

# 한국복지패널 가중치 및 발전방안

*The Weighting and Development Plan of KOWEPS*



손창균 한국보건사회연구원 부연구위원

한국복지패널 3차 자료의 가중치 조정 과정을 단계별로 살펴보고 있으며, 향후 복지패널의 발전 방안을 논의하였다. 2차 웨이브 종단면 가중치를 기준으로 3차 웨이브에서 종단면 가중치와 횡단면 가중치를 조정하는 절차를 소개하였고, 부여된 가중치의 적절성을 상대표 준오차를 이용하여 살펴보았다. 한편 현재 복지패널이 보다 개선되기 위해서 몇가지 개선 사항에 대해 논의하였다.

## 1. 들어가기

한국복지패널은 2009년 7월 조사를 완료하여 4차 웨이브까지 조사를 수행하였다. 패널의 시작부터 타 패널과 다르게 3개 패널을 통합하여 예산의 효율성을 극대화 하였고, 한국보건사회연구원의 유능한 자체조사인력을 활용함으로써 패널자료의 신뢰성을 높이도록 하였다. 패널의 생명은 무엇보다도 원표본(original sample)의 유지율이다. 이러한 관점에서 4차 웨이브까지 진행된 한국복지패널은 약 83.7%의 높은 원표본 유지율을 보여 성공적인 패널구축이라고 자평하고 싶다.

또한 매년 기초분석 보고서와 심층보고서를 발간하고, 패널학술대회를 개최함으로써 명실상부한 한국의 대표 패널로서의 위상을 정립해

가고 있다. 이와 더불어 패널자료의 생산과정에 연구진이 적극적으로 개입하여 보다 신뢰성 있는 패널자료를 생산하고 있으며, 특히 패널관리, 자료에 대한 내검(editing), 에러체크(error check), 가중치 조정(weighting adjustment) 등의 전반적인 사항을 내부연구진이 모두 수행함으로써 보다 정확한 자료 생산을 위해 노력하고 있다.

이러한 측면에서 패널의 가중치 조정은 패널자료의 대표성 확보 및 정확성을 높이기 위해 반드시 필요한 연구이다. 시간의 흐름에 따라 발생하는 패널표본의 탈락으로 인한 패널의 대표성 상실을 보완하고, 패널표본의 변동을 반영하여 조사기준 연도의 전국 인구 및 가구를 대표할 수 있도록 조정하는 작업이 가중치 조정이 다. 1차 웨이브에서는 횡단면 가중치만을 부여

하였으나, 2차년도 이후 패널가구의 변동을 종 단면과 횡단면상에 정확히 반영하기 위해 “로지스틱회귀분석(logistic regression analysis)”, “사 후층화조정(post-stratification)”, “다중대체(multiple imputation)” 등과 같은 다양한 통계적 기법을 활용하고 있다.

본고는 한국복지패널의 3차 웨이브 자료의 가중치 조정과정을 단계적으로 고찰하고, 이를 통해 한국복지패널의 대표성을 기초적인 분석을 통해 검증하며, 마지막으로 전반적인 한국복지패널의 향후 개선점에 대해 논의하고자 한다.

## 2. 가중치 부여 체계

### 1) 가중치 부여 체계

한국복지패널의 3차년도 가중치 산정은 기본적으로 2차 웨이브에 부여된 개인 가중치를 기준으로 하였다. 가구가중치는 조사차수가 증가함에 따라 가구원의 변동에 기인하여 1차 웨이브에 정의한 가구의 개념이 달라질 수 있기 때문에 기본가중치로 고려되지 않았다. 한국복지

패널의 가중치 부여 체계는 다음의 [그림 1]과 같다.

이와 같은 가중치 부여 체계에 따르면, 1차 웨이브에서 파악된 원표본가구원(original sample household members)이 분가하여 새로운 가구를 생성하거나, 1차 웨이브에 조사되지 않은 원표본가구원이 2차 웨이브에 새롭게 진입한 가구원, 1차 웨이브에 응답하였으나, 2차년도 이후 사망, 이민, 유학 등의 사유로 자연 소실되거나, 조사거절, 기관입소, 행방불명 등의 사유로 탈락한 모든 가구원에 대해 이를 보완할 수 있도록 가중치를 조정해야 한다.

이와 더불어 개별 가구에서 가구원 변동이 발생할 수 있는 유형과 그에 따른 가중치 산정방법을 요약하면 다음과 같다.

### 2) 가중치 산정 절차

2.1절에서 서술한 바와 같이 패널표본가구의 다양한 변동을 유형별로 분류하고, 그에 따라 가중치를 산정하게 된다. 기본적인 가중치 산정 절차를 구체적으로 살펴보면 다음과 같다.

○ 1단계: 2차 웨이브에서 응답한 개인이 3차

그림 1. 한국복지패널 가중치 부여 체계

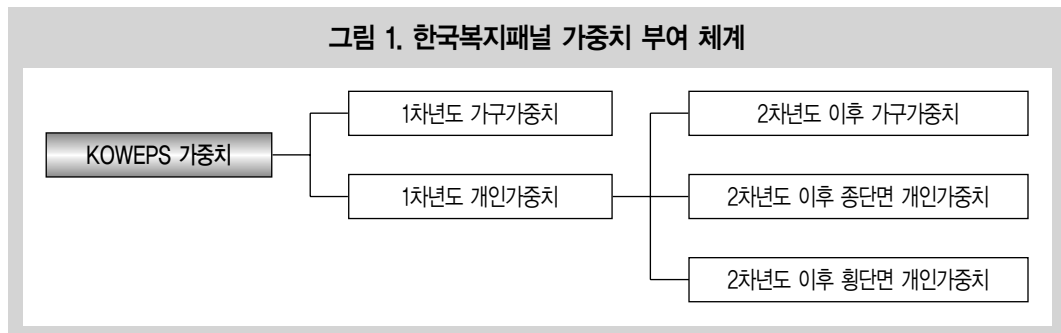


표 1. 가구원 변동과 가중치 산정

변동유형	개인 가중치 산정	
	종단면 가중치	횡단면 가중치
[유형1] 3차 웨이브에서 원 표본가구에 결혼 또는 출생 등으로 인하여 신규 가구원이 진입한 경우	0의 가중치	가구의 평균가중치
[유형2] 2차 웨이브에서 응답한 원표본 가구원 중 일부가 사망하여 3차 웨이브에서 탈락하는 경우	전체 가중치의 조정	전체 가중치 조정
[유형3] 1차 웨이브에서 원가구원인데 응답하지 않아 누락되었다가, 3차 웨이브에서 새롭게 응답한 경우	가구 가중치의 평균	종단면 가중치 부여
[유형4] 2차 웨이브에서 원표본가구에 포함되었던 가구원이 결혼 또는 이민 등으로 원표본가구에서 분리되어 가구원에서 제외된 경우	2차 웨이브 가중치부여	2차 웨이브 가중치부여
[유형5] 원표본가구와 비표본가구가의 합가한 경우 분가했던 자식들이 2차 웨이브에서는 원표본가구로 진입하여 3차 웨이브에 패널가구원으로 진입하는 경우	0의 가중치	가구원 가중치의 평균

웨이브에서 탈락한 경우 로지스틱회귀모형을 이용하여 응답확률을 추정한다.

- 2단계: 2차 웨이브의 종단면 가중치를 기본가중치로 고려하여 응답확률의 역수를 곱하여 패널 소실에 따른 가중치를 조정한다.
- 3단계: <표 1>에서 분류한 각 유형별로 자료를 분석하여 해당 표본에 대해 종단면 가중치를 조정한다.
- 4단계 : 조사기준 연도의 인구수에 따라 종단면 가중치와 횡단면 가중치를 사후증화 조정한다.

### 3. 가중치 분석

단계적으로 조정된 종단면 가중치와 횡단면 가중치에 대한 기술통계값을 구해보면 <표 2>와 같이 종단면 및 횡단면 가중치의 평균은 2,916.6으로서 개인 1명이 평균적으로 약 2,916명을 대표하고 있으며, 종단면 가중치의 변동이 횡단면 가중치의 변동에 비해 상대적으로 크게 나타나고 있다. 이러한 원인은 종단면 가중치는 가구내 개인의 변동을 반영하였기 때문이다.

한편 16개 지역별로 개인 종단면 가중치의 기술통계값을 산출해보면 <표 3>과 같이 경기

표 2. 개인 종단면 및 횡단면 가중치의 기술통계값

구분	평균	표준오차	중위수	제1사분위수	제3사분위수
종단면가중치	2,916.6	14.28	2,678.87	1,488.97	4,015.36
횡단면가중치	2,916.6	14.15	2,671.26	1,500.06	4,003.60

자료: KOWEPS 3차웨이브 원자료.

지역이 4,063.04 로 가장 크며, 전남지역이 2069.47로 가장 작게 나타나고 있다. 가중치 변동이 가장 큰 지역은 경북지역으로 가중치의 상대표준오차가 3.21%로 나타나고 있으며, 가장 변동이 적은 지역은 서울로 0.76%로 나타나고 있다.

개인의 종단면 가중치의 산포정도를 지역별로 살펴보면 서울을 비롯한 부산, 대구, 인천, 울산, 경기, 충남, 전남, 경북 등에서 매우 큰 가중치가 발견되고 있다(그림 2 참조). 이와 같이 가중치의 극단 값이 나타나는 이유로는 표본추출단계에서 지역별 조사구의 추출율의 차이가 1,2,3차 웨이브에 반영되고 있으며, 지역별 구의 변동이 심하기 때문인 것으로 사료된다.

한국복지패널에서는 이러한 극단가중치를 별도로 조정하지 않은 이유는 앞에서도 언급했지만, 가중치의 상대표준오차가 최대 3.2%정도로 작기 때문에 별도의 조정 작업을 수행하지 않았으며, 권역별로 상대표준오차를 구하더라도 적절한 수준으로 나타나기 때문이다(표 4 참조).

다음으로 <표 5>로 부터 개인 횡단면 가중치를 분석하면 거의 지역적으로 종단면 가중치와 차이가 없다. 이는 가구변동이 있는 표본에 대해서만 종단면 가중치와 횡단면 가중치가 차이가 나기 때문에 평균적으로는 거의 차이가 없는 것으로 나타나고 있다. 단지 가중치의 평균의 표준오차가 종단면 가중치에 비해 작게 나타나

표 3. 지역별 개인 종단면 가중치 분포

지 역	관측치 수	평균값	표준오차	중위수	제1사분위수	제3사분위수
서울	3,015	3325.26	25.32	3178.81	2317.77	4210.19
부산	1,195	2949.97	48.01	2622.35	1761.94	3714.84
대구	994	2485.16	40.23	2379.88	1620.83	3249.32
인천	1,057	2471.96	44.60	2323.98	1491.82	3405.82
광주	655	2206.21	43.44	1959.45	1300.87	2947.16
대전	525	2831.70	50.27	2785.12	1977.73	3677.29
울산	486	2221.86	58.27	1884.79	1190.61	3128.74
경기	2,717	4063.04	34.91	3902.52	2784.95	5075.81
강원	560	2625.80	84.08	1712.05	1039.57	4846.02
충북	532	2788.35	72.87	2655.67	1256.93	4014.27
충남	852	2272.61	69.14	1376.53	778.91	3407.67
전북	750	2360.93	64.47	1973.86	836.77	3363.12
전남	873	2069.47	40.74	1716.00	1313.70	2514.95
경북	1,078	2443.99	78.45	973.90	564.52	3596.76
경남	1,134	2755.66	75.46	1178.30	434.65	4952.36
제주	191	2850.97	76.71	2743.55	2048.51	3381.91

자료: KOWEPS 3차웨이브 원자료.

그림 2. 한국복지패널 종단가중치 분포

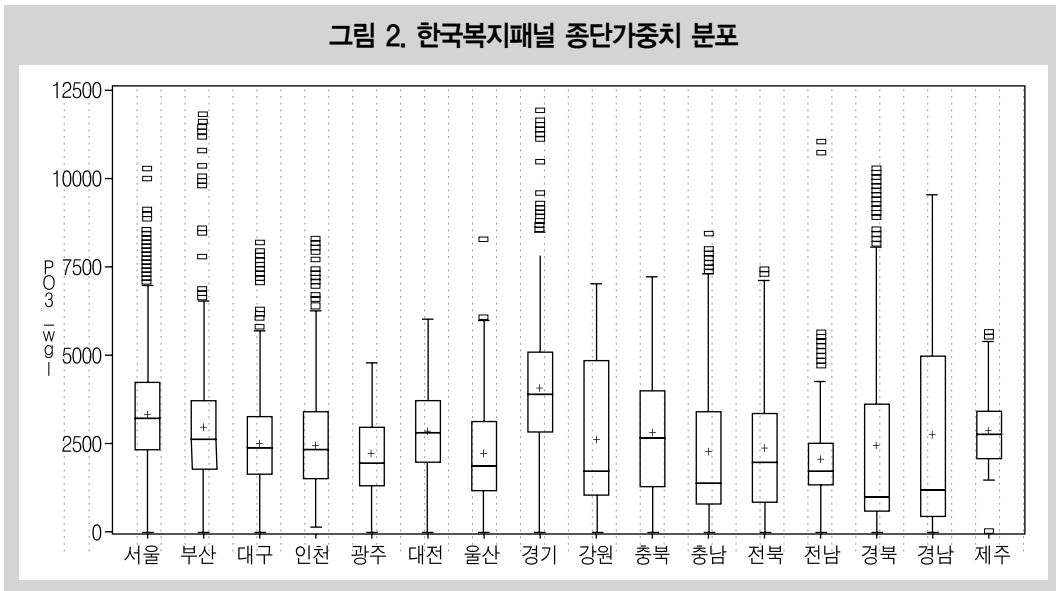


표 4. 7개 권역별 개인 종단면 가중치의 특성

권역별	평균	표준오차	상대표준오차(%)
서울	3325.263	25.32086	0.7615
수도권(인천/경기)	3617.424	30.37791	0.8398
부산/경남/울산	2745.992	38.24725	1.3928
대구/경북	2463.742	45.13807	1.8321
대전/충남	2485.773	47.43079	1.9081
강원/충북	2704.989	55.88071	2.0658
광주/전남/전북/제주	2254.737	27.85614	1.2354

자료: KOWEPS 3차 웨이브 원자료.

며, 이는 전술한 바와 같이 가구변동에 따른 종단면 가중치의 변동이 더 크게 나타나기 때문인 것으로 판단된다.

이와 같이 부여된 종단면 가중치를 이용하여 가구의 평균경상소득을 7개 권역별로 분석해보면 <표 6>과 같다. 권역별 경상소득의 상대표준오차를 계산해보면 전국 추정치의 상대표준오

차는 0.83%로 매우 작게 나타나며, 강원/충북 지역이 4.09%로 가장 변동이 크며, 서울지역 1.69%로 가장 작게 나타나고 있다. 전체적으로 종단면가중치를 적용한 경상소득의 추정치의 변동은 적음을 알 수 있다.

표 5. 지역별 개인 횡단면 가중치 분포

지역	관측치 수	평균값	표준오차	중위수	제1사분위수	제3사분위수
서울	3,015	3325.26	24.82	3167.81	2310.99	4194.24
부산	1,195	2949.98	47.46	2613.52	1758.70	3702.34
대구	994	2485.16	39.86	2378.00	1618.09	3240.89
인천	1,057	2471.96	44.60	2323.98	1491.82	3405.82
광주	655	2206.21	42.41	1950.42	1299.67	2928.04
대전	525	2831.70	49.00	2772.71	1968.92	3660.90
울산	486	2221.86	58.07	1882.69	1195.04	3125.25
경기	2,717	4063.04	34.33	3889.54	2774.88	5057.48
강원	560	2625.80	83.85	1710.20	1038.44	4840.78
충북	532	2788.35	72.33	2648.53	1256.74	4003.46
충남	852	2272.61	68.31	1397.23	780.39	3385.14
전북	750	2360.93	64.29	1973.38	836.58	3357.26
전남	873	2069.47	40.22	1709.44	1312.48	2505.34
경북	1,078	2444.00	78.03	1003.54	562.83	3583.76
경남	1,134	2755.66	75.26	1185.88	435.25	4944.68
제주	191	2850.97	74.97	2735.01	2042.13	3371.38

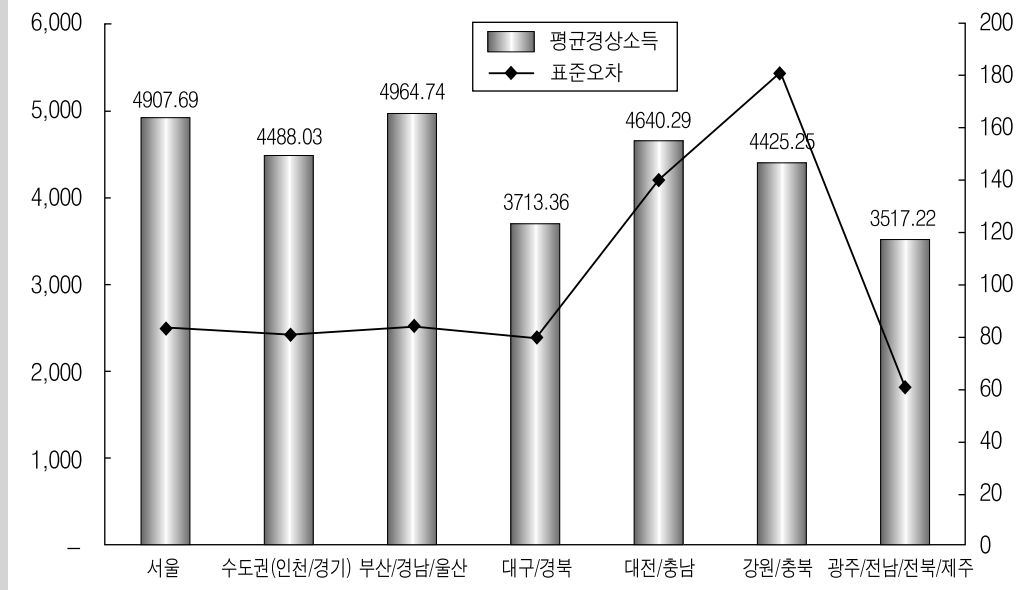
자료: KOWEPS 3차 웨이브 내부자료.

표 6. 권역별 평균 경상소득

권역별	평균(만원)	표준오차	상대표준오차(%)
전국	4,464.688	36.63821	0.8206
서울	4,907.689	82.94931	1.6902
수도권(인천/경기)	4,488.027	80.73300	1.7989
부산/경남/울산	4,964.736	84.14081	1.6948
대구/경북	3,713.363	79.60018	2.1436
대전/충남	4,640.294	140.5104	3.0280
강원/충북	4,425.254	180.9184	4.0883
광주/전남/전북/제주	3,517.220	60.71073	1.7261

자료: KOWEPS 3차 웨이브 원자료.

그림 3. 권역별 평균 경상소득 및 표준오차



자료: KOWEPS 3차 웨이브 원자료

마지막으로 가중치를 사용함으로써 분산의 증가분을 계산하기 위해 다음의 식을 이용한다 (Kish, 1965).

$$L = n' \frac{\sum w_h^2}{(\sum w_h)^2} = n' \frac{\sum n'_h w_h^2}{(\sum n'_h w_h)^2}$$

여기서  $L$ 은 가중치를 이용함으로써 발생하는 분산의 증가 값이며,  $n'$ 은 최종 관찰치수,  $n'_h$ 는  $h$ 층의 표본수,  $w_h$ 는  $h$ 층의 최종가중치를 나타낸다.

이 식으로부터 가중치를 적용함으로써 분산은  $L=1.0465$ 로서 전체분산은 약 4.7%정도 증가하여 매우 안정적으로 나타나고 있다.

## 4. 나가기

한국복지패널은 한국보건사회연구원과 서울대학교 사회복지연구소의 공동 컨소시엄으로 학제간 및 학연간의 공동연구로 수행되고 있는 대표적인 패널이다. 학연간 공동 연구를 통해 다양한 사회현상을 보다 과학적으로 담아낼 수 있으며, 따라서 학문적 연구자료 및 정책 기초 자료로서의 가치를 증대시킬 수 있는 장점을 가진다. 이러한 측면에서 복지패널은 매우 선도적인 연구 작업을 수행한다고 할 수 있다. 초기 패널구축과정에서 많은 논점들이 노정되었으나, 현재 안정된 패널을 유지하고 있으며, 패널 손실 또한 연차적으로 감소하는 경향을 보이고 있어 향후 한국의 대표 패널조사로서의 자리매김

할 수 있을 것으로 기대된다.

그러나 패널조사는 적절한 시간과 비용의 투입을 요구하는 매우 장기적인 사업이기 때문에 근시안적인 성과가 나타나길 기대하는 것은 패널 조사에 대한 잘못된 시각이다. 통상적으로 약 5차 웨이브 이상의 자료누적을 통해 패널자료로서의 유의미한 결과를 도출할 수 있기 때문이다.

패널표본으로 지속적으로 참여하는 가구들은 조사원의 다양한 접근 방식과 설득 과정을 통해 패널로 구축된 표본가구들이다. 따라서 이들에 대한 다양한 인센티브의 제공은 필수적이다. 패널조사에 참여하도록 유도하는 과정은 1차적으로 전화 컨택을 통해 조사시작을 공지하고, 조사현장에서 패널가구를 방문하여 가구조사표와 개인조사표에 대해 응답을 얻어내는 과정은 매우 전문적인 분야이다. 이러한 이유로 조사원에 대한 인센티브도 반드시 고려되어야 할 것이다.

조사방법의 개선도 요구된다. 1~4차 웨이브까지는 PAPI로 조사를 수행하였다. 2008년 정기 통계품질진단 결과로 부터 CAPI 조사로의 전환을 개선사항으로 요구하고 있다. 조사방법의 전환이 조사 자료의 품질에 얼마나 영향을 끼칠 것인지에 대해서는 많은 논쟁이 있다. 그러나 현실적으로 조사 자료의 품질을 보증하기 위해서는 CAPI 조사로의 전환이 바람직 할 것으로 본다.

현재 한국사회는 다른 나라에 비해 매우 동적

인 행태를 나타내고 있다. 과거에 경험해보지 못한 다양한 사회현상이 유발되고 있으며, 이 중 하나가 다문화 가구의 지속적인 증가추세이다. 이러한 다문화 가구에 대한 생성, 유지, 소멸에 대한 종단면적 연구를 위해 추가예산을 투입하여 약 1,500가구 정도의 다문화 부가표본 확보가 필요할 것이다.

마지막으로 복지패널은 매년 9월 말에 학술대회를 개최하고 있다. 2009년 현재 까지 2회에 걸쳐 국내 연구자들이 다양한 주제의 논문을 발표하였으며, 참가인원 또한 매년 급속히 증가하고 있는 추세이다. 연구영역의 다각화를 위해 국제학술대회의 개최도 고려할 필요가 있다. 이는 국내 사회복지관련 연구의 폭을 넓힐 수 있는 좋은 연구의 장이 될 것이다. 외국패널 학술대회 참여자를 분석해보면 세계 각국의 연구자들이 참여하여 매우 풍부한 연구결과를 산출하고 있다. 이러한 측면에서 향후 국제학술대회를 개최하여 한국의 사회상을 반영한 양질의 자료를 활용토록 함으로써 밖에서 보는 한국의 사회를 평가하는 좋은 기회가 될 수 있을 것으로 기대한다.

지금까지 한국복지패널의 가중치와 관련된 주제를 단계별로 살펴보았다. 가구의 변동을 적절히 반영하면서, 전국 대표성을 갖도록 하는 작업이 가중치 조정이다. 향후 이에 대한 적절한 매뉴얼을 구비하여 이용자들이 가중치체계에 대해 이해를 돕도록 할 계획이다. **본문**