

정책분석과 동향

한국의료패널조사 패널가구의 표본이탈 현황과 시사점

|이해정|



한국보건사회연구원
KOREA INSTITUTE FOR HEALTH AND SOCIAL AFFAIRS

한국의료패널조사 패널가구원의 표본이탈 현황과 시사점¹⁾

Characteristics and Implications of Sample Attrition in the Korea Health Panel Survey

이혜정 한국보건사회연구원 사회보장재정데이터연구실 부연구위원

패널조사는 보통 조사 차수가 거듭될수록 패널 대상이 이탈하는 현상이 발생하기 마련이다. 패널 대상의 이탈이 일어나면 표본이 모집단을 대표하지 못하게 됨에 따라 추정의 편의가 발생하는 등 잘못된 분석 결과를 도출할 수 있게 되어 패널조사에 대한 신뢰도 하락으로 이어질 수 있다. 이에 표본이탈이 발생하기 전에 원인을 파악하고 지속적으로 관리해야 한다.

이 글에서는 2기 한국의료패널조사 패널가구원의 응답 형태 현황 및 특성 분석, 응답 형태에 따른 영향 분석 등 통계적인 측면에서 다각적으로 살펴보았다. 이러한 연구 결과를 기반으로 패널 대상의 이탈을 최소화하기 위해서 패널 관리가 집중적으로 필요한 부분에 대해 네 가지를 제안하였다.

1 들어가며

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 대상의 이탈이 일어나면 표본이 모집단을 대표하지 못하게 됨에 따라 추정의 편의가 발생하

1) 이 글은 이혜정, 변중석, 황안나, 염아림. (2023). 패널조사 표본이탈 경향 분석과 추가표본 추출에 관한 연구-한국의료패널조사를 중심으로(한국보건사회연구원)에 수록된 내용을 바탕으로 재구성하여 작성한 것이다.

는 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 등이 있다. 패널 대상자의 인구사회학적 특성을 살펴보면 남성이 여성에 비해, 연령이 젊을수록, 학력이 높을수록, 배우자가 없는 경우, 임금근로자일수록 표본이탈의 경향이 높다는 결과를 보였다(이혜정 외(2023) 재인용. pp. 17-18.). 패널 가구의 가구주가 여성이거나, 연령이 젊을수록, 학력이 높을수록 표본이탈이 발생할 가능성이 높다고 나타났다(이혜정 외(2023) 재인용. pp. 17-18.). 또한 이사, 주거 형태가 월세인 경우 자가에 비해 표본이탈의 위험을 높인다고 하였다(이혜정 외(2023) 재인용. pp. 17-18.). 다음 방안은 패널조사가 일정한 시점이 지나면 표본을 추가하는 것이다. 1기 한국의료패널조사에서는 2007년에 원표본을 추출한 이후

2012년에 추가표본을 구축하였다.²⁾ 한국복지패널조사에서는 7차 조사(2012년)에서 표본을 추가³⁾하였고, 최근 표본 추가를 위한 사전 작업을 완료하였다(정은희 외, 2022). 한국노동패널조사에서도 12차 조사(2009년)와 21차 조사(2018년) 두 번에 걸쳐 표본을 추가하였다.⁴⁾ 보통 추가표본을 구축하는 목적은 마모된 표본을 보완하는 데 있다. 그래서 현시점에서의 패널 구조 파악, 조사 결과 분석 등을 기반으로 패널에 대해 진단하고, 표본이탈로 인해 추가표본이 필요한 부분을 보완하는 방향으로 추가표본을 구축하게 된다.

한편 한국의료패널조사는 2018년부터 2기 패널을 구축하기 위한 준비를 하였고, 2019년 사전 조사를 시작으로 매해 조사를 하고 있으며, 현재 2024년 조사를 진행하고 있다. 2기 한국의료패널조사도 조사 차수가 상당히 지남에 따라 현시점에서의 표본 구성에 대한 검토가 필요한 상황이라고 볼 수 있다.

이 글에서는 최근 조사가 완료된 데이터를 활용하여 2기 한국의료패널의 조사 완료 현황을 파악하고, 응답 형태에 따른 패널가구의 표본이탈 특성을 체계적으로 분석하고자 한다.⁵⁾ 응답 형태로

2) 한국의료패널. (n.d.). 표본추출 및 패널특성. <https://www.khp.re.kr:444/web/survey/sampling.do>

3) 한국복지패널. (n.d.). 표본추출. <https://www.koweps.re.kr:442/probe/sample.do>

4) 한국노동패널조사. (n.d.). 표본추출. <https://www.kli.re.kr/menu.es?mid=a40201000000>

5) 한국의료패널조사는 가구를 추출하여 해당 가구와 가구 내 가구원을 모두 조사하는 방식으로 가구의 표본이탈에 대한 특성을 살펴보는 것도 중요하다. 이 원고에서는 가구원의 표본이탈에 초점을 두었으나, 가구의 표본이탈 분석 결과는 '패널조사 표본이탈 경향 분석과 추가표본 추출에 관한 연구-한국의료패널조사를 중심으로' 보고서를 참고하기 바란다.

구분하여 패널가구원의 인구사회학적 특성에 따른 분포를 살펴본다. 응답 형태는 2019~2021년 연간 데이터를 사용하여 패널가구원의 응답 여부를 조합하여 연속응답, 2020년 이탈, 2021년 이탈, 재진입 4개 집단으로 구분한다. 또한 표본이탈에 영향을 미치는 요인을 살펴보기 위해서 응답 형태를 종속변수로 하여 다항 로지스틱 회귀모형을 분석한다.

2 2기 한국의료패널조사 개요 및 완료 현황

한국의료패널은 기존 행정 자료로 파악하기 어려운 개인과 가구 단위의 의료비 지출과 의료 이용 결정 요인 등을 심층적으로 파악하는 데 목적이 있다. 1기 한국의료패널조사가 2008년 처음 시작하여 2019년을 마지막으로 마쳤다. 이후 여

러 가지 검토 과정을 가진 후에 2기 한국의료패널 조사를 착수하였다. 주요 개편 방향은 2016년 등록 센서스를 기준으로 표본을 구축하여 표본의 대표성을 확보하는 것이었다. 그리고 매년 조사하는 항목과 주기적으로 조사하는 항목을 구분하고, 불필요한 설문 내용을 최소화하는 등으로 조사의 효율성을 제고하려고 하였다. 또한 환자 경험, 영양 및 간병 등에 대한 조사 내용을 추가하여 정책 활용도를 높이하고자 하였다.

2기 한국의료패널의 목표 모집단은 2018년 전국 17개 시도의 일반거주시설에 거주하는 일반가구 및 가구원이다. 표본 추출틀은 2016년 등록 센서스를 활용하였는데, 조사구 중에서 기숙사, 특수 시설 조사구, 섬지역 조사구를 제외한 조사구에 거주하는 일반가구가 표본 추출 대상이다. 표본 조사구 추출 방법은 2단계 층화집락추출 방법이다. 1차 층화 변수는 17개 시도로 지역별 가

[표 1] 2기 한국의료패널조사 연간 데이터 기준 완료 현황

(단위: 가구, 명)

	전체	
	가구	가구원
유치 조사	8,516	21,684
2019년 연간 데이터	6,748	16,587
2020년 연간 데이터	6,217	14,844
2021년 연간 데이터	5,907	13,799

출처: “패널조사 표본이탈 경향 분석과 추가표본 추출에 관한 연구-한국의료패널조사를 중심으로”, 이혜정 외, 2023, 한국보건사회연구원, p. 27. <표 2-1>.

구 규모에 따른 제공근비례 배분과 설계 효과를 고려하였다. 이때 설계 효과는 만성질환 수를 기준으로 하였다. 2차 층화 변수는 동부·읍면부에 따라 비례 배분하여 708개의 표본 조사구를 추출하였다. 그런 다음 조사구당 12가구를 계통 추출하여 가구와 가구 내 거주하는 가구원을 조사함에 따라 최종 8500가구를 패널로 구축하는 것을 목표로 하였다.

2기 한국의료패널조사 연간 데이터 기준 완료 현황을 살펴보면 <표 1>과 같다.⁶⁾ 유치 조사(8516가구)의 79.2%가 2019년 연간 데이터(6748가구)로 최종 구축되었다. 1기 한국의료패널의 경우 가구의 이탈을 고려하여 목표 가구인 7000~8000가구보다 과대 표집하여 9015가구

를 유지하였다. 이러한 과정을 거쳐 약 6개월 이후에 다시 방문한 7866가구를 한국의료패널의 원표본 가구로 구축하였는데,⁷⁾ 유치 조사의 87.3%가 2008년 원표본 가구였다. 이러한 비율의 차이는 1기의 경우 유치 조사 이후 가구를 재방문한 기간이 짧은 편이었고, 과대 표집 등을 했다는 점에 있는 것으로 보인다.

<표 2>는 패널가구의 응답 현황과 원표본 유지율이다. 2021년 연간 데이터의 원표본 유지율이 86%로 나타나 1기보다 높은 수준을 유지하고 있다. 1기 한국의료패널 데이터의 3차 연도 원표본 유지율은 77.4%⁸⁾였다. 한편 한국노동패널의 09 통합표본과 18통합표본과 비교하면 원표본 유지율이 낮게 나타났다. 그러나 한국복지패널과 장애

[표 2] 2기 한국의료패널조사 연간 데이터 기준-원표본 유지율

(단위: 가구, %)

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

주: 1) 원표본 유지율은 2019년 원표본을 기준으로 계산함.

2) 2020년에는 분가 가구 54가구가 응답하여 전체 응답 가구수는 6217가구임.

3) 2021년에는 분가 가구 103가구가 응답하여 전체 응답 가구수는 5907가구임.

출처: “패널조사 표본이탈 경향 분석과 추가표본 추출에 관한 연구-한국의료패널조사를 중심으로”, 이해정 외, 2023, 한국보건사회연구원, p. 28. <표 2-2>

6) 이 연구에서 분석에 활용한 데이터는 2019~2021년 연간 데이터이므로 이를 기준으로 작성하였다.

7) 정영호 외. (2009). 2008년 한국의료패널 기초분석보고서 (1). 한국보건사회연구원. p. 31.

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

인살패널의 3차 연도 원표본 유지율과는 비슷한 수준이라고 볼 수 있다.⁹⁾

3 패널가구원의 응답 형태 현황 및 특성 분석

패널가구원의 응답 형태 현황 및 특성에 대해 2019~2021년 연간 데이터를 대상으로 하여 살펴 보았다. 분석 대상은 분가 가구를 제외한 원가구단

을 대상으로 하였으며 개인정보 활용에 대해 미동 의한 가구원을 포함한 가구 및 가구원은 제외하였 다. 설문조사에 참여하여 응답했으나 개인정보 제 공에 동의하지 않아서 응답값을 결측으로 처리하였 기 때문이다. 해당 연도에서 ‘패널가구원’이 조사에 참여하면 ‘1’이고 조사에 참여하지 않으면 ‘0’으로 정의하였다. ‘연속응답’ 가구원은 3개 연도 모두 조 사에 참여하였는데, 기호는 111이 된다. ‘이탈’ 가 구원은 2019년 조사에 참여하였으나 이후 조사에

9) 타 패널조사 원표본 유지율은 다음과 같다. 3번째 조사에 대한 원패널 유지율을 살펴보면 1기 한국의료패널은 77.4%, 한국복지패널 은 86.7%, 한국노동패널은 80.9%(98표본), 90.5%(09표본), 92.7%(18표본), 장애인살패널은 85.9%로 나타났다.

〈표〉 2019년 기준 원표본 유지율

(단위: 가구, %)

조사	조사 연도	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	
	2010년 하반기	6,089	1,777	7,866	77.4	
한국복지패널	2006년	7,072	-	7,072	100	
	2007년	6,511	561	7,072	92.1	
	2008년	6,128	944	7,072	86.7	
한국노동패널	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
		2018년	6,121	-	6,121	100
장애인살패널	2019년	5,527	594	6,121	90.3	
	2020년	5,259	862	6,121	85.9	

주: 1) 한국의료패널: 한국의료패널 홈페이지. 조사 결과. <https://www.khp.re.kr:444/web/survey/results.do>
 2) 한국복지패널: 한국복지패널 홈페이지. 조사 결과. <https://www.koweps.re.kr:442/probe/result.do>
 3) 한국노동패널: 한국노동패널 홈페이지. 조사 결과. <https://www.kli.re.kr/menu.es?mid=a40204010000>
 4) 장애인살패널: 김현지 외. (2023). 장애인살 패널조사 이용자 안내서(1차~4차). 한국장애인개발원. p. 26.

참여하지 않은 경우로, 기호는 100(‘2020년 이탈’) 또는 110(‘2021년 이탈’)이 된다. 마지막으로 2019년 조사에 참여하고 2021년 조사에 다시 참여한 경우를 ‘재진입’ 가구원이라고 정의하였다(기호: 101).

〈표 3〉을 보면 패널가구원의 응답 형태는 연속응답이 81.5%로 가장 높았다. 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 패널가구의 응답 형태와 비교하면 패널가구원의 연속응답 비율이 더 낮았고(패널가구 85.7%), 첫 번째와 두 번째에서의 이탈 비율이 더 높았다(패널가구 8.4%, 5.6%). 이는 패널가구원의 조사 참여에 대한 충성도가 높지 않다고 볼 수 있다. 조사 참여를 높이기 위한 방안을 마련할 필요가 있다.

패널가구원의 응답 형태에 따른 성별 분포를 보면 연속응답의 비율은 여성이 83.8%로 남성(83.1%)에 비해 0.7%포인트 높게 나타났다(표

[표 3] 패널가구원의 응답 형태별 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.

출처: “패널조사 표본이탈 경향 분석과 추가표본 추출에 관한 연구-한국의료패널조사를 중심으로”, 이혜정 외, 2023, 한국보건사회연구원, p. 58. (표 3-9).

[표 4] 패널가구원의 응답 형태별 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’, 100은 ‘2020년 이탈’, 110은 ‘2021년 이탈’, 101은 ‘재진입’을 의미함.

출처: “패널조사 표본이탈 경향 분석과 추가표본 추출에 관한 연구-한국의료패널조사를 중심으로”, 이혜정 외, 2023, 한국보건사회연구원, p. 58. (표 3-10).

4). 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.

다음 <표 5>는 패널가구의 응답 형태에 따른 연령대 분포로 연속응답의 비율은 50세 이상~79세 이하가 높은 편, 49세 이하가 낮은 편에 속하였다. 이와 반대로 2020년과 2021년 이후 이탈은 50세 이상~79세 이하가 낮은 편, 49세 이하가 높

[표 5] 패널가구의 응답 형태별 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', 100은 '2020년 이탈', 110은 '2021년 이탈', 101은 '재진입'을 의미함.

출처: "패널조사 표본이탈 경향 분석과 추가표본 추출에 관한 연구-한국의료패널조사를 중심으로", 이해정 외, 2023, 한국보건사회연구원, p. 59. <표 3-11>.

[표 6] 패널가구의 응답 형태별 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', 100은 '2020년 이탈', 110은 '2021년 이탈', 101은 '재진입'을 의미함.

2) 응답 거부 2명은 제외하고 분석하였음.

출처: "패널조사 표본이탈 경향 분석과 추가표본 추출에 관한 연구-한국의료패널조사를 중심으로", 이해정 외, 2023, 한국보건사회연구원, p. 60. <표 3-12>.

은 편으로 나타났다. 80세 이상의 경우에는 연속응답 비율이 낮고 이탈 비율이 높은 경향을 보였는데, 이는 노환, 사망 등으로 인한 표본이탈로 볼 수 있다.

패널가구원의 응답 형태에 따른 혼인상태 분포를 보면 연속응답은 사별·실종(87.8%), 혼인 중(84.6%), 별거·이혼(83.6%)이 높게, 미혼은

79.4%로 가장 낮게 나타났다(표 6). 미혼의 경우 다른 혼인상태에 비해 표본이탈 경향이 높다고 볼 수 있다. 미혼 다음으로 별거·이혼인 경우도 2020년과 2021년 이탈 비율(각각 9.5%, 6.5%)이 높은 편으로 나타났다.

다음 <표 7>은 패널가구원의 응답 형태에 따른

[표 7] 패널가구원의 응답 형태별 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', 100은 '2020년 이탈', 110은 '2021년 이탈', 101은 '재진입'을 의미함.

2) 무학은 '미취학도 포함'하여 분석하였음.

출처: "패널조사 표본이탈 경향 분석과 추가표본 추출에 관한 연구-한국의료패널조사를 중심으로", 이혜정 외, 2023, 한국보건사회연구원, p. 60. <표 3-13>.

[표 8] 패널가구원의 응답 형태별 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', 100은 '2020년 이탈', 110은 '2021년 이탈', 101은 '재진입'을 의미함.

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

3) 해당 없음은 15세 미만 응답자임.

출처: "패널조사 표본이탈 경향 분석과 추가표본 추출에 관한 연구-한국의료패널조사를 중심으로", 이혜정 외, 2023, 한국보건사회연구원, p. 61. <표 3-14>.

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.

마지막으로 <표 8>은 패널가구원의 응답 형태에 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.

4 패널가구원의 응답 형태에 따른 영향 분석

패널가구원의 응답 형태에 따른 집단을 종속변수로 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

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

<표 9>는 모형의 분석 결과로, 연속응답을 준거 집단으로 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. 주거 형태가 단독주택(기준변수: 아파트)인 경우 2020년 이탈 집단일 가능성이 연속응답 집단일 가능성보다 25.8% 감소하고, 다

[표 9] 패널가구원의 응답 형태에 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
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
주거 형태_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_중소득						
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) 주거 형태에서 '기타'는 영업용 건물 내 주택, 오피스텔, 호텔, 여관, 기숙사 등이 해당함.

5) 연간 가구 총소득에서 저소득은 0~2000만 원 미만, 중소득은 2000만~5000만 원 미만, 고소득은 5000만 원 이상을 의미함.

출처: "패널조사 표본이탈 경향 분석과 추가표본 추출에 관한 연구-한국의료패널조사를 중심으로", 이혜정 외, 2023, 한국보건사회연구원, pp. 70-71. (표 3-21).

세대주택인 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 연속 응답을 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 비경제활동인구(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. 주거형태가 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 가능

성이 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, 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.¹⁰⁾ 한편 주거 형태의 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. 이 외에도 2021년 이탈 집단은 남성이 여성에 비해, 연간 가구 총소득의 경우 저소득이 중소득에 비해 표본이탈에 영향을 주는 것으로 나타났다.

10) 패널가구의 표본이탈 특성 분석 내용은 이혜정 외. (2023). 패널조사 표본이탈 경향 분석과 추가표본 추출에 관한 연구-한국의료패널조사를 중심으로(한국보건사회연구원)를 참고(pp. 47-56)하기 바란다.

5 나가며

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 한다. 2기 한국의료패널조사 패널 대상자의 현황 및 표본이탈의 경향을 통계적인 측면에서 다각적으로 살펴보았다.

패널가구원의 표본이탈 특성을 주요 변수를 중심으로 살펴보았다. 성별의 경우 2020년 이탈에서는 남성과 여성의 비율이 비슷했으나, 2021년 이탈에

서는 남성이 여성에 비해 높게 나타났다. 연령대의 경우 50세 이상~79세 이하가 표본이탈 비율이 낮은 편으로, 49세 이하가 높은 편으로 나타났다. 80세 이상의 경우에는 연속응답 비율이 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. 무학(미취학 포함)의 경우도 높게 나타나 고학력 부모가 표본이탈하면서 미취학인 자녀도 함께 이탈한 것으로 추정할 수 있다. 경제활동 상태에 따라서는 15세

[표 10] 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 (+)
주거 형태_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) 주거 형태에서 기타는 영업용 건물 내 주택(상가, 공장 등), 오피스텔, 호텔, 여관, 기숙사 등을 의미함.

4) 연간 가구 총소득에서 저소득은 0~2000만 원 미만, 중소득은 2000만~5000만 원 미만, 고소득은 5000만 원 이상을 의미함.

출처: "패널조사 표본이탈 경향 분석과 추가표본 추출에 관한 연구-한국의료패널조사를 중심으로", 이혜정 외, 2023, 한국보건사회연구원, p. 76. <표 3-23>.

미만 응답자인 '해당 없음'과 비경제활동인구의 표본이탈 비율이 높은 데 비해 기타와 고용주·자영업자의 비율은 낮았다.

패널가구원의 응답 형태에 따른 3개 집단(연속 응답, 2020년 이탈, 2021년 이탈)을 종속변수로 하여 다항 로지스틱 회귀모형을 분석하였다. 사용한 설명변수는 2019년 연간 데이터에서의 가구원 특성(성별, 연령, 혼인상태, 최종 학력, 경제활동 상태), 2019년에서의 가구 특성(거주 지역, 주거 형태, 가구원 수, 연간 가구 총소득), 그리고 2019~2021년에 걸친 가구의 상태 변화(이사 횟수)이다. <표 10>은 이 분석 결과를 요약한 것으로, 연속응답을 준거집단으로 하여 이탈 집단에 영향을 주는 요인에 관한 것이다. 즉 설명변수 중에서 통계적으로 유의한 결과를 나타난 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. 이사회 횟수는 패널가구의 표본이탈에 영향을 주는 요인에도 해당되었는데, 패널가구와 패널가구원에서 동일한

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

이러한 연구 결과를 기반으로 패널 대상의 이탈을 최소화하기 위해 패널 관리가 집중적으로 필요한 부분에 대해 네 가지를 제안하고자 한다.¹¹⁾¹²⁾

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 선행연구와 유사한 결과로, 주소 변경을 파악할 수 있다면 일시적으로 이탈이 발생하더라도 재복귀가 가능하다는 분석과 일맥

11) 패널가구의 표본이탈 특성 분석 내용은 이혜정 외. (2023). 패널조사 표본이탈 경향 분석과 추가표본 추출에 관한 연구-한국의료패널조사를 중심으로(한국보건사회연구원)를 참고(pp. 47-56)하기 바란다.

12) 패널 대상의 표본이탈 특성 및 표본 현황 분석 결과를 고려하여 추가표본 구축을 위한 표본추출 방안을 마련하였음. 이에 따라 2700가구 추가 패널 구축을 목표로 2023년 12월 조사를 완료하였고, 이를 통해 2기 한국의료패널조사의 표본 마모 및 변화된 모집단 특성을 반영하게 되어 안정적인 대표성 있는 패널을 유지할 수 있다고 생각한다.

상통한다고 볼 수 있다.

둘째, 아파트에 거주하는 패널가구에 대한 패널 관리가 필요하다. 패널가구의 주택 유형 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 이상, 기타를 제외한 나머지 경제활동 상태의 특성을 지녔다. 또한 3인 가구 이상의 응답 분포는 점점 줄어드는 양상을 보였고, 15세 미만 응답자의 비율도 점차 감소하는 추세를 나타냈다.

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 한다. 䄠

참고문헌

이혜정, 변종석, 황안나, 염아림. (2023). **패널조사 표본 이탈 경향 분석과 추가표본 추출에 관한 연구-한국의료패널조사를 중심으로**. 한국보건사회연구원.

정영호, 고숙자, 이은영, 진달래, 김성옥, 한준태, 이수연. (2009). **2008년 한국의료패널 기초분석보고서 (1)**. 한국보건사회연구원.

정은희, 신재동, 우선희, 하은솔, 김지원, 박나영, ..., 김건. (2022). **2022년 한국복지패널 조사·분석 보고서**. 한국보건사회연구원.

한국노동패널조사. (n.d.). 표본추출. <https://www.kli.re.kr/menu.es?mid=a40201000000>

한국복지패널. (n.d.). 표본추출. <https://www.koweps.re.kr:442/probe/sample.do>

한국의료패널. (n.d.). 표본추출 및 패널특성. <https://www.khp.re.kr:444/web/survey/sampling.do>

Characteristics and Implications of Sample Attrition in the Korea Health Panel Survey

Lee, Hyejung

(Korea Institute for Health and Social Affairs)

Panel surveys are prone to undergoing respondent dropout as the waves progress over time. Panel attrition in effect reduces the representativeness of the sample and can lead to inferential bias and erroneous conclusions, thereby rendering the credibility of the surveys compromised. Therefore, it is important to identify the causes of attrition and hold them in check in advance on an ongoing basis. In this article, I examined data from the 2nd Korea Health Panel Survey and analyzed the characteristics and response types of participating household members. I then statistically explored the effect of each different response-type group from various angles. Based on the findings, I suggest four areas for focused panel management to minimize sample attrition.