

발 간 등 록 번 호

11-1240000-000449-10

GOVP1200609247

T0006263

가구조사자료의 종단화 방안

2005. 8



통계청

Korea National Statistical Office

제 출 문

본 보고서를 2005년도 연구과제인 「가구조사자료의 종
단화 방안」의 연구결과 보고서로 제출합니다.

2005년 8월

통계개발팀장 최성욱

연구자 : 이지연

요 약 문

과 제 명	가구조사자료의 종단화방안		
중심단어	패널자료, 패널소실, 생존분석, 경제활동인구조사, 도시가계조사		
연구기관	통계정책국 통계개발팀	연구자	이지연
연구기간	2005. 1. ~ 2005. 8.		

본 연구의 목적은 통계청에서 생산하고 있는 가구단위 조사 중 반복횡단조사로 설계된 경제활동인구조사와 도시가계조사자료를 이용하여 1998년에서 2002년까지 패널자료를 구축하고, 이 자료가 가진 표본 대표성을 분석해 보는 것이다.

총 60회차 월별패널로 구축된 경향의 가구단위 최종 표본소실율은 53.5%, 총 20회차 분기별 패널로 구축된 도시가계의 최종 표본소실율은 63.3% 였다. 연도별 패널소실 추이는 첫해의 소실이 가장 많이 발생했고, 점차 감소하여 3년 이후 부터는 안정화되었다.

생존분석을 통해 가구와 가구원의 사회경제적 특성 차이가 패널의 잔류기간에 미친 영향을 분석한 결과 개인차원에서는 연령이, 가구 차원에서는 주택소유형태가 가장 중요한 변수로 작용했다. 개인의 특성별로는 여자보다는 남자가, 장년층보다는 젊은층이 패널소실 위험률이 높았다. 학력이 높을수록 패널소실확률도 함께 증가했으며, 취업자보다는 실업자의 패널소실확률이 높았다.

가구조사결과를 일반에게 패널자료로 제공하기 위해서는 먼저 패널소실의 계절적 요인이 분석결과에 미치게 될 영향과 무응답으로 인한 가중치 조정문제에 관한 심층적인 연구가 선행되어야 한다. 이러한 연구결과를 바탕으로 패널자료의 제공범위 및 유형에 관한 합리적인 기준을 설정할 수 있을 것으로 보인다.

목 차

I. 서론	2
1. 한국 패널조사의 역사	2
2. 패널조사 수요증가의 원인	3
3. 통계청 가구단위조사의 패널적 특성	3
4. 연구목적	4
II. 패널자료의 특성과 해외 패널조사 현황	6
1. 패널자료의 특성	6
1) 패널조사의 장점	6
2) 패널조사의 단점	6
2. 해외 패널조사의 현황과 주요 특징	7
III. 경찰 및 도시가계패널의 특성과 선행연구	10
1. 거처패널	10
2. 선행연구: 경찰 및 도시가계 패널구축 사례 및 방법	11
IV. 연구방법	14
1. 자료	14
2. 패널구축방법	14
3. 표본소실	15
V. 분석결과	16
1. 가구관리 경찰패널의 표본소실 추이 및 특성	16
1) 패널표본의 대표성	16

2) 경활패널의 표본소실을 추이	16
3) 경활패널의 최초 및 최종잔류표본의 특성	19
4) 도시가계패널의 표본소실 추이 및 특성	21
2. 경활 및 도시가계패널의 생존기간 분석	24
1) 경활패널의 가구원 특성별 생존곡선: 생명표 분석	24
2) 도시가계패널의 가구특성별 생존곡선	28
3. 경활 및 도시가계패널 생존기간 결정요인 분석: 콕스비례위험모형	30
VI. 결론 및 제언	35
참고문헌	39
부 록	40

I . 서 론

I. 서 론

1. 한국 패널조사의 역사

한국에서 특정한 사회경제적 현상을 시간의 흐름에 따라 반복적으로 측정하는 종단조사(longitudinal survey), 특히 동일한 대상을 반복적으로 조사하는 패널조사(panel survey)의 역사는 그리 길지 않다. 미국은 1960년대 후반부터 대표적인 패널조사인 Personal Survey of Income Dynamics (PSID)나 National Longitudinal Survey(NLS)등을 시작했다. 그 이후로 패널조사를 실시하는 나라들이 속속 늘어나면서 현재 전 세계적으로 20여개국에서 대규모 패널조사들을 정기적으로 실시하고 있다. 대다수의 패널조사들이 대학이나 연구기관을 통해서 실시되고 있지만 캐나다, 프랑스, 네델란드등에서는 국가 통계기관이 직접 패널조사를 실시하고 있다(Gong, 2004).

한국의 최초의 패널조사는 대우경제연구소가 1993년 부터 연단위로 실시한 ‘한국가구패널조사(Korean Household Panel Study)’이다. 이 조사는 가구 및 가구원의 경제활동과 이에 영향을 미치는 사회경제적 변수들을 측정하기 위해 전국의 4,547가구와 18세 이상의 가구원 10,460명을 대상으로 처음 실시되었다. 그러나 높은 패널 소실율과 외환위기등으로 인해서 1998년에는 약 54%의 가구만이 패널로 남게 되면서 6회차 조사를 마지막으로 조사가 중단되었다 (대우경제연구소, 1999).

그 이후 가장 대표적인 패널조사는 ‘한국노동패널조사(Korea Labor and Income Panel Study)’로서, 한국노동연구원이 1998년부터 매년마다 전국 5천 가구, 15세 이상 경제활동인구 13,321명을 대상으로 경제행위 유형과 가구의 경제적 상황변화를 조사하고 있다(한국노동연구원, 2005). 이 외에도 최근 2-3년 사이에 전국규모의 패널조사들이 급격히 늘고 있는데, 정부승인통계로 등록된 패널조사만도 한국직업능력개발원의 ‘한국교육고용패널조사(2004년)’, 보건복지부의 ‘한국복지패널(2005년)’등이 있다.

2. 패널조사 수요증가의 원인

패널조사를 포함하는 종단조사와 일반적인 횡단조사간의 가장 큰 차이점은 시간이 자료의 수집과 분석에 있어 중요한 변수인가 아닌가에 달려 있다 (Gong, 2004). 포괄적인 의미로는 시간의 흐름에 따른 변화가 측정되는 모든 조사를 종단조사라고 부를 수 있지만, Menard(2002)는 다음의 세 가지 요건을 언급했다. 첫째, 조사되는 각 변수나 항목이 적어도 두 시점 이상의 기간을 두고 수집되어야 한다. 둘째, 첫 번째 조사 시점에서 다음번 조사 시점까지 조사대상자가 동일하거나 비교 가능한 집단이어야 한다. 마지막으로, 종단 조사 자료는 상이한 시점들에서 비교 가능해야만 한다.

최근 패널조사에 관한 수요가 급격히 증가하고 있는 이유는 패널자료를 통해 특정정책이나 프로그램이 목표집단(target group)에 미친 인과적인 효과를 직접적으로 평가하거나 예측(forecasting)해 볼 수 있는 개인차원의 정보를 얻을 수 있기 때문이다. 근래 종단자료(longitudinal data) 분석 기법의 발달과 각종 통계분석패키지들이 널리 보급되었다는 점 또한 패널조사 수요를 확대시킨 한 요인으로 작용하고 있다.

3. 통계청 가구단위조사의 패널적 특성

통계청에서는 2005년 현재 총 64종의 통계를 작성하고 있는데, 최초로 패널조사로 설계된 조사들은 아니다. 그러나 가구를 단위로 하는 표본조사 중에서 반복횡단조사로 설계된 ‘경제활동인구조사(이하 경활)’, ‘도시가계조사(이하 도시가계)’등의 조사는 패널자료의 속성을 가지고 있다. 경활은 전국의 약 33,000가구를 대상으로 매월 15일 현재 만 15세 이상 인구 경제활동상태를 조사한다. 도시가계는 경활의 표본으로 선정된 가구중 도시지역에 거주하는 가구원수 2인 이상의 약 5천 가구를 대상으로 가구의 소득 및 소비실태

를 매월 조사하여 분기별로 발표하는 조사다¹⁾.

두 조사에서 표본선정은 인구주택총조사를 표본프레임으로 해서 지역, 거주, 가구주의 특성을 기준으로 추출된다. 한번 선정된 표본은 차기 인구주택총조사를 통해 표본프레임이 갱신되기 전까지 5년간은 유지된다. 즉 동일한 대상을 상대로 2002년 까지는 5년 동안, 연동표본이 도입된 2003년부터는 약 3년 동안 매월 조사가 반복적으로 실시되는 것이다. 이러한 특성들은 앞서 언급한 Menard(2002)의 중단조사의 세 가지 조건을 충족시키고 있다. 또한, 경찰과 유사한 조사인 Labour Force Survey를 실시하고 있는 캐나다, 영국, 뉴질랜드 통계청에서도 조사기간이 중복되는 가구들을 중단자료 형태로 구축해서 제공하고 있다(Rowe and Nguyen, 2004; Laux and Tonks 1998; and Kuzmicich and Wigbout 2001).

4. 연구목적

이 연구의 목적은 통계청에서 생산하고 있는 가구조사 중 반복횡단조사로 설계된 경제활동인구조사(이하 경찰)와 전국가계조사자료(이하 도시가계)를 이용해서 패널자료를 구축하고, 구축된 패널자료가 가진 표본대표성을 분석해 보는 것이다. 일반적으로 횡단조사로 설계된 자료를 패널로 구축해서 분석하게 될 경우 두 가지 방법론적인 문제들이 제기 된다: 표본의 체계적인 탈락으로 인해 발생하는 편향과 응답오차로 인해 발생하는 편향이다(Clarke and Tate, 1999). 본 연구는 전자의 문제에 중점을 두고자 한다.

먼저 본 연구는 패널조사가 가진 일반적 특성 및 장단점을 기술하고, 해외의 주요 패널조사 현황을 간략히 살펴본 후 통계청에서 작성된 가구단위조사가 가진 패널자료로서의 특성과 선행연구들을 살펴 볼 것이다. 다음으로 경찰 및 가계조사패널 마지막 차수까지 생존한 가구와 중도에 탈락한 가구

1) 도시가계조사는 2003년 부터 가구소비실태조사와 통합되면서, 표본의 크기를 7,500가구로 확대하고 조사범위를 농촌을 포함한 전국으로 확대시키면서 '전국가계조사'로 명칭을 개정하였다.

를 비교하여 패널표본으로서의 대표성을 평가해 볼 것이다. 만약 두 조사에서 체계적인 탈락이 발생하고 있다면 구축된 패널이 모집단을 특성을 대표한다고 말할 수 없을 것이다. 따라서, 패널자료로서의 유용성은 떨어질 수밖에 없다. 또한, 가구와 가구원의 사회경제적인 특성의 차이가 패널의 잔류기간에 과연 얼마만큼의 영향을 미치는지를 생존분석을 통해 심층적으로 분석해보고자 한다. 마지막으로 본 연구는 경찰 및 가계조사자료를 패널로 구축하여 일반에게 제공하는데 고려되어야만 할 사안들을 제언하고자 한다.

II. 패널자료의 특성과 해외 패널조사 현황

1. 패널자료의 특성

1) 패널조사의 장점

패널조사는 시간의 흐름에 따른 변수들의 변화 상태를 반복적으로 측정하기 때문에 변화과정 뿐만 아니라 특정한 원인에 대한 결과가 장기간에 걸쳐 발생할 경우에도 이를 파악할 수 있다는 장점이 있다. 이러한 특성때문에 패널조사는 사회변화 혹은 개인의 태도나 행위변화에 관한 인과관계 분석에 가장 적합한 통계자료를 제공한다(Davies, 1994).

또한 패널조사를 포함한 종단조사들은 연령효과, 코호트효과, 시기효과를 각각 분리해 낼 수 있다(Diggle, Heagerty, Liang, and Zeger, 2002). 예를 들어 한 조사를 통해 연령에 따라 희망자녀수가 다르다는 사실이 밝혀졌다고 하자. 하지만 그 차이가 개인의 연령이나 생애주기 단계상의 위치가 다르기 때문인지(연령 효과), 베이비붐이나 베이비버스트 세대와 같이 출생시점이 달라서 생기는 상이한 사회화 경험 때문인지(코호트 효과), 경제위기와 같은 특정 사건의 영향(시기효과) 때문 인지는 일회적인 측정으로 끝나는 횡단조사를 통해 파악해내기란 어렵기 때문이다.

2) 패널조사의 단점

종단조사는 특정현상을 여러 번에 걸쳐서 반복적으로 조사하기 때문에 횡단조사에 비해 비용이 많이 든다는 단점이 있다. 특히, 패널조사의 경우 동일인을 지속적으로 추적하기 위해서는 더 많은 시간과 경제적인 비용이 요구된다.

일반적으로 패널조사는 조사설계에서 공표까지의 전 과정에 시간이라는

차원이 개입되기 때문에 횡단조사에 비해 더 복잡하다. 패널조사의 계획과 관리가 어려운 근본적인 이유는 장기적으로 설정된 목표나 수요에 비해 단기간에 동원 가능한 자원들 사이에는 항상 갭이 존재하기 때문이다. 따라서 조사 초기부터 장기간의 목표와 자원이 충분히 고려된 상태에서 진행되지 않으면 패널조사는 지속되기 어렵다.

패널조사의 가장 큰 문제점은 패널 소실(panel attrition)이다. 시간이 지날수록 패널조사는 처음 조사 차수와 다음 차수 간에 동일한 표본을 유지하기가 점점 어려워진다. 조사대상자가 응답을 거절하거나 거주지 이전 및 사망 등의 이유로 조사에서 탈락되는 경우 등 절단사례(censored cases)들이 점차 많아지기 때문이다. 이 때 절단된 사례들이 무작위적으로 발생하는 것이 아니라, 일정한 소득이나 직업을 갖고 있지 않아서 추적하기 어려운 빈곤층과 같이 특정한 계층이나 집단에서 주로 발생했다면, 표본에 남아있는 패널이 모집단의 대표성을 유지하기 어려울 것이다. 이를 패널소실이라 한다. 패널조사에서는 조사 차수가 길어질수록 더 많은 정보를 수집할 수 있고, 조사간격이 짧을수록 기억오차가 줄어드는 장점이 있지만, 응답자의 부담은 그만큼 더 증가하게 되고 패널 소실의 가능성 또한 더 커진다는 단점이 있다.

2. 해외 패널조사의 현황과 주요 특징

현재 전 세계적으로 가구단위의 패널조사들이 다양한 형태로 진행되고 있는데, <부표 1>은 한국을 제외한 19개국 24개 주요 해외 패널조사들을 정리한 것이다. 표의 내용을 살펴보면 네가지 특징을 발견할 수 있다. 첫 번째는 패널조사의 약 4/5 정도는 대학이나 연구소에서 자체적인 조사조직이나 전문 여론조사기관을 통해서 조사를 진행하고 있다는 점이다. 반면에 국가통계기관에서 직접 패널조사를 실시하고 있는 나라는 캐나다, 프랑스, 네델란드이며, 미국과 헝가리는 국가통계기관과 대학이 공동으로 조사를 주관하고 있다.

두 번째 특징은 거의 대부분의 패널조사들이 가구를 조사대상으로 하고 있지만 다루어 지고 있는 주제는 상당히 다양하다는 점이다. 가족의 구성과 인구학적 특성 뿐만 아니라 가구경제, 교육과 취업, 주택이나 건강등의 정보 개인 및 가구차원 뿐만 아니라 지역차원에 관련된 정보들도 조사되고 있다 (Gong, 2004).

세 번째 특징은 패널조사의 유형들이 상당히 다양하다는 점이다. 먼저, 표본 프레임에 따라서도 조사유형을 나눌 수 있다. 일반적으로 표본조사들은 전국 단위의 정보가 실린 센서스자료를 표본프레임으로 선정하는 경우도 많다. 그러나 호주의 General Customer Survey는 Family and Community Service의 행정자료에 등록된 소득지원대상자만을 표본프레임으로 선정하거나, 프랑스의 French Household Panel처럼 특정지역 거주자만을 선정하는 경우도 있다.

다음으로 표본 선택의 기준에 따라, 패널조사는 대표 패널(representative panel), 코호트 패널(cohort panel)로 조사유형을 나눌 수 있다. 대표 패널이 일정지역의 전체 모집단 특성이 그대로 추론될 수 있도록 설계된 표본패널이라면, 코호트 패널은 출생시기와 같은 세대의 개념이 표본추출의 기준이 되는 패널이다. 예를 들어 미국의 PSID가 성, 연령, 인종구성에 있어서 전국을 대표하는 가구패널을 사용해서 실시된다. 반면에 NLS는 특정 출생 코호트를 먼저 선정한 후 성과 인종구성비가 전국 특성을 반영하도록 설계했다는 점에서 차이가 있다. 코호트 패널의 장점은 변화의 원인이 연령효과 때문인지, 특정 사건이나 시대적 유행과 같은 시기효과 때문인지를 구별할 수 있다는 점이다. 또한, 다른 코호트 패널이 존재할 경우 상호비교를 통해서 세대간의 차이도 분석할 수 있다.

마지막 특징은 조사의 응답자뿐만 아니라 조사자체도 시간에 따라 변화할 수 있다는 점이다. 모든 패널조사들이 설계초기에 패널종료시점, 표본규모, 조사항목등 조사특성에 대해서 완전한 계획을 세우고 진행되는 것은 아니다. 예를 들어, 미국의 Health and Retirement Survey처럼 처음에는 두 개의 개별적인 연구로 시작했다가 통합되는 경우도 있다. PSID는 처음 선정된 패널

가구 자녀들이 성장하여 분가하게 되면 새로운 패널가구로 진입하게 되면서 표본규모가 자연적으로 증가하게 된 사례이다. 또한 스웨덴의 Swedish Panel Survey Market and Non-market Activities는 공통적으로 조사되는 핵심항목이외에도 각 차수별로 시대적인 수요가 반영된 특성항목들을 부가적으로 조사하기도 한다. 이러한 시간에 따른 조사자체의 변화가능성 때문에 전반적인 계획이 조사초기에 설계된다 하더라도, 구체적인 내용들은 자료처리가 모두 끝나고, 자료가 공표되기 전까지는 확정될 수 없는 경우들도 발생하게 된다.

Ⅲ. 경찰 및 도시가게패널의 특성과 선행연구

1. 거처패널

경찰과 도시가게 조사결과를 패널자료로 사용하는 데는 두 가지 사항이 사전에 고려되어야 한다. 먼저 가구를 단위로 추적조사를 실시하는 일반적인 가구패널조사와는 달리 경찰과 도시가게는 거처를 표본선정의 단위로 하고 추적조사를 실시하지는 않기 때문에 ‘거처패널’이라고 부르는 것이 보다 정확한 표현일 것이다.

거처패널과 가구패널은 패널의 무응답 유형에서 차이가 날 수 있다. 일반적으로 패널조사에서 무응답은 다음의 세 가지 경우에 발생한다: 무응답, 부적격(non eligible), 그리고 미상(unknown)인 경우이다(Nathan, 1998). 무응답은 조사대상자가 집에 없거나, 응답을 거절한 경우이다. 부적격은 사망하거나, 이민가는 경우이다. 미상은 응답자가 이주한 장소를 파악할 수 없어서 추적이 불가능 할 때 발생한다. 경찰과 도시가게조사는 가구를 추적조사하지 않기 때문에 일반 가구단위 패널에 비해서 미상으로 인한 패널의 소실이 많이 발생하고, 따라서 패널응답률이 낮아 질 수 밖에 없다.

Nathan(1998)은 영국의 주요패널조사에서 나타나는 소실에 관한 연구에서 이론적으로는 패널조사 응답률은 조사완료 된 사례수를 조사되었어야만 하는 모든 사례수로 나눈 것이라고 했다. 예) $\text{응답률} = \frac{\text{조사완료사례}}{\text{조사완료} + \text{무응답} + \text{부적격} + \text{미상}} * 100$. 이때 현실적으로는 조사되어서는 안 될 부적격 사례를 분모에서 제외시키게 되면 응답률의 하한선이 설정된다. 만약 미상인 사례와 부적격 사례를 명확히 파악 할 수 없기 때문에 미상인 사례도 제외시키게 되면 응답률의 실제적인 상한선을 설정할 수 있다($\text{응답률 상한선} = \frac{\text{조사완료사례}}{\text{조사완료} + \text{무응답}} * 100$). 예를 들어 총 100개의 표본 중에서 70개 표본은 조사되었고, 무응답, 부적격, 미상이 각각 10개씩 발생했다고 하

자. 정확한 패널응답률은 70%이지만, 실제적으로는 부적격과 미상이 구분이 불분명하므로 응답하한선은 77%, 응답상한선은 87.5%가 되는 셈이다. Nathan의 방식을 적용해 보면 거처패널은 정확한 패널응답률 산출방식으로만 계산되어지는데 반해, 일반적인 가구패널은 실질적인 응답하한선 내지 응답상한선 방식으로도 계산될 수 있다.

두 번째는 거처가 표본선정의 단위이기 때문에 봄과 가을의 이사철에는 전출로 인한 무응답이 증가하면서 패널소실에 일정한 계절성이 나타날 것이라는 점이다. 전국규모의 조사들이 인구이동이 적은 달을 선정해 조사를 진행하는 이유도 이러한 계절성이 고려된 것이다. 하지만 사후적으로 패널로 구축된 경찰과 도시가계자료는 이러한 계절성을 조절할 수가 없고, 이것이 자료분석에 있어서 어떠한 영향을 미치는 지가 명확하지 않다는 문제점이 있다.

2. 선행연구: 경찰 및 도시가계 패널구축 사례 및 방법

한국에서 패널조사는 90년대 중반에 최초로 실시되었지만, 이보다 10여년 앞서서 사후적으로 패널자료를 구축한 연구가 시도되었다. 서로 다른 두 시점상의 경찰자료를 매칭시킨 최초의 사례는 류재우·배무기(1984)의 연구이다. 이들은 한 달 동안의 노동력상태 변화를 파악하기 위해서 1983년 4월과 5월의 2회차 경제활동인구조사 자료데이프상에서 조사구역, 가구번호, 성, 생년월일등을 추출해서 동일한 응답자의 것으로 간주되는 월별 자료를 연계시켰다. 이를 통해서 4월 기준으로 전국수준에서 생산가능인구는 94.21%, 농가 98.53%, 비농가 92.49%를 각각 대응시킬 수 있었던 것으로 보인다.

이후 장기간의 경찰자료들을 연계해서 본격적인 패널분석을 실시한 사람은 남재량(1997)이다. 실업률 추세변화를 분석하기 위해서 그는 1982년부터 1994년까지 경찰의 하반기 자료들을 이용하여 5년 단위 패널을 구축했다. 월

별 자료 테이프 상에서 각 개인의 조사구번호, 구역, 거처, 가구원번호, 그리고 출생년월일을 연계해서 각 개인의 패널 ID를 생성한 후 바로 이웃하는 두 시점끼리 비교해서 동일인을 식별해 낼 수 있었다. 패널기간중 1983년과 1984년 9월과 12월 자료는 개인의 생년월일 정보에 누락이 있어서 개인별 ID에 생년월일 대신 성과 연령을 포함시켰는데, 생년월일 없이도 거의 정확한 매칭이 가능했다고 한다.

이러한 방법을 통해서 이웃한 두 달의 자료들을 개인별로 매칭시킬 수 있었던 비율은 전체 13년간의 패널기간중 평균 91.71%였다고 한다. 최근에 올수록 개인별 ID를 통한 자료 매칭율이 높아지고 있는데, 1987년 10월 이후부터는 91%이상, 90년대에는 평균 97%정도 까지 매칭이 가능했다고 한다(남재량, 1997). 이후에 남재량·류근관(1999)은 패널조사로 설계된 조사자료들은 아니지만, 사후에 동일인을 식별해 낼 수 있는 정보들을 통해서 동일인으로 추정되는 자료들을 시간의 흐름에 따라 연계시킨 것을 ‘이동패널(moving panel)’이라고 불렀다.

경찰자료를 월별로 연계해서 구축한 이동패널을 이용하면 어떤 분석이 가능한지 잠시 살펴보자. 횡단조사로 설계된 경찰자료를 이용하면 취업률과 실업률은 알 수 있지만 개인들의 노동력 상태간(취업, 실업, 비경제활동)의 변화를 파악할 수는 없었다. 이에 반해 <참고 1>은 남재량·류근관(1999)이 1985년부터 1997년까지 경찰패널을 이용하여 계산한 노동력 상태별 유동률이다. 예를 들어 1985년부터 1989년까지 남성의 월별평균 취업유동률은 0.056이었다.

<참고 1>. 경찰의 노동력 상태별 유동률

패널기간	성별	취업유동률	실업유동률	비경제활동유동률
1985-1989년	남성	0.056	0.681	0.097
	여성	0.136	0.844	0.090
1993-1997년	남성	0.034	0.586	0.075
	여성	0.080	0.727	0.062

출처: 남재량, 류근관. 1999. “우리나라 여성 노동력 상태의 동태적 특성연구,” 「한국사회과학 21」에서 재인용.

이는 지난달에 취업되어 있던 사람들을 100명이라고 할 때 이번 달에 실업이나 비경제활동인구에서 새로 취업상태로 들어왔거나 취업상태에서 다른 노동력 상태로 빠져나간 남성은 5.6명이라는 의미이다. 여성의 유동률은 남성에 비해 높게 나타나고 있으며 실업상태로의 유출입정도를 측정하는 수치는 남성보다 훨씬 높게 나타나고 있음을 알 수 있다. 이렇듯 패널자료는 개인적인 상태의 변화과정을 측정할 수 있고, 이러한 변화에 영향을 미친 요인들간의 인과관계 분석이 가능한 정보를 제공할 수 있다는 장점이 있다.

개인별 자료를 패널로 구축하는 방법은 특정조사 내에서 뿐만 아니라 동일한 표본프레임으로부터 추출된 표본을 사용하는 상이한 조사간에도 가능했다. 남재량, 이창용(2000)은 경찰과 도시가계를 결합해서 1982년부터 1999년까지 5년 단위 패널을 만들어서 외환위기와 실업률 변화에 관한 연구를 수행하기도 했다.

한편, 도시가계 자료를 패널로 구축한 사례는 황덕순(2002)과 이병희·정재호(2002)의 연구가 있다. 도시가계는 가구원수 2인이상의 비농촌지역에 거주하는 가구 약 5,000여 가구를 대상으로 조사가 이루어진 자료이다. 두 연구 모두 1998-2000년 까지 총 36개월의 도시가계 월별자료를 가구별 ID를 패널자료를 만들었다. 먼저 3개월간의 자료를 연결하여 분기패널을 만들었다. 이 때 분기중에 누락이 발생한 가구는 패널에서 제외시켰는데, 분기별 평균 가구수는 3,839가구 였다고 한다. 다음으로 연속되는 2개의 분기패널을 연결시켜서 6개월간 누락이 한번이라도 발생한 가구는 제외시킬 경우, 평균 가구수는 3,437가구 였다고 한다. 1998년 부터 3년간 모두 조사된 가구만을 연결할 경우 1,475가구의 패널자료를 구축할 수 있었다고 한다. 패널소실의 유형은 자영자가구 보다는 근로자가구와 무업자가구에서, 소득이 낮은 계층에서 소실율이 약간 더 높게 나타났다고 한다.

IV. 연구방법

1. 자료

이 연구에서는 1998년 2002년까지 경찰과 도시가계의 가구관리명부 자료를 사용하였다. 가구관리명부는 특정조사의 표본으로 선정된 표본가구의 관리를 목적으로 매월 조사원에 의해 작성된 자료이다. 연구의 목적인 패널표본의 대표성을 파악하기 위해서는 다음의 두 가지 정보가 필수적이다. 첫 번째는 표본가구와 가구원 각각의 사회 인구학적 특성정보, 두 번째는 사망이나 전출, 조사표 미제출등 표본에서 제외된 사유에 대한 기록이다. 경찰과 도시도시가계의 원자료상에는 위의 두 가지 정보가 불충분했고, 가구관리명부는 원자료에 비해 자료의 용량이 상대적으로 적어 분석이 용이하므로 본 연구에 이용되었다.

2. 패널구축 방법

분석 기간은 2003년에 연동표본이 도입되기 전까지 최장 5년간의 패널구축이 가장 최근 자료인 1998년부터 5년간을 선정했다. 패널구축 방법은 먼저 가구관리명부상의 조사구번호, 구역, 거처, 가구원번호를 연계하여 일련의 개인별 ID를 만들었다. 월별자료들을 병합한 후 같은 ID별로 자료를 정리해서 가구관리패널을 구축했다. 이 가구관리명부패널은 경찰과 도시가계 모두 월별로 조사가 이루어졌기 때문에 원래 총 60회차의 패널차수를 갖는다. 그러나 도시가계의 경우 조사는 월별로 이루어지더라도 결과는 분기별로 발표되고 있기 때문에, 각 분기중에서 응답이 발생한 첫 달을 선정하여 총 20회차의 패널을 구축하였다.

3. 표본소실

이 연구에서 표본의 소실은 1회차에 응답한 표본중에서 응답거절, 전출, 사망등의 사유로 무응답이 발생한 최초의 시점으로 정의하였다. 이렇게 정의할 경우 표본소실율을 과대평가(overestimated)하게 되는 경향이 있다. 실제 조사에서는 한번 무응답이 발생했을지라도 전출여부가 명확하지 않는 한 그 가구는 계속 반복조사하게 되기 때문에 표본에서 완전히 탈락하는 것은 아니다. 이것은 무응답의 형태중 패널조사에서만 나타나는 차수 무응답(wave non-response)의 경우이다. 위의 패널소실의 정의를 따르면 차수무응답이 발생했을 지라도 그 이후의 정보들이 패널자료상에서는 제외된다. 극단적인 예를 들자면 한 가구가 총 60회차중 2회차에 부재등의 사유로 무응답이 발생했지만, 그 이후 58회차를 모두 응답했을지라도, 본 연구에서는 2회차에 표본 소실이 발생한 것으로 간주하고, 그 이후의 정보들은 패널로 구축된 자료상에서 제외시키게 된다. 이렇게 엄격하게 표본소실의 정의하게 된 이유는 1998년부터 5년간의 경찰 및 도시가계자료를 통한 각종 패널구축시 표본소실율의 최하한을 제공하기 위해서이다.

또 다른 사유로 인해서도 본 연구의 표본소실율은 과대측정될 수 있다. 경찰과 도시가계에서 가구원 번호는 첫 번째 조사에서는 가구주와의 관계 순으로 부여되지만, 다음번 조사부터 변동사항이 발생했을 경우 가구원이 신규 진입한 순서대로 부여된다. 이동이 없을 경우 동일한 번호가 부여되지만, 가구원중 일부가 일정기간 외지에 나가 있다가 다시 대상가구로 돌아 왔을 경우라도 원래 가구원번호가 아닌 새 번호를 부여하게 된다. 현재의 가구원번호 체계가 출입(in and out)여부를 자유롭게 반영할 수 없기 때문에 실제로는 동일인에 관한 기록일 수 있지만, 가구번호가 다를 경우 연결할 수 없기 때문에 패널로 구축된 사례에서 제외될 수 있다.

V. 분석결과

1. 가구관리 경찰패널의 표본소실 추이 및 특성

1) 패널표본의 대표성

추출된 표본이 모집단 전체의 특성을 얼마만큼 잘 대표하는가라는 표본대표성의 문제는 횡단조사에서는 최종표본의 추출확률이 동일했는가를 의미한다. 이 대표성의 문제가 패널조사에서는 더욱 복잡해지는데, 그 이유는 모집단 자체가 시간에 따라 변화하기 때문이다. 횡단조사의 목표모집단은 시점에 따라 달라지겠지만, 패널조사의 경우는 기간에 따라 달라지게 된다(Nathan, 1998)

동일한 대상을 일정한 간격을 두고 계속 추적 조사한다는 패널조사의 특성은 또 다른 문제를 야기시킨다. 패널조사 첫 차수의 표본이 대표성을 확보하고 있었더라도, 무응답이 무작위적으로 발생한 것이 아니라면 그 이후에 조사된 표본의 대표성을 보장할 수 없다. 응답자에 관한 기록이 시간에 따라 누적되는 패널자료의 특성상 첫차수의 응답률은 상당히 중요하다. 2회차 이후부터 표본의 소실율은 첫차수의 응답자를 분모로 계산되기 때문이다.

이하에서는 경찰패널과 도시가계패널의 연간 표본소실을 추이와 특성을 각각 살펴보고자 한다. 두 패널자료의 소실추이를 직접적으로 비교하는 것은 해석상에 주의를 요한다. 앞서 언급했듯이 통계청 가구단위 조사에서는 경찰과 도시가계는 동일한 표본프레임을 사용하고, 경찰표본의 일부가 도시가계 표본으로 이용된다. 그러나 경찰표본에서 비농가로 분류된 표본만이 도시가계의 대상이 되기 때문에 표본 표본소실 특성에서 차이를 발생시킬 수 있다.

2) 경찰패널의 표본소실율 추이

<표 2>는 경찰 가구관리명부자료를 바탕으로 가구와 개인표본의 소실율 추

이를 연도별로 살펴본 것이다. 최초표본의 의미는 각 연도 1월의 패널표본의 총 규모를 의미하고, 소실을 당해년도 12월까지 탈락된 사례들의 비율이다. 패널 1차(1998년 1월)에는 총 29,271가구, 93,726명의 가구원이 응답하였다. 이들 중 마지막 60차까지 모두 응답한 가구는 46.5%(13,616가구)이고, 가구원은 36.7%(34,386명)였다. 가구에 비해 가구원의 소실율은 높을 수 밖에 없다. 가구차원에서 한번 무응답이 발생하면 가구원 모두에게 무응답이 발생한 셈이 되지만, 한 가구가 마지막차수 까지 생존표본으로 잔류했을지라도 가구내 특정 가구원의 소실은 발생할 수 있기 때문이다.

연도별 패널소실율 추이를 살펴보면, 첫 1년간의 가구 소실율은 19.5%로 가장 높았고, 시간이 갈수록 점차 감소하다가 3년 이후 부터는 소실율이 일정 수준에서 안정화되는 것을 알 수 있다.

월별로 이루어지는 가구단위 패널조사가 드물기 때문에 직접적으로 비교할 수는 없지만 국내외 주요 패널조사와 비교해 볼 때 경찰패널의 소실율은 높지 않은 것으로 보인다. <참고 2>는 한국노동패널의 연도별 패널소실율 추이이다. 매년 조사가 이루어지는 한국노동패널의 경우, 1998년 5천 가구 패널로 시작해서 지난 6회차까지의 가구 소실율은 61.7%였는데, 이는 신규로 대체된 표본이 감안된 수치이다. 패널 첫 차수에 소실율이 높고, 3년 이후 부터 안정화되는 패턴도 경찰패널과 유사한 면이 있다.

이렇게 패널초기에 높은 소실율과 이후의 안정화 경향은 외국의 조사에서도 발견할 수 있다. 미국의 대표적인 패널조사인 PSID(<참고 3>)는 1968년에 18,191가구를 대상으로 시작되었는데, 20년 후인 1988년에서야 표본잔류율이 50%이하로 떨어질 만큼 상당히 성공적으로 패널이 정착된 조사이다. 이 조사도 첫 차수에 소실율이 11.9%로 가장 높다가 점차 떨어져서, 3년차 이후부터는 안정화되는 것을 알 수 있다²⁾.

2) 초기 차수들의 패널소실율에 영향을 미치는 첫 번째 요인은 가구의 무응답이었고, 두 번째는 이주로 인한 추적불능이다. <참고 3>을 보면 3년차 부터 가구무응답과 함께 이주로 인한 소실율이 안정화되는 것을 알 수 있다. 대부분의 패널조사들이 아주 다양한 방식으로 이주로 인한 추적불능을 줄일 수 있는 통로들을 마련해 놓고 있음을 감안한다면 조사 초기에 이주로 인한 소실은 가구에서 계속 조사에 적극적으로 참여할 의사가 없기 때문인 것으로 해석 할 수 있다.

<표 2>. 경찰패널의 연도별 패널소실을 추이

		가구					
		1998년 1월	1999년 1월	2000년 1월	2001년 1월	2002년 1월	2002년 12월
패널표본 규모		29,271	23,570	19,448	16,902	15,081	13,616
소실율			19.5	17.5	13.1	10.8	9.7
잔류율			80.5	66.4	57.7	51.5	46.5

		개인					
		1998년 1월	1999년 1월	2000년 1월	2001년 1월	2002년 1월	2002년 12월
패널표본 규모		93,726	72,563	57,125	47,221	39,931	34,386
소실율			22.6	21.3	17.3	15.4	13.9
잔류율			77.4	60.9	50.4	42.6	36.7

<참고 2>. 한국노동패널(KLIPS)의 표본 소실을 추이

		가구					
		1998	1999	2000	2001	2002	2003
최초 표본		5,000	4,508	4,266	4,248	4,298	4,592
연간소실율		12.4	11.4	11.0	7.8	7.5	
잔류율							61.7
*신규표본가구			273	485	719	926	

		개인					
		1998	1999	2000	2001	2002	2003
최초표본		13,107	12,537	12,186	12,678	13,264	14,961
연간소실율		16.5	14.3	12.5	8.9	7.6	7.6

출처: 김대일, 남재량, 류근관. 2000. "한국노동패널 표본의 대표성과 패널조사 표본이탈자의 특성연구," 노동경제학논집.

<참고 3>. 미국 Panel Study of Income Dynamics의 패널 소실율과 유형, 1968-72, 1987-89.

연도		1968	1969	1970	1971	1972	1987	1988	1989
잔류율		100.0	88.1	85.0	82.9	80.8	52.2	50.6	49.0
표본 유형	가구 무응답	-	9.9	2.2	1.3	1.3	2.2	1.9	2.3
	사망	-	0.5	0.5	0.6	0.8	0.1	0.1	0.9
	이주	-	1.6	1.1	0.7	0.8	0.5	0.4	0.3

출처: Fitzgerald, John, Peter Gottschalk, and Robert Moffitt. 1998. "An Analysis of Sample Attrition in Panel Data: The Michigan Panel Study of Income Dynamics." *Journal of Human Resources*.

3) 경찰패널의 최초 및 최종잔류표본의 특성

앞서 언급했듯이 패널자료에서 무응답이나 전출, 사망등으로 절단사례들이 발생할 때 이것이 응답자의 인구사회학적 특성에 관계없이 무작위로 발생한다면 패널표본의 대표성은 심각한 문제가 되지 않는다. 패널 1차에 추출된 표본의 특성과 최종 60차 까지 응답을 완료한 표본(잔류표본)의 특성이 얼마만큼의 차이를 보이는가는 패널자료의 대표성을 판단하는 기준이 될 수 있다.

<표 3>. 가구원 특성에 따른 경찰패널의 최초표본과 최종잔류표본 비교

		최초표본(P)	%	최종잔류 (R)	%	증감률
총개인		93,726	100	34,386	100	
성	남	45,632	48.7	15,545	45.2	-7.1
	여	48,094	51.3	18,841	54.8	6.8
연령	15-19	8,286	8.8			
	20-24	6,505	6.9	1,352	3.9	-55.5
	25-29	8,314	8.9	997	2.9	-58.2
	30-34	7,916	8.4	1,373	4.0	-55.0
	34-39	8,919	9.5	2,108	6.1	-27.4
	40-44	8,251	8.8	3,419	9.9	4.5
	45-49	5,615	6.0	3,313	9.6	9.4
	50-55	4,628	4.9	2,925	8.5	42.0
	55-59	4,552	4.9	2,663	7.7	56.8
	60+	10,967	11.7	2,805	8.2	68.0
	65+			5,924	17.2	47.2
교육	초졸이하	39,172	41.8	17,334	50.4	20.6
	중졸	14,165	15.1	5,139	14.9	-1.1
	고졸	25,558	27.3	8,321	24.2	-11.3
	초대졸	6,395	6.8	1,370	4.0	-41.6
	대졸이상	8,436	9.0	2,222	6.5	-28.2
가구주관계	가구주	29,399	31.4	11,807	34.3	9.5
	배우자	21,238	22.7	9,978	29.0	28.1
	미혼자녀	33,872	36.1	9,880	28.7	-20.5
	형제자매	1,550	1.7	208	0.6	-63.4
	부모	4,268	4.6	1,816	5.3	16.0
	조부모	119	0.1	40	0.1	-8.4
	기타인척	2,619	2.8	636	1.8	-33.8
	동거인	661	0.7	21	0.1	-91.3
경제활동상태	취업	39,794	42.5	15,299	44.5	4.8
	실업	1,987	2.1	476	1.4	-34.7
	비경찰	30,890	33.0	11,051	32.1	-2.5
	제외자	21,055	22.5	7,560	22.0	-2.1

가구원 특성

경찰패널에서 최초표본 가구원과 최종잔류표본 가구원의 특성을 성과 연령, 교육, 가구주와의 관계별로 살펴보자. <표 3>을 보면 동일 가구내에서도 가구원의 특성에 따라 표본탈락이 차별적으로 발생하고 있음을 알 수 있다. 여자보다는 남자가 표본소실이 많이 발생했다. 전체 표본에서 남자가 차지하는 비중은 최초표본에 비해 최종표본에서 7.1% 감소한데 비해, 여자의 구성비는 6.8% 증가했다. 연령별로 살펴보면 대체로 연령이 낮을수록 많은 소실이 발생했다. 1998년에 35세 이하 연령층에 속한 집단에서 가장 많은 표본소실이 발생했는데, 20-24세 연령층의 구성비는 58.2%나 감소했다. 35세 이후 연령부터는 점차 표본소실이 줄어들면서 40대 연령층에서는 별다른 변화가 없다가 그 이후 연령부터는 최종표본에 잔류하는 비중이 다른 연령층에 비해 상대적으로 높아지고 있다.

교육정도별로 두 표본간의 분포차이를 살펴보면, 초대졸자의 비중의 감소가 41.6%로 가장 많았고, 다음으로는 대졸자와 고졸자의 순으로 감소하고 있다. 가구주와의 관계별로 최초표본과 최종잔류표본을 살펴보면 가구주와 배우자가 표본에 잔류하는 비중은 상대적으로 높은 반면 형제자매와 미혼자녀의 경우 표본소실이 많이 발생했음을 알 수 있다.

가구특성

<표 4>는 경찰패널의 소실이 가구주의 직업, 거처의 종류등 가구의 특성에 따라 어떤 차이가 있는 지를 살펴본 것이다. 가구주의 직업면에서 보면 거의 대부분의 직업에서 표본소실로 인해 최종표본에서의 구성비가 감소한 반면, 농림어업종사자와 단순무직 근로자는 최초표본에 비해 상대적으로 구성비가 증가했다. 거처의 종류별로 살펴보면 주택거주자보다 주택이외의 거처에서 생활하는 가구에서 표본 소실이 많이 발생했다. 주택에 거주하는 경우에는 단독주택과 아파트의 거주가구 보다는 연립이나 다세대 주택거주 가

구에서 표본소실이 약간 더 많이 발생하고 있다.

통계청 가구단위 조사에서는 경찰 표본의 일부가 도시가계 표본으로도 이용되는데, 조사의 부담 정도가 응답자의 최종잔류 여부에 영향을 미칠 수도 있다. 응답가구의 조사유형에 따라 최초표본과 최종표본을 비교해 본 결과 도시가계와 경찰 모두를 조사하는 가구의 구성비는 -12.4%로 경찰만 조사하는 가구에 비해 더 많이 감소했음을 알 수 있다.

<표4>. 가구특성에 따른 경찰패널의 최초표본과 최종잔류표본 비교

		최초표본		최종잔류 표본		증감률
			%		%	
총가구		29,271	100.0	13,613	100.0	
가구주 직업	입법자/ 고위임직원/관리자	905	3.1	408	3.0	-3.1
	전문가	1,385	4.7	521	3.8	-19.1
	기술공/ 준전문가	2,130	7.3	793	5.8	-19.9
	사무직원	2,731	9.3	1,006	7.4	-20.8
	서비스/ 시장판매 근로자	4,809	16.4	1,897	13.9	-15.2
	농/임/어업종사자	3,468	11.8	2,754	20.2	70.8
	기능관련근로자	3,824	13.1	1,575	11.6	-11.4
	장치/기계조작	3,402	11.6	1,536	11.3	-2.9
	단순노무직 근로자	1,987	6.8	979	7.2	5.9
	군인/ 기타	4,630	15.8	2,144	15.7	-0.4
	거처종류	단독	17,713	60.5	8,289	60.9
아파트		7,782	26.6	3,686	27.1	1.8
연립		2,294	7.8	1,027	7.5	-3.7
다세대주택		524	1.8	238	1.7	-2.3
비주거용주택		933	3.2	366	2.7	-15.7
주택이외의 거처		39	0.1	10	0.1	-44.9
조사유형	도시가계+경찰	8,156	27.9	3,323	24.4	-12.4
	경찰	21,129	72.2	10,293	75.6	4.7

4) 도시가계패널의 표본소실 추이 및 특성

분기별로 발표되는 도시가계자료의 특성을 감안하여 1998년 1월에 응답한 4,612가구중에서 이후 각 분기별로 응답이 발생한 가장 첫 달을 선정하여 20 회차의 도시가계패널이 구축되었다. <표 5>에서 도시가계패널의 5년간 최종 잔류율은 가구차원에서 34.5%인 1,590가구, 가구원별로는 28.1%인 4,818명으로 경찰패널보다 훨씬 낮게 나타나고 있다. 이는 도시가계와 경찰간의 응답

부담의 차이 때문인 것으로 보인다. 도시가계는 한 달 동안의 가계소비 및 지출에 관한 모든 현황을 조사하기 때문에 응답부담이 높은 데 반해, 경활은 일주일간의 경제활동 여부만을 조사하므로 상대적으로 응답이 용이한 편이다.

도시가계패널의 가구단위 연간소실을 패턴은 일반 패널조사들과 마찬가지로 패널초기에 높고 점차 안정화되어가고 있다. 가구의 패널소실이 첫해와 그 이듬해까지 연간 22% 대로 높았고, 3년차부터 19%대로 낮아져서 마지막 해에는 16.3%까지 떨어졌다. 가구원의 연간 소실을 패턴은 가구와 유사하지만 소실율이 상대적으로 더 높아서 최종 60회차 잔류율은 28.1% 였다.

<표 5>. 도시가계패널의 연도별 패널소실을 추이

		가구					
		1998년 1월	1999년 1월	2000년 1월	2001년 1월	2002년 1월	2002년 12월
패널표본 규모		4,612	3,572	2,764	2,233	1,868	1,590
소실율			22.5	22.6	19.2	16.3	14.9
잔류율			77.5	59.9	48.4	40.5	34.5

		개인					
		1998년 1월	1999년 1월	2000년 1월	2001년 1월	2002년 1월	2002년 12월
패널표본 규모		17,158	12,769	9,506	7,360	5,906	4,818
소실율			25.6	25.6	22.6	19.8	18.4
잔류율			74.4	55.4	42.9	34.4	28.1

도시가계패널은 가구단위로 소득과 지출결과가 발표되기 때문에 가구특성에 따른 패널 소실 추이가 중요하다. <표 6>에서 거처의 종류별로 최초표본과 최종잔류표본의 구성비 변화를 살펴보면 비주거용 주택과 주택이외의 거처에 거주하는 가구와 단독주택 거주가구의 패널소실이 두드러진다. 단독주택 거주가구의 경우 최초표본에서는 53.2%를 차지한 반면 최종잔류표본에서는 47.7%로 10.2% 감소했다. 응답가구의 주택소유 형태에 따라서도 최초표본과 최종표본간의 구성비의 차이가 보이는데 주택을 소유하고 있는 가구는 가장 안정적으로 조사된 반면 전월세 거주 가구의 표본소실이 많았기 때문에 비중이 크게 감소했다.

가계의 주된 소득 유형에 따라서도 최종표본 잔류율은 차이가 있었다. 근로소득과 사업소득인 가구의 비중은 최초표본과 최종표본에서 별다른 차이를 보이지 않았다. 연금이나 재산소득이 주된 소득원인 가구가 가장 안정적으로 조사되었다면, 수증이나 보조에 의한 생계를 유지하는 가구의 표본소실율이 높았기 때문에 최종표본으로 잔류하는 비중이 가장 낮았다. 가계형태에 따른 최종표본 잔류율은 별다른 차이가 없었다. 다만 봉급자 가계는 최초표본에 비해 최종잔류 표본에서 차지하는 비중이 다른 가계형태에 비해 유일하게 감소하고 있다.

<표 6>. 가구특성에 따른 도시가계패널의 최초표본과 최종잔류표본 비교

		최초표본(P)	%	최종잔류(R)	%	증감율
총가구		4,612	100.0	1,590	100.0	
거처종류	단독	2,452	53.2	759	47.7	-10.2
	아파트	1,463	31.7	558	35.1	10.6
	연립	459	10.0	182	11.4	15.0
	다세대주택	140	3.0	57	3.6	18.1
	기타(비주거용, 주택이외)	98	2.1	24	1.5	-29.0
주택소유	자가	2,462	53.4	1,291	81.2	52.1
	전세	1,476	32.0	188	11.8	-63.1
	월세	572	12.4	80	5.0	-59.4
	기타	102	2.2	31	1.9	-11.8
주된소득	근로	2,852	61.8	967	60.8	-1.7
	사업	1,232	26.7	439	27.6	3.4
	연금	30	0.7	19	1.2	83.7
	재산	88	1.9	36	2.3	18.7
	수증/보조	244	5.3	69	4.3	-18.0
	기타	166	3.6	60	3.8	4.8
가계구분	봉급자	1,199	26.0	383	24.1	-7.3
	노무자	1,603	34.8	565	35.5	2.2
	근로자외	1,252	27.1	444	27.9	2.9
	무직	558	12.1	198	12.5	2.9

2. 경찰 및 도시가계패널의 생존기간 분석

1) 경찰패널의 가구원 특성별 생존곡선: 생명표 분석

표본의 특성에 따라 패널생존기간은 얼마나 차이가 날까? 패널자료를 분석하는데 주요한 개념은 위험율(hazard rate)이다. 위험율이란 사건발생 위험에 노출되어 있는 집단(위험집합)에 속한 한 개인이 특정 시점 전까지 사건을 경험하게 될 확률이다. 달리 말하자면 이것은 특정한 시간 간격내의 사건발생율(사건 발생 사례 수/ 위험집단에 포함된 사례수)이다. 1에서 위험율을 빼면 사건이 발생하지 않고 생존할 확률을 구할 수 있다.

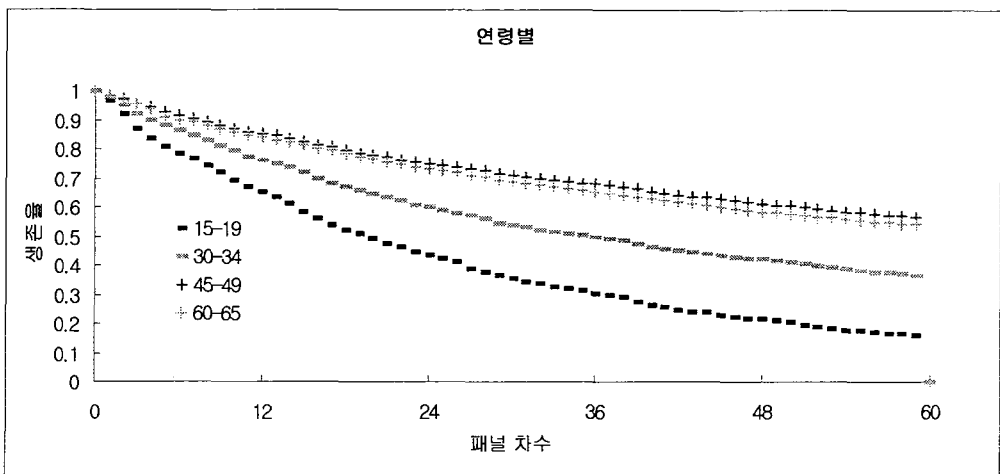
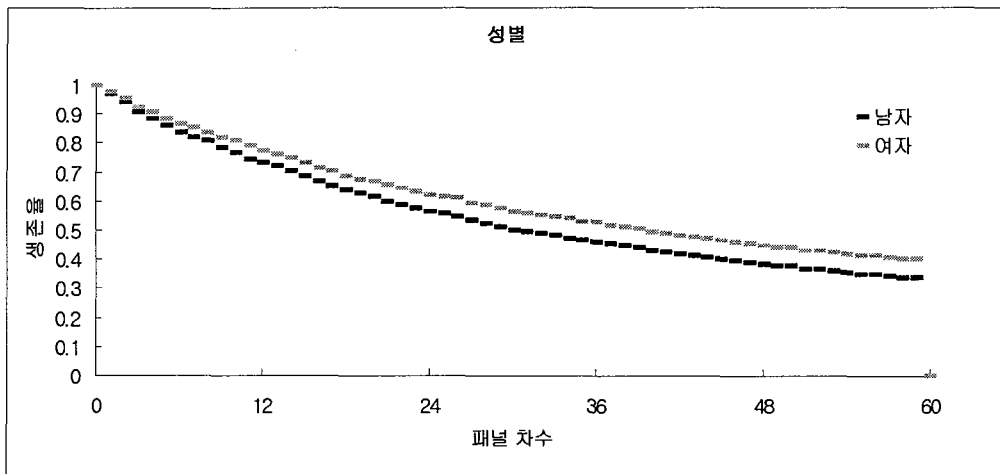
생명표분석 방법을 이용하면 패널표본의 생존 기간별 분포를 추정할 수 있다. 표본의 소실이라는 사건의 발생 시간을 경찰의 경우 한 달이라는 간격으로 도시가계는 3개월간의 간격으로 집합한 후 각 간격의 중간시점에서 누적생존율을 추정할 수 있다. 생명표 방법의 핵심은 일정시점과 그 다음 시점사이의 구간별 사건발생률을 구하고, 이를 통해 일정시점까지 사건이 발생하지 않고 생존할 확률을 구하는 것이다. <그림 1>에서는 경찰패널의 가구원 특성에 따른 생존곡선을, <그림 2>는 도시가계 패널에서 가구특성에 따른 생존곡선을 분석해 놓은 것이다.

경찰패널에서 응답자의 성별에 따른 생존기간 차이를 살펴보면 여자의 생존확률이 남자에 비해 항상 높았던 것을 알 수 있다. 남녀간의 생존확률의 차이는 6회차 이후부터 점차 벌어지기 시작해서 마지막 60차수에서는 6.4%까지 벌어진다. 남자의 경우 패널생존율은 3회차 조사에서 90.6%까지 떨어지고, 12회차에 73%, 30회차가 되면 50%밖엔 되지 않는다. 다시 말해서 경찰조사에 표본으로 선정된 사람이 남자라면 조사가 시작된 달로 부터 3개월 이내에 10명중 1명은 적어도 1회 이상은 조사에 응답하지 않았음을 의미이다.

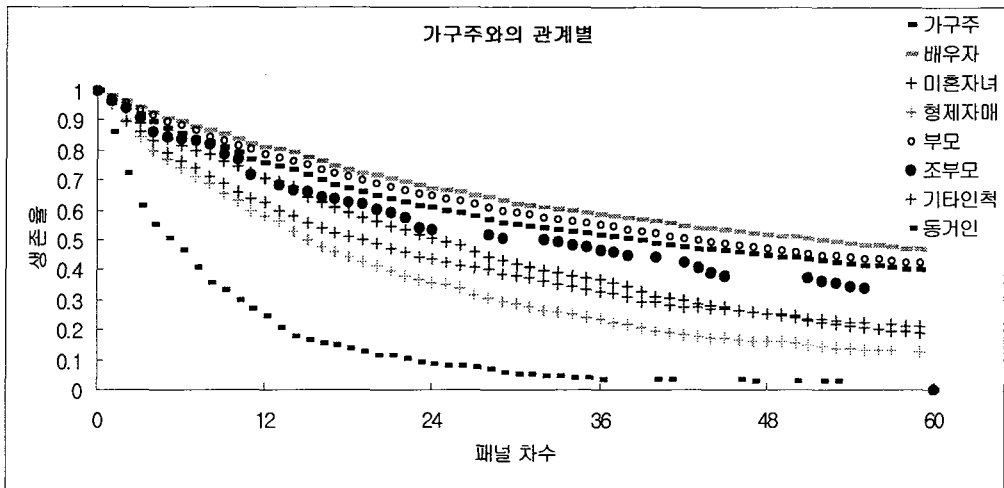
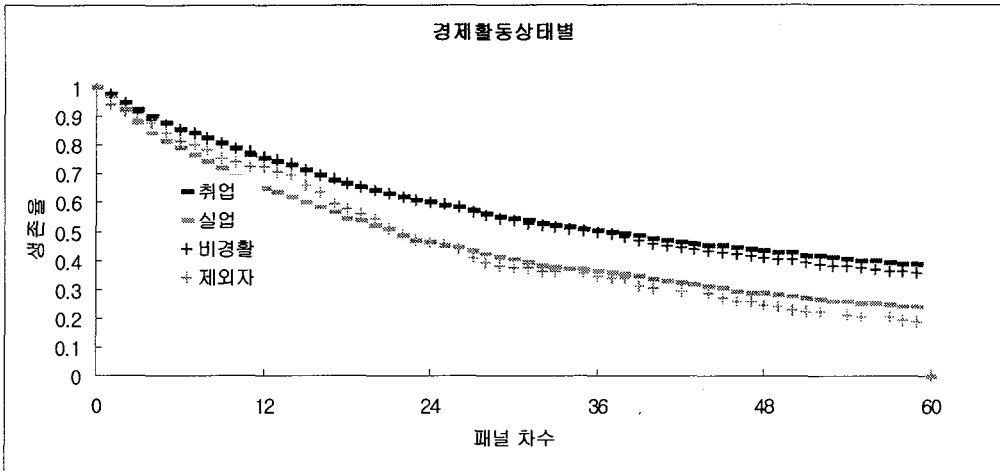
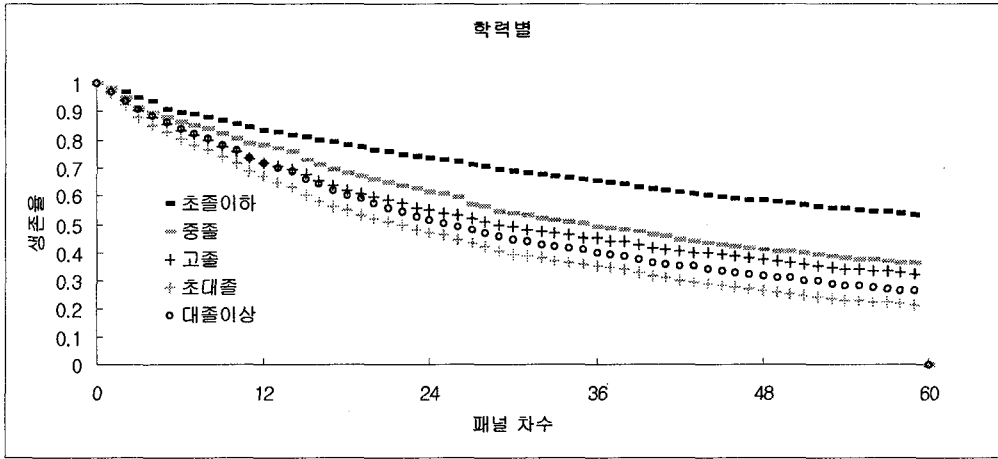
패널의 생존기간의 차이를 설명하는데 있어서 연령은 중요한 변수이다. 5세 계급별로 생존확률을 분석해 본 결과, 20대 전후의 연령대를 제외하고는 이웃한 연령대간에는 생존곡선이 크게 차이 나지 않았기 때문에 분석상의

편의를 위해 본문에는 연령계급을 4개 집단으로 구분한 자료가 수록되었다. 연령과 패널의 생존확률은 정의 상관관계를 나타내는데, 45세 이후의 연령부터는 생존곡선이 매우 유사한 것을 알 수 있다. 15-29세 집단의 생존곡선은 가장 급격히 떨어진다. 생존확률이 5회차에 80%, 20회차에 50%, 최종 60회차에는 16%에 이른다. 이에 비해서 30-34세 집단은 비교적 완만한 편으로 35회차까지는 50%의 생존확률을 보이고 있다. 이에 반해 60세 이상 집단의 경우는 16회차까지도 평균 5명중 1명 정도만 무응답을 일으킬 가능성이 있기 때문이며, 마지막 차수까지도 평균 2명중 1명만 무응답이 있었음을 알 수 있다.

<그림1>. 경찰패널의 가구원 특성별 생존곡선



<그림1>. 경찰패널의 가구원 특성별 생존곡선 (계속)



생존기간은 학력에 따라서도 차이가 있다. 패널의 생존확률과 학력은 연령과는 달리 전체적으로는 부의 상관관계를 보인다. 그러나 가장 생존확률이 낮은 집단은 대졸이상자가 아니라 초대졸자였다. 이것은 부분적으로는 연령효과로 인한 것으로 보이는데, 뒤에 소개될 콕스비례위험모형에서 연령효과가 통제되자 생존확률이 가장 낮은 집단은 대졸자임이 드러났다. 초졸이하의 학력소지자들은 6회차까지 90%이상의 생존확률을 보이고, 44회차에 60%, 최종 60차에 53%의 생존확률을 보인다. 이에 반해 대졸자의 생존확률은 4회차에 이미 88%까지 떨어지고, 18회차에는 60%, 최종 60회차에는 26%만이 생존하게 된다.

응답자의 경제활동상태 또한 패널의 생존기관과 밀접한 관련이 있다. <그림 1>의 경제활동상태별 생존곡선을 보면 취업자와 비경제활동자의 생존곡선의 패턴이 유사하고, 실업자와 제외자(군인, 집단시설 수감자등)의 생존곡선이 유사하다는 점을 알 수 있다. 12회차 이후 부터는 경제활동상태에 따른 집단간의 차이가 확연히 들어 나기 시작한다. 취업자의 경우 22회차까지의 생존확률이 90.3%인 반면, 실업자의 경우는 79%로 낮았다. 39회차에도 전자의 생존확률은 80.2%인 반면, 후자는 62.9%까지 낮아져 있다.

취업자의 생존확률은 60회차 모두 항상 높는데 반해, 실업자의 생존율은 처음에는 가장 낮지만 2년차 이후부터는 제외자 보다 높게 나타난다. 이러한 변화의 원인은 경제활동상태 변수가 시간에 따라 변화하는 변수(time-dependent covariate)인데 반해 이러한 특성이 생명표 분석에서 반영되지 않았기 때문이다. 따라서 실업자와 제외자의 생존곡선상의 교차는 실제로 두 집단의 생존확률이 변화했다기 보다는 실업자가 2년 정도 후에는 다른 경제활동상태로 변화를 경험하기 때문인 것으로 보이는데, 이는 별도의 검증이 이루어져야 명확히 설명될 수 있을 것이다.

가구주와의 관계 또한 생존기간의 차이가 있는데, 조부모의 생존곡선이 가장 완만한 반면, 기타인척의 생존곡선이 가장 급격하게 떨어졌다. 43회차까지도 조부모가 표본으로 생존할 확률은 90.2%인 반면, 기타친인척은 59% 밖

엔 되지 않았다. 생존확률의 차이는 있지만 생존곡선의 패턴이 서로 유사한 집단들이 있다. 가구주와 배우자의 곡선이 유사하고, 형제자매는 기타인척과 부모와 미혼자녀의 패턴이 유사한 것을 알 수 있다.

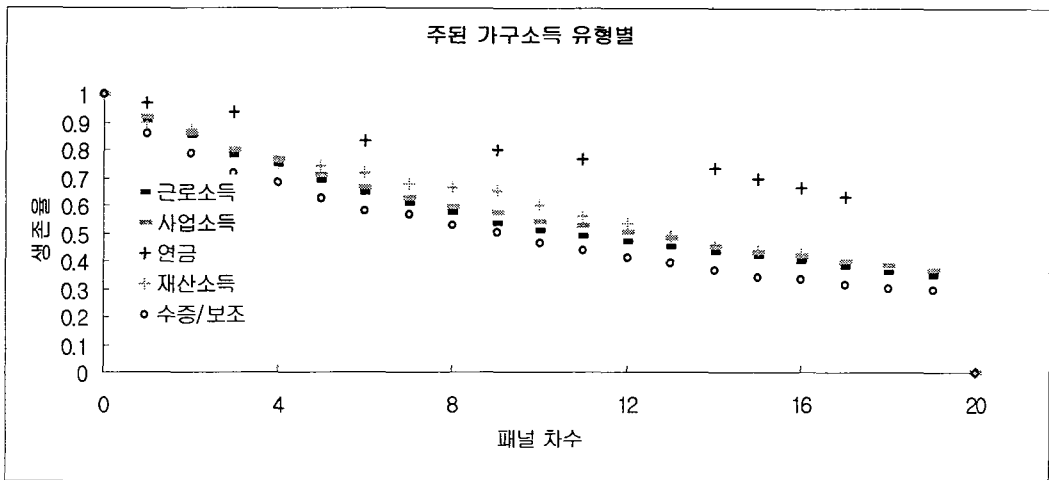
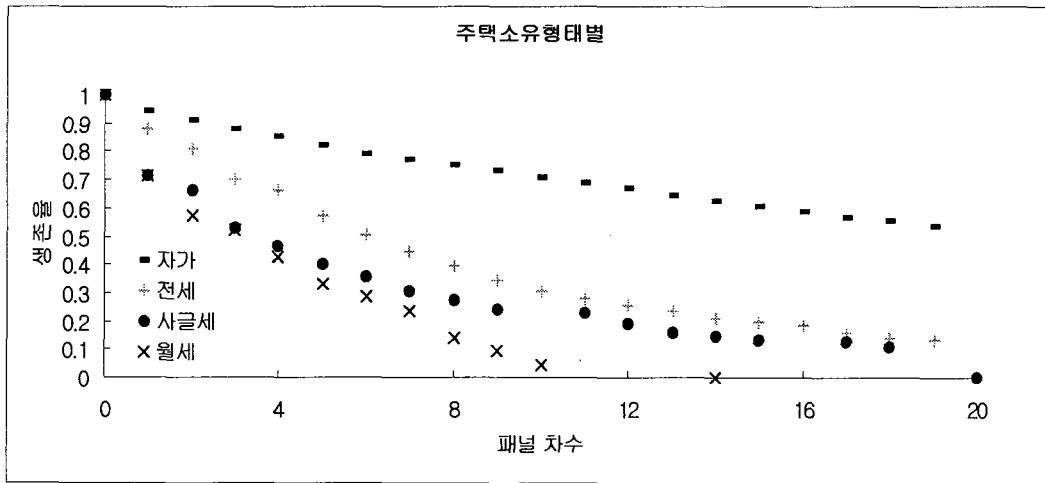
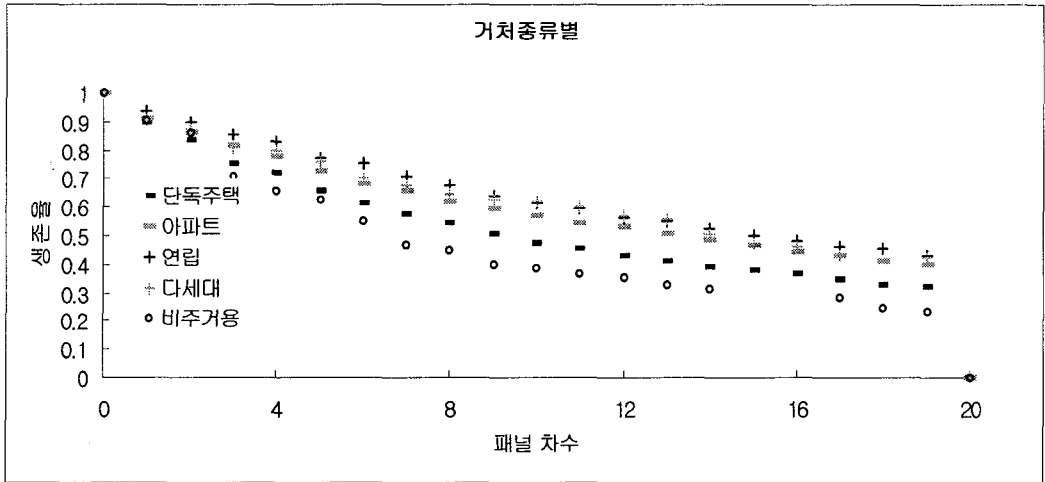
경찰응답자가 도시가계도 같이 응답하는 경우 패널생존확률에 유의미한 영향을 주고 있다. 두개의 조사를 같이 응답할 경우 패널 1년차 부터 경찰만 응답하는 가구와 차이가 벌어지기 시작하면서, 2년뒤인 25회차에는 경찰과 도시가계 응답가구의 생존확률이 88.2%인 반면, 경찰만 응답하는 가구는 90.8%를 유지하고 있다.

2) 도시가계패널의 가구특성별 생존곡선

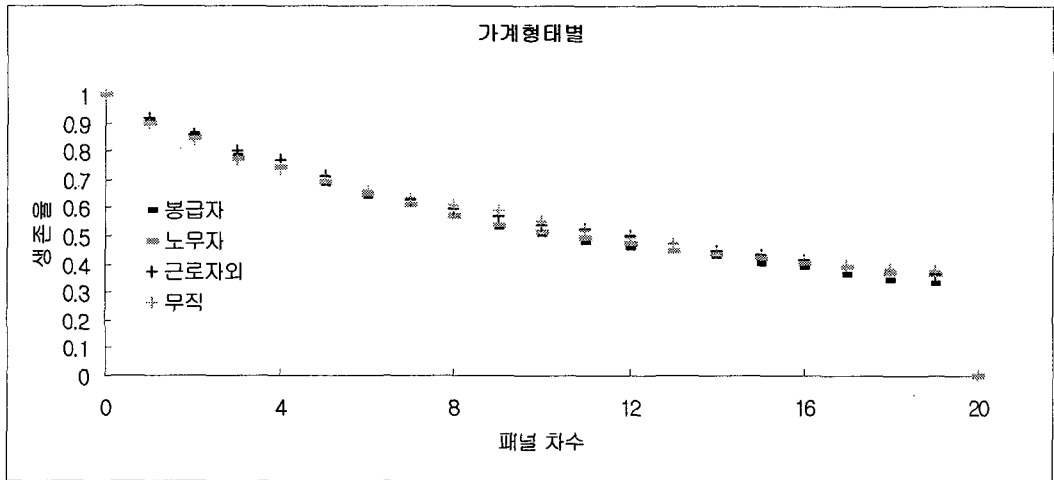
도시가계 패널은 가구특성이 주요한 변수이기 때문에, 거처의 종류, 주택 소유형태, 주된 가구소득 유형별로 생존곡선의 차이를 살펴보았다. <그림 2>를 보면 주택의 소유여부가 패널의 생존기간에 유의미한 차이를 주고 있었다. 가장 생존기간이 긴 집단은 연립주택거주 가구였는데, 첫회차 조사로부터 1년 후인 4회차 까지도 82.6%의 생존확률을 보였고, 최종 20회차에도 43.2%가 생존했다. 이에 반해서 단독주택과 비주거용주택소유자의 생존확률이 낮은 편이다. 2회차 조사까지는 단독주택의 생존확률이 83%로 가장 빨리 떨어졌지만 이후부터는 비주거용주택에 거주하는 가구의 생존확률이 더 빨리 떨어져서 이 집단의 생존확률은 조사 시작 후 1년 반이 지난 6회차에는 55% 정도 밖에 되지 않았다.

주택소유형태별로 생존곡선을 살펴보면, 자가주택소유, 전세, 사글세와 월세 순으로 생존확률이 높았다. 자가주택 소유가구의 생존확률은 2회차까지 90%를 유지했고, 최종 20회차 까지 53.7%가 생존했다. 이에 반해 전세 거주 가구는 2회차에 80%, 20회차에 13.1% 생존했고, 월세의 경우는 8회차에 이미 생존확률이 14%밖에 되지 않았다.

<그림2>. 도시가게패널의 가구 특성별 생존곡선



<그림2>. 도시가계패널의 가구 특성별 생존곡선 (계속)



가구의 주된소득 유형에 따라서도 도시가계패널의 생존확률은 차이가 났다. 연금소득 가구는 다른 가구와 달리 생존확률이 높았다. 그 외의 집단의 생존확률은 재산소득, 근로소득, 사업소득, 수증과 보조 가구순으로 낮아지지만, 생존곡선의 패턴이 서로 유사한 것을 알 수 있다.

경찰응답자가 도시가계도 같이 응답하는 경우 패널생존확률에 유의미한 영향을 주고 있다. 두개의 조사를 같이 응답할 경우 패널 1년차 부터 경찰만 응답하는 가구와 차이가 벌어지기 시작하면서, 2년뒤인 25회차에는 경찰과 도시가계 응답가구의 생존확률이 88.2%인 반면, 경찰만 응답하는 가구는 90.8%를 유지하고 있다.

3. 경찰 및 도시가계패널의 생존기간 결정요인 분석: 콕스비례위험모형

지금까지는 가구원 특성 변수별 경찰패널의 생존기간과 가구특성에 따른 도시가계패널의 생존기간 차이를 살펴보았다. 그러나 패널의 생존기간은 단일한 설명변수에 의해서 결정되는 현상이 아니다. 연령과 학력정도와 같이 설명변수들 간에도 상관관계가 존재하기 때문에 다른 설명변수의 값을 통제

한 상태에서 특정 설명변수가 표본의 생존기간이라는 종속변수에 독립적으로 행사하는 영향력을 회귀모형을 통해서 측정해 볼 필요가 있다.

본 연구에서는 가구원과 가구차원의 설명변수들이 경찰과 도시가게 패널의 생존기간에 미치는 상대적인 효과를 알아보기 위해서 콕스 비례위험 모형을 이용하였다. 한 개인이 특정시점에서 패널소실을 경험할 확률은 다음의 방정식으로 표현된다.

$$\log h_i(t) = \log h_0(t) + \beta_1 x_{i1} + \dots + \beta_k x_{ik}$$

이 때 $h_0(t)$ 는 설명변수들의 값에 의해서 변화하는 시간 t 에서의 기저위험 (baseline hazard)이다. 콕스의 비례위험 모형에서 회귀계수의 해석은 설명변수가 한 단위씩 증가함에 따른 종속변수의 위험도(hazard ratio) 백분율 변화량으로 해석된다.

<표7>은 콕스비례위험 모형을 이용하여 경찰패널에서 표본소실과 관련된 개인차원의 요인을 분석한 결과이다. 분석에 포함된 성, 연령, 교육정도, 가구주와의 관계, 경제활동 상태 변수 모두가 패널소실 위험에 통계적으로 유의미한 영향력을 행사하고 있음을 알 수 있다. 특히 연령에 의한 효과가 가장 두드러지는데, 연령이 증가할수록 패널소실 확률이 감소하고 있다. 예를 들어 30세 이하의 집단에 비해 30-44세 집단은 패널 차수가 한 단위씩 높아질 때마다 패널소실확률이 42.2%(1 - .578= .422) 감소하고 있다.

그러나 교육과 가구주와의 관계에서 각 항목집단별 패널소실위험은 앞의 생명표분석 결과와는 차이가 나는데, 이것은 다른 설명변수의 값들이 통제된 상태에서 한 변수가 표본의 생존기간에 독립적으로 행사하는 영향력을 보여주기 때문이다. 예를 들어 생명표분석에서는 학력이 높아지면서 생존확률이 점차 낮아져서 초대졸자에서 가장 낮고, 대졸자는 되려 높아지는 U자 형태를 보였다. 반면에 콕스비례위험모형에서는 연령효과가 통제되자 교육정도와

패널소실위험이 정의 상관관계를 갖고 있다는 점이 드러났다. 초졸이하의 학력소지자에 비해 대졸이상의 학력소지자의 패널소실위험은 패널차수가 한 단위 증가할 때 마다 1.28배씩 증가하고 있다.

가구주와의 관계에 따라 패널소실위험을 살펴보면, 가구주로 부터 관계가 멀수록 위험도가 높아지는 것을 알 수 있다. 생명표분석결과에는 배우자, 부모, 가구주의 순으로 생존확률이 높은 반면에 미혼자녀의 생존확률은 매우 낮았다. 그러나 콕스비례위험모형의 결과를 보면 다른 변수들을 통제할 때 가구주에 비해 미혼자녀의 패널소실위험은 오히려 가장 낮은 77.1% 수준이었다. 부모의 경우 또한 패널소실위험이 가구주의 124%에 해당되었다. 동거인의 패널 소실 위험은 가구주에 비해 3배가 넘는 것으로 나타났다.

<표7>. 경찰패널 생존기간의 가구원 차원 비례위험도

변수	β
성 (남)	
여	.933**
연령 (15-29)	
30-44	.578**
45-59	.443**
60+	.434**
교육 (초졸이하)	
중졸	1.097**
고졸	1.117**
초대졸	1.177*
대졸이상	1.282**
가구주관계 (가구주)	
배우자	.889**
미혼자녀	.771**
형제자매	1.071*
부모	1.249**
조부모	1.500**
기타인척	1.025
동거인	3.031**
경제활동상태 (취업)	
실업	1.134**
비경찰	1.063**
제외자	1.058
Chi-Square	7,894.5
DF	18
N	72,953

경제활동상태에 따른 패널소실위험을 살펴보면 취업자가 다른 집단에 비해 가장 위험률이 낮았다. 취업자에 비해 실업자의 상대적 위험도가 가장 높았고, 그 다음은 비경활자였다.

패널소실의 위험도를 설명하는데 있어서 가장 직접적인 요인은 가구차원 보다는 개인차원의 변수들이다. 이는 <표8>에서 보듯이 도시가계패널의 소실위험도를 설명하는데 있어 대부분의 가구차원 변수들이 가진 효과가 통계적으로 그다지 유의미하지 않은 점을 통해서도 알 수 있다.

<표8>. 도시가계패널 생존기간의 비례위험도

변수	β
거처종류	
(단독주택)	
아파트	1.02
연립	.957
다세대주택	.995
비주거용	1.099
기타	2.893
주택소유	
(자가)	
무상	1.512*
사택	1.146*
전세	2.205**
보증부월세	2.155**
사글세	2.841**
기타	4.888**
주된소득	
(근로)	
사업	.951
연금	.880
재산	1.209
수증/보조	1.274
기타	.989
가계구분	
(봉급자)	
노무자	.891*
근로자외	.971
무직	.840
Chi-Square	594.5
DF	19
N	4108

그러나 주택소유여부는 패널소실의 위험을 설명하는데 있어서 상당히 큰 설명력이 있는 변수임을 알 수 있다. 주택을 소유한 가구의 패널소실위험이 가장 낮았고, 세들어 사는 가구는 상대적으로 위험이 높았다. 주택을 소유한 가구에 비해 전세로 거주하는 가구는 패널 차수가 한 단위씩 높아질 때마다 패널소실확률이 2.2배 만큼 증가하고 있다. 사글세의 경우 패널소실 확률이 2.8배가 넘었고, 기타 형태의 경우는 패널소실위험이 무려 4.8배에 달하고 있다. 가계구분을 보면 봉급자가계에 비해 노무자가계의 낮은 위험도만이 통계적으로 유의미했다.

VI. 결론 및 제언

최근 국내에서도 사회적인 현상, 개인이나 집단의 특성이 시간에 따라 어떻게 변화하는지를 측정하고자 하는 연구들을 종종 찾아 볼 수 있다. 패널조사는 적어도 두 시점 이상의 기간 동안 동일한 대상을 상대로 같은 변수나 항목이 조사되는 것을 말한다. 이렇게 패널자료에 대한 수요가 증가하는 이유는 기존의 횡단조사가 특정 현상의 존재 여부나 변수들간의 현재적인 관계만을 파악하는 것과는 달리, 한 요인이 특정 현상에 미친 인과적인 효과를 직접적으로 평가해볼 수 있기 때문이다 (Ruspini, 2002).

60년대 미국에서 처음 시작된 패널조사는 종단자료 분석기법의 발달과 함께 현재 20여 개국에서 다양한 조사목적과 조사방법으로 실시되고 있다. 패널조사는 대학이나 연구기관이 주로 주관하는 경우들도 많지만, 국가통계기관에서 패널조사를 실시하는 경우들도 있다. 가구소득조사를 위한 캐나다의 Survey of Labor and Income Dynamics, 프랑스 Lorraine 지역 거주자들의 특성을 연구하는 French Household Panel, 개인과 가구의 사회경제적인 특성을 조사하는 Dutch Socio Economic Panel 등은 국가통계기관이 주관하는 대표적인 패널조사들이다.

패널자료가 가진 폭넓은 분석 가능성에도 불구하고, 동일대상을 지속적으로 추적할 때 발생하는 비용과 표본설계 및 조사과정의 복잡성은 패널조사를 개발하는데 가장 큰 장애요인으로 작용하고 있다. 그러나 반복 횡단조사로 설계된 자료를 가구 및 개인의 ID를 기준으로 상이한 시점들을 연계해서 사용할 수 있다면 별도의 비용 없이도 패널자료를 이용할 수 있을 것이다. 경찰과 도시가계의 경우 인구주택총조사에 의해 표본프레임이 확정된 후 특정가구가 표본으로 추출되면 다음번 표본개편 전까지는 동일한 대상을 상대로 반복적으로 조사가 실시된다. 따라서 조사결과를 최장 5년, 2003년 연동 표본이 도입된 이후에는 최장 3년간의 패널자료로 구축해 볼 수 있다.

본 연구에서는 1998년에서 2002년까지 경찰자료를 이용하여 60회차의 월별 패널을, 도시가계자료는 20회차의 분기별 패널로 구축한 후, 두 패널자료가 가진 표본대표성을 분석해 보았다. 경찰패널의 최종 표본소실율은 53.5%, 도시가계패널은 63.3%였다. 연도별 패널소실을 추이는 두 조사 모두 비슷한데, 첫해의 소실이 가장 많이 발생하고, 점차 감소하다가 3년 이후 부터는 안정화되는 것을 알 수 있다.

경찰과 도시가계 패널의 생존기간의 차이를 살펴보면, 패널의 소실에 특정한 유형이 있다는 것을 알 수 있다. 경찰의 경우 여자보다는 남자가 패널소실위험률이 높았고, 장년층보다는 젊은층이 패널소실 위험률이 높았다. 학력이 높을수록 패널소실확률도 함께 증가했으며, 취업자보다는 실업자의 패널소실확률이 높게 나타났다. 도시가계의 경우 세 들어 사는 가구의 패널소실 위험률이 주택을 소유한 가구에 비해 높게 나타났고, 노무자가계보다 봉급자가계가 패널소실위험이 더 높았다.

개인차원에서 패널소실기간의 차이를 설명하는 가장 효과적인 변수는 연령이었던 반면에, 가구 차원에서는 주택소유형태였다. 이러한 결과는 Clarke and Tate(1999)의 연구와도 일치한다. 이들은 횡단조사로 설계된 영국의 Labour Force Survey 자료를 이용해서 최장 12개월간의 패널자료를 구축한 후 표본소실의 발생하고 있는 유형을 분석한 결과 연령과 주택소유형태가 가장 중요한 변수였다고 주장했다. 정리하자면 패널소실은 주로 지리적 이동성이 높은 젊은 연령층과 주거상태가 상대적으로 불안정한 가구에서 주로 발생한다는 것을 알 수 있다. 또한 응답자 부담도 하나의 변수로 작용해서 경찰과 도시가계를 모두 응답하는 가구의 경우 패널생존확률이 낮았다.

이 상의 연구결과를 통해 경찰과 도시가계자료에서 패널의 소실이 무작위적으로 발생하고 있는 것은 아님을 알 수 있다. 그러나 5년간의 패널기간 중 경찰의 경우 총 60회차, 도시가계의 경우 20회차라는 패널차수를 고려해 볼 때 전체적인 표본소실의 수준이 표본의 대표성을 크게 위협할 정도로 심각한 것은 아니라고 판단된다. 그 이유는 본 연구에서 차수무응답까지도 패널

소실에 포함시키는 엄격한 기준을 적용했음에도 경찰의 경우 최초패널의 과반수 가까이 60회차 까지, 도시가계패널의 경우 최초패널의 1/3이상이 20회차 조사까지 생존했기 때문이다.

경찰과 도시가계자료를 패널자료로 이용하는데 있어서 다음의 두 가지 사항을 주의해야 할 것으로 보인다. 첫 번째는 경찰과 도시가계가 일종의 거처패널이기 때문에 봄과 가을의 이사철에 무응답이 증가로 인해 패널소실에 계절성이 나타날 것이라는 점이다. 전국규모의 조사들이 인구이동이 적은 달을 선정해 조사를 진행하는 이유도 이러한 계절성이 고려된 것이다. 하지만 사후적으로 패널로 구축된 경찰과 도시가계자료는 이러한 계절성을 조절할 수가 없기 때문에, 자료 분석에 어떠한 영향을 미치게 될지는 별도의 연구가 필요한 사안이다.

두 번째는 가중치문제이다. 횡단조사로 설계된 경찰과 도시가계조사에서는 무응답이나 사후층화로 인해서 표본의 추출확률이 불균등 해지는 것을 보정하기 위해서 사후에 가중치를 조정하고 있다. 두 조사를 패널로 구축하게 되면 가중치 조정문제가 더욱 복잡해진다. 예를 들어 표본소실이 위의 분석결과와 같이 연령과 주택소유 유형에 따라서 상이하게 발생하고 있다고 하자. 연령은 기존의 가중치 부여시 성과 지역과 함께 통제변수로 사용되고 있던 변수이다. 그러나 주택소유유형은 통제변수에 포함되지 않았다. 가중치를 부여해서 조정하기 위해서는 연령과 주택소유 유형분포에 관한 모수값을 알아야만 한다. 제 1차 조사가 실시된 시점의 주택소유 유형분포는 알 수 있겠지만, 모수값은 시간에 따라 변동하므로 제 60차의 조사시점의 모수값은 알 수가 없다. 이 때 제 1차 조사에서 사용된 가중치를 60차 조사까지 계속 사용할 경우 기존의 횡단설계의 가중치 체계를 통해 발표된 결과치와 패널자료를 통해 분석된 결과치가 다를 수 있다는 점이다. 따라서 패널자료 구축시 어떠한 방식으로 무응답에 대한 가중치를 조정해야 공표했던 결과치와 유사한 결과를 도출할 수 있을지는 별도의 심층적인 연구를 필요로 하는 사안이다. 현재로서는 조사된 원래의 값과 가중치가 부여된 값을 동시에 제공해야

할 것으로 보인다.

세 번째는 무응답을 처리하는 일관된 기준과 이에 대한 상세한 설명이 필요할 것으로 보인다. 예를 들어, 도시가계에서는 단위무응답으로 인해 발생하는 대표성의 문제를 완화시키기 위해서 응답가구의 값을 복제하여 일정 규모의 표본규모를 유지하는 방식을 사용해 왔다. 문제는 같은 표본주기 내에서도 복제방식이 다르다는 데 있다. 1998년과 1999년에는 응답한 가구의 항목별 응답값 뿐만 아니라 가구일련번호까지 모두 무응답가구 자료에 복제했기 때문에 가구일련번호가 중복되는 사례들이 있었다. 이러한 문제점을 보완하기 위하여 2000년부터는 응답가구의 값만 무응답 가구자료에 복제하고 무응답 가구가 가진 원래 일련번호는 유지시켰다. 따라서 도시가계자료를 제공할 때는 이러한 기준변화에 대한 설명과 함께 원래 값과 복제된 값을 데이터상에 표기해주어야만 할 것이다.

이상으로 경찰과 도시가계자료를 패널로 구축하고, 분석하는 과정에서 해결해야만 할 과제들을 살펴보았다. 그러나, 패널조사 자료를 통해 얻을 수 있는 풍부한 분석의 가능성과 정책이나 프로그램의 효과를 객관적으로 측정할 수 있는 수단으로서 패널자료에 대한 수요는 앞으로 점점 더 확대될 전망이다. 또한, 기존의 자료에 대한 새로운 이용수요를 창출한다는 면에서도 패널자료 구축에 대한 연구는 지속적으로 수행되어야 할 것으로 보인다.

참 고 문 헌

- Clarke, Paul and Pam Tate. 1999. "Methodological Issues in the Production and Analysis of Longitudinal Data from the Labour Force Survey" GSS Methodology Series No. 17. Office for National Statistics. U.K..
- Davies, Richard B. 1994. "From Cross-Sectional to Longitudinal Analysis", in *Analysing Social and Political Change*. Edited by Angela Dale and Richard B. Davies. London: Sage.
- Diggle, Peter J., Patrick J. Heagerty, Kung-Yee Liang, and Scott Zeger. 2002. *Analysis of Longitudinal Data*. NY: Oxford University Press.
- Fitzgerald, John, Peter Gottschalk, and Robert Moffitt. 1998. "An Analysis of Sample Attrition in Panel Data: The Michigan Panel Study of Income Dynamics." *Journal of Human Resources* 33:251-299.
- Gong, Greg X. 2004. "Planning and Design of Household Panel Surveys for Enhancing Poverty Statistics." Paper presented at the UNESCAP Subcommittee on Statistics, 1st session, Bangkok. Thailand.
- Kuzmich, Geoff and Wigbout Max 2001. "A Longitudinal Look at Some Data of the Household Labour Force Survey." Research Report 16. Statistics New Zealand.
- Laux, Richard and Emma Tonks. 1998. "Longitudinal Data from the Labour Force Survey." *Labour Market Trends* 106: 175-188.
- Menard, Scott. 2002. *Longitudinal Research*, 2nd Edition. London: Sage.
- Nathan, Gad. 1998. "A Review of Sample Attrition and Representativeness in Three Longitudinal Surveys," GSS Methodology Series No.13. Office for National Statistics. U.K..
- Rowe, Geoff and Huan Nguyen. 2004. "Longitudinal Analysis of Labour Force Survey Data." *Survey Methodology* 30:105-114.
- Ruspini, Elisabetta. 2002. *Introduction to Longitudinal Research*, London: Routledge.

- 김대일, 남재량, 류근관. 2000. “한국노동패널 표본의 대표성과 패널조사 표본 이탈자의 특성연구,” 「노동경제논집」 23:1-33
- 남재량, 1997. “우리나라 실업률 추세변화에 관한 연구,” 서울대학교 경제학박사 학위논문.
- 남재량, 류근관. 1999. “우리나라 여성 노동력 상태의 동태적 특성연구,” 「사회과학과 정책연구」 21:115-159.
- 대우경제연구소. 1999. 「대우패널데이터」. 대우경제연구소.
- 류재우·배무기. 1984. “한국의 노동시장 플로우와 실업” 「노동경제논집」 10: 55-75.
- 이병희·정재호. 2002. “경제위기 이후의 빈곤구조 분석: 반복빈곤 및 고용과의 관계를 중심으로.” 「동향과 전망」 52:128-150.
- 한국노동연구원. 2005. 「6차년도 패널 보고서: 한국 가구와 개인의 경제활동」. 한국노동연구원.
- 황덕순. 2002. “빈곤에 대한 동태적 분석.” 「소득불평등 및 빈곤의 실태와 정책 과제」. 한국노동연구원.

부 록

<부표 1>. 해외 주요 패널조사 현황

국가	조사명	내용	
호주	Household, Income and Labor Dynamics in Australia (HILDA)	주관 조사연도 표본특성 조사내용	University of Melbourne, Melbourne Institute of Applied Economic and Social Research 2001 호주전역의 488개 지역별로 총7,682 가구 추출 소득, 노동시장, 가족
	General Customer Survey	주관 조사연도 표본특성 조사내용	Department of Family and Community Services in Australia (FaCS); the Commonwealth 에서 자료수집 2000, 2001, 2002 Family and Community Services의 행정자료를 이용해서 소득지원 대상자 중 전국대표 표본을 선출. 샘플링은 소득지원유형에 따라 일차 사회보장 및 가족지원 수혜자 관련 정보: 인구학적 특성, 가족 및 가구, 교육, 어린이와 양육, 고용, 퇴직, 장애, 지역및 응급서비스 사용,
벨기에	Belgian Socio-economic Panel (SEP)	주관 조사연도 표본특성 조사내용	조사컨셉, 자료처리 및 분석은 Centre for Social Policy, University of Antwerp (UFSIA). Dimarso (Gallup Belgium)에서 조사수행 1985, 1988, 1992, 1997 전국대표 표본으로 1985년에 6,471가구, 1998에 3800가구, 1992년 38000(신규 900개 가구포함), 1997년에 4,632가구(신규2,375 가구 포 인구학적 특성, 활동상태, 고용상태, 가구의 월별소득, 가구소득, 주
캐나다	Survey of Labor Income Dynamics (SLID)	주관 조사연도 표본특성 조사내용	Statistics Canada 1993-2000 10개 지역별 총 35,000가구를 추출하여 동일인을 계속 추적조사하도 개인과 가족의 경제적 웰빙과 변화요인. 교육과 장애등 관련 항목도
중국	Chinese Longitudinal Healthy Longevity Survey (CLHLS)	주관 조사연도 표본특성 조사내용	National Archive of Computerized Data on Aging (NACDA) 1998-2000 중국 22개 도시지역과 지방의 대부분의 노인을 대상으로 건강과 장수의 결정요인에 관한 연구를 목적으로 설계. 노인관련 연구중 세계 최 노인의 건강, 사회경제적특성, 가족, 라이프스타일, 인구특성 프로하
프랑스	French Household Panel	주관 조사연도 표본특성 조사내용	ADEPS, INSEE 1985-1990 Lorraine 지역 거주자 (집단가구제외)로 첫차수에 715가구로 출발하 가구구성 및 인구학적 특성, 주택, 소득, 교육, 고용, 생애사건등의 표
서독	Sozio-oekonomisches Panel (SOEP)/German Socio Economic Panel (GSOEP)	주관 조사연도 표본특성 조사내용	DIW Berlin (German Institute for Economic Research) 1984- 개인가구를 대상으로 전국대표 표본으로 구성된 종단조사. 1998년에 6,600 가구 이상, 12,7000명을 표본으로 추출 가족구성, 직업사, 취업, 수입, 건강및 만족지표
헝가리	Hungarian Household Panel	주관 조사연도 표본특성 조사내용	Social Research Informatics Centre (TARKI), Sociology Department of Budapest University of Economics, Hungarian Central Statistical Office 1992-1996 2,600 헝가리인 가구 사회계층, 자산, 소득, 가구의 경제 및 재정전략, 인구학적 특성, 고용
인도네시아	Indonesia Family Life Survey	주관 조사연도 표본특성 조사내용	Demographic Institute of the University of Indonesia, the Population Studies Center of Gadjah Mada University, University of California--Los Angeles, Michigan State University 공동주관 1993/94, 1997/98, 2000 전국 26개 도지역중 13개 도에 살고 있는 30,000명이상으로 전인도네 개인 및 가구수준의 자산, 출산율, 건강상태, 가구원 관계, 가구의사결정, 세대간 이동성등

<부표 1>. 해외 주요 패널조사 현황(계속)

국가	조사명	내용	
일본	Japan Panel Study on Consumers (JPSC)	주관 조사연도 표본특성 조사내용	The Institute for Research on Household Economics 1994-2000 코호트 A는 1993년에 전국대표 표본 24-34세 여성 1,500명; 코호트 B 가구구성, 노동시장행태, 소득, 소비, 저축, 자산, 부채, 주택, 가정관
룩셈부르 그	Panel Socio-Economique "Liewen zu Lëtzebuerg" (PSELL)	주관 조사연도 표본특성 조사내용	CEPS/INSTEAD: A Centre for Socio-Economic Research 1985- PSELL I (1985-1994) 은 대표표본 2,012 가구(6,110 명); PSELL II 주택, 가족구성, 주거, 소비재, 가구의 일반적인 예산관리 및 문제점,
멕시코	Mexican Family Life Survey (MxFLS)	주관 조사연도 표본특성 조사내용	Department of Economics at the Universidad Iberoamericana (UIA); the Division of Economics at the Centro de Investigación y Docencia Econó 2001 and 2004 상이한 지역에서 추출된 8,000 가구를 기준으로 2001년 (MxFLS1) 과 2004년 (MxFLS2) 두 시점에서 각각 개인과 지역공동체에 관한 정 소득, 소비, 자산, 건강, 영양, 출산, 산전관리, 교육, 고용, 이주, 피해, 지역의 보건 및 교육시설, 가구내 소득분배 및 데이터 링크
네델란드	Dutch Socio-Economic Panel (ISEP)	주관 조사연도 표본특성 조사내용	Statistics Netherlands 1984-1997 5,000 가구의 16세이상 모든 가구원을 대상으로 개인의 사회-경제적 교육, 소득, 자산, 숙박, 웰빙수준 자가측정, 소비재 소유
폴란드	Polish Household Panel (PHP)	주관 조사연도 표본특성 조사내용 조사연도 표본특성 조사내용	Department of Economics, Warsaw University 1987-1990 개인가구 거주자로서 경찰, 군인, 고위관료는 제외 각 개인의 가구구성, 인구학적특성, 가구소득, 개인 소득, 노동력 변수 1992- 7,200 가구를 대상으로 러시아 개혁이 개인과 가구의 경제적 웰빙에 개인 건강 및 영양섭취; 소비와 서비스 활용 측정; 지역차원의 자료
남아프리카	KwaZulu-Natal Income Dynamics Study (KIDS)	주관 조사연도 표본특성 조사내용	University of Natal-Durban, University of Wisconsin-Madison, Data Research Africa 공동주관 1993, 1998 1993년 첫 차수에서 남아메리카 최대 지역인 Kwazulu-Natal 의 70개 농촌과 도시에서 약 1,400를 선출하여 조사한 후, 1998년 동일가구를 정책관련 이슈: Child Maintenance Grants의 변화, Old Age Pensions 설계, 교육에 고난한 정부투자수준, 특정 성중심의 조정, 정부가 사회자
스웨덴	The Swedish Panel Study Market and Nonmarket Activities (HUS)	주관 조사연도 표본특성 조사내용	Department of Economics, Gothenburg University 1984- 1984년에 18-74세의 스웨덴 거주자 (시설수용자 제외) 2000명이상 가족형성; 양육; 주택; 시장노동; 소득과 부; 세금개혁(1993), 환경개선을 위한 납부의지(1996); 지방세; 공공 서비스; 지하경제활동(1998)
스위스	Swiss Household Panel (SHP)	주관 조사연도 표본특성 조사내용	University of Neuchâtel의 SPH 팀과 여론조사기관 M.I.S. Trend 공동 1999- 개인가구의 전국대표표본으로 1999년 에 CATI를 통해 5,074 가구 가구차원: 가구조직, 주택, 소득, 가족생활; 개인차원: 생애사건, 가족 이슈, 건강, 교육, 고용, 소득, 네트워크, 종교, 레저, 미디어, 정치와 가치,

<부표 1>. 해외 주요 패널조사 현황(계속)

국가	조사명	내용	
대한	Panel Study of Family Dynamics (PSFD)	주관	Institute of Economics, Institute of Sociology, Office of Survey Research of Academia Sinica
		조사연도	1999-
		표본특성	첫 차수에서는 36-45세 성인 전국대표 표본 선정, 두번째 차수에서는 46-65세 성인 표본 추가
		조사내용	중국인 가족의 특수성 조사: 교육경험, 직업경험, 가족구성원 상호작용, 가족 자원의 분배, 거주형태와 가족내 권력분배
영국	European Community Household Panel Study (ECHP)	주관	The London School of Economics and Political Science (LSE)
	British Household Panel Survey (BHPS)	조사연도	1994-1996
		표본특성	유럽연합 국가들의 상호비교를 위해서 설계된 가구패널조사
		조사내용	이주, 주택, 소득, 경제활동, 소비, 사회 관계, 교육과 건강. LES만 자
미국	Health and Retirement Study (HRS)	주관	Institute for Social and Economic Research at the University of Essex
		조사연도	1991-
		표본특성	전국 대표표본 5,000가구
		조사내용	인구 및 가족특성, 가구조직, 노동시장, 건강, 교육, 주택, 과 소비, 소
		주관	The Institute for Social Research at the University of Michigan
		조사연도	1992-
		표본특성	HRS는 2개의 개별적인 연구로 부터 시작: ① 1992년 HRS 원표본 코호트는 1931-41년 사이에 태어난 51-61세 인구로 구성; ② 1993년 AHEAD 코호트는 1923년 이전에 태어난 70세 이상의 인구로 구성. 배우자들은 연령에 관계없이 조사에 포함. 1998년에 HRS와 AHEAD를 통합하고 1924-30년과 1942-47년생 새 표본을 추가하면서 50세 이상 노인인구 전체를 커버하는 21,000 명의 응답자로 구성됨. 2004년에는
		조사내용	소득과 취업, 소득과 부, 건강조건, 건강상태, 보험커버리지, 세대간 상속, 가족구조등
	Panel Study of Income Dynamics (PSID)	주관	Institute for Social Research at the University of Michigan
		조사연도	1968-
		표본특성	전국대표 표본으로 1948년 4,800가구에서 1999년 6,434가구로 증가
		조사내용	핵심항목: 소득원천, 빈곤상태, 주거 및 음식형태의 공적지원, 재정문제, 가족구조 및 인구특성, 노동시장, 가사노동 시간, 주택, 지리적 이동성, 사회경제적 배경, 건강 / 부가항목: 주택 및 이웃 특성, 성취동기, 위험수용력, 자녀양육, 아동지원, 아동 발달, 직업훈련 및 획득, 노후계획, 건강, 친척관계, 부, 교육, 전투참여 경험, 이민사, 시간활용
	National Longitudinal Survey (NLS)	주관	Census Bureau; the Center for Human Resource Research (CHRR) at Ohio
		조사연도	1966-
		표본특성	1966년에 시작된 NLS는 코호트의 특성별로 5개의 패널조사를 총칭: ① 1968년에 14-24세 젊은 여성과 1967년에 30-44세인 중년 여성 5,300여명으로 구성된 NLS of Young Women and Mature Women, ② 1981년에 14-24세인 젊은 남성과 45-59세 장년 남성으로 구성된 NLS of Young Men and Older Men, ③ 1979년에 14-21세 남녀 코호트로 구성된 NLSY79, ④ NLSY79
		조사내용	패널들의 노동시장 활동뿐만 아니라 코호트가 성장하면서 경험하게 되는 교육이나 가족구조의 변화, 양육과 같은 다양한 행동양태나 태도적인 측면 조사

_: 국가통계기관 수행 조사

출처: PSID, 2005. "Panel Study Around The World"

<<http://psidonline.isr.umich.edu/Guide/PanelStudies.aspx>>