

# 체질량지수와 비만이 개인의료비지출에 미치는 영향에 대한 분석: Control Function 방법

이 용 우  
(영남대학교)

본 연구에서는 관찰되지 않는 개인적인 이질성의 통제를 위한 control function 방법과 의료비지출변수의 특징을 잘 고려하는 two part 모형을 이용하여 우리나라에서 체질량지수와 비만상태가 개인의료비지출에 미치는 영향을 한국의료패널 2009~2015년간 데이터를 이용하여 분석하였다. 연구결과 전체표본에서 체질량지수 한 단위의 증가는 약 5만 원의 추가적인 연간의료비지출을, 비만상태는 비만이 아닌 상태와 비교했을 때 92만 원의 추가적인 연간의료비지출을 유발하는 것으로 나타났다. 더 나아가 분위수회귀의 결과를 볼 때 체질량지수와 비만상태는 의료비지출액분포의 상단으로 올라갈수록, 즉 건강이 좋지 않아 의료비지출이 큰 집단에 더 큰 영향을 미치는 것으로 나타난다. 또한 대부분의 결과에서 control function 변수에 대한 추정계수가 유의하게 음의 값을 가지며, 이는 내생성이 적절하게 통제되지 않는 경우 과소추정이 발생할 수 있음을 보여 준다.

주요 용어: 체질량지수, 비만, 개인의료비지출, Control Function, Two Part 모형

이 연구는 2018학년도 영남대학교 학술연구조성비에 의한 것임

■ 투고일: 2019.1.28 ■ 수정일: 2019.3.29 ■ 게재확정일: 2019.4.2

## I. 서론

전 세계적으로 비만은 지역을 가리지 않고 증가추세에 있으며<sup>1)</sup>, 그에 따라 지역별로 많은 사회경제적 비용을 초래하고 있다. 우리나라에서도 비만율은 지속적인 증가추세에 있다. 국가지표체계의 통계에 의하면 체질량지수 25이상으로 정의된 우리나라의 비만율은 [그림 1]의 파란색 선으로 나타난 것처럼 1998년 이래 지난 20년간 증가해오고 있으며, 2017년 현재 전체 비만율은 34.1%(남성 41.6%, 여성 25.6%)이다.<sup>2)</sup> 이러한 비만율의 증가는 상당히 심각한 문제인데 비만이 각종 만성질환과 치명적인 질병의 유발요인으로 알려져 있기 때문이다. Dixon(2010)은 이에 대한 세밀한 연구를 진행한 바 있다. 이 연구에 따르면 비만은 고혈압, 제2형 당뇨병, 각종 심장질환, 각종 암, 골관절염, 천식 등의 육체적인 질환뿐만 아니라 우울증을 포함한 각종 정신적 질병을 야기하는 것으로 나타난다. 더 나아가, 결근, 산업재해, 생산성 저하 등과 관련된 간접적인 사회적 비용까지도 초래하는 것으로 나타난다.

이러한 각종 비만관련 질병은 결국 관련 의료이용을 늘려 개인적인 의료비지출과 사회적인 의료비지출을 증가시키게 된다. 이러한 문제 때문에 그간 비만이 의료비지출에 미치는 영향에 대한 연구가 전 세계적으로 광범위하게 진행되었다. Kim and Basu(2016)는 체계적 문헌고찰을 통해 미국에서의 그간의 대표적 연구성과들에 대해 세밀하게 논의하였다. 이 문헌고찰에 따르면 그간의 대표적 연구는 데이터, 통계적 모형, 목표모집단을 통해 분류할 수 있다. 일단 데이터의 측면에서 대부분의 연구들이 미국 의료비패널조사(Medical Expenditure Panel Surveys)와 메디케어 현재수급자 서베이(Medicare Current Beneficiary Survey)를 이용하는 것으로 나타나며, 이러한 데이터들은 대부분 전국적인 대표성을 갖는 표본이다. 이러한 공통점에도 불구하고 연구들은 방법론에서 상당한 차이를 보인다. 초기연구는 주로 양의 값을 갖는 의료비를 대상으로 의료비지출액 변수를 로그 전환(Wolf et al., 2008, Bell et al., 2011, Yang & Hal, 2008)하여 단일 방정식 체계를 추정했다(Wolf et al., 2008, Onwudiwe et al., 2011, Alley et al., 2012, Cai et al., 2010 등 참조). 하지만 이에 대해서 의료비발생여부 자체

1) 세계비만연합(World Obesity Federation)의 global obesity observatory를 통해 이를 시각적으로 확인할 수 있다(<https://www.worldobesitydata.org/map/overview-adults>).

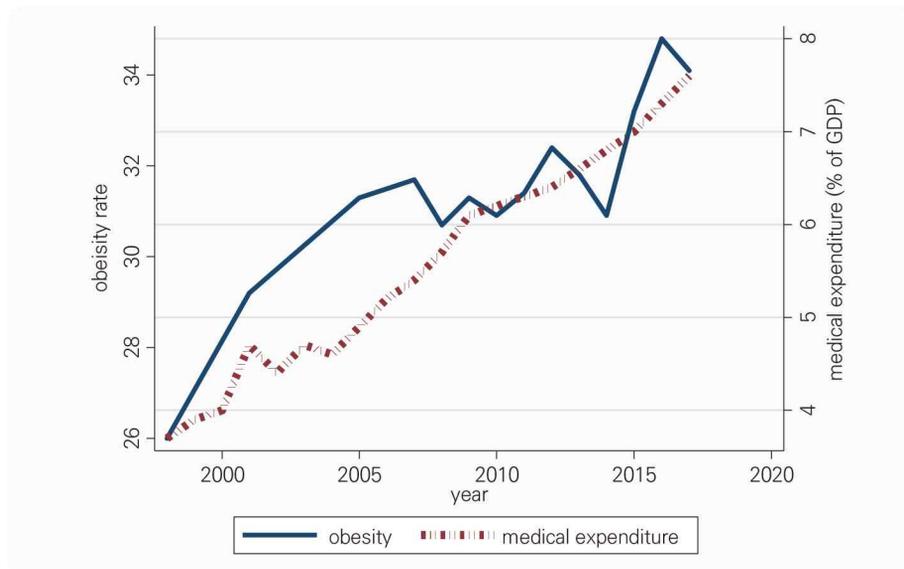
2) [그림 1]에서 1999~2000년, 2002~2004년, 2006년의 데이터는 결측치이다.

는 의료비수준의 변화와는 달리 별도로 동시에 추정되어야 한다는 문제가 제기되었다. 이러한 논의를 거쳐 현재는 의료비발생자체의 분석과 의료비지출액 수준을 동시에 분석하는 two part 모형이 광범위하게 이용되고 있다(Finkelstein et al., 2008, Finkelstein et al., 2009, Cawley & Meyerhoefer, 2012, Yang & Hall, 2008 등 참조). 더 나아가 최근에는 몸무게, 체질량지수, 비만도가 내생성을 가지는 변수이므로 이러한 내생성을 적절하게 통제하지 않으면 추정계수에 편의가 발생할 가능성이 존재한다는 사실에 주목하여 도구변수추정법이 광범위하게 이용되고 있는 추세이다(Cawley & Meyerhoefer, 2012, Cawley et al., 2015). 우리나라에서 의료비지출은 [그림 1]에서처럼 지속적인 증가세에 있다. [그림 1]의 빨간색 굵은 점선은 GDP대비 경상의료비지출을 보여준다. 이 데이터에서 GDP대비 의료비지출과 비만을 간의 상관계수는 0.8772로 두 변수 간에 상당한 정도의 양의 상관관계가 존재하여 이에 대한 연구가 필요함을 보여준다. 하지만 사안의 중요성을 감안할 때 국외의 연구성과들과 비교하여 국내의 연구성과는 그리 활발하지 못한 것이 사실이다. Kang 등(2011)은 국민건강보험공단의 2005년 데이터를 이용하여 20세 이상의 성인을 대상으로 과체중과 비만의 사회경제적비용을 추정하였다. 따라서 이 연구에서의 의료비는 국민건강보험 부담금이며 실제 개인들이 지출한 수납금액은 포함되지 않는다. 이들의 추정에 의하면 과체중과 비만으로 인한 총비용(의료이용 관련 직접비용과 사회경제적 간접비용의 합)은 2005년 현재 GDP의 0.22%, 우리나라 전체 의료비지출의 3.7%에 달하는 것으로 나타난다. 미시데이터를 이용한 비만관련 의료비지출의 연구는 윤난희와 권순만(2013)을 통해 이루어졌다. 이 연구는 한국의료패널 2009-2010년간 자료를 이용하여 two part 모형을 추정하였는데, 정상군과 비만군을 비교한 결과 비만군이 정상군에 비해 비만으로 인해 8,721원을 연간의료비로 더 지출하는 것으로 나타났다. 하지만 이들의 연구는 체질량지수와 비만도가 가지는 내생성에 대한 고려가 전혀 없어 추정결과에 편의가 존재할 가능성이 크다.<sup>3)</sup> 본 연구는 우리나라에서 비만이 의료비지출에 미치는 영향을 파악하기 위하여 한국의료패널 2009-2015년간 자료를 이용하여 분석을 한다. 특히, 현재까지의 연구성과를 반영하여 의료비지출액 변수의 특성을 잘 반영하는 two part 모형을 통해 분석을 시도한다. 여기서 의료비발생 자체에 대한 분석에는 probit 모형을 이용하며, 의료비지출액 수준에 대한 분석에는 일

3) 또한, 2절 분석방법에서 설명하듯이 Two part 모형의 일반화 선행모형 부분은 modified Park 검정 등을 이용하여 분포함수를 결정해야 하지만 이들의 연구에는 이에 대한 논의가 전혀 없다.

반화 선형모형(Generalized Linear Model; 이하 GLM)을 이용한다. 가장 중요하게는 체질량지수와 비만상태 변수의 내생성을 고려하여 도구변수추정법에서 배제조건(exclusion restriction)을 만족하는 변수로 여겨지는 자녀의 체질량지수와 비만상태를 이용한 control function 방법을 통하여 모형을 추정한다. [그림 1]에서 우리가 확인할 수 있는 것은 두 변수간의 단순한 상관관계이며, 실제 정책적 함의를 가지는 관계는 인과관계이다. 이러한 인과관계 분석을 위해서 미시계량분석에서는 누락변수(omitted variable)의 문제를 초래하는 개인들의 관찰되지 않는 이질성(unobserved individual heterogeneity)을 통제하는 것이 가장 중요하다. 이를 통제해야 우리가 관심을 가지는 설명변수인 체질량지수와 비만상태 변수의 추정계수에 편의가 발생하지 않기 때문이다. 본 연구는 비만이 의료비지출에 미치는 영향을 분석한 연구에서 개인들의 관찰되지 않는 이질성을 통제하기 위하여 control function 방법을 이용한 우리나라 최초의 연구이다. 본 논문의 전개는 다음과 같다. II절에서는 본 연구에서 사용되는 데이터에 대한 설명이 주어진다. 또한 의료비지출액, 체질량지수, 비만에 대한 다양한 기술

그림 1. 비만율과 의료비지출의 추이(1998-2017)



자료: 국가지표체계

통계적 분석이 이루어진다. 그 다음으로는 본 연구에서 이용하는 실증분석모형에 대해 자세하게 설명을 한다. III절에서는 연구결과가 주어지고 이에 대한 설명이 주어진다. 마지막으로 IV절에서는 결과에 대한 요약, 본 연구의 한계, 그리고 향후 연구방향 등이 논의된다.

## II. 연구방법

### 1. 연구대상

#### 가. 데이터와 변수들의 기술통계량

본 연구는 한국보건사회연구원과 국민건강보험공단이 공동으로 주관하는 한국의료패널 2009~2015년 연간데이터(Version 1.4)를 활용하였다.<sup>4)</sup> 양 기관은 컨소시엄을 구성하여 우리나라 가구의 보건의료실태와 의료비 지출수준, 건강수준 및 건강행태 등에 관한 기초자료를 생산하기 위하여 2008년 7,866가구를 선정하여 분석을 시작한 이후 전국 규모의 한국의료패널 사업을 공동으로 수행해 오고 있다. 2008년 한국의료패널 표본가구로 선정된 7,866가구 중 일부가 이사, 조사거부 등으로 탈락하였으며, 이를 보완하고자 2012년에 약 2,500가구를 추가적으로 선정하여 2015년 말 현재 분가가구를 포함하여 총 6,984가구가 등록, 관리되고 있다. 조사내용에는 가구 및 가구원의 사회경제적 특성 전반, 민영의료보험, 건강관련 인식 및 행태, 연간 응급입원·외래 의료이용 내역, 일반의약품 복용, 만성질환, 의료비지출액 등과 관련된 포괄적인 정보가 포함되어 있어 본 연구가 필요로 하는 주요 변수가 망라되어 있다. 본 연구의 실증분석에서 이용되는 데이터의 구축과정은 다음과 같다. 다음 절인 분석방법에서 자세한 논의를 하겠지만 본 연구를 위해서는 자녀의 체질량지수와 부모의 체질량지수를 연결시키는 것이 관건이다. 체질량지수를 구축하기 위해 필요한 정보인 신장과 몸무게에 대한 조사는 한국

4) 원 데이터는 2008년부터 시작을 하지만 2008년에는 본 연구에서 이용하는 주요 변수인 신장과 몸무게 등이 조사되지 않아 2009년 데이터부터 이용한다.

의료패널에서 만 18세 이상의 성인 가구를 대상으로 조사되었다. 그러므로 18세 이상의 동거 자녀가 있는 중장년층 이상을 중심으로 하여 표본이 구축되었다.<sup>5)</sup> 이런 과정을 통하여 2009~2015년간 5,429명에 대한 총 19,412개의 관측치로 구성된 표본이 만들어 졌다.<sup>6)</sup>

<표 1>은 실증분석에 사용된 변수들에 대한 기술통계량을 제시한다. 실증분석은 2009-2015년간의 통합데이터를 이용하므로 개인의료비지출액과 가구소득은 모두 소비자물가지수를 이용하여 실질변수로 전환하였다.<sup>7)</sup> 개인의료비지출액은 실제 수납금액(법정본인부담금+비급여항목)을 의미한다. 건보부담금을 묻는 설문항목도 있으나 모름/무응답의 비율이 너무 많아 이용가능하지가 않다. 가구소득의 경우는 가구소득을 가구원수의 제곱근으로 나눈 가구원수 보정 연간가구소득을 이용하였다.<sup>8)</sup> 체질량지수의 경우 부모들의 평균 BMI는 23.50으로 나타나며, 이들의 자녀들의 평균 BMI는 21.97로 부모들의 지수가 1.5정도 더 높은 것으로 나타난다. 결국 체질량지수를 반영한 표본의 평균 비만율은 27%로 나타나며, 이들의 자녀들의 비만율은 16%로 연령이 증가함에 따라 비만도가 증가하는 것으로 나타난다. 표본의 평균 연령은 56세인 것으로 나타나 자녀의 체질량지수를 연결시킬 수 있는 본 연구의 표본은 중장년층 표본임을 알 수 있다. 또한 이들의 대략 90%는 혼인 중이며, 20% 가량이 전문대 이상의 고등교육을 받은 것으로 나타난다. 또한 68%가 경제활동을 영위 중이며, 97%는 국민건강보험에 가입한 상태이다. dsmok2는 개인이 고위험흡연자인지를 나타내는 터미변수로서 고위험흡연자이면 1 아니면 0의 값을 가진다. 본 연구에서는 현재 흡연자 중 매일흡연자로서 일 15개 비 이상 흡연하고 있는 자를 고위험흡연자로 정의한다. 실제 건강에 유해한 행위는 가끔 흡연 혹은 소량흡연이라기보다는 각종 흡연억제정책에 비탄력적인 고위험흡연이라고

5) 자녀가 많은 경우 장남/장녀를 연결시키는 것을 원칙으로 하고, 장남/장녀가 없는 경우 그 다음 번 자녀를 연결하는 방식으로 표본을 구축하였다.

6) 일단 자녀를 연결시킬 수 있는 개인들은 5,432명으로 이에 대한 총 19,426개의 관측치가 선별되었으며, 이 중 가구소득이 전무한 2개의 관측치와 흡연, 음주, 신체활동에 대한 정보가 없는 12개의 관측치를 배제하였다.

7) 한국은행 소비자물가지수통계를 이용하여 2015년 가격을 100으로 하여 전환하였다.

8) 규모의 경제로 인하여 의료비지출을 포함한 소비에 있어서 가구 내 니즈는 가구의 규모가 커지면서 이에 비례하여 증가하지 않는다. 이러한 특징을 반영하기 위하여 가구원수에 대해 조정을 한 가구원수 보정 가구소득을 이용한다. 이러한 가구원수 보정 소득을 구성하는 방법에는 가구원수의 제곱근으로 나누는 방법 외에 몇 가지가 더 있다. 이에 대한 포괄적이고 자세한 논의는 d'Ercole과 Förster(2012)를 참조.

판단하여 이 변수를 구성하였다. 표본의 대략 12%정도가 고위험흡연자인 것으로 나타난다. drink2는 개인이 고위험음주자인지를 나타내는 더미변수로서 고위험음주자이면 1 아니면 0의 값을 가진다. 고위험음주자는 질병관리본부의 정의를 따라서 1회 평균 음주량이 7잔 이상(여성 5잔)이며 주 2회 이상 음주하는 자로 규정한다.

표 1. 변수들의 기술통계량(2009-2015 통합데이터)

변수	전체표본		의료비지출이 0인 표본	의료비지출이 양의 값인 표본	비고
	평균	표준편차	평균	평균	
rmexp	643,498	1,308,217			실질 개인의료비지출액(원)
rmexp>0	721,390	1,364,701			실질 개인의료비지출액이 0이 아닌 경우
BMI	23.50	2.88	23.31	23.53	몸무게/((신장/100)) <sup>2</sup>
dobese2	0.268		0.237	0.271	BMI>=25 기준그룹: 비만아님
rei	2,555.48	1,759.14	2,490.60	2,563.33	가구원수 보정 실질 연간가구소득(만원)
age	56.24	8.79	53.61	56.55	만 연령
dgen2	0.528		0.328	0.553	여성더미 기준그룹: 남성
dms2	0.104		0.073	0.108	결혼상태(이혼/사별/별거/미혼) 기준그룹: 결혼상태(혼인 중)
ded2	0.214		0.295	0.204	전문대학교육 이상 기준그룹: 고등학교졸업 이하
decon2	0.322		0.195	0.338	경제활동 하지 않음 기준그룹: 경제활동 하고 있음
dnhis2	0.031		0.025	0.032	의료급여 등 기준그룹: 국민건강보험 가입
dsmok2	0.123		0.258	0.106	고위험흡연 기준그룹: 고위험흡연 하지 않음
drink2	0.122		0.208	0.111	고위험음주 기준그룹: 고위험음주 하지 않음
dexer2	0.545		0.551	0.544	신체활동 미흡 기준그룹: 규칙적 신체활동
cbmi	21.97		21.52	22.03	자녀의 BMI
dcobe2	0.156		0.110	0.161	자녀의 비만여부 기준그룹: 자녀 비만아님

주: 범주변수의 평균값은 특정범주가 차지하는 %이며, 연속변수의 경우는 표준편차 값도 주어짐.

이 경우도 흡연과 마찬가지로 단순소량음주가 건강에 미치는 영향이 불투명하다는 일부 실증연구에 근거하여 구성된 변수이다. 음주의 경우에도 12%가 고위험음주자인 것으로 나타난다. dexter2는 신체활동 실천여부를 나타내는 더미변수이다. 이 변수의 구성도 음주의 경우와 마찬가지로 질병관리본부의 정의를 따라 격렬한 신체활동, 중등도 신체활동, 걷기 실천 중 한 가지라도 수행하면 0의 값을 가지며 그렇지 않으면 1의 값을 가진다. 여기서 격렬한 신체활동이란 최근 1주일 동안 평소보다 몸이 매우 힘들거나 숨이 많이 가쁜 격렬한 신체활동을 1회 10분 이상, 1일 총 20분 이상, 주 3일 이상 실천한 것이다. 중등도 신체활동은 최근 1주일 동안 평소보다 몸이 조금 힘들거나 숨이 약간 가쁜 중등도 신체활동을 1회 10분 이상, 1일 총 30분 이상, 주 5일 이상 실천한 것이다. 마지막으로 걷기는 최근 1주일 동안 걷기를 1회 10분 이상, 1일 총 30분 이상 주 5일 이상 실천한 것을 의미한다. 표본의 55%가 규칙적인 신체활동을 하지 않는 것으로 나타난다.

<표 1>의 4열과 5열은 전체표본을 의료비지출이 0인 그룹과 의료비지출이 양의 값을 갖는 그룹으로 나누어 각각의 기술통계량을 보여준다. 이를 통해 보면 의료비지출이 양의 값을 갖는 경우 체질량지수의 값과 비만도가 그렇지 않은 경우보다 더 높으며, 자녀의 체질량지수와 비만도에서도 비슷한 현상이 나타난다.

## 나. 개인의료비지출과 비만

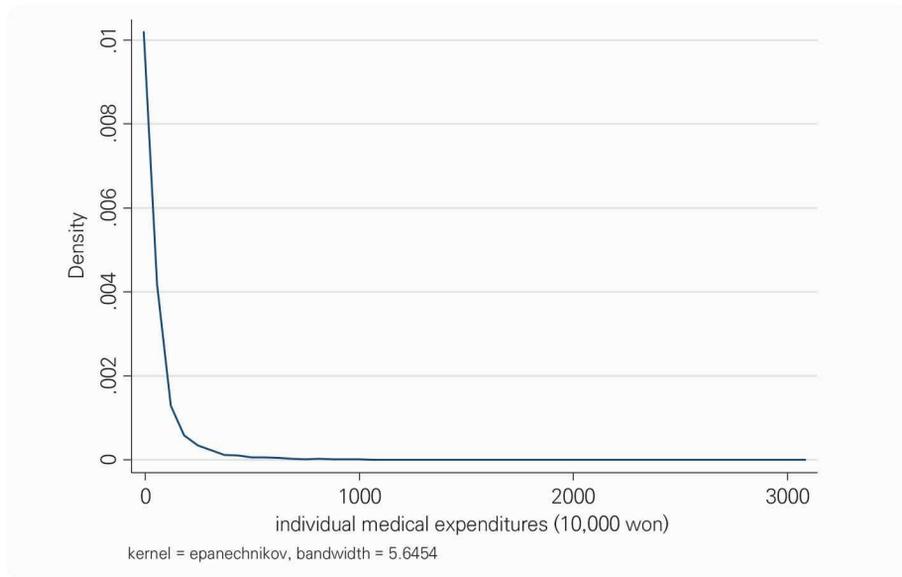
본 연구에서 가장 중요한 변수는 개인의료비지출액<sup>9)</sup>, 체질량지수, 그리고 체질량지수를 통해 구축된 비만상태 변수이다. 기존의 연구들에서 빈번하게 지적되었듯이 개인의료비지출액은 일반적으로 오른쪽 꼬리분포를 가지며 많은 경우 의료비지출이 없는 경우도 있는데, 이러한 특징은 실증분석방법의 채택에서 중요한 역할을 한다.<sup>10)</sup> [그림 2]에서 보이듯이 한국의료패널의 개인의료비지출 항목도 상당히 긴 오른쪽 꼬리분포를 가지며 또한 동시에 상당히 많은 0의 관측치를 가진다. 양의 의료비지출만을 고려하여

9) 한국의료패널은 두 가지의 개인의료비지출액 정보를 가지고 있다. 개인의료비1은 응급, 외래, 입원과 관련된 의료비와 처방약값을 합산한 금액이며, 개인의료비2는 개인의료비1에 교통비를 합산한 금액이다. 본 연구에서는 개인의료비1을 이용한다. 개인의료비2를 사용하는 경우 개인의료비1을 사용하는 경우와 실증분석결과와 정성적인 특징은 같다.

10) 이에 대한 자세한 논의는 Mullahy(2009)를 참조.

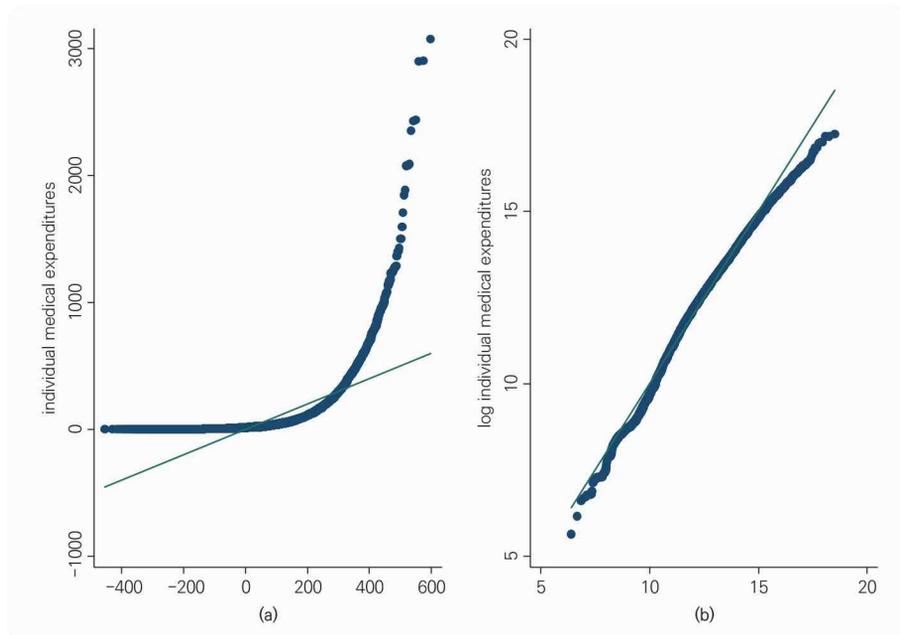
정규분포의 분위들에 대한 개인의료비지출액의 분위분표를 보면 [그림 3(a)]와 같고 개인의료비지출액을 로그값으로 전환하여 동일한 분표를 그려보면 [그림 3(b)]와 같다. 그림에서 보듯이 개인의료비지출이 양의 값을 갖는 경우 원 데이터는 전혀 정규분포를 따르지 않지만 로그변환을 하는 경우 거의 정규분포를 따르는 것을 알 수 있다.<sup>11)</sup>

그림 2. 개인의료비지출액 분포



11) 분표가 직선에 가까울수록 데이터는 정규분포를 따른다. 이 그래프는 Stata 명령어 *qnorm*을 통해 만들어졌다. 이 방법론은 Miller(1997)가 강하게 제안한 것으로 그는 이 방법을 probit plotting이라고 명명했다. 그에 의하면, “육안으로 뚜렷하게 정규분포로부터의 벗어남이 포착되지 않으면, 우리는 분표의 정규성에 대해 걱정할 필요가 없다.” Miller는 정규분포에 대한 Kolmogorov - Smirnov 테스트나 그와 비슷한 테스트들이 실제로 표본이 정규성으로부터 이탈했는지에 대해 거의 말해주는 바가 없다고 주장했다.

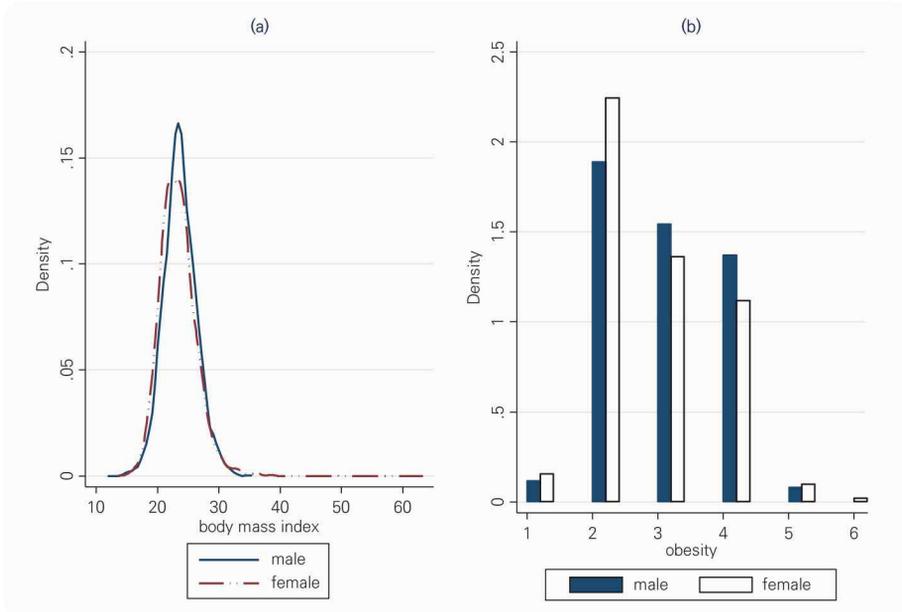
그림 3. 개인의료비지출액 분포(의료비지출)0



다음으로는 남성과 여성에 있어서 비만도의 분포를 살펴본다. [그림 4(a)]는 남성과 여성의 체질량지수 분포에 대한 kernel density 추정이다. 이를 통해 여성의 체질량지수 분포가 남성보다 약간 낮은 평균값을 가지지만 비만 쪽에 훨씬 극단치를 많이 가지는 것을 볼 수 있다. [그림 4(b)]는 대한비만학회(2018)가 체질량지수에 근거하여 제시하는 비만도의 6 범주별 남성과 여성의 히스토그램을 결합하여 제시한 것이다.<sup>12)</sup> 이를 통해 보면, 여성의 경우 저체중과 비만 쪽에 남성보다 많은 분포를 가지고 있는 것을 볼 수 있다.

12) 대한비만학회는 비만관련 세부척도를 저체중(BMI<18.5), 정상(18.5≤BMI<23), 비만전단계(23≤BMI<25), 1단계비만(25≤BMI<30), 2단계비만(30≤BMI<35), 3단계비만(BMI≥35) 등 6개의 범주로 구분하였다. 본 연구의 실증분석에서는 BMI 25를 기준으로 '비만'과 '비만이남'으로 이산 변수를 구성하여 분석을 실행한다.

그림 4. 여성과 남성의 체질량지수 분포 및 비만도 히스토그램

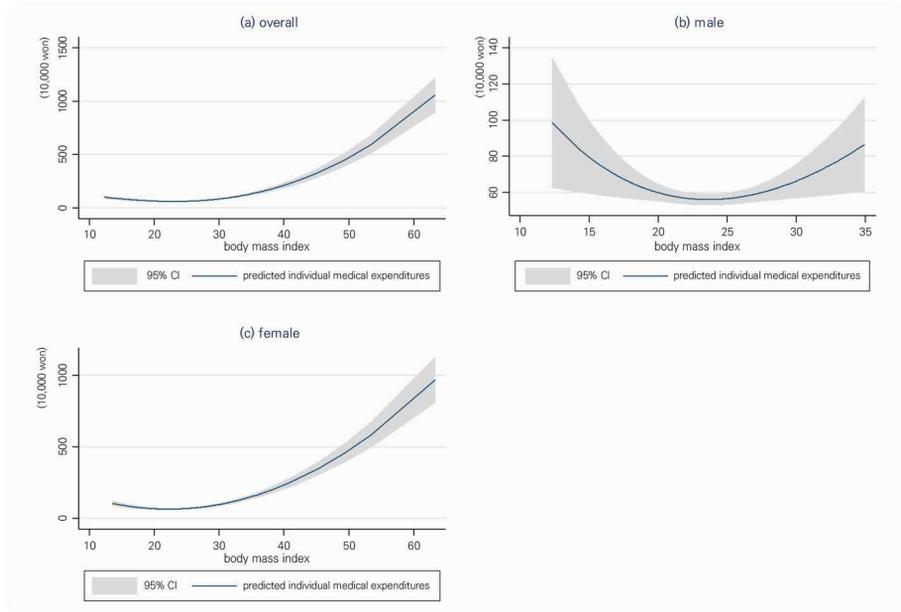


다음은 체질량지수와 이에 따른 개인의료비지출 간의 관계를 [그림 5]를 통해 살펴본다. 이는 fractional polynomial plot을 통해 두 변수 간의 관계를 추정한 것이다.<sup>13)</sup> 그림 (a)를 통해 나타나듯이 전체 개인의료비지출은 J-curve의 모습을 가진다. 즉, 저체중 그룹에서 정상체중보다 약간 많은 의료비지출 추이를 보이며 비만이 진행되면서 의료비지출이 급격하게 상승하는 양상을 보인다. 하지만 이를 남성과 여성의 두 그룹으로 분해하면 상당히 다른 그림과 만나게 된다. 그림 (b)는 남성의 경우 체질량지수와 의료비지출 간의 관계를 추정한 것이며, 이 경우 U-curve의 모양을 가진다. 즉, 저체중 그룹이 비만 그룹 만큼이나 의료비지출에 많은 영향을 주는 것을 알 수 있다. 그림 (c)는 여성의 경우 두 변수 간의 관계를 추정한 것인데 전체표본의 경우와 마찬가지로 J-curve의 모양을 가지며, 결국 여성에 있어서 두 변수 간의 관계가 전체관계를 추동하는 것을 알 수 있다.<sup>14)</sup>

13) 이는 Stata 명령어 *fpfitci*를 통해 만들어졌으며, [그림 3]에서처럼 추정에 대한 신뢰구간도 표시해 준다.

14) 흥미로운 것은 이러한 관계의 양상이 선행연구인 Cawley와 Meyerhoefer(2012)에서 나타난 양상과

그림 5. 체질량지수와 개인의료비지출액 간의 관계에 대한 추정



[그림 4]와 [그림 5]에서 드러나듯이 여성과 남성은 체질량지수와 비만도의 분포에서 상당한 차이를 보이며, 더 나아가 체질량지수와 의료비지출 간의 관계에서도 상당한 차이를 노정한다. 그러므로 전체 표본을 가지고 실증분석을 수행함과 동시에 남성과 여성을 각기 따로 구분하여 분석을 실행한다.

## 2. 분석방법

### 가. Control function 방법과 Two Part 모형

체질량지수와 비만이 개인의료비지출에 미치는 영향을 분석하기 위해서는 데이터의 특징에 따른 올바른 실증분석모형을 채택하는 것이 가장 중요하다. 본 연구의 데이터는

---

유사하다는 것이다. 미국의 의료비패널조사(Medical Expenditure Panel Survey)를 이용한 그들의 연구에서도 여성의 경우는 두 변수 간의 관계에 대한 추정에서 J-curve의 모양을 나타내며, 남성은 U-curve의 모습을 나타낸다. 또한, 전체적으로는 J-curve의 모양을 가진다.

도전적인 과제를 제시하는 두 가지 특징을 가진다. 첫 번째 과제는 데이터가 무작위대조 연구(randomized controlled trial)가 아니라 관찰연구(observational study)를 통해 구축되었다는 사실에서 파생한다. 이 경우 연구자가 관찰할 수 없는 개인적인 이질성(unobserved individual heterogeneity)이 존재하며, 이러한 관찰되지 않는 이질성을 적절하게 통제하지 않으면 관심변수(체질량지수 혹은 비만여부)가 개인의료비지출에 미치는 영향에 대한 추정계수는 필연적으로 (양의 혹은 음의) 편의를 갖게 된다. 그간 미시계량경제학에서는 도구변수추정법이나 패널데이터모형의 추정을 통해 이러한 문제를 극복하고자 했다.<sup>15)</sup> 본 연구에서는 도구변수추정법과 동일하게 배제조건을 충족시키는 변수에 기반한 control function 방법을 통해 이 문제에 대처하고자 한다. 이를 위해 본 연구에서는 아래에서 자세하게 설명하는 바와 같이 그간 배제조건을 만족시키는 변수로 비만과 의료비지출 관련 연구에서 많이 이용되었던 자녀의 체질량지수와 비만여부를 이용한다.<sup>16)</sup> 하지만 기존의 도구변수추정법을 이용하지 않고 control function 방법을 이용하는데, 그 이유는 우리가 이용하고자 하는 probit 모형과 일반선형화모형과 같은 비선형모형에서 도구변수추정법을 이용하는 경우 추정계수가 일치성을 가지지 않는 것으로 나타나기 때문이다.<sup>17)</sup> 두 번째 과제는 이미 살펴본 바와 같이 본 연구의 종속변수인 개인의료비지출액이 상당한 수의 0의 관측치를 가지며, 또한 동시에 분포가 오른쪽으로 긴 꼬리를 가진다는 사실에서 기인한다. 이 경우 문제를 해결하기 위해 다양한 방식이 제기되었다. 최초에는 양의 의료비지출만 고려하여 이를 로그변환한 변수에 통상최소사승법을 적용하는 방식이 이용되었다. 하지만 원래의 종속변수인 로그변환하지 않은 의료비지출액에 대해 추정계수를 적용 해석하는데 있어 몇 가지 쉽지 않은 이슈가 제기되어<sup>18)</sup> 점차적으로 일반화선형모형을 이용하기 시작하였다. 이 두 방법은 모두 한 개 방정식의 체계이다. 하지만 점차적으로 상당한 수의 0의 관측치를 고려하기 위하여

15) 비만이 의료비지출에 미치는 영향과 관련하여 도구변수추정법에 기반한 대표적인 연구는 Cawley와 Meyerhoefer(2012), 패널데이터모형의 추정은 Mora 등(2015) 참조.

16) Cawley 등(2015)에서도 자녀의 체질량지수와 비만여부를 부모의 체질량지수와 비만여부에 대한 도구변수로 사용하고 있으며, Biener 등(2017)에서는 청년층에 대한 동일한 연구를 위해 부모의 체질량지수와 비만여부를 도구변수로 이용하였다.

17) 이에 대한 자세한 논의는 control function 방법에 대한 포괄적인 서베이 연구인 Wooldridge(2015) 참조. 보건계량경제학의 틀에서 control function 방법을 최초로 이용한 논문은 Terza 등(2008)이다. 이후 Garrido 등(2012)도 의료비지출에 대한 분석에 있어 control function 방법을 적용하였다.

18) 이는 재전환문제(retransformation problem)라고 불리는데, 이에 대한 자세한 논의를 위해서는 Jones(2011), Duan(1983), Manning(1998), 그리고 Mullahy(1998)를 참조.

이러한 의료비 미발생에 대한 행태의 분석까지 포함하는 two part 모형이 광범위하게 이용되었다.<sup>19)</sup> 본 연구는 이러한 두 가지 문제의 해결을 위하여 two part 모형을 이용하며, 의료비발생여부에 대한 모형인 probit 모형과 양의 의료비지출에 대한 분석모형인 GLM이 모두 비선형모형이므로 관찰되지 않는 이질성의 통제를 위하여 control function 방법을 적용한다. Control function 방법은 두 단계를 거쳐서 작동한다.

$$y_1 = x_1\alpha_1 + \beta_1y_2 + u_1 \quad (1)$$

$$y_2 = x_1\gamma_{21} + x_2\gamma_{22} + v_2 \quad (2)$$

$$u_1 = \delta_1v_2 + e_1 \quad (3)$$

여기서 (1)은 의료비지출액 수준의 추정방정식이며, (2)는 체질량지수(혹은 비만)에 대한 추정방정식이다. 설명의 편의를 위하여 두 방정식 모두 선형인 경우를 고려한다. 여기서 (1)의 오차항은 (3)과 같이 주어진다. (3)을 (1)에 대입하면 다음의 식을 얻는다.

$$y_1 = x_1\alpha_1 + \beta_1y_2 + \delta_1v_2 + e_1 \quad (4)$$

위와 같이 주어진 구조에서 1단계에서는 내생성을 갖는 변수인 체질량지수모형 (2)를 2단계 실증모형에서 사용된 변수들인  $x_1$ 과 배제조건을 만족하는 변수, 즉 자녀의 체질량지수인  $x_2$ 를 이용하여 추정한다. 이러한 추정 후에 우리는 다음과 같이 잔차를 계산할 수 있다.

$$\hat{v}_2 = y_2 - x_1\hat{\gamma}_{21} - x_2\hat{\gamma}_{22}$$

2단계 의료비지출모형 (4)에서는 의료비지출액에 대해  $x_1, y_2$  그리고 잔차  $\hat{v}_2$ 를 이용하여 추정한다.<sup>20)</sup> 여기서 (4)의 추정에 이용된  $v_2$ 의 추정치  $\hat{v}_2$ 를 control function이라

19) 이러한 일련의 과정에 대한 논의는 Basu와 Manning(2009)을 참조.

20) 일반적인 도구변수추정법과의 차이가 이 지점에서 확연하게 드러난다. 도구변수추정법은 일반적으로 삽입법(plug-in method)이라고 불리는데 1단계에서 배제조건을 만족시키는 변수를 이용하여  $y_2$ 의 예측치,  $\hat{y}_2$ 를 구한 후 이 예측치를 2단계의 식에서  $y_2$  대신 사용하기 때문이다. 이에 반해 control function 방법에서는 (1)의 오차항과 상관관계에 있을지도 모르는 관찰되지 않는 이질성이

고 부르는데, 이 변수가 내생성을 통제하기 때문에 붙여진 이름이다. 이렇게 누락변수로써 내생성을 나타내는 변수가 직접 본 추정방정식에 포함되므로 control function 방법은 관찰되지 않는 이질성의 존재와 그것의 방향에 대한 직접적인 검증방법도 제시한다는 장점을 가진다. 우리의 경우 (1)이 비선형모형(probit과 GLM)이므로 위와는 약간 다르게 모형이 복잡해지지만 기본 작동원리는 동일하다. 다만 추정계수를 구한 후 한계 효과를 다시 구해야 한다(이러한 과정에 대한 모든 세부적인 이론적 논의는 Wooldridge(2015) 참조).<sup>21)</sup> 마지막으로 논의할 사항은 비만여부와 같이 방정식 (2)의 종속변수가 이산변수인 경우의 control function의 형태와 관련한 문제이다. 이 경우에는 잔차를 그대로 이용하지 않고 일반화된 잔차(generalized residual)을 이용하며, 이는 다음과 같은 형태를 가진다.

$$gr = y_2 \lambda(x\hat{\gamma}) - (1 - y_2) \lambda(-x\hat{\gamma})$$

여기서  $\lambda = \phi(\cdot)/\Phi(\cdot)$ 는 잘 알려진 inverse Mills ratio이며,  $x\hat{\gamma}$ 는 비만여부에 대한 probit모형 (2)의 추정을 통한 예측치이다.

Two part 모형에서 일반화 선형모형은 연구자가 분포함수와 연결함수를 결정해야 한다. 분포함수의 결정을 위해서는 modified Park 검정을 이용한다. 이를 위해서 단일 GLM 의료비지출액 방정식을 감마반응분포와 로그연결함수를 이용하여 추정하고, 이를 통해 종속변수에 대한 예측값을 구한다. 다음으로는 실제 의료비지출액에서 이 예측값을 빼서 잔차를 구한 후 이를 제공한다. 그리고 마지막 단계에서 잔차를 종속변수로 하고 의료비지출액의 예측값을 설명변수로 하여 감마반응분포와 로그연결함수를 가지고 GLM을 추정한다. 여기서 구한 예측치의 추정계수를 통해 반응분포를 결정하는데 이 추정계수가 0이면 정규분포, 1이면 포아송분포, 2이면 감마분포, 3이면 역정규분포를 채택한다. 우리의 경우 추정계수는 1.65이며 검정결과 2인 것으로 판정되어 감마분포가 적합한 것으로 드러난다.<sup>22)</sup> 연결함수와 관련하여 의료비지출 분석에서는 일반적으로 로그연결함수를 이용하는데 앞에서 살펴본 대로 의료비지출을 로그변화하는 경우

---

$\hat{v}_2$ 의 형태로 들어가기 때문에 2단계에서  $y_2$ 를 그대로 이용한다.

21) 이는 Stata에서 *mlx* 명령어를 통해 구할 수 있다.

22) 이러한 modified Park 검정의 절차에 대해서는 Deb and Norton(2018) 참조.

정규성을 가지기 때문이다. 따라서 본 연구의 GLM에서는 감마분포함수와 로그연결함수를 이용한다. 이와 같은 실증분석모형을 통하여 본 연구는 전체 표본, 여성, 남성에 대하여 각각 분석을 실행한다.

#### 나. control function 분위수회귀

위 소절에서 주어진 방법론을 통한 추정 외에 본 연구는 분위수회귀도 실행한다. 잘 알려져 있듯이 분위수회귀는 기존의 분석방법이 가지는 한계를 극복하고 좀 더 풍부하게 문제를 분석하기 위해 고안되었다. 기존의 분석은 주로 조건부 평균을 추정하는데 집중되었다. 하지만 이는 종속변수인 의료비지출의 전체 분포에 우리가 관심을 가지는 변수인 체질량지수와 비만상태가 어떻게 다른 방식으로 영향을 미치는지를 보여주지 못하는 한계를 가진다. 특히, 정책효과의 분석 등에서 이러한 이질적인 분포적 효과는 실제적인 중요성을 가지며 더 세부적인 정책의 고안에 도움을 주는 경우가 많다. 따라서 본 연구에서는 위의 소절에서 주어진 주요 분석에 더해 양의 의료비지출에 대한 control function 분위수회귀를 통해 체질량지수와 비만여부가 개인의료비지출액의 분포에 다르게 영향을 미치는지 살펴본다.<sup>23)</sup>

### III. 연구결과

#### 1. 체질량지수가 개인의료비지출에 미치는 영향

<표 2>는 체질량지수가 개인의료비지출에 미치는 영향을 control function 방법과 two part 모형을 이용하여 분석한 결과를 정리한 것이다.<sup>24)</sup> 의료비발생여부에 대한 probit 모형추정의 결과 전체 표본의 경우 체질량지수의 증가가 의료비발생의 확률을

23) 의료비지출에 대한 분석에서 조건부평균 접근법을 넘어서려는 여러 가지 시도(분위수회귀 포함)에 대한 서베이는 Jones 등(2015) 참조. 보건경제분야에서 분위수회귀의 구체적인 적용의 예는 복약지속성을 분석한 Borah와 Basu(2013) 참조.

24) 모든 실증모형 추정은 통계패키지 STATA 13/SE를 통해 이루어졌다.

증가시키는 것으로 나타난다. 표본을 남성과 여성으로 구분하는 경우 여성의 경우에는 체질량지수의 증가가 의료비발생확률에 영향을 주지 않지만 남성의 경우는 체질량지수의 증가가 의료비발생확률에 양의 영향을 주는 것으로 나타난다. 일단 의료비가 발생하는 경우 의료비수준에 대해 분석하는 GLM추정의 경우는 반대의 현상이 나타난다. 이 추정에서 전체 표본의 경우 체질량지수의 증가는 의료비지출액을 증가시키는 것으로 나타난다. 즉, 다른 설명변수들을 모두 통제해도 [그림 4(a)]의 관계가 유지되는 것으로 나타난다. 하지만 남성과 여성의 경우로 구분하는 경우에 두 그룹 간에는 차이가 존재한다. 여성의 경우에는 체질량지수의 증가가 의료비지출액의 증가를 초래하지만 남성의 경우는 영향이 없는 것으로 나타난다. 이는 [그림 4(b)]와 [그림 4(c)]를 통해 볼 때 유추할 수 있는 사실이다. 즉, 체질량지수와 의료비지출액의 관계가 J-curve인 여성의 경우와 달리 U-curve의 관계를 가지는 남성의 경우는 체질량지수의 의료비지출에 대한 영향이 분포의 양 극단에서 서로 상쇄하는 것으로 추측된다. 의료비지출액만을 보는 경우 체질량지수의 영향은 주로 여성에 의해 초래되는 것으로 나타난다. 다음으로 주목할 추정계수는 control function 변수의 추정계수이다. 남성의 GLM 방정식을 제외한 모든 추정에서 이 추정계수는 음의 값을 가지며 통계적으로 유의한 것으로 나타난다. 이는 관찰되지 않는 개인적인 이질성이 존재함을 나타내며 종속변수에 대해 음의 영향을 미친다는 것을 의미한다. 이 경우 Cawley and Meyerhoefer(2012)에서 지적되었듯이 누락변수의 존재로 인해 체질량지수에 대한 추정계수가 과소추정될 가능성이 있다. 일반적으로 내생성으로 인한 추정편의는 다음과 같이 적을 수 있다.

$$bias = \delta_1 \frac{cov(y_2, v_2)}{var(y_2)}.$$

우리의 추정결과 (4)번 식에서  $\hat{v}_2$ 에 대한 추정계수  $\hat{\delta}_1$ 는 음의 값을 가진다. 분모의  $var(y_2)$ 는 항상 양의 값이므로 분자인  $cov(y_2, v_2)$ 가 양의 값인 경우 내생성을 가지는 설명변수의 추정계수는 과소추정된다. 한계효과를 보여주는 <표 3>을 보면 본 연구에서 control function을 이용하여 관찰되지 않는 이질성을 통제하지 않으면 실제 과소추정이 발생함을 알 수 있다. Cawley and Meyerhoefer(2012)에서는 이러한 상황의 한 예로 사회경제적으로 열위에 있는 사회집단이 비만이 될 확률이 높고, 경제적인 이유로 병원 이용이 원활하지 못한 경우를 제시하고 있다. 본 연구의 추정을 통해서 보면 체질량지수

의 1단위 증가는 연간 대략 5만 원 정도의 추가적인 의료비지출을 초래하는 것으로 나타난다.

아래에서는 기타 설명변수들의 추정계수에 대해 의료비지출에 대한 영향을 중심으로 살펴본다. 가구원수 보정 실질소득의 증가는 의료비지출을 증가시키는 것으로 나타나며, 연령의 증가도 의료비지출의 상승을 가져온다. 결혼상태와 교육수준은 여성의 의료비지출에만 영향을 미치는 것으로 나타난다. 혼인 중이 아니거나 고등학교졸업이하의 교육수준인 경우 의료비지출이 증가하는 것으로 나타난다. 경제활동이 없는 경우 전체적으로 의료비지출이 증가하며, 건강보험과 의료급여 가입은 별도의 영향을 미치지 않는 것으로 나타난다. 마지막으로 생활습관과 관련하여 흡연, 음주, 신체활동은 남성의 경우에만 의료비지출에 영향을 미치는 것으로 나타난다. 여기서 의아한 점은 규칙적 신체활동이 없는 경우는 의료비지출을 증가시키지만 흡연과 음주의 경우 의료비지출을 감소시킨다는 것이다. 이러한 직관과 일견 상충하는 현상에 대한 설득력 있는 설명이 최성은(2016)에서 주어진다. 이 연구에 따르면 국민건강보험자료 표본코호트 데이터를 통해 본 흡연자의 평균 의료비 지출은 2002~2013년 기간 동안 비흡연자의 평균 의료비 지출에 비해 낮았다. 이에 대해 이 연구는 흡연자라 하더라도 고령에 달하게 되면 금연을 하는 경우가 발생하고, 건강상의 문제가 발생하여 금연을 하는 경우도 발생할 수 있는 가능성을 제시한다. 또한 생애의료비 지출의 경우에도 흡연자의 기대여명이 비흡연자에 비해 짧을 수 있기 때문에 흡연자의 의료비 지출이 비흡연자에 비해 평균적으로 높지 않을 수도 있음을 지적한다. 본 연구에서 나타나는 흡연과 음주의 의료비지출에 대한 영향은 최성은(2016)과 비슷한 맥락에 있을 가능성이 있다고 판단되지만 좀 더 세밀한 연구가 추후에 진행되어야 할 것으로 생각한다.

표 2. 체질량지수가 개인의료비지출에 미치는 영향

probit	전체	여성	남성
BMI	0.089*** (0.023)	0.053 (0.036)	0.115*** (0.031)
log(rei)	0.108*** (0.023)	0.147*** (0.035)	0.075** (0.031)
age	0.025*** (0.002)	0.022*** (0.033)	0.028*** (0.002)
dgen2	0.458*** (0.034)	-	-
dms2	-0.064 (0.049)	-0.008 (0.064)	-0.100 (0.081)
ded2	-0.111*** (0.031)	-0.041 (0.055)	-0.134*** (0.039)
decon2	0.097*** (0.032)	0.115*** (0.041)	0.050 (0.054)
dnhis2	0.077 (0.078)	0.260** (0.130)	-0.040 (0.100)
dsmok2	-0.255*** (0.038)	-0.023 (0.342)	-0.250*** (0.040)
drink2	-0.082** (0.037)	-0.267** (0.132)	-0.070* (0.039)
dexer2	-0.027 (0.025)	-0.071* (0.040)	0.002 (0.033)
control	-0.070*** (0.598)	-0.098*** (0.791)	-2.023** (0.891)
GLM	전체	여성	남성
BMI	0.073*** (0.025)	0.089*** (0.033)	0.055 (0.038)
log(rei)	0.214*** (0.027)	0.215*** (0.034)	0.229*** (0.043)
age	0.020*** (0.002)	0.020*** (0.002)	0.019*** (0.003)
dgen2	0.019 (0.038)	-	-
dms2	0.107** (0.050)	0.101* (0.056)	0.095 (0.108)
ded2	-0.132*** (0.039)	-0.205*** (0.056)	-0.079 (0.054)
decon2	0.274*** (0.033)	0.229*** (0.038)	0.371*** (0.063)
dnhis2	-0.051 (0.084)	0.002 (0.110)	-0.148 (0.127)
dsmok2	-0.120*** (0.052)	-0.012 (0.323)	-0.129** (0.057)
drink2	-0.180*** (0.050)	-0.224 (0.154)	-0.161*** (0.056)
dexer2	0.072** (0.029)	0.023 (0.037)	0.124*** (0.044)
control	-0.057** (0.026)	-0.067** (0.034)	-0.053 (0.039)

주: 1) 지면절약을 위하여 년도 더미변수들과 상수항에 대한 추정계수의 보고는 생략하였다.  
2) \*\*\* $p < 0.01$ , \*\* $p < 0.05$ , and \* $p < 0.1$ .  
3) 괄호 안은 표준오차임.

표 3. 체질량지수가 개인의료비지출에 미치는 영향(한계효과)

(단위: 원)

	전체	여성	남성
BMI(control)	49,771***	63,608***	35,670
BMI	12,219***	17,408***	2,719

주: 1) \*\*\* $p < 0.01$ , \*\* $p < 0.05$ , and \* $p < 0.1$ .  
2) cotrol은 control function을 통해 내생성을 통제한 모형의 추정치임을 나타냄.

## 2. 비만여부가 개인의료비지출에 미치는 영향

<표 4>는 비만여부가 개인의료비지출에 미치는 영향을 control function 방법과 two part 모형을 이용하여 분석한 결과를 정리한 것이다. 의료비발생여부에 대한 probit 추정의 결과 전체 표본의 경우 비만상태가 의료비발생의 확률을 증가시키는 것으로 나타난다. 표본을 남성과 여성으로 구분하는 경우 여성의 경우에는 체질량지수의 경우와 비슷하게 비만상태가 의료비발생확률에 영향을 주지 않지만 남성의 경우는 비만상태가

표 4. 비만여부가 개인의료비지출에 미치는 영향

probit	전체	여성	남성
dobese2	0.837*** (0.247)	-0.081 (0.379)	1.470*** (0.326)
log(rei)	0.120*** (0.023)	0.151*** (0.035)	0.093*** (0.031)
age	0.024*** (0.002)	0.021*** (0.003)	0.027*** (0.002)
dgen2	0.454*** (0.035)	-	-
dms2	-0.057 (0.049)	0.015 (0.063)	-0.113 (0.082)
ded2	-0.116*** (0.031)	-0.070 (0.055)	-0.127*** (0.038)
decon2	0.097*** (0.032)	0.115*** (0.041)	0.049 (0.054)
dnhis2	0.049 (0.078)	0.269** (0.131)	-0.087 (0.100)
dsmok2	-0.260*** (0.038)	-0.063 (0.342)	-0.234*** (0.040)
drink2	-0.102*** (0.039)	-0.253* (0.133)	-0.120*** (0.043)
dexer2	-0.039 (0.026)	-0.066 (0.040)	-0.021 (0.034)
gr	-0.429*** (0.147)	0.124 (0.224)	-0.811*** (0.196)
GLM	전체	여성	남성
dobese2	1.018*** (0.254)	1.276*** (0.343)	0.841** (0.380)
log(rei)	0.229*** (0.026)	0.233*** (0.033)	0.238*** (0.042)
age	0.019*** (0.002)	0.018*** (0.002)	0.018*** (0.003)
dgen2	0.036 (0.038)	-	-
dms2	0.105** (0.050)	0.100* (0.056)	0.086 (0.108)
ded2	-0.129*** (0.039)	-0.209*** (0.055)	-0.072 (0.054)
decon2	0.281*** (0.033)	0.242*** (0.038)	0.372*** (0.064)
dnhis2	-0.076 (0.083)	-0.016 (0.111)	-0.170 (0.127)
dsmok2	-0.106** (0.043)	-0.005 (0.324)	-0.114** (0.057)
drink2	-0.218*** (0.052)	-0.285* (0.155)	-0.194*** (0.059)
dexer2	0.057** (0.029)	0.008 (0.038)	0.108** (0.045)
gr	-0.555*** (0.249)	-0.682*** (0.203)	-0.482** (0.230)

주: 1) 지면절약을 위하여 년도 더미변수들과 상수항에 대한 추정계수의 보고는 생략하였다.

2) \*\*\* $p < 0.01$ , \*\* $p < 0.05$ , and \* $p < 0.1$ .

3) 괄호 안은 표준오차임.

의료비발생확률에 양의 영향을 주는 것으로 나타난다. 일단 의료비가 발생하는 경우를 분석하는 GLM추정의 경우는 체질량지수의 경우와 달리 비만상태가 남성표본과 여성표본 모두에서 의료비지출의 증가를 초래하는 것으로 나타난다. 또한 control function 변수의 추정계수가 모두 음의 값으로 통계학적으로 유의하게 나타나 관찰되지 않는 이질성을 통제하지 않는 경우 비만상태 변수는 내생성을 가져 추정계수에 편이가 발생할 가능성이 있음을 보여준다. 비만상태의 한계효과를 보여주는 <표 5>를 보면 체질량지수의 경우와 마찬가지로 실제 과소추정이 발생함을 볼 수 있다. 본 연구의 추정을 통해서 보면 비만상태는 비만상태가 아닌 경우와 대비하여 연간 대략 92만 원 정도의 추가적인 의료비지출을 초래하는 것으로 나타난다. 기타설명변수들의 의료비지출에 대한 영향은 체질량지수의 경우와 정성적으로 비슷한 효과를 가지는 것으로 나타나며, 따라서 별도의 설명은 생략한다.

표 5. 비만여부가 개인의료비지출에 미치는 영향(한계효과)

(단위: 원)

	전체	여성	남성
dobese2(control)	919,023***	1,340,808***	664,187**
dobese2	69,403***	99,158***	34,773

주: 1) \*\*\* $p < 0.01$ , \*\* $p < 0.05$ , and \* $p < 0.1$ .

2) cotrol은 control function을 통해 내생성을 통제한 모형의 추정치임을 나타냄.

### 3. 체질량지수와 비만여부가 개인의료비지출에 미치는 영향: 분위수회귀

<표 6>은 체질량지수와 비만여부가 개인의료비지출에 미치는 영향을 분위수회귀를 이용하여 분석한 결과를 정리한 것이다. 이를 보면 의료비지출액분포의 하단에서는 체질량지수와 비만상태의 추정계수가 상대적으로 작으며, 분포의 상단으로 올라갈수록 체질량지수의 90번째 백분위를 제외하고 추정계수가 단조증가하는 것으로 나타난다. 체질량지수와 비만상태는 적은 의료비를 지출하는 상대적으로 건강한 개인들의 의료비지출에는 작은 영향을 미치는 것으로 나타난다. 특히 비만상태의 경우 10번째 백분위와 30번째 백분위에서는 각각 연간 10만 원과 28만 원의 추가적인 의료비지출을 초래하지

만 분포의 상단에서는 상당한 금액의 지출을 초래하며 90번째 백분위에서는 거의 연간 2백만 원의 추가적인 의료비지출을 유발한다. 즉 의료비지출이 많은 집단에서 비만의 영향이 상당히 큰 것으로 드러난다.<sup>25)</sup>

양의 의료비지출에 대한 기타 설명변수의 영향을 살펴보면 가구원보정 실질가구소득과 연령은 분포의 모든 측면에 걸쳐서 통계적으로 유의하게 의료비지출에 양의 영향을 미치는 것으로 나타난다. 또한 이 변수들의 영향은 분포의 상단으로 올라갈수록 단조증가하며 영향을 미친다. 혼인 중이 아닌 결혼상태 변수는 건강이 좋지 않아 의료비지출이 큰 표본에만 영향을 미치는 것으로 나타난다. 교육수준과 경제활동유무도 각각 단조증가 단조감소적인 형태로 의료비지출에 영향을 미치는 것으로 나타난다. 전문대교육이상인 경우 분포의 상단으로 가면서 단조감소의 형태로 의료비지출에 영향을 미치며, 경제활동이 없는 경우 상단으로 가면서 단조증가의 형태로 영향을 미친다. 의료급여가입 등은 앞의 분석과 달리 국민건강보험가입과 비교하여 의료비지출에 단조감소적으로 영향을 미치는 것으로 나타난다. 이는 건강이 좋지 못한 그룹의 경우 국민건강보험에 비해 의료급여가입 등이 의료비지출에 상당한 부정적 영향을 받는 것을 의미한다. 생활습관 변수 중에서는 음주만이 통계적으로 유의미한 음의 값을 가지며, 이 변수도 분포의 상단으로 가면서 단조감소의 형태로 영향을 미치는 것으로 나타난다.

표 6. 체질량지수와 비만여부가 개인의료비지출에 미치는 영향(의료비지출>0): 분위수회귀

	$q(0.1 x)$	$q(0.3 x)$	$q(0.5 x)$	$q(0.7 x)$	$q(0.9 x)$
BMI	8,579*** (1,963)	25,778*** (4,109)	42,297*** (6,863)	63,152*** (14,040)	80,635 (54,326)
log(rei)	6,140*** (2,048)	25,894*** (4,288)	54,401*** (7,161)	116,318*** (14,650)	297,576*** (56,687)
age	1,988*** (146)	5,936*** (305)	9,717*** (509)	15,360*** (1,042)	26,214*** (4,031)
dgen2	13,787*** (2,939)	35,875*** (6,152)	52,891*** (10,275)	49,549** (21,020)	-91,448 (81,337)

25) 앞의 분석과 마찬가지로 분위수회귀에서도 여성과 남성을 따로 분석하였다. 이 경우 대부분의 결과가 <표 6>의 결과와 정성적으로 유사하다. 단, 남성의 경우 체질량지수가 의료비지출에 미치는 영향의 분석에서 70번째 백분위와 90번째 백분위에서 영향이 없는 것으로 드러나며, 이러한 결과가 전체표본에서 90번째 백분위에서 영향이 없는 것에 영향을 미친 것으로 파악된다. 남성과 여성의 별도분석은 지면관계 상 보고를 생략하였으며, 저자에게 요청 시 제공가능함.

	$q(0.1 x)$	$q(0.3 x)$	$q(0.5 x)$	$q(0.7 x)$	$q(0.9 x)$
dms2	-1,008 (3,902)	499 (8,168)	17,442 (13,641)	65,265** (27,905)	280,484*** (107,978)
ded2	-3,791 (2,992)	-15,889*** (6,262)	-29,676*** (10,459)	-83,074*** (21,397)	-254,018*** (82,794)
decon2	13,138*** (2,584)	48,855*** (5,409)	98,550*** (9,034)	183,368*** (18,481)	468,036*** (71,512)
dnhis2	-36,309*** (6,532)	-74,478*** (13,673)	-92,751*** (22,835)	-124,211*** (46,714)	-191,349 (180,759)
dsmok2	-4,136 (4,056)	-12,579 (8,489)	-14,464 (14,178)	5,027 (29,005)	-158,954 (112,233)
drink2	-4,577 (3,857)	-13,189 (8,073)	-37,144*** (13,483)	-107,691*** (27,582)	-339,256*** (106,731)
dexer2	336 (2,222)	184 (4,653)	3,962 (7,770)	15,257 (15,896)	158,456*** (61,510)
control	-6,541*** (2,002)	-19,516*** (4,190)	-32,026*** (6,998)	-50,192*** (14,315)	-70,530 (55,395)
dobese2	101,442*** (20,512)	282,004*** (39,531)	480,710*** (69,108)	745,807*** (149,316)	1,976,547*** (572,207)
log(rei)	7,177*** (2,129)	29,331*** (4,102)	61,938*** (7,171)	127,925*** (15,494)	301,281*** (59,378)
age	1,911*** (149)	5,764*** (288)	9,419*** (504)	14,749*** (1,088)	26,322*** (4,169)
dgen2	15,299*** (3,092)	38,816*** (5,959)	56,721*** (10,418)	58,358*** (22,508)	5,793 (86,257)
dms2	-727 (4,056)	212 (7,818)	20,864 (13,666)	54,474* (29,528)	221,618** (113,157)
ded2	-4,103 (3,103)	-17,628*** (5,981)	-29,559*** (10,456)	-79,954*** (22,591)	-200,205** (86,574)
decon2	12,515*** (2,698)	48,491*** (5,200)	97,512*** (9,090)	186,764*** (19,640)	458,831*** (75,265)
dnhis2	-39,944*** (6,826)	-81,875*** (13,155)	-105,929*** (22,997)	-142,904*** (49,688)	-197,799 (190,413)
dsmok2	-2,787 (4,235)	-10,438 (8,161)	-11,933 (14,268)	8,777 (30,825)	-40,100 (118,134)
drink2	-5,476 (4,153)	-20,681*** (8,004)	-48,892*** (13,993)	-129,731*** (30,233)	-451,181*** (115,858)
dexer2	-1,556 (2,362)	-3,125 (4,552)	-5,337 (7,957)	868 (17,193)	99,986 (65,885)
gr	-55,652*** (12,265)	-147,187*** (23,638)	-254,745*** (41,323)	-399,882*** (89,283)	-1,144,138*** (342,149)

주: 1) 지면절약을 위하여 년도 더미변수들과 상수항에 대한 추정계수의 보고는 생략하였음.

2) \*\*\* $p < 0.01$ , \*\* $p < 0.05$ , and \* $p < 0.1$ .

3) 괄호 안은 표준오차임.

## IV. 결론과 논의

본 연구에서는 관찰되지 않는 개인적인 이질성의 통제를 위한 control function 방법과 의료비지출변수의 특징을 잘 고려하는 two part 모형을 이용하여 우리나라에서 체질량지수와 비만상태가 개인의료비지출에 미치는 영향을 한국의료패널을 이용하여 분석하였다. 우리나라에서 비만이 의료비지출에 미치는 영향과 관련하여 본 논문은 도구변수를 이용하여 내생성을 통제하기 위해 노력한 최초의 논문이며, 도구변수를 이용한 control function 방법을 응용한 최초의 논문이기도 하다. 아래의 논의에서 서술되듯 비록 전체 모집단으로 연구결과를 확장해서 해석할 수는 없지만, 그간의 연구에서의 추정치가 편의를 가질 가능성이 매우 크다는 것을 보였으며 향후 비만과 의료비지출의 관계와 관련하여 의미 있는 정책적 논의의 출발점을 제시했다고 생각한다. 연구결과 전체표본에서 체질량지수 1단위의 증가는 약 5만 원의 추가적인 연간의료비지출을, 비만상태는 비만이 아닌 상태와 비교했을 때 92만 원의 추가적인 연간의료비지출을 유발하는 것으로 나타났다. 더 나아가 분위수회귀의 결과를 볼 때 체질량지수와 비만상태는 의료비지출액분포의 상단으로 올라갈수록, 즉 건강이 좋지 않아 의료비지출이 큰 집단에 더 큰 영향을 미치는 것으로 나타난다. 또한, 대부분의 결과에서 control function 변수에 대한 추정계수가 유의하게 음의 값을 가지며, 이는 내생성을 통제하지 않는 모형은 체질량지수와 비만여부 계수에 대한 과소추정을 유발하는 것을 의미한다.

아래에서는 본 연구의 한계와 추후연구의 방향을 논의하면서 글을 맺는다.

첫째, 본 연구는 도구변수추정법의 한계와 동일한 근본적인 한계를 가진다. 즉, 본 연구에서 이용한 표본은 전체 인구에서 확률적으로 임의 선별된 표본이 아니다. 본 연구의 표본은 한국의료패널이라는 확률적 표본에서 자녀의 신장과 몸무게 정보를 연결할 수 있는 일부 성인표본을 대상으로 분석을 했다. 따라서 본 연구의 결과는 전체 인구를 대상으로 확대적용 할 수 없으며, 18세 이상의 자녀를 둔 중장년 이상 부모표본에 국한된 결과라는 것을 명심해야 할 것이다.<sup>26)</sup>

둘째, 우리가 이용하는 체질량지수와 비만여부 변수는 모두 실제 계측이 아닌 서베이 응답자가 주관적으로 응답한 신장과 몸무게에 기반하여 구축되었다. 이러한 자가보고(self-reporting)된 변수는 잘 알려져 있다시피 일반적으로 측정오차(measurement

26) 도구변수추정법의 결과가 가지는 이러한 한계에 대한 논의는 Newhouse & McClellan(1998) 참조

error)를 지닌다. 측정오차가 고전적인 성격을 가지는 경우 추정계수에 대해 소위 희석 편향(attenuation bias)을 낳으며, 본 연구의 control function이 관찰되지 않는 이질성뿐만 아니라 이러한 고전적 측정오차를 적절하게 통제했다면 문제는 해결될 수 있었을 것이다. 하지만 O'Neill과 Sweetman(2013)이 제기하듯 측정오차가 비고전적인 성격을 가진다면 이러한 측정오차는 고전적 측정오차와 달리 내생변수의 추정계수에 과대추정을 초래한다. 만일 우리나라의 자가보고 변수들이 비고전적인 성격을 가진다면 본 연구의 control function은 이러한 비고전적 측정오차를 적절히 통제하지 못하고 추정계수는 과대추정 되었을 가능성을 배제할 수 없다. 이러한 문제들을 극복하기 위하여 최근에는 측정치와 보고치를 모두 가지고 있는 데이터를 이용하여 자가보고된 신장과 몸무게, 그리고 체질량지수를 교정하려는 연구들이 진행되고 있다. 예를 들어, Cawley(2004)는 실제 계측된 신장과 몸무게, 자가보고된 신장과 몸무게에 대한 모든 정보를 가지고 있는 미국 NHANES III 데이터를 이용하여 두 정보 간의 관계를 추정하고 이로부터 얻은 추정계수를 자가보고된 신장과 몸무게만 가지고 있는 데이터에 적용하여 측정오차를 교정하였다.<sup>27)</sup> 결국 이는 데이터의 문제로 귀착된다. 우리나라의 경우 실제 측정된 신장과 몸무게에 대한 정보를 가지고 있는 데이터는 국민건강영양조사(KNHANES)가 유일하다. 하지만 우리나라 국민건강영양조사에는 자가보고된 신장과 몸무게에 대한 정보가 없다.<sup>28)</sup> 향후 국민건강영양조사에서도 미국에서와 유사하게 명시적으로 본인의 신장과 몸무게를 자가보고하게 한다면 체질량지수와 비만상태 변수에 존재하는 측정오차를 제거할 수 있는 가능성이 생길 것이다.

셋째, 본 연구에서는 비만의 척도로 체질량지수를 이용했으나 그간 비만의 척도로써 이용되고 있는 체질량지수에 대한 많은 비판들이 있어 왔다. 체질량지수는 체지방과의 연관성은 좋으나 근육량이 많거나 신장이 작은 경우 정확하게 평가할 수 없고, 체지방분을 반영하지 못하여 복부비만을 선별할 수 없다는 단점이 있다(길주현 등(2010) 참조). 그러므로 향후에는 다양한 비만지표를 구성하여 연구를 하는 것이 바람직하다 하겠

---

27) 근래에는 Cawley(2004)의 방법론이 너무 강한 가정에 기반하고 있다는 문제의식 하에 자가보고된 신장과 몸무게의 수준변수 대신에 백분위 순위를 이용하여 오차를 교정하는 방법을 제시한 Courtemanche 등(2015)의 방법론이 많이 이용되고 있다. 두 방법의 비교평가는 Maclean과 Kessler(2015) 참조.

28) 비만관련 자가보고의 항목이 있지만 다음과 같은 형태이다. “현재 본인의 체형이 어떻다고 생각하십니까?” 이에 대한 응답은 매우 마른편이다/약간 마른 편이다/보통이다/약간 비만이다/매우 비만이다 가운데서 택일을 하게 되어 있다.

다.<sup>29)</sup> 현재 우리나라에서 다양한 비만의 척도를 구성할 수 있는 데이터는 국민건강영양 조사가 유일하다. 하지만 국민건강영양조사는 의료비에 대한 정보가 전무하며, 의료이용에 대한 정보도 상당히 제한적이다.<sup>30)</sup> 향후 비만관련 보건연구의 진행을 위해서는 국민건강영양조사에서 이 부분을 추가 조사할 필요가 있다고 생각한다.

마지막으로, 본 연구에서 사용한 개인의료비지출액은 수납금액 기준이며, 건보부담금은 별도로 알 수가 없는 한계를 가진다. 그러므로 비만이 사회 전체적으로 가지는 외부효과를 적절하게 추정할 수가 없다. 향후 한국의료패널의 조사과정에서 좀 더 철저한 추적이 있어야 한다고 생각한다. Bhattacharya와 Sood(2007, 2011)는 건강보험료의 책정에 있어 개인의 비만여부가 고려되지 않아 개인들이 비만을 유발하는 식습관 및 생활습관 등과 관련하여 총비용을 지불하지 않고 따라서 이를 통해 비만인 사람들이 그렇지 않은 사람들에 대해서 외부효과를 초래한다고 주장하였다. 특히 Bhattacharya와 Sood(2007)는 몸무게를 보험료계산에 반영하는 체제와 그렇지 않은 체제의 모형을 구성하고 미국의료비패널조사를 이용하여 외부효과를 추정하였다. 이들의 연구에 의하면 본인부담률이 17.5%인 경우 비만은 일인당 150달러의 후생비용을 창출하는 것으로 나타난다. 일부 연구자들은 이러한 연구결과를 토대로 비만자들을 대상으로 보험료를 높일 것을 주장한다. 하지만 비만이 사회경제적 요인들에 의하면 결정된다면 사회경제적 요인들의 결과일 수 있는 식습관 및 생활습관에 의해 결정되는 비만에 의해 건강보험료를 조정하는 것은 건강의 사회경제적 불평등을 더 확대시킬 수 있다는 것에 유념할 필요가 있다.<sup>31)</sup> 이 연구과제에 대한 실증분석이 향후 좀 더 세밀하게 이루어질 필요가 있다 하겠다.

29) 체질량지수 이외의 비만지표를 검토하는 논문으로는 2005년 소아청소년 신체 계측조사를 이용한 김주현 등(2010), 우리나라 국민건강영양조사를 이용한 Hong 등(2011), 그리고 미국의 NHANES III을 이용한 Burkhauser와 Cawley(2008) 참조. 다른 종류의 비만척도가 다른 정책적 함의를 가지는 것을 보여준 논문으로는 피하지방을 이용하여 비만척도를 구성한 후 이것과 체질량지수를 이용하는 경우 미국의 장기 비만추세가 달라짐을 보인 Burkhauser 등(2009) 참조.

30) 1년간 입원여부와 입원횟수, 2주간 외래이용여부와 횟수가 전부이다.

31) 사회경제적 요인 중 교육수준이 연령에 따라 비만에 미치는 영향을 분석한 연구로 백은정과 김진영(2013) 참조.

## 참고문헌

---

- 길주현, 이미나, 이해아, 박혜숙, 서정완. (2010). 한국 소아청소년 비만에서 허리둘레-신장비의 유용성. *대한소아소화기 영양학회지*, 13(2), pp.180-192.
- 대한비만학회. (2018). *비만진료지침 2018 요약본*. 서울: 동 기관.
- 백은정, 김진영. (2013). 한국 성인남녀의 교육 수준과 비만 간 관계. *보건교육·건강증진학회지*, 30(5), pp.91-100.
- 윤난희, 권순만. (2013). 비만이 의료이용과 의료비용에 미치는 영향. *보건경제와 정책연구*, 19(2), pp.61-80.
- 최성은. (2016). 흡연자의 의료비 지출과 흡연이 의료비 지출에 미치는 효과. *재정학연구*, 9(2), pp.1-21.
- Alley, D., Lloyd, J., Shaffer, T., & Stuart, B. (2012). Changes in the association between body mass index and medicare costs, 1997-2006. *Archives of Internal Medicine*, 172, pp.277-278.
- Basu, A., & Manning, W. (2009). Issues for the next generation of health care cost analyses. *Medical Care*, 47, pp.109-114.
- Bell, J. F., Zimmerman, F. J., Arterburn, D. E., & Maciejewski, M. L. (2011). Health-care expenditures of overweight and obese males and females in the medical expenditures panel survey by age cohort. *Obesity*, 19, pp.228-232.
- Bhattacharya, J., & Sood, N. (2007). Health insurance and obesity externality. *Advances in Health Economics and Health Services Research*, 17, pp.279-318.
- Bhattacharya, J., & Sood, N. (2011). Who pays for obesity?. *Journal of Economic Perspectives*, 25, pp.139-158.
- Biener, A., Cawley, J., & Meyerhoefer, C. (2017). *The medical care costs of youth obesity: An instrumental variables approach*. Cambridge: NBER.
- Borah, B., & Basu, A. (2013). Highlighting differences between conditional and unconditional quantile regression approaches through an application to assess medication adherence. *Health Economics*, 22, pp.1052-1070.

- Burkhauser, R. V., & Cawley, J. (2008). Beyond BMI: The value of more accurate measures of fatness and obesity in social science research. *Journal of Health Economics*, 27, pp.519-529.
- Burkhauser, R. V., Cawley, J., & Schmeiser, M. D. (2009). The timing of the rise in U.S. obesity varies with measure of fatness. *Economics and Human Biology*, 7, pp.307-318.
- Cai, L., Lubitz, J., Flegal, K. M., & Pamuk, E. R. (2010). The predicted effects of chronic obesity in middle age on medicare costs and mortality. *Medical Care*, 48, pp.510-517.
- Cawley, J. (2004). The impact of obesity on wages. *Journal of Human Resources*, 39, pp.451-474.
- Cawley, J., & Meyerhoefer, C. (2012). The medical care costs of obesity. *Journal of Health Economics*, 31, pp.219-230.
- Cawley, J., Meyerhoefer, C., Biener, A., Hammer, M., & Wintfeld, N. (2015). Savings in medical expenditures associated with reductions in body mass index among US adults with obesity, by diabetes status. *Pharmacoeconomics*, 33, pp.707-722.
- Courtemanche, C., Pinkston, J. C., & Stewart, J. (2015). Adjusting body mass index for measurement error with invalid validation data. *Economics and Human Biology*, 19, pp.275-293.
- d'Ercole, M. M., & Förster, M. (2012). The OECD approach to measuring income distribution and poverty: Strengths, limits and statistical issues. In Besharov D. J. and Couch K. A.(eds.), *European Measures of Income and Poverty: Lessons for the U.S.* (pp.27-58). New York: Oxford University Press.
- Deb, P., & Norton, E. C. (2018). Modeling health care expenditures and use. *Annual Review of Public Health*, 39, pp.489-505.
- Dixon, J. B. (2010). The effect of obesity on health outcomes. *Molecular and Cellular Endocrinology*, 316, pp.104-108.
- Duan, N. (1983). Smearing estimate: A nonparametric retransformation method.

- Journal of the American Statistical Association*, 78, pp.605-610.
- Finkelstein, E. A., Trogdon, J. G., Brown, D. S., Allaire, B. T., Dellea, P. S., & Kamal-Bahl, S. J. (2008). The lifetime medical cost burden of overweight and obesity: Implications for obesity prevention. *Obesity*, 16, pp.1843-1848.
- Finkelstein, E. A., Trogdon, J. G., Cohem, J. W., & Dietz, W. (2009). Annual medical spending attributable to obesity: Payer-and service-specific estimates. *Health Affairs*, 28, pp.822-831.
- Garrido, M., Deb, P., Burgess, J., & Penrod, J. (2012). Choosing models for health care cost analyses: Issues of nonlinearity and endogeneity. *Health Services Research*, 47, pp.2377-2397.
- Hong, S., Oh, H. J., Choi, H., Kim, J. G., Lim, S. K., Kim, E. K. et al. (2011). Characteristics of body fat, body fat percentage and other body composition for Koreans from KNHANES IV. *Journal of Korean Medical Science*, 26, pp.1599-1605.
- Jones, A. (2011). *Models for health care*. In Clements M. P. and Hendry D. F.(eds.), *The Oxford Handbook of Economic Forecasting*. (pp.625-654). New York: Oxford University Press.
- Jones, A., Lomas, J., & Rice, N. (2015). Healthcare cost regressions: Going beyond the mean to estimate the full distribution. *Health Economics*, 24, pp.1192-1212.
- Kang, J. H., Jeong, B. G., Cho, Y. G., Song, H. R., & Kim, K. A. (2011). Socioeconomic costs of overweight and obesity in Korean adults. *Journal of Korean Medical Science*, 26, pp.1533-1540.
- Kim, D., & Basu, A. (2016). Estimating the medical care costs of obesity in the United States: Systematic review, meta-analysis, and empirical analysis. *Value in Health*, 19, pp.602-613.
- Maclean, J. C., & Kessler, A. S. (2015). *Reporting error in weight and height among the elderly: Implications and recommendations for estimating healthcare costs*. Philadelphia: Temple University.
- Manning, W. (1998) The logged dependent variable, heteroscedasticity, and the

- retransformation problem. *Journal of Health Economics*, 17, pp.283-295.
- Miller, R. G. (1997). *Beyond ANOVA: Basics of Applied Statistics*. London: Chapman & Hall.
- Mora, T., Gil, J., & Sicras-Mainar, A. (2015). The influence of obesity and overweight on medical costs: A panel data perspective. *European Journal of Health Economics*, 16, pp.161-173.
- Mullahy, J. (1998). Much ado about two: Reconsidering retransformation and the two-part model in health econometrics. *Journal of Health Economics*, 17, pp.247-281.
- Mullahy, J. (2009). Econometric modeling of health care costs and expenditures: A survey of analytical issues and related policy considerations. *Medical Care*, 47, pp.104-108.
- Newhouse, J. P., & McClellan, M. (1998). Econometrics in outcomes research: The use of instrumental variables. *Annual Review of Public Health*, 19, pp.17-34.
- O'Neill, D., & Sweetman, O. (2103). The consequences of measurement error when estimating the impact of obesity on income. *IZA Journal of Labor Economics*, 2, pp.1-20.
- Onwudiwe, N. C., Stuart, B., Zuckerman, I. H., & Sorkin, J. D. (2011). Obesity and medicare expenditure: Accounting for age-related height loss. *Obesity*, 19, pp.204-211.
- Terza, J., Basu, A., & Rathouz, P. (2008). Two-stage residual inclusion estimation: Addressing endogeneity in health econometric modeling. *Journal of Health Economics*, 27, pp.531-543.
- Wolf, A. M., Finer, N., Allshouse, A. A., Pendergast, K. B., Sherrill, B. H., Caterson, I. et al. (2008). PROCEED: Prospective obesity cohort of economic evaluation and determinants: baseline health and healthcare utilization of the US sample. *Diabetes, Obesity and Metabolism*, 10, pp.1248-1260.
- Wooldridge, J. (2015). Control function methods in applied econometrics. *The Journal of Human Resources*, 50, pp.420-445.

Yang, Z., & Hall, A. G. (2008). The financial burden of overweight and obesity among elderly Americans: The dynamics of weight, longevity, and health care cost. *Health Services Research*, 43, pp.849-868.

이용우는 영국 UCL에서 경제학 박사학위를 받았으며, 현재 영남대학교 경제금융학부에서 부교수로 재직 중이다. 주요 관심분야는 보험경제, 보건경제, 노동경제이며, 현재 청년층의 건강과 지역 간 노동시장 이동, 건강이 사회이동성에 미치는 영향 등을 연구하고 있다.

(E-mail: leastsquares@yu.ac.kr)

# Estimating the Medical Costs of BMI and Obesity in Korea: A Control Function Approach

**Lee, Yong-Woo**

(Yeungnam University)

---

This study analyzed the effects of body mass index (BMI) and obesity status on individual medical expenditures in Korea, using the control function method to control for unobserved individual heterogeneity and the two-part model to consider the characteristics of medical expenditure variables. For this purpose, the Korea Health Panel (2009-2015) is utilized. The study found that an increase in one unit of BMI in the overall sample resulted in an additional annual medical expenditures of about 50,000 won, while being -obese, compared to being non-obese, caused an additional annual medical expenditure of 920,000 won. Furthermore, the results of the quantile regression show that BMI and being -obese have a greater effect on the upper tail of the health expenditure distribution. Also, in most of the results, the estimation coefficients for the control function variable are significantly negative, indicating that underestimation may occur if the endogeneity is not properly controlled.

---

**Keywords:** Body Mass Index and Obesity, Medical Expenditures, Control Function, Two Part Model