HIRA RESEARCH 2021;1(2):166-178 | pISSN 2765-6764 eISSN 2765-7353
https://doi.org/10.52937/hira.21.1.2.166

Original Article

# 새로운 환자표본자료 표본 추출 및 대표성 검증

**윤철영[1], 안재준[2], 이경민[3], 최용석[3], 김리현[3], 하대우[1], 유기봉[2,3]**
연세대학교 [1]응용통계학과, [2]데이터사이언스학부, [3]보건행정학과

# Developing the New National Patient Sample and Evaluating Representations

**Chul Young Yoon[1], Jae Joon Ahn[2], Gyeongmin Lee[3], Yongseok Choi[3], Lihyun Kim[3], Dae Yoo Ha[1], Ki-Bong Yoo[2,3]**
[1]Department of Applied Statistics, [2]Division of Data Science, and [3]Department of Health Administation, Yonsei University, Wonju, Korea

**Correspondence to:**
Ki-Bong Yoo
Department of Health Administation, Yonsei University, Changjo 406, 1 Yeonsedae-gil, Wonju 26493, Korea
Tel: +82-33-760-2458
Fax: +82-33-760-2919
E-mail: ykbong@yonsei.ac.kr

Received: October 14, 2021
Revised: November 2, 2021
Accepted after revision: November 2, 2021



**Background:** The sampling framework of the National Patient Sample of Health Insurance Review & Assessment Service is needed to be improved due to the current demographic structure. We proposed a sampling method and additional strata for extracting the National Patient Sample data due to the current demographic structure, such as low birth rate and aged population.
**Methods:** A total of 36 strata were set by adding four strata compared to the existing one. The maximum rate of minimal sample number was defined among the entire strata. Based on the rate, we extracted a small-scale sample dataset consisting of about 400,000 people and a large-scale sample dataset of more than 700,000 people.
**Results:** The representativeness of the high-frequency disease and the low-frequency disease was confirmed. For health expenditure, the representativeness of samples was confirmed in large-scale samples. However, the representativeness of small-scale samples was not confirmed in five strata.
**Conclusion:** Using the maximum rate of minimal sample number can reflect the demographic structure changes and diverse medical utilization. Although lack of representativeness in the five strata of the small-scale sample, both the small-scale sample and the large-scale sample are necessary to improve data accessibility and a sustainable data provision system. It will be helpful in establishing health policies and conducting medical research.

**Keywords:** National patient sample; Sampling studies; Validation study

## 서 론

우리나라의 건강보험 청구자료는 당연지정제라는 환경에서 대표성이 확보된 자료로 보건의료 관련 정책 및 기술 발전 연구에 기초자료로 이용

되고 있다. 청구자료는 일선의 의료기관이 건강보험 청구를 진행하며 축적된 자료로 건강보험심사평가원(심사평가원)과 국민건강보험공단이 각각 관리하고 있고 각 기관에서 여러 가지 방향으로 청구자료를 제공하고 있다.

수요자가 건강보험 청구자료를 이용하는 방법은 다음과 같다. 미리 구축된 표본자료를 이용하거나 맞춤형 자료로 신청하여 추출된 자료를 이용하는 방법이 있다. 심사평가원 및 국민건강보험공단에서 표본자료 및 맞춤형 자료를 제공하는 절차는 유사하다. 최근에는 보건의료기본법 제44조, 보건의료기술진흥법 제10조, 제26조에 근거하여 보건의료 빅데이터 플랫폼을 통해 연계된 자료를 제공하였으며[1], 데이터 3법을 바탕으로 심사평가원, 국민건강보험공단, 한국보건산업진흥원을 데이터 결합기관으로 지정하고 데이터를 결합하여 제공하는 절차를 수립하였다[2].

청구자료가 다양한 절차로 제공되고 있지만 맞춤형으로는 증가하는 수요를 감당하기 어렵다. 수요에 대응하고 접근성과 즉시성을 확보하기 위해서는 표본자료가 중요한 역할을 한다. 이에 심사평가원은 환자표본자료를 개발하고 이를 제공하기 시작하였다[3]. 마찬가지로 국민건강보험공단에서도 표본코호트를 개발하여 제공하기 시작하였다. 환자표본자료는 2010년부터 개발을 시작하였다. 환자표본자료는 전체환자데이터셋(national patient sample, NPS), 입원환자데이터셋(national inpatient sample, NIS), 소아청소년환자데이터셋(pediatric patient sample, PPS), 고령환자데이터셋(aged population sample, APS)으로 구성 및 제공되고 있다. 데이터 제공 연도는 2009년부터 매년 표본 추출 및 제공하고 있다. 국민건강보험공단 표본코호트는 2002년부터의 건강보험 자격 대상자 코호트 자료를 구축하여 제공하고 있다. 세부적인 표본 데이터로는 표본코호트DB, 건강검진코호트DB, 노인코호트DB, 직장여성코호트DB, 영유아검진코호트DB로 구성되어 있다. 환자표본자료와 표본코호트 모두 국민건강보험의 청구자료를 바탕으로 구성되어 있다. 건강보험 빅데이터의 20번 테이블(명세서 일반), 30번 테이블(진료내역), 40번 테이블(상병내역), 53번 테이블(원외처방내역), 요양기관 테이블은 환자표본자료와 표본코호트에 모두 존재한다. 세부 변수로는 요양기관 테이블에서 차이가 나고 나머지는 거의 동일하다. 두 표본자료의 큰 차이점은 자격테이블과 건강검진 테이블이다. 자격테이블은 건강보험 자격 대상자에 대한 정보로, 국민건강보험공단이 보험료 징수를 담당하고 있기에 표본코호트에만 존재한다. 자격 정보에는 소득수준의 대리변수로 이용할 수 있는 보험료 납부 분위와 환자의 거주지역, 장애등급 등의 정보가 담겨있다. 마찬가지로 검진 데이터도 국민건강보험공단이 국가검진을 관리하고 있기에 표본코호트에만 존재하고 있다.

환자표본자료가 비록 자격 및 검진 정보가 부족하고 단일 연도 데이터라는 한계가 있지만, 접근성 및 연도별 대표성 측면에서 사용자에게 선호되는 특성이 있다. 환자표본자료는 처방양상을 분석하거나[4,5] 질병부담 측정[6], 의료이용행태 보고 연구[7,8] 등을 수행하는 데 주로 이용되고 있으며, 환자표본자료를 이용하여 시장분석을 수행하는 벤처기업 및 제약회사가 있을 정도로 산학계 여러 방면에서 활용도가 높은 상황이다. 질병 위험요인 연구[9]에도 일부 이용되고 있지만, 주요 활용도는 연도별 대표성을 기반으로 신속하게 현황 및 추세를 보고하는 데 주로 사용되고 있기에 사회적으로 수요가 매우 높은 자료원이라 볼 수 있다.

다만 2010년부터 환자표본자료의 개발 및 추출을 시작한 이후로 추출 틀에 대한 변경이 없었기에 연도

별 대표성에 대한 보완의 필요성이 제기되고 있는 상황이다. 저출산 고령화 및 의료이용 패턴의 변화를 반영하여 새로운 층의 도입이 요구되고 있다. 2012년 의료비통계연보와 2020년 의료비통계연보에서 연령별 요양급여비용을 비교하였을 때, 만 0세 기준 2012년에는 634,087,591천 원이 지출되었지만 2020년에는 807,674,274천 원이 지출되었고, 명세서 건당 요양급여비용은 2012년에는 약 26,000원 수준이었지만 2020년에는 85,000원 수준으로 증가하였다[10,11]. 보장성 강화 및 수가 인상을 감안하더라도 2012년 출생아 수가 약 48만 명, 2020년 출생아 수가 약 27만 명으로 큰 폭으로 감소하였음에도 불구하고 영유아 층에서 의료비가 많이 발생한다는 것은 그만큼 의료비의 분산이 클 수 있다는 것을 의미한다[12,13]. 더불어 80세 이상 인구의 비율이 2012년 2.1%에서 2020년 3.6%로 큰 폭으로 늘었다. 노인 의료비의 경우 금액이 많고 분산이 크기 때문에 표본 추출 층을 조금 더 세분화할 필요가 있다[14]. 이렇게 인구구조가 변화하였지만, 기존의 환자표본자료는 영유아 층을 4세 이하, 5-9세로 구분하고 있으며 고령층의 경우 75세 이상을 한 개의 층으로 설정한 상황이기에 현재의 인구구조를 반영한 환자표본자료 표본추출 틀을 보완할 필요성이 제기된 상황이다.

따라서 이 연구에서는 저출산 고령화의 현황에 맞는 환자표본자료 추출방법과 층을 제안하고, 이를 토대로 추출한 환자표본자료를 제시하고자 한다. 대표성을 확보하여 보다 근거수준이 높은 연구를 수행하는 데 도움이 될 것이다.

## 방 법

저출산 고령화에 따라 지속적으로 변화하고 있는 인구구조를 반영하기 위해 표본 추출 층과 방법을 제안하고 이에 맞추어 표본 추출을 진행하고, 대표성을 검증하고자 한다.

### 1. 표본 추출 층

표본 추출 층화 변수로는 연령과 성을 선택하였다. 사전에 수행된 연구 프로젝트(G000FF8-2020-179)에 기초하였을 때 모집단 실제값과 표본의 모집단 추정값은 주로 저연령 및 고연령에서 크게 차이 나고 있기 때문에 이에 맞추어 연령층을 19세 이전의 5개 구간, 20세 이상은 5세 단위로 나눠 20-79세 12개 구간과 80세 이상 1개 구간, 총 18개 구간으로 구분하고자 한다. 19세 이전의 층은 기존 0-4세, 5-9세, 10-14세, 15-19세에서 0-2세, 3-5세, 6-9세, 10-12세, 13-19세로 구분하였다. 생애주기상 0-2세는 영아기로 구분하고 있으며 3-5세는 미취학 아동기인 유아기로 구분된다. 보통 6-12세는 아동기로 분류하는데, 6-9세는 초등학교 저학년, 10-12세는 초등학교 고학년이고 성장과정에 있기 때문에 두 개 층으로 구분하였다. 13-19세는 청소년기로 한 개의 층으로 설정하였다. 기존 PPS도 0-2세, 3-5세, 6-9세, 10-12세, 13-19세로 구분하고 있으나 NPS, NIS는 5세 단위로 연령을 구분하기에 기준을 통일하였다. 종합적으로 총 36개의 층을 설정하였다.

### 2. 표본 추출방법

이 연구에서는 2가지의 층화 변수, 총 36개의 층을 사용하는 층화임의추출법을 사용하였다. 각 층마다 표본의 최소 크기는 다음과 같이 진행하였다. 층별 대표 변수는 1인당 총의료비의 로그를 취한 값($X$)으로 설정하였다. 이는 과거 표본과의 일관성을 유지하기 위함이다[3]. 1인당 총의료비는 경제협력개발기구(Organization for Economic Cooperation

and Development, OECD) 등 국제 데이터와의 비교, 재정 추계, 정책 결정에 이용되는 가장 기본적인 변수이기에 대표 변수로 선정하였다. 각 층별로 log (1인당 총의료비)의 표본 평균 $\bar{X}$를 이용하여 모평균 $\mu$를 추정하기 위한 최소 표본의 크기를 오차한계 기준으로 계산하였다. 오차한계는 표본으로 허용할 수 있는 모집단과의 오차를 뜻하며 다음 수식 1로 계산하였다.

**수식 1**

$$d = Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

수식 1에서 α는 신뢰수준이며, $Z$는 신뢰수준에 대한 $Z$값(=1.96), $d$는 100(1−α)%의 오차한계(=5%), $\sigma$는 모 표준편차, $n$은 표본의 수를 뜻한다. 수식 1을 이용하여 100(1−α)%, 5%의 오차한계를 가지는 최소 표본의 수를 구하려면 수식 2를 거쳐 수식 3으로 변형해야 한다.

**수식 2**

$$d \geq Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

**수식 3**

$$n \geq (Z_{\alpha/2}\frac{\sigma}{d})^2$$

일반적으로 수식 1에서 문제가 되는 부분은 모 표준편차인 $\sigma$를 구할 수 없다는 것인데, 이 연구에서는 모분산을 구할 수 있으므로 모 표준편차 $\sigma$는 문제가 되지 않는다. 따라서 수식 3을 이용하여 모집단을 추정할 수 있는 층별 표본의 수를 구할 수 있다.

수식 3의 표본의 수(n)는 층별 모분산 값에 따라 달라지므로 층마다 필요한 최소 표본의 수(minimum sample number)가 달라진다. 각 층을 최소 표본의 수만큼 추출하는 경우 층마다 모집단 대비 표본의 비율이 달라질 수 있기 때문에 층별 대표성은 확보될 수 있지만 전체 집단에서의 대표성은 왜곡될 수 있다. 따라서 이 연구에서는 층별 최소 표본의 수를 계산한 이후, 층별 모집단 대비 최소 표본 수의 비율을 계산하고, 이 비율 중 최대값을 기준으로 전체 층의 표본을 추출하는 방법을 사용하였다. 그림 1은 각 층의 최소 표본 수의 비율 중 최대값을 기준으로 각 층의 표본 수를 고려하는 과정을 보여준다. Maximum rate of minimum sample number는

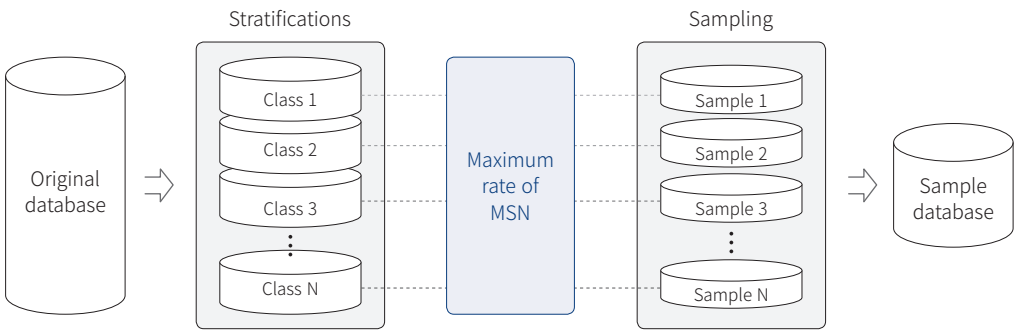


**그림 1.** 각 층의 최소 표본 수(minimum sample number, MSN)의 비율 중 최대값을 이용한 전체 표본 수 확보과정.

모집단에서 층을 나눌 때 사용되는 각 층의 최소 표본 수의 비율 중 최대값을 나타낸다.

종합하자면, 이 연구에서는 기존에 비해 연령층을 19세 이전과 고령층에서 층을 추가하여 총 36개 층으로 구성하였고, 표본 추출의 비율을 모집단 전체를 기준으로 선택한 것이 아닌 층별 표본 추출의 비율을 계산하고, 이 중 최대값을 선택하여 표본 추출의 비율로 결정하였다.

### 3. 신규 데이터 구축

이 연구에서 제안하는 표본은 기존 환자표본자료의 형태에 맞추어 전체 환자, 입원환자, 소아 · 청소년 환자 그리고 고령 환자로 나누고 단면 표본으로 구축하였다. 보안에 대한 우려가 있기 때문에 최소한의 대표성을 확보하는 소규모 표본자료와 기존의 심사평가원의 표본 데이터의 크기와 비슷한 100만 명 규모의 대규모 표본자료로 구분하여 구축하였다. 최소 표본 수의 비율 중 최대값을 적용하였을 때 NPS, NIS에서 약 40만 수준으로 도출되었기 때문에 PPS, APS 소규모 표본자료도 약 40만으로 구성하였다. 신규 데이터는 심사평가원 내부의 보안기준을 거친 2018년 청구자료 전체를 이용하였다.

### 4. 신규 데이터 대표성 검증

상병 빈도는 SAS ver. 9.4의 PROC SURVEY−FREQ (SAS Institute Inc., Cary, NC, USA), 1인당 총의료비 등의 연속형 자료의 검증은 SAS ver. 9.4의 PROC SURVEYMEANS (SAS Institute Inc.)를 이용하여 검정하였다. SAS ver. 9.4의 SURVEY 프로시저(SAS Institute Inc.)는 표본 데이터로부터 모집단의 추정값과 표준오차, 신뢰구간을 계산하는 데 이용한다. SURVEY 프로시저 수행 시에는 표본추출과정에서 계산된 층별 가중치를 적용하였다.

## 결 과

### 1. 층별 현황

층별 모집단 수와 최소 표본의 수, 추출비율은 표 1과 같다. 층별 분산을 바탕으로 최소 표본의 수를 계산하였고, 모집단 수에 최소 표본의 수를 나누어 추출비율을 결정하였다. NPS의 경우 추출비율 최대값이 남성 80세 이상 층 약 0.007795, NIS는 여성 10−12세 0.05204로, PPS는 여성 2세 이하 층 0.002863, APS는 남성 80세 이상 층 약 0.007795로 확인되었다(표 1).

**표 1.** 층별 모집단 수와 최소 표본의 수, 추출비율

| 표본 | 연령(세) | 모집단 수(명) | | 최소 표본수(명) | | 층별 추출 비율 | |
|---|---|---|---|---|---|---|---|
| | | 남 | 여 | 남 | 여 | 남 | 여 |
| NPS | ≤2 | 758,733 | 716,629 | 2,073 | 2,052 | 0.002732 | 0.002863 |
| | 3-5 | 887,767 | 843,281 | 1,578 | 1,530 | 0.001777 | 0.001814 |
| | 6-9 | 1,137,869 | 1,073,031 | 1,687 | 1,790 | 0.001483 | 0.001668 |
| | 10-12 | 851,996 | 803,075 | 2,208 | 2,137 | 0.002592 | 0.002661 |
| | 13-19 | 1,962,660 | 1,876,450 | 2,565 | 2,210 | 0.001307 | 0.001178 |
| | 20-24 | 1,712,151 | 1,737,361 | 2,846 | 2,444 | 0.001662 | 0.001407 |
| | 25-29 | 1,786,214 | 1,757,933 | 2,792 | 2,820 | 0.001563 | 0.001604 |

(다음 페이지에 계속)

**표 1.** 계속

| 표본 | 연령(세) | 모집단 수(명) | | 최소 표본수(명) | | 층별 추출 비율 | |
|---|---|---|---|---|---|---|---|
| | | 남 | 여 | 남 | 여 | 남 | 여 |
| | 30-34 | 1,720,477 | 1,766,241 | 2,847 | 3,192 | 0.001655 | 0.001807 |
| | 35-39 | 2,078,015 | 2,139,900 | 2,955 | 3,088 | 0.001422 | 0.001443 |
| | 40-44 | 2,058,297 | 2,112,279 | 3,112 | 3,018 | 0.001512 | 0.001429 |
| | 45-49 | 2,360,978 | 2,440,302 | 3,284 | 3,070 | 0.001391 | 0.001258 |
| | 50-54 | 2,236,672 | 2,329,915 | 3,523 | 3,036 | 0.001575 | 0.001303 |
| | 55-59 | 2,330,555 | 2,430,825 | 3,542 | 2,864 | 0.001520 | 0.001178 |
| | 60-64 | 1,875,847 | 1,983,045 | 3,757 | 2,975 | 0.002003 | 0.001500 |
| | 65-69 | 1,310,637 | 1,436,561 | 3,744 | 2,865 | 0.002857 | 0.001994 |
| | 70-74 | 980,208 | 1,161,800 | 3,904 | 3,020 | 0.003983 | 0.002599 |
| | 75-79 | 770,571 | 1,074,890 | 3,952 | 3,283 | 0.005129 | 0.003054 |
| | ≥80 | 586,105 | 1,209,991 | 4,569 | 4,801 | 0.007795 | 0.003968 |
| NIS | ≤2 | 200,186 | 173,128 | 1,835 | 1,806 | 0.009166 | 0.010432 |
| | 3-5 | 100,746 | 86,834 | 1,426 | 1,397 | 0.014155 | 0.016090 |
| | 6-9 | 79,173 | 64,938 | 1,509 | 1,453 | 0.019059 | 0.022378 |
| | 10-12 | 43,869 | 31,995 | 1,604 | 1,665 | 0.036568 | 0.052038 |
| | 13-19 | 125,924 | 97,732 | 1,706 | 1,705 | 0.013549 | 0.017445 |
| | 20-24 | 109,684 | 108,181 | 1,730 | 1,752 | 0.015773 | 0.016195 |
| | 25-29 | 106,583 | 165,022 | 1,939 | 1,543 | 0.018192 | 0.009353 |
| | 30-34 | 112,118 | 241,557 | 1,968 | 1,356 | 0.017552 | 0.005615 |
| | 35-39 | 151,530 | 230,518 | 2,089 | 1,858 | 0.013788 | 0.008058 |
| | 40-44 | 156,333 | 179,396 | 2,292 | 2,261 | 0.014659 | 0.012606 |
| | 45-49 | 195,720 | 226,763 | 2,467 | 2,343 | 0.012605 | 0.010331 |
| | 50-54 | 219,249 | 256,770 | 2,668 | 2,358 | 0.012170 | 0.009182 |
| | 55-59 | 276,240 | 302,891 | 2,717 | 2,282 | 0.009835 | 0.007535 |
| | 60-64 | 251,387 | 260,820 | 2,793 | 2,287 | 0.011110 | 0.008769 |
| | 65-69 | 196,671 | 201,035 | 2,542 | 2,097 | 0.012927 | 0.010431 |
| | 70-74 | 170,319 | 179,911 | 2,545 | 2,097 | 0.014943 | 0.011657 |
| | 75-79 | 161,005 | 215,945 | 2,496 | 2,307 | 0.015500 | 0.010684 |
| | ≥80 | 177,550 | 383,367 | 2,843 | 2,968 | 0.016011 | 0.007741 |
| PPS | ≤2 | 758,733 | 716,629 | 2,073 | 2,052 | 0.002732 | 0.002863 |
| | 3-5 | 887,767 | 843,281 | 1,578 | 1,530 | 0.001777 | 0.001814 |
| | 6-9 | 1,137,869 | 1,073,031 | 1,687 | 1,790 | 0.001483 | 0.001668 |
| | 10-12 | 851,996 | 803,075 | 2,208 | 2,137 | 0.002592 | 0.002661 |
| | 13-19 | 1,962,660 | 1,876,450 | 2,565 | 2,210 | 0.001307 | 0.001178 |
| APS | ≤69 | 1,310,637 | 1,436,561 | 3,744 | 2,865 | 0.002857 | 0.001994 |
| | 70-74 | 980,208 | 1,161,800 | 3,904 | 3,020 | 0.003983 | 0.002599 |
| | 75-79 | 770,571 | 1,074,890 | 3,952 | 3,283 | 0.005129 | 0.003054 |
| | ≥80 | 586,105 | 1,209,991 | 4,569 | 4,801 | 0.007795 | 0.003968 |

NPS, national patient sample; NIS, national inpatient sample; PPS, pediatric patient sample; APS, aged population sample.

## 2. 표본의 일반적 현황

소규모에서 NPS, NIS는 위 추출비율 중 최대값으로 하였을 때 약 40만 정도의 규모이고, APS와 PPS는 일관성을 위해 추출비율 중 최대값보다 더 크게 늘려 APS는 0.042945, PPS는 0.054565로 설정하여 표본 추출하였다. 대규모 데이터의 경우 NPS는 100만 규모, NIS와 PPS, APS는 기존 표본자료와 동일한 수치로 설정하여 NPS는 0.021, NIS, PPS, APS는 각 0.1로 추출 진행하였다. 진행된 표본에서의 환자 수 및 명세서 건수는 표 2와 같다.

## 3. 대표성 검증

NPS, NIS, PPS, APS의 소규모, 대규모 표본에 있어 상병 및 1인당 총의료비의 대표성을 검증하였다. 모집단의 통계값인 실제 수치가 표본자료의 신뢰구간에 포함이 된다면 대표성이 있다고 판단할 수 있다. 상병 중 다빈도와 저빈도 상병을 1개씩 선정하여 대표성을 확인하였고 표 3에 제시하였다.

NPS, NIS, PPS, APS의 소규모, 대규모 표본의 성별, 연령별 1인당 총의료비의 대표성을 검증하였고, 표 4–7에 각각 제시하였다. NPS의 경우, 소규

**표 2.** 표본자료의 데이터 수

| 표본 | 소규모 | | 대규모 | |
|---|---|---|---|---|
| | 환자 수(명) | 명세서 건수(건) | 환자 수(명) | 명세서 건수(건) |
| NPS | 385,454 | 7,638,865 | 1,038,388 | 20,546,234 |
| NIS | 394,910 | 13,701,712 | 758,870 | 26,347,433 |
| PPS | 408,054 | 6,800,288 | 950,169 | 15,793,300 |
| APS | 394,682 | 15,734,081 | 723,321 | 28,841,428 |

NPS, national patient sample; NIS, national inpatient sample; PPS, pediatric patient sample; APS, aged population sample.

**표 3.** 다빈도 및 저빈도 상병 대표성 검증

| 표본 | 상병 | 실제(건) | 소규모 | | | 대규모 | | |
|---|---|---|---|---|---|---|---|---|
| | | | 추정 | 신뢰구간 | | 추정 | 신뢰구간 | |
| | | | | 하한 | 상한 | | 하한 | 상한 |
| NPS | K05 | 16,232,820 | 16,265,564 | 16,194,422 | 16,336,705 | 16,236,848 | 16,193,520 | 16,280,176 |
| | C80 | 5,885 | 6,286 | 4,526 | 8,045 | 5,048 | 4,087 | 6,008 |
| NIS | A09 | 1,133,011 | 1,135,503 | 1,127,436 | 1,143,570 | 1,132,980 | 1,127,163 | 1,138,797 |
| | C78 | 12,908 | 13,125 | 12,142 | 14,108 | 13,230 | 12,518 | 13,942 |
| PPS | J30 | 2,807,208 | 2,800,148 | 2,787,323 | 2,812,973 | 2,803,517 | 2,795,113 | 2,811,922 |
| | C91 | 2,663 | 2,631 | 2,146 | 3,116 | 2,750 | 2,425 | 3,075 |
| APS | I10 | 2,795,066 | 2,797,204 | 2,786,256 | 2,808,152 | 2,800,206 | 2,792,118 | 2,808,295 |
| | C78 | 7,358 | 7,147 | 6,438 | 7,856 | 7,620 | 7,079 | 8,161 |

NPS, national patient sample; NIS, national inpatient sample; PPS, pediatric patient sample; APS, aged population sample.

**표 4.** NPS 표본 성별, 연령별 1인당 총의료비 대표성 검증

| 연령(세) | 성별 | 실제(원) | 소규모 | | | 대규모 | | |
|---|---|---|---|---|---|---|---|---|
| | | | 추정 | 하한 | 상한 | 추정 | 하한 | 상한 |
| ≤2 | 남 | 1,322,302 | 1,453,272 | 1,197,810 | 1,708,735 | 1,391,095 | 1,255,208 | 1,526,982 |
| | 여 | 1,173,350 | 1,231,344 | 1,052,635 | 1,410,052 | 1,168,563 | 1,068,155 | 1,268,972 |
| 3-5 | 남 | 733,924 | 709,332 | 677,806 | 740,858 | 738,289 | 694,678 | 781,899 |
| | 여 | 671,396 | 663,243 | 632,169 | 694,318 | 722,105 | 644,989 | 799,221 |
| 6-9 | 남 | 542,950 | 558,833 | 490,093 | 627,572 | 543,515 | 507,122 | 579,908 |
| | 여 | 533,463 | 512,236 | 496,056 | 528,416 | 522,641 | 506,211 | 539,071 |
| 10-12 | 남 | 460,592 | 459,516 | 413,492 | 505,539 | 435,666 | 403,173 | 468,158 |
| | 여 | 382,659 | 356,036 | 335,886 | 376,186 | 390,748 | 349,368 | 432,128 |
| 13-19 | 남 | 467,012 | 442,143 | 410,711 | 473,575 | 507,013 | 449,390 | 564,635 |
| | 여 | 388,055 | 378,898 | 362,418 | 395,378 | 389,456 | 374,195 | 404,717 |
| 20-24 | 남 | 477,387 | 499,964 | 402,072 | 597,856 | 473,840 | 450,790 | 496,890 |
| | 여 | 496,200 | 495,035 | 471,680 | 518,391 | 484,415 | 471,389 | 497,442 |
| 25-29 | 남 | 502,568 | 583,750 | 394,206 | 773,294 | 498,456 | 468,951 | 527,961 |
| | 여 | 702,024 | 677,819 | 650,341 | 705,297 | 693,885 | 673,286 | 714,483 |
| 30-34 | 남 | 581,526 | 553,662 | 504,377 | 602,948 | 628,955 | 519,464 | 738,446 |
| | 여 | 964,073 | 963,280 | 921,898 | 1,004,662 | 957,165 | 931,332 | 982,999 |
| 35-39 | 남 | 681,893 | 732,967 | 634,606 | 831,328 | 673,073 | 638,013 | 708,133 |
| | 여 | 897,088 | 895,233 | 823,791 | 966,675 | 917,905 | 887,527 | 948,284 |
| 40-44 | 남 | 824,574 | 779,297 | 735,674 | 822,919 | 865,217 | 821,812 | 908,622 |
| | 여 | 892,348 | 896,995 | 844,835 | 949,155 | 892,543 | 859,030 | 926,055 |
| 45-49 | 남 | 1,011,337 | 977,615 | 925,476 | 1,029,754 | 1,016,375 | 975,215 | 1,057,535 |
| | 여 | 1,045,452 | 1,051,535 | 991,740 | 1,111,329 | 1,027,281 | 999,067 | 1,055,494 |
| 50-54 | 남 | 1,295,986 | 1,265,238 | 1,185,552 | 1,344,925 | 1,273,605 | 1,227,215 | 1,319,996 |
| | 여 | 1,309,168 | 1,287,535 | 1,230,760 | 1,344,311 | 1,302,634 | 1,264,509 | 1,340,759 |
| 55-59 | 남 | 1,662,184 | 1,727,430 | 1,638,167 | 1,816,692 | 1,644,394 | 1,595,691 | 1,693,096 |
| | 여 | 1,567,572 | 1,555,387 | 1,493,564 | 1,617,211 | 1,579,303 | 1,540,361 | 1,618,246 |
| 60-64 | 남 | 2,130,135 | 2,194,704 | 2,083,513 | 2,305,896 | 2,118,668 | 2,050,616 | 2,186,721 |
| | 여 | 1,918,699 | 1,911,169 | 1,829,666 | 1,992,673 | 1,910,925 | 1,857,406 | 1,964,444 |
| 65-69 | 남 | 2,990,566 | 2,882,671 | 2,742,222 | 3,023,120 | 2,972,068 | 2,882,546 | 3,061,590 |
| | 여 | 2,665,034 | 2,725,941 | 2,610,891 | 2,840,990 | 2,636,517 | 2,563,943 | 2,709,091 |
| 70-74 | 남 | 3,665,582 | 3,881,361 | 3,655,913 | 4,106,809 | 3,628,779 | 3,510,891 | 3,746,668 |
| | 여 | 3,263,979 | 3,246,704 | 3,094,603 | 3,398,805 | 3,296,855 | 3,200,558 | 3,393,152 |
| 75-79 | 남 | 4,166,596 | 4,253,642 | 4,006,024 | 4,501,259 | 4,139,138 | 3,996,325 | 4,281,952 |
| | 여 | 4,034,107 | 4,044,903 | 3,862,518 | 4,227,287 | 4,072,169 | 3,957,638 | 4,186,701 |
| ≥80 | 남 | 5,102,505 | 5,163,580 | 4,846,841 | 5,480,318 | 5,117,117 | 4,936,382 | 5,297,851 |
| | 여 | 5,280,912 | 5,267,678 | 5,066,161 | 5,469,196 | 5,339,026 | 5,216,332 | 5,461,719 |

NPS, national patient sample.

**표 5.** NIS 표본 성별, 연령별 1인당 총의료비 대표성 검증

| 연령(세) | 성별 | 실제(원) | 소규모 | | | 대규모 | | |
|---|---|---|---|---|---|---|---|---|
| | | | 추정 | 하한 | 상한 | 추정 | 하한 | 상한 |
| ≤2 | 남 | 3,083,317 | 3,108,995 | 2,809,719 | 3,408,270 | 3,160,217 | 2,989,859 | 3,330,576 |
| | 여 | 2,900,984 | 2,836,512 | 2,613,786 | 3,059,238 | 2,837,314 | 2,675,626 | 2,999,003 |
| 3-5 | 남 | 2,051,805 | 2,120,317 | 1,934,264 | 2,306,371 | 2,015,049 | 1,924,036 | 2,106,062 |
| | 여 | 1,974,168 | 2,086,719 | 1,855,716 | 2,317,723 | 1,974,570 | 1,895,648 | 2,053,493 |
| 6-9 | 남 | 2,033,037 | 2,056,861 | 1,870,567 | 2,243,155 | 2,042,561 | 1,923,296 | 2,161,825 |
| | 여 | 1,991,468 | 1,899,812 | 1,796,787 | 2,002,838 | 2,013,343 | 1,847,932 | 2,178,755 |
| 10-12 | 남 | 2,288,919 | 2,005,456 | 1,774,761 | 2,236,152 | 2,371,823 | 2,039,812 | 2,703,835 |
| | 여 | 2,209,595 | 2,211,895 | 1,807,071 | 2,616,719 | 2,416,130 | 2,020,105 | 2,812,155 |
| 13-19 | 남 | 2,724,124 | 2,695,082 | 2,443,880 | 2,946,285 | 2,636,511 | 2,499,121 | 2,773,902 |
| | 여 | 2,376,644 | 2,383,844 | 2,227,721 | 2,539,967 | 2,423,582 | 2,303,976 | 2,543,188 |
| 20-24 | 남 | 2,699,991 | 2,804,289 | 2,647,138 | 2,961,441 | 2,813,188 | 2,652,342 | 2,974,034 |
| | 여 | 2,484,728 | 2,430,241 | 2,328,938 | 2,531,544 | 2,498,575 | 2,404,159 | 2,592,992 |
| 25-29 | 남 | 2,820,627 | 2,783,661 | 2,628,785 | 2,938,537 | 2,847,395 | 2,735,191 | 2,959,599 |
| | 여 | 2,838,457 | 2,826,234 | 2,729,252 | 2,923,215 | 2,836,817 | 2,737,963 | 2,935,672 |
| 30-34 | 남 | 3,014,368 | 3,078,189 | 2,904,887 | 3,251,490 | 3,091,308 | 2,956,061 | 3,226,555 |
| | 여 | 3,208,294 | 3,233,035 | 3,133,090 | 3,332,980 | 3,183,724 | 3,130,464 | 3,236,984 |
| 35-39 | 남 | 3,468,894 | 3,412,829 | 3,248,569 | 3,577,089 | 3,470,946 | 3,328,682 | 3,613,209 |
| | 여 | 3,589,463 | 3,571,119 | 3,479,851 | 3,662,387 | 3,637,555 | 3,555,674 | 3,719,437 |
| 40-44 | 남 | 4,236,566 | 4,317,453 | 4,124,269 | 4,510,637 | 4,206,519 | 4,073,823 | 4,339,214 |
| | 여 | 4,109,230 | 4,102,679 | 3,962,522 | 4,242,835 | 4,089,840 | 3,992,994 | 4,186,687 |
| 45-49 | 남 | 5,136,691 | 5,058,617 | 4,880,214 | 5,237,021 | 5,137,952 | 5,008,418 | 5,267,486 |
| | 여 | 4,516,914 | 4,577,213 | 4,431,532 | 4,722,893 | 4,562,075 | 4,464,789 | 4,659,361 |
| 50-54 | 남 | 5,993,194 | 6,190,556 | 5,989,368 | 6,391,744 | 5,930,617 | 5,802,996 | 6,058,239 |
| | 여 | 4,854,526 | 4,859,121 | 4,730,027 | 4,988,214 | 4,848,272 | 4,750,993 | 4,945,552 |
| 55-59 | 남 | 6,790,560 | 6,659,588 | 6,487,106 | 6,832,071 | 6,817,903 | 6,696,613 | 6,939,192 |
| | 여 | 5,323,464 | 5,273,582 | 5,153,400 | 5,393,764 | 5,280,338 | 5,193,119 | 5,367,557 |
| 60-64 | 남 | 7,834,976 | 8,038,019 | 7,830,867 | 8,245,171 | 7,775,876 | 7,637,741 | 7,914,010 |
| | 여 | 6,132,595 | 6,058,289 | 5,908,401 | 6,208,177 | 6,056,887 | 5,952,552 | 6,161,222 |
| 65-69 | 남 | 9,101,251 | 9,201,074 | 8,963,098 | 9,439,050 | 9,119,432 | 8,958,935 | 9,279,929 |
| | 여 | 7,378,341 | 7,409,823 | 7,239,492 | 7,580,153 | 7,341,618 | 7,223,179 | 7,460,057 |
| 70-74 | 남 | 10,046,251 | 10,183,799 | 9,932,430 | 10,435,169 | 9,903,032 | 9,726,165 | 10,079,900 |
| | 여 | 8,552,042 | 8,420,647 | 8,231,930 | 8,609,363 | 8,510,922 | 8,372,401 | 8,649,442 |
| 75-79 | 남 | 10,353,127 | 10,105,443 | 9,862,263 | 10,348,624 | 10,274,655 | 10,090,383 | 10,458,927 |
| | 여 | 9,738,153 | 9,928,643 | 9,735,176 | 10,122,110 | 9,705,922 | 9,568,464 | 9,843,380 |
| ≥80 | 남 | 11,021,637 | 11,049,590 | 10,799,531 | 11,299,649 | 11,005,646 | 10,825,033 | 11,186,259 |
| | 여 | 11,715,765 | 11,678,083 | 11,514,612 | 11,841,554 | 11,673,531 | 11,558,230 | 11,788,833 |

NIS, national inpatient sample.

**표 6.** PPS 표본 성별, 연령별 1인당 총의료비 대표성 검증

| 연령(세) | 성별 | 실제(원) | 소규모 | | | 대규모 | | |
|---|---|---|---|---|---|---|---|---|
| | | | 추정 | 하한 | 상한 | 추정 | 하한 | 상한 |
| ≤2 | 남 | 1,322,302 | 1,369,571 | 1,279,777 | 1,459,366 | 1,312,705 | 1,260,171 | 1,365,239 |
| | 여 | 1,173,350 | 1,223,787 | 1,141,599 | 1,305,975 | 1,205,830 | 1,149,342 | 1,262,317 |
| 3-5 | 남 | 733,924 | 716,154 | 694,180 | 738,127 | 730,976 | 714,488 | 747,464 |
| | 여 | 671,396 | 673,134 | 651,636 | 694,631 | 669,073 | 656,513 | 681,633 |
| 6-9 | 남 | 542,950 | 531,117 | 516,081 | 546,153 | 545,236 | 526,242 | 564,231 |
| | 여 | 533,463 | 548,032 | 523,843 | 572,221 | 535,585 | 521,852 | 549,319 |
| 10-12 | 남 | 460,592 | 455,561 | 419,898 | 491,224 | 446,940 | 429,893 | 463,986 |
| | 여 | 382,659 | 397,085 | 371,351 | 422,818 | 376,167 | 360,487 | 391,847 |
| 13-19 | 남 | 467,012 | 496,237 | 446,422 | 546,052 | 466,502 | 450,246 | 482,758 |
| | 여 | 388,055 | 389,531 | 372,431 | 406,631 | 395,753 | 383,302 | 408,204 |

PPS, pediatric patient sample.

**표 7.** APS 표본 성별, 연령별 1인당 총의료비 대표성 검증

| 연령(세) | 성별 | 실제(원) | 소규모 | | | 대규모 | | |
|---|---|---|---|---|---|---|---|---|
| | | | 추정 | 하한 | 상한 | 추정 | 하한 | 상한 |
| 65-69 | 남 | 2,990,566 | 2,939,154 | 2,882,797 | 2,995,512 | 2,998,595 | 2,955,275 | 3,041,914 |
| | 여 | 2,665,034 | 2,639,793 | 2,597,686 | 2,681,900 | 2,669,456 | 2,637,618 | 2,701,294 |
| 70-74 | 남 | 3,665,582 | 3,624,768 | 3,546,464 | 3,703,071 | 3,710,432 | 3,651,761 | 3,769,103 |
| | 여 | 3,263,979 | 3,301,617 | 3,240,571 | 3,362,664 | 3,252,375 | 3,210,392 | 3,294,358 |
| 75-79 | 남 | 4,166,596 | 4,132,768 | 4,041,682 | 4,223,854 | 4,205,333 | 4,137,055 | 4,273,611 |
| | 여 | 4,034,107 | 4,000,130 | 3,932,933 | 4,067,326 | 4,014,227 | 3,964,298 | 4,064,157 |
| ≥80 | 남 | 5,102,505 | 5,176,958 | 5,060,719 | 5,293,197 | 5,086,463 | 5,002,011 | 5,170,915 |
| | 여 | 5,280,912 | 5,244,668 | 5,168,822 | 5,320,514 | 5,308,977 | 5,252,931 | 5,365,022 |

APS, aged population sample.

모에서 6–9세, 10–12세 여성과 40–44세 남성의 경우 1인당 총의료비 실제값이 추정값의 신뢰구간 범위 내에 있지 않아 낮은 대표성을 확인하였고, 대규모에서는 모든 층에서 대표성을 확인하였다. NIS의 경우, 대규모에서는 모든 층에서 대표성을 확인하였으나 소규모에서 10–12세와 75–79세 남성의 경우 1인당 총의료비 실제값이 추정값의 신뢰구간 범위 내에 있지 않아 낮은 대표성을 확인하였다. PPS와 APS는 모든 성별, 연령 구간에서 1인당 총의료비 실제값이 추정값의 신뢰구간 내에 포함되기 때문에 대표성을 만족함을 확인하였다.

## 고 찰

이 연구에서는 새로운 표본 추출방법을 적용하여 소규모와 대규모의 표본자료를 구축하였다. 기존 표

본과의 차이점은 표본 추출비율을 산정하는 방식이 다르고 연령층이 다르다. 기존 표본 추출의 경우 전체 인구 모집단을 바탕으로 필요 표본의 수를 계산하고 이에 맞추어 각 연령, 성별 계층에서 표본을 추출하였다[3]. 각 연령, 성별 계층별로 대표성을 확보할 수 있는 최소 표본의 수 및 표본 추출비율을 계산하고, 층별 표본 추출비율 중 가장 큰 값을 기준으로 표본 추출을 진행하였다. 기존의 필요 표본의 수를 계산하는 방식을 적용해도 이론적으로 현재 대표성을 확보할 수 있는 수준이지만, 제안한 연구방법을 적용할 경우 다음과 같은 장점이 있다. 인구구조 및 의료이용행태가 급격하게 변하면 개별 층에서 대표변수인 의료비의 분산이 크게 변동될 여지가 있다. 이 경우 기존 표본 추출방식으로 진행하는 경우 일부 층에서 늘어난 분산을 담보할 수 있는 최소 표본의 수를 추출하지 못할 가능성이 있다. 이 연구에서 제안하는 표본 추출방식을 적용하는 경우 매번 층마다 분산을 계산하여 추출비율을 결정하는 과정을 거치지만, 층을 추가하지 않고도 대표성을 담보할 수 있다는 것이다. 더불어 인구의 저출산 고령화 추세를 반영하기 위해 저연령과 고령층에서 각각 1개 층씩 추가하여 기존 16개 연령층에서 18개 연령층으로 확대했다.

대규모 표본에서는 1인당 총의료비의 대표성이 모두 충족되었지만, NPS, NIS의 소규모 표본 중 일부 연령층에서 추정값의 신뢰구간이 실제값을 포함하지 못한 것이 확인되었다. 층별 1인당 총의료비는 의료급여, 요양병원, 희귀질환 등의 영향으로 분산이 크다. 소규모 표본에서는 무작위 표본 추출을 진행할 시 극단 치의 영향이 커, 이로 인해 대표성에서 벗어난 것으로 여겨진다. 대표성이 약간 어긋나 있다 하더라도 소규모 표본으로 자료의 접근성을 향상하는 것이 활용도를 높이는 데 도움이 될 것이기에 제안하고자 한다. 데이터 3법이 통과하고 가명 정보 방식으로 동의 없이 제3자에게 환자표본자료를 직접 제공하는 방식이 법적으로 가능하지만, 청구자료라는 특수성 때문에 소규모 표본과 대규모 표본으로 이원화하여 공개수준을 구분하는 것도 지속 가능한 데이터 제공을 위한 선택 전략일 것이다.

저빈도 상병 일부의 대표성을 확인하였지만, 무작위 추출로 구축한 표본자료라는 한계상 모집단에서 청구빈도가 적은 희귀질환의 경우 대표성이 떨어질 수 있다. 환자표본자료를 적절히 이용하기 위해서는 가급적 유병률이나 발생률이 어느 정도 수준이 되는 상병을 연구대상으로 하거나 어느 정도 규모가 있는 인구집단을 연구대상으로 선정하는 것이 중요하다.

국민건강보험공단의 표본코호트 데이터베이스의 경우 층화계통추출 비례배분법으로 표본을 구축하였다[15]. 층화 변수로는 연령 · 성별 보험료 분위 및 지역으로 총 1,476층을 포함하였다. 층은 표본코호트 2.0이 되면서 2,142층으로 더 추가하였다. 층별 대표 변수로 연간 의료비를 선택하였다. 층별 표본은 층의 모집단 의료비와 표본의 의료비의 오차비율이 5%가 넘지 않도록 반복 추출하였다. 대만의 National Health Insurance Research Database의 경우 연령, 성별, 지역을 대표 변수로 설정하였으며 2016년부터는 전체 인구 데이터셋을 제공하고 있다[16]. 위 선행연구 사례에 맞추어 소규모 표본의 대표성을 향상시키는 방법으로는 개인의 자격 정보를 연계시키는 방법이 가장 효과적일 것이다. 자격 정보에 있는 지역, 보험료 수준 및 장애 정보 등을 활용한다면 대표성을 더욱 강화할 수 있을 것이다. 의료비 변수는 왜도가 매우 높은 변수이기 때문에 층화계통추출을 사용하는 것이 효율적이라 볼 수 있다. 이 연구에서는 층화임의추출 방법을 사용하였는데, 자격 정보 없이 의료비 순으로 정렬할 경우 소득

분위 및 의료급여 등의 정보에 따라 패턴이 발생할 수 있으며, 층의 수도 표본코호트에 비해 적기 때문에 기존의 임의추출방법을 유지하였다. 미국 Agency for Healthcare Research and Quality (AHRQ)에서 제공하고 있는 National Inpatient Sample의 경우 층화계통추출을 진행하였으며, 대표 변수로 지역, 수련병원 여부, 의료기관 운영 주체 및 침상 수로 설정하였다[17]. AHRQ의 National Inpatient Sample의 경우 환자보다는 대표성 있는 입원자료 구축이 목적이기 때문에 환자표본자료와는 추구하는 방향이 달라 의료이용 변수를 추출 틀에 포함하지 않았다.

현재 우리나라에서 보건의료 관련 연구를 하는 데 있어 청구자료가 가지고 있는 가치는 상당히 높다. 청구자료 중 가장 접근성이 높은 환자표본자료를 지속적으로 관리 및 제공하는 것이 좀 더 세밀한 국가정책 수립 및 보건의료 연구를 수행하는 데 도움이 될 수 있을 것이다.

## 감사의 글

이 논문은 2020년 건강보험심사평가원의 지원을 받아 수행된 연구이다(환자표본자료 구축 방안 및 활용 실태 조사, 2020, G000FF8-2020-179).

## ORCID

ChulYoung Yoon: https://orcid.org/0000-0003-0162-1741
Jae Joon Ahn: https://orcid.org/0000-0001-7974-8027
Gyeongmin Lee: https://orcid.org/0000-0002-5052-2232
Yongseok Choi: https://orcid.org/0000-0002-6781-4522
Lihyun Kim: https://orcid.org/0000-0002-2074-0810
Dae Yoo Ha: https://orcid.org/0000-0002-6498-5051
Ki-Bong Yoo: https://orcid.org/0000-0002-2955-6948

## 참고문헌

1. 보건의료 빅데이터 플랫폼. 보건의료 빅데이터 플랫폼 사업 [Internet]. 청주: 한국보건산업진흥원; c2021 [cited 2021 Oct 10]. Available from: https://hcdl.mohw.go.kr/BD/Portal/Enterprise/DefaultPage.bzr?tabID=1093&ftab=1003.
2. 보건복지부. 보건의료분야 결합전문기관 소개[Internet]. 세종: 보건복지부; c2020 [cited 2021 Oct 10]. Available from: https://datalink.mohw.go.kr/intro.html.
3. Kim L, Sakong J, Kim Y, Kim S, Kim S, Tchoe B, et al. Developing the inpatient sample for the National Health Insurance claims data. Health Policy Manag. 2013;23(2):152-61. DOI: https://doi.org/10.4332/KJHPA.2013.23.2.152.
4. Hwang SG, Park H. An analysis on prescribing patterns of Alzheimer's dementia treatment and choline alfoscerate using HIRA claims data. Korean J Clin Pharm. 2019;29(1):1-8. DOI: https://doi.org/10.24304/kjcp.2019.29.1.1.
5. Jeon SM, Park S, Rhie SJ, Kwon JW. Prescribing patterns of polypharmacy in Korean pediatric patients. PLoS One. 2019;14(10):e0222781. DOI: https://doi.org/10.1371/journal.pone.0222781.
6. Cha YJ. The economic burden of stroke based on South Korea's national health insurance claims database. Int J Health Policy Manag. 2018;7(10):904-9. DOI: https://doi.org/10.15171/ijhpm.2018.42.
7. Rhee CK, Kim K, Yoon HK, Kim JA, Kim SH, Lee SH, et al. Natural course of early COPD. Int J Chron Obstruct Pulmon Dis. 2017;12:663-8. DOI: https://doi.org/10.2147/COPD.S122989.
8. Yuk JS, Baek JC, Park JE, Jo HC, Park JK, Cho IA. Incidence of gestational trophoblastic disease in South Korea: a longitudinal, population-based study. PeerJ. 2019;7:e6490. DOI: https://doi.org/10.7717/peerj.6490.

9. Lee JY, Lim NG, Chung CK, Lee JY, Kim HJ, Park SB. Parkinson's disease as risk factor in osteoporosis and osteoporotic vertebral fracture : prevalence study using National Inpatient Sample Database in Korea. J Korean Neurosurg Soc. 2019;62(1):71-82. DOI: https://doi.org/10.3340/jkns.2018.0012.
10. 건강보험심사평가원. 2012년 진료비통계지표. 원주: 건강보험심사평가원; 2013.
11. 건강보험심사평가원. 2020년 진료비통계지표. 원주: 건강보험심사평가원; 2021.
12. 통계청. 인구동향조사. 대전: 통계청; 2021.
13. 통계청. 장래인구조사. 대전: 통계청; 2021.
14. 이수연, 문용필. 국민건강보험의 노인의료비 지출추계 및 장기재정 전망. 비판사회정책. 2018;58:53-93. DOI: https://doi.org/10.47042/ACSW.2018.02.58.53.
15. Lee J, Lee JS, Park SH, Shin SA, Kim K. Cohort profile: the National Health Insurance Service-National Sample Cohort (NHIS-NSC), South Korea. Int J Epidemiol. 2017;46(2):e15. DOI: https://doi.org/10.1093/ije/dyv319.
16. Lin LY, Warren-Gash C, Smeeth L, Chen PC. Data resource profile: the National Health Insurance Research Database (NHIRD). Epidemiol Health. 2018;40:e2018062. DOI: https://doi.org/10.4178/epih.e2018062.
17. Healthcare Cost and Utilization Project. HCUP sample design: national databases-accessible version [Internet]. Rockville (MD): Agency for Healthcare Research and Quality; 2018 [cited 2021 Nov 1]. Available from: https://www.hcup-us.ahrq.gov/tech_assist/sampledesign/508_compliance/index508_2018.jsp#nissample.