

한국복지패널의 가중치 및 표본 특성¹⁾



Weighting and Sample Characteristics of the Korea Welfare Panel

오미애 | 한국보건사회연구원 연구위원

신재동 | 한국보건사회연구원 전문원

해마다 동일한 조사를 반복하는 패널조사의 특성상 패널 마모는 불가피하다. 한국복지패널 역시 원표본 유지율이 60%대로 낮아져 표본의 대표성 문제 및 표본 보완의 필요성이 제기되고 있다. 이에 따라 표본 관리 및 모집단 추정 정확도를 기하기 위한 마모패턴별 가구 특성 비교 분석 연구가 필요하다. 이러한 배경에서 가중치 산출 과정, 마모패턴별 가구 특성 비교, 가중치 변화 분석을 실시하였다. 분석 결과, 저소득 가구는 패널에 잔류할 가능성이 큰 반면, 일반 가구 혹은 3인 이상 가구는 패널에서 이탈할 가능성이 컸다. 시간에 따른 패널 마모로 인하여 가중치값이 증가함을 알 수 있고, 저소득 가구 가중치 증가량은 일반 가구에 비해 적음을 알 수 있다.

패널 연구에서 대표성을 확보하기 위해서는 마모패턴 분석 및 가중치 산출 과정에 대한 검토가 지속적으로 이루어져야 할 것이다. 한국복지패널과 관련해서는 품질 개선을 포함한 다양한 통계적 이슈를 다룰 수 있는 연구가 지속되어야 하겠다. 여러 이슈들에 대한 한국복지패널 연구가 활발히 이루어진다면, 기존 데이터 생산 및 관리 사업에 더해 향후 한국복지패널 조사 연구의 안정적인 수행에 기여할 수 있으며, 데이터 품질을 개선하고 향상시키는 데 중요한 역할을 할 수 있을 것이다.

1. 들어가며

한국복지패널조사(KOWEPS: Korea welfare

panel study)는 2006년 1차 조사를 시작으로 2019년에 14차 조사를 완료하였다. 1차 조사에 서는 7072가구가 조사되었고 6차인 2011년 조

1) 이 글은 오미애, 이해정, 신재동, 이계오, 박승환, 손창균(2019), 『2019년 한국복지패널 심층분석』에 수록된 내용을 일부 포함하고 있으며, 분석은 13차 조사까지의 결과를 바탕으로 이루어졌다.

사까지 5000가구 이상의 원패널 규모를 유지하여 약 75%의 원패널 유지율을 보였다. 그러나 조사 거부 및 자연 손실 등으로 원표본이 지속적으로 감소하는 데다 표본 탈락으로 인해 패널 표본 가구 분포가 치우치는 문제가 제기됨에 따라, 이러한 문제점을 개선하여 최초 구축 당시의 표본 규모와 대표성을 유지할 수 있도록 하기 위하여 7차 연도인 2012년에 1800가구의 신규 패널을 추가하였다.

패널조사는 해마다 동일한 조사를 반복하는 특성상 표본 마모가 불가피하다. 한국복지패널 13차 조사 기준 원표본 규모는 4997가구로 원표본 유지율이 60.32%로 낮아져 표본의 대표성 문제 및 표본 보완의 필요성이 제기됨에 따라, 표본 관리 및 모집단 추정 정확도를 기하기 위한 표본 특성 및 가중치 관련 연구가 필요한 상황이다. 이 글에서는 한국복지패널의 가중치에 대해 설명하고, 한국복지패널의 표본 특징을 살펴보고, 한국복지패널의 마모패턴별 가구 특성을 비교 분석하고자 한다.

2. 한국복지패널의 가중치

가. 한국복지패널의 표본 규모

1) 기존 패널 - 원패널(원표본) 가구 구축 과정

한국보건사회연구원에서는 ‘2006년 국민생활실태조사’에 필요한 표본 약 3만 가구를 조사하기 위해 2005년 인구주택총조사 조사구의 90%인 23만여 개 조사구 중 517개 조사구를 지역별

조사구 규모에 따라 층화 확률 비례 추출하였다.

전체 517개 표본 조사구 중에서 수해와 같은 천재지변으로 조사가 불가능한 지역을 제외하고 487개 조사구에 대한 조사를 완료하였다.

조사 대상 지역은 섬을 제외한 전국의 각 시도이며, 제주도를 포함한다. ‘2006년 국민생활실태조사’에서 최종적으로 조사가 완료된 2만 4711가구의 소득을 기준으로 7000가구를 2단계 표본추출하여 최종적으로 7072가구를 패널로 구축하였다.

1단계 표집 자료인 ‘2006년 국민생활실태조사’ 자료를 바탕으로 패널 가구의 대표성을 확보하기 위해 중위소득 60% 이하인 저소득층 가구 3500가구와 중위소득 60% 이상인 일반 가구 3500가구를 각각 표본으로 추출하여 조사를 수행하였다. 이때 저소득 가구와 일반 가구를 구분하는 기준에 대해서는 2006년 국민생활실태조사를 분석하여 3가지 대안을 고려하였고, 최종적으로 ‘공공부조 전 경상소득’의 중위소득 60%를 기준으로 저소득 가구와 일반 가구를 구분하였다.

소득 규모별로 구분된 2개의 층에서 지역별, 조사구별로 확률 비례 계통 추출에 따라 일반 가구와 저소득 가구를 표본으로 추출하였다. 상대적으로 규모가 작은 저소득층 가구에 대해서는 추출률을 상향 조정하여 일반 가구와 동일한 수준인 3500가구를 표본 가구로 선정하였다.

KOWEPS의 1차 조사에서 구축된 패널 가구를 소득 기준별로 살펴보면 <표 1>과 같다.

표 1. 일반 가구와 저소득 가구의 분류 기준

(단위: 가구, %)

중위소득	경상소득		가처분소득		공공부조 전 경상소득	
	가구 수	백분율	가구 수	백분율	가구 수	백분율
< 40%	2,481	10.00	2,489	10.09	3,477	13.96
< 50%	4,016	16.12	3,880	15.62	4,757	19.04
< 60%	5,227	22.56	5,473	22.25	6,128	24.76

자료: 김미곤, 여유진, 손창균, 김문길, 이봉주, 구인회, 윤민석. (2006). 2006 한국복지패널 기초분석 보고서. p. 80 <표 2-2-3>.

표 2. 한국복지패널조사(KOWEPS) 1차 조사의 지역별 조사구와 가구 분포 현황

(단위: 가구)

지역	조사구 수		일반 가구		저소득 가구		합계	
	표본 조사구 수	완료 조사구	표본 배분 가구 수	완료 가구 수	표본 배분 가구 수	완료 가구 수	표본 배분 가구 수	완료 가구 수
서울	93	93	811	886	506	449	1,317	1,335
부산	33	33	254	277	272	251	526	528
대구	25	25	187	204	227	218	414	422
인천	28	28	228	247	193	190	421	437
광주	15	15	114	123	130	121	244	244
대전	14	14	118	132	89	76	207	208
울산	14	14	120	124	82	79	202	203
경기	76	76	644	706	471	426	1,115	1,130
강원	14	14	102	118	131	118	233	236
충북	14	14	108	113	113	107	221	221
충남	20	20	153	161	168	160	321	321
전북	20	20	138	140	209	207	347	347
전남	19	19	104	109	273	268	377	377
경북	26	26	152	164	339	329	491	494
경남	30	30	229	240	254	248	483	488
제주	5	5	38	45	43	36	81	81
합계	446	446	3,500	3,789	3,500	3,283	7,000	7,072

자료: 김미곤, 여유진, 손창균, 김문길, 이봉주, 구인회, 윤민석. (2006). 2006 한국복지패널 기초분석 보고서. p. 81 <표 2-2-4>.

패널로 구축된 표본 가구를 소득 기준별로 살펴보면 당초 각 층별로 3500가구씩 배분하였으나 조사 거절, 패널 참여 거부 등의 사유로 저소

득층에서는 표본 설계 당시의 3500가구를 약간 밑도는 규모인 3283가구가 패널로 구축되었다. 일반 가구의 경우 평균적으로 100% 이상의

완료율을 보인 반면, 저소득층 가구는 약 94%의 완료율을 보여 저소득층 가구의 조사 거부가 많았던 것으로 나타났다. 완료 가구 수를 기준으로 지역별 목표 오차는 2.7~11.1%로 지역별 변동이 큰 것으로 나타났다. 패널 가구의 탈락률을 최소화하는 것은 종단면 조사인 패널조사에서 중요하게 고려해야 하는 사항으로, 패널 자료의 대표성을 유지하는 데 가장 중요한 부분이기도 하다. 패널조사의 가장 큰 문제점은 시간의 흐름에 따라 패널 대상 가구의 탈락률이 지속적으로 증가한다는 점이다. 이러한 점을 고려할 때 1차 조사 당시 패널 가구를 구축하는 과정에서 거부율과 비협조성을 최소화하는 방향으로 표본 가구를 선정하도록 하였다. 이때 조사구별 표본 가구는 계통 추출을 하여 선정하도록 하였으며, 만일 조사 대상 표본 가구가 조사를 거절할 경우 순서상 다음 가구를 조사 대상 가구로 선정하도록 하였다.

2) 신규 패널 - 신규 패널 가구 구축 과정

한국복지패널의 6차 연도 조사 이후에 원표본 가구 유지율이 감소하는 상황에서 신규 패널 가구를 추가할 필요성이 제기되었다. 저소득층 가구 및 가구원의 분포가 치우치는 현상이 발생하였고, 지역별 표본 규모는 잦은 이주와 탈락 등의 사유로 변동이 발생하였다. 따라서 7차 조사에서는 1차 조사 패널 표본 규모를 유지하고자 약 1800가구를 추가하여 신규 패널을 구축하도록 하였다.

한국복지패널의 경우 1차 조사 당시 인구주택

총조사에서 표본을 직접 추출하지 않고, '2006년 국민생활실태조사' 표본으로부터 이중 추출(double sampling) 방식으로 7072가구를 일반 가구와 저소득층 가구로 구분하여 추출하였다. 신규 패널 추가를 위한 표본추출 방법도 1차 조사와 동일한 방식으로 고려하였다. 패널의 특성을 유지하기 위해 가능한 한 저소득 및 일반 가구 표본을 6차 조사에서 탈락한 비율에 따라 선정하는 방안을 고려하였다. 이 경우 활용 가능한 표본은 '2011년 복지욕구실태조사(1만 5000여 가구 자료)'이다. 수집한 소득 자료를 이용하여 저소득 및 일반 가구를 1단계 표본으로 파악하였기에 복지패널 1차 조사와 동일한 속성의 패널 자료 구축이 용이하며, 현재 패널 가구의 가중치 조정에 맞추어 조정 가능하다는 장점이 있다. 일반 가구와 저소득 가구의 비율도 1차 조사와 동일하게 저소득 가구를 과대 표집하였다. 지역별 표본 배분 또한 1차 조사 당시의 지역별 가구 비율과 유사하게 표본 가구를 배분하여 표본의 대표성을 확보하고자 하였다. <표 3>은 7차 조사 당시 신규 패널의 지역별 표본 배분 결과를 나타낸다. '2011년 복지욕구실태조사'에서 일반 가구와 저소득 가구로 구분하여 지역별 가구 분포에 따라 1800가구를 비례 배분하였다. 서울과 경기도는 표본 가구가 가장 많이 배분된 지역이며, 일반 가구가 저소득 가구에 비해 상대적으로 많이 배분되었다. 대구, 강원, 충북, 충남, 전북, 전남, 경북, 경남 지역은 저소득 가구가 일반 가구에 비해 상대적으로 많이 배분되었다.

표 3. 한국복지패널조사(KOWEPS) 7차 조사 신규 패널 지역별 표본 배분 결과

(단위: 가구)

지역	2011년 복지욕구실태조사 표본			표본 배분		
	일반	저소득	계	일반	저소득	계
전국	10,745	4,298	15,043	900	900	1,800
서울	1,494	343	1,837	125	72	197
부산	821	330	1,151	69	69	138
대구	616	313	929	52	66	118
인천	706	243	949	59	51	110
광주	504	200	704	42	42	84
대전	535	172	707	45	36	81
울산	457	142	599	38	30	68
경기	1,510	342	1,852	126	72	198
강원	443	286	729	37	60	97
충북	490	235	725	41	49	90
충남	482	344	826	40	72	112
전북	524	278	802	44	58	102
전남	501	300	801	42	63	105
경북	600	351	951	50	73	123
경남	751	303	1,054	63	63	126
제주	311	116	427	27	24	51

자료: 이현주, 노대명, 오미애, 전지현, 이봉주, 김태성, ... 이혜림. (2013). 2013년 한국복지패널 기초분석 보고서. p. 39 <표 2-1-4>.

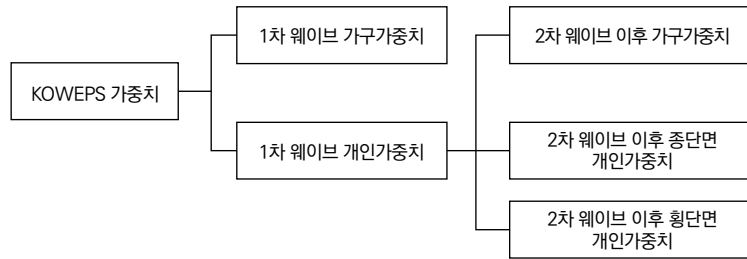
나. 한국복지패널의 가중치 산출 방법

한국복지패널 가중치 산출 방법은 13차 조사 가중치를 기준으로 설명하고자 한다. KOWEPS의 13차 조사 가중치 산정은 기본적으로 12차 조사에 부여된 개인 가중치를 기준으로 가구 가중치와 개인 가중치를 산정하는 방식으로 한다. 기존의 KOWEPS 가중치 부여 체계는 [그림 1]과 같다. 가구 가중치는 조사 치수가 증가함에 따라 가구원의 변동으로 인해 가구의 개념이 1차 조사 때의 정의와 상이해질 수 있으므로 2차 조사 이후의 가중치는 개인 가중치를 중심으로 산

정하도록 한다. 또한 가구는 개인과 달리 치수가 증가함에 따라 생성과 소멸이 반복적으로 이루어지기 때문에 종단면과 횡단면의 구분이 모호해져 이를 구분하는 것은 별 의미가 없다. 따라서 가구 가중치는 개인 가중치를 기준으로 종단면과 횡단면 가중치를 구분하지 않고 단일 가중치를 부여한다.

한국복지패널 7차 조사의 경우 기존의 1~6차 조사에서 구축된 패널 표본과 7차 조사에서 새롭게 추가된 신규 패널을 병합하여 새롭게 가중치를 조정할 필요가 있었다. 먼저 7차 조사의 가중치 조정 과정을 간략히 살펴보면 다음과 같다.

그림 1. 한국복지패널조사(KOWEPS)의 가중치 부여 체계



자료: 김미근, 여유진, 손창균, 김문길, 이봉주, 구인화,...지은정. (2007). 2007 한국복지패널 기초분석 보고서. p. 102 [그림 2-3-2].

1차 조사 이후 7차 조사까지 구축된 기존의 패널 표본의 경우 6차 조사의 응답률을 로지스틱 회귀모형에 적용해 응답 확률을 추정하였고, 6차 조사의 종단면 가중치에 기초하여 7차 조사에서 조사된 개인 종단면 가중치를 조정하였다. 7차 조사에서 추가된 신규 패널은 ‘2011년 복지욕구 실태조사’의 최종 가중치를 기본 가중치로 고려하되 1800가구의 추출 확률의 역수를 곱하여 설계 가중치를 조정하였다.

이와 같이 각각 독립적으로 조정된 가중치에서 신규 패널의 7차 조사 종단면 가중치는 2012년 새롭게 진입한 신규 가구이므로 모두 ‘0’으로 부여하였고, 7차 조사 횡단면 가중치의 경우 기존의 패널과 신규 패널을 병합하여 새롭게 가중치를 부여하였다.

8차 조사의 개인 가중치는 7차 조사의 개인 종단면 가중치를 기준으로 종단면과 횡단면 가중치를 부여한 것이다. 단 7차 조사에서 추가된 신규 패널 1800가구의 가구원은 개인 종단면 가중치

가 0이므로 7차 조사 개인 횡단면 가중치를 기준으로 원표본 패널 가중치와 병합하여 8차 조사의 종단면 가중치와 횡단면 가중치를 산정하였다.

9차 조사의 개인 가중치는 8차 조사의 개인 종단면 가중치를 기준으로 종단면과 횡단면 가중치를 부여한 것이며, 10차 조사의 개인 가중치는 9차 조사의 개인 종단면 가중치를 기준으로 종단면과 횡단면 가중치를 부여한 것이다. 13차 조사의 개인 가중치 역시 12차 조사의 개인 종단면 가중치를 기준으로 종단면과 횡단면 가중치를 부여하는 방식으로 산출하였다. 13차 조사의 개인 가중치에 대한 설명은 다음과 같다.

1) 개인 종단면 가중치 산출 과정

13차 조사의 개인 종단면 가중치는 13차 조사의 종단면 기본 가중치에 대해 무응답 보정과 개인별 변동 상황을 고려하여 조정된 뒤 사후 조정으로 계산한다.

13차 조사의 종단면 기본 가중치는 12차 조사

의 개인 종단면 가중치를 기본 가중치로 한다. 13차 조사의 무응답으로 인한 실제 표본 크기 감소를 반영하기 위한 무응답 보정은 다음과 같다.

13차 조사에서 조사가 완료된 개인의 가중치 조정은 패널 탈락으로 인한 무응답 보정을 위해 13차 조사의 종단면 응답 여부 변수와 12차 조사 변수들의 관계에 대해 로지스틱 회귀모형을 이용하여 응답 확률을 추정하였다. 로지스틱 모형의 설명 변수로는 응답자의 성별, 연령, 지역, 경제활동 상태 변수를 사용하였다. 13차 조사의 종단면 응답 여부 변수는 13차와 12차 조사 모두 응답하면 1, 그렇지 않으면 0의 값을 갖는다.

다음으로 13차 조사의 개인 종단면 기본 가중치에 로지스틱 모형으로 추정된 응답 확률의 역수를 곱하여 13차 조사의 개인 종단면 가중치를 조정한다. 13차 조사의 기본 가중치를 $w_{\text{기본}13\text{차}}$, 로지스틱 회귀 분석을 통해 예측된 예측 응답 확률을 $\hat{p}_{13\text{차}}$ 라고 하면 무응답 보정을 통해 얻어지는 가중치는

$$w_{13\text{차}} = w_{\text{기본}13\text{차}} \times \hat{p}_{13\text{차}}^{-1}$$

로 표현된다.

그리고 개인별 변동 상황에 따라 13차 종단면 가중치를 조정한 뒤, 지역 및 응답자의 인구학적 특성에 따라 2017년 인구 추계값을 이용하여 사후 조정을 하였다.

개인별 변동 상황의 예는 다음과 같다.

1차 조사 패널 가구원 중 군 입대 및 해외여행 등의 사유로 조사 당시 응답하지 않았던 가구원은 가구의 평균 가중치를 적용받고, 결혼 등의 사

유로 원패널 가구에 진입한 가구원은 0의 가중치를 부여받는다. 또한 13차 조사에서 원패널 가구로부터 분가한 가구원은 원래의 가구원 가중치를 부여받지만, 분가한 후 결혼 등의 사유로 새로 진입한 신규 가구의 신규 가구원은 0의 가중치를 부여받는다. 12차 조사에서 탈락했던 가구원이 13차 조사에 재진입한 경우에는 지역별 평균 가중치를 부여하였다.

지나치게 큰 가중치값들은 추정량의 분산을 크게 만들어 추정의 정확도에 영향을 미칠 수 있는데, 이러한 가중치의 과도한 변동을 줄이고자 13차 조사에서는 사후 조정 후 극단 가중치 조정을 하였다.

레이킹을 통한 사후 조정 후 지나치게 큰 가중치들(상위 1%)을 찾아 절단(trimming)한 뒤 절단값을 제외한 나머지 값을 극단 관측치가 속한 지역·성별·연령 내의 관측치에 동일하게 배분하였다.

2) 개인 횡단면 가중치 산출 과정

13차 조사의 개인 횡단면 가중치는 13차 조사의 종단면 기본 가중치에 대해 무응답 보정과 개인별 변동 상황을 고려하여 조정한 뒤, 13차 조사의 종단면 가중치값이 0인 가중치에 대해 가구별 평균 가중치를 적용하고 사후 조정으로 계산한다.

패널 탈락으로 인한 무응답 보정을 위해 13차 조사의 종단면 응답 여부 변수와 12차 조사 변수들의 관계에 대해 로지스틱 회귀모형을 이용하여

응답 확률을 추정하였다.

다음으로 13차 조사의 개인 종단면 기본 가중치에 로지스틱 모형으로 추정된 응답 확률의 역수를 곱하여 13차 조사의 개인 종단면 가중치를 조정한다.

그리고 개인별 변동 상황에 따라 13차 조사의 종단면 가중치를 조정한 뒤, 13차 조사의 종단면 가중치값이 0인 가중치에 대해 가구별 평균 가중치를 적용하고, 지역 및 응답자의 인구학적 특성에 따라 2017년 인구 추계값을 이용하여 사후 조정을 하였다.

13차 조사의 횡단면 가중치에서는 레이킹을 통한 사후 조정 후 13차 조사의 종단면 가중치와 마찬가지로 지나치게 큰 가중치들(상위 1%)을 찾아서 절단한 뒤 절단값을 제외한 나머지의 값을 극단 관측치가 속한 지역·성별·연령 내의 관측치에 동일하게 배분하였다.

3) 가구 가중치 산출 과정

13차 조사의 가구 가중치는 먼저 가구의 패널 진입 차수에 따라 13차 조사 개인 종단면 가중치의 가구 내 평균값을 계산하여 산출하였다. 13차 조사에서 새롭게 진입한 신규 가구의 경우 신규 가구 내 원표본(또는 신규 패널) 가구원들의 가중치 평균을 적용하였고, 원표본(또는 신규 패널) 가구원들만으로 구성된 가구는 이들의 평균으로 가구 가중치를 산출하였다. 다음으로 이와 같이 계산된 가구 가중치를 2017년 추계 가구 수에 따라 지역별로 사후 조정하고, 극단값 조정을 통

해 최종 가중치를 산출하였다.

극단값 조정은 레이킹을 통한 사후 조정 후 지나치게 큰 가중치들(상위 1%)을 찾아 절단한 뒤 절단값을 제외한 나머지 값을 극단 관측치가 속한 지역·가구원 수 내의 관측치에 동일하게 배분하였다.

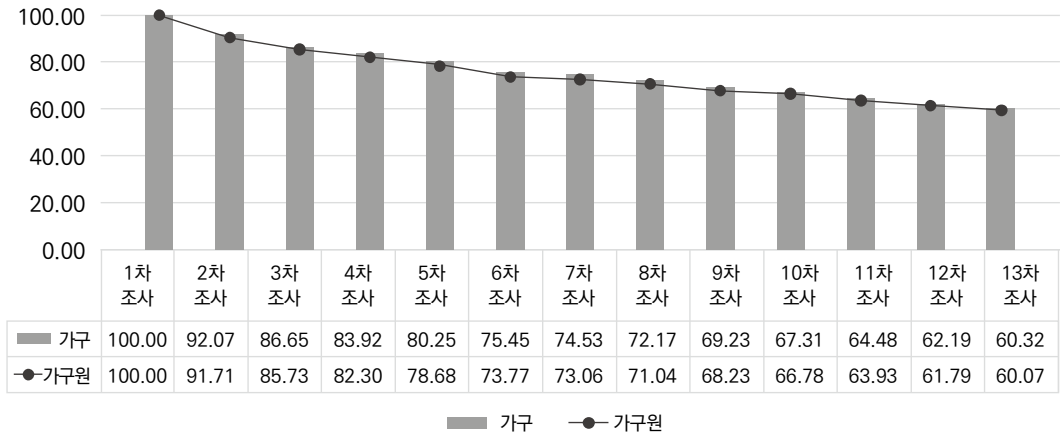
3. 표본 특성

KOWEPS 13차 조사(2018년)에서는 먼저 12차 조사(2017년) 때 조사된 가구와 가구원에 대한 조사를 하였다. 또한 12차 조사 당시 응답하지 않았던 가구 및 가구원에 대해 추가로 조사를 하였다. 따라서 원표본 가구 및 가구원 패널의 유지율을 파악하면 [그림 2]와 같다. 13차 조사에서 1차 조사 표본인 원표본의 유지율은 가구의 경우 60.32%, 가구원은 60.07%로 나타났다. 가구의 경우 전년 대비 약 1.87%포인트 떨어졌으며, 가구원은 1.71%포인트 떨어져 원표본 패널의 13차 조사 탈락률이 높지 않았음을 알 수 있다(그림 2).

KOWEPS 13차 조사에서는 원표본 가구 외에 2012년에 새롭게 조사된 신규 패널 1800가구에 대한 조사도 수행하였는데, 신규 패널 가구 및 가구원 패널의 유지율도 살펴볼 필요가 있다. 신규 패널 가구 및 가구원 패널의 유지율을 살펴보면 [그림 3]과 같다. 신규 패널 가구 유지율은 77.33%이고 가구원 유지율은 73.54%로 7차 조사의 원표본 유지율보다 높음을 알 수 있다.

그림 2. 1~13차 한국복지패널조사의 원표본 가구 및 가구원 유지율 추이

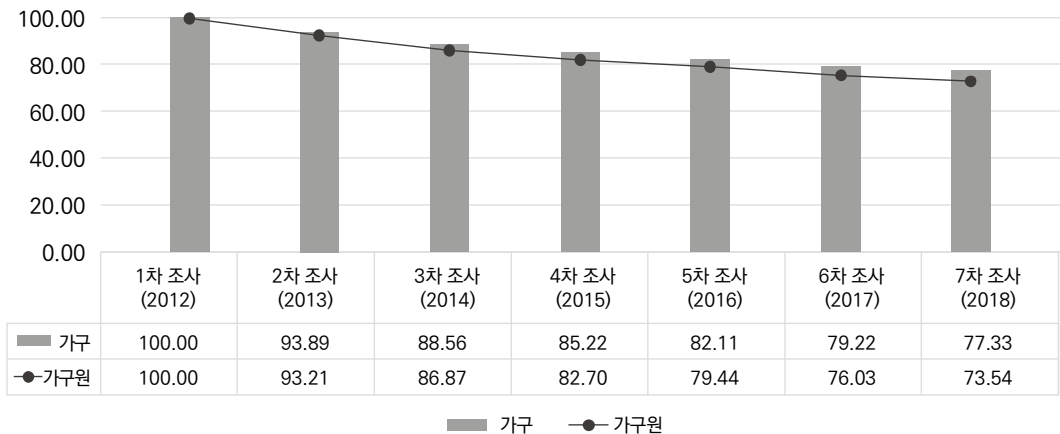
(단위: %)



자료: 김태완, 오미애, 이봉주, 강상경, 정원오, 박정민...김정욱. (2018). 2018년 한국복지패널 기초분석 보고서. p. 37 [그림 2-2-1].

그림 3. 1~7차 한국복지패널조사의 신규 패널 가구 및 가구원 유지율 추이

(단위: %)



자료: 김태완, 오미애, 이봉주, 강상경, 정원오, 박정민...김정욱. (2018). 2018년 한국복지패널 기초분석 보고서. p. 38 [그림 2-2-2].

표 4. 한국복지패널조사 1차 조사 대비 13차 조사의 조사 완료 패널 가구 및 가구원 수

(단위: 가구, 명, %)

구분	가구		가구원	
	표본 수	비율	표본 수	비율
원표본	4,266	60.32	8,688	60.07
신규 패널	1,392	19.68	2,482	17.16
신규	84	1.19	94	0.65
계	5,742	81.19	11,264	77.88

자료: 김태완, 오미애, 이봉주, 강상경, 정원오, 박정민, 김정욱. (2018). 2018년 한국복지패널 기초분석 보고서. p. 38 (표 2-2-3).

표 5. 한국복지패널조사 13차 조사의 원표본 가구 및 가구원 지역별 탈락률 비교

(단위: 가구, 명, %)

지역	가구			가구원		
	1차 조사	13차 조사	탈락률	1차 조사	13차 조사	탈락률
전국	7,072	4,266	39.68	14,463	8,688	39.93
서울	1,335	622	53.41	2,847	1,361	52.20
부산	528	300	43.18	1,090	593	45.60
대구	422	252	40.28	873	516	40.89
인천	437	258	40.96	914	531	41.90
광주	244	172	29.51	519	386	25.63
대전	208	116	44.23	435	255	41.38
울산	203	111	45.32	424	246	41.98
경기	1,130	651	42.39	2,324	1,422	38.81
강원	236	159	32.63	465	313	32.69
충북	221	163	26.24	457	329	28.01
충남	321	248	22.74	662	509	23.11
전북	347	237	31.70	655	431	34.20
전남	377	273	27.59	709	504	28.91
경북	494	327	33.81	942	563	40.23
경남	488	326	33.20	975	621	36.31
제주	81	51	37.04	172	108	37.21

자료: 김태완, 오미애, 이봉주, 강상경, 정원오, 박정민, 김정욱. (2018). 2018년 한국복지패널 기초분석 보고서. p. 39 (표 2-2-4).

1차 조사 대비 13차 조사의 조사 완료 패널 가구 및 가구원 수를 원표본, 신규 패널, 신규(13차 조사에 새롭게 진입한 가구 또는 가구원)로 나누어 살펴보면 다음과 같다.

〈표 5〉의 1차 조사 대비 13차 조사의 원표본 패널 가구 탈락률(attrition rate)을 살펴보면, 전체적으로 가구는 39.68%, 가구원은 39.93%의 탈락률을 보임을 알 수 있다.

표 6. 한국복지패널조사 13차 조사의 신규 패널 가구 및 가구원 지역별 탈락률 비교

(단위: 가구, 명, %)

지역	가구			가구원		
	1차 조사 (2012년)	7차 조사 (2018년)	탈락률	1차 조사 (2012년)	7차 조사 (2018년)	탈락률
전국	1,800	1,392	22.67	3,375	2,482	26.46
서울	194	129	33.51	397	247	37.78
부산	133	103	22.56	239	188	21.34
대구	126	95	24.60	223	167	25.11
인천	109	83	23.85	199	143	28.14
광주	87	75	13.79	159	140	11.95
대전	89	62	30.34	166	114	31.33
울산	77	56	27.27	164	116	29.27
경기	180	143	20.56	370	278	24.86
강원	95	77	18.95	174	126	27.59
충북	90	66	26.67	176	115	34.66
충남	106	100	5.66	196	177	9.69
전북	104	82	21.15	190	143	24.74
전남	102	77	24.51	178	126	29.21
경북	130	104	20.00	229	166	27.51
경남	127	99	22.05	218	155	28.90
제주	51	41	19.61	97	81	16.49

자료: 김태완, 오미애, 이봉주, 강상경, 정원오, 박정민, 김정욱, (2018). 2018년 한국복지패널 기초분석 보고서. p. 40 (표 2-2-5).

지역별 가구 탈락률은 서울이 33.51%로 가장 높고, 울산이 5.66%로 두 번째로 높다. 반면 충남은 22.74%로 탈락률이 가장 낮게 나타났다. 지역별 가구원 탈락률은 서울이 37.78%로 가장 높고, 부산이 21.34%로 두 번째로 높았다. 지역별 가구원 탈락률이 가장 낮은 곳은 충남으로 9.69%이다.

〈표 6〉에서 1차 조사 대비 7차 조사의 신규 패널 가구 탈락률을 살펴보면, 전체적으로 가구는 22.67%, 가구원은 26.46%의 탈락률을 보였다. 지역별 가구 탈락률은 서울이 33.51%로 가장 높고, 대전이 30.34%로 두 번째로 높았다. 반면 충

남은 5.66%로 가장 낮게 나타났다. 지역별 가구원 탈락률은 서울이 37.78%로 가장 높고, 충남이 34.66%로 두 번째로 높았다.

4. 마모패턴별 특성 비교 분석

가. 마모패턴 분석

한국복지패널 표본 가구의 마모패턴을 나타내는 변수를 생성하고, 기존 패널과 신규 패널의 원 가구에 대하여 마모패턴을 분석하였다. 마모패턴 변수는 패널 유지인 경우에 '1'의 값을 지정하고,

표 7. 기존 패널 마모패턴별 원가구 범주 구분

(단위: 가구)

가구 구분	범주	마모패턴	가구 수
원가구	p1	"11111111111111"	4,170
	p2	"11111111111110"	139
	p3	"11111111111100"	165
	p4	"11111111111000"	206
	p5	"11111111110000"	124
	p6	"11111111100000"	191
	p7	"11111110000000"	146
	p8	"11111100000000"	131
	p9	"11111000000000"	289
	p10	"11110000000000"	251
	p11	"11100000000000"	205
	p12	"11000000000000"	384
	p13	"10000000000000"	527

자료: 오미애, 이해정, 신재동, 이계오, 박승환, 손창균. (2019). 2019년 한국복지패널 심층분석 보고서, p. 48 (표 4-1-1) 재구성.

표 8. 신규 패널 마모패턴별 원가구 범주 구분

(단위: 가구)

가구 구분	범주	마모패턴	가구 수
원가구	np1	"1111111"	1,389
	np2	"1111110"	36
	np3	"1111100"	52
	np4	"1111000"	55
	np5	"1110000"	59
	np6	"1100000"	96
	np7	"1000000"	110

자료: 오미애, 이해정, 신재동, 이계오, 박승환, 손창균. (2019). 2019년 한국복지패널 심층분석 보고서, p. 48 (표 4-1-1) 재구성.

패널 마모인 경우는 '0'의 값을 지정하였다. 기존 패널의 경우 총 13차에 걸쳐 조사가 이루어졌으므로 자릿수는 13자리로 나타나며, 신규 패널의 경우는 7차부터 13차까지 조사가 이루어졌으므로 자릿수는 7자리로 나타난다. <표 7>은 기존 패널의 원가구에 대한 마모패턴을 정리한 것이

다. 원가구 외에도 분가 가구(1차 이후에 패널에 진입한 가구)와 기타 가구(마모패턴이 불규칙한 가구)가 있지만, 해당 가구 수가 적어 추후 마모 패턴별 가구 특성 분석은 각 차수의 원가구를 중심으로 진행하였다.

<표 8>은 7차부터 시행된 조사에서 신규 패널

가구에 대한 마모패턴을 정리한 것이다. 기존 패널과 마찬가지로 분가 가구와 기타 가구는 제외하고 추후 마모패턴별 가구 특성 분석은 각 차수의 원가구를 중심으로 진행하였다.

1) 균등화된 소득에 따른 가구 구분

일반 가구(1)와 저소득 가구(2)로 구분되어 있는 균등화소득에 따른 가구 구분 변수를 이용해 저소득층 가구의 비율을 마모패턴별로 비교하였

다. <표 9>는 기존 패널 및 신규 패널에 대해 마모패턴별 저소득 가구 비율을 계산한 것이다. 13차까지 모두 응답한 패널 가구(마모패턴 p1)의 저소득 가구 비율이 상대적으로 일찍 탈락한 패널에 비하여 높게 나타남을 알 수 있다. 이는 기존 패널과 신규 패널 모두 같은 경향으로 나타나고 있다. 또한 신규 패널의 저소득 가구 비율이 전체적으로 기존 패널의 저소득 가구 비율보다 높게 나타남을 알 수 있다.

표 9. 기존 패널과 신규 패널의 마모패턴별 저소득 가구 비율

패널 구분	마모 패턴	1차	2차	3차	4차	5차	6차	7차	8차	9차	10차	11차	12차	13차
기존 패널	p1	0.49	0.42	0.43	0.42	0.41	0.42	0.43	0.41	0.42	0.41	0.41	0.42	0.42
	p2	0.50	0.48	0.47	0.48	0.45	0.42	0.45	0.41	0.43	0.43	0.44	0.49	
	p3	0.52	0.49	0.49	0.45	0.46	0.47	0.46	0.47	0.48	0.43	0.45		
	p4	0.47	0.44	0.44	0.43	0.45	0.42	0.41	0.39	0.37	0.39			
	p5	0.52	0.44	0.49	0.47	0.46	0.45	0.50	0.44	0.44				
	p6	0.54	0.46	0.49	0.48	0.46	0.48	0.45	0.46					
	p7	0.57	0.50	0.55	0.53	0.51	0.51	0.48						
	p8	0.59	0.54	0.54	0.53	0.53	0.51							
	p9	0.47	0.38	0.41	0.43	0.40								
	p10	0.45	0.39	0.40	0.40									
	p11	0.42	0.35	0.35										
	p12	0.32	0.26											
	p13	0.27												
	전체	0.46	0.41	0.43	0.42	0.41	0.42	0.41	0.39	0.39	0.38	0.38	0.38	0.38
신규 패널	np1							0.54	0.53	0.55	0.54	0.54	0.54	0.55
	np2							0.53	0.61	0.56	0.50	0.56	0.44	
	np3							0.60	0.56	0.58	0.64	0.58		
	np4							0.58	0.58	0.56	0.55			
	np5							0.44	0.42	0.48				
	np6							0.43	0.41					
	np7							0.36						
		전체							0.53	0.52	0.55	0.54	0.54	0.52

자료: 오미애, 이해정, 신재동, 이계오, 박승환, 손창균. (2019). 2019년 한국복지패널 심층분석 보고서. p. 55-56 <표 4-2-2>, <표 4-2-3>.

2) 총 가구원 수

인 이상 가구의 비율은 감소하는 경향을 보였다.

〈표 10〉은 전체 가구를 대상으로 기존 패널 및 신규 패널에 대한 차수별 총 가구원 수 비율을 보여 준다. 기존 패널 및 신규 패널에서 차수가 진 행됨에 따라 1인, 2인 가구의 비율은 증가하고 3

3) 가구 형태

〈표 11〉은 전체 가구를 대상으로 기존 패널 및 신규 패널에 대한 차수별 가구 형태 범주의 비

표 10. 차수별 총 가구원 수 비율

(단위: %)

패널구분	총 가구원 수	1차	2차	3차	4차	5차	6차	7차	8차	9차	10차	11차	12차	13차
기존 패널	1인	21.5	22.1	22.8	23.3	24.4	25.4	25.7	26.1	26.7	27.5	28.4	28.8	30.1
	2인	29.6	29.4	29.9	29.9	29.8	29.6	30.2	29.9	30.1	30.1	30.3	30.9	30.8
	3인	19.2	19.1	18.6	18.2	18.1	17.9	17.6	17.8	17.8	17.1	17.0	17.3	17.0
	4인	22.0	21.5	20.8	20.5	19.9	19.4	18.9	18.9	18.3	18.2	17.7	16.8	16.2
	5인 이상	7.7	7.9	7.9	8.0	7.8	7.6	7.6	7.3	7.0	7.1	6.6	6.2	5.9
	합계	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
신규 패널	1인							29.8	31.4	33.3	34.8	36.2	37.9	39.9
	2인							38.6	37.6	37.5	37.2	35.9	35.2	33.9
	3인							14.8	14.4	13.5	12.9	13.2	12.9	12.1
	4인							13.2	12.6	12.1	11.7	11.7	11.3	11.4
	5인 이상							3.7	3.9	3.7	3.5	3.1	2.7	2.8
	합계							100.0	100.0	100.0	100.0	100.0	100.0	100.0

자료: 오미애, 이해정, 신재동, 이계오, 박승환, 손창균. (2019). 2019년 한국복지패널 심층분석 보고서, p. 60 〈표 4-2-5〉.

표 11. 차수별 가구 형태 비율

(단위: %)

패널 구분	가구 형태	1차	2차	3차	4차	5차	6차	7차	8차	9차	10차	11차	12차	13차
기존 패널	단독	21.5	22.1	22.8	23.3	24.5	25.4	25.7	26.1	26.7	27.5	28.4	28.8	30.1
	모자	1.6	1.6	1.4	1.2	1.2	1.1	1.2	1.2	1.2	1.3	1.3	1.3	1.0
	부자	0.6	0.6	0.5	0.5	0.5	0.5	0.4	0.6	0.5	0.5	0.5	0.7	0.5
	조손 가구 또는 소년소녀가장	0.6	0.6	0.9	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.6	0.6	0.5
	기타	75.8	75.3	74.4	74.2	73.2	72.4	72.0	71.5	70.8	70.0	69.2	68.8	68.0
	합계	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
신규 패널	단독							29.8	31.4	33.3	34.8	36.2	37.9	39.9
	모자							2.3	1.6	1.5	1.6	1.6	1.1	0.8
	부자							0.6	0.5	0.4	0.5	0.4	0.4	0.2
	조손 가구 또는 소년소녀가장							0.5	0.8	0.8	0.6	0.5	0.5	0.2
	기타							66.8	65.7	64.0	62.6	61.3	60.1	58.9
	합계							100.0	100.0	100.0	100.0	100.0	100.0	100.0

자료: 오미애, 이해정, 신재동, 이계오, 박승환, 손창균. (2019). 2019년 한국복지패널 심층분석 보고서, p. 63 〈표 4-2-7〉.

을 보여 준다. 분석 결과, 기존 패널 및 신규 패널에서 차수가 진행됨에 따라 단독 가구의 비율이 증가하는 경향을 보였고, 반대로 기타 가구의 비율은 감소하는 경향을 보였다.

나. 차수별 횡단가중치 분포 분석

전체 패널에 대한 가중치는 가구 통합 가중치를 사용하였다. 차수별로 전체 가구, 일반 가구, 저소득 가구를 구분하여 가중치의 분포를 비교하도록 하였다. <표 12>~<표 14>에서는 가중치에 대하여 최댓값, 99%, 95%, 90%, 75%, 50%, 25%, 10%, 5%, 1% 백분위 수, 최솟값, 평균, CV, 4분위 수 범위(3분위 수 - 1분위 수)를 구하여 가중치 분포를 분석하였다.

전체 가구의 가중치 분포를 보면 차수가 지남

에 따라 표본 수가 감소하게 되어 전체적으로 가중치값이 증가하는 것을 알 수 있다. 전반적인 값이 증가했을 뿐만 아니라, 4분위 수 범위를 보게 되면 차수가 지남에 따라 가중치의 산포도 증가했음을 알 수 있다. CV는 1차에 150% 정도였으나 3차 이후로는 120% 정도를 유지하고 있다. 또한 7차에 추가한 1800가구로 인해 가중치의 평균값 등 전반적인 값을 낮추는 효과를 보였다.

일반 가구의 가중치값 평균이 전체 가구에 비하여 크고 저소득 가구의 가중치값 평균은 전체 가구에 비하여 작음을 알 수 있다. 이는 차수가 지남에 따라 일반 가구의 탈락이 저소득 가구에 비해 높다는 것을 의미한다.

표 12. 차수별 전체 가구 가중치 분포

	1차	2차	3차	4차	5차	6차	7차	8차	9차	10차	11차	12차	13차
최대	9,321	15,053	28,136	20,714	21,593	21,633	34,994	10,399	11,282	11,795	12,171	12,643	13,300
99%	6,666	7,674	9,166	9,862	10,074	10,948	11,293	10,250	10,940	11,620	11,830	12,360	13,300
95%	4,939	5,483	6,454	6,614	7,062	7,639	6,966	7,486	7,866	8,297	8,634	8,895	9,295
90%	4,231	4,572	5,118	5,357	5,731	6,239	5,594	5,993	6,304	6,609	6,978	7,162	7,503
75%	3,090	3,340	3,578	3,730	3,885	4,164	3,709	4,022	4,216	4,283	4,548	4,766	4,829
50%	1,986	2,111	2,078	2,136	2,193	2,281	2,116	2,266	2,400	2,464	2,524	2,604	2,649
25%	1,138	1,264	1,140	1,133	1,150	1,178	1,116	1,177	1,221	1,245	1,250	1,293	1,294
10%	554	676	624	628	611	635	583	607	608	624	647	655	659
5%	386	515	450	451	440	445	419	429	438	445	455	468	452
1%	257	316	298	296	279	271	234	236	249	235	219	225	219
최소	160	131	95	48	83	75	49	87	66	71	65	69	70
평균	2,247	2,456	2,600	2,686	2,804	2,991	2,741	2,884	3,021	3,125	3,260	3,368	3,464
CV	154	151	131	128	126	122	118	129	128	125	125	125	122
4범위	1,953	2,075	2,438	2,598	2,736	2,986	2,592	2,845	2,994	3,038	3,298	3,473	3,535

자료: 오미애, 이해정, 신재동, 이계오, 박승환, 손창균. (2019). 2019년 한국복지패널 심층분석 보고서. p. 69 (표 4-3-1).

표 13. 차수별 일반 가구 가중치 분포

	1차	2차	3차	4차	5차	6차	7차	8차	9차	10차	11차	12차	13차
최대	9,321	15,053	28,136	20,714	21,593	21,633	34,994	10,399	11,282	11,795	12,111	12,643	13,300
99%	7,053	8,259	10,255	10,956	11,074	12,612	12,299	10,250	10,940	11,620	11,830	12,360	13,300
95%	5,691	6,220	7,369	7,459	7,908	8,783	8,013	8,373	8,813	9,403	9,576	10,024	10,501
90%	4,828	5,197	5,964	6,193	6,546	7,221	6,255	6,701	7,210	7,613	7,795	8,123	8,512
75%	3,837	3,903	4,346	4,510	4,832	5,214	4,411	4,840	5,079	5,269	5,504	5,683	5,920
50%	2,924	2,820	2,990	3,061	3,165	3,379	2,815	3,098	3,228	3,253	3,392	3,501	3,626
25%	2,066	1,819	1,810	1,775	1,774	1,857	1,571	1,745	1,835	1,815	1,862	1,974	2,027
10%	1,284	1,027	988	969	917	957	810	890	993	963	990	1,009	1,044
5%	893	691	682	651	603	671	551	602	676	635	643	658	662
1%	462	338	346	357	327	358	283	328	333	303	285	270	255
최소	229	136	95	48	83	91	49	116	94	104	101	97	70
평균	3,039	3,035	3,334	3,404	3,548	3,852	3,334	3,546	3,736	3,847	3,987	4,130	4,289
CV	212	177	157	153	150	145	133	152	151	145	147	147	144
4범위	1,772	2,084	2,536	2,735	3,058	3,357	2,840	3,095	3,244	3,454	3,641	3,710	3,892

자료: 오미애, 이해정, 신재동, 이계오, 박승환, 손창균. (2019). 2019년 한국복지패널 심층분석 보고서. p. 71 (표 4-3-2).

표 14. 차수별 저소득 가구 가중치 분포

	1차	2차	3차	4차	5차	6차	7차	8차	9차	10차	11차	12차	13차
최대	6,109	8,127	13,275	17,782	14,988	15,350	16,096	10,372	10,940	11,620	12,171	12,444	13,300
99%	3,598	5,007	6,576	6,988	6,862	7,042	7,205	8,448	8,487	8,466	9,506	10,051	9,809
95%	2,743	3,711	3,979	4,287	4,300	4,690	4,759	4,982	5,192	5,363	5,806	6,024	6,066
90%	2,371	3,035	3,198	3,368	3,440	3,658	3,492	3,734	4,068	4,191	4,504	4,726	4,577
75%	1,792	2,072	2,046	2,138	2,180	2,265	2,237	2,335	2,513	2,604	2,719	2,824	2,716
50%	1,265	1,456	1,308	1,325	1,334	1,376	1,330	1,379	1,418	1,467	1,475	1,519	1,501
25%	663	842	801	790	786	787	749	756	762	794	788	807	821
10%	385	535	473	470	463	455	439	448	439	461	471	488	462
5%	307	448	370	369	379	370	356	343	351	342	357	351	331
1%	241	263	251	239	246	213	192	180	218	201	188	186	188
최소	160	131	97	78	94	75	71	87	66	71	65	69	72
평균	1,329	1,635	1,632	1,693	1,721	1,784	1,753	1,824	1,924	1,989	2,089	2,149	2,145
CV	169	156	129	119	123	121	116	115	115	115	110	109	108
4범위	1,128	1,230	1,246	1,348	1,394	1,477	1,488	1,579	1,751	1,810	1,931	2,017	1,895

자료: 오미애, 이해정, 신재동, 이계오, 박승환, 손창균. (2019). 2019년 한국복지패널 심층분석 보고서. p. 73 (표 4-3-3).

5. 나가며

한국복지패널은 첫 조사 이후 15년 동안 이어져 오며 따라 사회복지 및 경제학적인 측면에서의 패널 연구와 더불어 패널 데이터에 대한 근본적인 고민의 하나로 통계학적인 측면에서 패널 가구의 진입과 소멸과 같은 패널 자체의 변동성에 대한 연구가 필요한 시점이 되었다. 한국복지패널은 조사 연차가 6년 이상 지나면서 패널 탈락률이 커짐에 따라 표본 대표성을 보완하기 위해 7차 조사 때 1800가구를 신규로 추가하였다. 그리고 2019년에는 한국복지패널의 마모 특성 및 종단적 결측 패턴을 살펴보고 가중치 작업을 검토하여 원표본에 대해 종단가중치를 부여한 심층 연구를 진행하였다.

여기에서는 연구 내용 중 일부인 가중치 산출 과정을 설명하고 표본 관리 및 모집단 추정의 정확도를 기하기 위한 마모패턴별 가구 특성 비교 분석을 다루었다. 이 연구는 패널 자료의 품질 개선에 근거 자료로 활용될 수 있다. 패널 연구에서 대표성을 확보하기 위해서는 마모패턴 분석 및 가중치 산출 과정에 대한 검토가 지속적으로 이루어져야 할 것이다. 이러한 분석이 단발성 검토로 끝난다면 앞으로 가중치 산출 과정의 방향성과 마모에 대한 대처 방안을 논의하는 데 한계가 있을 수 있다. 한국복지패널의 연속적인 결과물로 분석이 이루어지고 통계 전문가뿐만 아니라 복지 전문가도 함께 논의하는 구조가 만들어진다면 한국복지패널의 품질은 지속적으로 관리될 수

있을 것이다.

한국복지패널에 대해 품질 개선을 포함한 다양한 통계적 이슈를 다룰 수 있는 연구도 지속되어야 하겠다. 한국복지패널에서 분가 가구가 원가구와 비교할 때 어떠한 특성이 있는지에 대한 연구도 심층적으로 이루어져야 한다. 또한 향후 표본 탈락으로 인한 신규 패널 추가는 불가피한데, 이에 대한 준비도 철저히 해야 한다. 신규 패널 추가는 이전처럼 이중 추출 방식을 쓸 수 없기 때문에, 어떻게 보정해야 하는지에 대한 고민도 필요하다. 그리고 다른 패널들에서 패널을 운영하면서 겪어 온 경험과 이슈를 공유할 수 있는 네트워크 구성도 필수적이다. 한국복지패널의 품질 개선을 위해 파라데이터(paradata) 관리 틀도 재정립할 필요가 있다. 이러한 다양한 이슈들에 대한 한국복지패널 연구가 활발히 이루어진다면, 기존 데이터 생산 및 관리 사업에 더해 향후 한국복지패널조사 연구의 안정적인 수행에 기여할 수 있으며, 데이터 품질을 개선하고 향상시키는 데 중요한 역할을 할 수 있을 것이다. ■

참고문헌

- 김미곤, 여유진, 손창균, 김문길, 이봉주, 구인회, ... 윤민석. (2006). 2006 한국복지패널 기초분석 보고서. 서울: 한국보건사회연구원, 서울대학교 사회복지연구소.
- 김미곤, 여유진, 손창균, 김문길, 이봉주, 구인회, ... 지은정. (2007). 2007 한국복지패널 기초분석 보고서. 서울: 한국보건사회연구원.

- 원, 서울대학교 사회복지연구소.
- 이현주, 노대명, 오미애, 전지현, 이봉주, 김태성, ...이혜림. (2013). 2013년 한국복지패널 기초분석 보고서. 서울: 한국보건사회연구원.
- 김태완, 오미애, 이봉주, 강상경, 정원오, 박정민, ...김정욱. (2018). 2018년 한국복지패널 기초분석 보고서. 세종: 한국보건사회연구원.
- 오미애, 이혜정, 신재동, 이계오, 박승환, 손창균. (2019). 2019년 한국복지패널 심층분석 보고서.
- Duncan, G. (1995). A Simple Method for Weighting in Household Panel Survey. Working paper, Northwestern University.
- Kalton, G., & Brick, M. (1994). Weighting Schemes for Household Panel Survey. Proceedings of the Section on Survey research Method, American Statistical Association. 785-790.