

연구노트

가구조사 표집틀 작성:

한국종합사회조사*

Creating the Household Sampling Frame:
The Korean General Social Survey(KGSS)

구묘정^{a)} · 김솔이^{b)} · 김하영^{c)} · 김지범^{d)}

MyoJung Ku · Sori Kim · HaYoung Kim · Jibum Kim

연구 모집단의 표집틀은 데이터의 질과 관련되어 있다. 면접 대상 가구는 표집틀에 의해 추출되기 때문에 표집틀의 미포함 혹은 과포함은 조사 결과를 편향시킬 수 있다. 이처럼 표집틀이 중요함에도 불구하고 정부기관으로부터 표집틀을 제공받을 수 없는 경우, 가구 표집틀의 작성과정 및 문제점을 알 수 없다. 이 연구는 한국종합사회조사 표집틀 작성과정을 서술하고 2013년 한국종합사회조사의 표집틀을 작성한 표집원을 설문하여, 가구 표집틀 작성에 대한 실질적 지식을 제공하려고 한다.

주제어: 표집틀, 전통적 가구명부 작성, 의존적 가구명부 작성, 한국종합사회조사

* 지난 10년간 한국종합사회조사에 참여한 조사응답자, 조사원, 조사팀장, 전국 대학교수들께 감사드린다. 그리고, 2014년 2학기 사회조사방법론을 수강하는 대학원생들의 세심한 제언에 감사드린다. 또한, 이 연구노트는 한국종합사회조사를 이끌어 오신 석현호, 양종희, 차종천, 김상욱, 김석호 교수가 정리해 온 한국종합사회조사 표집틀 교육자료가 바탕이 되었다는 점을 명백히 밝힌다.

a) 성균관대학교 사회학과 석사과정

b) 성균관대학교 사회학과 석사과정

c) 성균관대학교 사회학과 석사과정

d) 교신저자(corresponding author): 성균관대학교 사회학과 부교수 김지범.

E-mail: jbk7000@skku.edu

In Korea, little is known about how to create a household sampling frame when survey organizations do not get the household sampling frame from the government. Using the Korean General Social Survey(KGSS), we describe how to create the household sampling frame, and identify the problems by surveying the listers who worked on the household listing for the 2013 KGSS. The sampling frame of the KGSS is based on both traditional and dependent listing, and the listers have difficulty getting cooperation from the staff of “dong office,” and of the management office of the apartment buildings. Given the increasing time and money costs for household listing, a reliable database of the address list is much needed, and cooperation between government and non-government survey organizations and among non-government survey organizations themselves needs to be sought.

Key words: sampling frame, traditional listing, dependent listing, Korean General Social Survey

I. 서론

목표 모집단(target population)의 표집틀(sampling frame)은 데이터의 질과 관련되어 있다. 포함오류(coverage error)는 표집틀과 목표 모집단이 상응하지 않을 때 생기는 오류이다. 포함오류는 표본오류(sampling error), 측정오류(measurement error), 무응답오류(nonresponse error)와 더불어 전체조사오류(total survey error)의 한 영역으로 조사의 질을 결정하는 중요한 요소이다(Groves & Lyberg 2010). 연구 모집단이 표집틀에 미포함(under-coverage) 혹은 과포함(over-coverage)되었을 때 조사결과는 편향될 수 있

다. 예로, 전화번호부를 기초로 한 표집틀은 전화번호부에 등재되지 않은 사람들을 제외시키는 것처럼 가구조사 표집틀에서 제외된 가구는 조사에 참여할 기회가 없다. 이로 인하여 표집틀의 정확성은 목표 모집단에 대한 추론의 타당도에 영향을 미친다. 표집틀의 중요성에도 불구하고 정부기관으로부터 표집틀을 제공받을 수 없는 경우, 가구 표집틀의 작성과정 및 문제점을 알 수 없다.

일반가구 조사에서 표집틀은 전통적 가구 명부작성(traditional listing) 혹은 의존적 가구 명부작성(dependent listing)을 통해 이루어졌다. 전통적 가구 명부작성은 조사원(표집원)이 추출된 표본 지역에 대한 현지 조사를 수행하여 표집틀을 작성하는 것을 의미한다(Manheimer & Hyman 1949; Simmons 1953). 미국에서는 오래전부터 전통적 가구 명부작성이 가구조사 표집틀 작성의 절대적 기준이었다(O'Muircheartaigh et al. 2006). 하지만 전통적 가구 명부작성은 시간과 비용이 많이 들기 때문에 이것을 보완 혹은 대체할 새로운 표집틀 작성 방법을 모색해 왔다(Kish & Hess 1958; English et al. 2009; Iannacchinone 2011). 보완의 경우 이를 의존적 가구 명부작성이라고 하는데, 이미 존재하는 가구 명부를 바탕으로 조사원이 해당지역을 방문하여 확인 그리고 수정하는 표집틀 작성방식이다(Eckman & Kreuter 2011). 예를 들면, 미국 우정공사(United States Postal Service)에서 제공하는 주소 명부(address list)를 가지고 전통적인 가구 명부작성을 대체하거나 아니면 이를 전통적 가구명부와 결합하는 방법이 활용되고 있다(Iannacchione, Staab, & Redden 2003; O'Muircheartaigh, Eckman, & Weiss 2003; O'Muircheartaigh, English, & Eckman 2007). 또한, 상업용 가구 데이터베이스의 활용 그리고 이들 데이터베이스의 적용 적정성에 대한 연구도 진행되고 있다(Eckman & English 2012). 미국의 사회과학데이터를 대표하는 종합사회조사(General Social Survey)는 2000년 표집틀부터 도시지역은 우정공사 주소 명부를 활용하고, 주소가 명확하지 않은 농촌지역은 전통적인 가구 명부작성을 이용하여 표집틀의 정확성을 높이고 표집틀 작성을 위한 비용을 줄이려고 했다(Smith et al. 2013).

한국에서도 조사의 증가와 함께 전화조사 그리고 인구주택총조사를 표집

틀로 이용하는 가구조사의 정확성에 대한 관심이 커지고 있다. 대부분의 여론조사가 전화조사라는 점에서 전화조사의 표집틀에 관련된 연구들(강현철 외 2008; 허명희·김영원 2008; 홍성준·박소형·김선웅 2009)이 활발하게 진행되어 왔다. 우정엽·김지윤·문종배(2011) 연구는 전화조사 실시 시의 문제점과 더불어 여론조사회사들이 가지고 있는 전화번호부 데이터베이스가 2009년 이후 갱신되지 않은 케이티스(KTIS, 구 한국인포서비스주식회사)를 이용함으로써 표집틀이 모집단을 제대로 반영하지 못하고 있다는 것을 밝혀 주었다. 예를 들면, 60대 이상 연령층에서는 등재집단이 비등재집단과 비교하여 약 15 퍼센티지 포인트(percentage point) 더 많고, 30대에서는 등재집단이 비등재집단과 비교해서 약 13 퍼센티지 포인트 정도 적다.

전화조사 표집틀의 문제점에 대한 인식과 더불어 공공기관 가구조사에서 표집틀로 가장 많이 사용되는 인구주택총조사 자료(강현철 외 2009; 박현아 외 2010)에 대한 표집틀의 문제점과 아울러 새로운 방향 모색도 이루어지고 있다(박진우·변종석·박민규 2010). 특히, 인구주택총조사 시점과 시간적으로 거리가 있는 조사의 경우 인구주택총조사 표집틀을 이용할 때 표집틀 오류(frame error)가 생기기 때문에 이를 줄이려는 노력으로 여러 개의 표집틀을 결합한 혁신적인 방법이 모색되었다. 강현철 외(2009)는 2008 서울서베이 표집틀로 주민등록DB와 과세대장DB를 통합해서 만든 표집틀을 이용하였다. 박현아 외(2010)에서는 인구주택총조사 표집틀 시점 이후부터 신규 분양된 아파트 정보를 KB국민은행 아파트 시세조사 사이트와 결합하는 방식을 택했다.

표집틀에 대한 관심에도 불구하고 한국에서의 기존 연구에서는 몇 가지 제한점이 있다. 첫째, 한글로 쓰여진 책 혹은 논문에서 전통적 표집틀 작성에 대해 언급하고 있지 않다. 한국의 대표적 조사 저널인 “조사연구”에서도 행정기관에서 표집틀을 제공받지 않았을 때 가구 표집틀 작성 방법과 관련된 논문을 발견할 수 없었다. 또한 조사방법과 관련된 책에서도 가구 표집틀 작성 과정이 언급되고 있지 않다(홍두승 2000; 김광웅 1996; 정창수 1996; 채서일 1998; 김승현 외 역 2001; 한승준 2003; 김구 2008; 김석용 2012; 신민철 2007; 정재욱 2006). 예외로는, 얼 바비(Earl Babbie) (고성호 외 2007: 293)의

표본 설계 단계 그림에서 “추출된 각 구역으로 가서 모든 가구들을 정렬한 목록을 만들어라.”라고 짧게 언급하였다. 둘째로, 표집틀 자체가 미비한 경우 기존 표집틀을 이용하였다. 예로, 박현아 외(2010: 107)는 “아파트를 제외한 일반가구는 대체적으로 변동사항이 적을 뿐 아니라 2005년 이후의 변동 상황을 파악할 수 있는 참고자료를 확보하기가 어려운 관계로 인구주택총조사 자료를 그대로 추출틀로 사용한다.”라고 언급하고 있다. 이처럼 국내 연구에서는 행정자료를 이용하지 않는 가구 표집틀을 작성할 때의 문제점 혹은 효과적인 방법을 언급하고 있지 않다.

따라서 본 연구는 지난 8년간 한국종합사회조사를 통해 축적된 표집틀 구축과정을 자세하게 서술하고 문제점 및 개선점을 파악하고자 한다. 이를 위해 먼저 한국종합사회조사 표집틀 작성과정을 정리하였다. 이는 시행착오를 거치면서 지속적으로 갱신되었던 내부분건을 재정리한 것으로 매년 표집틀을 작성하는 조사팀장(표집원) 교육에 이용하여 왔다(성균관대학교 서베이리서치 센터 2013). 그 다음은 이 자료로 교육받은 2013년 한국종합사회조사의 표집틀 담당 조사팀장을 설문 조사하여 문제점을 파악하도록 했다. 물론, 한국종합사회조사가 완전한 조사라거나, 한국종합사회조사의 표집틀 작성 방법이 절대적 기준은 아니다. 하지만, 한국종합사회조사는 가구 명부작성에 대한 경험이 축적되어 있기 때문에 조사기관/연구자들에게 가구 표집틀 작성 시 비용과 시간 절감을 가져올 수 있는 실질적 정보를 제공할 수 있다.

우리의 연구는 대면 가구조사 표집틀 분야 연구에 기여한다고 생각해 볼 수 있다. 왜냐하면, 조사기관에서 구전으로 전해지거나 내부 문건으로만 존재하는 표집틀 작성과정을 상세하게 기술하고, 표집틀 작성과정에 대한 새로운 대안을 토론하는 장을 마련하고자 했기 때문이다.

II. 한국종합사회조사

한국종합사회조사는 대한민국의 만 18세 이상의 사람 중 시설(예. 기숙사, 군부대, 양로원, 고시원, 호텔, 모텔, 하숙 등)에 살지 않고, 한국어로 의사소

통이 가능한 성인을 연구 모집단으로 한다. 한국종합사회조사는 다단계지역 확률표집방법을 이용하고 있다. 안전행정부의 가장 최근 ‘주민등록 인구통계’를 기초로(2013년 한국종합사회조사는 2012년 12월 자료 이용) 광역시·도를 층화변수로 이용하여 1차 표집단위(primary sampling unit)인 동·읍·면을 추출한 후, 2차 표집단위인(secondary sampling unit) 통·반·리 200개 조사구를 선정한다(김상욱 외 2012). 추출된 모든 조사구에서 조사팀장(각 대학 조사원을 대표하는)은 할당받은 지역을 직접 방문하여 표집틀을 작성한 후 이를 성균관대학교 서베이리서치센터에 보낸다. 서베이리서치센터는 각 해당 지역 표집틀에서 무작위로 가구를 선정하여 다시 조사팀장에게 보내준다. 조사원은 최종 표집틀을 가지고 해당된 가구를 방문하여 만 18세 이상 성인 가구원 중 태어난 월일이 가장 빠른(실제 생일 기준) 가구원을 최종 면접대상자로 선정한다. 예외적으로, KGSS는 2003년부터 2005년까지 안전행정부에서 자료를 받아 표집틀로 이용하였기 때문에 표집틀을 자체적으로 만들지 않았다.

1. 조사팀장(표집원)

한국종합사회조사는 성균관대학교의 서베이리서치센터와 전국 20여개 이상 대학의 공동 조사이다. 서베이리서치센터는 조사의 전 과정에 관여하고, 표집틀 작성과 실제 조사는 전국 참여 대학의 교수와 학생으로 구성된 조사팀이 담당하고 있다. 예로, 2013년은 26개 대학이 참여하고 있으며, 각 대학은 평균적으로 8개 최종 조사구를 담당하고 있다.

각 대학의 조사팀장은 한국종합사회조사의 데이터 질을 결정하는 아주 중요한 역할을 수행하고 있다. 각 대학의 학생 중 한 명은 조사팀장으로 6개월간 활동하며 가구 명부 및 표집틀을 작성하고, 해당 대학 조사원을 충원하여 감독하고 조사된 설문지를 검토한 후 서베이리서치센터에 보내는 역할을 한다. 표집틀을 작성하면서 조사팀장은 실제 조사 시 조사를 용이하게 하거나 조사비용 절감을 가져오는 유용한 정보를 수집하는 해당 지역 조사의 척후병 역할을 한다.

최종 표집 지역이 정해진 후 조사팀장을 대상으로 한 한국종합사회조사에 대한 전반적 교육 및 표집틀 교육은 2월에 실시된다. 2013년 조사를 위해서 2013년 2월 22~23일 1박 2일 동안 표집틀 작성을 위한 교육이 실시되었다. 표집틀 교육은 이론적인 설명 뿐 아니라, 팀장들이 몇 개의 팀으로 나뉘어 직접 표집틀을 작성해 본다. 서베이리서치센터에서는 실습지역에 대한 안내도를 팀장들에게 나누어 주고 팀장들은 해당 가구를 직접 방문하여 표집틀 작성을 위한 실습을 하게 된다. 표집틀 작성이 끝난 후 팀장들이 작성한 표집틀을 확인하여 잘못되었거나 문제가 있는 부분을 다시 한 번 설명하고 수정하여 표집틀에 대한 정확한 이해를 할 수 있도록 한다. 2013년 교육은 서울특별시 종로구 가회동 3통이었다. 교육에 참석하지 못한 팀장들에 대해서도 동일한 내용으로 추가 교육이 진행된다.

2. 표집틀 작성과정

<표 1>은 표집틀 작성과정이다. 각 단계별로 주요 행위자, 활동내용, 그리고 활동 시 주의사항으로 정리할 수 있다. 첫 단계는 서베이리서치 센터의 표본 추출과 표집틀을 작성하는 조사팀장에 대한 교육 내용으로 위에서 언급했기 때문에 두 번째 단계부터 서술한다.

<표 1> 표집틀 작성과정

행위자	활동내용	주의사항
서베이리서치센터	표본추출, 조사팀장 교육, 표본 지역 주민센터 접촉	
조사팀장	할당 지역 주민센터 방문 시 가구 수 확인 및 안내도 확보(인터넷 지도 활용)	적정 가구 수
조사팀장	할당 지역 표집틀 및 지도 작성	주택형태, 주소, 거주 확인 등
서베이리서치센터	표집틀 확인 및 최종가구 선정	

주민센터 방문 그리고 가구 수 및 번지 확인: 조사팀장은 표집지역에 나가 표집틀을 작성하기 전에 반드시 주민센터를 방문하여 해당 지역에 대한 정보를 확인한다. 주민센터를 방문하기 전에는 민원이 적은 시간을 확인하여 방문 약속을 잡는 것이 좋다. 또한 방문 시에는 공문, 한국종합사회조사, 그리고 성균관대 서베이리서치센터 브로슈어 등을 함께 활용한다. 전입담당자나 주민등록담당자로부터 해당 조사지역 반/리의 총 가구 수를 문의하고, 가능하면 상세한 번지 목록을 확보하도록 한다. 가구 수가 50가구 미만이거나 150가구 초과인 경우에는 조사지역의 조정이 필요함으로 서베이리서치센터와 상의하도록 한다. 50가구 이상이어야 하는 이유는 면접원이 가구를 방문했을 때, 해당 가구가 존재하지 않거나 응답자의 사정으로 인해 조사가 이루어지지 못하게 될 경우를 고려하여 충분한 가구를 포함하기 위해서이다. 이와 반대로 150가구 이하여야 하는 이유는 표집틀 작성 시간과 비용을 줄이기 위한 편의적 기준이다.

가구 수가 50가구에서 150가구 사이에 속하지 않을 때 서베이리서치센터에서 취하는 조치는 경우에 따라 다르다. 우선, 50가구 미만일 때에는 같은 통 안에서 새로운 반을 선정하게 된다. 하지만 아파트 지역일 경우 현재 선정된 통이 아파트 단지인 경우가 대다수이므로 새로운 반을 선정하는 것에 의미가 없어 같은 통 안에서 반을 뽑은 뒤, 기존의 반과 합하여 기준가구 수가 넘도록 한다. 이러한 방법을 적용하여도 가구 수의 문제가 해결되지 않으면 최종적으로 통과 반을 다시 선정한다. 150가구 초과인 경우는 조사팀장에게 해당 반의 관내도 또는 지도를 받아보고 큰 도로나 블록을 기준으로 기존 반의 구역을 새로이 정하여 가구 수가 50가구에서 150가구 사이가 되도록 조정한다.

관내도 복사: 조사팀장은 관내도 담당자를 찾아 해당 조사지역의 관내도를 확보한 후 주민센터 방문을 마치도록 한다. 주민센터 방문은 조사지역에 관련한 유용한 정보(해당 지역 특성- 주거 또는 상가, 재개발 가능성 등)를 미리 알 수 있는 기회가 되며, 가구 수, 번지목록, 관내도를 확보할 수도 있다. 관내도 등을 보면서 가령 20통 1반이 어느 쪽인가요?, 주거지역인가요, 상가

지역인가요?, 1반에는 총 몇 세대가 거주하나요?, 한 번지에 한 가구씩만 사나요?, 혹시 재개발될 가능성이 있나요? 등을 주민센터 직원에게 직접 물어본다. 번지목록은 주민들의 개인정보가 기재되어 있으므로 주민센터에서 협조적이지 않은 경우가 있다. 관내도를 구비하지 않거나 있다고 해도 알아보기 용이하지 않은 경우가 대다수이다. 만일 선정된 반이 재개발로 인해 세대가 전혀 살지 않거나 조사기간 내에 사람이 살지 않을 것으로 예상되는 경우에는 서베이리서치센터에 즉시 연락해서 새로운 통/반을 부여받아야 한다.

인터넷 지도 서비스 활용 정보수집: 주민센터에서 관내도를 얻지 못한 경우, 인터넷 포털사이트(예: 다음, 네이버, 구글 등)에 있는 지도를 이용하면 표집틀 작성에 소요되는 시간과 노력을 크게 절감할 수 있다. 이 때, 여러 가지 형태(주소가 표시된 지도, 건물모양이 표시된 지도, 위성지도 등)의 지도를 대조하여 사용하는 것이 더욱 효과적이다. 주의할 점은 인터넷 지도 정보는 유용하기는 하지만 최신정보가 아니기 때문에 모든 내용이 정확하다고 할 수 없다. 따라서 참고자료로만 활용하고 반드시 조사지역을 방문하여 구체적인 사항을 확인해야 하며 변동사항이 있을 때에는 지도에 표기하도록 한다. 이 때 가구뿐만 아니라 근처 다른 구조물들을 표기하여 후에 면접원이 해당 가구를 찾아가는 데 필요한 시간과 비용을 단축할 수 있도록 하는 것이 좋다.

표집틀 및 지도 작성: 서베이리서치센터에서는 각 대학 조사팀장들에게 이메일을 통해 표집틀 양식을 보낸다. 조사팀장은 조사지역에 대한 정보를 충분히 확보한 다음 해당 주소를 직접 방문하여 표집틀을 작성하며, 1차 워크숍이 끝난 직후(2013년의 경우에는 2월 23일)부터 약 보름 동안 표집틀을 작성해야 한다. 조사팀장은 표집틀 작성 시 크게 지역/주택형태와 가구에 대한 사항을 염두에 두어야 한다.

지역, 주택 형태:

(가) 다세대 주택: 한 번지에 여러 가구가 함께 살고 있는 주택형태를 다세대 주택이라고 한다. 조사팀장은 해당 번지에 거주하는 가구 수와 각 가구의 위치를 정확히 파악해야 한다. 예를 들어, 해당 번지에 5가구가 산다

는 정보만으로는 충분치 않다. 각 층의 가구 수와 동일한 층에 2가구 이상이 거주할 경우 각 가구를 구별할 수 있을 정도까지 파악해야 한다. 대안으로 지하와 지상을 구분하는 방법과 벽 또는 지붕의 색깔 명시, 좌·우측과 같은 방향 기록방법이 있으며, 번지수에 약도를 그려서 번호를 부여하는 것도 하나의 방법이 될 수 있다.

해당가구에 방문이 이루어지지 않거나 외관상으로 가구 수 파악이 어려운 경우 간접적으로 가구의 정보를 얻는 방법이 있다. 첫 번째, 전기계량기의 개수를 확인하고 가구 수를 파악하여 실제 가구에 사람이 거주하는지를 살펴본다. 단, 전기계량기의 작동여부와 전기계량기의 수치가 다른 가구와 비슷한지를 자세히 살펴보아야 한다. 두 번째, 현관 앞에 편지나 신문 등의 우편물이 많이 쌓여 있으면 가구에 사람이 거주하는지를 의심해 보아야 한다. 거주하더라도 장기간 집을 비울 가능성이 있기 때문이다. 세 번째, 근처 부동산의 도움을 받는다. 네 번째, 이웃집에 물어본다. 특히 주인집의 경우 세입자 가구에 대한 정보를 가장 정확하게 알고 있기 때문에 도움을 받을 수 있다. 실제로 2006~2012 KGSS 조사팀장의 경험에 의하면 직접 초인종을 눌러 확인하는 것이 가장 빠르고 쉬운 방법이었다.

- (나) 시골지역: 시골지역은 이장 또는 동네 주민들이 서로의 상황에 대해 매우 잘 알고 있는 경우가 많으므로 협조를 구하여 가구주소뿐 아니라 가구원의 이름과 연락처까지 알 수도 있다. 하지만 시골지역은 읍·면사무소에서 파악하고 있는 주소와 실제 주민들이 사용하고 있는 주소가 다른 경우가 많기 때문에 면접원이 표본가구를 찾기 쉽도록 표집틀을 작성한다. 또한 한 주소를 여러 가구가 함께 사용하는 경우에는 대문이나 지붕의 색깔 등 구별이 쉬운 특징을 함께 기재하는 것이 좋으나, 표집틀 작성과 실제 조사 기간의 차이를 염두에 두어 쉽게 변화되는 특징을 피하는 것이 좋다. 가령, ‘집 앞 담장에 개나리가 피어 있다’라는 특징은 조사기간에 면접원이 조사를 나갔을 때 그 특징을 찾기 힘들 수도 있다. 하지만 사납고 위협적인 큰 개가 있다든지, 집 앞에 큰 구덩이가 있어 위험할 것으로 예상되는 등의 특징들은 가구를 구별하는 데는 별 도움이 되지 않

지만, 면접원의 안전을 위해 기입해 주는 것이 좋다.

- (다) 상가지역/시장지역: 주거지와 상가가 혼재되어 있는 경우 사람이 거주하지 않는 상가는 표집틀에서 제외해야 하지만, 상가 안에서 주거생활(식사, 취침 등)을 동시에 하는 경우는 표집틀에 포함한다.
- (라) 아파트지역: 아파트는 직접 돌아보며 각 호수를 확인해야 한다. 간혹 우편함을 보거나 아파트의 라인과 층수만을 셈하여 작성해 오는 경우가 있다. 이는 잘못된 방법이다. 왜냐하면 주거용도가 아닌 다른 용도(경비실 또는 놀이방 등)로 사용되는 경우가 있고, 라인에 따라 층수와 배치가 다를 수 있기 때문이다. 이를 위해 일일이 직접 돌아보며 확인을 해야 한다. 도어락이 있는 아파트는 출입이 원활하지 않으므로 관리사무소에 가서 협조를 구해야 한다. 이때 관리사무소와 경비원에게 좋은 인상을 주는 것이 표집틀 작성뿐만 아니라 본조사의 성과에도 좋은 영향을 미칠 수 있다. 그러므로 처음 관리사무소를 방문할 때에 좋은 인상을 주기 위한 노력이 필요하다. 한편 관리사무소에 가면 해당 동과 호수에 주거 여부를 확인할 수 있는 현황표가 붙어 있으므로 이 현황표를 바탕으로 거주지 용도의 가구만 파악하여 정확한 표집틀을 작성하는 방법이 있다.

주소 및 가구:

- (가) 번지가 나타나 있지 않거나 체계적이지 않은 경우에는 주택의 배열 순서를 기준으로 임의의 번지를 지정한 후 표집틀에 기록한다. 이때 배열 순서에 대한 대략적인 정보를 반드시 표집틀에 기록해야 한다.
- (나) 주거지가 아닌 경우(공터, 주차장, 사무실, 가게, 도로 등) 또는 번지는 있지만 사람이 주거하지 않는 경우(폐가, 별장, 장기간 주거용도로 활용되지 않는 집 등) 등은 표집틀에서 제외시킨다. 그러나 일시적으로 비어 있는 집(이사, 장기출타 등)의 경우에는 표집틀에 포함시킨다.
- (다) 주소체계의 변화가 있을 때에는 구주소와 현주소를 함께 기록한다. 새주소(예를 들어 'OOO길 OO'로 정비된 지역)는 표집틀과 관내도에 새주소를 적어서 면접원이 찾기 쉽도록 한다.

- (라) 실제로 존재하는 모든 가구들이 포함되도록 해야 한다. 주민센터에서 알려주지 않은 가구가 있다면 추가하도록 하고, 주민센터에서 알려준 번지가 없을 때에는 실제로 존재하지 않는지 확인하기 위해 인근지역의 가구 또는 부동산에서 확인한 후에 제외한다. 이 같은 실질적인 정보를 파악하기 위해 가까운 부동산, 중국음식점, 치킨집, 미용실 등에 도움을 요청하는 것도 좋은 방법이다.
- (마) 한 건물에 여러 가구가 함께 거주하는 경우(다세대 주택, 아파트, 주거용 오피스텔 등)에 호수가 체계적으로 명시되어 있다면, 호수가 빠른 순서대로, 위에서 아래, 왼쪽에서 오른쪽으로 등의 가구 명부작성 기준을 동일하게 적용시켜 표집틀에 작성한다. 이때 호수가 명시되어 있지 않다면 임의로 번호를 부여한 후 앞의 기준을 따라 표집틀을 작성하되 반드시 간단한 그림을 첨부하여 면접원으로 하여금 혼란이 생기지 않도록 한다.



〈그림 1〉 조사팀장이 그린 지도

표집틀과 더불어 <그림 1>은 조사팀장이 그린 지도이다. 이 지도 혹은 관내도는 실제 조사 시에 조사원이 해당 지역에서 추출된 가구를 찾는 데 드는 시간과 비용을 줄여준다.

표집틀 목록 작성 및 정리: 각 대학 조사팀장들은 표집틀 양식에 표집틀을 작성한 후 파일로 서베이리서치센터에 제출한다. 서베이리서치센터에서는 각 대학 조사팀장들에게 받은 표집틀을 구글 지도와 네이버 로드뷰를 이용하여 비교 및 확인한다. 이러한 자료는 최신 정보가 아니라는 점과 우리가 검토하고자 하는 전 지역에 대한 정보를 모두 포함하고 있지 않다는 한계점이 있음에도 불구하고 해당 지역을 방문하지 않고도 확인할 수 있는 최선의 방법이다. 이때 한 지역을 검토하는 시간은 아파트 지역은 10~20분, 아파트 외의 지역은 30분에서 1시간이 소요된다. 조사팀장이 작성한 지도와 구글 지도가 일치하지 않는 등의 문제가 있는 표집틀은 다시 조사팀장에게 돌려보내 수정하도록 요청한다.

모든 표집틀에 대한 확인이 끝나면 서베이리서치센터에서는 각 지역의 표집틀 파일을 계통추출법(systematic sampling)을 이용하여 한 지역당 12개 또는 13개의 조사가구를 선정한다. 최종적으로 선정된 가구는 한국종합사회조사 면접원 오리엔테이션 기간(2013년의 경우 6월 21~22일에 걸쳐 이틀 동안 진행된다)에 조사팀장들에게 알려주고 오리엔테이션이 끝난 후 바로 조사가 시작될 수 있도록 한다.

Ⅲ. 조사팀장 설문

조사팀장 설문은 2013년 KGSS 조사팀장을 대상으로, 전국 26개 대학 26명의 팀장이 교육과정 마지막에 자기기입 방식으로 작성하였다. 설문지는 폐쇄형 10문항과 개방형 8문항으로 총 18문항으로 구성되었으며, 충실히 답을 준 3명의 학생을 선정해 문화상품권을 제공하는 인센티브를 주었다. 이 중 여덟 명은 이전에 조사원 경험, 다섯명은 표집틀 작성 경험이 있었다. 26명의

팀장 중 남자는 14명, 여자는 12명이었고, 전공은 사회학 15명, 사회복지학 3명, (경찰, 도시, 보건)행정학 3명, 윤리교육학, 청소년학, 정치외교학, 응용통계학, 컴퓨터공학 각 1명씩 5명이다. 학년은 1학년 3명, 2학년 5명, 3학년 8명, 4학년 7명, 대학원생 3명이었다.

아래는 조사팀장이 표집틀을 만들 때 주민센터와 아파트에서 봉착하는 문제점과 서베이리서치센터가 개선해야 할 사항에 대한 견해이다. 모든 조사팀장이 주민센터를 방문하였으며, 주민센터의 협조 정도에 대한 질문에 26명 중 20명(77%)이 비협조적이었다라고 응답하였다. 비협조적인 이유로 개인적인 정보라는 이유에서 자료 제공을 거부하는 곳이 많았으며, 한국종합사회조사에 대해 잘 모르거나 센터에서 제공한 공문과 팀장들의 설명에도 불구하고 조사에 대해 의심하는 곳이 많았다고 답하고 있다. 주민센터에서 조사에 대한 의심으로 정보 제공을 꺼려했다고 응답한 팀장은 11명(42.3%)이었다:

“주민센터에서 한국종합사회조사에 대해 처음 들어본다며 경계를 했습니다. 관내도는 줄 수 없다고 하는 경우도 있었습니다. 협조문을 가지고 가도 귀찮다는 식으로 설렁설렁 해주고 동사무소에서는 높은 직급의 사람이 와야 한다며 1~2시간 기다리게 하는 일이 있었습니다.”

가구 명부작성을 위해 접촉한 사람에 관한 질문에 26명의 모든 조사팀장이 주민센터 직원과 접촉하였다고 응답하였다. 이외 통장 및 이장과 접촉하였다고 응답한 팀장이 11명(42.3%), 관리사무소 및 경비원과의 접촉은 10명(38.5%)이었다:

“저는 표집틀 작성을 위해 각 지역 주민센터의 전/출입을 담당하시는 분과 만났고, 아파트에서는 관리사무소를 방문해 협조를 구하였습니다. 그리고 도시 외의 지역에서는 이장님을 만나 많은 도움을 받았습니다.”

표집틀에서 제외된 가구의 특성에 관한 질문에 19명 중 8명(30.8%)이 전입신고 여부와 실제 거주 여부가 상이한 경우라고 응답하였다. 그 다음으로 가구 구분이 명확하지 않은 경우가 6명(23.1%)이었다:

“주민센터에서 받은 목록에서는 세대가 거주하는 것으로 기입되어 있는데, 실제 조사지역을 방문하였을 때 폐가였던 경우가 있었고, 장기부채자가 거주하여서 확인이 어려운 경우도 있었다.”

아파트 가구 표집틀 작성의 문제점에 대한 응답으로 경비원이나 이웃 주민들의 도움을 받았다는 팀장이 25명 중 12명(46.2%)으로 가장 많았다:

“먼저 아파트 관리사무소에 들러 협조 요청과 현재 몇 세대가 거주하고 있는지 파악한 후 해당 아파트 위층부터 차례대로 내려오면서 계량기 사용 여부를 확인하고 각 가구마다 초인종을 눌러 거주 사실을 확인하였습니다.”

“도어락 또는 경비가 있는 경우 어떻게 조사하였습니까?”라는 질문에 대한 응답으로 주민들의 왕래를 기다렸다가 따라 들어갔다는 팀장이 26명 중 15명(57.7%)로 가장 많았는데, 조사팀장들의 고충과 이를 해결하려는 임기응변적 방법이 취해짐을 알 수 있다. 이 외에 경비원의 도움을 구해 출입했다는 응답도 있었다:

“경비실에 도움을 구했으나 허락을 받지 못해 최대한 눈에 띄지 않게 기다렸다가 이웃주민이 들어가는 틈에 따라 들어갔습니다.”

마지막으로 서베이리서치센터가 조사팀장의 표집틀 작성과정을 원활히 진행할 수 있도록 도움을 줄 수 있는 일에 대해 물어보았다:

“해당 반의 가구 수가 50가구 미만인 경우 다시 반을 부여받아 주민센터에서 정보를 구하는 것이 미안한 경우가 있었습니다. 그래서 경우의 수를 대비해 예비 반을 부여받아 한 번에 협조를 구할 수 있었으면 좋겠습니다.”

“리서치센터에서 미리 주민센터에 연락을 취해서 보다 원활한 사전조사가 될 수 있도록 해주셨으면 좋겠습니다. 너무 경비가 삼엄한 고급아파트

인 경우 관련 공문을 미리 만들어 협조요청 및 설득에 도움이 되도록 해주셨으면 좋겠습니다.”

IV. 결론

본 논문은 한국종합사회조사를 이용하여 표집틀의 구체적인 작성과정을 공개하고 표집원 조사를 통해 표집틀의 문제점을 파악하였다. 한국종합사회조사는 주민센터의 가구 관련 자료를 얻을 수 없는 경우 전통적 가구명단 작성방식을 이용하며, 주민센터의 가구 관련 자료를 얻을 수 있는 경우는 의존적 가구명단 작성방법을 이용하였다. 표집원 조사를 통해 표집원들이 주민센터로부터 가구 관련 자료 획득이 어려웠고, 가구 접근성과 관련된 두 가지 문제에 봉착하고 있으며, 서베이리서치센터는 표집원이 접촉하는 기관(주민센터, 아파트관리소 등) 책임자들과의 소통에 미진했다는 점을 알게 되었다.

비록 한국종합사회조사는 표집틀 작성에 대한 실질적인 지식과 경험을 바탕으로 표집원들을 교육시키고 표집틀을 작성하지만 아직 여러 면에서 제한점이 있다. 표집틀 작성과정에서의 문제점과 개선점은 다음과 같다.

첫째, 표집원이 주민센터 혹은 관리사무소 등과의 접촉에서 순조롭게 도움을 받지 못하고 있다. 현재는 협조 공문이 해당 기관 담당자에게 제대로 전달되었는지 확인하지 못하고 있으며, 협조 공문 또한 성균관대학교 서베이리서치센터의 공문으로 제한되어 있다. 공문을 받지 못 했다고 하는 경우에 다시 협조 공문을 재발송하고 있는데, 서베이리서치센터 공문 이후 시차를 두고 각 지역 대학 교수들의 협조 공문 발송도 필요하다.

둘째, 서베이리서치센터는 표집원이 표집틀을 작성할 때 발생하는 애로사항을 즉시 해결할 수 없는 경우가 있다. 표집원들이 각 지역 표집틀을 만드는 시점이 학기 중이므로 때때로 서베이리서치센터 조교들이 수업이 있는 경우에는 원활한 소통이 이루어지지 않고 있다. 이를 해결하기 위해서 적어도 표집틀 작성 시기동안 만이라도 표집원 연락 전담 담당자가 필요하다.

셋째, 각 대학 표집원이 기록한 표집틀은 서베이리서치센터에서 구글과 네이버 지도로 검토한다. 구글 위성지도는 위에서 내려다 본 모습만 보이므로 아파트의 경우 전체 층을 파악하기 어렵고, 이러한 경우에는 포털사이트인 ‘네이버’에서 제공하는 ‘로드뷰’를 참고하여 더욱 자세한 표집틀 지역의 정보를 얻어 조사팀장들이 그려온 표집틀을 검토하였다. 즉, 서베이리서치센터는 조사비용등의 현실적 문제로 인하여 전국의 표집 지역을 방문하여 조사팀장들이 작성한 지역 표집틀을 확인하지는 못했다. 따라서 조사팀장에게 전적으로 의존하는 가구 표집틀의 정확성에 문제가 있을 수 있다. 2009년부터 2011년까지, 지난 3년 동안 가구 표집틀 중 표집틀 내 대체가구 현황자료를 이용하여 실제 조사 시에 표집틀로부터 어떠한 문제가 나타나는지 알 수 있었다. “지난 3년간의 대체케이스 수는 390건(2009년 77건, 2010년 169건, 2011년 144건)으로 4.4%가 대체되었는데, 대체 케이스 중 빈집 57%, 거주지가 아님 8%, 이사 4%, 장기출장 3%와 기타(재개발 예정, 공사 중, 철거, 기숙사) 등이었으며, 빈집이 가장 많은 비중을 차지하고 있는 것으로 나타났다. 빈집의 대부분(76%)이 도시지역이기 때문에, 더욱 정교한 표집틀을 만들기 위해서는 수집된 표집틀 중 도시지역 몇 곳을 선정하여 사후답사와 같은 실질적인 재검토가 이루어져야 할 것이다.”

넷째, 표집틀 작성 교육 직후, 조사팀장들은 2주에 걸쳐 표집틀을 작성하게 된다. 그런데 실제 조사기간이 6월 말부터 8월 말까지 진행됨에 따라 표집틀 작성 시점과 실제 조사 시점 사이는 3~4개월의 차이가 생긴다. 이 기간 사이에 새로 생긴 가구는 표집틀 목록에서 제외되는 결과가 발생하고 있다. 즉, 조사가구 과포함(overcoverage)은 현지 조사 시 조사원이 문제를 확인할 수 있지만 조사가구 미포함(undercoverage)에 대한 대책은 없다.

마지막으로, 표집원과 조사원은 조사의 질과 관련된 표집틀을 작성하고, 추출된 가구를 방문하고, 추출된 가구에서의 가구원 선택(Manheimer & Hyman 1949)과 응답자와의 대면접촉을 통한 조사 수행이라는 핵심적 역할을 함에도 불구하고, 조사원에 대해서 거의 연구가 이루어지지 않고 있다. 학생들이 팀장으로 활동하는 한국종합사회조사에서 조사팀장을 선발할 때 조사팀장의 조사방법론에 대한 지식이 있는 학생, 사회조사에 관심이 높은 학생,

저학년부터는 고학년을 우선 고려하여 선정하는 것이 바람직하다는 결론을 내리게 되었다. 또한, 조사팀장 설문문을 통해 조사팀장들에 대한 교육이 충분치 않았다는 점도 파악할 수 있었다. 조사팀장들은 1박 2일의 짧은 교육시간으로 인해 실습시간이 짧다고 느꼈고, 한정된 주택 형태의 실습으로 인해, 다양한 가구를 방문할 기회가 없었다는 점과 개별적 실습 시간이 부족했다는 점을 언급했다. 이에 대한 개선책으로 교육 시간을 늘리는 것이 바람직하지만, 만약 교육 시간을 늘릴 수 없는 경우, 사전에 교육자료집을 팀장에게 배부·숙지시키며 대신 실습 시간을 늘릴 수 있다. 이로써 그룹 실습과 더불어 개인 실습을 강화하는 방법도 모색할 필요가 있음을 알게 되었다.

조사 데이터의 증가가 데이터의 질과 꼭 관련이 있는 것이 아닌 것처럼(한신갑 2012), 조사연구의 증가가 전체조사오류와 관련된 다양한 연구, 조사오류들 간의 연관성에 대한 논의로 발전되지 못한 것으로 판단된다. 특히, 조사원은 조사오류의 여러 면에 밀접히 관련되어 있다. 따라서 조사원이 겪는 어려운 점을 파악하고, 조사를 용이하게 수행할 수 있도록 적극적으로 도움을 주기 위한 조사원에 대한 연구가 요구된다. 궁극적으로 조사원의 부담을 줄여주려는 노력이 조사에 드는 비용과 시간을 절약하고 데이터의 질을 높일 수 있기 때문이다. "또한, 표집틀 작성 문제는 조사의 대표성과 관련되어 있다. 센서스와 비교하여, 한국종합사회조사에서는 토지가격이 높은 지역이 과소대표되고 있는데(김지범 외 2014), 이는 표집틀 작성문제와 연관지어 생각할 수 있다." 토지가격이 높은 지역일수록 여러 장애물(도어락, 삼엄한 경비 등)에 의해 가격이 낮은 지역에 비해 접근이 어려워져, 표집틀 작성 원칙에 따라 주거여부 확인이 제대로 이루어지지 않아, 빈집이 포함될 수 있다. 실지조사 시 빈집이 접촉불가능의 경우로 간주되어 토지가격이 높은 지역이 과소대표될 수 있다. 즉, 표집틀의 올바른 작성은 조사 결과에 영향을 미칠 수 있는 중요한 요인이 될 수 있다.

응답률 저조와 조사비용의 증가 등으로 조사환경이 더욱 어려워지고 있다. 외국의 경우 전통적 가구 명부작성이 우정공사(United States Postal Service) 혹은 상업용 주소 데이터베이스와의 결합으로 더 정확한 가구조사 표집틀을 만드는 방법의 추세를 고려할 때, 한국에서도 앞으로의 과제는 과

연 전통적 가구명부 작성을 대체할 혹은 보완할 방법이 있느냐이다. 강현철 외(2009)에서 주민등록DB와 과세대장DB를 통합해서 만든 표본추출틀을 이용한 2008 서울서베이는 일반적 조사기관에서 쉽게 활용할 수 없는 자료를 표본추출틀로 이용하고 있다. 분명 새로운 방법의 모색임에도 불구하고, 이런 주민등록DB와 과세대장DB 등과 같은 자료를 일반 조사기관이 채택할 수 없는 아쉬움이 있다. 아직 한국에서 민간기관이 표집틀 작성에 활용할 수 있는 데이터베이스가 없고 통계청에서도 표집틀을 제공받지 못하고 있다. 전통적 가구 명부작성이 아닌 의존적 가구 명부작성의 경우에 조사원은 처음에 제공된 표집틀을 그대로 받아들이는 확신편향(confirmation bias)이 있지만 (Eckman & Kreuter 2011), 적어도 통계청과의 협력은 표집틀 생성을 위한 시간과 비용을 줄일 수 있을 것으로 생각된다. 또한 민간기관끼리 공동협력으로 주소 데이터베이스를 만들거나, 표집지역을 같이 활용하는 방법도 생각할 수 있다.

이론적으로 표집틀 작성과정은 표집틀 작성교육을 받은 표집원들이 주민센터 방문을 시작으로 하여 할당받은 각 지역 표집틀을 작성하는 간단한 작업이다. 하지만 본 연구는 표집틀 작성이 주민센터의 협조 및 가구 접근성 문제로 인하여 완전한 표집틀 작성의 어려움을 보여주었으며 행정기관 표집틀을 제공받지 못하는 조사기관에게는 표집틀 작성에 관련된 실질적 지식을 제공하였다. 더불어, 표집틀 작성방법의 다양성, 행정기관에서 제공하는 표집틀의 적정성, 조사원에 대한 조사의 필요성 등에 대한 의의가 있다.

참고문헌

- 강현철 · 박승열 · 김지연 · 김인수 · 이동수 · 황재일 · 박민규. 2009. “2008 서울서베이 표본추출틀 구축 및 표본추출 사례 연구.” 《조사연구》 10(3): 157-172.
- 강현철 · 한상태 · 김지연 · 정용찬 · 허명희. 2008. “RDD 전화조사와 주요결과.” 《조사연구》 9(1): 1-22.

- 김광웅. 1996. 《방법론강의》 다우문화사.
- 김구. 2008. 《사회과학 연구조사 방법론의 이해: 양적연구와 질적 연구의 접근》. 비엔엠북스.
- 김상욱 김지범 · 문용갑 · 신승배 · 장상수. 2012. 《한국종합사회조사(KGSS) 2011》 성균관대학교 출판부.
- 김석용. 2012. 《Intro 조사 연구방법론》 탐북스.
- 김지범 · 최문희 · 강정한 · 한수진 · 김석호. 2014. “한국에서의 무응답과 조사하기 어려운 응답자.” 《조사연구》 15(3): 85-107.
- 박진우 · 변종석 · 박민규. 2010. “국민건강영양조사 표본설계를 위한 추출틀 구축.” 《응용통계연구》 23(5): 923-932.
- 박현아 · 박승환 · 전종우 · 박진우. 2010. “계속조사 표본설계에서 추출틀 변경에 따른 층화변수 선정: 국민여행실태조사 사례연구.” 《조사연구》 11(3): 103-114.
- 성균관대학교 서베이리서치센터 2013. 《KGSS 조사팀장 1차 워크숍 교육자료집》.
- 신민철. 2007. 《사회연구방법의 기초》. 창민사.
- 우정엽 · 김지윤 · 문종배. 2011. “표집틀 설정과 표본추출방법에 따른 정치성향 분석의 문제점: 임의번호걸기(Random Digit Dialing)과 전화번호부 추출 방법 비교.” 《조사연구》 12(1): 153-174.
- 정재욱. 2006. 《사회조사방법론의 이해》. 학고방.
- 정창수. 1996. 《사회과학 방법론》. 대영문화사.
- 채서일. 1998. 《사회과학 조사방법론》. 학현사.
- 한승준. 2003. 《사회조사방법론》. 대영문화사.
- 한신갑. 2012. “혼합식 조사와 웹패널의 (열은) 빛과 (길은) 그늘.” 《조사연구》 13(3): 1-31.
- 허명희 · 김영원. 2008. “RDD 표본 대 전화번호부 표본: 2007년 대통령 선거 예측 사례.” 《조사연구》 9(3): 55-69.
- 홍두승. 2000. 《사회조사분석》. 다산출판사.
- 홍성준 · 박소형 · 김선웅. 2009. “국내 유선 전화조사에서 표본추출틀의 포함률.” 《조사연구》 10(1):33-56.
- Earl Babbie. 2007. 《사회조사방법론》 고성호 외 역. 톰슨코포레이션코리아.
- Polkinghorne, Donald. 2001. 《사회과학방법론》 김승현 외 역. 일신사.
- Eckman, S. and N. English. 2012. “Creating Housing Unit Frames from

- Address Databases: Geocoding Precision and Net Coverage Rates.” *Field Methods* 24(4): 399-408.
- Eckman, S and F. Kreuter. 2011. “Confirmation Bias in Housing Unit Listing.” *Public Opinion Quarterly* 75(1): 1-12.
- English, N., C. O'Muircheartaigh, K. Dekker, M. Latterner, and S. Eckman, 2009. “Coverage Rates and Coverage Bias in Housing Unit Frame.” *Proceedings of the Section on Survey Research Method Section*. American Statistical Association.
- Groves, R.M., and L. Lyberg. 2010. “Total Survey Error: Past, Present, and Future.” *Public Opinion Quarterly* 74(5): 849 - 879.
- Iannacchione, V.G. 2011. “The Changing Role of Address-Based Sampling in Survey Research.” *Public Opinion Quarterly* 75(3): 556-575.
- Iannacchione, V.G., J.M., Staab., and D.T. Redden. 2003. “Evaluating the Use of Residential Mailing Addresses in a Metropolitan Household Survey.” *Public Opinion Quarterly* 67(2): 202-210.
- Kish, L. and I. Hess. 1958, “On Noncoverage of Sample Dwellings.” *Journal of the American Statistical Association* 53(282): 509-524.
- Manheimer, D. and H. Hyman. 1949. “Interviewer Performance in Area Sampling.” *The Public Opinion Quarterly* 13(1): 83-92.
- O'Muircheartaigh, C., S. Eckman, and C. Weiss. 2003. “Traditional and Enhanced Field Listing for Probability Sampling.” *Proceedings of the Section on Survey Research Method Section*. American Statistical Association.
- O'Muircheartaigh, C.N. English, S. Eckman, H. Upchurch, E. Garcia, and J. Lepkowski. 2006. “Validating a Sampling Revolution: Benchmarking Address Lists against Traditional Listing.” *Proceedings of the Section on Survey Research Method Section*. American Statistical Association.
- O'Muircheartaigh, C., N. English, and S. Eckman. 2007. “Predicting the Relative Quality of Alternative Sampling Frames.” *Proceedings of the Section on Survey Research Method Section*. American

Statistical Association.

Simmons, W.R. 1953. "Prelisting in Market or Media Surveys." *Journal of Marketing* 18(1): 6-17.

Smith, T.W., P. Marsden, M. Hout, and J. Kim. 2013. *General Social Surveys, 1972-2012: Cumulative Codebook*. Chicago: NORC at the University of Chicago.

<접수 2014/9/27, 수정 2014/11/17, 게재확정 2014/11/21>