ‘proc syslin 2sls’ 등)을 이용하여 수행할 수 있다.

이 연구의 모든 통계적 유의성 검정은 p -value < 0.05를 기준으로 이루어졌다.

연구결과

연구 대상의 특징을 요약한 Table 1을 살펴보면, 2001년 현재, 조사 대상자인 15세 이상 인구 10,766명 중 38.2%가 사보험에 가입해 있으며 가입자와 비가입자는 다음과 같은 여러 측면에서 이질적임을 알 수 있다. 가입자들은 비가입자들에 비해 연령이 더 낮고 유배우 결혼 상태의 비율이 높았다. 또한 교육수준이 전반적으로 더 높으며, 직업 분류상 관리/전문직, 사무직 비율이 더 높고 개인 소득도 더 많았을 뿐 아니라, 가구 소득 및 자산도 더 많은 것으로 나타났다. 그리고 장애와 만성질환이 있다고 보고한 사람의 비율은 상대적으로 낮았다. 한편, 도구변수로 활용하고자 하는 두 가지 변수에서도 차이를 보였는데, 가입자군의 경우, 직장에서 ‘개인의료, 상해보험료 지원’을 제공하는 비율이 나 가구원이 ‘저축성보험(재테크보험, 교육보험 등)’에 가입해 있는 비율이 현저히 높았다. 요약하자면, 전반적으로 사회경제적 위치가 더 높고 건강한 이들이 사보험에 더 많이 가입해 있다고 할 수 있다 (Table 1). 이러한 결과는 Appendix Table 1에서도 다시 확인할 수 있다. 왼쪽 열은 성향점수를 추정하기 위한 로지스틱 모형(식 1)의 모수들로서, 여러 공변수들과 사보험 가입의 연관성을 보여준다. 한편 오른쪽 열은 사보험 가입 확률을 예측하기 위한 선형확률모형의 추정 결과로서, 도구변수를 활용한 2SLS 모형 중 1단계 결과에 해당한다 (Appendix Table 1).

한편 최종 결과 변수인 입원 이용 확률을 살펴보면, 가입자군의 입원 이용 빈도가 0.4%point 높은 것으로 나타나지만 통계적으로 유의한 차이는 관찰할 수 없었다 (Table 1).

1) 성향점수 방법

Appendix Figure 1은 사보험가입 여부에

다른 성향점수의 분포를 보여주고 있다. 원래 표본의 경우(1a), 예상할 수 있듯, 실제 가입자들은 성향점수가 높은 쪽에 몰려 있고, 비가입자들은 그 반대편에 몰려 있다. 그러면서도 두 집단의 분포는 상당 부분 중첩되어 있는데, 이는 다른 조건들이 매우 유사함에도 불구하고 실제 가입여부는 다른, 즉 ‘교환가능한’ 표본의 존재를 의미하며, 만일 이렇게 중첩된 부분이 없다면 어떠한 통계적 분석도 불가능할 것이다. 1:1, 1:2, 1:3 방식에 의한 짝짓기 결과 성향점수의 분포는 (b)~(d)와 같이 변화하며 표본 수가 감소한다. 이 때, 1:2나 1:3 짝짓기의 경우, 해당 짝이 존재하지 않는 영역, 이 자료에서는 높은 성향점수 쪽 대상자들이 대거 탈락함으로써 상대적으로 낮은 성향점수, 즉 사보험 가입 성향이 비교적 낮은 대상자들이 과다 대표되는 현상을 관찰할 수 있다(Appendix Figure 1).

한편, 층화와 짝짓기로 인한 공변수들의 분포 변화, % bias reduction 결과는 지면의 제한으로 인해 1:3 짝짓기 결과만을 Appendix Table 2에 제시하였다. 표준화 차이가 10이상인 경우를 공변수의 불균형이라고 간주할 때, 짝짓기를 통해서 대부분의 공변수들이 균형을 이루게 되었으며 그로 인한 편견 감소도 최고 99%에 이르는 것을 확인할 수 있다(Appendix Table 2).

Table 2는 여러가지 유형의 로지스틱 회귀모형을 통해 추정된 ‘사보험 가입이 임원의료 이용에 미치는 효과’를 보여주고 있다. 우선, 원래의 표본을 대상으로 로지스틱 회귀분석을 실시한 경우(panel A), 여타의 공변수를 포함하지 않은 상태에서 추정된 사보험 가입의 1년간 임원 증가비는 1.062에 불과하며 통계적으로 유의하지도 않다. 공변수를 모형에 포함하는 경우 OR은 1.174로 약간 증가하지만 통계적 유의성은 여전히 관찰되지 않는다. 성향점수에 따라 표본을 5개 집단(quintile)으로 층화분석한 경우(panel B), 성향점수의 평균이 가장 낮은(즉, 사보험 가입의 경향이 가장 적은) 첫번째 5분위군에서만 유의한 임원률 증가가 관찰되었고 (OR=2.319, 95% CI=1.250-4.301), 층화분석의 요약 추정치(stratified logistic summary estimates)를 살펴

Table 2. Relative odds to have at least one admission in the past year, estimated from a series of logistic models, using propensity scores

Analytical approach		Odds ratios	95% CI	p-value
(A) The original sample				
Crude	(N=10,766)	1.062	(0.918 - 1.230)	0.4179
Adjusted	(N=10,766)	1.174	(0.982 - 1.405)	0.0785
(B) Stratified analysis based on propensity scores (covariate-adjusted logistic regression model)				
Stratum	Q1 (PS mean =0.059) (N =2,153)	2.319	(1.250 - 4.301)	0.0076
	Q2 (PS mean =0.181) (N= 2,153)	1.480	(0.959 - 2.283)	0.0766
	Q3 (PS mean =0.364) (N= 2,154)	1.089	(0.760 - 1.559)	0.6421
	Q4 (PS mean =0.556) (N= 2,153)	1.084	(0.774 - 1.517)	0.6385
	Q5 (PS mean =0.749) (N= 2,153)	1.368	(0.891 - 2.098)	0.1520
Summary	(N=10,766)	1.230	(1.027 - 1.474)	0.0248
(C) Matched analysis based on propensity scores				
1:1 match	(N=2,714 pairs)	1.126	(0.918 - 1.382)	0.2530
1:2 match	(N=1,048 pairs)	1.237	(0.937 - 1.634)	0.1338
1:3 match	(N= 819 pairs)	1.375	(1.025 - 1.843)	0.0335

Notes: Covariates included age, gender, marital status, residence, education, job, individual income, disability or chronic disease, household equivalized income, financial & estate asset, number of household members, and basic livelihood support status, which were the same used in the model for estimating propensity scores (Equation 1). CI: confidence interval.

Source: Korea Labor Income and Panel Study, 4th wave.

Table 3. Regression analysis of having at least one admission (N=10,766)

Variable	OLS		IV-2SLS	
	Coeff.	95% CI	Coeff.	95% CI
Intercept	12.53	(7.73, 17.33)	14.58	(9.23, 19.93)
Having private health insurance	1.07	(-0.09, 2.23)	17.16	(1.28, 33.04)
Demographic				
Age (year)	-0.51	(-0.73, -0.29)	-0.63	(-0.89, -0.38)
Age squared	0.00	(0.00, 0.01)	0.01	(0.00, 0.01)
Female	-0.35	(-1.46, 0.76)	-0.57	(-1.74, 0.60)
Marital status				
Never married (reference)				
Married currently	8.89	(6.91, 10.86)	2.99	(-3.16, 9.14)
Married in the past	6.33	(3.57, 9.10)	0.71	(-5.53, 6.94)
Reside in metropolitan area	-1.20	(-2.20, -0.21)	-0.84	(-1.93, 0.25)
Socioeconomic position				
Education				
<Middle school graduation	-0.03	(-1.94, 1.89)	1.10	(-1.17, 3.38)
<High school graduation	-0.94	(-2.52, 0.65)	-0.53	(-2.22, 1.16)
High school graduation	0.25	(-1.13, 1.64)	-0.50	(-2.12, 1.11)
College or more (reference)				
Job classification				
Administrative/Professional	0.21	(-2.95, 3.37)	-0.63	(-4.00, 2.74)
White collar	-1.67	(-4.11, 0.78)	-3.10	(-5.99, -0.20)
Pink collar	-2.95	(-5.12, -0.78)	-3.58	(-5.91, -1.25)
Blue collar	-2.13	(-4.20, -0.06)	-1.61	(-3.81, 0.60)
No job (reference)				
Ln(annual individual income)	-0.26	(-0.69, 0.16)	-0.69	(-1.30, -0.08)
Having any disability	9.84	(8.37, 11.30)	10.34	(8.75, 11.93)
Having any chronic disease	13.25	(10.46, 16.04)	14.06	(11.07, 17.04)
Household characteristics				
Ln(equivalized annual income)	0.43	(0.05, 0.81)	0.03	(-0.53, 0.59)
Ln(financial assets)	-0.03	(-0.18, 0.13)	-0.29	(-0.59, 0.01)
Ln(real estate and house price)	-0.11	(-0.23, 0.01)	-0.16	(-0.29, -0.02)
Number of household members	-0.25	(-0.65, 0.14)	-0.04	(-0.50, 0.42)
Basic Livelihood Support	2.26	(-1.41, 5.93)	2.08	(-1.72, 5.88)
NR2 test, $\chi^2_{df=1}$			0.0095 (p=0.9220)	
Hausman test, $\chi^2_{df=1}$			4.24 (p=0.0395)	

Notes: OLS: ordinary least squares, IV-2SLS: instrumental variable-two stage least squares, CI: confidence interval.

Source: Korea Labor Income and Panel Study, 4th wave.

보면 전체적으로는 사보험 가입이 유의한 임원 확률 증가를 가져오는 것으로 나타났다 (OR=1.230, 95% CI=1.027-1.474). 한편, 성향점수에 의해 짝지은 표본의 조건부 로지스틱 회귀분석 결과를 살펴보면

(panel C), 1:3 짝짓기의 경우에만 유의한 임원 확률 증가를 관찰할 수 있다 (OR=1.375, 95% CI=1.025-1.843). 이는 1:3 짝지은 표본이 상대적으로 낮은 성향점수에 집중되어 있었고, 층화분석 시 첫번

제 5분위군에서 유의한 위험 증가를 보였던 것과 관련된다.

2) 도구변수 방법

도구변수 방법을 적용하기 위해서는 우선 도구변수가 사보험 가입 확률을 충분히 설명할 수 있는지 확인할 필요가 있다. 이 결과는 Appendix Table 1의 오른쪽 열에 제시되어 있는데, 두 가지 도구변수는 각각 통계적으로 유의하며 또한 F-test 결과 집합적으로도 유의하였다 (F statistic=30.81, $p < 0.001$).

이 1단계 모형을 통해 예측된 사보험 가입 확률을 2단계 주 모형에 대체하여 추정한 결과는 Table 3의 오른쪽 열에 제시되어 있다. 비교를 위해 먼저 왼쪽 열의 통상적인 선형 다중회귀분석(ordinary least squares: OLS) 결과를 살펴보면, 사보험 가입이 입원의료 이용에 미치는 효과는 1.07% point로 나타나며 통계적으로 유의하지 않았다. 그러나 도구변수를 이용한 추정에서 사보험 가입은 입원의료 이용을 17.16% point 증가시키며 통계적으로도 유의하였다. 그리고 Hausman test의 결과($\chi^2_{df=1}=4.24$, $p=0.0395$)는 OLS 추정 결과와 2SLS 추정 결과 사이에 차이가 없다는 귀무가설을 기각함으로써 OLS 추정치는 사보험이 입원의료이용에 미치는 효과를 과소추정함을 시사한다.

마지막으로, 주 모형에서 도구변수의 통계적 설명력이 없음을 확인함으로써 (Table 3, $NR2\ test\ \chi^2_{df=1}=0.0095$, $p=0.922$) [37], 다른 공변수를 포함하였을 때 도구변수가 결과변수(즉, 입원의료 이용)에 직접 영향을 미쳐서는 안 된다는 조건을 만족함을 확인할 수 있다.

고 찰

이 논문은 성향점수와 도구변수 방법을 이용하여 비실험자료 분석에서 인과성 추론을 강화할 수 있는 방법을 실제 분석을 통해 제시하고자 하였다. 두 가지 방법을 이용하여 사보험이 입원의료 이용에 미치는 영향을 분석한 결과, 통상적인 선형, 비선형 다중회귀분석에서는 통계적으로 유

의하게 나타나지 않았던 사보험의 입원의료 이용 증가 효과를 성향점수 분석의 일부 결과와 IV-2SLS 결과에서 확인할 수 있었다. 이는 통상적인 다중회귀분석의 전제인, 사보험 가입을 제외한 다른 조건이 비교 집단 간에 동일하다는 가정이 충분히 만족되지 않았음을 시사한다. 또한 내용적인 측면에서 보자면, 사보험 가입 가능성이 낮거나(사회경제적 위치가 낮고 건강 수준이 낮음) 혹은 보조금 지급 여부에 따라 가입 여부가 달라질 수 있는 이들이 사보험에 가입하는 경우 입원의료이용의 확률이 커진다는 점을 확인할 수 있었다. 이는 사보험 가입의 효과가 인구집단에서 이질적일 가능성을 시사하며(heterogeneous treatment effect), 따라서 인구집단 전체를 대상으로 사보험의 평균적 효과를 추정하는 것이 논리적으로 부적절함을 보여준다.

이어지는 글에서는 성향점수와 도구변수 방법론을 추정의 틀, 결과의 해석, 추정의 효율성이라는 세 가지 측면으로 나누어 서술하고자 한다. 이 세 가지 측면은 두 방법론이 지니는 장단점과도 관계 있으며 서로 밀접하게 관련되어 있다.

첫째, 두 가지 방법 모두 2단계에 걸쳐 반사실적 조건 하에서의 효과를 추정하는데, 성향점수 방법은 관찰된 공변수의 함수로 예측된 성향점수를 토대로 층화 또는 짝짓기를 함으로써 반사실적 조건을 구축하는데 비해, 도구변수 방법은 도구변수가 초래하는 예측 확률의 임의적 변이를 통해 반사실적 조건을 형성한다. 그렇기 때문에 각 방법론은 다음과 같은 장단점을 갖는다. 성향점수 방법론은 성향점수 예측에 적절하다고 판단되는 공변수들이 존재한다면 일단 적용할 수 있는데 비해, 도구변수 방법은 적절한 도구변수를 찾을 수 없다면 분석 자체가 불가능하다. 적절한 도구변수를 자료 내에서 구하는 것이 쉽지 않기 때문에 활용 자체가 제한적일 수 밖에 없으나, 만일 적절한 도구변수를 구할 수만 있다면 성향점수 방법에 비해 상대적인 잇점이 있다. 도구변수 방법은 자연실험 조건을 창출함으로써 비교 집단 간의 관찰되지 않는 이질성까지 통제할

수 있는 반면, 성향점수는 관찰된 공변수의 집합을 토대로 예측된 것이기 때문에 관찰되지 않은 변수에 의한 숨겨진 편견(hidden bias)을 통제하는 것이 상대적으로 어렵다 [27]. 또한, 성향점수를 예측하기 위한 모형에 어떤 변수를 포함하는가에 관한 명확한 기준이 없으며 포함된 변수에 따라 성향점수의 추정 결과가 달라질 수 있다는 점이 문제가 된다 [38]. 그리고 통상적인 다중회귀모형이 많은 공변수를 통해 집단 간의 이질성을 충분히 통제할 수 있다면, 이들 공변수에 근거한 성향점수 방법은 추가적인 이득이 별로 없을 가능성이 크다. Shah에 의하면 통상적인 회귀모형과 성향점수를 이용한 분석결과들 사이에 차이가 발견되지 않았고 [39], 이 연구에서도 성향점수 예측을 위해 사용된 변수들과 다중회귀분석에 포함된 변수들을 동일하게 하였을 때 1:1 짝지는 표본의 odds ratio 추정값이 크게 다르지 않았었다. 하지만, 성향점수를 이용한 층화분석이나 짝짓기를 하지 않더라도 일단 성향점수의 분포를 살펴보고 비교가능성(comparability)을 확인하는 것은 지지범위를 벗어난 추론을 예방하고 추론의 투명성을 확보하는데 큰 도움이 될 수 있다 [7].

둘째, 추정 결과의 해석에서도 차이가 있다. 성향점수 방법을 이용한 추정치는(다른 모든 조건이 동일할 때) 실제 사보험 가입자군과 비가입자군 사이의 차이의 평균, 즉 폭로군의 폭로평균효과(Average Treatment Effect on the Treated, ATT)를 나타낸다. 반면, 도구변수 방법의 추정치는 도구변수가 초래하는 사보험 가입 확률의 변이 범위 내에서 일어나는 입원 확률의 차이, 즉 일부 집단에만 적용되는 국지적 평균 폭로효과(Local Average Treatment Effect, LATE)를 나타낸다 [40]. 따라서 두 가지 방법에 의한 추정치는 모두 RCT의 Intention-To-Treat (ITT)에 의한 효과 추정치와는 다른 의미를 지닌다. 성향점수에 의한 층화 분석 결과나 짝지는 자료를 이용한 조건부 로지스틱 회귀분석의 결과는 통상적인 로지스틱 회귀모형과 마찬가지로 실제 비가입군에 대한 가입군의 입원의료이용 상대 증가비로 해석될 수 있다. 한편 도구변

수 방법에서 추정된 사보험의 입원 확률 증가효과 17.16% point는 사보험 가입에 관한 확률 할당을 통해 얻은 두 집단 전체 차이의 평균이 아니라, 도구변수, 즉 기업의 사보험 가입 보조금 지급 여부와 가구의 위험회피 성향 차이가 초래하는 사보험 가입 확률의 변이 범위 내에서(국지적으로) 추정되는 효과이다. 따라서 사보험 가입 확률이 원래 매우 높은 군이나 표본 전체에는 이 추정치가 적용될 수 없다. 성향점수 방법 적용 결과에서 성향점수가 낮은 제 1/5분위군이나 1:3 짝지은 자료에서만 입원 확률의 차이가 관찰되는 것도 이런 맥락에서 이해할 수 있다. 연구 결과를 표본 전체 혹은 인구집단 전체로 일반화하는 측면에서는 상대적으로 성향점수 방법이 낫다고 할 수 있으나, 도구변수의 LATE 추정이 반드시 단점이라고 할 수는 없다. 보건정책이나 사회역학 분야에서는, 효능(efficacy)을 평가하여 치료법 도입 여부를 결정하는 임상시험과는 다른 차원의 문제를 다루기 때문이다. 이를테면, 이 연구에서 나타나듯 사보험 가입 확률이 높은 사람(실제 가입자와는 다른 개념)과 낮은 사람은 실제 사보험에 가입 시에도 입원의료 이용 확률에 차이가 있을 가능성이 높다 [40]. 또한, 어떤 중재의 도입이나 확대/축소를 결정하는 정책은 대개 그러한 중재의 결과로 변이가 초래되는 인구집단에 미치는 실제 효과(effectiveness)에 초점을 둘 필요가 있다. 이런 점에서 LATE 해석은 개인의 선택을 벗어난 정책적/제도적 요인의 효과에 대한 질문에 좀더 적절한 답을 제시할 수 있다 [1]. 이 연구에서라면, 기업의 사보험 보조금 지급이 사보험 가입 증가군의 입원을 증가시키는 것으로 나타나는데, 따라서 기업의 사보험 보조금 지급을 정책적으로 장려할 것인가 말 것인가에 대한 근거를 제공할 수 있다.

셋째, 추정의 효율성은 흔히 간과되지만 중요한 문제 중 하나이다. 두 가지 방법 모두 내적 타당도(internal validity) 향상을 위해 효율성이 희생되는 경향이 있다. 먼저 성향점수 방법에서는 집단 간 교환가능성 보장을 위해 짝짓기가 이루어진 일부 표본만을 사용함으로써 통계적 검정력이 감

소할 수 있다. 하지만 인과적 대조(causal contrast)를 극대화시키고, 여러 개의 공변수를 하나의 단일한 변수로 총화함으로써 혼란변수의 개수를 감소시킨다는 점에서 표본 수의 감소를 반드시 단점으로 여길 필요는 없다. 한편, 도구변수 방법에서는 도구변수가 초래하는 변이의 범위가 크지 않으면 추정의 효율성이 현저히 감소한다. 이 연구에서도 도구변수를 적용하였을 때의 모수 추정치는 통상적인 회귀모형의 추정치에 비해 매우 넓은 신뢰구간을 보이고 있다. 따라서 효과의 크기가 비교적 큰 경우에도 귀무가설을 기각하지 못하는 제2종 오류의 가능성이 증가한다. 이 점은 성향점수나 도구변수 방법론 적용을 위해서 상대적으로 대규모의 자료가 필요함을 시사한다. 하지만 이들 방법론이 대규모 이차자료의 분석에 활용된다면 이는 큰 문제가 되지 않을 수 있다.

이와 같은 장단점을 고려할 때, 비실험자료를 이용한 연구에서 이러한 방법론을 적용해야 하는지, 그리고 양자 중에서 어떤 것을 선택하는가 하는 문제는 신중하게 결정되어야 한다. 무엇보다도 연구 문제에 반사실적 조건 개념을 적용한 방법론이 이론적으로 필요하고 적절한 것인지를 검토하는 것이 선행되어야 하며, 자료의 성격과 크기, 적절한 도구변수의 확보 가능성이 실질적인 결정 기준이 될 것이다.

위에서 기술한 성향점수와 도구변수 방법론 상의 문제 이외에 연구의 제한점을 지적한다면 노동패널 자료의 한계를 들 수 있다. 이는 전국적 대표성을 지니고 있으며 고용 및 경제활동과 관련한 상세한 정보를 담고 있다는 장점이 있으나 건강이나 의료이용이 연구의 일차적인 목적이 아니기 때문에 사보험 가입, 질병, 건강상태, 의료이용에 관한 정보가 매우 제한적이다. 특히 이 연구에서는, 현실에서 매우 다양한 형태의 사보험 상품이 존재하고 있음에도 불구하고 단일 문항에 근거하여 사보험 가입 여부만을 독립변수로 구성하였고, 입원의료 이용에 영향을 미칠 수 있는 건강 관련 정보도 매우 제한적이었다. 그러나 다수의 행정자료나 전국 표본조사 자료 등의 이차 자료들은 특정 주제의 연

구를 위해 설계된 것이 아니기 때문에, 이러한 자료의 한계는 이 연구만의 제한점이라기보다 일반적인 현상으로 바라 보아야 할 것이다.

이 글에서는 비실험자료를 이용한 연구에서 인과성 추론을 강화하기 위해 쓰이는 대표적 방법으로서 성향점수와 도구변수 방법을 소개하고 실제 분석에 적용해 보였다. 사보험의 입원의료 이용에 미치는 효과를 분석한 결과에서 성향점수와 도구변수 방법은 통상적인 선형, 비선형 다중회귀분석에서는 나타나지 않은, 사보험의 입원의료 이용 증가 효과를 보여주었다. 두 가지 방법론은 반사실적 조건 개념을 적용한다는 면에서는 유사하나, 추정의 기본 틀, 추정 결과의 해석, 추정의 효율성 면에서 차이가 있다. 비실험자료를 이용한 연구에서 이들 방법론을 적용하기 위해서는, 연구 문제에 반사실적 조건 개념을 적용한 방법론이 이론적으로 필요하고 적절한 것인지를 검토하는 것이 선행되어야 하며, 자료의 성격과 크기, 적절한 도구변수의 확보 가능성이 실질적인 기준으로 고려되어야 한다. 이 두 가지 방법이 인과성 추론에서 대두하는 모든 문제들을 해결할 수는 없으나, 적절하게 활용될 경우 비실험자료를 이용한 연구에서 인과성 추론을 강화하는 유용한 접근법이 될 수 있을 것이다.

감사의 말씀

이 연구에 사용된 한국노동패널 자료를 제공해 주신 한국노동연구원에 감사드립니다.

참고문헌

1. Newhouse JP, McClellan M. Econometrics in outcomes research: The use of instrumental variables. *Annu Rev Pub Health* 1998; 19: 17-34
2. Kaufman JS, Kaufman S, Poole C. Causal inference from randomized trials in social epidemiology. *Soc Sci Med* 2003; 57(12): 2397-2409
3. Kaufman JS, Cooper RS. Seeking causal explanations in social epidemiology *Am J Epidemiol* 1999; 150(2): 113-120
4. Jary D, Jary J. HarperCollins Dictionary of

- Sociology. New York: HarperCollins Publishers, Ltd.; 1991
5. Lewis D. Causation. *J Philos* 1973; 70(17): 556-567
 6. Maldonado G, Greenland S. Estimating causal effects. *Int J Epidemiol* 2002; 31(2): 422-429
 7. Oakes JM, Johnson PJ. Propensity score matching for social epidemiology. In: Oakes JM, Kaufman JS, editors. *Methods for Social Epidemiology*. San Francisco: Jossey-Bass; 2006. p. 370-392
 8. McClellan M, McNeil BJ, Newhouse JP. Does more intensive treatment of acute myocardial infarction in the elderly reduce mortality? Analysis using instrumental variables. *JAMA* 1994; 272(11): 859-866
 9. Zohoori N, Savitz DA. Econometric approaches to epidemiologic data: Relating endogeneity and unobserved heterogeneity to confounding. *Ann Epidemiol* 1997; 7(4): 251-257
 10. Luft HS, Hunt SS, Maerki SC. The volume-outcome relationship: Practice-makes-perfect or selective-referral patterns? *Health Serv Res* 1987; 22(2): 157-182
 11. Lee JY, Rozier RG, Norton EC, Vann WF Jr. Addressing selection bias in dental health services research. *J Dent Res* 2005; 84(10): 942-946
 12. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983; 70(1): 41-55
 13. D'Agostino RB Jr. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Stat Med* 1998; 17(19): 2265-2281
 14. Rubin DB. Estimating causal effects from large data sets using propensity scores. *Ann Intern Med* 1997; 127(8 Pt 2): 757-763
 15. Newhouse JP, McClellan M. Econometrics in outcomes research: The use of instrumental variables. *Annu Rev Pub Health* 1998; 19(1): 17-34
 16. McClellan MB, Newhouse JP. Overview of the special supplement issue. *Health Serv Res* 2000; 35(5 Pt 2): 1061-1069
 17. Menemeyer ST. Can econometrics rescue epidemiology? *Ann Epidemiol* 1997; 7(4): 249-250
 18. Zohoori N, Savitz DA. Econometric approaches to epidemiologic data: Relating endogeneity and unobserved heterogeneity to confounding. *Ann Epidemiol* 1997; 7(4): 251-257
 19. Greenland S. An introduction to instrumental variables for epidemiologists. *Int J Epidemiol* 2000; 29(4): 722-729
 20. Martens EP, Pestman WR, de Boer A, Belitser SV, Klungel OH. Instrumental variables: Application and limitations. *Epidemiology* 2006; 17(3): 260-267
 21. Glymour MM. Natural Experiments and Instrumental Variable Analyses in Social Epidemiology. In: Oakes JM, Kaufman JS, editors. *Methods for Social Epidemiology*. San Francisco: Jossey-Bass; 2006. p. 429-460
 22. Zeliadt SB, Potosky AL, Penson DF, Etzioni R. Survival benefit associated with adjuvant androgen deprivation therapy combined with radiotherapy for high- and low-risk patients with nonmetastatic prostate cancer. *Int J Radiat Oncol Biol Phys* 2006; 66(2): 395-402
 23. Brooks JM, Chrischilles EA, Scott SD, Chen-Hardee SS. Was breast conserving surgery underutilized for early stage breast cancer? Instrumental variables evidence for stage II patients from Iowa. *Health Serv Res* 2003; 38(6): 1385-1402
 24. Earle CC, Tsai JS, Gelber RD, Weinstein MC, Neumann PJ, Weeks JC. Effectiveness of chemotherapy for advanced lung cancer in the elderly: Instrumental variable and propensity analysis. *J Clin Oncol* 2001; 19(4): 1064-1070
 25. Hadley J, Polsky D, Mandelblatt JS, Mitchell JM, Weeks JC, Wang Q, Hwang YT. An exploratory instrumental variable analysis of the outcomes of localized breast cancer treatments in a Medicare population. *Health Econ* 2003; 12(3): 171-186
 26. Bao Y, Duan N, Fox SA. Is some provider advice on smoking cessation better than no advice? An instrumental variable analysis of the 2001 National Health Interview Survey. *Health Serv Res* 2006; 41(6): 2114-2135
 27. Schwartz M, Ash AS. Estimating the effect of an intervention from observational data. In: Iezzoni LI, editor. *Risk Adjustment For Measuring Health Care Outcomes*. 3rd ed. Ann Arbor: AcademyHealth/Health Administration Press; 2003
 28. Landrum MB, Ayanian JZ. Causal effect of ambulatory specialty care on mortality following myocardial infarction: A comparison of propensity score and instrumental variable analyses. *Health Serv Outcome Res Meth* 2001; 2(3-4): 221-245
 29. Posner MA, Ash AS, Freund KM, Moskowitz MA, Shwartz M. Comparing standard regression, propensity score matching, and instrumental variables methods for determining the influence of mammography on stage of diagnosis. *Health Serv Outcome Res Meth* 2001; 2(3-4): 279-290
 30. Stukel TA, Fisher ES, Wennberg DE, Alter DA, Gottlieb DJ, Vermeulen MJ. Analysis of observational studies in the presence of treatment selection bias: Effects of invasive cardiac management on AMI survival using propensity score and instrumental variable methods. *JAMA* 2007; 297(3): 278-285
 31. Yoon T, Hwang I, Sohn H, Koh K, Jeong B. The determinants of private health insurance purchasing decisions under national health insurance system in Korea. *Kor J Health Pol Admin* 2005; 15(4): 161-175 (Korean)
 32. Kang SW, Kwon YD, You CH. Effects of supplemental insurance on health care utilization and expenditures among cancer patients in Korea. *Kor J Health Pol Admin* 2005; 15(4): 65-80 (Korean)
 33. Lim JH, Kim SG, Lee EM, Bae SY, Park JH, Choi KS, Hahm MI, Park EC. The determinants of purchasing private health insurance in Korean cancer patients. *J Prev Med Public Health* 2007; 40(2) : 150-154 (Korean)
 34. Korea Labor Institute. *Korean Labor and Income Panel Study (KLIPS) User's Guide*. Korea Labor Institute; 2006 (Korean)
 35. Rubin DB. Estimating causal effects from large data sets using propensity scores. *Ann Intern Med* 1997; 127(8 Pt 2): 757-763
 36. Parsons LS. Performing a 1:N case-control match on propensity score. Proceedings of the Twenty-Ninth Annual SAS® Users Group International Conference. Montreal: SAS Institute Inc.; 2004
 37. Greene WH. *Econometric Analysis*. 5th ed. Upper Saddle River: Prentice Hall; 2003
 38. Smith JA, Todd PE. Does matching overcome LaLonde's critique of nonexperimental estimators? *J Econom* 2005; 125(1-2) : 305-353
 39. 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(6): 550-559
 40. Angrist JD, Krueger AB. Instrumental variables and the search for identification: From supply and demand to natural experiment. *J Econ Persp* 2001; 15(1): 69-85

Appendix Table 1. Regression analysis of having private health insurance (N=10,766)

Variable	Logistic model		OLS model	
	Odds ratio	95% CI	Coeff.	95% CI
Intercept			-12.64	(-20.45, -4.84)
Demographic				
Age (year)	1.185	(1.153, 1.217)	0.85	(0.49, 1.20)
Age squared	0.998	(0.997, 0.998)	-0.02	(-0.02, -0.01)
Female	1.063	(0.956, 1.181)	1.44	(-0.36, 3.24)
Marital status				
Never married (reference)				
Married currently	4.070	(3.416, 4.849)	35.61	(32.47, 38.76)
Married in the past	3.417	(2.605, 4.482)	33.89	(29.43, 38.35)
Reside in metropolitan area	0.884	(0.805, 0.971)	-2.22	(-3.84, -0.60)
Socioeconomic position				
Education				
<Middle school grad	0.792	(0.658, 0.954)	-6.85	(-9.96, -3.74)
<High school graduation	0.989	(0.850, 1.152)	-2.16	(-4.74, 0.43)
High school graduation	1.261	(1.113, 1.430)	4.59	(2.34, 6.84)
Some college or more (reference)				
Job classification				
Administrative/Professional	1.288	(0.974, 1.702)	4.39	(-0.75, 9.53)
White collar	1.527	(1.228, 1.898)	7.91	(3.91, 11.91)
Pink collar	1.259	(1.039, 1.526)	4.04	(0.51, 7.57)
Blue collar	0.919	(0.761, 1.111)	-3.37	(-6.74, 0.00)
No job (reference)				
Ln(annual individual income)	1.102	(1.061, 1.145)	2.50	(1.81, 3.20)
Having any disability	0.675	(0.502, 0.907)	-3.18	(-5.55, -0.80)
Having any chronic disease	0.906	(0.785, 1.046)	-4.99	(-9.52, -0.46)
Household characteristics				
Ln(equivalized annual income)	1.174	(1.127, 1.224)	2.35	(1.73, 2.97)
Ln(financial assets)	1.083	(1.067, 1.099)	1.40	(1.14, 1.65)
Ln(real estate and house price)	1.022	(1.010, 1.033)	0.29	(0.09, 0.49)
Number of household members	0.897	(0.863, 0.933)	-1.49	(-2.13, -0.84)
Basic Livelihood Support	0.898	(0.570, 1.414)	1.18	(-4.79, 7.15)
Instrumental variables (IVs)				
Employment providing subsidy for PHI	1.297	(1.092, 1.539)	5.55	(2.35, 8.74)
Household having saving insurance	1.538	(1.335, 1.772)	9.13	(6.59, 11.68)
<i>F</i> -test on IVs, <i>F</i> (2, 10742)		-	30.81	(<i>p</i> <0.001)
Goodness-of-fit				
Hosmer-Lemeshow test, $\chi^2_{df=8}$	5.1355	(<i>p</i> =0.7430)	-	-
c-index		0.81	-	-
Adjusted <i>R</i> -squared		-		25.75%

Notes: Logistic regression model was applied to predict propensity scores, and ordinary least square (OLS) model was used for the 1st stage equation in the two stage least square estimation. Parameter estimates are presented in odds ratios and linear coefficients, respectively. CI: confidence interval.

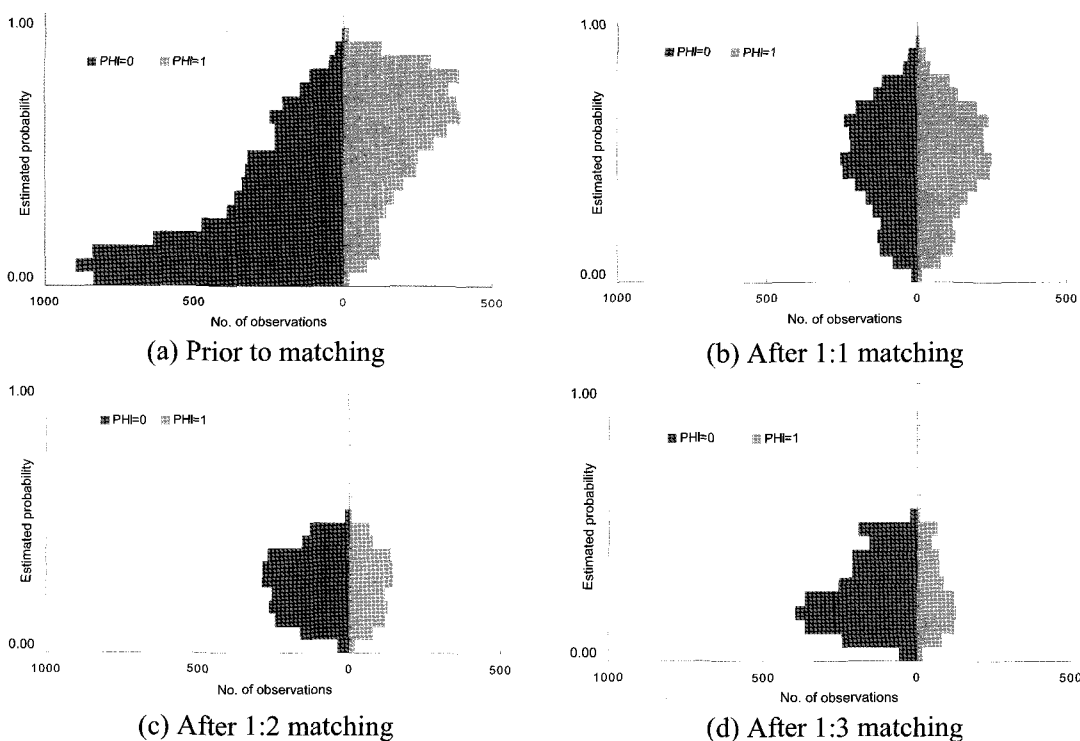
Source: Korea Labor Income and Panel Study, 4th wave.

Appendix Table 2. Covariates distribution and bias reduction by matching (1:3 match)

Variable	Matched pairs		Standardized difference ^c	% bias reduction [†]	T-stat	p-value
	Without PHI (n1=2,457)	With PHI (n2=819)				
<i>Demographic</i>						
Age (year)	40.1	39.4	3.935	81.9	0.98	0.329
Age squared	1,914.0	1,860.4	3.654	90.3	0.90	0.368
Female	51.4	52.5	2.117	68.1	0.52	0.560
<i>Marital status</i>						
Married currently	45.7	45.2	1.144	98.2	0.28	0.777
Married in the past	9.6	9.4	0.555	98.2	0.14	0.891
Never married	44.7	45.4	1.472	97.0	-0.36	0.715
Reside in metropolitan area	58.5	60.0	2.981	37.6	-0.74	0.461
<i>Socioeconomic position</i>						
<i>Education</i>						
<Middle school graduation	26.3	24.5	3.926	91.7	0.97	0.333
<High school graduation	22.0	22.1	0.294	98.0	-0.07	0.942
High school graduation	22.8	24.3	3.548	91.0	-0.88	0.377
Some college or more	29.0	29.1	0.179	98.6	-0.04	0.965
<i>Job classification</i>						
Administrative/Professional	2.8	2.9	0.731	95.6	-0.18	0.856
White collar	6.1	7.1	3.766	89.3	-0.92	0.359
Pink collar	9.6	10.7	3.631	84.1	-0.91	0.362
Blue collar	24.4	21.7	6.282	40.9	1.54	0.123
No job	57.0	57.5	0.987	98.3	-0.24	0.807
Ln(annual individual income)	1.5	1.5	0.648	99.0	0.16	0.872
Having any disability	4.2	4.3	0.404	97.2	-0.10	0.920
Having any chronic disease	22.0	19.8	5.407	82.3	1.33	0.185
<i>Household characteristics</i>						
Ln(equivalized annual income)	6.4	6.4	0.500	98.8	0.12	0.906
Ln(financial assets)	3.7	3.7	2.000	95.7	0.50	0.619
Ln(real estate and house price)	6.1	6.2	1.666	18.1	-0.41	0.680
Number of household members	3.9	4.0	5.448	-42.9	-1.35	0.176
Basic Livelihood Support	2.2	1.7	3.528	77.0	0.90	0.367

Notes: PHI: private health insurance.
Source: Korea Labor Income and Panel Study, 4th wave.

$$^c \text{standardized difference} = \frac{100(\bar{x}_{\text{exposed}} - \bar{x}_{\text{unexposed}})}{\sqrt{\frac{s_{\text{exposed}}^2 + s_{\text{unexposed}}^2}{2}}} \quad ^\dagger \% \text{ bias reduction} = 1 - \left(\frac{|\text{Standardized Difference}_{\text{matched}}|}{|\text{Standardized Difference}_{\text{unmatched}}|} \right)$$



Appendix Figure 1. Distribution of propensity scores prior to and after matching.