

연구노트

사전투표 득표율 예측에 대한 연구

The Study on the Advance Voting Forecasting

이찬복^{a)}

Chan Bok Lee

지난 2014년 6월 지방선거 때 처음 전국적으로 실시된 사전투표제도로 인해 전국 어디에서든 유권자가 선거일에 앞서 미리 투표하는 것이 가능해졌다. 일부 유권자들이 선거일에 앞서 미리 투표한다는 점에서, 이는 그간 선거 당일에 투표하는 유권자들만을 대상으로 한 출구조사로 선거예측을 해왔던 선거예측조사방식에 대한 큰 도전이었다.

본 연구는 2014년 6월 지방선거 개표결과 데이터를 사용하여 사전투표 결과를 분석하고, 사전투표 득표율을 예측하기 위한 조사방법으로 출구조사 및 전화조사방법론을 검토하고자 한다. 또한, 각 방법론에 기초한 예측 모형과 예측 시뮬레이션을 통해서 각 방법론의 특징과 적합성을 분석한다.

결론적으로 사전투표 득표율을 예측하는 방법 중에서는 현재 출구조사를 이용한 방법이 가장 안정된 예측치를 산출하고 있는 것으로 나타났으며, 그럼에도 일관되게 발견되는 여권 과대 편향에 대해서는 향후 보완적 연구가 필요한 것으로 나타났다. 더불어 사전투표 득표율을 예측하는 대안적인 방법론에 대한 향후 연구 또한 필요한 것으로 나타났다.

주제어 : 지방선거, 사전투표, 출구조사, 득표율 예측, 예측 오차

Advance voting was held nationwide in 2014 local election. It made it possible for the electorate to vote ahead of local election. It

b) TNS 사회조사본부장/이사 이찬복.

E-mail: chanbok.lee@tnsglobal.com

caused serious problems in election forecasting through exitpoll because exitpoll has been just conducted on the election day.

This study analyzes the vote count result of advance voting and examine alternative methodologies to forecast the result of advance voting through forecasting simulations.

In conclusion, the methodology using exitpoll data is more stable and accurate than any other options. But the conservative biased errors were discovered consistantly in the methodoly using exitpoll data and this would be investigated by additional studies.

Key words: local election, advance voting, exitpoll, vote count forecasting, forecasting error

I. 들어가며: 사전투표제도의 확대와 선거예측

지난 2013년 4.24 재보궐선거에서 사전투표제도가 처음 도입되어 시범적으로 실시된 후 2014년 6.4 지방선거에서 사전투표제도가 전국 단위로 확대되어 실시되었다. 사전투표제도의 확대 실시는 출구조사를 통해 지방선거결과를 예측하여 보도하는 지방선거 방송사 공동예측조사위원회(Korea Election Pool: KEP)에 어려운 과제를 안겨주었다.

사전투표제도는 유권자가 별도의 신고없이 사전투표 기간(선거일 전 5일부터 2일간) 동안 전국의 읍/면/동에 설치된 어느 사전투표소에서나 투표할 수 있는 제도로 기존 부재자투표의 편의성을 높이기 위해 수정·도입된 것이다. 기존 부재자투표는 사전에 부재자신고가 필요했으며 투표소는 각 시/군/구마다 설치됐었다. 반면 사전투표제는 사전신고를 하지 않아도 되고, 투표소는 관할구역 안의 읍/면/동마다 설치되었다.

시범적으로 실시된 지난 2013년 노원병 재보궐선거에서 유권자의 사전투표율(전체 유권자 중 사전투표에 참여한 유권자의 비율)은 8.4%로 나타났다. 당시 노원병 재보궐선거 투표율이 29.5%였으므로 당시 전체 투표자 중에서 사전투표자가 차지하는 비중은 28.5%나 되었다. 만일 지방선거 투표율이 50%라고 가정하고, 노원병 재보궐선거 때와 동일하게 유권자의 8.4%가 사전투표에 참여한다면 사전투표자들이 전체 투표자에서 차지하는 비중이 16.8%에 이를 것으로 추정할 수 있다. 특히, 지방선거는 유권자의 관심도가 높은 전국적 선거이고 사전투표 또한 전국적으로 실시되기 때문에, 사전투표에 대한 선관위와 각 당의 홍보가 강화되고 있었던 점을 고려할 때 사전투표율이 10%를 상회할 것으로 전망되고 있었다. 10%의 사전투표율은 전체 투표율 50%를 가정할 경우에 전체 투표자의 1/5을 넘는 비중을 차지하여, 선거 승패에 중요한 영향을 미칠 것으로 전문가들은 전망하고 있었다.

출구조사를 통한 선거예측을 준비하고 있던 KEP는 우선적으로 사전투표자에 대한 출구조사가 가능한지를 검토한 결과, 사전투표자 출구조사에는 몇 가지 한계가 있다는 점을 발견하였다. 첫째, 사전투표는 본인의 주민등록거주지가 아닐지라도 전국 어디서나 가까운 읍/면/동 내에 설치된 3,506개 사전투표소에서 자유롭게 투표할 수 있다. 따라서 서울시장 선거의 사전투표에 참여하는 유권자는 극단적으로 전국의 3,506개의 사전투표소에서 서울시장 선거에 참여할 수 있기 때문에, 사전투표자에 대한 출구조사를 제대로 진행하려면 전국의 투표소에서 서울 주민등록 유권자를 추출해서 조사를 진행해야 하는데, 이는 현실적으로 불가능하다.

둘째, 공직선거법상 선거 당일에 출구조사를 진행하는 것은 가능하지만, 선거일 4~5일 전에 사전투표소에서 투표를 마치고 나온 유권자에게 출구조사를 진행하는 것은 불가능하다. 공직선거법 167조 2항에 따르면, '선거인은 투표한 후보자의 성명이나 정당명을 누구에게도 또한 어떠한 경우에도 진술할 의무가 없으며, 누구든지 선거일의 투표마감시각까지 이를 질문하거나 그 진술을 요구할 수 없다. 다만, 텔레비전방송국·라디오방송국·『신문 등의 진흥에 관한 법률』 제2조 제1호 가목 및 나목에 따른 일간신문사가 선거의 결과를 예상하기 위하여 선거일에 투표소로부터 50미터 밖에서 투표의 비밀이 침해되

지 않는 방법으로 질문하는 경우에는 그러하지 아니하며 이 경우 투표마감시까지 그 경위와 결과를 공표할 수 없다.’

한편, 대안적인 방법으로 사전투표가 종료된 직후 전화조사를 통해 사전투표자 여부와 사전투표 득표율에 대해 예측하는 방법이 있는데, 이에 대해서는 다음의 단점들이 제기될 수 있다. 첫째, 투표참여 여부는 ‘사회적으로 바람직하게 보이려는 편향(social desirability bias)’이 있기 때문에 전화조사에서 ‘사전투표에 참여했다’는 응답자가 실제 사전투표자와 다를 수 있다. 미국 선거조사 경험에 따르면, 투표참여 여부를 질문한 결과, 투표하지 않고서도 투표를 했다고 허위 응답한 사람의 비율이 20%~30%에 이르는 것으로 나타났다. 투표는 시민의 의무이기 때문에 바람직한 것이라고 생각하여, ‘사회적으로 바람직하게 보이려는 편향’ 요인이 작용해 과대응답을 촉진시키는 경향이 있을 수 있다는 것이다(오인환 1992).

둘째, 실제 사전투표율이 만약 10%라면, 2,000명을 전화로 조사할 경우 약 200여명이 사전투표자일 것으로 기대할 수 있는데, 이를 근거로 한 예측치의 허용오차는 95% 신뢰수준에서 $\pm 7.0\%$ 로 매우 클 수밖에 없을 것이다.

사전투표자 예측치 산출을 위한 여러 가지 대안적인 방법을 검토한 결과, 최종적으로 KEP는 선거 당일 출구조사의 성/연령별 후보 득표율에 선관위가 발표한 사전투표자의 성/연령별 구성비를 반영하여 사전투표자 예측치를 산출하기로 최종 결정하였다. 이러한 결정은 사전투표에 참여한 사전투표자의 성/연령별 후보 지지율이 선거 당일 참여한 투표자의 성/연령별 지지율과 큰 차이를 보이지 않을 것이라는 가정과 함께 이 방법이 현실적으로 선택 가능한 대안들 중에서 가장 안정적이라는 판단에 기초하고 있다. 특히 출구조사 데이터는 투표 직후에 투표행동에 대해 응답하는 행동 데이터(behavior data)이기 때문에, 완벽히 통제하기 어려운 비표본오차나 편향이 개입되는 전화조사와 비교해서 일반적으로 신뢰도가 높다고 볼 수 있다(김정훈 2003).

본 연구는 제2장에서 KEP가 적용했던 사전투표자의 후보 득표율 예측방법이 적절했는지를 살펴보고, 제3장에서는 사전에 실시되었던 전화조사 결과를 분석하여, 전화조사를 통한 사전투표자의 후보 득표율 예측방법론을 검토하고 각각의 사전투표 득표율 예측방법론을 비교·분석해 보았다. 마지막으로 제4장

과 제5장에서는 사전투표자의 성향을 분석해 보고, 향후 개선방향과 대안을 모색해 보고자 한다.

II. 2014년 6.4 지방선거 사전투표자 후보 득표율 예측

KEP의 2014년 6.4 지방선거 사전투표자 후보 득표율 예측방법은 선거 당일 투표한 투표자의 성/연령별 투표성향이 사전투표에 참여한 투표자의 성/연령별 투표성향과 유사할 것이라는 가정에 기초하고 있다. 다시 말하자면, 선거 당일 투표에 참여한 남자 20대 투표자의 후보 지지율은 선거 5일~4일 전에 실시한 사전투표에 참여한 남자 20대 투표자의 후보 지지율과 유사할 것이라는 가정이다.

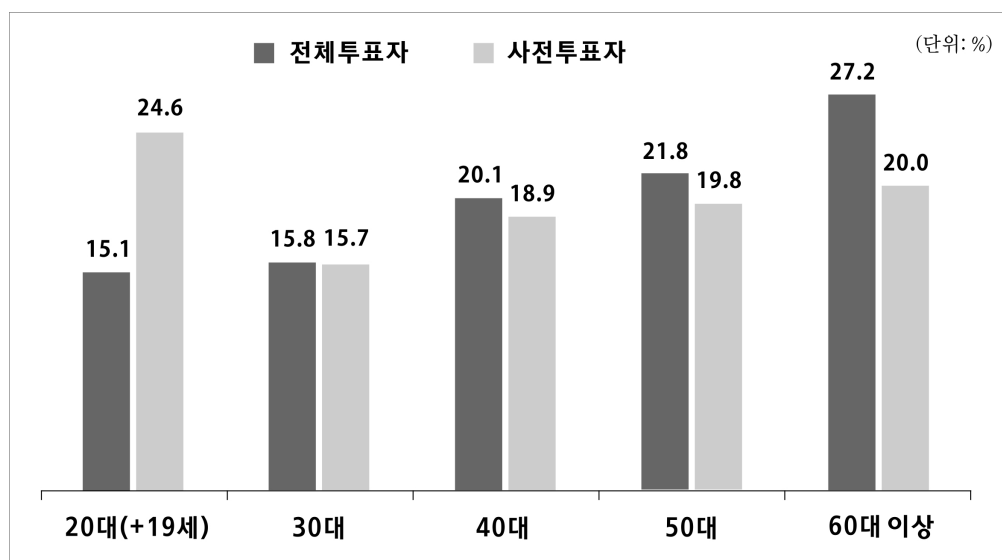
이러한 가정에 기초해 선관위가 사전투표 마감 후 발표한 사전투표자의 성/연령별 투표참여율 자료를 출구조사에서 분석된 성/연령별 후보 득표율에 반영하여 목표한 사전투표자의 후보별 득표율 예측치를 산출하였다. 이렇게 산출된 사전투표자 예측치와 선거 당일 출구조사에서 산출된 예측치가 결합되어 출구조사 최종 예측치가 산출되었다.

이처럼 후보 득표율 최종 예측치를 산출하기 위해 필요한 사전투표 득표율 예측치 산출이 가능했던 것은 중앙선관위가 사전투표 직후 사전투표자의 성/연령별 투표율과 투표자 구성비 자료를 공개하기로 결정했기 때문이었다. 먼저 실제 선관위가 발표한 2014년 6.4 지방선거 사전투표 현황을 살펴보았다. 투표결과, 17개 시/도별 사전투표율과 전체 투표자 중 사전투표자의 비율은 <표 1>과 같다.

<표 1>에서 보는 바와 같이 중앙선관위가 발표한 2014년 지방선거 투표결과를 살펴보면, 사전투표율이 전국 평균 11.5%로 나타났다. 전국 평균투표율이 56.8%인 점을 감안하면, 전체 투표자 중에서 사전투표자가 차지하는 비율은 전국 평균 20.2%인 셈이다. 이는 전문가들이 예상한 대로 2013년 두 번의 재보궐 선거에서 나타난 사전투표율보다 높은 수치였고, 따라서 사전투표자는 해당 선거구의 선거결과에 중요한 영향을 미칠 것으로 전망되었다.

<표 1> 2014년 6.4 지방선거 17개 시/도 사전투표율과 전체 투표자 중 사전투표자 비율
(단위: %)

지역	사전투표율	투표자 중 사전투표자 비율	지역	사전투표율	투표자 중 사전투표자 비율
서울	11.1	18.9	강원	14.2	22.8
부산	8.9	15.8	충북	13.3	22.6
대구	8.0	15.3	충남	11.9	21.4
인천	11.3	21.0	전북	16.1	26.5
광주	13.3	23.5	전남	18.1	27.3
대전	11.2	20.2	경북	13.1	21.8
울산	10.7	20.2	경남	11.9	19.8
세종	15.1	24.2	제주	11.1	17.6
경기	10.3	19.1	-	-	-



<그림 1> 2014년 6.4 지방선거의 전체 투표자와 사전투표자의 연령 구성비 비교

한편 지방선거는 선거 당일이 휴일이었기 때문에, 휴일 계획 때문에 투표에 참여하기 어려운 젊은 층들이 사전투표에 많이 참여할 것으로 예상되었다. 당시 '6월5일 휴가내면 5일간 연휴... 지방선거 투표율 비상'이라는 동아일보

2014년 1월 13일자 기사에 따르면, 지방선거가 수요일이고 이틀 후인 금요일(6월 6일)은 현충일이므로 목요일만 휴가를 내면 수요일부터 일요일까지 5일 연휴가 된다는 이유로, 많은 젊은이들이 금요일에 휴가를 내고 선거일부터 휴가를 떠날 것이라는 전망이 제기되었다. 이러한 젊은이들이 미리 사전투표에 많이 참여하여 젊은이들의 사전투표율이 높아질 것이라고 보았는데, 실제로 <그림 1>에서 보이는 것처럼 선거결과 투표자 구성비를 분석해 보면, 사전투표에 젊은 층이 상대적으로 많이 참여한 것으로 나타났다.

〈표 2〉 사전투표의 개표결과와 선거 당일 개표결과 비교 - 새누리당 후보 기준 (단위: %)

지역	사전투표 새누리당 후보 개표결과	당일투표 새누리당 후보 개표결과	차이 (사전-당일)
서울	39.7	43.8	-4.1
부산	43.9	51.9	-8.1
대구	49.1	57.2	-8.1
인천	48.0	50.5	-2.5
광주*	52.9	59.4	-6.5
대전	44.1	47.4	-3.4
울산	58.6	67.2	-8.6
세종	38.8	43.3	-4.5
경기	47.2	51.2	-4.0
강원	42.2	49.9	-7.7
충북	42.7	49.2	-6.5
충남	39.7	45.1	-5.5
전북*	71.4	68.5	2.9
전남*	78.8	77.7	1.1
경북	74.4	78.7	-4.3
경남	55.0	59.8	-4.9
제주	58.0	60.4	-2.3

※ 단, 광주시장, 전북도지사, 전남도지사 선거의 경우는 새정치민주연합 후보 득표율.

일반적으로 젊은 층에서 야권 지지가 높기 때문에 선거 당일 투표결과와 비교해 사전투표 결과는 야권후보에게 유리할 것으로 예상되었다. 결국 사전투표로 인해 젊은 층의 투표율이 얼마나 높아질 것인가와 이것이 선거결과에 어떤 영향을 미칠 것인가가 사전투표와 관련한 최대 이슈였다. 실제 개표결과를 살펴보면, 사전투표 결과는 야권에 유리한 결과를 보여주었다. <표 2>는 새누리당 후보의 사전투표 득표율과 선거 당일 득표율을 비교한 것이다.

전문가들의 예상대로 선거 당일 결과와 대비하여 사전투표 결과의 새누리당 후보 득표율이 낮게 나타난 것을 발견할 수 있다. 이는 야권 지지성향이 높은 유권자가 사전투표에 더 많이 참여한 것으로 해석할 수 있다. 이는 또한 젊은 층이 선거 당일과 비교해 사전투표에 높은 비중으로 참여한 데 기인한 결과로 볼 수 있으며, KEP는 이와 같이 젊은 층의 높은 투표자 구성비를 반영하여, 사전투표자 후보 득표율을 유사하게 예측할 수 있을 것으로 기대하고 있었다.

선거 당일 출구조사 결과를 활용한 사전투표 후보 득표율 예측결과를 살펴보기에 앞서 먼저 2014년 6.4 지방선거 출구조사 설계를 간략히 요약하면 <표 3>과 같다.

<표 3> 2014년 6.4 지방선거 출구조사 설계

조사 대상	전국에서 추출된 654 개 투표소에서 투표를 마치고 나오는 투표자
조사 시점	6 월 4 일 06:00~18:00
투표소 추출	층화 후 정렬계통추출
응답자 추출	출구를 나오는 투표자 기준으로 매 5 번째 투표자를 등간격으로 체계적 추출
표본 크기	134,366 명 (17 개 시도별로 3,642 명~14,444 명)
조사 기관	TNS-MBMR-R&R

KEP의 사전투표자 후보 득표율 예측방법론을 적용한 각 지역별 새누리당 후보의 사전투표 득표율 예측치와 실제 개표결과를 비교하면 <표 4>와 같다.

예측치 비교의 준거로 새누리당 후보 득표율을 사용한 것은 지방선거의 정당 경쟁구도에서 새누리당이 보수성향을 대표하는 유일한 정당으로 유권자의 투표행태를 분석하는 데 있어서 좀 더 단순 명료한 분석이 가능하기 때문이다.

<표 4> 사전투표의 득표율 예측치와 실제 개표결과 득표율 비교-새누리당 후보 기준
(단위: %)

지역	사전투표자의 새누리당 후보 예측치	사전투표자의 새누리당 후보 득표율 개표결과	오차 (예측치-개표결과)
서울	44.6	39.7	+4.9
부산	46.5	43.9	+2.6
대구	50.1	49.1	+1.0
인천	48.5	48.0	+0.5
광주*	58.9	52.9	+6.0
대전	45.7	44.1	+1.6
울산	62.6	58.6	+4.0
세종	41.5	38.8	+2.7
경기	47.4	47.2	+0.2
강원	42.7	42.2	+0.5
충북	44.7	42.7	+2.0
충남	45.1	39.7	+5.4
전북*	73.2	71.4	+1.8
전남*	77.6	78.8	-1.2
경북	73.5	74.4	-0.9
경남	55.1	55.0	+0.1
제주	58.1	58.0	+0.1

※ 단, 광주시장, 전북도지사, 전남도지사 선거의 경우는 새정치민주연합 후보 득표율

사전투표자의 새누리당 후보 득표율 예측결과와 실제 개표결과 간 오차를 살펴보면(<표 4> 참고), 새누리당 후보 득표율 예측치의 전체 오차평균은 2.1%p

였으며, 광주(새정치민주연합 후보 득표율)가 6.0%p로 가장 큰 오차를, 경남과 제주가 0.1%p로 가장 작은 오차를 기록하여 비교적 정확한 예측을 한 것으로 나타났다. 한편 오차의 방향성을 살펴본 결과, 전체적으로 KEP의 예측치가 새누리당 후보의 득표율을 소폭 과대예측한 것으로 분석되었다.

사전투표자의 젊은 층 비율이 높은 점을 반영해 예측했음에도 오차의 방향성 측면에서 새누리당 후보에 대한 과대예측 경향이, 소폭이지만 전체 선거구에서 일관되게 발견되었다. 결과적으로 같은 연령대의 유권자일지라도 선거 당일보다 사전투표일에 야권 지지성향 투표자가 더 많았다고 추정할 수 있다. 즉, 동일한 남자 20대이더라도 야권 지지성향이 있는 남자 20대 투표자들이 선거 당일보다 사전투표일에 투표에 더 많이 참여했다고 가정할 수 있다.

현실적 경험적 한계 속에서 처음 도입해 본 방법론으로 오차 평균이 2.1%p였다는 점에서 KEP의 사전투표자 후보 득표율 예측방법론에 일정 부분 성과가 있었다고 생각하지만, 동일한 성/연령대라 하더라도 사전투표에서 투표한 사람이 선거 당일 투표한 사람보다 상대적으로 야권 지지성향이 높다는 점에서 선거 당일의 성/연령별 후보 득표율을 그대로 활용할 때, 소폭의 오차가 발생함을 발견하였다.

제한된 정보와 현실적 장애 속에서 KEP가 적용한 사전투표율 예측방법론이 최선이라는 평가를 내리는 것은 성급한 일일 것이다. 선거 종료 후 중앙선거관리위원회에서 사전투표자의 후보 득표율 개표결과를 공개했으므로, KEP가 도입했던 사전투표 예측방법론과 앞서 검토했던 사전투표 직후 실시된 전화조사를 활용한 예측방법론도 본 연구에서 비교·검토해 보고자 한다.

Ⅲ. 사전투표자 후보 득표율 예측방법론 비교

사전투표자의 후보 득표율을 예측하는 대안적 방법론으로 전화조사를 활용한 방법론을 검토해 볼 필요가 있다. 전화조사가 원칙대로 실시되는 이상적인 상황을 가정한다면, 사전투표 종료 후에 사전투표를 마친 투표자를 전화로 찾아서 조사하는 것이 이론적으로는 가능하다.

그러나 서두에서 언급한 대로 전화조사로 사전투표자를 찾아서 조사를 하는 데는 몇 가지 난점이 따른다. 앞서 언급한 대로 그 첫 번째 난점이 ‘사회적으로 바람직하게 보이려는 편향’ 때문에 전화조사에서 사전투표했다는 응답에 대한 신뢰도가 떨어진다는 것이다(오인환 1992).

먼저 사전투표 직후 실시된 전화조사 설계를 살펴보면, 조사 표본규모는 선거구의 경합도를 반영하여 1,000명~3,000명이었으며, 유무선 병행 RDD 조사로 수행되었다. TNS, MBMR, R&R 3개사가 공동으로 실시하였으며 구체적인 설계는 <표 5>와 같다.

<표 5> 2014년 6.4 지방선거 전화조사 설계

조사 대상	17 개 시/도 거주 19 세 이상 유권자
표본 프레임	사전에 구축된 유/무선 RDD 전화번호 리스트
표본 추출	지역/성/연령 할당 추출
표본 크기	경합도에 따라 각 시/도별로 1,000 명~3,000 명
조사 일시	6 월 1 일~3 일 (D-3~D-1)
조사 기관	TNS·MBMR·R&R

<표 6>은 투표참여 관련 전화조사 결과를 분석한 것이다. 그간 일반적인 선거여론조사에서 투표참여를 묻는 질문에 ‘사회적으로 바람직하게 보이려는 편향’이 반영되어 ‘투표하겠다’는 응답이 실제 투표율보다 과대 응답되는 것(오인환 1992)과 달리, 이번 조사에서 사전투표참여 여부에 대한 응답은 실제 선관위가 발표한 사전투표 결과와 큰 오차를 보이지 않았다.

당시 설문조사 시점이 선거일 이전이고, 투표참여 관련 응답항목이 ‘① 꼭 투표할 것이다 ② 웬만하면 투표할 생각이다 ③ 별로 투표할 생각이 없다 ④ 전혀 투표할 생각이 없다 ⑤ 이미 사전투표를 했다’였기 때문에, ‘① 꼭 투표할 것이다 ② 웬만하면 투표할 생각이다’ 항목에 이미 ‘사회적으로 바람직하게 보이려는 편향’이 반영되어 상대적으로 ‘⑤ 이미 사전투표를 했다’는 항목에 이

러한 편향이 최소화되었다고 가정할 수도 있다.

<표 6> 2014년 6.4 지방선거일 3일~1일 전에 실시된 전화조사 사전투표참여응답 분석
(단위: %)

지역	전화조사 표본크기	사전투표 응답표본	표본오차 (95%신뢰수준)	전화조사 사전투표율	선관위 발표 사전투표율	차이
서울	3,000	368	±3.4	12.3	11.1	+1.2
부산	2,000	203	±4.2	10.2	8.9	+1.3
인천	2,000	236	±4.1	11.8	11.3	+0.5
울산	1,000	117	±5.8	11.7	10.7	+1.0
세종	1,000	152	±5.7	15.2	15.1	+0.1
경기	3,000	329	±3.4	11.0	10.3	+0.7
강원	2,000	282	±4.1	14.1	14.2	-0.1
충북	2,000	275	±4.1	13.8	13.3	+0.5
경북	1,000	125	±5.8	12.5	13.1	-0.6
제주	1,000	84	±5.9	8.4	11.1	-2.7

※ 전화조사 분석을 위한 세부 데이터가 확보된 10개 지역 대상 분석

이번 전화조사 결과만 볼 때, 사전투표율 예측치에 '사회적으로 바람직하게 보이려는 편향'이 반영되어 응답이 과대하게 나타났다는 판단을 내리기는 어려웠다. 다만, 두 수치가 유사하다 하더라도 전화에서 '이미 사전투표했다'고 응답한 응답자가 '실제로 사전투표일에 사전투표했다' 라고 단정할 수는 없을 것이다.

전화조사 사전투표율 예측치에 외견상 일반적으로 나타나는 '사회적으로 바람직하게 보이려는 편향'이 명확하게 드러나지 않았다고 하더라도, 두 번째 문제는 남는다. 그것은 사전투표자의 득표율 예측에 사용할 사전투표자 표본 규모가 너무 작다는 것이다. 서울의 경우, 전체 유권자 3,000명을 조사하여, 이 중 사전투표자 368명만이 사전투표자 후보 득표율 예측치 분석에 사용될 수 있다. 전체 조사표본규모가 이보다 줄어들 경우, 사전투표자 분석에 사용할 수 있는 표본 수는 100~200표본 정도 규모가 된다.

제한된 조사기간과 예산이라는 현실적인 여건들을 고려할 때, 전화조사 표본규모를 크게 늘리는 것은 어려우므로 아무리 많이 표본규모를 늘린다고 해도 사전투표자 후보 득표율 분석에 사용할 수 있는 분석 표본규모는 300명 내외가 될 것이다. 표본크기가 300이면 95% 신뢰수준에서 허용오차가 $\pm 5.7\%p$ 에 이르기 때문에, 박빙의 선거판세가 벌어지는 우리 선거에서 의미 있는 표본규모라고 보기 힘들다.

세 번째로, 지지후보를 밝히지 않는 사람들을 어떻게 처리할 것인가 또한 매우 중요하다. 일반적으로 선거가 며칠 남지 않은 시점까지도 전화조사에서 지지후보 무응답이 10%~20% 발견되는데, 이들 무응답층의 향배가 선거 승패에 중요한 영향을 미칠 수 있다.

전화조사를 활용한 사전투표자 후보 득표율 예측방법론이 안고 있는 이러한 한계를 인정한 채, 사전투표자 후보 득표율 예측에 있어서 전화조사방법론이 산출하는 예측치의 경향성과 특징을 살펴본다는 측면에서 전화조사를 활용한 사전투표자 예측을 진행해 보고자 한다.

분석에 사용될 데이터는 <표 5> 전화조사 조사설계에서 언급한 것과 동일한 전화조사 데이터로 전화조사 표본규모는 <표 6>과 같다. 전화조사를 활용한 사전투표자 후보 득표율 예측방법론은 분석자의 시각에 따라 다양할 수 있지만, 여기에서는 다음 3가지 방법론으로 한정하여 분석을 진행한다(<표 7>참고).

<표 7>의 방법1은 가장 단순한 방법으로 사전투표 종료 직후 실시된 전화조사에서 ‘이미 사전투표를 했다’는 응답자의 후보 득표율을 분석하는 것이다. 방법2는 방법1의 ‘이미 사전투표를 했다’는 응답자의 후보 득표율을 그대로 사용하는 것이 아니라, 이를 성/연령별 득표율로 세분화하여 여기에 선관위에서 발표한 사전투표자의 성/연령별 구성비를 반영하여 최종 후보 득표율을 산출하는 것이다. 방법3은 방법2와 절차는 유사하나, 방법1과 방법2의 표본규모가 작아서 데이터의 안정성이 떨어질 위험성이 있다고 본다. 그래서 전화조사 설문에서 ‘이미 사전투표를 했다’는 사람과 함께, 선거일에 ‘꼭 투표할 예정’이라고 답한 응답자까지 모두 포함하여 이들을 ‘투표확실층’으로 분류하고 그들의 성/연령별 후보 득표율에다가 선관위가 발표한 사전투표자의 성/연령별 구성

비를 반영하여 최종 후보 득표율을 산출하는 것이다.

<표 7> 사전 전화조사를 활용한 사전투표자의 후보 득표율 예측방법론

	방법론 설명	지지후보 무응답 분석방법	비고
방법 1	설문에서 ‘이미 사전투표를 했다’는 응답자들의 후보 득표율 분석	무응답 제외 100% 재환산	전화조사 결과만 활용
방법 2	설문에서 ‘이미 사전투표를 했다’는 응답자들의 성/연령별 후보 득표율에 선관위에서 발표한 사전투표자의 성/연령별 구성비를 반영	무응답 제외 100% 재환산	전화조사 결과와 사전투표자 성/연령별 구성비 자료 활용
방법 3	설문에서 ‘투표확실층’으로 분류된 응답자들의 후보 득표율에 선관위에서 발표한 사전투표자의 성/연령 구성비를 반영	무응답 제외 100% 재환산	전화조사 결과와 사전투표자 성/연령별 구성비 자료 활용

한편 ‘어느 후보에게 투표할 것인지’를 묻는 질문에서 불가피하게 나타나는 지지후보 무응답에 대한 분석도 ‘무응답을 제외한 100% 재환산 방법(무응답을 제외하고 후보 지지율의 합을 100%로 재계산하는 방법으로, 이는 지지하는 후보를 밝히지 않은 무응답층도 지지후보를 밝힌 층과 동일한 후보 지지율을 보일 것이라는 가정에 기초하고 있다.)’, 열세후보 회피효과를 보정하기 위해 ‘열세인 후보에게 좀더 배분하는 방법(이는 자신이 지지하는 후보가 현재 득표율에서 열세에 처할 경우에 유권자들은 자신들이 지지하는 후보를 밝히지 않는 경향이 있다는 가정에 기초하고 있다. 특히, 열세후보 회피효과를 보정하는 방법은 어느 정도로 보정할 것인가에 따라 여러 가지 방법이 있을 수 있다)’, ‘판별분석방법’ 등과 같이 다양한 방법론을 상정할 수 있다.

그간 학계에서는 전화조사와 출구조사의 지지후보 무응답 분석방법에 대한 많은 검토가 있어 왔다(김영원·김지현 2007; 조성겸 2009; 김영원·황다솜 2014). 특정한 하나의 방법론이 모든 선거에서 다른 방법보다 우세하다고 이야기할 수는 없다. 선거의 판세, 구도, 이슈에 따라, 그리고 그것의 변동에 따

라 선거마다 적합한 방법이 달라지곤 한다. 본 연구에서는 가장 일반적으로 사용되는 방법인 ‘무응답을 제외한 100% 재환산 방법’을 사용하여 분석하였다(장덕현·조성겸 2015; 최필선·민인식 2013; 홍내리·허명희 2001).

<표 8>은 이 3가지 분석방법을 사용하여 시뮬레이션한 전화조사의 사전투표자 후보 득표율 예측치를 비교한 것이다. 앞서와 마찬가지로 새누리당 후보의 득표율을 비교하였다.

<표 8> 사전전화조사를 활용한 사전투표 득표율 예측방법 비교-새누리당 후보 득표율 기준
(단위:%)

지역	새누리당 후보 개표결과	전화조사 방법1	전화조사 방법2	전화조사 방법3	방법1 오차	방법2 오차	방법3 오차
서울	39.7	31.7	34.8	37.4	-8.0	-4.9	-2.3
부산	43.9	25.3	32.3	43.0	-18.6	-11.6	-0.9
인천	48.0	40.6	43.8	42.8	-7.4	-4.2	-5.2
울산	58.6	62.9	64.5	64.0	4.3	5.9	5.4
세종	38.8	46.7	44.7	46.6	7.9	5.9	7.8
경기	47.2	39.2	43.2	47.6	-8.0	-4.0	0.4
강원	42.2	32.1	37.3	40.7	-10.1	-4.9	-1.5
충북	42.7	36.5	38.9	42.3	-6.2	-3.8	-0.4
충남	39.7	31.0	36.0	38.2	-8.7	-3.7	-1.5
경북	74.4	85.9	87.8	82.7	11.5	13.4	8.3

※ 전화조사 분석을 위한 세부 데이터가 확보된 10개 지역 대상 분석

분석의 결과를 보면, 우려했던 것처럼 ‘이미 사전투표를 했다’는 응답자의 후보 득표율을 분석한 방법1과 방법2의 경우는 분석에 사용된 표본크기가 너무 작아 데이터의 안정성이 떨어져, 추정치와 실제값 간의 오차가 큰 것으로 나타났다. 다만, 방법3의 경우에 분석표본의 크기를 늘린 ‘투표확실층’을 분석에 사용했으므로 방법1과 방법2와 비교해 상대적으로 오차가 작았다. 다만 지역에 따라 오차의 변동폭이 커서 데이터의 불안정성은 여전한 것으로 나타났다.

앞선 분석결과들을 종합하여, 출구조사와 전화조사를 활용한 사전투표자 후보 득표율 예측방법론의 오차를 정리해 보면 <표 9>와 같다.

<표 9> 사전투표자 득표율 예측방법론의 오차 비교 - 새누리당 후보 득표율 기준 (단위: %)

지역	오차 평균	오차의 표준편차	오차의 최댓값	오차의 최솟값
출구조사 방법	2.38	1.91	5.43	0.20
전화조사－방법 1	9.06	3.87	18.56	4.32
전화조사－방법 2	6.22	3.43	13.45	3.67
전화조사－방법 3	3.36	3.06	8.35	0.37

결론적으로, 출구조사를 활용한 사전투표자의 후보 득표율 예측방법이 오차의 크기나 표준편차의 안정성 측면에서 전화조사방법보다 나은 것으로 보인다. 한편, 오차가 컸던 전화조사방법 중에서는 ‘투표확실층’을 분석한 예측치가 ‘사전투표층’을 분석한 예측치보다 상대적으로 오차가 작은 것으로 나타났다.

그러나 전화조사 투표확실층을 대상으로 한 사전투표 후보득표율 예측방법론은 ‘사회적으로 바람직하게 보이려는 편향’ 때문에 투표확실층이 실제 투표자 규모보다 과대추정되어 나타난다는 점에서 문제가 있으며, 또한 투표참여 예상자의 성/연령별 지지율을 사용한다는 점에서, 투표 예상자가 아닌 투표자의 성/연령별 지지율 자료를 사용하는 기존 KEP의 방법론보다 신뢰도가 떨어진다고 볼 수 있다.

IV. 사전투표자 성향 분석 및 향후 개선방향 모색

현실적 문제들과 조사결과의 안정성 등을 고려할 때, 현 시점에서 사전투표자의 후보 득표율을 예측하는 최선의 방법은 당일 출구조사 결과를 활용하

는 것이다. 하지만, 그럼에도 앞서 살펴본 것처럼, 출구조사를 이용한 방법이 기대고 있는 가정, 즉, 동일한 성/연령대에 속하는 투표자라면 사전투표일에 투표하는 사람들과 선거 당일에 투표하는 사람들의 투표행태에 큰 차이는 없을 것이라는 가정은 어디까지나 하나의 가설일 뿐이다.

이번 선거결과만 놓고 보면, 동일한 성/연령 그룹에 속하더라도, 사전투표일에 투표하는 사람들의 야권 지지성향이 좀 더 강하게 나타났다. 물론, 모든 연령층에서 사전투표에 참여한 유권자의 야권 지지성향이 더 강하다고 볼 수는 없을 것이다. 사전투표자의 구체적인 성/연령별 후보 득표율 분석이 가능하다면, 어떤 성/연령층에서 이러한 경향이 특히 강하게 나타나는지를 알 수 있을 것이지만, 사전투표자들의 성/연령별 후보 득표율 자료는 집계되지 않기 때문에 분석이 불가능하다.

다만, 젊은 연령층에서 이러한 경향이 좀 더 두드러지게 나타나지 않을까 하는 가정을 해볼 수 있다. 평일인 사전투표일에 투표하는 사람들은 주로 선거 당일에 투표가 어려운 사람들이라는 가정이 가능할 텐데, 이는 고령층보다는 20~30대 젊은 층일 가능성이 높다. 특히 선거를 포기하지 않고 평일에 시간을 내어 능동적으로 사전투표소를 찾는 수고를 마다하지 않는 이들 20~30대 사전투표자들은 선거 당일에 투표하는 젊은 층들보다 강한 투표의지를 보인다고 해석할 수 있다. 따라서 야권 지지성향이 높은 젊은 층들을 중심으로 사전투표에 많이 참여했을 것으로 가정해 볼 수 있다.

이 같은 경향은 사전투표자를 구분하여 분석하여도 발견된다. 사전투표는 유권자 본인이 거주하는 주소지 시/군/구에 설치된 사전투표소에서 투표하는 ‘관내 사전투표’와 거주하는 주소지 시/군/구가 아닌 곳에서 투표하는 ‘관외 사전투표’로 나뉜다. 참고로 지방선거의 관내 투표자와 관외 투표자 비율은 각각 56.1%와 43.9%였다.

관외 사전투표의 경우에는 평일 대부분의 시간에 주로 거주하는 주소지 시/군/구 밖에 머무르는 직장인이나 대학생들의 비중이 높을 것으로 가정할 수 있고, 관내 사전투표의 경우는 평일에도 거주하는 주소지에 주로 머물고 있는 고연령층의 비중이 높을 것으로 가정할 수 있다. 이를 반영하듯이, 사전투표자를 세분화해서 새누리당 후보 득표율을 분석해 보면 <표 10>에서 보는 것처럼

럼 관내 사전투표자의 새누리당 후보 득표율이 관외 사전투표자의 새누리당 후보 득표율보다 대체로 높다.

〈표 10〉 관외 사전투표와 관내 사전투표의 새누리당 후보 득표율 비교 (단위: %)

지역	관외 사전투표자 새누리당 득표율	관내 사전투표 새누리당 득표율
서울	38.7	40.4
부산	41.9	45.4
대구	47.9	50.2
인천	46.2	49.2
광주	5.2	3.3
대전	42.6	45.2
울산	60.0	57.6
세종	39.1	38.5
경기	46.3	47.9
강원	38.7	45.2
충북	38.7	45.5
충남	36.2	42.5
전북	18.1	19.5
전남	9.5	9.6
경북	69.8	77.7
경남	51.6	57.6
제주	55.7	60.9

출처: 중앙선거관리위원회, 2014. 《제6회 전국동시지방선거 읍면동별 개표결과》.

〈표 10〉에서 보는 것처럼 사전투표에서 야권 지지 투표행태는 대체로 평일에 능동적으로 관외에서 사전투표하는 20~30대 직장인과 대학생을 중심으로 나타나는 것으로 보여지며, 출구조사 당일 결과를 사용한 사전투표자 후보

득표율 예측치에서 소폭의 여권 과대예측 효과가 나타나는 것도 이와 같은 이유라고 가정해 볼 수 있다.

현행 사전선거 후보득표율 예측방법론에서 나타난 소폭의 여권 과대예측 경향이 매 선거 때마다 지속적으로 발견되는 현상인지, 지방선거의 특수성에 따른 일시적인 현상인지에 대해서는 단언하기 어렵다. 만일 이것이 지속적으로 나타나는 현상이라면, 현재 분석에 사용된 성/연령 이외에 직업 등과 같이 사전투표자의 투표성향에 영향을 미치는 변수는 어떤 것이 있으며, 이를 사전투표자 예측치 산출에 활용할 방법에 대한 검토도 필요할 것이다. 이 모든 것이 앞으로 진행될 선거예측조사에서 지속적으로 검토되고, 그것에 대한 연구가 진행되어야 할 것이다.

본 연구를 진행하는 도중, 4.29 재보선을 앞두고 선관위의 발표가 있었다. 지난 지방선거 사전투표 직후에 공개되어 KEP의 사전투표 예측을 가능하게 했던 사전투표자 투표율 관련 통계자료를 앞으로 선거 전에 공개하지 않겠다는 것이다. 통계에 각 지역별 투표자의 성별과 연령별 숫자가 담겨 있어 이를 바탕으로 표심 분석이 가능하기 때문에 이것이 본 선거에 영향을 줄 수 있다고 판단했다는 것이다. 앞으로는 사전투표가 끝난 직후에 공개하지 않고, 본투표가 끝난 후에 공개하겠다는 것이다.

결과적으로 다가올 선거예측조사에서는 지난 2014년 6.4 지방선거에서 진행했던 방식의 출구조사를 활용한 사전투표자 예측은 어렵게 되었다. 즉, 당해 선거의 사전투표자 성/연령별 실제 투표자 구성비 자료를 활용하는 것이 불가능해졌기 때문에, 대안적으로 우선 사전 전화조사에서 ‘사전투표했다’고 응답한 층의 성/연령별 구성비의 사용을 검토해 볼 수 있다. <표 11>은 지방선거 일 3일~1일 전에 실시된 사전 전화조사에서 ‘이미 사전투표했다’고 응답한 계층의 성/연령별 구성비 자료이다. 대표적으로 수도권인 서울과 경기, 행정도 지역인 강원도와 충청북도를 대상으로 분석해 보았다.

사전 전화조사에서 나타난 사전투표율 예측치가 실제 사전투표율과 유사했었기 때문에 성/연령별 구성비 자료의 유사성에 대한 기대감이 있었지만, 전화조사에서 예측된 사전투표층의 성/연령 구성비는 실제 선관위가 발표한 사전투표자 성/연령 구성비와는 다소 차이를 보였다.

<표11> 사전 전화조사 사전투표층 성/연령 구성비와 실제 사전투표자 성/연령 구성비 비교 (단위: %)

	전화조사에 나타난 사전투표층 성/연령별 구성비				선관위 발표 사전투표자 성/연령별 구성비			
	서울	경기	강원	충북	서울	경기	강원	충북
표본크기	368 명	329 명	282 명	275 명	-	-	-	-
남 20 대	12.4	10.7	17.4	12.3	15.8	17.8	17.7	16.5
남 30 대	11.9	11.4	11.4	9.9	10.4	9.2	8.4	7.5
남 40 대	12.4	13.9	9.1	12.9	10.6	12.2	11.7	10.9
남 50 대	11.3	13.8	11.1	8.9	11.2	12.2	12.7	12.4
남 60 대 이상	8.0	7.1	4.4	7.9	12.2	10.3	9.9	11.5
여 20 대	9.3	10.3	9.6	11.0	7.4	7.1	8.2	7.9
여 30 대	11.4	7.1	11.8	8.7	8.9	7.8	6.2	6.0
여 40 대	8.3	13.0	12.2	7.3	7.4	8.4	8.0	7.9
여 50 대	6.8	6.2	5.6	10.5	7.6	7.4	8.3	8.3
여 60 대 이상	8.2	6.4	7.4	10.6	8.4	7.6	9.0	10.9

특히, 전화조사에서 남자 20대, 남자60대 이상의 사전투표자 구성비가 실제보다 낮게 나타난 반면, 여자 20대, 30대, 40대의 사전투표자 구성비가 실제사전투표자보다 소폭 높게 나타났다. 남자 20대의 구성비가 실제보다 낮게 나타난 점은 군 부재자의 사전투표가 전화조사에서 반영되지 못한 점 때문인 것으로 판단되며, 전화조사를 통한 사전투표자 구성비 예측에서 군 부재자의 사전투표에 대한 보정작업이 필요하다는 점이 발견되었다.

군 부재자에 대한 보정작업이 진행될 경우, 서울, 경기, 충북에서 남자 20대 구성비는 실제와 근접하는 방향으로 보정되는 데 반해 강원 지역의 남자 20대 오차는 더 벌어지는 현상이 발견되며, 특히 군 부재자 보정으로 남자 20대의 구성비가 증가할 경우, 현재 실제보다 낮은 사전투표자 구성비를 보이는 남자 60대 이상과 여자 50대 이상의 사전투표자 구성비는 더 작아져 실제와 비교해 오차가 더 벌어지는 현상이 발견된다.

V. 나가며: 대안 검토 종합

사전 전화조사를 통한 사전투표자 구성비 예측 또한, 앞서 사전 전화조사를 통한 사전투표 득표율 예측이 보여주었던 불안정성의 한계를 극복하지 못하고 있는 것처럼 보인다. 전화조사를 활용한 사전투표자의 구성비 및 득표율 예측방법론을 제대로 평가하기 위해서는 지금보다 좀더 안정된 표본 수의 확보가 선행되어야 하며, 분석방법의 정교화에 대한 연구가 추가적으로 진행되어야 할 것이다.

방법론적 연구와는 별개로, 가장 손쉬운 방법으로는 선거가 종료된 이후에 공개된 지난 선거 또는 지난 유사 선거의 사전투표 성/연령별 자료를 활용하여 당해 선거의 성/연령 구성비를 추정하는 방법도 있다. 그러나 선거환경의 변화와 역동적인 선거판세 등을 고려할 때, 이는 과거의 선거자료를 사용한다는 점에서 한계가 있다.

마지막으로 각 선거구에서 추출된 일부 투표소를 대상으로 하여, 사전투표가 진행되는 시점에서 지지도를 묻는 출구조사가 아닌 성/연령을 파악하는 출구조사를 진행하여 각 선거구의 사전투표자 성/연령 구성비를 파악하는 방법을 생각해 볼 수 있다. 하지만, 이미 막대한 예산이 소모되는 당일 출구조사에 더해서 사전투표자 성/연령 구성비 파악을 위한 출구조사의 예산이 추가적으로 필요하다는 점에서 현실적인 어려움이 발생한다. 또한 유권자가 전국 어디서나 사전투표에 참여할 수 있다는 점에서, 특정 선거구에서 출구조사에 참여한 사전투표자가 해당 선거구의 유권자가 아닐 수 있다는 점이 문제로 제기될 수도 있다.

앞으로 선거예측조사가 정확히 수행되기 위해서는 사전투표자의 성/연령별 후보 득표율을 어떻게 추정할 것인가와 함께 성/연령 구성비를 어떠한 방식으로 추정할 것인가에 대한 연구가 지속적으로 수행되어야 할 것이다.

2014년 지방선거는 전국적으로 실시된 사전투표제도로 인해 제한된 여건

속에서 사전투표자에 대한 예측치 산출이 처음으로 진행되었다. 물론, 처음 수행된 사전투표자 예측치 산출 경험에 근거하여 현재의 사전선거예측방법론을 평가하고 성급한 결론을 내리는 것은 무리이다. 이번 사전선거 후보 득표율 예측방법론에서 발견된 특징들을 염두에 두고, 향후 축적된 연구를 통해 사전투표자의 후보 득표율 예측치 및 성/연령 구성비 추정방법론이 개선되어 갈 수 있을 것으로 기대해 본다.

참고문헌

- 경향신문. 2014. “사전투표제, 지방선거의 핵심 변수로.” 2014년 2월 7일자.
- 김영원·김지현. 2007. “2006년 지방선거 출구조사 현황 및 예측오차.” 《조사연구》 8(1): 55-79.
- 김영원·황다솜. 2014. “2014년 지방선거 여론조사 전화조사방법에 따른 예측오차 및 편향.” 《조사연구》 15(4): 1-32.
- 김정훈. 2003. “선거예측과 출구조사 : 16대 대선을 중심으로.” 《조사연구》 4(2): 88.
- 동아일보. 2014. “6월5일 휴가내면 5일간 연휴... 지방선거 투표율 비상.” 2014년 1월 13일자.
- 법제처 국가법령정보센터. 《공직선거법 제167조 2항》.
- 오인환. 1992. 《사회조사방법론-오차요인 집중연구》. 나남.
- 장덕현·이계오·홍영택. 2014. “RDD 전화여론조사의 무응답 편향 보정방법: 2014 지방선거 여론조사 사례 연구.” 《조사연구》 15(4): 57-58.
- 장덕현·조성겸. 2015. “2014 지방선거에서의 여론조사의 정확성 연구 : 조사방법별 정밀성과 편향 비교.” 《조사연구》 16(1): 137.
- 조성겸. 2009. “전화여론조사의 예측정확도 분석.” 《조사연구》 10(1): 57-72.
- 중앙선거관리위원회. 2014. 《제6회 전국동시지방선거 읍면동별 개표결과》.
- 중앙선거관리위원회. 2014. 《제6회 전국동시지방선거 투표율 분석》.
- 최필선·민인식. 2013. “18대 대통령 선거 여론조사의 기관별 정확성 측정 및 비교.” 《조사연구》 14(2): 6.

한국방송협회. 2014. 《2014 지방선거 방송사 공동 예측조사위원회(KEP: Korea Election Pool) 백서》.

홍내리·허명희. 2001. “제16대 국회의원선거의 예측조사에 대한 사후적 검증.”
《조사연구》 2(1): 8..

<접수 2015/6/5, 수정 2015/7/29, 게재확정 2015/8/5>