

연구보고서(수시) 2023-02

패널조사 표본이탈 경향 분석과 추가표본 추출에 관한 연구

- 한국의료패널조사를 중심으로

이혜정
변종석·황안나·염아림



KOREA INSTITUTE FOR HEALTH AND SOCIAL AFFAIRS



한국보건사회연구원
KOREA INSTITUTE FOR HEALTH AND SOCIAL AFFAIRS



■ 연구진

연구책임자	이혜정	한국보건사회연구원 부연구위원
공동연구진	변종석	한신대학교 교수
	황안나	한국보건사회연구원 전문연구원
	염아림	한국보건사회연구원 전문원

연구보고서(수시) 2023-02

패널조사 표본이탈 경향 분석과 추가표본 추출에 관한 연구

- 한국의료패널조사를 중심으로

발행일 2023년 12월
발행인 이태수
발행처 한국보건사회연구원
주소 [30147]세종특별자치시 시청대로 370
세종국책연구단지 사회정책동(1~5층)
전화 대표전화: 044)287-8000
홈페이지 <http://www.kihasa.re.kr>
등록 1999년 4월 27일(제2015-000007호)
인쇄처 (주)정인애드

© 한국보건사회연구원 2023
ISBN 978-89-6827-952-2 [93510]
<https://doi.org/10.23060/kihasa.b.2023.02>

발|간|사

2기 한국의료패널조사도 어느덧 5년 동안 실시됨에 따라 패널가구가 점점 줄어드는 추세가 나타났다. 이렇듯 패널 마모가 지속적으로 발생하게 된다면 표본 대표성에 문제가 생길 수 있으며 더 나가서는 데이터 품질에 대한 신뢰성도 담보하기 어렵게 될 수 있다. 이러한 상황에 대비하기 위하여 표본 구성에 대한 다각적이고 면밀한 검토가 필요한 상황이라고 볼 수 있다.

이 연구에서는 최근 조사가 완료된 연간데이터를 활용하여 응답형태에 따른 패널가구와 패널가구원의 표본이탈의 특성을 파악하였고, 표본이탈에 영향을 주는 요인에 대해 살펴보았다. 또 현시점에서의 패널 구조를 진단하여 표본 대표성을 만족할 수 있도록 추가표본 추출 방안을 마련하였다.

이 보고서는 이해정 부연구위원의 책임 하에 황안나 전문연구원, 염아립 전문원이 원내 연구진으로 참여하였다. 외부 연구진으로 한신대학교 변종석 교수가 참여하였다. 모든 연구진의 노고에 감사드린다. 보고서 작성과 관련하여 유익한 의견을 주신 원내 배재용 연구위원, 원외 농촌진흥청 고현석 농업연구원, 그리고 익명의 검토위원들에게도 감사의 마음을 전한다. 마지막으로, 본 보고서의 내용은 본원의 공식적인 의견이 아님을 밝힌다.

2023년 11월

한국보건사회연구원 원장

이 태 수



Abstract	1
요 약	5
제1장 서론	15
제1절 연구 배경 및 목적	17
제2절 연구 내용 및 방법	20
제2장 패널 대상 기초 분석	23
제1절 2기 한국의료패널조사 개요 및 완료 현황	25
제2절 응답 여부에 따른 패널 대상 특성 분석	29
제3절 소결	40
제3장 패널 대상의 표본이탈 특성 분석	43
제1절 표본이탈의 정의	45
제2절 패널가구의 표본이탈 특성 분석	47
제3절 패널가구원의 표본이탈 특성 분석	57
제4절 소결	72
제4장 2기 한국의료패널조사 추가표본 추출 방안	77
제1절 모집단 및 층화변수 현황 분석	79
제2절 2기 한국의료패널조사의 주요 변수 분석	86
제3절 추가표본 추출을 위한 표본설계	95
제4절 소결	121



제5장 결론	125
제1절 주요 연구 결과 요약	127
제2절 패널 유지를 위한 패널 대상의 관리 전략	135
참고문헌	137

표 목차

KOREA INSTITUTE FOR HEALTH AND SOCIAL AFFAIRS



〈표 1-1〉 한국의료패널조사 조사 차수 기준 완료 현황	19
〈표 2-1〉 2기 한국의료패널조사 연간데이터 기준 완료 현황	27
〈표 2-2〉 2기 한국의료패널조사 연간데이터 기준 - 원표본 유지율	28
〈표 2-3〉 패널 원가구 조사 완료 여부에 따른 분포-빈도	38
〈표 2-4〉 패널가구의 조사 미완료 사유	40
〈표 3-1〉 패널가구의 응답형태별 분포	48
〈표 3-2〉 패널가구의 응답형태별 거주 지역 분포	49
〈표 3-3〉 패널가구의 응답형태별 동부/읍면부 분포	50
〈표 3-4〉 패널가구의 응답형태별 가구원 수 분포	50
〈표 3-5〉 패널가구의 응답형태별 세대구성 분포	51
〈표 3-6〉 패널가구의 응답형태별 주거형태 분포	52
〈표 3-7〉 패널가구의 응답형태별 연간 가구 총소득 분포	52
〈표 3-8〉 패널가구의 응답형태에 대한 다항 로지스틱 회귀모형 분석 결과	55
〈표 3-9〉 패널가구원의 응답형태별 분포	58
〈표 3-10〉 패널가구원의 응답형태별 성별 분포	58
〈표 3-11〉 패널가구원의 응답형태별 연령대 분포	59
〈표 3-12〉 패널가구원의 응답형태별 혼인상태 분포	60
〈표 3-13〉 패널가구원의 응답형태별 최종 학력 분포	60
〈표 3-14〉 패널가구원의 응답형태별 경제활동 상태 분포	61
〈표 3-15〉 패널가구 및 패널가구원의 응답형태별 분포	62
〈표 3-16〉 4개 집단별 패널가구원의 응답형태별 성별 분포	63
〈표 3-17〉 4개 집단별 패널가구원의 응답형태별 연령대 분포	64
〈표 3-18〉 4개 집단별 패널가구원의 응답형태별 혼인상태 분포	65
〈표 3-19〉 4개 집단별 패널가구원의 응답형태별 최종 학력 분포	65
〈표 3-20〉 4개 집단별 패널가구원의 응답형태별 경제활동 상태 분포	66
〈표 3-21〉 패널가구원의 응답형태에 대한 다항 로지스틱 회귀모형 분석 결과	70
〈표 3-22〉 패널가구의 다항 로지스틱 회귀모형 분석 결과 요약	73

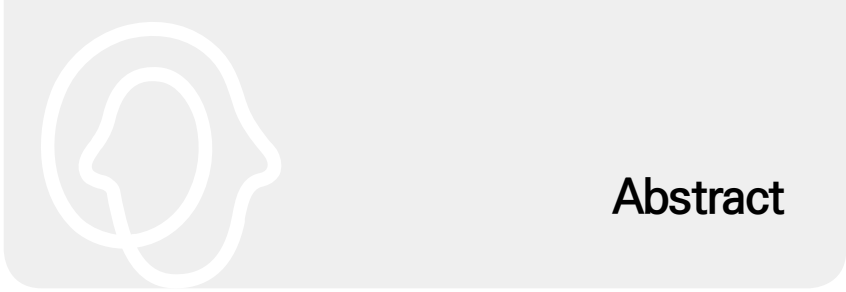
〈표 3-23〉 패널가구원의 다항 로지스틱 회귀모형 분석 결과 요약	76
〈표 4-1〉 2021년 인구주택총조사 표본추출틀의 조사구 특성별 비중	81
〈표 4-2〉 조사구 내 평균 가구수 및 표준오차	81
〈표 4-3〉 17개 시도 및 조사구 특성에 따른 가구 현황	83
〈표 4-4〉 17개 시도 및 동부·읍면부 특성에 따른 가구 현황	85
〈표 4-5〉 17개 시도 및 동부읍면부의 가구 현황	87
〈표 4-6〉 17개 시도의 만성질환 수 평균 및 표준오차	88
〈표 4-7〉 동부읍면부의 만성질환 수 평균 및 표준오차	89
〈표 4-8〉 17개 시도 및 동부읍면부의 만성질환 수 상대표준오차	90
〈표 4-9〉 17개 시도의 연간 가구 총소득 평균 및 표준오차	92
〈표 4-10〉 동부읍면부의 연간 가구 총소득 평균 및 표준오차	93
〈표 4-11〉 17개 시도 및 동부읍면부의 연간 가구 총소득 상대표준오차	94
〈표 4-12〉 주요 변수에 대한 예상 상대표준오차	96
〈표 4-13〉 17개 시도에 대한 조사구 배분 결과	98
〈표 4-14〉 17개 시도에 대한 동부읍면부 조사구 배분 결과	100
〈표 4-15〉 17개 시도 및 동부읍면부에 대한 최종 조사구 배분 결과	101
〈표 4-16〉 17개 시도 및 동부읍면부의 목표 가구수	102
〈표 4-17〉 통합표본을 구축을 위한 과부족 현황	103
〈표 4-18〉 17개 시도 및 동부읍면부의 최종 추출 가구수	104
〈표 4-19〉 17개 시도 및 동부읍면부의 최종 추출 조사구 수	105
〈표 4-20〉 통합패널의 최종 목표 가구수	106
〈표 4-21〉 통합패널의 17개 시도에 대해 연간 가구 총소득으로 파악한 목표 CV	109
〈표 4-22〉 통합패널의 17개 시도에 대해 만성질환 수로 파악한 목표 CV	110
〈표 4-23〉 통합패널의 17개 시도 및 동부읍면부에 대해 연간 가구 총소득으로 파악한 목표 CV	111
〈표 4-24〉 통합패널의 17개 시도 및 동부읍면부에 대해 만성질환 수로 파악한 목표 CV	112
〈표 5-1〉 패널가구의 다항 로지스틱 회귀모형 분석 결과 요약	130
〈표 5-2〉 패널가구원의 다항 로지스틱 회귀모형 분석 결과 요약	133

그림 목차

KOREA INSTITUTE FOR HEALTH AND SOCIAL AFFAIRS



[그림 2-1] 유치조사, 2019년~2021년 연간데이터의 가구 특성 - 거주 지역	30
[그림 2-2] 유치조사, 2019년~2021년 연간데이터의 가구 특성 - 동부/읍면부	31
[그림 2-3] 유치조사, 2019년~2021년 연간데이터의 가구 특성 - 가구원 수	32
[그림 2-4] 유치조사, 2019년~2021년 연간데이터의 가구 특성 - 세대구성	32
[그림 2-5] 유치조사, 2019년~2021년 연간데이터의 가구 특성 - 주택유형	33
[그림 2-6] 유치조사, 2019년~2021년 연간데이터의 가구원 특성 - 성별	33
[그림 2-7] 유치조사, 2019년~2021년 연간데이터의 가구원 특성 - 연령대	34
[그림 2-8] 유치조사, 2019년~2021년 연간데이터의 가구원 특성 - 혼인상태	35
[그림 2-9] 유치조사, 2019년~2021년 연간데이터의 가구원 특성 - 최종 학력	35
[그림 2-10] 유치조사, 2019년~2021년 연간데이터의 가구원 특성 - 경제활동 상태	36
[그림 2-11] 패널가구의 유지 및 이탈 흐름도	37
[그림 2-12] 패널가구 조사 완료 여부에 따른 분포-비율	39
[그림 3-1] 패널조사의 무응답 구성	46



Abstract

A Study of Sample Attrition and Supplementary Samples in Panel Surveys

– Based on the Korea Health Panel Survey

Project Head: Lee, Hyejung

This study used data from the 2nd Korea Health Panel Survey to evaluate survey completion rates and analyze the characteristics of participating households and their members. We identified key characteristics of panel households and sample attrition among household members based on response types. We also looked at factors affecting sample dropout, diagnosed the current panel structure, and proposed a sample design to extract additional samples.

The following are the findings of our variable-focused analysis of the characteristics of sample attribution among panel households. Daegu and Incheon had a high sample attrition rate in 2020 and 2021. As the number of household members increased, the rate of attrition increased. The sample attrition rate was higher for households living in apartments. However, sample attrition among panel households living in detached houses was low. In 2020, the higher the total annual household income of the household, the higher the sample attrition rate. In 2021, panel households with an annual income

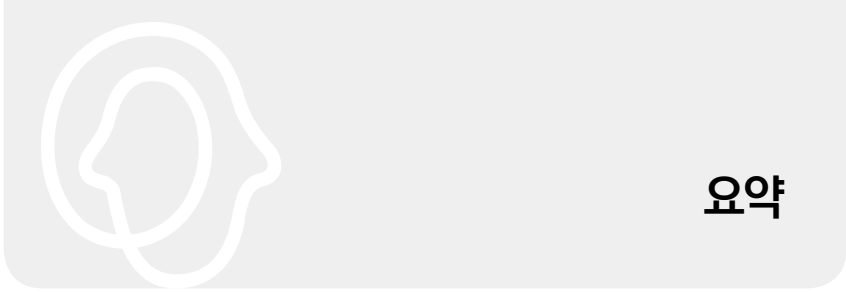
over 60 million won had the highest sample attribution rate, and those less than 20 million won the next highest.

A multinomial logistic regression model was used to analyze three response-type groups as dependent variables. The three groups were 'continuous response', 'sample dropout in 2020', and 'sample dropout in 2021'. Factors that commonly affect the sample dropout groups in 2020 and 2021 are as follows. It was found that panel households living in apartments rather than detached houses, moving less often, and having younger household heads were more likely to be in the sample dropout group. In addition, the sample dropout group in 2021 was found to affect sample attrition more among low-income households than among middle-income households and more among households with a head with high school education than among households with a head with elementary school education.

From a statistical perspective, we examined sample attrition among panel households in various ways. Our research results suggest four key areas for focused panel management to minimize sample dropout. First, there is a need for continuous tracking and management of the panel households on the move. Second, panel management is necessary for those living in apartments. Third, measures are necessary in panel management to prevent families consisting of middle-aged parents and children from dropping out of the sample. Lastly,

management is necessary for young and unmarried panel household members.

Keyword : Korea Health Panel Survey, sample attrition, supplementary samples, sampling design



1. 연구의 배경 및 목적

패널조사에서 패널 대상의 이탈이 일어나면 표본이 모집단을 대표하지 못하게 됨에 따라 추정의 편이가 발생하는 등 잘못된 분석 결과를 도출할 수 있게 되어 패널조사에 대한 신뢰도 하락으로 이어질 수 있다. 이를 해결하기 위한 첫 번째 방안은 패널조사의 표본이탈 현상을 파악하고 표본 유지를 위한 대책을 마련하는 것이다. 그다음 방안은 패널조사가 일정한 시점이 지나면 표본을 추가하는 것이다.

한국의료패널조사는 2018년부터 2기 패널을 구축하기 위한 준비를 하였고, 2019년 사전 조사를 시작으로 매해 조사를 실시하고 있으며, 올해 8월에 2023년 조사를 완료하였다. 조사 차수가 지남에 따라 패널가구가 점점 줄어드는 추세가 나타났다. 이렇듯 2기 한국의료패널조사도 지속적으로 패널 마모가 발생하게 된다면 표본 대표성에 문제가 생길 수 있다. 더 나가서는 데이터 품질에 대한 신뢰성도 담보하기 어렵게 될 수 있다. 또한, 패널조사의 추가표본에 대한 유지조사를 실시하기 위해서는 최소 1년 이상의 준비 기간이 필요하고, 사전 조사를 거친 후에 1차 조사가 진행되어야 하는 등의 상황도 고려해야 한다. 이에 5년간의 조사 데이터를 분석하여 도출된 표본이탈 추세를 고려하여 2~3년 이후의 표본이탈 및 표본유지 상황을 예측한 뒤 표본을 추가해야 한다. 이를 위하여 2기 한국의료패널조사의 현시점에서의 표본 구성에 대한 다각적이고 면밀한 검토가 필요한 상황이라고 볼 수 있다.

이 연구의 목적은 최근 조사가 완료된 데이터를 활용하여 2기 한국의료패널의 조사 완료 현황 및 패널 대상자 특성을 분석하는 것이다. 그리고

응답형태에 따라 패널가구와 패널가구원의 표본이탈의 특성을 파악하고, 표본이탈에 영향을 주는 요인을 살펴보고자 한다. 마지막으로 현시점에서 패널 구조를 진단한 다음에 표본 대표성을 만족할 수 있도록 추가 표본 추출 방안을 마련하고자 한다.

2. 주요 연구 결과

2기 한국의료패널의 원표본 유지율은 2020년 연간데이터 기준으로 91.3%이고, 2021년은 86%로 하락하는 추이를 보였다. 1기 한국의료패널보다 높은 수준을 유지하고 있었다. 타 패널의 3차 연도 원표본 유지율과 비교했을 때 한국노동패널의 09통합표본과 18통합표본에 비해서는 낮으나 한국복지패널과 장애인삶패널과는 비슷한 수준으로 나타났다.

패널가구의 표본이탈 특성을 주요 변수를 중심으로 살펴본 결과, 거주 지역이 대구와 인천이면 2020년과 2021년에 이탈 비율이 높았다. 세종의 경우 2020년에 이탈 비율이 가장 높게 나타났다. 가구원 수가 많아질수록 이탈 비율이 높았다. 아파트에 거주하면 이탈 비율이 높은 편이었으나, 단독주택은 낮은 편에 속하였다. 연간 가구 총소득은 2020년 이탈의 경우 소득이 높을수록 이탈 비율이 높은 반면에, 2021년 이탈의 경우 6,000만원 이상이 가장 높았고, 그다음은 2,000만원 미만으로 나타났다.

패널가구의 응답형태에 따른 3개 집단(연속응답, 2020년 이탈, 2021년 이탈)을 종속 변수로 하여 다항 로지스틱 회귀모형을 분석하였다. 사용한 설명 변수는 2019년 연간데이터에서의 가구 특성(거주 지역, 주거형태, 가구원 수, 연간 가구 총소득), 2019년~2021년에 걸친 가구의 상태 변화(이사 횟수), 그리고 2019년에서의 가구주 특성(성별, 연령, 혼인상태, 최종 학력)이다. <요약표 1>은 이 분석 결과를 요약한 것으로, 연속응답을

준거집단으로 하여 이탈 집단에 영향을 주는 요인, 즉 설명 변수 중에서 통계적으로 유의한 결과를 나타낸 변수를 정리하였다. 2020년과 2021년 이탈 집단에 공통으로 영향을 주는 요인을 살펴보면, 주거형태가 아파트인 경우 단독주택에 비해 이탈 가능성이 높았고, 이사 횟수가 적을수록, 가구주가 젊을수록 이탈 집단에 있을 가능성이 높은 것으로 나타났다. 이외에도 2021년 이탈 집단은 연간 가구 총소득이 저소득인 경우 중소득에 비해, 가구주의 최종 학력이 고졸인 경우 초졸에 비해 표본이탈에 영향을 주었다.

〈요약표 1〉 패널가구의 다항 로지스틱 회귀모형 분석 결과 요약

변수	2020년 이탈	2021년 이탈
거주 지역_서울		
주거형태_아파트	단독주택 (-) 다세대주택 (+)	단독주택 (-)
가구원 수		
연간 가구 총소득_중소득		저소득 (+)
이사 횟수	(-)	(-)
가구주 성별_남성		
가구주 연령	(-)	(-)
가구주 혼인상태_유배우자		
가구주 최종 학력_고졸		초졸 (-)

주: 1) 괄호 안의 내용은 각 변수에 해당하는 계수(coefficient)값의 부호를 의미함.

2) 연간 가구 총소득에서 저소득은 0원~2,000만 원 미만, 중소득은 2,000만 원~5,000만 원 미만, 고소득은 5,000만 원 이상을 의미함.

자료: 한국보건사회연구원·국민건강보험공단. (2023). 2기 한국의료패널조사 2019년~2021년 연간데이터(2023년 학술대회용 버전)를 분석함.

다음으로 패널가구원의 표본이탈 특성을 주요 변수를 중심으로 살펴 보았다. 성별의 경우 2020년 이탈에서는 남성과 여성의 비율이 비슷했으나, 2021년 이탈에서는 남성이 여성에 비해 높게 나타났다. 연령대의 경우 50세 이상~79세 이하가 표본이탈 비율이 낮은 편이었고, 49세 이하가

높은 편으로 나타났다. 한편, 80세 이상의 경우에는 연속응답 비율은 낮고 이탈 비율은 높은 경향을 보였는데, 이는 노환, 사망 등으로 인한 표본 이탈로 볼 수 있다. 미혼의 경우 다른 혼인상태에 비해 표본이탈의 경향이 높은 편이었다. 초졸과 중졸의 표본이탈 비율은 낮은 반면에, 고졸 이상 일수록 비율이 높아지는 편이었다. 한편, 미취학 포함인 무학의 경우도 이탈 비율이 높게 나타났는데, 이는 고학력 부모가 표본이탈하면서 미취학인 자녀도 함께 이탈한 것으로 추정할 수 있다. 경제활동 상태에 따라서는 15세 미만 응답자인 '해당 없음'과 비경제활동인구의 표본이탈 비율이 높은 데 비해서, 기타와 고용주/자영업자의 비율은 낮았다.

한편, 패널가구와 패널가구원의 응답형태를 조합하여 4개 집단의 주요 특성을 살펴보았다. 패널조사 시에 패널가구와 패널가구원이 모두 응답한 집단이 81.4%로 가장 높았다. 다음으로 패널가구와 패널가구원이 모두 무응답한 집단은 15.8%이고, 패널가구는 응답하였으나 패널가구원이 무응답한 집단은 2.4%였으며, 재진입 등 기타는 0.4%로 미미하였다. 이를 통해 패널가구의 응답 여부가 패널을 유지하는 데 중요하다는 것을 알 수 있다. 패널가구는 응답하였으나 패널가구원은 무응답한 집단의 인구사회학적 특성을 보면 남성이 여성에 비해, 80세 이상이 다른 연령대에 비해, 사별/실종이 다른 혼인상태에 비해, 초졸과 중졸인 경우가 다른 최종 학력에 비해, 비경제활동인구가 다른 경제활동 상태에 비해 조사에 참여하지 않을 가능성이 높은 편으로 나타났다. 패널가구 내에서 고령의 가구원은 패널에서 이탈할 가능성이 높은 편이라고 볼 수 있다. 패널가구와 패널가구원 모두 무응답한 집단의 인구사회학적 특성을 보면, 성별에 따른 차이는 없는 편이었다. 0세 이상~19세 이하 및 40세 이상~49세 이하가 다른 연령대에 비해, 미혼이 다른 혼인상태에 비해, 미취학이 포함된 무학과 대졸 이상이 다른 최종 학력에 비해, 기타를 제외한 나머지 경제

활동 상태가 조사에 참여하지 않을 가능성이 높은 편으로 나타났다. 미혼 이거나, 중년의 부모와 자녀로 구성된 가족이 패널에서 이탈할 가능성이 높다고 볼 수 있다.

패널가구원의 응답형태에 따른 3개 집단(연속응답, 2020년 이탈, 2021년 이탈)을 종속 변수로 하여 다항 로지스틱 회귀모형을 분석하였다. 사용한 설명 변수는 2019년 연간데이터에서의 가구원 특성(성별, 연령, 혼인상태, 최종 학력, 경제활동 상태), 2019년에서의 가구 특성(거주 지역, 주거형태, 가구원 수, 연간 가구 총소득), 그리고 2019년~2021년에 걸친 가구의 상태 변화(이사 횟수)이다. <요약표 2>는 이 분석 결과를 요약한 것으로, 연속응답을 준거집단으로 하여 이탈 집단에 영향을 주는 요인에 관한 것이다. 즉, 설명 변수 중에서 통계적으로 유의한 결과를 나타낸 변수를 정리하였다. 2020년과 2021년 패널가구원의 이탈에 영향을 주는 공통적인 요인은 가구원의 연령, 최종 학력, 경제활동 상태, 그리고 가구원 수, 이사 횟수였다. 가구원의 연령은 젊을수록, 최종 학력은 고졸이 초졸에 비해, 경제활동 상태는 임금근로자가 기타에 비해, 가구원 수는 많을수록, 이사 횟수는 적을수록 패널가구원의 이탈 경향이 높은 편으로 나타났다. 이사 횟수는 패널가구의 표본이탈에 영향을 주는 요인에도 해당되었는데, 패널가구와 패널가구원에서 동일한 결과가 나왔다. 한편, 주거형태의 경우 2020년은 아파트가 단독주택에 비해, 다세대주택이 아파트에 비해 표본이탈 경향이 높았고, 2021년은 기타가 아파트에 비해 표본이탈 경향이 높은 편에 속하였다. 2020년 결과는 패널가구의 표본이탈에 영향을 주는 요인과 동일하였다. 거주 지역의 경우 광역시(준거 집단: 서울)가 통계적으로 유의한 영향을 주었는데 2020년 이탈 집단에서는 음(-)의 효과로, 2021년 이탈 집단에서는 양(+)의 효과로 상반되게 나타났다. 이외에도 2021년 이탈 집단은 남성이 여성에 비해, 연간 가구

총소득의 경우 저소득이 중소득에 비해 표본이탈에 영향을 주는 것으로 나타났다.

〈요약표 2〉 패널가구원의 다항 로지스틱 회귀모형 분석 결과 요약

변수	2020년 이탈	2021년 이탈
가구원 성별_남성		여성 (-)
가구원 연령	(-)	(-)
가구원 혼인상태_유배우자		
가구원 최종 학력_고졸	초졸 (-)	무학 (+) 초졸 (-)
가구원 경제활동 상태_임금근로자	기타 (-)	비경제활동인구 (+) 기타 (-)
거주 지역_서울	광역시 (-) 기타 도지역 (-)	광역시 (+)
주거형태_아파트	단독주택 (-) 다세대주택 (+)	기타 (+)
가구원 수	(+)	(+)
연간 가구 총소득_중소득		저소득 (+)
이사 횟수	(-)	(-)

- 주: 1) 괄호 안의 내용은 각 변수에 해당하는 계수(coefficient)값의 부호를 의미함.
 2) 가구원 경제활동 상태에서 기타는 무급가족종사자, 자활근로, 공공근로, 희망근로, 노인 일자리가 해당함.
 3) 주거형태에서 기타는 영업용 건물 내 주택(상가, 공장 등), 오피스텔, 호텔, 여관, 기숙사 등을 의미함.
 4) 연간 가구 총소득에서 저소득은 0원~2,000만 원 미만, 중소득은 2,000만 원~5,000만 원 미만, 고소득은 5,000만 원 이상을 의미함.
 자료: 한국보건사회연구원·국민건강보험공단. (2023). 2기 한국의료패널조사 2019년~2021년 연간데이터(2023년 학술대회용 버전)를 분석함.

마지막으로 2기 원패널을 보정하고 변화된 새로운 모집단 특성을 반영한 통합표본의 패널을 구축하려는 방향에 맞춰, 추가표본을 추출하는 방안을 마련하였다. 이때 현재 2기 원패널은 그대로 유지한다. 우선, 목표모집단과 조사모집단이 일치하는지 비교한 결과, 큰 차이 없이 일치하는 것을 확인하였다. 다음으로 지역 특성에 따른 층화변수인 17개 시도와 동부·읍면

부에 따른 가구 분포를 분석하였는데, 층화변수 간 가구수 차이가 큰 것으로 나타났다. 표본설계 시 적절한 배분방법을 선정하여 가구가 적은 층에서 과소표집이 되지 않도록 하였다. 기존 조사 결과 분석에서는 만성질환 수와 연간 가구 총소득에 대해 살펴보았다. 2021년 연간데이터 기준 시도별 및 동부·읍면부별 만성질환 수의 상대표준오차는 전체적으로 25% 이하로 안정적으로 유지되고 있었다. 그러나 부산광역시 읍면부, 대구광역시 읍면부와 울산광역시 읍면부 층에서는 25%를 초과하는 것으로 나타났다. 연간 가구 총소득도 만성질환 수와 유사한 결과를 보였으나, 대구광역시 읍면부만 25%를 초과했으며 나머지 지역은 전반적으로 안정적이라고 볼 수 있다. 현시점에서 최신 모집단인 2021년 인구주택총조사를 활용하여 추가표본을 구축하는 방안을 마련하였다. 최종 목표 표본 규모는 8,400가구로 결정하였고, 이에 따라 추가표본은 2,700가구를 추출할 수 있도록 표본설계를 하였다. 이때 2기 한국의료패널 구축 당시에 활용하였던 층화변수, 표본 배분방법 등은 동일하게 적용하였다. 표본 배분은 1차 시도의 경우 가구수에 따라 제곱근 비례배분방법을, 2차 동부·읍면부의 경우 비례배분방법을 적용하였다. 패널조사의 특성상 패널을 지속적으로 유지하는 것이 중요하므로 목표 가구수보다 조사 완료 가구수가 더 큰 층에서는 추가표본을 추출하지 않고, 부족한 층에 대해서만 표본을 추출하도록 설계하였다. 이때 패널 마모가 심한 4개 지역이 있었는데 부산광역시 읍면부, 대구광역시 읍면부, 울산광역시 읍면부의 패널은 5가구 이하였고, 인천광역시 읍면부는 0가구였다. 이는 제3장 패널가구의 표본 이탈 특성에서 대구와 인천의 경우 이탈 비율이 높게 나타난 것과 유사한 결과였다. 인천광역시 읍면부의 경우 충분한 표본을 확보하기 위해 최소 3개 조사구를 배분할 수 있도록 조정하였고, 3개 지역도 목표 가구수에 따라 조사구를 배분하였다. 통합표본에서의 예상 상대표준오차를 만성

질환 수와 연간 가구 총소득에 대해 검토한 결과, 표본을 추가할 경우 상당히 개선되는 것으로 나타났다. 표본조사구 추출 시 내재적 층화변수는 층별 조사구 특성, 유배우 인구수, 아파트 거주 가구수를 활용하였다. 2기 한국의료패널 구축 당시에는 조사구 특성, 점유형태, 혼인상태, 거처 종류를 활용하였으나, 현재 혼인상태와 거처 종류는 사용할 수 없기 때문에 새롭게 선정하였다. 조사구당 8~12가구(평균 10가구)를 조사 완료하여 최종 추가표본으로 구축하는 것을 목표로 하였다. 여기서 제안한 추가표본 구축을 위한 표본추출 방안에 따라 2023년 12월에 조사를 착수할 계획이다.

3. 결론 및 시사점

패널 대상의 이탈 특성에 대해 통계적인 측면에서 다각적으로 살펴 보았다. 이러한 연구 결과를 기반으로 패널 대상의 이탈을 최소화하기 위해서 패널 관리가 집중적으로 필요한 부분에 대해 네 가지를 제안하고자 한다.

첫째, 이사하는 패널가구에 대한 꾸준한 추적 관리가 필요하다. 이사로 인해 패널가구에 대한 추적이 불가능한 비율이 점차 증가하는 추세를 확인하였다. 하지만 이사 변동에 대한 추적이 관리될 수 있다면 패널에서 이탈할 가능성이 낮은 편이라는 분석 결과를 얻었다. 이는 선행 연구와 유사한 결과로, 주소 변경을 파악할 수 있다면 일시적으로 이탈이 발생하더라도 재복귀가 가능하다는 분석과 일맥상통한다고 볼 수 있다.

둘째, 아파트에 거주하는 패널가구에 대한 패널 관리가 필요하다. 패널 가구의 주택유형 분포는 아파트와 단독주택이 높은 편이었으나, 시간이 지남에 따라 아파트의 비율은 낮아지고 단독주택은 높아지는 추세를

보였다. 아파트에 거주하면 이탈 비율은 높은 편이었으나, 단독주택은 낮은 편에 속하였다. 또한, 패널가구 이탈에 영향을 주는 요인에서도 아파트가 단독주택에 비해 가능성이 있는 것으로 나타났다.

셋째, 중년의 부모와 자녀로 구성된 가족에 대한 패널 관리가 필요하다. 패널가구와 패널가구원 모두 무응답한 집단은 0세 이상~19세 이하 및 40세 이상~49세 이하인 연령대이며, 미취학이 포함된 무학과 대졸 이상이며, 기타를 제외한 나머지 경제활동 상태의 특성을 지녔다. 또한, 3인 가구 이상의 응답 분포는 점점 줄어드는 양상을 보였고, 15세 미만 응답자의 비율도 점차 감소하는 추세였다.

마지막으로 연령이 젊은 패널가구원과 혼인상태가 미혼인 패널가구원에 대한 패널 관리가 필요하다. 패널가구와 패널가구원 모두 무응답한 집단의 인구사회학적 특성을 보면 미혼이 다른 혼인상태에 비해 조사에 참여하지 않을 가능성이 높은 편으로 나타났다. 또한, 패널가구원의 이탈에 영향을 주는 요인으로 패널가구원의 연령이 해당하였다.

패널조사에서 패널 대상의 이탈 증가 현상은 패널조사의 안정성을 위협하는 위험 요인이라고 볼 수 있다. 패널 관리가 집중적으로 필요한 부분에 대해 제안한 네 가지를 중심으로 사전에 대처함으로써 패널 대상의 이탈을 막을 수 있도록 노력해야 할 것이다.

주요 용어 : 한국의료패널조사, 표본이탈, 추가표본, 표본설계

사람을
생각하는
사람들



KOREA INSTITUTE FOR HEALTH AND SOCIAL AFFAIRS



제 1 장

서론

제1절 연구 배경 및 목적

제2절 연구 내용 및 방법

제 1 장 서론

제1절 연구 배경 및 목적

동일한 대상에 대해 동일한 문항을 반복적으로 측정하는 패널조사는 보통 조사 차수가 거듭될수록 패널 대상이 이탈하는 현상이 발생하기 마련이다. 패널에서 대상이 이탈하는 주요 요인은 다음과 같다. 패널 대상이 사람이면 이사, 이민 등으로 인한 연락두절, 반복되는 조사 참여에 대한 응답자 피로도, 거부감 등으로, 사업체이면 담당자의 변경, 사업체의 폐업 등으로 볼 수 있다.

패널조사에서 패널 대상(표본)의 이탈이 일어나면 표본이 모집단을 대표하지 못하게 됨에 따라 추정의 편이가 발생하는 등 잘못된 분석 결과를 도출할 수 있게 되어 패널조사에 대한 신뢰도 하락으로 이어질 수 있다. 이를 해결하기 위한 첫 번째 방안은 패널조사의 표본이탈 현상을 파악하고 표본 유지를 위한 대책을 마련하는 것이다. 그래서 이러한 연구는 활발히 수행되고 있는 편으로 선행연구 결과를 살펴보면 다음과 같다. 표본이탈에 영향을 미치는 요인은 패널 대상자의 인구사회학적 특성, 이사 등의 주거지 변동, 조사체계의 특성(면접원 특성, 조사 참여 사례, 조사도구) 등이 있다. 패널 대상자의 인구사회학적 특성을 살펴보면 남성이 여성에 비해, 연령이 젊을수록, 학력이 높을수록, 배우자가 없는 경우, 임금근로자일수록 표본 이탈의 경향이 높다는 결과를 보였다(주재선, 임연규(2018), 이상협, 박찬용, 정성석, 최혜미(2011), 이지연, 김진(2006)). 또한 패널 가구의 가구주가 여성이거나, 연령이 젊을수록, 학력이 높을수록 표본이탈이 발생할 가능성이 높다고 나타났다(이경희, 민인식(2016), 송헌재(2012),

심영상, 허명희(2002), 김대일, 남재량, 류근관(2000)). 남기성·천영민(2012)의 연구 결과에서는 이사, 주거형태가 월세인 경우 자가에 비해 표본이탈의 위험을 높인다고 하였다. 마지막으로 조사체계의 특성에 대한 연구에서 신선옥(2008)은 면접원의 학력이 높고 면접 경력이 길수록 응답의 협조에 긍정적이라고 하였다(이상호, 이해정, 2017 재인용). 그리고 성재민(2007)은 응답자에 대한 사례(현금 보상)가 조사 참여에 상당한 영향력을 가지며, 특히 저소득 가구의 응답률을 높이는 데 긍정적인 작용을 한다고 하였다(이상호, 이해정, 2017 재인용).

다음 방안은 패널조사가 일정한 시점이 지나면 표본을 추가하는 것이다. 1기 한국의료패널조사는 2007년에 원표본을 추출한 이후 2012년에 추가 표본을 구축하였다.¹⁾ 한국복지패널조사는 7차 조사(2012년)에서 표본을 추가²⁾하였고, 최근 표본 추가를 위한 사전 작업을 완료하였다(정은희 외, 2022). 한국노동패널조사도 12차 조사(2009년)와 21차 조사(2018년) 두 번에 걸쳐 표본을 추가하였다.³⁾ 보통 추가표본을 구축하는 목적은 마모된 표본을 보완하는 데 있다. 그래서 현시점에서의 패널 구조 파악, 조사 결과 분석 등을 기반으로 패널에 대해 진단한다. 그런 다음 표본 이탈로 인해 추가표본이 필요한 부분을 보완하는 방향으로 추가표본을 구축하게 되며, 추가표본 추출은 활용 가능한 최신 인구주택총조사를 표본추출틀로 활용한다. 표본설계를 할 때 원패널과 통합패널의 체계적 차이가 발생하지 않도록 최초 패널조사 구축 당시의 표본설계 방법을 따를 수 있게 함께 고려한다. 추가표본을 구축한 시점의 원표본 조사 완료 가구와 추가표본 가구가 합쳐져 새로운 통합표본의 패널로 구축된다. 또한, 통합표본의 패널은 현재 시점의 모집단을 대표하도록 구성되어 횡단면

1) <https://www.khp.re.kr:444/web/survey/sampling.do> 2023.9.30. 인출함.

2) <https://www.koweps.re.kr:442/probe/sample.do> 2023.9.30. 인출함.

3) <https://www.kli.re.kr/menu.es?mid=a40201000000> 2023.9.30. 인출함.

데이터의 특성도 만족할 수 있도록 한다. 이때 원표본의 패널은 그대로 유지되어, 통합표본과 원표본으로 관리하게 된다.

한편, 한국의료패널조사는 2018년부터 2기 패널을 구축하기 위한 준비를 하였고, 2019년 사전 조사를 시작으로 매해 조사를 실시하고 있으며, 올해 8월에 2023년 조사를 완료하였다. <표 1-1>은 2기 한국의료패널조사의 조사 차수를 기준으로 한 완료 현황으로 2018년 유치조사에서 8,516가구가 조사되었다. 그러나 조사 차수가 지남에 따라 패널 가구가 점점 줄어드는 추세가 나타나며 2022년 조사에서는 5,974가구를 조사 완료하였다. 특히 2019년 사전 조사 완료 가구는 유치조사 완료 가구 기준 대비 87.9%로 나타나, 가구 응답률이 가장 낮은 편에 속하였다. 보통 패널조사의 경우 첫 번째 조사 이후 바로 다음 조사에서 패널가구의 조사 참여도가 가장 저조한 편이고, 그 이후부터는 안정적인 추세를 가지게 된다.

<표 1-1> 한국의료패널조사 조사 차수 기준 완료 현황

(단위: 가구, 명)

	가구	가구원
2018년 유치 조사	8,516	21,684
2019년 사전 조사	7,485	18,630
2020년 1차 조사	6,748	16,587
2021년 2차 조사	6,217	14,844
2022년 3차 조사	5,974	13,966

자료: 한국보건사회연구원·국민건강보험공단. (2023). 2기 한국의료패널조사 2019년~2021년 연간데이터(2023년 학술대회용 버전) 및 비공개 자료를 분석함.

2기 한국의료패널조사도 지속적으로 패널 마모가 발생하게 된다면 표본 대표성에 문제가 생길 수 있다. 더 나아가서는 데이터 품질에 대한 신뢰성도 담보하기 어렵게 될 수 있다. 또한, 패널조사의 추가표본에 대한 유치

조사를 실시하기 위해서는 최소 1년 이상의 준비 기간이 필요하고, 사전 조사를 거친 후에 1차 조사가 진행되어야 하는 등의 상황도 고려해야 한다. 이에 5년간의 조사 데이터 분석을 통해 도출된 표본이탈 추세를 고려하여 2~3년 이후의 표본이탈 및 표본유지 상황을 예측하여 표본을 추가해야 한다. 이를 위하여 2기 한국의료패널조사의 현시점에서의 표본 구성에 대한 다각적이고 면밀한 검토가 필요한 상황이라고 볼 수 있다.

이 연구의 목적은 최근 조사가 완료된 데이터를 활용하여 2기 한국의료패널의 조사 완료 현황 및 패널 대상자 특성을 분석하는 것이다. 그리고 응답형태에 따라 패널가구와 패널가구원의 표본이탈의 특성을 파악하고, 표본이탈에 영향을 주는 요인을 살펴보고자 한다. 마지막으로 현시점에서의 패널 구조를 진단한 다음에 표본 대표성을 만족할 수 있도록 추가표본 추출 방안을 마련하고자 한다.

제2절 연구 내용 및 방법

1. 연구 내용

이 연구는 패널조사의 표본이탈 경향 분석과 추가표본 추출에 대한 연구로 2기 한국의료패널조사를 중심으로 살펴본다. 각 장의 연구 내용은 다음과 같다.

제2장은 2기 한국의료패널 대상에 대한 기초 분석을 실시하고자 한다. 우선 조사 개요와 2기 패널 구축 당시의 표본설계에 대해 살펴본다. 이는 제4장에서 추가표본 구축 방안을 마련하기 전에 2기 한국의료패널 구축 당시의 표본설계를 살펴볼 필요가 있기 때문이다. 다음은 조사 완료 현황과 원표본 유지율 등을 살펴본다. 그리고 응답한 가구와 가구원의 인구사회

학적 특성에 따른 응답 추이를 분석한다.

제3장은 2기 한국의료패널의 표본이탈에 대해 체계적으로 분석하여 표본이탈의 경향을 파악하고자 한다. 응답형태로 구분하여 패널가구와 패널가구의 인구사회학적 특성에 따른 분포를 살펴본다. 응답형태는 2019년~2021년 연간데이터를 사용하여 패널 대상의 응답 여부를 조합하여 연속응답, 2020년 이탈, 2021년 이탈, 재진입 4개 집단으로 구분한다. 또한, 표본이탈에 영향을 미치는 요인을 살펴보기 위해서 응답형태를 종속 변수로 하여 다항 로지스틱 회귀모형을 분석한다. 사용한 설명 변수는 2019년 연간데이터의 가구, 가구주 및 가구원 특성, 가구의 상태 변화 등이다. 이때 패널가구와 패널가구에 대해 각각 분석하여 살펴본다.

제4장은 2기 한국의료패널의 추가표본 구축을 위한 표본 추출 방안에 대해 살펴본다. 먼저 2021년 인구주택총조사 데이터를 사용하여 모집단의 특성에 대해 파악한다. 모집단 특성을 정확하게 분석하여 추가표본 구축을 위한 표본설계에 활용하기 위함이다. 다음으로 기존 조사 결과에 대해 분석하는데, 주요 변수는 만성질환 수와 연간 가구 총소득을 중심으로 설정하였다. 이 변수는 2기 한국의료패널을 구축할 때도 활용할 변수여서 동일하게 선정하였다. 기존 표본설계의 적절성을 검토하고 최적의 추가 표본 구축 방안을 제안하고자 한다. 이를 통해 2기 한국의료패널의 표본 마모를 보완하여 표본 대표성을 확보하고자 한다.

마지막 제5장은 결론으로 주요 연구 결과를 정리하고, 패널 대상의 이탈을 최소화하기 위해서 패널 관리가 집중적으로 필요한 부분은 무엇인지 제안하고자 한다.

2. 연구 방법

국내외 문헌 연구, 조사 데이터 분석, 전문가 자문회의 등을 다양하게 활용하여 연구한다. 이 연구에서 사용한 조사 데이터는 2기 한국의료패널 조사로 유치조사, 2019년~2021년 연간데이터(2023년 학술대회 버전)이다. 필요시 조사 실사 관리를 위해 정리한 내부 데이터를 사용한다. 그리고 제4장 추가표본 구축을 위한 표본설계에서는 통계청의 2021년 인구주택총조사 데이터도 함께 활용하였다. 통계패키지는 SAS를 사용한다.

사람을
생각하는
사람들



KOREA INSTITUTE FOR HEALTH AND SOCIAL AFFAIRS



제2장

패널 대상 기초 분석

제1절 2기 한국의료패널조사 개요 및 완료 현황

제2절 응답 여부에 따른 패널 대상 특성 분석

제3절 소결

제 2 장 패널 대상 기초 분석

2기 한국의료패널조사의 패널가구와 패널가구원에 대한 기초 분석을 실시하였다. 제1절에서는 조사 개요와 표본설계를 정리하였고, 조사 완료 현황을 파악하였다. 이는 제4장에서 추가표본 구축 방안을 마련하기 전에 2기 한국의료패널조사 구축 당시의 표본설계를 살펴볼 필요가 있기 때문이다. 제2절에서는 유치조사, 2019년~2021년 연간데이터⁴⁾를 사용하여 응답한 가구와 가구원의 인구사회학적 특성에 따른 응답 추이를 분석하였고, 응답하지 않은 가구의 경우 조사에 참여하지 않은 사유에 대해 살펴보았다.

제1절 2기 한국의료패널조사 개요 및 완료 현황

1. 조사 개요 및 표본설계 현황⁵⁾

한국의료패널은 기존 행정 자료로 파악이 어려운 개인과 가구 단위의 의료비 지출과 의료 이용 결정 요인 등을 심층적으로 파악하는 데 있다. 1기 한국의료패널조사가 2008년에 처음 시작하여 2019년을 마지막으로 마쳤다. 이후 여러 가지 검토 과정을 가진 후에 2기 한국의료패널조사를 착수하였다. 주요 개편 방향은 2016년 등록센서스를 기준으로 표본을

4) 한국의료패널 데이터는 2개의 조사 응답 시점을 결합하여 생성하는 연간데이터로 구성되어 있음.

5) 박은자 외(2019)의 내용을 기반으로 정리하였음.

구축하여 표본의 대표성을 확보하는 것이었다. 그리고 매년 조사하는 항목과 주기적으로 조사하는 항목을 구분하고, 불필요한 설문 내용을 최소화하는 등으로 조사 효율성을 제고하는 데 있다. 또한 환자 경험, 요양 및 간병 등에 대한 조사 내용을 추가하여 정책 활용도를 높이고자 하였다.

다음은 2기 한국의료패널을 구축할 때 작성했던 표본설계 내용에 대해 살펴보았다. 목표모집단은 2018년 전국 17개 시도의 일반거주시설에 거주하는 일반가구 및 가구원이다. 표본추출틀은 2016년 등록센서스를 활용하였는데, 조사구 중에서 기숙사, 특수 시설 조사구, 섬지역 조사구는 제외한 조사구에 거주하는 일반가구가 표본 추출 대상이다. 표본조사구 추출 방법은 2단계 층화집락추출방법이다. 1차 층화변수는 17개 시도로 지역별 가구 규모에 따른 제곱근비례배분과 설계 효과를 고려하였다. 이때 설계 효과는 만성질환 수를 기준으로 하였다. 2차 층화변수는 동부/읍면부에 따라 비례배분하여 708개의 표본조사구를 추출하였다. 그런 다음 조사구당 12가구를 계통추출하여 가구와 가구 내 거주하는 가구원을 조사함에 따라 최종 8,500가구를 패널로 구축하는 것을 목표로 하였다.

2. 조사 완료 현황

2기 한국의료패널조사 연간데이터 기준 완료 현황을 살펴보면 <표 2-1>과 같다.⁶⁾ 유치조사(8,516가구)의 79.2%가 2019년 연간데이터(6,748가구)로 최종 구축되었다. 1기 한국의료패널의 경우, 가구의 이탈을 고려하여 목표가구인 7,000가구~8,000가구보다 과대표집하여 9,015가구를 유치하였다. 이러한 과정을 거쳐 약 6개월 이후에 다시 방문한

6) 이 연구에서 분석에 활용한 데이터는 2019년~2021년 연간데이터이므로, 이를 기준으로 작성하였음.

7,866가구를 한국의료패널의 원표본 가구로 구축하였고,⁷⁾ 이는 유치 조사의 87.3%가 2008년 원표본 가구였다. 이러한 비율의 차이는 1기의 경우 유치조사 이후 가구를 재방문한 기간이 짧은 편이었고, 과대표집 등을 했다는 점에 있다고 보인다.

〈표 2-1〉 2기 한국의료패널조사 연간데이터 기준 원료 현황

(단위: 가구, 명)

	전체	
	가구	가구원
유치조사	8,516	21,684
2019년 연간데이터	6,748	16,587
2020년 연간데이터	6,217	14,844
2021년 연간데이터	5,907	13,799

자료: 한국보건사회연구원·국민건강보험공단. (2023). 2기 한국의료패널조사 2019년~2021년 연간 데이터(2023년 학술대회용 버전)를 분석함.

다음은 패널가구의 응답 현황과 원표본 유지율을 살펴보았다. 〈표 2-2〉를 보면 2021년 연간데이터의 원표본 유지율은 86%로 나타나, 1기 보다 높은 수준을 유지하고 있다. 1기 한국의료패널 데이터의 3차 연도 원표본 유지율은 77.4%⁸⁾였다. 한편, 한국노동패널의 09통합표본과 18 통합표본과 비교하면 원표본 유지율은 낮게 나타났다. 그러나 한국복지패널과 장애인삶패널의 3차 연도 원표본 유지율과는 비슷한 수준이라고 볼 수 있다.⁹⁾

7) 정영호 외. (2009). 2008년 한국의료패널 기초분석보고서 (I). 한국보건사회연구원. p. 31.

8) 1기 의료패널의 경우 2번(상반기, 하반기)에 걸쳐 조사를 실시하였는데, 본문에서는 2010년 하반기의 값을 3차 연도 원표본 유지율로 보았음.

9) 타 패널조사 원표본 유지율은 다음과 같음. 3번째 조사에 대한 원표본 유지율을 살펴보면 1기 한국의료패널은 77.4%, 한국복지패널은 86.7%, 한국노동패널은 80.9%(98표본), 90.5%(09표본), 92.7%(18표본), 장애인삶패널은 85.9%로 나타남.

28 패널조사 표본이탈 경향 분석과 추가표본 추출에 관한 연구- 한국의료패널조사를 중심으로

〈표 2-2〉 2기 한국의료패널조사 연간데이터 기준 - 원표본 유지율

(단위: 가구, %)

	응답	무응답	전체	원표본 유지율
2019년 연간데이터	6,748	-	6,748	100
2020년 연간데이터	6,163	585	6,748	91.3
2021년 연간데이터	5,804	944	6,748	86.0

주: 1) 원표본 유지율은 2019년 원표본을 기준으로 계산함.

2) 2020년에는 분가 가구 54가구가 응답하여, 전체 응답 가구수는 6,217가구임.

3) 2021년에는 분가 가구 103가구가 응답하여, 전체 응답 가구수는 5,907가구임.

자료: 한국보건사회연구원·국민건강보험공단. (2023). 2기 한국의료패널조사 2019년~2021년 연간데이터(2023년 학술대회용 버전을 분석함).

〈표〉 2019년 기준 원표본 유지율

(단위: 가구, %)

조사	조사연도	응답	무응답	전체	원표본 유지율	
한국의료패널 (1기)	2008년 상반기	7,866	-	7,866	100	
	2008년 하반기	7,169	697	7,866	91.1	
	2009년	6,727	1,139	7,866	85.5	
	2010년 상반기	6,313	1,553	7,866	80.3	
한국복지패널	2010년 하반기	6,089	1,777	7,866	77.4	
	2006년	7,072	-	7,072	100	
	2007년	6,511	561	7,072	92.1	
한국노동패널	2008년	6,128	944	7,072	86.7	
	98표본	1998년	5,000	-	5,000	100
		1999년	4,378	622	5,000	87.6
		2000년	4,044	956	5,000	80.9
	09통합표본	2009년	6,721	-	6,721	100
		2010년	6,232	489	6,721	92.7
		2011년	6,082	639	6,721	90.5
		2018년	12,134	-	12,134	100
	18통합표본	2019년	11,471	663	12,134	94.5
		2020년	11,253	881	12,134	92.7
장애인삶패널	2018년	6,121	-	6,121	100	
	2019년	5,527	594	6,121	90.3	
	2020년	5,259	862	6,121	85.9	

주: 1) 한국의료패널: 한국의료패널 홈페이지 조사결과.

<https://www.khp.re.kr:444/web/survey/results.do>에서 2023.08.10. 인출.

2) 한국복지패널: 한국복지패널 홈페이지 조사결과.

<https://www.koweps.re.kr:442/probe/result.do>에서 2023.08.10. 인출.

3) 한국노동패널: 한국노동패널 홈페이지 조사결과.

<https://www.kli.re.kr/menu.es?mid=a40204010000>에서 2023.08.10. 인출.

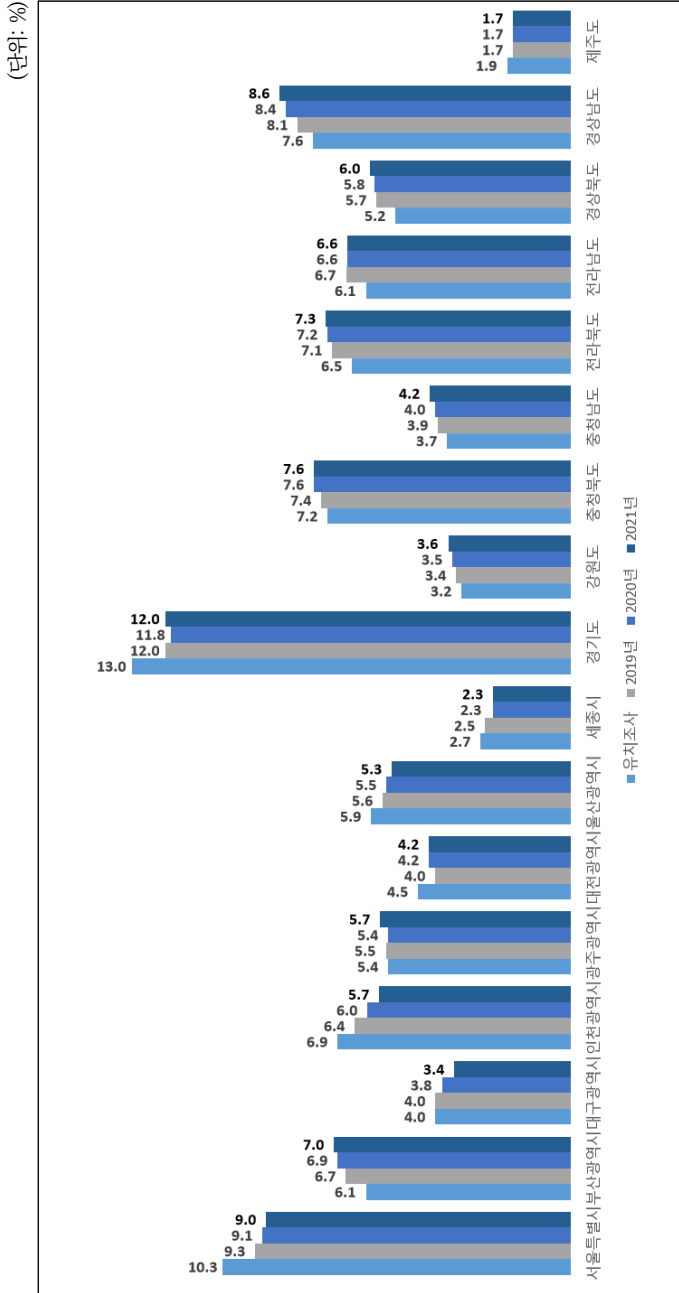
4) 장애인삶패널: 장애인삶패널조사 이용자 안내서(1차-4차). 김현지 외(2023). 한국장애인개발원. p.26.

제2절 응답 여부에 따른 패널 대상 특성 분석

1. 응답 가구 및 가구원 특성 분석

유치조사, 2019년~2021년 연간데이터에 담긴 응답 가구 및 가구원의 인구사회학적 특성 분포를 살펴보았다. 먼저, 응답 가구의 주요 변수에 따른 특성을 살펴보면 다음과 같다([그림 2-1] 참조). 2019년 구축된 패널 가구의 거주 지역을 17개 시도별 분포로 살펴보면 경기와 서울의 비율이 다른 지역에 비해 높은 편에 속한다. 그런데 유치조사와 2019년의 비율 차이는 다른 지역에 비해 가장 큰 차이를 가졌으나, 1%p 이하로 크지 않은 편이었다. 경기의 경우 2019년에는 12%로 유치조사(13%)에 비해 낮았고, 서울도 2019년이 9.3%로 유치조사(10.3%)보다 낮게 나타났다. 조사 차수가 거듭될수록 서울, 대구, 인천, 대전, 울산, 세종, 경기, 제주의 비율은 점차 감소하는 추세를 보였다.

[그림 2-1] 유치조사, 2019년~2021년 연간테이터의 가구 특성 - 거주 지역

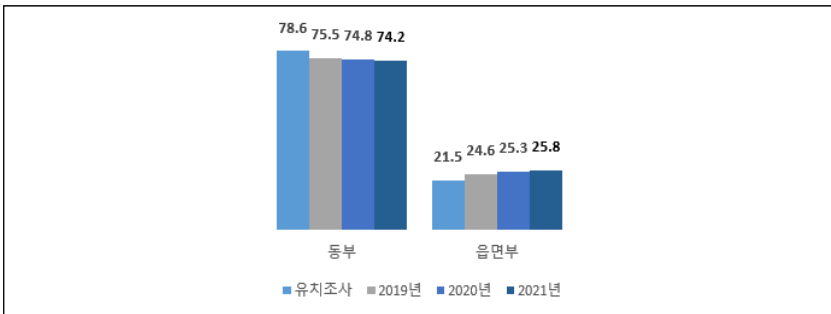


자료: 한국보건사회연구원·국민건강보험공단. (2023). 2기 한국의료패널조사 2019년~2021년 연간테이터(2023년 학술대회용 버전)를 분석함.

[그림 2-2]의 동부/읍면부를 보면 동부의 경우가 2019년 75.5%로 유치조사(78.6%)에 비해 3.1%p 낮은 편이었다. 시간이 지날수록 동부가 읍면부에 비해 비율이 점점 낮아지는 것으로 나타났다.

[그림 2-2] 유치조사, 2019년~2021년 연간데이터의 가구 특성 - 동부/읍면부

(단위: %)

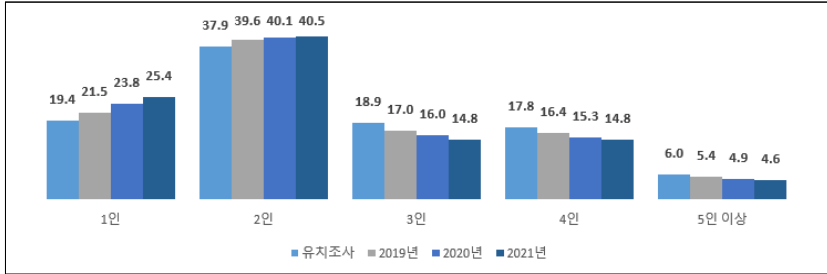


자료: 한국보건사회연구원·국민건강보험공단. (2023). 2기 한국의료패널조사 2019년~2021년 연간데이터(2023년 학술대회용 버전)를 분석함.

다음 [그림 2-3]의 가구원 수를 보면 유치조사, 모든 연간데이터에서 2인 가구의 비율이 가장 높고 다음으로 1인 가구였다. 1인과 2인 가구의 경우 2019년 이후로 지속적으로 증가하는 추세가 나타났다. 이에 반해 3인 이상은 2019년 이후로 유치조사에 비해 점점 줄어드는 양상을 보였다. 시간이 지남에 따라 3인 이상 가구의 조사 참여도가 낮아진다고 볼 수 있다. 이는 세대구성의 결과에서도 유사하게 나타났다([그림 2-4] 참조). 2019년의 경우 2세대가 41.1%로 유치조사(44.9%)에 비해 3.8%p 낮았던 반면에, 1세대는 55.5%로 유치조사(51.3%)에 비해 4.2%p 높았다.

[그림 2-3] 유치조사, 2019년~2021년 연간데이터의 가구 특성 - 가구원 수

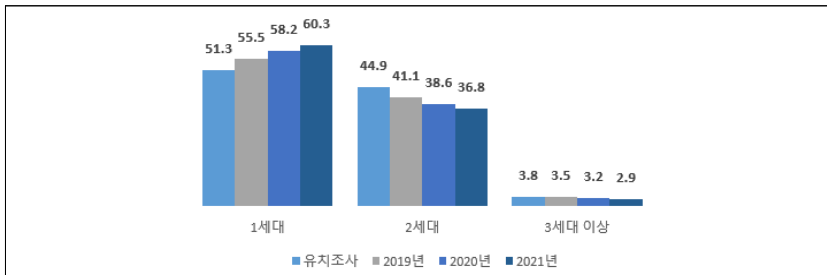
(단위: %)



자료: 한국보건사회연구원·국민건강보험공단. (2023). 2기 한국의료패널조사 2019년~2021년 연간데이터(2023년 학술대회용 버전)를 분석함.

[그림 2-4] 유치조사, 2019년~2021년 연간데이터의 가구 특성 - 세대구성

(단위: %)

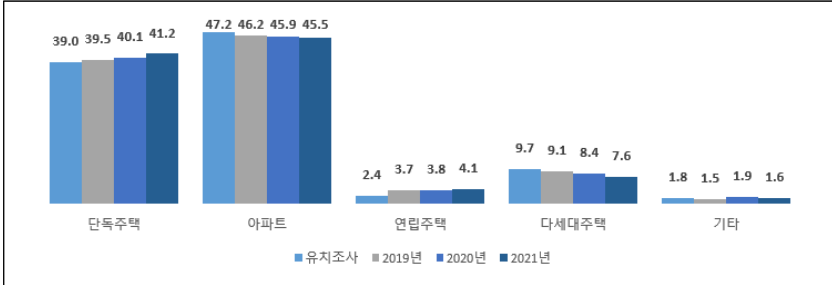


자료: 한국보건사회연구원·국민건강보험공단. (2023). 2기 한국의료패널조사 2019년~2021년 연간데이터(2023년 학술대회용 버전)를 분석함.

마지막으로 [그림 2-5]의 주택형태를 보면 아파트와 단독주택의 비율이 높은 편인데, 시간이 지남에 따라 아파트의 비율은 낮아지고 단독주택은 높아지는 추세를 보였다. 그리고 아파트는 2019년 46.2%로 유치조사(47.2%)에 비해 1%p 낮았다. 이에 비해 연립주택은 2019년 3.7%로 유치조사(2.4%)에 비해 1.3%p 높게 나타났다. 조사 차수가 거듭됨에 따라 아파트에 거주하는 가구의 비율이 낮아진다고 볼 수 있다.

[그림 2-5] 유치조사, 2019년~2021년 연간데이터의 가구 특성 - 주택유형

(단위: %)

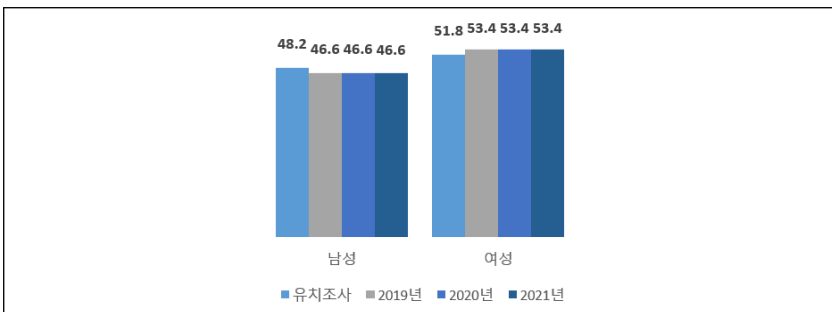


자료: 한국보건사회연구원·국민건강보험공단. (2023). 2기 한국의료패널조사 2019년~2021년 연간데이터(2023년 학술대회용 버전)를 분석함.

다음으로, 유치조사, 2019년~2021년 연간데이터에서 응답 가구원의 인구사회학적 특성 분포를 살펴보았다. [그림 2-6]은 가구원의 성별 분포인데 조사 차수가 거듭되더라도 동일하게 유지되고 있다. 다만 남성의 비율이 유치조사에서 48.2%였으나 2019년에는 46.6%로 1.6%p 낮게 나타났다.

[그림 2-6] 유치조사, 2019년~2021년 연간데이터의 가구원 특성 - 성별

(단위: %)



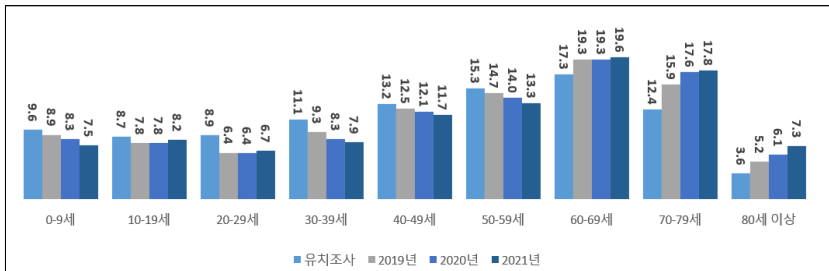
자료: 한국보건사회연구원·국민건강보험공단. (2023). 2기 한국의료패널조사 2019년~2021년 연간데이터(2023년 학술대회용 버전)를 분석함.

34 패널조사 표본이탈 경향 분석과 추가표본 추출에 관한 연구- 한국의료패널조사를 중심으로

다음 [그림 2-7]의 가구원 연령대 분포를 보면 60세 이상의 비율은 시간이 지날수록 점점 증가하는 추세를 보이는 반면에, 0세 이상~9세 이하 및 30세 이상~59세 이하의 비율은 감소하는 경향을 보였다. 2019년의 경우 60세 미만 가구원의 비율이 유치조사에 비해 낮게 나타났다. 특히 20세 이상~29세 이하 가구원은 유치조사에서 8.9%였으나 2019년에 6.4%로 2.5%p 감소하였다. 다음으로 30세 이상~39세 미만 가구원이 유치조사에서 11.1%로 2019년(9.3%)에 비해 1.8%p 감소하였다. 이를 통해 청년층, 중장년층의 응답 참여도가 노년층에 비해 낮다고 볼 수 있다.

[그림 2-7] 유치조사, 2019년~2021년 연간데이터의 가구원 특성 - 연령대

(단위: %)

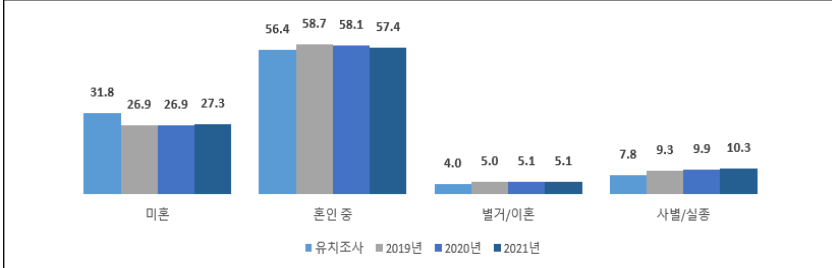


자료: 한국보건사회연구원·국민건강보험공단. (2023). 2기 한국의료패널조사 2019년~2021년 연간데이터(2023년 학술대회용 버전)를 분석함.

가구원의 혼인상태를 보면 혼인 중인 경우가 가장 높고, 다음으로 미혼, 사별/실종, 별거/이혼의 순이었다([그림 2-8] 참조). 미혼인 경우 유치조사 (31.8%)와 2019년(26.9%)의 차이가 4.9%p로 다른 혼인상태에 비해 가장 큰 차이를 보였다.

[그림 2-8] 유치조사, 2019년~2021년 연간데이터의 가구원 특성 - 혼인상태

(단위: %)



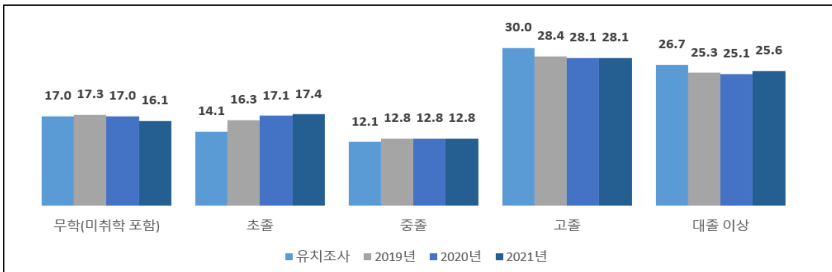
주: 2019년 연간데이터의 경우, 응답거부 2명 제외하고 분석하였음.

자료: 한국보건사회연구원·국민건강보험공단. (2023). 2기 한국의료패널조사 2019년~2021년 연간데이터(2023년 학술대화용 버전)를 분석함.

가구원의 최종 학력이 고졸 이상인 경우가 절반 이상으로 나타났다([그림 2-9] 참조). 한편, 2019년 고졸 이상 가구원의 비율(고졸 28.4%, 대졸 이상 25.3%)은 유치조사(고졸 30%, 대졸 이상 26.7%)에 비해 감소하여 학력이 높을수록 조사 참여도가 낮아진다고 볼 수 있다.

[그림 2-9] 유치조사, 2019년~2021년 연간데이터의 가구원 특성 - 최종 학력

(단위: %)



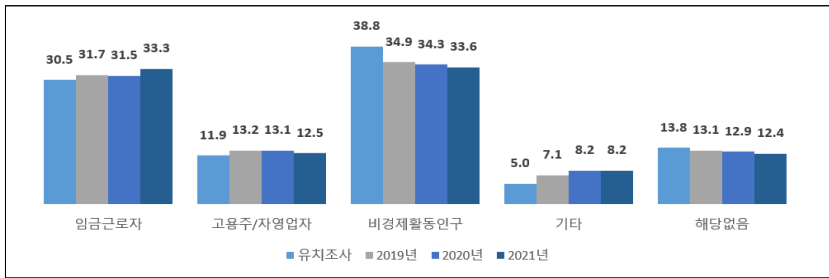
자료: 한국보건사회연구원·국민건강보험공단. (2023). 2기 한국의료패널조사 2019년~2021년 연간데이터(2023년 학술대화용 버전)를 분석함.

마지막으로 가구원의 경제활동 상태에 따른 분포를 보면 비경제활동 인구나 임금근로자의 비율이 가장 높았다([그림 2-10] 참조). 비경제

활동인구의 경우 2019년 34.9%로 유치조사(38.8%)에 비해 3.9%p 낮게 나타났으며 2020년 이후로 점점 감소하였다. 15세 미만 응답자인 ‘해당 없음’의 경우에는 유치조사에서 13.8%로 가장 높았으나 2019년 이후로 점차 감소하여 2021년에는 12.4%로 나타났다.

[그림 2-10] 유치조사, 2019년~2021년 연간데이터의 가구원 특성 - 경제활동 상태

(단위: %)



주: 1) 기타는 무급가족종사자, 자활근로, 공공근로, 희망근로, 노인 일자리가 해당함.

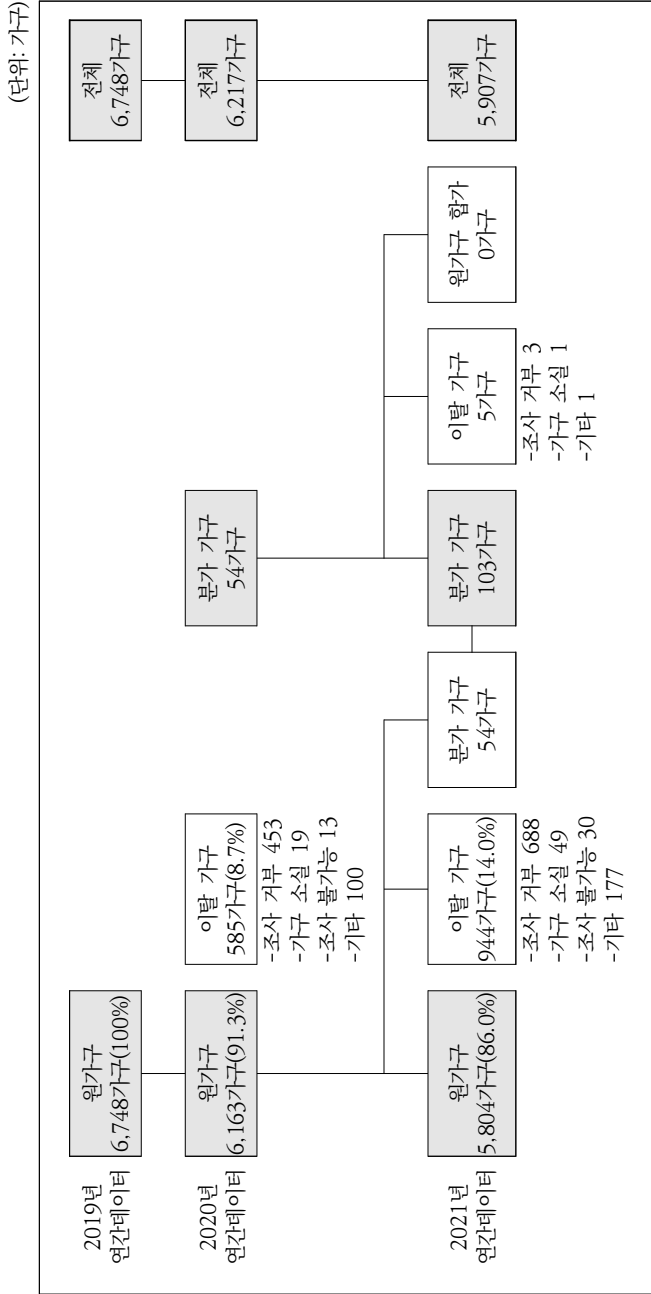
2) 해당 없음은 15세 미만 응답자임.

자료: 한국보건사회연구원·국민건강보험공단. (2023). 2기 한국의료패널조사 2019년~2021년 연간데이터(2023년 학술대회용 버전)를 분석함.

2. 무응답 가구의 조사 미참여 사유 분석

[그림 2-11]은 패널가구의 유지 및 이탈 흐름도로 원가구와 분가 가구로 구분하여 도식화하였다. 2020년 원가구는 6,163가구이고 분가 가구는 54가구였다. 2021년 원가구는 5,804가구이고 분가 가구는 103가구로 나타났다. 분가 가구는 원가구에서 분가한 원가구원이 있는 경우를 의미하며 패널 추적 대상에 해당하는데, 2기 한국의료패널에서도 분가 가구가 증가하는 양상을 보였다.

[그림 2-11] 패널가구의 유지 및 이탈 흐름도



자료: 필자 작성

다음으로 패널가구의 조사 완료 여부에 따른 분포를 원가구를 중심으로 살펴보면 조사 연도가 지날수록 조사를 완료하지 못하는 가구가 점차 증가하고 있다(〈표 2-3〉 및 [그림 2-12] 참조). 조사를 완료하지 못하는 경우는 두 가지로, ‘조사 가능하나 응답하지 않음’과 ‘조사 불가능함’으로 구분할 수 있다. 2020년의 경우 조사가 가능하나 응답하지 않은 가구는 553가구(8.2%)이고 조사 불가능한 가구는 32가구(0.5%)였다. 2021년의 경우 조사 가능하나 응답하지 않은 가구는 865가구(12.8%)이고 조사가 불가능한 가구는 79가구(1.2%)였다. 조사 불가능한 가구의 비율도 전년 대비 0.7%p 상승하여 2021년 1.2%로, 자연스럽게 패널에서 이탈하는 비율도 증가하고 있다고 볼 수 있다.

〈표 2-3〉 패널 원가구 조사 완료 여부에 따른 분포-빈도

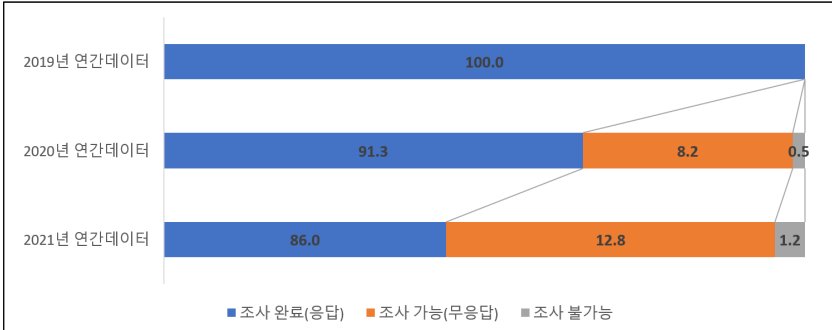
(단위: 가구)

	조사 완료(응답)	조사 미완료	
		조사 가능(무응답)	조사 불가능
2019년 연간데이터	6,748	-	-
2020년 연간데이터	6,163	553	32
2021년 연간데이터	5,804	865	79

주: 패널가구 중에서 원가구를 기준으로 작성함(분가 가구 제외).
 자료: 비공개 자료에 따른 출처 생략함.

[그림 2-12] 패널가구 조사 완료 여부에 따른 분포-비율

(단위: %)



주: 패널가구 중에서 원가구를 기준으로 작성함(분가 가구 제외).
 자료: 비공개 자료에 따른 출처 생략함.

조사가 가능하지만 무응답한 가구와 조사 불가능한 가구의 사유에 대해 살펴보면 다음 <표 2-4>와 같다. 두 개 연도 모두 조사 가능하나 응답하지 않은 가구의 비율이 90% 이상으로 조사 불가능한 가구에 비해 높게 나타났다. 조사 가능하나 응답하지 않은 가구의 조사 미완료 사유를 보면 2020년의 경우 조사 거부(77.4%)로 가장 높았으며 다음으로 접촉 회피(11.8%), 추적 불가(2.6%) 순으로 나타났다. 2021년의 경우도 동일한 결과로 조사 거부(72.9%), 접촉 회피(11.5%), 추적 불가(4.4%)로 나타났다. 조사를 거부하거나 접촉을 회피하는 가구에 대해 조사 참여를 유인할 수 있는 방안을 마련해야 할 필요가 있다. 또한, 이사 등의 추적 불가의 비율이 지속적으로 증가하는 양상을 보이고 있어 이에 대한 대책 마련이 필요하다고 생각한다. 한편, 조사 불가능한 사유의 경우 두 개 연도 모두 가구 소실(2020년 3.2%, 2021년 5.2%), 조사 불가능(2020년 2.2%, 2021년 3.2%) 순으로 높게 나타났으며 각각의 비율도 모두 증가하는 추세를 보였다.

〈표 2-4〉 패널가구의 조사 미완료 사유

(단위: 가구, %, %)

조사 미완료 사유		2020년 연간데이터			2021년 연간데이터		
		가구수	비율	비율	가구수	비율	비율
조사 가능 (무응답)	조사 거부	453	81.9	77.4	688	79.5	72.9
	접촉 회피	69	12.5	11.8	109	12.6	11.5
	추적 불가1)	15	2.7	2.6	42	4.9	4.4
	늦은 귀가, 장기 출타	7	1.3	1.2	11	1.3	1.2
	기타	9	1.6	1.5	15	1.7	1.6
소계		553	100.0	94.5	865	100.0	91.6
조사 불가능	가구 소실2)	19	59.4	3.2	49	62.0	5.2
	조사 불가능3)	13	40.6	2.2	30	38.0	3.2
	소계	32	100.0	5.4	79	100.0	8.4
총계		585	-	100.0	944	-	100.0

주: 1) 추적 불가능은 이사 등이 해당함.
 2) 가구 소실은 사망, 이민, 합가 등이 해당함.
 3) 조사 불가능은 해외 취업, 부적합 가구 등이 해당함.
 자료: 비공개 자료에 따른 출처 생략함.

제3절 소결

이 장에서는 2기 한국의료패널조사의 패널 대상자에 대한 기초 분석을 실시하였다. 유치조사, 2019년~2021년 연간데이터를 사용하여 응답한 가구와 가구원의 인구사회학적 특성에 따른 응답 추이를 분석하였다. 또한, 응답하지 않은 가구의 경우 조사를 참여하지 않은 사유에 대해 살펴보았다.

주요 결과를 보면 2기 한국의료패널의 원표본 유지율은 2020년 연간 데이터 기준으로 91.3%이고, 2021년은 86%로 하락하는 추이를 보였다. 이는 1기 한국의료패널보다 높은 수준을 유지하고 있다. 타 패널의 3차 연도 원표본 유지율과 비교했을 때 한국노동패널의 09통합표본과 18통합 표본에 비해서는 낮으나, 한국복지패널과 장애인삶패널과는 비슷한 수준

으로 나타났다.

응답 가구의 주요 변수에 따른 특성을 살펴보면, 2019년에 구축된 패널 가구의 거주 지역 분포를 유치조사와 비교했을 때 경기와 서울이 다른 지역에 비해 가장 큰 차이를 보이거나 1%p 이하로 크지 않은 편이었다. 2021년으로 갈수록 서울, 대구, 인천, 대전, 울산, 세종, 경기, 제주의 비율은 점차 감소하는 추세를 보였다. 2019년 이후에는 지역별로 안정적인 추세를 보였다. 가구원 수 경우에는 2019년 이후로 1인과 2인 가구는 지속적으로 증가하는 반면에, 3인 가구 이상은 유치조사에 비해 점점 줄어드는 양상을 보였다. 주택형태는 아파트와 단독주택의 비율이 높은 편이었으나, 시간이 지남에 따라 아파트의 비율은 낮아지고 단독주택은 높아지는 추세였다.

응답 가구원의 인구사회학적 특성을 중심으로 살펴본 결과는 다음과 같다. 성별은 조사 차수가 거듭되더라도 동일하게 유지하는 편이었는데, 여성의 비율이 남성보다 높게 나타났다. 연령대 분포의 경우 60세 이상의 비율은 시간이 지날수록 점점 증가하는 추세여서 노년층의 조사 참여율은 높은 데 비해 청년층과 중장년층은 낮았다. 혼인상태를 보면, 혼인 중인 경우가 가장 높은 비율을 차지하며, 미혼인 경우 다른 혼인상태에 비해 유치조사와 2019년의 차이가 가장 컸다. 최종 학력은 고졸 이상인 경우가 절반 이상이었는데, 2019년 고졸 이상 가구원의 비율이 유치조사에 비해 감소하여 학력이 높을수록 조사 참여도가 낮아진다고 볼 수 있다. 마지막으로 경제활동 상태는 비경제활동인구와 임금근로자의 비율이 가장 높았다. 15세 미만 응답자의 경우에는 2019년 이후로 점차 감소하는 추세였다.

패널가구의 조사 완료 여부에 따른 분포를 원가구를 중심으로 살펴보았는데 조사 연도가 지날수록 조사를 완료하지 못하는 가구가 점점 증가하고 있는 것으로 나타났다. 조사를 완료하지 못하는 경우는 두 가지로, '조사 가능하나 응답하지 않음'과 '조사 불가능함'으로 구분할 수 있다.

조사 가능하나 응답하지 않은 가구는 2021년 12.8%로 전년 대비 4.6%p 증가하였다. 조사 불가능한 가구의 비율도 전년 대비 0.7%p 상승하여 2021년 1.2%로 나타나, 자연스럽게 패널에서 이탈하는 비율도 증가하고 있다고 볼 수 있다. 조사가 가능하지만 응답하지 않은 가구와 조사 불가능한 가구의 사유에 대해 각각 살펴보았다. 조사 가능하지만 응답하지 않은 가구의 조사 미완료 사유를 보면 2020년의 경우 조사 거부(77.4%)가 가장 높았으며 다음으로 접촉 회피(11.8%), 추적 불가(2.6%) 순으로 나타났다. 2021년의 경우도 조사 거부(72.9%), 접촉 회피(11.5%), 추적불가(4.4%)로 나타나 조사 미완료의 사유 순위가 동일한 결과를 보였다. 조사를 거부하거나 접촉을 회피하는 가구에 대해 조사 참여를 유인할 수 있는 방안을 마련해야 할 필요가 있다. 또한, 이사 등의 추적 불가의 비율이 지속적으로 증가하는 양상을 보이고 있어 이에 대한 대책 마련이 필요하다고 생각한다. 한편, 조사 불가능한 사유의 경우 두 개 연도 모두 사망, 이민, 합가 등의 가구 소실(2020년 3.2%, 2021년 5.2%), 해외 취업, 부적합 가구 등의 조사 불가능(2020년 2.2%, 2021년 3.2%) 순으로 높게 나타났으며 각각의 비율도 모두 증가하는 추세를 보였다.



제3장

패널 대상의 표본이탈 특성 분석

제1절 표본이탈의 정의

제2절 패널가구의 표본이탈 특성 분석

제3절 패널가구원의 표본이탈 특성 분석

제4절 소결

제 3 장 패널 대상의 표본이탈 특성 분석

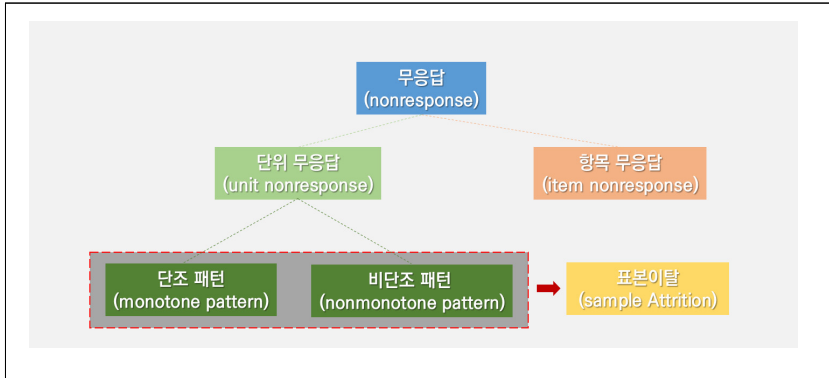
2기 한국의료패널조사의 표본이탈 경향에 대해 2019년~2021년 연간 데이터를 대상으로 하여 살펴보았다. 먼저 제1절에서는 표본이탈의 개념에 대해 정리하였고, 제2절과 제3절에서는 패널가구와 패널가구원의 이탈 특성 및 표본이탈에 영향을 주는 요인을 분석하였다.

제1절 표본이탈의 정의

표본이탈이란 개인, 가구, 사업체 등의 표본 대상이 조사에 참여하지 않아 더 이상 자료를 수집할 수 없게 되는 영구적인 손실을 의미한다 (Uhrig, 2008; Lynn, Buck, Burton, Jäckle & Laurie. 2005; Zabel 1998 재인용).

패널조사의 단위무응답에서 표본이탈이 나타나는데, 무응답 패턴은 단조 패턴과 비단조 패턴으로 구성된다([그림 3-1] 참조). 단조 패턴은 패널 응답자가 조사 참여를 중단한 이후 다시는 참여하지 않는 경우를 의미한다(이혜정, 2018, p. 9). 그러나 비단조 패턴은 단위무응답이 간헐적으로 발생하는 형태로, 즉 현 조사 차수에 참여하지 않았더라도 다음 조사 차수에서는 참여하는 등의 이탈과 재진입을 반복하는 것을 의미한다 (이혜정, 2018, p. 9).

[그림 3-1] 패널조사의 무응답 구성



자료: 이상호, 이해정. (2017). 한국노동패널조사(KLIPS)의 패널이탈 분석-비단조 이탈을 중심으로. p.5를 재구성함.

패널조사에서 표본이탈은 두 가지 이유로 문제가 될 수 있다(Uhrig, 2008, p. 1). 조사 차수가 거듭되면서 패널 표본 크기가 감소함에 따라 해당 패널 표본에서 산출된 추정치의 정밀도(신뢰도)가 낮아지게 된다 (Branden, Gritz & Pergamit, 1995; Watson 2003 재인용). 더 중요한 점은 표본이탈이 무작위가 아닐 수 있는데, 비무작위 이탈(non-random attrition)이면 시간이 지남에 따라 패널 표본은 대표성을 가지지 못한다. 또한, 관심 있는 연구 변수가 이탈과 관련된 요인만큼 편향된 연구 결과를 나타낼 수 있다.

제2절 패널가구의 표본이탈 특성 분석

1. 패널가구의 응답형태 현황 및 특성 분석

여기에서는 3개 연간데이터의 패널가구의 응답형태에 따른 주요 특성을 살펴보았다. 분석 대상은 분가 가구를 제외한 원가구만을 대상으로 하였으며 개인정보 활용에 대해 미동의한 가구원을 포함한 가구 및 가구원은 제외하였다. 설문조사에 참여하여 응답했으나, 개인정보 제공에 동의하지 않아서 응답값을 결측으로 처리하였기 때문이다.

해당 연도에서 ‘패널가구’가 조사에 참여하면 ‘1’이고 참여하지 않으면 ‘0’이라고 정의하였다. 예를 들면 ‘연속응답’ 가구는 3개 연도 모두 조사에 참여하였으며, 기호는 111이 된다. ‘이탈’ 가구는 2019년 조사를 참여하였으나 이후 조사에서 참여하지 않은 경우로, 기호는 100(‘2020년 이탈’) 또는 110(‘2021년 이탈’)이 해당된다. 3개 연도 분석이어서 이탈 가구라고 보기 어려울 수 있지만, 용어의 명칭으로 사용하였다. 한편, 2019년 조사는 참여한 다음에 2021년에 다시 조사에 참여한 경우를 ‘재진입’이라고 정의하였다(기호: 101). 2020년과 2021년 분가 가구는 표본이탈 특성 분석에서 제외하였다.

〈표 3-1〉을 보면 패널가구의 응답형태는 연속응답이 85.7%로 가장 높았으며 다음으로 첫 번째 조사 이후 두 번째에서 이탈이 8.4%, 세 번째에서 이탈이 5.6%로 나타났다. 두 번째 조사에서 표본이탈을 많이 한다는 것을 확인할 수 있다. 한편, 첫 번째 조사 이후 세 번째 조사에서 다시 응답한 형태도 있는데 이는 0.3%로 낮은 편에 속하였다.

〈표 3-1〉 패널가구의 응답형태별 분포

(단위: 가구, %)

구분	기호	가구	비율
연속응답	111	5,784	85.7
이탈	2020년	565	8.4
	2021년	379	5.6
재진입	101	20	0.3
전체		6,748	100

주: '기호'에서 '1'은 응답에 참여한 가구, '0'은 응답에 참여하지 않은 이탈 가구임.
 자료: 한국보건사회연구원·국민건강보험공단. (2023). 2기 한국의료패널조사 2019년~2021년 연
 간데이터(2023년 학술대회용 버전)를 분석함.

패널가구의 응답형태별 거주 지역, 동부/읍면부, 가구원 수, 세대구성, 주택형태, 연간 가구 총소득 분포에 대해 살펴보았다. 〈표 3-2〉를 보면 패널가구의 응답형태에 따른 거주 지역으로 연속응답의 경우 대구와 인천이 각각 75.5%, 78.3%로 가장 낮게 나타났다. 반면에 강원(91.2%), 경남(90.6%)과 충남(90.5%)이 높은 편에 속하였다. 2020년 이탈은 세종이 15.4%로 가장 높았고, 대구(12.6%)와 인천(12.4%) 순으로 높게 나타났다. 또한 2021년 이탈에서도 대구(11.9%)와 인천(9.1%)이 가장 높았다. 대구와 인천은 다른 지역에 비해 패널가구의 이탈률이 높다고 볼 수 있다.

〈표 3-2〉 패널가구의 응답형태별 거주 지역 분포

(단위: 가구, %)

	111	100	110	101	합계	111	100	110	101	합계
서울	526	61	36	2	625	84.2	9.8	5.8	0.3	100.0
부산	398	23	22	3	446	89.2	5.2	4.9	0.7	100.0
대구	203	34	32	0	269	75.5	12.6	11.9	0.0	100.0
인천	335	53	39	1	428	78.3	12.4	9.1	0.2	100.0
광주	319	30	15	2	366	87.2	8.2	4.1	0.5	100.0
대전	237	16	15	1	269	88.1	5.9	5.6	0.4	100.0
울산	311	35	25	1	372	83.6	9.4	6.7	0.3	100.0
세종	136	26	7	0	169	80.5	15.4	4.1	0.0	100.0
경기	667	83	49	3	802	83.2	10.3	6.1	0.4	100.0
강원	207	13	7	0	227	91.2	5.7	3.1	0.0	100.0
충북	431	32	30	1	494	87.2	6.5	6.1	0.2	100.0
충남	238	14	9	2	263	90.5	5.3	3.4	0.8	100.0
전북	420	30	22	1	473	88.8	6.3	4.7	0.2	100.0
전남	386	36	23	0	445	86.7	8.1	5.2	0.0	100.0
경북	343	24	15	2	384	89.3	6.3	3.9	0.5	100.0
경남	490	30	21	0	541	90.6	5.5	3.9	0.0	100.0
제주	96	10	9	1	116	82.8	8.6	7.8	0.9	100.0
전체	5,743	550	376	20	6,689	85.9	8.2	5.6	0.3	100.0

주: 111은 '연속응답'이고 100은 '2020년 이탈'이고 110은 '2021년 이탈'이고 101은 '재진입'을 의미함.

자료: 한국보건사회연구원·국민건강보험공단. (2023). 2기 한국의료패널조사 2019년~2021년 연간데이터(2023년 학술대회용 버전)를 분석함.

패널가구의 응답형태에 따른 동부/읍면부 분포를 보면 연속응답은 읍면부(90.1%)가 동부(84.5%)에 비해 5.6%p 높았으나, 2019년 응답 이후 이탈률은 동부가 높게 나타났다(〈표 3-3〉 참조). 보통 도시 지역의 조사 참여도가 비도시 지역에 비해 낮은 편인데 이와 동일한 결과를 보였다.

〈표 3-3〉 패널가구의 응답형태별 동부/읍면부 분포

(단위: 가구, %)

	111	100	110	101	합계	111	100	110	101	합계
동부	4,264	455	311	17	5,047	84.5	9.0	6.2	0.3	100.0
읍면부	1,479	95	65	3	1,642	90.1	5.8	4.0	0.2	100.0
전체	5,743	550	376	20	6,689	85.9	8.2	5.6	0.3	100.0

주: 111은 '연속응답'이고 100은 '2020년 이탈'이고 110은 '2021년 이탈'이고 101은 '재진입'을 의미함.

자료: 한국보건사회연구원·국민건강보험공단. (2023). 2기 한국의료패널조사 2019년~2021년 연간데이터(2023년 학술대회용 버전)를 분석함.

패널가구의 응답형태에 따른 가구원 수의 분포를 보면 연속응답의 비율은 3인(83%)을 기준으로 가구원 수가 적을수록 커지며, 가구원 수가 많아질수록 작게 나타났다(〈표 3-4〉 참조). 이와 반대로 2019년 이후 이탈 비율은 가구원 수가 적을수록 낮으며, 가구원 수가 많아질수록 높게 나타났다. 한국의료패널조사의 경우 혼자 사는 고령 가구의 구성이 상대적으로 높은 편이어서 이러한 결과가 나온 것으로 보인다.

〈표 3-4〉 패널가구의 응답형태별 가구원 수 분포

(단위: 가구, %)

	111	100	110	101	합계	111	100	110	101	합계
1인	1,286	68	82	4	1,440	89.3	4.7	5.7	0.3	100.0
2인	2,358	167	119	6	2,650	89.0	6.3	4.5	0.2	100.0
3인	942	122	67	4	1,135	83.0	10.7	5.9	0.4	100.0
4인	869	145	81	5	1,100	79.0	13.2	7.4	0.5	100.0
5인 이상	288	48	27	1	364	79.1	13.2	7.4	0.3	100.0
전체	5,743	550	376	20	6,689	85.9	8.2	5.6	0.3	100.0

주: 111은 '연속응답'이고 100은 '2020년 이탈'이고 110은 '2021년 이탈'이고 101은 '재진입'을 의미함.

자료: 한국보건사회연구원·국민건강보험공단. (2023). 2기 한국의료패널조사 2019년~2021년 연간데이터(2023년 학술대회용 버전)를 분석함.

다음 <표 3-5>는 패널가구의 응답형태에 따른 세대구성이며 연속응답의 경우 1세대가 89.4%로 가장 높고 3세대 이상(82.4%), 2세대(81.4%) 순으로 나타났다. 2019년 이후 이탈 비율은 1세대가 가장 낮고, 2세대와 3세대가 높았다. 이러한 결과는 앞에서 살펴본 가구원 수 분포와 일맥상통한다고 볼 수 있다.

<표 3-5> 패널가구의 응답형태별 세대구성 분포

(단위: 가구, %)

	111	100	110	101	합계	111	100	110	101	합계
1세대	3,316	203	181	9	3,709	89.4	5.5	4.9	0.2	100.0
2세대	2,235	320	181	11	2,747	81.4	11.6	6.6	0.4	100.0
3세대 이상	192	27	14	0	233	82.4	11.6	6.0	0.0	100.0
전체	5,743	550	376	20	6,689	85.9	8.2	5.6	0.3	100.0

주: 111은 '연속응답'이고 100은 '2020년 이탈'이고 110은 '2021년 이탈'이고 101은 '재진입'을 의미함.

자료: 한국보건사회연구원·국민건강보험공단. (2023). 2기 한국의료패널조사 2019년~2021년 연간데이터(2023년 학술대회용 버전)를 분석함.

다음 <표 3-6>은 패널가구의 응답형태에 따른 주거형태로 연속응답의 경우 주거형태가 단독주택일 때 90%로 가장 높고, 아파트가 82.9%로 가장 낮았다. 2020년 이탈의 경우 다세대주택(10.8%)과 아파트(10.1%)가 가장 높고, 기타(5%)와 단독주택(5.6%)이 낮게 나타났다. 2021년 이탈의 경우 기타(7.9%)가 가장 높고, 단독주택(4.1%)이 가장 낮았다. 단독주택인 경우 다른 주거형태에 비해 조사참여도가 높다고 볼 수 있다.

52 패널조사 표본이탈 경향 분석과 추가표본 추출에 관한 연구- 한국의료패널조사를 중심으로

〈표 3-6〉 패널가구의 응답형태별 주거형태 분포

(단위: 가구, %)

	111	100	110	101	합계	111	100	110	101	합계
단독주택	2,378	148	108	9	2,643	90.0	5.6	4.1	0.3	100.0
아파트	2,560	313	207	8	3,088	82.9	10.1	6.7	0.3	100.0
연립주택	211	18	17	1	247	85.4	7.3	6.9	0.4	100.0
다세대주택	506	66	36	2	610	83.0	10.8	5.9	0.3	100.0
기타	88	5	8	0	101	87.1	5.0	7.9	0.0	100.0
전체	5,743	550	376	20	6,689	85.9	8.2	5.6	0.3	100.0

주: 1) 111은 '연속응답'이고 100은 '2020년 이탈'이고 110은 '2021년 이탈'이고 101은 '재진입'을 의미함.

2) 기타는 영업용 건물 내 주택(상가, 공장 등), 오피스텔, 호텔, 여관, 기숙사 등을 의미함.

자료: 한국보건사회연구원·국민건강보험공단. (2023). 2기 한국의료패널조사 2019년~2021년 연간데이터(2023년 학술대회용 버전)를 분석함.

패널가구의 응답형태에 따른 연간 가구 총소득을 보면 〈표 3-7〉과 같다. 연간 가구 총소득이 낮을수록 연속응답의 비율이 높았다. 2,000만 원 미만인 경우 87.3%로 가장 높았으나 6,000만 원 이상인 경우는 83.2%로 가장 낮게 나타났다. 2020년 이탈의 경우 소득이 높을수록 이탈할 비율이 높고, 소득이 낮을수록 낮았다. 2021년 이탈의 경우 6,000만 원 이상이 7%로 가장 높았고, 4,000만 원 이상~6,000만 원 미만이 4.8%로 가장 낮게 나타났다.

〈표 3-7〉 패널가구의 응답형태별 연간 가구 총소득 분포

(단위: 가구, %)

	111	100	110	101	합계	111	100	110	101	합계
2,000만 원 미만	2,048	162	133	3	2,346	87.3	6.9	5.7	0.1	100.0
4,000만 원 미만	1,569	151	93	8	1,821	86.2	8.3	5.1	0.4	100.0
6,000만 원 미만	971	98	54	6	1,129	86.0	8.7	4.8	0.5	100.0
6,000만 원 이상	1,081	125	91	3	1,300	83.2	9.6	7.0	0.2	100.0
전체	5,669	536	371	20	6,596	85.9	8.1	5.6	0.3	100.0

주: 1) 111은 '연속응답'이고 100은 '2020년 이탈'이고 110은 '2021년 이탈'이고 101은 '재진입'을 의미함.

2) 모름/무응답이라고 응답한 93가구는 제외하고 분석함.

자료: 한국보건사회연구원·국민건강보험공단. (2023). 2기 한국의료패널조사 2019년~2021년 연간데이터(2023년 학술대회용 버전)를 분석함.

2. 패널가구의 응답형태에 따른 영향 분석

패널가구의 응답형태에 따른 집단을 종속 변수로 하여 다항 로지스틱 회귀모형(multinomial logistic regression model)을 분석하였다. 이때 재진입을 제외한 3개 집단(연속응답, 2020년 이탈, 2021년 이탈)에 대해서만 살펴보았다. 재진입의 경우 관측치가 20가구로 작아 모형 적합이 되지 않을 수 있기 때문이다. 준거집단은 연속응답으로 설정하였다. 사용한 설명 변수는 2019년 연간데이터에서의 가구 특성(거주 지역, 주거형태, 가구원 수, 연간 가구 총소득), 2019년~2021년에 걸친 가구의 상태 변화(이사 횟수), 그리고 2019년에서의 가구주 특성(성별, 연령, 혼인상태, 최종 학력)이다.

〈표 3-8〉은 모형의 분석 결과로, 연속응답을 준거집단으로 하여 2020년 이탈 집단에 영향을 주는 요인을 보면 주거형태, 이사 횟수, 가구주 연령이 통계적으로 유의한 결과를 나타내고 있다. 주거형태가 단독주택(기준 변수: 아파트)인 경우 2020년 이탈 집단일 가능성이 연속응답 집단일 가능성(odds ratio, 오즈비)보다 26.8% 감소하고, 다세대주택인 경우 39.6% 증가한다고 볼 수 있다. 이사 횟수는 한 단위 증가할수록 2020년 이탈 집단일 가능성이 79.2% 감소한다고 볼 수 있다. 가구주 연령이 한 단위 증가할수록 2020년 이탈 집단일 가능성이 2.2% 감소한다고 볼 수 있다.

다음은 2021년 이탈에 영향을 주는 요인을 연속응답을 준거집단으로 하여 보면 주거형태, 연간 가구 총소득, 이사 횟수, 가구주 연령, 가구주 최종 학력이 통계적으로 유의한 결과를 나타내고 있다. 주거형태가 단독주택(기준변수: 아파트)인 경우 2021년 이탈 집단일 가능성이 연속응답 집단일 가능성보다 29.8% 감소한다고 볼 수 있다. 연간 가구 총소득이 저소득(기준변수: 중소득)인 경우 2021년 이탈 집단일 가능성이 37%

증가하고, 이사 횟수는 한 단위 증가할수록 2021년 이탈 집단일 가능성이 39.4% 감소한다고 볼 수 있다. 가구주 연령이 한 단위 증가할수록 2021년 이탈 집단일 가능성이 1.9% 감소한다고 볼 수 있다. 가구주 최종 학력이 초졸(기준변수: 고졸)인 경우 2021년 이탈 집단일 가능성이 40.6% 감소한다고 볼 수 있다.

2020년과 2021년 패널가구의 이탈에 영향을 주는 공통적인 요인은 주거형태, 이사 횟수, 가구주 연령이었다. 아파트가 단독주택에 비해, 이사 횟수는 적을수록, 가구주 연령이 젊을수록 패널가구의 이탈 경향이 높은 편으로 나타났다. 이사 횟수의 경우는 이상호, 이해정(2017)의 연구 결과와 유사하였고, 남기성, 천영민(2012)의 연구에 따르면 이사의 경우 주소 변경을 파악할 수 있다면 일시적으로 이탈이 발생하더라도 재복귀가 가능하다고 하였다. 이를 통해 패널 대상에 대한 꾸준한 추적 관리가 중요함을 알 수 있다.

〈표 3-8〉 패널가구의 응답형태에 대한 다항 로지스틱 회귀모형 분석 결과

변수	2020년 이탈		2021년 이탈	
	coefficient	SE	coefficient	SE
상수항	-1.030 (***)	0.447	-1.446 (***)	0.504
거주 지역_서울				
광역시	-0.200	0.181	0.028	0.204
경기	-0.005	0.207	-0.053	0.241
기타 도지역	-0.237	0.178	-0.207	0.204
주거형태_아파트				
단독주택	-0.312 (***)	0.128	-0.354 (**)	0.142
연립주택	-0.263	0.300	0.162	0.270
다세대주택	0.334 (**)	0.162	-0.079	0.201
기타	-0.198	0.473	0.243	0.387
가구원 수	0.094	0.057	0.035	0.068
연간 가구 총소득_중소득				
저소득	-0.143	0.150	0.315 (**)	0.160
고소득	0.064	0.133	0.078	0.156
이사 횟수	-1.570 (***)	0.229	-0.502 (***)	0.155
가구주 성별_남성				
여성	0.224	0.201	-0.256	0.204
가구주 연령	-0.023 (***)	0.005	-0.020 (***)	0.006
				exp(b)
				-
				1.029
				0.948
				0.813
				0.702
				1.176
				0.925
				1.275
				1.035
				1.370
				1.081
				0.606
				0.774
				0.981

변수	2020년 이탈		2021년 이탈	
	coefficient	SE	coefficient	SE
기구조 혼인상태_유배우자				
미혼	-0.391	0.300	0.676	0.277
기타	0.025	0.207	1.025	0.214
기구조 최종 학력_고졸				
무학	0.005	0.264	1.005	0.263
초졸	-0.158	0.186	0.854	0.215
중졸	0.201	0.170	1.222	0.191
대졸 이상	0.007	0.132	1.007	0.147

주: 1) coefficient에서 () 안의 값은 p-value에 대한 유의성을 나타내며 (****)은 0.01, (***)은 0.05, (**)은 0.1에서 통계적으로 유의하다고 할 수 있음.
 2) 주거형태에서 기타는 영업용 건물 내 주택, 오피스텔, 호텔, 여관, 기숙사 등이 해당함.
 3) 혼인상태에서 기타는 별거/이혼, 사별/실종이 해당함.
 4) 연간 가구 총소득에서 저소득은 0원~2,000만 원 미만, 중소득은 2,000만 원~5,000만 원 미만, 고소득은 5,000만 원 이상을 의미함.
 자료: 한국보건사회연구원·국민건강보험공단. (2023). 2기 한국의료패널조사 2019년~2021년 연간데이터(2023년 학술대회용 버전)를 분석함.

제3절 패널가구원의 표본이탈 특성 분석

1. 패널가구원의 응답형태 현황 및 특성 분석

패널가구원의 응답형태 현황 및 특성에 대해 살펴보았다. 앞의 패널 가구의 응답형태와 동일하게 정의하였다. 단, 해당 연도에서 ‘패널가구원’이 조사에 참여하면 ‘1’이고 조사에 참여하지 않으면 ‘0’으로 정의하였다. ‘연속응답’ 가구원은 3개 연도 모두 조사에 참여하였으며, 기호는 111이 된다. ‘이탈’ 가구원은 2019년 조사를 참여하였으나 이후 조사에서 참여하지 않은 경우로, 기호는 100(‘2020년 이탈’) 또는 110(‘2021년 이탈’)이 된다. 마지막으로 2019년 조사는 참여한 다음에 2021년 다시 조사를 참여한 경우를 ‘재진입’ 가구원이라고 정의하였다(기호: 101).

〈표 3-9〉를 보면 패널가구원의 응답형태가 연속응답이 81.5%로 가장 높았으며 다음으로 첫 번째 조사 이후 두 번째에서의 이탈이 11%, 세 번째에서의 이탈이 7.2%로 나타났다. 한편, 첫 번째 조사 이후 세 번째 조사에서 다시 응답한 형태도 있는데 0.3%로 낮은 편에 속하였다. 앞의 패널 가구의 응답형태와 비교하면 패널가구원의 연속응답의 비율은 더 낮았고(패널가구 85.7%), 첫 번째와 두 번째에서 이탈 비율은 더 높았다(패널가구 8.4%, 5.6%). 이는 패널가구원의 조사 참여에 대한 충성도가 높지 않다고 볼 수 있으며 조사 참여를 높이기 위한 방안을 마련할 필요가 있다.

〈표 3-9〉 패널가구의 응답형태별 분포

(단위: 명, %)

구분	기호	가구원	비율
연속응답	111	13,520	81.5
이탈	2020년	1,821	11.0
	2021년	1,193	7.2
재진입	101	53	0.3
전체		16,587	100

주: '기호'에서 '1'은 응답에 참여한 가구, '0'은 응답에 참여하지 않은 이탈 가구임.
 자료: 한국보건사회연구원·국민건강보험공단. (2023). 2기 한국의료패널조사 2019년~2021년 연간데이터(2023년 학술대회용 버전)를 분석함.

패널가구의 응답형태에 따른 성별 분포를 보면 연속응답의 비율은 여성이 83.8%로 남성(83.1%)에 비해 0.7%p 높게 나타났다(〈표 3-10〉 참조). 2020년 이후 이탈은 성별에 따른 차이가 크지 않았으나, 2021년 이후 이탈 비율은 남성(7.2%)이 여성(6.3%)에 비해 높게 나타났다. 전반적으로 보면 여성이 남성에 비해 조사 참여도가 높은 편이라고 볼 수 있다.

〈표 3-10〉 패널가구의 응답형태별 성별 분포

(단위: 명, %)

	111	100	110	101	합계	111	100	110	101	합계
남성	5,706	646	497	21	6,870	83.1	9.4	7.2	0.3	100.0
여성	6,593	752	498	28	7,871	83.8	9.6	6.3	0.4	100.0
전체	12,299	1,398	995	49	14,741	83.4	9.5	6.7	0.3	100.0

주: 111은 '연속응답'이고 100은 '2020년 이탈'이고 110은 '2021년 이탈'이고 101은 '재진입'을 의미함.
 자료: 한국보건사회연구원·국민건강보험공단. (2023). 2기 한국의료패널조사 2019년~2021년 연간데이터(2023년 학술대회용 버전)를 분석함.

다음 〈표 3-11〉은 패널가구의 응답형태에 따른 연령대 분포로 연속응답의 비율은 50세 이상~79세 이하가 높은 편이고, 49세 이하가 낮은 편에 속하였다. 이와 반대로 2020년과 2021년 이후 이탈은 50세 이상~

79세 이하가 낮은 편이고, 49세 이하가 높은 편으로 나타났다. 한편, 80세 이상의 경우에는 연속응답 비율은 낮고 이탈 비율은 높은 경향을 보였는데, 이는 노환, 사망 등으로 인한 표본이탈로 볼 수 있다.

〈표 3-11〉 패널가구원의 응답형태별 연령대 분포

(단위: 명, %)

	111	100	110	101	합계	111	100	110	101	합계
0-9세	1,043	164	99	7	1,313	79.4	12.5	7.5	0.5	100.0
10-19세	906	146	88	6	1,146	79.1	12.7	7.7	0.5	100.0
20-29세	738	111	98	2	949	77.8	11.7	10.3	0.2	100.0
30-39세	1,107	139	113	7	1,366	81.0	10.2	8.3	0.5	100.0
40-49세	1,462	221	153	5	1,841	79.4	12.0	8.3	0.3	100.0
50-59세	1,808	211	139	13	2,171	83.3	9.7	6.4	0.6	100.0
60-69세	2,503	215	115	7	2,840	88.1	7.6	4.0	0.2	100.0
70-79세	2,116	116	108	2	2,342	90.4	5.0	4.6	0.1	100.0
80세 이상	616	75	82	0	773	79.7	9.7	10.6	0.0	100.0
전체	12,299	1,398	995	49	14,741	83.4	9.5	6.7	0.3	100.0

주: 111은 '연속응답'이고 100은 '2020년 이탈'이고 110은 '2021년 이탈'이고 101은 '재진입'을 의미함.

자료: 한국보건사회연구원·국민건강보험공단. (2023). 2기 한국의료패널조사 2019년~2021년 연간데이터(2023년 학술대회용 버전)를 분석함.

패널가구원의 응답형태에 따른 혼인상태 분포를 보면 연속응답은 사별/실종(87.8%), 혼인 중(84.6%), 별거/이혼(83.6%)으로 높은 편이며 미혼은 79.4%로 가장 낮게 나타났다(〈표 3-12〉 참조). 이는 미혼의 경우 다른 혼인상태에 비해 표본이탈 경향이 높다고 볼 수 있다. 한편, 미혼 다음으로 별거/이혼인 경우도 2020년과 2021년 이탈 비율(각각 9.5%, 6.5%)이 높은 편으로 나타났다.

〈표 3-12〉 패널가구원의 응답형태별 혼인상태 분포

(단위: 명, %)

	111	100	110	101	합계	111	100	110	101	합계
미혼	3,154	467	333	17	3,971	79.4	11.8	8.4	0.4	100.0
혼인 중	7,324	774	534	26	8,658	84.6	8.9	6.2	0.3	100.0
별거/이혼	616	70	48	3	737	83.6	9.5	6.5	0.4	100.0
사별/실종	1,205	85	80	3	1,373	87.8	6.2	5.8	0.2	100.0
전체	12,299	1,396	995	49	14,739	83.4	9.5	6.8	0.3	100.0

주: 1) 111은 '연속응답'이고 100은 '2020년 이탈'이고 110은 '2021년 이탈'이고 101은 '재진입'을 의미함.

2) 응답 거부 2명은 제외하고 분석하였음.

자료: 한국보건사회연구원·국민건강보험공단. (2023). 2기 한국의료패널조사 2019년~2021년 연간데이터(2023년 학술대회용 버전)를 분석함.

다음 〈표 3-13〉은 패널가구원의 응답형태에 따른 최종 학력 분포로 연속응답의 경우 초졸과 중졸의 비율이 높은 반면에, 무학과 고졸 이상의 비율은 낮게 나타났다. 이와 반대로 표본이탈 비율은 초졸과 중졸의 비율이 낮은 반면에, 무학과 고졸 이상의 비율이 높게 나타났다. 학력이 높아질수록 표본이탈 경향성이 높아진다고 볼 수 있다. 여기서, 무학의 경우에는 미취학도 포함되어 있어 고학력 부모가 표본이탈하면서 그의 자녀도 함께 표본이탈하게 되었기 때문으로 보인다.

〈표 3-13〉 패널가구원의 응답형태별 최종 학력 분포

(단위: 명, %)

	111	100	110	101	합계	111	100	110	101	합계
무학	2,082	275	189	9	2,555	81.5	10.8	7.4	0.4	100.0
초졸	2,102	170	119	4	2,395	87.8	7.1	5.0	0.2	100.0
중졸	1,606	156	115	6	1,883	85.3	8.3	6.1	0.3	100.0
고졸	3,458	406	301	20	4,185	82.6	9.7	7.2	0.5	100.0
대졸 이상	3,051	391	271	10	3,723	82.0	10.5	7.3	0.3	100.0
전체	12,299	1,398	995	49	14,741	83.4	9.5	6.7	0.3	100.0

주: 1) 111은 '연속응답'이고 100은 '2020년 이탈'이고 110은 '2021년 이탈'이고 101은 '재진입'을 의미함.

2) 무학은 '미취학도 포함'하여 분석하였음.

자료: 한국보건사회연구원·국민건강보험공단. (2023). 2기 한국의료패널조사 2019년~2021년 연간데이터(2023년 학술대회용 버전)를 분석함.

마지막으로 <표 3-14>는 패널가구원의 응답형태에 따른 경제활동 상태 분포로 연속응답의 경우 기타(무급가족종사자, 자활근로, 공공근로, 희망근로, 노인 일자리)가 91.8%이고 고용주/자영업자가 97%로 높은 편이었다. 이에 반해 15세 미만 응답자인 ‘해당 없음’은 79.5%로 가장 낮게 나타났다. 2020년과 2021년 표본이탈 경향을 보면 ‘해당 없음’과 비경제활동인구의 비율은 높은 반면에, 기타와 고용주/자영업자의 비율은 낮았다.

<표 3-14> 패널가구원의 응답형태별 경제활동 상태 분포

(단위: 명, %)

	111	100	110	101	합계	111	100	110	101	합계
임금근로자	3,880	459	316	17	4,672	83.0	9.8	6.8	0.4	100.0
고용주/ 자영업자	1,698	133	112	9	1,952	87.0	6.8	5.7	0.5	100.0
비경제 활동인구	4,226	514	388	12	5,140	82.2	10.0	7.5	0.2	100.0
기타	960	51	34	1	1,046	91.8	4.9	3.3	0.1	100.0
해당 없음	1,535	241	145	10	1,931	79.5	12.5	7.5	0.5	100.0
전체	12,299	1,398	995	49	14,741	83.4	9.5	6.7	0.3	100.0

주: 1) 111은 ‘연속응답’이고 100은 ‘2020년 이탈’이고 110은 ‘2021년 이탈’이고 101은 ‘재진입’을 의미함.

2) 기타는 무급가족종사자, 자활근로, 공공근로, 희망근로, 노인 일자리가 해당함.

3) 해당 없음은 15세 미만 응답자임.

자료: 한국보건사회연구원·국민건강보험공단. (2023). 2기 한국의료패널조사 2019년~2021년 연간데이터(2023년 학술대회용 버전)를 분석함.

한편, 앞에서 3개 연간데이터의 패널가구와 패널가구원의 응답형태에 따른 분포를 살펴보았는데(<표 3-1> 및 <표 3-9> 참조), 이를 조합하면 4개 집단으로 구분할 수 있다. 패널가구와 패널가구원이 모두 응답한 경우는 ‘패널가구 응답 & 패널가구원 응답’으로, 패널가구는 응답하였으나 패널가구원이 무응답한 경우는 ‘패널가구 응답 & 패널가구원 무응답’으로, 패널가구와 패널가구원 모두 무응답한 경우는 ‘패널가구 무응답 & 패널

가구원 무응답'으로 표기하였다. 또한 재진입 등 기타는 '재진입 등 기타'로 동일하게 표기하였다.

4개 집단에 따라 2019년 기준 패널가구원의 인구사회학적 특성에 대해 살펴보았다. <표 3-15>를 보면 '패널가구 응답 & 패널가구원 응답'이 81.4%로 가장 높았다. 다음으로 '패널가구 무응답 & 패널가구원 무응답'이 15.8%이고 '패널가구 응답 & 패널가구원 무응답'이 2.4%의 순이었으며 재진입 등 기타는 0.4%로 미미하였다. 이를 통해 패널가구의 응답 여부가 패널을 유지하는 데 중요하다는 것을 알 수 있다.

<표 3-15> 패널가구 및 패널가구원의 응답형태별 분포

(단위: 명, %)

구분		가구원	비율
1	패널가구 응답 & 패널가구원 응답	13,510	81.4
2	패널가구 응답 & 패널가구원 무응답	398	2.4
3	패널가구 무응답 & 패널가구원 무응답	2,618	15.8
4	재진입 등 기타	61	0.4
전체		16,587	100

자료: 한국보건사회연구원·국민건강보험공단. (2023). 2기 한국의료패널조사 2019년~2021년 연간데이터(2023년 학술대회용 버전)를 분석함.

다음으로 4개 집단별 패널가구원의 응답형태에 따른 성별 분포를 보면 '패널가구 응답 & 패널가구원 무응답'의 비율은 남성이 2.2%로 여성(1.1%)에 비해 1.1%p 높게 나타났다(<표 3-16> 참조). 패널가구가 응답한 경우 패널가구원 중에서 남성이 응답하지 않을 가능성이 높다고 볼 수 있다. 나머지 집단은 성별에 따른 차이가 크지 않은 편에 속하였다.

〈표 3-16〉 4개 집단별 패널가구원의 응답형태별 성별 분포

(단위: 명, %)

	1	2	3	4	합계	1	2	3	4	합계
남성	5,703	154	990	23	6,870	83.0	2.2	14.4	0.3	100.0
여성	6,588	87	1,163	33	7,871	83.7	1.1	14.8	0.4	100.0
전체	12,291	241	2,153	56	14,741	83.4	1.6	14.6	0.4	100.0

주: 1은 '패널가구 응답 & 패널가구원 응답'이고 2는 '패널가구 응답 & 패널가구원 무응답'이고 3은 '패널가구 무응답 & 패널가구원 무응답'이고 4는 '재진입 등 기타'를 의미함.

자료: 한국보건사회연구원·국민건강보험공단. (2023). 2기 한국의료패널조사 2019년~2021년 연간데이터(2023년 학술대회용 버전)를 분석함.

다음 〈표 3-17〉은 4개 집단별 패널가구원의 응답형태에 따른 연령대 분포로 '패널가구 응답 & 패널가구원 무응답'의 비율은 80세 이상의 비율이 8.7%로 가장 높았다. 이는 앞에서 살펴보았던 경향과 유사했으며 노환, 사망 등으로 인한 무응답이라고 볼 수 있다(〈표 3-11〉 참조). 다음으로 20세 이상~29세 이하가 4.1%이고 70세 이상~79세 이하의 비율이 2.3%의 순이었다. '패널가구 무응답 & 패널가구원 무응답'의 비율은 0세 이상~19세 이하 및 40세 이상~49세 이하가 높은 편에 속하였다. 이를 통해 중년의 부모와 자녀로 구성된 가족이 패널에서 이탈할 가능성이 가장 높다고 볼 수 있다.

64 패널조사 표본이탈 경향 분석과 추가표본 추출에 관한 연구- 한국의료패널조사를 중심으로

〈표 3-17〉 4개 집단별 패널가구원의 응답형태별 연령대 분포

(단위: 명, %)

	1	2	3	4	합계	1	2	3	4	합계
0-9세	1,043	4	259	7	1,313	79.4	0.3	19.7	0.5	100.0
10-19세	906	9	225	6	1,146	79.1	0.8	19.6	0.5	100.0
20-29세	735	39	170	5	949	77.4	4.1	17.9	0.5	100.0
30-39세	1,104	23	230	9	1,366	80.8	1.7	16.8	0.7	100.0
40-49세	1,462	14	360	5	1,841	79.4	0.8	19.6	0.3	100.0
50-59세	1,808	10	340	13	2,171	83.3	0.5	15.7	0.6	100.0
60-69세	2,502	21	309	8	2,840	88.1	0.7	10.9	0.3	100.0
70-79세	2,115	54	170	3	2,342	90.3	2.3	7.3	0.1	100.0
80세 이상	616	67	90	0	773	79.7	8.7	11.6	0.0	100.0
전체	12,291	241	2,153	56	14,741	83.4	1.6	14.6	0.4	100.0

주: 1은 '패널가구 응답 & 패널가구원 응답'이고 2는 '패널가구 응답 & 패널가구원 무응답'이고 3은 '패널가구 무응답 & 패널가구원 무응답'이고 4는 '재진입 등 기타'를 의미함.

자료: 한국보건사회연구원·국민건강보험공단. (2023). 2기 한국의료패널조사 2019년~2021년 연간데이터(2023년 학술대회용 버전)를 분석함.

〈표 3-18〉은 4개 집단별 패널가구원의 응답형태에 따른 혼인상태 분포이며 '패널가구 응답 & 패널가구원 무응답'의 비율은 사별/실종이 2.5%로 가장 높게 나타났다. 이에 반해 '패널가구 무응답 & 패널가구원 무응답'의 비율은 미혼이 18.2%로 가장 높게 나타났다. 이는 미혼인 경우 다른 혼인상태에 비해 패널에서 이탈할 경향이 높으며 앞에서 살펴보았던 결과와 유사하다고 볼 수 있다(〈표 3-12〉 참조).

〈표 3-18〉 4개 집단별 패널가구원의 응답형태별 혼인상태 분포

(단위: 명, %)

	1	2	3	4	합계	1	2	3	4	합계
미혼	3,148	77	724	22	3,971	79.3	1.9	18.2	0.6	100.0
혼인 중	7,324	118	1,190	26	8,658	84.6	1.4	13.7	0.3	100.0
별거/이혼	615	10	108	4	737	83.4	1.4	14.7	0.5	100.0
사별/실종	1,204	35	130	4	1,373	87.7	2.5	9.5	0.3	100.0
전체	12,291	240	2,152	56	14,739	83.4	1.6	14.6	0.4	100.0

주: 1) 1은 '패널가구 응답 & 패널가구원 응답'이고 2는 '패널가구 응답 & 패널가구원 무응답'이고 3은 '패널가구 무응답 & 패널가구원 무응답'이고 4는 '재진입 등 기타'를 의미함.

2) 응답거부 2명은 제외하고 분석하였음.

자료: 한국보건사회연구원·국민건강보험공단. (2023). 2기 한국의료패널조사 2019년~2021년 연간데이터(2023년 학술대회용 버전)를 분석함.

다음 〈표 3-19〉는 4개 집단별 패널가구원의 응답형태에 따른 최종 학력 분포로 '패널가구 응답 & 패널가구원 무응답'의 비율은 초졸(2.0%)과 중졸(1.9%)이 높은 편에 속하였다. 이에 비해 '패널가구 무응답 & 패널가구원 무응답'의 비율은 무학(16.6%)과 대졸 이상(16.4%)이 높게 나타났다. 이는 〈표 3-17〉에서 보았듯이 중년의 부모와 자녀로 구성된 가족이 패널에서 이탈할 가능성이 가장 높다고 볼 수 있다.

〈표 3-19〉 4개 집단별 패널가구원의 응답형태별 최종 학력 분포

(단위: 명, %)

	1	2	3	4	합계	1	2	3	4	합계
무학	2,082	41	423	9	2,555	81.5	1.6	16.6	0.4	100.0
초졸	2,100	48	241	6	2,395	87.7	2.0	10.1	0.3	100.0
중졸	1,606	36	235	6	1,883	85.3	1.9	12.5	0.3	100.0
고졸	3,456	65	642	22	4,185	82.6	1.6	15.3	0.5	100.0
대졸 이상	3,047	51	612	13	3,723	81.8	1.4	16.4	0.3	100.0
전체	12,291	241	2,153	56	14,741	83.4	1.6	14.6	0.4	100.0

주: 1) 1은 '패널가구 응답 & 패널가구원 응답'이고 2는 '패널가구 응답 & 패널가구원 무응답'이고 3은 '패널가구 무응답 & 패널가구원 무응답'이고 4는 '재진입 등 기타'를 의미함.

2) 무학은 '미취학도 포함'하여 분석하였음.

자료: 한국보건사회연구원·국민건강보험공단. (2023). 2기 한국의료패널조사 2019년~2021년 연간데이터(2023년 학술대회용 버전)를 분석함.

마지막으로 <표 3-20>은 4개 집단별 패널가구원의 응답형태에 따른 경제활동 상태 분포로 ‘패널가구 응답 & 패널가구원 무응답’의 비율은 비경제활동인구가 2.8%로 가장 높았다. 한편, ‘패널가구 무응답 & 패널가구원 무응답’의 비율은 해당 없음이 19.5%로 가장 높게 나타났다. 다음으로 임금근로자가 15.3%, 비경제활동인구가 14.8%이며 고용주/자영업자가 11.5%의 순이었다.

<표 3-20> 4개 집단별 패널가구원의 응답형태별 경제활동 상태 분포

(단위: 명, %)

	1	2	3	4	합계	1	2	3	4	합계
임금근로자	3,874	62	714	22	4,672	82.9	1.3	15.3	0.5	100.0
고용주/ 자영업자	1,698	21	224	9	1,952	87.0	1.1	11.5	0.5	100.0
비경제 활동인구	4,224	142	760	14	5,140	82.2	2.8	14.8	0.3	100.0
기타	960	6	79	1	1,046	91.8	0.6	7.6	0.1	100.0
해당 없음	1,535	10	376	10	1,931	79.5	0.5	19.5	0.5	100.0
전체	12,291	241	2,153	56	14,741	83.4	1.6	14.6	0.4	100.0

주: 1) 1은 ‘패널가구 응답 & 패널가구원 응답’이고 2는 ‘패널가구 응답 & 패널가구원 무응답’이고 3은 ‘패널가구 무응답 & 패널가구원 무응답’이고 4는 ‘재진입 등 기타’를 의미함.

2) 기타는 무급가족종사자, 자활근로, 공공근로, 희망근로, 노인 일자리가 해당함.

3) 해당 없음은 15세 미만 응답자임.

자료: 한국보건사회연구원·국민건강보험공단. (2023). 2기 한국의료패널조사 2019년~2021년 연간데이터(2023년 학술대회용 버전)를 분석함.

2. 패널가구원의 응답형태에 따른 영향 분석

패널가구원의 응답형태에 따른 집단을 종속 변수로 하여 다항 로지스틱 회귀모형을 분석하였다. 이때 재진입을 제외한 3개 집단(연속응답, 2020년 이탈, 2021년 이탈)에 대해서만 살펴보았다. 재진입의 경우 관측치가 49명으로 작아 모형 적합이 되지 않을 수 있기 때문이다. 준거집단은 연속응답

으로 설정하였다. 사용한 설명 변수는 2019년에서의 가구원 특성(성별, 연령, 혼인상태, 최종 학력, 경제활동 상태), 2019년 연간데이터에서의 가구 특성(거주 지역, 주거형태, 가구원 수, 연간 가구 총소득), 그리고 2019년~2021년에 걸친 가구의 상태 변화(이사 횟수)이다. 분석 대상은 15세 이상 가구원으로 한정하였다. 설명 변수 중에서 가구원의 경제활동 상태는 15세 미만의 경우에는 해당하지 않기 때문에 분석에서 제외하였다.

〈표 3-21〉은 모형의 분석 결과로, 연속응답을 준거집단으로 하여 2020년 이탈에 영향을 주는 요인을 보면 가구원의 성별과 혼인상태, 연간 가구 총소득을 제외한 나머지 변수들은 통계적으로 유의한 결과를 나타내고 있다.

가구원의 연령이 한 단위 증가할수록 2020년 이탈 집단일 가능성이 연속응답 집단일 가능성보다 1.4% 감소한다고 볼 수 있다. 즉, 가구원이 젊을수록 이탈 가능성이 높은 편에 속하였다. 가구원의 최종 학력이 초졸(기준변수: 고졸)인 경우 2020년 이탈 집단일 가능성이 22.7% 감소하고, 가구원의 경제활동 상태가 기타(기준변수: 임금근로자)인 경우 25.8% 감소한다고 볼 수 있다.

거주 지역이 광역시(기준변수: 서울)인 경우 2020년 이탈 집단일 가능성이 연속응답 집단일 가능성보다 25% 감소하고, 기타 도지역인 경우 27.2% 감소한다고 볼 수 있다. 주거형태가 단독주택(기준변수: 아파트)인 경우 2020년 이탈 집단일 가능성이 연속응답 집단일 가능성보다 25.8% 감소하고, 다세대주택인 경우 43% 증가한다고 볼 수 있다. 가구원 수가 한 단위 증가할수록 2020년 이탈 집단일 가능성이 18.8% 증가하고, 이사 횟수는 한 단위 증가할수록 73.4% 감소한다고 볼 수 있다.

다음은 2021년 이탈에 영향을 주는 요인을 연속응답을 준거집단으로 하여 보면 가구원의 혼인상태를 제외한 나머지 변수들은 모두 통계적으로

유의한 결과를 나타내고 있다.

가구원의 성별이 여성(기준변수: 남성)인 경우 2021년 이탈 집단일 가능성이 연속응답 집단일 가능성보다 17.4% 감소하고, 가구원의 연령이 한 단위 증가할수록 1.5% 감소한다고 볼 수 있다. 즉, 가구원이 젊을수록 이탈 가능성이 높은 편에 속하였다. 가구원의 최종 학력이 무학(기준변수: 고졸)인 경우 2021년 이탈 집단일 가능성이 42.5% 증가하고, 초졸인 경우 23.4% 감소한다고 볼 수 있다. 그리고 가구원의 경제활동 상태가 비경제 활동인구(기준변수: 임금근로자)인 경우 21.6% 증가하고, 기타인 경우 36.3% 감소한다고 볼 수 있다.-

거주 지역이 광역시(기준변수: 서울)인 경우 2021년 이탈 집단일 가능성이 연속응답 집단일 가능성보다 28.6% 증가한다고 볼 수 있다. 주거 형태가 기타(기준변수: 아파트)인 경우 2021년 이탈 집단일 가능성이 연속응답 집단일 가능성보다 72.8% 증가한다고 볼 수 있다. 연간 가구 총소득이 저소득(기준변수: 중소득)인 경우 2021년 이탈 집단일 가능성이 34.6% 증가하고, 가구원 수가 한 단위 증가할수록 9.8% 증가한다고 볼 수 있다. 이사 횟수가 한 단위 증가할수록 2021년 이탈 집단일 가능성이 38% 감소한다고 볼 수 있다.

2020년과 2021년 패널가구원의 이탈에 영향을 주는 공통적인 요인은 가구원의 연령, 최종 학력, 경제활동 상태, 그리고 가구원 수, 이사 횟수였다. 가구원의 연령은 젊을수록, 최종 학력은 고졸이 초졸에 비해, 경제 활동 상태는 임금근로자가 기타에 비해, 가구원 수는 많을수록, 이사 횟수는 적을수록 패널가구원의 이탈 경향이 높은 편으로 나타났다. 이사 횟수는 패널가구의 표본이탈에 영향을 주는 요인에도 해당되었는데, 패널가구와 패널가구원에서 동일한 결과가 나왔다. 한편, 주거형태의 경우 2020년은 아파트가 단독주택에 비해, 다세대주택이 아파트에 비해 표본

이탈 경향이 높았고, 2021년은 기타가 아파트에 비해 표본이탈 경향이 높은 편에 속하였다. 2020년 결과는 패널가구의 표본이탈에 영향을 주는 요인과 동일하게 나타났다. 거주 지역의 경우 광역시(준거집단: 서울)가 통계적으로 유의한 영향을 주었는데 2020년 이탈 집단에서는 음(-)의 효과로, 2021년 이탈 집단에서는 양(+)의 효과로 상반되게 나타났다. 이외에도 2021년 이탈 집단은 남성이 여성에 비해, 연간 가구 총소득의 경우 저소득이 중소득에 비해 표본이탈에 영향을 주는 것으로 나타났다.

〈표 3-21〉 패널가구의 응답형태에 대한 다항 로지스틱 회귀모형 분석 결과

변수	2020년 이탈			2021년 이탈		
	coefficient	SE	exp(b)	coefficient	SE	exp(b)
상수항	-1.555 (***)	0.302	-	-2.072 (***)	0.342	-
가구원 성별_남성						
여성	0.101	0.078	1.106	-0.191 (**)	0.087	0.826
가구원 연령	-0.014 (***)	0.004	0.986	-0.016 (***)	0.004	0.985
가구원 혼인상태_유배우자						
미혼	-0.137	0.132	0.872	0.023	0.145	1.023
기타	-0.072	0.117	0.930	0.070	0.125	1.072
가구원 최종 학력_고졸						
무학	-0.198	0.200	0.820	0.354 (*)	0.189	1.425
초졸	-0.258 (*)	0.134	0.773	-0.267 (*)	0.151	0.766
중졸	-0.035	0.127	0.966	-0.009	0.143	0.991
대졸 이상	-0.023	0.087	0.977	-0.037	0.099	0.963
가구원 경제활동 상태_임금근로자						
고용주/자영업자	-0.178	0.116	0.837	-0.084	0.128	0.919
비경제활동인구	0.009	0.088	1.009	0.196 (**)	0.099	1.216
기타	-0.299 (*)	0.168	0.742	-0.451 (**)	0.207	0.637
거주 지역_서울						
광역시	-0.288 (**)	0.121	0.750	0.251 (*)	0.148	1.286

변수	2020년 이월			2021년 이월		
	coefficient	SE	exp(b)	coefficient	SE	exp(b)
경기	-0.063	0.137	0.939	0.143	0.170	1.154
기타 도지역	-0.317 (***)	0.120	0.728	-0.046	0.150	0.955
주거형태_아파트						
단독주택	-0.298 (***)	0.090	0.742	-0.080	0.099	0.923
연립주택	-0.235	0.203	0.790	0.259	0.192	1.296
다세대주택	0.358 (***)	0.113	1.430	0.091	0.141	1.095
기타	0.062	0.324	1.064	0.547 (*)	0.293	1.728
가구원 수	0.172 (***)	0.035	1.188	0.094 (**)	0.041	1.098
연간 가구 총소득_총소득						
저소득	0.166	0.103	1.181	0.297 (***)	0.114	1.346
고소득	-0.106	0.088	0.900	0.085	0.102	1.089
이사 횟수	-1.325 (***)	0.147	0.266	-0.478 (***)	0.113	0.620

주: 1) coefficient에서 () 안의 값은 p-value에 대한 유의성을 나타내며 (***)은 0.01, (***)은 0.05, (*)은 0.1에서 통계적으로 유의하다고 할 수 있음.

2) 혼인상태에서 '기타'는 별거/이혼, 사별/실종이 해당함.

3) 경제활동 상태에서 '기타'는 무급가족종사자, 자활/공공/희망/노인 일자리가 해당함.

4) 주거형태에서 '기타'는 영입용 건물 내 주택, 오피스텔, 호텔, 여관, 기숙사 등이 해당함.

5) 연간 가구 총소득에서 저소득은 0원~2,000만 원 미만, 중소득은 2,000만 원~5,000만 원 미만, 고소득은 5,000만 원 이상을 의미함.

자료: 한국보건사회연구원·국민건강보험공단. (2023). 2기 한국의료패널조사 2019년~2021년 연간데이터(2023년 학술대회용 버전)를 분석함.

제4절 소결

패널조사에서 표본이탈은 표본 대표성과 밀접한 관련이 있다. 표본 대표성을 만족하지 못한다면 모집단의 특성을 반영하지 않게 되기 때문이다. 그러므로 표본이탈이 발생하기 전에 원인을 파악하고 지속적으로 관리해야 할 것이다. 이에 따라 여러 번 조사가 진행된 2기 한국의료패널조사에 대해 표본이탈의 경향을 다각적으로 살펴보았다.

2019년 연간데이터의 패널 대상을 기준으로, 2019년~2021년에 대한 응답형태를 네 가지(연속응답, 2020년 이탈, 2021년 이탈, 재진입)로 구분하였다. 패널가구와 패널가구의 표본이탈 특성을 파악한 다음에, 응답형태에 따라 표본이탈에 영향을 주는 요인에 대해 각각 살펴보았다.

패널가구의 표본이탈 특성을 주요 변수를 중심으로 살펴본 결과, 거주 지역이 대구와 인천이면 2020년과 2021년에서 이탈 비율이 높았다. 세종의 경우 2020년에 이탈 비율이 가장 높게 나타났다. 가구원 수가 많아질수록 이탈 비율이 높게 나타났다. 아파트에 거주하면 이탈 비율이 높은 편이었으나, 단독주택은 낮은 편에 속하였다. 연간 가구 총소득은 2020년 이탈의 경우 소득이 높을수록 이탈 비율이 높은 반면에, 2021년 이탈의 경우 6,000만 원 이상이 가장 높았고, 그다음은 2,000만 원 미만으로 나타났다.

패널가구의 응답형태에 따른 3개 집단(연속응답, 2020년 이탈, 2021년 이탈)을 종속 변수로 하여 다항 로지스틱 회귀모형을 분석하였다. 사용한 설명 변수는 2019년 연간데이터에서 가구 특성(거주 지역, 주거형태, 가구원 수, 연간 가구 총소득), 2019년~2021년에 걸친 가구의 상태 변화(이사 횟수), 그리고 2019년에서의 가구주 특성(성별, 연령, 혼인상태, 최종 학력)이다. <표 3-22>는 이 분석 결과를 요약한 것으로, 연속응답을

준거집단으로 하여 이탈 집단에 영향을 주는 요인, 즉 설명 변수 중에서 통계적으로 유의한 결과를 나타낸 변수를 정리하였다. 2020년과 2021년 이탈 집단에 공통으로 영향을 주는 요인을 살펴보면, 주택형태가 아파트인 경우 단독주택에 비해 이탈 가능성이 높았고, 이사 횟수가 적을수록, 가구주가 젊을수록 이탈 집단에 있을 가능성이 높은 것으로 나타났다. 이 외에도 2021년 이탈 집단은 연간 가구 총소득이 저소득인 경우 중소득에 비해, 가구주의 최종 학력이 고졸인 경우 초졸에 비해 표본이탈에 영향을 주었다.

〈표 3-22〉 패널가구의 다항 로지스틱 회귀모형 분석 결과 요약

변수	2020년 이탈	2021년 이탈
거주 지역_서울		
주거형태_아파트	단독주택 (-) 다세대주택 (+)	단독주택 (-)
가구원 수		
연간 가구 총소득_중소득		저소득 (+)
이사 횟수	(-)	(-)
가구주 성별_남성		
가구주 연령	(-)	(-)
가구주 혼인상태_유배우자		
가구주 최종 학력_고졸		초졸 (-)

주: 1) 괄호 안의 내용은 각 변수에 해당하는 계수(coefficient)값의 부호를 의미함.

2) 연간 가구 총소득에서 저소득은 0원~2,000만 원 미만, 중소득은 2,000만 원~5,000만 원 미만, 고소득은 5,000만 원 이상을 의미함.

자료: 한국보건사회연구원·국민건강보험공단. (2023). 2기 한국의료패널조사 2019년~2021년 연간데이터(2023년 학술대회용 버전)를 분석함.

패널가구원의 표본이탈 특성을 주요 변수를 중심으로 살펴보았다. 성별의 경우 2020년 이탈에서는 남성과 여성의 비율이 비슷했으나, 2021년 이탈에서는 남성이 여성에 비해 높게 나타났다. 연령대의 경우 50세 이상~

79세 이하가 표본이탈 비율이 낮은 편이었고, 49세 이하가 높은 편으로 나타났다. 한편, 80세 이상의 경우에는 연속응답 비율은 낮고 이탈 비율이 높은 경향을 보였는데, 이는 노환, 사망 등으로 인한 표본이탈로 볼 수 있다. 미혼의 경우 다른 혼인상태에 비해 표본이탈의 경향이 높은 편이었다. 초졸과 중졸의 표본이탈 비율은 낮은 반면에, 고졸 이상일수록 비율이 높아지는 편이었다. 한편, 무학(미취학 포함)의 경우도 높게 나타나, 고학력 부모가 표본이탈하면서 미취학인 자녀도 함께 이탈한 것으로 추정할 수 있다. 경제활동 상태에 따라서는 15세 미만 응답자인 '해당 없음'과 비경제활동인구의 표본이탈 비율이 높은 데 비해서, 기타와 고용주/자영업자의 비율은 낮았다.

한편, 패널가구와 패널가구원의 응답형태를 조합하여 4개 집단의 주요 특성을 살펴보았다. 패널조사 시에 패널가구와 패널가구원이 모두 응답한 집단이 81.4%로 가장 높았다. 다음으로 패널가구와 패널가구원이 모두 무응답한 집단은 15.8%이고, 패널가구는 응답하였으나 패널 가구원은 무응답한 집단은 2.4%의 순이었으며 재진입 등 기타는 0.4%로 미미하였다. 이를 통해 패널가구의 응답 여부가 패널을 유지하는 데 중요함을 알 수 있다. 패널가구는 응답하였으나 패널가구원이 무응답한 집단의 인구사회학적 특성을 보면 남성이 여성에 비해, 80세 이상이 다른 연령대에 비해, 사별/실종이 다른 혼인상태에 비해, 초졸과 중졸인 경우가 다른 최종 학력에 비해, 비경제활동인구가 다른 경제활동 상태에 비해 조사에 참여하지 않을 가능성이 높은 편으로 나타났다. 패널가구 내에서 고령의 가구원은 패널에서 이탈할 가능성이 높은 편이라고 볼 수 있다. 패널가구와 패널가구원 모두 무응답한 집단의 인구사회학적 특성을 보면 성별에 따른 차이는 없는 편이었다. 0세 이상~19세 이하 및 40세 이상~49세 이하가 다른 연령대에 비해, 미혼이 다른 혼인상태에 비해, 미취학이 포함된

무학과 대졸 이상이 다른 최종 학력에 비해, 기타를 제외한 나머지 경제활동 상태가 조사에 참여하지 않을 가능성이 높은 편으로 나타났다. 미혼 이거나, 중년의 부모와 자녀로 구성된 가족이 패널에서 이탈할 가능성이 높다고 볼 수 있다.

패널가구원의 응답형태에 따른 3개 집단(연속응답, 2020년 이탈, 2021년 이탈)을 종속 변수로 하여 다항 로지스틱 회귀모형을 분석하였다. 사용한 설명 변수는 2019년 연간데이터에서의 가구원 특성(성별, 연령, 혼인상태, 최종 학력, 경제활동 상태), 2019년에서의 가구 특성(거주 지역, 주거형태, 가구원 수, 연간 가구 총소득), 그리고 2019년~2021년에 걸친 가구의 상태 변화(이사 횟수)이다. <표 3-23>은 이 분석 결과를 요약한 것으로, 연속응답을 준거집단으로 하여 이탈 집단에 영향을 주는 요인에 관한 것이다. 즉, 설명 변수 중에서 통계적으로 유의한 결과를 나타낸 변수를 정리하였다. 2020년과 2021년 패널가구원의 이탈에 영향을 주는 공통적인 요인은 가구원의 연령, 최종 학력, 경제활동 상태, 그리고 가구원 수, 이사 횟수였다. 가구원의 연령은 젊을수록, 최종 학력은 고졸이 초졸에 비해, 경제활동 상태는 임금근로자가 기타에 비해, 가구원 수는 많을수록, 이사 횟수는 적을수록 패널가구원의 이탈 경향이 높은 편으로 나타났다. 이사 횟수는 패널가구의 표본이탈에 영향을 주는 요인에도 해당되었는데, 패널가구와 패널가구원에서 동일한 결과가 나왔다. 한편, 주거형태의 경우 2020년은 아파트가 단독주택에 비해, 다세대주택이 아파트에 비해 표본이탈 경향이 높았고, 2021년은 기타가 아파트에 비해 표본이탈 경향이 높은 편에 속하였다. 2020년 결과는 패널가구의 표본이탈에 영향을 주는 요인과 동일하게 나타났다. 거주 지역의 경우 광역시(준거집단: 서울)가 통계적으로 유의한 영향을 주었는데 2020년 이탈 집단에서는 음(-)의 효과로, 2021년 이탈 집단에서는 양(+의 효과로 상반되게 나타났다.

이외에도 2021년 이탈 집단은 남성이 여성에 비해, 연간 가구 총소득의 경우 저소득이 중소득에 비해 표본이탈에 영향을 주는 것으로 나타났다.

〈표 3-23〉 패널가구원의 다항 로지스틱 회귀모형 분석 결과 요약

변수	2020년 이탈	2021년 이탈
가구원 성별_남성		여성 (-)
가구원 연령	(-)	(-)
가구원 혼인상태_유배우자		
가구원 최종 학력_고졸	초졸 (-)	무학 (+) 초졸 (-)
가구원 경제활동 상태_임금근로자	기타 (-)	비경제활동인구 (+) 기타 (-)
거주 지역_서울	광역시 (-) 기타 도지역 (-)	광역시 (+)
주거형태_아파트	단독주택 (-) 다세대주택 (+)	기타 (+)
가구원 수	(+)	(+)
연간 가구 총소득_중소득		저소득 (+)
이사 횟수	(-)	(-)

- 주: 1) 괄호 안의 내용은 각 변수에 해당하는 계수(coefficient)값의 부호를 의미함.
 2) 가구원 경제활동 상태에서 기타는 무급가족종사자, 자활근로, 공공근로, 희망근로, 노인 일자리가 해당함.
 3) 주거형태에서 기타는 영업용 건물 내 주택(상가, 공장 등), 오피스텔, 호텔, 여관, 기숙사 등을 의미함.
 4) 연간 가구 총소득에서 저소득은 0원~2,000만 원 미만, 중소득은 2,000만 원~5,000만 원 미만, 고소득은 5,000만 원 이상을 의미함.
 자료: 한국보건사회연구원·국민건강보험공단. (2023). 2기 한국의료패널조사 2019년~2021년 연간데이터(2023년 학술대회용 버전)를 분석함.



제4장

2기 한국의료패널조사 추가표본 추출 방안

제1절 모집단 및 층화변수 현황 분석

제2절 2기 한국의료패널조사의 주요 변수 분석

제3절 추가표본 추출을 위한 표본설계

제4절 소결

제 4 장

2기 한국의료패널조사 추가표본 추출 방안

이 장에서는 2기 한국의료패널조사의 추가표본 추출 방안을 마련하고자 한다. 먼저, 모집단 분석을 실시하고 2기 한국의료패널조사의 2021년 연간데이터의 주요 변수에 대해 분석한다. 적정 표본 규모를 검토하고 층화변수, 배분방법 등을 고려하여 표본설계 방안을 마련하고자 한다. 이를 기반으로 새로운 표본을 추가함으로써 표본의 마모를 보완하고자 한다. 마지막으로 2기 한국의료패널조사의 시도 단위 통계 산출을 위해서 제안한 표본설계 내용이 반영된 가중치 산출 방안을 제시하고자 한다.

제1절 모집단 및 층화변수 현황 분석

1. 목표모집단 및 조사모집단 현황 분석

통계조사에서는 이론적으로 통계를 산출하고자 하는 목표모집단과 실제 조사를 위해서 활용할 수 있는 조사모집단을 각각 설정한다. 여기서 중요한 점은 목표모집단과 조사모집단 간에 큰 차이가 없이 일치해야 한다는 것이다. 따라서 목표모집단과 조사모집단 간에 차이가 있는지 살펴보았다. 전체적으로 모집단 간에 차이가 없다면 조사모집단을 적절하게 설정한 것이며, 차이가 크게 발생한다면 차이의 원인을 찾을 필요가 있다.

한국의료패널조사의 목표모집단은 우리나라에 거주하는 전체 가구이고, 조사모집단은 인구주택총조사의 섬조사구, 시설조사구 등을 제외한 조사구에 거주하는 가구이다. 인구주택총조사에서는 2015년 이후 일반조사구,

아파트조사구, 섬조사구 3개로 구분하여 조사하고 있다. 따라서 전체 조사구 대비 조사구 특성별 비중을 살펴보고 조사구 특성에 대한 비교 분석을 통해 목표모집단과 조사모집단이 일치하는지 확인하였다.

먼저, 인구주택총조사의 조사구 특성을 분석하기 전에 인구주택총조사의 표본추출틀에 대해 알아볼 필요가 있다. 인구주택총조사는 기본적으로 인구, 가구, 주택에 대해 각각의 데이터로 구분하고 있다. 당연히 인구 데이터의 단위는 인구로, 가구데이터는 가구로, 주택데이터는 주택으로 구성되어 있다. 다만, 인구주택총조사는 조사 기반이 아닌 주민등록, 가족관계증명서 등 다양한 행정데이터를 가공하여 만든 데이터이다. 그래서 조사구에 대한 정보가 없으므로 통계조사의 조사구 추출을 위해 사용할 수 없다. 이러한 단점을 보완하기 위해 통계청에서는 인구주택총조사에 대한 표본추출틀을 별도로 제공하고 있다. 인구주택총조사를 모집단으로 하는 통계는 이 표본추출틀을 활용하여 조사구를 추출하는 방식으로 표본을 설계하고 있다.

〈표 4-1〉을 보면 2021년 인구주택총조사의 표본추출틀에서 전체 조사구는 약 39만 개이고 가구는 약 2,193만 가구가 있다. 이 중 일반조사구는 약 19만 개 조사구이고, 약 1,072만 가구가 있다. 아파트조사구는 약 20만 개 조사구이고, 약 1,128만 가구가 있다. 섬조사구는 906개 조사구이고 37,261가구로 나타나, 일반조사구와 아파트조사구에 비해 매우 적은 비중을 가졌다. 전체 조사구 대비 0.2%로 나타났다.

〈표 4-1〉 2021년 인구주택총조사 표본추출틀의 조사구 특성별 비중

(단위: 개, 가구, %)

구분	계	일반조사구		섬조사구		아파트조사구	
		개	구성비	개	구성비	개	구성비
조사구 수	389,510	189,132	48.6	906	0.2	199,472	51.2
가구수	21,932,739	10,619,436	48.4	37,261	0.2	11,276,042	51.4

자료: 통계청. (2022). 2021년 인구총조사.

통계청에서는 조사구를 설정할 때 1조사구 내 평균 60가구가 배치 되도록 조정하고 있다. 표본추출틀에서 조사구 내 가구수를 분석해보면 〈표 4-2〉와 같다. 전체 평균 56.3가구가 있으며, 동부는 평균 57.3가구이고 읍면부는 평균 52.5가구가 있어 비교적 동부가 읍면부에 비해서 조사구 내 가구가 더 많았다.

조사구 특성별로 살펴보면 일반조사구는 평균 56.2가구, 아파트조사구는 평균 56.5가구로, 조사구 내 가구수가 거의 비슷한 편이었다. 한편, 섬조사구는 평균 41.1가구로 다른 조사구 특성과 비교했을 때 가구수가 적었다.

〈표 4-2〉 조사구 내 평균 가구수 및 표준오차

(단위: 가구)

동/ 읍면부	계		일반조사구		섬조사구		아파트조사구	
	평균	표준 오차	평균	표준 오차	평균	표준 오차	평균	표준 오차
계	56.3	15.7	56.2	14.5	41.1	21.7	56.5	16.8
동부	57.3	15.6	57.5	14.3	35.8	22.5	57.1	16.6
읍면부	52.5	15.7	52.5	14.3	41.3	21.7	52.7	17.7

자료: 통계청. (2022). 2021년 인구총조사.

〈표 4-3〉은 17개 시도 및 조사구 특성에 따른 가구 현황으로, 섬조사구는 극단적인 특성을 지녔다. 섬조사구는 기본적으로 섬지역의 조사구이므로 섬지역이 있는 시도에만 존재함을 알 수 있다. 17개 시도 중에서 섬조사구가 있는 시도는 부산광역시, 인천광역시, 경기도, 충청남도, 전라북도, 전라남도, 경상북도, 경상남도, 제주특별자치도로, 총 9개 시도에만 분포되어 있다. 이 중 전라남도에 22,762가구로 가장 많이 분포되어 있으며, 다음으로 인천광역시 6,266가구, 경상남도 3,997가구 순으로 나타났다. 특이한 점은 경상북도에는 8가구만 있는 것을 확인할 수 있다.

전체 가구 대비 섬조사구 내 가구 비중은 0.2%로 매우 작으며 섬조사구 내 가구가 가장 많은 전라남도의 경우 전체 가구 대비 2.9%를 차지하고 있다. 전라남도를 제외한 섬조사구가 있는 8개 시도에서 섬조사구 내 가구의 비중은 1%보다 작아 섬조사구의 비중이 매우 낮음을 알 수 있다.

이러한 점을 종합적으로 고려하면 섬조사구에 분포되어 있는 가구는 전체 가구 대비 비중이 매우 낮으며, 특히 섬조사구가 없는 8개 시도에서는 목표모집단과 조사모집단이 100% 일치함을 알 수 있다. 섬조사구가 있는 9개 시도에서도 섬조사구 내 가구가 가장 많은 전라남도도 2.9%에 불과하므로 섬조사구를 제외하더라도 목표모집단과 조사모집단 간에 큰 차이가 없이 일치함을 알 수 있다.

이러한 분석을 통해 목표모집단과 조사모집단을 비교한 결과, 한국 의료패널조사는 목표모집단과 조사모집단이 큰 차이 없이 일치한다고 볼 수 있다.

〈표 4-3〉 17개 시도 및 조사구 특성에 따른 가구 현황

(단위: 가구, %)

지역	계	일반조사구		섬조사구		아파트조사구	
		구성비	구성비	구성비	구성비		
전국	21,932,739	10,619,436	48.4	37,261	0.2	11,276,042	51.4
서울	4,178,144	2,393,323	57.3	0	0.0	1,784,821	42.7
부산	1,442,240	609,157	42.2	2	0.0	833,081	57.8
대구	1,011,013	428,498	42.4	0	0.0	582,515	57.6
인천	1,212,267	538,283	44.4	6,266	0.5	667,718	55.1
광주	624,069	207,448	33.2	0	0.0	416,621	66.8
대전	646,492	290,562	44.9	0	0.0	355,930	55.1
울산	458,242	182,259	39.8	0	0.0	275,983	60.2
세종	146,883	33,915	23.1	0	0.0	112,968	76.9
경기	5,465,007	2,333,804	42.7	94	0.0	3,131,109	57.3
강원	679,642	357,149	52.5	0	0.0	322,493	47.5
충북	710,084	357,766	50.4	0	0.0	352,318	49.6
충남	941,894	501,493	53.2	1,108	0.1	439,293	46.6
전북	782,986	401,785	51.3	1,225	0.2	379,976	48.5
전남	785,527	443,056	56.4	22,762	2.9	319,709	40.7
경북	1,171,641	678,197	57.9	8	0.0	493,436	42.1
경남	1,400,131	657,989	47.0	3,997	0.3	738,145	52.7
제주	276,477	204,752	74.1	1,799	0.7	69,926	25.3

자료: 통계청. (2022). 2021년 인구총조사.

2. 총화변수 현황 분석

한국의료패널조사의 총화변수는 17개 광역시·도와 동읍면부, 두 개의 변수를 활용한다. 한국의료패널조사의 작성 목적인 우리나라 보건의료의 이용과 비용에 대한 정보를 조사하기 위해서는 지역적 특성이 큰 영향을 주고 있다.

이러한 지역적 특성에 따른 총화변수를 설정하기 위해서 먼저 시도 변수를 고려할 수 있다. 지역에 따라 인구구조, 의료인프라 등에 큰 차이가 발생하므로 시도를 먼저 고려하고, 도시지역과 농어촌 지역을 구분할 수 있는 동읍면부 변수를 추가로 고려하고자 한다.

시도 변수에 따른 가구 분포를 살펴보면, 경기도의 가구수는 546만 가구로 전체 가구 대비 25%를 차지하고 있으며 서울특별시의 가구수는 418만 가구로 전체 가구 대비 19.1%를 차지하고 있다. 가구가 가장 많은 경기도와 서울특별시 두 지역에서 전체 가구 대비 44.1%를 차지하고 있어 전체 가구의 절반에 가까운 가구가 서울, 경기도에 분포되어 있다. 이와 반대로 가장 적은 시도인 세종특별자치시는 15만 가구로 전체 가구 대비 0.7%, 제주특별자치도는 27만 가구로 전체 가구 대비 1.3%를 차지하고 있어 타 시도 대비 매우 적은 가구가 분포되어 있다.

동읍면부를 살펴보면 동부에는 1,772만 가구가 있어 전체 가구 대비 80.9%를 차지하고 있으며 읍면부는 418만 가구로 전체 가구 대비 19.1%를 차지하고 있어 동부의 비중이 매우 높은 것을 알 수 있다. 또한 3개 시인 서울특별시, 광주광역시, 대전광역시의 경우에는 읍면부가 없이 동부만 있다.

두 총화변수인 시도 변수와 동읍면부 변수에 따른 가구 분포를 분석해 본 결과 총화변수 간 가구수 차이가 매우 심하므로 표본설계 시 가구가

적은 층에서 과소표집되지 않도록 적절한 배분방법을 활용할 필요가 있다.

〈표 4-4〉 17개 시도 및 동부·읍면부 특성에 따른 가구 현황

(단위: 가구, %)

지역	계		동부		읍면부	
	계	구성비	계	구성비	계	구성비
전국	21,895,478	100.0	17,719,004	100.0	4,176,474	100.0
서울	4,178,144	19.1	4,178,144	23.6	0	0.0
부산	1,442,238	6.6	1,372,541	7.7	69,697	1.7
대구	1,011,013	4.6	906,181	5.1	104,832	2.5
인천	1,206,001	5.5	1,174,189	6.6	31,812	0.8
광주	624,069	2.9	624,069	3.5	0	0.0
대전	646,492	3.0	646,492	3.6	0	0.0
울산	458,242	2.1	366,538	2.1	91,704	2.2
세종	146,883	0.7	100,746	0.6	46,137	1.1
경기	5,464,913	25.0	4,530,302	25.6	934,611	22.4
강원	679,642	3.1	407,629	2.3	272,013	6.5
충북	710,084	3.2	405,658	2.3	304,426	7.3
충남	940,786	4.3	434,222	2.5	506,564	12.1
전북	781,761	3.6	551,674	3.1	230,087	5.5
전남	762,765	3.5	330,185	1.9	432,580	10.4
경북	1,171,633	5.4	607,402	3.4	564,231	13.5
경남	1,396,134	6.4	885,800	5.0	510,334	12.2
제주	274,678	1.3	197,232	1.1	77,446	1.9

자료: 통계청. (2022). 2021년 인구총조사.

제2절 2기 한국의료패널조사의 주요 변수 분석

1. 2기 한국의료패널조사 구축에 대한 표본 현황

2기 한국의료패널조사는 2016년 인구주택총조사를 모집단으로 하여 17개 시도와 동부·읍면부를 층화변수로 사용하여 층화계통추출방법을 활용하였다. 층화변수 외에 내재적 층화변수로 조사구 특성, 점유형태, 혼인상태, 거처 종류 총 4개 변수를 사용하였다. 배분방법으로는 시도 변수에 대해서는 제곱근 비례배분, 동읍면부 변수에 대해서는 비례배분 방법을 활용하였다.

전체 표본 크기는 8,500가구이며 경기도 1,104가구, 서울특별시 864가구, 경상남도 648가구 순으로 많았으며 제주특별자치도가 160가구로 가장 적고, 세종특별자치시가 228가구로 그 다음으로 적었다(〈표 4-5〉 참조). 시도 변수에 대해서는 제곱근 비례배분방법을 활용하여 전체 가구 대비 비중이 높은 경기도, 서울특별시에는 비교적 적은 가구가 배분되었으며, 가구 비중이 낮은 제주특별자치도, 세종특별자치시에는 비교적 많은 가구가 배분되었음을 알 수 있다.

동부의 경우 6,494가구가 배분되었으며 읍면부에는 2,006가구가 배분되어 비례배분방법을 활용하여 전체 가구 비중에 따라 배분되었음을 알 수 있다.

층 내 가구수가 가장 적은 층인 부산광역시 읍면부와 인천광역시 읍면부에는 24가구가 배분되었으며, 층 내 가구수가 가장 큰 층은 경기도 동부이며 912가구가 배분되었다.

〈표 4-5〉 17개 시도 및 동부읍면부의 가구 현황

(단위: 가구)

지역	계	동부	읍면부
전국	8,500	6,494	2,006
서울	864	864	0
부산	516	492	24
대구	336	300	36
인천	588	564	24
광주	456	456	0
대전	384	384	0
울산	504	408	96
세종	228	120	108
경기	1,104	912	192
강원	276	168	108
충북	612	360	252
충남	312	132	180
전북	552	384	168
전남	516	216	300
경북	444	228	216
경남	648	408	240
제주	160	98	62

자료: 2기 한국의료패널조사 표본설계 보고서를 인용하여 작성함.

2. 만성질환 수 및 연간 가구 총소득에 대한 분석

한국의료패널조사의 주요 변수로는 만성질환 수와 연간 가구 총소득
이므로 각 변수별로 분석을 하고자 한다.

가. 만성질환 수

17개 시도의 만성질환 수의 평균과 표준오차를 살펴보면, 먼저 조사의 차수가 진행될수록 만성질환 수가 증가하고 있는 것을 알 수 있다(표 4-6) 참조). 또한 17개 시도의 경우 만성질환 수의 가장 큰 특징은 광역시와 경기도 지역과 도 지역 간에 차이가 있다는 점이다. 비교적 도시화의 비중이 높은 광역시와 경기도 지역에서는 만성질환 수가 낮고, 상대적으로 농어촌 비중이 높은 도 지역에서는 만성질환 수가 많다는 특징을 보였다.

만성질환 수의 표준오차는 제주특별자치도와 세종특별자치시의 표준오차가 타 지역에 비해서 높은 것으로 나타났다. 이는 두 지역의 가구수가 적으므로 상대적으로 적은 표본이 배분된 결과인 것으로 보였다.

〈표 4-6〉 17개 시도의 만성질환 수 평균 및 표준오차

(단위: 개)

지역	2019년 연간데이터		2020년 연간데이터		2021년 연간데이터	
	평균	표준오차	평균	표준오차	평균	표준오차
전국	0.73	0.01	0.77	0.01	0.83	0.01
서울	0.66	0.03	0.69	0.03	0.78	0.03
부산	0.78	0.04	0.84	0.04	0.94	0.04
대구	0.72	0.05	0.69	0.05	0.75	0.06
인천	0.65	0.04	0.69	0.04	0.76	0.05
광주	0.73	0.04	0.76	0.04	0.79	0.05
대전	0.79	0.05	0.82	0.05	0.88	0.06
울산	0.64	0.04	0.69	0.05	0.80	0.05
세종	0.76	0.06	0.85	0.06	0.80	0.08
경기	0.72	0.03	0.75	0.03	0.79	0.03
강원	0.92	0.05	0.92	0.05	1.01	0.05

지역	2019년 연간데이터		2020년 연간데이터		2021년 연간데이터	
	평균	표준오차	평균	표준오차	평균	표준오차
충북	0.81	0.04	0.83	0.04	0.89	0.04
충남	0.80	0.05	0.85	0.06	0.87	0.06
전북	0.76	0.04	0.81	0.04	0.86	0.04
전남	0.89	0.04	0.93	0.05	1.00	0.05
경북	0.80	0.04	0.87	0.04	0.87	0.05
경남	0.74	0.03	0.80	0.04	0.87	0.04
제주	0.75	0.07	0.72	0.07	0.70	0.08

자료: 한국보건사회연구원·국민건강보험공단. (2023). 2기 한국의료패널조사 2019년~2021년 연간데이터(2023년 학술대회용 버전)를 분석함.

동부·읍면부의 만성질환 수를 보면 동부와 읍면부 모두 조사 차수가 진행될수록 만성질환 수가 많아지는 공통적인 특성을 보인다(〈표 4-7〉 참조). 동부와 읍면부 간에는 큰 차이가 있다고 볼 수 있다. 2021년 연간 데이터 기준으로 보면 동부는 0.79이고 읍면부는 1.02로 나타나, 읍면부가 동부보다 훨씬 큰 값을 가졌다.

〈표 4-7〉 동부·읍면부의 만성질환 수 평균 및 표준오차

(단위: 개)

지역	2019년 연간데이터		2020년 연간데이터		2021년 연간데이터	
	평균	표준오차	평균	표준오차	평균	표준오차
전국	0.73	0.01	0.77	0.01	0.83	0.01
동부	0.69	0.01	0.73	0.01	0.79	0.01
읍면부	0.91	0.02	0.97	0.03	1.02	0.03

자료: 한국보건사회연구원·국민건강보험공단. (2023). 2기 한국의료패널조사 2019년~2021년 연간데이터(2023년 학술대회용 버전)를 분석함.

설계 층에 대해 만성질환 수의 상대표준오차를 확인해보고자 한다. 2021년 연간데이터를 기준으로 보면 전국의 상대표준오차는 1.5%로 매우 낮은 수준을 유지하고 있다(〈표 4-8〉 참조). 17개 시도에서 세종특별자치시와 제주특별자치도만 10%를 초과하고 있으나, 나머지 지역은 25% 이하 수준을 유지하고 있어 높은 수준의 신뢰도를 보여주고 있다.

17개 시도 및 동부·읍면부의 만성질환 수의 상대표준오차를 살펴보면 전반적으로 25% 이하로 나타나 안정적으로 유지되고 있다. 그러나 부산광역시 읍면부, 대구광역시 읍면부와 울산광역시 읍면부는 25%를 초과하였다. 특히 부산광역시 읍면부와 울산광역시 읍면부는 100%를 초과하여 매우 높게 나타났다.

〈표 4-8〉 17개 시도 및 동부·읍면부의 만성질환 수 상대표준오차

(단위: %)

지역	2019년 연간데이터			2020년 연간데이터			2021년 연간데이터		
	계	동부	읍면부	계	동부	읍면부	계	동부	읍면부
전국	1.49	1.75	2.70	1.53	1.80	2.69	1.50	1.76	2.65
서울	4.48	4.48	-	4.62	4.62	-	4.36	4.36	-
부산	4.83	4.82	130.04	4.79	4.77	126.21	4.52	4.50	131.59
대구	6.49	6.58	0.00	7.15	7.26	0.00	7.40	7.53	36.12
인천	6.06	6.06	0.00	6.23	6.23	-	6.42	6.42	-
광주	5.85	5.85	-	5.74	5.74	-	6.06	6.06	-
대전	6.83	6.83	-	6.57	6.57	-	6.71	6.71	-
울산	6.21	6.21	0.00	6.58	6.60	88.49	5.62	5.62	104.08
세종	7.70	14.69	8.57	7.34	16.49	6.65	10.31	20.56	6.63
경기	3.69	4.17	7.83	3.80	4.34	7.70	3.74	4.27	7.52
강원	5.76	7.70	8.58	5.83	8.11	8.09	5.24	7.18	7.54
충북	4.79	7.19	5.64	4.80	7.76	4.42	4.67	7.12	4.85

지역	2019년 연간데이터			2020년 연간데이터			2021년 연간데이터		
	계	동부	읍면부	계	동부	읍면부	계	동부	읍면부
충남	6.78	12.55	7.64	6.78	12.26	7.89	6.72	12.63	7.14
전북	4.86	6.16	7.56	4.72	5.78	8.10	4.71	6.03	6.87
전남	4.99	7.33	6.75	5.80	8.69	7.70	5.35	8.55	6.86
경북	5.49	7.95	7.57	5.12	7.45	6.96	5.62	7.97	7.94
경남	4.57	6.60	5.56	4.45	6.44	5.32	4.39	6.11	5.80
제주	9.66	13.14	13.41	9.66	14.36	10.69	11.25	15.58	13.50

자료: 한국보건사회연구원·국민건강보험공단. (2023). 2기 한국의료패널조사 2019년~2021년 연간데이터(2023년 학술대회용 버전)를 분석함.

나. 연간 가구 총소득

17개 시도의 연간 가구 총소득의 평균과 표준오차는 지역마다 약간 차이가 있지만, 대체로 시간이 지남에 따라 연간 가구 총소득이 증가하고 있는 것을 알 수 있다(〈표 4-9〉 참조). 또한 17개 시도의 연간 가구 총소득의 평균은 서울이 가장 높았고, 경북이 가장 낮았고, 광역시와 경기도가 다른 도에 비해 약간 높은 편이었다. 표준오차는 제주특별자치도를 제외한 도지역이 낮은 편에 속했다.

동부·읍면부의 연간 가구 총소득을 보면 동부와 읍면부 모두 조사 차수가 진행될수록 커지는 공통적인 특성을 보이고 있다(〈표 4-10〉 참조). 동부와 읍면부 간에는 큰 차이가 있다고 볼 수 있다. 2021년 연간데이터 기준으로 보면 동부는 5,230만 원 정도이고 읍면부는 4,056만 원 정도로 나타났다. 표준오차는 읍면부가 동부에 비해 큰 값을 가졌다.

92 패널조사 표본이탈 경향 분석과 추가표본 추출에 관한 연구- 한국의료패널조사를 중심으로

〈표 4-9〉 17개 시도의 연간 가구 총소득 평균 및 표준오차

(단위: 만 원)

지역	2019년 연간데이터		2020년 연간데이터		2021년 연간데이터	
	평균	표준오차	평균	표준오차	평균	표준오차
전국	4,038.07	63.64	4,467.21	97.02	5,018.16	79.05
서울	4,602.59	242.35	5,547.08	434.16	5,907.64	303.00
부산	3,876.19	179.46	4,232.49	209.07	4,487.52	211.88
대구	2,967.89	216.40	3,992.46	365.96	4,320.68	240.99
인천	3,846.21	167.67	4,277.20	196.13	4,929.54	219.07
광주	4,050.39	222.99	4,176.60	234.29	5,282.02	362.57
대전	3,983.11	245.60	4,058.34	238.66	4,202.54	238.62
울산	4,080.29	236.85	4,503.81	222.72	5,406.14	282.49
세종	5,015.91	353.01	4,502.05	336.61	5,170.54	433.24
경기	4,464.03	130.80	4,705.16	142.20	5,506.60	152.81
강원	3,776.19	202.72	3,847.59	214.88	4,496.34	246.22
충북	3,787.65	171.90	4,123.02	184.15	4,664.01	216.22
충남	3,719.34	216.98	3,864.43	200.53	4,650.99	355.53
전북	3,418.61	157.20	3,786.48	169.43	4,206.94	195.65
전남	3,262.24	144.73	3,849.01	187.59	4,197.58	214.64
경북	3,281.25	173.55	3,427.85	172.44	3,949.81	192.61
경남	3,724.54	143.03	4,097.25	163.41	4,359.38	158.07
제주	4,115.06	336.25	3,526.61	259.21	4,618.93	377.97

자료: 한국보건사회연구원·국민건강보험공단. (2023). 2기 한국의료패널조사 2019년~2021년 연간데이터(2023년 학술대회용 버전)를 분석함.

〈표 4-10〉 동부·읍면부의 연간 가구 총소득 평균 및 표준오차

(단위: 만 원)

지역	2019년 연간데이터		2020년 연간데이터		2021년 연간데이터	
	평균	표준오차	평균	표준오차	평균	표준오차
계	4,038.07	63.64	4,467.21	97.02	5,018.16	79.05
동부	4,190.04	74.89	4,646.02	116.43	5,229.82	93.16
읍면부	3,338.40	89.50	3,654.93	92.96	4,056.49	106.35

자료: 한국보건사회연구원·국민건강보험공단. (2023). 2기 한국의료패널조사 2019년~2021년 연간데이터(2023년 학술대회용 버전)를 분석함.

〈표 4-11〉은 설계 층의 연간 가구 총소득의 상대표준오차를 나타낸 것이다. 만성질환 수와 유사한 결과를 보이고 있으나, 25%를 넘는 층은 대구광역시 읍면부만 해당하였다. 이를 통해서 만성질환 수의 상대표준 오차보다는 안정적이라고 볼 수 있다.

〈표 4-11〉 17개 시도 및 동부읍면부의 연간 가구 총소득 상대표준오차

(단위: %)

지역	2019년 연간데이터			2020년 연간데이터			2021년 연간데이터		
	계	동부	읍면부	계	동부	읍면부	계	동부	읍면부
전국	1.58	1.79	2.68	2.17	2.51	2.54	1.58	1.78	2.62
서울	5.27	5.27	-	7.83	7.83	-	5.13	5.13	-
부산	4.63	4.62	9.60	4.94	4.96	15.74	4.72	4.73	7.94
대구	7.29	7.35	18.34	9.17	9.21	13.00	5.58	5.63	25.48
인천	4.36	4.36	0.00	4.59	4.59	-	4.44	4.44	-
광주	5.51	5.51	-	5.61	5.61	-	6.86	6.86	-
대전	6.17	6.17	-	5.88	5.88	-	5.68	5.68	-
울산	5.81	5.85	36.82	4.95	4.97	49.04	5.23	5.27	8.38
세종	7.04	8.69	8.29	7.48	9.96	9.54	8.38	12.21	9.82
경기	2.93	3.29	6.05	3.02	3.48	5.38	2.78	3.07	6.27
강원	5.37	7.22	7.94	5.59	7.66	7.94	5.48	7.68	7.50
충북	4.54	5.61	6.87	4.47	5.67	6.24	4.64	5.69	6.94
충남	5.83	8.00	8.19	5.19	6.90	7.06	7.64	10.90	6.91
전북	4.60	5.53	5.45	4.48	5.31	6.03	4.65	5.55	6.76
전남	4.44	6.57	5.91	4.87	7.33	6.52	5.11	8.36	6.44
경북	5.29	7.44	7.49	5.03	7.23	6.62	4.88	7.20	6.39
경남	3.84	4.71	6.18	3.99	4.91	5.95	3.63	4.54	5.89
제주	8.17	8.92	14.99	7.35	8.30	14.31	8.18	8.94	17.95

자료: 한국보건사회연구원·국민건강보험공단. (2023). 2기 한국의료패널조사 2019년~2021년 연간데이터(2023년 학술대회용 버전)를 분석함.

제3절 추가표본 추출을 위한 표본설계

1. 표본설계의 목적

한국의료패널조사는 패널조사로 표본을 지속적으로 관리하며 유지하는 것이 중요하다. 따라서 2023년 표본설계에서는 2기 한국의료패널조사에서 유지되고 있는 표본을 그대로 유지하면서 5년간의 조사로 인해 마모된 표본의 현황을 파악하고 마모된 표본을 보완하는 것을 목표로 한다. 또한, 최신의 모집단을 활용하여 통합패널의 가중치를 새롭게 부여함으로써 추정의 정확도를 높이고자 한다.

2. 모집단 층화

추가표본 추출을 위한 표본설계는 2기 한국의료패널조사의 마모된 표본을 추가로 추출하는 데 목적이 있으므로 2기 조사 구축 당시와 동일한 층화 기준을 적용하고자 한다. 2기 한국의료패널조사에서는 먼저 17개 시도로 1차 층화한 다음에 동부·읍면부로 2차 층화하였으므로 동일한 기준을 적용하고자 한다. 모집단 층화에 따른 가구 현황은 앞의 <표 4-5>에서 제시하였다.

3. 표본 규모

일반적으로 통계조사의 표본의 크기는 통계조사의 정확도와 통계조사에 대한 예산과 인력 등을 종합적으로 판단하여 결정된다. 여기에서는 통계조사의 정확도에 초점을 맞춰서 표본의 크기를 결정하고자 하며, 2기 한국의료패널조사의 표본 크기(8,500개)의 표본 규모를 크게 변화하지

않고 적정 수준의 표본 크기를 결정하고자 한다.

표본 크기를 결정하기 위해 2기 한국의료패널조사의 상대표준오차를 활용하여 표본 크기에 따른 예상 상대표준오차를 계산했는데, 그 공식과 계산 결과는 아래와 같다.

$$N_1 CV_1^2 = N_2 CV_2^2$$

N_1 : 2기 한국의료패널의 2022년 연간데이터 표본규모(완료 수 기준)
 CV_1^2 : 2기 한국의료패널의 2022년 연간데이터 조사 결과 상대표준오차
 N_2 : 2기 한국의료패널의 통합패널 조사 산출된 표본규모(완료 수 기준)
 CV_2^2 : 2기 한국의료패널의 통합패널 조사 목표 상대표준오차

〈표 4-12〉 주요 변수에 대한 예상 상대표준오차

(단위: 가구, %)

항목	2021년 연간데이터		통합패널(예상)	
	가구	상대표준오차	표본 규모	예상 상대표준오차
연간 가구 총소득	5,878	1.575	9,000	1.27
			8,400	1.32
			8,000	1.35
			7,400	1.40
만성질환 수	5,878	1.498	9,000	1.27
			8,400	1.32
			8,000	1.35
			7,400	1.40

자료: 한국보건사회연구원·국민건강보험공단. (2023). 2기 한국의료패널조사 2019년~2021년 연간데이터(2023년 학술대회용 버전)를 분석함.

표본의 크기를 8,400개로 할 경우 주요 변수에 대한 예상 상대표준 오차가 1.32%로 매우 안정적이면서 높은 신뢰의 통계를 작성할 수 있을 것으로 예상된다.

그러므로 인구주택총조사의 추출 단위인 조사구를 기준으로 조사구 1개당 평균 10가구를 추출하는 것을 목표로 하여 총 840개의 조사구를 표본 크기로 정하고자 한다.

4. 표본 배분

표본이론에서 층화 추출 설계의 층별 표본 배분방법 중 전체 모수 추정을 목적으로 하는 배분방법으로는 조사비용, 층별 모분산과 부모집단 크기 등을 고려한 최적배분, 네이만배분, 비례배분이 적절하다. 이에 추가하여 층별 모수 추정 및 분석을 하기 위해서는 층별 분석을 위해 필요한 표본 크기 확보가 중요한데, 이들 배분방법은 층별 분석을 위해서는 적절한 배분을 하지 못하는 경우가 발생한다. 이를 위해 층별 분석이 가능한 표본 배분을 고려해야 한다. 일반적으로 층별 분석이 가능한 표본 배분방법은 보조정보를 이용한 역배분(power allocation)을 이용한다. 역배분을 위한 보조 정보로는 층별 모집단 크기, 층별 모분산, 층별 대리변수 등을 널리 사용한다. 조사비용 및 층별 모분산의 차이가 크지 않으면 보조 정보로 부모집단 크기를 이용하는 제곱근 비례배분방법을 널리 사용하고 있다. 앞에서 주요 항목으로 검토한 연간 가구 총소득과 만성질환 수의 17개 시도에 대한 표준오차는 시도별로 약간 차이가 있었지만 큰 차이를 보이지 않아 제곱근 비례배분방법을 추가적으로 고려하고자 한다.

2기 한국의료패널조사에서는 1차 층화변수인 17개 시도에 대해서는 제곱근 비례배분, 2차 층화변수인 동부·읍면부에는 비례배분을 적용하였다.

여기서는 현행 표본 체계를 유지하면서 시도별 추정이 가능하도록 최종 배분방법을 결정하고자 한다.

먼저, 1차 배분인 17개 시도의 배분방법을 결정하기 위해서 검토한 17개 시도에 대한 비례배분과 제공근 비례배분의 배분 결과는 <표 4-13>과 같다.

<표 4-13> 17개 시도에 대한 조사구 배분 결과

(단위: 개)

지역	표본 크기: 8,400가구	
	비례배분	제공근 비례배분
전국	840	840
서울	160	98
부산	55	58
대구	39	48
인천	46	53
광주	24	38
대전	25	39
울산	18	32
세종	6	18
경기	209	112
강원	26	40
충북	27	40
충남	36	46
전북	30	42
전남	29	42
경북	45	52
경남	54	57
제주	11	25

자료: 필자 작성

모집단의 분석 결과에 따르면 17개 시도 변수는 가구수에 대한 비중의 차이가 매우 큰 변수이다. 서울특별시, 경기도와 같이 가구수가 많은 지역과 세종특별자치시, 제주특별자치도와 같이 가구수가 적은 지역 간의 편차가 매우 크다. 따라서 가구수에 따른 비율로 배분하는 비례배분의 경우 경기도는 209개의 조사구가 배분되는 반면, 세종특별자치시는 6개 조사구만 배분되어 격차가 매우 크다.

따라서 한국의료패널조사는 조사목적에 충실하기 위해서 17개 시도에 대한 통계산출이 필요하고, 이를 위해서는 적절한 수준의 표본이 확보되어야 한다. 비례배분방법을 활용하면 일부 지역은 통계산출이 어려울 수 있으므로, 17개 시도에 대한 배분방법은 제곱근 비례배분방법을 활용하고자 한다.

〈표 4-14〉 17개 시도에 대한 동부·읍면부 조사구 배분 결과

(단위: 개)

지역	표본 크기: 8,400가구			
	비례배분		제공근 비례배분	
	동부	읍면부	동부	읍면부
전국	653	187	597	243
서울	98	-	98	-
부산	55	3	47	11
대구	43	5	36	12
인천	52	1	46	7
광주	38	-	38	-
대전	39	-	39	-
울산	26	6	21	11
세종	12	6	11	7
경기	93	19	77	35
강원	24	16	22	18
충북	23	17	21	19
충남	21	25	22	24
전북	30	12	26	16
전남	18	24	20	22
경북	27	25	26	26
경남	36	21	32	25
제주	18	7	15	10

자료: 필자 작성

2차 배분은 동부·읍면부로, 동부·읍면부에 대한 비례배분과 제공근 비례배분을 검토하였다. 동부의 가구 비중은 80.9%이고, 읍면부의 가구 비중은 19.1%로 나타났다. 층이 2개로 단순하며 19.1%의 가구 비중은 비례배분을 활용하기에 충분하다고 판단된다. 제공근 비례배분의 경우 비중이 높은 층에서 과소표집되는 단점이 있으므로 동부·읍면부에 대해서는

비례배분을 적용하는 것으로 결정하였다.

앞의 2기 한국의료패널조사 분석 결과에서 보았듯이, 인천광역시 읍면부의 경우 2019년 연간데이터에서 1가구이고 그 이후로는 0가구였다. 패널조사를 위해서는 충분한 표본이 필요하다. 인천광역시에 비례배분을 하면 1개 조사구(10가구)만 배분되어 추후 표본이 탈락할 위험이 있으므로, 최소 3개 조사구를 배분할 수 있도록 조정하였고 최종 배분 결과는 <표 4-15>와 같다.

<표 4-15> 17개 시도 및 동부읍면부에 대한 최종 조사구 배분 결과

(단위: 개)

지역	표본 크기: 8,400가구		
	계	동부	읍면부
전국	840	651	189
서울	98	98	0
부산	58	55	3
대구	48	43	5
인천	53	50	3
광주	38	38	0
대전	39	39	0
울산	32	26	6
세종	18	12	6
경기	112	93	19
강원	40	24	16
충북	40	23	17
충남	46	21	25
전북	42	30	12
전남	42	18	24
경북	52	27	25
경남	57	36	21
제주	25	18	7

자료: 필자 작성

5. 추출 표본 규모

표본의 크기 및 층화변수에 대한 표본을 배분한 결과와 2021년 연간 데이터의 가구수를 비교하여 실질적으로 추출해야 할 추가표본의 규모를 결정하였다. 조사구 1개당 평균 10가구를 조사하므로 배분한 조사구 수를 기반으로 했을 때 층별 조사 가구수는 <표 4-16>과 같다.

<표 4-16> 17개 시도 및 동부읍면부의 목표 가구수

(단위: 가구)

지역	표본 크기: 8,400가구		
	계	동부	읍면부
전국	8,400	6,510	1,890
서울	980	980	0
부산	580	550	30
대구	480	430	50
인천	530	500	30
광주	380	380	0
대전	390	390	0
울산	320	260	60
세종	180	120	60
경기	1,120	930	190
강원	400	240	160
충북	400	230	170
충남	460	210	250
전북	420	300	120
전남	420	180	240
경북	520	270	250
경남	570	360	210
제주	250	180	70

자료: 필자 작성

2021년 연간데이터의 가구수, 목표 가구수와 과부족 가구수를 각각 산출하였고, 이를 기반으로 추출해야 할 추가 가구수를 결정하였다.

〈표 4-17〉 통합표본을 구축을 위한 과부족 현황

(단위: 가구)

지역	2019년 연간데이터			목표 가구수			과부족 가구수		
	계	동부	읍면부	계	동부	읍면부	계	동부	읍면부
전국	5,878	4,364	1,514	8,400	6,510	1,890	2,522	2,166	356
서울	531	531	-	980	980	0	449	449	-
부산	409	407	2*	580	550	30	171	143	28
대구	202	197	5*	480	430	50	278	233	45
인천	333	333	-*	530	500	30	197	167	30
광주	332	332	-	380	380	0	48	48	-
대전	245	245	-	390	390	0	145	145	-
울산	311	309	2*	320	260	60	9	-49	58
세종	136	71	65	180	120	60	44	49	-5
경기	707	577	130	1,120	930	190	413	353	60
강원	213	113	100	400	240	160	187	127	60
충북	444	241	203	400	230	170	-44	-11	-33
충남	245	88	157	460	210	250	215	122	93
전북	426	274	152	420	300	120	-6	26	-32
전남	390	139	251	420	180	240	30	41	-11
경북	350	165	185	520	270	250	170	105	65
경남	507	285	222	570	360	210	63	75	-12
제주	97	57	40	250	180	70	153	123	30

주: *는 패널마모가 심해 현재 패널 자료가 없거나 패널 규모가 5가구 이하인 지역을 의미함.
자료: 필자 작성

〈표 4-17〉을 보면 대부분의 층에서 표본이 부족하여 추가로 표본 추출이 필요한 상황이다. 다만 충청북도, 전라북도 읍면부, 전라남도 읍면부, 경상남도 읍면부 등과 같은 층에서는 목표 가구수보다 2021년 연간데이

터의 가구수가 더 많은 것으로 나타났다.

패널조사의 특성상 패널을 지속적으로 유지하는 것이 중요하므로 목표 가구수보다 조사 완료 가구수가 더 큰 층에서는 추가표본을 추출하지 않으며, 부족한 층에 대해서만 표본을 추출하였다.

따라서 최종적으로 추출한 표본 현황은 <표 4-18>과 같다. 다만, 일부 층의 경우 음(-)의 값을 가지는 층을 제외하였으므로 전체적으로 추출한 가구수는 <표 4-17>보다 커진 것을 볼 수 있다.

<표 4-18> 17개 시도 및 동부읍면부의 최종 추출 가구수

(단위: 가구)

지역	계	동부	읍면부
전국	2,675	2,206	469
서울	449	449	-
부산	171	143	28
대구	278	233	45
인천	197	167	30
광주	48	48	-
대전	145	145	-
울산	58	-	58
세종	49	49	-
경기	413	353	60
강원	187	127	60
충북	-	-	-
충남	215	122	93
전북	26	26	-
전남	41	41	-
경북	170	105	65
경남	75	75	-
제주	153	123	30

자료: 필자 작성

모집단의 추출 단위인 조사구로 변환하기 위해서 평균 10가구를 적용하여 조사구 수로 변환하였다. 조사구 내 가구수는 편차가 매우 크므로 조사구 내 가구수에 따라 조사할 가구수를 조정할 필요가 있다. 우선 최소 조사구 1개당 8가구를 조사하며, 60가구를 기준으로 10가구를, 80가구 이상일 경우 12가구를 조사하도록 조사구 내 조사할 가구수를 탄력적으로 적용하는 것을 제안한다. 이러한 규모를 반영한 추출 조사구는 <표 4-19>와 같다.

<표 4-19> 17개 시도 및 동부읍면부의 최종 추출 조사구 수

(단위: 개)

지역	계	동부	읍면부
전국	270	222	48
서울	45	45	-
부산	17	14	3
대구	28	23	5
인천	20	17	3
광주	5	5	-
대전	15	15	-
울산	6	-	6
세종	5	5	-
경기	41	35	6
강원	19	13	6
충북	-	-	-
충남	21	12	9
전북	3	3	-
전남	4	4	-
경북	18	11	7
경남	8	8	-
제주	15	12	3

자료: 필자 작성

6. 통합패널 규모

2기 한국의료패널조사는 2016년의 모집단을 활용하여 표본이 설계되었으므로 최신의 모집단을 반영하여 통계의 정확도를 향상시키고자 한다. 또한, 5년간의 패널조사 과정에서 표본 탈락 등으로 인해 표본 규모가 최초의 표본설계와는 차이가 생겼으므로 이를 보정하기 위해서 2021년 인구주택총조사 결과를 활용한 추가표본을 투입하여 통합패널로 구축하고자 한다.

따라서 2기 한국의료패널조사에서 유지되고 있는 표본을 유지하면서 마모된 층에 대해서만 표본을 추가적으로 투입하고, 목표 표본보다 큰 층에 대해서는 표본을 축소하지 않고 유지한다. 그러므로 통합패널의 전체적인 표본 크기는 앞에서 배분한 표본보다 많이 확보될 것으로 예상된다. 통합패널의 최종 표본 규모는 <표 4-20>과 같다.

<표 4-20> 통합패널의 최종 목표 가구수

(단위: 가구)

지역	계	동부	읍면부
전국	8,553	6,570	1,983
서울	980	980	-
부산	580	550	30
대구	480	430	50
인천	530	500	30
광주	380	380	-
대전	390	390	-
울산	369	309	60
세종	185	120	65
경기	1,120	930	190
강원	400	240	160

지역	계	동부	읍면부
충북	444	241	203
충남	460	210	250
전북	452	300	152
전남	431	180	251
경북	520	270	250
경남	582	360	222
제주	250	180	70

자료: 필자 작성

7. 통합패널의 주요 변수에 대한 예상 상대표준오차 검토

앞에서 2기 한국의료패널조사 2021년 연간데이터에 마모된 표본을 추가적으로 추출하여 구축할 통합패널의 적절성을 평가하기 위해 2021년 연간데이터의 주요 변수에 대한 상대표준오차를 활용하여 층별 예상 상대표준오차를 분석하였다.

상대표준오차는 25% 이상일 경우 해당 층에 대한 통계의 신뢰가 낮아 층별 예상 상대표준오차를 통해 통합패널이 적절하게 설계되었는지 확인하였다. 그리고 층별 예상 상대표준오차가 25%를 초과할 경우 그 원인을 찾아 필요할 경우 추가적인 표본 조정을 통해 예상 상대표준오차를 개선하였다.

먼저, 전국의 통합패널에 대한 예상 상대표준오차를 살펴보면 표본 추가로 인해 현행 표본에 비해 안정적인 결과를 제공하게 되었다. 특히 연간 가구 총소득의 경우 1.3%이고, 만성질환 수는 1.2%로 현재보다 정확도 높은 통계를 산출하게 되므로 통계 신뢰도가 높아질 것으로 기대한다(〈표 4-21〉 및 〈표 4-22〉 참조).

17개 시도에 대한 예상 상대표준오차를 검토하면, 17개 시도의 연간

가구 총소득은 경기도에서 2.2%로 가장 낮은 반면에, 세종특별자치시가 7.2%, 광주광역시 6.4%로 가장 높은 편에 속하였다. 하지만 모든 시도가 10% 이내로 나타나 안정적인 결과가 산출될 것으로 예상된다. 17개 시도의 만성질환 수는 세종특별자치시가 8.8%, 제주특별자치도가 7%로 다른 시도에 비해 예상 상대표준오차가 높은 편으로 나타났다. 그러나 모든 시도가 10% 이내로 나타나 안정적인 결과가 산출될 것으로 예상된다.

동부·읍면부의 예상 상대표준오차를 검토하면, 연간 가구 총소득은 동부의 경우 1.5%이고 읍면부의 경우 2.3%이다(〈표 4-23〉 참조). 만성질환 수는 동부의 경우 1.4%이고 읍면부의 경우 2.3%로 예상되어 정확도가 높은 통계를 산출하게 되므로 통계 신뢰도가 높아질 것으로 기대한다(〈표 4-24〉 참조).

17개 시도의 동부·읍면부별 예상 상대표준오차를 보면, 연간 가구 총소득은 제주특별자치도 읍면부가 13.6%로 약간 높지만, 제주특별자치도 읍면부를 제외한 나머지 시도 및 동부·읍면부의 예상 상대표준오차는 10% 이내로 안정적으로 나타났다. 만성질환 수의 예상 상대표준오차는 연간 가구 총소득보다 약간 높은 편이며 세종특별자치시 동부, 부산광역시 읍면부, 대구광역시 읍면부, 울산 읍면부, 제주 읍면부가 상대적으로 높게 나타났다. 부산광역시, 대구광역시, 울산광역시는 광역시여서 읍면부의 표본이 작으므로, 광역시는 읍면부를 통합해 추정하는 것이 바람직하다고 볼 수 있다. 세종특별자치시 동부와 제주특별자치도 읍면부도 약간 높게 나타났다. 동부·읍면부 통계 산출이 가능하지만, 특별자치시도의 특성상 동부·읍면부를 통합하여 산출하는 것을 권유한다.

결론적으로 17개 시도 및 동부·읍면부의 추가표본에 대한 예상 상대표준 오차는 현행보다 크게 개선됨에 따라 전반적인 통계 품질도 개선되어 통계 정확성 및 신뢰도가 향상될 것으로 기대한다.

〈표 4-21〉 통합패널의 17개 시도에 대해 연간 가구 총소득으로 파악한 목표 CV

(단위: 가구, %)

지역	2021년 연간데이터 (조사 결과)		2기 한국의료패널의 통합패널 (예정)	
	가구	(CV)	가구	목표 정도 (CV)
전국	5,878	1.6	8,553	1.3
서울	531	5.1	980	3.8
부산	409	4.7	580	4.0
대구	202	5.6	480	3.6
인천	333	4.4	530	3.5
광주	332	6.9	380	6.4
대전	245	5.7	390	4.5
울산	311	5.2	369	4.8
세종	136	8.4	185	7.2
경기	707	2.8	1,120	2.2
강원	213	5.5	400	4.0
충북	444	4.6	444	4.6
충남	245	7.6	460	5.6
전북	426	4.7	452	4.5
전남	390	5.1	431	4.9
경북	350	4.9	520	4.0
경남	507	3.6	582	3.4
제주	97	8.2	250	5.1

자료: 한국보건사회연구원·국민건강보험공단. (2023). 2기 한국의료패널조사 2019년~2021년 연간데이터(2023년 학술대회용 버전)를 분석함.

110 패널조사 표본이탈 경향 분석과 추가표본 추출에 관한 연구- 한국의료패널조사를 중심으로

〈표 4-22〉 통합패널의 17개 시도에 대해 만성질환 수로 파악한 목표 CV

(단위: 가구, %)

지역	2021년 연간데이터 (조사 결과)		2기 한국의료패널의 통합패널 (예정)	
	가구	(CV)%	가구	목표 정도 (CV)%
전국	5,878	1.5	8,553	1.2
서울	531	4.4	980	3.2
부산	409	4.5	580	3.8
대구	202	7.4	480	4.8
인천	333	6.4	530	5.1
광주	332	6.1	380	5.7
대전	245	6.7	390	5.3
울산	311	5.6	369	5.2
세종	136	10.3	185	8.8
경기	707	3.7	1,120	3.0
강원	213	5.2	400	3.8
충북	444	4.7	444	4.7
충남	245	6.7	460	4.9
전북	426	4.7	452	4.6
전남	390	5.3	431	5.1
경북	350	5.6	520	4.6
경남	507	4.4	582	4.1
제주	97	11.3	250	7.0

자료: 한국보건사회연구원·국민건강보험공단. (2023). 2기 한국의료패널조사 2019년~2021년 연간데이터(2023년 학술대회용 버전)를 분석함.

〈표 4-23〉 통합패널의 17개 시도 및 동부·읍면부에 대해 연간 가구 총소득으로 파악한
 목표 CV

(단위: 가구, %)

지역	가구 총소득 목표 cv(%)		
	계	동부	읍면부
전국	1.3	1.5	2.3
서울	3.8	3.8	-
부산	4.0	4.1	2.1*
대구	3.6	3.8	8.1*
인천	3.5	3.6	*
광주	6.4	6.4	-
대전	4.5	4.5	-
울산	4.8	5.3	1.5*
세종	7.2	9.4	9.8
경기	2.2	2.4	5.2
강원	4.0	5.3	5.9
충북	4.6	5.7	6.9
충남	5.6	7.1	5.5
전북	4.5	5.3	6.8
전남	4.9	7.3	6.4
경북	4.0	5.6	5.5
경남	3.4	4.0	5.9
제주	5.1	5.0	13.6

주: *는 패널마모가 심해 현재 패널 자료가 없거나 패널 규모가 5가구 이하인 지역을 의미함.

자료: 한국보건사회연구원·국민건강보험공단. (2023). 2기 한국의료패널조사 2019년~2021년
 연간데이터(2023년 학술대회용 버전)를 분석함.

112 패널조사 표본이탈 경향 분석과 추가표본 추출에 관한 연구- 한국의료패널조사를 중심으로

〈표 4-24〉 통합패널의 17개 시도 및 동부·읍면부에 대해 만성질환 수로 파악한 목표 CV
(단위: 가구, %)

지역	만성질환 수 목표 cv(%)		
	계	동부	읍면부
전국	1.2	1.4	2.3
서울	3.2	3.2	-
부산	3.8	3.9	34.0*
대구	4.8	5.1	11.4*
인천	5.1	5.2	*
광주	5.7	5.7	-
대전	5.3	5.3	-
울산	5.2	5.6	19.0*
세종	8.8	15.8	6.6
경기	3.0	3.4	6.2
강원	3.8	4.9	6.0
충북	4.7	7.1	4.9
충남	4.9	8.2	5.7
전북	4.6	5.8	6.9
전남	5.1	7.5	6.9
경북	4.6	6.2	6.8
경남	4.1	5.4	5.8
제주	7.0	8.8	10.2

주: *는 패널마모가 심해 현재 패널 자료가 없거나 패널 규모가 5가구 이하인 지역을 의미함.
 자료: 한국보건사회연구원·국민건강보험공단. (2023). 2기 한국의료패널조사 2019년~2021년
 연간데이터(2023년 학술대회용 버전)를 분석함.

8. 표본조사구 추출

추가표본의 표본설계에서는 2기 한국의료패널조사의 2021년 연간 데이터의 표본을 유지하면서 통합패널을 구성하기 위한 마모된 표본을 추가적으로 추출하고자 한다. 인구주택총조사 표본추출틀은 조사구 단위로 구축된 데이터이므로 조사구를 먼저 추출하고, 조사구 내에서 가구를 추출한다. 1개 조사구당 평균 10가구를 추출하도록 하고, 조사구의 가구 수 규모에 따라 8~12가구를 추출하도록 제안한다.

조사구를 추출하기 전에 계통추출을 위해 내재적 층화변수를 결정하고, 내재적 층화변수로 정렬한 다음에 가구수를 기준으로 확률비례계통추출로 조사구를 추출한다. 2기 한국의료패널조사 구축 당시의 내재적 층화변수는 조사구 특성, 점유형태, 혼인상태, 거처 종류인 4개 변수를 활용하였다. 내재적 층화변수 중에서 ‘점유형태’는 등록센서스 방식으로 변경된 이후부터는 5년마다 조사되는 인구주택총조사 표본조사에서만 조사하고 있다. 그래서 모집단으로 활용하는 등록센서스 방식의 인구주택총조사의 전수 조사 부문에서는 이용할 수 없으므로 내재적 층화변수로 활용할 수 없다. ‘혼인상태’와 ‘거처 종류’의 경우 가구 데이터에서는 각각 변수로 있으나, 표본추출틀은 조사구 데이터이므로 혼인상태와 거처 종류의 각각에 대한 인구수 및 가구수 변수로 구성되어 있다. 예를 들어 거처 종류는 단독주택, 아파트, 연립주택 등으로 구분되는데, 표본추출틀에서는 단독주택 거주 가구수, 아파트 거주 가구수, 연립주택 거주 가구수 등으로 구성되어 있으므로 이 모든 변수를 내재적 층화변수로 활용할 경우 내재적 층화의 효과가 감소한다는 문제가 있다. 이러한 모집단의 변화로 인해 통합패널의 일치성을 위해 2기 한국의료패널조사 구축 당시의 내재적 층화변수와 유사한 효과를 주면서 적절한 수준을 갖춘 내재적 층화변수를 새롭게

결정할 필요가 있다.

먼저 혼인상태를 대체할 내재적 총화변수로 '유배우 인구수'를 활용하고자 한다. 혼인상태에서 혼인한 가구와 미혼, 이혼, 사별 가구는 가구 구성 측면에서 차이가 발생하는데, 이를 반영하기 위해 유배우 인구수를 활용하고자 한다. 거처 종류는 주거의 특성을 나타내는 변수로 가구 소득 등을 간접적으로 추정할 수 있는 변수이다. 가구 소득을 간접적으로 알 수 있는 조사구의 아파트 거주 가구수를 내재적 총화변수로 활용하고자 한다.

내재적 총화변수로 정렬 후 계통추출로 조사구를 추출하며, 이 과정에서 지역에 따른 추가적인 고려를 하고자 한다. 총화변수로 17개 시도 변수를 활용하고 있으나, 17개 시도 내에 시군구별로 지역 간 차이가 있으므로 이를 최대한 반영하기 위해 2021년 연간데이터에서 조사되지 않는 시군구 위주로 통합패널을 위한 추가표본을 추출하고자 한다.

먼저 시군구별 가구수를 제공근 비례배분을 활용하여 17개 시도에 대한 추출 가구수를 배분한 후, 동부·읍면부에서는 비례배분방법을 통해 시군구별 동부·읍면부의 표본을 배분한다. 이 배분 결과, 동부·읍면부의 경우 일부 동부 혹은 읍면부로만 이루어진 시군구가 다수 있어서, 배분 결과가 표본설계 시의 17개 시도 및 동읍면부 배분 결과와 차이가 발생하여, 이를 조정하였다. 조정 시 2021년 연간데이터에서 조사되지 않은 조사구, 2021년 연간데이터의 가구수와 목표 가구수와의 차이가 큰 시군구 순으로 조정하였다.

9. 표본가구 추출

추출된 표본조사구에서 표본가구를 추출하기 위해 조사구 내 가구에 대한 가구명부에서 계통추출로 표본가구를 선택한다. 가구명부는 주소

등으로 단순하게 구성되어 표본가구에 대한 다른 추출방법을 적용할 수 없으므로 가구명부상에서 계통추출을 활용하여 표본가구를 추출 후 해당 가구에 조사한다.

10. 표본 대체

조사에 활용되는 가구명부는 2021년 11월 1일 기준 명부이므로 재개발 혹은 천재지변 등으로 인해 조사구가 없어지거나, 조사가 불가능한 경우가 발생할 수 있어 현재 시점의 명부와는 차이가 있을 수 있다. 이 경우 예비 표본에서 동일한 층 내 조사구로 대체하여 조사를 진행한다.

단위무응답이 발생할 경우 1차적으로 조사구 내 다른 가구를 확률 추출하여 조사하며, 조사구 내 모든 가구를 소모할 경우 예비표본에서 동일한 층 내 조사구로 가구를 추출하여 조사를 진행한다.

11. 가중치

2기 한국의료패널조사의 통합패널은 기존 2021년 연간데이터의 패널을 유지하면서, 표본을 추가하여 구축한 것이다. 2기 한국의료패널조사의 원패널은 기존의 표본설계에서 부여한 가중치를 가지고 있다. 2기 통합패널에 대한 가중치를 부여하기 위해 기존 2기 한국의료패널조사의 원패널 가중치는 기존의 부여 방식으로 계속 부여하고, 통합패널은 1차를 기준으로 새로운 가중치를 부여한다. 단, 통합패널의 새로운 가중치는 기존 표본설계에서 부여한 가중치 부여 방식과 동일하게 부여한다.

2기 한국의료패널조사는 기존 조사에서 유지하고 있는 원패널과 이 장에서 층별 마모된 표본에 추가한 통합패널로 분리하여 운영한다. 따라서

원패널과 통합패널에 대한 가중치는 다음과 같이 부여하여 운영할 필요가 있다.

- 2기 한국의료패널조사 원패널: 현행 가중치 부여 방식으로 계속 부여함
- 2기 한국의료패널조사 통합패널: 구축 시점을 1차 조사로 설정하여 가중치 부여함(원패널 가중치 부여 방식과 동일하게 적용)

가. 기본 가중치 부여 방안

먼저 표본설계에 따른 기본 가중치를 부여하기 위해서 가중표본 합계(weighted sample total)를 산출하고, 이를 사용하여 평균, 구성비 등의 각종 통계를 산출하도록 한다. 또한, 추정치 편향을 없애기 위해 사후층화(post stratification)를 하여 가중치 조정을 실시하고자 한다.

조사 대상인 가구의 추출은 1단계에서 조사구를 추출하고, 2단계에서 가구를 추출하므로, 가중치 부여도 또한 동일한 순서로 조사구 추출이 1단계이고 추출된 조사구에서 가구 추출이 2단계이다. 각 조사구 내 가구들은 동일한 가중값(설계 가중치)을 가지며, 조사 완료 후 추출 확률을 고려하여 가중값을 부여한다.

$$\text{설계 가중치} = \frac{S_h}{n_h m_{hi}}$$

L : 층의 수

N_h : 층 h 의 모집단 조사구 수

n_h : 층 h 의 표본 조사구 수

S_{hi} : 층 h 의 i 번째 조사구의 총가구수

$S_h = \sum_{i=1}^{N_h} S_{hi}$: 층 h 에서 총가구수

M_{hi} : 층 h 의 i 번째 조사구 내 가구수

m_{hi} : 층 h 의 i 번째 조사구 내 조사 착수 가구수

r_{hi} : 층 h 의 i 번째 표본조사구 내 조사 완료 가구수

조사 과정에서 발생한 무응답을 조정하기 위해서 무응답 조정 가중치를 산출한다. 무응답 조정은 최초 표본 단위인 가구가 단위무응답인 경우 이를 보정하기 위해 최초 목표 표본 크기를 유효 표본 크기로 나누어 조정한다. 최초 표본 단위를 대체표본으로 교체(substitution)하여 조사하는 경우 무응답 보정은 필요하지 않다.

$$\text{무응답 조정 가중치} = \frac{m_{hi}}{r_{hi}}$$

사후층화는 레이킹 비 방법(Raking Ratio Method)을 적용하여 사후 층화 가중치를 부여한다. 설계 가중치와 무응답 조정 가중치로 계산된 가구 가중치는 17개 시도(동부·읍면부 고려), 가구원 수 정보를 이용하여 기준시점에서 주민등록상 인구 집계 결과로 사후 조정한다.

최종 가중치의 산술식은 다음과 같다.

$$\text{최종 가중치} = \text{설계 가중치} \times \text{무응답 조정 가중치} \times \text{사후층화 가중치}$$

나. 종단면 및 횡단면 가중치 부여 방안¹⁰⁾

한국의료패널에서 가구의 횡단면 가중치, 가구원의 종단면 및 횡단면 가중치는 다음의 순서로 산출한다.

먼저 산출하는 것은 가구원 종단면 가중치이다. 이전 연도 가구원 종단면 가중치를 기저로 패널마모를 반영하여 가구원 종단면 가중치를 먼저 산출한다. 다음으로 가구원 종단면 가중치를 기저로 가구원 횡단면 가중치를 산출한다. 마지막으로 가구원 횡단면 가중치를 기저로 가구원 횡단면 가중치를 산출한다.

개인 종단면 가중치 산출 과정을 보면, t차 조사(2차 이후) 개인 종단면 가중치는 t차 조사 종단면 기본 가중치에 무응답 보정과 개인별 변동 상황을 고려하여 조정된 뒤, 사후조정하여 보정한다. t차 조사 종단면 기본 가중치는 (t-1)차 조사의 개인 종단면 가중치를 기본 가중치로 한다.

t차 조사에서 무응답으로 인한 실제 표본 크기 감소를 반영하기 위한 무응답 보정은 다음과 같다.

t차 조사에서 조사가 완료된 개인 가중치 조정은 패널 탈락으로 인한 무응답 보정을 위해, t차 조사 종단면 응답 여부 변수와 (t-1)차 조사 변수들의 관계를 로지스틱 회귀모형을 이용하여 응답 확률을 추정한다. 로지스틱 회귀모형의 설명 변수는 응답자의 성별, 교육, 연령, 지역이고, t차 조사 종단면 응답 여부 변수는 t차와 (t-1)차 조사 모두 응답하면 1, 그렇지

10) 2기 한국의료패널조사 표본설계 보고서의 가중치 부여 방안을 인용하여 작성함.

않은 경우는 0을 가진다. 다음으로 t차 조사의 개인 종단면 기본 가중치에 로지스틱 회귀모형으로 추정된 응답 확률의 역수를 곱하여 t차 조사의 개인 종단면 가중치를 조정한다. t차 웨이브의 기본 가중치를 $w_{\text{기본}t\text{차}}$, 로지스틱 회귀분석을 통해 예측된 예측 응답 확률을 $\hat{p}_{t\text{차}}$ 라고 하면 무응답 보정을 통해 얻어지는 가중치는 $w_{t\text{차}} = w_{\text{기본}t\text{차}} \times \hat{p}_{t\text{차}}^{-1}$ 로 표현된다.

그리고 개인별 변동 상황에 따라 t차 조사 종단면 가중치를 조정한다. t차 조사에 새롭게 태어난 가구원의 경우 원가구의 평균 가중값을 부여한다. 분가 가구원의 경우에는 원가구(HHIDWON)의 가구 평균 가중값을 적용한다. t차 웨이브에 새롭게 진입한 신규 가구원에 대해서는 0의 종단면 가중치를 부여한다. 마지막으로 지역 및 응답자의 인구학적 특성에 따라 인구 추계 값을 이용하여 사후조정을 실시한다. 지나치게 큰 가중치 값들은 추정량의 분산을 크게 만들어 추정의 정확도에 영향을 미칠 수 있는데, 이러한 가중치의 과도한 변동을 줄이기 위해 t차 조사에서는 사후조정 후 극단 가중치 조정을 실시한다.

12. 추정¹¹⁾

추정식은 표본설계 과정을 반영한 가중치를 활용하여 가중치가 반영된 추정식을 제공하는 것이 필요하다. 또한 현행 조사와의 시계열 연속성 등을 고려하여 현행 추정식에 근거하여 수식을 제공하는 것이 필요하다.

한국의료패널조사에서 만성질환 수, 연간 가구 총소득과 같이 주요한 변수의 요약을 위한 모수(평균, 비율 및 분산) 추정식은 다음과 같다.

11) 2기 한국의료패널조사 표본설계 보고서를 인용하여 작성함.

- 전체 가구의 평균 산출

$$\widehat{Y} = \sum_{h=1}^L \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} w_{hij} y_{hij} / \sum \sum \sum w_{hij}$$

w_{hij} : 각 응답자에 부여된 가중치

y_{hij} : 각 응답 결과

L : 층의 수

n_h : 층 h 에서의 표본 조사구의 수

m_{hi} : 층 h 내 i 번째 표본 조사구의 응답자 수

$w_{...} = \sum_{h=1}^L \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} w_{hij}$: 전체 응답자에 대한 가중치의 합계

- 추정량의 분산

$$var(\widehat{Y}) = \sum_{h=1}^L \frac{n_h(1-f_h)}{n_h-1} \sum_{i=1}^{n_h} (e_{hi} - \overline{e_{h..}})^2$$

L : 층의 수

n_h : 층 h 에서의 조사구 수, m_{hi} 는 층 h 내 i 번째 표본조사구의 응답자 수

$$f_h = n_h / N_h$$

$$e_{hi} = \left(\sum_{j=1}^{m_{hi}} w_{hij} (y_{hij} - \bar{y}) \right) / w_{...}$$

$$\overline{e_{h..}} = \left(\sum_{i=1}^{n_h} e_{hi} \right) / n_h$$

- 표준오차, 상대표준오차, 오차한계

$s.e(\bar{y}) = \sqrt{var(\bar{y})}$: 모평균 및 모비율 추정에 대한 표준오차 (standard error)

$rse(\bar{y}) = \frac{s.e(\bar{y})}{\bar{y}} \times 100(\%)$: 상대표준오차(relative standard error)

오차의 한계 = $1.96 \times \sqrt{var(\bar{y})}$: 95% 신뢰수준 오차의 한계

또한 제공된 추정 식은 가중치를 반영한 추정 방법으로 분석하는 통계 패키지를 활용하는 것이 바람직하다. 한국의료패널조사의 데이터 분석은 복합표본이므로 한국의료패널조사에서 제공하는 추정 및 분석을 참고하여 복합표본조사 데이터 분석을 위한 패키지(SAS survey procedure, SPSS complex samples 등)로 분석할 것을 권유한다.

제4절 소결

한국의료패널조사는 패널조사로 표본을 지속적으로 관리하며 유지하는 것이 중요하다. 하지만 최근 2기 한국의료패널조사는 모집단 특성의 급속한 변화뿐만 아니라 5년간의 조사로 표본이 마모되어 모집단 특성과 괴리가 있었다. 이렇듯 표본 특성에 변화가 발생했기 때문에 모집단의 변화를 반영하고, 마모된 표본을 보정하는 것이 필요하였다. 이 장에서는 현재 유지되고 있는 패널을 지속 관리하면서 변화된 모집단 특성과 표본 마모를 보정하는 방안을 제시하였다. 2기 한국의료패널조사의 패널에 대한 추가 보정 방안 결과를 간략하게 요약하면 다음과 같다.

첫째, 추가 보정 방안을 위한 표본설계는 변화된 모집단 특성을 반영하고, 2기 한국의료패널조사의 마모된 표본의 추가 보정을 목적으로 하기에 2기 조사의 표본설계와 동일하게 설계하였다. 모집단 분석을 통해 동일한 층화 기준을 적용하여 1차 층화는 17개 시도, 2차 층화는 동부·읍면부로 층화하였다.

둘째, 추가표본 규모는 2기 한국의료패널조사의 목표 표본 규모를 고려하여 표본 규모를 결정하였으며, 최종 목표 표본 규모를 적정 수준인 8,500가구로 결정하였다. 추가표본 보정을 위한 표본 규모를 결정하기 위해서 2기 한국의료패널조사의 주요 변수에 대한 상대표준오차를 활용하여 표본 규모에 따른 예상 상대표준오차를 계산하여 결정하였다.

셋째, 표본배분은 현행 표본 체계를 유지하면서 시도별 추정이 가능하도록 최종 배분방법을 결정하고자 하였다. 2기 한국의료패널조사와 마찬가지로 1차 층화변수인 17개 시도는 제곱근 비례배분, 2차 층화변수인 동부·읍면부는 비례배분을 적용하였다.

넷째, 추가표본 규모는 목표 표본 규모를 기준으로 배분된 17개 시도 및 동부·읍면부 표본 규모와 2021년 연간데이터의 표본 규모를 비교하여 부족한 층에 대해서 보정하는 방식으로 추가표본 규모를 산정하였다. 추가 보정한 표본 규모는 270개 조사구이고 2,700가구이다.

다섯째, 추가 보정된 통합패널에 대한 예상 상대표준오차는 연간 가구 총소득의 경우 1.3%이며 만성질환 수는 1.2%로 안정적으로 나타났다. 연간 가구 총소득으로 본 상대표준오차는 동부 1.5%, 읍면부 2.3%이고, 만성질환 수는 동부 1.4%, 읍면부 2.3%로 통계 정확성이 제고되어 통계 신뢰도가 높아질 것으로 기대된다.

여섯째, 추가표본의 표본추출은 등록센서스 기반의 모집단에서 제공하는 변수인 유배우 인구수, 아파트 거주 가구수 등을 활용하였다. 기본적으로

2기 한국의료패널조사 구축 당시와 동일한 정렬순서를 이용하였고, 계통 추출방법으로 가구를 추출하는 것으로 제안하였다.

일곱째, 가중치는 현행 2기 한국의료패널조사와 동일한 방법을 적용하여 조사 결과의 산출 과정에 대한 일관성을 유지하도록 하였다.

결론적으로 2기 한국의료패널조사의 표본 마모 및 변화된 모집단 특성을 반영하면서 현재 2기 원패널은 유지하되 통합패널을 구축하는 방안인 것이다. 통합패널은 횡단 모집단을 대표할 수 있도록 구축하여 통계 품질을 제고하였으며, 동시에 2기 한국의료패널조사 원패널 데이터와 통합패널 데이터를 제공하는 것이 가능하다.



제5장

결론

제1절 주요 연구 결과 요약

제2절 패널 유지를 위한 패널 대상의 관리 전략

제 5 장 결론

이 연구는 2기 한국의료패널조사의 표본이탈 경향을 분석하고, 마모된 표본을 보완하기 위하여 표본을 추가하는 방안에 대해 살펴보았다. 주요 연구 결과를 정리하고, 패널 대상의 이탈을 최소화하기 위해서 패널 관리가 집중적으로 필요한 부분에 대해 제안하고자 한다.

제1절 주요 연구 결과 요약

제2장의 주요 연구 결과를 보면 2기 한국의료패널의 원표본 유지율은 2020년 연간데이터 기준으로 91.3%이고, 2021년은 86%로 하락하는 추이를 보였다. 1기 한국의료패널보다 높은 수준을 유지하고 있었다. 타 패널의 3차 연도 원표본 유지율과 비교했을 때 한국노동패널의 09통합 표본과 18통합표본에 비해서는 낮았으나, 한국복지패널과 장애인삶패널과는 비슷한 수준으로 나타났다.

응답 가구의 주요 변수에 따른 특성을 살펴보면, 2019년에 구축된 패널 가구의 거주 지역 분포를 유치조사와 비교했을 때 경기와 서울이 다른 지역에 비해 가장 큰 차이를 보이거나 1%p 이하로 크지 않은 편이었다. 2021년으로 갈수록 서울, 대구, 인천, 대전, 울산, 세종, 경기, 제주의 비율은 점차 감소하는 추세를 보였다. 가구원 수의 경우에는 2019년 이후로 1인과 2인 가구는 지속적으로 증가하는 반면에, 3인 가구 이상은 유치조사에 비해 점점 줄어드는 양상을 보였다. 주택형태는 아파트와 단독주택의 비율이 높은 편이었으나, 시간이 지남에 따라 아파트의 비율은

낮아지고 단독주택은 높아지는 추세였다.

응답 가구원의 인구사회학적 특성을 중심으로 살펴본 결과는 다음과 같았다. 성별은 조사 차수가 거듭되더라도 동일하게 유지되는 편이었는데, 여성의 비율이 남성보다 높게 나타났다. 연령대 분포의 경우 60세 이상의 비율은 시간이 지날수록 점점 증가하는 추세여서 노년층의 조사 참여율은 높은 데 비해 청년층과 중장년층은 낮았다. 혼인상태를 보면 혼인 중인 경우가 가장 높은 비율을 차지하며, 미혼인 경우는 다른 혼인상태에 비해 유치조사와 2019년의 차이가 가장 크게 나타났다. 최종 학력은 고졸 이상인 경우가 절반 이상이었는데, 2019년 고졸 이상 가구원의 비율이 유치조사에 비해 감소하여 학력이 높을수록 조사 참여도가 낮아진다고 볼 수 있다. 마지막으로 경제활동 상태는 비경제활동인구와 임금근로자의 비율이 가장 높았다. 15세 미만 응답자의 경우에는 2019년 이후로 점차 감소하는 추세였다.

패널가구의 조사 완료 여부에 따른 분포를 원가구를 중심으로 살펴 보았는데, 조사 연도가 지날수록 조사를 완료하지 못하는 가구가 점점 증가하고 있는 것으로 나타났다. 조사를 완료하지 못하는 경우는 두 가지로, '조사 가능하나 응답하지 않음'과 '조사 불가능함'으로 구분할 수 있다. 조사 가능하나 응답하지 않은 가구는 2021년 12.8%로 전년 대비 4.6%p 증가하였다. 조사 불가능한 가구의 비율도 전년 대비 0.7%p 상승하여 2021년에 1.2%로 나타나, 자연스럽게 패널에서 이탈하는 비율도 증가하고 있다고 볼 수 있다. 조사 가능하지만 응답하지 않은 가구의 조사 미완료 사유를 보면 2020년과 2021년 모두 조사 거부가 가장 높았으며 다음으로 접촉 회피, 추적 불가 순으로 나타났다. 한편, 2021년의 경우 2020년에 비해 조사 거부는 줄어드는 반면에, 추적 불가는 증가하는 양상을 보였다. 조사를 거부하거나 접촉을 회피하는 가구에 대해 조사

참여를 유인할 수 있는 방안을 마련해야 할 필요가 있다. 또한, 이사 등의 추적 불가의 비율이 지속적으로 증가하는 양상을 보이고 있어 이에 대한 대책 마련이 필요하다고 생각한다.

제3장에서는 패널가구와 패널가구원의 표본이탈 특성을 파악한 다음에, 응답형태에 따라 표본이탈에 영향을 주는 요인을 각각 살펴보았다. 2019년 연간데이터의 패널 대상을 기준으로 하여 2019년~2021년에 대한 응답 형태를 네 가지 집단(연속응답, 2020년 이탈, 2021년 이탈, 재진입)으로 구분하였다.

패널가구의 표본이탈 특성을 주요 변수를 중심으로 살펴본 결과, 거주 지역이 대구와 인천이면 2020년과 2021년에 이탈 비율이 높았다. 세종의 경우 2020년에 이탈 비율이 가장 높게 나타났다. 가구원 수가 많아질수록 이탈 비율이 높게 나타났다. 아파트에 거주하면 이탈 비율이 높은 편이었으나, 단독주택은 낮은 편에 속하였다. 연간 가구 총소득은 2020년 이탈의 경우 소득이 높을수록 이탈할 비율이 높은 반면에, 2021년 이탈의 경우 6,000만 원 이상이 가장 높았고, 그다음은 2,000만 원 미만으로 나타났다.

패널가구의 응답형태에 따른 3개 집단(연속응답, 2020년 이탈, 2021년 이탈)을 종속 변수로 하여 다항 로지스틱 회귀모형을 분석하였다. 사용한 설명 변수는 2019년 연간데이터에서의 가구 특성(거주 지역, 주거형태, 가구원 수, 연간 가구 총소득), 2019년~2021년에 걸친 가구의 상태 변화(이사 횟수), 그리고 2019년에서의 가구주 특성(성별, 연령, 혼인상태, 최종 학력)이다. <표 5-1>은 이 분석 결과를 요약한 것으로, 연속응답을 준거집단으로 하여 이탈 집단에 영향을 주는 요인, 즉 설명 변수 중에서 통계적으로 유의한 결과를 나타낸 변수를 정리하였다. 2020년과 2021년 이탈 집단에 공통으로 영향을 주는 요인을 살펴보면, 주거형태가 아파트인 경우 단독주택에 비해 이탈 가능성이 높았고, 이사 횟수가 적을수록,

가구주가 젊을수록 이탈 집단에 있을 가능성이 높은 것으로 나타났다. 이외에도 2021년 이탈 집단은 연간 가구 총소득이 저소득인 경우 중소득에 비해, 가구주의 최종 학력이 고졸인 경우 초졸에 비해 표본이탈에 영향을 주었다.

〈표 5-1〉 패널가구의 다항 로지스틱 회귀모형 분석 결과 요약

변수	2020년 이탈	2021년 이탈
거주 지역_서울		
주거형태_아파트	단독주택 (-) 다세대주택 (+)	단독주택 (-)
가구원 수		
연간 가구 총소득_중소득		저소득 (+)
이사 횟수	(-)	(-)
가구주 성별_남성		
가구주 연령	(-)	(-)
가구주 혼인상태_유배우자		
가구주 최종 학력_고졸		초졸 (-)

주: 1) 괄호 안의 내용은 각 변수에 해당하는 계수(coefficient)값의 부호를 의미함.

2) 연간 가구 총소득에서 저소득은 0원~2,000만 원 미만, 중소득은 2,000만 원~5,000만 원 미만, 고소득은 5,000만 원 이상을 의미함.

자료: 한국보건사회연구원·국민건강보험공단. (2023). 2기 한국의료패널조사 2019년~2021년 연간데이터(2023년 학술대회용 버전)를 분석함.

다음으로 패널가구원의 표본이탈 특성을 주요 변수를 중심으로 살펴 보았다. 성별의 경우 2020년 이탈에서는 남성과 여성의 비율이 비슷했으나, 2021년 이탈에서는 남성이 여성에 비해 높게 나타났다. 연령대의 경우 50세 이상~79세 이하가 표본이탈 비율이 낮은 편이었고, 49세 이하가 높은 편으로 나타났다. 한편, 80세 이상의 경우에는 연속응답 비율은 낮고 이탈 비율이 높은 경향을 보였는데, 이는 노환, 사망 등으로 인한 표본 이탈로 볼 수 있다. 미혼의 경우 다른 혼인상태에 비해 표본이탈의 경향이

높은 편이었다. 초졸과 중졸의 표본이탈 비율은 낮은 반면에, 고졸 이상일수록 비율이 높아지는 편이었다. 한편, 미취학 포함인 무학의 경우도 높게 나타나, 고학력 부모가 표본이탈하면서 미취학인 자녀도 함께 이탈한 것으로 추정할 수 있다. 경제활동 상태에 따라서는 15세 미만 응답자인 해당 없음과 비경제 활동인구의 표본이탈 비율이 높은 데 비해서, 기타와 고용주/자영업자의 비율은 낮았다.

한편, 패널가구와 패널가구원의 응답형태를 조합하여 4개 집단의 주요 특성을 살펴보았다. 패널조사 시에 패널가구와 패널가구원이 모두 응답한 집단이 81.4%로 가장 높았다. 다음으로 패널가구와 패널가구원이 모두 무응답한 집단은 15.8%이고, 패널가구는 응답하였으나 패널가구원은 무응답한 집단은 2.4%였으며, 재진입 등 기타는 0.4%로 미미하였다. 이를 통해 패널가구의 응답 여부가 패널을 유지하는 데 중요하다는 것을 알 수 있다. 패널가구는 응답하였으나 패널가구원은 무응답한 집단의 인구사회학적 특성을 보면 남성이 여성에 비해, 80세 이상이 다른 연령대에 비해, 사별/실종이 다른 혼인상태에 비해, 초졸과 중졸인 경우가 다른 최종 학력에 비해, 비경제활동인구가 다른 경제활동 상태에 비해 조사에 참여하지 않을 가능성이 높은 편으로 나타났다. 패널가구 내에서 고령의 가구원은 패널에서 이탈할 가능성이 높은 편이라고 볼 수 있다. 패널가구와 패널가구원 모두 무응답한 집단의 인구사회학적 특성을 보면 성별에 따른 차이는 없는 편이었다. 0세 이상~19세 이하 및 40세 이상~49세 이하가 다른 연령대에 비해, 미혼이 다른 혼인상태에 비해, 미취학이 포함된 무학과 대졸 이상이 다른 최종 학력에 비해, 기타를 제외한 나머지 경제활동 상태가 조사에 참여하지 않을 가능성이 높은 편으로 나타났다. 미혼이거나, 중년의 부모와 자녀로 구성된 가족이 패널에서 이탈할 가능성이 높다고 볼 수 있다.

패널가구원의 응답형태에 따른 3개 집단(연속응답, 2020년 이탈, 2021년 이탈)을 종속 변수로 하여 다항 로지스틱 회귀모형을 분석하였다. 사용한 설명 변수는 2019년 연간데이터에서의 가구원 특성(성별, 연령, 혼인상태, 최종 학력, 경제활동 상태), 2019년에서의 가구 특성(거주 지역, 주거형태, 가구원 수, 연간 가구 총소득), 그리고 2019년~2021년에 걸친 가구의 상태 변화(이사 횟수)이다. <표 5-2>는 이 분석 결과를 요약한 것으로, 연속응답을 준거집단으로 하여 이탈 집단에 영향을 주는 요인에 관한 것이다. 즉, 설명 변수 중에서 통계적으로 유의한 결과를 나타낸 변수를 정리하였다. 2020년과 2021년 패널가구원의 이탈에 영향을 주는 공통적인 요인은 가구원의 연령, 최종 학력, 경제활동 상태, 그리고 가구원 수, 이사 횟수였다. 가구원의 연령은 젊을수록, 최종 학력은 고졸이 초졸에 비해, 경제활동 상태는 임금근로자가 기타에 비해, 가구원 수는 많을수록, 이사 횟수는 적을수록 패널가구원의 이탈 경향이 높은 편으로 나타났다. 이사 횟수는 패널가구의 표본이탈에 영향을 주는 요인에도 해당되었는데, 패널가구와 패널가구원에서 동일한 결과가 나왔다. 한편, 주거형태의 경우 2020년은 아파트가 단독주택에 비해, 다세대주택이 아파트에 비해 표본이탈 경향이 높았고, 2021년은 기타가 아파트에 비해 표본이탈 경향이 높은 편에 속하였다. 2020년 결과는 패널가구의 표본이탈에 영향을 주는 요인과 동일하게 나타났다. 거주 지역의 경우 광역시(준거집단: 서울)가 통계적으로 유의한 영향을 주었는데 2020년 이탈 집단에서는 음(-)의 효과로, 2021년 이탈 집단에서는 양(+)의 효과로 상반되게 나타났다. 이외에도 2021년 이탈 집단은 남성이 여성에 비해, 연간 가구 총소득의 경우 저소득이 중소득에 비해 표본이탈에 영향을 주는 것으로 나타났다.

〈표 5-2〉 패널가구원의 다항 로지스틱 회귀모형 분석 결과 요약

변수	2020년 이탈	2021년 이탈
가구원 성별_남성		여성 (-)
가구원 연령	(-)	(-)
가구원 혼인상태_유배우자		
가구원 최종 학력_고졸	초졸 (-)	무학 (+) 초졸 (-)
가구원 경제활동 상태_임금근로자	기타 (-)	비경제활동인구 (+) 기타 (-)
거주 지역_서울	광역시 (-) 기타 도지역 (-)	광역시 (+)
주거형태_아파트	단독주택 (-) 다세대주택 (+)	기타 (+)
가구원 수	(+)	(+)
연간 가구 총소득_중소득		저소득 (+)
이사 횟수	(-)	(-)

주: 1) 괄호 안의 내용은 각 변수에 해당하는 계수(coefficient)값의 부호를 의미함.

2) 가구원 경제활동 상태에서 기타는 무급가족종사자, 자활근로, 공공근로, 희망근로, 노인 일자리가 해당함.

3) 주거형태에서 기타는 영업용 건물 내 주택(상가, 공장 등), 오피스텔, 호텔, 여관, 기숙사 등을 의미함.

4) 연간 가구 총소득에서 저소득은 0원~2,000만 원 미만, 중소득은 2,000만 원~5,000만 원 미만, 고소득은 5,000만 원 이상을 의미함.

자료: 한국보건사회연구원·국민건강보험공단. (2023). 2기 한국의료패널조사 2019년~2021년 연간데이터(2023년 학술대회용 버전)를 분석함.

제4장에서는 2기 한국의료패널이 2019년 유치조사를 시작으로 5년간의 조사를 실시하는 동안 마모된 표본을 보완하기 위해서 모집단 및 기존 조사 분석 결과를 기반으로 추가표본 추출 방안을 마련하였다. 2기 원패널을 보정하고 변화된 새로운 모집단 특성을 반영한 통합표본의 패널을 구축하려는 방향에 맞춰 표본설계를 하였다. 이때 현재 2기 원패널은 그대로 유지한다. 우선 목표모집단과 조사모집단이 일치하는지 비교한 결과, 큰 차이 없이 일치하는 것을 확인하였다. 다음은 지역 특성에 따른 층화변수인 17개 시도와 동부·읍면부에 따른 가구 분포를 분석하였는데 층화변수 간

가구수 차이가 큰 것으로 나타났다. 표본설계 시 적절한 배분방법을 선정하여 가구가 적은 층에서 과소표집이 되지 않도록 하였다. 기존 조사 결과 분석에서는 만성질환 수와 연간 가구 총소득에 대해 살펴보았다. 2021년 연간데이터 기준 시도별 및 동부·읍면부별 만성질환 수의 상대표준오차는 전체적으로 25% 이하로 안정적으로 유지되고 있었다. 그러나 부산광역시 읍면부, 대구광역시 읍면부와 울산광역시 읍면부 층에서는 25%를 초과하는 것으로 나타났다. 연간 가구 총소득도 만성질환 수와 유사한 결과를 보였으나, 대구광역시 읍면부만 25%를 초과했으며 나머지 지역은 전반적으로 안정적이라고 볼 수 있다. 현시점에서 최신 모집단인 2021년 인구주택총조사를 활용하여 추가표본을 구축하는 방안을 마련하였다. 최종 목표 표본 규모는 8,400가구로 결정하였고, 이에 따라 추가표본은 2,700가구를 추출할 수 있도록 표본을 설계하였다. 이때 2기 한국의료패널 구축 당시에 활용하였던 층화변수, 표본 배분방법 등은 동일하게 적용하였다. 표본 배분은 1차 시도의 경우 가구수에 따라 제곱근 비례배분방법을, 2차 동부·읍면부의 경우 비례배분방법을 적용하였다. 패널조사의 특성상 패널을 지속적으로 유지하는 것이 중요하므로 목표 가구수보다 조사 완료 가구수가 더 큰 층에서는 추가표본을 추출하지 않고, 부족한 층에 대해서만 표본을 추출하도록 설계하였다. 이때 패널 마모가 심한 4개 지역이 있었는데 부산광역시 읍면부, 대구광역시 읍면부, 울산광역시 읍면부의 패널은 5가구 이하였고, 인천광역시 읍면부는 0가구였다. 이는 앞에서 살펴보았던 제3장 패널가구의 표본이탈 특성에서 대구와 인천의 경우 이탈 비율이 높게 나타난 것과 유사한 결과였다. 인천광역시 읍면부의 경우 충분한 표본을 확보하기 위해 최소 3개 조사구를 배분할 수 있도록 조정하였고, 3개 지역도 목표 가구수에 따라 조사구를 배분하였다. 통합표본에서의 예상 상대표준오차를 만성질환 수와 연간 가구 총소득에 대해 검토한 결과,

표본을 추가할 경우 상당히 개선되는 것으로 나타났다. 표본조사구 추출 시 내재적 층화변수는 층별 조사구 특성, 유배우 인구수, 아파트 거주 가구수를 활용하였다. 2기 한국의료패널 구축 당시에는 조사구 특성, 점유 형태, 혼인상태, 거처 종류를 활용하였으나, 현재 혼인상태와 거처 종류는 사용할 수 없기 때문에 새롭게 선정하였다. 조사구당 8~12가구(평균 10 가구)를 조사 완료하여 최종 추가표본으로 구축하는 것을 목표로 하였다. 이 장에서 제안한 추가표본 구축을 위한 표본추출 방안에 따라 2023년 12월에 조사를 착수할 계획이다.

제2절 패널 유지를 위한 패널 대상의 관리 전략

패널 대상의 이탈 특성에 대해 통계적인 측면에서 다각적으로 살펴 보았다. 이러한 연구 결과를 기반으로 패널 대상의 이탈을 최소화하기 위해서 패널 관리가 집중적으로 필요한 부분에 대해 네 가지를 제안하고자 한다.

첫째, 이사하는 패널가구에 대한 꾸준한 추적 관리가 필요하다. 이사로 인해 패널가구에 대한 추적이 불가능한 비율이 점차 증가하는 추세를 확인하였다. 하지만 이사 변동에 대한 추적이 관리될 수 있다면 패널에서 이탈할 가능성이 낮은 편이라는 분석 결과를 얻었다. 이는 선행 연구와 유사한 결과로, 주소 변경을 파악할 수 있다면 일시적으로 이탈이 발생 하더라도 재복귀가 가능하다는 분석과 일맥상통한다고 볼 수 있다.

둘째, 아파트에 거주하는 패널가구에 대한 패널 관리가 필요하다. 패널 가구의 주택유형 분포는 아파트와 단독주택이 높은 편이었으나, 시간이 지남에 따라 아파트의 비율은 낮아지고 단독주택은 높아지는 추세를

보였다. 아파트에 거주하면 이탈 비율이 높은 편이었으나, 단독주택은 낮은 편에 속하였다. 또한, 패널가구 이탈에 영향을 주는 요인에서도 아파트가 단독주택에 비해 가능성이 있는 것으로 나타났다.

셋째, 중년의 부모와 자녀로 구성된 가족에 대한 패널 관리가 필요하다. 패널가구와 패널가구원 모두 무응답한 집단은 0세 이상~19세 이하 및 40세 이상~49세 이하인 연령대이며, 미취학이 포함된 무학과 대졸 이상이며, 기타를 제외한 나머지 경제활동 상태의 특성을 지녔다. 또한, 3인 가구 이상의 응답 분포는 점점 줄어드는 양상을 보였고, 15세 미만 응답자의 비율도 점차 감소하는 추세였다.

마지막으로 연령이 젊은 패널가구원과 혼인상태가 미혼인 패널가구원에 대한 패널 관리가 필요하다. 패널가구와 패널가구원 모두 무응답한 집단의 인구사회학적 특성을 보면 미혼이 다른 혼인상태에 비해 조사에 참여하지 않을 가능성이 높은 편으로 나타났다. 또한, 패널가구원의 이탈에 영향을 주는 요인으로 패널가구원의 연령이 해당하였다.

패널조사에서 패널 대상의 이탈 증가 현상은 패널조사의 안정성을 위협하는 위험 요인이라고 볼 수 있다. 패널 관리가 집중적으로 필요한 부분에 대해 제안한 네 가지를 중심으로 사전에 대처함으로써 패널 대상의 이탈을 막을 수 있도록 노력해야 할 것이다.



- 김대일, 남재량, 류근관. (2000). 한국노동패널 표본의 대표성과 패널조사 표본 이탈자의 특성 연구. 노동경제논집. 23(3). 1-33.
- 김현지, 김태용, 최효임. (2023). 장애인삶 패널조사 이용자 안내서(1차-4차). 서울: 한국장애인개발원.
- 남기성, 천영민. (2012). An Analysis of Response Pattern and Panel Attrition in KLIPS. 응용통계연구 25(6). 933-945.
- 박은자, 정영, 서제희, 배정은, 이나경, 김은주, ..., 박현아. (2019). 제2기 한국 의료패널 구축운영을 위한 연구. 세종: 한국보건사회연구원.
- 성재민. (2007). 응답자 사례금, 조사에 영향이 있는가. 노동리뷰. 34, 67-75.
- 송헌재. (2012). 재정패널 2차년도 이탈가구와 대체가구의 특성 분석. 재정학 연구. 5(1), 97-133.
- 신선옥. (2008). 한국노동패널조사의 응답자 태도에 면접원이 미치는 효과. 노동리뷰. 37, 74-82.
- 심영상, 허명희. (2002). 한국노동패널조사에서 표본이탈에 대한 통계적 모형화 및 활용. 응용통계. 17, 31-46.
- 이경희, 민인식. (2016). 패널조사 응답지속성에 관한 연구: 한국노동패널조사를 중심으로. 조사연구. 17(3), 1-24.
- 이상호, 이혜정. (2017). 한국노동패널조사(KIPS)의 패널이탈 분석-비단조 이탈을 중심으로. 2017 한국노동패널학술대회, 세종: 한국노동연구원.
- 이상협, 박찬용, 정성석, 최혜미. (2011). 한국노동패널 탈락 분석. 한국데이터정보과학회지. 22(1), 1-8.
- 이지연, 김진. (2006). 횡단조사자료 종단화의 가치와 한계: 경제활동인구조사와 도시가계조사. 한국인구학. 29(3), 159-188.
- 이혜정. (2018). 함수적 모형을 이용한 연속형 패널자료의 항목무응답 대체 방법. (박사학위논문). 서울: 고려대학교 대학원.
- 정영호, 고숙자, 이은영, 진달래, 김성욱, 한준태, 이수연. (2009). 2008년 한국

의료패널 기초분석보고서 (I). 세종: 한국보건사회연구원.

주재선, 임연규. (2018). 여성가족패널 이탈 분석과 표본추가 영향 분석. 2018 한국여성정책연구원 패널조사 학술대회. 서울:한국여성정책연구원.

정은희, 신재동, 우선희, 하은솔, 김지원, 박나영, ..., 김진. (2022). 2022년 한국 복지패널 조사·분석 보고서. 세종: 한국보건사회연구원.

Branden, L., Gritz, R. and Pergamit, M. (1995). The Effect of Interview Length on Attrition in the National Longitudinal Survey of Youth. No. NLS 95-28, US Department of Labour.

Lynn, P., Buck, N., Burton, J., Jäckle, A., & Laurie, H. (2005). A review of methodological research pertinent to longitudinal survey design and data collection. *IDEAS Working Paper Series from RePEc*, IDEAS Working Paper Series from RePEc, 2005.

Uhrig, S. (2008). The nature and causes of attrition in the British Household Panel Study. *IDEAS Working Paper Series from RePEc*, IDEAS Working Paper Series from RePEc, 2008.

Watson, D. (2003). Sample Attrition between Waves 1 and 5 in the European Community Household Panel. *European Sociological Review*, 19(4), 361-378.

Zabel, J. (1998). An Analysis of Attrition in the Panel Study of Income Dynamics and the Survey of Income and Program Participation with an Application to a Model of Labor Market Behavior. *The Journal of Human Resources*, 33(2), 479-506.

<웹페이지>

한국노동패널 홈페이지. 조사결과.

<https://www.kli.re.kr/menu.es?mid=a40204010000>에서
2023.08.10. 인출.

한국복지패널 홈페이지. 조사결과.

<https://www.koweps.re.kr:442/probe/result.do>에서 2023.08.10.
인출.

한국의료패널 홈페이지. 조사결과.

<https://www.khp.re.kr:444/web/survey/results.do>에서
2023.08.10. 인출.

〈데이터〉

통계청. (2022). 2021년 인구총조사[데이터파일]. mdis 원격접근서비스 활용
하여 분석.

<https://mdis.kostat.go.kr/index.do>에서 2023. 7. 1. 인출.

한국보건사회연구원. (2023). 한국의료패널조사 내부자료.

한국보건사회연구원·국민건강보험공단. (2023). 2기 한국의료패널조사 2019년
~2021년 연간데이터[데이터파일].

<https://www.khp.re.kr:444/>에서 2023. 5. 1. 인출.