

연구노트

표집틀 설정과 표본추출방법에 따른 정치성향 분석의 문제점:**임의번호걸기(Random Digit Dialing)과 전화번호부 추출방법 비교**

Study on Sampling Frame and Methods for Analyzing Political Attitudes :
A Comparison of RDD and Direct Sampling

우정엽* · 김지윤** · 문종배***

Jung-yeop Woo · Jiyeon Kim · Jong-bae Moon

이 연구는 현재 한국의 정치사회여론조사에서 쓰이고 있는 표집틀 설정과 표본추출방법의 문제점을 지적하고 이로 인해 발생하는 여론조사의 비정확성에 대해 논하고자 한다. 현재 한국의 여론조사기관에서 가장 보편적으로 쓰이고 있는 표본추출방법은 KT에서 제공한 전화번호부에 등재되어 있는 번호를 표집틀로 사용하는 것이다. 이러한 방법의 문제점은 전화번호부에 등재되어 있지 않은 전화번호들은 표본추출작업의 초기부터 제외된다는 것이다. 이 제외된 번호들과 등재된 번호들 간에 사회인구학적 또는 정치경제학적으로 구조적인 차이가 존재한다면, 전화번호부에 등재되어 있는 전화번호들만을 대상으로 표본을 구축하는 방법은 한국의 전체 성인 인구라는 모집단을 적절히 대변하지 못하는 문제점이 있다.

2011년 1월에 발표된 아산정책연구원의 데이터를 통해 분석한 결과, 전화번호부에 등재된 번호들과 등재되지 않은 번호들 간에는 사회인구학적, 정치경제학적인 차이가 엄연히 존재하며, 이는 대통령 지지도나 정당 지지도에도 영향을 미치는 것으로 나타났다. 본 연구에 따르면, 전화번호부에 등재된 표본은 등재되지 않은 표본에 비해 친한나라당적인 태도를 보이며 대통령 지지도도 높게 나타나는 것으로 나타났다.

주제어: 표집틀, 표본추출법, 임의번호걸기, 확률추출, 할당추출

This research aims to discuss the causes of inaccuracy in public opinion polls currently conducted in Korea. In particular, identifying the problems in sampling frame and sampling methods in political and social public opinion polls is an important

* 교신저자(corresponding author): 아산정책연구원 연구위원 우정엽.

E-mail: woo@asaninst.org

** 아산정책연구원 연구위원

*** 서울대학교 국제대학원 석사과정

question. Currently, most polling organizations operating in Korea are using phone number directories provided by Korea Telecom(KT) as its sampling frame for conducting most political polls. A critical problem of using a phone number directory as a sampling frame is that unlisted phone numbers can never be included in the sample. If a systematic difference in socio-demographic or politico-economic characteristics exists between the listed number using group and the unlisted group, using a phone number directory as a sampling frame cannot produce a sample that can represent the whole adult population in Korea. According to the poll result commissioned by the Asan Institute for Policy Studies in January 2011, there are statistically significant differences in socio-demographic and politico-economic characteristics between those two groups, and those differences led to the differences in the presidential job approval rating and party support. Our findings include that the listed number using group is more pro-Grand National Party and show stronger support for the president than the unlisted group.

key words: sampling frame, sampling method, random digit dialing, probability sampling, quota sampling

I. 서론

한국 정치에서 여론조사는 현재 각 당의 대통령 선거 후보자를 선정하는 데에 이용될 만큼 큰 위치를 차지하고 있다. 2002년 대통령 선거에서 노무현 후보와 정몽준 후보 간의 단일화과정에 여론조사가 사용된 이후, 대부분의 정당들이 상향식 공천의 방편으로 여론조사를 사용해 왔다. 이렇듯 큰 정치적 의미를 차지하고 있음에도 불구하고, 여론조사가 실행되는 방법이나 정확성에 대해서는 논란의 여지가 많다. 일례로, 지난 2010년 6월 2일 실시된 지방선거 과정에서 발표되었던 여론조사 결과는 여론조사의 신뢰성에 대해 심각한 의문을 갖게 하였다. 특히 서울시장 선거에서 선거 일주일 전 이루어졌던 YTN과 조선일보의 여론조사에 의하면 18%p 정도의 차이로 민주당의 한명숙 후보를 따돌리는 것으로 나타났던 한나라당의 오세훈 후보가, 실제 투표결과에서는 0.6%p밖에 차이가 나지 않았다. 뿐만 아니라, 타 지역에서도 예상 당선자와 실제 당선자가 차이가 나면서 한국의 여론조사는 그 신뢰성에 있어서 심각한 타격을 입었다.

이러한 결과의 원인에 대해서는 여러 의견들이 제시되었다. 일부에서는 ‘샤이 토리 효과(shy Tory effect)’에 빗대어 빗나간 여론조사를 설명하기도 했다. ‘샤이 토리 효과’란 1992년 영국 의회 선거에서 노동당에 뒤지고 있었던 것으로 알려졌던 보수당이 투표 결과에서는 제법 큰 차이로 노동당을 누르고 재집권에 성공한 것을 두고, 빗나간 여론조사의 원인을 실제로 보수당을 지지하면서도 사회적 분위기가 지지 의사를 명확히 밝히지 못했던 보수당 지지자들의 침묵 현상을 일컬으며 생긴 용어이다(Curtis 1997). 이는 곧 ‘침묵의 나선효과(spiral of silence effect)’와 일맥상통하는 개념으로, 주변의 다수가 자신과 다른 의견을 가지고 있다고 여길 때 자신의 의견을 피력하지 않고 침묵을 지키는 현상을 일컫는다(Noelle-Neumann 1984). 이러한 현상에 빗대어 한명숙 서울시장 후보나 다른 민주당 후보들을 지지하면서 여론조사 시에는 야당후보 지지를 밝히지 않았던 유권자들이 많았다는 분석도 있었다. 다른 한편으로는 현재 행해지고 있는 여론조사들이 구조적으로 친여 성향을 지닌 유권자들에 치우쳐 표본수집을 할 수밖에 없다는 지적이 있었다. 즉, 여론조사에서 사용되는 표본이 한국의 유권자들을 제대로 대표하지 못하기 때문에 이를 통한 조사결과가 편향되어 나타날 수밖에 없다는 것이다.

이 연구는 이들 중 두 번째 주장을 검증하는 데 목적을 두고 있다. 현재 한국에서 행해지고 있는 대부분의 여론조사에서 쓰이고 있는 표본추출방법은 전화번호부를 기준으로 한 할당추출법(quota sampling)이다. 이 방법은 엄밀한 의미에서 무작위추출이라고 보기 어려울뿐만 아니라 표본추출의 시작인 표집틀(sampling frame) 설정에 있어서 그 표집틀이 모집단을 대표하지 못한다는 문제를 가지고 있다. 기존의 연구들 역시 전화번호부를 이용한 표본추출방법에 대해 의문을 제기하고 있다(허명희 2007; 허명희·김영원 2008; 강현철 외 2008; 조성겸 외 2007; Glasser & Metzger 1972; Cummings 1979; Waksberg 1978; Crespi 1988).

이미 미국이나 캐나다, 영국 같은 여론조사 선진국에서는 임의번호걸기방법(RDD: Random Digit Dialing)을 널리 적용하고 있으나, 많은 비용과 시간 제약 등의 이유로 한국의 여론조사기관들과 여론조사 발주자들은 이 방법을 사용하는데 소극적인 모습을 보여 왔다. 이에 2011년 1월에 이루어진 아산정책연구원 여론조사는 임의번호걸기방법과 더불어 여론조사의 정확성을 높이기 위해 확률표본추출(probability sampling)의 기본적인 요소들을 적용하였다. 이 연구에서는 아산정책연구원의 여론조사 결과를 바탕으로 하여 기존의 전화번호부를 바탕으로 한 표본추출방법을 통해 도출되는 표본 집단들과 사회인구학적으로 어떤 차이가 얼마나 나는지, 그리고 그 차이가 대통령 지지도와 같은 민감

한 정치적 사안에도 영향을 미치는지를 면밀히 분석해 보도록 한다.

이 논문은 다음과 같은 순서로 구성되어 있다. 첫 번째 장에서는 기존의 임의번호결기사용에 관한 연구들을 소개하고, 다음 장에서 기존 한국의 여론조사에서 이용되는 표집틀에 대한 간략한 설명과 함께 현재 한국에서 주로 사용되고 있는 전화번호부를 이용한 할당추출법의 문제점과 이를 개선하기 위해 선진국에서 사용되고 있는 임의번호결기방법에 대해 간략하게 소개해 보도록 한다. 다음 장에서는 아산정책연구원에서 실시한 표본추출에 대한 간략한 설명과 함께 몇 가지 기술통계를 보여준다. 이어, 대통령 지지도와 정당 지지도에 대한 비교와 함께 기존의 방법을 채택할 경우 발생할 수 있는 오차에 대해 점검해 보도록 한다. 마지막 장에서는 새로운 여론조사기법의 당위성에 대해 논하도록 한다.

II. 전화번호부 방법과 임의번호결기방법

기존의 여론조사들이 많은 문제점을 노출하면서 학계를 중심으로 이를 극복하고자 하는 노력이 많았으나, 연구환경 등의 문제로 인하여 정작 임의번호결기방법을 통한 여론조사가 기존 전화번호부 이용방법의 여론조사와 어떤 차이가 있는지에 대한 연구가 많은 편은 아니었다. 임의번호결기의 여론조사에 대한 연구에서 두 가지 측면을 구별하여 살펴볼 필요가 있다. 첫째는 전화번호부에 번호가 등재되어 있는 집단과 등재되어 있지 않은 집단 간에 인구통계학적 특성의 차이가 있는가 하는 문제이고, 둘째는 이 두 집단 간에 정치적 사안에 대한 의견의 차이가 존재하는가 하는 문제이다.

두 집단 간에 특성의 차이가 존재하지 않는다고 하면, 전화번호부를 사용하는 표집틀과 임의전화번호를 사용하는 표집틀 간에 특성의 차이가 없다는 것을 의미하며, 결론적으로 더 많은 비용과 시간을 들이면서 임의번호결기에 의한 표본추출을 할 실제적 필요성이 적다는 것을 의미한다. 설령 두 집단 간에 특성의 차이가 있다고 하더라도, 이 특성의 차이가 두 집단 사이의 정치적 견해에 대한 차이를 이끌어내지 않는다면, 결과적으로 전화번호부에 의존하는 표본추출에 의한 결과나 임의번호결기에 의한 표본추출에 따른 조사결과가 큰 차이가 없기 때문에 조사를 실무적으로 대하는 조사 발주처나 조사업체에서는 비용과 시간이 더 드는 임의번호결기를 쓸 당위적 필요성이 줄어든다. 따라서 이 두 가지 측면에서 기존의 연구는 어땠는지 살펴보는 것이 의미가 있다고 할 수 있다.

이 문제는 외국의 학자들 사이에서 의견이 나뉘어 논의되고 있다. 1960년대와 1970년대 Cooper나 Brunner & Brunner, Glasser & Metzger 등은 임의번호결기에 의한 표본과 전화번호부에 의한 표본 사이에 인구통계학적 차이가 있음을 밝혀냈다(Cooper 1964; Brunner & Brunner 1971; Leuthod & Scheele 1971; Glasser & Metzger 1972, 1975; Guterbock et al. 2003). 이들의 연구는 이러한 인구통계학적 차이가 사회정치적 여론의 차이 또한 이끌어내는지에 대해서까지 심도있는 연구를 하지는 않았다. 하지만 이들 연구에서 나타난 차이 중 인종적 구성비의 차이가 매우 큼을 주목할 필요가 있다. 대체로 민주당(Democratic Party)을 지지하는 경향이 강한 흑인 그룹이 전화번호부에 등재되어 있지 않을 확률이 높다는 점에서 보건대, 전화번호부에 의한 표본에는 민주당 지지자들이 실제보다 적게 포함되었을 가능성이 높음을 암시하였다. Guterbock et al.(2003)은 등재 응답자와 비등재 응답자 간에 인구통계학적 차이도 존재할 뿐만 아니라, 미국 자동차 등록국(DMV: Department of Motor Vehicle)의 서비스에 대한 만족도 평가에 있어서도 차이가 난다는 사실을 밝혀냈다. 그들은 전화번호부에 등재되지 않은 응답자들이 조사에 포함된 자동차 등록국의 서비스 만족도에 대한 평가가 현저하게 낮음을 발견했고, 따라서 표본추출방법에 의해 중요한 여론조사결과가 편향될 수 있다는 점을 강조하면서, 정확성을 위해 임의번호결기를 사용할 것을 주장하였다(Guterbock et al. 2003).

한편 Curtis & Sparrow(1997)는 1992년과 1997년의 영국 의회 선거의 예측조사를 다시 한 번 면밀히 검사하면서, 당시 영국에서 주로 사용되던 할당표본추출법을 통한 조사는 노동당(Labour Party) 우위의 결과를 보여주는 편향적인 표본이 추출될 수밖에 없다고 주장하였다.

위의 연구들과는 달리 할당표본추출법에 의한 표본과 임의번호결기 표본 사이의 차이에 큰 의미를 두지 않아도 된다는 주장도 있다. Wilson et al.(1999)은 호주 주민들의 건강상태를 살펴보는 여론조사를 통해 임의번호결기 표본과 전화번호부 표본을 비교한 결과 약간의 차이는 존재하지만 크게 유의미한 정도는 아니며, 효율성 측면에서 굳이 시간과 비용이 많이 드는 임의번호결기를 사용할 이유가 없다고 주장하였다.

한국에 관한 연구 중 강현철 외(2008)는 임의번호결기 조사를 통해 전화번호부 등재 응답자와 비등재 응답자 간의 사회인구적 특성과 사회심리학적 성향을 비교하였다. 그들은 전화번호부에 기초한 표본추출이 사회경제적 측면에서 편향된 결과를 산출한다는 결론을 내렸다. 그러한 결론을 내린 이유는 비등재번호 응답자의 가구소득 수준이 등재번호 응답자에 비해 높았고, 학력수준 역시 비등재번호 응답자 집단이 더 높았을 뿐만 아

나라, 개인정보 침해에 대해서 더 큰 우려를 표시하는 등 보다 진보적인 경향을 보이기 때문이라고 하였다. 그들의 연구는 두 집단 간의 인구통계학적 특성의 차이는 밝혀냈으나, 주요 정치적 사안에 대한 질문이 조사에 포함되지 않아서 과연 두 집단 간에 정치적 견해의 차이가 존재하는가를 밝혀낼 수는 없었다.

허명회·김영원(2008)은 임의번호결기에 의한 표본과 전화번호부에 의한 표본 사이의 특성과 정치적 선택에 관한 비교를 하였다. 그러나 허명회·김영원의 연구의 한계는 그들이 사용한 임의번호결기 표본과 전화번호부 표본이 각기 다른 여론조사기관에 의해 추출되었다는 점이다. 즉, 각기 다른 조사기관에 의해 수집된 표본이므로, 두 표본 사이의 정치적 의견 차이가 추출방법의 차이에 의한 것인지 아니면 조사기관 차이에 따른 것인지 구분하기 어려운 점이 있다. 그러나 허명회·김영원은 일단 그러한 차이는 없는 것으로 가정하고 연구를 진행하였다. 그들의 조사 결과, 강현철 외(2008)과 마찬가지로 비등재번호 응답자 집단이 보다 고소득, 고학력인 것으로 나타나 두 집단 간에 인구사회학적 차이는 존재하지만, 대통령 후보 지지에 대해서는 두 집단 간에 유의미한 차이가 없는 것으로 나타났다. 따라서 등재 여부가 교육수준과 직업 및 소득 등에서는 차이를 가져오지만, 지지후보 선택에 있어서는 연관성이 없는 것으로 결론을 짓고 있다.

홍성준 외(2009)는 전화번호부 등재 표본의 한계를 인식해야 한다고 지적하고 있다. 최근 사생활보호 차원에서 전화번호부 등재를 원하지 않거나 유선전화를 아예 보유하지 않은 가구들의 증가로 인해 전화번호부를 통한 표본추출방식은 특정 집단을 편향적으로 포함할 위험성이 있으며, 이는 정확도가 떨어지는 조사결과를 낳을 수 있다고 주장한다.

요약하면, 기존의 연구는 전화번호부에 의한 여론조사와 임의번호결기에 의한 여론조사의 차이점에 대한 문제의식은 있었으나, 그러한 조사방법상의 차이가 정치적 사안에 있어서 구체적으로 어떠한 차이를 가져오는지에 대한 정확한 평가가 부재했다. 다음 장에서는 이러한 기존연구들이 주는 함의를 바탕으로 하여 기존의 한국 여론조사가 사용해 오던 표본추출방법에 대해 검토해 본다.

III. 한국의 여론조사 기법

이번 연구에서 사용된 임의번호결기방법에 따른 표본추출법을 설명하기에 앞서, 기존 한국의 여론조사에서 대부분 사용되던 표본추출방법에 대한 이해가 선행되어야 할 것

이다. 여론조사 결과가 사회의 여론을 대표할 수 있기 위해서 가장 중요한 점은 여론조사에 사용되는 표본이 모집단을 대표할 수 있어야 한다. 이를 위해서는 여론조사의 예비 응답자들이 모두 같은 확률을 가지고 최종적인 표본에 추출될 수 있어야 한다. 이와 관련하여 필자들이 지적하고자 하는 기존 한국의 여론조사 표본추출의 문제점은 크게 두 가지로 나눌 수 있는데, 첫 번째는 표집틀의 제한성이고 두 번째는 표본추출방법의 문제점이다.

우선 제한된 표집틀 사용과 관련한 문제를 살펴보도록 하겠다. 기존 한국의 여론조사는 사단법인 한국조사협회(KORA: Korea Research Association)에 데이터베이스 구입 비용을 지불하고 KORA로부터 전화번호부 데이터베이스를 제공받아 표집틀로 사용하여 왔다. KORA에서 제공하는 데이터베이스는 케이티스(KTIS, 구 한국인포서비스주식회사)에서 구입한 전화번호부 CD에 기초하고 있는데, ‘정보통신망 이용촉진 및 정보 보호 등에 관한 법률(이하 정보통신망법)’ 등에 의한 제약 등 여러 가지 이유로 인해 2009년부터는 더 이상 업데이트된 전화번호부 CD를 케이티스로부터 구매하지 못하게 되었다. 이에 따라, 여론조사회사들은 각자 구축하고 있는 데이터베이스에 의존하여 여론조사 표본 추출에 사용하고 있다.

문제는 여론조사회사들이 2009년까지 사용해 오던 케이티스 제공의 전화번호 데이터베이스와, 2009년 이후 자신들의 데이터베이스 구축에 기본적으로 사용한 번호가 큰 차이가 없다는 것이다. 필자들의 조사에 따르면, 대부분의 한국 여론조사회사들이 공식적으로 2009년까지 사용해 온 이 전화번호 데이터베이스에는 상업용 전화번호를 제외하고 약 762만 여개의 일반 전화번호가 올라와 있는 것으로 나타났다.

이 762만 여개의 전화번호에서 추출되어 여론조사의 표본으로 사용되는 번호들이 한국 여론을 대변하는 대표성을 갖기 위해서는 우선 이 762만 여개의 전화번호가 모집단을 충실히 반영하는 표집틀인가 하는 문제를 검토해야 한다. 2010년 인구주택총조사의 잠정적 집계에 따르면, 2010년 현재 우리나라의 총 가구 수는 1,733만 4천여 가구인 것으로 나타났다(통계청 2010). 1가구에서 2개 이상의 전화번호를 사용하고 있는 경우도 많이 있기 때문에, 가구 수를 기준으로 하면, 현재 여론조사회사들이 사용하고 있는 전화번호 데이터베이스의 번호가 1가구를 대표한다고 가정하더라도 총 가구 수의 43.9%밖에 포함하지 못하는 문제가 발생한다. 이렇게 되면, 표본추출 과정에서 전화번호부 등재번호 사용자와 비등재번호 사용자 사이에 표본으로 추출될 확률이 동일하지 않게 된다. 본 연구가 지적하고자 하는 것처럼, 만약 등재번호를 사용하고 있는 가구와 비등재번호를 사용

하고 있는 가구 사이에 인구통계학적, 정치사회적 특성의 차이가 존재한다면 기존의 데이터베이스에 의존한 표본추출은 정확한 여론조사 결과를 담보할 수 없는 것이다. 따라서 기존의 여론조사들이 표본추출에 사용해왔던 전화번호부 데이터베이스에 등재된 사용자와 전화번호부에 등재되어 있지 않은 번호를 사용하는 사람들 사이에 어떠한 차이가 존재하는지에 대한 연구가 필요한 것이다.

두 번째, 기존의 여론조사에서 응답자들을 대상으로 실사를 할 때 생기는 문제점으로서, 표본추출의 무작위성(randomness)이 위배되는 경우이다. 이것과 연관되어 기존의 여론조사가 가지는 세 가지 복합적인 문제점이 있다. 첫 번째는 기존의 여론조사에서는 무작위 확률표본추출법이 아닌 할당추출법(quota sampling)을 사용한다는 것이고, 두 번째는 시간과 비용의 제약을 이유로 전화 통화 시도 시 전화를 받는 사람만을 조사한다는 것, 그리고 세 번째는 재전화(call back) 시도를 하지 않는다는 점이다. 할당표본추출법은 대표적인 비확률적 표본추출법으로서, 조사결과 해석에 있어서 표본의 범위를 넘어서는 모집단으로의 일반화가 불가능한 것이다.

한국의 여론조사에서 사용하는 할당표본추출법은 확률적 표본추출방법의 일종인 층화표본추출법(stratified sampling)과 일견 비슷한 면이 있으나, 조사대상 선정에 있어서 무작위성 원칙을 지키지 않고 재전화를 시도하지 않기 때문에, 할당표본추출법으로 간주된다. 일부 한국의 여론조사에서 층화무작위표본추출법을 사용하였다고 하는 경우가 있으나, 대부분의 경우 수화기를 드는 응답자를 바로 조사하고 재전화 시도를 하지 않기 때문에, 엄밀히 말해서 확률적 표본추출법의 일종인 층화표본추출법이 아닌 할당추출법이라고 해야만 옳은 표현이다. 위에서도 언급하였지만, 표본이 대표성을 갖기 위해서는 표본에 선택될 수 있는 성인들이 모두 같은 확률을 가지고 표본에 선택될 수 있어야 한다. 그러나 전화 시도에서 전화를 받는 사람만을 대상으로 조사를 실시하게 될 경우, 직장이나 학교에서 시간을 보내는 사람들의 경우에는 조사대상에 포함될 확률이 낮아지는 반면, 집에서 보내는 시간이 많은 계층, 예를 들면, 주부나 노년층의 경우에는 조사대상에 포함될 확률이 높아지게 된다. 그렇게 된다면, 두 집단, 다시 말해 집에서 보내는 시간이 많은 계층과 대부분의 시간을 집 외부에서 보내는 계층 사이에는 최종 표본에 포함될 확률이 동일할 수가 없게 되기 때문에, 표본추출의 원칙인 동일 확률에 의한 표본 선택을 위배하게 된다.

이러한 점을 시정하기 위해 사용할 수 있는 방법 중에는 예를 들면, '생일법'이 있다. 이 방법은 전화를 걸어 처음 전화를 받는 사람(수신자)을 대상으로 조사하는 것이 아니

고, 수화기를 든 사람에게 그 가구 내에서 최근 생일이 지난 사람을 바꿔 달라고 요청하여 그 사람에 대해 조사를 하는 것이다. 그럼으로써 최초 수화기를 드는 사람에 대해 조사를 하면서 발생하는 표본 선택의 편향성을 극복하게 된다. 여기에서 최근 생일이 지난 사람이 외부에 있거나 전화 통화를 할 수 없는 상황인 경우, 그 선택된 가구를 포기하고 다른 가구로 대체하는 것이 아니라, 그 가구에 대하여 재통화를 시도해야만 처음 추출된 표본을 유지하여 표본추출의 무작위성을 지킬 수 있게 된다.

그러나 기존의 여론조사는 이러한 절차를 지키지 않은 채, 전화를 받는 사람을 바로 조사대상으로 삼고, 전화를 받지 않는 경우에는 바로 다음 표본으로 대체하면서, 예비표본으로 그 조사대상을 바꾸곤 했다. 이렇게 할 수밖에 없는 이유 중의 하나는, 위에서 언급한 것처럼, 애당초 표본추출이 확률표본추출이 아닌 할당표본추출이라는 점에 있다. 할당표본추출법은 기존의 연구에서도 그 문제점이 지적되어 왔다(허명희·김영원 2008; 허명희·황진모 2006; 조성겸 외 2007; 현경보 2005; 홍내리·허명희 2001; 조성겸 1997).

할당표본추출법에 따른 표본추출은 우리나라의 성, 연령, 지역에 따른 인구비례에 따라 목표된 표본할당량을 가지고 시작한다. 할당표본추출법에 따르는 조사인 경우에는 특정 계층의 목표량이 충분히 확보되어 정해진 할당 매트릭스(quota matrix)상의 셀(cell)이 채워지면, 그 셀에 해당하는 계층의 조사를 중단하고 목표량이 충분히 확보되지 않은 계층을 찾아 나서게 된다. 만약 경기도에 거주하는 20대 남성이 충분히 채워지지 않으면, 그 특정 계층을 찾을 때까지 조사를 하게 되는 것이다. 할당표본추출은 비용과 시간상 여러 가지 장점에도 불구하고, 조사자가 피조사자를 선택하는 문제가 발생하여 표본추출의 무작위성을 위배하게 된다. 이러한 문제가 있기 때문에 할당표본추출법은 대표적인 비확률표본추출법이 되는 것이다.

IV. 데이터와 방법론

이 연구에 쓰인 데이터는 2011년 1월 아산정책연구원에서 실시한 대통령 지지도 설문조사이다. 이 여론조사는 ㈜리서치앤리서치에 의뢰해 2011년 1월 11일부터 17일까지 진행되었다. 전국의 만 19세 이상 성인남녀 1,003명을 대상으로 하여 컴퓨터를 이용한 전화면접조사(Computer Assisted Telephone Interviewing: CATI)로 이루어졌다.

아산정책연구원에서 실시한 이 연구는 위에서 언급한 세 가지 문제점들을 각각 다음과 같이 개선함으로써 표본의 무작위성을 확보하였다

먼저 컴퓨터 프로그램을 이용한 임의번호결기를 통해 표본을 추출하여 표본의 무작위성(randomness)을 최대화하였다. 임의번호결기방법은 현재 전화번호부에 등재된 번호를 대상으로 조사하는게 아니라 국번 이외의 마지막 네 자리 번호를 컴퓨터에서 무작위로 생성한 뒤 조사하는 방법이다. 기존 여론조사들이 전화번호부에 등재된 번호들을 모집단으로 하여 표본을 추출하는 방법을 사용하기 때문에 생기는 비등재 번호들의 누락을 피할 수 있다. 이와 같은 방법을 사용할 경우 비용과 시간이 많이 소요된다는 단점이 있지만 조사의 정확성을 더 높일 수 있다. 표본추출 후 성별, 연령별, 지역별로 사후층화(post-stratification)를 적용함으로써 조사결과를 보정하였다. 추후 표본에 포함된 응답자들이 전화번호부에 등재되어 있는지 여부를 기록하여 등재자와 비등재자 간의 비교를 가능하게 하였다.

둘째로, 이 조사에서는 처음 작성한 무작위성 확보를 위해 확률표본추출의 두 가지 요소를 충실히 이행하려고 하였다. 그 하나는 조사대상 선정에 있어서 생일법을 적용하는 것이고, 두 번째는 재통화를 하는 것이다. 조사대상 선정의 임의성이 확률적 표본추출의 기본적 요소임에도 불구하고 기존의 많은 조사들이 시간과 비용 문제 때문에 전화를 처음 받는 사람을 조사한 것에 비해 보다 조사방법론에 충실한 방법이다. 무작위 표본에 해당하는 응답자가 전화를 받지 않거나 적격 응답자와 접촉이 가능하지 않을 경우 시차를 두고 5회 재통화(call back)를 실시하여, 초기에 구성된 무작위 표본에 충실하도록 하였다. 조사결과 응답률은 평균 22%를 나타냈으며 95% 신뢰수준에서 $\pm 3.1\%$ 오차수준을 나타냈다. 응답률은 응답자가 수화기를 든 경우 중에서 조사가 완료된 경우의 비율을 의미한다.

마지막으로 응답순서효과(response order effect)에 따라 생기는 편향성을 방지하고 응답자들이 한 답변의 순차적 패턴을 따르지 않도록 하기 위해 답변들의 순서를 바꾸가며 질문을 하는 방법을 도입하였다. 편향성은 설문문항의 지문이 조사대상자에게 제시되는 순서에 따라 조사대상자가 첫 번째 나온 지문에 많이 응답을 하는 초두효과(primacy effect)와 제일 마지막에 들은 지문에 많이 응답을 하는 최신효과(recency effect)가 혼재하여 발생할 수 있다. 이러한 편향성을 극복하기 위해 대부분의 설문에서 지문 제시 순서를 바꾸어 무작위화하였다. 이러한 방법을 적용하여 표본 모집의 무작위성과 포괄성을 최대한 지키도록 하였고, 이를 통해 좀더 정확한 분석이 가능하도록 하였다.

V. 분석결과

〈표 1〉은 등재번호와 비등재번호 응답자 그룹 간의 사회인구학적 차이를 정리해 놓았다. 전체 조사된 응답자들의 전화번호 중 37%만이 전화번호부에 등재되어 있었고, 나머지 63%에 달하는 응답자들은 전화번호부에 등재되어 있지 않은 것으로 나타났다. 이러한 등재자와 비등재자 간에 사회인구학적으로 차이가 있는지를 살펴보기 위해 연령, 학력, 소득수준, 직업군별로 비등재 그룹과 등재 그룹 간의 비율을 비교해 보았다.

연령별로는, 등재자 그룹은 확실히 고령층이 높은 비율을 차지하고 있었다. 등재자 그룹 중 30%에 달하는 응답자들이 60대 이상으로, 이는 비등재자 그룹에선 60대 이상의 고령층이 15%에 불과하는 것과 비교되는 수치이다. 특히 3,40대 연령층이 등재자 그룹에 비해 비등재자 그룹에서 상대적으로 비율이 높고, 30대 연령층이 비등재자 그룹에서 차지하는 비율은 등재자 그룹에서의 비율의 두 배를 넘는 수치다. 다시 말해, 등재자 그룹에서 가장 높은 비율을 차지하는 연령층은 60대 이상의 노년층인 반면, 비등재자 그룹에서 가장 높은 비율을 차지하는 연령층은 30대 연령층이라는 차이가 존재한다.

학력수준별로도 등재자 그룹과 비등재자 그룹은 차이를 보인다. 예를 들어, 비등재자 그룹의 11% 정도만 중학교 졸업자인데 반해 등재자 그룹의 약 27%가 중학교 졸업의 학력을 지닌 것으로 나타났다. 등재자 그룹에서는 고등학교 졸업의 학력을 지닌 응답자들이 25.5%를 차지하고 있었고, 비등재자 그룹에서는 33.9%였다. 또한 전문대 이상 재학의 학력을 가진 응답자들이 비등재자 그룹에서는 과반이 넘는 55%인데 비해 등재자 그룹에서는 47%로 그보다 낮게 나타났다. 전체적으로 비등재자 그룹의 90% 가까이가 고등학교 졸업 이상의 학력을 가진데 반해, 등재자 그룹에서 고등학교 졸업 이상의 학력을 가진 응답자들의 비율은 72.5%로 비등재자 그룹보다 약 18%정도 낮았다.

소득수준별로도 역시 구성비에서 차이가 난다. 월 소득 200만원 이하의 응답자들의 비율이 등재자 그룹에서 높은 반면, 월 소득 200만원 이상인 응답자들의 비율은 비등재자 그룹에서 높게 나타났다.

마지막으로 직업군으로 분류하여 살펴보았다. 비등재 그룹과 등재 그룹이 가장 큰 차이를 보이는 직업군은 화이트 칼라로, 등재 그룹의 화이트 칼라 계층 비율은 6.2%에 지나지 않는 반면 비등재 그룹의 화이트 칼라 계층 비율은 16%나 되었다. 다음으로 큰 차이를 보이는 직업군은 농업과 임업에 종사하는 1차 산업 종사자 계층이었다. 등재자 그룹의 1차 산업 종사자 비율은 10%로 높은 반면, 비등재자 그룹에서는 1%밖에 되지 않았다.

〈표 1〉 전화번호 등재 여부에 따른 사회인구학적 특성

(단위: %)

		등재	비등재
전체		37.0	63.0
연령	20대	20.6	19.0
	30대	11.5	25.1
	40대	19.2	24.8
	50대	19.2	16.3
	60대 이상	29.5	14.8
학력	중학교	26.9	11.0
	고등학교	25.5	33.9
	전문대 이상 재학	47.0	55.0
소득수준	100만원 이하	25.9	10.4
	101만원~200만원 이하	13.9	11.8
	201만원~300만원 이하	13.0	21.7
	301만원~400만원 이하	8.5	17.1
	401만원~500만원 이하	10.0	12.5
	501만원 이상	11.1	14.5
직업군	1차 산업	10.0	1.0
	자영업	15.3	16.1
	영업직	2.0	4.0
	기술직	2.9	2.4
	블루 칼라	3.1	2.5
	화이트 칼라	6.2	16.0
	전문직	1.6	1.2
	주부	27.1	31.7
	학생	16.0	13.1
	미취업/기타	20.0	11.9

참고: 위 표의 수치들은 모두 지역, 성별, 연령층에 따라 사후 가중치가 주어진 수치임.

위와 같은 전화번호부 등재 여부와 사회인구학적 차이가 대통령 지지율과 정당 지지도의 차이로 확대될 수 있는지를 알기 위해, 먼저 사회인구학적 차이가 대통령 지지율과 정당 지지도 차이를 도출해 내는지를 살펴보았다. <표 2>를 보면 대통령에 대한 긍정적 평가 그룹과 부정적 평가 그룹 간의 사회인구학적 비율이 나와 있다. 전체적으로 긍정적 평가가 44.4%이고 부정적 평가가 46.7%인 가운데 보통 혹은 무응답 비율이 8.9%를 차지하였다.

연령대별로 살펴보면 20대와 30대 연령층에서 대통령에 대한 부정적 평가가 긍정적 평가보다 훨씬 높음을 알 수 있다. 20대 응답자들 중 긍정적으로 평가한 응답자는 32.3%에 지나지 않는 반면, 부정적 평가는 이의 두 배 가까이 되는 63.2%에 달한다. 30대의 경우는 긍정적 평가자가 27.6%밖에 되지 않았고, 부정적 평가자는 두 배를 넘는 63.8%나 되었다. 반면, 50대와 60대 이상 응답자들은 대통령에 대해 긍정적으로 평가하는 경향이 뚜렷했다. 50대 응답자의 55.6%가 긍정적으로 평가한 반면 부정적으로 평가한 응답자는 35.1%에 지나지 않았다. 60대 응답자의 경우는 무려 62.5%가 긍정적으로 평가했지만, 부정적으로 평가한 응답자는 이의 3분의 1 수준인 21.7%였다. 이처럼 연령층별로 대통령의 국정 운영에 대한 의견이 뚜렷이 갈리는 현상을 보여주고 있다.

이러한 현상은 학력이나 소득수준에서도 볼 수 있다. 고학력자일수록 부정적으로 평가하는 응답자가 많았고 학력 수준이 낮아질수록 긍정적으로 평가하는 응답자가 많았다. 또한 고소득자일수록 부정적으로 평가하는 경향이 있었으며, 저소득계층일수록 긍정적으로 평가하는 경향이 강했다.

다음으로는 정당지지도와 사회인구학적 분류를 연관지어 살펴보았다. <표 3>에 나타나 있듯이 전체적으로 한나라당 지지자는 30% 정도였고 민주당 지지자는 17.95를 차지했다. 연령대별로 살펴보았을 때, 20대에서는 민주당 지지 비율이 한나라당 지지 비율보다 높은 반면에 40대 이상일수록 민주당보다는 한나라당을 지지하는 비율이 높았다. 전반적으로 연령을 제외한 학력, 소득수준, 직업군 모두에서 한나라당 지지자의 비율이 높게 나타났다. 민주당 지지율은 소득 수준이 400만원과 500만원 사이인 중상위층과 1차 산업 종사자들, 학생, 그리고 기술직에 종사하는 응답자들 사이에서 유난히 높게 나타났다.

〈표 2〉 사회인구학적 요인에 따른 대통령 지지율

(단위: %)

		긍정적 평가	부정적 평가	보통
전체		44.4	46.7	8.9
연령	20대	32.3	63.2	4.5
	30대	27.6	63.8	8.6
	40대	45.1	48.3	6.7
	50대	55.6	35.1	9.3
	60대 이상	62.5	21.7	15.9
학력	중학교	55.7	25.9	18.5
	고등학교	50.3	41.2	8.5
	전문대 이상 재학	37.2	56.8	6.0
소득수준	100만원 이하	52.1	31.0	16.9
	101만원~200만원 이하	53.6	39.1	7.3
	201만원~300만원 이하	39.7	52.8	7.4
	301만원~400만원 이하	43.9	51.4	4.8
	401만원~500만원 이하	35.3	56.0	8.8
	501만원 이상	49.6	42.2	8.2
직업군	1차 산업	37.7	43.1	19.2
	자영업	50.4	37.3	12.4
	블루 칼라	33.3	61.1	5.6
	화이트 칼라	39.8	57.5	2.7
	전문직	51.9	36.9	11.2
	주부	27.8	67.3	4.9
	학생	52.5	37.8	9.7
	미취업/기타	32.9	67.1	0.0

참고: 위 표의 수치들은 모두 지역, 성별, 연령층에 따라 사후 가중치가 주어진 수치임.

〈표 3〉 사회 인구학적 요인에 따른 정당 지지도

(단위: %)

		한나라당 지지	민주당 지지
전 체		30.0	17.9
연 령	20대	18.9	23.7
	30대	18.8	19.3
	40대	28.7	19.8
	50대	37.2	16.4
	60대 이상	47.3	10.3
학 력	중학교	36.2	15.4
	고등학교	31.0	17.4
	전문대 이상 재학	27.3	19.2
소득수준	100만원 이하	34.6	17.2
	101만원~200만원 이하	29.2	18.1
	201만원~300만원 이하	28.2	20.1
	301만원~400만원 이하	35.5	17.6
	401만원~500만원 이하	18.4	23.6
	501만원 이상	38.8	14.3
직업군	1차 산업	25.9	35.3
	자영업	38.2	11.9
	영업직	33.1	18.2
	기술직	5.5	22.1
	블루 칼라	22.3	25.5
	화이트 칼라	22.8	14.7
	전문직	51.8	0.0
	주 부	33.0	17.5
	학 생	17.6	30.4
	미취업	37.4	7.4
	기 타	46.7	13.6

참고: • 기타 정당 지지자 비율과 무당과 비율은 위의 표에서 생략되었다.

• 위 표의 수치들은 모두 지역, 성별, 연령층에 따라 사후 가중치가 주어진 수치임.

이처럼 사회인구학적으로 대통령 지지도나 정당 지지율이 다르게 나타난다면, 사회인구학적으로 다른 특성을 보이는 등재 집단과 비등재 집단도 대통령 지지도나 정당 지지율에서 차이를 보일 것이라 예측된다. 이를 위해 t -검정을 통해 등재자와 비등재자 집단 간에 대통령 지지도와 정당, 그리고 집권당에 대한 지지율의 차이가 존재하는지, 만일 그렇다면 이 차이가 통계학적으로 의미를 갖는 것인지를 분석해 보았다. 분석의 편의를 위해서 대통령의 국정운영에 대한 답변을 ‘지지한다’와 ‘지지하지 않는다’는 두 척도로 간편화하였다. ‘지지하지 않는다’라고 답변한 응답자들은 0으로, ‘지지한다’고 답변한 응답자들은 1로 재코딩하였다. 따라서 평균치는 곧 지지율을 의미한다고 할 수 있다. 또한 무응답 혹은 ‘잘 모르겠다’라고 답변한 집단은 분석에서 제외하였다.

분석결과는 <표 4>와 <그림 1>에 나타나 있다. 먼저 ‘잘 모르겠다’나 무응답으로 답변한 집단을 포함해서 국정운영 지지율을 비교해 보면, 등재집단의 대통령 지지율은 약 48%이고 비등재 집단의 대통령 지지율은 42%이다. 대통령의 국정운영에 대해 부정적으로 평가하는 비율은 등재 집단에서는 41%이고 비등재 집단에서는 50%로 나타났다(<그림 1>). ‘잘 모르겠다’나 무응답 답변 집단을 제외하였을 경우 등재 집단의 대통령 지지율은 약 54%(.54)인 반면, 비등재 집단의 대통령 지지율은 45%(.45)였다. 따라서 대통령 지지율에 대한 등재 집단과 비등재 집단의 차이는 약 .09로 9% 정도 되었다. 차이가 통계적으로 유의미한 것인지를 살펴보기 위해 t -검정을 사용했다. 그 결과, t -값은 2.326로 산출되었고(자유도 872), p -값이 임계값 0.05보다 낮은 .02로 통계적으로 유의미한 차이라고 할 수 있다.

다음으로는 2011년 현재 집권당인 한나라당 지지도가 비등재 그룹과 등재 그룹 간에 차이를 보이는 지를 살펴보았다. <표 5>에 기술되어 있듯이, 집권당인 한나라당 지지율은 전화번호부 등재 집단에서는 약 36%인데 반해 비등재 집단에서는 27%에 지나지 않아 그 차이가 약 9%, 즉 .09정도 되었다. t -검정을 통해 분석한 결과, t -값은 2.963(자유도 958), p -값은 .003으로 나타나 등재 집단과 비등재 집단 간의 한나라당 지지율의 차이 역시 통계적으로 매우 유의미하다. 따라서 이 두 집단은 대통령 지지와 더불어 한나라당 지지에 있어 중요한 차이를 보여준다고 할 수 있다. 다시 말해 전화번호부 등재 집단에 비해 비등재 집단은 대통령과 집권당에 훨씬 낮은 지지를 보내고 있는데, 이는 우연히 도출된 결과가 아니라 통계상으로 의미가 있는 차이임이 입증되었다.

〈표 4〉 전화번호부 등재 여부에 따른 대통령 지지도

(단위: 명)

전화번호부 등재 여부	표본수	평균	표준편차	t-score	자유도	p-값
등재	318	0.54	0.5	2.326	872	0.02
비등재	556	0.45	0.5			

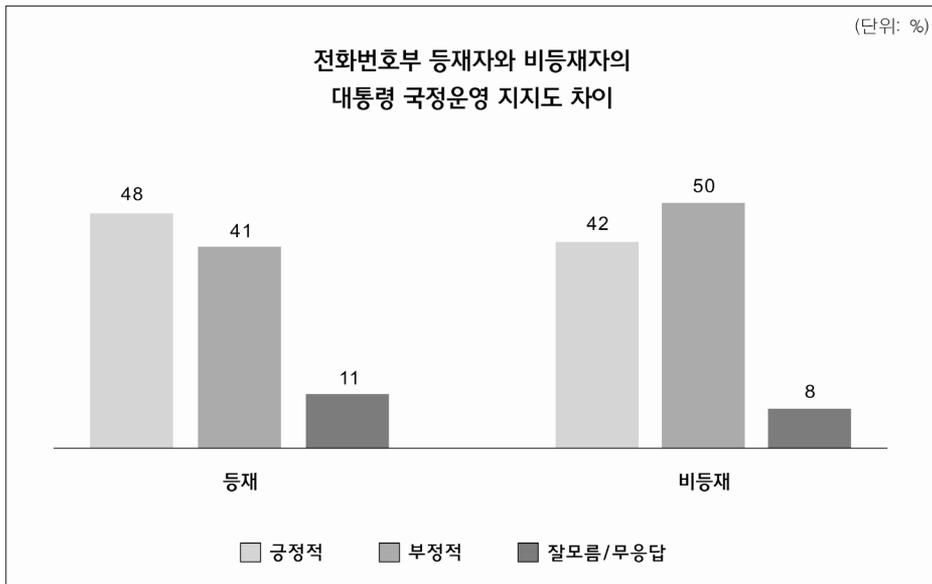
참고: 위 표의 수치들은 모두 지역, 성별, 연령층에 따라 사후 가중치가 주어진 수치임.

〈표 5〉. 전화번호부 등재 여부와 집권당(한나라당) 지지도

(단위: 명)

전화번호부 등재 여부	표본수	평균	표준편차	t-score	자유도	p-값
등재	356	0.36	0.48	2.963	958	0.003
비등재	604	0.27	0.44			

참고: 위 표의 수치들은 모두 지역, 성별, 연령층에 따라 사후 가중치가 주어진 수치임.



〈그림 1〉 전화번호부 등재 여부와 대통령 국정운영 지지도

〈표 6〉 이항 로짓 분석 결과

	로짓모형[1]	로짓모형[2]
전화번호부 등재	0.42** (0.16)	0.14 (0.16)
연 령		0.45*** (0.08)
교육수준		-0.07 (0.14)
남 성		0.07 (0.16)
소득수준		0.00 (0.06)
서 울	1.50*** (0.26)	1.71*** (0.28)
영 남	1.88*** (0.25)	2.13*** (0.27)
경 기	1.33*** (0.25)	1.57*** (0.27)
충 청	1.82*** (0.40)	1.50*** (0.32)
강원 및 제주	1.32*** (0.30)	1.92*** (0.40)
상 수	-1.34*** (0.22)	-2.97*** (0.63)
유사 R 제공	0.07	0.13
카이 자승	63.33	102.05
로그 우도	-515.77	-483.69
사례수	801	801

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

마지막으로 이러한 등재 집단과 비등재 집단 간의 대통령 지지율 차이가 두 집단을 구성하는 응답자들의 사회인구학적 특성에 기인한 것인지, 아니면 사회인구학적 특성을 통해 설명할 수 없는 비등재 집단과 등재 집단 간의 구조적 차이가 존재하는지를 살펴볼 필요가 있다. 이를 위해 이항 로짓 분석을 시도하였다. 종속변수는 이명박 대통령에 대한 지지 여부로 이명박 대통령이 대통령직을 잘 수행한다고 생각하는 '긍정적 집단'과 잘 수행하지 못하고 있다고 생각하는 '부정적 집단'으로 이분화한 변수를 사용하였다. 가장 중요한 독립변수는 전화번호부 등재 여부이다. 등재 집단을 '1'로, 비등재 집단을 '0'으로 코딩하여 더미 변수화하였다. 이 외에 연령, 성별, 교육 수준, 소득 수준 및 거주지역 등 다양한 사회인구학적 변수들을 회귀분석모형에 포함하였다.

첫 번째 로짓 모형은 여타 변수들을 고려하지 않은 채 전화번호부 등재 여부와 거주지역만을 독립변수로 포함하여 종속변수인 이명박 대통령지지 여부를 설명하려 하였다. <표 6>에서 회귀계수는 0.42, 표준오차는 0.16으로 유의수준 0.05에서 통계학적으로 유의미한 결과로 나타났다. 즉, 전화번호부에 등재된 응답자들은 등재되지 않은 응답자들에 비해 이명박 대통령을 지지할 확률이 높다고 할 수 있다.

두 번째 로짓 모형은 첫 번째 모형에 그 외의 사회인구학적 변수들을 포함한 것이다. 만일 이러한 사회인구학적 변수들이 통제된 후에도 전화번호부 등재 여부 변수의 회귀계수가 통계학적으로 유의미한 것으로 나타난다면, 전화번호부 등재 집단과 비등재 집단 사이에는 단순히 사회인구학적 요소가 아닌 다른 정치 성향적인 구조적 특성의 차이가 존재한다고 할 수 있다. 그러나 사회인구학적 변수들이 포함된 결과에서 전화번호부 등재 여부 회귀계수가 통계학적으로 더 이상 유의미하지 않은 것으로 나타난다면 전화번호부 등재 여부로 인한 이명박 대통령 지지 여부의 차이는 사실상 사회인구학적 요인의 차이에 기인한다고 유추할 수 있다.

<표 6>의 두 번째 로짓 모형 결과를 보면 알 수 있듯이, 전화번호부 등재 여부 변수의 회귀계수는 더 이상 통계학적으로 유의미하지 않게 나타난다. 회귀계수의 수치가 반 이상 줄었으며, 그로 인해 전화번호부에의 등재 여부는 이명박 대통령을 지지할 확률을 효과적으로 설명하지 못한다. 대신 연령 변수가 매우 큰 설명력을 가진 것으로 나타났다. 나이가 많은 응답자들일수록 이명박 대통령의 대통령직 수행에 대해 긍정적으로 평가하고 있었다. 이러한 결과를 통해 전화번호부 등재 여부와 대통령의 대통령직 수행평가 사

이에 나타났던 상관관계는 사실상 등재 집단과 비등재 집단 사이에 존재하는 연령층의 구성비 차이로 인한 것이었다고 할 수 있다.

VI. 결론

그동안 한국의 여론조사는 비용과 시간의 제약 때문에, 그리고 임의번호걸기에 의한 표본과 전화번호부에 의한 표본 사이에 조사결과가 유사한 점이 있다는 이유로 그 이론적 적합성에도 불구하고 전화번호부에 의한 할당표본추출에 따른 여론조사를 계속해 왔다. 전화번호부에 등재된 번호를 사용하는 응답자와 비등재 번호를 사용하는 응답자 사이에 차이가 없다면, 이론적으로 아무리 임의번호걸기방법이 우수하다고 하더라도 실제 적용에 있어서 임의번호걸기방법을 사용해야 한다고 주장하기란 어려운 것이다.

그러나 이번 연구에 의하면, 등재번호 응답자와 비등재번호 응답자 사이에는 인구통계학적 특성이 존재할 뿐만 아니라 중요한 정치적 견해에 있어서도 유의미한 차이를 보이고 있음이 입증되었다. 이번 연구에 사용된 여론조사 방법은 크게 두 가지 측면에서 기존의 여론조사들과 차이를 보이고 있는데, 첫째, 전화번호부를 이용한 표본추출이 아닌 임의번호걸기방법에 의한 표본추출이라는 점, 둘째, 생일법과 재통화를 엄격히 실시함으로써 표본의 무작위성을 확보하려고 노력했다는 점이다.

조사결과, 비등재번호 응답자들은 등재번호 응답자들에 비해 소득수준과 교육수준이 높은 것으로 나타났고, 직업군 구성에 있어서도 차이를 보였으며, 젊은 연령층이 많이 포함되어 있었다. 이러한 결과는 기존의 연구결과들과 많은 부분 일치하는 것이다. 여기서 더 나아가, 이번 연구결과는 두 집단 간에 정치적 견해에 있어서도 유의미한 차이를 보이고 있다는 것이다. 등재번호 응답자들이 전통적으로 친여성향의 인구통계학적 특성을 가지고 있는 것으로 나타나고, 반대로 비등재번호 응답자들이 전통적으로 친야성향의 인구통계학적 특성을 가지고 있는 것으로 나타나는 것과 일치하여, 두 집단의 정치적 견해 선택에도 그와 유사한 차이가 나타난 것이다. 등재번호 응답자들이 보다 대통령과 한나라당에 우호적인 의견을 표명하였고, 비등재번호 응답자들은 그렇지 않은 것으로 나타났다. 따라서 한국의 통신환경을 감안하였을 때 보다 정확한 정치 여론조사를 위해서는 전화번호부를 이용한 할당표집에서 벗어나 임의번호걸기에 의한 확률표집법을 사용하는 것이 바람직하다고 할 수 있다.

물론 임의번호걸기에 의한 여론조사가 장점만 있는 것은 아니다. 조사설계에서 보듯이, 임의번호걸기에 의한 확률표본추출법을 사용하게 되면 더 많은 시간이 소요되고 그에 따라 더 많은 비용이 들게 된다. 임의번호걸기에 의해 생성된 번호 중에는 많은 무효 번호가 존재하기 때문에 그에 따른 시간 소모가 있고, 표본의 무작위성 확보를 위해 생일법을 사용하고 또 재통화를 하게 되면, 기존의 전화번호부에 의한 할당추출법보다 많은 시간이 소요될 수 있다. 그에 따라 그만큼의 비용이 더 소모되는 것은 물론이다.

따라서 조사발주자와 조사업계, 그리고 조사결과의 소비자들은 이러한 차이점을 유의하여 조사결과를 소비하는 데 있어서 비용 효과 분석을 통해 조사 의도에 알맞은 조사 방법을 선택하는 것이 중요하다고 할 수 있다.

참고문헌

- 강현철 · 한상태 · 김지연 · 정용찬 · 허명희, 2008. “임의번호 걸기 전화조사와 주요결과.” 《조사연구》 9권 1호: 1-22.
- 조성겸 · 김지연 · 나윤정 · 이명진, 2007. “선거여론조사의 문제점과 개선방향: 2006년 지방선거를 중심으로.” 《조사연구》 8권 1호: 31-54.
- 통계청, 2010. “2010 인구주택 총 조사 잠정집계 결과 보도자료.”
- 허명희 · 김영원, 2008. “임의번호 걸기 표본 대 전화번호부 표본: 2007년 대통령 선거 예측 사례.” 《조사연구》 9권 3호: 55-69.
- 허명희 · 황진모, 2006. “전화조사를 위한 시간균형할당표본추출.” 《조사연구》 7권 2호: 39-52
- 허명희, 2007. “여론조사 방법론: 과제와 전망.” 대한통계협회 《통계》 33권 1호: 27-36.
- 현경보, 2005. “선거 예측조사의 정확도와 그 영향요인 연구: 17대 총선 예측조사 결과의 사후분석.” 《언론과학연구》 5권 1호: 301-336.
- 홍내리 · 허명희, 2001. “제 16대 국회의원 선거의 예측조사에 대한 사후적 검증.” 《조사연구》 2권 1호: 1-35.
- 홍성준 · 박소형 · 김선웅, 2009. “국내 유선 전화조사에서 표본추출틀의 포함률.” 《조사연구》 10권 1호: 33-56.
- Brunner, A. James and Brunner G. Brunner, 1971. “Are Voluntarily Unlisted Telephone Subscribers Really Different?” *Journal of Marketing Research* Vol. 8, No.1: 121-124.

- Cooper, L. Sanford. 1964. "Random Sampling by Telephone: An Improved Method." *Journal of Marketing Research* Vol.1, No. 4: 45-48.
- Crespi, Irving. 1988. *Pre-election Polling: Sources of Accuracy and Error*. New York: Russell Sage Foundation.
- Cummings, K. Michael. 1979. "Random Digit Dialing: A Sample Technique for Telephone Surveys." *The Public Opinion Quarterly* Vol. 43, No. 2: 233-244.
- Curtis, J. 1997. "So How Well Did They Do? The Polls in the 1997 Election." *Journal of the Market Research Society* 39(3): 449-461.
- Curtis, J.K., and N. Saproow. 1997. "How Accurate Are Traditional Quota Opinion Polls?" *International Journal of Market Research* 39(3): 433-448.
- Glasser, J. Gerald. and D. Gale Metzger. 1972. "Random-Digit Dialing as a Method of Telephone Sampling." *Journal of Marketing Research* Vol. 9, No. 1: 59-64.
- Glasser, J. Gerald. and D. Gale Metzger. 1975. "National Estimates of Nonlisted Telephone Households and Their Characteristics." *Journal of Marketing Research* Vol. 12, No. 3: 359-361
- Guterbock, Thomas M., David E. Hartman and Ryan A. Hubbard. 2003. "RDD vs. Listed: An Experimental Study of Coverage Error, Costs and Non-Response in a Statewide Telephone Survey." *Paper Presented at the Annual Meeting of the American Association for Public Opinion Research*, Nashville, TN. May.
- Leuthold, A. David. and Scheele, Raymond. 1971. "Patterns of Bias in Samples Based on Telephone Directories." *The Public Opinion Quarterly* Vol. 35, No. 2: 249-257.
- Noell-Neumann, Elisabeth. 1984. *The Spiral of Silence: Public Opinion - Our Social Skin*. Chicago: University of Chicago.
- Waksberg, J. 1978. "Sampling Methods for Random Digit Dialing." *Journal of the American Statistical Association* 73: 40-46
- Wilson, D.H., Starr, G.J., A.W. Talyor and E.D. Grande. 1999. "Random Digit Dialing and Electronic White Pages Samples Compared: Demographic Profiles and Health Estimates." *Australian and New Zealand Journal of Public Health* 23: 627 - 633.

<접수 2011/3/3, 1차수정 2011/3/11, 2차수정 2011/3/17, 게재확정 2011/3/18>