

조사동향

유·무선전화 병행조사에 대한 연구:

2011년 서울시장 보궐선거 여론조사 사례

A Study on Mixed-Mode Survey which Combine the Landline and Mobile Telephone Interviews: The Case of Special Election for the Mayor of Seoul

이경택* · 이화정** · 현경보***

Kyoungtaeg Lee · Hwajeong Lee · Kyungbo Hyun

국내 선거여론조사에서 면접대상인 가구(또는 개인) 표본을 추출하기 위해 유선전화 전화번호부 또는 임의번호걸기(RDD) 방식이 주로 이용되고 있다. 하지만 유선전화를 보유하지 않은 가구가 늘어나고 있고, 유선전화가 있더라도 전화번호부 등재를 꺼리는 가구가 점차 많아지고 있다. 또한 젊은 층이나 직장인의 경우 여론조사가 실시되는 낮 시간대에 주로 외부에서 활동하므로 유선전화를 통한 접촉이 매우 어려운 실정이다.

상술한 문제들로 인하여 선거 여론조사의 예측력이 점점 떨어지고 있으며, 특히 조사시간대에 주로 외부에서 활동하는 사람들에 대한 낮은 접근성은 보수 성향 후보에게 긍정적인 예측결과를 내놓는 편향으로 이어지고 있다. 이러한 문제점을 해소할 수 있는 한 가지 방법으로 이동전화를 함께 활용하는 조사를 생각해 볼 수 있다. 즉, 낮 시간대 채택성향이 높은 사람들에 대해서는 유선전화를 활용한 조사를 수행하고, 부채성향이 높은 사람들에 대해서는 이동전화조사를 수행한 후, 두 결과를 혼합하는 방식(유·무선전화 병행조사)이다.

유·무선전화 병행조사를 실시하기 위해서는 1)유선전화와 이동전화 조사를 위한 표집틀이 확보되어야 하고, 2)유선전화와 이동전화로 조사할 비중을 사전에 결정해야 한다. 본 연구에서는 유·무선전화 병행조사를 실시하기 위한 경험적(heuristic) 방법론을 제안한다. 제안된 방법에서는 유선전화조사를 위해 임의번호걸기 방식을 이용하고, 이동전화조사를 위해 조사회사에서 모집한 조사패널을 활용한다. 또한, 유선전화와 이동전화로 조사할 표본의 비중은 통계청 생활시간조사 결과를 이용해서 계산한 채택율과 부채율을 활용한다.

* 교신저자(corresponding author): 엠브레인 리서치연구소 상무이사 이경택.

E-mail: ktlee@embrain.com

** 엠브레인 리서치연구소 부장

*** SBS 보도국 여론조사 전문기자, 언론학 박사

제안된 조사방법을 활용하여 10.26 서울시장 보궐선거에 대한 여론조사를 실시하였다. 총 4회의 여론조사가 실시되었는데, 처음 3회의 조사는 관세분석용 조사이고, 최종 조사는 선거결과 예측용 조사이다. 관세분석조사의 경우 조사시점에 발생한 이슈에 대한 반응이 타당성 있게 조사되었고, 선거예측조사의 경우 실제 선거결과에 매우 근접한 예측능력을 보였다.

주제어: 전화번호부, 임의번호걸기(RDD), 유선전화조사, 무선전화조사, 병행조사, 생활시간조사

Korean telephone surveys have been based on landline telephone directory or RDD(Random Digit Dialing) method. These days, however, there has been an increase of the households with no landline, or households with the line but not willing to register in the directory. Moreover, it is hard to contact young people or office workers who are usually staying out of home in the daytime.

Due to these issues above, the predictability of election polls gets weaker. Especially, low accessibility to those who stay out of home when the poll's done, results in predictions with positive inclination toward conservatism. A solution to resolve this problem is to contact respondents by using both mobile and landline phones—via landline phone to those who are at home and via mobile phone to those who are out of home in the daytime(Mixed Mode Survey, hereafter MMS).

To conduct MMS, 1) we need to obtain the sampling frames for the landline and mobile surveys, and 2) we need to decide the proportion of sample size of both. In this paper, we propose a heuristic method for conducting MMS. The method uses RDD for the landline phone survey, and the access panel list for the mobile phone survey. The proportion of sample sizes between landline and mobile phones are determined based on the 'Lifestyle and Time Use Study' conducted by Statistics Korea.

As a case study, 4 election polls were conducted in the periods of the special election for the mayor of Seoul on Oct 26th, 2011. From the initial 3 polls, reactions and responses regarding the issues raised during the survey period were appropriately covered, and the final poll showed a very close prediction to the real election result.

Key words: telephone directory, RDD(Random Digit Dialing), landline telephone survey, mobile telephone survey, MMS(Mixed Mode Survey), lifestyle and time use study

I. 서론

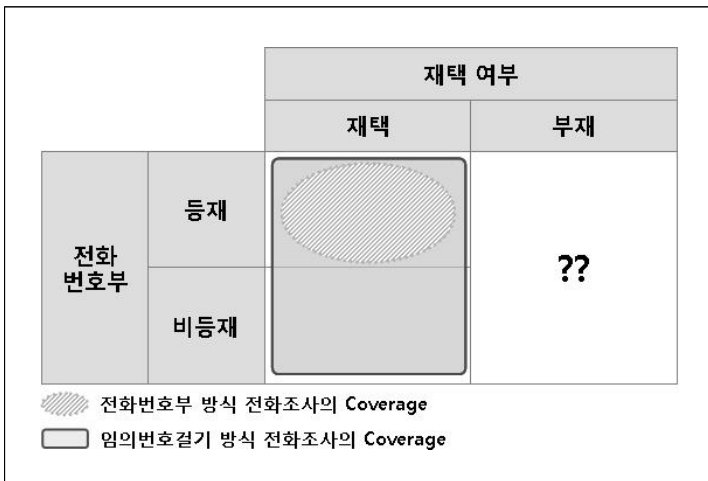
2010년에 실시된 6.2 지방선거 이전에는 대다수의 국내 선거여론조사가 전화번호부를 표집틀로 하여 진행되었다. 6.2 지방선거기간 동안 실시된 서울시장 선거 관련 여론조사에서 오세훈 후보가 한명숙 후보보다 크게 우세하다는 결과가 많았으나, 최종 선거 결과는 오세훈 후보가 47.43%, 한명숙 후보가 46.83%의 득표율을 보여 그 차이가 1%P도 안 되는 박빙이었다. 그 밖에 경기, 인천, 강원, 충북 등 대다수 시·도지사 선거 결과도 전화조사 예측과는 커다란 차이를 보였다. 이로 인해 전화번호부를 이용한 전화 여론조사가 여론을 정확히 반영하지 못한다는 비난의 목소리가 많았다.

전화번호부를 이용하는 전화 여론조사가 실제 여론을 제대로 반영하지 못하게 된 원인으로 1) 최근 사생활 보호 차원에서 유선전화를 보유하고도 전화번호부에 등재하지 않는 가구가 증가하였다는 점, 2) 젊은 층이나 1인 가구를 중심으로 유선전화를 보유하지 않은 가구가 많아졌다는 점, 3) 070 인터넷전화 보유가구가 늘어났다는 점 등이 주로 언급되고 있다. 또한 개인정보보호법의 영향으로 조사회사들이 2009년 이후의 전화번호부 데이터베이스를 제공받지 못하여, 2009년 이후에 개설된 유선전화 보유 가구에 대해서는 접근 자체가 원천적으로 불가능하다는 점도 주요 원인 중의 하나이다.

최근 들어 전화번호부를 이용한 전화조사의 문제점을 개선하고자 하는 연구가 다양하게 진행되고 있다. 강현철 외(2008)는 전화번호부의 등재율이 낮아 포함오류(coverage error)가 심각하다는 점에 착안하여, 임의번호걸기(RDD; Random Digit Dialing) 방식을 이용할 경우 표집편향을 일부 제어할 수 있음을 보였다. 이들의 연구에 의하면 전화번호부 등재가구의 비중이 60% 이하였고, 등재가구의 소득 수준이 다소 낮았으며, 이념성향 측면에서는 보수적인 성향을 띠었다. 허명희·김영원(2008)은 2007년 대통령 선거 예측조사 사례를 통해 전화번호부 기반 표본과 임의번호걸기 기반 표본이 어떻게 다른가를 연구하였다. 우정엽 외(2011)는 임의번호걸기 방식과 더불어 확률표본추출(probability sampling)의 기본적인 요소들을 함께 적용하였다. 이 연구에서도 등재가구와 비등재가구 간에 체계적인 차이가 존재했고, 특히 대통령 국정운영과 정당지지도 등에서 두 집단 간 뚜렷한 차이를 보였다.

임의번호걸기 방식을 활용하면 포함오류를 일부 제어할 수는 있지만, <그림 1>에 나타난 바와 같이 유선전화로만 조사가 진행될 경우 전화조사가 주로 진행되는 낮 시간대에 집에 있는 사람들에 대해서만 응답을 받을 수 있는 문제가 있다. 즉, 직장이나 학교에서 대부분의 시간을 보내는 사람들은 조사대상에 포함될 확률이 낮아지는 반면, 전업주부나 노년층처럼 집에서 보내는 시간이 많은 계층은 조사대상에 포함될 확률이 높아지게 된다. 이 문제와 관련하여 조성겸 외(2007)는 재택률의 차이로 인해 유선전화로만 여론 조사를 수행할 때에는 조사시간대를 잘못 설정하면 표본과 조사결과에 편향을 가져올 수 있다는 점을 지적하였다.

재택성향의 불균형을 개선하기 위해 허명희·황진모(2006)는 시간균형할당표집(time-balanced quota sampling) 방법을 제안했는데, 통계청의 생활시간조사 자료를 토대로 할당변수에 조사 시간대를 추가함으로써 편향이 일부 제거됨을 보였다. 허명희 외(2011)는 시간균형할당표집과 할당표집에 의한 임의번호걸기 방식의 전화조사 결과를 비교했는데, 재택성향과 상관관계가 높을 것으로 예상되는 가구소득과 학력 측면에서 유의한 차이가 있음을 보였다. 하지만 허명희 외(2011)는 시간균형할당표집 방법을 활용하더라도 가구원수 3인 이하의 비중이 모집단 분포에 비해 과소 표집된다는 점을 밝혔는데, 이는 유선전화만을 이용할 경우에는 일정 정도의 편향성을 피할 수 없음을 의미한다.



<그림 1> 유선전화조사의 포함오류

본 연구에서는 유선전화조사의 문제점을 해결할 수 있는 유·무선전화 병행조사에 대해 다루고자 한다. 본 연구의 구성은 다음과 같다. 먼저 제 2절에서는 유·무선전화 병행조사의 일반적인 사항에 대해 살펴본 후, 현실적인 조사환경에서 사용 가능한 경험적(heuristic) 방법을 제안한다. 제3절에서는 2011년 10월에 실시된 서울시장 보궐선거에 맞춰 실시된 유·무선전화 병행조사와 임의번호걸기 방식의 유선전화조사결과를 비교한다. 또한 제안된 조사방법을 활용한 선거결과 예측치와 실제 득표율을 비교한다. 끝으로 제안된 방법의 한계와 추후 과제를 도출한다.

II. 유·무선전화 병행조사

유·무선전화 병행조사는 유선전화와 이동전화를 함께 활용하는 조사방법이다. 유·무선전화 병행조사 수행을 위한 가장 이상적인 상황은 유선전화만 보유한 가구(A), 유선전화와 이동전화를 모두 보유한 가구(B) 및 이동전화만 보유한 가구(C)의 비중을 모두 알고 있고, A, B, C에 해당되는 모든 가구의 전화번호를 100% 포함하는 전화번호부가 존재하는 경우이다. 이럴 경우, A, B, C의 비중을 따라 각 전화번호부에서 조사대상을 무작위로 추출하여 조사를 진행하면 된다.

국내의 경우 A, B, C의 비중이 알려져 있지 않고, 유선전화의 경우 전화번호부에 포함된 가구의 비중이 60%를 넘지 않으며, 이동전화의 경우에는 전화번호부 자체가 존재하지 않는 문제가 있다. 또한 Link et al.(2007)의 연구에 의하면, B집단의 경우 유선전화로 접촉하느냐 이동전화로 접촉하느냐에 따라 조사결과에 유의한 차이가 나타나, B 집단에 대한 접근방식의 결정도 쉬운 문제는 아니다.

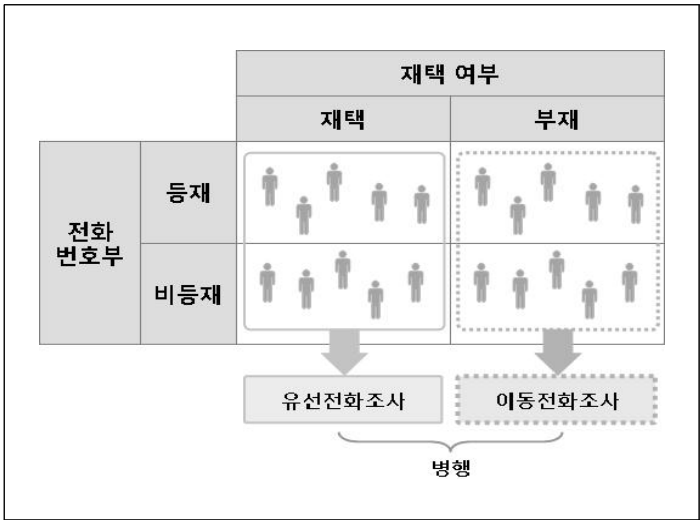
상술한 여러 문제들로 인하여 아직까지는 이론적으로 완벽한 유·무선전화 병행조사를 실시하는 데에 많은 어려움이 있다. 따라서 본 연구에서는 현실적인 조사환경에서 사용 가능한 경험적(heuristic) 방법론을 제안하고자 한다.

1. 제안 조사방법(유선RDD + 무선List)

유·무선전화 병행조사를 실시하기 위해서는 1) 유선전화와 이동전화조사를 위한 표집틀이 확보되어야 하고, 2) 유선전화와 이동전화로 조사할 표본의 비중을 사전에 결정해야 한다. 유선전화 표집틀에 대해 살펴보면, 많은 사전연구에서 전화번호부 방식에 문

제가 있는 것으로 밝혀졌기에 임의번호걸기 방식을 이용하는 것이 타당하다. 이동전화의 경우에는 전화번호부 자체가 존재하지 않기에 임의번호걸기 방식을 고려해 볼 수 있다. 하지만 대다수의 전화조사가 지역을 기반으로 진행되는데, 이동전화번호의 경우에는 국번과 지역 간의 관련성이 없어서 활용하기 쉽지 않다는 문제가 있다. 예를 들어, 시·군·구 단위의 국회의원 총선 관련 여론조사를 위해 임의번호걸기 방식의 이동전화조사를 실시하기 위해서는 엄청난 시간과 비용이 소요될 것이다. 대신 주소 정보가 포함된 이동전화번호 리스트(list)가 존재한다면, 이 리스트를 활용하여 이동전화조사를 수행할 수 있을 것이다. 이 연구에서는 유선전화조사를 위해 임의번호걸기 방식을 활용하고, 이동전화조사를 위해 리스트를 이용하는 방법을 제안한다. 지금부터 제안하는 조사방식을 <유선RDD + 무선List>로 지칭하기로 한다.

이동전화번호 리스트가 갖춰야 할 요건은 시·군·구 단위의 조사가 가능할 정도로 충분한 규모여야 하고, 리스트에 포함된 가망 응답자들에게 접촉했을 때 응답률이 높아야 한다. 또한, <그림 2>에 나타난 바와 같이, 이동전화를 도입하는 이유가 재택성향 위주의 표집결과를 방지하기 위한 것이므로 전화조사가 실시되는 평일 낮 시간대에 주로 외부에서 생활하는 사람들이 많아야 효율적이다. 이를 위해, 본 연구의 사례연구에서는 국내 최대 규모의 조사패널을 확보하고 있는 엠브레인의 온라인조사 패널을 활용하였다.

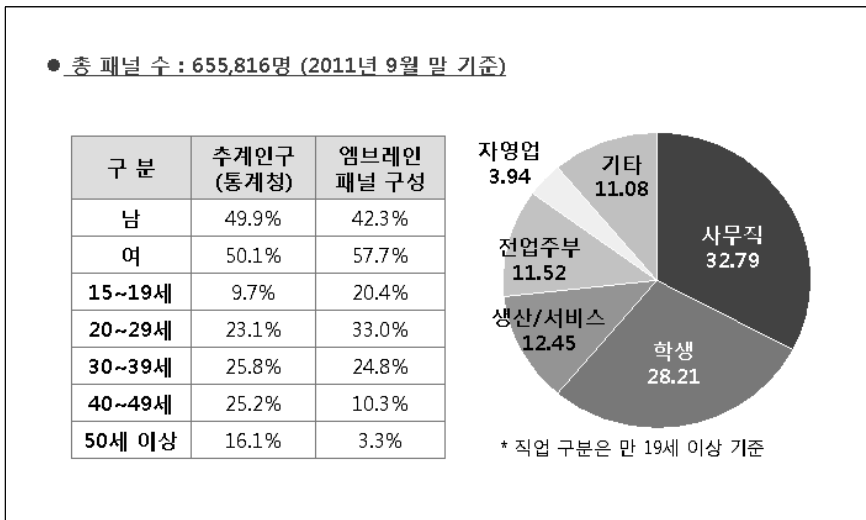


<그림 2> 유·무선전화 병행조사

엠브레인은 홈페이지를 통한 모집, 임의번호걸기 방식의 전화모집, 길거리에서의 가두 모집, 인터넷 제휴 사이트를 통한 모집 등 다양한 방식으로 응답자 집단을 모집하고 있다. 패널(access panel)이란 사전에 참여의사를 밝히고, 조사회사와의 계약에 의하여 개인에 대한 기초정보를 제공한 조사응답자 집단을 말한다. 엠브레인의 조사패널로 가입하려면 본인의 주소, 이메일 주소, 이동전화번호 등을 사전에 회사에 제공해야 하고, 가입한 회원들에게는 온라인조사를 포함한 다양한 조사에 참여할 수 있는 기회가 제공되며, 조사 참여율은 비교적 높은 편이다. 2011년 9월말 기준으로 655,816명이 등록되어 있어 시·군·구 단위의 조사가 가능할 정도로 충분한 규모이다. 특히 <그림 3>에 나타난 바와 같이, 낮 시간대에 주로 외부에서 활동할 가능성이 높은 젊은 직장인이나 학생의 비중이 높아 유·무선전화 병행조사를 위한 표집틀로 적당하다고 생각된다.

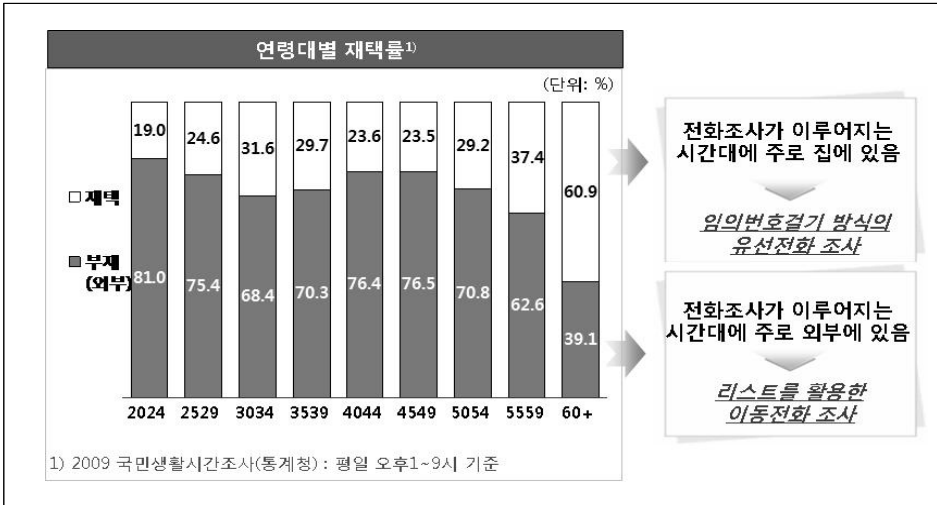
다음으로, 유선전화와 이동전화로 조사할 표본수의 결정에는 채택과 부채성향을 반영하는 것이 바람직하며, 이를 위한 기초자료는 공인된 자료일수록 좋을 것이다. 본 연구에서는 국가 승인통계인 2009년 통계청 생활시간조사 결과를 활용하였다.

통계청의 2009년 생활시간조사는 전국에서 표본추출된 8,100가구의 10세 이상 가구원 21,000명의 2일간 시간일지 기록을 담고 있다. 시간일지(time diary)는 10분 단위로 응답자가 연속 48시간 동안 한 일을 9개 대분류, 50개 중분류, 144개 소분류로 코딩한 것이

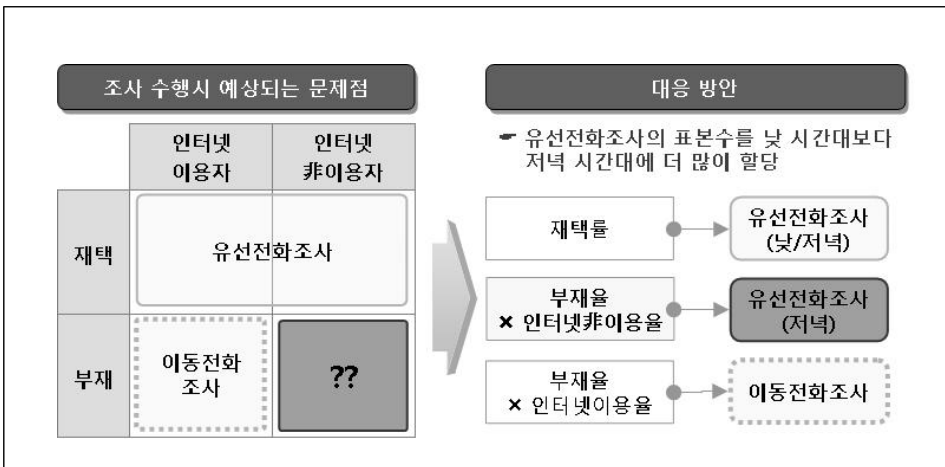


<그림 3> 엠브레인 온라인 조사패널 현황

다. 본 연구에서는 생활시간조사 결과를 바탕으로 하여 전화조사가 주로 실시되는 평일 오후 1시부터 9시까지의 재택율과 부재율을 계산한 다음, 재택율만큼 유선전화조사를 실시하고 부재율만큼 이동전화조사를 실시하는 방법을 활용하였다. <그림 4>는 전국의 만 19세 이상 성인남녀를 대상으로 계산한 재택율과 부재율 예시이다.



<그림 4> 생활시간조사를 활용한 재택률 및 부재율 계산 예시



<그림 5> 부재성향의 인터넷비이용자에 대한 대응방안

그런데 이동전화조사를 위한 리스트로 온라인조사 패널을 활용할 경우에는 혼합비중의 결정을 위해 약간의 조정이 필요하다. 즉 <그림 5>에 나타난 바와 같이, 온라인조사를 목적으로 모집된 패널 회원은 모두 인터넷 이용자이므로, 낮 시간대에 주로 외부에서 생활을 하는 인터넷 비이용자를 조사할 수 없다는 문제가 발생된다. 물론 국내의 인터넷 이용률이 매우 높고 외부에서 주로 활동하는 사람들의 대다수가 인터넷을 이용할 것으로 예상되므로, 이 문제가 중요하지 않을 수도 있다. 하지만 본 연구에서는 시간균형할당표집 방법의 아이디어를 응용하여, 부재율과 인터넷 비이용률을 곱한 비중만큼을 저녁시간대에 조사함으로써 최종적인 혼합비중을 결정하였다.

전술한 표본비중 결정방법의 이해를 돕기 위해 만 19세 이상 서울지역 거주자 1,000명을 대상으로 하는 여론조사를 가정하여 표본설계 방법을 설명한다.

먼저 행정안전부에서 발표하는 주민등록 인구통계 중 서울지역 자료를 활용하면, <표 1>과 같이 성별, 연령별로 조사해야 할 표본수를 구할 수 있다.

다음으로 통계청 생활시간조사 결과 중 서울지역 자료를 활용하여 재택율과 부재율을 계산하면 <표 2>의 결과를 얻게 된다. <표 2>를 보면 여성에 비해 남성의 부재율이 높고, 60세 이상의 재택율이 매우 높음을 알 수 있다.

<표 1> 인구비례 할당

(단위: 명)

연 령		19~29세	30~39세	40~49세	50~59세	60세~	합 계
성별	남성	102	113	104	89	82	490
	여성	101	109	104	95	101	510
합계		203	222	208	184	183	1,000

<표 2> 재택율 및 부재율

(단위: %)

연 령		19~29세	30~39세	40~49세	50~59세	60세~
재택율	남성	19.8	11.9	9.2	18.6	44.5
	여성	21.4	49.8	38.2	52.0	75.7
부재율	남성	80.2	88.1	90.8	81.4	55.5
	여성	78.6	50.2	61.8	48.0	24.3

〈표 3〉 인터넷이용률

(단위: %)

연 령		19~29세	30~39세	40~49세	50~59세	60세~
인터넷 사용률	남성	99.8	99.3	91.8	67.3	30.5
	여성	99.9	99.2	82.7	43.0	15.2

〈표 4〉 조사방법별 표본비중

(단위: %)

연 령		19~29세	30~39세	40~49세	50~59세	60세~
유선전화 조사비중	남성	20.0	12.5	16.7	45.2	83.1
	여성	21.4	50.2	48.9	79.3	96.3
이동전화 조사비중	남성	80.0	87.5	83.3	54.8	16.9
	여성	78.6	49.8	51.1	20.7	3.7

〈표 5〉 표본설계 예시

(단위: 명)

연 령		19~29세	30~39세	40~49세	50~59세	60세~	합 계
유선전화 조사	남성	20	14	17	40	68	160
	여성	22	55	51	76	97	300
	소계	42	69	68	116	165	460
이동전화 조사	남성	81	98	87	49	14	329
	여성	80	54	53	20	4	211
	소계	161	153	140	69	18	540

방송통신위원회에서 발표하는 인터넷이용률(〈표 3〉)과 〈표 2〉를 이용하여, 〈그림 5〉에서 설명한 대로 유선전화와 이동전화로 조사할 표본의 비중을 계산한 것이 〈표 4〉이다.

끝으로 〈표 1〉에서 구한 성별, 연령별 할당 표본 수에 〈표 4〉의 조사방법별 표본비중을 곱하면, 〈표 5〉와 같은 최종적인 표본설계표를 구할 수 있다.

2. 유선RDD + 무선RDD

조사지역이 전국 단위일 경우에는 유·무선전화 병행조사의 또 다른 대안으로 유선전화와 이동전화조사 모두를 임의번호걸기 방식으로 진행할 수 있다. 실제로 이 방법은 2011년 10월에 실시된 서울시장 보궐선거 기간 중 몇몇 조사회사에 의해 시도되었는데,

이는 우리나라 전체 인구 대비 서울시민의 비중이 1/4에 해당되기에 가능한 시도였다. 참고로 당시 진행된 유선전화와 이동전화간 표본의 혼합비중은 각각 50%였다. 지금부터는 이 조사방식을 <유선RDD + 무선RDD>로 지칭하기로 한다.

III. 사례 연구

이 절에서는 2011년 10월에 있었던 서울시장 보궐선거 기간 동안 실시된 유·무선전화 병행조사 결과를 활용하여 제안된 조사방법의 타당성을 살펴본다. 먼저 조사설계를 간략히 살펴본 후, 제안된 조사방법에 의한 유선전화조사결과와 이동전화조사결과를 비교한다. 다음으로, 선거기간 동안 언론에 발표된 다수의 여론조사 결과들과 제안된 조사방법을 비교하고, 끝으로 판세분석 결과와 선거예측 결과를 통해 제안된 조사방법의 타당성에 대해 살펴본다.

1. 조사설계

본 연구를 위해 서울시장 보궐선거 기간 동안 총 4회의 <유선RDD + 무선List>조사를 실시하였다. 1차는 9월 19~20일, 2차는 10월 10~11일, 3차는 10월 17~18일, 4차는 10월 24~25일에 걸쳐 진행되었다. 각 차수별 조사대상, 표본 수 및 표본구성은 모두 동일하다.

조사대상은 서울시에 거주하는 만19세 이상 유권자이며, 표본수는 각 차수별 1천 명이다. 표본구성은 행정안전부 주민등록 인구통계를 기준으로 권역별(서북권/강북권/서남권/강남권), 성별(남/녀), 연령대별(19~29 / 30~39 / 40~49 / 50~59 / 60세 이상)로 인구비례 할당방법을 이용하였고, 유선전화와 이동전화를 통해 조사할 표본의 수는 통계청 생활시간조사와 방송통신위원회 인터넷이용률 결과를 이용하여 분배하였다(<표 6> 참고).

<표 6>을 살펴보면 나이가 적을수록 이동전화조사의 비중이 높고 나이가 많을수록 유선전화조사의 비중이 높는데, 이는 표본설계 과정에 채택과 부재 성향을 반영했기 때문이다. 젊은 층일수록 가구 내 유선전화 보유율 및 낮 시간대 채택율이 낮아 유선전화를 통한 조사가 어려운 경향이 있는데, 유·무선전화 병행조사를 실시하게 되면 이런 어려움을 일정 부분 해소할 수 있다.

〈표 6〉 〈유선RDD + 무선List〉 조사 표본설계

(단위: 명)

			19~29	30~39	40~49	50~59	60~	합 계	
전체 표본	서북권	남	16	18	18	15	17	170	1,000
		여	16	18	17	15	20		
	강북권	남	33	35	35	29	24	318	
		여	31	34	34	30	33		
	서남권	남	31	37	31	27	24	304	
		여	31	34	30	30	29		
	강남권	남	21	22	21	18	17	208	
		여	23	23	23	21	19		
	전체	남	101	112	105	89	82		
		여	101	109	104	96	101		
유선전화 표본	서북권	남	3	2	3	7	14	80	461
		여	3	9	8	12	19		
	강북권	남	7	4	6	13	20	147	
		여	7	17	17	24	32		
	서남권	남	6	5	5	12	20	138	
		여	7	17	15	23	28		
	강남권	남	4	3	4	8	14	96	
		여	5	12	11	17	18		
	전체	남	20	14	18	40	68		
		여	22	55	51	76	97		
이동전화 표본	서북권	남	13	16	15	8	3	90	539
		여	13	9	9	3	1		
	강북권	남	26	31	29	16	4	171	
		여	24	17	17	6	1		
	서남권	남	25	32	26	15	4	166	
		여	24	17	15	7	1		
	강남권	남	17	19	17	10	3	112	
		여	18	11	12	4	1		
	전체	남	81	98	87	49	14		
		여	79	54	53	20	4		

* 서북권(6) : 은평구, 서대문구, 마포구, 용산구, 종로구, 중구
 강북권(8) : 광진구, 성동구, 동대문구, 도봉구, 강북구, 성북구, 노원구, 중랑구
 서남권(7) : 구로구, 금천구, 관악구, 동작구, 영등포구, 강서구, 양천구
 강남권(4) : 강남구, 서초구, 송파구, 강동구

2. 유선전화조사 결과와 이동전화조사 결과 비교

제안된 조사방법을 통해 조사된 유선전화조사결과와 이동전화조사결과 간에 어떤 차이가 존재하는지를 살펴보기 위해, 선거 1주일 전에 시행한 3차 조사결과를 분석하였다.

1) 응답자 특성 비교

〈표 7〉을 살펴보면, 유선전화 응답자와 이동전화 응답자의 직업과 가구소득에 유의한 차이가 존재함을 알 수 있다. 즉 유선전화 응답자의 경우 자영업자나 전업주부가 많고 월평균 가구소득이 낮은 특징을 보이는 반면, 이동전화 응답자의 경우는 사무직종사자 및 학생이 많고 가구소득이 다소 높은 특징을 보인다. 이러한 결과는 표본설계 과정에서 채택율과 부재율을 반영하였기에 당연한 결과라 할 수 있다. 따라서 여기에서는 공정한 비교를 위해 연령대를 기준변수로 하여, 동일 연령대에서 유선전화와 이동전화 응답자 간에 어떤 차이가 있는지를 살펴보기로 한다.

〈표 7〉 유선전화 및 이동전화 응답자의 특성

(단위: 명)

변 수	범 주	유선전화 (n=461,%)	이동전화 (n=539,%)	χ^2 검증
직업	자영업자	19.3	9.1	$\chi^2 = 249.950$ p-value = 0.000
	육체노동자	4.6	8.9	
	사무직종사자	14.3	47.5	
	전업주부	41.4	13.9	
	학생	4.6	15.6	
	기타/무직	15.8	5.0	
월평균 가구소득	200만원 미만	19.1	3.2	$\chi^2 = 128.884$ p-value = 0.000
	200만원대	14.3	12.4	
	300만원대	17.6	21.0	
	400만원대	12.6	19.5	
	500만원대	10.2	18.9	
	600만원 이상	9.8	19.3	
	무응답	16.5	5.8	

연령대별로 유선전화와 이동전화조사 응답자들의 직업을 비교해 보면, 20대의 경우에는 큰 차이를 보이지 않았으나($\chi^2 = 6.418$, $p\text{-value} = 0.601$), 30대 이상에서는 유선전화 집단에서 전업주부의 비중이 높았고, 이동전화 집단에서 사무직종사자의 비중이 높았다(<표 8> 참고). 20대에서 차이가 나지 않는 이유는 유선전화로 할당된 표본의 수가

<표 8> 연령대별 유선전화 및 이동전화 응답자의 직업 비교

(단위: 명)

		유선전화(%)	이동전화(%)	χ^2 검증
20대	표본수	(n=42)	(n=160)	$\chi^2 = 6.418$ $p\text{-value} = 0.601$
	자영업자	7.1	2.5	
	육체노동자	2.4	8.1	
	사무직종사자	31.0	30.6	
	전업 주부	4.8	1.9	
	학생	50.0	51.3	
	기타/무직	4.8	5.6	
30대	표본수	(n=69)	(n=152)	$\chi^2 = 22.802$ $p\text{-value} = 0.019$
	자영업자	8.7	6.6	
	육체노동자	7.2	9.9	
	사무직종사자	44.9	64.5	
	전업 주부	30.4	15.1	
	학생	0.0	1.3	
	기타/무직	8.7	2.6	
40대	표본수	(n=69)	(n=140)	$\chi^2 = 33.728$ $p\text{-value} = 0.000$
	자영업자	31.9	12.1	
	육체노동자	10.1	7.9	
	사무직종사자	15.9	55.0	
	전업 주부	40.6	22.9	
	학생	0.0	0.0	
	기타/무직	1.4	2.1	
50대 이상	표본수	(n=281)	(n=87)	$\chi^2 = 99.158$ $p\text{-value} = 0.000$
	자영업자	20.6	20.7	
	육체노동자	2.8	10.3	
	사무직종사자	3.9	36.8	
	전업 주부	49.8	19.5	
	학생	0.0	0.0	
	기타/무직	22.8	12.6	

단 42명에 그쳐, 조사 당일 우연히 학교나 직장에 가지 않은 정상적인 응답자가 표집되었기 때문으로 추측된다. 20대의 경우, 채택율 대비 많은 수의 표본이 유선전화조사로 할당될 경우에는 조사결과의 타당도가 떨어질 것으로 예상되며, 이에 대해서는 추후 연구가 필요할 것으로 생각된다.

〈표 9〉 연령대별 유선전화 및 이동전화 응답자의 월평균 가구소득 비교

		유선전화(%)	이동전화(%)	χ^2 검증
20대	표본수	(n=42)	(n=160)	$\chi^2 = 6.755$ p-value = 0.344
	200만원 미만	2.4	2.5	
	200만원대	23.8	16.9	
	300만원대	21.4	16.9	
	400만원대	19.0	20.6	
	500만원대	4.8	15.0	
	600만원대	9.5	16.9	
	무응답	19.0	11.3	
30대	표본수	(n=69)	(n=152)	$\chi^2 = 10.492$ p-value = 0.162
	200만원 미만	4.3	2.6	
	200만원대	17.4	12.5	
	300만원대	26.1	28.3	
	400만원대	20.3	17.1	
	500만원대	13.0	21.7	
	600만원대	11.6	16.4	
	무응답	7.2	1.3	
40대	표본수	(n=69)	(n=140)	$\chi^2 = 14.318$ p-value = 0.046
	200만원 미만	7.2	2.1	
	200만원대	10.1	8.6	
	300만원대	26.1	20.0	
	400만원대	17.4	23.6	
	500만원대	18.8	20.7	
	600만원대	8.7	20.7	
	무응답	11.6	4.3	
50대 이상	표본수	(n=281)	(n=87)	$\chi^2 = 45.813$ p-value = 0.000
	200만원 미만	28.1	6.9	
	200만원대	13.2	10.3	
	300만원대	12.8	17.2	
	400만원대	8.5	14.9	
	500만원대	8.2	18.4	
	600만원대	9.6	26.4	
	무응답	19.6	5.7	

〈표 9〉를 보면, 나이가 많을수록 유선전화 대비 이동전화조사 응답자의 가구소득이 높음을 알 수 있다. 이는 이동전화조사에 참여한 상당수 사무직종사자 집단이 맞벌이를 하고 있기 때문으로 해석된다¹⁾. 다수의 기존 연구에서 저소득층과 저학력층이 과다하게 표집됨으로써 결과가 편향될 수 있다는 것이 유선전화조사의 문제점 중 하나로 지적되어 왔는데(허명희(2007), 홍성준 외(2009), 우정엽 외(2011)), 상기 결과에 의하면 이동전화 조사를 함께 이용함으로써 이 문제를 일정 부분 해소할 수 있을 것으로 기대된다.

2) 후보지지율 비교

이동전화 응답자의 경우 진보 후보인 박원순 후보에 대한 지지율이 높았고, 유선전화 응답자의 경우 보수 후보인 나경원 후보에 대한 지지율이 높았다. 하지만 이 경우에도 재택성향이 반영된 표본설계의 효과를 배제하기 위해 연령대와 직업군을 기준변수로 하여 비교했다.

〈표 10〉을 살펴보면, 20대를 제외한 전 연령층에서 이동전화조사 응답자의 박원순 후보 지지율이 나경원 후보 지지율보다 높은 것을 알 수 있다. 이는 유선전화만을 이용한 여론조사를 실시할 경우, 보수 성향의 후보자에게 유리한 조사결과가 나올 수 있음을 의미한다. 한 가지 주목할 점은 유·무선 모두에서 나이가 들수록 나경원 후보에 대한 지지율이 높고, 젊을수록 박원순 후보의 지지율이 높은 동일한 경향을 보인다는 점이다.

재택성향이 높은 전업주부 층에서 나경원 후보에 대한 지지율이 상대적으로 높은 반면, 부재성향이 높은 사무직종사자와 학생층에서는 박원순 후보의 지지율이 높음을 알 수 있다. 그러나 동일한 직업이라 하더라도 이동전화 응답자의 경우 박원순 후보에 대한 지지율이 상대적으로 높은 데 비해, 유선전화 응답자의 경우에는 그 반대의 결과가 나왔다. 이는 유선전화나 이동전화 한 가지로만 조사를 수행한 후, 재택성향과 상관관계가 높을 것으로 예상되는 ‘직업’ 변수로 사후층화를 하더라도 대표성을 담보하기 힘들다는 점을 의미한다.

1) 본 논문에는 제시하지 않았지만, 이 조사에 참여한 여성만을 대상으로 살펴본 결과, 유선전화 조사에 참여한 여성 중 약 64%가 전업주부였음에 반해, 이동전화조사에 참여한 여성 중에는 약 36%만이 전업주부였다.

〈표 10〉 유선전화 및 이동전화 응답자의 후보자 지지율 비교

		유선전화(%)	이동전화(%)	χ^2 검증	
전 체	나경원	49.0	37.7	$\chi^2 = 46.299$ p-value = 0.000	
	박원순	35.6	56.8		
	기타/무응답	15.4	5.5		
연령	20대	나경원	21.4	34.4	$\chi^2 = 2.855$ p-value = 0.414
		박원순	61.9	56.3	
		기타/무응답	16.7	9.3	
	30대	나경원	36.2	28.9	$\chi^2 = 4.585$ p-value = 0.205
		박원순	52.2	67.1	
		기타/무응답	11.6	4.0	
	40대	나경원	36.2	40.0	$\chi^2 = 6.013$ p-value = 0.049
		박원순	44.9	57.1	
		기타/무응답	18.9	2.9	
	50대 이상	나경원	59.4	55.2	$\chi^2 = 17.518$ p-value = 0.000
		박원순	25.3	39.1	
		기타/무응답	15.3	5.7	
직업	자영업자	나경원	49.4	44.9	$\chi^2 = 2.193$ p-value = 0.334
		박원순	36.0	49.0	
		기타/무응답	14.6	6.1	
	육체노동자	나경원	42.9	39.6	$\chi^2 = 6.597$ p-value = 0.086
		박원순	38.1	52.1	
		기타/무응답	19.0	8.3	
	사무직종사자	나경원	39.4	31.6	$\chi^2 = 3.924$ p-value = 0.270
		박원순	50.0	64.1	
		기타/무응답	10.6	4.3	
	전업주부	나경원	53.9	48.0	$\chi^2 = 16.569$ p-value = 0.000
		박원순	31.4	48.0	
		기타/무응답	14.7	4.0	
	학생	나경원	23.8	39.3	$\chi^2 = 0.835$ p-value = 0.841
		박원순	52.4	53.6	
		기타/무응답	23.8	7.1	
기타/무직	나경원	53.4	44.4	$\chi^2 = 4.511$ p-value = 0.105	
	박원순	27.4	44.4		
	기타/무응답	19.2	11.2		

3. 유선전화조사 결과와의 비교

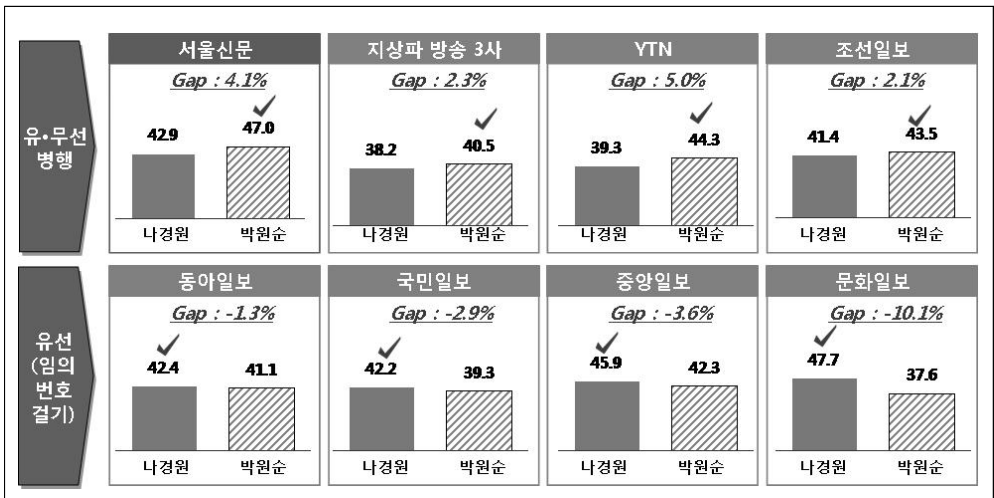
본 연구의 조사결과를 서울시장 보궐선거 기간 중 실시되어 언론을 통해 발표된 조사 결과들과 비교하였다. 이를 위해서 3차 조사와 비슷한 시기에 실시된 조사결과들을 이용하였고, 비교대상 조사는 <표 11>에 정리되어 있다.

<표 11> 비교대상 주요 조사¹⁾

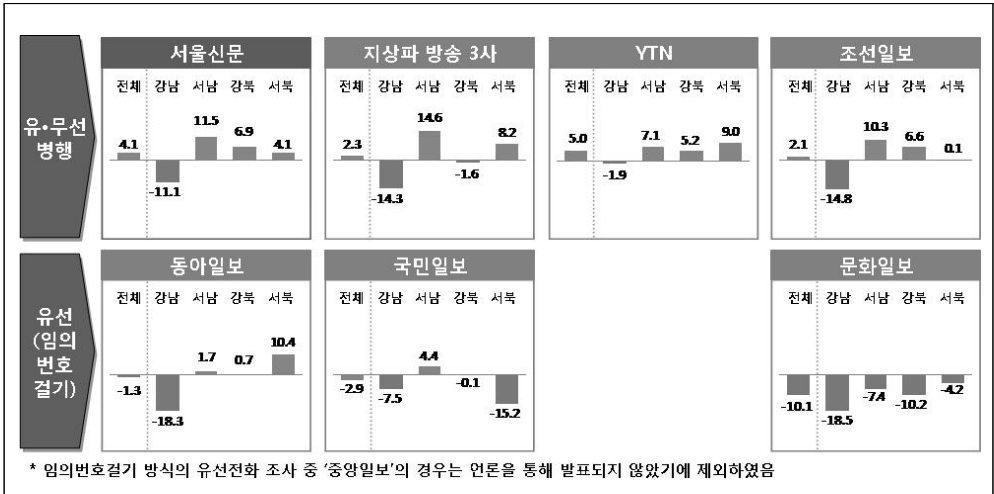
(단위: 명)

조사방식	언론사	조사 수행기관	조사시기	표본크기
유·무선 병행 ²⁾	서울신문	엠브레인	10월 17~18일	1,000
	지상파방송 3사	TNS, KRC, 미디어리서치	10월 16~17일	1,000
	YTN	한국리서치	10월 17~19일	3,950
	조선일보	미디어리서치	10월 19일	1,000
유선 (임의번호결기)	동아일보	KRC	10월 16~17일	1,000
	국민일보	GH코리아	10월 18일	800
	중앙일보	중앙일보 조사연구팀	10월 17~18일	1,695
	문화일보	디오피니언	10월 19일	1,000

- 1) 조사방식, 조사시기 및 표본크기는 언론사에서 공개한 대로 정리하였음
- 2) '서울신문'을 통해 발표된 내용은 본 연구에서 제안한 <유선RDD+무선List> 방식이고, '지상파방송 3사', 'YTN', '조선일보'는 <유선RDD + 무선RDD> 방식임



<그림 6> 후보지지율 비교



〈그림 7〉 권역별 후보간 지지율 Gap 비교

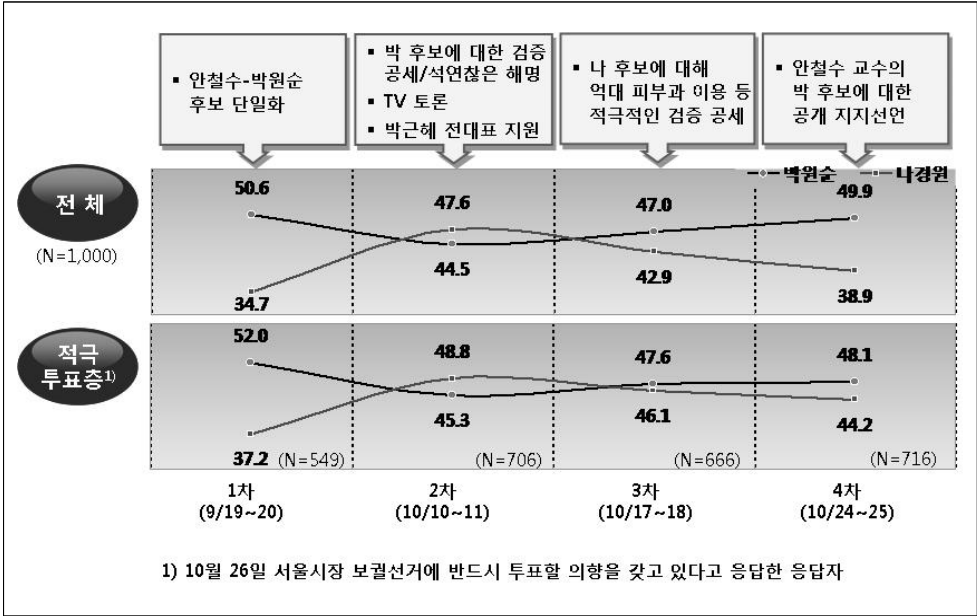
전체응답자를 기준으로 후보 지지율을 비교해 본 결과, 4개의 유·무선전화 병행조사에서는 모두 박원순 후보의 지지율이 높은 데 반해, 임의번호걸기 방식을 활용한 유선전화조사에서는 모두 나경원 후보의 지지율이 높게 나타났다(〈그림 6〉 참고). 이로부터, 전화조사에 이동전화를 포함시키느냐의 여부에 따라 조사결과에 체계적인 차이가 발생됨을 알 수 있다.

〈그림 7〉은 박원순 후보와 나경원 후보의 지지율 차이를 권역별로 계산한 결과이다. 유·무선전화 병행조사의 경우에는 강남권 이외의 모든 지역에서 박원순 후보의 지지율이 일관되게 높은 특징을 보였으며, 〈유선RDD + 무선List〉와 〈유선RDD + 무선RDD〉 간에 큰 차이를 보이지 않았다. 반면, 임의번호걸기 방식의 유선전화조사들 간에는 일정한 경향성이 존재하지 않았다.

4. 판세분석 조사 결과

선거 여론조사의 목적은 크게 1) 판세분석과 2) 투표결과예측의 2가지로 구분할 수 있다. 판세분석을 목적으로 하는 여론조사는 조사결과를 활용하여 선거 캠페인의 전략을 수정·보완하기 위해 실시되며, 주로 선거기간의 초, 중반에 행해지는 조사이다. 반면, 결과예측을 목적으로 하는 여론조사는 선거 당일의 투표결과를 예측하기 위해 진행되는

조사로, 선거 막바지에 실시되는 것이 보통이며 출구조사도 예측조사로 볼 수 있다. 본 연구에서 실시된 4차례의 여론조사 중 1~3차는 판세분석용 조사이고, 4차는 결과예측용 조사이다.



〈그림 8〉 차수별 박원순 후보와 나경원 후보의 지지율 변화

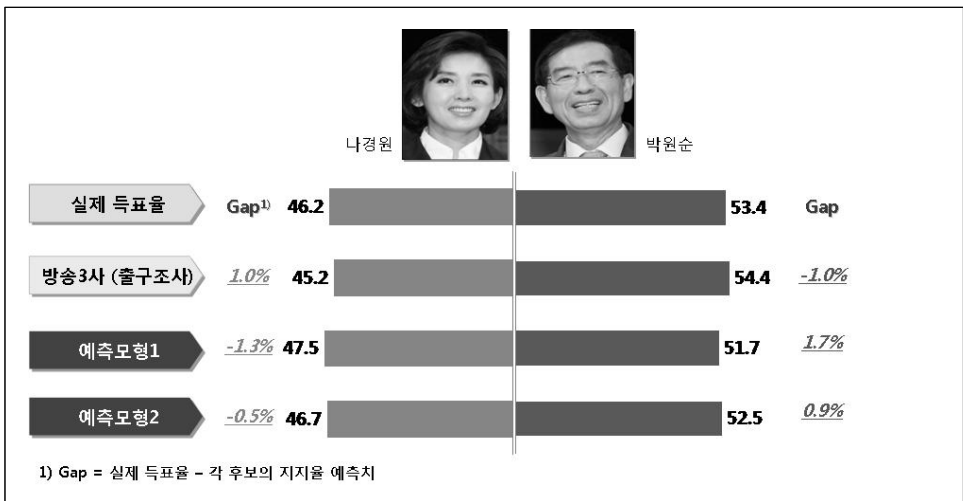
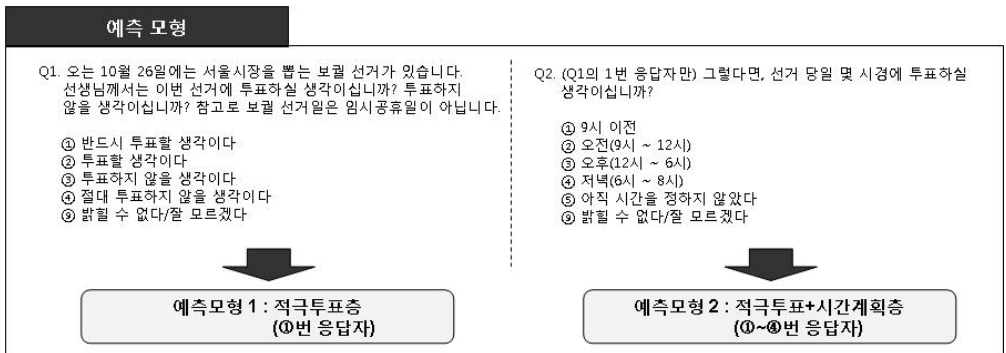
〈그림 8〉은 4차례 실시된 조사에서의 후보별 지지율을 나타낸 것으로, 해당 시점의 이슈와 이에 반응한 여론의 추이가 조사결과에 비교적 잘 반영되었음을 알 수 있다. 즉 안철수 후보와 박원순 후보의 단일화가 이루어진 9월의 1차 조사에서는 박원순 후보의 지지율이 나경원 후보의 지지율 대비 15%p 이상 높게 나타났다. 박원순 후보에 대한 검증 공세와 TV토론에서 밀리는 모습 등의 영향으로 인해 지지율이 한 때 역전되기도 했으나(2차 조사), 나경원 후보에 대한 검증 공세가 시작되면서 박원순 후보의 지지율이 회복되었다(3차 조사).

〈그림 8〉에 제시된 후보별 지지율에는 무응답층(1차 14.7%, 2차 6.2%, 3차 8.8%, 4차 10.0%)에 대한 별도의 처리를 하지 않았으며, '적극 투표층'은 선거 당일 반드시 투표에 참여하겠다고 응답한 사람들의 지지율을 계산한 값이다.

5. 예측조사 결과

선거결과는 전체 유권자 중에서 투표에 참여한 사람들의 지지율에 따라 결정된다. 따라서 전화 여론조사를 이용한 투표결과 예측에도 조사에 참여한 사람들 중 투표 참여 의향이 높은 사람들만의 결과를 이용하는 것이 합리적이고, 이들 의향자를 어떻게 선별하느냐가 핵심이다.

본 연구에서는 아래와 같이 ‘적극투표층’(예측모형1)과 ‘적극투표 + 시간계획층’(예측모형2)을 선별하여 투표결과를 예측하였다.



〈그림 9〉 서울시장 보궐선거 결과 예측

	실제 결과		방송3사 출구조사		예측모형1		예측모형2	
	박원순	나경원	박원순	나경원	박원순	나경원	박원순	나경원
강북동	55.1%	43.8%	56.1%	43.6%	54.7%	44.4%	55.1%	43.9%
강북서	54.8%	44.3%	57.8%	41.8%	55.4%	44.6%	55.6%	44.4%
서남권	56.3%	42.7%	58.2%	41.4%	53.3%	45.8%	53.9%	45.1%
강남권	44.5%	54.7%	45.8%	53.7%	41.6%	57.3%	43.6%	55.2%
Gap (실제 - 예측)			-1.0%	0.2%	-0.4%	0.6%	0.0%	0.1%
			-3.0%	2.5%	0.6%	0.3%	0.8%	0.1%
			-1.9%	1.3%	-3.0%	3.1%	-2.4%	2.4%
			-1.3%	1.0%	-2.9%	2.6%	-0.9%	0.5%

〈그림 10〉 서울시장 보궐선거 결과 예측(권역별)

〈그림 9〉는 4차 조사에서 ‘적극투표층’과 ‘적극투표 + 시간계획층’의 결과만을 활용하여 계산한 선거결과 예측치와 실제 선거결과 및 출구조사 결과를 비교한 것으로, 예측결과가 실제 득표율 및 출구조사 결과에 상당히 근접했음을 알 수 있다. 특히 ‘예측모형1’에 비해 ‘예측모형2’의 예측 정확도가 더 높아, 투표에 참여할 것으로 예상되는 유권자를 얼마나 정확히 가려낼 수 있는지가 선거결과 예측의 핵심사항인 것으로 생각된다.

권역별 예측 결과에서도 상당히 높은 예측력을 보였는데, 특히 예측모형2는 서남권을 제외한 모든 권역에서 1%P 미만의 예측 오차를 보였다(〈그림 10〉 참고).

IV. 맺음말

2010년에 실시된 6.2 지방선거 기간 동안 실시된 전화번호부 방식의 전화 여론조사 결과와 실제 선거결과 간에 큰 차이를 보임에 따라 조사업계는 많은 비난에 직면하게 되었고, 이를 개선하기 위해 많은 노력들을 기울여 왔다. 그 일환으로 과거 일부 조사에서만 활용되던 임의번호걸기 방식이 적극 도입되기 시작하였고, 불과 1년도 지나지 않은

2011년 4.23 재보궐 선거에서는 대다수의 전화 여론조사가 임의번호걸기 방식으로 진행 되기에 이르렀다.

임의번호걸기 방식은 전화번호부 비등재가구에 대해서도 접근할 수 있기 때문에 대표성 측면에서 분명 진일보한 방법이라 할 수 있다. 하지만 특히 젊은 층을 중심으로 일어나고 있는 유선전화 비보유 및 인터넷전화 보유가구의 증가와 낮 시간대 부재성향 등 전화조사 환경의 변화로 인해, 임의번호걸기 방식을 활용하더라도 유선전화만을 이용한 전화조사에서는 무시할 수 없는 과소포함(under-coverage) 문제가 존재할 수밖에 없다(허명희 외 2011).

이러한 점에 근거하여, 본 연구에서는 임의번호걸기 방식의 유선전화와 리스트 방식의 이동전화를 함께 이용하는 유·무선전화 병행조사를 제안하고, 제안된 조사방법을 2011년 10월에 있었던 서울시장 보궐선거에 적용해 보았다. 사례연구의 결과, 제안된 조사방법이 기존 유선전화조사의 문제점으로 제기되었던 저소득, 저학력, 보수층의 과대표집 문제를 일정 부분 해결할 수 있고 선거예측 결과도 실제 득표율과 큰 차이를 보이지 않아, 기존 전화조사의 문제를 해결할 수 있는 새로운 대안으로 활용될 수 있을 것으로 기대된다. 하지만 제안된 조사방법의 폭넓은 활용을 위해 몇 가지 선결되어야 할 문제 또한 존재한다.

첫째, 이동전화조사를 위한 표집틀의 대표성 문제이다. 본 연구에서는 표집틀로 올라 인조사를 위해 모집된 조사패널을 이용했는데, 비록 국내 최대 규모의 패널이라고 하더라도 자원자 집단이라는 측면에서 이동전화를 위한 표집틀로 적절할가에 대한 의문이 있을 수 있다. 이 문제에 대해서는 패널을 이용하는 이동전화조사와 임의번호걸기 방식의 이동전화조사를 동시에 수행하여 비교함으로써 실증적으로 검증해 볼 수 있을 것이다.

둘째, 유선전화조사와 이동전화조사 표본수 결정을 위해 통계청 생활시간 조사 결과를 활용했는데, 생활시간조사 결과가 시·도 단위로만 제공되므로 시·군·구 단위의 조사를 수행할 경우 현실과 동떨어진 표본설계를 해야 하는 경우가 발생할 수 있다. 예를 들어 경기도 '분당'과 '연천군' 거주자의 재택성향이 분명히 다름에도 불구하고, 생활시간 조사 결과를 활용할 경우에는 동일한 재택율과 부재율을 이용할 수밖에 없다.

마지막으로, 본 연구의 사례분석을 통해 유·무선전화 병행조사의 선거결과 예측 능력이 뛰어남을 보였다. 하지만 한 번의 결과로 일반화하기에는 어려움이 따르므로, 향후 다수의 실증적 연구를 바탕으로 일반화를 위한 논의가 필요할 것으로 생각된다.

참고문헌

- 강현철·한상태·김지연·정용찬·허명희. 2008. “RDD전화조사와 주요결과.” 《조사연구》 9(1): 1-22.
- 우정엽·김지윤·문종배. 2011. “표집틀 설정과 표본추출방법에 따른 정치성향 분석의 문제점.” 《조사연구》 12(1): 155-176.
- 조성겸·김지연·나운정·이명진. 2007. “선거여론조사의 문제점과 개선방향.” 《조사연구》 8(1): 31-54.
- 허명희. 2007. “여론조사 방법론: 과제와 전망.” 대한통계협회 《통계》 33(1): 27-36.
- 허명희·김영원. 2008. “RDD 표본 대 전화번호부 표본: 2007년 대통령 선거예측 사례.” 《조사연구》 9(2): 55-69.
- 허명희·한상태·김지연·성은아·강현철. 2011. “임의번호걸기와 시간균형할당표집에 의한 전화 조사의 주요결과.” 《조사연구》 12(2): 77-88.
- 허명희·황진모, 2006. “전화조사를 위한 시간균형할당 표본추출.” 《조사연구》 7(2): 39-52.
- 홍성준·박소형·김선웅. 2009. “국내 유선 전화조사에서 표본추출틀의 포함률.” 《조사연구》 10(1): 33-56.
- Link, Michael W. Link, Michael P. Battaglia, Martin R. Frankel, Larry Osborn, and Ali H. Mokdad. 2007. “Reaching the U.S. Cell Phone Generation: Comparison of Cell Phone Survey Results with an Ongoing Landline Telephone Survey.” *Public Opinion Quarterly* 71: 814-839.

<접수 2012/2/13, 수정 2012/3/6, 게재확정 2012/3/12>