

# 분석 보고서

본 공모전은 국내외에 공개된 보건의료 빅데이터를 활용하여 데이터 분석 과정을 상세히 기술하고, 분석 결과를 재연 가능하도록 제시하여 빅데이터를 이용해서 분석하는 경험 공유하는 것을 목표로 합니다. 심사는 제출된 서류로만 진행되오니, 각 항목을 충실히 작성해주시기 바랍니다.

<b>1. 제목</b>			
경제활동인구의 만성질환 유무가 의료비 과부담에 미치는 영향: 국내 보건 빅데이터를 활용한 인과분석			
<b>2. 분석 주제</b>			
경제활동인구의 만성질환이 의료비 과부담에 미치는 인과적 효과 분석			
<b>3. 제출자 정보</b>			
<b>구분</b>	<b>성명</b>	<b>소속</b>	<b>직위</b>
책임자	강재서	이화여자대학교 인공지능대학 이지영교수님 연구실	대학생
공동연구자	신아영	이화여자대학교 인공지능대학 이지영교수님 연구실	대학생
공동연구자	양인경	이화여자대학교 인공지능대학 이지영교수님 연구실	대학생
공동연구자	유보현	이화여자대학교 인공지능대학 이지영교수님 연구실	대학생
공동연구자	주수현	이화여자대학교 인공지능대학 이지영교수님 연구실	대학생
<b>1) 책임자 연락처</b> - 이메일 : kkabool1235@naver.com - 전화번호(핸드폰) : 010-9088-0367			
<b>4. 분석 데이터</b>			

<p>1) 사용 데이터 : 한국의료패널 2019-2022년 연간데이터 (Version 2.3)</p> <p>2) 자료가 일반인에게 공개되어 누구든지 얻을 수 있는 자료입니까?</p> <p><input type="checkbox"/> 일반 공개 (누구나 곧바로 다운로드 가능)</p> <p><input type="checkbox"/> 신청 필요 (가입 또는 신청 후 다운로드 가능)</p> <p><input checked="" type="checkbox"/> 허가 필요 (심사 및 허가 절차를 거쳐 다운로드 가능)</p> <p><input type="checkbox"/> 비공개</p> <p>3) 데이터 출처 : <a href="https://www.khp.re.kr:444/web/data/data.do">https://www.khp.re.kr:444/web/data/data.do</a></p> <p>4) 데이터 특징 :</p> <p>첫째, 동일한 가구와 가구원을 장기간 추적 조사하는 종단 자료(longitudinal data)이다. 이는 특정 시점이 아닌 시간의 흐름에 따른 개인의 건강 상태, 의료 이용, 사회경제적 지위 등의 동태적 변화를 관찰하고 그 인과 관계를 분석할 수 있는 기반이 된다.</p> <p>둘째, 대한민국 전체 인구를 모집단으로 하여 설계된 전국 단위의 대표 표본이다. 따라서 이 자료에서 파생된 통계치는 높은 외적 타당도(external validity)를 가지며, 대한민국 인구 전체의 특성으로 일반화하여 해석될 수 있다.</p> <p>셋째, 개인 수준의 인구 사회학적 정보, 건강 행태, 질병 상태와 가구 수준의 소득, 지출, 자산 등 사회경제적 정보를 모두 포함하는 다층적(multi-level) 데이터 구조를 갖추고 있다. 또한, 상세한 의료 이용 및 의료비 지출 내역을 포함하고 있어 개인의 보건의료 행태를 가구의 사회경제적 맥락과 연계하는 다각적 분석을 가능하게 한다.</p>
<p><b>5. 공모제출 분석 내용 (해당하는 내용 모두 선택)</b></p>
<p>1) 데이터 특성 및 전처리 과정의 문제점 해결.</p> <p>[<input checked="" type="checkbox"/>] 데이터 수집, 저장, 불러오기</p> <p>[<input checked="" type="checkbox"/>] 데이터 정제 (Data Cleaning)</p> <p>[<input checked="" type="checkbox"/>] 데이터 변환 (Data Transformation)</p> <p>[<input checked="" type="checkbox"/>] 데이터 병합 또는 분할</p> <p>[ ] 다른 데이터와의 연계</p> <p>2) 데이터를 통한 새로운 인사이트 도출</p> <p>[ ] 새로운 시각화 생성</p> <p>[<input checked="" type="checkbox"/>] 새로운 분석 결과</p> <p>3) 분석법 적용</p> <p>[ ] LLM을 이용한 분석</p> <p>[<input checked="" type="checkbox"/>] 머신러닝을 이용한 분석</p> <p>[<input checked="" type="checkbox"/>] 기타 (분석법 : IPTW, PSM, CRE)</p> <p>4) 프로그램 개발</p> <p>해당사항 '없음'</p> <p>[ ] 대시보드 프로그램 개발</p> <p>[ ] 탐색적 데이터 분석 (Exploratory Data Analysis - EDA)</p>

<p><input type="checkbox"/> 리포트 작성 프로그램 개발</p> <p><input type="checkbox"/> 새로운 시각화 방법 개발</p> <p>5) 기타</p> <p>해당사항 '없음'</p> <p><input type="checkbox"/> 간단한 제목</p>
<p><b>6. 분석에 사용한 프로그램 (해당하는 내용 모두 선택)</b></p> <p>1) 프로그램</p> <p><input checked="" type="checkbox"/> R</p> <p><input checked="" type="checkbox"/> Python</p> <p><input type="checkbox"/> Excel</p> <p><input type="checkbox"/> SQL</p> <p><input type="checkbox"/> SAS</p> <p><input type="checkbox"/> SPSS</p> <p><input checked="" type="checkbox"/> 기타 ( 프로그램명 : Colab )</p> <p>2) 프로그램 버전</p> <p>R 4.5.1</p> <p>3) 같이 사용한 주요 외부 라이브러리 또는 모듈</p> <ul style="list-style-type: none"> <li>• WeightIt: 성향점수 기반의 가중치 추정을 위한 패키지)</li> <li>• cobalt: 공변량의 균형 상태를 시각화 · 진단하는 패키지. 매칭이나 가중치 적용 후의 공변량 균형을 쉽게 비교 가능</li> <li>• MatchIt: 성향점수 매칭(PSM)을 지원하는 대표적인 패키지. 최근접매칭, 칼리퍼 매칭, 서브클래스 매칭 등 다양한 매칭 방법 제공</li> <li>• readxl: Excel 파일을 불러오는 패키지. 데이터 불러오기용</li> <li>• dplyr: 데이터 전처리 · 가공에 특화된 패키지. filter, select, mutate, summarise 등 파이프라인 방식으로 직관적 데이터 처리 가능</li> <li>• ggplot2: 데이터 시각화 패키지. 통계 그래프, 분포도, 회귀선 등 다양한 시각화 가능</li> <li>• gridExtra: 여러 개의 ggplot 그래프를 하나의 화면에 배열해서 출력할 수 있게 해주는 패키지</li> <li>• writexl: 데이터프레임을 Excel 파일로 저장할 수 있게 해주는 패키지</li> <li>• grf: 기계학습 기반의 인과추론 패키지. Casual Forest, Instrumental Forest, Quantile Forest 등 다양한 임의효과 추정 기법 제공</li> <li>• CRE: Casual Random Forest 기반의 Conditional Randomization-based</li> </ul>

Estimator 관련 패키지. 이질적 처리 효과(HTE, 이질적 인과효과)를 추정하는 데 사용됨
<b>7. 소스코드 및 별첨 자료</b>
<p>1) 소스코드 제출 및 공개 여부</p> <p>[ ] 제출함. 공개 가능      [x] 제출함. 평가에만 이용. 비공개      [ ] 미제출</p> <p>2) 소스코드 링크 또는 첨부</p> <ul style="list-style-type: none"> <li>- Code.zip(추가로 제출)</li> <li>- final_dataset.xlsx</li> </ul> <p>3) 별첨 자료명</p> <p>1) 전처리</p> <ul style="list-style-type: none"> <li>- 데이터 전처리.ipynb</li> </ul> <p>2) TableOne</p> <ul style="list-style-type: none"> <li>- Table1_final_code.R</li> </ul> <p>3) 방법론 적용</p> <ul style="list-style-type: none"> <li>- CRE_final_code.R</li> <li>- PSM_final_code.R</li> <li>- IPTW_final_code.R</li> </ul>
<b>8. 요약</b>
<p>본 연구는 국내 보건 빅데이터를 활용하여 경제활동인구의 만성질환 유무가 가구 의료비 과부담에 미치는 영향을 인과적으로 규명하고자 하였다. 기존 연구는 만성질환과 의료비 지출 간의 단순한 연관성에 집중해 왔으나, 선택 편의(selection bias)로 인해 인과효과 추정에 한계가 존재하였다. 이에 본 연구는 성별, 연령, 교육 수준, 소득 등 사회경제적 요인을 통제된 상태에서 성향점수매칭(Propensity Score Matching, PSM), 역확률가중치(Inverse Probability of Treatment Weighting, IPTW), 그리고 Causal Rule Ensemble(CRE) 방법론을 적용하여 분석을 수행하였다.</p> <p>분석 대상은 경제활동인구 중 근로·사업소득이 존재하는 20세 이상 61세 이하인 인구로 한정하였다. 처리변수(treatment)는 만성질환 유무로 정의하였고, 결과변수(outcome)는 가구 의료비 지출 대비 소득 비율로 설정하였다. 결측치는 소득 및 지출 특성에 따라 0 대체 혹은 제외 처리하여 최종 표본을 구성하였다.</p>

분석 결과, 만성질환 보유 집단은 비보유 집단에 비해 의료비 과부담 위험이 유의하게 높았으며, 이는 PSM과 IPTW 분석에서 일관되게 확인되었다. CRE를 통한 하위집단별 분석 결과, 특히 여성, 지역 건강보험 세대주, 의료급여 1·2종 세대주 집단에서 만성질환으로 인한 과부담 위험이 더 크게 나타나는 것을 보아 이들 집단이 정책적 개입이 필요한 취약 집단임을 확인할 수 있었다.

이러한 결과는 만성질환 관리 정책이 건강 증진을 넘어 가계 재정 안정성 확보와 사회적 불평등 완화에 중요한 기여를 할 수 있음을 시사한다. 본 연구는 인과추론적 방법론을 적용하여 기존 연구의 한계를 보완하였으며, 향후 보건·경제 정책 수립에 실질적인 근거를 제공할 수 있을 것으로 기대된다.

**Keywords:** 만성질환, 의료비 과부담, 인과추론, PSM, IPTW, Causal Rule Ensemble

## 9. 본문

### I. 서론

만성질환은 개인의 건강뿐 아니라 가계 재정에도 중대한 영향을 미친다. 특히 만성질환 보유 가구는 의료비 지출이 장기적으로 증가하며, 이로 인해 소득 대비 과도한 의료비 부담(Catastrophic Health Expenditure, CHE)을 경험할 가능성이 높다. 국내 연구에서도 만성질환이 CHE 발생과 관련이 있음이 확인되었으나(조운민, 2015), 다수가 교차단면 분석에 기반하여 인과적 경로를 충분히 규명하지 못했다.

최근에는 한국의 패널 데이터를 활용한 보건경제 연구에서 준실험적 방법(예: 성향점수 매칭, 역확률가중법 등)을 적용해 인과추론을 강화하려는 시도가 증가하고 있다. 그러나 여전히 평균 처치 효과에 집중하는 경향이 강하며, 연령·소득·보험 보장성과 같은 사회경제적 요인에 따른 이질적 효과는 충분히 검토되지 않았다.

이에 본 연구는 국내 보건 빅데이터를 활용하여 경제활동인구를 대상으로 만성질환 유무가 의료비 과부담에 미치는 인과적 효과를 분석하고자 한다. 특히 준실험적 방법을 통해 평균 효과를 추정하는 동시에, 하위집단별 이질적 처치효과를 탐색함으로써 보다 구체적인 정책적 함의를 도출하고자 한다.

### II. 이론적 배경 및 선행연구 검토

#### 1. 이론적 배경

의료비 과부담(Catastrophic Health Expenditure, CHE)은 세계보건기구(WHO)에서 정의한 바와 같이, 가계의 지출 중 의료비가 일정 비율(예: 가처분소득 대비 10% 혹은 비식료 지출 대비 40%)을 초과하는 경우로 측정된다. 이는 가계의 재정적 파탄과 빈곤으로 이어질 수 있는 중요한 보건경제 지표이다. 만성질환은 지속적 치료와 약제 사용을 필요로 하기 때문에 CHE 발생의 주요 결정 요인으로 주목받아 왔다. 그러나 이러한 효과는 모든 집단에서 동일하게 나타나지 않으며, 연령, 소득, 건강보험 보장성 등 사회경제적 요인에 따라 달라질 수 있다는 점에서 이질적 처치 효과(HTE) 개념이 중요하다.

## 2. 선행연구 검토

국내외 연구는 만성질환과 의료비 과부담 간의 관련성을 반복적으로 확인해왔다(조운민, 2015). 그러나 다수의 연구는 단면자료와 회귀모형에 의존하여 인과적 해석이 제한적이었다. 최근에는 한국의 패널 데이터를 활용해 PSM이나 IPTW를 적용한 연구들이 등장하였으며(김동욱 외, 2025), 이는 공변량 불균형을 완화하여 평균 효과를 추정하는 데 기여하였다. 다만 이들 연구는 평균 효과 추정에 집중하여, 정책적 개입의 우선순위를 결정하는 데 필요한 하위집단별 정보는 충분히 제공하지 못했다는 한계가 있다. 국제적으로는 하위집단별 차이를 분석하기 위한 HTE 연구가 증가하고 있으며, 특히 Causal Rule Ensemble(CRE)과 같은 트리 기반 모형은 해석 가능한 규칙을 제공하여 정책 설계에 활용도가 높다. 그러나 국내 보건 의료 분야에서는 이러한 접근이 활발히 적용되지 않고 있다..

## 3. 연구의 차별성

따라서 본 연구는 IPTW와 PSM을 활용해 평균 인과효과를 검증하는 동시에, CRE를 적용해 하위집단별 이질적 처치효과를 규명하고자 한다. 이를 통해 기존 연구의 한계를 보완하고, 정책적으로 의미 있는 취약 집단을 식별하여 맞춤형 지원 방안을 제시하는 데 기여할 수 있다.

# III. 연구 방법

## 1. 분석 자료 및 대상

본 연구는 한국보건사회연구원과 국민건강보험공단이 주관하는 한국의료패널(Korea Health Panel, KHP) 2019-2022년 연간 데이터(Version 2.3)중 2020년 데이터를 활용하였다. 연구의 초점을 경제활동인구로 명확히 하기 위해, 20세 이상 61세 이하 성인 중 근로소득이나 사업소득이 존재하는 표본으로 제한한 3,605명을 최종 분석 표본으로 구성하였다. 이러한 표본 선정은 인과추론 분석에서 성별, 연령, 소득 수준 등 공변량을 통제하고, 편향 없는 처치 효과 추정을 가능하게 하는 것을 목적으로 한다.

## 2. 주요 변수 정의

처치 변수(Treatment Variable)는 만성질환 보유 여부(CD\_total)로 정의하였다. 조사 대상자가 고혈압, 당뇨병, 고지혈증 등 만성질환 중 하나 이상을 진단받은 경험이 있는 경우 '1'(처치 집단), 그렇지 않은 경우 '0'(통제 집단)으로 설정하였다. 이를 통해 만성질환 유무가 의료비 과부담에 미치는 인과적 영향을 평가할 수 있도록 하였다.

결과 변수(Outcome Variable)는 개인 의료비 과부담(overburden\_yn)으로 설정하였다. 세계보건기구(WHO)의 정의를 참고하여 정의하였다. 본 연구는 연구의 초점을 경제활동인구 개인에 맞추고자, 개인 연간 총소득(근로소득+사업소득) 대비 본인부담의료비(MED\_EXP) 지출이 10% 이상인 경우 '1', 그렇지 않은 경우 '0'으로 측정하였다. 이는 가구 전체의 부담을 넘어 개인이 경제활동을 통해 벌어들인 소득으로 의료비를 감당하는데 있어 과도한 부담을 경험하는지를 이진 변수 형태로 평가하는데 의의가 있다.

공변량(Covariates)은 인과추론의 핵심 가정인 조건부 독립 가정(Conditional Independence Assumption)을 충족시키고, 선택 편의를 통제하기 위해 선행연구를 바탕으로 총 15개의 잠재적 교란 변수를 선정하였다. 주요 공변량은 다음과 같이 조작적 정의를 적용하였다.

(1) 이진 변수(Binary variable): 성별(SEX, 남성=1), 혼인 여부(MARR, 기혼=1), 교육 수준(EDU, 고등학교 졸업 이상=1), 스트레스 인지 정도(HS1, 인지함=1), 흡연 여부(S3, 현재 흡연함=1), 음주 빈도(D1, 한달 두 번 이상=1), 경제활동 참여 상태(ECO1, 취업자=1), 장애 유무(DISA\_YN, 장애=1), 민간의료보험 가입 여부(I\_PHI1\_YN, 가입=1), 우울감 유무(HS2\_YN, 우울=1), 규칙적 운동 여부(P1, 실천=1).

(2) 다중 범주 변수(Multicategory → Dummy variable): 의료보장 형태(HEALTH\_INS)는 각 범주를 개별 가변수(dummy variable)로 변환하여 분석에 포함하였으며, 다중공선성 방지를 위해 첫 번째 범주를 기준으로 설정하였다.

(3) 연속형 변수(Continuous variable): 연령(AGE), 체질량지수(BMI), 소득(INCOME)은 연속형 변수로 활용하였다. 이 중 연령과 체질량지수는 다음과 같이 산출하여 변수로 활용하였다.

- 연령(AGE): 출생 연도(BIRTH\_Y)를 이용하여 세는나이를 계산하였다. 기준 연도는 자료가 수집된 2020년으로 설정하였다.

$$AGE = 2020 - BIRTH_{year} + 1$$

- 체질량지수(BMI): 체중(WT, kg)을 신장(HT, m)의 제곱으로 나누어 산출하였다.

$$BMI(kg/m^2) = \frac{WT(kg)}{(HT(m))^2}$$

- 소득(INCOME): 근로소득(I\_INC1)과 사업소득(I\_INC2)의 합으로 산출하였다.

$$INCOME = INCOME_{earn} + INCOME_{business}$$

### 3. 결측치 처리

본 연구는 분석 데이터의 정확성과 신뢰도를 확보하기 위해 다음과 같은 기준에 따라 결측치를 처리하였다.

- 주요 소득 변수인 근로소득과 사업소득에서 결측치는 소득이 발생하지 않은 것으로 정의가 되어 있으므로 0으로 대체하였다. 분석의 정확성을 위해 근로소득과 사업소득의 합이 음수(-)인 경우와 소득 중 하나라도 응답이 없는 경우(-9)는 분석 대상에서 제외하였다.
- 의료비 지출(MED\_EXP) 항목의 결측치는 해당 기간 의료 비용이 없는 것으로 정의가 되어 있어 0으로 대체하였다. 분석의 일관성을 위해 비논리적인 음수(-)값이나 '모름/무응답'으로 코딩된 값(-9)은 분석 대상에서 제외하였다.
- 연구를 위해 필요한 공변량에 '모름/무응답'으로 코딩된 값(-9)이 존재하는 표본은 분석의 완전성을 위해 분석 대상에서 제외하였다.
- 연구의 핵심 변수인 처리 변수(만성질환 보유 여부)와 결과 변수(의료비 과부담)에 '모름/무응답'으로 코딩된 값(-9)이 존재하는 표본은 분석의 완전성을 위해 분석 대상에서 제외하였다.

이러한 데이터 전처리 과정을 통해 분석 표본의 품질을 확보하였으며, 이는 후속 분석의 결과 해석에 대한 타당성을 높이는 데 기여하였다.

### 4. 분석의 주요 가정

본 연구에서 사용하는 IPTW, PSM, CRE 등 성향점수 기반의 인과추론 방법론은 그 타당성을 위해 다음과 같은 세 가지 핵심 가정을 전제한다. 이에 본 연구는 연구 설계 및 데이터 분석 과정에서 해당 가정들을 최대한 만족시키고자 하였다.

#### (1) 조건부 독립 가정 (Conditional Independence Assumption, CIA)

조건부 독립 가정은 처리 여부가 관측되지 않은 잠재적 결과와는 독립적이며, 공변량을 통제했을 때만 처리 집단과 통제 집단 간의 차이가 무작위적이라고 볼 수 있다는 전제를 의미한다. 본 연구에서는 선행 연구와 이론적 배경을 바탕으로 만성질환 발생과 의료비 과부담에 동시에 영향을 줄 수 있는 주요 공변량들을 모형에 포함하여 통제하였다. 이를 통해 관측 가능한 요인에 기초한 선택 편향(selection bias)을 최소화하고, CIA가 합리적으로 성립된다는 전제하에 분석을 수행하였다. 다만 관측되지 않은 교란변수의 존재 가능성을 완전히 배제할 수는 없다는 점에서 이는 본 연구의 한계로 남는다.

#### (2) 공통 지지 영역 가정 (Common Support / Overlap)



공통 지지 영역 가정은 처치 집단과 통제 집단의 성향 점수 분포가 충분히 겹쳐, 비교 가능한 영역이 존재해야 함을 의미한다. 본 연구에서는 로지스틱 회귀를 통해 추정된 성향 점수의 분포를 시각적으로 검토한 결과, 두 집단 간의 분포가 일정 부분 겹치는 공통 지지 영역(Propensity Score: 0.033~0.915)이 확인되었다. 이에 따라 해당 영역 내 표본만을 분석에 활용하여 극단적인 표본의 영향을 배제하였으며, 이로써 공통 지지 영역 가정이 충족되었음을 확인하였다.

### (3) 안정적 단위 처리 값 가정 (Stable Unit Treatment Value Assumption, SUTVA)

SUTVA는 각 개인의 결과가 자신의 처치 여부에만 의존하고, 다른 개인의 처치 여부에 영향을 받지 않아야 한다는 것을 전제로 한다. 본 연구는 개인 단위 패널 자료를 활용하였으며, 개별 응답자의 만성질환 여부가 무관한 다른 개인의 의료비 과부담에 직접적인 영향을 줄 가능성은 낮다고 판단하였다. 다만 동일 가구 내에서 여러 구성원이 동시에 표본에 포함될 수 있어 가구 내 간섭의 가능성을 완전히 배제하기는 어렵다. 본 연구에서는 이러한 한계를 인지하고 있으나, 분석 단위를 개인으로 설정하여 가구 구성원 간 상호작용이 결과에 미치는 영향은 제한적일 것으로 판단하였다. 따라서 본 연구는 SUTVA가 대체로 충족된다는 전제 하에 분석을 수행하였다.

## 5. 분석 방법

### 5.1. 분석 전략 개요

본 연구는 만성질환 유무가 의료비 과부담 발생 여부에 미치는 인과적 효과를 다각도로 검증하기 위해 단계적 분석 전략을 적용하였다. 먼저, 평균 처치 효과(ATE)를 추정하기 위해 성향점수 매칭(Propensity Score Matching, PSM)과 가중치 기법(Inverse Probability of Treatment Weighting, IPTW)을 활용하였다. 이는 표본 내 공변량 분포를 균형화함으로써, 만성질환이 전체 인구에서 의료비 과부담에 미치는 평균적 영향을 확인하기 위함이다.

이후, 하위집단별 이질적 처치 효과(HTE)를 탐색하기 위해 두 가지 상보적 접근법을 사용하였다. 첫째, 로지스틱 회귀 모형을 적용하여 사전에 고려된 주요 인구학적·사회경제적 요인에 따른 효과 차이를 추정하였다. 이는 전통적 회귀 기반 방법을 통해 가설에 따른 집단 간 차이를 확인하려는 목적이다. 둘째, 인과 규칙 앙상블(Causal Rule Ensemble, CRE)을 활용하여 데이터 기반 탐색을 수행하였다. CRE는 무작위 포레스트와 LASSO 회귀를 결합한 방법으로, 사전 가정에 의존하지 않고도 통계적으로 유의한 규칙을 도출하여 정책적으로 의미 있는 취약 집단을 식별할 수 있다.

따라서 본 연구는 (1) 전체 평균 효과(ATE) → (2) 가설 기반 하위집단 효과(HTE) → (3) 데이터 기반 탐색적 하위집단 효과(HTE)라는 분석 구조를 통하여 만성질환이 의료비 과부담에 미치는 인과적 영향을 다층적으로 규명하고자 한다.

## 5.2. 1단계: 평균 처치 효과(ATE) 추정

만성질환자와 비질환자 간의 관찰 가능한 특성 차이를 보정하여 만성질환의 평균적인 효과를 추정하고자 하였다.

### 5.2.1. 역확률 가중치(IPTW)를 이용한 처치효과 분석

본 연구는 만성질환 보유 여부가 개인의 의료비 과부담 발생에 미치는 평균 처치 효과를 추정하기 위해 Inverse Probability of Treatment Weighting(IPTW) 방법을 적용하였다. IPTW는 성향점수를 기반으로 각 개체에 가중치를 부여하여 관찰 연구에서 발생하는 선택 편향을 통제하고 처치군과 대조군 간 공변량 분포를 균형화함으로써, 무작위 배정 임상시험의 환경을 근사할 수 있도록 설계된 방법론이다.

#### (1) 성향점수 추정 (Propensity Score Estimation)

처치 변수인 만성질환 보유 여부( $T$ )를 종속 변수로 하고, 선행 연구와 도메인 지식을 바탕으로 잠재적 교란 변수로 작용할 수 있는 인구 사회학적·경제적·건강 관련 공변량( $X$ )를 독립변수로 포함하여 로지스틱 회귀 모형을 적합하였다. 구체적으로 성별, 연령, 교육 수준, 결혼 상태, 건강보험 및 민간보험 가입 여부, 체질량지수(BMI) 등의 변수가 포함되었다. 각 개체의 성향점수 ( $PS_i = P(T_i = 1 | X_i)$ )는 해당 모형으로부터 산출되었으며, 해당 개체가 처치군에 속할 확률을 의미한다.

#### (2) 역확률 가중치 산출

산출된 성향점수를 이용해 역확률 가중치를 정의하였다. 구체적으로 만성질환이 있는 집단(처치군)에 속한 개인의 가중치는 해당 개인이 처치군에 속할 확률(성향점수)의 역수로 계산된다. 반대로, 만성질환이 없는 집단(대조군)에 속한 개인의 가중치는 1에서 성향점수를 뺀 값의 역수로 계산된다. 즉, 성향점수가 높을수록 처치군에서는 낮은 가중치가, 대조군에서는 높은 가중치가 부여되며, 이를 통해 두 집단 간 공변량 분포가 균형을 이룰 수 있도록 조정한다.

#### (3) 균형성 검증

가중치 적용의 타당성을 검증하기 위해 IPTW 적용 전후의 공변량 균형성을 평가하였다. 공변량 균형성 평가는 단순히 집단 간 평균의 통계적 유의성 차이를 검정하는 것보다, 표본 크기에 의존하지 않고 집단 간 차이를 정량적으로 보여주는 효과 크기 지표를 활용하는 것이 바람직하다. 이에 본 연구에서는 각 공변량에 대해 표준화 평균 차이(Standardized Mean Difference, SMD)를 산출하였다. SMD는 처치군과 대조군의 평균 차이를 두 집단의 표준편차로 나누어 산출함으로써, 변수의 측정 단위와 분산의 크기와 관계없이 비교 가능하다. 일반적으로  $|SMD| < 0.1$ 일 때 두 집단의 공변량 분포가 충분히 균형을 이루었다고 간주한다. 또한 균형성 평가 과정에서는 러브 플롯(Love plot)을 통해 각 공변량의 SMD 변화를 시각적으로 확인하였으며, 성향점수 밀도 그림(density plot)을 통해 IPTW 적용 전후 두 집단의 분포 중첩 정도를 점검하였다.

#### (4) 평균 처치효과(ATE) 추정

최종적으로 IPTW 가중치를 반영한 가중 로지스틱 회귀 모형을 통해 만성질환 보유 여

부가 의료비 과부담 발생 확률에 미치는 평균 처치 효과(ATE)를 추정하였다. IPTW는 공변량 분포를 균형화하여 마치 무작위 배정과 유사한 비교 집단을 구성하므로, 이 방법을 통해 산출된 추정치는 관찰연구임에도 불구하고 인과적 해석에 근접한 의미를 가진다.

구체적으로, 본 연구에서 추정된 ATE는 “만성질환이 없었더라면 해당 개인이 의료비 과부담을 경험했을 확률”이라는 반사실적(counterfactual) 확률을 추론할 수 있게 한다. 이는 단순한 상관관계 분석이 아닌, 처치 변수의 효과를 인과적으로 해석할 수 있는 준실험적 방법론적 토대를 제공한다. 다만, 이러한 해석은 모든 관련 교란변수가 관찰되어 성향점수 모형에 포함되었다는 ‘관측되지 않은 교란변수가 없다 (No unmeasured confounding)’는 가정이 충족된다는 조건하에서만 성립한다. 그럼에도 불구하고 IPTW 분석을 통해 관찰 가능한 범위 내에서 교란 요인이 충분히 보정됨을 확인하였고, 따라서 본 연구의 ATE의 추정치는 정책적, 학술적으로 해석 가능한 신뢰성 있는 근거를 제공한다.

### 5.2.2. 성향점수 매칭(PSM)을 이용한 처치효과 분석

본 연구는 성향점수 매칭(P propensity Score Matching, PSM)을 활용하여 선택 편향(selection bias)을 통제하고, 특정 처치(treatment)가 결과에 미치는 인과적 효과를 추정하였다. PSM은 관찰 연구에서 처치 집단과 통제 집단 간의 공변량(covariates) 분포 차이를 보정하여, 마치 무작위 배정 실험(RCT)과 유사한 환경을 구현하는 통계적 방법론이다. 이 방법은 관찰된 특성(공변량)이 유사하여 처치 배정 확률(성향점수)이 비슷한 개인들을 서로 짝지어줌으로써, 처치 여부 외 다른 요인들의 영향을 최소화하고 순수한 처치 효과를 분석하는 데 강점을 가진다. 본 연구의 PSM 분석은 다음과 같이 크게 두 단계의 과정을 통해 수행되었다.

#### (1) 성향점수 추정 및 매칭

첫 번째 단계에서는 각 개체가 처치 집단에 속할 조건부 확률인 성향점수(P propensity Score)를 추정하고, 이를 바탕으로 처치 집단과 통제 집단 간의 매칭을 수행한다. 먼저, 처치 여부에 영향을 미치면서 동시에 결과 변수에도 영향을 줄 수 있는 주요 공변량(confounder)들을 이론적 배경과 선행 연구를 기반으로 신중하게 선택하였다. 본 연구에서는 로지스틱 회귀 모형(Logistic Regression)을 사용하여 성향점수를 추정하였는데, 이는 가장 전통적이고 널리 사용되는 방법이다. 추정된 성향점수를 기반으로, 처치 집단의 각 개인에 대해 가장 유사한 성향점수 값을 갖는 통제 집단의 개인을 짝짓는 최근접 이웃 매칭(Nearest Neighbor Matching) 방법을 적용해 1:1 매칭을 수행하였다. 이를 통해 공변량 분포가 유사한 두 비교 집단을 구성하였다.

#### (2) 매칭 균형 평가 및 효과 추정

두 번째 단계에서는 매칭이 성공적으로 이루어졌는지 평가하고, 최종적으로 처치 효과를 추정한다. 먼저, 매칭 후 처치 집단과 통제 집단 간의 공변량 분포가 균형을 이루는지 확인하는 균형 평가(Balancing Check)를 수행하였다. 각 공변량에 대해 표준화된 평균 차이(Standardized Mean Difference)를 계산하여 집단 간 차이가 유의미하게 감소하였는지 확인하고, 히스토그램과 같은 시각적 진단 그래프를 통해 두 집단의 분포가 유사해

졌는지 점검하였다. 균형 평가를 통해 매칭의 질이 확보된 것을 확인한 후, 매칭된 표본을 대상으로 두 집단 간 결과 변수의 평균 차이를 비교해 평균 처치 효과(Average Treatment Effect on the Treated)를 추정하고 통계적 유의성을 검증하였다. 이와 같은 엄격한 분석 절차를 통해 선택 편향의 영향을 최소화하고 신뢰성 있는 인과적 효과를 도출하고자 하였다.

### 5.3. 2단계: 이질적 처치 효과(HTE) 추정

본 연구는 만성질환의 영향이 모든 인구 집단에서 동일하지 않을 수 있다는 가정 하에, 처치 효과의 이질성을 탐색하는 것을 핵심 목표로 설정하였다. 이를 위해 가설 기반 접근법과 데이터 기반 탐색 접근법을 상호 보완적으로 사용하였다.

#### 5.3.1. 로지스틱 회귀를 이용한 집단별 처치효과(HTE) 분석

본 연구는 전통적 로지스틱 회귀 모델을 활용하여 사전에 정의된 하위집단별 만성질환이 의료비 과부담에 미치는 효과를 정량적으로 평가하였다. 구체적으로 성별, 연령대, 소득수준 등 주요 공변량을 고려한 로지스틱 회귀를 각 하위집단별로 수행하였으며, 처치 변수(만성질환 유무)와 결과 변수(의료비 과부담) 간 관계를 추정하였다. 분석 결과는 각 하위집단에서의 처치 효과가 log odds 형태로 산출되었으며, 유의성을 검증하기 위해 표준오차, z값, p값을 함께 산출하였다. 하위집단별 효과 차이를 직관적으로 확인하기 위해 막대그래프 형태로 시각화하였으며, 이를 통해 정책적으로 주목할 필요가 있는 집단을 사전 가설 기반으로 도출할 수 있었다.

#### 5.3.2. Causal Rule Ensemble(CRE)를 이용한 하위집단별 처치효과 분석

본 연구는 Causal Rule Ensemble (CRE) 모델을 활용하여 만성질환 유무가 의료비 과부담 발생에 미치는 하위집단별 처치효과(CATE)를 탐색하고 추정하였다. CRE는 기계 학습 기반 탐색적 인과 규칙 학습 방법으로, 다수의 잠재적 하위집단 규칙을 생성하고, LASSO 회귀를 통해 통계적으로 유의미한 규칙을 선별함으로써 이질적 처치 효과(Heterogeneous Treatment Effect, HTE)를 탐지한다. 본 연구에서 CRE 분석은 다음과 같이 수행되었다.

##### (1) 규칙 후보 추출

먼저 의사결정나무(decision tree)를 기반으로 10,000개의 나무(ntrees = 10000)를 생성하여 다양한 잠재적 하위집단 규칙을 탐색하였다. 각 나무의 최대 깊이(max\_depth = 5)를 설정하고, 최소 노드 크기(node\_size = 5)를 적용하여 복합 규칙과 작은 하위집단까지 포착할 수 있도록 하였다. 또한, 안정성 평가를 위해 부트스트랩 반복(B = 200)과 subsample = 0.5를 적용하였다. 한편, 본 연구에서 사용된 공변량 중 소득(INCOME)과 경제활동 참여상태(ECO1)은 모형에 포함되었으나, 규칙 후보 추출 과정에서는 제외하였다. 소득(INCOME)과 경제활동 참여상태(ECO1)은 만성질환과 의료비 과부담 모두에 영향을 미칠 수 있는 교란 변수(confounding variables)일 가능성이 높다. 따라서 이러한 변수들을 규칙 생성 과정에서 배제함으로써, 만성질환 유무가 의료비 과부담에 미치는 순수

한 처치 효과를 보다 정확하게 추정할 수 있다.

## (2) 규칙 선택 및 효과 추정

추출된 규칙 후보 중 LASSO 회귀를 통해 결과 변수와의 관련성이 높은 규칙을 선별하였다. 본 연구에서는 stability\_selection을 'no'로 설정하여 LASSO 규칙 선택 과정에서 안정성 평가를 제한하고, 모든 잠재적 하위집단을 탐색할 수 있도록 하였다. 이를 통해 CRE 모형이 발견한 각 하위집단의 처치효과를 충실히 반영하고, 정책적 함의를 도출할 수 있도록 하였다. 안정성 임계값(cutoff = 0.7)을 만족하는 규칙만 최종 선택하여 과적합을 방지하고 일반화 가능성을 확보하였다. 아울러 교차 타겟 학습기(Cross-Target Learner) 방식을 채택하여 처치 변수 예측에는 SL.glmnet을, 결과 변수 예측에는 SL.xgboost를 사용하여 예측 성능을 최적화하였다.

## (3) 결과 해석

최종 선정된 규칙을 기반으로 하위집단별 평균 처치효과(CATE)를 산출하였다. CRE 분석 결과, 전체 평균 대비 통계적으로 유의미한 효과를 보이는 특정 하위집단을 식별할 수 있었으며, 이를 정책적 개입이 필요한 취약 집단으로 정의하였다. 분석 결과는 summary(cre\_results) 및 plot(cre\_results)를 통해 하위집단 규칙과 효과 크기에 대한 해석 가능한 인사이트를 확인하였다.

# IV. 분석 결과

## 1. 연구 대상자의 일반적 특성

본 연구의 최종 분석 대상은 3,605명이었다. 전체 연구 대상자의 평균 연령은 45.8세였으며, 남성이 56%(2,009명)를 차지하였다. 분석 대상자의 특성을 만성질환 보유 여부에 따라 비교한 결과, 두 집단 간에 통계적으로 유의한 차이가 다수 발견되었다 (Table1 참조).

만성질환 보유 집단(N=1,214)은 미보유 집단(N=2,391)에 비해 평균 연령(51.0세 vs 43.1세)과 평균 체질량지수(BMI)(24.6 vs 23.6)가 통계적으로 유의하게 높았다(모두  $p < .001$ ). 사회경제적으로는 만성질환 보유 집단에서 기혼자(76% vs 69%), 장애 보유자(4.3% vs 1.0%), 지역 건강보험 가입자(21% vs 17%)의 비율이 더 높은 반면, 교육 수준과 평균 소득은 상대적으로 낮은 경향을 보였다(모두  $p < .001$ ). 또한, 이들은 우울감 경험률은 더 높았으나( $p = .004$ ), 음주 빈도는 더 낮았다( $p = .007$ ).

특히 연구의 주요 결과 변수인 의료비 과부담 경험률은 만성질환 보유 집단에서 15%로, 미보유 집단의 5.9%에 비해 월등히 높았다( $p < .001$ ). 반면, 두 집단 간 성별, 경제활동 상태, 민간의료보험 가입 여부 등에서는 유의한 차이가 관찰되지 않았다.

이처럼 두 집단 간 관찰된 주요 특성에서 차이를 보이므로, 선택 편향(selection bias)을 보정하고 순수한 인과적 효과를 추정하기 위해 성향점수 기반 분석이 필수적임을 확인하였다.

[Table 1] 연구 대상자의 일반적 특성

특성	Overall N = 3,605 <sup>1</sup>	만성질환 보유 여부		p-value <sup>2</sup>	민간의료보험 가입			
		만성질환 없음 N = 2,391 <sup>1</sup>	만성질환 보유 N = 1,214 <sup>1</sup>		0.5			
연령 (세)	45.8 ± 10.3	43.1 ± 10.3	51.0 ± 8.2	<0.001	미가입	331 (9.2%)	226 (9.5%)	105 (8.6%)
체질량지수 (kg/m <sup>2</sup> )	23.9 ± 3.5	23.6 ± 3.4	24.6 ± 3.5	<0.001	가입	3,274 (91%)	2,165 (91%)	1,109 (91%)
성별				0.9	장애 유무	<0.001		
여성	1,596 (44%)	1,061 (44%)	535 (44%)		아니오	3,528 (98%)	2,366 (99%)	1,162 (96%)
남성	2,009 (56%)	1,330 (56%)	679 (56%)		예	77 (2.1%)	25 (1.0%)	52 (4.3%)
혼인상태				<0.001	스트레스 인지	0.2		
미혼	1,029 (29%)	739 (31%)	290 (24%)		아니오	465 (13%)	320 (13%)	145 (12%)
기혼	2,576 (71%)	1,652 (69%)	924 (76%)		예	3,140 (87%)	2,071 (87%)	1,069 (88%)
교육수준				<0.001	현재 흡연 여부	0.4		
고졸 미만	458 (13%)	235 (9.8%)	223 (18%)		비흡연	2,664 (74%)	1,756 (73%)	908 (75%)
고졸 이상	3,147 (87%)	2,156 (90%)	991 (82%)		현재 흡연	941 (26%)	635 (27%)	306 (25%)
의료보장 형태				<0.001	음주 빈도	0.007		
직장가입자(본인)	2,195 (61%)	1,500 (63%)	695 (57%)		월 2회 미만	1,637 (45%)	1,047 (44%)	590 (49%)
건보: 직장피부양자	475 (13%)	310 (13%)	165 (14%)		월 2회 이상	1,968 (55%)	1,344 (56%)	624 (51%)
건보: 지역세대주	669 (19%)	417 (17%)	252 (21%)		규칙적 운동	0.2		
건보: 지역세대	211 (5.9%)	142 (5.9%)	69 (5.7%)		아니오	1,958 (54%)	1,317 (55%)	641 (53%)
의료급여 1,2종 세대주	41 (1.1%)	16 (0.7%)	25 (2.1%)		예	1,647 (46%)	1,074 (45%)	573 (47%)
의료급여 세대원	13 (0.4%)	6 (0.3%)	7 (0.6%)		우울감 경험	0.004		
미가입	1 (<0.1%)	0 (0%)	1 (<0.1%)		아니오	3,367 (93%)	2,254 (94%)	1,113 (92%)
경제활동 상태				0.3	예	238 (6.6%)	137 (5.7%)	101 (8.3%)
비취업	71 (2.0%)	52 (2.2%)	19 (1.6%)		의료비 과부담 여부	322 (8.9%)	141 (5.9%)	181 (15%)
취업	3,534 (98%)	2,339 (98%)	1,195 (98%)		개인 소득	3,398.7 ± 2,634.6	3,431.0 ± 2,480.0	3,335.3 ± 2,915.3

<sup>1</sup> Mean ± SD; n (%)  
<sup>2</sup> Wilcoxon rank sum test; Pearson's Chi-squared test

## 2. 방법론별 분석 결과

### 2.1. 평균 처치 효과(ATE) 추정 결과

#### 2.1.1. IPTW

##### (1) IPTW 가중치 생성 및 분포

본 연구의 목적은 만성질환 보유 여부가 가구의 의료비 과부담 발생에 미치는 인과적 효과를 추정하는 것이다. 단순 회귀분석만으로는 처치군과 대조군 간의 기저 특성 차이에 따른 교란을 충분히 제거하기 어렵다. 이에 따라, 본 연구에서는 Inverse Probability of Treatment Weighting (IPTW) 방법을 적용하여 평균처치효과(ATE)를 추정하였다. 만성질환 보유 여부를 처치 변수로 하여 성향점수를 로지스틱 회귀 기반으로 추정한 후, 역확률 가중치(Inverse Probability of Treatment Weight, IPTW)를 산출하였다. 가중치의 분포를 살펴본 결과, 대부분의 표본은 1~5 구간에 집중되어 있어 안정적인 형태를 보였으며, 일부 극단값(10 이상)이 확인되었다. 치료군의 가중치 범위는 1.0~28.5, 대조군은 1.0~12.2로 나타났으며, 이는 관리 가능한 수준(30 미만)에 해당하였다. 변동계수(CV)는 치료군 0.86, 대조군 0.41로 산출되어 두 집단 모두 허용 가능한 변동성(CV < 1) 내에 있었다. 유효 표본 크기(Effective Sample Size, ESS)는 치료군에서 1214명에서 699명으로 감소했으나, 분석력 확보에는 충분한 수준을 유지하였다.

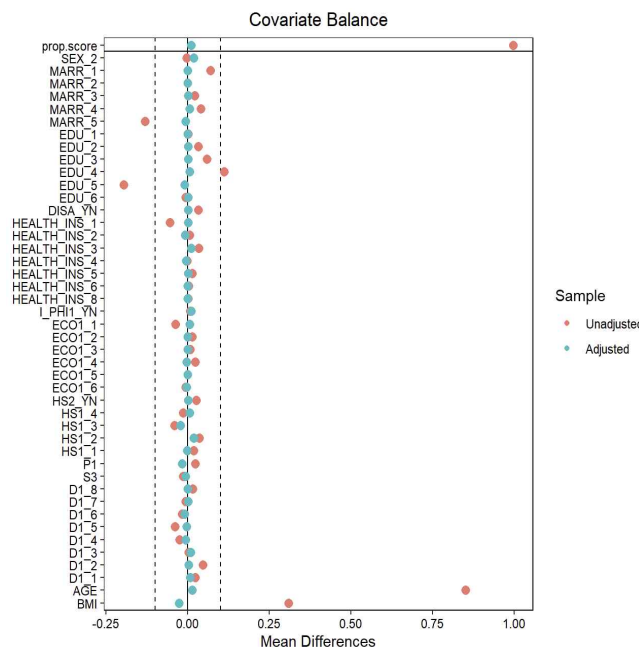
##### (2) 가중치 분포 확인



IPTW에서 성향점수가 0 또는 1에 근접한 경우 극단적으로 큰 가중치가 발생할 수 있으므로, 분석 전 가중치 분포를 확인하였다. 대부분의 표본은 안정적인 범위(약 1~5 구간)에 집중되어 있었으며, 일부 큰 값이 존재했으나 관리 가능한 수준으로 판단되었다. 따라서 별도의 극단값 조정은 적용하지 않았다.

### (3) 공변량 균형성 검증

공변량 균형성은 표준화 차이(Standardized Mean Difference, SMD)를 이용하여 평가하였다. IPTW 적용 전에는 일부 변수(예: BMI, AGE 등)에서  $\pm 0.25 \sim 0.90$  수준의 불균형이 확인되었다. 그러나 IPTW 적용 후 모든 변수의 SMD가  $\pm 0.10$  이내로 수렴하고, 가장 큰 잔차였던 BMI도  $-0.03$  수준으로 무시 가능한 크기이다. 이는 IPTW가 공변량 분포의 균형을 효과적으로 확보했음을 의미한다 (Figure 1 참조).



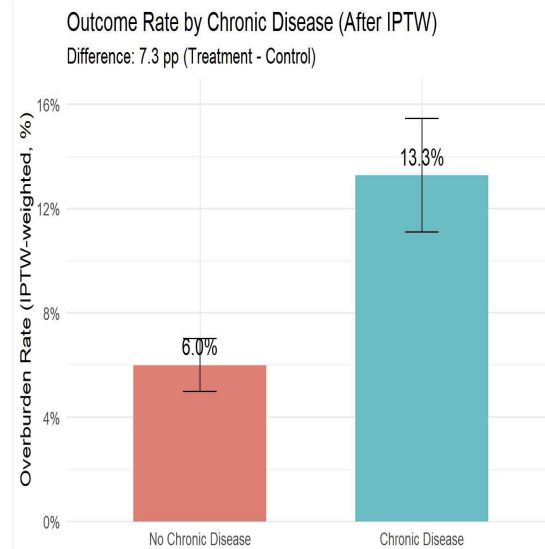
[Figure 1] 공변량 균형성 검증 그래프

### (4) ATE 추정 결과

가중치를 적용한 표본에서 만성질환 보유 여부가 의료비 과부담 발생 확률에 미치는 평균 처치 효과(ATE)를 추정하였다. 회귀분석 결과, 절편(Intercept) 계수는 0.05979로 나타났으며 이는 만성질환을 보유하지 않은 가구에서 의료비 과부담이 발생할 확률이 약 5.98%임을 의미한다. 이는 대조군의 기준선 발생 확률로 해석된다. 반면 만성질환 보유 여부(CD\_total)의 계수는 0.07326으로, 만성질환을 보유한 가구는 보유하지 않은 가구에 비해 의료비 과부담 발생 확률이 평균적으로 약 7.3%p 더 높았다. 즉, 만성질환 보유 가구의 과부담 발생 확률은 약 13.3%로 추정된다. 통계적 유의성 검정 결과, 해당 효과는  $t = 5.59$ ,  $p < 0.001$  수준에서 유의함을 확인했다.

다음은 IPTW 가중치를 적용한 후 만성질환 보유 여부별 과부담 발생률을 막대그래프로 나타낸 것이다.(Figure 2 참조.) 막대 위에 표시된 수치는 발생률 추정치, 에러바는 95% 신뢰구간을 의미한다. 시각적으로도 두 집단 간 발생률 격차가 뚜렷하게 드러나며,

만성질환 보유가 가구의 의료비 부담 위험을 실질적으로 증가시킨다는 분석 결과를 뒷받침한다.



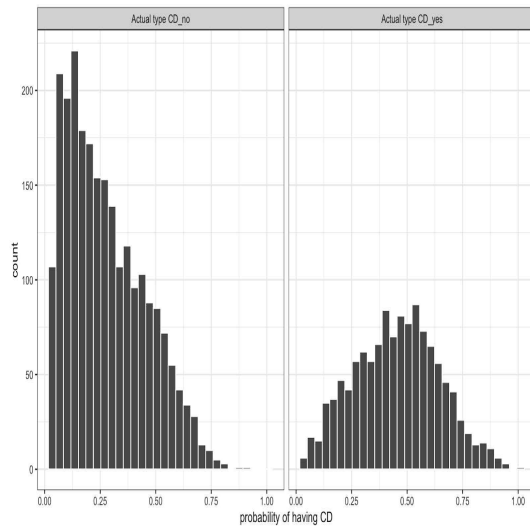
[Figure 2] IPTW 가중치 적용 후 만성질환 보유 여부별 과부담 발생률

## 2.1.2. PSM

### (1) 성향 점수 추정 및 매칭 절차

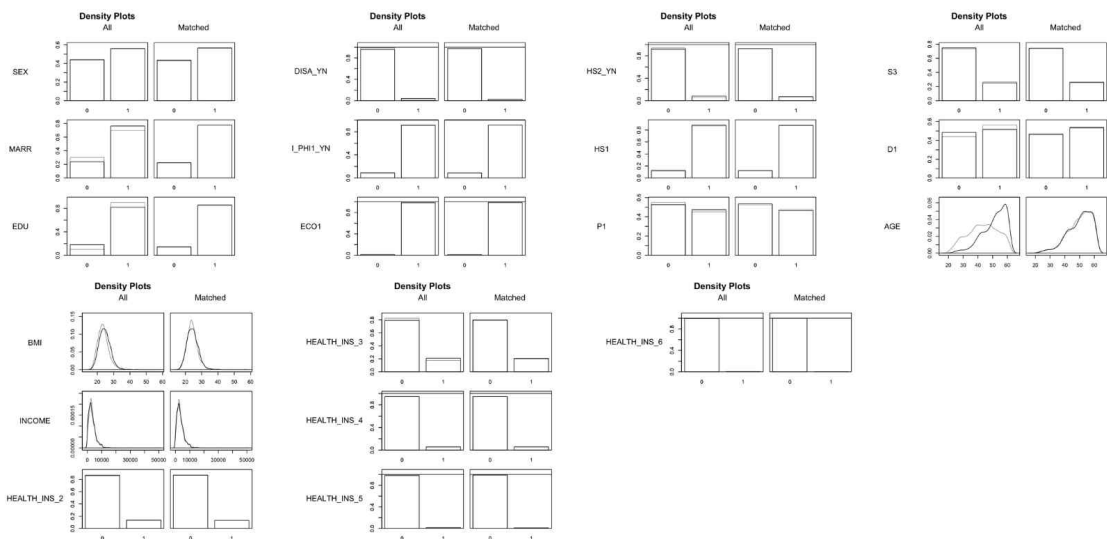
본 연구에서는 만성질환 보유 여부가 의료비 과부담 발생에 미치는 인과적 효과를 추정하기 위해 성향 점수 매칭(PSM)을 적용하였다. 만성질환 보유 여부를 처치 변수로 설정하고, 사전 공변량들을 이용한 로지스틱 회귀를 통해 각 개인의 성향 점수(Propensity Score, PS)를 추정하였다. 추정된 PS 분포는 두 집단(만성질환 유·무) 간에 약 0.03~0.92 구간에서 공통 지원 영역이 확보되었으며, 이 구간을 중심으로 잘라낸 후(Trimming) 매칭을 수행하였다 (Figure 3, 4, 5 참조). 매칭 알고리즘은 MatchIt 패키지의 최근접 이웃 매칭(Nearest Neighbor Matching)을 적용하였고, 매칭 후 총 2,116명(1,058쌍)이 분석에 활용되었다. 매칭 시에는 처리군과 비교군의 성향 점수가 과도하게 차이나는 사례를 배제하고 두 집단 간의 공변량 분포 균형을 개선하기 위하여 caliper(=0.1)를 적용하였다. 이는 성향 점수가 일정 범위 이상 차이 나는 개체 간의 매칭을 제한함으로써, 유사한 특성을 가진 표본들만을 매칭하도록 유도하기 위한 조치이다.





[Figure 3] 추정된 PS 분포

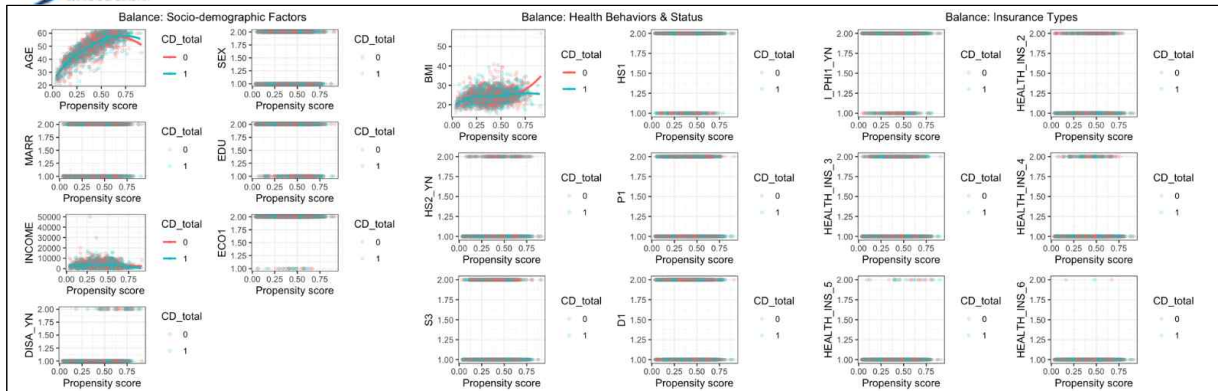
[Figure 4] 두 집단의 공통 지원 영역



[Figure 5] Trimming 이후 매칭 알고리즘 실행 결과

## (2) 공변량 균형성 검증

매칭 전에는 연령, 보험 가입 여부, 소득 수준 등 일부 변수에서 집단 간 분포 차이가 뚜렷하였다. 그러나 매칭 이후에는 시각적 공변량 균형 검사와 t-검정을 통해 두 집단의 분포가 대부분 겹쳐 나타났으며, 모든 공변량의 평균 차이가 통계적으로 유의하지 않게 수렴하였다. 이는 PSM을 통해 처치군과 대조군의 기저 특성이 효과적으로 균형을 이루었음을 시사한다 (Figure 6 참조).



[Figure 6] 매칭 전과 후 두 집단의 분포 비교

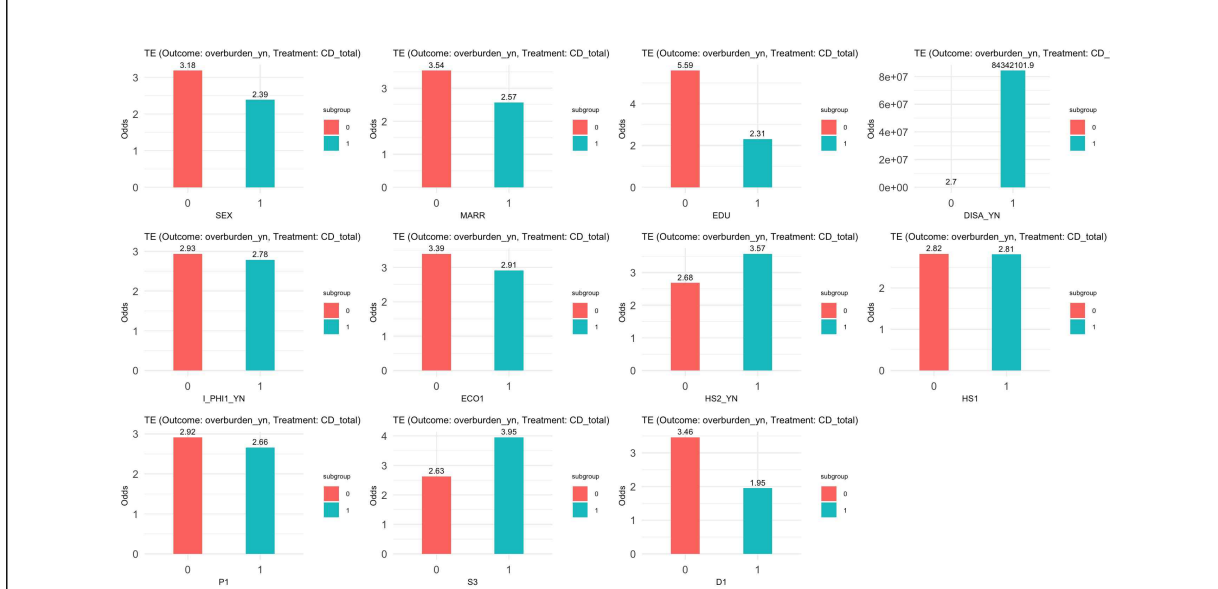
### (3) 평균 처치 효과(ATE) 추정 결과

매칭된 표본에서 만성질환 보유 여부가 의료비 과부담 발생 확률에 미치는 평균처치효과(ATE)를 추정하였다. 분석 결과, 만성질환이 없는 집단의 과부담 발생률은 약 5.9%였으며, 만성질환이 있는 집단에서는 약 13.8%로 나타나 두 집단 간 발생률 차이는 약 8%p였다. t-검정 결과 통계적으로 유의하였으며( $p < 0.001$ ), 이는 IPTW 분석과 일관된 방향성을 보였다. 즉, 동일한 배경 특성을 지닌 두 개인을 비교했을 때, 만성질환을 가진 경우 의료비 과부담을 경험할 확률이 유의하게 더 높다는 점을 확인하였다.

## 2.2. 이질적 처치 효과(HTE) 추정 결과

### 2.2.1. 로지스틱 회귀분석을 통한 집단별 처치 효과

만성질환의 영향이 인구사회학적 특성에 따라 어떻게 달라지는지 확인하기 위해, 처치 변수와 주요 공변량 간의 상호작용항을 포함한 로지스틱 회귀분석을 시행하였다. 이후 주요 하위집단별 처치 효과(odds)를 시각화하였다 (Figure 6 참조).

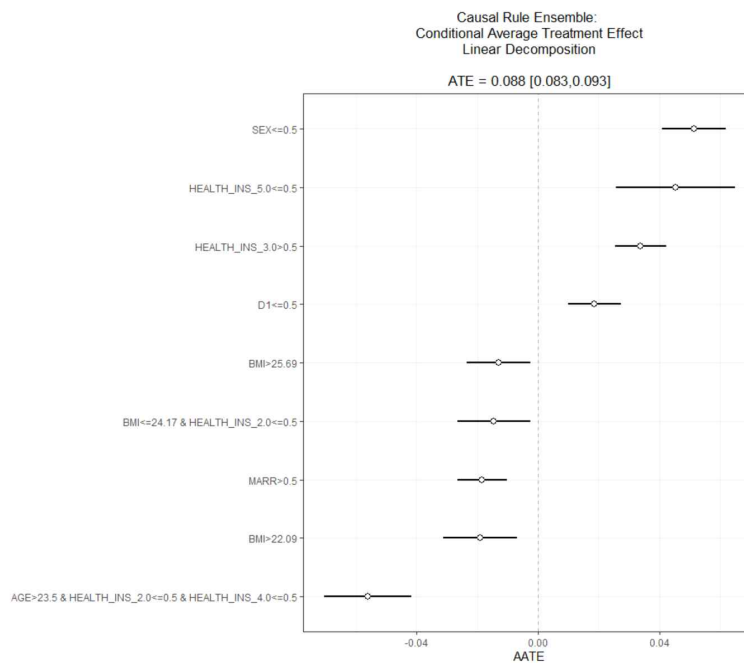


[Figure 6] 각 하위집단별 처치효과 차이

- 여성(SEX = 0): 여성 집단에서 만성질환이 의료비 과부담에 미치는 영향이 남성 집단에 비해 통계적으로 유의하게 더 컸다. 이는 CRE 분석 결과와 일관되며, 여성의 사회경제적 취약성이 반영된 결과로 해석된다.
- 저학력 집단(EDU = 0): 고등학교 졸업 미만의 저학력 집단에서 만성질환의 효과가 고학력 집단보다 두드러지게 높게 나타났다. 이는 학력이 소득 및 직업 안정성과 밀접한 관련이 있음을 고려할 때, 낮은 교육수준이 의료비 부담에 대한 취약성을 높이는 요인임을 시사한다.

### 2.2.2. CRE(Causal Rule Ensemble)를 이용한 하위집단 분석

본 연구는 Causal Rule Ensemble(CRE) 모델을 활용하여 만성질환 유무가 의료비 과부담 발생 여부에 미치는 하위집단별 처치효과(CATE)를 추정하였다. CRE 모델은 무작위 포레스트 기반으로 다수의 잠재적 규칙을 생성한 후, LASSO 회귀를 통해 결과 변수와 가장 관련성이 높은 규칙들을 선별하는 인과 규칙 학습(Causal Rule Learning) 방법론이다. 이는 복잡한 상호작용 효과를 탐지하고, 전통적인 회귀 분석으로는 파악하기 어려운 정책 대상 집단을 명확하게 식별하는 데 효과적이다. 전체 표본을 대상으로 분석한 결과, 최종적으로 9개의 규칙이 선택되었다 (Figure 7 참조).



[Figure 7] CRE 분석 결과 summary

#### (1) 전체 평균 처치효과(ATE) 분석 결과

전체 평균 처치효과(ATE)는 0.088(95% CI: [0.083, 0.093])로 산출되었다. 이는 만성질환 보유 여부가 의료비 과부담 발생 확률을 평균적으로 약 8.8%p 증가시킨다는 것을 의

미한다.

## (2) 주요 하위집단별 처치효과(CATE) 분석 결과

탐색 결과, 총 9개의 규칙이 유의하게 선택되었으며, 일부는 단일 변수 기반, 일부는 다 변수 상호작용 기반 규칙이었다. 이는 데이터 내 특정 조건에서 만성질환의 부담 효과가 강화되거나 완화되는 이질적 양상을 보여준다. 주요 규칙은 다음과 같다.

- 규칙 1: 여성 집단( $SEX \leq 0.5$ )

여성은 전체 평균 대비 약 5.1%p 더 높은 의료비 과부담 효과를 보였다. 이는 기존 로지스틱 회귀 결과와도 일관된다.

- 규칙 2: 지역 건강보험 세대주 집단( $HEALTH\_INS\_3.0 > 0.5$ )

지역 건강보험 세대주인 경우, 의료비 과부담 확률이 평균 대비 약 3.4%p 더 높았다. 이는 직장 기반 가입자에 비해 경제적 취약성이 상대적으로 크게 작용할 수 있음을 시사한다.

- 규칙 3: 중상위 수준의 체질량지수( $BMI > 22.09$ ,  $BMI > 25.69$ )

BMI가 22.09 이상 또는 25.69 이상인 집단에서는 처치효과가 각각 -1.9%p, -1.3%p 낮았다. 이는 비만 혹은 과체중 집단에서 만성질환 효과가 상대적으로 약화될 수 있음을 보여준다.

- 규칙 4: 장애 없음 집단( $D1 \leq 0.5$ )

장애가 없는 집단에서는 만성질환의 효과가 평균 대비 약 1.9%p 높았다. 즉, 건강상 제약이 적은 집단에서도 질환 자체가 과부담 위험을 키운다는 점을 보여준다.

- 규칙 5: 결혼한 집단( $MARR > 0.5$ )

혼인 상태인 경우, 효과가 -1.8%p 낮게 나타났다. 이는 가구 단위의 소득·부담 분산 효과로 해석할 수 있다.

- 규칙 6: 의료급여(1·2종) 세대주가 아닌 집단( $HEALTH\_INS\_5.0 \leq 0.5$ )

의료급여(1·2종) 세대주가 아닌 경우, 즉 다른 보험 유형에 속한 집단은 만성질환 효과가 약 4.5%p 더 높았다. 이는 의료급여 세대주가 오히려 제도적 지원을 받아 과부담 위험이 완화되는 반면, 그렇지 않은 집단은 부담이 상대적으로 크다는 점을 시사한다.

- 규칙 7: 다중 조건 집단( $AGE > 23.5$  &  $HEALTH\_INS\_2.0 \leq 0.5$  &  $HEALTH\_INS\_4.0 \leq 0.5$ )

23세 이상이면서, 직장 건강보험 피부양자가 아니고 지역 건강보험 세대원도 아닌 집단에서 효과가 -5.6%p 낮게 나타났다. 이는 특정 연령·보험 조건의 결합이 만성질환 부담을 완화할 수 있음을 보여준다.

종합적으로, CRE 분석은 만성질환의 경제적 부담이 성별·체질량지수·장애 여부·혼인 상태·보험 유형 등 다양한 요인에 따라 달라짐을 보여주었다. 특히 보험 유형(직장 피부양자, 지역 세대주, 의료급여 세대주 등)이 하위집단 효과를 크게 설명하는 변수로 나타나, 제도적 보장 구조가 의료비 과부담에 중요한 영향을 미친다는 점을 시사한다.

## V. 결론 및 고찰

## 1. 연구 요약

본 연구는 만성질환 보유 여부가 경제활동인구 개인의 의료비 과부담 발생에 미치는 인과적 효과를 분석하였다. 최종 분석 대상은 3,605명으로, 만성질환 보유 집단은 비보유 집단에 비해 연령과 BMI가 높고, 기혼자 및 장애 비율이 상대적으로 높았다. 이러한 집단 간 기저 특성 차이를 보정하기 위해 IPTW 및 PSM 기반 성향점수 분석을 적용하였다. 더불어 로지스틱 회귀 기반 HTE 분석과 CRE 모형을 활용하여 하위집단별 처치효과 (CATE)를 탐색함으로써, 단순 평균 효과를 넘어 정책적으로 의미 있는 취약 집단을 규명하고자 하였다.

## 2. 주요 분석 결과

### 2.1. 평균처치효과 (ATE)

IPTW 및 PSM 분석 결과, 만성질환 보유는 의료비 과부담 발생 확률을 평균적으로 약 8.0%p 증가시키는 것으로 나타났다( $p < 0.001$ ). 이는 만성질환이 가계 경제적 부담의 주요 위험 요인임을 강력히 시사한다.

### 2.2. 하위집단별 처치효과 (HTE/CATE)

로지스틱 회귀 기반 HTE 및 CRE 분석 결과, 만성질환의 영향은 인구사회학적 특성에 따라 달라지는 것으로 확인되었다. 특히 CRE 분석 결과에서 만성질환의 영향은 성별, 혼인 상태, 체질량지수(BMI), 그리고 건강보험 유형 등 인구사회학적 특성에 따라 달라지는 것으로 확인되었다. 특히, 다음 세 가지 집단으로 구분할 수 있었다.

#### 2.2.1. 경제적 취약성 집단

- 여성: 남성보다 만성질환으로 인한 의료비 과부담 위험이 평균 대비 더 크게 나타났다. 이는 한국 보건·노동 연구에서 일관되게 나타나는 여성의 낮은 소득 수준, 비정규직 비중, 높은 유병률과 맞닿아 있다.

- 지역 건강보험 세대주: 직장 기반 가입자에 비해 고용 안정성이 낮고, 소득·재산에 따른 보험료 산정으로 인해 동일 의료비라도 체감 부담이 크다. CRE 결과에서도 과부담 위험이 유의하게 높았다.

- 의료급여 세대주가 아닌 집단: 의료급여 세대주 집단은 오히려 제도적 보호 효과로 과부담 위험이 완화되는 반면, 그렇지 않은 집단에서는 부담이 상대적으로 더 크게 나타났다.

#### 2.2.2. 경제적·사회적 안정성 집단

- 기혼자: 만성질환으로 인한 과부담 위험이 낮았다. 이는 배우자와의 소득·돌봄 공유를 통한 부담 완화 효과로 해석된다.

- 비피부양자 집단: 직장 건강보험 피부양자가 아닌 경우, 경제적으로 독립하여 일정 수준 이상의 소득·고용 기반을 보유한 집단으로 볼 수 있으며, CRE 결과에서도 과부담 위

힘이 낮게 나타났다.

- 젊은 독립적 집단: 특정 연령 이상이면서 직장 피부양자나 지역 세대원에 해당하지 않는 집단에서 효과가 완화되었는데, 이는 사회적·경제적 자립성이 일정 수준 확보된 집단의 특성과 연결된다.

### 2.2.3. BMI 관련 효과 (비선형적 패턴)

BMI가 일정 구간(22 이상~25 이상)에서는 오히려 만성질환의 추가 부담이 낮게 나타났다. 이는 단순히 체중 증가가 위험을 선형적으로 높이는 것이 아님을 보여준다.

이러한 결과는 기존 문헌에서 보고된 Obesity paradox(비만 역설, 약간의 과체중 구간에서는 건강위험이 급격히 커지지 않고 오히려 보호 효과가 관찰되는 현상)과도 일맥상통한다.(Zhang et al., 2024)

## 3. 정책적 시사점

본 연구 결과를 토대로 다음과 같은 정책적 제언을 제시한다.

- 취약 집단 맞춤형 의료비 지원: 여성, 지역보험 세대주 등 경제적 취약성이 큰 집단을 우선 대상으로 하여 재정 지원 및 건강관리 프로그램을 배치할 필요가 있다.
- 만성질환 관리 및 예방 강화: 단순 치료를 넘어 조기 검진·관리 체계 강화와 의료비 부담 완화를 위한 재정적 안전망 마련이 필요하다.
- 보험제도 기반 보호 한계 보완: 피부양자 제도나 의료급여와 같은 제도적 틀만으로는 충분치 않으며, 실제 경제적 취약성을 반영한 타겟팅이 요구된다.

## 4. 연구 의의 및 한계

본 연구는 만성질환이 의료비 과부담에 미치는 인과적 영향을 IPTW·PSM 기반 분석과 CRE 모형을 결합하여 엄밀히 검토하였으나, 다음과 같은 한계점을 지닌다.

첫째, 일부 희소 표본 집단에서 추정치가 불안정할 수 있다는 점이다. 하위집단 분석 과정에서 관측치 수가 충분히 확보되지 않은 경우, 추정된 처치 효과가 과대 혹은 과소 추정될 가능성이 존재한다.

둘째, 본 연구는 정신건강, 소득, 직업 안정성 등 의료비 과부담과 밀접하게 연관될 수 있는 추가 사회경제적 요인을 충분히 포함하지 못하였다. 따라서 분석 결과가 실제 상황을 완전히 반영하지 못할 수 있다.

셋째, 본 연구는 패널 데이터를 활용했음에도 불구하고 특정 연도(2020년)의 횡단면 분석에 머물렀다는 점이다. 이로 인해 만성질환 상태의 변화나 의료비 지출의 장기적인 동태성을 반영하지 못하는 한계가 있다.

이러한 한계점에도 불구하고, 본 연구는 만성질환이 의료비 과부담에 미치는 인과적 효과를 다양한 방법론을 통해 검증하고, 하위집단별 정책적 함의를 제시했다는 점에서 의의가 있다.

## 5. 향후 연구 방향

향후 연구에서는 몇 가지 보완 방향을 고려할 수 있다. 첫째, 장기 추적 자료를 활용하여 만성질환 관리가 시간이 지남에 따라 의료비 부담에 어떤 변화를 가져오는지 분석함으로써, 보다 동태적인 인과 효과를 규명할 필요가 있다. 둘째, 정신건강, 소득, 직업 안정성과 같은 사회경제적 요인을 포함한 다층적 이질적 처치 효과(HTE) 분석을 통해 정책적 개입이 필요한 세부 집단을 보다 정밀하게 파악할 수 있을 것이다. 셋째, CRE와 로지스틱 회귀 기반 HTE 결과를 결합한 정책 시뮬레이션 연구를 통해 실제 정책 설계와 적용 가능성을 검증함으로써 맞춤형 지원 방안을 제시할 수 있다.

## 6. 결론

만성질환은 의료비 과부담을 유의하게 증가시키며, 그 영향은 성별, 혼인 상태, 보험 유형, 체질량지수 등 인구 사회학적 특성에 따라 달라진다. 따라서 의료비 지원 및 만성질환 관리 정책은 평균 효과를 넘어서 하위집단별 취약성을 고려한 맞춤형 전략이 필요하다. 본 연구는 인과적 분석과 CRE 기반 하위집단 탐색을 통해 이러한 정책적 근거를 제시하였다는 점에서 의의가 있다.