

연구논문

2014년 지방선거 여론조사 전화조사방법에 따른 예측오차 및 편향*

Forecasting Error and Bias of Telephone Survey for 2014 Local Elections

김영원^{a)} · 황다솜^{b)}

Young-Won Kim · Dasom Hwang

선거여론조사가 얼마나 정확한지 그리고 어떤 조사가 다른 조사에 비해 왜 더 정확한지 등에 대해서 여론조사 전문가들은 많은 관심을 갖고 있다. 본 연구에서는 이러한 질문에 대한 해답을 찾기 위해 2014년 제6회 전국동시지방선거에 대한 전화여론조사자료를 토대로 전화조사방법에 따른 편향과 예측오차를 분석하였다. 이를 위해 중앙선거여론조사공정심의위원회 홈페이지에 등록된 선거여론조사자료를 활용하였으며, Martin et al. (2005)이 제안한 정확성 척도 A 를 예측오차 및 편향의 판단 기준으로 사용하였다. 2014년 지방선거 선거여론조사 중 광역단체장 선거에 관한 150건의 전화조사를 분석해 본 결과, ARS전화조사는 전화면접조사에 비해 새누리당을 과대 추정하는 편향을 발생시키는 경향이 있다는 것을 확인할 수 있었고, 조사방법과 함께 가중치 변동 폭, 조사기간, 그리고 응답률 등도 전화여론조사에서 편향이나 오차 발생에 영향을 주는 요인이라는 결론을 내릴 수 있었다.

주제어 : 선거여론조사, 전화조사, 예측오차, 편향, 2014년 지방선거

* 본 연구는 숙명여자대학교 2012학년도 교내연구비 지원에 의해 수행되었음.

a) 교신저자(corresponding author): 숙명여자대학교 통계학과 교수 김영원.

E-Mail: ywkim@sookmyung.ac.kr

b) 숙명여자대학교 통계학과 대학원생

If late-campaign polls are to be used as forecast, it is important to ask, how well do the polls do and why are some polls better forecast than others? To answer the questions, we investigate the forecasting error and bias of telephone survey conducted for 2014 local election of Korea. To resolve the problem, we use the telephone survey para-data released by National Fair Election Survey Deliberation Commission. The aim of this study is to find main cause of the forecasting error and bias of telephone survey. By analyzing the 150 election telephone survey and using the measure of predictive accuracy(A) proposed by Martin et al.(2005), we found telephone survey method, period of poll, variation level of weight, and response rate are the factors to cause the forecasting error and bias in telephone survey. Also we found that ARS telephone survey as well as large variation of weight bring overestimation of candidate of Saenuri Party.

Key words: election poll, telephone survey, forecasting error, bias, 2014 local election of Korea

I . 서론

여론조사는 각종 정치, 사회적 문제나 국가 정책 등에 대한 국민들의 의견을 파악할 수 있는 매우 유효한 방법이며, 특히 선거에서는 유권자들이 자신들을 대표해 줄 수 있는 사람을 선택하는 데 필요한 정보를 얻고 난립하는 후보자를 거를 수 있는 순기능도 갖고 있다. 하지만 같은 시기에 실시된 여론조사의 경우에도 후보자의 지지율이 크게 차이가 나서 여론조사에 대한 신뢰

성을 의심하게 하는 경우가 흔히 발생한다. 선거여론조사는 조사방법, 표본추출방법, 조사일시나 기간, 그리고 설문지 구성 등 실제 어떤 방식으로 어떤 조사회사가 조사를 수행하는지에 따라 조사결과가 크게 달라질 수 있기 때문이다.

이런 관점에서 국내에서도 조사방법에 따른 전화조사의 정확성을 비교하기 위한 연구들이 수행되어 왔다. 강현철 외(2008)는 RDD(Random Digit Dialing) 방식의 전화조사와 전화번호부 데이터베이스를 기반으로 전화번호를 추출하는 전화조사의 차이점을 심층적으로 분석한 결과를 제시하였고, 정형철(2007)은 ARS(Automatic Response System) 전화조사와 면접전화조사가 특성상 어떤 차이가 있는지 분석하기 위해 2007년 한나라당 대선후보에 대한 전화조사를 두 가지 방법으로 별도로 수행해 얻는 자료를 기초로 분석한 결과를 제시하고 있다.

전화조사방법과 관련해 조성겸(1997)은 전화여론조사의 취약점으로 자주 거론되는 지역·성·연령대를 기준으로 한 할당표본추출(quota sampling)이 갖고 있는 문제점을 지적하였으며, 이런 문제점을 해결하기 위해 허명희·황진모(2006)는 시간대별로 할당을 하는 동시에 부재 가구에 대한 재통화를 강화하여 기존 할당추출법보다 편향을 줄이는 시간균형 할당표본추출방법을 제안하고 효율성을 검토한 바 있다. 또한 허명희·김영원(2008)은 2007년 대통령 선거예측을 위해 KBS·MBC에서 공동으로 수행한 전화조사자료를 이용해 RDD 표본과 전화번호부 표본을 대비시켜 교육수준, 직업, 소득 등에 있어서 어떤 차이가 있는지 살펴보고, 특히 후보자 지지율 관점에서 전화번호부 표본과 RDD 표본에서 어떤 차이가 나는지 밝혀주고 있다.

전화조사의 체계적인 편향과 관련해서 김세용·허명희(2009)는 2007년 대통령선거 여론조사들에 대한 메타분석을 통해 지역, 성과 연령대를 고려한 할당추출을 하는 경우, 전화조사 표본이 학력과 직업 분포 등에 있어서 대표성이 떨어져 결국 이명박 후보에 대한 지지율이 과다 추정되었다는 사실을 밝혔다. 또한 2012년 대통령 선거 관련 전화여론조사에서 나타난 조사기관들의 편향과 관련해 박종희(2013)는 개별 조사자료의 중심성과 여론의 연속성

4 조사연구

을 반영하는 베이지안 상태공간모형을 기반으로 한 추세분석을 토대로 조사기관 편향에 대한 분석 결과를 제시하였고, 최필선·민인식(2013)도 공통확률추세라는 시계열 방법론과 Martin et al.(2005)이 제안한 정확성 척도를 사용하여 조사기관별 조사오차에 대한 비교 결과를 제시하였다. 대통령선거와 관련한 이들 전화조사의 편향에 대한 연구에서는 대선인 경우 후보자가 확정된 상태에서 여론조사가 상당히 오랜 기간 실시되고, 각 후보자의 지지율이 선거운동 기간 동안 여러 가지 요인으로 인하여 변동이 생길 수 있다고 볼 수 있기 때문에 최종 개표결과와 비교하는 것 대신 지지율 흐름을 시계열 관점에서 추세로 보고 오차를 분석하는 접근방식을 취하고 있다. 이에 반해 본 연구에서는 후보 등록이 마감된 후 선거를 앞 둔 시점에 집중적으로 수행된 전화조사만을 분석대상으로 삼고 있기 때문에 최종 개표결과와 직접 비교하는 방식을 통해 전화조사의 정확성을 평가하고 있다는 점을 밝혀둔다.

아울러 우리나라 국회의원선거에서 의석수 예측 문제와 관련해 출구조사를 포함한 방송사 선거예측조사를 중심으로 특정 정당에 대한 체계적인 편향에 대한 몇 편의 연구가 있었다. 이준웅(2004)은 17대 총선 선거예측조사 사후분석을 통해 표본추출이나 조사과정에 내재하는 무작위(random) 오차보다는 체계적으로 기울어지는 편향이 예측 실패의 주요한 원인이 되고 있다는 것과 함께, 우리나라 총선 예측조사에서 그동안 지속적으로 발생해 온 ‘여당과대평가’라는 체계적인 편향 문제를 다루었다. 김영원·곽은선(2010)은 2008년 총선 출구조사자료를 토대로 총조사오차라는 개념을 도입해 출구조사에서 발생하는 오차를 투표소 추출오차와 응답과정에서의 오차로 구분하고 응답과정에서 발생하는 응답 편향 문제를 다루었다. 또한 김영원·최윤정(2011)은 2010년과 2006년 지방선거와 2007년 대통령선거 출구조사자료를 토대로 하여 지역별로 발생하는 특정 정당에 대한 체계적인 편향 발생 양상을 밝혔다.

한편, 관련 학계와 조사업계에서는 전화조사방법이 달라지면 조사의 정확성에 구체적으로 어떤 영향을 주는지에 대해 많은 관심을 가져왔다. 하지만 그동안 우리나라에서 개별 조사기관에서 수행하는 전화여론조사가 구체적으로 어떤 방식으로 이루어지는지 파악할 수 있는 정보가 공개되지 않았기 때

문에 아주 제한적인 경우를 제외하고는 이런 궁금증에 대한 해답을 줄 수 있는 연구가 이루어지지 못 하였다. 그러나 2014년 제6회 전국동시지방선거를 앞두고 2014년 2월 「공직선거법」 개정을 통해 공표 또는 보도를 목적으로 하는 선거여론조사의 경우 개별 여론조사가 어떻게 수행되었는지 파악할 수 있는 정보공개가 의무화 되었다. 결과적으로 2014년 지방선거부터는 중앙선거여론조사공정심의위원회 홈페이지(www.nesdc.go.kr)에 등록된 선거여론조사 자료를 활용하면 조사방식 등에 따른 예측오차나 편향을 분석하는 것이 가능하게 된 것이다. 이와 같이 최근 여론조사 정보공개 제도가 도입됨에 따라 본 연구가 수행될 수 있었으며, 향후 본 연구와 같은 여론조사방법론에 대한 계량적인 연구가 다양한 관점에서 수행될 수 있다면, 우리나라 여론조사 수준을 한 단계 높이는 데 크게 기여할 수 있을 것으로 기대된다.

본 연구에서는 우선 2절에서 여론조사의 정확성을 평가하기 위해 본 연구에서 사용한 정확성 척도(accuracy measure)와 분석대상 전화여론조사자료에 대해 설명한다. 3절에서는 중앙선거여론조사공정심의위원회 홈페이지에 등록된 광역단체장 선거 관련 전화여론조사들을 대상으로 하여 조사방법이나 지역, 조사회사, 조사기간, 가중치 변동 폭 등이 달라짐에 따라 여론조사 결과에 어떤 편향이나 오차가 발생하는지 살펴보고, 4절에서는 선형모형을 활용하여 조사방법이나 조사기간 등 다양한 요인들이 전화여론조사의 오차 발생 패턴에 어떤 영향을 주는지 분석하기로 한다.

II. 예측 정확성 척도와 분석대상 여론조사

1. 선거조사 예측 정확성과 편향 평가 척도

전화조사라고 하더라도 실제로 매우 다양한 형식의 조사가 가능하기 때문에 조사방식에 따른 전화여론조사의 예측 정확성이나 편향 발생 여부 등에 대해 학계나 조사업계에서 관심이 많다. 우리나라에서 선거예측을 위해 사용되는 전화조사방법을 보면 우선 크게 ARS와 전화면접방식이 있고, 표본추출

을 위해서는 유선전화, 무선전화 또는 유·무선 병합조사 방식 등이 있다. 조사기간 등도 조사회사나 조사방법에 따라 차이가 나는 등 전화조사방법이 다양해짐에 따라 조사방법에 따른 예측 정확성 또는 편향 발생 가능성에 대한 분석이 필요하다.

선거조사 예측 정확성 평가를 위해서는 적절한 정확성 척도를 사용하여 여론조사결과들을 분석하는 것이 필수적이다. Mitofsky(1998)는 여론조사 정확성 평가 척도들의 장·단점을 잘 정리해 주고 있으며, 김영원·최윤정(2011)은 어떤 척도가 우리나라 선거예측조사 관련 연구에서 주로 사용되고 있는지 언급하고 있다. 합리적인 전화여론조사 정확성 평가를 위해서는 Martin et al.(2005)이 지적한 것처럼 ‘부동층(undecided vote)’ 처리방법에 따라 정확성 측정 결과가 달라지면 혼란이 생길 수 있다는 점을 고려해야 한다. 또한 우리나라 예측오차 관련 연구에서 흔히 사용하는 1위와 2위 후보자 예측값 차이와 지지율 차이를 비교하는 $|(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)|$ 와 같은 절대값을 취하는 방식의 척도는 그 의미를 해석하기 어려울 뿐만 아니라 오차의 방향성, 즉 편향 발생 여부를 판정하기 어렵다는 점도 고려해 적절한 정확성 척도를 선택할 필요가 있다.

이런 점들을 고려해 Martin et al.(2005)은 선거여론조사의 예측 정확성 측정을 위해 미국 공화당과 민주당 소속 정당 후보자의 득표율에 대한 오즈비(odds ratio)를 기초로 한 식(1)과 같은 예측 정확성 척도(predictive accuracy measure) A 를 제시하였다. 최근 여론조사 정확성 평가를 목적으로 하는 국내·외 연구에서는 척도 A 가 폭 넓게 사용되고 있다(최필선·민인식 2013; 김영원·최윤정 2011; Mokrzycki et al. 2009; Traugott & Christopher 2009). 본 연구에서도 개별 전화조사의 정확성 평가를 위해 척도 A 를 사용하기로 한다. 여기서는 편의상 새누리당(Saenuri Party)을 ‘S’ 새정치민주연합(The New Politics Alliance for Democracy; 이하 새정연)을 ‘D’로 표기하기로 한다.

$$A = \log\left(\frac{s/d}{S/D}\right) \quad (1)$$

여기서 s 와 d 는 여론조사를 통해 얻은 득표율 예측값, S 와 D 는 개표결과 두 후보자들의 최종 득표율을 나타낸다.

정확성 척도 A 는 오즈비(odds ratio)에 로그(log)를 취한 범주형 자료분석에서 자주 사용되는 통계량으로 흔히 “로그오즈비”라고 부른다. Martin et al.(2005)이 언급한 대로 척도 A 는 오즈(r/d)를 사용하기 때문에 여론조사에서 부동층이나 기타 정당 후보자의 득표율을 무시하는 경우나 응답자 지지율에 따라 1, 2위 후보자에게 비례배분하는 경우 모두 같은 값을 갖게 된다. 이는 다른 척도들이 부동층 처리방법에 따라 정확성이 달리 계산된다는 것과 비교하면 A 가 갖는 큰 장점이 될 수 있다. 아울러 A 가 0이면(즉, 오즈비가 1이면) 여론조사결과가 개표결과를 정확히 예측한 것을 나타내며, A 가 0에서 멀어질수록 정확성이 떨어지는 것으로 보면 된다. 아울러 A 가 0보다 커지면(작아지면) 여론조사결과가 S정당(D정당) 후보를 과대 추정한 것을 의미하므로 편향에 대한 해석이 매우 용이하다. 아울러 4절에서와 같이 선형모형 등을 이용한 분석이 가능하다.

한편 A 값이 표본추출오차로 설명될 수 있는 수준을 벗어나는 경우 해당 여론조사에서 특정 정당에 대해 편향이 발생한 것으로 해석할 수 있는데, 여기서 오즈비를 로그(log) 변환한 이유는 A 가 대칭성을 갖는 동시에 분산추정이 쉽다는 측면도 고려한 것이다. 단순확률추출을 가정하는 경우 A 의 분산은 다음과 같이 추정될 수 있다(Martin et al. 2005).

$$Var(A) = Var\left(\log\left(\frac{s/d}{S/D}\right)\right) \cong Var\left(\frac{s/d}{S/D}\right) \hat{=} \frac{1}{nsd} \quad (2)$$

한편 특정 여론조사에서 오차가 발생하면, 다시 말해 어떤 여론조사에서 A 가 0과 차이가 있으면 예측오차(forecasting error)가 발생한 것으로 볼 수 있는데, 이런 오차를 어떤 경우 랜덤(random) 오차를 벗어난 조사방식 등에 구조적인 문제가 있어서 발생한 편향으로 볼 것인지에 대해 정리가 필요하다. A 를 사용한 선행 연구들(Martin et al. 2005; 최필선·민인식 2013; 김영원·최윤정 2011)을 보면 예측값이 참값으로부터 지나치게 크거나 작아지면,

다시 말해 일상적인 표본추출오차로 설명되기 어려운 수준으로 오차가 특정 방향으로 발생하는 경우 이는 편향으로 해석하고 있다. 따라서 본 연구에서도 척도 A 가 식(2)를 기준으로 설정한 95% 신뢰수준을 벗어나는 경우 편향이 발생한 것으로 간주하기로 한다.

이와 관련해 우리나라 전화여론조사의 경우 실제로는 지역·성·연령대를 기준으로 한 할당추출법을 적용하고 있으며, 할당이 차지 않는 경우 사후적으로 가중치를 적용해 보정하는 방식을 사용하기 때문에 단순확률추출을 가정한 식(2)의 분산추정이 이론적으로 적절한 방법이 아닐 수도 있다. 특히 가중치가 커지게 되면 가중치의 변동계수(CV)제곱에 비례해 분산이 커지게 되지만, 본 연구에서 사용한 자료에서는 개별 자료에 대한 가중치를 파악할 수 없기 때문에 식(2)를 이용해 구한 신뢰구간을 그대로 사용하기로 한다.

2. 여론조사 현황 및 정확성 분석 대상 자료

여론조사 관련 학계와 조사업계에서는 선거여론조사의 품질을 향상시키기 위해서는 다양한 여론조사가 과연 어떤 방식으로 이루어지고 있고, 구체적으로 어떤 차이가 있는지 비교해 볼 수 있도록 여론조사 수행과정에 관한 정보(paradata)가 투명하게 공개되고, 이를 기반으로 시장에서 관련 여론조사의 신뢰성을 평가하는 것이 필요하다는 주장이 꾸준히 있어 왔다. 특히 여론조사과정에서 수집된 전화접촉기록, 조사일시 및 기간, 표본추출과정, 응답률, 가중치 부여방법 등 다양한 형태의 파라데이터를 체계적으로 공개함으로써 우리나라 조사방법의 발전과 여론조사에 대한 사회적 신뢰를 지속적으로 축적해 나갈 수 있다는 점에 학계와 조사업계가 인식을 같이하고 있다. 이런 취지에서 한국조사연구학회에서는 여론조사정보공개위원회를 구성하고 관련 지침을 제시하기도 했다(이기재 외 2014).

정치권에서도 선거여론조사가 선거결과에 미치는 영향이 매우 크다는 것과 함께 잘못된 여론조사에 의해 피해를 볼 수 있다는 생각을 여야가 공유하고 있었기 때문에 선거여론조사의 객관성과 신뢰성을 높이는 방안을 모색하

는 데 의견을 함께하고, 2014년 2월 「공직선거법」 개정을 통해 공표 또는 보도를 목적으로 하는 선거여론조사의 객관성과 신뢰성을 확보하는 것을 목적으로 선거여론조사공정심의위원회(이하 ‘여심위’)를 설치·운영하도록 하였다.

이에 따라 중앙 여심위에서는 2014년 제6회 전국동시지방선거를 앞두고 ‘선거여론조사기준’을 공표했다(중앙선거여론조사공정심의위원회 2014a). 이 기준에 따르면 선거여론조사를 실시한 기관이나 단체가 해당 여론조사를 공표·보도하기 위해서는 조사 의뢰자 및 조사기관, 조사방법, 표본의 크기(연령대별·성별 표본의 크기를 포함), 피조사자 선정방법, 피조사자 접촉 현황, 응답률, 연령대별·성별 가중치, 표집오차, 질문지, 분석 결과 등을 여심위 홈페이지에 등록하여야 한다. 결과적으로 2014년 지방선거부터는 여심위 홈페이지에 등록된 선거여론조사 관련 정보를 활용하면 조사방식 등에 따른 예측오차나 편향을 분석하는 연구가 가능하게 되었다.

중앙선거여론조사심의위원회(2014b)에 의하면 제6회 전국동시지방선거와 관련해 여심위 홈페이지에 816건의 선거여론조사가 등록되었으며, 이 중 기초단체장을 대상으로 한 조사가 390건(47.8%)이고, 시·도지사에게 대한 조사가 323건(39.6%)를 차지하고 있으며, 그 외에 교육감, 광역의원, 기초의원 등에 대한 조사를 포함하고 있다.

조사방법에 따라 이들 여론조사를 구분해 보면 전화면접조사가 322건(39.5%), ARS전화조사가 458건(56.1%), 그밖에 유선 또는 무선 전화조사와 스마트폰앱을 병합하는 방식의 조사가 36건(4.4%) 있었다. 피조사자 선정방법에 있어서는 유선 RDD가 523건(64.1%), 유·무선 RDD 방식의 조사가 110건(13.5%) 등으로 대부분의 조사가 RDD 방식으로 이루어지고 있다.

일반적으로 학계나 국내 주요 조사회사 등에서 신뢰성과 타당성에 문제가 있다는 지적과 함께 사용을 자제할 것을 권장¹⁾하는 ARS 조사가 차지하는 비중이 56.1%로 상당히 높다. 이렇게 많은 여론조사가 ARS 방식으로 이루어진

1) 참고로 국내 주요 41개 여론조사회사가 회원으로 가입하고 있는 ‘한국조사협회’는 2014년 국가주요정책 수립에 이용되는 여론조사의 신뢰성을 제고하기 위해 전체 회원사가 향후 ARS조사를 수행하지 않을 것을 결의한 바 있다.

것은 이번 선거가 동시지방선거이기 때문에 기초장이나 지방의회 의원을 선출하는 작은 규모의 선거가 동시에 이루어졌고, 이런 조사의 경우 정확성보다는 저렴한 비용을 우선시하는 경향이 강하기 때문에 나타난 현상으로 볼 수 있다. 실제 대다수 유권자가 관심을 갖는 시·도지사 선거만을 떼어서 보면 전체 323건 중 ARS조사는 92건으로 28.5% 수준인 것으로 나타났다.

본 연구에서는 여심위 홈페이지에 등록된 2014년 지방선거 여론조사 중 광역단체장(특·광역시장 및 도지사) 선거예측을 위한 전화조사만을 분석 대상으로 하기로 한다. 이 중 편향 분석을 위해 새누리당과 새정치민주연합 후보자가 경합을 하지 않은 광주광역시장(새정연 윤장현 대 무소속 강운태) 선거와 후보 단일화가 늦게 이루어져 실제 등록 입후보자에 대한 여론조사가 거의 이루어지지 못한 울산광역시장(새누리당 김기현 대 정의당 조승수) 선거에 대한 여론조사는 분석대상에서 제외하기로 한다. 한편 부산의 무소속 오거돈 후보의 경우 정치성향을 고려해 새정연 후보로 간주하고 편향을 분석하기로 한다. 또한 후보등록 기간이 15~16일이었던 점을 고려해 정당별 입후보자가 확정된 이후의 양당 후보자에 대한 여론조사만을 분석하기 위해 17일 이후에 조사가 수행된 여론조사만을 대상으로 하였으며, 응답률이 1% 이하인 것으로 등록된 5개 여론조사는 분석대상에서 제외하였다. 최종적으로 15개 특·광역시장 및 도지사 선거에 대한 전화면접조사 또는 ARS전화조사 방식으로 조사가 수행된 150개 전화조사를 분석대상으로 하였다.

Ⅲ. 선거예측 전화조사 정확성 비교

우선 150개 분석대상 전화여론조사를 조사방법, 지역, 조사기관, 조사일시 및 기간 등에 따라 몇 개의 범주로 구분하고 각 범주별로 편향과 예측오차에 있어 어떤 차이가 있는지 살펴봄으로써 조사방법에 따른 예측오차를 비교해 보기로 한다. 여기서는 해당 범주에 속하는 전체 여론조사 수를 n 으로 표시하고, 해당 범주에 속한 개별 여론조사의 정확성 척도 A_i 의 평균($\bar{A}$), A_i 의

절대값들의 평균($\overline{|A|}$), A_i 들의 표준편차($sd(A)$), 또한 해당 범주에 속하는 여론조사 n 개 중 식 (2)를 기준으로 설정한 95% 신뢰구간에서 양 또는 음의 방향으로 벗어난 편향된 여론조사의 수($n(+)$ 및 $n(-)$) 등을 제시하고 이를 기준으로 편향이나 예측오차를 분석하기로 한다.

1. 지역별 예측오차 비교

분석대상 150개 여론조사를 지역(특·광역시 및 도)별로 구분하고, 척도 A_i 를 토대로 산출된 요약값들을 정리하면 <표 1>과 같다. 우선 대전, 세종, 전북, 경남, 제주 등 5개 지역의 경우 모든 조사가 신뢰구간을 벗어나는 편향이 발생한 것으로 나타났다. 이 중 전북 지역의 경우(분석대상 조사가 1개) A_i 가 음의 방향으로 신뢰구간을 벗어나는 음의 편향(실제보다 새정연 후보를 과대추정하는 경향)을 보였으며, 나머지 4개 지역에서는 모든 조사가 양의 편향(실제보다 새누리당 후보를 과대추정하는 경향)이 발생했다. 아울러 $\overline{A}$ 를 보면 전북의 경우 -0.4505로 다른 지역에 비해 상대적으로 큰 음수이고, 나머지 양의 편향을 보인 4개 지역은 0.3492~0.4849로 상당히 큰 양수값을 갖는다. 이들 지역의 경우 평균 예측오차로 해석할 수 있는 $\overline{|A|}$ 역시 다른 지역에 비해 상당히 크다는 것을 확인할 수 있다.

다른 지역의 경우를 살펴보면, 인천과 충남의 경우 음의 편향이 발생한 지역으로 볼 수 있으며, 경기의 경우 양의 편향이 발생한 지역으로 볼 수 있다. 전체적으로 28개 조사회사가 150개의 여론조사를 수행하였으며, 150개 전화조사 중 양의 편향이 발생한 것이 44개(29.33%)이고 음의 편향이 발생한 것이 22개(14.67%)로 2014년 전화여론조사에서는 전반적으로 새누리당을 실제보다 과대추정하는 양의 편향이 발생하는 경향이 있었다는 것을 알 수 있다.

<표 1> 지역별 편향 및 예측오차

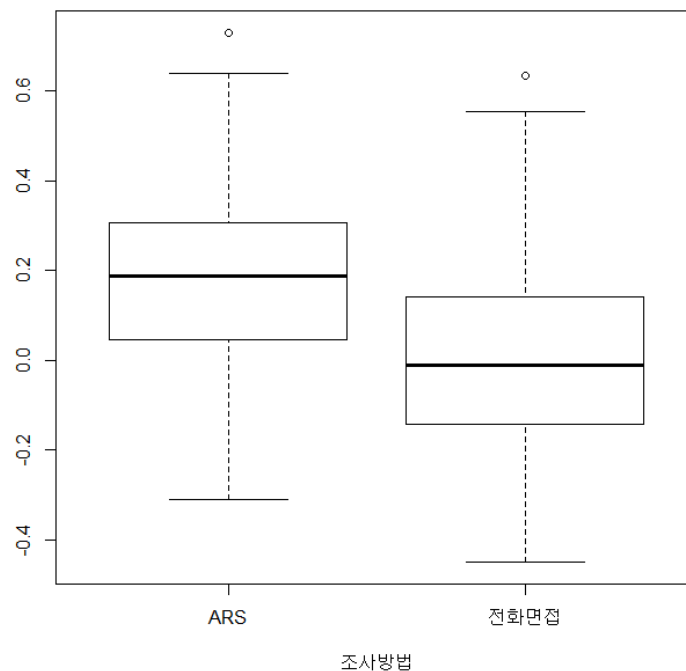
지역(선거명)	$\bar{A}$	$ \bar{A} $	$sd(A)$	n	$n(+)$	$n(-)$	편향비율	조사회사 수
서울시장	-0.0562	0.1196	0.1387	17	2	4	0.3529	13
부산시장	0.0409	0.1821	0.2480	17	4	2	0.3529	13
대구시장	-0.0174	0.0949	0.1431	7	0	1	0.1429	6
인천시장	-0.1265	0.1710	0.1543	19	1	8	0.4737	14
대전시장	0.3548	0.3548	0.1558	6	6	0	1.0000	3
세종시장	0.3849	0.3849	0.1330	12	12	0	1.0000	6
경기도지사	0.1263	0.1331	0.1069	24	7	0	0.2917	16
강원도지사	-0.0352	0.0478	0.0663	13	0	1	0.0769	9
충북도지사	-0.0742	0.1225	0.1148	12	1	2	0.2500	10
충남도지사	-0.1403	0.1895	0.1980	7	0	3	0.4286	3
전북도지사	-0.4505	0.4505	-	1	0	1	1.0000	1
전남도지사	0.0642	0.1875	0.2652	2	0	0	0.0000	2
경북도지사	0.1074	0.1146	0.1620	2	0	0	0.0000	2
경남도지사	0.4849	0.4849	0.1124	4	4	0	1.0000	2
제주도지사	0.3492	0.3492	0.0960	7	7	0	1.0000	5
전체	0.0596	0.1856	0.2326	150	44	22	0.4400	28

2. 조사방법에 따른 예측오차 비교

분석대상 150개 전화조사를 조사방법에 따라 크게 구분해 보면 전화면접 조사와 ARS전화조사로 나눌 수 있다. 분석대상 150개 전화조사 중 ARS전화 조사는 33개이고 전화면접조사는 117개였다. 전체 전화조사를 면접조사와 ARS조사 두 개 범주로 나누어 조사방법에 따른 예측 정확성을 비교해 보면 <표 2>와 같고, 각 범주에 속하는 여론조사의 정확성을 비교해 보기 위해 A_i 들의 상자그림을 그리면 <그림 1>과 같다.

<표 2> 조사방법별 편향 및 예측오차

조사방법	$\bar{A}$	$ \bar{A} $	$sd(A)$	n	$n(+)$	$n(-)$	편향비율	조사회사 수
ARS 전화조사	0.0596	0.2322	0.2196	33	20	1	0.6364	11
전화면접조사	0.0206	0.1725	0.2219	117	24	21	0.3846	18

<그림 1> 조사방법에 따른 정확성 척도(A_i) 분포 비교

우선, 전화면접조사와 ARS조사에서 편향이 발생한 여론조사 비율을 <표 2>에서 보면, ARS 전화조사의 경우 전체 33개 중 21개가 편향이 발생해 편향비율이 0.6364이었으며, 편향 대부분(21개 중 20개)이 양의 방향으로 발생했다는 점에 특히 유의할 필요가 있다. 다시 말해 ARS전화조사의 경우 새누리당 후보를 과대추정하는 체계적인 편향이 매우 강하게 발생하였다는 것을 알 수 있다. 반면 전화면접조사의 경우 117개 중 45개의 여론조사에서 편향이 발생해 편향비율은 0.3846인 것으로 나타나 ARS에 비해 편향비율이 낮았

고, 편향이 양의 방향으로 24개, 음의 방향으로 21개가 발생하여 ARS 조사와는 달리 특정 정당에 유리한 체계적인 편향이 발생하지는 않은 것으로 보인다.

실제 각 여론조사에서 산출된 A_i 값을 하나의 관측값처럼 생각하고 전화면접조사와 ARS조사에서 발생하는 예측오차 평균에 대한 차이를 $\overline{A}$ 를 기준으로 t -검정을 통해 비교해 보면, t -값은 4.0669이고, 유의확률 p -값은 0.0001 보다 작은 것으로 나타났다. 따라서 ARS조사와 전화면접조사는 편향 발생 측면에서 통계적으로 유의한 차이가 있다고 결론내릴 수 있다. <그림 1>을 보면 전화면접조사의 경우 ARS조사에 비해 A_i 값들이 상대적으로 0을 중심으로 집중되어 분포해 있어 정확성 측면에서 전화면접조사가 ARS조사에 비해 우수하다는 것을 확인할 수 있다.

아울러 <표 2>에서 $\overline{A}$ 는 A_i 들을 평균한 것으로 양과 음의 편향이 균형을 이루는 경우 예측오차가 크더라도 0에 가까운 값이 나올 수도 있기 때문에, A_i 를 기초로 ARS조사와 전화면접조사의 예측오차를 비교하기 위해서는 절대값을 취한 A_i 들의 평균인 $|\overline{A}|$ 를 살펴보는 것이 의미가 있다. $|\overline{A}|$ 를 기준으로 보더라도 ARS조사의 경우 0.2322, 전화면접조사의 경우 0.1725로 상대적으로 전화면접의 경우 예측오차가 적어진다는 것을 확인할 수 있다. 이 경우 정규분포 가정에 문제는 있지만, t -검정을 해보면, 유의확률 p -값이 0.0456으로 통계적으로 유의한 차이가 있다고 결론내릴 수 있다.

3. 조사회사별 예측오차 비교

동일한 전화면접방식으로 조사가 수행되었다고 해도 실제 어떤 조사회사가 조사를 실시하는지에 따라 조사수행 시간대 및 기간, RDD 방식, 부재자 재통화실시 여부, 패널 사용 여부 등에 있어서 차이가 있기 때문에 표본추출 오차로 설명될 수 없는 차이가 발생할 수 있다. 흔히 이런 차이를 조사기관 효과(housing effect) 또는 조사기관 편향(housing bias)이라고 부른다(박종희 2013; 최필선·민인식 2013). 2014년 지방선거에서도 조사기관이 달라짐에

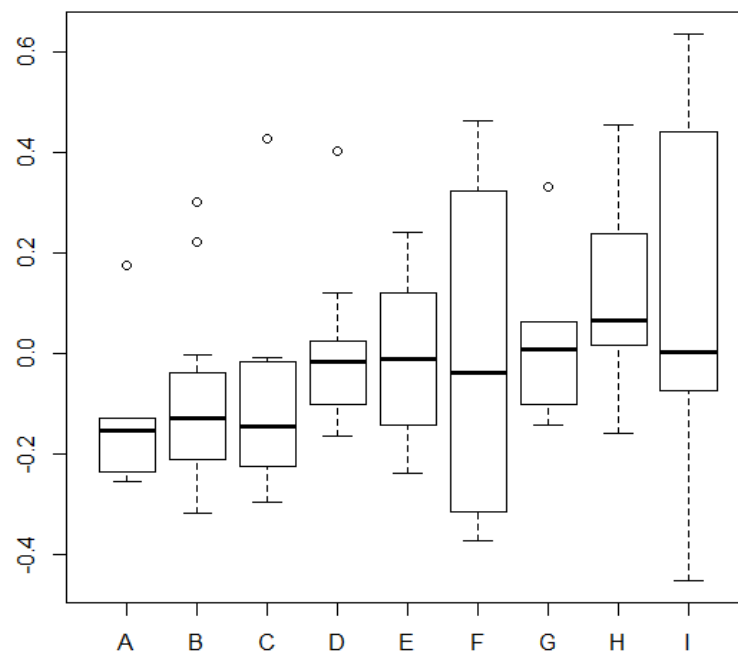
따라 예측오차에 차이가 생겼는지 보기 위해 분석대상 150개 여론조사 중 5회 이상 전화여론조사를 수행한 9개 조사회사의 결과만을 대상으로 조사기관별 예측 정확성을 비교해 보면 <표 3>과 같다. 아울러 조사회사별 A_i 값들에 대한 상자그림인 <그림 2>를 보면 조사회사별로 예측오차 패턴에 있어서 상당히 큰 차이가 있다는 것을 볼 수 있다. 제시된 A_i 를 기준으로 한 정확성 평가 결과가 조사기관의 순위를 매기는 절대적인 기준으로는 볼 수 없기 때문에 여기서 조사회사 이름은 익명으로 처리하였다는 점을 참고하기 바란다.

<표 3>과 <그림 2>를 보면, 우선 편향이라는 관점에서 D사가 가장 우수한 것으로 볼 수 있으며, 이 회사의 경우 이상값에 해당하는 여론조사(양의 편향에 해당함)가 한 지역에서 발생했지만 이 지역을 제외하면 전반적으로 상당히 안정적이고 정확한 조사결과를 산출하고 있다. D사의 경우 예측오차를 설명해 줄 수 있는 $\overline{|A|}$ 도 0.1103으로 9개 조사회사 중 가장 작은 것으로 나타났으며, D사가 수행한 10개 조사 중 편향이 발생한 것도 1개로 다른 회사들에 비해 편향비율도 가장 낮았다.

편향의 관점에서 조사회사들을 전체적으로 비교해 보면 A사의 경우 체계적인 음의 편향이 발생하였으며, 반면에 H사와 I사의 경우 다른 회사에 비해 상대적으로 양의 편향이 발생하는 경향이 있다. 특정 조사회사가 수행한 조사 중 편향이 발생한 조사의 비율을 나타내는 편향비율로 보면 F, C, I, E사 등이 정확성이 떨어졌고, 평균오차를 설명해 주는 $\overline{|A|}$ 를 기준으로 보면 F와 I사가 상대적으로 정확성이 떨어진다. 특히 F사의 경우 10개 조사 중 9개 조사에서 편향이 발생해 편향비율이 가장 높을 뿐만 아니라 $\overline{|A|}$ 를 기준으로 하는 경우에도 예측오차가 가장 큰 것으로 나타났다. <그림 2>의 상자그림에서도 F와 I사의 경우 오차의 발생폭이 상당히 크고 동시에 음과 양의 방향으로 편차가 심하게 요동치는 것을 볼 수 있다. 이 회사들의 경우 상대적으로 다른 조사회사에 비해 조사방법의 안정성이 매우 떨어진다는 것을 의미한다. 또한 상자그림에서 A사의 경우를 보면, 오차의 변동이 적어 상당히 안정적인 조사결과를 내고 있지만 체계적인 음의 편향이 발생하고 있다는 특징을 갖고 있다.

<표 3> 조사회사별 편향 및 예측오차

조사회사	$\bar{A}$	$ \bar{A} $	$sd(A)$	n	$n(+)$	$n(-)$	편향비율
A	-0.1249	0.1830	0.1551	6	0	2	0.3333
B	-0.0886	0.1758	0.1848	12	1	3	0.3333
C	-0.0705	0.1924	0.2439	7	1	3	0.5714
D	0.0045	0.1103	0.1635	10	1	0	0.1000
E	0.0076	0.1362	0.1688	13	3	3	0.4615
F	0.0116	0.3250	0.3561	10	4	5	0.9000
G	0.0281	0.1195	0.1692	6	1	0	0.1667
H	0.1126	0.1555	0.1687	24	9	1	0.4167
I	0.1225	0.2459	0.3089	19	7	2	0.4737

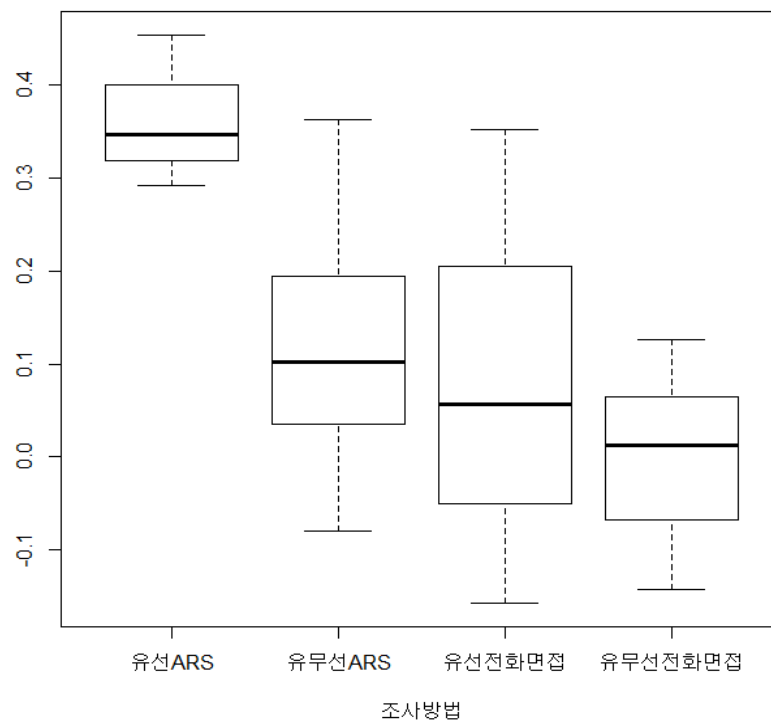
<그림 2> 조사회사별 정확성 척도(A_i) 분포 비교

참고로, 음의 편향이 가장 크게 발생한 A사가 수행한 6개 전화조사는 모두 유·무선전화 병합조사였고, A사에서는 무선전화 조사를 위해서 해당 조사회사가 보유하고 있는 패널을 사용하였다는 점에 특히 유의할 필요가 있다. 본 연구에서는 패널 사용 여부 또는 유·무선전화 병합조사 여부 등을 별도의 조사방법으로 구분해 분석하지 않았다. 왜냐하면 지방선거의 경우 현실적으로 일반적인 유·무선전화 병합조사가 가능한 지역이 서울이나 경기 등 매우 제한적이고, 이런 측면으로 인하여 유·무선전화 병합조사를 수행한 회사 중 상당수가 자체적으로 보유한 패널을 사용한 경우가 다수이기 때문에 패널 사용 여부를 구분하지 않고 유·무선전화 병합조사로 등록된 조사들을 하나의 범주로 묶는 것은 타당성이 결여될 수 있다는 점을 고려했기 때문이다. 패널 사용 여부가 예측오차에 미치는 영향에 대해서는 3절의 일반선행모형을 사용한 분석에서 개략적으로 다시 살펴보기로 한다.

한편 <표 3>에 있는 H사의 경우 다른 조사회사와는 달리 유선 ARS, 유·무선 ARS, 유선 전화면접 및 유·무선 전화면접 등 4가지 조사방법을 모두 사용하고 있는 회사이다. 따라서 동일한 회사에서 수행된 이들 4가지 조사방법에 따른 차이를 비교해 보는 것은 매우 흥미로운 분석이 될 수 있다. 물론 이 회사(또는 조사의뢰자)가 이들 조사방법을 선택하는 과정에서 조사지역의 경합도 등을 고려해 해당 지역에서 어떤 조사방법을 사용할 것인지를 결정했을 것으로 추측되기 때문에 지역 효과와 조사방법 효과가 일부 교란이 발생할 수 있다는 한계는 있겠지만, 무엇보다 조사기관 효과를 제거하고 순수하게 조사방법에 따른 예측 정확성을 객관적으로 비교해 볼 수 있다는 측면에서 상당히 의미가 있을 것이다. H사에서 수행한 24개의 광역단체장 여론조사를 조사방법에 따라 4가지 범주로 구분하여 예측오차를 비교한 결과는 <표 4>와 같고, 조사방법별 A_i 값들의 분포를 보기 위한 상자그림은 <그림 3>과 같다.

<표 4> H사 여론조사 조사방법별 편향 및 예측오차

조사방법	$\bar{A}$	$ \bar{A} $	$sd(A)$	n	$n(+)$	$n(-)$	편향비율
유선 ARS	0.3641	0.3641	0.0824	3	3	0	1.0000
유·무선 ARS	0.1245	0.1390	0.1294	11	5	0	0.4545
유선 전화면접	0.0838	0.1887	0.2560	3	1	1	0.6667
유·무선 전화면접	-0.0014	0.0777	0.0978	7	0	0	0.0000

<그림 3> H사 여론조사 조사방법별 정확성 척도(A_i) 분포 비교

H사에서 실시한 전화조사의 경우 <표 4>와 <그림 3>을 보면 대체적으로 양의 편향이 발생하고 있으며, 특히 유선 ARS조사의 경우 수행된 3개 조사

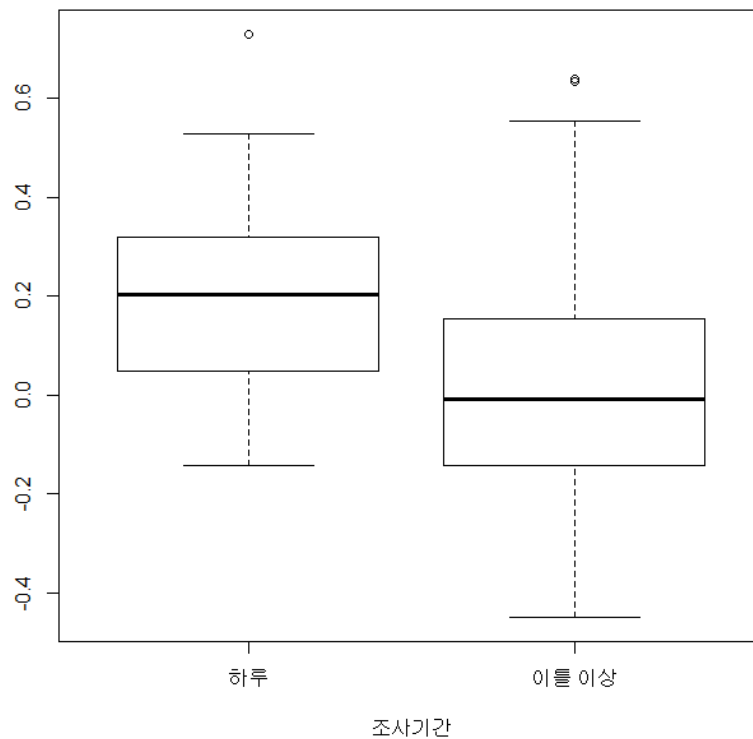
모두 신뢰구간을 벗어나는 심한 양의 편향이 발생하고 있다. 평균오차를 나타내는 $\overline{|A|}$ 와 편향비율을 보면 다른 조사에 비해 유·무선 전화면접 조사가 예측오차나 편향 측면에서 모두 가장 안정적으로 정확한 예측을 해주는 조사 방법이라는 것을 알 수 있다. 여기서 각 범주에 속하는 전화조사 개수가 별로 많지 않지만 분산분석을 해보면 F -값은 5.14이고 p -값은 0.0085로 조사방법에 따른 차이가 있는 것으로 나타났다. <그림 3>을 보면, 일반적으로 알려진 것처럼 유·무선 전화면접이 편향이나 예측오차 관점에서 가장 우수하고, 그 다음 유선 전화면접, 유·무선 ARS, 유선 ARS 순서로 예측오차나 편향이 점점 커지는 경향이 있다는 것을 볼 수 있다.

4. 조사기간 및 조사시점에 따른 예측오차 비교

재통화 규칙 등을 준수하여 이틀 이상의 기간 동안 수행되는 전화조사는 당일치기 전화조사에 비해서 비용은 더 많이 들지만, 보다 신뢰할 수 있는 결과를 얻을 수 있다고 일반적으로 알려져 있다. 이런 가설이 2014년 지방선거 여론조사들에서도 성립하는지 확인해 보기 위해 150개 여론조사를 하루 동안 수행된 31개 조사와 조사기간이 이틀 이상인 119개 조사로 구분하여 예측 정확성을 비교한 결과는 <표 5>에 제시되어 있다. 조사기간을 기준으로 두 범주를 구분해 A_i 들에 대한 상자그림은 <그림 4>와 같다.

<표 5> 조사기간에 따른 편향 및 예측오차

조사기간	$\overline{A}$	$\overline{ A }$	$sd(A)$	n	$n(+)$	$n(-)$	편향비율
하루	0.1889	0.2125	0.2027	31	17	0	0.5484
이틀 이상	0.0259	0.1786	0.2288	119	27	22	0.4118



〈그림 4〉 조사기간에 따른 정확성 척도(A_i) 분포 비교

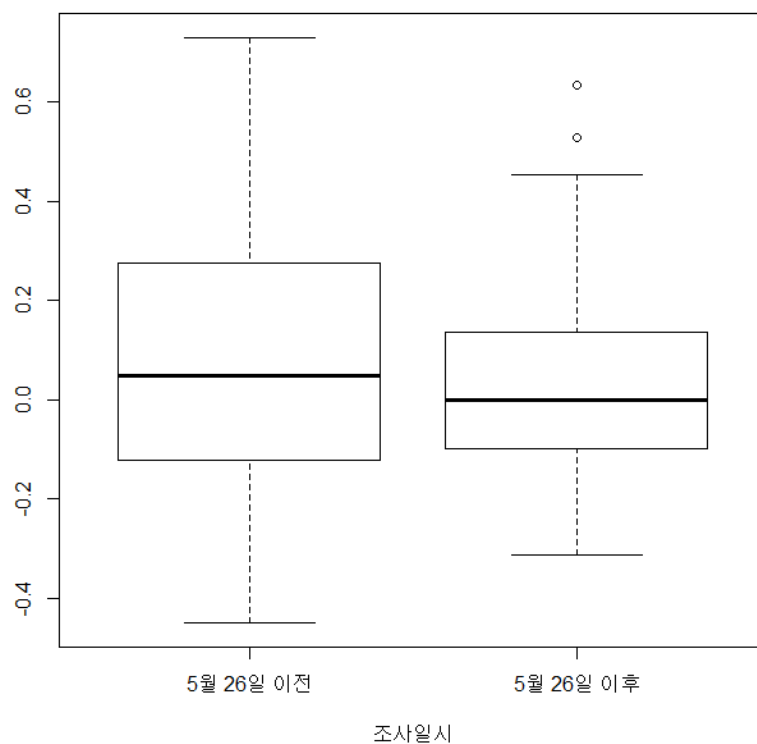
조사기간이 하루인 조사와 이틀 이상인 조사에서 발생한 예측오차를 비교해 보면 우선 하루만 조사하는 전화조사의 경우 $\bar{A} = 0.1889$ 로 상대적으로 심한 양의 편향이 발생하게 된다는 것을 알 수 있다. 편향비율의 측면에서 보면 31개 조사 중 17개 조사에서 편향이 발생해 편향비율이 0.5484이고 특히 이들 17개 조사의 편향이 모두 양의 방향으로 나타나고 있다는 점을 참고로 하면, 2014년 지방선거에서 하루 동안 조사하는 당일치기 조사방식이 체계적인 양의 편향을 발생시키는 요인이었다는 것을 확인할 수 있다. 실제 두 범주에서 $\bar{A}$ 를 기준으로 한 t -검정 결과 t -값은 3.6111이고 유의확률 p -값은 0.0004로 통계적으로 유의한 차이가 있다는 결론을 내릴 수 있다.

한편, 일반적으로 투표일에 가까운 시점에 이루어진 조사와 투표일에서 멀리 떨어진 시점에 수행된 조사의 예측정확성에 대해서도 차이가 있다고 흔히 생각한다. 따라서 선거법에 의해 공표가 가능한 여론조사 수행시점이 5월 28

일까지라는 것을 고려해 전체 150개 조사를 26일 이전에 완료한 조사들과 26일 이후(26일 포함)에 수행된 조사로 구분하여 예측오차를 비교해 본 결과는 <표 6>과 같고, 실제 26일 이후에 수행된 조사는 거의 모든 조사회사에서 선거를 앞두고 지역별로 마지막으로 수행한 조사로 보면 된다. 조사일시를 기준으로 구분해 A_i 들의 분포를 살펴보기 위한 상자그림은 <그림 5>와 같다.

<표 6> 조사일시 구분에 따른 편향 및 예측오차

조사일시	$\bar{A}$	$ \bar{A} $	$sd(A)$	n	$n(+)$	$n(-)$	편향비율
5.26 이전	0.0759	0.2072	0.2500	90	32	16	0.5333
5.26 이후	0.0352	0.1532	0.2035	60	12	6	0.3000



<그림 5> 조사일시 구분에 따른 정확성 척도(A_i) 분포 비교

예측오차 분석결과인 <표 6>을 참고로 5.26 이전과 이후에 수행된 조사가 편향에 있어서 차이가 있는지 $\overline{A}$ 를 기준으로 t -검정을 하면 p -값이 0.2951로 통계적으로 유의한 차이가 있다고 볼 수는 없다. 하지만 평균오차를 나타내는 $|\overline{A}|$ 는 상대적으로 5.26일 이후 조사에서 작아졌고, <그림 5>에서도 5.26일 이후 조사가 상대적으로 0을 중심으로 안정적인 분포를 보이고 있다는 것을 알 수 있다. 5.26 이전과 이후 두 범주 간에 오차가 차이가 있는지 통계적으로 검정하기 위해 $|\overline{A}|$ 를 기준으로 t -검정을 하면 유의확률 p -값은 0.0640으로 통계적으로 유의한 차이가 있다고 볼 수 있다. 다시 말해 편향에 있어서는 차이가 없지만 투표일에 임박해서 수행된 조사의 경우 상대적으로 예측오차가 작아진다고 말할 수 있다.

5. 응답률과 가중치 적용에 따른 예측오차 비교

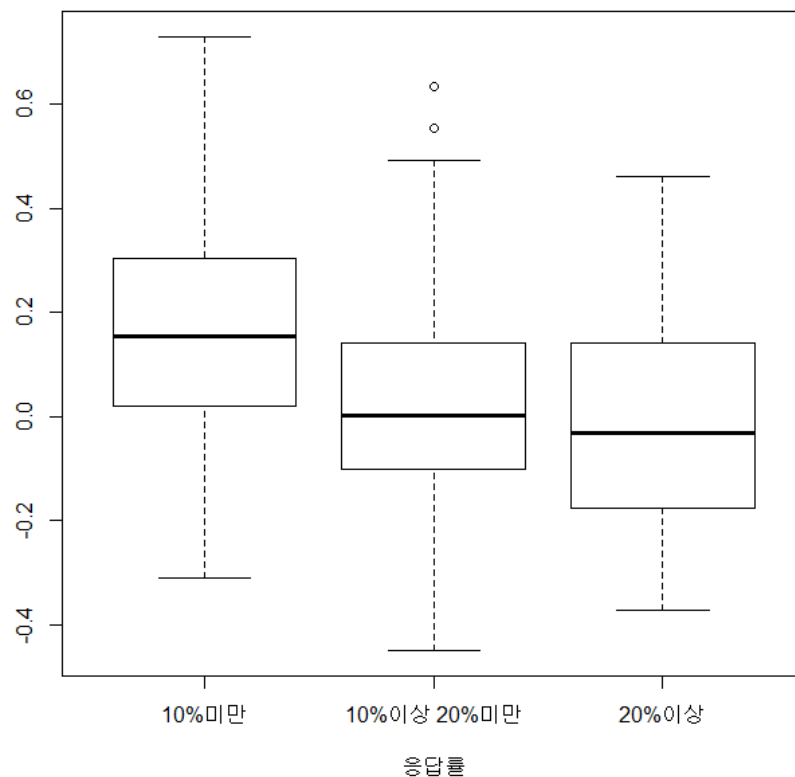
여론조사의 품질을 평가함에 있어 전통적으로 응답률을 중요하게 생각해 왔다. 따라서 2014년 지방선거 전화조사 결과들을 대상으로 응답률이 높아짐에 따라 과연 조사의 정확성이 높아진다고 볼 수 있는지 살펴보기로 한다.

우선 150개 분석대상 여론조사에서의 응답률과 정확성 척도 A_i 와의 상관계수를 구해 보면 -0.2215 이고, p -값은 0.0065으로 음의 상관관계가 있는 것으로 나타났다. 다시 말해 응답률이 증가하면 편향이 음의 방향으로 움직이는 경향이 있다는 것을 의미한다. 하지만 이런 음의 상관관계가 응답률이 높아지면 오차가 줄어든다는 것을 의미하는 것은 아니다. 응답률에 따라 예측오차가 어떤 차이를 보이는지 살펴보기 위해 응답률을 10%, 20%를 기준으로 3개 범주로 구분해 정확성을 비교해 보면 <표 7>과 같고, 각 범주별로 A_i 들의 분포를 보기 위한 상자그림은 <그림 6>과 같다.

<표 7>을 보면 응답률이 10% 미만인 경우 $\overline{A} = 0.1723$ 으로 양의 편향이 발생하지만, 응답률이 10% 이상으로 증가하면 이런 양의 편향이 발생하는 현상이 없어짐을 알 수 있다. 편향비율을 보면 응답률이 10% 미만인 38개 조사 중 양의 편향이 20개, 음의 편향이 1개 발생했다는 것에서도 체계적인 양의 편향이 발생하였다는 것을 확인할 수 있다. 실제 이들 3개 범주에 대한 분

<표 7> 응답률 구분에 따른 편향 및 예측오차

응답률	$\bar{A}$	$ \bar{A} $	$sd(A)$	n	$n(+)$	$n(-)$	편향비율
10% 미만	0.1723	0.2121	0.2184	38	20	1	0.5526
10%~20% 미만	0.0414	0.1750	0.2270	78	17	11	0.3590
20% 이상	-0.0244	0.1804	0.2190	34	7	10	0.5000

<그림 6> 응답률 구분에 따른 정확성 척도(A_i) 분포 비교

산분석(ANOVA)을 하면 F -값이 7.51이고 p -값이 0.0008로 범주에 따라 통계적으로 유의한 차이가 있다고 말할 수 있다. 또한 이들 3개 범주에 대한 Tukey의 다중비교를 해보면 응답률이 10% 미만인 경우와 나머지 두 가지 범주는 통계적으로 유의한 차이가 있다는 결론을 내릴 수 있다.

제시된 응답률 분석결과와 관련해 한 가지 주의해야 할 사항이 있다. 앞서도 언급했지만 분석대상이 되는 150개 여론조사 중 유·무선전화 병합조사를 했다고 등록한 다수의 전화조사의 경우 사전에 구축해 놓은 패널에서 무선전화번호를 추출해 조사를 수행했다. 이들 패널을 사용한 조사의 경우도 현행 여심위 선거여론조사기준에 따르면 일반 전화조사와 동일하게 통화에 성공한 전화회선 중에서 응답이 완료된 회선의 비율(이런 비율은 조사 응답률보다는 협조율로 부르는 것이 더 적절함)로 계산하게 되어 있기 때문에 패널을 사용한 경우 응답률이 높게 계산된다.

따라서 <표 7>의 결과는 패널을 사용한 조사의 응답률과 패널을 사용하지 않은 경우의 응답률을 동일한 것으로 보고 분석했기 때문에 실제적으로 응답률에 따른 여론조사의 정확성을 평가하는 데 한계가 있다. 그럼에도 불구하고 10% 미만의 응답률을 갖는 경우와 10% 이상의 응답률을 갖는 여론조사(모든 패널조사는 이 범주에 포함됨)를 비교했을 때, 편향이란 관점에서 통계적으로 유의한 차이가 있다는 것은 응답률이 10% 이상으로 높아지게 되면(패널 사용 여부와 상관없이) 편향이란 관점에서 정확성을 제고하는 데 도움이 된다고 이해하면 된다.

한편 할당추출의 경우 특정 범주에 할당된 표본을 채우지 못하는 경우 최종적으로 가중치를 주어 보정을 하게 된다. 결국 조사의 가중치 변동이 커진다는 것은 할당을 못 채운 정도가 심하다는 것을 의미하기 때문에 편향 발생 가능성이 높아진다고 알려져 있으며, 이론적으로 보면 할당을 제대로 채운 경우에 비해 표본추출오차도 커지게 된다.

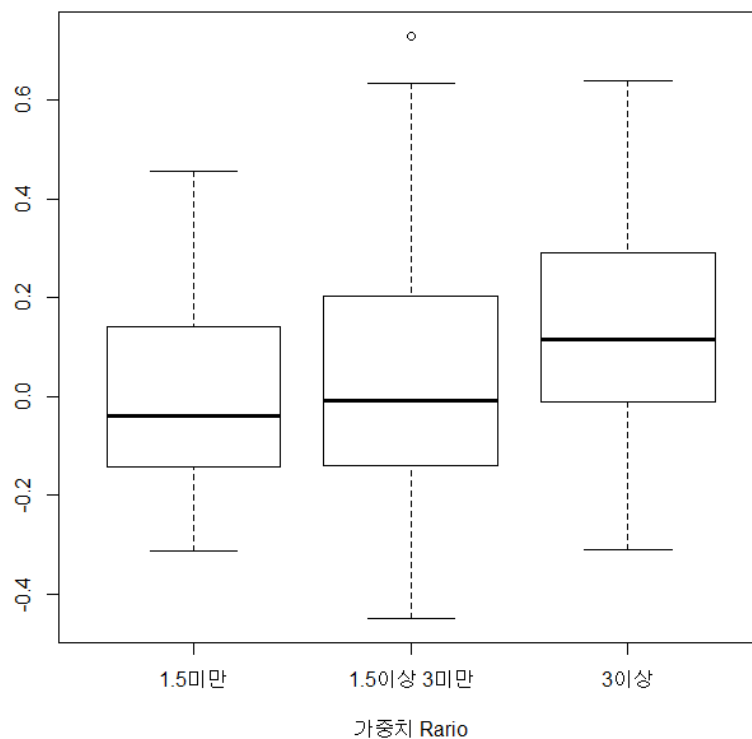
여심위 등록 자료에는 성별 가중치 현황과 연령대별 가중치 현황만이 별도로 등록되어 있기 때문에 실제 개별 자료에 대한 가중치는 확인이 불가능하다. 따라서 여기서는 가중치 변동 폭이 큰 연령대별 가중치에서 최대값을 최소값으로 나눈 일종의 최소가중치 대비 최대가중치의 비(ratio)를 ‘가중치 비’로 부르기로 하고 이를 기준으로 가중치 변동이 커지게 되면 예측 정확성 A_i 에 어떤 영향을 미치는지 살펴보기로 한다. 우선 가중치 비와 A_i 와의 상관계수를 구해 보면 0.1492이고 p -값은 0.0684로 어느 정도 양의 상관관계가 있는 것으로 볼 수 있다. 즉, 가중치 변동이 커지면 A_i 가 양의 값을 갖는 경

향을 보인다는 것을 의미한다.

응답률 분석과 동일하게 가중치 비의 변화가 정확성 척도 A_i 에 미치는 영향을 살펴보기 위해 전체 여론조사를 가중치 비 1.5와 3.0을 기준으로 3개 범주로 구분해 예측오차를 정리해 보면 <표 8>과 같고, 범주별 A_i 들의 분포를 보기 위한 상자그림은 <그림 7>과 같다.

<표 8> 가중치 비 구분에 따른 편향 및 예측오차

가중치 비	$\bar{A}$	$ \bar{A} $	$sd(A)$	n	$n(+)$	$n(-)$	편향비율
1.5 미만	-0.0038	0.1566	0.1872	30	6	6	0.4000
1.5 이상~ 3 미만	0.0340	0.1976	0.2595	70	16	14	0.4286
3 이상	0.1335	0.1862	0.2001	50	22	2	0.4800



<그림 7> 가중치 비 구분에 따른 정확성 척도(A_i) 분포 비교

<표 8>과 <그림 7>을 보면 가중치 비가 커짐에 따라 양의 편향이 발생한다는 것을 알 수 있다. 상자그림에서 가중치 비가 작은 범주에 속하는 A_i 들의 분포가 다른 범주에 비해 편향도 0에 가깝고 변동도 작기 때문에 상대적으로 안정적이고 정확한 예측 결과를 산출할 수 있다는 것도 알 수 있다. 실제 A_i 에 대한 분산분석(ANOVA)을 해보면 F -값이 4.24이고 p -값이 0.0163으로 범주별로 유의한 차이가 있다는 결론을 내릴 수 있다. 다시 말해 가중치 비가 3.0 이상이 되는 경우 다른 두 개의 범주에 비해 양의 편향이 발생할 가능성이 높다. 또한 <표 8>을 보면 가중치 비가 3 이상인 경우 50개 조사 중 24개 조사에서 편향이 발생했는데 이 중 대부분(22개)이 양의 편향이었다. 따라서 가중치 변동이 커지면 양의 편향이 체계적으로 발생한다고 말할 수 있다. 반면에 다른 두 개 범주에서는 양과 음의 편향이 골고루 발생하고 있어 특정 정당에 대한 체계적인 편향이 발생하지 않는다. 여기서도 패널을 사용한 경우 가중치 비가 작아진다는 점을 함께 고려하기 바란다.

한편, <표 7>에 제시된 응답률에 대한 분석결과와 <표 8>의 가중치 비에 대한 분석결과를 비교해 보면 응답률이 낮아지는 것과 가중치 비가 커지는 것이 동일한 방식으로 A_i 에 영향을 주고 있다는 것을 알 수 있다. 이는 전화조사에서 보통 응답률이 낮은 경우 가중치 비가 커지는 경향을 보이기 때문에 나타나는 현상이다. 분석대상 150개 전화조사에서 응답률과 가중치 비의 상관계수는 -0.4807이고, p -값은 0.0001 이하로 매우 강한 음의 상관관계가 있음을 확인할 수 있다. 따라서 2014년 지방선거의 경우 응답률이 낮은 전화조사의 경우 동시에 가중치 비가 커지는 경향을 보이고, 이 경우 양의 편향(즉, 새누리당을 실제보다 과대추정하는 경향)이 발생하게 되며 결과적으로 예측오차가 커지는 원인이 된다는 것을 확인할 수 있다.

IV. 선형모형을 이용한 예측오차 발생 요인 분석

일반적으로 앞 절에서 다룬 조사방법이나 조사기간, 응답률, 가중치 비 등이 전화조사에서 예측오차나 편향 발생에 영향을 주는 요인으로 알려져 있

다. 따라서 예측 정확성을 나타내는 A_i 를 반응변수로 하고 연속형 변수인 응답률과 가중치비, 범주형 변수인 조사방법(ARS조사/전화면접조사), 지역(15개 시·도), 조사기간(하루/이틀 이상), 조사일시(5.26 이전/이후), 패널사용 여부 등 7개 설명변수를 사용한 일반선형모형을 이용해 조사의 정확성 척도 A_i 에 어떤 요인들이 많은 영향을 주는지 분석하였고, 이를 통해 전화조사에서 예측오차 또는 편향에 영향을 주는 주요 요인이 무엇이고, 각 요인이 변함에 따라 어떤 패턴의 예측오차가 발생하게 되는지 분석해 보기로 한다.

SAS의 Proc GLM 프로시저를 이용하여 앞에서 설명한 대로 7개 설명변수를 사용한 선형모형을 적합한 결과는 <표 9>와 같다. 결정계수(R^2)는 0.7288이고 F -값이 17.33으로 나타나 제시된 선형모형이 매우 적합이 잘되었다는 것을 알 수 있다. 다시 말해 예측 정확성 A_i 는 조사방법, 지역, 조사기간 등 제시된 설명변수들에 의해 영향을 많이 받는다는 것을 확인할 수 있다.

<표 9> 일반선형모형 적합 결과

Source	<i>d.f.</i>	Sum of Squares	Mean Square	<i>F</i> Value	Pr > <i>F</i>	<i>R</i> -Square
Model	20	5.8767	0.2938	17.33	< .0001	
Error	129	2.1873	0.0170			0.7288
Corrected Total	149	8.0640				

Source	<i>d.f.</i>	Type III SS	Mean Square	<i>F</i> Value	Pr > <i>F</i>
응답률	1	0.0406	0.0406	2.39	0.1244
가중치 비	1	0.1119	0.1119	6.60	0.0114
지역	14	4.2507	0.3036	17.91	< .0001
조사방법	1	0.0637	0.0637	3.75	0.0549
조사일시	1	0.0607	0.0607	3.58	0.0608
조사기간	1	0.0642	0.0642	3.79	0.0538
패널사용 여부	1	0.5128	0.5128	30.24	< .0001

<표 9>를 보면 설명변수 중 유의수준 10%를 기준으로 하면 응답률만 제외하고, 나머지 6개 설명변수는 모두 통계적으로 유의하다. 응답률도 통상적인 유의수준(5%나 10%)에서는 유의하다고 볼 수 없지만 p -값이 0.1244로 어느 정도 예측 정확성 척도 A_i 에 영향을 주는 변수라고 볼 수 있다. 여기서는 A_i 를 하나의 관측값으로 간주하고 오차항이 동일한 분산을 갖는다고 가정하고 모형을 적합시켰지만, 좀 더 정교한 모형 설정을 위해서는 A_i 에 대한 식(2)에 의한 분산을 고려해 모형을 적합시키는 방법도 가능할 것이다.

한편 유의한 변수들이 실제로 A_i 에 어떤 영향을 주는지 살펴보기 위해서 제시된 선형모형에서 변수들에 대한 계수 추정값을 정리해 보면 <표 10>과 같다. 지역에 대한 계수 추정값은 전화조사 방식에 따른 예측오차의 변동을 살펴보는 데 별 의미가 없기 때문에 여기서는 지역변수에 대한 추정값은 제외하고 나머지 변수들에 대한 추정값만을 정리하였다.

<표 10>에 정리된 각 변수에 대한 계수 추정결과를 부호를 중심으로 살펴보면, 우선 가중치 비가 증가하는 경우 양의 편향을 증가시킨다. 다시 말해 2014년 지방선거 전화조사의 경우 할당추출 과정에서 특정 성·연령대에 따른 할당 표본크기를 크게 채우지 못해 할당 범주에 따른 가중치 변동이 심해지는 경우, 새누리당을 과대추정하는 양의 편향이 발생시키는 요인이 되었다고 보면 된다. 이미 3절에서 가중치 비에 대한 분석결과에서 이런 현상을 설명한 바 있다.

<표 10> 일반선형모형 변수별 계수 추정 결과

Source	Estimate	Standard Error	t Value	Pr > t
절편	-0.0730	0.0502	-1.45	0.1484
응답률	0.0031	0.0020	1.55	0.1244
가중치 비	0.0107	0.0042	2.57	0.0114
조사방법(ARS)	0.0834	0.0430	1.94	0.0549
조사일시(5.26 이후)	-0.0458	0.0242	-1.89	0.0608
조사기간(1 일)	0.0599	0.0308	1.95	0.0538
패널 사용	-0.1617	0.0294	-5.50	< .0001

또한 전화면접조사 대신 ARS조사를 하거나 이를 이상 조사하는 대신 하루만 조사하면 새누리당을 실제보다 유리하게 추정(양의 편향)하게 된다는 것을 알 수 있다. 반면 조사일시가 5.26일 이후에 실시된 경우 또는 패널을 사용하는 경우 새정연을 유리하게 추정(음의 편향)하는 경향이 발생한다는 것도 알 수 있다.

V. 결론

본 연구에서는 2014년 제6회 동시지방선거 광역단체장 선거에 대한 전화조사의 예측오차 또는 편향에 영향을 주는 요인이 무엇인지, 또한 각 요인들이 변함에 따라 어떤 정당에 유리한 전화조사 결과가 산출되는 경향이 있는지 살펴보았다. 이를 위해 여심위 홈페이지에 등록된 개별 전화조사 수행과정에 대한 구체적인 정보(조사방법, 응답률, 가중치, 조사기간, 조사일시 등의 파라미터)를 활용하였다.

150개 사전 전화여론조사자료에 대한 분석결과, 조사방법에 있어서 ARS조사에 비해서는 전화면접조사가 보다 정확한 조사결과를 산출해 주는 것으로 결론을 내릴 수 있었으며, 이번 2014년 지방선거에서는 ARS조사의 경우 새누리당을 과대추정하는 편향을 발생시키는 원인이 되었다는 것도 밝힐 수 있었다. 또한 조사기관별로 예측오차에 있어서 상당히 큰 차이가 있고 일부 조사회사의 경우 다른 조사회사에 비해 매우 정확도가 떨어지는 조사를 수행하고 있다는 것도 알 수 있었다. 특히 단편적인 사례이지만 동일한 조사회사에서도 실제 조사방법을 달리하면 조사의 정확성이 크게 달라질 수 있다는 것을 확인할 수 있었으며, 유·무선 전화면접조사가 다른 조사방식에 비해 정확성이 높다는 것도 볼 수 있었다. 이를 통해 전화조사의 정확성 제고를 위해서는 어떤 조사방법을 사용할 것인지에 대해 조사회사는 물론 조사 이용자도 다시 한 번 깊이 고민해 볼 필요가 있다는 시사점도 얻을 수 있었다.

또한 하루 동안 조사하는 것보다(재통화 여부에 대해서는 확인할 수 없지

만) 이를 이상 조사하는 전화조사 방식이 정확성 제고에 기여할 수 있다는 것을 알 수 있었고, 응답률이 떨어지거나 가중치 변동폭이 커지면 2014년 지방선거에서는 새누리당을 과대추정하는 편향을 발생시키는 요인이 되었다는 점도 확인할 수 있었다.

그동안 학계나 조사업계에서는 전화면접이나 ARS조사 등 조사방법, 과도한 가중치 적용, 조사기간, 응답률 등이 선거여론조사의 예측오차나 편향에 어떤 영향을 주는지에 대해 많은 관심을 가져왔지만, 이런 궁금증에 대한 해답을 제공할 만한 개별 전화조사에 대한 세부 자료가 공개되지 않아 계량적인 분석이 제대로 이루어지지 못하였다. 다행히 최근 선거법이 개정되어 선거여론조사와 관련된 세부적인 파라미터들이 여심위 홈페이지를 통해 공개되고 있기 때문에 향후 본 연구와 유사한, 여론조사 방법론의 장·단점을 밝히는 연구가 꾸준히 수행될 수 있을 것으로 보이며, 이런 연구들이 우리나라 전화조사 방법을 개선하는 계기가 될 수 있을 것으로 보인다.

본 연구에서 얻어진 결과들을 종합적으로 정리해 보면, ARS조사보다는 전화면접조사, 당일치기 조사보다는 이를 이상의 기간을 통한 조사, 할당된 표본크기를 채우지 못 해 과도한 가중치를 적용하는 것보다는 가중치 변동폭이 적어질 수 있도록 할당된 범주별 표본크기를 꼼꼼히 채우는 조사를 통해 보다 정확한 여론조사 결과를 얻을 수 있다는 것을 알 수 있었다. 결국 이런 바람직한 조사를 위해서는 상대적으로 비용이 많이 드는 조사를 수행할 필요가 있다는 것을 시사해 주고 있다. 따라서 우리나라 정치권이나 언론기관 등 여론조사 이용자들이 이제는 비용보다 조사의 정확성과 방법의 안정성을 고려한 슬기로운 선택을 해주기를 바란다. 또한 시장에서 저렴한 엽터리 조사와 고비용 고품질 전화조사를 구분하고, 많은 노력과 시간이 들어간 양질의 조사를 있는 그대로 인정해 주는 환경이 조성되는 데 본 연구가 조금이나마 기여할 수 있기를 기대해 본다.

참고문헌

- 강현철 · 한상태 · 김지연 · 정용찬 · 허명희. 2008. “RDD 전화조사와 주요결과.”
《조사연구》 9(1): 1-22.
- 김세용 · 허명희. 2009. “전화조사의 체계적 편향 -2007년 대통령선거 여론조사들
에 대한 메타분석-” 《응용통계연구》 22(2): 375-385.
- 김영원 · 곽은선. 2010. “2008 총선 출구조사의 총조사오차 분석.” 《조사연구》
11(3): 33-55.
- 김영원 · 최윤정. 2011. “출구조사의 체계적인 예측 편향에 대한 분석: 2010년 지
방선거 출구조사를 중심으로.” 《조사연구》 12(3): 25-48
- 박종희. 2013. “제18대 대선 여론조사에서 나타난 조사기관 편향.” 《조사연구》
14(1): 1-30.
- 이기재 · 김도연 · 이기홍 · 조성겸. 2014. “한국여론조사 정보공개 제도화에 대한
연구: 한국조사연구학회 여론조사정보공개지침의 제정에 부쳐.” 《조사연구
》 15(1): 123-148.
- 이준웅. 2004. “제17대 총선 예측조사의 문제: 조사 거절자와 응답불성실자 편향
을 중심으로.” 《언론정보연구》 41(1): 110-136.
- 정형철. 2007. “ARS 조사와 전화여론 조사의 비교 - 2007년 한나라당 대선후보
에 대한 지지율 사례를 중심으로-.” 《한국자료분석학회지》 9(4): 1737-
1756.
- 조성겸. 1997. “대통령선거 여론조사와 할당표집 방법의 문제점.” 《언론과 사회》
18(12): 29-53.
- 중앙선거여론조사공정심의위원회. 2014a. 《제6회 전국동시지방선거 선거여론조
사 가이드북》.
- 중앙선거여론조사공정심의위원회. 2014b. 《제6회 전국동시지방선거 선거여론조
사 심의백서》.
- 최필선 · 민인식. 2013. “18대 대통령 선거 여론조사의 기관별 정확성 측정 및 비
교.” 《조사연구》 14(2): 1-18.

- 허명희 · 김영원. 2008. “RDD 표본 대 전화번호부 표본: 2007년 대통령 선거 예측 사례.” 《조사연구》 9(3): 55-69.
- 허명희 · 황진모. 2006. “전화조사를 위한 시간균형할당표본추출.” 《조사연구》 7(2): 39-52.
- Martin, E.A., M.W. Traugott, and C. Kennedy. 2005. “A Review and Proposal for a New Measure of Poll Accuracy.” *Public Opinion Quarterly* 69(3): 342-369
- Mitofsky, Warren. 1998. “The Polls-Review: Was 1996 a Worse Year for Polls than 1948?” *Public Opinion Quarterly* 62: 230-249
- Mokrzycki, M., S. Keeper, and C. Kennedy. 2009. “Cell-phone-only Voters in the 2008 Exit Poll and Implications for Future Noncoverage Bias.” *Public Opinion Quarterly* 73(5): 845-865.
- Traugott, M.W. and W. Christopher. 2009. “The Dynamics of Poll Performance During the 2008 Presidential Nomination Contest.” *Public Opinion Quarterly* 73(5): 866-894.

<접수 2014/11/04, 게재확정 2014/11/17>