

조사보고서 2010-01 (2/2)

제3차(2009년도) 우리나라 중·고령자 의 경제생활 및 노후준비 실태

-국민노후보장패널조사(KReIS) 심층분석보고서-

석상훈·최옥금·송현주·이은영·오지연·신혜리

머 리 말

최근 우리사회는 저출산 및 기대수명의 급속한 증가로 인해 인구구조가 급격히 고령화되고 있다. 이러한 고령화에 따른 사회경제적 파급효과에 대처하기 위한 다양한 방안이 연구 중에 있으나 관련 연구를 뒷받침해줄 수 있는 기초자료는 아직까지 미흡한 실정이다.

지금까지 고령화 연구를 위한 실태조사가 다양하게 실시되고 있으나 이들 조사는 당시의 정책적 목적에 의해서 일회성으로 실시되거나 조사내용이 특정한 항목에 제한되는 등의 문제점이 있었다. 따라서 이러한 실태조사를 통하여 고령화 환경에서 시간의 흐름에 따라 변동하는 개인들의 총체적 생활실태를 파악할 수 없었던 것이 현실이다.

이러한 측면에서 2005년부터 격년으로 실시되고 있는 국민노후보장패널조사는 우리나라 중·고령자들의 노후소득보장정책을 수립하기 위한 매우 중요한 자료가 될 것이다. 본 패널조사는 2005년 당시 만 50세 이상의 중·고령자가 있는 5,110가구를 대상으로 패널을 구축하였으며, 2007년 2차년도 조사에는 89.0%, 2009년 3차년 조사에는 83.0%의 원표본 유지율을 기록하였다. 완전한 패널자료를 구축하기 위해서는 수년이 더 소요되겠지만 현재의 자료를 통해서도 우리나라 중·고령자들의 생활실태와 조사기간 동안에 변동한 모습들을 다양한 측면에서 연구할 수 있는 유용한 정보를 포함하고 있다.

국민노후보장패널조사의 자료를 바탕으로 학문적으로 더 풍성한 연구가 수행되기를 바라며, 많은 연구자들의 아낌없는 성원과 조언도 기대한다.

이 보고서는 국민연금연구원의 석상훈 부연구위원의 책임 하에 최옥금 부연구위원, 송현주 주임연구원, 이은영 주임연구원, 오지연 연구원, 신혜리 연구원의 참여로 이루어졌다. 특히 조사과정에서부터 본 보고서 완료 단계에 이르기까지 굳은 업무를 도맡아 한 오지연 연구원과 신혜리 연구원의 노고에 고마움을 전한다. 뿐만 아니라 이 과정에서 조언을 아끼지 않은 많은 공단 내·외부의 다양한 전문가와 전문조사기관, 아울러 검독과정에서 본 보고서에 대한 유익한 조언을 해 준 관계전문가에게 감사드린다.

끝으로 보고서에 수록된 모든 내용은 어디까지나 저자들의 의견이며, 공단의 공식견해가 아님을 밝혀 둔다.

2010년 12월
이사장 전 광 우
원 장 백 화 중

<제목 차례>

요약	1
I. 서론	8
II. 노인의 가구유형 변화 및 영향을 미치는 요인	11
1. 문제제기	11
2. 연구방법	13
3. 분석결과	17
4. 논의	31
III. 노인가구의 빈곤여부에 따른 가구 지출 유형화 및 영향 요인 분석	33
1. 문제제기	33
2. 연구방법	35
3. 분석결과	37
4. 논의	49
IV. 우리나라 중·고령자의 은퇴경로 유형과 은퇴 후 소득 비교	51
1. 문제제기	51
2. 분석방법 및 자료 소개	53
3. 은퇴경로의 유형화 및 유형별 분포	57
4. 은퇴경로별 은퇴 후 소득 비교	71
5. 결론	82

V. KReIS 표본 특성 및 가중치 부여 방식에 따른 자료 활용 방법	84
1. 서론	84
2. KReIS 표본 추출 방법	86
3. 가중치 부여 방법 검토	89
4. 표본 특성에 따른 가중치 활용 방법	99
5. 논의	114
<부록1> 국민노후보장패널조사 표본 추가 방향 설정 연구	115
<부록2> 국민노후보장패널조사 무응답대체 적용방법에 관한 연구	148
<참고문헌>	193

<표 차례>

<표Ⅱ-1> 연구대상의 인구사회학적 특성	19
<표Ⅱ-2> 2005~2009년 노인의 가구유형 변화	21
<표Ⅱ-3> 노인 가구유형 조건부 전이확률 분석: 전체	22
<표Ⅱ-4> 노인 가구유형 조건부 전이확률 분석: 성별	23
<표Ⅱ-5> 노인 가구유형 조건부 전이확률 분석: 연령대	24
<표Ⅱ-6> 노인 가구유형 조건부 전이확률 분석: 공적연금 수급여부	25
<표Ⅱ-7> 노인 가구유형에 영향을 미치는 요인 분석	29
<표Ⅲ-1> 군집유형별 소비지출 구성비	40
<표Ⅲ-2> 군집유형별 가구지출 구성집의 타군집 지출구성비 평균 대비 비	41
<표Ⅲ-3> 가구 지출 유형별 가구 및 가구원 특성의 기술통계치	43
<표Ⅲ-4> 가구 지출 유형에 영향을 미치는 요인 분석	48
<표Ⅳ-1> 의사결정나무의 구성요소	54
<표Ⅳ-2> 은퇴경로별 평균 퇴직 연령 및 비중	64
<표Ⅳ-3> 은퇴경로별 인구학적 특성 및 생애주된 일자리 특성 ...	69
<표Ⅳ-4> 은퇴경로별 소득원천	74
<표Ⅳ-5> 분석에 사용된 변수	77
<표Ⅳ-6> 은퇴경로별 은퇴 후 소득의 이익도표	82
<표Ⅴ-1> KReIS 조사구 현황	87
<표Ⅴ-2> 유효 조사가구 및 참여가구 현황	88
<표Ⅴ-3> 모집단과 표본자료의 분포	90
<표Ⅴ-4> 연령 및 성별 가구가중치 보정 factor	93

<표 V-5> 연령 및 성별 개인가중치 보정 factor	93
<표 V-6> 3차까지 진행된 패널조사의 무응답 패턴	95
<표 V-7> tree모형에 쓰인 변수	97
<표 V-8> 조사 차수에 따른 가구의 거주지 및 주택 분포 비교	105
<표 V-9> 조사 차수에 따른 가구의 특성 비교	106
<표 V-10> 조사 차수에 따른 가구주의 특성 비교	108
<표 V-11> 조사 차수에 따른 개인조사응답자의 특성 비교	109

<그림 차례>

[그림 1] 2005~2009년 노인의 가구유형 변화	21
[그림 2] 노인빈곤가구 및 비빈곤가구 지출구성	38
[그림 3] 노인빈곤가구 및 비빈곤가구 소득구성	39
[그림 4] 군집별 소득구성	45
[그림 5] 은퇴경로의 유형화	61
[그림 6] 은퇴경로별 은퇴 후 소득액(2008년 기준)	72
[그림 7] 은퇴경로별 은퇴 후 소득의 결정요인	81
[그림 8] 만 50세 이상 가구주 연령 분포 비교	101
[그림 9] 1차년도와 3차년도 만 50세 이상 가구주 연령 분포 비교	102
[그림 10] 만 50세 이상 개인조사응답자의 연령 분포 비교	103
[그림 11] 종단면 가중치 적용 후 개인조사대상자 연령 분포	112
[그림 12] 종단면 가중치 적용 후 전체 개인조사대상자 연령 분포	113

<요 약>

I. 서론

○ 우리나라는 2028년 고령화 사회로 진입할 것으로 예상됨. 따라서 고령사회에 대비하기 위한 기초자료가 필요한 상황임.

- 이에 국민연금연구원에서는 우리나라 중·고령층의 노후준비 및 노후생활을 파악하여 노후소득보장과 관련된 자료를 구축하고자 2005년부터 국민노후보장패널조사를 실시함.

○ 본 보고서는 2009년 조사된 3차년도 국민노후보장패널에 대한 심층보고서로, 다음과 같이 구성됨.

- II. 노인의 가구유형 변화 및 영향을 미치는 요인
- III. 노인가구의 빈곤여부에 따른 가구 지출 유형화 및 영향 요인 분석
- IV. 우리나라 중·고령자의 은퇴경로 유형과 은퇴 후 소득 비교
- V. KReIS 표본 특성 및 가중치 부여 방식에 따른 자료 활용 방법
- <부록1> 국민노후보장패널조사 표본 추가 방향 설정 연구
- <부록2> 국민노후보장패널조사 무응답대체 적용방법에 관한 연구

II. 노인의 가구유형 변화 및 영향을 미치는 요인

○ 노인의 가구유형 변화는 가족구성원에 의한 부양서비스 제공 여부, 동거에 따른 경제적 지원 정도의 변화 등으로 노인들의 복지에 큰 영향을 미치며, 노인 단독가구의 경우 가구구성원에 의한 부

양서비스를 제공받기 어렵기 때문에 학문적·정책적 관심이 필요함.

○ 이러한 연구 필요성에 따라 1차년도 조사 당시 60세 이상 노인이 포함되어 있는 가구를 추출하여 노인단독가구, 노인부부가구, 성인가구원 동거, 기타가구의 가구유형으로 분류하고, 1차부터 3차에 걸쳐 노인의 가구유형 변화를 살펴보았음.

- 노인의 가구유형 변화는 조건부 전이확률을 통해 분석하였으며, 패널 프로빗 분석을 통해 노인의 가구유형에 영향을 미치는 요인(노인의 성별, 연령대, 질환 유무, 자녀의 수, 돌봐주는 손자녀 유무 등)을 살펴보았음.

○ 노인의 가구유형은 부부가구의 비중이 가장 높았으며, 성인가구원 동거, 단독가구, 기타가구 순으로 분포하는 것으로 나타남.

○ 가구유형의 이동확률은 대부분의 가구유형에서 큰 변화는 나타나지 않았으나 단독가구의 경우 다른 가구유형으로 이동할 확률이 가장 낮고, 기타가구의 경우 가구유형에서 이동이 가장 활발하게 나타나고 있었음. 이를 성별, 연령대, 공적연금 수급여부로 구분하여 각 집단의 조건부 이행확률을 살펴본 결과 집단의 특성에 따라 조건부 이행확률이 다르게 나타나고 있었음.

○ 한편 각 가구유형에 속하는 노인의 특성을 살펴본 결과, 노인 단독가구에 속할 가능성은 여성이며, 연령대가 높고, 질환을 갖고 있을 경우, 경제적으로도 매우 취약한 상태의 노인들이 단독가구에 속할 가능성이 높은 것으로 나타남.

- 따라서 단독가구에 속한 노인의 소득보장정책뿐만 아니라 심리적·사회적 서비스가 요구됨.

○ 부부가구의 경우 단독가구에 비해 상대적인 상황은 나은 편이었으나 가구의 경제적 상황이 좋지 않은 것으로 나타났음. 따라서 단독가구로의 변화 가능성이 크고, 사회서비스의 취약계층에 속할

가능성이 높은 이들을 위한 여러 소득보장정책의 확대가 요구됨.

○ 성인가구원 동거 가구의 경우 돌봄 손자녀가 있을수록 이 유형에 속할 가능성이 높아 노인의 필요에 의한 동거라기보다 자녀의 필요에 의한 동거가 증가하고 있음을 판단할 수 있음.

○ 기타가구원의 경우 중소도시에 거주할수록, 돌보는 손자녀가 있을수록 이 유형에 속할 가능성이 높은 것으로 나타나 대부분 조손가구일 것으로 추정됨. 특히 가구소득이 낮아 취업하고 있을수록 기타가구에 속할 가능성이 높은 것으로 나타나 조손가구의 경제적 상황을 지원할 수 있는 사회복지정책의 필요성이 제기됨.

Ⅲ. 노인가구의 빈곤여부에 따른 가구 지출 유형화 및 영향 요인 분석

○ 본 연구는 노인가구의 가구 지출 및 이들에 영향을 미치는 요인이 무엇인지 살펴보고자 하며, 이때 노인으로만 이루어진 가구를 분석대상으로 함.

○ 국민노후보장패널 3차년도 자료를 분석하였으며, 65세 이상 노인단독가구와 노인부부가구의 가계지출 구조를 분석함.

○ 노인가구의 '작년 한해'의 가구 지출을 활용하였으며, 가구 지출 유형별 소득 구성도 함께 고려. 이때 소득 및 지출의 특성은 총 지출 대비 각 지출 비중과 총 소득 대비 각 소득 비중으로 살펴봄.

- 노인가구의 빈곤여부에 따른 가구 지출 비교를 위해 가구 총 소득이 최저생계비 미만인 경우 빈곤, 이상인 경우 비빈곤으로 상정하여 분석.

- 로짓분석에서는 k-평균 군집분석을 통해 분류된 각 군집의 값을

1, 0으로 부여하여 군집에 따른 분석을 실시하였고 인구사회학적 요인, 가구특성 요인을 투입하여 분석함.

○ 분석결과는 다음과 같음.

- 첫째, 빈곤가구는 비빈곤가구에 비해 식비나 주거비, 광열수도비와 같은 영역이 전체 가계지출에서 더 많은 비중을 차지하며, 비빈곤가구의 경우 기타소비지출이나 비소비지출 영역이 더 많음. 보건의료비의 경우 노인가구에서 많은 비중 차지함. 빈곤가구 및 비빈곤가구의 소득 구성에서 빈곤가구는 비빈곤가구보다 이전소득의 비중이 높고 비빈곤가구는 시장소득의 비중이 높게 나타남.

- 둘째, 노인가구의 가구 지출 유형을 살펴보기 위해 군집분석을 살펴본 결과, 4가지 유형으로 구분 가능하였으며, 이를 각각 기본생활지출 중심형, 주거지출 중심형, 관계지출 중심형, 의료지출 중심형으로 명명함. 주거지출 중심형의 경우 가장 열악한 특성을 가진 집단으로, 빈곤하면서 대도시에 거주하는, 자가를 소유하지 않은 노인가구의 특성을 보임. 기본생활지출 중심형의 경우 소득 측면에서는 저소득층의 특성을 나타내지만 중소도시에 거주하면서 자가를 소유한 노인가구이며, 의료지출 중심형의 경우 차상위계층이면서 건강이 좋지 않아 의료비에 많은 지출을 하는 노인가구를 나타냄. 관계지출 중심형의 경우 비교적 여유로운 집단을 나타냄. 사례수가 가장 많은 군집은 기본생활지출 중심형임.

○ 결론적으로 주거지출 중심형의 경우 국민기초생활보장제도에서의 주거급여 현실화 및 대도시에 자가를 소유하지 않은 독거노인에 대한 주거관련 정책 및 서비스의 필요성이 제기되며, 의료지출 중심형의 경우 차상위계층에 대한 의료급여의 확대 및 노인을 대상으로 한 각종 의료서비스의 확대 필요성이 제기됨.

IV. 우리나라 중·고령자의 은퇴경로 유형과 은퇴 후 소득 비교

○ 은퇴에 관한 연구들이 은퇴를 하나의 사건(event)만으로 바라보고 있으나 은퇴는 생애사에 있어 하나의 과정(process)으로 바라보아야 함. 그리고 현재까지 은퇴를 다양한 측면에서 바라보고 있으나 생애 근로경험과 은퇴 이후의 경제적 복지수준을 연계하여 파악하는 은퇴에 대한 연구가 드문 것이 현실임.

○ 이에 본 연구에서는 국민연금연구원에서 수행하고 있는 KReIS의 직업력 자료와 3차년도 개인자료의 소득정보를 활용하여 우리나라 중·고령자 세대에서 발견되는 다양한 은퇴경로를 유형화하고 이에 따른 은퇴 후 소득 수준의 차이를 살펴보았음.

○ 분석결과는 다음과 같음.

- 첫째, 우리나라 2009년 현재 만 50세 이상의 중·고령자 중 미은퇴자는 44.7%이며, 생애 주된 일자리에서 은퇴한 완전은퇴자는 36.0%, 가교일 자리를 통한 점진적 은퇴자는 19.3%인 것으로 나타남. 이는 완전은퇴와 더불어 점진적 은퇴도 하나의 주된 은퇴 유형으로 자리 잡고 있음.

- 둘째, 은퇴 후 소득 수준은 은퇴경로에 따라 큰 차이를 보이고 있음. 은퇴 후 소득 수준은 가교 일 자리를 통해 은퇴한 경우보다 생애 주된 일자리에서 은퇴한 경우가 높으며 그리고 이들 중에서는 공적연금을 수급 받고 있는 집단의 소득 수준이 높게 나타나고 있음. 이는 은퇴를 하기 이전에 일자리에서의 고용상태와 연금 가입여부가 은퇴 후 소득을 결정함에 있어 중요한 요인으로 작용하고 있음을 보여줌.

- 셋째, 의사결정나무모형으로 은퇴 후 소득에 대한 결정요인을

살펴본 결과에 의하면, 은퇴 후 소득수준에 가장 영향력이 있는 변수는 최종학력이고 그 다음이 생애 근로 경험에 따른 은퇴경로의 유형이며 이들 가운데에서는 공적연금 수급여부가 개인소득 수준의 차이를 가져온 중요한 요인인 것으로 나타남.

V. KReIS 표본 특성 및 가중치 부여 방식에 따른 자료 활용 방법

○ KReIS 1차~3차년도 자료 표본의 특성을 분석하고, 이러한 표본이 대표성을 갖기 위해 적용하여 분석할 수 있는 가중치 활용 방법에 대하여 고찰함.

○ KReIS는 2005년 만 50세 이상의 가구원이 있는 전국 5,110 가구를 추출하고, 그 가구에 속하는 50세 이상의 가구원 및 그의 배우자를 개인조사대상으로 함. 표본설계 및 추출은 우리나라 50세 이상 인구 모집단을 대표하고 연령별, 성별, 도·농 지역별 독립분석이 가능하도록 설계함.

○ 기본 가중치는 표본 설계에서 사용된 표본 추출 확률을 바탕으로 하되, 표본 설계 당시 고려하지 못하였던 50세 이상 가구원을 가지는 가구 수의 분포를 이용하여 50세 이상의 가구에 대해 대표성을 가지도록 기본 가중치를 결정함.

○ 통계청에서 발표되는 지역별, 연령별, 성별 분포를 국민노후보장패널조사 표본에 시의성을 반영하도록 최종 가중치를 결정하였으며, calibration을 실제적으로 구현하는 방법으로 raking ratio adjustment 방법을 사용함.

○ 중·고령자 표본이라는 특수성 때문에 자료를 이용하여 분석을

할 때 대표하고자 하는 모집단을 설정함에 있어 주의가 필요함.

○ KReIS에서 종단면 가중치는 1~3차년도 계속 응답 가구 및 계속응답 개인조사대상자에 한하여 부여함.

○ KReIS의 자료의 경우 1차년도, 2차년도, 3차년도 자료가 각각 해당 조사년도인 2005, 2007년, 2009년을 대표하는 표본이 아니기 때문에 횡단면 분석과 종단면 분석 시 주의가 필요함.

○ 종단면 가중치를 적용하면 1차년도 조사 이후 2차년도, 3차년도 시간에 걸쳐 어떠한 이행을 보이는지 분석하기 용이하며, 특히 원표본이 대표하는 모집단이 50세 이상을 대표하므로 이후 은퇴 및 가족구성의 변화, 경제생활 및 노후준비 실태 관련하여 어떠한 행태를 보이는지 분석하기 위한 충분한 표본 수를 확보하였기 때문에 조사 목적에 부합한 연구가 가능함.

I. 서론

우리나라는 2008년에 전체 인구에서 65세 이상 노인이 차지하는 비율이 10.3%로 이미 UN이 규정한 고령화 사회로 진입했으며 유년인구의 감소로 인하여 세계에서 가장 빠르게 고령사회로 이행할 것으로 예상되고 있다. 이에 따라 우리나라는 재정안정화, 형평성 문제 등 국민연금 관련 문제가 크게 대두되고 있으나 노후소득보장과 관련한 통계적인 분석과 전망자료가 구축되어 있지 않아 저축과 근로, 세대 내, 세대 간 분배 등의 국민경제를 파악하기 어려운 상태이다. 또한 노후소득보장을 위한 연금정책의 결정 과정에서 국민들의 수요에 대한 명확한 판단 없이 공급자가 주된 정책결정을 내려 재정 불안정 등을 야기하고 있다.

급속히 진행되는 인구구조의 고령화는 근로인구의 감소와 저축의 감소 등으로 경제성장의 둔화를 야기할 수 있다. 뿐만 아니라 평균수명의 연장으로 개인의 노후설계와 복지수준에 상당한 영향을 미쳐 향후 연금 및 의료보장 등 사회보장 재정에 큰 부담으로 작용할 것이다. 이렇듯 고령화과정에서는 연금(공적연금, 기업연금, 개인연금 등), 저축과 자산, 가족, 건강, 퇴직, 근로, 소득 등에 있어 복잡한 인과관계가 존재하므로, 고령사회를 대비하기 위해서는 이에 대한 기초 자료가 필요하다.

이에 국민연금연구원은 우리나라 중·고령층의 노후준비 및 노후생활을 지속적으로 파악하여 노후소득보장과 관련된 정책을 위한 체계적인 자료를 구축하고자, 만 50세 이상 가구원을 가진 전국 5,000여 가구와 그 가구에 속하는 만 50세 이상 개인 8,600여 명을

추출하여 2005년부터 격년으로 국민노후보장패널조사(KReIS)를 실시하고 있다. 본 조사에서는 표본가구의 경제상황, 중·고령자의 고용현황 및 퇴직, 건강, 가족관계, 노후보장 현황 등 노후생활 전반에 걸친 주제들을 포괄적으로 조사한다. 현재 1차 본조사(2005년), 2차 본조사(2007년), 3차 본조사(2009년) 자료구축을 완료한 상태이다. 또한 격년으로 진행되는 본조사 사이에는 부가조사 형태로 노후소득보장과 관련된 당해 연도 이슈조사를 진행하고 있다. 1차 부가조사(2006년)에서는 50대 가구주 가구를 대상으로 가계지출 및 소득현황에 대해 심층조사를 실시하였으며, 2차 부가조사(2008년)에서는 2차 본조사 개인응답자를 대상으로 그들의 만 15세 이후 일자리 변동이력과 공적연금 가입/수급이력에 대해 심층조사를 실시하였으며, 현재 진행되고 있는 3차 부가조사(2010년)에서는 개인조사대상의 주관적 노후시기 해당여부를 판별하고, 노후준비와 노후관련 제도 인식에 대해 심층적으로 조사를 실시하고 있다.

본 연구는 다음과 같이 구성되어 있다. 제1장 머리말에 이어, 제2장부터는 국민노후보장패널 3차년도 조사항목 가운데 연구진에서 몇 가지 주제를 선택하여 심층분석을 실시하였다. 심층분석에서는 노인의 가구유형 변화 및 영향을 미치는 요인, 노인가구의 빈곤여부에 따른 가구 지출 유형화 및 영향 요인 분석, 그리고 중·고령자의 은퇴경로 유형과 은퇴 후 소득 비교, 마지막으로 국민노후보장패널 표본 특성 및 가중치 부여 방식에 따른 자료 활용 방법으로 구성하였다.

노인의 가구유형 변화 및 영향을 미치는 요인, 노인가구의 가구 지출 유형화 및 영향 요인 분석과 같은 주제는 국민노후보장패널에서 활용 가능한 주제이지만 이전 연구에서 많이 활용되지 않아 이에 대한 소개의 의미로, 은퇴 경로와 관련된 연구는 제2차 부가

조사와 제3차 본조사에서 제공하고 있는 직업력 조사를 활용한 결과의 의미로, 마지막으로 가중치와 관련된 영역은 국민노후보장패널 활용자들에게 관련 정보를 제공하는 것을 목적으로 기술되었다.

Ⅱ. 노인의 가구유형 변화 및 영향을 미치는 요인

1. 문제제기

2009년 현재 우리나라 총 인구 가운데 65세 이상 인구가 차지하는 비중은 10.7%로, 향후 9년 후인 2018년에는 고령사회(aged society), 2026년에는 초고령사회(super-aged society)에 도달할 것으로 전망(통계청, 2009)되는 등 우리나라는 급속한 고령화를 경험하고 있다. 이러한 빠른 고령화와 더불어 노인 가구유형의 변화도 급속한 변화를 맞고 있는데, 2008년 노인실태조사에서 60세 이상 노인이 포함된 가구의 유형을 살펴본 결과, 노인독거가구가 25.7%, 노인부부가구가 39.4%, 자녀 동거가구가 30.2%, 기타 노인가구가 14.7%로, 노인 3명 가운데 2명은 노인들로만 이루어진 가구에 살고 있는 것으로 나타났다.¹⁾

이러한 노인의 가구유형 변화는 가족구성원에 의한 부양서비스 제공 여부, 동거에 따른 경제적 지원 정도의 변화 등으로 노인들의

1) 1995년 인구주택총조사 보고서 고령자편에 나타난 60세 이상 고령자들의 세대 구성을 살펴본 결과 노인부부로만 구성된 가구는 24.6%, 노인 단독가구가 11.7%로 나타난 것(김경혜, 1998)과 이를 비교하면, 약 10년 동안 노인으로만 이루어진 가구의 비중이 크게 증가했음을 확인할 수 있다.

생활과 안녕에 지대한 영향을 미친다는 점에서 노인복지와 관련하여 매우 중요하게 다루어져야 할 주제라 할 수 있다. 특히 노인의 가구유형 중 노인 1인만 거주하는 노인 단독가구의 경우, 다른 가구유형의 노인들에 비해 가구구성원에 의한 부양서비스를 제공받기 어렵다는 측면에서(이민아·Weber, 2000) 이를 대체할 제도적 장치의 마련이 필요할 뿐 아니라, 빈곤과 같은 가구의 경제상황, 소외감 및 고독감과 같은 사회심리적 영역 등 노후생활의 여러 측면에서 취약점을 보이는 것으로 나타나기 때문에(Iacovou, 2000) 더 많은 사회서비스 제공을 위한 학문적·정책적 관심이 필요하다. 또한 노인을 대상으로 하는 주택정책이나 2008년부터 시행되고 있는 장기요양보험과 같은 사회복지서비스 수요에 중요한 영향을 미칠 것이라는 점에서 노인의 가구유형이 어떻게 변화하고 있는지, 그리고 노인의 가구유형에 영향을 미치는 요인은 무엇인지 살펴보고 이것이 시사하는 바를 정책적으로 제시하는 것이 필요하다.

그러나 주제의 중요성과 시의성에도 불구하고, 이와 관련된 선행 연구는 매우 부족한 상황이다. 노인의 가구유형에 관한 연구는 노인빈곤 문제를 살펴보면서 노인가구의 취약성을 분석하거나 독거노인의 건강 및 심리적 상태를 살펴보는 것과 같이 다른 주제를 분석하기 위한 도구로 다루어진 것이 대부분이다. 한편 본 연구와 같이 노인의 가구 유형을 살펴본 연구(김경혜, 1998)가 있으나 서울시에 국한되어 조사한 자료를 활용하였고, 조사가 한 시점에만 이루어져 노인가구 유형의 변화가 어떠한지, 노인의 특성에 따른 가구 유형의 변화가 어떠한지 살펴보기 어렵다는 한계가 있다. 그러므로 본 고에서는 KReIS의 1~3차(2005년, 2007년, 2009년) 자료를 활용하여 노인의 가구유형 변화를 살펴보고, 동일한 개인을 추적하여 가구 유형이 어떻게 변화하는지 그 이행율을 살펴볼 것이다. 그

리고 마지막으로 가구유형에 영향을 미치는 요인을 살펴보고, 이것이 함의하는 바를 제시하고자 한다.

2. 연구방법

가. 분석자료 및 연구대상

본 고의 분석자료는 KReIS 1~3차이다. KReIS는 2005년 기준으로 만 50세 이상 가구원이 있는 전국 5,000여 가구와 그 가구에 속하는 8,600여명의 가구원을 조사대상으로 이들의 경제활동 및 소득, 노후 준비, 노후소득보장 실태 등을 추적 조사하는 패널조사이다. KReIS의 본조사는 격년으로 실시하며, 본조사가 실시되지 않는 해에는 노후소득보장과 관련된 이슈를 조사하는 부가조사가 실시되고 있다(국민연금연구원, 2010).

본 고에서는 2005년에 60세 이상 노인이 포함되어 있는 가구를 추출하고, 이 가구를 노인단독가구, 노인부부가구, 성인가구원 동거, 기타가구로 구분하였다. 이후에 60세 이상 개인의 특성을 개인자료에서 추출하여 가구자료와 병합하였다. 이렇게 1차년도에 60세 이상의 가구원이 속한 가구 및 개인을 추출한 이후에 2차년도, 3차년도까지 모두 응답한 사례를 연구대상으로 선정하여 분석하였다.

나. 변수의 정의

본 고에서는 조사대상 노인들의 연령증가에 따른 가구유형 변화를 살펴보고, 특정한 가구유형에 속할 확률이 높은 노인의 특성은

어떠한 것인지를 파악하고자 하였다. 이는 노인이 속한 가구유형은 노인의 자원(resources)요인²⁾에 의해 결정되는 것이라는 연구결과(김정석, 2005)에서 기인한 것으로, 가구유형에 따라 노인의 특성이 이질적이며(이선형·김근홍, 2003) 가구구성원과의 상호지원교류정도에 차이가 있으므로(이민아·Weber, 2000) 노인의 노후생활에 대해 더욱 면밀한 정보를 제시해 줄 수 있을 것으로 생각된다.

노인의 가구유형은 노인가구의 소득, 소비, 자산, 빈곤을 분석한 많은 선행연구(김연정, 1998; 이선형, 2000; 심영, 2002; 이희숙·신상미, 2003; 최현수·류연구, 2003; 이윤정·김순미, 2004; 한수진·서지수, 2008)에 의해 다양한 형태로 분류되었다. 그 분류의 기준을 살펴보면 가구주의 연령이 일정한 기준 연령³⁾ 이상인 경우 노인가구로 정의하여 가구주의 성별, 배우자 유무, 취업여부, 은퇴여부 등에 따라 가구를 유형화하여 분석하였다. 그러나 이러한 기준으로 노인가구를 정의할 경우 가구의 경제상황 변화가 노인가구주 요인에 기인한 것인지, 다른 가구구성원에 의한 것인지를 파악하기 어렵다는 한계를 가지며, 가구 내에서 노인의 위치를 동일하게 분석함으로써 노인이 포함된 가구의 경제적 상황의 이질적 측면에 대한 정보를

2) 노인의 자원은 인구학적 자원(배우자와 자녀의 가용성), 건강자원(일상활동(ADL)의 제한여부), 경제적 자원(경제적 생활수단, 수입의 정도) 등으로 구분되며, 노인들의 가치 및 규범과 더불어 세대 간 거주형태의 중요한 요인이다(Choi, 1999; Da Vanzo and Chan, 1994; Goldscheider and Jones, 1989; Logan and Fuqin, 1999; Martin, 1989; 유성호, 1996)(김정석, 2005에서 재인용).

3) 대개의 연구들에서 노인을 정의하는 연령은 만55세 또는 만60세 또는 만65세이며, 만55세로 정의한 연구들에서는 고령자 고용촉진법과 통계청조사의 기준을 잣대로 삼았고(김연정, 1998; 이희숙·신상미, 2003), 만65세로 정의한 연구들에서는 노인복지법과 평균수명연장에 따른 노인인식 연령의 증가를 이유로 하였으며(이선형, 2000; 최현수·류연구, 2003), 만60세로 정의한 연구들에서는 국민연금법상 연금수급연령과 통상적 사회퇴직연령을 기준으로 하였다(심영, 2002; 이윤정·김순미, 2004, 한수진·서지수, 2008). 이에 따라 본 연구에서는 선행 연구 흐름과 본 연구의 목적을 모두 고려하여 만 60세를 노인으로 정의하는 연령으로 설정하였다.

충분히 제시할 수 없게 된다.

이에 따라 본 고에서는 개별노인이 누구와 함께 동거하는지를 기준으로 가구유형을 구분하였다. 특히 가구구성원 중 경제활동연령대(만18~59세)의 비노인가구원이 한 명이라도 같이 살고 있는 가구의 경우와 그 반대의 경우를 구분하여 분석하였는데, 최현수와 류연구(2003)의 연구에 의하면 경제활동연령대의 가구원이 한 명이라도 같이 살고 있는 가구의 경우 노인빈곤 구성에 차이가 있으며, 그러한 가구유형의 절대빈곤율과 상대빈곤율이 다른 가구에 비해 월등히 낮고(각 14.5%, 16.9%), 소득충족률은 상대적으로 높은 결과(85%)를 보였기 때문이다. 홍백의(2005)의 노인빈곤의 원인에 관한 연구에서도 노인의 경제적 상태나 빈곤에는 경제활동을 하는 사람과의 동거여부가 유의한 영향을 미치는 것으로 나타나 본 연구에서는 이를 구분하여 살펴보고자 하였다.

다. 변수의 측정

본 고에서는 앞서 언급한 정의에 따라 가구유형변수를 노인 1인 가구인 '노인단독가구', 부부 중 한명이라도 노인에 속하는 '노인부부가구', 18세 이상 60세 미만의 가구원이 동거하는 '성인가구원 동거가구', 60세 이상의 노인들과 18세 미만의 가구원이 동거하는 '미성년가구원 동거가구'의 4가지 유형으로 분류하였다. 이후 패널프로빗 분석에서는 4가지 노인 가구유형의 각 값에 해당하면 1로, 나머지는 0으로 부여하여 분석을 실시하였다⁴⁾. 패널프로빗 분석에 투

4) 일반적으로 종속변수가 3개 이상의 명목변수일 때에는 준거집단(1집단)을 기준으로 한 다항로짓분석(Multinomial Logit Analysis)을 실시한다. 그러나 이러한 다항로짓분석은 준거집단과 다른 집단의 차이를 보기에는 적합하지만 비준거집단들 사이의 차이를 밝히지 못하는 한계가 있다는 점이 있기 때문에, 여기에서

입된 독립변수들은 인구사회학적 요인, 가구특성 요인으로 구분된다.

우선 연구 대상의 인구사회학적 요인으로 성별은 여성을 0, 남성은 1의 값을 부여하였으며, 연령의 경우 연령대를 구분하여 60대의 경우 1, 70대는 2, 80대는 3으로 투입하여 분석하였고 교육년수는 연속변수 형태로 분석에 투입했다. 또한 취업유무는 비취업을 0, 취업을 1로 측정하였고, 질환유무 변수는 질환이 없는 경우 0, 질환이 있는 경우 1의 값을 부여하여 분석하였다. 그리고 공적연금 수급 여부 또한 노인의 가구유형 선택에 영향을 미칠 수 있는데, 수급자의 경우 노인의 독립적인 경제력을 반영하는 것으로 간주하여 수급자의 경우에는 1, 공적연금 비수급자의 경우에는 0을 부여하였으며 취업 상태이면 1, 비취업 상태인 경우에는 0 값을 부여하였다. 마지막으로 자녀 수 또한 노인의 가구유형 선택에 영향을 미칠 수 있으므로 자녀의 수를 투입하였다. 한편 노인의 필요에 의한 가구유형의 선택 뿐 아니라 최근 맞벌이의 증가로 손자녀 돌봄과 같은 영역에서 자녀들의 필요에 의해 노인과 동거할 가능성이 있으므로, 이를 반영하기 위한 변수로 돌보고 있는 손자녀가 있는 경우 1의 값을, 그렇지 않은 경우 0의 값을 투입하여 분석하였다.

다음으로 가구특성 요인으로 주택점유형태는 전월세가구를 0, 자가점유가구인 경우 1의 값을 부여하여 측정하였다. 거주 지역은 서울을 기준으로 광역시 더미, 중소도시 더미를 생성하여 분석하였으며, 가구균등화 지수를 적용한 가구총소득을 분석에 투입하였다.

는 프로빗 분석을 실시하였다.

- 5) KReIS의 개인자료에서는 개인조사대상자의 개인소득과 자산을 조사하고 있으나, 1~2차 자료의 경우 세대단위로 측정된 문항이 있어 개인의 총소득과 총자산을 산출하기 어렵다는 문제가 있다(물론 3차부터는 개인의 소득 및 자산을 산출가능하다). 따라서 여기에서는 공적연금 수급 여부를 노인 개인의 경제력을 반영하는 변수로 투입하여 분석하였다.

라. 분석방법

본 고에서는 2005년~2009년 노인의 가구유형 변화 및 이에 영향을 미치는 요인을 살펴보기 위해 2005년~2009년 노인의 가구유형 변화를 살펴보고, 2005년~2009년 노인의 가구유형 변화를 조건부 전이확률을 통해 분석하도록 하였다. 마지막으로 패널 프로빗 분석을 통해 노인의 가구유형에 영향을 미치는 요인을 살펴보았는데, 여기에는 노인의 성별, 연령대, 최종학력, 질환 유무, 자녀의 수, 돌봐주는 손자녀 유무, 공적연금 수급 여부, 취업 유무, 주택점유 형태 및 가구의 소득, 거주 지역을 독립변수로 투입하였다.

3. 분석결과

가. 연구대상의 인구사회학적 특성

연구대상의 인구사회학적 특성은 다음 <표 II-1>에 나타나 있다. 우선 전체 연구대상자의 특성을 살펴보면, 여성의 비중이 61.93%로 38.07%인 남성보다 높은 것을 알 수 있는데 이는 본 고의 대상이 노인인 것에서 비롯된 것으로 보인다. 연령대는 60대가 54.99%, 70대가 35.97%, 80대 이상이 9.04%로 분포하고 있으며, 교육년수의 평균은 6.35년으로 초등학교 졸업 수준임을 확인할 수 있다. 다음으로 비취업자의 비중은 61.29%로 취업자의 38.71%보다 많으며, 질환이 있다고 응답한 비중이 62.53%로 없다고 응답한 37.47%보다 많았다. 자녀 수의 평균은 3.76명이었으며, 현재 손자녀를 돌보고 있는지의 여부에는 8.48%만이 '그렇다'라고 응답하였고, 마지막으로

공적연금을 수급받고 있는 비중은 전체의 22.73%로 나타났다.

다음으로 가구특성을 살펴보면, 가구균등화 소득의 평균은 연 1,207만원으로 나타났으며, 자가에 거주하는 가구의 비중이 77.72%로 전월세에 거주하는 가구의 비중(22.28%)에 비해 3.5배 가량 높게 나타났다. 마지막으로 거주 지역을 살펴본 결과, 중소도시에 거주하는 비중이 59.36%로 가장 높았으며, 광역시에 거주하는 비중이 24.33%, 서울에 거주하는 비중이 16.31%로 가장 낮게 나타났다.

다음으로 이를 가구유형에 따라 구분하여 살펴보도록 하자. 먼저 단독가구의 특성을 살펴보면, 여성의 비중이 다른 가구유형에 비해 매우 높았으며, 80대 이상의 비중이 높고, 교육년수가 가장 낮았다. 또한 비취업자의 비중이 높고 질환을 갖고 있는 비중과 전월세의 비중이 높으며, 가구소득도 다른 유형에 비해 가장 낮은 것으로 나타났다. 이러한 결과는 한국노동패널조사를 활용하여 노인단독가구의 높은 상대적 빈곤율을 분석한 박능후·송미영(2006)의 연구결과를 일반적 특성을 통해 살펴보는 것이라고 할 수 있는데, 우리나라 노인단독가구에 속한 노인의 인적자본 및 경제적 상황은 매우 열악하다고 할 수 있다. 반면 부부가구에 속한 노인들의 경우 남성의 비중이 가장 높고 비교적 연령대가 낮은 것으로 나타났다. 또한 취업자의 비중도 높고 공적연금을 수급 받는 노인의 비중이 높은 것으로 나타나 단독가구에 비해 경제적 상황 및 노인 개인의 인적특성도 좋은 것으로 나타났다. 곧 노인만으로 이루어진 가구라 하더라도 노인 단독가구와 노인 부부가구의 경제적 상황은 차이가 있었으며, 본 고에서 살펴본 지표들로 볼 때 노인 부부가구의 상황이 더 나은 것으로 판단된다.

<표 II-1> 연구대상의 인구사회학적 특성

(단위: %, 만원)

구분		단독 가구	부부 가구	성인 가구원 동거	기타 가구	전체
성별	여	86.63	51.23	63.53	64.87	61.93
	남	13.37	48.77	36.47	35.13	38.07
연령대	~60대	33.22	55.29	64.81	55.90	54.99
	70대	51.38	38.24	26.42	30.00	35.97
	80대~	15.40	6.47	8.77	14.10	9.04
교육년수		3.69	6.88	6.82	5.52	6.35
취업유무	비취업	72.36	54.76	64.60	58.46	61.29
	취업	27.64	45.24	35.40	41.54	38.71
질환유무	없음	26.51	36.81	43.80	34.36	37.47
	있음	73.49	63.19	56.20	65.64	62.53
자녀수		4.11	3.89	3.62	3.90	3.76
돌봄 손자녀 여부	없음	97.80	95.27	86.40	66.41	91.52
	있음	2.20	4.73	13.60	33.59	8.48
공적연금 수급	비수급	78.68	76.22	78.01	76.41	77.27
	수급	21.32	23.78	21.99	23.59	22.73
가구균등화 소득		774.46	1108.76	1579.54	796.42	1207.09
주택집유형태	전월세	46.36	16.08	18.40	25.64	22.28
	자가	53.64	83.92	81.60	74.36	77.72
거주 지역	서울	10.89	15.45	20.61	10.26	16.31
	광역시	24.87	20.56	29.06	22.56	24.33
	도	64.24	63.99	50.33	67.18	59.36

주 1) 모든 수치는 가중치 적용

2) 모든 수치는 패널자료의 기술통계값에서 overall 값을 제시함.

한편 성인가구원과 동거하는 노인의 특성을 살펴보면, 60대의 비중이 가장 높은 것으로 나타났으며, 질환을 갖고 있는 노인의 비중이 다른 가구유형과 비교할 때 가장 높았다. 이를 통해 혼자 독립적으로 생활하기 어려운 노인들이 자녀들과 동거하는 것으로 파악할 수 있다. 또한 돌봄 손자녀가 있다고 응답한 경우도 단독가구와

부부가구에 비해 높은 것으로 나타났다. 반면 18세 이하의 가구원과 노인으로 이루어진 기타가구에서 가장 두드러지는 특성은 돌봄 손자녀가 있다고 응답한 비중이 33.59%로 다른 가구 유형과 비교할 때 가장 높다는 점이다. 또한 가구소득에 있어 단독가구보다 약간 높은 편이고, 대부분 중소도시에 거주하는 것으로 나타났는데, 이러한 특징들은 기타가구에 조손가구가 상당수 포함되어 있음을 추정케 한다. 이후 어떠한 특성을 가진 노인들이 각 가구유형에 속할 확률이 높은지 패널프로빗 분석을 통해 살펴보도록 한다.

나. 노인의 가구유형 변화

1차년도(2005년)에 60세 이상인 사람들이 속한 가구유형이 2007년과 2009년에 어떻게 변했는지 살펴본 결과는 다음 <표 II-2>와 같다. 이 때, 본 고의 연구목적에 따라 KReIS 2005년~2009년에 응답한 노인이 속한 가구유형 변화가 어떠한지 파악하기 위해 분석의 단위는 개인으로 설정하여 분석하였다.

우선 노인이 속한 가구의 유형은 부부가구의 비중이 가장 높고 성인가구원 가구, 단독가구, 기타가구의 순으로 나타났다. 다음으로 각 개인이 속한 가구의 유형을 시간의 흐름에 따라 살펴본 결과, 단독가구는 2005년 13.5%에서 2007년 15.2%, 2009년 18.0%로 증가하고 있으며, 부부가구 또한 42.6%에서 43.6%, 44.5%로 증가하고 있음을 알 수 있다. 이는 곧 시간이 지남에 따라 노인으로만 이루어진 가구가 증가하고 있음을 알 수 있는 결과이다. 반면 성인가구원이 동거하는 가구의 경우 2005년 40.3%에서 2007년 37.7%, 2009년 33.8%로 감소하고 있다. 이는 시간이 흐름에 따라 자녀의 혼인 등으로 가구원이 분가하는 것과 관련이 있는 것으로 보인다. 기타

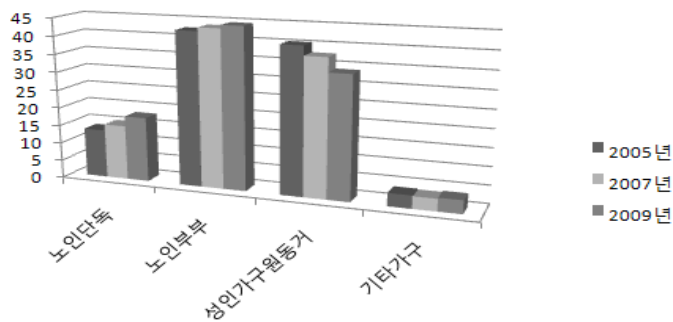
가구의 경우에는 2005년 3.6%에서 2007년 3.4%로 감소하였다가 2009년 다시 3.7%로 증가하고 있다. [그림 1]은 이러한 변화를 그래프로 나타낸 것이다.

<표 II-2> 2005~2009년 노인의 가구유형 변화

가구유형	1차(2005년)	2차(2007년)	3차(2009년)
노인단독가구	13.5	15.2	18.0
노인부부가구	42.6	43.6	44.5
성인가구원 동거	40.3	37.7	33.8
기타가구	3.6	3.4	3.7

주 1) 각년도 개인가중치를 적용한 결과임

[그림 1] 2005~2009년 노인의 가구유형 변화



다음으로 패널데이터를 활용하여 노인의 가구유형이 시간에 따라 어떻게 변했는지 살펴보도록 한다. 여기에서는 현재(t기) 상태가 다음 시기(t+1기)에 그 상태가 될 확률을 의미하는 조건부 전이확률(conditional transition probability)을 통해 살펴보았다(민인식·최필선, 2009). 우선 여기에서는 전체 노인을 대상으로 분석하고, 노인의 특성 가운데 성별, 연령대와 공적연금 수급여부를 가지고 집단을 구분하여 조건부 전이확률을 분석하였다.

우선 전체 노인을 대상으로 조건부 전이확률을 분석한 결과는 다음 <표 II-3>과 같다. 이를 살펴보면, 대부분의 가구유형에서 큰 변화는 나타나지 않았으나 단독가구에 속했던 노인의 경우 다른 가구유형에 비해 상대적으로 다른 가구유형으로 이동할 확률이 매우 낮은 것으로 나타났다.

<표 II-3> 노인 가구유형 조건부 전이확률 분석: 전체

t	t+1				
	단독가구	부부가구	성인가구원 동거	기타가구	전체
노인단독가구	0.9735	0.0027	0.0192	0.0046	1.00
노인부부가구	0.0374	0.9292	0.0295	0.0038	1.00
성인가구원동거	0.0249	0.0905	0.8631	0.0215	1.00
기타가구	0.0606	0.1402	0.1174	0.6818	1.00
전체	0.1776	0.4503	0.3366	0.0356	1.00

다른 가구유형에 비해 변화가 가장 적긴 하지만 단독가구에 속했던 노인의 경우 성인가구원 동거로 이동할 확률이 높은 것으로 나타났다. 부부가구에 속했던 노인의 경우에는 단독가구로 이동할 확률이 높은 것으로 나타났는데, 이는 부부 중 한 사람의 사망으로 인한 가능성이 큰 것으로 보인다. 다음으로 성인가구원 동거 가구에 속했던 노인의 경우에는 부부가구로 이동할 확률이 높았으며, 기타가구에 속했던 노인의 경우 가구유형의 변화가 가장 컸는데, 부부가구, 성인가구원 동거가구로 이동할 확률이 높았다.

다음으로 노인의 특성에 따라 노인이 속한 가구유형의 이동 확률을 분석한 결과는 다음 <표 II-4>와 같다. 먼저 성별에 따라 살펴본 결과, 여성의 경우 남성에 비해 단독가구를 유지하는 확률이 높았으며 다른 가구유형에서 단독가구로 이동하는 확률이 높은 것은

로 나타났다. 또한 전체 비중에 있어서도 남성에 비해 단독가구에 속할 확률이 높은 것으로 나타났다. 이는 앞선 가구유형에 따른 일반적 특징에서 설명한 바와 같이 경제적으로 어려움을 경험할 가능성이 큰 여성노인이 단독가구를 이루며 생활할 경우 ①빈곤을 지속하는 경우가 많으며(배성우·손지아·박순미, 2008) ②가족구성원과 함께 사는 여성노인이 빈곤으로부터 벗어날 가능성이 높다는 연구결과(Dodge, 1995)에 비춰볼 때, 사회적 서비스 제공에 있어 경제적 어려움에서 벗어날 가능성이 적은 여성노인단독가구를 위한 선별적 정책마련의 필요성이 제시되는 결과라 할 수 있다.

<표 II-4> 노인 가구유형 조건부 전이확률 분석: 성별

성 별	t	t+1				전체
		단독가구	부부가구	성인가구 원동거	기타가구	
남	단독가구	0.9639	0.0066	0.0230	0.0066	1.00
	부부가구	0.0242	0.9287	0.0419	0.0052	1.00
	성인가구원동거	0.0200	0.1113	0.8431	0.0256	1.00
	기타가구	0.0342	0.1966	0.1453	0.6239	1.00
	전체	0.1034	0.4569	0.4048	0.0349	1.00
여	단독가구	0.9759	0.0021	0.0168	0.0052	1.00
	부부가구	0.0557	0.9089	0.0318	0.0037	1.00
	성인가구원동거	0.0334	0.0723	0.8737	0.0206	1.00
	기타가구	0.0809	0.1040	0.1329	0.6821	1.00
	전체	0.2473	0.3686	0.3471	0.0370	1.00

다음으로 연령대별 가구유형 이동을 살펴보면, 전체 비중에서는 연령대가 가장 낮은 60대의 성인가구원 동거 가구유형과 부부가구의 비중이 다른 가구유형과 비교할 때 가장 높은 것으로 나타났다. 이는 연령대가 낮을수록 배우자가 생존해 있을 가능성이 크고 시계적으로 성인가구원이 혼인 등으로 분가하기 전이기 때문인 것과

관련이 있는 것으로 보인다. 그런데 연령대가 높아질수록 단독가구를 유지하는 확률이 높아지고 성인가구원 동거 가구유형을 유지하는 확률이 높아지고 있음을 확인할 수 있다. 이는 연령대가 높아질수록 배우자의 사망 등으로 단독가구가 될 가능성이 높으며, 성인가구원과 동거했던 노인들의 경우 신체적 기능의 쇠퇴와 장애를 경험할 가능성의 증가, 경제적 자원의 감소 등으로 성인가구원으로부터 부양서비스를 계속 제공받기 위해 가구유형의 변화 없이 지속적으로 유지하는 것으로 판단된다.

<표 II-5> 노인 가구유형 조건부 전이확률 분석: 연령대

연령대	t	t+1				전체
		단독가구	부부가구	성인가구 원동거	기타가구	
60대	단독가구	0.9639	0.0066	0.0230	0.0066	1.00
	부부가구	0.0242	0.9287	0.0419	0.0052	1.00
	성인가구원동거	0.0200	0.1113	0.8431	0.0256	1.00
	기타가구	0.0342	0.1966	0.1453	0.6239	1.00
	전체	0.1034	0.4569	0.4048	0.0349	1.00
70대	단독가구	0.9746	0	0.0211	0.0042	1.00
	부부가구	0.0541	0.9316	0.0143	0	1.00
	가구원동거	0.0302	0.0698	0.8981	0.0019	1.00
	기타가구	0.1250	0.0972	0.0972	0.6806	1.00
	전체	0.2624	0.4654	0.2468	0.0253	1.00
80대 이상	단독가구	0.9853	0	0.0074	0.0074	1.00
	부부가구	0.0390	0.9286	0.0195	0.0130	1.00
	가구원동거	0.0223	0.0112	0.9162	0.0503	1.00
	기타가구	0	0.0645	0.0645	0.8710	1.00
	전체	0.2880	0.2940	0.3400	0.0780	1.00

다음으로 공적연금 수급여부에 따른 노인의 가구유형 조건부 전이확률을 살펴보도록 하겠다. 노인은 다른 소득원이 부족하기 때문에 공적연금의 수급여부가 노인 개인의 경제력을 반영하는 지표라

할 수 있을 것이다. 분석결과 공적연금을 수급 상태의 노인의 경우 비수급 노인에 비해 성인가구원 동거를 유지하는 비중이 낮은 것으로 나타났다. 곧 공적연금 수급 노인의 경우 성인가구원 동거 유형에서 단독가구나 부부가구로 이동할 확률이 비수급 노인에 비해 높다는 것이다. 이러한 결과는 노인의 세대간 상호 지원과 거주형태를 연구한 이민아·Weber(2000)의 연구에서 비슷한 결과를 찾을 수 있는데, 자립적인 노인들의 경우 가까운 장래에 노인단독가구로 변화할 계획이 있다는 것이다. 그러나 이는 경제적으로 충분한 원조가 있는 경우와 자기 자산을 관리할 능력이 있는 노인단독가구에 한하는 연구결과라 할 수 있다.

<표 II-6> 노인 가구유형 조건부 전이확률 분석: 공적연금 수급여부

수급	t	t+1				전체
		단독가구	부부가구	성인가구 원동거	기타가구	
수급	단독가구	0.9659	0.0098	0.0244	0	1.00
	부부가구	0.0258	0.9484	0.0230	0.0029	1.00
	성인가구원동거	0.0316	0.1022	0.8364	0.0297	1.00
	기타가구	0.1207	0.1897	0.0690	0.6207	1.00
	전체	0.1602	0.1897	0.3171	0.0360	1.00
비수 급	단독가구	0.9742	0.0012	0.0188	0.0059	1.00
	부부가구	0.0349	0.9318	0.0290	0.0043	1.00
	성인가구원동거	0.0228	0.0873	0.8727	0.0173	1.00
	기타가구	0.0448	0.1244	0.1294	0.7015	1.00
	전체	0.1800	0.4440	0.3407	0.0354	1.00

다. 노인의 가구유형에 영향을 미치는 요인

여기에서는 노인의 가구유형에 영향을 미치는 요인을 살펴보았다. 분석모델은 노인 가구유형에 따라 단독가구 모델, 부부가구 모

텔, 성인가구원 동거 모델, 기타가구원 모델로 구분하였으며, 각 가구에 해당하는 경우 1의 값을, 그렇지 않은 경우 0의 값을 부여하여 패널프로빗 분석을 실시하였다. 이를 통해 각 가구유형에 속할 확률이 높은 노인의 특성을 살펴보도록 한다.

우선 단독가구 모델을 살펴보면, 성별, 연령대, 교육년수, 취업유무, 질환유무, 자녀 수, 돌보는 손자녀 여부, 공적연금 수급여부, 가구소득, 주택점유형태 및 거주 지역 변수 모두 통계적으로 유의미했다. 이를 살펴보면 남성보다는 여성일수록, 연령대가 높을수록, 최종학력이 낮을수록, 취업자보다는 비취업자일수록, 질환이 있을수록 단독가구에 속할 가능성이 높은 것으로 나타났다. 또한 자녀가 적을수록, 돌보는 손자녀가 있는 것보다 없을수록, 공적연금을 수급 받을수록, 가구소득이 낮을수록, 자가에 비해 전월세 가구에 거주할수록, 서울에 거주하는 것보다는 광역시나 중소도시에 거주할수록 단독가구에 속할 가능성이 높은 것으로 나타났다. 이는 앞서 살펴본 조사대상 가구유형별 일반적 특성에서 나타난 바와 같이 단독가구의 경우 대부분 여성이면서 연령대도 높고 질환을 갖고 있는 것으로 나타나 인적자본이 낮으며, 가구의 경제적 상황도 그리 좋지 않다는 결과를 뒷받침하는 것이라 하겠다. 이는 노인과 관련된 여러 선행연구에서 지적하는 바와 같이 단독가구의 경우 노후생활의 안녕을 위한 자본의 구성 측면에서 취약한 특성을 갖고 있음을 확인할 수 있다.

다음으로 부부가구 모델에서는 성별, 연령대, 교육년수, 자녀 수, 돌보는 손자녀 여부, 공적연금 수급 여부, 가구소득, 주택점유형태, 거주 지역에서 통계적으로 유의미한 결과가 나타났다. 여성보다는 남성일수록, 연령대가 낮을수록, 그리고 교육년수가 높을수록 부부가구 모델에 속할 확률이 높았으며, 자녀수가 많을수록, 돌보는 손

자녀가 없을수록, 공적연금을 수급 받을수록, 가구소득이 낮을수록, 자가일수록, 서울보다는 중소도시에 거주할수록 부부가구 모델에 속할 확률이 높았다. 이를 통해 볼 때 부부가구유형의 경우 연령대도 비교적 낮고 자가의 주택을 점유하고 있는 점으로 볼 때 여러 상황이 단독가구에 비해 좋다고 할 수 있으나 단독가구와 마찬가지로 가구소득이 낮을수록 부부가구에 속할 가능성이 높기 때문에 정책적 관심이 필요한 가구유형이라 할 수 있을 것이다.

세 번째는 성인가구원 동거 모델인데, 여기에서는 연령대, 질환유무, 돌보는 손자녀 여부, 공적연금 수급여부, 가구소득, 주택점유형태와 거주 지역이 통계적으로 유의미한 것으로 나타났다. 이를 좀 더 자세히 살펴보면 연령대가 낮을수록, 질환이 있는 사람보다는 질환이 없는 사람이, 돌보는 손자녀가 있을수록, 공적연금을 수급하지 않을수록 성인가구원과 동거하는 모델에 속할 가능성이 높은 것으로 나타났다. 또한 가구소득이 많을수록, 주택점유형태가 자가일수록, 중소도시에 거주하는 것보다는 서울에 거주할수록 성인가구원 동거 모델에 속할 가능성이 높은 것으로 나타났다.

이러한 결과 가운데 질환이 있는 경우보다 없는 경우 성인가구원 동거 모델에 속할 가능성이 높다는 결과는 다소 의외의 결과인데, 이는 서울시 노인을 대상으로 노인의 동거형태 결정요인을 살펴본 김경혜(1998)의 연구에서도 언급된 내용으로, 이 연구결과에서는 건강이 나빠질수록 독신이나 부부, 자녀동거 가구보다 기타 가구원과 동거할 가능성이 높다는 것을 지적하고 있다. 이러한 결과가 나타난 이유에 대해 본 연구의 결과를 통해서도 정확히 추정하기 어렵지만, 가구유형에 따라 동년배와 비교한 주관적 건강상태 인지정도에 차이가 있으며, 만성질환의 수 등 건강상태 지표들의 차이가 있다는 전예슬(2009)의 연구결과와 비교하여 볼 때, 노인의 건강상태

가 가구유형을 결정하는 하나의 요인이라는 사실은 확인 할 수 있다. 다만, 건강상태에 따라 가구유형에 달라지는 이유는 무엇인지, 그리고 가구유형에 따라 노인이 갖는 건강상의 취약점은 무엇인지에 대해서는 좀 더 자세히 살펴볼 필요가 있겠다. 반면 돌보는 손자녀가 있을수록 성인가구원 동거 모델에 속할 가능성이 높다는 것은 노인의 가구유형이 노인의 필요뿐만 아니라 자녀의 필요에 의해서도 결정될 수 있다는 것을 제시한다고 할 수 있다.⁶⁾

6) KReIS에서는 손자녀를 동거 또는 비동거하는 초등학교 3학년 이하의 손자녀(친손자녀, 외손자녀 모두 포함)로 정의하고, 돌봐주고 있는 손자녀는 거주와 생계를 같이 하고 있는 손자녀와 가까운 곳에 거주하는 비동거 손자녀를 포함하여 응답자 및 응답자의 배우자가 거의 도맡아 돌봐주는 손자녀로 정의하고 있다.

<표 II-7> 노인 가구유형에 영향을 미치는 요인 분석

변수	단독가구 모델	부부가구 모델	성인가구원 동거 모델	기타가구원 모델
	B(S.E)	B(S.E)	B(S.E)	B(S.E)
성별(=남성)	-3.4546*** (.4882)	2.4717*** (.1860)	-.1534 (.1629)	-.1794 (.1671)
연령대	1.5057*** (.1543)	-.2933** (.0893)	-.6215*** (.0900)	.0637 (.1061)
교육년수	-.1731*** (.0332)	.1442*** (.0207)	-.0133 (.0181)	-.0102 (.0408)
취업유무(=취업)	-.3279* (.1579)	.2039 (.1176)	-.1345 (.1125)	.2763* (.1387)
질환 유무(=질환 있음)	.5612*** (.1302)	.0329 (.0902)	-.4059*** (.0905)	.0437 (.1164)
자녀 수	-.3623*** (.0728)	.2503*** (.0458)	-.0113 (.0379)	-.0120 (.0408)
돌보는 손자녀 여부 (=있음)	-1.7574*** (.3302)	-1.3334*** (.1405)	.9202*** (.1286)	1.4480*** (.1439)
공적연금 수급(=수급)	2.4369*** (.3232)	.6732*** (.1440)	-.6435*** (.1451)	.3878* (.1655)
ln 가구소득	-.8422*** (.1672)	-.9594*** (.1220)	1.7557*** (.1356)	-.9099*** (.1688)
주택점유형태(=자가)	-1.2623*** (.1711)	.3857** (.1239)	.4374*** (.1232)	-.0500 (.1521)
거주 지역(기준:서울)				
광역시 더미	1.1491** (.3633)	-.0519 (.2049)	-.4677 (.2937)	.3165 (.2576)
도 더미	1.5669*** (.3574)	1.0520*** (.1963)	-1.8939*** (.2621)	.4858* (.2343)
상수	-1.9901* (.8220)	-.9666 (.5412)	-6.7718*** (.5537)	-1.6239* (.6830)
number of obs	10,668			
number of groups	3557			
LR chi2(10)	167.61***	433.92***	335.40***	130.29***

주 1) *=p<.05. **=p<.01. ***=p<.001.

그리고 공적연금을 수급하지 않는 경우에 성인가구원 동거 가구 유형에 속할 가능성이 높다는 것은 노인이 독립적인 경제력이 없

는 경우 성인가구원과 동거하면서 경제적으로 의존할 가능성이 높다는 것을 의미하는 것이라 할 수 있겠다. 또한 가구소득이 높을수록 성인가구원 동거 모델에 속할 가능성이 높은 것은 노인의 소득보다는 다른 성인가구원의 소득원에서 비롯된 결과인 것으로 보인다.

한편 다른 가구유형과는 달리 중소도시보다는 서울에 거주할수록 성인가구원 동거 모델에 속할 가능성이 높은 것으로 나타났는데, 이는 노인의 특성에서 비롯된 것이라기보다는 노인과 비교할 때 경제활동을 비교적 활발히 하는 성인가구원들의 특성에서 비롯된 결과로 보인다.

마지막으로 기타가구원 모델의 경우에는 취업유무, 돌보는 손자녀 여부, 공적연금 수급 여부, 가구소득, 거주 지역 변수가 통계적으로 유의미하게 나타났다. 이를 살펴보면, 노인이 비취업자보다는 취업자일수록, 돌보는 손자녀가 없는 것보다는 있을수록, 공적연금을 수급 받을수록, 가구소득이 낮을수록, 서울보다는 중소도시에 거주할수록 기타가구원 모델에 속할 가능성이 높은 것으로 나타났다. 기타가구원 동거 모델의 경우 노인과 미성년자 가구원이 동거하는 가구를 의미할 뿐 아니라 돌보는 손자녀가 있을수록 이 가구유형에 속할 가능성이 높아 대부분 조손가구인 것으로 판단되는데, 취업자일수록 이 가구유형에 속할 가능성이 높은 반면 가구소득이 낮은 것으로 나타나 열악한 가구의 상황을 반영한다고 할 수 있다. 특히 대부분의 조손가구가 농어촌 지역에 거주하는 것으로 나타나는데, 본 고의 분석결과에서도 서울보다는 중소도시에 거주할수록 이 가구유형에 속할 가능성이 높은 것으로 나타났다.

4. 논 의

본 고에서는 노인의 가구유형 변화를 살펴봄에 있어 동일한 개인을 추적하여 가구 유형이 어떻게 변화하는지 그 이행율을 살펴보고, 특정한 가구유형에 속할 가능성이 높은 노인의 특성은 어떠한 것인지 살펴보았다.

분석결과, 노인의 가구유형은 부부가구의 비중이 가장 높았으며 성인가구원 동거, 단독가구, 기타가구의 순으로 분포하는 것으로 나타났다. 이를 2년 간격으로 시간의 흐름에 따른 변화를 살펴본 결과, 부부가구 및 단독가구의 비중은 높아지는 반면 성인가구원 동거의 비중은 낮아지는 것으로 나타났다. 또한 패널데이터를 활용하여 현재 상태가 다음 시기에 그 상태가 될 확률을 의미하는 조건부 전이확률을 통해 가구유형 이동확률을 살펴본 결과 대부분의 가구유형에서 큰 변화는 나타나지 않았으나 단독가구의 경우 다른 가구유형으로 이동할 확률이 가장 낮았고 기타가구의 경우 가구유형에서 이동이 가장 활발하게 나타나고 있다는 것을 확인할 수 있었다. 그리고 이를 성별, 연령대, 공적연금 수급여부로 구분하여 각 집단의 조건부 이행확률을 살펴본 결과 집단의 특성에 따라 조건부 이행확률이 다르게 나타나는 것을 확인할 수 있었다.

한편 각 가구유형에 속하는 노인의 특성을 살펴본 결과, 대부분 여성이면서 연령대가 높을수록, 또한 질환을 갖고 있을수록 노인 단독가구에 속할 가능성이 높아 노인 개인의 인적자본도 낮을 뿐 아니라 가구소득이 낮을수록, 전월세 가구일수록 단독가구에 속할 가능성이 높은 것으로 나타나 경제적으로 매우 취약한 상태임을 알 수 있었다. 따라서 단독가구에 속한 노인의 소득보장정책 뿐 아

나라 이들의 수발을 도울 수 있는 노인서비스의 확대 및 심리·사회적 서비스 또한 요구된다고 할 수 있다. 뿐만 아니라 부부가구의 경우에도 단독가구에 비해 상대적인 상황은 조금 낫다고 할 수 있겠으나, 가구의 경제적 상황은 그리 좋지 않은 것으로 나타나 이들의 노후생활을 보장할 수 있는 여러 정책의 확대가 요구됨을 제시한다.

다음으로 성인가구원 동거 가구의 경우에는 돌봄 손자녀가 있을수록 이 유형에 속할 가능성이 높아 노인의 필요에 의한 동거보다는 자녀의 필요에 의한 동거 또한 증가하는 것을 알 수 있었다. 반면 질병이 있는 것보다 질병이 없을수록 성인가구원 동거 가구에 속할 가능성이 높아 이에 대한 보다 심층적인 분석의 필요성을 제기하였다. 마지막으로 기타가구원의 경우 서울보다 중소도시에 거주할수록, 돌보는 손자녀가 있을수록 이 유형에 속할 가능성이 높은 것으로 나타나 대부분 조손가구일 것으로 추정된다. 특히 가구소득이 낮아 취업하고 있을수록 기타가구에 속할 가능성이 높은 것으로 나타났다. 따라서 조손가구의 경제적 상황을 지원할 수 있는 여러 사회복지정책의 필요성을 제기한다.

Ⅲ. 노인가구의 빈곤여부에 따른 가구 지출 유형화 및 영향 요인 분석

1. 문제제기

우리나라의 급속한 고령화에 따라 노인 가구의 경제적 상태가 어떠한지 측정하고자 하는 연구들이 매우 다양하게 이루어졌다. 본 고에서는 노인가구의 경제적 상태가 어떠한지 살펴보기 위하여 노인가구의 가구지출 및 이들에 영향을 미치는 요인이 무엇인지 살펴보고자 한다. 여기에서 가구 지출을 통해 이를 살펴보고자 하는 이유는 다음과 같다.

첫째, 일반적으로 경제적 복지는 소득을 지표로 이루어진 경향이 있지만 소득보다 소비가 더 나은 지표라는 점이 제시되고 있다는 점이다. Friedman의 항상소득가설에 따르면 가구의 경제 상태를 평가하기 위해서는 소비가 소득보다 더 나은 지표임을 시사하고 있는데, 소비는 소득에 비해 일시적인 요소에 의해 영향을 덜 받기 때문이다(여운경, 2003). 뿐만 아니라 노인과 같이 소득의 급속한 감소를 겪는 집단이나 취약계층의 경우 소득보다는 소비가 이들의 경제적 복지를 측정하기 위한 더 좋은 지표로 제시되고 있다(Meyer & Sullivan, 2007; 백학영, 2010에서 재인용). 따라서 노인가

구의 지출이 어떻게 구성되어 있는지 살펴보고, 이를 이들의 빈곤 지위 및 소득의 구성을 통해 총체적으로 살펴보고자 한다.

둘째, 가계지출 유형화를 통해 노인가구의 경제적 상태 뿐 아니라 이들의 욕구도 파악할 수 있다는 점이다. 개별 가구의 욕구는 결국 소비로 구현된다. 경제학에서는 소비를 '인간의 욕구를 충족시키기 위해 사용하는 재화와 서비스'로 정의하고 있다. 따라서 가구의 지출을 유형화함으로써 노인 가구의 욕구를 파악할 수 있을 것이라 판단된다. 따라서 노인가구의 지출유형을 파악함으로써 이들의 욕구 및 경제적 상태가 어떠한지 살펴볼 수 있다는 장점이 있다.

특히 여기에서는 노인으로만 이루어진 가구를 대상으로 이를 살펴보고자 하는데, 이는 노인으로만 이루어진 독립적인 가구가 증가하고 있다는 현상과 관련이 있다. 우리나라는 노인단독가구가 1990년에는 28%였으나 2000년에는 35%, 2020년에는 60%로 증가할 것으로 예상되고 있다(삼성경제연구소, 2000). 이렇게 노인으로만 이루어진 가구의 경우 여러 측면에서 매우 취약한 특성을 갖는 것으로 제시되고 있는 바, 가구 지출 및 소득 구성을 통해 이들의 경제적 상태를 살펴보고, 가구 지출을 유형화하여 이들의 욕구를 살펴보고자 한다.

2. 분석방법

가. 분석자료 및 연구대상

본 고의 분석자료는 KReIS 3차년도 자료이다. KReIS는 2005년 기준으로 만 50세 이상 가구원이 있는 전국 5,000여 가구와 그 가구에 속하는 8,600여명의 가구원을 조사대상으로 이들의 경제활동 및 소득, 노후 준비, 노후소득보장 실태 등을 추적 조사하는 패널 조사이다. KReIS의 본조사는 격년으로 실시하며, 본조사가 실시되지 않는 해에는 노후소득보장과 관련된 이슈를 조사하는 부가조사가 실시되고 있다(국민연금연구원, 2010).

본 고에서는 65세 이상 노인단독가구, 노인부부가구의 가계지출 구조를 분석하고자 3차 가구자료에서 65세 이상 노인이 속한 가구를 추출하고, 이 가운데 단독가구와 부부가구만을 추출하여 개인자료를 병합(merge)하여 활용하였다.

나. 변수측정

KReIS에서는 지출을 식비, 주거비, 광열수도비, 피복비, 교통비, 통신비, 문화생활비, 보건의료비, 가구집기 및 가사서비스, 교육비, 보육비, 기타소비지출, 비소비지출로 측정하고 있으며, '작년 한 해'와 '지난 달'로 구분하여 측정하고 있다. 여기에서는 '작년 한 해'의 가구 지출을 활용하였으며, 노인으로만 이루어진 가구에서 상대적으로 그 중요성이 덜한 교육비와 보육비를 묶어서 활용했다. 또한 여기에서는 가구 지출 유형별 소득 구성이 어떻게 되어 있는지도

살펴볼 것인데, 소득의 경우 KReIS의 분류에 따라 근로소득, 금융소득, 부동산소득, 공적이전소득, 사적이전소득, 기타소득으로 구분하였다. 소득의 경우에도 지출과 마찬가지로 '작년 한 해'를 기준으로 측정한 변수를 활용하였다. 이 때, 소득 및 지출의 특성은 총지출별 각 지출 비중, 총소득별 각 소득 비중으로 살펴보았다. 이는 절대적인 규모보다는 서로 다른 수준에서의 비중이 가구의 지출 및 소득 특성을 더욱 잘 보여주기 때문인 것과 관련이 있다.

한편, 본 고에서는 노인가구의 빈곤여부에 따른 가구 지출이 어떠한지 비교할 것인데, 이를 위해 2008년도 기준 최저생계비를 활용⁷⁾하였으며, 가구의 총소득이 최저생계비 미만인 경우 빈곤, 최저생계비 이상인 경우에는 비빈곤으로 상정하여 분석하였다.⁸⁾

이후 로짓분석에서는 군집분석을 통해 분류된 각 군집의 값을 1로, 나머지는 0으로 부여하여 군집에 따른 분석을 실시하였으며, 투입된 독립변수들은 인구사회학적 요인, 가구특성 요인으로 구분된다. 우선 연구 대상의 인구사회학적 요인으로 성별은 여성을 0, 남성은 1의 값을 부여하였으며, 연령과 교육년수는 연속변수 형태로 분석에 투입했다. 또한 취업유무는 비취업을 0, 취업을 1로 측정하였고, 질환유무 변수는 질환이 없는 경우 0, 질환이 있는 경우 1의 값을 부여하여 분석하였다.

다음으로 가구특성 요인으로 가구유형은 노인단독가구를 0, 노인부부가구를 1의 값을 부여하여 측정하였으며, 주택점유형태는 전월세가구를 0, 자가점유가구를 1의 값을 부여하여 측정하였다. 그리고 거주 지역은 서울을 기준으로 광역시 더미, 중소도시 더미를 생

7) KReIS의 경우 연구대상이 50세 이상 가구원을 포함한 가구이므로, 상대빈곤선을 활용하기 어렵다는 단점이 있다.

8) 분석결과, 65세 이상 노인이 포함된 가구의 빈곤율은 35.8%인 것으로 나타났다.

성하여 분석하였고, 빈곤인 경우 1, 비빈곤은 0을 적용하여 분석에 투입하였다.

다. 분석방법

본 고에서는 노인가구의 가계지출 특성을 분석하기 위해 우선 가구의 가계지출 및 소득구성이 어떠한지 기술통계 등을 통해 살펴보고, 빈곤가구와 비빈곤가구로 구분하여 그 특성을 비교하였다. 이후 가계지출의 유형을 분석하기 위해 군집분석을 실시하였는데, 여기에서는 k-평균 군집분석(k-means clustering)으로 불리는 비계층적 군집분석 방법을 활용하여 분석하였다. 이후 가구원의 특성을 통제된 상태에서 가계 소비패턴에 미치는 영향을 분석하기 위해 로짓 분석을 실시하였으며, 독립 변수로는 선행연구에서 가구의 소비구조와 관련되어 있다고 지적된 가구원의 성별, 연령, 교육년수, 경제활동상태, 건강상태, 가구유형, 주택점유형태, 거주 지역, 총소득을 투입하였다.

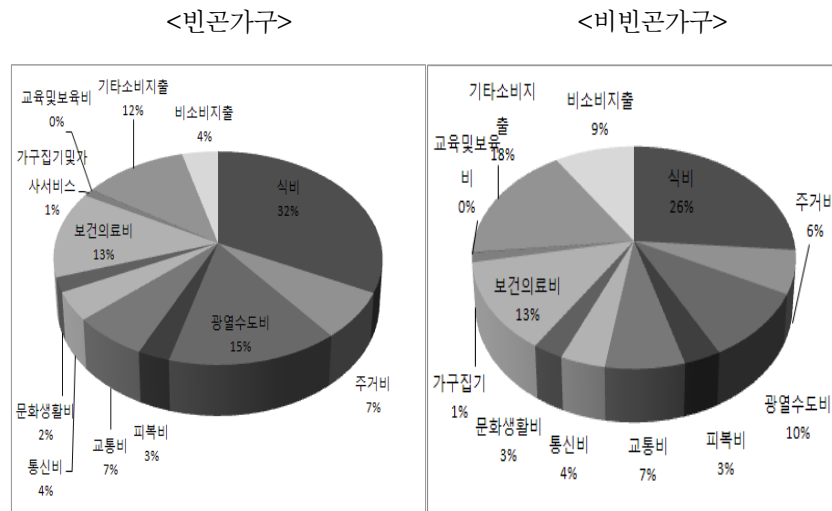
3. 분석결과

가. 노인가구의 지출 및 소득 구성

우선 노인빈곤가구 및 비빈곤가구의 지출 및 소득 구성이 어떠한지 살펴본 결과가 다음 [그림 2]에 나타나있는데, 빈곤가구와 비빈곤가구의 지출 구성을 살펴보면, 빈곤가구는 식비 32%, 주거비 %, 광열수도비 15%, 피복비 3%, 교통비 7%, 통신비 4%, 문화생활비

2%, 보건의료비 13%, 가구집기 및 가사서비스 1%, 기타소비지출 12%, 비소비지출 4%로 나타났으며, 비빈곤가구는 식비 26%, 주거비 6%, 광열수도비 10%, 피복비 3%, 교통비 7%, 통신비 4%, 문화생활비 3%, 보건의료비 13%, 가구집기 및 가사서비스 1%, 기타소비지출 18%, 비소비지출 9%인 것으로 나타났다.

[그림 2] 노인빈곤가구 및 비빈곤가구 지출구성

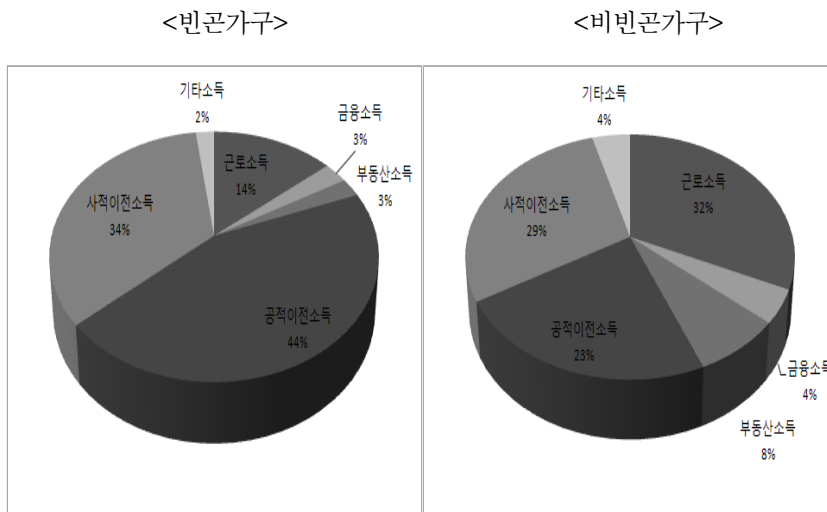


이를 정리하면, 빈곤가구는 비빈곤가구에 비해 식비나 주거비, 광열수도비와 같은 영역이 전체 가계지출에서 더 많은 비중을 차지했고, 비빈곤가구의 경우 기타소비지출이나 비소비지출과 같은 영역이 빈곤가구보다 더 많은 비중을 차지했다. 그리고 보건의료비와 같은 영역은 빈곤가구, 비빈곤가구에 상관없이 가구 지출에서 많은 비중을 차지하고 있어 노인의 높은 의료욕구를 반영하고 있음을 확인할 수 있다.

III. 노인가구의 빈곤여부에 따른 가구 지출 유형화 및 영향 요인 분석 39

다음으로 노인 빈곤가구 및 비빈곤가구의 소득 구성이 어떠한지 살펴보았다. 그 결과, 빈곤가구의 소득은 근로소득 14%, 금융소득 3%, 부동산소득 3%, 공적이전소득 44%, 사적이전소득 34%, 기타소득 2%로 구성되어 있으며, 비빈곤가구의 경우에는 근로소득 32%, 금융소득 4%, 부동산소득 5%, 공적이전소득 23%, 사적이전소득 29%, 기타소득 4%로 구성되어 있었다.

[그림 3] 노인빈곤가구 및 비빈곤가구 소득구성



곧 빈곤가구는 공적이전소득과 사적이전소득과 같은 영역의 소득 비중이 비빈곤가구보다 많았으며, 비빈곤가구의 경우 근로소득과 금융소득, 부동산소득의 비중이 빈곤가구보다 많은 것을 확인할 수 있다. 빈곤가구의 경우 국민기초생활보장제도나 기초노령연금과 같은 공적이전소득의 비중 및 사적이전소득의 비중이 높은 것과 관련이 있는 것으로 보인다.

나. 노인가구의 지출 유형 분석

다음으로 노인가구의 가구 지출 유형이 어떠한지 분석하기 위해 노인단독가구 및 부부가구를 대상으로 군집분석을 실시하였다. 그 결과는 <표 III-1>과 같다. 분석결과, 지출 유형에 따라 군집은 4가지로 구분할 수 있었다. 우선 군집 I의 경우 노인단독가구 및 부부가구에서 가장 많은 비중인 55.32%를 차지했으며 전체 가구 지출에서 식비의 비중이 36.20%로 가장 많았고, 광열수도비가 14.55%, 기타소비지출의 비중이 11.78%로 높게 나타났다.

<표 III-1> 군집유형별 소비지출 구성비

	I	II	III	IV
식비	36.20	25.95	17.38	19.73
주거비	2.90	30.36	2.38	2.77
광열수도비	14.55	9.47	8.36	9.58
피복비	3.46	2.26	2.82	2.46
교통비	7.54	5.29	6.98	5.74
통신비	4.58	3.69	3.45	3.30
문화생활비	2.89	1.61	2.38	1.97
보건의료비	9.69	7.08	6.12	39.00
가구집기및가사서비스	1.28	1.98	1.24	1.13
교육및보육비	0.09	0.11	0.02	0
기타소비지출	11.78	8.36	32.52	10.44
비소비지출	4.99	3.78	16.29	3.82
사례수(=1381)(%)	764(55.32)	158(11.44)	274(19.84)	185(13.39)

군집 II의 경우에는 전체 노인가구에서 그 비중이 11.44% 정도로 나타났으며, 지출에서 주거비의 비중이 30.36%로 가장 높았고

식비의 비중도 25.95%로 높게 나타났다. 군집 III의 경우 전체 노인 가구 가운데 19.84%를 차지했고 지출에서 기타소비지출의 비중이 32.52%로 가장 높게 나타났으며 식비가 17.38%, 비소비지출이 16.29%로 높게 나타났다. 마지막으로 군집 IV의 경우 전체 노인 가구 가운데 13.39%를 차지했으며, 보건의료비의 지출이 전체 가구 지출에서 39%를 차지하였다. 또한 식비의 비중도 19.73%로 높은 비중을 차지하고 있었다.

<표 III-2> 군집유형별 가구지출 구성비의 타군집 지출구성비 평균 대비 비

지출	I	II	III	IV
식비	1.72	1.06	0.63	0.74
주거비	0.24	11.32	0.19	0.23
광열수도비	1.59	0.87	0.74	0.88
피복비	1.37	0.77	1.03	0.86
교통비	1.25	0.78	1.12	0.86
통신비	1.31	0.97	0.89	0.84
문화생활비	1.45	0.66	1.1	0.86
보건의료비	0.55	0.38	0.32	5.11
가구집기및가사서비스	0.88	1.63	0.84	0.75
교육및보육비 ⁹⁾	2.25	3.66	0.33	0
기타소비지출	0.68	0.45	3.19	0.59
비소비지출	0.62	0.45	3.88	0.45
유형 분류	기본생활 지출 중심형	주거지출 중심형	관계지출 중심형	의료지출 중심형

다음으로 각 군집의 가구 지출 구성 특징을 좀 더 자세히 살펴보

9) 교육 및 보육비의 경우 노인가구 전체에서 가구 전체 지출 가운데 0~1% 정도로 매우 미미하였다. 따라서 여기에서는 이를 제외하고 서술하도록 한다.

기 위해 타 군집의 지출 구성비의 평균과 대비할 때 지출 구성 비율이 어떠한지 살펴보았다.

분석결과, 군집 I의 경우 다른 군집에 비해 식비와 광열수도비의 비중이 높게 나타났으며, 피복비, 교통통신비, 문화생활비의 비중도 높은 것으로 나타났다. 따라서 군집 I은 '기본생활지출 중심형'으로 명명하였다. 군집 II의 경우에는 다른 유형과 비교할 때 주거비의 비중이 11.32로 매우 높아, '주거지출 중심형'으로 명명하였으며, 군집 III의 경우에는 기타소비지출과 비소비지출의 비중이 높게 나타나 '관계지출 중심형'으로 명명하였다.¹⁰⁾ 마지막으로 군집 IV는 다른 군집과 비교할 때 보건의료비의 비중이 5.11로 높게 나타나 '의료지출 중심형'으로 명명하였다.

다음으로 이렇게 도출된 가구 지출 유형별 가구원 특성 및 가구 특성을 살펴보도록 하자. 우선 각 유형에 속한 가구의 가구원의 특성을 살펴보면 모든 가구에서 여성의 비중이 남성보다 높은 것을 확인할 수 있는데, 이는 대상이 노인인 것에서 비롯된 것으로 보인다. 그런데 군집 III의 경우 다른 군집에 비해 남성의 비중이 높고, 군집 II의 경우 여성의 비중이 높은 것으로 나타난다. 또한 가구원의 평균 연령은 의료 욕구가 높게 나타난 군집 IV에서 가장 높은 것으로 나타났으며, 교육년수의 평균은 대부분 비슷한 것으로 나타나지만 군집 III의 평균 교육년수가 가장 높은 것으로 나타났다. 취업유무를 살펴본 결과, 취업자의 비중은 군집 III에서 가장 높았고 군집 II에서 가장 낮게 나타났으며, 건강상태를 질환 유무로 살펴

10) 기타소비지출은 '경조사 부조금/기부금, 귀금속 구입비, 가구원 잡비(담배 및 미용비), 목욕료, 화장품 구입비, 교제비, 모임 회비 등 통상적인 생활유지비, 타가구원에게 주는 일시적인 용돈, 결혼 비용, 장례비용, 공공기관 관련 수수료, 사회/종교단체 지원금, 보장성 보험료 불입액 등을 포함하는 영역으로 사회적 관계와 직,간접적으로 관련이 있기 때문에 노인가구의 소비 특성을 분석한 이소정(2009), 백학영(2010) 등의 연구에서 '관계지출형'으로 명명한 바 있다.

보았는데, 군집 III에서 질환을 갖고 있는 가구원이 가장 적었으며 의료 욕구가 높은 군집 IV에서 질환이 있다고 응답한 비중이 높게 나타났다. 가구원의 인구사회학적 요인을 통해 살펴볼 때 남성의 비중이 높고 연령도 낮으며 교육년수도 높은 군집 III에서 질환을 갖고 있는 가구원도 낮고 취업한 사람이 많다는 것을 확인할 수 있다. 곧 가구원의 인적자본은 군집 III의 경우가 다른 군집과 비교할 때 가장 좋다고 할 수 있겠다.

<표 III-3> 가구 지출 유형별 가구 및 가구원 특성의 기술통계치
(단위: %, 세, 년)

구분		I	II	III	IV	전체
성별	남	40.50	38.25	44.01	42.20	41.23
	여	59.49	61.74	55.98	57.79	58.76
연령		72.41	72.86	70.02	73.15	72.15
교육년수		5.90	6.12	6.71	5.12	5.98
취업유무	비취업	69.31	74.29	41.78	68.68	64.08
	취업	30.68	25.70	58.21	31.31	35.91
질환유무	없음	29.03	24.59	35.55	15.86	28.07
	있음	70.96	75.40	64.44	84.13	71.92
가구유형	부부가구	71.59	56.35	79.72	72.37	71.81
	단독가구	28.40	43.64	20.27	27.62	28.18
주택점유형태	전월세	21.97	55.37	9.74	17.94	22.30
	자가	78.02	44.62	90.25	82.05	77.69
거주 지역	서울	16.05	20.80	6.52	10.79	13.84
	광역시	22.14	29.29	19.66	13.41	21.14
	도	61.80	49.90	73.80	75.78	65.00
빈곤율		53.91	55.16	20.05	42.03	46.06

주 1) 모든 수치는 가중치 적용

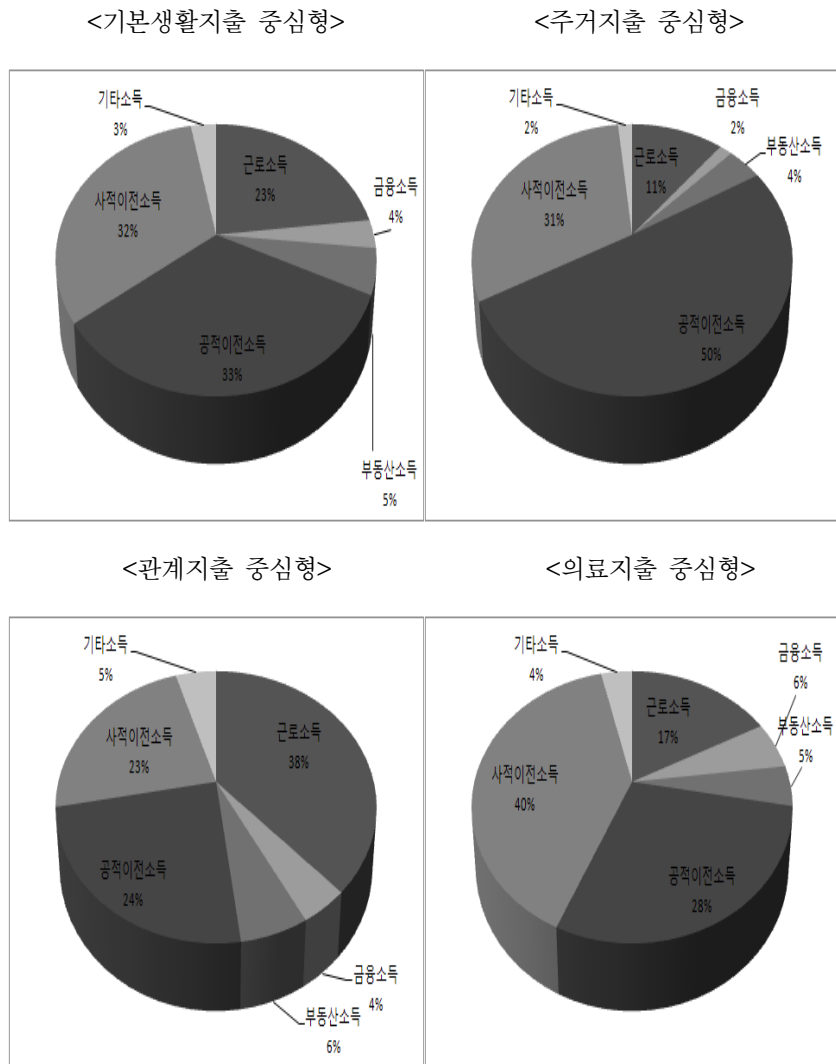
다음으로 가구 특성을 통해 군집의 특성을 좀 더 자세히 살펴볼 도록 하자. 가구유형을 살펴본 결과 단독가구의 비중은 주거욕구가

높은 군집 II에서 가장 높게 나타났으며, 군집 III에서 가장 낮게 나타났다. 또한 주택 점유형태를 살펴본 결과 노인가구의 특성상 대부분 자가의 비중이 높았으나, 주거욕구가 높게 나타났던 군집 II에서 전월세의 비중이 다른 군집과 비교할 때 매우 높은 것으로 나타나, 높은 주거비 지출이 여기에서 비롯되었음을 파악할 수 있다. 거주 지역을 통해 살펴보아도 이를 확인할 수 있는데, 대부분 중소도시에서 거주하는 데 비해 군집 II의 경우에는 서울과 광역시 거주 비중이 높은 편이다. 곧, 군집 II는 대도시 거주 비중이 높고, 전월세이기 때문에 가구 지출의 많은 부분을 주거비로 사용하고 있음을 확인할 수 있다. 마지막으로 빈곤율을 살펴본 결과, 주거지출 중심형의 군집 II의 빈곤율이 55.16%로 가장 높게 나타나고 있음을 확인할 수 있다. 다음으로 기본생활 욕구 중심형인 군집 I이 53.91%로 나타났으며, 의료 욕구 중심형인 군집 IV가 42.03%, 관계 중심형인 군집 III이 20.05%로 빈곤율이 가장 낮았다.

이후 각 군집별 경제적 상황을 좀 더 자세히 살펴보기 위하여 소득 구성이 어떻게 이루어져 있는지 살펴보았다. 분석결과, 빈곤율도 가장 높고 가구원의 인구사회학적 특성 및 가구 특성에서도 가장 열악한 것으로 나타나는 주거지출 중심형인 군집 II에서 전체 소득 가운데 공적이전소득이 50%를 차지하는 것으로 나타났다. 또한 군집 II의 경우 사적이전소득의 비중도 31%로 높아, 이전소득이 전체 소득의 81%를 차지하는 것으로 나타났다. 반면 근로소득과 금융소득, 부동산소득과 같은 시장소득은 17% 정도인 것으로 나타났다. 또한 빈곤율이 그 다음으로 높았던 군집 I의 경우 공적이전소득이 전체 소득의 33%를 차지하였고, 사적이전소득은 32%를 차지하여 이전소득이 전체 소득의 65%를 차지하였다. 또한 근로소득이 27%, 금융소득 4%, 부동산소득이 5%를 차지하여, 군집 II보

다는 시장소득의 비중이 높은 것으로 나타났다.

[그림 4] 군집별 소득 구성



한편, 빈곤율이 세 번째로 높았던 군집 IV의 경우에는 공적이전소득이 전체 소득의 28%, 사적이전소득이 40%로 나타나 이전소득은 전체 소득의 68%를 차지하는 것으로 나타났으며, 근로소득 17%, 금융소득 6%, 부동산소득 5%로 시장소득은 전체 소득의 28%를 차지하였다. 마지막으로 가구의 경제적 상황이 가장 좋은 것으로 나타난 군집 III의 경우 공적이전소득이 24%, 사적이전소득이 23%로 이전소득이 군집 가운데 가장 낮은 47%를 차지하였으며, 근로소득의 비중이 38%로 군집 가운데 가장 높게 나타났다.

각 군집별 소득 구성을 통해 볼 때에도 군집 II의 경제적 상황이 가장 열악하며, 대부분을 공적이전소득에 의존하여 살아가고 있는 것을 알 수 있다. 군집 I과 군집 IV의 경우도 대부분을 이전소득에 의존하는 것을 알 수 있는데, 군집 IV의 경우 사적이전소득의 의존도가 가장 높은 것으로 나타나 차상위층 이상에 많이 분포하고 있는 것으로 파악되었다. 관계육구 중심의 군집 III의 경우 노동시장에서 아직까지 일하고 있는 사람들이 많이 포함되어 있어, 시장소득의 비중이 높고 이전소득의 비중이 낮다는 것을 확인할 수 있다.

다. 노인가구의 지출 유형에 영향을 미치는 요인 분석

노인가구의 지출 유형에 영향을 미치는 요인을 분석하기 위해, 각 유형에 해당하는 사례를 1로, 그렇지 않은 사례를 0으로 코딩하여 4가지 모형에 따른 로짓 분석을 실시하였다.

분석결과, 기본생활지출 중심형 결정 모형에서는 취업유무, 주택점유형태, 광역시 더미, 중소도시 더미, 빈곤여부 변수가 통계적으로 유의미했다. 곧 가구원이 취업보다 비취업일수록, 주택 점유형

태가 전월세보다 자가일수록, 광역시에 거주하는 것보다는 서울에 거주할수록, 중소도시에 거주하는 것보다는 서울지역에 거주할수록, 빈곤가구일수록 기본생활지출 중심형에 속할 확률이 높아지는 것을 확인할 수 있다. 이를 통해 살펴볼 때 기본생활지출 중심형에 속한 가구의 경우 서울에 거주하면서 빈곤할 가능성이 크지만 자가를 소유하고 있기 때문에 주거비보다는 식비나 기타 기본생활유지에 들어가는 비용에 지출을 많이 하는 가구일 확률이 크다는 것을 확인할 수 있다.

주거지출 중심형 결정 모형에서는 주택 점유형태 변수와 중소도시 더미만이 통계적으로 유의미한 것으로 나타났는데, 주택 점유형태가 자가보다는 전월세일수록, 중소도시보다는 서울에 거주할수록 주거지출 중심형에 속할 확률이 높아지는 것을 확인할 수 있다. 주거지출 중심형 가구의 경우 앞에서 살펴본 결과 빈곤한 가구의 비중이 가장 높았는데, 다른 요인들을 통제한 결과 빈곤여부는 통계적으로 유의미하지 않았으며, 주택 점유형태와 거주 지역만이 유의미했다. 곧 서울에 거주하는 전월세 가구인 경우 주거비 지출이 많으며, 주거지출 중심형 모형에 속할 가능성이 크다는 것을 제시한다고 하겠다.

다음으로 관계지출 중심형 결정 모형에서는 연령, 교육년수, 취업유무, 질환유무, 주택점유 형태, 광역시 더미, 중·소도시 더미, 빈곤여부 변수가 모두 통계적으로 유의미하게 나타났다. 이를 살펴보면, 연령이 낮을수록, 교육년수가 높을수록, 비취업보다는 취업일수록, 질환이 있는 것보다는 없을수록, 서울보다는 광역시와 중·소도시에 거주할수록, 빈곤보다는 비빈곤가구일수록 관계지출 중심형에 속할 확률이 높아지는 것을 확인할 수 있다.

<표 III-4> 가구 지출 유형에 영향을 미치는 요인 분석

변수	기본생활지출		주거지출		관계지출		의료지출	
	중심형		중심형		중심형		중심형	
	B (S.E)	Exp (B)	B (S.E)	Exp (B)	B (S.E)	Exp (B)	B (S.E)	Exp (B)
성별(=남성)	-.0608 (.1234)	.9409	.0777 (.2017)	1.0808	-.1091 (.1596)	.8966	.2314 (.1735)	1.2604
연령	.0047 (.0091)	1.0048	.0055 (.0150)	1.0018	-.0264* (.0126)	.9738	.0153 (.0117)	1.0154
교육년수	.0004 (.0132)	1.0004	.0265 (.0209)	1.0269	.0398* (.0176)	1.0406	-.0622*** (.0178)	.9396
취업유무 (=취업)	-.3548** (.1172)	.7013	-.0114 (.2142)	.9886	.8624*** (.1409)	2.3688	-.4999** (.1696)	.6065
질환 유무 (=질환 있음)	-.1331 (.1156)	.8753	.0630 (.1989)	1.0651	-.3273* (.1411)	.7208	.7701*** (.1872)	2.1601
가구유형 (=노인 단독)	-.1089 (.1319)	.8967	.4074 (.2179)	1.5030	.0762 (.1756)	1.0792	-.2048 (.1769)	.8147
주택점유형태 (=자가)	.3294* (.1334)	1.3901	-1.5432*** (.1917)	.2136	.6405*** (.1948)	1.8975	.3692 (.1896)	1.4466
거주 지역 (기준: 서울)								
광역시 더미	-.3730* (.1807)	.6886	-.0551 (.2520)	.9463	.9802*** (.2637)	2.6650	-.2396 (.2732)	.7869
도 더미	-.4779** (.1615)	.6200	-.5382* (.2315)	.5837	.9978*** (.2284)	2.7125	.3724 (.2303)	1.4513
빈곤여부 (=빈곤)	.7878*** (.1111)	2.1987	.0012 (.1933)	1.0012	-1.1763*** (.1591)	.3083	-.3112* (.1574)	.7325
상수	-.0451 (.6891)	-	-1.6262 (1.1043)	-	-.8982 (.9197)	-	-3.4242*** (.8918)	-
사례수	2137 ¹¹⁾							
LR chi2(10)	88.71***		136.07***		185.97***		60.35***	

* = p < .05. ** = p < .01. *** = p < .001.

11) 각 가구에 속한 가구원을 단위로 분석한 결과이다.

이를 통해 볼 때 이 모형에 속할 가능성이 높은 가구는 비교적 다른 노인가구들에 비해 인적자본이 높고 건강상태가 비교적 양호하여 노동시장에서 일하고 있으며, 서울보다는 광역시나 중소도시에 거주하면서 자가를 소유하고 있는 비교적 여유 있는 생활을 유지하는 가구인 것으로 파악된다. 따라서 다른 군집에 속한 가구들과는 달리 기본적인 생활유지에 필요한 지출보다는 기타 관계 유지에 요구되는 기타소비지출의 비중이 높은 것으로 파악된다.

마지막으로 의료지출 중심형 결정 모형을 살펴보면, 교육연수, 취업유무, 질환유무, 빈곤여부 변수가 통계적으로 유의미하게 나타났는데, 교육연수가 낮을수록, 취업보다는 비취업일수록, 질환이 있을수록, 비빈곤가구보다는 빈곤가구일수록 의료지출 중심형에 속할 확률이 높아지는 것으로 나타났다. 이를 통해 살펴볼 때 의료지출 중심형 모형에 속할 가능성이 큰 가구는 건강상태가 좋지 않아 비취업일 가능성이 높고 빈곤하지 않아 의료 급여 등의 보호를 받지 못해 의료비 지출이 많은 가구일 가능성이 클 것이라 파악된다.

4. 논의

본 고에서는 노인만으로 이루어진 가구의 경제적 상태 및 욕구를 살펴보기 위하여 노인빈곤가구 및 비빈곤가구의 소득구성을 살펴보고, 노인가구의 가구지출 유형 및 이에 영향을 미치는 요인을 살펴보았다.

분석결과, 노인빈곤가구의 경우 공적이전소득 및 사적이전소득의 비중이 비빈곤가구보다 높았으며, 근로소득, 부동산소득, 금융소득의 비중은 노인비빈곤가구보다 낮게 나타난다는 것을 확인할 수

있었다. 한편 노인가구의 가구지출 유형은 기본생활지출 중심형, 주거지출 중심형, 관계지출 중심형, 의료지출 중심형 4가지로 구분할 수 있었다. 기본생활지출 중심형은 식비를 중심으로 주거비, 광열수도비 등과 같이 기본생활유지에 필요한 지출들의 비중이 높게 나타나는 유형이었으며, 주거지출 중심형은 주거비의 비중이 높게 나타났다. 또한 관계지출 중심형은 기타소비의 지출 비중이 높은 유형이었으며 의료지출 중심형은 의료비 지출이 높게 나타나는 유형이었다.

이 가운데 기본생활지출 중심형과 주거지출 중심형은 가구지출 유형 가운데 빈곤율도 높고 전체 소득에서 공적이전소득의 비중이 높게 나타나 정책적 관심이 필요한 대상임을 확인할 수 있었다. 특히 주거지출 중심형은 서울, 광역시와 같은 대도시 거주 비중이 높고 대부분 전월세에서 거주하는 것으로 나타났으며, 부부가구의 비중보다는 단독가구의 비중이 높아 국민기초생활보장제도에서의 주거급여 현실화 및 대도시에 자가를 소유하고 있지 않은 독거노인에 대한 주거관련정책 및 서비스의 필요성을 제기한다고 할 수 있겠다.

한편, 의료지출 중심형의 경우 빈곤율은 기본생활지출 중심형과 주거지출 중심형에 비해 낮게 나타났지만, 소득 구성을 통해 살펴볼 때 사적이전소득의 비중이 상당히 높게 나타나는 것을 확인할 수 있다. 따라서 이들은 주로 빈곤선 위에 있거나, 혹은 소득이 빈곤선 이하이더라도 부양의무자 기준에 의해 공공부조에서 배제된 집단일 가능성이 높다. 이들의 의료욕구가 높다는 점을 고려할 때, 차상위계층에 대한 의료급여의 확대 및 노인을 대상으로 한 각종 의료서비스의 확대가 요구된다고 할 수 있겠다.

IV. 우리나라 중·고령자의 은퇴경로 유형과 은퇴 후 소득 비교

1. 문제제기

고령사회에 대비하여 최근 우리나라에서는 중·고령자들의 점진적 퇴직(gradual retirement)에 대한 논의가 활발하게 진행되고 있다. 서구에서 논의되는 점진적 퇴직제도란, 정년을 전후로 일정 연령 시점 이후에는 근로시간을 부분적으로 줄이는 대신, 줄어드는 임금에 대해서 고용보험이나 공적연금에서 부분연금의 형태로 소득을 지원하는 방식을 말한다. 이는 경직적인 정년퇴직제도를 유지하고 있는 우리나라에 많은 시사점을 주고 있는 것이 사실이다.

그러나 여기서 고려해야 할 점은 서구 유럽 국가들에서 고령자들의 자발적 조기은퇴가 경제사회적으로 큰 부담을 주고 있는 것이 현실이라면 우리나라의 경우에는 그와 반대로 중·고령자들이 본인의 의사와 반하는 강제 조기퇴직이 문제라는 것이다. 또한 우리나라는 서구와는 달리 국민연금제도가 미성숙 단계에 있으며 아직까지 가입에서 배제된 집단이 많다는 문제가 있다. 그리고 국민연금에 가입이 되어 있다고 하더라도 국민연금 수급연령이 현행 60세에서 65세로 상향 조정되어 있어 현실적으로 정년이 65세까지 연장되지 않거나 이 기간까지 근로소득을 대체할 소득원이 없다면 빈곤에 처할 위기가 높은 것이 현실이다.

이와 같은 이유로 현세대 중·고령자들은 생애 주된 일자리에서

바로 은퇴하기 보다는 일정기간동안 다른 가교 일자리(파트타임이나 영세 자영업 등)를 거쳐서 점진적으로 은퇴하는 경우를 주변에서 많이 접할 수 있다. 이러한 은퇴과정은 서구의 점진적 퇴직과 모습은 비슷할 수 있으나 노후에 대한 개인적 준비와 제도적 보호장치가 완전하게 구축되어 있지 않은 현실에서 고령임에도 불구하고 노후소득원 마련을 위해 할 수 없이 선택하는 경우가 대다수이다. 이러한 상황에서 생애 근로를 마감하게 되면 은퇴자는 그나마 유지되었던 근로소득을 상실하게 되어 생활을 유지하는데 필요한 소득을 근로소득이 아닌 다른 소득원으로 조달해야 하나, 앞서 언급했듯이 노후소득보장을 위한 제도적 시스템이나 개인적인 준비는 아직까지 미흡한 실정이다.

이에 대한 관심으로 최근 중·고령자의 은퇴에 대한 다양한 연구가 진행되고 있다. 그러나 여기서 문제는 각 연구마다 은퇴자의 개념이 다르고 일부에서는 은퇴를 연령이나 공적연금 수급여부 등을 기준으로 정의하고 있다는 점이다. 이와 같은 정의는 은퇴자를 특정 연령 이상의 노인인구 전체를 지칭하거나 연금을 받는 협소한 범위내의 고령자로 제한하는 한계가 있을 뿐만 아니라 노령으로 인한 은퇴만을 고려한 개념이기 때문에, 외환위기 이후 증가되고 있는 중장년층의 은퇴를 포함할 수 없으므로 현 시점에서의 은퇴와 관련된 이슈를 논하는데 있어 적절하지 못한 것으로 지적되고 있다. 더욱이 현재까지 꾸준히 이어오고 있는 은퇴에 관한 연구들이 은퇴를 하나의 사건(event)만으로 바라보고 있으나, 은퇴는 생애사에 있어 하나의 과정(process)으로 바라보는 것이 옳을 것이다.

그리고 현재까지 은퇴를 다양한 측면에서 바라보고 있으나 생애 근로경험과 은퇴 이후의 경제적 복지수준을 연계하여 파악하는 연구가 드문 것이 현실이다. 이에 본 연구에서는 국민연금공단에서

수행 중에 있는 『KReIS』의 직업력 자료와 3차년도(2009년) 개인 자료에서 측정되고 있는 소득정보를 활용하여 우리나라 중·고령자의 다양한 은퇴경로를 유형화하고 이에 따른 은퇴 후 소득 수준의 차이를 살펴보고자 한다. 이는 은퇴를 하기 이전에 생애 근로경험과 노후준비가 은퇴 후 소득에 어떠한 영향을 미치는지를 보여줌으로써 은퇴 이후 경제적 복지와 관련된 정책을 수립하는데 있어 보다 구체적인 정보를 제공하고자 한다.

2. 분석방법 및 자료 소개

가. 의사결정나무모형

의사결정나무모형은 자료 내에 존재하는 관계, 패턴, 규칙 등을 탐색하고 찾아내어 모형화 하는 데이터마이닝 기법 중의 하나이다. 나무구조에 의하여 모형이 표현되기 때문에 해석이 용이하고 나무구조로부터 어떤 입력변수(독립변수)가 목표변수(종속변수)를 설명하기 위하여 중요한지 쉽게 파악할 수 있다. 즉, 유용한 입력변수를 찾아내고 입력변수간의 다양한 교호효과 즉, 두 개 이상의 변수가 결합하여 목표변수에 어떻게 영향을 주는지를 찾아내는 알고리즘이다¹²⁾.

12) 의사결정나무분석은 개별 독립변수가 종속변수에 미치는 효과를 측정하는 것이 아니라, 여러 독립변수들의 상호 결합이 종속변수에 미치는 효과를 파악한다는 점에서 회귀분석과는 관점의 차이가 있다. 또한 회귀분석과는 달리 선형성(linearity)이나 정규성(normality) 또는 등분산성(equal variance) 등의 가정은 필요로 하지 않는 비모수적 방법이다.

<표 IV-1> 의사결정나무의 구성요소

마 디	내 용
뿌리마디	나무구조가 시작되는 마디로서 전체자료로 이루어져 있음.
자식마디	하나의 마디로부터 분리되어 나간 2개 이상의 마디
부모마디	자식마디의 상위마디
끝마디	각 나무줄기의 끝에 위치한 마디로 잎(leaf)이라고도 함.
중간마디	나무구조의 중간에 있는 끝마디가 아닌 마디들
가지	하나의 마디로부터 끝마디까지 연결된 일련의 마디들

의사결정나무는 하나의 나무구조를 이루고 있으며, 마디(node)라고 불리는 구성요소들로 이루어져 있다. 마디는 그 기능에 따라 <표 IV-1>과 같이 분류된다. 의사결정나무모형은 분석의 목적과 자료구조에 따라서 적절한 분리기준¹³⁾과 분리규칙¹⁴⁾ 그리고 정지규칙¹⁵⁾을 지정하고 의사결정나무를 얻으면 분류오류를 크게 할 위험이 높거나 부적절한 추론규칙을 가지고 있는 가지를 제거하고 이익도표(gain chart)나 위험도표(risk chart) 또는 검증용 자료에 의한 교차타당성(cross validation) 등을 이용하여 타당성 평가를 한 후 분석의 결과를 해석하게 된다.

13) 분리기준은 하나의 부모마디로부터 자식마디들이 형성될 때 예측변수의 선택과 범주의 병합이 이루어 질 기준의 의미한다. 즉, 어떤 입력변수를 이용하여 어떻게 분리하는 것이 목표변수의 분포를 가장 잘 구별해 주는지를 파악하여 자식마디가 형성되는데, 목표변수가 이산형이나 연속형이나에 따라 사용되는 분리기준이 달라진다.

14) 분리규칙은 부모마디로부터 형성되는 부모마디로부터 형성되는 자식마디의 개수가 2개로 제한되는 이진분리(binary split)와 3개 이상으로 분리되는 다지분리(multiway split)로 나뉘어 진다.

15) 정지규칙은 뿌리마디를 최대 몇 단계의 수준으로 분리할 것인가와 부모마디와 자식마디의 최소 크기는 어떻게 할 것인가를 지정하여 나무 키우기를 제한한다.

이러한 의사결정나무모형에는 CHAID(Chi-squared Automatic Interaction Detection), CART(Classification And Regression Trees), C4.5와 같은 다양한 알고리즘들이 제안되어 있으나 본 연구에서는 가장 보편적으로 사용되고 있는 CART 알고리즘을 이용하여 분석을 수행한다¹⁶⁾. CART는 지니지수(Gini Index; 이산형 목표변수인 경우) 또는 분산의 감소량(연속형 목표변수인 경우)을 이용하여 마디 내 동질성을 최대화시키는 알고리즘이다. 본 연구에서 살펴보고자 하는 목표변수는 은퇴 후 소득 수준으로 연속형 변수이기에 때문에 분리기준은 분산의 감소량을 사용하여 나무를 형성한다. 분산의 감소량은 부모마디의 분산에서 자식마디들의 마디 내 분산의 가중합을 뺀 값이다. 분산을 마디 내 다양도의 의미로 해석하면 감소량이 최대화되는 예측변수를 선택하고 그 때의 최적분리를 하게 된다. 즉, CART 알고리즘은 불순도(impurity)라는 마디 내 이질성(heterogeneity)을 최소화하여 자식마디 내 목표변수값들이 최대한 동질적으로 되도록 부모마디를 분리한다(최종후·서두성, 1999).

나. 분석자료

본 연구에서 사용되는 KReIS는 국민연금연구원에서 중·고령자들의 노후준비 및 노후생활을 지속적으로 파악하기 위하여 2005년 당시 전국의 만 50세 이상 가구원이 있는 약 5,000여 가구와 그 가구에 속하는 만 50세 이상 개인 8,600여명을 추출하여 2005년부터 격년으로 조사하는 종단면 자료이다. KReIS는 2005년 1차 본조사를 시작으로 2007년 2차 본조사를 실시하였으며, 현재 3차 본조사를 완료한 자료를 구축하고 있는 상황이다. 본조사에서는 가구의 경제

16) 각 알고리즘의 비교는 허명희(2007)를 참조

상황, 중·고령자의 고용현황 및 퇴직, 건강, 가족관계, 노후소득보장 현황 등 노후생활 전반에 걸친 주제들을 포괄적으로 조사한다. 본 패널조사는 소득정보를 제공하는 일반 패널조사와 다르게 가구 단위에서의 소득 원천에 대한 정보뿐만 아니라 개인단위의 다양한 소득원천에 대한 정보도 제공하고 있다.

또한 본조사가 시행되지 않는 연도에는 노후소득보장과 관련된 당해 연도의 이슈를 조사하는 부가조사가 진행되고 있는데, 2006년 1차 부가조사에서는 50대 가구주 가구를 대상으로 가계지출 및 소득현황에 대한 심층조사를 실시하였으며, 2008년 2차 부가조사에서는 2차년도 본조사(2007년)까지 정보가 파악된 가구의 개인조사 대상자들을 대상으로 이들의 만 15세 이후 일자리 변동이력과 공적연금 이력에 대한 심층조사를 실시하였다. 이들 부가조사 자료는 한 개인의 생애에 걸친 직업이력 및 연금의 가입 및 수급 이력을 파악할 수 있을 뿐 아니라 이미 조사가 완료된 본조사의 정보와 연계하면 앞으로 우리나라 중·고령자들이 노동시장의 진입과 퇴장 및 이동이 어떠한지, 공적연금을 통한 노후준비는 어느 정도이며 그 수준은 적절한지에 대한 심층 분석이 가능해질 것으로 기대된다¹⁷⁾.

본 연구에서는 2008년에 실시된 직업력 자료를 기본 자료로 하고 2009년 3차년도 본조사에서 입수된 직업 정보를 업데이트 하였으며 은퇴 후 소득은 3차년도 개인자료에서 제공하고 있는 2008년 현재 시점의 개인소득 정보를 사용한다.

17) 노동이력과 관련해서는 국내 다른 패널조사(한국노동패널, 고령화연구패널 등)에서도 조사를 수행하고 있으나 노동이력과 연금가입이력을 함께 연계하여 조사하는 KReIS가 유일하다.

3. 은퇴경로의 유형화 및 유형별 분포

가. 은퇴경로의 유형화

은퇴에 대한 정의는 연구주제와 연구자에 따라 다양하게 사용된다. 기존 연구에서 사용되는 은퇴에 대한 정의는 크게 분석대상자의 주관적 평가에 의한 정의와 연구자에 의한 객관적인 기준에 따른 정의로 나누어진다. 먼저, 사회조사에서 응답대상자에게 은퇴에 대한 판단기준¹⁸⁾을 설명하고 그 기준에 따라 본인이 은퇴자인지를 판단하는 주관적인 방법이 있다. 이 방법은 노동시장에서의 과거의 경험과 현재의 주된 경제활동참여 여부를 적절하게 고려되기 때문에 은퇴와 관련된 실태분석에서 주로 활용되고 있다(안중범 외, 2004; 김지경, 2005; 이승렬, 2008; 이승렬 외, 2008; 백은영, 2008; 홍백의 외 2009 등). 그러나 이 방법은 본인이 은퇴를 주관적으로 평가하여 결정하기 때문에 종단연구에서 이 방법이 적용되기 위해서는 은퇴에 대한 문항이 지속적으로 조사되어야 하고 과거에 응답했던 결과와의 오류체크가 필수적이다.

이와는 달리 분석대상자의 객관적인 정보를 통하여 은퇴를 정의하는 방법이 있다. 우선 특정 연령대 이상을 은퇴자로 판별하는 방법이 있다(원종욱, 2000; 안중범 외, 2002, 전승훈 외 2009 등). 이 경우 아주 간단하게 은퇴자를 판별할 수 있다는 장점이 있으나 이 방법은 분석결과를 해석함에 있어 주의를 요하게 된다. 왜냐하면,

18) KReIS 본조사에서는 은퇴와 관련한 문항을 측정하였다. 여기서 은퇴는 '본격적인 소득활동을 그만두고 조사당시 시점에 일을 하지 않고 있거나 소일거리 정도의 일을 하고 있는 경우' 또한 '앞으로도 특별한 변화가 없는 한 소일거리 정도의 일 이외의 일을 할 의사가 없는 상태'로 정의하고 있다.

연령¹⁹⁾을 기준으로 은퇴자를 판별한다는 것은 '고령자 = 은퇴자'라는 등식이 성립하지만 우리나라 노인대상의 연구와 은퇴자대상의 연구는 소득구조면에서 큰 차이를 보이고 있다. 60세 또는 65세 이상 노인을 대상으로 그들의 소득구성을 살펴본 기존의 연구결과에서 상당히 큰 비중을 차지하고 있는 것은 노인 가구주의 근로 및 사업소득인 것으로 나타나고 있다. 따라서 근로 및 사업소득이 전혀 없는 완전 은퇴가구의 경우에는 이들 연령집단 중에서 일부뿐이기 때문에 단순히 가구주의 특정 연령을 근거로 은퇴자를 정의할 경우에는 은퇴자를 과대추정하는 문제가 있으며, 또한 특정 연령대 이상만을 대상으로 선별하기 때문에 외환위기 이후 우리사회에 만연해 있는 조기퇴직이나 명예퇴직의 문제 등은 고려되지 않는 문제가 발생한다.

한편 노동시간이나 임금수준의 일정 수준 이하로 급격히 감소한 경험여부에 따라 은퇴자를 구분하기도 한다. 그 기준은 연구자에 따라 다소 차이가 있는데, 연간 1,000시간미만(Parnes & Nestel, 1974) 또는 연간 1,200시간미만(Gustman & Steinmeier, 2000) 노동하는 사람을 은퇴자로 보는 경우도 있고, 주당 30시간미만(Burtless & Moffitt, 1985) 또는 주당 25시간미만(Gustman & Steinmeier, 2000) 일하는 자를 은퇴자로 정의하는 경우도 있다. 또한 Gustman & Steinmeier(2000)는 임금수준을 기준으로 최고급여 수준의 60% 미만으로 급격히 감소하는 경우를 은퇴자로 정의하기도 한다. 그러나 국내에서 활용가능한 직업력 자료²⁰⁾는 개별 근로자들의 과거

19) 은퇴자로 판별하는 연령으로는 연구자에 따라 55세, 60세, 또는 65세의 연령기준이 사용된다. 이를 판별하는 기준은 관련법에 따른 것으로 고령자 고용촉진법에서는 고령자를 55세 이상-64세 미만으로 정의하고 있다. 노인복지법에서는 60세 기준으로 고령자를 정의하고 있으며, 국민연금법에서는 완전노령연금의 수급 시기를 60세 이상으로 정해져 있다. 그리고 고용보험법에서는 고용보험피보험자를 64세 이하로 정하였지만, 60세 이상 재취업자는 제외하고 있다.

직업력에 대한 정보를 회고적 방법에 의존하여 획득되기 때문에 과거에 경험하였던 직업들 중 근로시간이나 임금의 상당 부분이 결측되어 있고 근로시간 변수는 신뢰성의 문제도 제기되고 있다.

다음의 기준으로 10년 또는 20년 이상 일한 일자리에서 45세 이후 해당 일자리를 그만두는 경우를 은퇴로 정의하기도 한다(Ruhm, 1990; Quinn 1999; Gustman & Steinmeier, 2000; 신동균, 2009; 우석진, 2010 등). 이러한 은퇴의 정의는 은퇴를 노동시장의 경험이나 현재의 경제활동과 관련하여 개념화되어야 한다는 측면을 충족시킬 수 있으나 10년이나 20년이라는 제약으로 인하여 취약한 근로경험을 한 계층은 표본에서 제외되는 문제가 있다.

마지막으로 공적연금 수급여부에 따라 은퇴자를 판별하는 방법 또한 은퇴자를 판별하는 간단한 방법이지만 하나 국민연금제도의 도입역사가 짧은 우리나라에서는 공적연금 수급대상자를 은퇴자로 판별하게 되면 대부분 특수직역연금 수급자이나 일부 국민연금에서 특례노령연금수급자나 조기노령연금수급자만을 포함하기 때문에 이들 또한 특정 연령집단이나 퇴직자군의 특성을 담고 있기 때문에 아직까지 우리나라에서는 은퇴자에 대한 일반적인 연구에서 적합하지 않다.

이와 같이 연구자에 따라 은퇴의 정의는 다양하게 존재하며 정의된 은퇴에 따라 분석 결과가 상당히 달라질 수도 있다²¹⁾. 이에 대

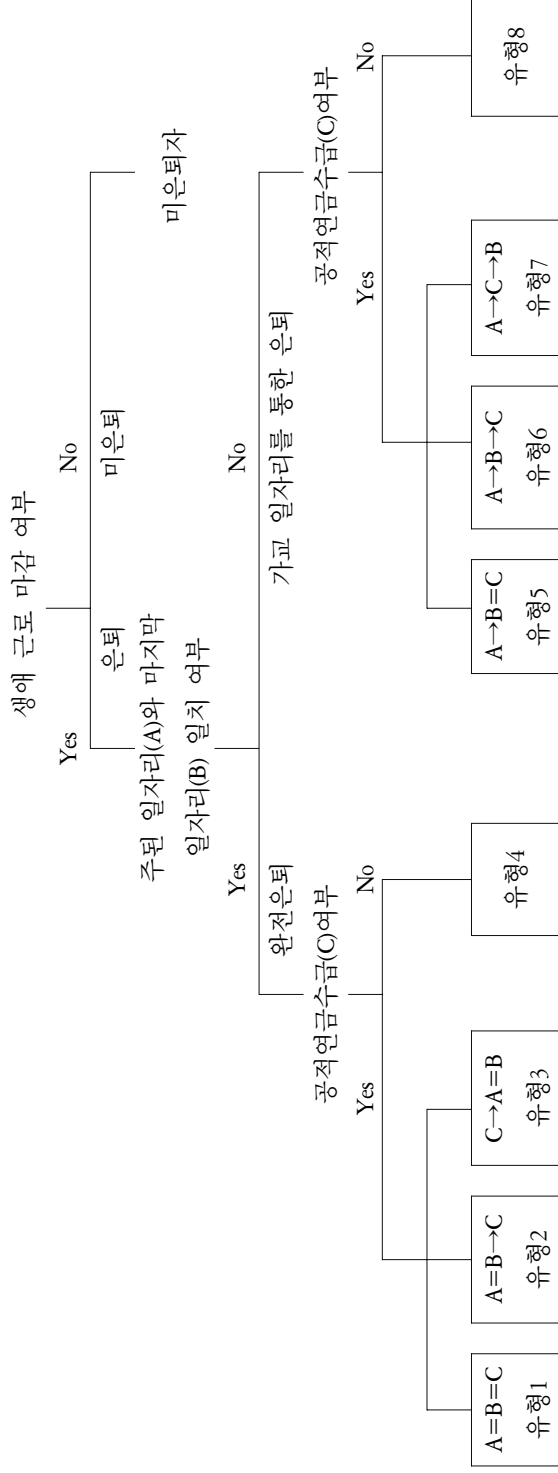
20) 국내에서 직업력 정보가 담긴 자료는 한국노동연구원의 『한국노동패널』과 『고령화연구패널』 그리고 국민연금연구원의 KReIS 등이 있다.

21) Gustman and Steinmeier(2000)는 HRS(Health and Retirement Study) 1차년도 자료와 4차년도 자료를 이용하여 통계를 산출하여 본 결과, 은퇴의 정의에 따라 은퇴자 비율이 상당히 다르게 나타남을 알 수 있었다. 예를 들어, 본인의 신고에 따른 은퇴자 비율은 1차년도에 12.3%였으나 시간급과 주급으로 정의한 경우의 은퇴자 비율은 32.3%에 이르러 무려 20.0% 포인트의 차이를 보인다.

해 Gustman and Steinmeier(2000)는 어느 한 가지 정의에 따라 은퇴를 분류할 것이 아니라 여러 정의를 조합하여 분류하는 것이 바람직한 것으로 제시하고 있다. 이에 본 연구에서는 은퇴란 생애 근로경험에 바탕을 두고 '생애 근로를 그만두는 시점 이후'로 정의하고 생애 주된 일자리에 대한 제한을 두지 않고 생애 근로 가운데 근속기간이 가장 긴 일자리를 생애 주된 일자리로 정의하여 다양한 고용형태를 포괄하고자 한다.

그리고 은퇴를 생애 근로를 마감하는 하나의 사건이 아니라 [그림 5]와 같이 생애 근로경험과 공적연금 수급과의 관계 속에서 논의를 하고자 한다. 이를 위해 먼저, 생애 근로경험이 있는 개인 중에서 생애 근로를 그만두었는지를 판별하여 생애 근로를 그만 둔 개인을 은퇴자로 그리고 현재까지도 소득활동을 하고 있는 개인은 '미은퇴자'로 구분한다. 이후 은퇴자는 생애 마지막 일자리가 생애 주된 일자리와 동일한지에 따라 생애 주된 일자리에서 생애 근로를 마감하는 완전 은퇴자와 생애 주된 일자리 이후 가교 일자리(bridge job)를 통해 생애 근로가 끝나는 점진적 은퇴자로 구분하도록 한다. 이후 공적연금 수급여부에 따라 연금을 수급받고 있는 집단과 수급 받지 못하는 집단 구분하고 연금을 수급 받는 집단에서는 생애 근로경험과 공적연금 수급의 사건 발생 순서에 따라 다시 유형을 구분한다. 이러한 방법에 따라 은퇴경로는 총 8가지의 유형으로 나누어진다.

[그림 5] 은퇴경로의 유형화



- 주 1) A : 주된 일자리 끝난 시기
- 2) B : 마지막 일자리 끝난 시기
- 3) C : 공적연금 수급 시작시기
- 4) 생애 무직자는 분석에서 제외함.

나. 은퇴경로별 분포

앞서 언급한 KReIS 직업력 조사에서 생애 근로경험이 존재하는 개인은 총 7,339명이며 이들 가운데 55.3%에 해당하는 4,060명이 2009년 현재 생애 근로를 마감한 것으로 나타났다. 이들 은퇴자를 대상으로 생애 근로경험과 공적연금 수급의 시간적 발생 순서에 따라 구분한 8가지 은퇴경로별 사건이 발생한 시점과 각 유형별 비중은 <표 IV-2>와 같다. 우선 생애 주된 일자리에서 생애 근로를 마감하는 경우는 전체 은퇴경로에서 65.1%를 차지하고 있다. 이를 구체적으로 살펴보면, 유형1은 생애 주된 일자리가 곧 생애 마지막 일자리인 근로자로 59.8세에 생애 주된 일자리에서 퇴직과 동시에 공적연금 수급을 받고 있는 경우로 이들은 전체 은퇴경로에서 3.6%를 차지하고 있다. 유형2는 생애 주된 일자리에서 생애 근로를 53.9세에 마감한 이후 61.4세에 공적연금을 수급 받고 있는 경우로 이들은 전체 은퇴경로에서 7.7%를 차지하고 있다. 유형3은 59.1세에 공적연금을 먼저 수급을 받고 65.0세에 생애 주된 일자리에서 생애 근로를 마감하는 경우로 이들은 전체 은퇴경로에서 3.5%를 차지하고 있다. 유형4는 55.2세에 생애 주된 일자리에서 생애 근로를 마감하지만 공적연금은 수급을 받지 못하는 경우로 이들은 전체 은퇴경로에서 50.2%를 차지하고 있다.

다음으로 생애 주된 일자리에서 퇴직 이후 가교 일자리를 통해 생애 근로가 끝나게 되는 경우로 이들은 전체 은퇴경로에서 34.9%를 차지하고 있다. 이를 구체적으로 살펴보면, 유형5는 48.4세에 생애 주된 일자리에서 퇴직 이후 가교 일자리를 통해 61.2세에 생애 근로를 그만두는 동시에 공적연금을 수급 받는 경우로 이들은 전체 은퇴경로에서 1.0%를 차지하고 있다. 유형6은 43.9세에 생애 주

된 일자리를 그만두고 55.5세에 가교 일자리를 통하여 생애 근로를 그만 둔 이후 62.3세에 공적연금을 수급 받는 경우로 이들은 전체 은퇴경로에서 3.3%를 차지하고 있다. 유형7은 50.7세에 생애 주된 일자리에서 퇴직 이후에 60.1세에 공적연금을 수급 받고 생애 근로는 64.7세에 마감하는 경우로 이들은 전체 은퇴경로에서 4.1%를 차지하고 있다. 그리고 유형8은 47.2세에 생애 주된 일자리를 그만 둔 이후 가교 일자리를 통하여 60.6세에 생애 근로를 마감하지만 공적연금을 수급 받지 못하는 경우로 이들은 전체 은퇴경로에서 26.5%를 차지하고 있다.

이상을 정리하면, 현세대 노인 은퇴자의 대부분은 공적연금제도 시행의 역사가 짧아 공적연금에 대한 혜택을 받지 못하고 생애 주된 일자리 또는 가교 일자리를 통하여 생애 근로를 마감하고 있는 것을 알 수 있다.

<표 IV-2> 은퇴경로별 평균 퇴직 연령 및 비중

		주된 일자리 끝난 시기(A)	마지막 일자리 끝난 시기(B)	공적연금 수급시작 시기(C)	비중
주된 일자리에서 은퇴 (A=B)	유형1	① 59.8세	① 59.8세	① 59.8세	3.6
	유형2	① 53.9세	① 53.9세	② 61.4세	7.7
	유형3	② 65.0세	② 65.0세	① 59.1세	3.5
	유형4	① 55.2세	① 55.2세	-	50.2
	소계				65.1
가교 일자리에서 은퇴 (A<B)	유형5	① 48.4세	② 61.2세	② 61.2세	1.0
	유형6	① 43.9세	② 55.5세	③ 62.3세	3.3
	유형7	① 50.7세	③ 64.7세	② 60.1세	4.1
	유형8	① 47.2세	② 60.6세	-	26.5
	소계				34.9
합계					100.0

주 1) 사건의 발생 시기는 2009년 현재 시점으로 산정한 연령을 기준으로 함.

다. 은퇴경로별 인구학적 특성 및 생애 주된 일자리 특성

본격적으로 은퇴경로에 따른 은퇴 후 소득 수준을 살펴보기에 앞서 본 절에서는 은퇴경로에 따라 분류된 은퇴자의 인구학적 특성과 생애 주된 일자리 특성을 살펴보고자 한다(<표 IV-3>참조). 우선 성별의 경우, 생애 주된 일자리에서 생애 근로를 마감한 집단 중 생애 주된 일자리에서 생애 근로를 마감하는 동시에 공적연금을 수급 받는 유형1은 남성이 79.1%로 대다수를 차지하고 있으며, 생애 주된 일자리에서 생애 근로를 마감한 이후 시차를 두고 공적연

금 수급을 받는 유형2는 남성이 50.9%로 남녀의 비율이 유사하며, 공적연금 수급을 받은 이후 생애 주된 일자리에서 생애 근로를 마감한 유형3은 남성이 61.4%로 다소 높은 반면, 생애 주된 일자리에서 생애 근로를 마감하지만 공적연금 수급권이 없는 유형4는 여성이 73.5%로 높게 나타났다.

다음으로 가교 일자리를 통하여 은퇴한 집단을 살펴보면 생애 주된 일자리 이후 가교 일자리를 통해 생애 근로를 마감함과 동시에 공적연금을 수급 받는 유형5는 남성이 66.7%, 생애 주된 일자리 이후 가교 일자리를 통해 생애 근로를 마감한 후 시차를 두고 공적연금 수급을 받는 유형6은 남성이 52.6%, 그리고 생애 주된 일자리 이후 시차를 두고 공적연금 수급을 받은 이후에 가교 일자를 통해 생애 근로를 마감한 유형7은 남성이 66.1%로 많은 반면에 생애 주된 일자리 이후 가교 일자리를 통해 생애 근로를 마감하지만 공적연금 수급권이 없는 유형8은 여성이 64.6%로 높게 나타났다. 즉, 공적연금을 수급 받지 못하는 은퇴경로의 경우 여성의 비율이 높은 것을 알 수 있다.

그리고 연령의 경우, 공적연금 수급을 받지 못하는 유형4와 유형8이 각각 71.6세와 70.3세이며 공적연금 수급을 받는 집단 중에서는 유형3이 70.6세로 다소 높은 반면에 나머지 유형은 68세 내외로 다소 낮은 것을 알 수 있다. 한편, 최종학력으로 측정한 최종학력의 경우, 생애 주된 일자리에서 생애 근로를 마감하는 유형1은 대졸이상 43.4%, 고졸 30.9% 순으로 최종학력이 상대적으로 높게 나타났다고 유형2는 초졸 28.7%, 고졸 25.3% 순이고 유형3은 초졸 28.0%, 무학 24.5% 순이며 유형4는 무학 32.7%, 초졸 32.1% 순으로 최종학력이 상대적으로 낮은 것으로 나타나고 있다. 한편 가교 일자리에서 생애 근로 마감하는 유형5는 대졸이상과 초졸이 각각

29.7%로 나타났고 유형6은 초졸 30.4%, 무학 23.0%, 고졸 20.0% 순이고 유형7은 초졸 31.7%, 고졸 24.4% 순이며 유형8은 초졸 33.6%, 무학 32.5% 순으로 최종학력이 상대적으로 낮게 나타났다. 여기서 유형1과 유형5는 다른 유형 보다 최종학력이 상대적으로 높으며 유형4와 유형8은 다른 유형에 비해 최종학력이 상대적으로 낮은 집단으로 구성되어 있음을 알 수 있다.

다음으로 은퇴경로별 주된 일자리에서의 특징을 살펴보자. 먼저, 생애 주된 일자리에서의 직업의 구성을 살펴보면, 생애 주된 일자리에서 생애 근로를 마감하는 유형1은 대부분 전문사무직으로 이들은 74.1%를 차지하고 있고 유형2는 전문사무직 27.2%, 생산직 22.4% 순이고 유형3은 농림어업직 27.8%, 전문사무직 24.3% 순이며 유형4는 농림어업직 28.7%, 판매서비스직, 21.7% 순으로 나타났다. 또한 가교 일자리에서 생애 근로를 마감하는 유형5는 전문사무직 35.9%, 농림어업직 25.6% 단순노무직, 20.5% 순으로 나타났고 유형6은 농림어업직 29.1%, 생산직 24.6%, 판매서비스직 20.9% 순이고 유형7은 생산직 31.3%, 농림어업직 26.3% 순이며 유형8은 농림어업직 36.8%, 판매서비스직 20.8% 순인 것으로 나타났다.

여기서 유형1은 전문사무직이 상대적으로 높게 나타났으며 유형4와 유형8은 상대적으로 농림어업직이나 판매서비스직이 많은 것을 알 수 있다. 그리고 생애 주된 일자리에서의 종사상지위를 보면, 생애 주된 일자리에서 생애 근로를 마감하는 유형1의 경우 상용직 임금근로자가 83.1%로 가장 많았고 유형2는 상용직임금근로자 45.8%, 자영업자 15.7%, 무급가족종사자 14% 순이고 유형3은 상용직임금근로자가 34.8%, 자영업자가 29% 순이며 유형4는 무급가족종사자 27.4%, 자영업자 22.9%, 상용임금근로자 19.7% 순으로 나타났다. 또한 가교 일자리에서 생애 근로를 마감하는 유형5의 경우에

IV. 우리나라 중·고령자의 은퇴경로 유형과 은퇴 후 소득 비교 67

는 상용직임금근로자가 56.4%, 무급가족종사자가 20.5% 순으로 많았고 유형6은 상용직임금근로자 32.6%, 무급가족종사자 23.7%, 자영업자가 21.5% 순이고 유형7은 상용직임금근로자 50.3%, 무급가족종사자 18.2%, 자영업자 16.4% 순이며 유형8은 무급가족종사자 31.9%, 상용직임금근로자 22.0%, 자영업자 19.9% 순으로 나타났다. 여기서 유형4와 유형8의 경우에는 무급가족종사자이거나 자영업자가 많은 것을 알 수 있다. 그리고 생애 주된 일자리에서의 평균 근속기간을 살펴보면, 생애 주된 일자리에서 생애 근로를 마감하는 유형1이 32.1년으로 가장 길며 그 다음이 유형3으로 27.9년, 그 다음은 유형2, 5, 7, 8로 20년 내외인 것으로 나타났으며 유형6은 16.7년으로 가장 짧았다.

마지막으로 생애 주된 일자리에서의 퇴직사유를 살펴보면, 생애 주된 일자리에서 생애 근로를 마감하는 유형1은 대부분 정년퇴직으로 주된 일자리를 물러났으며 유형2는 건강 및 고령으로 근로유지가 어려워져 은퇴한 경우가 29.2%, 정년이 26.4%순이며 유형3은 건강 및 고령이 38.3%, 정년이 21.8%, 파산, 폐업, 명예퇴직, 해고 등 시장 및 기업의 상황에 따라 은퇴한 경우가 21.1%이며, 유형4는 건강 및 고령으로 은퇴한 경우가 43.4%, 결혼, 출산, 육아 등의 가사문제로 은퇴한 경우가 25.7%인 것으로 나타났다. 가교 일자리에서 생애 근로를 마감한 유형5의 경우 52.8%가 생애 주된 일자리에서 소득이나 장래성, 적성, 근로환경 등 근로조건 불만으로 인하여 퇴직한 것으로 나타났고 그 다음이 파산, 폐업, 명예퇴직, 해고 등이 22.2%의 순으로 나타났다. 유형6의 경우에도 생애 주된 일자리에서 근로조건 불만으로 49.6%가 퇴직하였으며 그 다음이 파산, 폐업, 명예퇴직, 해고 등으로 23.3%를 차지하였다. 유형7의 경우에도 생애 주된 일자리에서 37.9%가 근로조건 불만으로 퇴직하였으

며 그 다음이 정년으로 23.0%를 차지하였다. 그리고 유형8의 경우에도 42.4%가 생애 주된 일자리에서 근로조건의 불만으로 퇴직하였으며, 그 다음이 파산, 폐업, 명예퇴직, 해고 등으로 21.6%를 차지하였다. 이는 생애 주된 일자리에서 생애 근로를 마감하는 경우 생애 주된 일자리에서의 퇴직사유는 대부분 정년이거나 건강 및 고령 등의 사유이며 예외적으로 상대적으로 여성 은퇴자가 많은 유형4의 경우는 결혼, 출산, 육아 등 가사문제로 은퇴한 경우가 많은 것으로 나타났다. 그리고 가교 일자리에서 생애 근로를 마감하는 경우에는 생애 주된 일자리의 퇴직사유는 대부분 근로조건의 불만인 경우가 많은 것으로 나타났다.

IV. 우리나라 중·고령자의 은퇴경로 유형과 은퇴 후 소득 비교 69

<표 IV-3> 은퇴경로별 인구학적 특성 및 생애 주된 일자리 특성

구분		주된 일자리에서 은퇴			
		유형1	유형2	유형3	유형4
성별	남성	79.1	50.9	61.4	26.5
	여성	20.9	49.1	38.6	73.5
연령	평균	68.4	67.9	70.6	71.6
최종 학력	무학	5.1	14.2	24.5	32.7
	초등학교	8.8	28.7	28.0	32.1
	중학교	11.8	19.9	18.2	14.8
	고등학교	30.9	25.3	16.8	15.4
	대학교 이상	43.4	11.8	12.6	4.9
직업	전문사무직	74.1	27.2	24.3	19.4
	판매서비스직	4.9	18.9	13.9	21.7
	농림어업직	5.6	15.4	27.8	28.7
	생산직	11.2	22.4	16.7	13.9
	단순노무직	4.2	16.0	17.4	16.3
종사상 지위	상용직근로자	83.1	45.8	34.8	19.7
	임시직근로자	1.5	7.3	6.5	8.7
	일용직근로자	2.9	11.5	13.0	15.9
	고용주	2.9	5.6	4.3	5.3
	자영업자	5.1	15.7	29.0	22.9
	무급가족종사자	4.4	14.0	12.3	27.4
근속 기간	10년미만	2.7	23.1	14.5	29.6
	10-20년미만	6.1	27.8	20.7	20.5
	20-30년미만	15.5	22.8	20.0	16.8
	30년 이상	75.7	26.3	44.8	33.1
	평균	32.1	19.9	27.9	22.6
퇴직 사유	정년	64.4	26.4	21.8	3.9
	파산/폐업/ 명예퇴직/해고	17.0	16.9	21.1	14.2
	근로조건 불만	0.7	6.0	3.0	9.1
	건강/고령	12.6	29.2	38.3	43.4
	가계여유/여가	3.0	3.5	4.5	2.2
	결혼/출산/육아	2.2	17.3	9.8	25.7
기타	-	0.7	1.5	1.4	

<표 IV-3> 계속

구분		가교 일자리에 은퇴			
		유형5	유형6	유형7	유형8
성별	남성	66.7	52.6	66.1	35.4
	여성	33.3	47.4	33.9	64.6
연령	평균	67.2	67.3	68.6	70.3
최종 학력	무학	10.8	23.0	18.3	32.5
	초등학교	29.7	30.4	31.7	33.6
	중학교	16.2	17.0	17.1	15.6
	고등학교	13.5	20.7	24.4	14.5
	대학교 이상	29.7	8.9	8.5	3.7
직업	전문사무직	35.9	15.7	18.1	10.5
	판매서비스직	5.1	20.9	14.4	20.8
	농림어업직	25.6	29.1	26.3	36.8
	생산직	12.8	24.6	31.3	15.6
	단순노무직	20.5	9.7	10.0	16.3
종사 상 지위	상용직근로자	56.4	32.6	50.3	22.0
	임시직근로자	2.6	6.7	3.6	7.7
	일용직근로자	12.8	8.9	7.3	13.1
	고용주	2.6	6.7	4.2	5.4
	자영업자	5.1	21.5	16.4	19.9
	무급가족종사자	20.5	23.7	18.2	31.9
근속 기간	10년미만	10.3	23.0	8.5	23.3
	10-20년미만	33.3	40.7	33.9	31.2
	20-30년미만	38.5	26.7	32.1	23.6
	30년 이상	17.9	9.6	25.5	21.9
	평균	21.3	16.7	22.0	19.7
퇴 직 사 유	정년	5.6	3.0	23.0	2.8
	파산/폐업/ 명예퇴직/해고	22.2	23.3	19.3	21.6
	근로조건 불만	52.8	49.6	37.9	42.4
	건강/고령	8.3	8.3	11.2	14.2
	가계여유/여가	-	-	-	0.6
	결혼/출산/육아	5.6	13.5	6.2	15.3
	기타	5.6	2.3	2.5	3.2

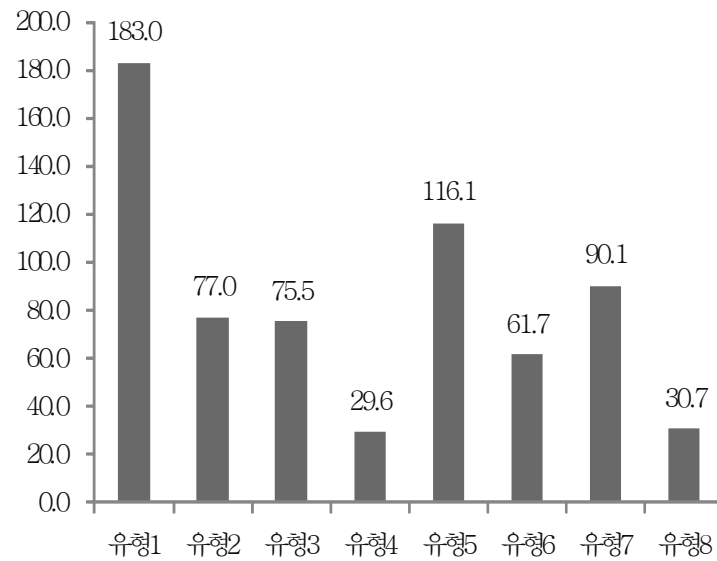
4. 은퇴경로별 은퇴 후 소득 비교

가. 은퇴경로별 은퇴 후 소득 분포

이와 같이 은퇴경로를 다양하게 살펴보는 이유는 은퇴경로에 따라 은퇴 후 소득 수준이 차이를 보이기 때문이다. 직관적으로 개인의 생애 근로이력에 따라 정의된 은퇴경로에 따른 은퇴 후 소득 수준은 [그림 6]과 같다. 다양한 은퇴경로 중 은퇴 후 소득 수준이 가장 높은 집단은 생애 주된 일자리에서 생애 근로를 완전히 그만둠과 동시에 공적연금을 수급한 유형1로 이들의 평균 개인소득은 183만원이다²²⁾. 그 다음으로 개인소득이 높은 집단은 생애 주된 일자리에서 퇴직 이후 가교 일자리에서 생애 근로를 마감함과 동시에 공적연금을 수급한 유형5로 이들의 평균 개인소득은 116만원이며, 그 다음은 주된 일자리에서 퇴직하고 공적연금을 수급 받은 이후에 재취업을 통해 가교 일자리에서 생애근로를 그만 둔 유형7로 이들의 평균 개인소득은 90만원인 것으로 나타났다.

22) 은퇴 후 소득에는 은퇴이후 경과시간의 효과도 있으나 본 절에서는 이를 고려하지 않은 상태에서 비교한 수치임을 주의해야 한다.

[그림 6] 은퇴경로별 은퇴 후 소득액(2008년 기준)
(단위 : 만원/월, 경상소득 기준)



주 1) 분석대상은 은퇴시기가 2008년 이전인 은퇴자로 한정함.

이와 반대로 은퇴경로 가운데 은퇴 후 소득 수준이 가장 낮은 집단은 생애 주된 일자리에서 생애 근로를 마감하거나 생애 주된 일자리에서 퇴직 이후 가교 일자리 통해 생애 근로를 그만두었지만 공적연금을 받지 못하는 유형4와 유형8이며, 이들의 소득수준은 정부에서 공식적으로 발표한 2008년 1인 가구 최저생계비보다도 낮은 30만원 내외의 수준이다²³⁾. 이들의 은퇴 후 소득 수준은 소득이 가장 높은 유형1의 16% 수준이며, 이들이 문제가 되는 것은 이들 집단이 은퇴경로 가운데 76.7%로 대다수를 차지하고 있다는 것이다. 이는 우리나라 노인 빈곤율이 높은 문제와도 연관되어 있다. 이들

²³⁾ 2008년 1인 가구의 최저생계비는 463,047원이다.

보다 조금 나은 집단은 생애 주된 일자리에서 물러나 가교 일자리를 통해 생애 근로를 마감한 이후 몇 년간의 시차를 두고 연금을 수급 받는 유형6으로 이들의 평균 개인소득은 62만원이며 생애 주된 일자리에서 생애 근로를 마감한 이후 시차를 두고 공적연금을 수급 받는 유형2와 공적연금을 수급 받은 이후 생애 주된 일자리에서 생애 근로를 마감하는 유형3의 평균 개인소득은 각각 77만원과 76만원의 수준인 것으로 나타났다. 이러한 은퇴경로에 따른 은퇴 후 소득수준은 소득구성에 있어서도 차이를 보이고 있다. 상대적으로 소득이 높은 집단의 경우에는 공적연금의 구성이 높으며 소득수준이 낮은 유형에서는 자녀 등으로부터의 사적이전소득의 의존율이 상대적으로 높은 것을 알 수 있다.

<표 IV-4> 은퇴경로별 소득원천

구분	주된 일자리에서 은퇴			
	유형1	유형2	유형3	유형4
연금소득	76.3	48.4	45.9	0.1
공적연금	76.3	48.4	45.9	0.0
개인연금	0.0	0.0	0.0	0.1
이전소득	6.2	24.7	23.7	74.9
공적이전	1.6	4.8	5.4	22.7
사적이전	4.6	19.9	18.3	52.2
가구내 ²⁴⁾	0.8	4.3	2.4	7.6
가구간	3.8	14.9	15.9	44.0
기타	0.0	0.6	0.1	0.6
자산소득	17.5	26.9	30.4	25.0
금융자산	9.5	7.4	6.2	5.8
부동산자산	8.1	21.3	23.6	21.8
전체	100.0	100.0	100.0	100.0

주 1) 분석대상은 은퇴시기가 2008년 이전인 은퇴자로 한정함.

24) 동거하는 가족으로부터 받게 되는 이전소득은 은퇴한 노인세대에게 중요한 소득원으로 작동하는 것이 일반적이기 때문에 이들로부터 받는 이전소득을 소득으로 포함하였다.

<표 IV-4> 계속

구분	가교 일자리에 은퇴			
	유형5	유형6	유형7	유형8
연금소득	60.6	35.8	45.6	0.5
공적연금	60.6	34.1	45.6	0.0
개인연금	0.0	1.6	0.0	0.5
이전소득	18.7	34.7	31.5	82.4
공적이전	2.0	12.4	4.0	27.3
사적이전	16.7	22.1	27.5	55.0
가구내 ²⁵⁾	0.3	5.7	3.8	7.4
가구간	16.4	16.3	23.7	47.2
기타	0.0	0.0	0.0	0.3
자산소득	20.7	29.5	22.9	17.1
금융자산	11.9	5.0	3.8	5.0
부동산자산	8.5	24.0	18.2	13.9
전체	100.0	100.0	100.0	100.0

이상을 정리하면, 은퇴경로와 은퇴 후 소득 수준과의 관계에서 생애 주된 일자리를 이후 가교 일자리를 통해 생애 근로를 마감하는 점진적 은퇴자 보다는 생애 주된 일자리에 생애 근로를 마감하는 완전 은퇴자의 소득 수준이 상대적으로 높으며, 이들 중에서는 공적연금 수급을 받고 있는 경우에 은퇴 후 소득수준이 높다는 사실을 알 수 있다. 그리고 은퇴 시점과 공적연금 수급시기에 있어서 은퇴와 동시에 공적연금 수급을 받고 있는 경우가 이들 사이에 시차가 있는 경우보다 은퇴 후 소득 수준이 높다는 사실도 알 수

25) 동거하는 가족으로부터 받게 되는 이전소득은 은퇴한 노인세대에게 중요한 소득원으로 작동하는 것이 일반적이기 때문에 이들로 부터 받는 이전소득을 소득으로 포함하였다.

있다. 이는 생애 주된 일자리에서 고용상태의 안정도가 높아 생애 근로를 마감할 수 있는가와 공적연금에 가입되어 있는가에 따라 은퇴 후 소득 수준에 차이를 보이고 있음을 시사한다.

나. 은퇴경로별 은퇴 후 소득의 결정요인 분석

앞에서 우리는 은퇴경로에 따라 은퇴 후 소득 수준에 차이가 있음을 보았지만, 이는 다시 계량모형을 통하여 엄밀하게 분석될 필요가 있다. 본 연구에서는 이를 위해 앞서 설명한 데이터 마이닝(data mining)의 한 방법인 의사결정나무모형을 통하여 분석을 시도한다. 이를 위하여 본 연구에서는 여러 데이터 마이닝 알고리즘 가운데 가장 간단하고 보편적으로 사용되고 있는 CART 알고리즘을 사용하며, 정지규칙으로는 최대나무깊이(maximum tree depth=5)와 최소 케이스수(minimum number of cases: parent node=50, child node=25)로 설정하였다. 의사결정나무모형에 사용된 변수와 분류 단위는 <표 IV-5>와 같다.

<표 IV-5> 분석에 사용된 변수

변수명		측정단위
개인특성	소득	2008년 경상소득(단위 : 만원/월)
	성별	1. 남성, 2. 여성
	나이	1. 50대, 2. 60대, 3. 70대 4. 80대 이상
	최종학력	1. 무학, 2. 초등학교, 3. 중학교, 4. 고등학교, 5. 대학교 이상
은퇴경로		유형1~8
주된일자리특성	직업	1. 전문사무직, 2. 판매서비스직, 3. 농림어업직, 4. 생산직, 5. 단순노무직
	종사상지위	1. 상용직 임금근로자, 2. 임시직 임금근로자, 3. 일용직 임금근로자, 4. 고용주, 5. 자영업자, 6. 무급가족종사자
	근속기간	1. 10년미만, 2. 10-20년미만, 3. 20-30년미만, 4. 30년 이상
	퇴직시기	1. 60세 이전, 2. 60세 이후
	퇴직사유	1. 정년, 2. 파산/폐업/명예퇴직/해고 등, 3. 근로조건불만, 4. 건강/고령, 5. 가계여유/여가, 6. 결혼/출산/유아 등, 7. 기타
은퇴경과기간		은퇴 이후 2008년까지의 소요기간(년)

본 연구에서 종속변수는 은퇴 이후의 개인소득을 사용하였으며, 독립변수로는 성별, 나이, 최종학력 등의 인구학적 정보와 생애 주된 일자리에서의 직업, 종사상지위, 근속기간, 퇴직 시기, 퇴직사유 등과 같은 일자리 정보 그리고 다양한 은퇴경로와 은퇴경과기간 등의 변수를 사용하였다. 이들 변수를 사용하여 은퇴 후 소득의 결정요인을 의사결정나무모형을 통하여 살펴본 결과는 [그림 7]과 <표 IV-6>과 같다. 분석결과에 의하면, 우선 은퇴 후 소득에 가장 영향력이 있는 변수는 최종학력으로 이로 인해 1단계 가지분리가 나타나고 2단계에서는 은퇴경로에 따라, 3단계에서는 성별과 생애

주된 일자리에서의 퇴직사유 등으로, 4단계에서는 생애 주된 일자리에서의 종사상지위와 직업, 그리고 연령과 최종학력 등으로, 마지막 5단계에서는 생애 주된 일자리에서의 직업, 근속기간, 은퇴경로 그리고 연령 등으로 가지가 분리되었다. 즉, 최종학력에 따라 생애 근로경험은 차이가 생기게 되어 은퇴 이후의 소득이 결정되고 이는 다시 은퇴자의 인구학적 특성에 따라 차이를 보이게 된다. 이에 따라 은퇴 후 소득이 가장 높은 집단은 마디6이고 이어서 마디17, 마디26의 순이고 작아지며 소득 수준이 가장 낮은 집단은 마디23인 것으로 나타났다. 이를 구체적으로 살펴보면, 생애 근로경험이 있는 은퇴자소득에서 2008년 이전에 은퇴한 개인은 총 2,980명이며 이들 은퇴자의 평균 소득은 45만원인 것으로 나타났다(뿌리마디=마디0).

먼저, 은퇴자 평균 소득에 비하여 가장 높은 소득 수준을 보이는 집단은 마디6으로 최종학력은 대졸 이상이면서 생애 근로를 그만두는 동시에 공적연금을 수급 받는 유형1, 생애 근로를 그만둔 이후 시차를 두고 공적연금을 수급받은 유형2, 생애 주된 일자리를 그만두고 가교 일자리를 통해 생애 근로를 마감함과 동시에 공적연금을 수급 받는 유형5, 그리고 생애 주된 일자리를 그만둔 이후 시차를 두고 공적연금을 수급 받은 이후 가교 일자리를 통하여 생애 근로를 마감하는 유형7로 이들의 은퇴 후 평균 소득은 228만원인 것으로 나타났다. 그 다음으로 은퇴 후 소득 수준이 높은 집단은 마디17로 이들은 최종학력이 고졸 이하이면서 생애 주된 일자리에서의 직업은 전문사무직이면서 퇴직사유가 정년이거나 가계여유 또는 여가를 위해 퇴직하였으며 공적연금을 수급 받는 집단(유형1, 2, 3, 5, 6, 7)으로 이들의 평균 소득은 176만원인 것으로 나타났다. 이들 다음으로 소득 수준이 높은 집단은 마디26으로 이들은

최종학력이 고졸이면서 생애 근로를 그만둠과 동시에 공적연금을 수급 받는 집단(유형1, 유형5)이거나 공적연금을 수급 받고 난 이후에 생애 근로를 마감하는 집단(유형3, 유형7)이면서 생애 주된 일자리에서의 퇴직사유가 시장이나 기업의 상황에 따라 퇴직하거나, 근로조건에 대한 불만이거나 결혼, 출산, 육아 등의 가사문제로 퇴직하는 경우로 이들의 평균 소득은 129만원인 것으로 나타났다. 이상과 같이 은퇴자의 평균 소득 보다 높은 집단은 공적연금 수급을 받고 있다는 공통점이 존재한다.

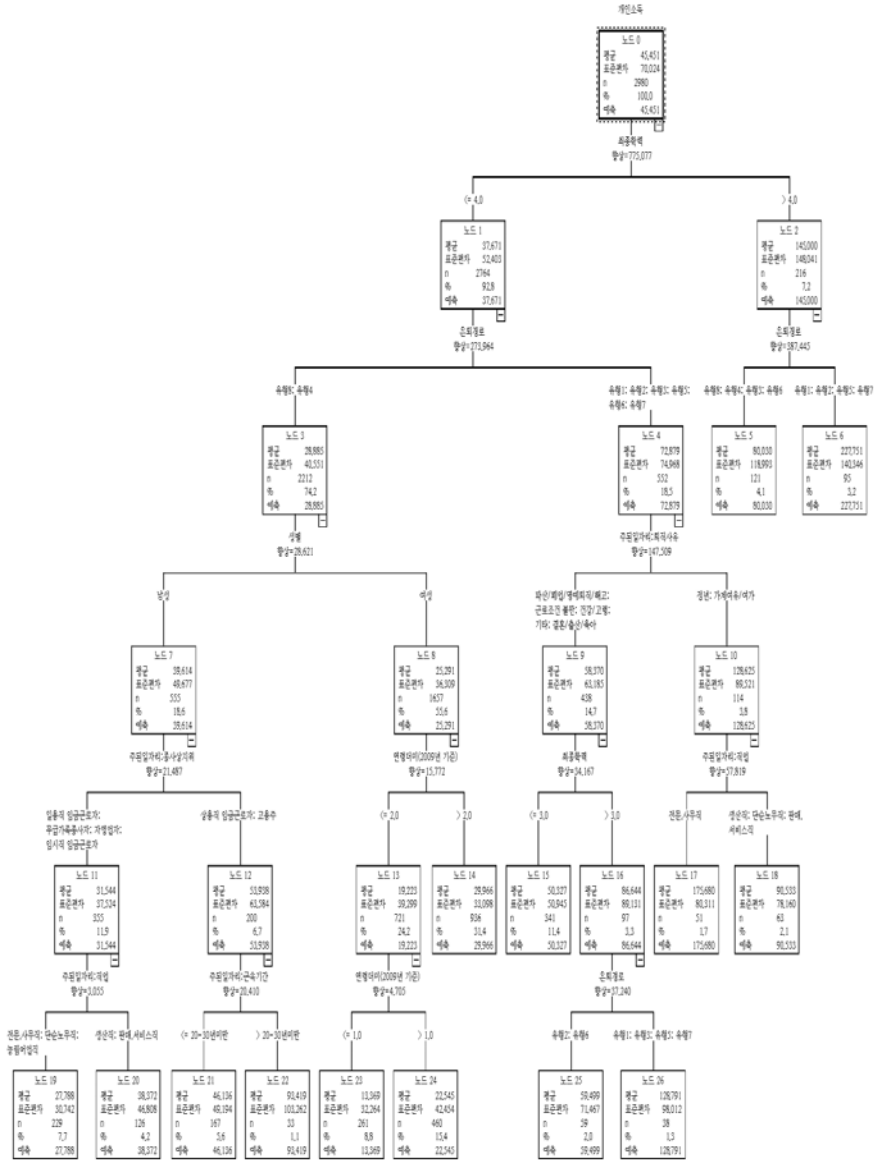
이와 반대로 은퇴자의 평균 소득 수준 보다 낮은 집단은 최종학력이 고졸이하이면서 공적연금 수급을 받지 못하는 집단인 것으로 나타났다. 이들 중 소득 수준이 가장 낮은 집단은 마디23으로 최종학력이 고졸이하이면서 은퇴경로는 공적연금 수급을 받지 않고 있는 유형4(생애 주된 일자리에서 생애 근로를 마감하는 은퇴자)와 유형8(생애 주된 일자리 이후 가교 일자리를 통해 생애 근로를 마감한 은퇴자)인 50대 여성 집단으로 이들의 평균 소득은 13만원인 것으로 나타났다. 이보다 조금 나은 집단은 마디24로 이들은 마디23과 같이 최종학력이 고졸 이하이면서 은퇴경로는 유형4와 유형8인 60대 여성 집단으로 이들의 평균 소득은 22만원인 것으로 나타났다. 이들이 50대 여성 은퇴자 집단 보다 소득수준이 다소 높은 것은 이들 집단에서 자녀 등으로부터의 사적이전소득이 발생할 가능성이 높기 때문이다. 이와 같이 최종학력이 낮고 은퇴 이후 공적연금 수급을 받지 못하는 집단의 경우는 은퇴자의 평균소득 보다 대부분 낮은 수준인 것으로 나타나고 있다. 그러나 예외적으로 생애 주된 일자리에서의 종사상 지위가 상용직 임금근로자 또는 고용주이면서 근속기간이 높은 남성 은퇴자 집단의 경우는 최종학력이 낮고 은퇴 이후 공적연금을 수급 받지 못하더라도 은퇴자의 평

균소득 보다는 높게 나타나고 있다. 이는 상대적으로 안정된 고용형태의 경험이 은퇴 후 소득 결정에 있어서 중요함을 보여주고 있다.

이상의 결과를 정리하면, 은퇴 후 소득 수준은 최종학력에 따라 크게 영향을 보이고 있으며 그 다음으로 생애 근로경험의 차이나 공적연금 수급을 받고 있는지에 따라 차이를 보이고 있는 것을 알 수 있다.

IV. 우리나라 중·고령자의 은퇴경로 유형과 은퇴 후 소득 비교 81

[그림 7] 은퇴경로별 은퇴 후 소득의 결정요인



<표 IV-6> 은퇴경로별 은퇴 후 소득의 이익도표

Node	Gain(N)	Gain(%)	Mean
6	95	3.2	227.8
17	51	1.7	175.7
26	38	1.3	128.8
22	33	1.1	93.4
18	63	2.1	90.5
5	121	4.1	80.0
25	59	2.0	59.5
15	341	11.4	50.3
21	167	5.6	46.1
20	126	4.2	38.4
14	936	31.4	30.0
19	229	7.7	27.8
24	460	15.4	22.5
23	261	8.8	13.4

주 1) Node=마디(집단) 번호, Gain(N)=해당 노드에서 목표 범주의 수, Gain(%)=100×(해당 노드에서 목표 범주의 수/전체 목표 범주의 수)

5. 결론

본 연구에서는 국민연금연구원에서 수행하고 있는 KReIS의 직업력 자료와 3차년도(2009년) 개인자료의 소득정보를 활용하여 우리나라 중·고령자 세대에서 발견되는 다양한 은퇴경로를 유형화하고 이에 따른 은퇴 후 소득 수준의 차이를 살펴보았다.

분석결과에 의하면, 첫째, 우리나라 2009년 현재 만 50세 이상의 중·고령자 중 경제활동상태에 있는 인구(미은퇴자)는 44.7%이며, 생애 주된 일자리에서 은퇴를 경험한 인구(완전 은퇴자)는 36.0%,

생애 주된 일자리에서 퇴직이후 가교 일자리를 통하여 생애 근로를 마감하는 인구(점진적 은퇴자)는 19.3%인 것으로 나타났다. 이는 전통적인 은퇴경로라고 생각되어 온 주된 일자리에서 생애 근로를 마감하는 완전은퇴의 경향과 더불어 가교 일자리를 통한 점진적 은퇴도 하나의 주된 은퇴 유형으로 자리 잡고 있음을 알 수 있다.

둘째, 은퇴 후 소득 수준은 은퇴경로에 따라 큰 차이를 보이고 있다. 은퇴 후 소득 수준은 가교 일자리를 통해 은퇴한 경우보다 생애 주된 일자리에서 은퇴한 경우가 높으며 그리고 이들 중에서는 공적연금 수급을 받고 있는 집단의 소득 수준이 높게 나타나고 있다. 이는 은퇴를 하기 이전에 일자리에서의 고용상태와 연금 가입여부가 은퇴 후 소득을 결정함에 있어 중요한 요인으로 작용하고 있음을 보여준다.

셋째, 의사결정나무모형으로 은퇴 후 소득에 대한 결정요인을 살펴본 결과에 의하면, 은퇴 후 소득수준에 가장 영향력이 있는 변수는 최종학력이고 그 다음이 생애 근로 경험에 따른 은퇴경로의 유형이며 이들 가운데에서는 공적연금 수급여부가 개인소득 수준의 차이를 가져온 중요한 요인인 것으로 나타났다.

따라서 현재 우리나라의 중·고령자 세대의 은퇴와 관련한 연구에서 생애 주된 일자리에서 생애 근로를 마감하는 완전 은퇴와 더불어 주된 은퇴경로로 자리 잡고 있는 가교 일자리를 통한 점진적 은퇴를 감안하여 이에 대한 연구는 앞으로 활발히 진행되어야 하며, 이는 노동시장의 변화로만 인식될 것이 아니라 노후소득보장의 측면에서도 바라볼 필요가 있음을 보여준다.

V. KReIS 표본 특성 및 가중치 부여 방식에 따른 자료 활용 방법

1. 서론

어느 집단의 특성을 알고자 할 때, 대부분의 경우 시간, 비용 등의 제약으로 인해 표본 조사를 실시하게 된다. 이 때 전체 모집단 중 일부 표본을 추출해 조사함으로써 발생하는 표본오차(sampling error)와 표본 조사 과정에서 응답자나 조사원, 조사 방법 등 다양한 요인에 의해 비표본오차(non-sampling error)가 발생한다.

본 고에서 분석하고자 하는 KReIS 자료 또한 표본조사 자료로, 일정한 시간적 간격을 두고 같은 표본에 대하여 반복 조사를 한다는 점에서 횡단면 자료와 시계열 자료의 장점을 동시에 갖고 있다. 하지만 계속 조사를 실행하기 위해 표본을 관리, 유지해야 하는 어려움과 시간의 흐름에 따라 변화해가는 모집단의 변동을 반영하기 위한 조정 작업 등 여러 가지 사항을 고려해야 한다. 모든 표본조사에서 그렇듯이 패널조사 역시 수집된 자료가 모집단의 특성을 잘 대표할 수 있어야 하며, 패널조사의 경우 추출된 표본에 대한 조사가 1회에 그치지 않고 계속 조사를 해야 하기 때문에 확보된 대표성을 유지하는 것이 중요하다. 그러나 최초에 추출한 표본이 대표성을 지니고 있어도 조사가 반복됨에 따라 응답거부, 사망, 이사 등의 이유로 표본 이탈이 발생하여 표본의 대표성을 유지하는

문제가 발생한다.

특히 중·고령자를 대상으로 한 KReIS 자료의 경우, 다음과 같은 이유로 자료를 이용하여 모수 추정 시 주의를 기울여야 한다. 첫째, 표본설계 당시 중·고령층이라는 특정 연령층을 대표하는 표본을 추출하였기 때문에 시간이 경과함에 따라 연령이 상승하여 표본 추출틀(sampling frame)을 벗어난다. 둘째, 다른 패널조사와 마찬가지로, 조사 횟수가 반복됨에 따라 응답에 대한 피로감이 누적되어 응답 기피를 하거나 접촉 불능 등으로 발생하는 무응답으로 인한 표본마모(sample attrition) 이외에도, 조사대상에 고령층이 차지하는 비중이 높기 때문에 매 조사마다 사망으로 인해 표본수가 감소하는 폭이 크다. 셋째, 중·고령자를 대표하는 표본 설계를 하였음에도 불구하고 표본 유지조사 당시 고령자가 모집단의 비율보다 많이 진입하였고, 50대의 표본 이탈률이 60, 70대의 표본 이탈률보다 크기 때문에 표본에 고령층이 차지하는 비율이 높아졌다.

따라서 본 고에서는 이러한 KReIS 1차~3차년도 자료 표본의 특성을 분석하고, 이러한 표본이 대표성을 갖기 위해 적용하여 분석할 수 있는 가중치 활용 방법에 대하여 고찰해 보기로 한다. 이를 위해 먼저 KReIS의 1차년도(2005년) 당시 표본 추출 방법을 검토하고, 각 조사년도의 가중치 보정 작업을 통한 횡단면 가중치와 종단면 가중치 부여 방법을 알아본다. 마지막으로 1차~3차년도 표본의 특성을 분석하여, 이러한 표본 특성에 따라 가중치를 적용하여 분석하는 방법을 논의해 보고자 한다.

2. KReIS 표본 추출 방법²⁶⁾

KReIS는 2005년 만 50세 이상의 가구원이 있는 전국 5,110 가구를 추출하고, 그 가구에 속하는 50세 이상의 가구원 및 그의 배우자를 개인조사대상으로 한다. 따라서 표본설계 및 추출은 우리나라 50세 이상 인구 모집단을 대표하고 연령별, 성별, 도·농 지역별 독립분석이 가능하도록 설계하였다. 본 장에서는 KReIS의 표본설계 및 추출 방법과 표본 조사구에 대한 실사 결과를 검토하여 원표본 구축 방법을 살펴보고자 한다.

가. 표본설계 및 표본추출과정 개요

KReIS의 목표 모집단은 전국의 모든 50세 이상 가구원이 있는 일반가구이며, 추출틀(sampling frame)은 2000년 인구주택총조사 10% 표본조사구이다.

목표 표본 수는 5,000가구의 약 10,000명의 개인이고, 한 조사구당 평균 30가구가 존재하고, 50세 이상 가구원이 있는 가구가 조사구당 평균적으로 10가구 정도 조사한다고 가정하여²⁷⁾, 확률비례추출법(Probability Proportional to Size Sampling)을 사용하여 총 500개 표본조사구를 추출하였다. 한편, 2000년 인구주택총조사 실시 이후, 아파트 1,039,946호가 신축되었는데, 2000년 센서스 10%

26) 본 장의 내용 및 관련 자료는 학술연구용역 “연금·퇴직패널조사 표본설계”, 대한통계협회, 2005.7 보고서를 요약한 내용이며 자세한 사항은 위 보고서에 기술되어 있다.

27) 실제로는 조사구마다 50세 이상 가구원이 있는 가구 분포는 매우 상이할 것이나 여기서는 평균적으로 10가구를 상정한 것이다.

표본조사구를 기준으로 설계한 표본과 1차년도(2005년) 표본 구축 시점과의 괴리를 보완하고자, 34개의 신축아파트 리스트를 별도의 조사구로 포함하였다.

나. 조사가구 유치 현황

앞서 살펴보았듯이 KReIS의 표본 추출을 위하여 2000년 인구주택총조사의 10%표본에서 확률비례추출을 활용하여 일반표본 500조사구와 신축아파트표본 34지역을 추출하였다. 그러나 추출된 표본 조사구의 가구명부와 요도는 5년 전 센서스 조사 당시의 기준이고, 조사구내 50세 이상 가구원의 분포를 알 수 없었기 때문에 먼저 표본 조사구에 대한 사전조사를 실시하였다.

<표 V-1> KReIS 조사구 현황

지역	표본설계시 배분된 조사구		조사완료된 조사구		차이 (조사완료-표본설계)	
	조사구수	비율(%)	조사구수	비율(%)	조사구수	비율(%)
합계	534	100.0	571	100.0	37	0.0
동부	420	78.7	457	80.0	37	1.4
읍면부	114	21.3	114	20.0	0	-1.4

이때, 표본설계시 배분된 표본 조사구수는 동부 420개(78.7%), 읍면부 114개(21.3%)로 총 534개(신축아파트 포함)였으나, 실사를 진행하면서 표본조사구의 투입과 대체과정을 거쳐 최종적으로 조사 완료된 표본 조사구수는 동부 457개(80.0%), 읍면부 114개(20.0%)로 총 571개²⁸⁾로 조사구를 결정하였다.

28) 현장 조사 전에 사용된 표본조사구는 2000년 기준으로 작성된 것이므로 약 5년간의 조사구 변화가 발생하였으므로 대한 보완을 수행하여 최종 571개 조사

<표 V-2> 유효 조사가구 및 참여가구 현황

지역	만 50세 이상 가구원이 있는 유효 가구		조사참여 가구수			
	총 가구수	평균가구수	총 가구수	유효가구 대비비율(%)	평균	비율(%)
서울	2435	17.64	1,046	43.0	7.58	20.4
부산	1006	21.40	388	38.6	8.26	7.6
대구	604	22.37	231	38.2	8.56	4.5
인천	485	17.32	272	56.1	9.71	5.3
광주	464	19.33	157	33.8	6.54	3.1
대전	316	18.59	174	55.1	10.24	3.4
울산	242	18.62	150	62.0	11.54	2.9
경기	1890	17.03	939	49.7	8.46	18.3
강원	470	29.38	170	36.2	10.63	3.3
충북	492	28.94	189	38.4	11.12	3.7
충남	597	28.43	231	38.7	11.00	4.5
전북	596	25.91	212	35.6	9.22	4.1
전남	649	32.45	214	33.0	10.70	4.2
경북	696	24.00	323	46.4	11.14	6.3
경남	942	26.91	372	39.5	10.63	7.2
제주	137	27.40	65	47.4	13.00	1.3
합계	12,021	21.05	5,133	42.7	8.99	100.0

<표 V-2>에서는 추출된 표본 조사가구에서 만 50세 이상 가구원이 있는 유효 가구 중 KReIS의 패널로 참여한 가구수가 총 5,133개 가구임을 알 수 있다. 유효 가구 중 42.7%(응답률)가 참여한 것으로 나타났으며, 1개 표본조사구당 평균 8.99개 가구가 패널로 참여하였으며, 조사에 참여한 5,133개 가구의 개인조사 대상자 수는 총 8,703명이었고 이중 8,602명이 응답하였다.²⁹⁾

구로 수정하게 되었다.

29) 1차년도(2005년) 조사가 끝난 후 조사표 검수를 마치고, 응답이 부실한 가구

3. 가중치 부여 방법 검토³⁰⁾

가. 1차년도 조사의 가중치 부여방법

앞 장에서 살펴보았듯이 KReIS 표본은 각각 표본 추출 확률이 상이하고 각 조사구 내에서의 50세 이상 가구원을 가지는 가구의 분포 또한 다르기 때문에 적절한 가중치를 사용하여 대표성을 확보할 수 있다. 따라서 본 장에서 표본 추출 확률을 바탕으로 한 기본 가중치 결정과 인구학적 특성이 모집단의 분포와 근접하도록 보정한 최종 가중치 부여 방법을 살펴보기로 한다.

1) 기본 가중치 결정

기본 가중치는 표본 설계에서 사용된 표본 추출 확률을 바탕으로 하되 표본 설계 당시 고려하지 못하였던 50세 이상 가구원을 가지는 가구 수의 분포를 이용하여 50세 이상의 가구에 대해 대표성을 가지도록 기본 가중치를 결정한다. 또한 표본 설계 당시에 추출되었던 표본 가구들이 접촉 불능, 부적격, 거절 등의 현실적인 이유로 최종 표본으로 선정되지 못하고 현장에서 유사 가구로 대체되는 경우가 종종 있으므로 이러한 현실들을 반영하여 기본 가

(23가구) 및 개인을 제외하였으며 이때의 최종 원표본 가구는 5,110가구이며 개인조사 원표본가구원은 8,567명이다. 1차년도에 조사되었어야 하나, 2차년도에 처음 조사된 개인조사대상자 122명은 2차년도부터 원표본으로 구축하였다.

30) 본 장에서는 학술연구용역 “1차 본조사 가중치 연구용역 보고서(연세대학교 김재광교수)”, “2차 본조사 가중치 연구용역 보고서(연세대학교 김재광교수)”, “3차 본조사 가중치 연구용역 보고서(고려대학교 안형진교수)”를 요약한 내용으로 자세한 사항은 위 보고서에 기술되어 있다.

중치가 현재의 목표 모집단에 최대한 대표성을 가지도록 결정해야 한다.

<표 V-3> 모집단과 표본자료의 분포

지역		가구			개인		
		모집단	표본	비율(%)	모집단	표본	비율(%)
서울특별시	동부	1,307,522	1,046	0.08	11,387,806	1,809	0.02
부산광역시	동부	556,365	388	0.07	938,292	614	0.07
대구광역시	동부	331,235	231	0.07	571,444	382	0.07
인천광역시	동부	287,280	272	0.09	509,476	452	0.09
광주광역시	동부	167,984	157	0.09	285,916	266	0.09
대전광역시	동부	166,657	174	0.10	292,828	285	0.10
울산광역시	동부	113,520	150	0.13	191,823	249	0.13
경기도	동부	859,082	662	0.08	1,566,375	1,104	0.07
	읍면부	245,801	277	0.11	464,757	468	0.10
강원도	동부	123,664	63	0.05	214,758	94	0.04
	읍면부	120,004	107	0.09	211,697	186	0.09
충청북도	동부	102,274	95	0.09	181,938	160	0.09
	읍면부	116,888	94	0.08	210,086	159	0.08
충청남도	동부	72,158	60	0.08	127,124	108	0.09
	읍면부	243,515	171	0.07	433,639	315	0.07
전라북도	동부	164,710	84	0.05	280,174	141	0.05
	읍면부	157,642	128	0.08	266,731	208	0.08
전라남도	동부	102,003	56	0.05	169,875	97	0.06
	읍면부	287,264	158	0.06	472,611	272	0.06
경상북도	동부	165,528	150	0.09	283,251	244	0.09
	읍면부	306,131	173	0.06	522,627	283	0.05
경상남도	동부	211,377	138	0.07	358,905	227	0.06
	읍면부	255,290	234	0.09	427,089	380	0.09
제주도	동부	48,033	22	0.05	79,744	34	0.04
	읍면부	31,314	43	0.14	50,996	69	0.14
계		6,543,241	5,133	0.08	20,499,962	8,606	0.04

<표 V-3>에서 모집단은 2000년 센서스 자료로부터 얻어진 자료이고 표본은 2005년 표본 설계로부터 얻어진 자료이며, 대전광역시나 울산광역시에서는 표본이 과대표집이 되었고 강원도 동부나 전라남도 읍면부, 경상북도 읍면부, 제주도 동부 등은 표본이 과소표집이 되었음을 알 수 있다. 이는 2000년과 2005년의 5년의 기간 동안 지역 간 인구이동에 기인한 것으로 볼 수 있고 또한 표본 추출 방법이 50세 이상의 가구원 정보를 사용하지 않은 것 때문으로도 볼 수 있다.

이와 같이 표본조사구의 분포와 표본조사구 내 가구의 분포를 알면, 기본 가중치는 조사구 추출확률과 조사구 내 표본 추출확률의 곱의 역수로 산출할 수 있다. 여기서 조사구 추출확률은 명부 가구 수에 비례하며, 조사구내 표본 추출확률은 조사구 내 유효가구수중 표본가구수의 비율이므로 그 식은 다음과 같다. 단, 조사구내유효가구는 조사구 내에서 50세 이상 가구원이 실제 거주하고 있는 가구를 말한다.

$$\text{◆기본가중치} = \frac{1}{\text{조사구추출 확률}} \times \frac{1}{\text{조사구내 표본추출확률}}$$

- 조사구추출 확률 \propto 명부가구수
- 조사구내 표본가구추출 확률 = $\frac{\text{표본가구수}}{\text{조사구내 유효(50세 이상)가구수}}$

이때, 조사구 추출확률이 명부 가구 수에 비례하므로 기본가중치의 합이 모집단 수와 일치하도록 비례상수를 결정해야 한다. 가구 가중치의 경우 지역별로 50세 이상 가구주를 가진 표본 가구들의 가중치 합이 50세 이상 가구주를 가진 센서스 가구들의 수와 같아

지도록 비례 상수를 결정하였다. 이렇게 해서 얻어진 비례 상수를 50세 미만의 가구주 가구에 지역별로 그대로 적용하였다. 마찬가지로, 개인 가중치 경우 지역별로 50세 이상의 표본들의 가중치 합이 50세 이상의 센서스 인구와 같아지도록 비례 상수를 결정하였다. 이렇게 해서 얻어진 비례 상수를 50세 이하 표본들의 기본 가중치 계산에 지역별로 그대로 적용하였다. 신축 아파트의 경우에는 지역별로 단순히 표본 추출 확률을 계산하여 그 역수로 기본 가중치를 결정하였다.

2) 가중치 보정 방법

위에서 얻어진 기본 가중치는 표본 설계 당시의 모집단인 2000년도의 모집단의 분포를 반영한 것이다. 이러한 자료가 현재년도의 모집단의 분포를 반영하기 위해서는 기본 가중치에 보정 (adjustment)을 하는 것이 중요하다. 이를 위해서 통계청에서 발표되는 지역별 연령별 성별 분포를 국민노후보장 패널조사 표본에 시의성 있게 반영하도록 최종 가중치를 결정하여야 할 것이다. 이러한 calibration 을 실제적으로 구현하는 방법으로 raking ratio adjustment 방법을 사용하였으며, 각 cell 별 가중치 보정 factor를 구하였다.

한 가지 특이한 점은 연령대별 성별 가중치 보정 factor의 분포가 다르다는 것이다. 아래의 <표 V-4>과 <표 V-5>에서 보면 가구주의 연령대가 낮을수록 가중 factor 가 높고 여자보다는 남자가 높는데 특히 50~54세의 경우에는 상당히 가중 factor 가 높음을 알 수 있었다. 이는 일반적으로 연령대가 낮을수록 응답률이 낮아짐을 보여주는데 특별히 50~54세의 경우가 응답률이 낮았다기보다는 특

별히 과소표집된 것으로 볼 수 있는데, 이는 50~54세의 경우 응답을 회피하기 위하여 나이를 낮추어 대답하여 유효 가구에 포함되지 않도록 한 것으로 해석된다.

<표 V-4> 연령 및 성별 가구가중치 보정 factor

구분		factor		
		평균	최소	최대
연령대	50 - 54세	1.66	0.34	6.44
	55 - 59세	1.16	0.25	5.35
	60 - 64세	0.90	0.15	1.33
	65 - 69세	0.77	0.29	2.47
	70 - 74세	0.78	0.25	1.55
	75세이상	0.84	0.23	1.92
성별	남	1.09	0.41	6.44
	여	0.85	0.15	2.60
계		1.02	0.15	6.44

<표 V-5> 연령 및 성별 개인가중치 보정 factor

구분		factor		
		평균	최소	최대
연령대	50 - 54세	1.61	0.63	8.73
	55 - 59세	1.08	0.77	3.88
	60 - 64세	0.86	0.57	1.29
	65 - 69세	0.79	0.51	2.41
	70 - 74세	0.83	0.38	1.40
	75세이상	0.90	0.40	1.61
성별	남	1.07	0.53	8.73
	여	0.96	0.38	4.70
계		1.01	0.38	8.73

3) 최종 가중치

기본 가중치에 가중치 보정 factor를 곱하면 최종가중치가 산출된다. 각 영역별로 기본 가중치의 합과 최종 가중치의 합이 계산되고 모집단 참값(센서스 값)과 비교해보면 대부분의 지역에서 최종 가중치의 합은 센서스에서 얻어진 모집단 가구 수(세대수, 인구수)에 매우 근접한다.³¹⁾

나. 2차년도 조사 이후 가중치 부여 방법³²⁾

1) 횡단면 가중치와 종단면 가중치 부여 대상

KReIS 2차년도 이후의 원표본 가구 및 개인조사 대상자 자료는 탈락(attrition)이나 단위 무응답(unit nonresponse)으로 인해 1차년도 자료의 부분 집합으로 구성되어 있다. 이러한 단위 무응답으로 인해 첫째, 표본수가 줄어들어 효율이 떨어지고 둘째, 응답가구 표본이 전체 모집단을 대표하지 못하는 대표성의 문제를 발생할 수 있다는 것이다. 이렇게 표본이 추출 방법의 문제로 모집단을 대표하지 못하는 문제를 추출편향 (selection bias) 이라고 한다.

따라서 이러한 무응답 경향을 반영한 가중치 보정 작업을 수행

31) 통계청 2005년 인구주택총조사 결과를 반영하여 2005년 시점으로 가중치를 수정하였다.

32) KReIS 자료에서 가중치를 부여하는 방법은 크게 1차년도 조사에 응답한 가구 및 개인조사 응답자와 2차년도 이후 새롭게 응답한 가구 및 개인조사 응답자에 대한 방법으로 나뉘어 진다. 본 장에서는 이와 같이 1차년도 조사에 응답한 원표본 가구 및 원표본 개인조사 응답자에 대한 2차년도 조사 이후 가중치 부여 방법에 대해 검토한 후, 다음 장에서 분가가구 및 신규로 진입한 개인 조사 대상자에 대한 가중치 부여 방안을 검토하도록 한다.

하였고 1차년도 자료에서 사용되었던 2005년 센서스 자료의 벤치마킹을 통한 추가적인 가중치 보정 작업이 2차년도 이후 자료에도 적용되어서 조사 차수마다 가중치 간 일관성을 유지하도록 하였다. 아래 <표 V-6>에서 횡단면 및 종단면 가중치를 산출할 조사대상자의 패턴을 고려해 볼 수 있다.

<표 V-6> 3차까지 진행된 패널조사의 무응답 패턴

패턴	응답상태	조사시점		
		1차조사	2차조사	3차조사
1	완전응답	○	○	○
2	감소	○	○	X
3		○	X	X
4	간헐적 무응답	○	X	○

주 1) ○ : 응답, X : 무응답

<표 V-6>는 분석을 위하여 가능한 응답자의 형태를 보여준다. 먼저 각 조사시점의 횡단면 분석 대상은 다음과 같다. 예를 들어, 1차년도 조사의 분석은 패턴 1에서 4까지의 모든 표본을 포함한다. 2차년도 조사의 횡단면 분석은 <표 V-6>의 패턴 1과 2에 관련된 조사 대상자들을 이용한다. 이 때 패턴 3과 4에 속하는 조사대상자는 무응답자로 간주하여 가중값을 구한다. 또한 2차년도 조사의 횡단면 가중값은 2차년도 조사에서 무응답이나 1차년도 조사에서는 응답한 조사대상자의 정보를 이용할 수 있다. 3차년도 조사의 횡단면 분석은 패턴 1과 4에 속하는 조사 대상자들을 이용할 수 있다. 3차년도 조사의 횡단면 가중값은 3차년도 조사에서 무응답이나 1차년도 조사에서는 응답한 조사 대상자의 정보를 이용할 수 있다.

한편, 패널자료의 종단 가중치의 경우, 3차년도의 종단면 분석은

<표 V-6>의 패턴 1에 속하는 표본을 이용한다. 종단 가중치는 1차, 2차, 3차년도 모두 응답한 개인을 대상으로 산출한다. 종단 가중치는 1차년도 조사의 가중치, 2차년도 조사의 횡단가중치, 그리고 1차년도 조사와 2차년도 조사에서 모두 응답했으나 3차년도 조사에서는 무응답한 조사 대상자의 정보를 이용하여 구한 가중치의 곱으로 구하게 된다.

이렇게 횡단면 및 종단면 가중치를 산출할 조사 대상자를 정한 후, 각 각의 가중치를 구하는 작업을 한다. 이때 자료 분석의 일관성을 유지하기 위하여 2차년도 조사 자료의 횡단면 가중치와 3차년도 조사 자료의 횡단면 가중치 및 종단면 가중치 계산할 때 개발한 방법 및 보조변수를 동일하게 적용한다. 또한 3차년도 종단 가중치 계산 시 1차 조사 때의 보조정보만을 이용하였는데 1차년도와 2차년도 모두 응답한 표본에서 얻을 수 있는 보조정보가 더 존재할 때 이를 층 구분에 이용하는 경우 무응답 편의를 좀 더 줄일 수 있으므로 이러한 보조정보를 이용할 수 있는 가능성을 고려할 수 있을 것이다.

2) 무응답 가중치 보정 방법

전체적으로 가중치 작업은 나무모형 중 응답률의 동질성을 분류하는 기법인 CHAID (Chi-Square Automatic Interaction Detection)³³⁾ 알고리즘을 이용하여 무응답층을 분류한 후, 이러한 무응답층을 이용한 무응답 조정 가중치 작업, 가중치 보정(calibration), 그리고 가중치 이상점 제거(weight trimming)의 세 단계로 진행한다.

33) Kass, G.V. (1980) "An exploratory techniques for investigating large quantities of categorical data", *Applied Statistics*, 29, p. 119-127.

<표 V-7> tree모형에 쓰인 변수

변 수	설 명
지역	1. 서울 2. 광역시 3.도
동부 읍면부	1. 동부 2. 읍면부
거주하는 주택형태	1. 단독주택 2. 아파트 3. 연립/다세대주택 4. 점포주택 등 복합용도주택 5. 비거주용 건물 내 주택 6. 기타
거주하는 주택소유형태	1. 자가 2. 전세 3. 월세 4. 기타
가구원수	
50세이상 가구원수	
가구주 연령	
가구주 성별	1. 남 2. 여
가구주 최종학력	1. 무학 2. 초등학교 재학/중퇴/졸업 3. 중학교 재학/중퇴/졸업 4. 고등학교 재학/중퇴/졸업 5. 대학교(2,4년제) 재학/중퇴/졸업 6. 대학원석사 재학/중퇴/졸업 7. 대학원박사 재학/중퇴/졸업 8. 미조사 및 무응답
가구 내 취업자 수	

<표 V-7>는 무응답층을 만들기 위한 나무모형에 쓰인 변수에 대한 설명이다. 이와 같이 인구학적, 경제학적, 지역적 변수를 이용

하여 전체 가구를 약 20~30여개의 cell로 나누었다.³⁴⁾ 개인 무응답 보정승수를 계산할 때는 각 개인들이 속한 가구에 해당되는 cell이 개인의 무응답 cell이 되었고 각 cell별로 무응답 보정승수를 계산하였다. 무응답 보정승수의 식은 아래와 같다.

$$f_g = \frac{\sum d_i}{\sum d_i R_i} \quad \begin{array}{l} d = 1\text{차년도 가구 최종가중치,} \\ R = \text{응답여부 지시변수 (1:응답, 0:무응답)} \end{array}$$

이렇게 해서 계산된 무응답 보정승수를 1차년도 가구 최종가중치에 곱하면 무응답 조정 가중치를 얻게 된다.

가구 가중치의 보정에는 1차년도와 마찬가지로 다음과 같은 벤치마킹 변수가 사용되었다.

- (a) 지역별 동부/읍면부
- (b) 지역별 가구주 연령대 : 30대 이하 / 40대 / 50대 / 60대 / 70대 이상
- (c) 지역별 가구주 성별 : 남 / 여
- (d) 지역별 가구원 수 : 1인 / 2인 / 3인 / 4인 이상
- (e) 거주종류 : 아파트 / 단독주택 / 다세대 또는 연립주택 / 기타
- (f) 점유형태 : 자가 / 전세 / 월세 / 기타

이렇게 해서 얻어진 최종 가중치는 각 지역별로 평균 가중치의

³⁴⁾ 2차년도에는 34개의 cell로 나뉘어졌으며, 3차년도의 경우 횡단면 가중치 보정 시에는 30개의 cell로, 종단면 가중치 보정시에는 23개의 cell로 구분되었다.

1/3 이하 이거나 3배 이상이 되는 경우를 가중치 이상점으로 규정하고 가중치 절삭을 실시하였다. 가중치 절삭 후, 다시 가중치 평균을 계산하여 절삭 전과 절삭 후의 평균의 비를 곱해 줌으로써 지역별 가중치 합이 같아지도록 하였다.

3) 분가가구 및 신규개인조사대상자 가중치 부여 방법

KReIS에서는 원가구의 개인조사대상자가 분가하여 새로운 가구를 구성한 경우를 분가가구로 정의한다. 이때 분가가구의 횡단면 가중치는 분가하기 이전 원가구의 1차년도 가중치를 동일하게 부여하며, 종단면 가중치는 부여하지 않는다. 한편, 1차년도 개인조사대상자 이외에 2차년도 이후 신규개인조사대상자가 된 경우, 해당 신규개인조사대상자 거주지역의 전체 개인조사대상자 가중치 평균값을 부여한다.

4. 표본 특성에 따른 가중치 활용 방법

가. KReIS 표본의 특성

지금까지 살펴본바와 같이, KReIS는 2005년 만 50세 이상의 가구원이 있는 가구를 추출하고, 그 가구에 속하는 50세 이상의 가구원 및 그의 배우자를 개인조사대상으로 하며, 표본 설계 당시 고려하지 못하였던 50세 이상의 가구에 대해 대표성을 가지도록 가중치를 부여하였다. 또한 조사차수가 더해감에 따라 생기는 무응답으로 인해 생기는 표본 대표성의 문제를 해결하고자 2, 3차년도에 무응

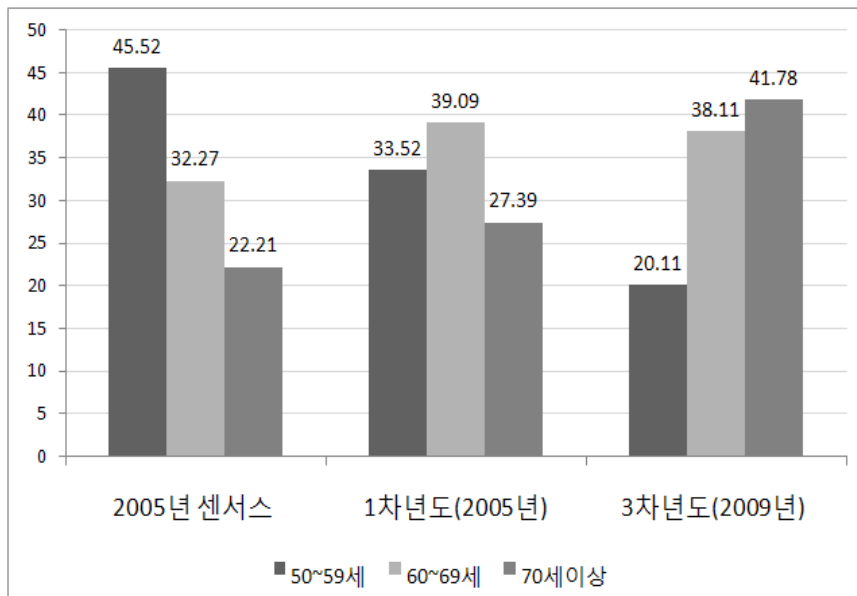
답 가중치 조정을 하여, 횡단면 가중치와 종단면 가중치를 개발하였다. 그렇다면 근본적인 문제로 돌아가서 KReIS 표본이 대표하고자 하는 모집단은 무엇일까?

[그림 8]은 2005년 인구주택총조사(이하 2005년 센서스)자료의 50대 이상 가구주의 분포와 KReIS 1차년도와 3차년도에서의 50대 이상 가구주의 분포를 비교해 볼 수 있다. 먼저 1차년도 자료와 비교해 보면, 2005년 센서스에서 50대 가구주는 45.52% 정도를 차지하지만 KReIS 1차년도 자료의 50대 가구주는 33.52%이고, 60대 가구주는 각각 32.27%, 39.09%였고, 70대 이상의 가구주는 2005년 센서스에서 22.21%, 1차년도 자료에서 27.39%로 KReIS의 가구주 연령대가 센서스 자료에 비해 고령층의 비중이 크다는 것을 알 수 있다. 또한 4년 뒤, 3차년도 자료에서는 50대 가구주가 20.11%, 60대 가구주가 38.11%, 70대 이상 가구주가 41.8%로 그 차이가 더 커지고 있음을 볼 수 있다.

이와 같이 2005년 센서스 자료와 1차년도 자료의 50세 이상 분포가 차이나는 이유는 표본 설계 당시 50세 이상 가구주 또는 50세 이상의 가구원의 분포를 고려하지 않고, 표본유치조사시 조사의 편의에 의해 60, 70대 고령자의 응답률이 50대의 응답률 보다 높았기 때문인 것으로 생각된다.

[그림 8] 만 50세 이상 가구주 연령 분포 비교

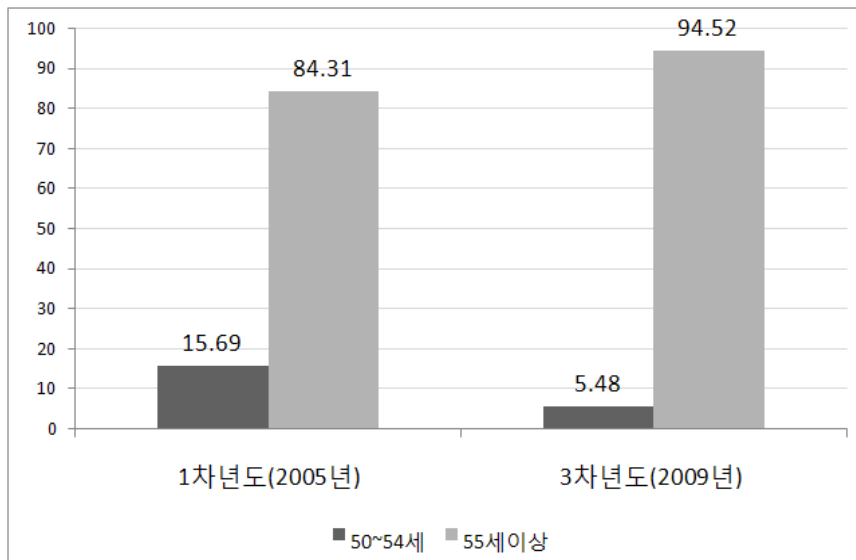
(단위 : %)



주 1) 가중치를 적용하지 않고 가구주의 연령이 만 50세 이상인 경우를 100으로 놓았을 때 각 사례 수의 백분율을 나타냄.

또한 [그림 9]에서 볼 수 있듯이, 1차년도 당시 50세 이상이라는 특정 연령층을 고려하여 표본을 추출하였기 때문에 50세 미만의 가구주 및 가구원이 상당히 과소 표집되어, 4년이 지나자 만 50~53세인 가구주의 비율이 1차년도 당시 15.69%에서 3차년도에 5.48%로 급격히 감소하였다.

[그림 9] 1차년도와 3차년도 만 50세 이상 가구주 연령 분포 비교
(단위 : %)



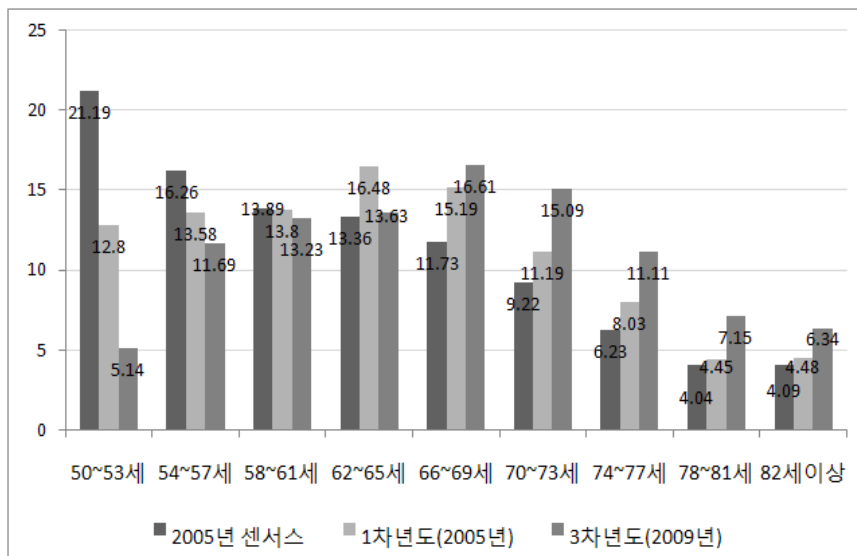
주 1) 가중치를 적용하지 않고 가구주의 연령이 만 50세 이상인 경우를 100으로 놓고, 사례 수의 백분율을 나타냄.

개인조사응답자의 연령 분포를 살펴보면, [그림 10]에서 가구주의 연령 분포를 비교했을 때와 마찬가지로 50대까지는 센서스의 자료에서의 비중이 더 크지만, 60세 이후부터는 KReIS에서의 비중이 더 크고, 1차년도(2005년)에서 3차년도(2009년)로 갈수록 고령자가 차지하는 비중이 커지고 있다. 특히 표본에 고령자가 많이 추출되어 조사대상자의 사망으로 인한 표본 소실이 3차년도 당시 원표본 개인응답자의 7.8%에 해당하였음에도 불구하고 3차년도의 고령자 비중이 여전히 센서스자료나 1차년도 자료에 비하여 크다는 사실을 볼 수 있다. 또한 50~53세의 경우, 그 차이가 다른 연령에 비하

여 매우 크다.

[그림 10] 만 50세 이상 개인조사응답자의 연령 분포 비교

(단위 : %)



주 1) 가중치를 적용하지 않고 개인조사응답자의 연령이 만 50세 이상인 경우를 100으로 놓았을 때 각 사례 수의 백분율을 나타냄.

나. 조사 차수에 따른 표본의 변화

앞 서에서는 특정 연령층을 대상으로 한 목표 모집단이 시간이 흘러 연령이 상승함에 따라 추출틀에 얼마나 벗어나는지 정도를 살펴보았다면, 본 장에서는 패널 조사에서 표본 이탈로 인해 그 분포가 얼마나 달라지는지 인구학적 분포를 통해 분석해 보고자 한다. 또한 변화를 살펴보기 위하여 1차년도, 2차년도, 3차년도 자료를 분석하고, 3차년도의 경우 원표본의 분포와 1차년도부터 3차년

도까지 계속 응답한 가구에 대한 분포를 비교해 보도록 한다.

KReIS의 조사결과는 1차 본조사(2005년)에서는 5,109가구를 조사 성공하였고, 2차 본조사(2007년)에서는 4,590가구, 3차 본조사(2009년)에서는 4,308가구를 조사에 성공하였으며 3차년도 조사당시 원표본가구는 4,240가구이고, 1차년도부터 3차년도까지 계속 응답한 가구는 4,171가구이다.

<표 V-8>에서는 조사 차수별 조사 성공 가구들의 특성 중 거주지 관련 변수들에 대한 분포를 볼 수 있다. 거주지의 경우 1차년도 당시 서울시에 거주했던 가구의 비중은 20.4% 정도로 나타났으나, 2차년도에 18.9%, 3차년도에 17.2%로 점차 감소하며, 반대로 도에 거주하는 가구에 비중은 2차년도 52.8%에서 2차년도 54.4%, 3차년도 55.9%로 점차 증가하고 있는 것을 볼 수 있다. 이는 일반적으로 표본 조사의 경우 서울의 응답률이 그밖에 지역의 응답률에 비하여 낮고, 고령층일수록 특별시나 광역시보다 도지역에 거주하는 비중이 크기 때문인 것으로 생각된다. 점유형태별로 살펴보면, 자가의 비율이 1차년도 73.2%에서 3차년도에는 77.6% 정도로 증가하였다.

가구 특성별로 살펴보면 <표 V-9>에서 세대구성별로 1인가구가 1차년도 17.8%에서 2차년도 20.4%, 3차년도 23.6%로 점점 증가하였고, 1세대 가구의 경우에도 증가하였으며, 반면 2세대 이상의 가구의 비중은 감소하였음을 볼 수 있다. 이와 마찬가지로 가구원수별로 보면, 가구원 수가 1인, 2인 가구는 점차 증가하는 반면 3인 이상이 가구는 점점 감소하는 것을 볼 수 있다.

이때 3차년도 전체 가구의 분포와 3차년도 가구 중 원표본 가구, 1차~3차 계속응답 가구의 분포를 비교해 보면 거의 분포가 비슷하다는 사실을 알 수 있다.

<표 V-8> 조사 차수에 따른 가구의 거주지 및 주택 분포 비교

(단위: 개, %)

구 분		1차년도 (2005)	2차년도 (2007)	3차년도 (2009)	3차년도 원표본가 구	계속 응답가구
시도	특별시	1,043 (20.4)	869 (18.9)	741 (17.2)	729 (17.2)	721 (17.3)
	광역시	1,371 (26.8)	1,226 (26.7)	1,158 (26.9)	1,144 (27.0)	1,113 (26.7)
	도	2,696 (52.8)	2,495 (54.4)	2,409 (55.9)	2,367(55. 8)	2,337 (56.0)
거주 형태	단독주택	2,674 (52.3)	2,381 (51.9)	2,228 (51.7)	2,200 (51.9)	2,170 (52.0)
	일반 아파트	1,529 (29.9)	1,382 (30.1)	1,323 (30.7)	1,293 (30.5)	1,268 (30.4)
	연립/ 다세대	703 (13.6)	606 (13.2)	555 (12.9)	549 (13.0)	539 (12.9)
	복합 용도주택	133 (2.6)	159 (3.5)	160 (3.7)	158 (3.7)	156 (3.7)
	비거주용 건물 내 주택	44 (0.9)	37 (0.8)	26 (0.6)	25 (0.6)	24 (0.6)
	기타	26 (0.5)	26 (0.6)	11 (0.3)	11 (0.3)	11 (0.3)
	모름/ 무응답	-	-	5 (0.1)	4 (0.1)	3 (0.1)
점유 형태	자가	3,742 (73.2)	3,308 (72.1)	3,343 (77.6)	3,304 (77.9)	3,252 (78.0)
	전세	541 (10.6)	466 (10.2)	430 (10.0)	420 (9.9)	411 (9.9)
	월세	478 (9.4)	406 (8.8)	387 (9.0)	293 (8.8)	366 (8.8)
	기타	347 (6.8)	410 (8.9)	148 (3.4)	143 (3.4)	142 (3.4)
	모름/ 무응답	1 (0.0)	-	-	-	-
전 체	5,109 (100.0)	4,590 (100.0)	4,308 (100.0)	4,240 (100.0)	4,171 (100.0)	

<표 V-9> 조사 차수에 따른 가구의 특성 비교 (단위 : 개, %)

구 분		1차년도 (2005)	2차년도 (2007)	3차년도 (2009)	3차년도 원표본 가구	계속 응답가구
세대 구성	1인	908 (17.8)	934 (20.4)	1,016 (23.6)	996 (23.5)	984 (23.6)
	1세대	1,409 (27.6)	1,355 (29.6)	1,360 (31.6)	1,352 (31.9)	1,337 (32.0)
	2세대	2,014 (39.5)	1,619 (35.4)	1,347 (31.3)	1,330 (31.4)	1,308 (31.4)
	3세대	765 (15.0)	656 (14.4)	571 (13.3)	548 (12.9)	528 (12.7)
	4세대	8 (0.2)	7 (0.2)	14 (0.3)	14 (0.3)	14 (0.3)
가구 원수	1인	907 (17.8)	934 (20.4)	1,016 (23.6)	996 (23.5)	984 (23.6)
	2인	1,759 (34.4)	1,687 (36.8)	1,663 (38.6)	1,649 (38.9)	1,630 (39.1)
	3인	1,015 (19.9)	876 (19.1)	785 (18.2)	772 (18.2)	757 (18.1)
	4인	765 (15.0)	570 (12.4)	442 (10.3)	431 (10.2)	420 (10.1)
	5인	428 (8.4)	317 (6.9)	235 (5.5)	227 (5.4)	221 (5.3)
	6인 이상	236 (4.6)	206 (4.5)	167 (3.9)	165 (3.9)	159 (3.8)
	평균	2.8명	2.6명	2.48명	2.5명	2.5명
전 체	5,109 (100.0)	4,590 (100.0)	4,308 (100.0)	4,240 (100.0)	4,171 (100.0)	

<표 V-10>과 <표 V-11>에서는 조사 차수에 따른 가구주의 특성과 개인조사 대상자의 특성 분포를 비교할 수 있다. 먼저 가구주의 특성을 성별로 살펴보면, 남성의 경우 1차년도 당시 71.1%에서

2차년도 69.5%, 3차년도에 67.7%로 감소하였다. 연령대별로 살펴보면 50세 미만의 가구주가 1차년도 17.6%에서 2차년도에 14.6%, 3차년도에 12.3% 정도로 감소하였고, 50~54세의 경우 1차년도 당시 13.0%에서 3차년도에 4.8%로 그 비중이 급속히 감소하였음을 볼 수 있다. 반면 60세 이상인 경우는 비중이 점점 증가하고 있으며 특히 75세 이상이 차지하는 비중은 1차년도 10.8%에서 3차년도에 19.2%로 두 배 정도 증가하였음을 알 수 있다.

<표 V-11>에서는 조사 차수에 따른 개인응답자의 분포를 비교해 볼 수 있다. KReIS의 개인조사대상자는 조사 가구의 만 50세 이상의 가구원 및 그의 배우자로 배우자의 경우, 연령이 만 50세 미만일 수 있다. 그러므로 개인조사 자료로 분석을 할 때 응답 자료를 모두 사용한다면, 그 표본이 대표하는 연령이 모호해지므로 자료 분석하기 전에 분석하고자 하는 연령대에 대한 대표성이 있는 자료를 구성해야 할 것이다.

개인조사결과를 살펴보면 1차 본조사(2005년)에서는 8,689명을 조사 성공하였고, 2차 본조사(2007년)에서는 7,608명, 3차 본조사(2009년)에서는 7,023명을 조사에 성공하였으며 3차년도 조사당시 원표본 개인조사응답자는 6,809명이고, 1차년도부터 3차년도까지 계속 응답한 개인조사응답자는 6,629명이다. 성별 분포는 큰 차이가 없었으나, 연령별로 보았을 때 <표 III-33>에서 가구주의 특성별로 살펴본 바와 같이 50세 미만의 비중이 1차년도 17.6%에서 3차년도 12.3%로 감소하였고 50~54세의 비중이 1차년도 13.0%에서 2차년도 8.1%, 3차년도 4.8% 정도로 크게 감소하였다. 이 때 2차년도와 3차년도의 50~54세인 신규 가구원의 경우 각각 49명, 72명으로 표본이 시간이 지남에 따라 50세 이상의 대표성을 유지하기에 그 수가 부족하다는 것을 알 수 있다.

<표 V-10> 조사 차수에 따른 가구주의 특성 비교 (단위: 명, %)

구분		1차 년도	2차 년도	3차 년도	원표본 가구 (3차)	계속 응답 가구	신규 응답 (2차)	신규 응답 (3차)
성별	남성	3,633 (71.1)	3,192 (69.5)	2,918 (67.7)	2,876 (67.8)	2,821 (67.6)	23 (65.7)	19 (57.6)
	여성	1,477 (28.9)	1,398 (30.5)	1,390 (26.8)	1,364 (32.2)	1,350 (32.4)	12 (34.3)	14 (42.4)
연령대	50세 미만	900 (17.6)	668 (14.6)	529 (12.3)	509 (12.0)	497 (11.9)	14 (40.0)	6 (18.2)
	50~ 54세	662 (13.0)	370 (8.1)	207 (4.8)	199 (4.7)	194 (4.7)	3 (8.6)	5 (15.2)
	55~ 59세	752 (14.7)	653 (14.2)	553 (12.8)	542 (12.8)	530 (12.7)	4 (11.4)	7 (21.2)
	60~ 64세	816 (16.0)	713 (15.5)	659 (15.3)	652 (15.4)	640 (15.3)	4 (11.4)	3 (9.1)
	65~ 69세	830 (16.2)	819 (17.8)	781 (18.1)	776 (18.3)	763 (18.3)	3 (8.6)	2 (6.1)
	70~ 74세	598 (11.7)	679 (14.8)	750 (17.4)	740 (17.5)	733 (17.6)	3 (8.6)	7 (21.2)
	75세 이상	552 (10.8)	688 (15.0)	829 (19.2)	822 (19.4)	814 (19.5)	4 (11.4)	3 (9.1)
최종 학력	무학	1,019 (19.9)	855 (18.6)	819 (19.0)	812 (19.2)	799 (19.2)	4 (11.4)	3 (9.14)
	초등 학교	1,223 (23.9)	1,192 (26.0)	1,165 (27.0)	1,149 (27.1)	1,133 (27.2)	7 (20.0)	9 (27.3)
	중학교	862 (16.9)	689 (15.0)	651 (15.1)	640 (15.1)	631 (15.1)	5 (14.3)	6 (18.2)
	고등 학교	1,279 (25.0)	1,176 (25.6)	1,098 (25.5)	1,078 (25.4)	1,061 (25.4)	11 (31.4)	9 (27.3)
	대학교	616 (12.1)	613 (13.4)	519 (12.1)	508 (11.9)	494 (11.9)	6 (17.1)	5 (15.1)
	대학원	60 (1.2)	65 (1.4)	51 (1.2)	49 (1.2)	49 (1.2)	1 (2.9)	1 (3.0)
	모름/ 무응답	51 (1.0)	-	5 (0.1)	4 (0.1)	4 (0.1)	1 (2.9)	-
전 체	5,109	4,590	4,308	4,240	4,171	35	33	

<표 V -11> 조사 차수에 따른 개인조사응답자의 특성 비교

(단위: 명, %)

구분		1차 년도	2차 년도	3차 년도	원 표본 (3차)	계속 응답 가구	신규 응답 (2차)	신규 응답 (3차)
성별	남성	3,552(40.9)	3,094(40.7)	2,826(40.2)	2,717(39.9)	2,651(40.0)	43(50.0)	66(51.6)
	여성	5,137(59.1)	4,514(59.3)	4,197(59.8)	4,092(60.1)	3,978(60.0)	43(50.0)	62(48.4)
연령대	50세 미만	439(5.1)	198(2.6)	113(1.6)	69(1.0)	65(1.0)	15(17.4)	29(22.7)
	50~54세	1,307(15.5)	830(10.9)	520(7.4)	399(5.9)	383(5.8)	49(57.0)	72(56.3)
	55~59세	1,459(16.8)	1,234(16.2)	1,063(15.1)	1,049(15.4)	1,012(15.3)	4(4.7)	10(7.8)
	60~64세	1,578(18.2)	1,310(17.2)	1,191(17.0)	1,181(17.3)	1,155(17.4)	6(7.0)	4(3.1)
	65~69세	1,571(18.1)	1,544(20.3)	1,393(19.8)	1,387(20.4)	1,357(20.5)	5(5.8)	1(0.8)
	70~74세	1,103(12.7)	1,183(15.5)	1,276(18.2)	1,275(18.7)	1,253(18.9)	-	1(0.8)
	75세 이상	1,232(14.2)	1,309(17.2)	1,467(20.9)	1,449(21.3)	1,404(21.2)	7(8.1)	11(8.6)
교육 수준	무학	2,310(26.6)	1,778(23.4)	1,588(22.6)	1,566(23.0)	1,520(22.9)	10(11.6)	12(9.4)
	초등학교	2,645(30.4)	2,467(32.4)	2,329(33.2)	2,296(33.7)	2,228(33.6)	9(10.5)	24(18.8)
	중학교	1,566(18.0)	1,315(17.3)	1,210(17.2)	1,176(17.3)	1,151(17.4)	20(23.3)	14(10.9)
	고등학교	1,586(18.3)	1,490(19.6)	1,400(19.9)	1,309(19.2)	1,287(19.4)	35(40.7)	56(43.8)
	대학교	503(5.8)	499(6.6)	446(6.4)	416(6.1)	397(6.0)	12(14.0)	18(14.0)
	대학원	59(0.7)	59(0.8)	44(0.6)	98(1.4)	91(1.4)	4(4.7)	3(2.3)
	모름/무응답	20(0.2)	-	6(0.1)	4(0.1)	4(0.1)	-	2(1.6)
전 체		8,689	7,608	7,023	6,809	6,629	86	128

한편, 앞서 살펴보았듯이, 3차년도에 조사 성공한 전체 가구와 3차년도 조사 성공 가구 중 원표본 가구, 1차~3차년도의 계속응답 가구의 특성 분포가 거의 비슷한 것과 마찬가지로, 가구주 및 개인 조사응답자의 경우에도 그 분포가 크게 다르지 않음을 알 수 있다. 이 사실은 앞에서 살펴본 KReIS의 가중치 부여 방법과 관련하여 큰 시사점을 갖는다. 대부분의 패널자료로 횡단면 분석을 할 때, 해당 조사 년도를 대표하는 표본이라고 생각하고 분석을 하는데 KReIS의 경우, 3차년도 횡단면 자료 분포와 종단면 자료 분포가 비슷하다는 사실을 볼 수 있기 때문에 이러한 분석을 할 때 자료 해석에 주의해야 한다.

다. 가중치 활용 방법

지금까지 KReIS의 표본 설계 및 표본 추출 방법에 따른 표본 구축의 개요와 각 조사의 결과를 살펴보고, 조사에 성공한 표본이 모집단의 대표성을 갖기 위해 가중치를 부여하고 보정하는 작업과정에 대하여 검토해 보았다. 또한 이렇게 구축된 표본이 조사차수가 지나감에 따라 어떻게 변화하는지에 대하여 살펴봄으로써 KReIS의 표본의 특성을 살펴본 결과, 중·고령자 표본이라는 특수성으로 인하여 자료를 이용하여 분석을 할 때 대표하고자 하는 모집단을 설정함에 있어 주의가 필요하다는 점을 인지하였다. 이때 표본이 대표성을 갖기 위한 방법으로 가중치를 부여하였는데 이러한 가중치를 활용하는 방법에 대하여 고찰해보도록 한다.

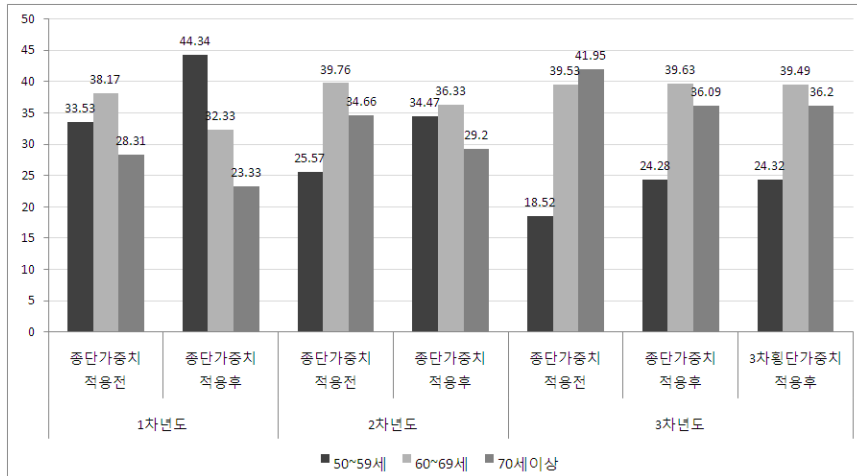
우선 앞에서 살펴본 바와 같이 KReIS에서 종단면 가중치는 1~3차년도 계속 응답 가구 및 계속응답 개인조사대상자에 한하여 부여하였다. 한편, 한국노동패널조사와 복지패널조사의 경우 계속 응

답여부와 관계없이 원표본가구 및 원표본가구원에 대하여 종단면 가중치를 부여하였고, 조사 특성마다 가중치를 부여하는 방법이 다르므로 자료 분석 시 이러한 가중치 부여 방법에 대한 검토가 필요하다.

[그림 11]은 KReIS의 1차년도, 2차년도, 3차년도 자료에 3차년도 종단면 가중치를 적용하였을 때 가구주의 연령분포를 보여준다. 앞에서 살펴본 2005년 센서스 자료의 경우 50세 이상의 연령 분포를 살펴보면 50대가 45.1%, 60대가 31.3%, 70세 이상이 25.6%로, 1차년도의 경우 종단면 가중치를 적용하였을 때 센서스 자료와 어느 정도 근접한 것을 볼 수 있다. 그러나 2차년도와 3차년도 자료의 경우 가중치를 적용하였을 때 어느 정도 보정을 하고 있기는 하지만 센서스 자료와 분포 차이가 여전히 크다. 이렇게 2차년도와 3차년도의 가중치 적용 결과가 가중치 보정작업을 하였음에도 센서스 자료와 큰 차이를 보이는 이유는 1차년도 당시 50세 미만 연령대에 대한 대표성이 없는 표본을 추출하고, 가중치 부여 및 보정을 할 때 센서스의 연령대 분포와 같게 해주었으나 이 또한 50세 이상의 연령대만 보정을 하였기 때문에, 2차년도 이후 가중치 보정 방법은 1차년도 당시의 가중치에 무응답 확률만을 고려하였으므로 가중치를 적용한다고 하더라도 표본의 대표성이 떨어진다.

한편 3차년도의 경우 종단가중치와 횡단가중치를 적용한 결과가 거의 차이가 없는 것을 확인할 수 있다. 이는 종단면가중치 부여 대상이 1차년도부터 3차년도에 모두 응답한 원표본가구 및 원표본개인조사응답자이기 때문에 결국 이는 3차년도 조사성공 가구 및 개인조사대상자의 특성에 근접할 수밖에 없는 것이다.

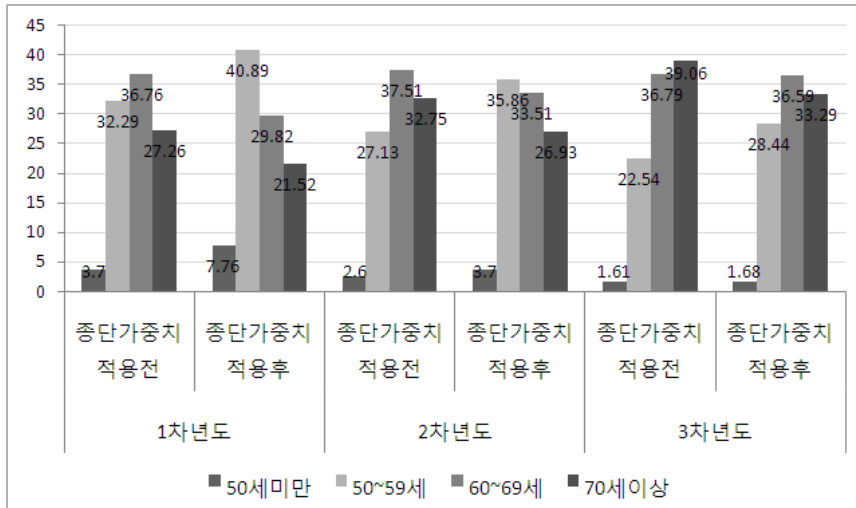
[그림 11] 종단면 가중치 적용 후 개인조사대상자 연령 분포



주 1) 3차년도 종단면 가중치를 적용하였고 가중치를 적용한 분석대상은 1차년도 당시 만 50세 이상인 개인응답자에 한하여 분석함.

[그림 12]에서는 전체 개인조사응답자에 대한 연령분포와 종단면 가중치 적용 결과를 살펴볼 수 있다. 이 그림에서는 50세 미만의 개인조사대상자에 대한 분포 및 가중치를 적용하였을 때의 결과를 비교하고자 한다. KReIS의 개인조사대상자는 조사 당시 만 50세 이상의 가구원과 그의 배우자로, 배우자의 경우 연령과 관계없이 개인조사대상자가 되어 개인자료에 제공되며 1차년도 당시 응답을 하고 2차, 3차년도에 계속 응답 하였다면 종단면가중치를 부여한다. 그러나 조사 설계 당시 50세 이상의 조사 대상 연령을 설정하고 연령에 관계없이 그 배우자를 조사하는 이유는 세대별 분석을 하기 위함이었고, 부부단위의 노후준비 실태를 분석하고자 함이었다. 따라서 이러한 목적으로 분석을 할 때 이외에 개인단위로 분석을 할 때에는 50세 미만의 개인조사 대상자의 경우 표본의 대표성이 없으므로 분석에서 제외해야 할 것이다.

[그림 12] 종단면 가중치 적용 후 전체 개인조사대상자 연령 분포



주 1) 3차년도 종단면 가중치를 적용하였고 1차년도 당시 가중치 적용한 분석대상은 1차년도 당시 만 50세 이상인 개인응답자에 한하여 분석함.

정리하면, KReIS의 자료의 경우 1차년도(2005년), 2차년도(2007년), 3차년도(2009년) 자료가 각각 해당 조사년도인 2005년과 2007년, 2009년을 대표하는 표본이 아니기 때문에 횡단면 분석과 종단면 분석 시 각 연도별 특성을 대표하는 자료로써 해석을 할 때에는 주의가 필요하다. 그러나 지금까지 조사에 참여한 대상에만 종단면 가중치를 부여하였기 때문에 1차년도 조사이후 2차년도, 3차년도 시간에 걸쳐 어떠한 이행을 보이는지 분석하기 용이하다. 특히 원표본이 대표하는 모집단이 50세 이상을 대표하므로 이후 은퇴 및 가족구성의 변화, 경제생활 및 노후준비 실태 관련하여 어떠한 행태를 보이는지 분석을 하기 위해 충분한 표본 수를 확보하였기 때문에 조사 목적에 부합한 연구가 가능할 것이다.

5. 논의

연구자가 어느 집단의 특성을 알고자 할 때, 시간, 비용 등의 제약으로 전체를 조사 할 수 없기 때문에 표본조사를 실시한다. 즉, 동전의 양면과 같이 표본이 있으면 그 표본이 대표하고자 하는 모집단이 반드시 존재하기 마련이다. 본 고에서는 KReIS의 표본 설계 및 표본 추출 방법과 표본 구축 결과, 가중치 적용 및 보정 방법 등에 대하여 살펴봄으로써 이러한 궁극적인 표본조사에 대한 목적을 재고찰하였다.

연구자가 자료를 활용하여 분석할 때, 아무리 분석 방법이 좋다 하더라도 편향된 자료를 이용하여 분석한다면 잘못된 결과를 도출하기 쉽다. 특히 KReIS의 경우 다른 국내 패널조사와는 달리 중·고령자를 대상으로 하는 조사이기 때문에 발생하는 여러 특성들을 고려하지 않고 분석을 한다면 분석하고자 하는 대상이 모호해지고 분석결과에 대한 올바른 해석이 어려워진다. 그러므로 자료에 대한 이해와 자료 구축 방법에 대한 충분한 검토가 사전에 이루어져야 할 것이다.

또한 조사 자료 생산자는 표본 대표성에 대한 심도있는 고민을 통해 다양한 패널 자료의 형태 중 해당 자료의 목적에 맞는 패널 자료 형태를 갖추어 나가는 연구를 지속적으로 해나가야 할 것이다.

<부록 1> 국민노후보장패널조사 표본 추가 방향 설정 연구³⁵⁾

1. 서론

본 연구는 국민연금연구원에서 실시하는 국민노후보장패널조사 자료의 대표성을 분석하고 이를 보완하기 위한 표본 추가 방법론에 대한 선행 연구를 통하여 향후 표본 추출과 표본 관리를 진행하는데 필요한 통계학적 방법론을 제공하는 것을 목적으로 한다.

국민노후보장패널조사는 일반적인 가구단위 패널조사와는 달리 조사 모집단이 가구 자료의 경우 50세 이상 가구원이 있는 가구로 정의하며, 개인 자료의 경우 만 50세 이상 가구원이 있는 가구의 만 50세 이상인 자와 그의 배우자로 정의한다. 이러한 모집단의 특수성 때문에 표본 추출 당시에 상당한 어려움이 있었고, 이러한 특수성을 표본 설계 당시에 완전히 반영하는 것이 기술적으로 어려운 점이 있었기에 이러한 대표성을 보완하기 위하여 가중치 보정이 필수적으로 사용되었다. 또한 2005년에 표본이 추출된 이후로 새로운 표본이 유입되지 않았기에 시간이 지남에 따라 표본이 대표하는 모집단(2005년 당시의 50세 이상 가구원을 가지는 가구)이 현재의 모집단(2010년 현재 50세 이상 가구원을 가지는 가구)과 달라지는 포함 오차 (coverage error)를 극복해야 하는 문제가 발생하였다. 마지막으로 시간이 경과함에 따라 패널 표본이 마모

35) 본 연구는 미국 아이오와 주립대학교 통계학과 김재광 교수에 의해 수행되었다.

(attrition)가 되는 표본 탈락을 보정해야 하는 문제도 발생하게 된다.

즉, 현재 국민노후보장패널 표본은 아래와 같은 세 가지 측면에서 대표성의 문제를 가지게 된다.

- (1) 표본 설계 당시의 대표성 문제
- (2) 모집단의 변화(2005년 50세 이상 -> 2010년 50세 이상)를 반영하지 못하는 대표성 문제
- (3) 패널 표본 마모 (또는 표본 탈락)로 인한 대표성 문제

첫 번째 표본 설계 당시의 대표성 문제는 심각할 수도 있는 것이지만, 이 후 가중치 작업을 통해서 어느 정도 보완하였기 때문에 이러한 표본 설계 자체의 대표성 문제는 일단 해결된 것으로 간주하였다. 둘째 모집단의 변화로 인한 부분은 반드시 해결되어야 하는 부분이며, 셋째 표본 마모로 인한 대표성 문제는 최종년도(3차년도) 가중치를 통해서 어느 정도 해결된 것으로 간주하였다. 다만 특정 지역이나 특정 연령대의 마모가 심해지면 가중치로 해결하는 것이 한계를 가지기 때문에 이를 파악하여 보완하여야 할 것이다.

또한 패널 조사 자체는 횡단면적 모수보다는 추이 변화와 같은 종단면적 모수에 보다 관심을 가지는 조사이기 때문에 표본의 대표성 문제는 그리 우려하지 않을 수도 있다. 그러나 이 자료가 일반 연구자들에게 배포되고, 사용자들이 여러 가지 다양한 분석을 시도하기 때문에 횡단면적 모수를 추정하게 되는 일도 빈번할 것이므로 대표성에 대한 면밀한 점검과 이를 보완하려는 노력은 계속되어야 할 것이다.

이 보고서가 이러한 목표를 달성하는데 도움을 주기 위하여 다

음과 같은 측면에서 연구를 진행하였다.

첫째, 노인 통계에 대한 기본적인 문헌 연구 및 해외 사례 연구를 통해 노인 통계에 대한 조사 방법론에 대한 기본 내용을 정리하고 향후 국민노후보장패널조사 표본에 대한 방향을 제시하고자 하였다.

둘째, 1-3차년도 자료를 분석하여 모집단 대비 대표성을 분석하고 이를 통해 표본 추가를 실시하게 될 경우 필요한 표본 배분과 표본 추출법을 제안하였다.

셋째, 보다 장기적인 관점에서 모집단의 변화를 보다 적절하게 반영할 수 있는 현실적인 방법론을 제안하고 향후 국민노후보장패널조사의 표본 추가 연구 방향을 제시하였다.

2. 연구내용

가. 해외 사례 분석

해외 사례는 미국의 HRS (Health and Retirement Survey), 영국의 ELSA (English Longitudinal Study of Ageing), 유럽연합의 SHARE (Survey of Health, Aging and Retirement in Europe)를 대상으로 하였으며, 표본마모 및 표본 편향과 관련된 내용을 중심으로 정리하였다.

나. 패널 자료 분석

패널 자료 분석은 각 연도별로 자료의 분포를 분석하는 연도별 분석과 1차년도 대비 3차년도의 변화를 분석하는 추이(transition) 분석으로 구분할 수 있다. 연도별 분석은 해당 연도 자료만을 가지

고 분석을 하는 것으로 이차원 분석에 해당하며, 추이 분석은 연도별 자료를 결합하여 실제 변화의 분포를 분석하는 것으로 이차원 분석에 해당한다.

다. 표본 추가 방법론 제안 및 기타 제안

이렇게 얻어진 분석 결과를 바탕으로 코호트 패널 표본을 추가할 경우 적절한 표본수와 표본 배정 방법을 제안한다. 또한 그 밖의 패널 표본 추가 방법 및 모집단 변화를 표본에 반영할 수 있는 다양한 방법론을 제안하였고 보다 장기적인 측면에서 고려해 볼만한 연구 주제들도 함께 간단하게 소개하였다.

3. 연구방법

가. 해외 사례 연구

해외 사례는 미국의 HRS (Health and Retirement Study), 영국의 ELSA (English Longitudinal Study of Ageing), 유럽연합의 SHARE (Survey of Health, Aging and Retirement in Europe)를 대상으로 하였다. 전반적인 내용은 선행 연구에서 실시하였기 때문에 본 보고서에는 총괄적인 내용을 요약하고 표본마모 및 표본편향과 관련된 내용을 중점으로 정리하였다.

1) HRS 사례 분석

가) 사례 요약

HRS는 미국에서 50세 이상 인구의 건강 및 재정 상태 등을 점검하기 위하여 1992년부터 실시하여 1996년부터 격년제로 실시하고 있는 조사이다. 실질적으로 HRS는 미국 대공황 이후 출생한 일부 연령층의 인구를 분석하기 위한 조사였으나 추후에 대상 연령군이 확대되었다. 조사 대상 통합에 대한 내용은 다음 <표 1>에 정리 하였다.

표본 설계 보고서 (Heeringa and Connor, 1995) 에 따르면 표본 추출은 4단계에 걸쳐서 실시된다. 제 1단계에서는 미국 Metropolitan Statistical Areas (MSA)와 non-MSA 지역의 county (또는 몇 개의 county를 합친 것)를 일차 표본 추출단위(PSU, primary sampling unit)로 하는 층화 PPS 추출 (probability proportional to size)을 사용하였다.

<표 1> HRS 패널 조사 자료 구조

	'92	'93	'94	'95	'96	'98	'00	'02	'04	'06	'08	'10
HRS	▷		▷		▷	▷	▷	▷	▷	▷	▷	▷
AHEAD		▷		▷		▷	▷	▷	▷	▷	▷	▷
CODA						▷	▷	▷	▷	▷	▷	▷
WB						▷	▷	▷	▷	▷	▷	▷
EBB									▷	▷	▷	▷
MBB												▷

- 주) 1. HRS: original Health & Retirement (1931~1941)
 2. AHEAD: Aging & Health Dynamics (<1923)
 3. CODA: Children of the Depression (1923~1930)
 4. WB: War Baby (1942~1947)
 5. EBB: Early Boomers (1948~1953)
 6. MBB: Mid Boomers (1954~1959)

제 2단계에서는 추출된 PSU 내에서 census block(일종의 조사구)을 추출단위(SSU: secondary sampling unit)로 하고 가구 수를 집락 크기로 하는 PPS 추출을 하는데 흑인(black)과 Hispanic 가구를 과다추출 하기 위하여 core sample 외에 추가적으로 Black supplement sample SSU와 Hispanic supplement sample SSU를 뽑는다. 이를 위해서 추출된 PSU 마다 별도의 조사구 frame을 구성하고(10% 이상의 black 가구를 가지는 조사구) 이것으로부터 조사구별 black 가구 수에 비례하는 PPS를 뽑아 black supplemental sample SSU를 추출한다.

이렇게 해서 표본 조사구들이 뽑히면 제 3단계 추출이 이루어지는데 각 조사구로부터 가구 명부를 얻어내어서 그 가구 명부에서 가구들을 계통 추출을 실시한다. 계통 추출을 하게 되는 표본 수는 예상 적격 가구 수를 계산하여 결정하게 된다. 표본 가구가 선정되면 마지막 단계에서 적격/응답 가구를 선정하게 되는데 조사원이 각 표본 가구를 방문하여 각 가구 내의 모든 가구원에 대한 이름

과 성별과 나이, 그리고 응답자와의 관계를 조사한다. 그 후 가구원들의 태어난 년도를 바탕으로 가구의 적격 여부를 판단한 후 가구 내에 적격 대상자가 한명인 경우 그 사람을 조사하고 여러 명인 경우 한명을 Kish 의 random assignment 방법을 사용하여 응답자를 결정한다. 이러한 표본 추출법은 현행 국민노후보장패널조사와 어느 정도 유사하지만 마지막 단계에서 적격 판정을 하는 부분이 더 명확하다. 특정 인종을 과대표집하고 응답자를 가구에서 한명만 선택하는 것이 국민노후보장패널조사와 상이한 부분이다.

나) 사례 특이점 요약

HRS의 경우는 조사 대상 군이 통합되는 형태로 되어 있다. 즉 6년 단위로 새로운 연령 군(Cohort)이 포함 되서 조사되기 때문에 표본 유입에 대한 관리와 대처가 용이한 편이다. 또한 추가 정보 소득, 연금 등과 연계하여 결측치에 대한 가상 값(imputed value)을 삽입한 데이터를 제공하고 있다. 데이터 작성 과정에서 다른 경로로 획득 가능한 정보를 반영하고 있다.

2) ELSA 사례 분석

가) 사례 요약

ELSA 조사는 기본적으로 영국보건조사(HSE: Health Survey for England)와 연계하여 표본(50세 이상이 가구 구성원으로 있는 가구)을 선택한다. 우편주소를 통하여 조사구를 추출 한 후에 HSE(단위:가구조사) 응답자 중에서 50세 이상의 구성원이 있는 가구를 모

집단으로 하여 대상 가구를 조사한다. ELSA 조사도 격년제로 시행 되는데 미국의 HRS의 경우와 마찬가지로 새로 유입되는 대상 연령 집단을 분리하여 조사를 진행한다.

나) 사례 특이점 요약

KReIS의 경우처럼 조사 과정에서 대상의 적격성 여부를 판단할 필요가 없고 주소 정보가 있기 때문에 조사 이전에 사전 동의를 구할 수도 있다는 것이 특징이다. 하지만 HSE 응답자만을 모집단으로 하기 때문에 응답하지 않은 가구에 대한 정보 손실은 여전히 발생한다. 이 내용은 향후 국민연금공단 전산 정보와 연동하는 표본 설계를 할 경우 벤치마킹 할 수 있는 부분으로 고려된다.

미국의 HRS 사례와 영국의 ELSA 사례처럼 Cohort를 추가하는 경우 대상이 겹치는 경우가 발생할 수 있다. 예를 들어, 특정 Cohort에서 조사된 대상의 배우자(조사 당시 50 세 이하)가 추후에 50세 이상이 되어 새로운 Cohort 편입되는 경우가 발생하기 이에 대한 처리 기준도 필요하다.

또한 현장 조사에서 조사 대상에게 건강보험/연금/세금/사망 등과 관련된 통계의 접근 허용을 물어 추가적인 자료와 연계하여 관리할 수 있도록 한 것이 특징이다. 만약 조사 대상자 사망의 정보가 열람된다면 KReIS의 조사모집단에 이를 반영하여 사망과 관련된 표본마모로 인한 오차를 줄일 수 있을 것이다.

3) SHARE 사례 분석

가) 사례 요약

SHARE는 유럽연합 전체 노령인구에 대한 조사 연구로 2005년과 2008년 두 번에 걸쳐서 시행한 상태이다. 현재는 20 개 국가의 조사 결과를 통합한 형태로 정리하여 발표하고 있다. 표본설계 방법은 다른 조사들과 마찬가지로 조사구가 1단계 혹은 2단계로 추출된 후에 표본을 뽑은 후에 적격성 여부를 판단하는 형태이나, 국가별로 기준이 상이하다. 이 때 가구 정보를 가지고 있는 상업회사의 데이터를 이용한다는 특징이 있다.

나) 사례 특이점 요약

SHARE의 경우는 국가별 통합이라는 한계와 국가 간 비교의 중단면 분석이 기본 목표로 인하여 표본편향에 관련하여 비교적 자유로운 편이다. 국가별 비교를 위하여 미리 특정 집단(Vignettes, www.compare-project.org)을 고정시켜 조사를 하는 방식을 일부 채택하고 있는 것이 특징이 있다. 또한 SHARE는 결측치(missing value)에 대하여 가상 값(imputed value)을 삽입한 데이터를 작성하고 있다.

4) HRS 및 ELSA 관련 연구 사례 요약

본 연구 용역 대상 내용 중에서 해외 사례 조사와 부합하는 것은 표본마모에 관한 내용이다. 따라서 표본마모와 관련된 연구 사

례를 중점적으로 조사하였다. 만약 표본마모 현상이 조사 변수들과 관련이 있다면 모집단 대표성에 심각한 문제를 가져올 수 있기 때문에 표본마모 현상이 실제로 관심 변수(Socioeconomic Status)와 관련(non-ignorable)이 있는지를 확인하는 것은 중요한 문제이다.

JAMES BANKS 외 2인(2010)의 연구 보고서는 HRS와 ELSA에서 발생한 표본마모가 조사 변수와 관련이 있는지를 분석하였다. 보고서에 의하면 HRS와 ELSA 두 조사에서 모두 유의한 표본마모가 존재하며 ELSA의 경우에 더 높다고 되어 있다. 또한 특정 연령 군과 관련하여 표본마모가 음의 상관관계가 있음을 파악했는데 이는 조사의 관심 변수에서 기인된 것으로 분석했다. 그리고 표본마모의 형태를 분류하고 그 중에서 모형화가 쉬운 질병 및 사망으로 인한 표본마모 내용을 probit 모델 등을 사용하여 분석했지만 크게 유의한 의미를 주지는 않는 것으로 나타났다. 그러나 보고서에서 사용된 대부분의 분석방법이 데이터 자체를 조사하고 정리하는 방법으로 통계 모형을 통한 검정 부분은 미흡한 것이 한계이다.

또한 Pierr-Carl Michaud 외 3인(2009)도 HRS를 대상으로 표본마모로 인한 효과분석을 실시하였는데 표본마모를 형태별로 구분하여 이들을 변수로 한 통계모형을 통하여 표본마모가 실제로 임의적으로 발생하였는지, 조사 변수와 관련되어 발생하였는지를 조사하였다. 하지만 표본마모로 인한 현황 파악이 주 분석내용이고 해결 방안에 대한 내용은 미흡했다.

Giuseppe De Luca(2009)의 보고서는 위 내용 중에서 표본마모에 영향을 주는 요인별로 정리했는데 내용 중 일부를 소개하면 다음과 같다.

- 응답 대상자의 나이에 밀접하게 연관되어 있다.
- 사회·인구학적 변수 (거주 지역, 주택 점유 형태, 소득) 등과 깊게 연관되어 있다.
- 자신이 상대적으로 가난하다고 생각하는 사람들의 마모율이 상대적으로 높다
- 표본마모는 조사원과 연관이 깊은데 조사원의 나이와 학력과 음의 상관관계가 있다.
- 조사과정의 강도와 음의 상관관계가 있다.

이러한 내용들은 국민노후보장패널조사의 결과와 유사하다. 하지만 대부분의 내용은 표본설계 보다는 다른 요인에 의하여 발생한 표본마모 형태들로 본 연구의 대상과는 거리가 있다.

공통적으로 대부분의 연구사례는 여러 가지 요인에 의하여 표본마모가 실제적으로 발생하고 있으며 모집단 대표성에 문제가 되고 있음을 지적했다. 하지만 대부분의 분석의 경우 현황 파악이 주된 내용이거나 혹은 표본설계 과정에서 발생하는 표본마모보다는 조사 과정 혹은 통계적으로 고려되기 힘든 외부 변수에 의하여 발생하는 내용들을 정리하고 있다. 그리고 그에 대한 대책으로 현행 국민노후보장패널 조사에서도 사용되는 가중치 조정 방법을 제시하고 있다. 즉 표본 설계 과정의 개선을 통한 대안연구는 부족함을 확인할 수 있다.

나. 패널 자료 분석 결과

조사 모집단은 50세 이상 가구원을 포함하는 가구 및 그 가구에 속한 50세 이상인 자와 그의 배우자이다. 모집단 가구와 표본 가구

의 지역별 분포는 <표 2>와 같다. 모집단의 지역별 인구 분포와 표본의 지역별 인구분포는 <표 3>과 같다. 여기서 모집단은 2005년 인구 주택 총 조사 자료로부터 얻어진 것이다. 이 자료에서 보면 대전광역시나 울산광역시에서는 표본이 과대표집되었고 전라북도나 전라남도 등은 표본이 과소표집되었음을 알 수 있다. 이는 표본 추출 방법이 50세 이상의 가구원 정보를 사용하지 않은 것 때문으로 볼 수 있다. 가중치를 적용한 자료를 보면 이러한 과대 표집이나 과소표집이 어느 정도 보정된 것으로 보인다. <표 2>와 <표 3>의 2차년도와 3차년도 결과는 해당 연도에 응답한 가구만을 가지고 분포를 계산하였다.

<표 2> 모집단과 표본자료의 지역별 분포 (가구)

지역	모집단 가구 수	1차년도		2차년도		3차년도	
		가구 수	가중치 적용	가구 수	가중치 적용	가구 수	가중치 적용
서울특별시 (%)	1,465,827 (19.86)	1,043 (20.41)	1,465,342 (19.86)	869 (18.93)	1,460,6 (19.64)	741 (17.20)	1,435,257 (19.22)
부산광역시 (%)	617,686 (8.37)	388 (7.59)	617,686 (8.37)	335 (7.30)	621,517 (8.36)	314 (7.29)	626,772 (8.39)
대구광역시 (%)	372,650 (5.05)	231 (4.52)	372,650 (5.05)	207 (4.51)	368,235 (4.95)	199 (4.62)	363,617 (4.87)
인천광역시 (%)	337,390 (4.57)	271 (5.30)	337,390 (4.57)	240 (5.23)	339,681 (4.57)	218 (5.06)	338,724 (4.54)
광주광역시 (%)	187,612 (2.54)	157 (3.07)	187,612 (2.54)	149 (3.25)	190,710 (2.56)	147 (3.41)	193,200 (2.59)
대전광역시 (%)	190,375 (2.58)	174 (3.41)	190,375 (2.58)	158 (3.44)	190,594 (2.56)	148 (3.44)	192,373 (2.58)
울산광역시 (%)	128,130 (1.74)	150 (2.94)	128,130 (1.74)	137 (2.98)	129,012 (1.74)	132 (3.06)	126,357 (1.69)
경기도 (%)	1,316,803 (17.84)	920 (18.00)	1,316,803 (17.85)	812 (17.69)	1,337,767 (17.99)	766 (17.78)	1,353,44 (18.12)
강원도 (%)	274,398 (3.72)	170 (3.33)	274,398 (3.72)	165 (3.59)	278,055 (3.74)	162 (3.76)	285,262 (3.82)
충청북도 (%)	248,489 (3.37)	189 (3.70)	248,489 (3.37)	178 (3.88)	254,261 (3.42)	169 (3.92)	253,371 (3.39)
충청남도 (%)	351,038 (4.76)	231 (4.52)	351,038 (4.76)	224 (4.88)	356,231 (4.79)	227 (5.27)	365,131 (4.89)
전라북도 (%)	351,162 (4.76)	212 (4.15)	351,162 (4.76)	196 (4.27)	354,257 (4.76)	195 (4.53)	372,571 (4.99)
전라남도 (%)	416,508 (5.64)	214 (4.19)	416,508 (5.64)	205 (4.47)	416,620 (5.60)	200 (4.64)	414,789 (5.55)
경상북도 (%)	515,934 (6.99)	323 (6.32)	515,934 (6.99)	311 (6.78)	529,297 (7.12)	297 (6.89)	537,983 (7.20)
경상남도 (%)	518,778 (7.03)	372 (7.28)	518,778 (7.03)	347 (7.56)	521,996 (7.02)	340 (7.89)	526,908 (7.06)
제주도 (%)	86,511 (1.17)	65 (1.27)	86,511 (1.17)	57 (1.24)	86,511 (1.16)	53 (1.23)	82,774 (1.11)
계	7,379,291	5,110	7,378,806	4,590	7,435,374	4,308	7,468,534

주 1) 각 년도 가중치를 적용함.

<표 3> 모집단과 표본 자료의 지역별 분포 (개인)

지역	모집단 개인 수	1차년도		2차년도		3차년도	
		개인 수	가중치 적용	개인 수	가중치 적용	개인 수	가중치 적용
서울특별시 (%)	2,485,197 (20.19)	1,804 (21.06)	2,434,731 (20.09)	1,496 (19.57)	2,473,159 (19.83)	1,264 (17.84)	2,446,149 (19.26)
부산광역시 (%)	1,024,248 (8.32)	614 (7.17)	1,007,249 (8.31)	525 (6.87)	1,039,182 (8.33)	489 (6.9)	1,077,152 (8.48)
대구광역시 (%)	620,918 (5.04)	382 (4.46)	608,180 (5.02)	346 (4.53)	618,371 (4.96)	336 (4.74)	629,662 (4.96)
인천광역시 (%)	560,892 (4.56)	450 (5.25)	554,713 (4.58)	400 (5.23)	579,215 (4.64)	366 (5.17)	581,577 (4.58)
광주광역시 (%)	312,217 (2.54)	266 (3.1)	312,217 (2.58)	250 (3.27)	322,013 (2.58)	248 (3.5)	332,480 (2.62)
대전광역시 (%)	319,335 (2.59)	285 (3.33)	317,084 (2.62)	255 (3.34)	320,647 (2.57)	246 (3.47)	335,430 (2.64)
울산광역시 (%)	214,763 (1.74)	249 (2.91)	213,508 (1.76)	221 (2.89)	217,538 (1.74)	209 (2.95)	216,565 (1.7)
경기도 (%)	2,214,292 (17.99)	1,541 (17.99)	2,180,970 (17.99)	1,350 (17.66)	2,253,988 (18.07)	1,251 (17.66)	2,341,280 (18.43)
강원도 (%)	454,893 (3.7)	280 (3.27)	452,870 (3.74)	271 (3.54)	472,404 (3.79)	264 (3.73)	485,788 (3.82)
충청북도 (%)	415,619 (3.38)	319 (3.72)	412,539 (3.4)	299 (3.91)	433,464 (3.48)	288 (4.06)	443,122 (3.49)
충청남도 (%)	592,155 (4.81)	423 (4.94)	578,145 (4.77)	401 (5.24)	605,762 (4.86)	390 (5.5)	625,772 (4.93)
전라북도 (%)	580,710 (4.71)	349 (4.07)	576,368 (4.76)	318 (4.16)	592,372 (4.75)	304 (4.29)	617,455 (4.86)
전라남도 (%)	679,202 (5.52)	369 (4.31)	676,918 (5.59)	347 (4.54)	678,756 (5.44)	328 (4.63)	688,635 (5.42)
경상북도 (%)	853,472 (6.93)	526 (6.14)	835,060 (6.89)	504 (6.59)	871,355 (6.99)	474 (6.69)	886,225 (6.98)
경상남도 (%)	843,397 (6.85)	607 (7.09)	821,223 (6.78)	573 (7.49)	857,022 (6.87)	546 (7.71)	861,354 (6.78)
제주도 (%)	138,837 (1.13)	103 (1.2)	138,157 (1.14)	90 (1.18)	137,855 (1.11)	82 (1.16)	133,143 (1.05)
계	12,310,147	8,567	12,119,931	7,608	12,473,102	7,022	12,701,790

주 1) 각 년도 가중치를 적용함.

<표 2>와 <표 3>은 표본 마모에 대해서도 정보를 제공하고 있는데 서울이나 부산과 같은 대도시에서 표본 마모가 발생한 것으로 보인다. 이는 대도시에 살던 사람이 노후에 귀향하는 경향이 있는 것과 관계가 있는 것으로 판단된다.

다음으로는 가구주의 성별 및 연령대별 분포를 살펴볼 수 있다. <표 4>에서는 가구 자료의 가구주 성별, 연령대별 분포를 나타내고 있고 <표 5>에서는 개인 자료의 성별, 연령대별 분포를 나타내고 있다. <표 4>와 <표 5>에서 볼 수 있듯이 연령대 분포의 모집단과 표본의 괴리는 매우 심각한 상태이다. 특히, 50대 가구주의 분포가 모집단에서는 40%에 달하는 반면 1차년도 표본에서는 28%, 3차년도 표본에서는 24%로, 이는 표본 마모와 함께 모집단 유입을 표본이 반영하지 못하기 때문에 생기는 현상으로 설명할 수 있을 것이다.

<표 4> 가구주 성별, 연령대별 분포 (가구)

가구주 성별 /연령 대	모집단 가구 수	1차년도		2차년도		3차년도	
		가구 수	가중치 적용	가구 수	가중치 적용	가구 수	가중치 적용
남	5,523,111	3,633	5,261,533	3,192	5,195,609	2,918	5,090,737
(%)	(74.85)	(71.10)	(71.31)	(69.54)	(69.88)	(67.73)	(68.16)
여	1,856,180	1,477	2,117,273	1,398	2,239,76	1,390	2,377,797
(%)	(25.15)	(28.90)	(28.69)	(30.46)	(30.12)	(32.27)	(31.84)
40세 미만	357,221	477	435,418	295	276,605	173	182,433
(%)	(4.84)	(9.33)	(5.90)	(6.43)	(3.72)	(4.02)	(2.44)
40대	478,817	423	451,350	373	643,076	356	604,000
(%)	(6.49)	(8.28)	(6.12)	(8.13)	(8.65)	(8.26)	(8.09)
50대	2,978,742	1,414	2,780,550	1,023	2,305,060	760	1,804,322
(%)	(40.37)	(27.67)	(37.68)	(22.29)	(31.00)	(17.64)	(24.16)
60대	2,111,271	1,646	2,199,386	1,532	2,358,960	1,440	2,581,565
(%)	(28.61)	(32.21)	(29.81)	(33.38)	(31.73)	(33.43)	(34.57)
70세 이상	1,453,240	1,150	1,512,102	1,367	1,851,673	1,579	2,296,21
(%)	(19.69)	(22.50)	(20.49)	(29.78)	(24.90)	(36.65)	(30.75)

주 1) 각 년도 가중치를 적용함.

<표 5> 성별, 연령대별 분포 (개인)

성별 /연령 대	모집단 개인 수	1차년도		2차년도		3차년도	
		개인 수	가중치 적용	개인 수	가중치 적용	개인 수	가중치 적용
남	5,192,155	3,508	5,090,872	3,094	5,221,284	2,825	5,328,112
(%)	(42.18)	(40.95)	(42.00)	(40.67)	(41.86)	(40.23)	(41.95)
여	7,117,992	5,059	7,029,059	4,514	7,251,818	4,197	7,373,676
(%)	(57.82)	(59.05)	(58.00)	(59.33)	(58.14)	(59.77)	(58.05)
40세 미만	14,678	9	8788.018	3	3048.931	1	1955.355
(%)	(0.12)	(0.11)	(0.07)	(0.04)	(0.02)	(0.01)	(0.02)
40대	992,957	415	910,818	195	488,421	112	285,633
(%)	(8.07)	(4.84)	(7.52)	(2.56)	(3.92)	(1.59)	(2.25)
50대	5,106,821	2,725	4,917,412	2,064	4,484,962	1,583	3,745,492
(%)	(41.48)	(31.81)	(40.57)	(27.13)	(35.96)	(22.54)	(29.49)
60대	3,551,533	3,126	3,658,401	2,854	4,132,771	2,584	4,521,932
(%)	(28.85)	(36.49)	(30.19)	(37.51)	(33.13)	(36.79)	(35.6)
70세 이상	2,644,158	2,292	2,624,511	2,492	3,363,900	2,743	4,146,774
(%)	(21.48)	(26.75)	(21.65)	(32.75)	(26.97)	(39.06)	(32.65)

주 1) 각 년도 가중치를 적용함.

다음으로는 거처 종류별, 점유 형태별, 가구원 수별 가구 분포를 살펴볼 수 있다. <표 5>에서는 이러한 분포를 연도별로 보여주고 있다. 단독주택과 다세대 주택은 조금씩 줄어들고 아파트가 조금씩 늘어나는 경향을 보여주고 있고 자가 거주 비율이 1차년도에 상당히 높는데 이는 1차년도에 응답 오차(거짓 응답)가 있었음을 시사

한다. 1인 가구가 점점 늘어나는 경향이 있는 것은 본 패널 자료의 특성상 배우자의 사망률이 높기 때문일 것으로 추측된다.

<표 6> 가구 거처 종류별, 점유 형태별, 가구원 수별 분포 (가구)

거처종류 /점유형태 /가구원수	모집단	1차년도		2차년도		3차년도	
		원자료	가중치 적용	원자료	가중치 적용	원자료	가중치 적용
단독주택 (%)	3,947,356 (53.49)	2,674 (52.34)	3,978,071 (53.91)	2,381 (51.86)	3,800,914 (51.12)	2,228 (51.72)	3,749,802 (50.21)
아파트 (%)	2,493,447 (33.79)	1,529 (29.93)	2,439,23 (33.06)	1,382 (30.10)	2,521,289 (33.91)	1,323 (30.71)	2,589,159 (34.67)
다세대 주택 (%)	749,174 (10.15)	703 (13.76)	771,231 (10.45)	606 (13.20)	785,660 (10.57)	555 (12.88)	805,855 (10.79)
기타 (%)	189,187 (2.56)	203 (3.97)	190,267 (2.58)	222 (4.84)	327,510 (4.40)	202 (4.69)	323,717 (4.33)
자가 (%)	5,317,305 (72.06)	3,742 (73.24)	5,319,007 (72.07)	3,308 (72.07)	5,243,529 (70.52)	3,343 (77.60)	5,550,140 (74.31)
전세 (%)	1,007,310 (13.65)	541 (10.59)	979,515 (13.27)	466 (10.15)	917,387 (12.34)	430 (9.98)	896,914 (12.01)
월세 (%)	868,968 (11.78)	478 (9.36)	862,149 (11.68)	406 (8.85)	834,615 (11.22)	387 (8.98)	828,284 (11.09)
무상 (%)	185,708 (2.52)	348 (6.79)	218,136 (2.96)	410 (8.93)	439,842 (5.92)	148 (3.44)	193,196 (2.59)
1인 가구 (%)	1,343,871 (18.21)	907 (17.75)	1,357,993 (18.40)	934 (20.35)	1,518,40 (20.42)	1,016 (23.58)	1,729,887 (23.16)
2인 가구 (%)	2,233,572 (30.27)	1,759 (34.42)	2,292,94 (31.07)	1,687 (36.75)	2,433,754 (32.73)	1,663 (38.60)	2,656,621 (35.57)
3인 가구 (%)	1,580,748 (21.42)	1,015 (19.86)	1,577,244 (21.38)	876 (19.08)	1,532,042 (20.60)	785 (18.22)	1,461,660 (19.57)
4인 이상 (%)	2,221,100 (30.1)	1,429 (27.96)	2,150,62 (29.15)	1,093 (23.81)	1,951,17 (26.24)	844 (19.59)	1,620,366 (21.70)

이러한 일차원적 분석 외에도 여러 가지 이차원적 분석도 실시하였으나 표본 추가 작업을 위해서 굳이 필요한 작업이 아니었으므로 본 보고서에는 분석 결과를 수록하지 않았다.

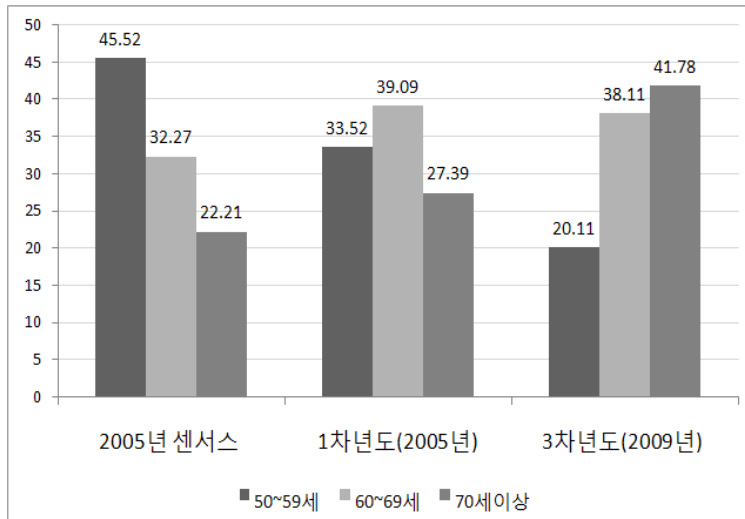
4. 연구결과

가. 표본 추가 방향

국민노후보장패널의 표본 추가 방향은 표본 설계 당시의 문제점을 극복하는 한편 표본 설계 당시의 목표 모집단이 지금의 목표 모집단과 달라진 것을 보완하고 표본의 마모를 보완하는 방향으로 진행되어야 한다. 표본 설계 당시의 대표적인 문제점으로는 50세 이상 가구원이 있는 가구 중 50대 가구주를 가지는 가구가 과소 추출되었다는 점이다.

[그림 1]에서 볼 수 있듯이 센서스 자료에 비해 50대 가구주 가구가 상당히 과소 추출되었고 이는 3차년도 자료에서 더욱 심화되었다. 이는 표본 설계 당시 50세 이상 가구주 또는 50세 이상의 가구원의 분포를 고려하지 않고, 표본유치 조사 시 조사의 편의에 의해 60, 70대의 고령자의 참여율이 50대의 참여율 보다 높았기 때문인 것으로 생각된다.

[그림 1] 만 50세 이상 가구주 연령 분포 비교 (단위 : %)



주 1) 가중치를 적용하지 않고 가구주의 연령이 만 50세 이상인 경우를 100으로 놓고, 사례 수의 백분율을 나타냄.

<표5>의 모집단 분포를 이용하면 50대, 60대, 70세 이상의 인구 비율은 각각 45%, 32%, 23%로써 <그림 1>의 모집단 가구주 연령대 분포와 크게 상이하지 않게 된다. 따라서 5,000 가구를 표본으로 뽑는다고 하였을 때 50대 가구주는 대략 2,250 가구가 뽑혔어야 한다. 하지만 실제로는 1,670 가구 정도가 뽑혔으므로 표본 마모를 고려하지 않더라도 일단 580 가구가 50대 가구주가 있는 가구 중에서 추가될 필요가 있다.

여기에 모집단의 변화와 표본 마모를 고려하면 50대의 표본 추가 가구 수는 더 늘어나야 할 것이다. 이를 좀 더 구체적으로 알아보기 위해서 5,000 가구를 최종 표본가구로 뽑는다고 하고, 모집단의 연령대별 분포가 2005년 당시와 비슷하다고 가정하면 <표 7>과

같은 목표 표본수를 계산할 수 있을 것이다. 여기서 50세 미만의 가구주는 고려하지 않았다.

<표 7> 가구주 연령대별 목표 표본 수와 추가 표본 수

가구주 연령대	최종 표본 (3차년도)	목표 표본수 (%)	추가 표본수
50대	845	2250	1405
60대	1,600	1,600	0
70세 이상	1,755	1,150	-605
계	4,000	5,000	

따라서 현실적으로 표본 추가를 실시할 경우에는 50대 가구주를 추가하는 것을 고려해야 할 것인데 70세 이상이 과대표집 되었지만 이를 인위적으로 버리지 않는다면 50대에서 대략 1,000 가구 정도를 추가하는 방향을 생각해 볼 수 있다.

실제로 50대 가구를 추출하기 위해서는 기존의 표본 설계 방법과 비슷한 과정을 따르게 된다. 즉, 2010년 센서스 자료를 구입하여 각 조사구 별로 연령대별 인구수 분포를 구한다. 이 중 50대 가구원이 있는 가구 수 (이를 유효가구수라고 하자)를 각 조사구 별로 계산한 후 이 유효가구수를 집락 크기 (measure of size; MOS)로 하는 층화 이단계 집락 추출을 실시한다. 여기서 층은 시도가 층이 되고 최적 표본 배정을 위해서는 각 시도별로 필요한 표본 유효 가구 수를 계산해야 한다.

이러한 추가 표본 유효 가구 수를 시도 별로 계산하기 위해서는 시도별 목표 표본수를 계산하고 기존 표본에서의 50대 가구 수를

계산하여 그 차이만큼이 추가 표본 유효 가구 수가 되는데 이를 정확히 계산하기 위해서는 2010년 센서스 자료를 분석하는 작업이 필요하게 된다. 따라서 이와 관련된 표본 배정 작업은 향후 과제로 남겨둔다.

여기에 향후 모집단의 변화를 고려하면 50대 가구만을 뽑을 것이 아니라 40대 가구도 추가적으로 뽑아서 향후 모집단 유입을 시의성 있게 반영하는 방법을 고려하여야 할 것이다. 이에 대해서는 다음의 4.2 절에서 자세하게 다루기로 한다.

나. 추정 방법 연구

여기서는 모집단의 특수성을 반영하는 추정 방법에 대해 다루고자 한다. 먼저 국민노후보장 패널조사는 조사모집단이 일반모집단과는 다른 특성을 가지고 있다. 따라서 표본 추출 당시 일반 모집단에서 표본을 추출한 후 적격 여부를 판단하여 적격으로 판정되는 가구만을 표본 가구로 포함하였다. 그러나 부적격 가구 (50세 미만 가구원으로만 구성된 가구)라 할지라도 이러한 가구들은 시간이 지나면 적격 가구가 되기 때문에 이러한 부적격 가구들을 표본으로 포함시킨 후에 조사를 하고 있지 않다가 적격가구가 되는 시점에서부터 조사를 실시하는 방법을 고려할 수도 있다.

즉, 다음과 같은 두 가지 시나리오를 생각해 볼 수 있다.

(시나리오 1) 일반 모집단 프레임 (센서스)에서 표본 조사구 추출
-> 표본 조사구내 모든 가구에 대한 적격 여부 판정 -> 적격 가구를 대상으로 최종 표본 채택 -> 최종 표본을 대상으로 조사 실시

(시나리오 2) 일반 모집단 프레임 (센서스)에서 표본 조사구 추출

-> 표본 조사구 내에서 표본 가구 추출 -> 표본 가구 중에서 적격 여부 판정 -> 표본 가구 중에서 적격 가구를 대상으로 조사 실시

지금까지의 조사 방법은 시나리오 1 이었다. 이러한 시나리오 1 은 두 가지 문제점을 가지고 있다. 첫째, 이단계 추출에서 일단계는 표본 조사구를 추출하고 이단계에서 최종 표본 가구를 추출하는데 이 최종 표본가구를 추출할 때 표본 조사구 내의 모든 가구에 대한 적격 여부를 판정하고 이러한 적격 가구 중에서 랜덤하게 일정 표본을 추출했어야 하는데 그러지 않고 자의적으로 조사자의 편의로 임의로 일정수의 적격가구를 모집하는 일종의 할당 추출이 되었다는 점이다. 따라서 제 이단계 추출 확률이 정확하게 계산되지 않았고 이는 추정의 편향을 가져오게 되는 결과 및 50대 가구 주 가구가 과소 표집되는 결과를 가져오게 되었다. 둘째 이러한 시나리오 1은 모집단의 변화 (또는 모집단 유입)를 시의성 있게 반영하지 못한다는 단점이 있다. 시나리오 1 하에서의 기본 가중치는 다음과 같이 계산된다.

$$\text{최종 가중치 (시나리오 1)} = (1/\text{조사구 추출 확률}) * (\text{조사구내 적격 가구 수} / \text{조사구내 표본 적격 가구 수})$$

여기서 조사구내 적격 가구 수는 조사구내의 모든 가구 수 중에서 적격 가구에 해당되는 가구 수를 의미한다. 실제로 이를 정확히 계산하는 것은 어려운 작업이 될 것이다.

시나리오 2는 시나리오 1과는 달리 표본 가구가 미리 결정되기

때문에 조사자가 자의적으로 할 수 있는 부분이 줄어들어 표본 추출의 객관성이 확보되고 모집단의 변화에도 신속하게 대응할 수 있다는 장점이 있다. 다만, 조사하지 않는 가구를 계속해서 관리해야 한다는 실사상의 어려움이 있을 수도 있다.

시나리오 2를 통해서 표본 가구를 추출하게 되면 추정은 적격 가구에 대해서만 얻어지므로 가중치가 다음과 같이 결정된다.

최종 가중치 (시나리오2) =
 $(1/\text{조사구 추출 확률}) * (\text{조사구내 표본 가구 수} / \text{조사구내 표본 적격 가구 수})$

이 최종 가중치 공식을 보면 조사구내 표본 가구 수와 조사구내 표본 적격 가구 수를 계산하는 것은 조사원의 자의적인 선택이 개입될 여지가 적으므로 보다 객관적인 추정이 가능하게 된다.

5. 결론

본 연구에서는 해외 사례 연구와 기존 패널 조사 자료의 분석을 바탕으로 지금까지 실시되어온 국민노후보장 패널조사의 대표성을 증진시키고 보다 객관적인 추정을 위한 표본 추가 방법을 제시하였다. 현재 표본에 과소 추출된 50대 가구 (가구주 연령=50대)를 보완하고 표본 추출 당시에 비해서 조사 대상이 바뀌었으므로 이를 반영하기 위해서는 가구주 연령이 50대인 가구를 코호트 표본으로 추출하여 보완하는 것이 필요한 실정이다. 본 연구에서는 50대 가구를 대략 1000 가구 정도 추가 추출하는 것이 적절한 것으로 판정하였다.

아울러 모집단 유입에 보다 많은 시의성을 반영하기 위해서는 여력이 되는 범위 안에서 40대 가구(가구주 연령=40대)를 함께 추출하여 표본으로 선정하되, 적격 가구로 판정된 후부터 실제 조사를 하는 방안도 함께 제안하였다. 이러한 추정 방법은 최종 표본 선정이 조사원이 개입될 여지를 최소화함으로써 조사구내 적격 표본 가구 선정 과정에서 발생할 수 있는 자의성을 줄여주고 표본 추출이 보다 객관적으로 진행될 수 있다는 장점이 있다. 이러한 경우 최종 표본에는 포함되지만 적격 가구로 선정되지 않는 조사 제외 가구에 대해서 어떻게 실사 관리를 하게 되는지는 실무진과 보다 깊이 있는 연구가 진행되어야 할 것이다.

또한 추가된 표본에 가장 적절한 가중치를 개발하는 것은 상당한 통계학적 지식과 이 패널 조사의 배경 지식이 요구된다. 이러한 가중치 작업에 대한 연구는 향후 과제로 남겨둔다.

<참 고>

1. 표본 마모 관련 연구

표본마모(attrition)가 존재하는 경우, 관심 모수를 추정하는 일반적인 방법은 크게 두 가지로 나뉜다. 첫 번째는 표본마모 현상이 조사 시점의 관심 변수에 의존하지 않는 경우(MAR, Missing at Random)를 가정한 모형이고 두 번째는 조사 시점의 관심 변수에 의존하는 경우(Non-ignorable)를 가정한 모형이다. MAR 가정의 경우에는 조사 대상자의 응답이 조사 시점의 관심 변수에 의존하지 않기 때문에 이전 조사 결과를 이용하여 모수 추정을 손쉽게 할 수 있다. 하지만 Non-ignorable한 경우에는 응답이 소득이나 건강 상태 등 조사 시점의 관심 변수에 따라서 달라지기 때문에 추가적인 가정 하에서만 추정이 가능하다. 하지만 위 방법 모두 기존의 패널 결과만으로는 어떤 방법이 모수를 추정하는데 더 정확한 방법인지는 판단할 수 없다. 두 방법의 적합성 여부는 관측되지 않은 조사 대상자의 응답에 의존하기 때문이다.

이러한 문제점을 효과적으로 해결할 수 있는 방법 중의 하나가 표본 추가이다. Hirano 외 3인(2001)은 표본추가가 존재하는 경우, 위의 두 가지 방법을 모두 포함하는 일반적인 모형 (AN, Additive Non-ignorable model)을 제안하였다. Hirano 외 3인(2001)이 제안한 모형은 다음과 같다.

$$P(\text{Response}) = f(k_0(x) + k_1(z_1, x) + k_2(z_2, x))$$

X : 시간에 영향을 받지 않는 변수 (성별 등)

Z : 시간에 영향을 받는 변수 (소득, 건강상태 등) $Z = (Z_1, Z_2)$

Z_1 : 이전 조사시점의 관심 변수

Z_2 : 해당 조사시점의 관심 변수

k : 해당 변수를 포함하는 임의의 함수

위 모형은 조사 대상자의 응답 확률이 인구학적 고정 변수, 과거 조사된 관심 변수와 현재 조사 대상의 관심 변수에 영향을 받는 것으로 되어 있다. 만약 k_2 값이 0에 가깝다면 조사 대상자의 응답이 해당 조사 시점의 관심 변수의 값에는 영향을 받지 않는 것으로 MAR 가정에 부합되고 k_1 이 0에 가깝고 k_2 값이 0이 아닌 의미 있는 값을 가진다면 Non-ignorable 한 가정과 일치한다. 즉 위 모형은 표본마모가 발생할 경우의 일반적인 두 가지 방법을 모두 포함한 일반적인 모형이다. 그리고 이 모형은 표본추가가 존재할 경우, 임의의 함수 k 에 대하여 유일한 응답확률 모형(Response model) 결과 값을 제공해준다. 즉 추가된 표본의 조사 결과가 손실된 정보를 복원해주는 역할을 하여 실제 관심 모수를 추정하는데 적합한 응답 확률 모형을 찾게 해주는 것이다.

다음은 총 조사가 2번 이루어진 패널 조사에 표본 추가가 존재하고 관심 모수가 평균 자가 주택 소유 현황인 경우에 해당하는 예이다. 관심 모수 EZ_2 은 다음의 평균 계산식으로 구할 수 있다.

$$\begin{aligned}
 EZ_2 = P(Z_2 = 1) &= \sum_{z,r} P(Z_2 = 1 | Z_1 = z, R = r, X) P(Z_1 = z, R = r, X) \\
 &= \sum_{z,r} p_{zr} q_{zr}
 \end{aligned}$$

(1)

$$p_{zr} = P(Z_2 = 1 | Z_1 = z, R = r, X)$$

$$q_{zr} = P(Z_1 = z, R = r, X)$$

X : 시간에 영향을 받지 않는 변수 (성별 등)

Z_1 : 첫 번째 조사에서의 자가 주택 여부 (1: 자가 주택, 0: 자가 주택 아님)

Z_2 : 두 번째 조사에서의 자가 주택 여부 (1: 자가 주택, 0: 자가 주택 아님)

R : 기존의 표본 대상자 중에서 두 번째 조사의 응답 결과
(1: 조사 응답, 0: 조사 비응답 (표본마모))

※ 계산의 편의를 위하여 X 는 영향을 주지 않는 것으로 가정하였다.

표본	Z_1	Z_2	R	관측 수
패널 표본	0	0	1	832
	0	1	1	66
	1	0	1	53
	1	1	1	80
표본 마모	0	-	0	518
	1	-	0	215
추가 표본	-	0	-	520
	-	1	-	136

그런데 위의 조사 결과에서 확인할 수 있듯이 표본마모가 발생하여 Z_2 의 값 일부분이 관측되지 않았다. 즉 위의 평균 계산식을 구하기 위해서는 $q_{00}, q_{01}, q_{10}, q_{11}, r_{00}, r_{01}, r_{10}, r_{11}$ 의 총 8가지 형태의 확률 값을 알아야 한다. 하지만 표본마모가 있을 경우에는 q_{00} 과 q_{10} 의 추정치를 구할 수 없다. 따라서 두 확률에 대한 추정치를 구하기 위해서는 추가적인 통계적인 가정이 필요한데, MAR 가정의 경우에는 $q_{00} = q_{01}, q_{10} = q_{11}$ 형태로 표현할 수 있다. 따라서 이 가정을 이용한 두 확률에 대한 추정치는

$$\hat{q}_{00} = \hat{q}_{01} = \frac{66}{832 + 66} = 0.074$$

$$\hat{q}_{10} = \hat{q}_{11} = \frac{80}{53 + 80} = 0.602$$

이 성립하여 최종적으로 $P(\hat{Z}_2 = 1)$ 을 구할 수 있다. 즉 조사 시점

의 관심변수에 영향을 받지 않고 이전의 조사 결과 추정값에만 영향을 받기 때문에 이전 조사 결과의 정보를 그대로 사용하면 된다.

응답확률이 조사 시점의 관심변수에 영향을 받는 Non-ignorable 한 경우에는 몇 가지 추가적인 조건이 필요한데 Hirano 외 3인(2001)에서는 Hausman 과 Wise(1979)이 제안한 제약 조건 q_{00}, q_{11}

$$q_{00} = \frac{r_{10}r_{01}(1-q_{01}) - r_{11}r_{00}(1-q_{11})}{r_{00}r_{11}q_{11}(1-q_{01})/q_{01} - r_{11}r_{00}(1-q_{11})}$$

$$q_{10} = \frac{q_{00}r_{00}q_{11}r_{11}}{q_{01}r_{01}r_{10}}$$

을 사용하여 \hat{q}_{00} 과 \hat{q}_{10} 을 계산하였다. 주어진 제약 조건하에서 결과 값을 계산하면 ,

$$\hat{r}_{00} = \frac{518}{832 + 66 + 53 + 80 + 518 + 215} = 0.294$$

$$\hat{r}_{01} = \frac{66 + 832}{832 + 66 + 53 + 80 + 518 + 215} = 0.509$$

$$\hat{r}_{10} = \frac{215}{832 + 66 + 53 + 80 + 518 + 215} = 0.122$$

$$\hat{r}_{11} = \frac{53 + 80}{832 + 66 + 53 + 80 + 518 + 215} = 0.075$$

$$\hat{q}_{01} = \frac{66}{832 + 66} = 0.074$$

$$\hat{q}_{11} = \frac{80}{80+53} = 0.602$$

이 되어 $\hat{q}_{00} = 0.306$, $\hat{q}_{10} = 0.894$ 가 된다.

또한 Hirano 외 3인(2001)이 제안한 AN 모형의 경우 추가된 표본을 이용하여 응답 확률을 로지스틱 모형

$$P(R=1|Z_1 = z_1, Z_2 = z_2) = \frac{\exp(a_0 + a_1z_1 + a_2z_2)}{1 + \exp(a_0 + a_1z_1 + a_2z_2)}$$

a_i $i=0,1,2$: 변수들의 계수 값을 가정하고 모수를 추정하였다. 응답확률이 얻어지면 몇 가지 수학적인 가정 하에서 Z_1 과 Z_2 의 결합 확률 분포가 구해진다. 이를 통하여 q_{00} 와 q_{11} 의 값을 도출할 수 있다.

$$P(Z_1 = z_1, Z_2 = z_2) = \frac{P(Z_1 = z_1, Z_2 = z_2|R=1)P(R=1)}{P(R=1|Z_1 = z_1, Z_2 = z_2)}$$

$$q_{00} = \frac{P(Z_1 = 0, Z_2 = 1) - P(Z_2 = 1|Z_1 = 0, R=1)P(Z_1 = 0, R=1)}{P(Z_1 = 0, R=0)}$$

$$q_{10} = \frac{P(Z_1 = 1, Z_2 = 1) - P(Z_2 = 1|Z_1 = 1, R=1)P(Z_1 = 1, R=1)}{P(Z_1 = 1, R=0)}$$

위의 응답확률 모형과 공식을 사용하면 $\hat{q}_{00} = 0.123$, $\hat{q}_{10} = 0.727$

의 값을 얻을 수 있다.

제시된 세 가지 방법에 따른 모수 추정 결과는 다음의 표에 요약되었다.

MAR	HW	AN
$P(\hat{Z}_2=1)=0.178$	$P(\hat{Z}_2=1)=0.282$	$P(\hat{Z}_2=1)=0.207$

추가된 표본에 따르면 추정 자가 주택 소유율은 $\frac{136}{136+520}=0.207$ 이다. 이를 대조군으로 비교하면 위의 패널 데이터에 관한 모수 추정 방법은 Non-ignorable한 가정보다는 MAR 가정에 적합하다는 것을 확인할 수 있다. Hirano 외 3인(2001)이 제안한 AN 모형은 두 방법과는 다르게 추가된 표본을 기존 조사 결과에 합쳐서 추정을 하였기 때문에 추가된 표본과의 비교에서 다른 두 방법보다 더 정확함을 알 수 있다.

표본마모가 존재하더라도 응답확률을 알 수 있다면 관심 모수의 추정 값을 계산할 수 있다. 하지만 응답확률은 관측되지 않은 결과에 영향을 받기 때문에 통계적인 가정 없이는 모수 추정이 어렵다. MAR 가정 혹은 Non-ignorable 가정을 사용하면 표본마모 하에서도 모수 추정이 가능하지만 사용된 가정이 적합한지에 대한 검토는 불가능하다. 하지만 표본추가를 통하여 이러한 문제점을 어느 정도는 해결할 수 있다. Hirano 외 3인(2001)이 제안한 AN 모형은 앞선 두 가지 모형을 모두 포함하는 일반적인 모형을 제공해 줄 뿐만 아니라 적합한 모형을 선택 검증하는데 이점이 있다.

2. 국민연금 가입자 표본의 리스트 연계 방향

국민노후보장패널 표본 자료 중에서 국민연금 가입자들은 국민연금공단의 가입자 리스트에서 몇 가지 항목(월 소득과 장애등급, 타 연금 가입 상태, 국민연금 가입 및 수급 여부, 가입기간, 수급기간 등)에 대해 자료를 얻어낼 수 있다. 따라서 가입자 리스트를 활용하여 노후 보장 패널 표본의 추정 값을 체크하고 표본의 대표성을 점검 및 보완하는데 사용할 수 있는 방법을 제안해 보고자 한다.

두 가지 자료(패널 샘플 자료, 국민연금 리스트)는 서로 다른 집단을 대상으로 하지만 두 집단에 공통으로 속해 있는 그룹인 가입자 집단(50세 이상 가구원을 가지는 가구에 속하면서 국민연금에 가입한 사람)에 대해서는 국민연금 리스트를 통해 정확한 참값을 얻어내고 패널 샘플 자료에서는 추정 값을 얻어낼 수 있으므로 그 두 값의 차이는 패널 샘플 자료 추정치의 오차가 된다. 이러한 오차는 해당 연도마다 별도로 구해 낼 수 있고 또 지역별, 연령대별, 성별로도 계산이 가능하다.

이러한 오차 값은 내부적으로 패널 샘플 자료의 신뢰성과 정확성을 판단하는 근거자료가 될 수 있고 또한 특정 표본 가구의 응답에 대한 신뢰성을 판단하고 이러한 신뢰성이 매우 떨어지는 경우에는 표본 교체의 근거자료로 사용될 수 있다. 더 나아가 항목 무응답이나 표본 마모의 경우에도 이러한 가입자 리스트를 이용하면 가입자 가구에 대해서는 보다 정확한 대체 값을 찾아낼 수 있을 것으로 판단된다.

<부록 2> 국민노후보장패널조사 무응답대체 적용방법에 관한 연구³⁶⁾

1. 서론

본 연구는 국민연금연구원에서 실시하는 국민노후보장패널 자료를 활용하여 무응답 대체를 적용하려는 연구자들에게 매뉴얼을 제공하는 것을 목적으로 한다.

설문조사 연구에서 결측값이 발생하는 경우는 매우 흔하다. 결측이 발생하는 원인은 설문 대상자가 응답을 회피하는 경우에서부터 문항 자체가 응답자에게 해당되지 않아 답을 할 수 없는 경우까지 그 이유가 다양하다. 특히, 국민노후보장패널조사의 경우처럼 장기간 동일한 연구참여자를 추적조사(follow-up study)하는 패널조사 연구에서는 시간이 흐름에 따라 연구에서 탈락(attrition)하여 결측이 발생하게 된다. 즉, 처음 조사에 참여한 연구자의 일부가 두 번째 조사나 그 이후 조사에서 여러 가지 이유로 연구 참여를 포기하기 때문에 시간이 지날수록 결측의 비율은 증가하는 경향을 보인다.

결측이 발생하는 많은 경우 결측값이 발생한 개체는 제외하고 완전히 관찰된 자료만을 이용하여 분석을 실시하게 되는데, 이는 자료의 손실로 인한 검정력의 약화와 추정치 및 추론에 편향(bias)이 발생할 수 있으므로 조심스럽게 접근해야 한다. 결측이 완전하

36) 본 연구는 고려대학교 의과대학 의학통계학과 안형진 교수에 의해 수행되었다.

게 임의로 발생하는 경우에도 결측으로 인한 표본의 대표성에 문제가 있을 수 있으므로 편향을 최소화하는 모수 추정방법들을 사용하여야 한다. 하지만 결측을 고려하여 보정하는 분석 방법은 매우 복잡하고 활용할 수 있는 통계분석 프로그램들이 제한적이어서 일반 연구자들이 손쉽게 사용할 수 없는 단점이 있다. 이는 결측값의 적절한 분석 필요성을 인지한 연구자들조차 결측값을 고려한 분석을 하는 것을 포기하게 하는 결과를 낳게 된다.

패널조사에서의 결측은 개체무응답(unit nonresponse)과 항목무응답(item nonresponse)으로 인하여 발생한다. 여기서 개체무응답은 연구참여자가 조사에 응답하지 않음으로써 발생하며, 이는 모든 설문 문항에 결측값이 발생한다. 항목무응답은 연구참여자가 조사에는 응하지만 일부 설문 문항을 답하지 않음으로써 발생하는 결측을 의미한다. 일반적으로 개체무응답은 가중치 보정(weight adjustment)을 통하여 무응답을 고려하며 항목무응답은 대체(imputation)방법을 통하여 무응답을 고려한다.

대체법은 결측값을 그럴듯한 값으로 대체하는 분석기법으로 결측값이 대체된 후에는 완전한 자료가 되므로 기존의 완전한 자료를 분석하는 통계 분석방법 및 프로그램을 이용할 수 있다는 장점이 있다. 특히 국민노후보장패널조사와 같은 대규모 연구조사는 자료를 국민연금연구원 내부에서 분석할 뿐만 아니라 관심 있는 일반 연구자에게도 제공하여 다양한 통계분석이 가능하게 할 수 있다. 이 경우 결측값이 대체된 자료를 일반 연구자에게 제공한다면 일반 연구자들이 좀 더 과학적인 분석을 시행할 수 있다. 그러므로 올바른 대체모형을 사용하여 결측을 대체하는 것은 매우 중요하다.

대체방법은 연구의 목적, 자료의 종류, 결측의 원인, 자료에 대한 정보 등 많은 요인에 따라 적절한 대체모형을 결정하고 모형을 통

해 결측값을 예측하고 대체하는 과정을 거치게 된다. 따라서 본 연구는 국민노후보장패널의 결측값을 대체할 수 있는 적절한 대체 모형 및 방법을 탐색하고, 그 매뉴얼을 작성하는 데 목적이 있다. 좀 더 구체적으로 다음과 같이 연구내용을 정리할 수 있다.

- 국내·외 대표 패널연구에서 사용되는 대체방법을 조사하고 정리한다.
- 패널자료에서 사용할 수 있는 대체방법들을 정리하고 소개하며 각 방법의 장단점을 제시한다.
- 국민노후보장패널에 적합한 대체방법을 제시하고 매뉴얼을 제시한다.

2. 연구내용 및 방법

가. 무응답 대체 사례 연구

국내·외 주요 패널조사를 소개하고 패널조사에서 사용되는 무응답 대체방법에 대하여 국내와 해외 사례로 나누어 간단히 살펴보도록 한다. 국내의 주요 패널조사에는 고령화연구패널조사, 사업체패널조사, 한국노동패널조사가 있으며, 추가로 국민노후보장패널조사 1차년도에 무응답 대체방법도 함께 살펴보도록 하겠다. 해외의 주요 패널조사에는 Survey of Income and Program Participation, European Community Household Panel, Health and Retirement Study, National Health Interview Survey, National Survey of Family Growth가 있다.

1) 국내패널

가) 고령화연구패널조사

고령화연구패널조사(Korean Longitudinal Study of Ageing: KLoSA)는 고령사회를 대비한 제도 개혁 및 정책 결정을 위해서 고령자의 노동시장 참여, 소득 및 자산 현황, 개인의 은퇴결정, 사회보장제도의 효과, 건강, 가족 내 소득 이전 등을 파악하기 위하여 2006년부터 매년 시행되고 있다. 조사대상은 제주도를 제외한 지역에 거주하는 45세 이상 중고령자로 표본수는 약 10,000명이다 (<http://www.kli.re.kr/klosa>).

고령자연구패널조사에서 사용된 대체방법은 다중대체(Multiple Imputation)와 수정된 예측 평균에 근거한 핫덱 방법(Hot deck based on a modified predictive mean matching)을 사용하여 인구, 건강, 고용, 소득, 자산, 가족의 순서로 대체하였다.

나) 사업체패널조사

사업체패널조사(Workplace Panel Survey: WPS)는 우리나라의 사업체를 대표하는 패널 구성원(2,000개)을 대상으로 전반적인 경영환경 및 인적자원관리체계, 조사관계의 현황 및 임금교섭과정 등에 관한 정보를 추적조사하는 종단면 조사(longitudinal survey)이다. 2006년부터 조사표본을 구성하여 동일한 표본을 대상으로 격년마다 조사가 이루어지며, 조사항목에는 근로자 고용현황, 재무정보, 노사관계 등이 포함되어 있다(<http://www.kli.re.kr/wps>).

사업체패널조사에서는 재무현황 관련 변수 및 무응답 변수에 대

하여 각 변수의 특성별로 평균 대체, 핫덱(Hot deck) 대체, 비율 추정(Ratio Estimation) 등의 대체 방법을 사용하였다.

다) 한국노동패널조사

한국노동패널조사(Korean Labor & Income Panel Study)는 비농촌지역에 거주하는 한국의 가구와 가구원을 대표하는 패널표본구성원(5,000 가구에 거주하는 가구원)을 대상으로 1년 1회 경제활동 및 노동시장이동, 소득활동 및 소비, 교육 및 직업훈련, 사회생활 등에 관하여 추적조사하는 종단면조사(longitudinal survey)이다(<http://www.kli.re.kr/klips>).

한국노동패널조사에서는 소득, 저축, 부채, 정규직 여부, 남부세액, 희망임금, 사업장 규모, 취업/퇴직시기, 공교육비 등의 변수를 대상으로 각 변수의 특성별로 중위수(median) 대체, 비율 조정(ratio adjustment), 핫덱(Hot deck), Stochastic EM generation(SEM)을 사용하여 무응답 대체를 하였다.

라) 국민노후보장패널조사

국민노후보장패널조사(Korean Retirement and Income Study: KReIS)는 우리나라 중·고령층의 노후준비 및 노후생활을 지속적으로 파악하여 노후소득보장과 관련된 정책을 위한 체계적인 자료를 구축하고자, 만 50세 이상 가구원을 가진 전국 5,000여 가구와 그 가구에 속하는 만 50세 이상 개인 8,600여명을 대상으로 가구의 경제상황, 중·고령자의 고용 현황 및 퇴직, 건강, 가족관계, 노후보장 현황 등에 관한 내용을 2005년부터 격년으로 조사하고 있는 종단

면 조사(longitudinal survey)이다(<http://www.nps.or.kr/jsppage>).

국민노후보장패널조사는 제1차 본조사(2005년)에서 가구소득 6개 변수, 가구지출 10개 변수 등 총 97개 변수를 대상으로 변수의 특성별로 최근방 대체, 랜덤 핫덱 대체, 이변량 핫덱 대체, 회귀(regression) 대체를 사용하였다.

2) 해외 패널

가) Survey of Income and Program Participation(SIPP)

Survey of Income and Program Participation(SIPP)은 소득의 원천과 금액, 노동관련 정보, 사회복지 프로그램 참여 정도와 수혜자 여부 및 기타 인구사회학적 특성들을 조사하여 현존하는 정부의 복지프로그램의 효과를 검증하고 미래의 대상자 규모와 소요비용을 추계하기 위한 조사이다. 특히 SIPP는 약 70개 항목의 현금소득 및 현물소득에 대한 다양한 정보를 제공하며 조세, 자산, 부채 및 정부의 소득 이전 프로그램의 수혜 여부와 급여수준 등에 대해서 조사하고 있다(U.S. Bureau of the Census 2001).

SIPP에서는 나이, 성별, 인종 등에 따라 층을 나누어 핫덱(Hot deck) 방법으로 대체하였으며, 덧셈 또는 비율적으로 변하는 값에 대해서는 다른 회차의 값을 이용하여 대체하였다.

나) European Community Household Panel(ECHP)

European Community Household Panel(ECHP)는 유럽 단위에서 1992년부터 매년 수행된 종단면 조사(longitudinal study)이다.

Statistical Office of the European Communities (Eurostat)에 의해 설계되었으며, 인구 통계, 노동력 행위, 소득, 보건, 교육과 훈련, 가구, 이주 등의 내용이 포함되어 있다(Eurostat 1996b).

ECHP는 가구소득 변수의 경우 비율 대체(ratio imputation)를 하였으며 개인소득의 경우 초기에는 예측 평균짜짓기(predictive mean matching)와 같은 층에서 랜덤 핫덱(random hot deck)방법을 사용하였다. 최근에는 Imputation and variance estimation(IVE)라 불리는 새로운 대체방법을 사용하였다.

다) Health and Retirement Study(HRS)

Health and Retirement Study(HRS)는 50세 이상의 미국인 22,000명 이상을 대상으로 격년으로 시행하는 종단면 조사(longitudinal survey)로 1992년부터 소득, 근로, 자산, 개인의 계획, 건강 보험, 장애, 신체적 건강과 기능, 인지 능력, 의료비 지출 등에 대한 정보를 수집해왔다(<http://hrsonline.isr.umich.edu>).

HRS는 혼합 방법(Mixed method)으로 핫덱(Hot deck)과 회귀(regression)를 혼합하였다. 구간을 나누어 하한 구간 또는 막힌 구간에는 핫덱(Hot deck)대체를 시행하고 상한 구간이나 범주형 전환 문장에 대해서는 점수(Score) 대체를 시행하였다.

라) National Health Interview Survey(NHIS)

National Health Interview Survey(NHIS)는 여러 목적을 가진 보건 조사이며 미국의 제도화되지 않은 가구모집단, 일반시민의 보건에 대해 조사한다(National Center for Health Statistics 2008). 이

조사는 1957년부터 넓은 범위의 보건에 대한 데이터 분석과 수집을 통해 미국 모집단의 보건을 감시하기 위해 시행해왔다. 조사내용은 기본적인 보건과 인구통계, 보험, 의료비 지출, 의료 행위 등이 있다(<http://www.cdc.gov/nchs/nhis>).

NHIS는 순차적 회귀(Sequential regression)를 이용한 다중대체(Multiple Imputation)방법을 사용하였으며, 최근 2007년에 가구소득과 개인소득에 대해 순차적 회귀를 이용한 다중대체를 소프트웨어 SUDAAN version 9.0과 IVE ware로 대체하였다.

마) National Survey of Family Growth(NSFG)

National Survey of Family Growth(NSFG)는 가족의 삶, 결혼 그리고 이혼, 임신, 불임, 피임의 사용과 남자와 여자의 건강에 대해 조사한다. 15세~44세의 미국의 남자와 여자를 대상으로 4년마다 48주간 진행된다. 2006년부터 시행되어 왔으며, 2010년에는 22,600명을 대상으로 조사하였다. 이 조사에서 제공하는 데이터는 미국의 조건부와 인적자원부 그리고 다른 보건서비스와 보건 교육프로그램에 이용된다(<http://www.cdc.gov/nchs/nsfg.htm>).

NSFG는 2007년에 순차적 회귀(Sequential regression)를 이용한 다중대체(Multiple imputation)를 사용하였으며, 소프트웨어는 IVE ware를 이용하였다.

나. 무응답 자료의 패턴

무응답의 발생 패턴은 매우 다양하게 나타나는데, 대표적인 몇 가지 패턴은 <그림 1>처럼 표현할 수 있다(Little and Rubin, 2002).

패턴은 응답과 무응답을 재정렬하여 나타낼 수 있으며, 각 패턴에 대한 설명은 다음과 같다.

<그림 1> 대표적인 무응답 발생 패턴

(1) 두 가지 패턴

		변 수					
		1	...	p	$p+1$...	$p+c$
관 측 값	1						
	2						
	3						
	⋮						
	m						
	$m+1$?	?	?
	⋮				⋮	⋮	⋮
	n				?	?	?

(2) 단조 패턴

		변 수					
		1	2	3	4	...	p
관 측 값	1						
	2						
	3						?
	4						?
	⋮						?
	$m+1$?
	⋮						?
	n						?

(3) 자료 짝짓기

		변 수								
		1	...	p	$p+1$...	$p+c$	$p+c+1$...	$p+2c$
관 측 값	1							?	...	?
	2							?	...	?
	3							?	...	?
	⋮							⋮		⋮
	m							?	...	?
	$m+1$?	...	?
	⋮							⋮		⋮
	n							?	...	?

(4) 일반적인 패턴

		변 수					
		1	2	3	4	...	p
관 측 값	1						
	2			?	?		
	3						?
	4		?				
	⋮						
	$m+1$?
	⋮						
	n						?

1) 두 가지 패턴(two pattern)

두 가지 패턴은 일부 변수에서만 무응답이 나타나는 경우의 결측패턴이다. 즉, m 개의 개체에서는 모든 변수에 대하여 관찰값이 존재하지만 나머지 $n-m$ 개의 개체에서는 처음 p 개의 변수에 대한 응답만 존재하고, 나머지 q 개의 변수에서는 결측이 발생한 경우이다. 이처럼 무응답 패턴은 전체 표본에 대하여 조사를 시행한 후 일부 표본(m 개)을 다시 추출하여 더 세부적인 조사를 시행하는 경우에 대표적으로 나타나는 무응답 발생 형태이다. 예를 들면, 통계청에서 인구주택총조사를 실시할 때 모든 사람들을 대상으로 전수조사(short form)를 실시하고 표본으로 뽑힌 사람들을 대상으로 심층조사(long form)를 통해 세부 정보를 추가 획득하는 경우에 이런 패턴이 나타난다. 이런 무응답 발생 패턴 중 가장 간단한 패턴은 변수 한 개에서만 무응답이 발생하는 일변량 결측 패턴(univariate missing pattern)이라고 한다.

2) 단조 패턴(monotone pattern)

처음 측정된 변수들에 대해서는 모든 관찰값에서 응답이 이루어지지만 두 번째 측정된 변수들에서부터 결측이 발생하는데, 한 번 결측이 발생한 관찰값은 이후 측정된 모든 변수에 대한 모든 관찰값이 무응답으로 남게 된다. 즉, 일단 무응답이 발생한 변수의 이후 변수에서는 모두 무응답으로 남게 되어 결측이 비율이 지속적으로 증가하는 경우로, 단조패턴이라고 부른다. 이 단조패턴은 패널조사(panel study)와 같은 경시적 연구 자료에서 흔하게 발생하는 패턴이다. 이는 처음 시점에서는 모든 개체에서 응답이 이루어

지지만 이후 시점에서의 몇몇 개체는 이어나 사망 등으로 인하여 연구에서 중도 탈락하여 결측이 발생하는 경우에 흔히 나타나는 패턴이다.

3) 자료 짝짓기(file matching)

자료 짝짓기는 모든 개체에서 기초 변수들에 대하여는 응답이 모두 존재하지만 나머지 변수들에 대하여 개체마다 응답하는 개체가 서로 다른 경우에 나타나는 패턴이다. 즉, 총 n 개의 개체 중에서 처음 m 개의 개체는 처음 $p+q$ 개의 변수들에 대한 응답값이 존재하는 반면 나머지 r 개의 변수에 대한 값은 결측으로 되어 있고, 뒤 $n-m$ 개의 관측값에서는 처음 p 개의 변수 및 마지막 r 개의 변수들에 대한 응답은 존재하지만 중간 q 개의 변수들의 값은 결측으로 나타난다. 더구나 처음 m 개의 개체와 뒤 $n-m$ 개의 개체는 처음 p 개의 공통 변수들에 대하여 응답값이 모두 존재하지만 나머지 $q+r$ 개의 변수들에서는 동시에 응답되지 않는다. 이와 같은 자료 형태는 인구통계학적 변수들(demographic variables)은 일치하지만 주요 관심 변수들이 각각 다른 두 자료를 병합(combine)할 때 주로 발생하는데 두 개의 다른 자료를 합하여 한 개의 자료로 합하는 과정에서 발생하는 무응답 패턴이라는 의미로 자료 짝짓기라 부른다.

4) 일반적인 패턴(general pattern)

무응답은 어떤 개체의 어떤 변수에서도 발생할 수 있으며 무응답의 비율도 변수별로 각각 다를 수 있다. 이 형태의 무응답은 어

떤 특별한 형태를 지니지 않고 가장 일반적으로 나타날 수 있다는 의미로 일반적인 패턴이라 부른다. 즉, 이 무응답 발생 패턴은 앞의 세 가지 무응답 발생 패턴이 특별한 형태의 무응답 발생 형태를 요구하는 데 반하여 전혀 제약을 가지지 않는다는 의미로 가장 일반적인 패턴이라 말할 수 있다. 이런 패턴은 관찰 연구에서 흔히 발생하게 된다.

무응답의 패턴에 따라 분석 기법이 달라지는 데, 즉, 두 가지 패턴이나 단조 패턴의 무응답 형태를 보이는 경우와 같이 특별한 형태의 무응답 형태를 요구하는 경우 무응답을 대체할 때 손쉬운 방법을 사용하여 편향을 제거할 수 있지만 무응답 발생 형태에 대한 제약이 적은 일반적인 패턴을 가지는 경우 더 복잡한 분석 기법이 요구된다. 또한, 자료 짝짓기 패턴을 가진 경우 동시에 측정되지 않는 변수들 사이의 연관성은 추정이 안 된다는 점을 유의하여야 한다.

다. 무응답 자료의 메커니즘

무응답 자료의 메커니즘은 결측이 발생한 원인을 밝히는 것으로 메커니즘에 따라 분석법이 다르기 때문에 이를 정확히 파악하는 것은 무응답 자료의 분석에 있어 매우 중요하다. Little and Rubin(2002)에서는 무응답 자료의 메커니즘을 다음과 같이 세 가지로 분류하였다.

1) 완전임의결측(Missing Completely At Random: MCAR)

자료행렬 Y 에서 무응답이 발생할 확률은 어떤 변수와도 상관없

이 완전히 무작위적이다. 먼저 응답지시 행렬 R 을 다음과 같이 정의한다. R 은 Y 와 동일하게 n 개의 행과 p 개의 열을 가지는 직사각형 행렬로서 $R = (r_{ij})$, $i = 1, \dots, n$, $j = 1, \dots, p$ 로 표현할 수 있는데, 여기서, i 는 n 개의 개체를 나타내고 j 는 p 개의 변수를 나타낸다. R 은 응답 및 무응답이 발생한 위치를 나타내므로 i 번째 개체의 j 번째 변수에서 응답값이 존재한다면 r_{ij} 가 1의 값을 가지고 무응답이 발생한다면 r_{ij} 가 0의 값을 가지도록 다음과 같이 표현한다.

$$r_{ij} = \begin{cases} i\text{번째 개체의 } j\text{번째 변수가 응답이면 } 1 \\ i\text{번째 개체의 } j\text{번째 변수가 무응답이면 } 0 \end{cases}$$

응답 지시행렬 R 의 자료행렬 Y 에 대한 조건부 분포(conditional distribution)를 $f(R|Y, \phi)$ 라고 한다. 여기서 ϕ 는 R 의 조건부 분포에 관련된 모수들(parameters)이다. 이 때, 완전임의결측은 모든 Y 와 ϕ 의 값에 대하여 다음과 같이 식으로 나타낼 수 있다.

$$f(R|Y, \phi) = f(R|\phi)$$

즉, 무응답의 발생은 자료 Y 와 상관없이 발생한다는 것을 의미한다. 여기서 $Y = (Y_{obs}, Y_{mis})$ 이다. Y 는 전체 자료이고 Y 는 응답된 자료(Y_{obs})와 무응답 자료 (Y_{mis})로 이루어져 있다. 완전임의결측은 무응답의 발생은 관측되지 않은 자료인 Y_{mis} 뿐 아니라 관측된 자료인 Y_{obs} 에도 의존하지 않는다. 이 완전임의결측은 매우 강한 가정으로 실제 자료에서 일어나는 결측은 완전임의결측인 경우가 거의 없다.

2) 임의결측(Missing At Random: MAR)

자료행렬 Y 에서 무응답이 발생할 확률은 자료의 응답된 자료에는 연관되지만 자료의 관측되지 않은 부분, 즉 결측과는 연관이 없다. 즉, 응답 지시행렬 R 의 자료행렬 Y 에 대한 조건부 분포 $f(R|Y, \phi)$ 는 모든 Y_{mis} 와 ϕ 의 값에 대하여

$$f(R|Y, \phi) = f(R|Y_{obs}, \phi)$$

로 표현할 수 있다. 예를 들어, 국민노후보장패널조사에서 가구별 수입 변수에서 무응답이 발생하였고 수입에 관한 무응답은 수입이 높은 가구에서 많이 발생하는 경향이 있다고 가정하자. 하지만 우리가 수입과 매우 높은 연관이 있는 가구별 세금 정보를 구할 수 있고 가구의 수입에 대한 응답 여부는 세금 액수에 따라 다르게 나타나지만 동일한 액수의 세금을 납부한 가구들 중에서 응답 여부가 완전히 임의로 결정된다면 분석에 세금 정보를 변수로 포함 시킴으로써 가구수입에 관한 무응답 자료 메커니즘을 임의결측을 따르도록 할 수 있다.

3) 비임의결측(Not Missing At Random: NMAR)

자료행렬 Y 에서 무응답이 발생할 확률이 응답 자료 Y_{obs} 뿐 아니라 결측자료인 Y_{mis} 와도 연관되어 있는 경우를 비임의 결측이라고 한다. 즉, 응답 지시행렬 R 의 자료행렬 Y 에 대한 조건부 분포 $f(R|Y, \phi)$ 는 아래와 같은 식으로 나타낼 수 있다.

$$f(R|Y, \phi) = f(R|Y_{obs}, Y_{mis}, \phi)$$

예를 들어, 국민노후보장패널조사에서 가구별 수입 변수에서 무응답이 발생하였고 수입에 관한 무응답은 수입이 높은 가구에서 많이 발생하는 경향이 있다고 가정하자. 우리가 수입과 매우 높은 연관성을 가지는 가구별 세금 정보를 구할 수 없고 대신 가구의 학력 정보를 가지고 있는 경우를 생각하자. 가구의 학력은 수입과 연관되지만 그 연관성이 매우 높지 않아 동일한 학력을 가진 가구들 중에서도 여전히 가구 수입이 높은 경우 무응답이 많이 발생한다면 수입 응답 여부는 여전히 무응답인 수입액 자체에 의존하게 되므로 이 때 가구별 수입액의 무응답 자료 메커니즘은 비임의결측을 따른다고 볼 수 있다.

대부분의 무응답 처리 기법은 무응답 자료 메커니즘이 임의결측이라 가정한다. 이는 완전임의결측의 가정은 너무 비현실적이고 비임의 결측의 가정은 분석법이 어려워지는 데 기인한다. 이런 임의결측의 메커니즘을 가정하기 위해서는 무응답 발생과 연관된 변수들을 될 수 있는 한 많이 포함시켜 분석을 시행해야 한다. 본 연구에서는 임의결측을 가정한 대체방법을 고려한다.

라. 무응답 대체 방법

결측자료 대체 방법은 이전에 언급한 바와 같이 연구의 목적, 자료의 종류, 결측의 원인, 자료에 대한 정보 등 많은 요인에 따라 가장 적절한 대체모형을 결정하여 한다. 조사연구 및 패널연구의 항목무응답을 고려하기 위하여 많이 사용하는 대체방법으로는 다변량 정규분포를 가정하거나 순차적 회귀모형을 이용하는 뚜렷한

모형(explicit model)을 이용하여 대체하는 방법과 핫덱(hot deck) 또는 콜드덱(cold deck)같이 내포적인 모형(implicit model)을 이용하는 방법 등이 있다. 본 연구에서는 이러한 대체 방법론을 소개하고 각 방법의 장단점을 평가한다.

1) 다변량 정규분포 가정하에서의 대체방법

변수들이 연속형이면서 정규분포를 따른다고 하면 다변량 정규분포를 가정하여 대체를 시행할 수 있다. 이 방법은 변수들이 연속형인 경우 흔하게 사용되는 모수적 대체방법으로 SAS PROC MI와 SPSS Missing Module과 같이 통계프로그램에 포함되어 있어 연구자가 쉽게 적용할 수 있다. 이 방법은 자료가 정규분포를 따르는 경우 가장 좋은 결과를 제공하며, 변수가 정규분포를 따르지 않는 경우에도 변수의 변환을 통해 정규 분포에 근사가 가능하거나 명목형 또는 이산형 변수에서 항목에 따른 빈도가 한쪽에 크게 치우치지 않으며 결측의 빈도가 높지 않은 경우 이 방법은 매우 강건하다(robust)고 알려져 있다. 또한, 이 방법은 선형회귀모형을 이용한 회귀대체 방법과 동일하다. 이 절에서는 일변량 패턴 결측에 관한 회귀대체에 관해 설명한다.

먼저 일변량 패턴 무응답 자료를 가정하자. 즉, Y_1, Y_2, \dots, Y_{K-1} 의 변수에는 무응답이 없고 오직 Y_K 변수에서 처음 r 개체가 관측되었고 다음 $(n-r)$ 개체에서는 무응답이 발생한 경우이다. 회귀 대체방법은 먼저 r 개의 완전히 관측된 자료를 이용하여 Y_K 를 종속 변수로 Y_1, Y_2, \dots, Y_{K-1} 를 독립변수로 하여 회귀분석을 하여 다음과 같은 회귀식을 적합한다.

$$\hat{y}_{iK} = \hat{\beta}_0 + \hat{\beta}_1 y_{i1} + \hat{\beta}_2 y_{i2} + \dots + \hat{\beta}_{K-1} y_{iK-1}$$

이 적합된 회귀식에 $(n-r)$ 개의 무응답 개체의 y_1, y_2, \dots, y_{K-1} 값을 이용하여 y_K 의 예측값을 구하고 그 예측값으로 결측값을 대체한다. 이 회귀대체법은 회귀식에 연속형 변수, 범주형 변수, 상호작용, 그리고 이차항 또는 스플라인과 같은 비선형 회귀식을 고려함으로써 예측력을 향상시킬 수 있다.

2) 순차적 회귀모형 (Sequential Regression)에 의한 대체방법

조사자료의 항목은 연속형 변수만으로 이루어진 경우가 거의 없고 많은 경우 명목형(nominal) 또는 계수형(counting) 변수들로 구성되어 있다. 따라서 다변량 정규분포의 가정보다는 연속형과 명목형, 계수형 변수 모두를 포함하는 모형을 고려하여 대체모형을 적합하여야 한다. 순차적 회귀 모형(sequential regression)을 이용한 대체방법(Raghunathan et al. 2001)은 이런 다양한 종류의 변수를 동시에 포함하는 모수적 모형을 설정하여 대체를 실시하는 방법이다. 이 방법은 또한 변수들이 제한된 범위 내에서 측정되거나 일부의 표본에서만 측정되는 경우도 분석 모형에 포함하는 것을 가능하게 한다. 이 방법은 SAS의 macro형태 또는 단독 프로그램으로 미시간 대학교 사회연구센터(institute of social research)가 제공하고 있어 실제 자료의 대체방법으로 많이 사용되고 있다 (<http://www.isr.umich.edu/src/smp/ive>). 이 모형은 미국의 National Health Interview Survey에서 주요 관심 변수인 소득의 결측값을 대체하는데 사용되는 등 다양한 분야에 적용되어 왔으며 그 유용성이 인정을 받고 있다. (Schenker et al. 2006).

이 방법을 좀 더 구체적으로 살펴보면 다음과 같다. Raghunathan, et. al. (2001)은 여러 가지 형태의 변수들을 가진 자료에 대한 대체 방법인 순차회귀 다중대체법(sequential regression multivariate imputation)을 제안하였다. 무응답을 포함한 자료를 자료행렬 Y 로 나타내고 n 개의 관측값에 대한 k 개의 설명변수들의 값을 행렬 X 로 표현하다. 여기서, Y 의 p 개의 변수 Y_1, Y_2, \dots, Y_p 는 각각 다른 변수 종류를 가질 수 있고 k 개의 설명변수들 또한 여러 가지 다른 종류의 변수 사용이 가능하다. 이 자료에 대한 모수적 모형에 근거한 대체를 실시하기 위하여 설명변수 X 의 값이 주어졌을 때 p 개의 변수 Y_1, Y_2, \dots, Y_p 들의 결합 조건부 분포(joint conditional density)는 다음과 같다.

$$f(Y_1, Y_2, \dots, Y_p | X, \theta_1, \theta_2, \dots, \theta_p)$$

이 때, 변수들의 종류가 다양하므로 이 결합 조건부 분포를 알 수 없기 때문에 표본을 결합 조건부 분포로부터 직접 추출할 수는 없다. 하지만 이 결합 조건부 분포는

$$\begin{aligned} & f(Y_1, Y_2, \dots, Y_p | X, \theta_1, \theta_2, \dots, \theta_p) \\ &= f(Y_1 | X, \theta_1) f(Y_2 | X, Y_1, \theta_2) \cdots f(Y_p | X, Y_1, \dots, Y_{p-1}, \theta_p) \end{aligned}$$

와 같이 여러 개의 조건부 분포(conditional density)들의 곱으로 표현될 수 있다. 여기서, $\theta_i, i = 1, \dots, p$,는 각 조건부 분포의 모수들의 벡터를 의미한다. p 개의 변수 Y_1, Y_2, \dots, Y_p 는 각각 다른 변수 타입을 가지므로 각 조건부 분포는 적절한 모형을 가지도록 선택한

다. 예를 들어 첫 번째 변수 Y_1 이 이산형 변수라면 로지스틱 회귀 분석 모형(logistic regression model)을, 두 번째 변수 Y_2 가 정규 분포를 따르는 연속형 변수라면 일반회귀모형(regression model)을, 세 번째 변수 Y_3 가 계수형 변수라면 포아송 회귀모형(Poisson regression model)을 가지고 적합할 수 있다. Raghunathan, et. al. (2001)은 위 식의 조건부 분포를 대신하여 $f(Y_j|X, Y_1^{(c)}, \dots, Y_{j-1}^{(c)}, Y_{j+1}^{(c)}, \dots, Y_p^{(c)}, \theta_p), j=1, \dots, p$,에서부터 모수인 $\theta_i, i=1, \dots, p$,를 마르코프 체인 몬테칼로 (Markov Chain Monte Carlo: MCMC)기법을 이용하여 값들을 추출하고 관찰된 자료 Y_{obs} 와 추출된 모수들 $\theta_i, i=1, \dots, p$,이 주어졌다는 가정 하에서 Y_{mis} 의 예측분포(predictive distribution)로부터 Y_{mis} 를 추출하는 과정을 반복 시행함으로써 대체를 실시하는 방법을 제안하였다.

이 방법에서 대체는 다음과 같은 순서로 이루어진다. 전체 C 번의 반복으로 이루어지는데 첫 번째 반복 $c=1$ 에 대하여 다음의 (1) - (J) 단계를 반복한다.

(1) $f(Y_1|X, \theta_1)$ 모형을 사용하여 모수 θ_1 을 추출하고 이 추출된 θ_1 과 설명변수행렬 X , 그리고 관찰된 Y_1 값이 주어졌다는 조건하에서 Y_1 의 무응답 값을 예측분포함수(또는 예측 사후분포함수)로부터 추출한다.

(2) 단계 (1)에서 추출된 Y_1 의 결측값을 Y_1 의 관찰된 값과 합하면 변수 Y_1 은 결측이 없도록 대체된다. 이 값을 $Y_1^{(1)}$ 이라 하고 이 값이 주어졌다고 가정한 후 $f(Y_2|X, Y_1^{(1)}, \theta_2)$ 에 대한 적절한 모형을 사용하여 모수 θ_2 를 추출하고 이 추출된 θ_2 과 설명변수행렬 X , 무응답이 대체된 Y_1 , 관찰된 Y_2 값이 주어졌다는 조건하에서 Y_2 의

결측을 예측분포함수(또는 예측 사후분포함수)로부터 추출된 값으로 대체한다.

⋮

(J) 단계 (1)부터 단계 (J-1)에서 대체된 $Y_1^{(1)}, \dots, Y_{p-1}^{(1)}$ 을 사용하여 $f(Y_p|X, Y_1^{(1)}, \dots, Y_{p-1}^{(1)}, \theta_p)$ 에 대한 적절한 모형(예: 로지스틱, 포아송 회귀모형)을 사용하여 모수 θ_p 를 추출하고 이 추출된 θ_p 와 설명변수 행렬 $X, Y_1^{(1)}, \dots, Y_{p-1}^{(1)}$, 관찰된 Y_p 값이 주어졌다는 조건하에서 Y_p 의 결측값을 예측분포함수(또는 예측 사후분포함수)로부터 추출된 값으로 대체한다.

두 번째 이후의 반복 $c=2, \dots, C$ 까지는 위의 (1)-(J) 단계에서 설명변수로 다른 모든 변수들을 수정하도록 아래와 같이 변경된다.

(1) $f(Y_1|X, Y_2^{(c-1)}, \dots, Y_p^{(c-1)}, \theta_1)$ 의 모형을 사용하여 모수 θ_1 을 추출하고 이 추출된 θ_1 과 설명변수 행렬 X , 이전 단계에서 대체된 $Y_2^{(c-1)}, \dots, Y_p^{(c-1)}$, 그리고 관찰된 Y_1 값이 주어졌다는 조건하에서 Y_1 의 결측값을 예측분포함수(또는 예측 사후분포함수)로부터 추출된 값으로 대체한다.

(2) 단계 (1)에서 추출된 Y_1 의 결측값을 Y_1 의 관찰된 값과 합하면 변수 Y_1 은 결측값이 없도록 대체된다. 이 값을 $Y_1^{(c)}$ 이라 가정한 후 $f(Y_2|X, Y_1^{(c)}, Y_3^{(c-1)}, \dots, Y_p^{(c-1)}, \theta_2)$ 에 대한 적절한 모형을 사용하여 모수 θ_2 를 추출하고 이 추출된 θ_2 와 설명변수 행렬 X , 대체된 $Y_1^{(c)}, Y_3^{(c-1)}, \dots, Y_p^{(c-1)}$, 관찰된 Y_2 값이 주어졌다는 조건하에서 Y_2 의 결측값을 예측분포함수(또는 예측 사후분포함수)로부터 추출된 값으로 대체한다.

(p) 단계 (1)부터 단계 (J-1)에서 대체된 $Y_1^{(c)}, \dots, Y_{p-1}^{(c)}$ 를 사용하여 $f(Y_p|X, Y_1^{(c)}, \dots, Y_{p-1}^{(c)}, \theta_p)$ 에 대한 적절한 모형을 사용하여 모수 θ_p 를 추출하고 이 추출된 θ_p 과 설명변수행렬 X , 대체된 $Y_1^{(c)}, \dots, Y_{p-1}^{(c)}$, 관찰된 Y_p 값이 주어졌다는 조건하에서 Y_p 의 결측값을 예측분포함수(또는 예측 사후분포함수)로부터 추출한 값으로 대체한다.

이 때 반복의 수 C 는 안정된 대체값을 얻을 수 있도록 결정한다. 이 방법은 추정된 함수

$\hat{f}(Y_j|X, Y_1^{(c)}, \dots, Y_{j-1}^{(c)}, Y_{j+1}^{(c)}, \dots, Y_p^{(c)}, \theta_p)$ 을 가지고

$$f(Y_j|X, Y_1^{(c)}, \dots, Y_{j-1}^{(c)}, Y_{j+1}^{(c)}, \dots, Y_p^{(c)}, \theta_p)$$

을 근사(approximate)하므로 SIR algorithm (Rubin, 1987b)이나 rejection algorithm(Gelman et. al., 2004)을 사용하여 근사를 개선하는 것이 바람직하다.

3) 핫덱(hot deck) 대체방법

핫덱 대체법은 결측값을 현재 응답된 다른 개체의 자료를 사용하여 대체하는 방법이다. 이 방법은 결측값이 관찰된 다른 응답자의 값으로 대체되기 때문에 자료의 형태가 관찰된 값과 동일하며 다른 분포 가정 하에서 적용되는 방법보다 현실적인 값들이 대체된다는 점에서 유용하다 할 수 있다. 결측값이 있는 개체의 관찰된

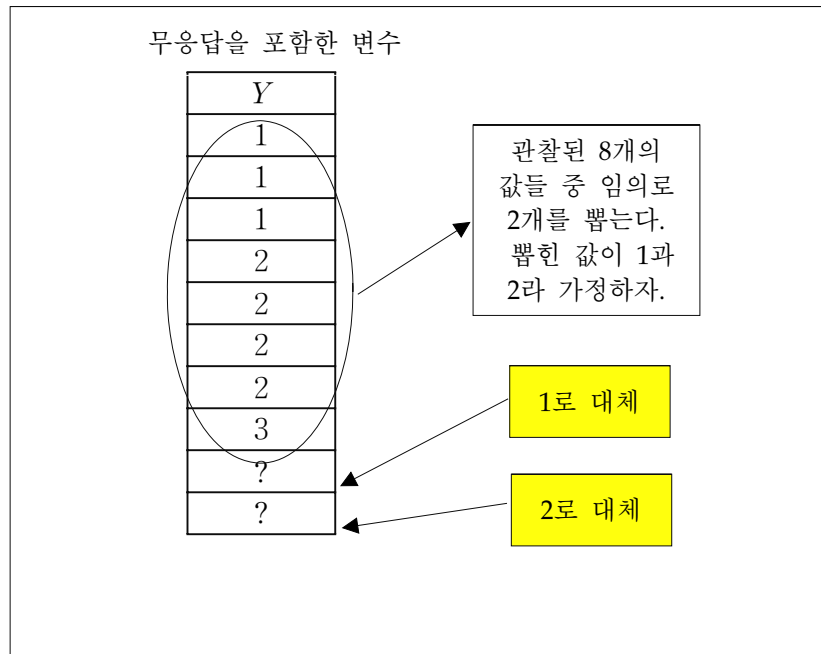
다른 변수들의 응답과 비슷한 응답을 한 다른 개체의 응답으로 결측값을 대체하는 방법으로 변수의 수의 많지 않은 경우 층화(stratification)를 통하여 같은 층에 속하는 연구 대상자 중 적절한 제공자(donor)를 선택하여 이 제공자의 응답값을 대체에 이용한다. 핫덱 대체에서 무응답값은 자료내의 응답된 값을 가지고 대체되므로 응답값이 무응답값에 기증되었다는 의미로 대체에 사용된 응답 개체를 기증자(donor)로, 무응답이 발생하여 응답값의 기증을 받은 개체를 수증자(donee)라 부른다. 핫덱은 기증자를 선택하는 방법이 매우 중요하며 여러 방법이 개발되었다. 여기서는 흔히 사용되는 몇 가지 핫덱 대체 방법들을 리뷰한다.

가) 단순임의 핫덱대체 방법(Hotdeck by Simple Random Sampling)

단순임의 핫덱 대체 방법은 자료내 각각의 무응답 값에 대하여 한 개의 응답값을 임의로 선택하여 대체하는 가장 간단한 형태의 방법이다. <그림 2>는 한 개의 변수에서 무응답이 발생할 때 단순임의 핫덱 대체를 시행하는 간단한 예를 설명한다. 무응답을 포함한 자료 Y 가 한 개의 변수를 포함하고 이 변수에 무응답이 발생하는 경우를 고려하자. 전체 관측값의 숫자는 10개인데 그 중 2개의 관측값에서 무응답이 발생한다면 응답된 8개의 자료가 2개의 무응답 개체를 대체하기 위하여 기증자로 사용된다. 8개의 기증자 중에서 임의로 2개의 응답값이 복원(with replacement) 또는 비복원(without replacement)으로 추출된다. 핫덱대체를 실제로 적용할 때에는 동일한 응답값이 여러 결측값에 대하여 기증자로 사용되는 것을 방지하기 위하여 복원추출보다는 비복원추출이 선호되는 경

향이 있다. 예를 들어, 8개의 응답값 중에서 2개를 추출할 때 추출된 값이 1과 2라면 첫 번째 무응답값에 1의 값을 두 번째 무응답값에 2를 대체한다.

<그림 2> 단순임의 핫덱대체 방법



자료 Y가 여러 개의 변수를 포함한 경우에도 이 방법을 확장할 수 있다. 각각의 변수별로 단순임의로 무응답값의 개수만큼 응답값으로부터 임의 추출한 후 그 변수의 무응답값에 대한 기증자로 사용하여 무응답을 대체한다. 이 방법은 대체에 관찰된 자료를 이용하지 않았으므로 결측 자료 메커니즘이 완전임의결측이 아니라면 모수의 추정치에 편향이 발생한다. 따라서 이 방법의 적용은 제한적이다.

나) 대체군을 이용한 핫덱대체 방법(Hotdeck Within Adjustment Cells)

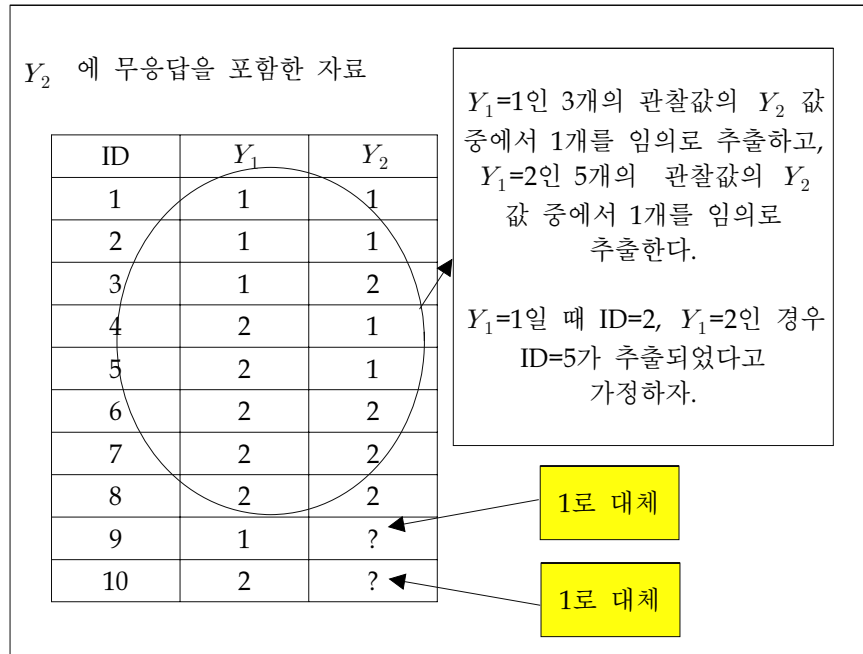
무응답 자료에 대한 대체를 실시할 때 관찰된 다른 변수들이 동일하거나 비슷한 응답값이 증거자로 선정된다면 완전임의로 응답값을 선정하여 대체를 실시하는 것보다 좀 더 나은 대체를 실시할 수 있다. 예를 들어 소득 변수에서 무응답이 발생한 경우 응답자 중 완전 임의로 증거자를 선정하는 것 보다 소득에 대하여 응답하지 않은 사람과 동일한 성별, 나이, 자산 수준을 가진 응답자들로 대체군을 형성하고, 이 대체군 내에서 임의로 한 응답자를 추출하여 이 응답자의 소득을 가지고 무응답자의 소득을 대체하는 것이 더 적절하다 할 수 있다.

<그림 3>은 대체군을 이용한 핫덱대체의 간단한 예이다. 자료가 관찰값 식별번호인 ID 변수와 두 개의 변수(Y_1 과 Y_2)를 포함하고 두 변수 중 Y_2 에서만 무응답이 발생할 때 완전히 응답된 Y_1 의 값을 가지고 대체군을 형성하고 이 대체군을 이용한 핫덱대체를 시행하는 방법을 설명한다. 무응답은 9번째와 10번째 관찰값(ID = 9와 10)에서 발생하였다. 즉, ID = 9는 $Y_1 = 1$ 의 값을, ID = 10은 $Y_1 = 2$ 을 가지는 것으로 관측되었으나 Y_2 변수의 값은 무응답으로 관측되지 않았다. 이 경우 ID = 9는 $Y_1 = 1$ 이므로 $Y_1 = 1$ 인 관찰단위 ID = 1, 2, 3 세 응답자의 값으로 대체군을 형성하고 이 대체군 안에서 임의로 한 개의 ID를 추출한다. 예를 들어 ID = 2가 추출되었다면 ID = 2의 Y_2 값인 1을 가지고 ID = 9의 Y_2 값을 대체한다. 마찬가지로 ID = 10은 $Y_1 = 2$ 이므로 $Y_1 = 2$ 인 관찰단위 ID = 4부터 ID = 8의 다섯 명의 응답자의 값으로 대체군을 형성하고 이 대체군 안에서 임의로 한 개의 ID를 추출한다. 예를 들어 ID = 5

가 추출되었다면 ID = 5의 Y_2 값인 1을 가지고 ID = 10의 Y_2 값을 대체한다.

대체군을 이용한 핫덱대체의 성능(performance)은 대체군을 어떻게 형성하느냐에 달려있다. 즉, 대체군 내에서 응답군과 무응답군의 분포가 동일하도록, 즉 대체군을 형성한 변수들이 주어졌을 때 무응답 발생 메커니즘이 무시할 수 있는 메커니즘이 되도록 대체군을 형성한다면 대체로 인한 편향(bias)이 발생하지 않는다. 즉, 대체군을 형성하기 위하여 무응답이 발생한 변수와 연관되어 있는 변수들을 포함하여야 한다. 이는 가능한 한 많은 관련 변수를 고려하여 대체군을 형성할수록 편향을 최소화할 수 있다. 하지만 대체군을 형성하기 위하여 변수가 추가될수록 대체군의 숫자가 기하급수적으로 늘어나게 되는 문제점을 가진다. 예를 들어 소득 변수에서 발생하는 무응답을 대체하기 위하여 대체군을 형성 하는 변수로 성별("남", "여" 구분)만을 사용하면 대체군의 수는 2개뿐이지만 연령 ("0-10세," "11-20세," "21-30세," "31-40세," "41-50세", "51-60세", "61-70세", "71세 이상"으로 구분)도 포함시키면 대체군의 숫자는 $2 \times 8 = 16$ 개로 늘어나며 거주지 구분("시," "도," "군") 및 자산정도("1천만원 미만," "1천 초과 ~ 1억," "1억 초과 ~ 5억," "5억 초과 ~ 10억," "10억 초과")를 추가하면 $2 \times 8 \times 3 \times 5 = 240$ 개로 크게 늘어난다.

<그림 3> 대체군을 이용한 핫덱대체 방법



한편, 대체군의 숫자가 늘어나면 일부 대체군에 속하는 응답값을 가진 개체의 수가 너무 적거나 일부 대체군에는 무응답은 있으나 관찰된 개체값이 없어 무응답에 대한 증거를 찾지 못하는 문제점이 발생할 수 있다. 또한 특정 대체군 내에서 응답자의 숫자가 무응답자의 숫자보다 적은 경우도 발생할 수 있으며 이 경우 이 대체군에 속하는 일부 무응답자는 증거가 없어 대체될 수 없다. 물론, 비복원추출 대신 복원추출을 사용하는 경우 이러한 문제는 덜 발생하지만 한 명의 응답자가 여러 무응답자에 대하여 대체되어 추정치의 표준오차가 부정확하게 추정되는 문제점이 발생할 수 있다.

대체군을 이용한 핫덱 대체 기법에서 이와 같이 대체군을 만드는 변수들이 많아져 기증자를 찾기 어려운 경우에 흔히 사용되는 방법으로 대체군을 형성하는 변수 일부를 생략하고 기증자를 찾는 방식이 있다. 예를 들어 “여,” “21-30세,” “도”에 거주하고 자산이 “10억 초과”인 경우 무응답이 발생하였는데 대체군에 속하는 응답이 하나도 없는 경우 연령을 대체군 형성에서 제외시키고 나머지 세 변수들로만 다시 대체군을 만든 다음 기증자를 찾는다. 이 단계에서도 기증자를 찾을 수 없다면 다시 거주지 구분 변수를 대체군 형성 변수에서 제외시키고 성별과 자산 정도만을 가지고 다시 대체군을 형성한 다음 기증자를 찾는다. 이 방법은 기증자를 찾을 때까지 대체군을 형성하는 변수의 숫자를 줄여가는 것이다.

대체군을 이용한 핫덱대체 방법의 문제점은 대체군을 형성하는 변수를 어떻게 설정해야 하는가와 기증자를 찾기 위하여 포기해야 하는 대체군 형성 변수의 순서를 결정하는 것이다. 우선 대체군을 형성하는 변수들은 무응답 발생 메커니즘이 임의결측 메커니즘이 되도록 만들어주는 변수들이 되어야 한다. Collins, Schafer, and Kam (2001)은 모의실험을 통하여 대체를 실시할 때 무응답이 발생한 변수와 밀접한 연관성을 가지는 변수들이 포함되어야 편향이 발생하지 않는다는 것을 보였다. 따라서 무응답이 발생한 변수와 연관성을 가지는 것으로 생각되는 변수들을 가능한 한 많이 포함하도록 대체군을 형성해야 편향을 최소화할 수 있다. 하지만 이와 같이 대체군을 형성한 경우 대체군의 숫자가 기하급수적으로 증가하여 일부 무응답값에 대한 기증자를 찾을 수 없어 대체군 형성 변수 일부를 포기해야 하는 경우에 발생한다. 연관성 정도가 약한 변수를 포기하는 방법을 고려할 수도 있지만 특정 변수 때문에 기증자를 구할 수 없는 경우도 종종 발생한다. 즉, 특정 변수에서 동

일한 항목에 속하는 응답자를 발견할 수 없다면 그 변수를 포기해야 기증자를 발견할 수 있다. 예를 들어 앞의 소득에 대한 대체에서 연령이 동일한 응답자를 발견할 수 없다면 다른 대체 형성 변수를 생략한다고 하여 기증자를 찾아낼 수 있지 않다는 것이다. 어느 변수를 먼저 포기해야 하는지에 대한 통일된 의견이 제시되기 어려운 이유이다. 따라서 대체군 형성은 자료에 대한 풍부한 정보와 경험에 근거하여 조심스럽게 선택되어야 하며 대체군 형성 변수의 포기 순서를 결정하는 것도 이와 마찬가지로 조심스럽게 결정되어야 한다.

다) 최근접 이웃 핫덱대체 방법(Nearest Neighbor Hotdeck)

위에서 다른 대체군을 이용한 핫덱대체를 실시할 때 대체군을 형성하는 변수는 범주형 변수이다. 연속형 변수가 대체군을 형성하는데 포함되기 위해서는 연속형 변수를 범주형 변수로 범주화(categorization)하여야 한다. 연속형 변수는 범주형 변수보다 더 정확한 정보를 포함하고 연속형 범주를 어떻게 범주화하느냐에 따라서 분석의 결과가 달라지는 등 민감한 문제를 포함하고 있다. 대체를 실시할 때 정확도를 높이기 위하여 고려하고자 하는 변수가 연속형이라면 최근접이웃 핫덱대체 방법을 실시할 수 있다. 이 방법은 대체를 실시할 때 고려하고자 하는 변수들의 값에 근거하여 관찰값 사이의 거리(distance metric)를 정의하고, 무응답값과 이 거리가 가장 가까운 응답자를 선택하여 이 응답자의 값을 대체에 사용하는 방법을 의미한다. 가장 간단한 예로 소득에 있어서 무응답이 발생한 경우 동일한 연령을 가진 응답자들 중에서 한 명을 임의로 선택하고 그 기증자의 소득을 가지고 대체를 실시하는 것이다. 물

론 여기서 동일한 연령은 정확한 연령이 될 수도 있고 무응답자와 비슷한 연령이 될 수도 있다.

대체에 고려하고자 하는 k 개의 변수가 있다고 가정하자. 이 변수들을 $X = (X_1, \dots, X_k)$ 라 하면 i 번째 관찰값의 X 값인 $x_i = (x_{i1}, \dots, x_{ik})'$, $i = 1, \dots, n$, 는 i 번째 응답자에 대한 k 개의 변수의 측정값들로 구성된 공변량 벡터를 표현한다. 이 k 개의 변수들을 사용하여 대체를 시행하기 위하여 i 번째 응답자와 j 번째 응답자 간의 거리는

$$d(i, j) = \begin{cases} i\text{번째 응답자와 } j\text{번째 응답자가 동일한 대체군에 속하면 } 0 \\ i\text{번째 응답자와 } j\text{번째 응답자가 동일한 대체군에 속하지 않으면 } 1 \end{cases}$$

와 같이 정의한다. 이 때 거리 $d(i, j)$ 는 여러 가지 방법으로 정해질 수 있는데 흔히 선택되는 방법들은 다음을 포함한다.

(가) 변수값들의 최대 편차(maximum deviation): i 번째 응답자와 j 번째 응답자의 각 변수별 측정값의 차이 중 최대 차이로서

$$d(i, j) = \max_k |x_{ik} - x_{jk}|$$

로 나타난다.

(나) 마할라노비스 거리(Mahalanobis distance): i 번째 응답자와 j 번째 응답자 사이의 마할라노비스 거리

$$d(i,j) = (x_i - x_j)' S_{xx}^{-1} (x_i - x_j)$$

를 사용한다. 여기서, S_{xx} 는 변수 X_1, \dots, X_k 들 간의 추정된 분산공분산행렬(the estimate of the covariance matrix)을 의미한다.

(다) 예측평균 (predictive mean): 무응답을 포함한 변수 Y 를 반응변수로, 변수 X_1, \dots, X_k 를 설명변수로 고려한 회귀분석에서 Y 의 예측값(predictive value)에 근거하여 i 번째 응답자와 j 번째 응답자 간 예측값의 차이의 제곱인

$$d(i,j) = [\hat{y}(x_i) - \hat{y}(x_j)]^2$$

을 사용한다. 여기서, $\hat{y}(x_i)$ 는 X_1, \dots, X_k 를 설명변수로 고려한 Y 의 회귀분석에서 i 번째 응답자의 Y 의 예측값(predictive value)을 의미한다.

i 번째 관찰값의 Y 변수의 값이 무응답이라면 응답자인 j 번째 관찰값들 중에서 $d(i,j) < d_0$ 인 응답값들 중에서 한 개의 관측값을 선택하여 그 개체의 Y 변수의 값을 가지고 대체한다. 여기서 d_0 을 크게 잡으면 잡을수록 더 많은 기증 가능한 응답자 대체군이 형성되게 된다.

최근접 이웃 핫덱대체 방법 중 처음 두 방법인 (1) 변수값들의 최대 편차를 이용하는 방법과 (2) 마할라노비스 거리를 사용하는 방법은 고려하는 변수들이 범주형인 경우 적용될 수 없다. 하지만 이 방법들은 이전에 설명한 대체군을 사용한 핫덱대체와 혼합하여 사용하는 것이 가능하다. 예를 들어 범주형 변수들을 사용하여 대

체군을 형성하고, 각 대체군 내에서 연속형 변수들을 공변량 벡터로 사용하여 거리 $d(i, j)$ 를 정의한 다음 이 거리에 근거하여 i 번째 무응답에 대하여 최근접이웃인 j 번째 응답자의 값을 가지고 대체를 실시할 수 있다. 이렇게 대체된 값은 무응답인 개체와 범주형 변수들에 대하여 동일한 값을 가지면서 연속형 변수들 간의 거리도 최근접인 응답값이므로 대체의 정확성을 추구할 수 있다. 한편, 예측평균을 이용한 핫덱대체 방법의 경우 범주형 변수를 지시변수(indicator variable)로 만들어 설명변수에 포함시키면 범주형 변수 및 연속형 변수를 모두 고려하여 대체를 실시할 수 있다.

핫덱대체는 명시적 형태의 모형(explicit model)을 정의하지 않고 대체를 실시한다는 의미로 내재적 모형(implicit model)하에서의 대체 방법이라 할 수 있다. 내재적 모형에는 가정도 명시적이지 않으므로 추정값의 편향을 수리적으로 계산하기 어렵다. 더구나 단순임의 핫덱대체보다는 대체군에 근거한 핫덱대체나 최근접이웃 핫덱대체, 또는 혼합 등 고려된 변수들의 복잡한 함수들(complex functions)에 근거하여 무응답에 대한 대체가 실시되므로 대체된 자료에 근거한 추정량의 성질을 평가하기 힘든 어려움이 존재한다. 따라서 핫덱대체의 성능(performance)은 대부분 비슷한 상황 하에서의 모의실험을 통해 평가되는데 상당수 모의실험에서 신중하게 선택된 방법에 근거한 핫덱대체는 정확성 높은 대체를 가능하게 하고 대체된 자료에 근거한 추정량에서 편향이 발생하지 않는다고 보고하고 있다.

4) 혼합적 모형에 근거한 대체 방법

위에서 살펴본 명시적 모형에 근거한 대체 방법들은 자료가 모

형의 가정을 만족하는 경우 우수한 성능을 보이는 것으로 나타나고 있다. 하지만 이 방법은 자료가 모형의 가정을 만족시키지 못하는 경우에도 그 결과가 그리 나쁘지 않다. 즉, 이 방법은 어느 정도의 모형이탈에도 강건하다(robust)라고 할 수 있다. 이는 대체가 전체 자료에 대하여 이루어지는 것이 아니라 무응답값에만 이루어지므로 무응답의 비율이 높지 않다면 잘못된 모형을 이용한 대체로 인한 편향이 그리 크지 않다. 하지만 모형의 분포 가정이 심하게 위배되는 경우 무응답의 비율이 높지 않아도 편향이 발생할 수 있다 (Tang et. al., 2005).

핫덱대체 방법은 일반적으로 자료에 대한 분포 가정을 명시적으로 하지 않으므로 여러 종류의 변수 대체에 유연하게 적용할 수 있으며 대체군을 잘 형성한다면 매우 적절한 대체를 할 수 있다. 하지만 대체군을 이용한 대체의 경우 대체군을 형성하는 변수들의 숫자가 늘어남에 따라 대체군의 숫자가 기하급수적으로 늘어나 기증자를 찾을 수 없는 무응답들이 생기게 되고 이 무응답들에 대하여 대체 변수의 숫자를 줄여 대체를 실시하면 무응답값들마다 대체 기준이 달라지는 문제점은 해결되어야 할 과제이다. 따라서 핫덱대체에서는 대체군 선정 및 조정 등 많은 노력이 필요하다. 이러한 문제점을 해결하기 위하여 핫덱 대체 방법들의 장점을 유지하면서 모수적 대체 모형을 동시에 고려하는 여러가지 모형들이 제시되어 왔다. 본 절에서는 핫덱 대체에 모수적 모형의 기법을 접목시킨 대체 방법과 모수적 모형의 예측력을 높이기 위한 비선형 회귀모형(nonlinear model)에 근거한 대체방법에 관하여 살펴본다.

가) 예측평균값(predictive mean value)에 근거한 핫덱대체 방법

무응답 대체를 위하여 무응답이 발생한 변수의 값과 비슷한 응답값을 갖는 개체들을 찾아내서 무응답을 대체하는 방법으로 Little(1988b)은 예측평균에 근거한 짝짓기 방법(predictive mean matching method)을 제안하였고 Bell(1999)은 수정된 예측평균에 근거한 대체 방법을 제안하였다. 이러한 방법의 주요 아이디어는 예측평균의 값(predictive mean)을 계산한 후 무응답값의 예측평균과 가까운 예측평균을 갖는 응답값들을 짝지어(match) 대체를 실시하는 것이다.

무응답을 포함한 변수 Y 와 이와 연관된 p 개의 공변량, X_1, X_2, \dots, X_p 가 존재한다고 가정하자. 회귀대체 방법에서처럼 무응답이 발생한 Y 변수를 종속변수로, 공변량 X_1, X_2, \dots, X_p 를 독립변수로 하여 회귀분석을 실시한 후 Y 변수의 예측값을 구한다. 이 예측값을 이용하여 자료를 몇 개의 대체군으로 범주화하고 각 대체군 내에서 무응답값을 같은 대체군 내의 응답자의 값으로 대체시키는 방법이다. 응답성향점수에 근거한 대체군의 숫자는 Rosenbaum and Rubin(1984)에서 제안한 바와 같이 전체 자료의 숫자에 따라 4개 또는 그 이상으로 결정될 수 있다. 이 방법은 예측평균을 구하기 위하여 모수적 모형을 적합시키지만 이 모형은 핫덱대체를 실시하기 위한 대체군의 형성에만 사용되므로 모형의 오지정(misspecification)에 영향을 덜 받는 장점을 지닌다.

이 방법은 공변량 X_1, X_2, \dots, X_p 에도 무응답이 발생하는 경우에도 사용이 가능하도록 확장이 가능하다. 즉, 공변량 X_1, X_2, \dots, X_p 에 무응답이 발생하는 경우 응답된 변수들만에 근거하여 회귀모형을 적합하고 모형의 예측값에 근거하여 대체군을 형성할 수 있다.

나) 비선형 회귀모형에 근거한 대체 방법

회귀대체는 일반적으로 선형회귀모형을 대체 모형으로 사용한다. 하지만 변수들이 선형적 관련을 갖지 않는 경우, 즉 모형이 잘못 지정되는 경우는 추정치의 편향이 심각해질 수 있다. 이러한 선형 모형(parametric linear model)을 가정하는 대체법의 단점을 보완하는 비선형 모형(nonlinear model) 하에서의 대체법을 고려할 수 있다. 무응답을 포함한 변수 Y 와 설명 변수 행렬 X 의 비선형적 관계(nonlinear relationship)는 X 의 이차식(quadratic equation)이나 삼차식(cubic equation) 등 다항식(polynomial equation)을 통해 회귀모형에서 고려할 수도 있지만 좀 더 강건한(robust) 방법으로 비모수적 회귀식을 고려할 수도 있다. 이 방법은 회귀식 $Y = g(X) + \epsilon$ 에서 함수 $g(\cdot)$ 를 스플라인(spline)이나 커널(kernel)을 이용하여 적합하는 방법이다(Cheng, 1994; Little and An, 2004). 이 방법 또한 만일 공변수의 개수가 늘어나면, 곧 공변수의 차원이 늘어나면 자료의 수가 무한대로 필요한 "차원의 저주(curse of dimensionality)"의 문제가 생기게 된다(Bellman, 1957). 이러한 다차원의 문제를 해결하기 위한 대체모형으로 일반가법모형(generalized additive model)을 대체모형으로 이용할 수 있다(Hastie and Tibshirani, 1990). 무응답을 포함한 변수 Y 에 대하여 p 개의 결측이 없는 공변수 X_1, X_2, \dots, X_p 가 있다면

$$Y = g_1(X_1) + g_2(X_2) + \dots + g_p(X_p) + \epsilon, \quad \epsilon \sim N(0, \sigma^2)$$

의 회귀식을 고려하고 함수 $g_i(\cdot)$, $i = 1, 2, \dots, p$ 를 스플라인(spline)이나 커널(kernel)을 이용하여 완전히 관측된 자료를 사용하

여 예측값을 추정하는 방법이다.

다차원의 문제를 해결하기 위한 다른 방법으로 Little and An(2004)은 응답성향(response propensity)을 이용한 차원의 축소를 제안하고 이를 벌점스플라인 대체(penalized spline imputation) 방법이라고 하였다. 자료에 결측이 없는 p 개의 변수, X_1, X_2, \dots, X_p 가 존재하고 1개의 변수 Y 에서 결측이 발생한다고 하면 결측자료 메커니즘이 임의결측일 때 응답성향은 $P(R=1|X_1, \dots, X_p) = X_1^*$ 로 정의되고 이 응답성향이 주어진 경우 Y_{obs} 와 Y_{mis} 의 분포는 같다. 즉, $f(Y_{obs}|X_1^*) = f(Y_{mis}|X_1^*) = f(Y|X_1^*)$ 이다. 그러므로 Y 를 X_1, X_2, \dots, X_p 로 회귀하는 대신 Y 를 X_1^* 로 회귀하여도 편향이 없는 추정치를 얻을 수 있게 된다. 즉 p 차원을 1차원으로 축소한 후 비모수적인 회귀식을 적합하는 방법이다. 이 방법은 강건한 모형을 바탕으로 대체값을 예측하기 때문에 변수 사이의 관계가 선형이 아니더라도 또는 변수의 차원이 높더라도 응답성향의 모형이 올바르게 선택되고 결측 메커니즘이 무시할 수 있는 메커니즘이라면 편향이 없는 추정치를 구할 수 있다.

마. 다중대체방법

지금까지 결측값을 한 개의 값으로 대체하는 방법에 관하여 설명하였다. 이렇듯 하나의 값으로 대체하는 방법을 단일대체(single imputation)라고 한다. 이 방법의 장점은 대체된 자료는 더 이상 무응답을 포함하고 있지 않으므로 연구자가 원하는 분석을 마음대로 시행할 수 있다는 점이다. 하지만 단일대체를 시행하면 대체로 인한 불확실성을 고려하지 않으므로 추정치의 표준편차가 과소추

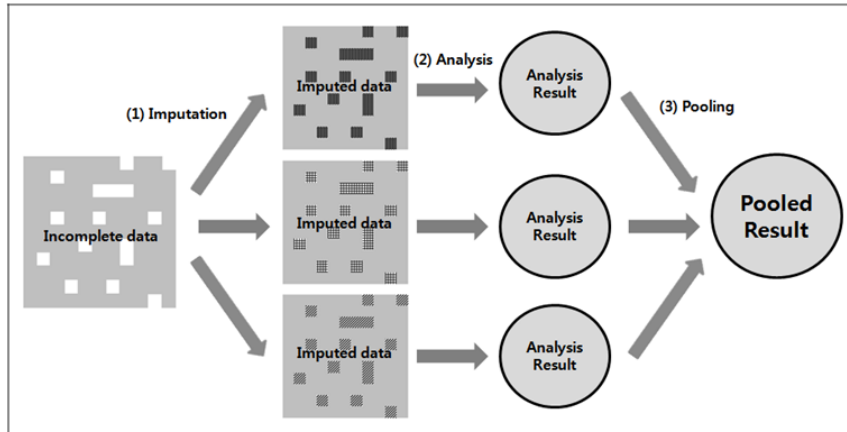
정되어 신뢰구간이 좁아지고 제1종의 오류율이 실제보다 작아지는 문제가 발생한다. 즉, 결측된 자료도 마치 실제로 관찰한 자료처럼 여기기 때문에 표준오차의 계산에 있어서 분모항이 실제보다 더 커지게 된다. 아주 적절한 대체모형을 사용하여 대체를 실시하더라도 대체된 값 모두가 무응답이 발생한 원래의 값과 동일할 가능성은 희박하기 때문에 이러한 불확실성을 표준오차 계산시 고려해 주어야 한다. 이는 추론을 실시할 때 기각하지 않아야 하는 모수를 기각으로 이끄는 문제점을 야기할 수 있다는 점에서 주의가 필요하다.

다중대체는 단일대체의 문제점을 해결하기 위하여 한 개의 무응답을 두 개 이상의 값을 가지고 대체를 실시하는 방법을 통틀어 칭한다. 다중대체에서는 한 개의 무응답을 그럴듯한(plausible) 여러 개의 값으로 대체하고 여러 개의 대체된 값들은 무응답이 발생한 값에 대한 불확실성(uncertainty)을 반영하게 된다. 즉, 무응답값을 정확하게 아는 것이 거의 불가능하므로 무응답값을 여러 개의 그럴듯한 값으로 표현함으로써 무응답값에 대한 우리의 불확실성을 모형에 포함시키는 통계적 기법이다. 다중대체를 실시하면 무응답 값들이 여러 개의 값들로 대체되고 이 값들간 차이에서 오는 분산이 추정량의 분산을 계산할 때 추가되어 분산이 과소추정되는 문제점을 방지하게 된다. Rubin(1987a)은 한 개의 무응답을 무한개의 값으로 대체한다면 추정량의 분산이 정확하게 추정되는 점을 이론적으로 보였다. 물론 한 개의 무응답을 무한개의 값으로 대체하는 것은 현실적으로 불가능하므로 실제로는 유한개의 대체를 실시하게 된다. 특히 무응답으로 인하여 손실된 모수에 대한 정보량(missing information)이 아주 크지 않다면 작은 숫자의 대체를 통해서도 모수의 분산이 거의 비슷하게 추정된다는 것을 보였다.

무응답을 포함한 자료에 대하여 다중대체를 실시하면 결과로서 단일대체와는 달리 한 개의 대체된 자료 대신 여러 개의 대체된 자료가 생성되게 된다. 즉, 무응답을 포함한 자료에 대하여 다중대체를 시행한 후 생성된 대체된 자료는 다음의 <그림 2.4>에 나타난 것과 같이 여러 개가 존재하게 된다. 이 여러 개의 대체된 자료들은 관찰된 값들은 모두 동일하지만 대체된 무응답값은 같기도 하고 다르기도 한 형태를 지닌다.

다중대체에서 대체의 숫자가 m 이면 연구자는 m 개의 대체된 자료 각각에 대하여 원하는 분석을 반복적으로 시행할 수 있다. 각 자료에 대하여 독립적으로 동일한 분석이 시행된 후 분석 결과는 유사하지만 동일하지 않은 m 개의 통계량 및 관련 분산(또는 표준오차)으로 나타나는데 연구자는 m 개의 각각 다른 통계량이 아닌 하나의 통합된 통계량을 구하거나 통합된 추론을 실시하는 데 목적이 있다. 이 목적은 (1) 다중대체된 각 자료의 분석, 그리고 (2) 분석된 자료를 통합한 결과 도출의 두 단계를 통하여 달성된다. 이 다중대체 방법은 아직 제한적이기는 하지만 SAS 9.2의 PROC MI와 PROC MIANALYZE 또는 SPSS 19.0에서 제공하고 있다. 아직은 변수가 모두 연속형인 경우 다변량 정규분포를 가정하여 다중대체를 하는 방법만 제공하고 있다. 일변량인 경우는 로지스틱 회귀 다중대체 방법을 제공하기도 한다.

<그림 4> 다중대체방법



1) 다중대체(multiple imputation)된 자료의 분석

다중대체된 자료 각각은 무응답이 대체되어 무응답이 없는 완전한 형태의 자료를 가지게 되므로 자료 각각에 대하여 연구목적에 알맞은 분석을 시행하면 된다. 예를 들어, 회귀분석을 시행하고자 한다면 관심있는 반응변수에 대하여 관심있는 설명변수를 포함하여 m 개의 대체된 자료 각각에 대하여 회귀분석을 실시하면 된다. 이렇게 m 번의 분석을 실시하면 추정된 회귀계수(regression coefficients), 표준오차(standard errors), 그리고 검정통계량(test statistics)은 m 개 자료 각각으로부터 약간씩 다르게 나타나는데 이는 관심 변수가 완전하게 측정되지 못하고 무응답을 포함하므로 무응답값에 대한 불확실성에 근거한 차이를 나타내는 것이다. 하지만 연구자의 분석 목적은 관심 자료에 대한 m 개의 서로 다른 결론이 아니라 한 개의 통합된 결론을 내리는 것이므로 m 개 분석의 결과를 통합하여 한 개의 결론을 도출하기 위하여 다음의 통합 과정

을 거쳐야 한다.

2) 다중대체 자료를 분석한 후 결과의 통합

다중대체를 m 번 시행한 후 대체된 자료 각각에 대하여 알맞은 분석을 시행한 후 얻어진 모수의 추정값들을 $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_m$ 이라 정의하고 이 모수들의 추정된 분산을 각각 W_1, W_2, \dots, W_m 이라 정의하자. 예를 들어, 회귀분석을 실시한 경우 i 번째 대체자료에 근거한 회귀 분석에서 관심 설명 변수의 회귀계수의 추정값 벡터는 $\hat{\theta}_i$ 이 되고 그 회귀계수 벡터의 표준오차의 추정값의 제곱이 W_i 행렬로 표현된다. 이 경우 통합된 모수의 추정값은

$$\bar{\theta} = \frac{1}{m} \sum_{i=1}^m \hat{\theta}_i$$

으로 표현될 수 있다. 즉, 모수 추정값들의 평균값이 통합된 모수의 추정값이 된다.

통합된 모수의 분산의 추정값은 다음의 두 개의 분산 성분의 합으로 표현된다. 첫 번째 분산 성분은

$$\overline{W}_m = \frac{1}{m} \sum_{i=1}^m W_i$$

로서 대체된 자료들로부터 추정된 m 개의 모수의 분산들의 평균이다. 이 분산 성분은 대체내분산(within-imputation variance)이라 부른다. 두 번째 분산 성분은

$$B_m = \frac{1}{m-1} \sum_{i=1}^m (\hat{\theta}_i - \bar{\theta})^2$$

으로 표현되는데 이는 m 개의 대체된 자료로부터의 m 개의 모수의 추정값들 사이의 분산을 나타내므로 대체간분산 (between-imputation variance)이라 부른다. 이 두 개의 분산 성분을 종합하여 계산한 모수의 분산에 대한 통합된 추정값은

$$T_m = \overline{W}_m + \frac{m+1}{m} B_m$$

으로 구할 수 있다.

관측값의 개수가 충분히 큰 경우, 일변량 모수(scalar parameter)에 대한 분포는 다음의 t -분포를 따른다.

$$(\theta - \bar{\theta}) T_m^{-\frac{1}{2}} \sim t_\nu$$

여기서, t -분포의 자유도 ν 는 $\nu = (\nu_0^{-1} + \widehat{\nu}_{obs}^{-1})^{-1}$ 로 계산되는데

이 중 ν_0^{-1} 는 $\nu_0 = (m-1) \left(1 + \frac{1}{m+1} \frac{\overline{W}_m}{B_m} \right)^2$ 으로 계산되고

$$\widehat{\nu}_{obs} = (1 - \widehat{\gamma}_m) \left(\frac{\nu_{com} + 1}{\nu_{com} + 3} \right) \nu_{com}$$

으로 나타나는데

이 때 ν_{com} 은 결측값이 없을 때 모수 θ 에 대한 추정의 자유도

(degree of freedom)를 나타내고 $\hat{\gamma}_m = \left(1 + \frac{1}{m}\right) \frac{B_m}{T_m}$ 으로서 무응답으로 인하여 손실된 모수 θ 에 대한 정보량(fraction of missing information about θ due to nonresponse)의 추정값이다.

바. 대체방법 프로그램 소개

1) SAS PROC MI 와 PROC MIANALYZE

간단한 형태의 PROC MI 프로그램

```
PROC MI DATA=temp2 SEED=85416 NIMPUTE=10 OUT=outmi
  MCMC CHAIN=multiple NBITER=2000 DISPLAYINIT
initial=em(itprint);
  VAR weight logtg hdl logfolate logpb loghg logcrp
RUN
```

위의 예제에서 사용된 PROC MI의 옵션을 살펴보면 다음과 같다.

- SEED: random number generator에 사용되는 양의 정수, 이 숫자를 같게 하면 항상 같은 결과를 얻을 수 있다.
- NIMPUTE=10: 10개의 impute된 데이터셋을 의미한다. 디폴트는 5이다.
- OUT: imputation 결과를 담은 SAS 데이터셋을 저장한다.
- MCMC: 대치를 위한 MCMC방법의 다양한 옵션들을 지정한다. MCMC 방법에 쓰이는 디폴트 prior는 non-informative

Jeffrey's prior이다.

- CHAIN = MULTIPLE: MCMC방법에서 다중 체인을 이용한다. 즉, 하나의 대치된 데이터셋은 MCMC iteration에서 얻어진다.
- NBITER=2000: 2000번의 초기(burn-in) 반복(iteration)후에 imputation을 시작하라는 옵션이다. 디폴트는 200이다.
- INITIAL=EM(ITPRINT): MCMC를 위한 초기 평균값과 공분산 EM방법을 통해 얻는다. 또한 "EM(Posterior Mode) Iteration History"결과를 얻을 수 있다.

이 외에도 다른 옵션들을 정할 수 있다. SAS 매뉴얼을 참고한다. PROC MI는 다중대체된 데이터셋을 만든다. 이 데이터셋을 이용하여 분석하고자하는 분석을 한 후 그 결과를 PROC MIANALYZE를 이용하여 하나로 합친다.

2) IVEware 프로그램

SAS내에서 IVEware를 사용하여 무응답 대체를 실시하는 매크로 모듈 IMPUTE의 사용법을 설명한다. 다음은 IVEware IMPUTE 모듈의 기본 명령문(syntax)을 나타낸다.

IVEware IMPUTE 모듈

```

%IMPUTE (NAME=filename, DIR=);
  DATAIN filename;
  DATAOUT filename;
  DEFAULT variable type;
  CATEGORICAL variables;
  RESTRICT variable(logical expression);
  BOUNDS variable(logical expression);
  ITERATIONS number;
  MULTIPLES number;
  SEED number;
RUN;

```

매크로 %IMPUTE에서 사용되는 키워드(keyword) 및 설명은 다음과 같다.

NAME=filename	setup file의 이름을 지정
DIR=directory	setup file과 output file이 저장되는 컴퓨터 내의 directory 지정

DATAIN은 대체될 무응답 자료를 지정한다. DATAOUT은 대체가 실시된 후 대체된 자료의 파일명을 지정한다. DEFAULT는 default로 생각될 변수의 타입을 지정하는데 일반적으로 가장 많은 변수 타입을 선택한다. 그 외에 CONTINUOUS, CATEGORICAL, COUNT, MIXED 문은 각각 연속형, 범주형, 가산형, 혼합형 변수들을 지정할 수 있다. RESTRICT 문에서는 제약을 둘 변수들 및 각 변수별 제약식을 지정한다. BOUNDS 문에서는 경계값을 가지

는 변수들 및 각 변수별 경계를 식으로 지정한다. ITERATIONS 문에서는 결측값을 추출하는 깃스샘플러의 반복의 수 C 를 지정한다. 2 이상의 값을 지정할 수 있다. MULTIPLES 문에서는 대체의 숫자를 지정하는데 default는 단일대체인 1이 된다. SEED 문은 난수생성 시 사용하는 0 또는 양수값을 지정하는 데 동일한 양수를 사용하면 동일한 결과를 얻을 수 있다. 0을 지정하면 예측값이나 회귀계수를 분포로부터 추출하는 대신 값 자체를 사용하고 default는 컴퓨터 시간에 의해 결정되는 임의의 값이 된다.

3. 고찰

본 연구에서는 국민연금연구원에서 실시하는 국민노후보장패널 자료를 활용하여 무응답 대체를 적용하려는 연구자들에게 매뉴얼을 제공하기 위한 여러 가지 대체 방법 및 프로그램에 관하여 살펴보았다. 대체는 무응답값이 채워져서 마치 자료가 실제로 관찰된 것처럼 보이기 때문에 기존의 무응답이 없는 자료의 분석방법들을 모두 사용할 수 있어 매우 매력적인 방법이다. 하지만 연구목적과 자료의 구조 등을 잘 반영하는 적절한 대체모형을 선택하는 것이 매우 중요하다. 만일 잘못된 모형을 이용하여 대체를 실시하게 되면 무응답을 제외하고 분석하는 완전히 응답한 자료만을 이용한 분석법의 결과보다도 더 심한 편향을 가져올 수 있다. 따라서 무응답의 대체는 조심스럽게 접근해야 하며, 여러 경우에도 민감도 분석 등을 통하여 이 방법이 강건한(robust) 결과를 보이는 지 확인해야 한다. 국민노후보장패널조사도 시간이 경과함에 따라 누적 무

응답의 비율이 높아질 것으로 횡단면 및 종단면 분석에서 편향을 최소화할 수 있는 대체 방법을 찾고, 이를 적용하여 자료 이용자들에게 과학적이고 신뢰할 수 있는 결과를 얻을 수 있도록 대체된 자료의 배포를 고려하여야 할 것이다.

기존 패널자료에서 사용한 무응답 대체 방법 등을 고려해 볼 때 국민노후보장패널도 순차적 다변량 회귀 모형을 적용한 IVEware를 이용한 다중 대체 방법이나 핫덱 다중 대체 방법을 적용하는 것이 적절할 것으로 생각된다.

참고문헌

- 김경자, 도시가계 생계비 선정기준의 다양화를 위한 연구, 한국가정관리학회지, 15(4), 137~148, 1997.
- 김대일·남재량·류근관, 한국노동패널 표본의 대표성과 패널조사 표본 이탈자의 특성연구, 한국노동경제학회, 23, 1~33, 2000.
- 김대회, 건강증진 정책의 방향 및 융통성 검토, 한국보건교육·건강증진학회, 16(2), 1~271, 1999.
- 김수완, 공적연금에서 파생적 수급권의 의미와 형태에 관한 연구: 배우자 급여와 유족연금을 중심으로, 사회복지연구, 20, 5~30, 2002.
- 김수완·조유미, 우리나라 노인가구 소득원의 실태분석, 국민연금연구원, 2005.
- 김연정, 은퇴노인가계와 취업노인가계의 소득, 지출 및 자산의 비교분석, 대한가정학회지, 36(7), 1998.
- 김영범, 한국 고령자의 경제활동참여 결정요인에 관한 연구, 사회복지연구, 24(3): 29~51, 2008.
- 김영원·류제복·박진우 역, Richard L. Scheaggar 저, 표본조사의 이해와 활용, 교우사, 2006.
- 김정석, 초고령 노부모들의 세대구성별 가구형태, 가족과 문화, 17(3), 3~18, 2005.
- 김재광, 표본조사론, 자유아카데미, 2008.
- 김지경, 성별 은퇴사유와 은퇴 후 소득 비교, 사회복지장연구, 21(2), 55-76, 2005.

- 김지경, 은퇴자의 은퇴사유 및 은퇴후 소득원천, KLIPS Research Brief, 3, 2004.
- 국민연금공단(내부자료), KReIS 가중치 연구 보고서, 2009.
- 국민연금공단(내부자료), KReIS 가중치 연구 보고서, 2008.
- 국민연금공단(내부자료), KReIS 연금·퇴직패널조사 표본설계 연구 보고서, 2008.
- 국민연금연구원, KReIS 유저가이드, 국민연금연구원 홈페이지, 2009.
- 남재량·이상호·최효미·신선옥·배기준, 제 10차 (2007)년도 한국 가구와 개인의 경제활동: 한국 노동패널 기초분석 보고서, 한국 노동연구원, 2009.
- 노경란, 중·고령자의 고용가능성과 고용촉진정책, 노동연구, 17, 5~42, 2009. “
- 박경숙, 중·장년기 종사상 지위와 은퇴 과정의 다양성, 노동경제논집, 24(1), 177~205, 2001.
- 박경숙, 55세 이상 고령자의 노동시장 이탈과정: 은퇴의 유형화에 대한 함의, 노동정책연구, 3(1), 103~140, 2003.
- 박능후·송미영, 노인가구 유형별 빈곤상태 변환에 대한 연구, 노인복지연구, 31, 7~26, 2006.
- 박태규, 한국의 최저생계비 추계에 관한 연구-「라이덴」방법을 중심으로, 한국재정학회 재정논집, 5(1), 297~387, 1991.
- 반정호·김경휘, 근로빈곤가구의 소비특성과 소비패턴 결정요인, 사회보장연구, 24(3), 1-28, 2008.
- 반정호, 가구유형에 따른 소득재분배와 빈곤완화 효과-고용·가족구성·세대효과를 중심으로-, 월간노동리뷰, 67~83, 2009.
- 방하남·신동균·이성균·한준·김지경·신인철, 베이비붐 세대의 경제적 부양부담, 한국 베이비붐 세대의 근로생애(Work life) 연

- 구, 한국노동연구원, 2010.
- 배성우·손지아·박순미, 빈곤노인가구의 특성과 빈곤탈피에 영향을 미치는 요인, 노인복지연구, 42, 291~318, 2008.
- 백은영, 은퇴 후 소득 적정성 및 지속성에 영향을 미치는 요인, 사회보장연구, 24(2), 27-51, 2008.
- _____, 여성노인가계의 적정 소득대체율 충족에 영향을 미치는 요인 분석, 여성경제연구, 4(2), 1~24, 2007.
- 백학영, 빈곤지위와 가구유형에 따른 노인가구의 소비특성 차이 분석, 한국노년학, 30(3), 911~931, 2010.
- 삼성경제연구소, 고령화 사회의 도래에 따른 기회와 위협, 삼성경제연구소 Issue Paper, 2002.
- 석재은·임정기, 여성노인과 남성노인의 소득수준 격차 및 소득원 차이와 결정요인, 한국노년학, 27(1), 1~22, 2007.
- 석재은, 연금의 성별 격차와 여성의 연금보장 방안, 보건사회연구, 24(10), 93~129, 2004.
- _____, 공적연금의 사각지대: 실태, 원인과 정책방안, 한국사회복지학, 53, 285~309, 2003.
- 석재은·김태완, 노인의 소득실태분석과 소득보장체계 개선방안연구, 한국보건사회연구원, 2000.
- 손병돈, 노인 소득의 불평등 추이와 불평등 요인분해, 한국노년학, 29(4), 1445~1461, 2009.
- 신동균, 중고령 남성 근로자들의 점진적 은퇴행위에 대한 연구, 노동정책연구, 9(2), 1-41, 2009.
- 심영, 독인가구의 경제생활에 대한 탐색적 연구, 한국가정관리학회지, 20(6), 197~208, 2002.
- 안중범·전승훈, 은퇴 결정과 은퇴 전·후 소비의 상호작용, 노동경제

- 논집, 27(3), 1~23, 2004.
- _____, 은퇴자가구의 적정소득대체율, 한국경제연구, 15, 5~33, 2005.
- 안중범, 인구고령화와 소비, 인구고령화와 거시경제, 한국개발연구원, 2004.
- 엄동욱, 중·고령자의 취업결정요인: KReIS 1차년도 자료를 중심으로, 노동정책연구, 8(3), 17~38, 2008.
- 여운경·정순희·문숙재, 한국가계의 은퇴준비에 관한 연구, 소비문화연구, 10(3), 29~155, 2005.
- 여운경, 노인부부가계와 노인독신가계의 소비패턴 비교, 한국가정관리학회지, 21(5), 1~12, 2003.
- _____, 목표소득대체율을 통한 은퇴소비의 추정, 대한가정학회지, 40(3), 83~97, 2002.
- _____, 한국 중산층 가계의 노후자금 적정성, 한국노년학, 25(1), 21~36, 2005.
- 우석진, 한국 고령노동자의 부분은퇴와 노동시장 복귀, 한국노년학, 30(2), 499-513, 2010.
- 윤재호·김현정, 은퇴와 가계소비간 관계 분석, 금융경제연구, 한국은행 금융경제연구원, 417.
- 원종욱, 국민연금의 적정소득대체율 분석, 보건복지포럼, 6(7), 32~42, 2000.
- 이미숙, 고령화연구패널자료를 이용한 노년기 건강불평등의 구조 분석, 한국보건사회학회, 보건과사회과학, 25, 5~32, 2009.
- 이민아·Margaret J. Weber, 우리나라 노인의 세대간 상호지원과 거주형태, 한국노년학, 20(3), 129~142, 2000.
- 이선형, 노인부부가계를 위한 최저 생계비 산정, 노인복지학회, 10(1),

- 61~88, 2000.
- 이선형·김근홍, 노인 가계지출구조 분석을 통한 집단 내 이질성에 대한 연구, 사회복지정책, 16, 43~60, 2003.
- 이성용·방하남, 중·고령 노동자의 퇴직 과정에 관한 일 연구, 사회보장연구, 25(3): 59~82, 2009.
- 이소정, 노인가구의 소비불평등 분석, 사회복지연구, 40(1), 235~260, 2009a.
- _____, Lifestyles of Korean Older Adults, 사회복지연구, 40(3), 327~348, 2009b.
- _____, 저소득 가구의 소비패턴과 경제적 복지의 안정성, 사회보장연구, 25(3), 317~336, 2009c.
- 이승렬·조수기, 중·고령자의 건강과 노동공급, 한국노동연구원, 2008.
- 이윤정·김순미, 남녀 노인독신가구의 경제상태와 소비지출의 영향요인 분석, 대한가정학회지, 42(2), 93~106, 2004.
- 이희숙·신상미, 은퇴노인가계의 경제구조 분석, 한국가정관리학회지, 21(4), 2003.
- 장지연, 중·고령자의 경력이동: 대안적 은퇴과정으로서의 근로시간단축과 자영업창업의 비중, 한국사회학, 37(2): 95~121, 2003.
- 정영숙, 노인가계의 소비패턴과 복지정책적 함의, 소비자학연구, 11(1), 59~74, 2000.
- 조추용, 노인이 가족에게 기대하는 부양 기대감에 영향을 미치는 요인에 관한 연구, 노인복지연구, 24, 71~99, 2004.
- 전승훈·강성호·임병인, 은퇴 후 필요소득수준과 국민연금 및 퇴직연금의 자산충분성, 경제학연구, 57(3), 34~67, 2009.
- 최종후·서두성, 데이터마이닝 의사결정나무의 응용, 통계분석연구,

- 4(1), 61~83, 1999.
- 최현수 · 류연구, 우리나라의 노인빈곤 동향 및 빈곤구성에 대한 연구, 한국노년학, 23(3), 143~160, 2003.
- 최현자 · 주소현 · 김민정 · 김정현, 한국의 은퇴준비지수산정에 관한 연구, 소비자학연구, 20(3), 189~214, 2009.
- 최효미, 노인가구의 소득원천, 노동리뷰, 69~79, 2007.
- 한수진 · 서지수, 노인가계 유형별 경제구조 분석, 한국가정관리학회지, 26(2), 2008.
- 홍백의 · 김혜연, 중 · 고령자의 퇴직 과정 유형과 그 결정요인에 관한 연구, 제1회 고령화연구패널 학술대회 자료집, 98~117, 2010.
- 홍백의 · 신유미 · 공주영, 은퇴결정에 대한 배우자의 영향 연구, 사회복지정책, 36(2), 503~522, 2009.
- Abdel-Ghany, Mohamed & Deanna L. Sharpe, Consumption patterns among the young-old and old-old, The Journal of Consumer Affairs 31(1), 90-112, 1997.
- Ando, A, and Modigliani, F, The Life Cycle Hypothesis of Saving: Aggregate Implication and Tests, American Economic Review, 53, 55-84, 1963.
- Banks, James, Richard Blundell, and Sarah Tanner, Is There a Retirement-Savings Puzzle? American Economic Review, 88(4), 769-88, 1998.
- Bernheim, B. Douglas, Jonathan Skinner, and Steven Weinberg, What Accounts for the Variation in Retirement Wealth Among U.S. Households? American Economic Review, 91(4) , 832-57, 2001.
- Bell R, Depression PORT Methods Workshop (I). RAND: Santa

- monica, CA, 1999.
- Bellman, R, *Dynamic Programming*, Princeton University Press, 1957.
- Burtless, G. and R. Moffitt, The Effect of Social Security Benefits on the Labor Supply of the Aged, In H. Aaron and G. Burtless(eds.), *Retirement and Economic Behavior*, 135-174, 1985.
- Cahill, Kevin E., Giandrea, Michael D., and Quinn, Joseph F. Retirement Patterns from Career Employment, *The Gerontologist*. 46(4), 514-523, 2006.
- Cheng, P. E, Nonparametric estimation of mean functionals with data missing at random, *Journal of the American Statistical Association*, 89, 81-87, 1994.
- Collins, L. M., Schafer, J. L., Kam, C. M, A comparison of inclusive and restrictive strategies in modern missing-data procedures, *Psychological Methods*, 6:330 - .351, 2001.
- Danziger, Aheldon, Jacques Van Der Gaag, Eugene Smolensky, and Michael K. Taussig, The life cycle hypothesis and the consumption behavior of the elderly, *Journal of Post Keynesian Economics* V(2), 208-227, 1983.
- Delsen, Lei., & Genevieve Reday-Mulvey, *Gradual retirement in the OECD countries: macro and micro issues and policies*. Dartmouth, 1996.
- Ermisch, John F., & Stephen P. Jenkins, Retirement and housing adjustment in later life: evidence from the British Household Panel Survey, *Labour Economics* 6, 311-333, 1999.

- Eurostat, The European Community Household Panel(ECHP): Survey methodology and implementation. Volume 1, Office for Official Publications of the European Communityied, Luxembourg, 1996.
- Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. Bayesian Data Analysis, Chapman and Hall, 2004.
- Gustman, Alan L. and Thomas L. Steinmeier, Retirement Outcomes in the Health and Retirement Study, Social Security Bulletin, 63(4), 57-71, 2000.
- Hamermesh, Daniel S, Consumption during Retirement: The Missing Link in the Life Cycle, Review of Economics and Statistics, 66(1), 1-7, 1984.
- Han, Shin-Kap., Moen, Phyllis, Clocking Out: Temporal Patterning of Retirement, American Journal of Sociology, 105(1), 191-236, 1999.
- Hastie, T. J. and Tibshirani, R. J, Generalized Additive Models, Chapman and Hall, 1990.
- Hausman & Wise, Attrition bias in experimental and panel data: The Gary income maintenance experiment. Econometrica, 47, 455-473, 1979.
- Havighurst. R. J. . Personality and patterns of aging, Gerontologist 8, 20-23, 1968.
- Heeringa, S.G. and Connor, J.H, Technical description of the Health and Retirement Survey Sample Design, 1995
- Hirano et al, Combining panel data sets with attrition and refreshment samples. Econometrica, 69, 1645-1659, 2001.

- Iacovou, Maria, Health, wealth and progeny: Explaining the living arrangements of older european women. Institute for Social Economic Research, Essex University, 2000..
- Little R. J. A. "Missing data adjustments in large surveys," *Journal of Business and Economic Statistics*, 6, 287-301, 1988.
- Little, R. J. A., and An, H. Robust likelihood-based analysis of multivariate data with missing values, *Statistica Sinica*, 14, 949-968, 2004.
- Little, R. J. A. and Rubin, D. B, *Statistical Analysis with Missing Data*, Wiley: New York, 2002.
- Mariger, Randall P, A Life-Cycle Consumption Model with Liquidity Constraints: Theory and Empirical Results, *Econometrica*, 55(3), 533-57, 1987.
- Meyer, Bruce D., & James X. Sullivan, Consumption and income poverty for those 65 and over. Harris school working paper series, 2007.
- National Center for Health Statistics 2007 National Health Interview Survey (NHIS) Public Use Data Release: NHIS Survey Description, Division of Health Interview Statistics, National Center for Health Statistics, Centers for Disease Control and Prevention, 2008.
- Oh, H. L and Scheuren, F. S, Weighting adjustments for unit nonresponse, in *incomplete Data in Sample Surveys*, 2, W. G. Madow, I. Olkin and D.B. Rubin, eds. New York: Academic Press, 1983.
- Palmer, B. A., Retirement Income Replacement ratios: An update,

- Benefits Quarterly, Second Quarter, 59-75, 1994.
- Raghunathan, T. E., Lepkowski, J. M., Hoewyk, J. V., and Solenberger, P, A Multivariate Technique for Multiply Imputing Missing Values Using a Sequence of Regression Models," Survey Methodology, 85, 85-95, 2001.
- Raghunathan, T. E., Solenberger, P., and Hoewyk, J. V., and IVEware: Imputation and Variance Estimation Software User Guide, Survey Research Center, Institute for Social Research, University of Michigan, available at <http://www.isr.umich.edu/src/smp/ive/>, 2002.
- Rosenbaum, P. R. and Rubin, D. B, Reducing Bias in Observational Studies Using Subclassification on the Propensity Score, Journal of American Statistical Association, 516-524, 1984.
- Rubin D. B, Multiple Imputation for Nonresponse in Surveys, John Wiley: New York, 1987a.
- Rubin, D. B, A Noniterative Sampling/Importance Resampling Alternative to the Data Augmentation Algorithm for Creating a Few Imputations When Fractions of Missing Information are Modest: The SIR-algorithm," A discussion of Tanner and Wong's "The Calculation of Posterior Distributions by Data Augmentation," Journal of American Statistical Association, 82, 543-546, 1987b.
- Ruhm, Christopher J. Bridge Jobs and Partial retirement, Journal of Labor Economics 8, 482-501, 1990.
- Schenker N., Raghunathan T.E., Chiu P-L, Makuc D.M., Zhang G., and Cohen A.J, Multiple Imputation of Family Income and

- Personal Earnings in the National Health Interview Survey: Methods and Examples, Technical Report, National Center for Health Statistics, Available online at <http://www.cdc.gov/nchs/about/major/nhis/2004imputedincome.htm>, (2006),
- Tang, L., Song, J, Belin, T. R. and Unützer, J. A Comparison of Imputation Methods in a Longitudinal Randomized Clinical Trial, *Statistics in Medicine*, 24, 2111-2128, 2004.
- U.S. Bureau of the Census. Survey of Income and Program Participation User's Guide, 2001.
- World Health Organization, Constitution of the World Health Organization. Chronicle of the World Health Organization, Annex 1. s.29. Technical Secretariat, WHO Copenhagen, 1947.
- Quinn, Joseph F. The Role of Bridge Jobs in the Retirement Patterns of Older Americans in the 1990s. Working papers in Economics, 1996.
- Quinn, Joseph F., Retirement patterns and bridge jobs in the 1990s, EBRI Issue Brief, 206, 1999.
- <http://www.kli.re.kr/klosa> (고령화연구패널조사 웹사이트)
- <http://www.kli.re.kr/wps> (사업체패널조사 웹사이트)
- <http://www.kli.re.kr/klips> (한국노동패널조사 웹사이트)
- <http://www.nps.or.kr/jsppage> (국민노후보장패널조사 웹사이트)
- <http://hrsonline.isr.umich.edu> (HRS 웹사이트)
- <http://www.cdc.gov/nchs/nsfg.htm> (NSFG 웹사이트)
- <http://www.isr.umich.edu/src/smp/ive> (IVEware 웹사이트)
- <http://www.cdc.gov/nchs/nhis> (NHIS 웹사이트)

國民年金研究院 發刊 報告書

□ 2006년도

정책보고서	국민연금 중기재정전망(2007~2011)	박성민 외	2006.7
2006-01			
Working Paper	국민연금 재정추계를 위한 거시경제 변수 장기 전망 방법론 연구 - 성장모형을 중심으로	최기홍	2006.9
2006-01			
Working Paper	증감 노동생명표에 의한 경제활동 기대여명의 측정과 전망	최기홍	2006.9
2006-02			
정책보고서	국민연금 수급자 관리체계 개선방안	김성숙 외	2006.9
2006-03			
연구보고서	연금 및 경제부문을 연계한 정책시뮬레이션 모형구축	박무환	2006.10
2006-01			
연구보고서	국민연금기금 위탁운용 주식벤치마크에 관한 연구	정문경 외	2006.10
2006-02			
연구보고서	국민연금기금 직접운용 주식벤치마크 설정에 관한 연구	남재우 외	2006.10
2006-03			
연구보고서	공적 연기금의 사회적 책임투자 방안	유상현	2006.12
2006-05			
연구보고서	국민연금기금 사모투자 성과평가에 관한 연구	정문경	2006.12
2006-06			
연구보고서	국민연금기금의 동태적 ALM 구축의 타당성 연구	원종현	2006.12
2006-07			
연구보고서	국민연금의 신용파생상품 투자에 관한 연구	남재우 외	2006.12
2006-08			
연구보고서	국민연금 재정평가지표와 필요보험료를 추계	김순옥 외	2006.12
2006-09			
연구보고서	다층노후소득보장체계 구축 연구(II)	김원섭 외	2006.12
2006-11			
연구보고서	자본시장발전을 위한 국민연금기금의 역할	박무환	2006.12
2006-12			

연구보고서 2006-13	국내외 추가수익률 프리미엄에 관한 연구	한덕희	2006.12
연구보고서 2006-14	국민연금기금과 국내장기채시장 발전방향	박태영 외	2006.12
연구보고서 2006-15	국민연금 보험료수입 및 신규수급자 급여액 추계	김순옥 외	2006.12
연차보고서 05	2005년도 국민연금 기금운용 연차보고서 (국문, 영문)	김성숙 외	2006.12
연구보고서 2006-16	국민연금 가입자 특성과 수급부담구조 분석	최기홍 외	2006.12
연구보고서 2006-17	국민연금 노령연금 및 반환일시금 수급자 추계	신경혜	2006.12
연구보고서 2006-18	국민연금 장애 및 유족연금 수급자 추계	신경혜	2006.12
연구보고서 2006-19	국민연금기금의 해외부동산 투자 방안	유상현	2006.12
연구보고서 2006-20	국민연금 가입기간별 가입자 추계	박성민	2006.12
연구보고서 2006-21	국민연금 장기재정추계모형 2006	박성민	2006.12

□ 2007년도

연차보고서 2007-01	국민연금 중기재정전망(2008~2012)	박성민 외	2007.7
정책보고서 2007-01	노령기초소득보장제도의 국제비교연구 -미국, 일본, 캐나다. 호주를 중심으로	김성숙 외	2007.11
연구보고서 2007-01	연금저축이 국민저축에 미치는 영향	박무환	2007.12
연구보고서 2007-02	여성의 연금수급권 강화를 위한 국민연금 개선방안 연구	강성호 외	2007.12
정책보고서 2007-03	국민연금기금의 성과분해 시스템 구축	정문경	2007.12

연구보고서 2007-04	국민연금과 타사회보험간 중복급여조정 개선방안	이용하 외	2007.12
연구보고서 2007-05	일반균형모형을 통한 국민연금 정책시뮬레이션	최기홍	2007.12
연구보고서 2007-06	공적연금연계 재정추계모형 구축-국민연금과 공무원연금의 연결통산 연계모형	박성민 외	2007.12
연구보고서 2007-07	국민연금 잠재부채(I)	김순옥	2007.12
연구보고서 2007-08	2007년 연금법 개정을 반영한 추계모형 재구축 및 재정시뮬레이션	김순옥 외	2007.12
연구보고서 2007-09	신규수급자 특성분석	신경혜	2007.12
연구보고서 2007-10	국민연금이 노년층의 소득 빈곤과 불평등에 미치는 영향	김경아	2007.12
연구보고서 2007-11	연금부채를 고려한 국민연금 성과평가에 관한 연구	정문경 외	2007.12
연구보고서 2007-12	근로유인 제고를 위한 국민연금제도 개선방안	김원섭	2007.12
연구보고서 2007-13	국민연금기금의 환관리 방안	한덕희	2007.12
정책보고서 2007-04	국민연금 부동산투자의 공정가 산출 및 성과평가 방안	남재우	2007.12
연구보고서 2007-14	국민연금기금의 국내주식 액티브/패시브 운용방안연구	박태영	2007.12
연구보고서 2007-15	국민연금기금의 대체투자 운용방안	유상현	2007.12
연구보고서 2007-16	국민연금 기금운용체계 개선방안에 대한 연구	박태영, 원종현	2007.12
연구보고서 2007-17	국민연금 급여수준의 국제비교 연구	이용하	2007.12
Working Paper 2007-02	퇴직급여(연금)제도의 문제점과 발전방안	이용하	2007.12
Working Paper 2007-03	Analyses on the Introduction and Reform of the National Pension Scheme in Korea from the Perspective of Social Rights Theories	김원섭	2007.12

연차보고서 2007-04	2006년도 국민연금 기금운용 연차보고서	국민연금 연구원	2007.12
연차보고서 2007-05	2006 Annual Report on National Pension Fund Management	국민연금 연구원	2007.12
정책보고서 2007-05	국민연금의 거시경제 과급효과와 기금의 장기운용 방향	국민연금 기금장기 운용전략 기획단	2007.12
정책보고서 2007-06	장애 판정체계 개선 및 재활서비스 구축방안 연구	가톨릭대 학교	2008.4

□ 2008년도

연차보고서 2008-01	국민연금 중기재정전망(2009~2013)	박성민 외	2008.4
연차보고서 2008-02	국민연금 기금운용 성과평가	정문경 외	2008.8
연차보고서 2008-03	2007년도 국민연금 기금운용 연차보고서	기금정책 팀	2008.12
연차보고서 2008-04	2007 Annual Report on National Pension Fund Management	기금정책 팀	2008.12
연구보고서 2008-01	국민연금 수급부담구조 분석	재정추계 분석팀	2008.12
연구보고서 2008-03	국민연금기금 사모투자지분 가치평가에 관한 연구	정문경	2008.12
정책자료 2008-02	2008년 상반기 국민연금 기금운용 성과평가 보고서	김대호, 황정욱	2008.12
연구보고서 2008-04	국민연금과 퇴직연금의 보완적 발전방안	이용하, 정해식	2008.12
연구보고서 2008-05	중기재정추계모형 개선 및 중기재정전망 (2010~2014)	신경혜	2008.12
연구보고서 2008-06	국민연금기금의 분할 운용에 관한 연구	이재현, 이준행	2008.12
연구보고서 2008-07	주요거시경제변수 전망을 위한 거시경제계량모형 개발	박무환	2008.12

연구보고서 2008-08	가입자통합 재정추계모형 타당성 연구	박성민	2008.12
연구보고서 2008-09	국민연금 ALM을 반영한 기금운용전략 수립방안	원종현, 신성환외	2008.12
연구보고서 2008-10	국민연금기금의 자본시장에 대한 파급효과 분석	김경아	2008.12
연구보고서 2008-11	국민연금기금의 금융신상품 투자에 관한 연구	박태영, 이재현	2008.12
정책자료 2008-03	수급권 소멸시효제도 개선방안	김성숙, 유호선외	2008.12
연구보고서 2008-12	비정규직 근로자의 국민연금 가입제고를 위한 정책방향 검토	권혁진, 최은아	2008.12
연구보고서 2008-13	국민연금이 거시경제 및 경제행태에 미치는 영향분석(I)	박무환, 최기홍외	2008.12
교육교재 2008-01	공적연금의 이해	김성숙 외	2008.12
조사자료 2008-01	주요국 공적연금의 재정계산제도	김순옥 외	2008.12
연구보고서 2008-14	장래인구추계의 방법과 현황	우해봉	2008.12
정책보고서 2008-01	국민연금기금의 위험관리체계 개선 -대체투자군 중심으로-	정문경, 남재우외	2008.12
연구보고서 2008-15	국민연금 기금운용의 허용위험기준에 관한 연구	박태영	2008.12
연구보고서 2008-16	국민연금의 재정정책의 후생효과 분석을 위한 중복세대모형의 개발	최기홍, 신성희	2008.12
연구보고서 2008-17	국내위탁주식 운용유형의 설정 및 관리방안	남재우, 이인형	2008.12
조사보고서 2008-01	제2차 우리나라 중·고령자 경제생활 및 노후준비 실태 -KReIS 기초분석보고서-	석상훈, 송현주외	2008.12
정책보고서 2008-02	국민연금 미수급자 규모추정과 지역가입자 관리 개선방향	강성호, 김태완외	2008.12
교육교재 2008-02	공적연금기금운용론	정문경, 박태영외	2008.12

정책자료 2008-04	외국인 가입자 및 수급자의 효율적 관리방안	이용하	2008.12
정책자료 2008-05	최소 가입기간 단축의 적절성 검토	권혁진	2008.12
정책자료 2008-06	기초노령연금 관리체계 개선방안	정해식	2008.12

□ 2009년도

연차보고서 2009-01	2008년 국민연금 기금운용 성과평가	남재우 외	2009.8
용역보고서 2009-02	사회보험료 사후정산제도 시행에 따른 건설 일용근로자 국민연금 가입제고 방안	심규범, 김지혜 등	2009.11
정책보고서 2009-01	국민연금의 재활서비스 구축방안 연구	이용하	2009.12
정책보고서 2009-02	부양가족제도의 개선방안 검토	김현수	2009.12
정책보고서 2009-03	사망관련 급여제도 개선방안 검토	유호선, 홍성우	2009.12
연구보고서 2009-01	공적연금 연계모형 구축(II) -공적연금연계를 반영한 국민연금 재정추계	박성민	2009.12
연구보고서 2009-02	국민연금의 확률적 재정추계방법 연구	신경혜, 이재현	2009.12
연차보고서 2009-02	국민연금 중기재정전망(2010-2014)	박성민 외	2009.12
정책보고서 2009-04	국민연금 기금운용 프로세스 현황 분석 및 개선방향	박태영	2009.12
연구보고서 2009-03	시장배수를 이용한 국민연금기금의 국내사 모펀드 가치평가에 관한 연구	정문경	2009.12
연구보고서 2009-04	동태적 최적화모형 개발(I)	박무환, 이태정, 양준모	2009.12
정책보고서 2009-05	장애연금 급여수준 적정성 분석	강성호, 홍성우	2009.12

연구보고서 2009-05	국민연금기금의 액티브 위험 관리방안	정문경, 남재우 외	2009.12
정책보고서 2009-06	국민연금기금 위탁운용사 선정 및 관리체계 개선-국내위탁주식을 중심으로	남재우 외	2009.12
연구보고서 2009-06	인구고령화가 경제성장과 국민연금에 미치는 영향	최기홍, 신성희	2009.12
연구보고서 2009-07	글로벌 금융위기가 주요국 및 국내 금융시장 변동성에 미치는 영향 분석	김경아	2009.12
연구보고서 2009-08	국민연금이 거시경제 및 경제행태에 미치는 영향 분석(II)-국민연금이 경제성장 및 개인 소비에 미치는 영향 중심으로	박무환, 최기홍, 김경아	2009.12
조사보고서 2009-01	우리나라 중·고령자의 생애 직업이력 및 연금이력 실태	석상훈 외	2009.12
연구보고서 2009-09	외환시장과급효과를 고려한 국민연금의 해외투자규모에 관한 연구	류두진, 박성준	2009.12
연차보고서 2009-03	2009년 상반기 국민연금 기금운용 성과평가	김대호, 황정욱	2009.12
정책보고서 2009-07	국민연금 재정계산의 개선방안	김순옥, 신승희	2009.12
정책보고서 2009-09	수급권 변동관리 개선방안	김성숙	2009.12
정책보고서 2009-10	국민연금 적용제외자 현황과 노후 소득보장 대책	우해봉, 최은아	2009.12
연구보고서 2009-10	국민연금기금의 투자자산군간 위험전이와 전술적 자산배분전략에 관한 연구	이재현	2009.12
정책보고서 2009-11	국민연금의 미시모의실험 모형 연구 및 개발	권혁진, 한정립	2009.12
용역보고서 2009-03	공·사 연금 정보연계의 타당성과 추진방안에 대한 연구	김원섭	2009.12
용역보고서 2009-04	국민연금 장애심사규정에 관한 연구	박병배 외	2009.12
조사자료 2009-01	장애인의 기초 소득보장 -호주, 스웨덴, 일본, 영국, 네덜란드	권혁진, 김경아 외	2009.12

정책자료 2009-01	국민연금 사업장관리 효율화 방안 검토	이용하, 권혁진	2009.12
정책자료 2009-02	양육 크레딧 도입 방안	유호선	2009.12

□ 저자 약력 □

• 석 상 훈

성균관대학교 졸업

성균관대학교 박사 (경제학)

현 국민연금연구원 부연구위원

• 최 옥 금

성균관대학교 졸업

성균관대학교 박사 (사회복지학)

현 국민연금연구원 부연구위원

• 송 현 주

인하대학교 졸업

인하대학교 박사수료(소비자학)

현 국민연금연구원 주임연구원

• 이 은 영

한국외국어대학교 졸업

연세대학교 석사(통계학)

현 국민연금연구원 주임연구원

• 오 지 연

인천대학교 졸업

인천대학교 석사(가계경제학)

현 국민연금연구원 연구원

• 신 혜 리

숭실대학교 졸업

연세대학교 석사(사회복지학)

현 국민연금연구원 연구원

조사보고서 2010-02 (2/2)

제3차(2009년도) 우리나라 중·고령자의
경제생활 및 노후준비 실태
-국민노후보장패널조사(KReIS) 심층분석보고서-

2010年 10月 日 印刷

2010年 10月 日 發行

發行人 : 전 광 우

編輯人 : 백 화 중

發行處 : 國民年金公團

國民年金研究院

서울특별시 강남구 논현동 4-15

TEL : 02-3218-8619 / FAX : 02-541-2901

ISBN