

연구논문

## 선거 여론조사의 품질과 대표성에 관한 분석: 제21대·제22대 국회의원선거를 중심으로

김인균\*·이승주\*\*·임종호\*\*\*

본 연구는 현재 한국에서 실시되고 있는 선거 여론조사의 문제점을 분석하고, 이를 개선하기 위한 방안을 모색한다. 한국의 선거 여론조사는 할당표집 방법을 통해 자료를 수집하며, 이 과정에서 발생하는 표집오차로 인해 대표성의 왜곡이 발생하고 있다. 또한, 할당표집의 특성상 검증가능성과 재현가능성을 확보하는 것이 사실상 불가능하여 개선이 이루어지기 어려운 측면이 있다. 선관위의 엄격한 규제와 적은 조사비용으로 인해 활용할 수 있는 조사방법이 제한되고 협조율이 낮아 대표성 왜곡이 더욱 심화되고 있다. 이러한 문제의식에서 본 연구는 할당표집과 과학적 표집 방법인 확률표집을 비교하는 시뮬레이션 분석을 실시하였다. 첫 번째 시뮬레이션에서는 협조율의 차이가 할당표집과 확률표집에 미치는 영향을 살펴보고, 두 번째 시뮬레이션에서는 유사한 협조율 조건하에서 할당표집, 확률표집, 재접촉을 포함한 확률표집 간의 대표성 차이를 비교하였다. 또한 2020년 4월 8일 제21대 총선과 2024년 4월 10일 제22대 총선 전 실시된 여론조사 자료를 바탕으로 예측정확성과 편향에 초점을 맞추어 분석하였다. 분석의 결과, 할당표집은 확률표집에 비해 대표성 왜곡이 더 크게 나타난다는 것, 할당표집과 낮은 협조율로 인해 결과의 일관성이 없다는 것, 그리고 할당표집이 재현가능성을 확보하지 못한다는 점을 확인하였다.

주제어: 선거 여론조사, 여론조사 품질, 대표성, 협조율, 제21대 국회의원선거, 제22대 국회의원선거

\* 연세대학교 통계데이터사이언스학과 통합과정(kiminkyun7@gmail.com), 제1저자.

\*\* 연세대학교 통계데이터사이언스학과 석사과정(lsj3232@yonsei.ac.kr), 제2저자.

\*\*\* 연세대학교 통계데이터사이언스학과 교수(ijh38@yonsei.ac.kr), 교신저자.

## I. 서론

본 연구는 현재 선거 여론조사에서 주로 활용되고 있는 비확률표집(non-probability sampling)의 일종인 할당표집(quota sampling)이 여러 가지 문제를 발생시키고 있으며, 이를 해결하기 위해서는 확률표집(probability sampling)에 기반한 과학적 방법으로 여론조사가 수행될 필요가 있다는 문제의식에서 출발하였다. 선거 여론조사의 예측정확성을 높이기 위해서는 표본을 과학적으로 확보하는 것이 중요하다. 이를 위해 제2장에서는 선거 여론조사를 평가하기 위한 품질의 기준들에 대해 검토하고, 제3장에서는 시뮬레이션 분석을 통해 할당표집과 확률표집을 비교·분석한다. 제4장에서는 2020년에 실시된 제21대 국회의원선거와 2024년에 실시된 제22대 국회의원선거(이하 ‘총선’)에서 나타난 특성을 파악한다. 제5장에서는 이러한 결과들을 정리하면서 개선방안을 제시한다. 다수의 선거 여론조사에서 체계적인 편향이 나타나고 있는데, 그 주요 원인은 여론조사에 관한 강한 법적 규제, 적은 조사비용, 할당표집의 활용, 낮은 협조율로 정리할 수 있다. 강한 법적 규제와 적은 조사비용은 할당표집의 활용과 낮은 협조율을 초래하며, 할당표집과 낮은 협조율은 여론조사의 품질을 저해한다.

중앙선거관리위원회(이하 ‘선관위’)와 그 산하의 중앙선거여론조사심의위원회(이하 ‘여론조사심의위’)는 조사기관에서 수집한 선거 여론조사 자료의 품질에 관한 최소 기준을 제시하며, 이를 충족하지 못한 경우에는 공표할 수 없게 하거나 시정, 경고, 정보도문의 게재 등의 조치를 취한다. 여론조사심의위의 규정과 제재는 선거 여론조사의 기본적인 체계를 잡고 일정 수준의 조사품질을 확보하는 데 도움이 되었다. 그러나 조사기관들의 조사품질은 그 수준 이상으로 향상되고 있지 않다. 매 선거마다 비슷한 시기에 서로 다른 조사기관이 발표한 결과가 큰 차이를 보이는 경우와 여론조사 결과와 실제 선거 결과가 상이한 경우가 적지 않다. 이로 인해 국민들의 선거 여론조사에 관한 신뢰가 높지 않은 상황이다(구본상 외 2023).

여론조사의 품질을 향상시키기 위해서는 표본 측면에서 표본의 대표성을 제고해야 하고, 통계량 측면에서 조사결과에 해당하는 통계량의 검증가능성과 재현가능성이 확보되어야 한다. 구체적으로, 확률표집에 근거한 과학적인 방법으로 표본을 수집하고 통계량을 작성해야 한다. 여론조사가 과학적인 방법으로 수행된다는 것은

조사의 목적에 맞게 논리적으로 설계되고 세부적인 내용들이 정합성을 갖고 있어야 하며 그 과정이 검증가능하고 재현가능해야 한다는 것이다. 선거 여론조사를 설계할 때에는 모집단의 특성이 잘 나타날 수 있도록 대표성을 높일 수 있는 표집방법 등을 채택해야 한다. 또한, 선거 여론조사는 분석대상으로 삼는 선거의 특성상 주기적으로 자료를 수집할 수 있는데, 이러한 상황을 고려하여 개선을 할 수 있도록 피드백 체계(feedback system)를 구축할 필요가 있다. 표본의 대표성을 높이기 위해서는 조사가 목적에 맞게 잘 설계되어야 하며, 현재와 같은 낮은 수준의 접촉률, 응답률, 협조율을 크게 높이고 응답의 품질이 제고되어야 한다. 선거 여론조사가 대표성을 잘 나타낼 수 있는 방식으로 설계되더라도, 조사대상의 접촉률, 응답률, 협조율이 낮으면 조사결과가 모집단의 특성을 잘 나타내지 못할 수 있기 때문이다.

그동안 학계, 정부(또는 선관위), 조사기관 등은 여론조사의 품질을 향상시키기 위해 다양한 노력을 해왔다. 학계는 여론조사의 품질을 높이기 위해 선거 전후로 여론조사방법에 관해 다양한 논의를 하거나 개선방안을 제시하였다. 정부는 학계의 논의들 중에서 공정한 선거의 실시, 선거결과에 영향을 미칠 수 있는 여론의 왜곡 등과 관련된 사항들을 적극적으로 수용하여 「공직선거법」과 「선거여론조사기준」 등을 통해 조사기관에 규제를 가하였다. 다수의 조사기관들은 선관위의 규제사항을 수용하고 이를 충족하는 수준으로 여론조사의 품질을 구상하였다.

그동안 학계에서 여론조사 방법과 관련하여 논의한 주요 사항들은 다음과 같다. 김선웅(2004), 강현철 외(2008), 허명희·김영원(2008)은 전화번호부를 표본추출틀로 사용할 때 미등재 가구가 누락되어 편향이 발생할 수 있으므로 임의번호걸기(RDD: Random Digit Dialing)를 활용해야 한다고 주장하였다. 이러한 학계에서의 논의를 바탕으로 2010년에 구성된 방송사공동예측조사위원회(KEP: Korea Election Pool)에서는 임의번호걸기 방식을 활용하여 여론조사를 실시하였다. 이준웅(2011), 이경택 외(2012), 조성겸·장덕현(2012)은 유선전화를 통해 조사한 표본과 휴대전화를 통해 조사한 표본 간에 체계적 차이가 존재하며, 휴대전화를 적극 활용해야 한다고 주장했다. 곽은선·김영원(2022)은 휴대전화 가상번호의 도입 이후 편향이 감소하고 여론조사의 예측정확성이 높아졌다고 한다. 이러한 논의를 바탕으로, 2016년 총선과 2017년 대선에서는 유선전화와 휴대전화를 병행하여 활용하는 조사가 증가하였고, 2020년 총선부터는 휴대전화를 중심으로 조사를 실시하면서 가상번호를 적극적으로 활용하였다. 한편 허명희·황진모(2006)는 기존의 할당표집방법이 평일 낮에 집중되고 비접촉자 및 응답이탈자에 대한 재접촉이 이뤄지지 않기 때문에

체계적 편향이 발생한다고 주장하면서 재접촉을 강화해야 한다고 주장하였다. 조사기관들은 최근 들어 응답률이 매우 낮아짐에 따라 재접촉을 실시하는 경우가 점점 더 많아지고 있다.

여론조사의 품질을 높이기 위한 학계, 선관위, 조사기관의 노력에도 불구하고, 현재 선거 여론조사는 대표성이 낮고, 심지어 과학적인 방법으로 진행되지 않는 경우도 존재한다. 가령, 대부분의 선거 여론조사들이 할당표집을 활용하고 있다. 할당표집을 활용할 경우에 표본의 대표성을 확보하기 어렵고 조사 과정을 재현할 수 없기 때문에 개선되기 어렵다는 문제가 있지만(조성겸 1997; 허명희·황진모 2006; 조성겸 외 2007; 김지윤·강충구 2014; 박인호 외 2019), 여전히 대부분의 선거 여론조사가 할당표집으로 수행되고 있다. 또한 허명희·홍내리(2001)는 무응답 비율이 높아지면 편향이 커지고 오차가 증가한다는 것을 밝혔고 김영원·황다솜(2014)과 김영원·윤지혜(2016)는 응답률 또는 협조율이 높아지면 오차가 감소한다고 했지만, 조사기관들이 응답률을 높이기 위한 획기적인 변화를 모색하고 있지 않다.

## II. 이론적 논의

통계 품질은 전통적으로 정확하고 신속한 통계에 주목했으나 오늘날에는 통계의 다차원적 속성을 바탕으로 이해하고 있다(UNECE 2003). OECD는 통계 품질의 속성으로 관련성(relevance), 정확성(accuracy)<sup>1)</sup>, 적시성(timeliness), 신뢰성(reliability), 접근성(accessibility), 해석가능성(interpretability), 일관성(coherence), 효율성(efficiency) 등 8가지 차원을 제시하였고(OECD 2002; 2011), 이 개념은 미국여론조사학회 및 세계여론조사학회(AAPOR/WAPOR 2021), 유럽 통계청(Eurostat 2000), 캐나다 통계청(Statistics Canada 2002), 한국 통계청<sup>2)</sup> 등에서 전부 또는 일부를 수용하고 있다.

<표 1>은 OECD가 제시한 통계 품질을 구성하는 8가지 차원의 개념을 바탕으로 선거 여론조사를 평가했을 때 확인할 수 있는 문제점을 정리한 것이다. 통계 품질

---

1) 통계 품질의 정확성은 분포의 개념으로서 실제값 주변에서 대칭적으로 잘 추정되는지를 판단하는 기준이지만, 선거 여론조사에서의 정확성은 여론조사가 실제 선거결과를 정확하게 맞혔는지를 판단하는 기준이므로, 두 개념이 일치하지 않는다. 따라서 개념상의 혼동을 방지하기 위해 선거 여론조사에서의 정확성을 다루는 부분에서는 ‘예측정확성’으로 표기한다.

2) 출처: 통계청 홈페이지(<https://kostat.go.kr/menu.es?mid=a10409010000>, 2024.10.28. 확인).

의 평가기준으로 볼 때, 선거 여론조사에서는 예측정확성, 일관성, 효율성, 신뢰성의 차원에서 문제점이 나타나고 있으며, 특히 그중에서 가장 큰 문제로 볼 수 있는 것은 예측정확성이다. 선거 여론조사에서는 체계적인 오차, 편향, 분산이 발생하고 있으며, 따라서 제4장에서 예측정확성의 기준을 활용하여 선거 여론조사 자료를 분석하고자 한다. 일관성의 측면에서, 현행 여론조사가 큰 문제를 갖고 있는 것은 아니다. 선거 여론조사는 주로 할당표집을 활용하고 있는데, 할당표집의 가장 큰 문제점은 검증가능성과 재현가능성이 부족하다는 것이다. 효율성의 측면에서, 통계 품질은 일반적으로 정보의 수집과 이용에서 비롯되는 효용에 비해 과도한 비용이 투입되는 것을 문제점으로 간주한다. 그런데 현행 선거 여론조사에서는 오히려 과도한 효율성의 추구로 조사비용이 저렴하게 책정되고 그 결과 여론조사의 품질이 높지 않다는 문제점이 발생하고 있다.

<표 1> 통계 품질의 평가 기준과 현행 선거 여론조사에서의 문제점

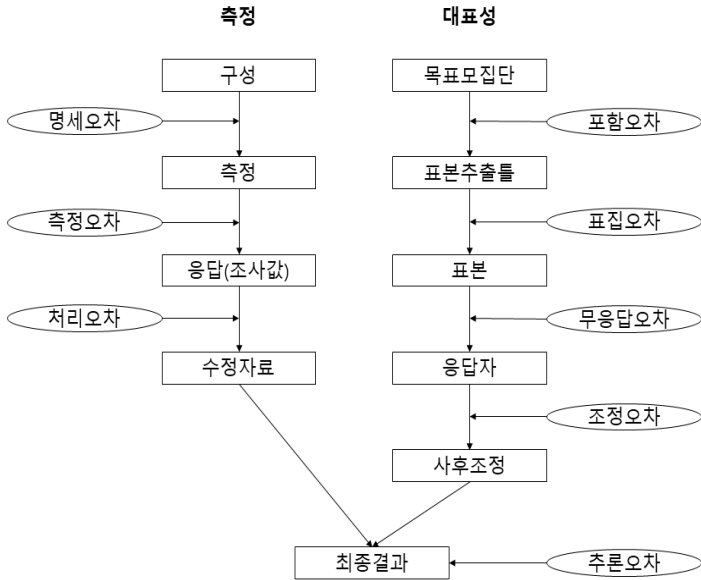
평가 기준	개념	현행 선거 여론조사에서의 문제점
<b>정확성</b> (예측정확성)	추정값이 측정하고자 하는 모집단의 참값을 정확하게 나타내는 것	오차, 편향, 분산이 발생하고 있음
<b>일관성</b>	서로 다른 자료들을 비교할 수 있도록 일관성이 나타나는 것	일부 문제점이 있음. 가령, 할당표집은 검증가능성, 재현가능성 부족함
<b>효율성</b>	여론조사에 필요한 비용을 최소화하면서 여론조사의 이용을 최대화하는 것	일부 문제점이 있음. 조사비용이 저렴하여 여론조사의 품질을 높이기 어려움
<b>신뢰성</b>	이용자가 여론조사 결과가 제시하는 정보에 대해 신뢰를 갖도록 하는 것	일관성, 효율성, 정확성 등에서 발생하는 문제들로 인해 여론조사 신뢰도가 높지 않을 것으로 사료됨
<b>관련성</b>	이용자들의 요구를 충족하는 것	현재 큰 문제점은 없음
<b>접근성</b>	이용자들이 여론조사에 쉽게 접근하고 이용할 수 있는 것	현재 큰 문제점은 없음
<b>적시성</b>	여론조사가 적절한 시점에 수행되고 공표되는 것	현재 큰 문제점은 없음
<b>해석가능성</b>	이용자들이 여론조사 결과의 의미를 정확히 이해할 수 있도록 하는 것	현재 큰 문제점은 없음

신뢰성의 측면에서, 예측정확성, 일관성, 효율성 등에서 발생하고 있는 문제점들로 인해 선거 여론조사의 이용자들로부터 충분한 신뢰를 받지 못하고 있다는 문제점이 나타나고 있다. 이와 관련하여 더 중요한 문제는 정부가 법적 규제를 통해 선거 여론조사에 대한 신뢰성을 높이고자 한다는 것이다. 이로부터 한편으로는 각 조사기관에서 활용하는 조사방법이 획일화된다는 문제점이 파생되고, 다른 한편으로는 조사기관들이 법적 규제를 충족하는 것에 초점을 맞추면서 다른 중요한 문제들에 대해서는 등한시할 수 있다는 문제점이 발생한다. 본 연구에서는 신뢰성을 크게 다루지는 않지만, 현행 선거 여론조사에서 문제를 초래하고 있으므로 제5장에서 해결방안을 제시할 때 일부 다루고자 한다. 이외의 관련성, 접근성, 적시성, 해석가능성은 현행 선거 여론조사에서 별다른 문제점이 나타나고 있지 않으므로 본 논문에서는 다루지 않는다.

한편 선거 여론조사의 경우에는 일반적인 품질 기준과 함께 과학성이라는 차원을 더 고려할 필요가 있다. 선거 여론조사에서 과학성은 이론적으로 체계가 구성되어 목적에 맞는 방법을 취하도록 설계되어야 하고, 각 단계가 전체 과정과 정합성을 가져야 하며 과학적 기준을 바탕으로 여론조사의 품질을 평가할 수 있어야 한다는 것을 의미한다. 선거 여론조사는 대표성이 매우 중요하고 비슷한 시기에 여러 조사기관에서 조사를 실시하며, 선거가 주기적으로 시행되어 여론조사 또한 반복적으로 실시된다는 특성을 갖는다. 선거 여론조사는 현재 사용되고 있는 방법이 옳지 않을 수 있다는 가능성을 염두에 두고 현재의 여론조사가 잘 수행되었는지를 평가할 수 있어야 한다. 특히 여론조사 결과의 좋고 나쁨이 우연에 의한 것인지를 판단할 수 있어야 한다. 이를 바탕으로 다음 조사에서는 대표성을 더욱 잘 확보할 수 있도록 개선할 수 있어야 한다(Converse 1987; Crespi 1987; Herbst 1993; Verba 1996; Newport 2004; Keeter 2012). 과학적 여론조사는 확률표본설계에 기반을 두고 수행하는 것으로 요약할 수 있다(박인호 외 2019). 과학성은 본 연구를 관통하는 가장 핵심적인 기준이며, 특히 제3장에서 시뮬레이션 분석을 수행할 때 적극적으로 활용한다.

여론조사 품질과 관련하여, 현재 수행되고 있는 선거 여론조사의 품질을 정량적으로 평가하기 위해서는 예측정확성에 주목하는 것이 용이하다. 예측정확성은 총조사오차 프레임(total survey error frame, Groves 1989; Biemer & Lyberg 2003)에 따라 오차를 중심으로 파악할 수 있다. 실제로, 예측정확성의 측면에서 국내의 선거 여론조사를 분석하는 다수의 연구들도 총조사오차 프레임을 활용하여 접근했다(곽

은선·김영원 2010; 이승희 2010; 최필선·민인식 2013; 김영원·황다솜 2014; 장덕현·조성겸 2015; 김영원·윤지혜 2016; 김정훈 2022).



<그림 1> 조사주기 관점에서의 총조사오차의 구조

출처: Groves et al. 2004; 박인호 외 2019

총조사오차는 조사설계, 자료수집, 자료처리, 자료분석 등 일련의 과정에서 발생할 수 있는 모든 오차를 의미한다(Biemer 2010). 총조사오차는 측정과 관련된 오류와 대표성과 관련된 오류로 구분할 수 있다. 선거 여론조사는 모집단 전체가 아닌 표본을 추출하여 조사하므로 그 과정에서 발생하는 표집오차(sampling error)가 중요하다. 현행 선거 여론조사에서는 할당표집이 활용되고 있는데, 이 방법으로 인해 대표성이 왜곡되어 표집오차가 발생하고 있다. 이와 함께, 접촉률, 응답률, 협조율이 매우 낮게 나타나고 있으며, 이는 표집오차를 커지게 하는 원인으로 볼 수 있다.

본 연구에서는 여론조사의 예측정확성을 파악하기 위한 척도로 마틴 외(Martin et al. 2005)의 예측정확성 척도  $A$ (predictive accuracy  $A$ )를 활용한다.<sup>3)</sup> 예측정확성

3) 마틴 외(Martin et al. 2005)의 예측정확성 척도  $A$ 는 여론조사의 정확성을 분석하는 다수의 연구들에서 활용된 바 있다(Traugott 2005; Callegaro et al. 2008; Mokrzycki et al. 2009;

척도  $A$ 를 토대로 제4장에서 실제 관측자료 즉, 2020년 총선과 2024년 총선의 예측 정확성을 살펴보고자 하며, 협조율, 보정계수, 조사시간에 따라 예측정확성이 어떻게 다른지를 확인한다. 예측정확성 척도  $A$ 는 경쟁하는 두 정당 후보의 득표율에 대한 승산비(odds ratio)에 로그(log)를 취한 것으로 수식 (1)과 같이 표현된다.

$$A_{ij} = \log\left[\frac{c_{ij}/d_{ij}}{C_i/D_i}\right] \quad (1)$$

$C_i$ 는  $i$ 번째 지역구에서 보수계열 정당(이하 ‘보수당’<sup>4)</sup>)의 후보가 실제 선거에서 획득한 득표율이고,  $c_{ij}$ 는 그중  $j$ 번째 여론조사에서 해당 후보가 얻은 지지율이다.  $D_i$ 는  $i$ 번째 지역구에서 민주계열 정당(이하 ‘민주당’<sup>5)</sup>)의 후보가 실제 선거에서 획득한 득표율이고,  $d_{ij}$ 는 그중  $j$ 번째 여론조사에서 해당 후보가 얻은 지지율이다.  $A_{ij}$ 가 0이면, 여론조사 결과와 선거 결과가 일치한다는 것을 의미한다. 반면에,  $A_{ij}$ 가 0과 차이를 보이면 예측오차가 발생한 것으로 볼 수 있다.  $A_{ij}$ 가 0보다 크면, 보수당 후보의 여론조사 지지율이 실제 선거 득표율에 비해 높게 나타났다는 것을 의미한다.  $A_{ij}$ 가 0보다 작으면, 민주당 후보의 여론조사 지지율이 실제 선거 득표율에 비해 높게 나타났다는 것을 의미한다.  $A_{ij}$ 의 절댓값이 클수록 예측정확성이 낮아진다. 해당 범주에 해당하는  $A_{ij}$ 값들을 평균한 것이  $\bar{A}$ 이며,  $A_{ij}$ 의 절댓값들을 평균한 것이  $|\bar{A}|$ 이다.

예측정확성 척도  $A$ 는 분산을 추정하기가 비교적 쉬우며, 단순확률표집을 가정할 경우에 수식(2)과 같이 추정할 수 있다(Martin et al. 2005; 김영원·황다솜 2014).

$$\text{Var}(A) = \text{Var}\left(\log\left(\frac{c/d}{C/D}\right)\right) \cong \text{Var}\left(\frac{c/d}{C/D}\right) = \frac{1}{ncd} \quad (2)$$

식(2)에서  $n$ 은 여론조사의 수를 의미한다. 본 연구에서는 선행연구(Martin et al. 2005; 김영원·최윤정 2011; 최필선·민인식 2013)를 따라 예측정확성 척도  $A$ 가 수식 (2)를 토대로 계산한 95% 신뢰수준을 벗어나는 경우를 편향으로 간주하며,  $n^+$ 는 양의 편향을,  $n^-$ 는 음의 편향을 의미한다. 편향비율은  $n^+$ 와  $n^-$ 의 합을 해당 범

Panagopoulos 2009; Arzheimer & Evans 2014; 김영원·최윤정 2011; 최필선·민인식 2013; 김영원·황다솜 2014; 김영원·윤지혜 2016; 류제복 2018; 곽은선·김영원 2022).

4) 2020년 총선의 미래통합당, 2024년 총선의 국민의힘이 이에 해당한다.

5) 2020년 총선과 2024년 총선의 더불어민주당이 이에 해당한다.

주에 해당하는 여론조사의 개수  $n$ 으로 나눈 것이다.

한편 본 연구는 응답률 또는 협조율의 높고 낮음이 대표성의 차이를 발생시키는 지를 파악하고자 한다. 김영원·황다솜(2014)과 김영원·윤지혜(2016)는 응답률이 낮을수록 여론조사의 예측정확성이 감소한다고 주장했다. 장덕현 외(2014)는 조사 방법에 따라 응답률이 달라지며, 특히 전화면접 대비 ARS의 비율이 높아질수록 20대와 여성의 응답률이 크게 낮아지고 보수당 지지율이 높아진다는 점에서 대표성의 문제가 발생한다고 주장했다. 반면에, 박인서·백영민(2021)은 선거기간에 접촉률과 협조율이 상승한다고 하면서, 이 시기에 상승한 접촉률과 협조율은 설문조사 표본의 대표성을 개선시킬 가능성보다 악화시킬 가능성이 더 높다고 주장한다.

응답완료자와 접촉실패자, 접촉 후 거절 또는 응답이탈자 간에 체계적인 편향이 존재할 경우에, 그러한 여론조사 결과를 바탕으로 추론할 때 오차가 발생할 가능성이 크다. 또한, 협조율의 차이가 존재함에도 불구하고 모집단과 유사한 특성을 갖는다는 결과가 나오더라도, 그것이 우연에서 비롯된 것이 아닌지 신중하게 확인할 필요가 있다. 왜냐하면 특정 시기에 연령대, 성별 등의 조사단위의 협조율이 상이하게 나타나면서 상쇄되어 우연히 여론조사 결과가 좋게 나타날 수 있기 때문이다.

본 연구는 여론조사심의위가 「선거여론조사기준」(중앙선거여론조사심의위원회고시 제2023-1호, 2023.11.2. 개정)에서 제시하고 있는 접촉률, 응답률, 협조율에 관한 개념을 따른다. 구체적인 내용은 <표 2>에 정리되어 있다. 현재 여론조사심의위는 조사방법에 따른 협조율을 각각 따로 공개하고 있는데, 이를 고려하여 본 연구에서는 조사방법별 가중평균 협조율을 사용한다.

<표 2> 접촉률, 응답률, 협조율의 개념과 구성성분

개념	내용
접촉률(%)	$\frac{I+R}{I+R+eU} \times 100$
응답률(%)	$\frac{I}{I+R} \times 100$
협조율(%)	접촉률 $\times$ 응답률 $\times 100 = \frac{I}{I+R+eU} \times 100$
접촉 실패 사례 중 응답 적격 비율 ( $e$ )	$e = \frac{I+R}{I+R+NE}$

비적격 사례 수( $NE$ )	결번, 사업체번호, 팩스번호, 대상지역 아님, 비적격자, 할당 초과 등을 말하며, 스마트폰앱을 이용한 조사의 경우 휴대전화 전원 차단, 휴대전화의 해당 앱 삭제 또는 알림 거부 설정 등
접촉 실패 사례 수( $U$ )	통화 중, 부재 중, 접촉 안 됨 등
접촉 후 거절 또는 중도이탈 사례 수( $R$ )	스마트폰앱 조사의 경우 해당 자원 위치 지정자에 접속한 후 여론조사 참여도중 이탈, 푸시알림을 수신한 전체 조사대상자 중 해당 자원 위치 지정자에 접속하지 않았거나 접속한 후 응답에 미참여한 경우를 포함
접촉 후 응답 완료 사례 수( $I$ )	조사대상에 해당하며 조사에 응답한 최종적인 완료 사례 수

### III. 시뮬레이션 분석

본 연구의 목적은 현행 선거 여론조사의 한계 및 문제점을 검토하고 개선방안을 논의하는 것이며, 본격적으로 여론조사 자료를 분석하기에 앞서 시뮬레이션 분석을 통해 할당표집과 확률표집을 활용할 때 나타나는 특성의 차이를 비교하고자 한다. 시뮬레이션 분석을 통해서는 크게 두 가지를 다루고자 한다. 시뮬레이션 분석1에서는 할당표집과 확률표집이 협조율에 따라 어떠한 차이를 보이는지에 주목한다. 협조율 10%인 경우와 협조율 5%인 경우를 비교하여, 할당표집을 실시할 때와 확률표집을 실시할 때 협조율과 대표성이 얼마나 다르게 나타나는지를 살펴본다. 시뮬레이션 분석2에서는 협조율 5%인 경우를 대상으로 할당표집, 확률표집, 재접촉을 실시하는 확률표집의 세 가지 조사방법에 따라 대표성이 다르게 나타나는지를 비교한다.

#### 1. 시뮬레이션 자료 형성 및 분석 개요

시뮬레이션 자료와 분석은 현재 실시되고 있는 여론조사와 비슷하게 설계한다. <표 3>은 시뮬레이션 분석에서 활용하는 자료의 형성과 표집 방법을 정리한 것이

다. 시뮬레이션 분석을 위한 자료에서는 지역구 1개의 모집단을 150,000명으로 설정한다.<sup>6)</sup>

<표 3> 시뮬레이션의 자료 형성, 표집, 분석 개요

자료 형성	<p><b>모집단 형성</b></p> <p>150,000명                  성별×연령대: 목표할당 표본 수를 고려하여 임의로 부여함.                  이념: [-1, +1]에서 균등분포를 따르도록 임의로 부여함.</p>	<p><b>확률표집틀 형성</b></p> <p>모집단에서 일부를 임의로 추출함. 예상협조율 10%인 경우 5,000명 예상협조율 5%인 경우 10,000명</p>
표집 방법	<p><b>확률표집</b></p> <p>확률표집틀을 대상으로, 주어진 성별×연령대의 협조율로 추출함.</p>	<p><b>할당표집</b></p> <p>⇒</p> <p>1. 확률표집의 결과를 성별×연령대로 집계함.                  2. (목표할당 표본 수가 충족되지 못한 성별×연령대가 있는 경우) 충족할 때까지, 확률표집틀로 선정되지 않은 모집단을 대상으로, 주어진 성별×연령대의 협조율로 추출함.</p>
	<p><b>재접촉을 실시하는 확률표집</b></p> <p>확률표집틀에서 이전 차수에 협조되지 않은 표본들을 대상으로, 주어진 성별×연령대의 협조율로 추출함.                  성별×연령대의 협조율은 차수가 높아질수록 소폭 감소하도록 설계함.</p>	<p><b>비고</b></p> <p>각 표집은 1,000회 실시하여 평균과 표준편차를 보고함.                  시뮬레이션 분석2에서는 성별×연령대와 함께 이념도 협조율에 영향을 미치도록 설계함.</p>
시뮬레이션 분석 1	<p>확률표집: 예상협조율 10%와 5%의 성별×연령대의 협조 표본 수를 비교함                  할당표집: 예상협조율 10%와 5%에서 목표할당 표본 수를 충족하기 위해 추가적으로 필요한 표본 수를 비교함.</p>	
시뮬레이션 분석 2	<p>확률표집, 재접촉을 실시하는 확률표집, 할당표집의 협조자 이념을 비교함.</p>	

6) 제22대 국회의원선거의 선거인 수 평균은 174,330명(총 선거인 수 44,280,011명, 총 지역구 254개)이다. 편의상 1개 지역구의 모집단을 150,000명으로 설정한다. 출처: 중앙선거관리위원회 선거통계시스템(<http://info.nec.go.kr/>)의 제22대 국회의원선거 선거인명부 확정사항.

현재 조사기관들에서 지역구 여론조사를 실시할 때 목표할당 표본 수의 합계를 주로 500개로 설정하고 각 성별×연령대에 목표할당 표본 수를 나누어 할당하고 있다.<sup>7)</sup> 시뮬레이션 자료에서도 목표할당 표본 수의 합계를 약 500개로 하고 행정안전부 주민등록 인구통계<sup>8)</sup>의 2024년 4월 전국 합계를 기준으로 하여, <표 4>와 같이 각 성별×연령대의 목표할당 표본 수를 할당한다.

<표 4> 목표할당 표본 수 500개 기준 각 성별×연령대의 목표할당 표본 수

성별	연령대					합계
	20대 이하	30대	40대	50대	60대 이상	
여성	38	36	43	49	86	500
남성	41	39	45	50	73	

현재 총선 여론조사의 협조율은 매우 낮은 상황이며, 조사기관들은 모집단에서 확률표집들을 확보할 때 목표할당 표본 수와 예상 협조율을 고려한다. 본 연구에서는 평균 협조율이 10%인 경우와 5%인 경우를 다루는데,<sup>9)</sup> 확률표집들의 크기는 협조율이 10%인 경우에 5,000개, 5%인 경우에 10,000개로 설정한다. 기본적으로 연령대에 따라 협조율에서 차이가 나타난다고 가정한다. 평균 협조율이 5%인 경우에는 20대 이하 3.9%, 30대 4.6%, 40대 5.0%, 50대 5.3%, 60대 이상 5.5%로 설정하고, 평균 협조율이 10%인 경우에는 20대 이하 8.9%, 30대 10.0%, 50대 10.3%, 60대 이상 10.5%로 설정한다. 실제 여론조사에서는 남성이 여성에 비해 평균적으로 약 10~20% p 정도 협조율이 높지만, 편의상 시뮬레이션 분석에서는 성별에 따른 차이를 고려하지 않는다.

또한 본 연구에서는 할당표집, 확률표집과 함께 재접촉을 실시하는 확률표집을 분석한다. 이때 재접촉을 실시하는 확률표집의 경우에 회차를 거듭할수록 협조율이 소폭 감소하도록 설정한다. 할당표집의 경우에는 확률표집과 동일한 협조율을 사용

7) 여론조사심의위원회의 「선거여론조사」 제5조에서 성별, 연령대별, 지역별 가중값 배율을 0.7-1.5로 설정하여 목표할당 표본 수의 충족을 완화하고 있는데, 본 연구에서는 이를 따르지 않고 할당표집을 할 때 목표할당 표본 수를 충족하도록 설계한다.

8) 출처: 행정안전부 주민등록 인구통계(<https://jumin.mois.go.kr/>, 2024.10.28. 확인).

9) 2020년 총선 및 2024년 총선의 여론조사 협조율 평균이 약 4%인데, 편의상 5%를 활용한다.

하여, 확률표집을 실시한 후에 충족되지 않은 각 성별×연령대의 목표할당 표본 수가 모두 충족될 때까지 표집을 진행하며, 이미 충족된 성별×연령대의 표본이 추출될 때에는 접촉 여부만 표시한 후에 다음 표본으로 넘어가는 방식을 취한다.

시뮬레이션 분석 2에서는 평균 협조율을 5%로 하고 연령대와 함께 이념 또한 협조율에 영향을 미치는 것으로 설정한다. 이념은 모집단 150,000명을 대상으로 [-1, +1]의 구간에서 균등분포를 따르도록 부여한 후에 0에서 협조율이 가장 낮고, -1과 +1에서 협조율이 가장 높게 설정한다. 시뮬레이션 자료에서는 거짓 응답이 없다고 가정한다.

## 2. 시뮬레이션 분석 1

이 절에서는 할당표집과 확률표집을 활용하여 평균 협조율이 10%인 경우와 5%인 경우를 비교한다. <표 5>는 평균 협조율이 10%와 5%인 경우에 확률표집을 통해 각각 5,000개, 10,000개의 확률표집들에서 표본을 추출하여 최종적으로 협조한 표본들을 각 성별×연령대로 합산한 것이다. 각 성별×연령대 칸에는 시뮬레이션을 1,000회 실시한 후 협조자 수의 평균값과 함께 괄호안에 표준편차를 기재하였다. 각 성별×연령대의 협조율을 설정할 때 전체 표본 수 합계가 500명 정도가 되도록 설계하였고, 실제로 1,000회 실시의 평균은 평균 협조율이 10%일 때 500.7명, 5%일 때 499.9명이다.

시뮬레이션을 1,000회 실시한 결과, 협조율이 10%일 때 평균값이 목표할당 표본 수를 충족한 성별×연령대는 40대 여성, 50대 여성, 50대 남성, 60대 이상 여성, 60대 이상 남성이었고, 충족하지 못한 성별×연령대는 20대 여성, 20대 남성, 30대 여성, 30대 남성, 40대 남성이었다. 평균적으로 20대 이하 여성은 4.2명(목표할당 38명), 20대 이하 남성은 4.5명(목표할당 41명), 30대 여성은 1.1명(목표할당 36명), 30대 남성은 1.0명(목표할당 39명), 40대 남성은 0.1명(목표할당 45명)이 적게 표집되었다.

<표 5> 평균 협조율 10%와 5%일 때 확률표집의 협조 표본 수 평균

평균 협조율	성별	연령대					합계
		20대 이하	30대	40대	50대	60대 이상	
10%	여성	<b>33.8</b> <b>(5.9)</b>	<b>34.9</b> <b>(6.0)</b>	43.1 (6.7)	50.2 (6.8)	91.0 (9.9)	500.7 (20.9)
	남성	<b>36.5</b> <b>(6.1)</b>	<b>38.0</b> <b>(6.1)</b>	<b>44.9</b> <b>(6.7)</b>	51.5 (7.2)	76.9 (8.5)	
5%	여성	<b>29.4</b> <b>(5.4)</b>	<b>33.4</b> <b>(5.8)</b>	43.3 (6.5)	52.0 (7.3)	94.8 (9.7)	499.9 (21.6)
	남성	<b>32.0</b> <b>(5.5)</b>	<b>36.1</b> <b>(6.1)</b>	45.3 (6.7)	53.2 (7.0)	80.5 (8.4)	

주1: 시뮬레이션을 1,000회 실시한 후에 각 해당 항목의 평균(표준편차)을 기재함.

2: 목표할당을 채우지 못한 성별×연령대는 진하게 표시함.

시뮬레이션을 1,000회 실시한 결과, 협조율이 5%일 때 평균값이 목표할당 표본 수를 충족한 성별×연령대는 40대 여성, 40대 남성, 50대 여성, 50대 남성, 60대 이상 여성, 60대 이상 남성이었고, 충족하지 못한 성별×연령대는 20대 여성, 20대 남성, 30대 여성, 30대 남성이었다. 평균적으로 20대 이하 여성은 8.6명(목표할당 38명), 20대 이하 남성은 9.0명(목표할당 41명), 30대 여성은 2.6명(목표할당 36명), 30대 남성은 2.9명(목표할당 39명) 적게 표집되었다.

연령대별 협조율의 차이가 존재한다고 가정된 상황에서, 확률표집의 평균 협조율이 10%인 경우와 5%인 경우를 비교하면, 평균 협조율이 10%인 경우에 비해 5%인 경우가 평균 협조율보다 낮은 20대 이하와 30대는 적게 표집되고, 평균 협조율보다 높은 50대와 60대 이상은 많이 표집된다는 것을 알 수 있다. 평균 협조율이 큰 경우에 비해 낮은 경우에, 20대 이하 여성, 20대 이하 남성, 30대 여성, 30대 남성과 같은 성별×연령대에서 표본 수가 더 적게 확보될 수 있으며, 목표할당 표본 수를 크게 하회할 수 있다. 그렇게 되면 해당 성별×연령대의 확보된 표본 수가 적기 때문에 평균제곱오차가 크게 나타날 수 있으며, 이후에 사후층화를 한다면 표집된 표본의 특성이 과도하게 많이 반영될 수 있다. 즉, 해당 성별×연령대는 확보된 표본 수가 적기 때문에 표본들의 편향이나 분산이 크게 나타날 수 있고, 해당 표본들에 큰 가중치를 부여하여 사후층화를 하게 되면 그 표본들의 특성이 과도하게 많이 나타날 수 있다.

할당표집을 통해 확률표집에서 충족하지 못한 성별×연령대를 확률표집틀 밖의 모집단에서 추가적으로 추출하였다. 각 성별×연령대의 목표할당 표본 수를 모두 충족하기 위해 추가적으로 접촉한 표본 수는 평균 협조율이 10%일 때 평균적으로 1478.3명(표준편차 640.1)이었고, 5%일 때 평균적으로 4244.5명(표준편차 1562.2)이었다. 평균 협조율이 5%인 경우는 10%인 경우에 비해 20대 이하와 30대에서 목표할당 표본 수를 크게 하회하였는데, 그 결과 목표할당 표본 수를 충족하기 위해 추가적으로 접촉하는 표본 수가 많아졌다.

<표 6> 평균 협조율 5%일 때 재접촉을 실시하는 확률표집의 협조 표본 수 평균

재접촉	성별	연령대					합계
		20대 이하	30대	40대	50대	60대 이상	
1차	여성	56.2 (7.5)	64.6 (7.6)	83.1 (9.2)	100.2 (9.8)	184.2 (13.9)	966.0 (29.0)
	남성	61.3 (7.7)	69.7 (8.4)	87.7 (9.4)	103.1 (10.2)	155.9 (11.8)	
2차	여성	80.0 (9.1)	93.7 (9.2)	121.0 (10.8)	145.8 (11.7)	267.3 (16.5)	1,400.2 (33.6)
	남성	87.2 (9.0)	100.8 (9.9)	127.2 (11.3)	150.1 (12.3)	227.2 (14.8)	
3차	여성	100.3 (10.2)	120.5 (10.4)	156.2 (12.5)	188.9 (13.2)	345.6 (18.4)	1,801.8 (37.3)
	남성	109.1 (10.2)	129.6 (11.2)	164.3 (12.7)	193.5 (13.5)	293.8 (16.5)	
4차	여성	115.2 (11.0)	145.3 (11.3)	189.1 (13.5)	229.0 (14.6)	419.2 (20.2)	2,169.0 (40.6)
	남성	125.4 (11.0)	156.3 (12.1)	199.2 (14.0)	234.0 (14.6)	356.5 (18.3)	

주: 시물레이션을 1,000회 실시한 후에 각 해당 항목의 평균(표준편차)을 기재함.

한편 <표 6>은 평균 협조율이 5%인 경우에 재접촉을 실시하는 확률표집의 재접촉 횟수별 협조 표본 수를 정리한 것이다. 1차 재접촉의 평균 협조율은 4.94%, 2차 재접촉의 평균 협조율은 4.87%, 3차 재접촉의 평균 협조율은 4.79%, 4차 재접촉의

평균 협조율은 4.70%로 설정했다. 1차 재접촉의 추가 협조자 수는 466.1명(총 966.0명), 2차 재접촉의 추가 협조자 수는 434.2명(총 1400.2명), 3차 재접촉의 추가 협조자 수는 401.6명(총 1801.8명), 4차 재접촉의 추가 협조자 수는 367.2명(총 2169.0명)이다. 1차 재접촉의 협조 표본 수를 보면, 목표할당 표본 수에 비해 20대 이하 여성 18.2명, 20대 이하 남성 20.3명, 30대 여성 18.6명, 30대 남성 30.7명, 40대 여성 40.1명, 40대 남성 42.7명, 50대 여성 51.2명, 50대 남성 53.1명, 60대 이상 여성 96.2명, 60대 이상 남성 82.9명이 상회하는 것으로 나타났다. 즉, 1차 재접촉을 완료할 경우에 모든 성별×연령대에서 목표할당 표본 수를 크게 웃도는 것으로 나타났다.

총 접촉 표본 수가 확률표집은 10,000명, 할당표집은 14,244.5명, 1차 재접촉을 실시하는 확률표집은 19,500.1명으로 차이가 나타나기는 했지만, 주어진 확률하에서 1차 재접촉을 실시하는 확률표집은 설계기반에 의하여 대표성이 왜곡되지 않을 뿐만 아니라 평균적으로 각 성별×연령대의 목표할당 표본 수를 모두 충족한다는 것을 확인할 수 있다. 또한, 확률표집과 재접촉을 실시하는 확률표집은 시뮬레이션 간의 편차가 크지 않기 때문에, 예측 가능한 범위 내의 표본 수를 확보할 가능성이 높다.

반면에 할당표집은 각 성별×연령대의 목표할당 표본 수를 충족하기 위해 추가적으로 접촉해야 하는 표본 수가 적지 않을 뿐만 아니라, 시뮬레이션 간의 총 접촉 표본 수의 편차가 상당히 크다는 것을 통해 볼 때 할당표집을 실시하기 전에 몇 명의 표본에게 접촉해야 하는지를 정확히 파악하기 어렵다. 무엇보다도, 충족되지 않은 성별×연령대의 표본을 추출하는 과정에서 표집오차가 크게 발생할 수 있으며, 결과적으로 대표성이 왜곡될 수 있다.

### 3. 시뮬레이션 분석 2

시뮬레이션 분석 2에서는 평균 협조율이 5%인 경우로 한정하여 이념에 따라 접촉률을 다르게 설정할 때, 할당표집, 확률표집, 재접촉을 실시하는 확률표집에서 어떠한 결과가 나타나는지를 비교한다. 모집단을 대상으로 [-1, +1]의 구간에서 가장 진보적인 사람에게 -1, 가장 보수적인 사람에게 +1을 부여하고, 협조율이 -1에서 0까지 선형적으로 감소하다가 0에서 +1까지 선형적으로 증가하도록 하여, 양극단의 협조율이 중도에 비해 약 0.30%p 크게 나타나도록 한다.

&lt;표 7&gt; 할당표집, 확률표집, 재접촉을 실시하는 확률표집의 협조자 이념 비교

	전체		[-1, 0] 구간		[0, +1] 구간	
	평균	분산	평균	분산	평균	분산
할당표집	-0.001 (0.026)	0.350 (0.013)	-0.519 (0.018)	0.082 (0.004)	0.518 (0.018)	0.082 (0.005)
확률표집	-0.001 (0.027)	0.349 (0.014)	-0.517 (0.018)	0.082 (0.005)	0.516 (0.018)	0.082 (0.005)
1차 재접촉	-0.001 (0.019)	0.351 (0.010)	-0.519 (0.013)	0.081 (0.003)	0.519 (0.013)	0.082 (0.004)
2차 재접촉	-0.001 (0.016)	0.352 (0.008)	-0.521 (0.011)	0.081 (0.003)	0.520 (0.011)	0.081 (0.003)
3차 재접촉	-0.001 (0.014)	0.354 (0.007)	-0.522 (0.009)	0.081 (0.002)	0.522 (0.010)	0.081 (0.003)
4차 재접촉	-0.001 (0.012)	0.355 (0.007)	-0.524 (0.009)	0.081 (0.002)	0.523 (0.009)	0.081 (0.002)

주: 시뮬레이션을 1,000회 실시한 후에 각 해당 항목의 평균(표준편차)을 기재함.

<표 7>은 할당표집, 확률표집, 재접촉을 실시하는 확률표집의 협조자 이념을 전체, [-1, 0] 구간, [0, +1] 구간으로 나누어 정리한 것이다. 각 칸에는 평균과 분산을 기재했고, 괄호 안에는 1,000회 실시한 시뮬레이션 간의 표준편차를 기입했다. 우선, 확률표집과 재접촉을 실시하는 확률표집을 비교한다. 시뮬레이션 간의 표준편차는 재접촉 횟수가 증가할수록 감소하였는데, 이는 총 협조자 수가 증가하는 것에서 기인한 결과로 볼 수 있다. 전체 협조자의 이념 평균은 큰 차이가 없었지만, [-1, 0] 구간과 [0, +1] 구간을 나눠서 이념 평균을 보면 일관성 있는 특성이 나타났다. 즉, 재접촉 횟수가 증가할수록 [-1, 0] 구간의 협조자 이념의 평균은 감소했고, [0, +1] 구간의 협조자의 이념 평균은 증가했으며, 두 구간에서 모두 협조자 이념의 분산이 감소했다. 다시 말하면, 재접촉 횟수가 증가할수록 더 진보적인 협조자와 더 보수적인 협조자가 상대적으로 많이 표집된다는 것이다.

다음으로, 할당표집의 특성을 파악한다. 전체 협조자의 이념 평균은 할당표집과 확률표집이 -0.001로 같았지만, 분산은 0.350으로 확률표집의 0.349에 비해 조금 더 컸다. 할당표집과 확률표집의 차이는 구간별로 나눠서 볼 때 더욱 분명하게 파악할 수 있다. 할당표집 협조자의 이념 평균은 [-1, 0] 구간에서 -0.519이고, [0, +1] 구간

에서 0.518이다. 확률표집 협조자의 이념 평균은 [-1, 0] 구간에서 -0.517이고, [0, +1] 구간에서 0.516이다. 할당표집 협조자의 이념 평균은 확률표집 협조자의 이념 평균에 비해 조금 더 절댓값이 더 컸다. 이러한 결과는 할당표집을 활용할 경우에 조사방법, 조사주체, 조사대상별로 응답자들이 편향되게 표집될 수 있으며 선거별로 패턴이 달라질 수 있음을 의미한다.

한편 시뮬레이션 간 표준편차를 보면, 전체와 [-1, 0] 구간, [0, +1] 구간 모두에서 확률표집이 가장 컸고, 할당표집이 그보다 약간 작았지만 큰 차이가 없었으며, 1차 재접촉부터 크게 감소했다. 다시 말하면, 재접촉을 실시하는 확률표집은 재접촉을 실시하지 않는 확률표집과 할당표집에 비해 시뮬레이션 간 차이가 적게 나타났으며, 대표성의 측면에서 더 안정적인 결과를 산출할 수 있다.

## IV. 여론조사 자료 분석

### 1. 자료

본 연구에서는 예측정확성에 주목하여 선거 여론조사에서 나타나는 특성을 살펴 보고자 하며, 이를 위해 2020년 제21대 총선과 2024년 제22대 총선의 지역구 여론조사 자료를 활용한다. 현재 총선 여론조사는 몇 가지 문제점을 갖고 있다. 첫째, 총선은 250여 개의 지역구 단위를 대상으로 하는데, 그중 다수의 지역구에서 여론조사가 충실히 진행되지 않고 있다. 총선 여론조사는 유력 국회의원 후보자가 출마한 지역구에 관심도가 집중되며, 그렇지 않은 지역구는 여론조사가 불과 몇 차례 실시되거나 심지어 실시되지 않는 경우도 많다. <표 8>은 2020년 총선 및 2024년 총선의 여론조사 실시횟수별 지역구의 수와 비중을 정리한 것이다. 2020년 총선은 253개 중 110개 지역구에서, 2024년 총선은 254개 중 100개 지역구에서 총선 기간에 여론조사가 실시되지 않았거나, 비공표 목적으로 실시되어 여론조사심의위에 보고되지 않았다.<sup>10)</sup>

10) 「공직선거법」 제108조 제3항에 따라 제3자로부터 여론조사를 의뢰받은 여론조사 기관·단체, 정당, 방송사업자, 신문사업자 및 정기간행물사업자, 뉴스통신사업자, 인터넷언론사 등 중에서 관계 법령의 특정 조건을 충족한 자들은 여론조사심의위에 서면 신고를 하지 않을 수 있다. 이들은

&lt;표 8&gt; 2020년 총선 및 2024년 총선의 여론조사 실시횟수별 지역구의 수와 비중

	미실시	1회	2회	3회	4회	5회 이상	합계
2020년 총선	110 (43.5)	52 (20.6)	38 (15.0)	28 (11.1)	7 (2.8)	18 (7.1)	253 (100)
2024년 총선	100 (39.4)	70 (27.6)	37 (14.6)	21 (8.3)	5 (2.0)	21 (8.3)	254 (100)

둘째, 대선을 비롯한 전국 단위의 여론조사가 표본크기를 약 1,000개 또는 2,000개로 설정하는 반면에, 지역구 단위의 여론조사는 대부분 500개 정도로 설정하며, 그 결과 할당표집을 활용할 때 표본크기가 특정 성별과 연령대에 매우 적게 배정되어 왜곡이 크게 발생할 수 있다. <표 9>는 2020년 총선 및 2024년 총선의 표본크기별 여론조사의 빈도수와 비중을 정리한 것이다. 500-599개의 표본크기가 활용된 여론조사는 2020년 총선에서 총 365개 중 297개였고, 2024년 총선에서 총 373개 중 348개였다.

&lt;표 9&gt; 2020년 총선 및 2024년 총선의 표본크기별 여론조사의 빈도수와 비중

	500-599	600-699	700-799	800-899	900-999	1,000 이상	합계
2020년 총선	297 (81.4)	6 (1.6)	27 (7.4)	9 (2.5)	0 (0.0)	26 (7.1)	365 (100)
2024년 총선	348 (93.3)	9 (2.4)	12 (3.2)	1 (0.3)	0 (0.0)	3 (0.8)	373 (100)

셋째, 현재 총선 여론조사에서 활용하고 있는 조사방법이 획일적이다. 2020년 총선의 365개 여론조사와 2024년 총선의 373개 여론조사 전부는 유선면접, 유선ARS, 무선면접, 무선ARS의 네 가지 방법을 사용했다. 네 가지 조사방법의 평균 비율을 보면, 2020년 총선 여론조사에서는 유선ARS 15.6%, 유선면접 6.0%, 무선ARS 38.9%, 무선면접 39.6%가 활용되었고, 2024년 총선 여론조사에서는 유선ARS 2.4%, 무선ARS 48.1%, 무선면접 49.5%가 활용되었다. 대부분의 여론조사가 무선ARS와 무선면접으로 실시되었다. 이외의 방법들은 거의 활용되지 않았다.

여론조사심의위에 신고하지 않고 「공직선거법」과 「선거여론조사기준」을 따르지 않는 여론조사들을 실시할 수 있는데, 이 여론조사들에 의해 여론이 왜곡될 수 있다.

&lt;표 10&gt; 2020년 총선 및 2024년 총선에서 활용한 조사방법의 비율

	유선ARS	유선면접	무선ARS	무선면접	합계
2020년 총선	15.6	6.0	38.9	39.6	100
2024년 총선	2.4	0	48.1	49.5	100

총선 여론조사에서 나타나는 문제점들은 조사비용의 측면에서 비롯되는 것으로 볼 수 있다. 신문사, 방송사 등의 여론조사 의뢰자들은 주목을 끌 수 있는 지역구를 중심으로 의뢰를 하며 그렇지 않은 지역구에 대해서는 관심을 갖지 않는다. 조사기관들은 선관위가 제시하는 최소한의 조건을 충족하는 선에서 표본크기를 설정하고 비용이 적게 드는 조사방법을 활용하여 여론조사를 실시하고 있다.

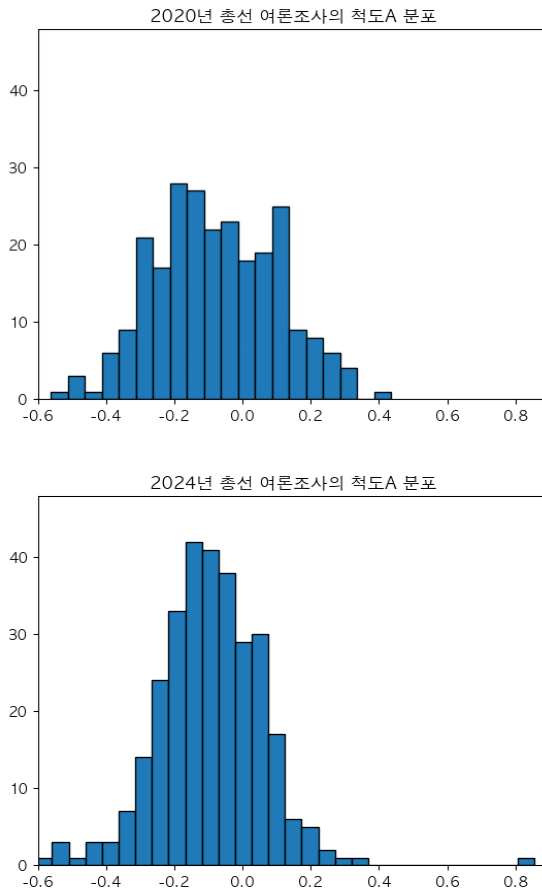
2020년 총선과 2024년 총선을 대상으로 분석하는 이유는 다음과 같다. 공식적으로 한국의 여론조사 결과는 여론조사심의위가 출범한 직후인 2014년 3월부터 수집되었다. 그 이후로 2016년 총선, 2020년 총선, 2024년 총선이 실시되었는데, 2016년 총선은 유선전화를 활용한 여론조사가 다수였고, 가중값 배율 규정이 이후의 선거들과 크게 달랐으며 접촉률을 공개하지 않는 등 2020년 총선, 2024년 총선과 상당한 차이를 보였다. 반면에, 2020년 총선과 2024년 총선은 일부 차이가 존재하지만 대체로 비슷한 선거 경쟁의 상황에서 비슷한 조사방법을 활용하여 여론조사가 실시되었다.

본 연구는 2020년 총선과 2024년 총선이 전반적으로 양대 정당 즉, 보수당과 민주당의 경쟁구도로 치러졌다는 점을 고려하여, 양대 정당의 경쟁이 강하게 나타난 지역구들을 중심으로 분석했다. 이는 양대 정당 후보를 중심으로 경쟁이 이뤄지는 지역구와 그렇지 않은 지역구 간에 여론조사의 여러 가지 특성들에서 차이가 나타날 수 있기 때문이다. 구체적으로, ① 양대 정당 외의 후보가 1위 또는 2위를 기록한 지역구를 제외했고, ② 양대 정당 후보의 득표율 차이가 25%p 이상인 곳을 제외했다. 2020년 총선의 경우에는 여론조사가 실시된 143개 지역구 중에서 위의 두 가지 기준에 따라 46개 지역구를 제외하고 남은 97개 지역구와 248건의 여론조사를 대상으로 한다. 2024년 총선의 경우에는 여론조사가 실시된 154개 지역구 중에서 위의 두 가지 기준에 따라 29개 지역구를 제외하고 남은 125개 지역구와 302건의 여론조사를 대상으로 한다. 여론조사의 특성을 파악하기 위하여, 여론조사심의위

홈페이지에 공개된 조사기관, 조사일시, 조사대상, 조사방법, 표본크기, 피조사자 선정방법, 조사시간, 보정계수 등에 관한 정보와 결과분석 자료를 병합하여 사용한다.

## 2. 분석

2020년 및 2024년 총선 여론조사를 대상으로 마틴 외(Martin et al. 2005)의 예측 정확성 척도  $A$ 를 활용하여 분석한다. <그림 2>는 2020년 및 2024년 총선 여론조사의 예측정확성 척도  $A$ 의 분포를 보여준다.



<그림 2> 2020년 및 2024년 총선 여론조사의 척도  $A$  분포

<표 11>은 예측정확성 척도  $A$ 를 바탕으로 2020년 및 2024년 총선 여론조사의 예측정확성 및 편향을 정리한 것이다. 2020년 총선 여론조사와 2024년 총선 여론조사 모두 예측정확성 척도  $A$ 의 중앙값과 평균이 0을 약간 하회하는 것을 확인할 수 있다. 각 여론조사의 예측정확성 척도  $A_{ij}$ 가 95% 신뢰수준을 벗어나는 경우의 비율을 계산한 편향비율이 2020년 및 2024년 총선 여론조사에서 각각 4.4%, 4.6%로 크지는 않았다. 그러나 전반적으로는 실제 선거결과에 비해 여론조사에서 민주당 후보의 지지율이 보수당 후보의 지지율에 비해 높게 나타났다(민주당 편향). 한편, 2020년 총선 여론조사에 비해 2024년 총선 여론조사는  $\bar{A}$ ,  $|\bar{A}|$ ,  $s.d.(A)$ 가 모두 소폭 감소했다.

<표 11> 2020년 및 2024년 총선 여론조사의 예측정확성 및 편향

선거	$\bar{A}$	$ \bar{A} $	$s.d.(A)$	$n$	$n^+$	$n^-$	편향비율
2020년 총선	-0.076	0.161	0.180	248	7	4	0.044
2024년 총선	-0.092	0.141	0.157	302	6	8	0.046

이와 같은 결과는 2016년 총선 여론조사와 큰 차이를 보인다. 김영원·윤지혜(2016)에 따르면 2016년 총선 여론조사에서는  $\bar{A}$ 과  $|\bar{A}|$ 은 각각 약 0.30, 약 0.31로 여론조사들에서 전반적으로 보수당 편향이 나타났고  $s.d.(A)$ 도 0.20을 상회했다. 편향비율은 경기 0.426, 서울 0.474, 인천 0.316로 나타났으며 각 여론조사의 예측정확성 척도  $A_{ij}$ 가 95% 신뢰수준을 벗어나는 경우가 상당히 많았다.

이처럼 2016년 총선 여론조사와 2020년 및 2024년 총선 여론조사가 큰 차이를 보인 것은 피조사자 선정방법과 조사방법 때문으로 볼 수 있다. 2016년 총선에서는 다수의 여론조사가 유선전화를 활용하여 조사를 실시하였고 전화번호부와 임의번호걸기를 중심으로 피조사자를 선정하였다. 반면에 2020년 및 2024년 총선 여론조사에서는 주로 무선전화를 활용하여 조사를 실시하였고 안심번호를 활용하여 피조사자를 선정하였다. 각 여론조사의 유선전화 사용 비율의 평균은 2016년 총선에서 92.3%이었는데, 2020년 총선에서 20.6%, 2024년 총선에서 2.7%로 크게 감소하였다. 안심번호는 2016년 총선에서 활용되지 않았고, 2020년 총선에서 79.7%, 2024년 총선에서 97.3%가 활용되었다.

<표 12>와 <표 13>은 2020년 및 2024년 총선 지역구 여론조사의 예측정확성 및 편향을 시·도별 또는 권역별로 묶어서 제시한 것이다. 예측정확성과 편향의 측면에서 특별히 다른 특성을 보이는 시·도 또는 권역은 없는 것으로 나타났다. 2020년 총선에서는 대구·경북과 강원에서 실시된 여론조사들이 다른 시·도 또는 권역에서 실시된 여론조사들에 비해 민주당 편향이 강하게 나타났다. 2024년 총선에서는 부산·울산·경남과 제주에서 실시된 여론조사들이 다른 시·도 또는 권역에서 실시된 여론조사들에 비해 민주당 편향이 강하게 나타났다. 강원에서 실시된 여론조사들은 보수당 편향이 나타났다.

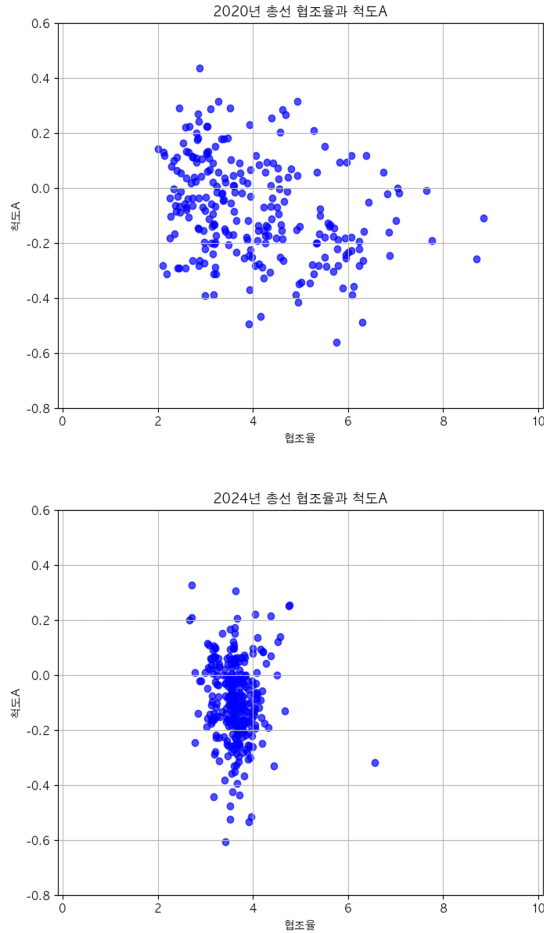
<표 12> 2020년 총선 여론조사의 시·도별, 권역별 예측정확성 및 편향

시·도 또는 권역	$\bar{A}$	$ \bar{A} $	$s.d.(A)$	$n$	$n^+$	$n^-$	편향비율
서울	-0.081	0.155	0.171	43	1	1	0.047
인천·경기	-0.048	0.157	0.178	56	0	0	0
대전·세종·충북·충남	-0.089	0.149	0.158	41	0	0	0
부산·울산·경남	-0.038	0.157	0.190	57	1	1	0.035
대구·경북	-0.135	0.202	0.206	22	1	0	0.046
강원	-0.160	0.197	0.179	19	0	0	0
제주	-0.089	0.132	0.137	10	0	0	0

<표 13> 2024년 총선 여론조사의 시·도별, 권역별 예측정확성 및 편향

시·도 또는 권역	$\bar{A}$	$ \bar{A} $	$s.d.(A)$	$n$	$n^+$	$n^-$	편향비율
서울	-0.105	0.125	0.112	70	1	2	0.043
인천·경기	-0.095	0.138	0.158	124	2	3	0.040
대전·세종·충북·충남	-0.009	0.105	0.131	48	1	0	0.021
부산·울산·경남	-0.199	0.220	0.173	43	2	1	0.070
강원	0.054	0.111	0.129	14	0	0	0
제주	-0.171	0.171	0.135	3	0	0	0

한편 전통적으로 지역주의 성향이 강하게 나타나는 권역들 즉, 2020년 총선의 부산·울산·경남과 대구·경북, 2024년 총선의 부산·울산·경남에서는  $s.d.(A)$ 가 다른 시·도 및 권역들에 비해 크게 나타났다. 이와 같은 결과는 양대 정당이 경쟁하는 지역구에 비해  $A_{ij}$ 가 상대적으로 더 큰 차이를 보이고 있다는 것을 의미한다.



<그림 3> 2020년 및 2024년 총선 여론조사의 협조율과 척도 A의 산점도

<그림 3>은 2020년 및 2024년 총선 여론조사의 협조율과 예측정확성 척도 A의 산점도이다. <표 14>와 <표 15>는 각각 2020년 및 2024년 총선 여론조사의 협조

율에 따른 예측정확성 및 편향의 특성을 정리한 것이다. 2020년 총선 여론조사의 경우에는 협조율이 주로 2%에서 6%까지 퍼져 있으며, 협조율이 높을수록  $A_{ij}$ 의 값이 감소하는 경향성이 나타나고 있다. 반면에, 2024년 총선 여론조사의 경우에는 대부분의 여론조사에서 협조율이 3% 이상 4% 미만에 집중되어 있으면서도  $A_{ij}$ 가 특정 지점에 몰려 있지 않은 것을 확인할 수 있다.

<표 14> 2020년 총선 여론조사의 협조율별 예측정확성 및 편향

협조율	$\bar{A}$	$ \bar{A} $	$s.d.(A)$	$n$	$n^+$	$n^-$	편향비율
2% 이상 3% 미만	-0.012	0.139	0.169	65	1	0	0.015
3% 이상 4% 미만	-0.052	0.145	0.175	75	1	1	0.027
4% 이상 5% 미만	-0.089	0.168	0.184	51	2	1	0.059
5% 이상 6% 미만	-0.182	0.218	0.160	33	2	1	0.091
6% 이상	-0.153	0.177	0.155	24	0	1	0.042

<표 15> 2024년 총선 여론조사의 협조율별 예측정확성 및 편향

협조율	$\bar{A}$	$ \bar{A} $	$s.d.(A)$	$n$	$n^+$	$n^-$	편향비율
2% 이상 3% 미만	0.036	0.132	0.180	9	0	0	0
3% 이상 4% 미만	-0.105	0.142	0.153	256	3	8	0.043
4% 이상 5% 미만	-0.028	0.129	0.148	36	0	1	0.028
6% 이상	-0.320	0.320		1	0	0	0

2020년 총선 여론조사에서는 상대적으로 협조율이 낮은 2% 이상 3% 미만, 3% 이상 4% 미만, 4% 이상 5% 미만에 비해, 상대적으로 협조율이 높은 5% 이상 6% 미만, 6% 이상의 경우에  $\bar{A}$ 가 작았다. 전자의  $|\bar{A}|$ 에 비해 후자의  $|\bar{A}|$ 가 컸지만 그 차이는  $\bar{A}$ 에서 나타난 차이에 비해 작았다. 이러한 결과는 2020년 총선 여론조사에서 전반적으로 협조율이 낮은 여론조사는 민주당 편향이 작게 나타나고 일부 여론조사는 보수당 편향이 나타났다는 것과 협조율이 높은 여론조사는 대체로 민주당 편향이 크게 나타났다는 것에서 기인한다.

2024년 총선 여론조사에서는 협조율에 따른  $\bar{A}$ ,  $|\bar{A}|$ 이 일관성을 보이지 않았다. 대부분의 여론조사 협조율이 3% 이상 4% 미만에 해당하였고, 이보다 높은 4% 이상 5% 미만의 경우에는  $\bar{A}$ 가 컸고  $|\bar{A}|$ 는 작았다.

이처럼 2020년 총선 여론조사에서 나타난 결과가 직관에 부합하지 않고, 2020년 총선 여론조사와 2024년 총선 여론조사의 결과가 일관성 있게 나타나지 않는 것은 할당표집과 낮은 협조율에서 발생한 선택편향(selection bias)에서 기인하는 것으로 사료된다. 설계에 기반한 확률표집이 아닌 할당표집을 활용하여 표본을 추출하고 대부분의 여론조사의 협조율이 5% 이하이기 때문에 모집단의 대표성을 확보하기 어려울 수 있다. 전수조사를 실시하지 않는 이상 선택편향이 발생할 수 있는데, 현재의 총선 여론조사와 같이 협조율이 매우 낮은 상황에서는 협조율이 소폭 증가하더라도 예측정확성이 반드시 향상되지는 않을 수 있다.

선거 여론조사에서는 연령대에 따라 협조율이 다르게 나타날 수 있다. 2019년부터 여론조사심의위는 가중값 배율을 0.7-1.5로 의무화하고 있으며, 조사기관들은 가중값 배율을 충족하지 못한 연령대의 표본 수를 확보할 때까지 할당표집을 통해 조사를 실시하고 있다. 선거 여론조사에서는 대체로 20대 이하와 30대가 낮은 협조율로 인해 과소표집되는 반면에, 60대 이상이 상대적으로 높은 협조율로 인해 과다표집되고 있다.

<표 16> 2020년 총선 여론조사의 20대 이하 보정계수별 예측정확성 및 편향

20대 보정계수	$\bar{A}$	$ \bar{A} $	$s.d.(A)$	$n$	$n^+$	$n^-$	편향비율
0.9 미만	-0.128	0.160	0.159	15	0	0	0
0.9 이상 1.1 미만	-0.108	0.175	0.177	88	3	3	0.046
1.1 이상	-0.051	0.153	0.180	145	2	3	0.034

20대 이하의 보정계수와 협조율에 따라 예측정확성이 다르게 나타나는지를 확인할 필요가 있다. 20대 이하의 보정계수가 0.9 이상 1.1 미만에 해당하는 경우는 20대 이하가 적정 수준으로 표집된 것이고, 보정계수가 0.9 미만에 해당하는 경우는 20대 이하가 과다표집된 것이고, 보정계수가 1.1 이상에 해당하는 경우는 20대 이하가 과소표집된 것이다. <표 16>은 2020년 총선 여론조사의 20대 이하 보정계수

별 예측정확성 및 편향을 정리한 것이다. 20대 이하 보정계수가 0.9 미만인 경우에 비해 0.9 이상 1.1 미만인 경우에, 0.9 이상 1.1 미만인 경우에 비해 1.1 이상인 경우에  $\bar{A}$ 가 작게 나타났고, 이는 전자에 비해 후자의 민주당 편향이 작게 나타난 것으로 볼 수 있다. 이와 같은 결과는 30대 보정계수별 예측정확성 및 편향에서도 유사하게 나타났다.

<표 17> 2024년 총선 여론조사의 20대 이하 보정계수별 예측정확성 및 편향

20대 이하 보정계수	$\bar{A}$	$ \bar{A} $	<i>s.d.</i> ( <i>A</i> )	<i>n</i>	<i>n</i> <sup>+</sup>	<i>n</i> <sup>-</sup>	편향비율
0.9 미만							
0.9 이상 1.1 미만	-0.142	0.183	0.171	29	2	1	0.103
1.1 이상	-0.087	0.136	0.155	273	4	8	0.044

<표 17>은 2024년 총선 여론조사의 20대 이하 보정계수별 예측정확성 및 편향을 정리한 것이다. 2024년 총선 여론조사에서는 20대 이하 보정계수가 0.9 미만인 사례는 없었고, 2020년 총선 여론조사와 유사하게 0.9 이상 1.1 미만인 경우에 비해 1.1 이상인 경우에  $\bar{A}$ 가 작게 나타났다.

이번에는 20대 이하 보정계수와 함께 협조율을 함께 고려하여 예측정확성 및 편향을 살펴보고자 한다. <표 18>과 <표 19>는 2020년 및 2024년 총선 여론조사의 협조율과 20대 이하 보정계수에 따른 예측정확성 및 편향을 보여준다. 우선, 2020년 총선 여론조사의 20대 이하 보정계수가 0.9 이상 1.1 미만으로 적정표집된 것을 보면, 협조율이 낮은 경우에 비해 협조율이 높은 경우의  $\bar{A}$ 가 더 작다는 것을 확인할 수 있다. 다음으로, 20대 이하 보정계수가 0.9 이상 1.1 미만인 여론조사들과 1.1 이상인 여론조사들은 협조율 5% 미만인 경우와 5% 이상인 경우에서  $\bar{A}$ 가 큰 차이를 보였는데 그 패턴은 상이했다. 협조율 2% 이상 3% 미만, 3% 이상 4% 미만, 4% 이상 5% 미만의 조건에서는 20대 이하 보정계수가 0.9 이상 1.1 미만인 것에 비해 1.1 이상인 것의 민주당 편향이 크게 감소했다. 반면에, 협조율 5% 이상 6% 미만, 6% 이상의 조건에서는 20대 이하 보정계수가 0.9 이상 1.1 미만인 것에 비해 1.1 이상인 것의 민주당 편향이 크게 증가했다. 이러한 결과들은 단순히 협조율이 소폭 높아진다고 해서 대표성이 제고되는 것은 아니라는 점을 보여준다.

한편 2024년 총선 여론조사를 보면,  $\bar{A}$ ,  $|\bar{A}|$ ,  $s.d.(A)$ 에서 모두 일관성 있는 특성이 나타나지 않았다. 대부분의 여론조사의 협조율이 3% 이상 4% 미만에 해당했는데, 그중 20대 이하 보정계수가 0.9 이상 1.1 미만인 것이 1.1 이상인 것에 비해  $\bar{A}$ 에서 민주당 편향이 더 크게 나타났다. 4% 이상 5% 미만의 경우에는 20대 이하 보정계수가 0.9 이상 1.1 미만인 것의  $\bar{A}$ 에서 보수당 편향이 나타난 반면에 1.1 이상인 것의  $\bar{A}$ 에서는 민주당 편향이 나타났다.

<표 18> 2020년 총선 여론조사의 협조율과 20대 이하 보정계수의 예측정확성 및 편향

협조율	20대 이하 보정계수	$\bar{A}$	$ \bar{A} $	$s.d.(A)$	$n$	$n^+$	$n^-$	편향비율
	0.9 미만	-0.088	0.203	0.238	3	0	0	0
2% 이상 3% 미만	0.9 이상 1.1 미만	-0.019	0.136	0.172	27	1	0	0.037
	1.1 이상	-0.000	0.136	0.166	35	0	0	0
3% 이상 4% 미만	0.9 미만	-0.117	0.134	0.148	8	0	0	0
	0.9 이상 1.1 미만	-0.106	0.214	0.216	10	0	0	0
	1.1 이상	-0.033	0.135	0.170	57	1	3	0.070
4% 이상 5% 미만	0.9 미만	-0.350	0.350		1	0	0	0
	0.9 이상 1.1 미만	-0.183	0.200	0.158	22	0	0	0
	1.1 이상	-0.006	0.137	0.162	28	1	0	0.036
5% 이상 6% 미만	0.9 미만	-0.254	0.254		1	0	0	0
	0.9 이상 1.1 미만	-0.127	0.197	0.176	13	0	0	0
	1.1 이상	-0.215	0.230	0.146	19	1	1	0.105
6% 이상	0.9 미만	-0.055	0.055	0.076	2	0	0	0
	0.9 이상 1.1 미만	-0.142	0.164	0.140	16	0	0	0
	1.1 이상	-0.212	0.251	0.205	6	0	0	0

&lt;표 19&gt; 2024년 총선 여론조사의 협조율과 20대 이하 보정계수의 예측정확성 및 편향

협조율	20대 이하 보정계수	$\bar{A}$	$ \bar{A} $	$s.d.(A)$	$n$	$n^+$	$n^-$	편향비율
2% 이상 3% 미만	1.1 이상	0.036	0.132	0.180	9	0	0	0
3% 이상 4% 미만	0.9 이상 1.1 미만	-0.182	0.186	0.137	22	0	1	0.045
	1.1 이상	-0.097	0.138	0.153	234	3	7	0.043
4% 이상 5% 미만	0.9 이상 1.1 미만	0.034	0.148	0.185	6	0	0	0
	1.1 이상	-0.041	0.125	0.140	30	0	1	0.033
6% 이상	0.9 이상 1.1 미만	-0.320	0.320		1	0	0	0

## V. 결론

본 연구는 선거 여론조사가 과학적으로 수행될 필요가 있다는 관점을 취하고 있다. 시뮬레이션 분석을 통해 비확률표집의 일종인 할당표집, 확률표집, 재접촉을 실시하는 확률표집의 결과를 비교하였다. 시뮬레이션 분석 1에서는 성별과 연령대를 고려하여 분석을 진행했으며 표본 수의 차원에서 할당표집과 확률표집의 차이를 비교했고, 시뮬레이션 분석 2에서는 성별, 연령대, 이념을 고려하여 분석을 진행하였다. 시뮬레이션 분석을 통해 확률표집, 재접촉을 실시하는 확률표집에 비해 할당표집은 각 성별×연령대의 목표할당 표본 수를 채우기 위해 표본을 추출하는 과정에서 대표성이 왜곡될 수 있다는 것을 확인했다. 2020년 및 2024년 총선 여론조사 자료에 관한 분석에서는 할당표집과 낮은 협조율에서 비롯된 선택편향으로 인해 대표성에 관한 결과들이 일정하지 않게 나타난다는 것을 확인할 수 있었다. 여론조사 자료를 수집할 때 전수조사를 실시하지 않으면 선택편향이 발생할 수 있고 총선 여론조사와 같이 협조율이 매우 낮은 상황에서는 협조율이 소폭 증가하더라도 예측정확성이 반드시 향상되지는 않을 수 있는데, 이로 인해 나타난 결과로 사료된다.

현재 한국의 선거 여론조사에서 나타나는 문제점과 개선방안에 대해 정부의 규제, 조사비용, 조사방법의 세 가지 측면에서 검토하였다. 정부가 선거에 대해 규제 중심으로 접근하면서 선관위와 여론조사심의위의 엄격한 규정과 제재가 부과되고 있다. 선관위와 여론조사심의위의 노력들로 인해 선거 여론조사 품질이 일정 수준을 확보할 수 있었다. 그러나 그 이상의 품질 향상을 위해 노력하는 조사기관들은 많지 않은 실정이다. 한편, 선관위와 여론조사심의위의 규정과 제재가 형식적인 측면이 있다. 가령, 조사의 형식에 관한 규정과 제재가 주를 이루는 반면에, 조사의 내용에 관한 규정과 제재는 충분치 않다. 조사기관의 표집방법 및 표집과정에 관한 관리·감독이 이뤄지고 있지 않다 보니, 조사기관이 어떠한 조사방법을 어떻게 활용하고 있는지에 관해서는 관리가 안 되고 있다. 이러한 점들을 고려하여, 여론조사의 품질이 향상될 수 있는 방식으로 선거법과 여론조사 관련 규정들이 개정될 필요가 있다.

또한, 정치 영역에서 여론조사가 수행하고 있는 기능과 역할에 비해 그 가치를 충분히 인정받지 못하면서 여론조사 비용이 적게 책정되고 있으며 그 결과 여론조사의 품질을 높이기 어려운 측면이 있다. 현재 여론조사 시장은 진입이 쉬운 반면에 여론조사심의위의 규정을 충족하기만 한다면 퇴출이 어려운 상황이다. 여론조사심의위의 규정을 충족하지만 낮은 품질의 여론조사를 산출하는 조사기관들을 정부가 추려낼 수 없다면, 시장에서 높은 품질의 여론조사를 산출하는 조사기관들이 선택받고 그렇지 않은 조사기관들이 선택받지 않는 구조가 되어야 전반적인 여론조사의 품질이 향상될 수 있을 것이다. 그런데 업계 내부와 시장에서 여론조사 품질에 근거한 조사기관에 대한 평가가 이루어지고 있지 않으며, 오히려 수요자들 중 상당수가 여론조사 품질과 관계없이 자신이 원하는 결과가 산출되기를 원하면서 낮은 품질의 여론조사를 산출하는 조사기관들이 계속해서 선택을 받고 있다. 조사비용이 적은 수준으로 책정되면서, 조사기관별로 다양한 조사방법을 활용하기보다는 ARS와 같은 비용이 적게 드는 특정한 방식을 많이 활용하고 있다.

이와 함께 조사비용이 현실화되어야 한다. 여론조사가 공직후보자의 공천과 선거에 큰 영향을 미치고 있다는 점을 고려하여, 조사비용이 지금보다 많이 책정될 필요가 있다. 그리고 이를 위해서는 조사기관들의 여론조사들에 대한 평가가 선행되어야 한다. 낮은 품질의 여론조사를 산출하는 조사기관들이 시장에서 퇴출되도록 함으로써 조사기관들이 가격경쟁을 우선시하는 상황을 해결할 필요가 있다.

정부의 강한 규제와 적은 여론조사 비용으로 인해 현재 조사기관들은 선거 여론조사에서 거의 비슷한 조사방법을 활용하고 있으며, 특히 편향이 발생할 가능성이 큰 할당표집을 주로 활용하고 있다. 그러나 할당표집은 표본 측면에서 대표성 왜곡이 나타날 수 있고, 통계량 측면에서 검증가능성과 재현가능성을 확보하기 어려운 측면이 있다. 게다가 현재 선거 여론조사에 만연해 있는 낮은 협조율은 대표성 왜곡을 심화시킬 수 있으며, 특정한 성별×연령대의 협조율이 현저히 낮을 때에는 그러한 왜곡이 더욱더 심하게 나타날 수 있다. 현재 선거 여론조사 결과에서 같은 시기 같은 지역구에서 실시된 여론조사 결과의 차이 등은 이와 같은 할당표집과 낮은 협조율에 의한 대표성 왜곡으로 인해 나타난 결과일 가능성이 높다.

조사방법의 측면에서 확률표집을 적극적으로 활용할 필요가 있다. 확률표집은 비확률표집에 비해 표집오차가 매우 작게 나타나며 따라서 대표성을 제고할 수 있다. 협조율이 낮을 때에도 할당표집에 비해 대표성 왜곡이 작게 나타난다. 특히 재접촉을 실시할 경우에는 대표성 왜곡을 크게 줄일 수 있다. 조사기관들은 여론조사를 설계할 때 협조율을 고려해서 표본추출틀을 넉넉히 확보하는 방식으로 설계하고, 성별×연령대의 협조율이 크게 다를 수 있다는 점을 염두에 두어야 한다. 선관위와 여론조사심의위원회에서 응답률과 협조율에 관해 규제를 가하는 것은 각 여론조사가 계획대로 여론조사를 실시했음에도 규제의 기준을 충족하지 못했을 때 폐기를 하기 위한 것이지, 조사기관이 애초의 설계와 계획을 바꿔서라도 이 기준을 충족하기만 하면 된다는 것이 아니다. 과학적 방법에 근거한 확률표집을 제대로 실시한다면 대표성 왜곡을 줄일 수 있고 높은 품질의 여론조사를 산출할 가능성이 높아진다. 물론, 이처럼 충분한 표본크기를 확보하고 확보된 표본추출틀에 대해 확률표집을 활용하여 조사를 충실히 하기 위해서는 앞서 언급한 조사비용의 현실화가 선행되어야 한다.

확률표집을 선택한다면 검증가능성과 재현가능성을 확보할 수 있다는 점에서 선거가 반복적으로 실시됨에 따라 개선이 이뤄질 수 있다. 연장선상에서, 현재는 직접통계량(direct estimator)을 중심으로 지지율을 추정하는 방식을 주로 활용하고 있는데, 과학적으로 재현가능하고 검증가능하다면 조사품질의 향상을 위해 선진 기법에 기반한 다양항 통계량 작성이 장려되어야 할 것이다. 선거 여론조사로 인한 여론 왜곡을 최소화하고, 보다 신속하고 정확한 정보를 전달하기 위해서는 정부(선관위), 학계, 조사기관들의 지속적인 연구와 개발이 필요하며 정부와 선관위의 전문성을 지속적으로 확보하고 증대할 수 있는 방안을 제시할 필요가 있다.

## 참고문헌

- 강현철·한상태·김지연·정용찬·허명희. 2008. “RDD 전화조사와 주요결과.” 《조사연구》 9(1): 1-22.
- 고길곤·김대중. 2018. “선거 여론조사의 비관측 오차 영향요인에 관한 연구: 20대 총선 여론조사 결과를 중심으로.” 《조사연구》 19(1): 27-60.
- 곽은선·김영원. 2010. “2008 총선 출구조사의 총조사오차 분석.” 《조사연구》 11(3): 33-55.
- 곽은선·김영원. 2022. “현행 선거여론조사 방법의 정확성과 선거결과 예측 가능성.” 《조사연구》 23(1): 131-153.
- 구본상. 2018. “제7회 전국동시지방선거 여론조사 평가와 제안.” 《제7회 전국동시지방선거 선거여론조사 심의 백서》. 중앙선거여론조사심의위원회.
- 구본상·이재묵·이나경·이병재·김현. 2023. “선거여론조사 품질평가 활성화 및 신뢰성 제고 방안 연구.” 중앙선거여론조사심의위원회.
- 김선용. 2004. “이동전화 확산에 따른 유선전화 가구보유율의 변화: 한국을 포함한 주요 국가들을 중심으로.” 《조사연구》 5(1): 27-49.
- 김영원·황다솜. 2014. “2014년 지방선거 여론조사 전화조사방법에 따른 예측오차 및 편향.” 《조사연구》 15(4): 1-32.
- 김영원·윤지혜. 2016. “2016년 국회의원선거 전화여론조사 정확성 분석.” 《조사연구》 17(3): 109-137.
- 김정훈. 2022. “표본추출틀에 따른 선거여론조사의 정확성 평가에 대한 연구.” 《한국자료분석학회》 24(6): 2185-2198.
- 김지윤·강충구. 2014. “여론조사의 대표성.” 《평화연구》 22(2): 360-397.
- 박인서·백영민. 2021. “선거기간 설문조사의 응답률 변화 연구: 설문응답 모형 및 설문조사방식을 기반으로.” 《언론정보연구》 58(3): 5-45.
- 박인호·임중호·박민규. 2019. “선거여론조사의 객관성·신뢰성 제고를 위한 조사방법론 개선방안 연구.” 중앙선거여론조사심의위원회.
- 이경택·이화정·현경보. 2012. “유·무선전화 병행조사에 대한 연구: 2011년 서울시장 보궐선거 여론조사 사례.” 《조사연구》 13(1): 135-158.
- 이승희. 2010. “총조사 오차의 패러다임으로 이해하는 표본조사.” 《통계연구》 15(1): 44-74.
- 이준용. 2011. “선거 여론조사의 점검과 성찰: ‘판세분석 여론조사’를 중심으로” 한국언론학회 심포지움 및 세미나.

- 장덕현·이계오·홍영택. 2014. “RDD 전화여론조사의 무응답 편향 보정방법: 2014년 지방선거 여론조사 사례연구.” 《조사연구》 15(4): 33-60.
- 장덕현·조성겸. 2015. “2014 지방선거에서의 여론조사의 정확성 연구: 조사방법별 정확성과 편향 비교.” 《조사연구》 16(1): 129-153.
- 조성겸. 1997. “대통령 선거 여론조사와 할당표집 방법의 문제점.” 《언론과 사회》 18: 29-53.
- 조성겸·김지연·나윤정·이명진. 2007. “선거여론조사의 문제점과 개선방향: 2006년 지방선거 전화조사를 중심으로.” 《조사연구》 8(1): 31-54.
- 조성겸·장덕현. 2012. “선거시기 여론조사 정확성 제고 방안.” 《언론중재》 123: 20-29.
- 중앙선거여론조사심의위원회. 2022. 《2022년 양대선거 선거여론조사 백서》.
- 최종호·김태균·백다예·왕정하·한강욱. 2022. “조사방법 차이가 정치적 태도 조사결과에 미치는 영향: 중앙여심위 선거여론조사 통합자료 분석.” 《한국정당학회보》 21(3): 5-40.
- 최필선·민인식. 2013. “18대 대통령 선거 여론조사의 기관별 정확성 측정 및 비교.” 《조사연구》 14(2): 1-18
- 허명희·황진모. 2006. “전화조사를 위한 시간균형할당표본추출.” 《조사연구》 7(2): 39-52.
- 허순영·이수철. 2016. “전화 선거여론조사에서 무응답률 증가로 인한 편의와 응답률 제고 방안.” 《한국데이터정보과학회지》 27(2): 315-325.
- AAPOR/WAPOR. 2021. “Task Force Report on Quality in Comparative Surveys.”
- Arzheimer, Kai and Jocelyn Evans. 2014. “A New Multinomial Accuracy Measure for Polling Bias.” *Political Analysis* 22(1): 31-44.
- Biemer, Paul P. 2010. “Total Survey Error: Design, Implementation, and Evaluation.” *Public Opinion Quarterly* 74(5): 817-848.
- Biemer, Paul P. and Lars E. Lyberg. 2003. *Introduction to Survey Quality*. New Jersey: John Wiley & Sons.
- Callegaro, Mario and Giancarlo Gasperoni. 2008. “Accuracy of Pre-election Polls for the 2006 Italian Parliamentary Election: Too Close to Call.” *International Journal of Public Opinion Research* 20(2): 148-170.
- Converse, Jean. 1987. *Survey Research in the United States*. Berkeley, California: University of California Press.
- Crespi, Irving. 1987. *Public Opinion Polls and Democracy*. Boulder, Colorado: Westview Press.
- Eurostat. 2000. “Assessment of the Quality in Statistics.” *Eurostat General Standard*

- Report*. Luxembourg.
- Groves, Robert M. 1989. *Survey Errors and Survey Costs*. New York: Wiley.
- Groves, Robert M., Floyd Fowler, Mick Couper, Eleanor Singer, and Roger Tourangeau. 2004. *Survey Methodology*. New York: Wiley.
- Groves, Robert M. 2006. “Nonresponse Rates and Nonresponse Bias in Household Surveys.” *Public Opinion Quarterly* 70(5): 646-675.
- Groves, Robert M. and Emilia Peytcheva. 2008. “The Impact of Nonresponse Rates on Nonresponse Bias: A Meta-Analysis.” *Public Opinion Quarterly* 72(2): 167-189.
- Herbst, Susan. 1993. *Numbered Voices: How Opinion Polling Has Shaped American Politics*. Illinois. Chicago: University of Chicago Press.
- Keeter, Scott. 2012. “Presidential Address: Survey Research, Its New Frontiers, and Democracy.” *Public Opinion Quarterly* 76(3): 600-608.
- OECD. 2011. “Quality Dimensions, Core Values for OECD Statistics and Procedures for Planning and Evaluating Statistical Activities.” OECD.
- Mokrzycki, Michael, Scott Keeter, and Courtney Kennedy. 2009. “Cell-phone-only Voters in the 2008 Exit Poll and Implications for Future Noncoverage Bias.” *Public Opinion Quarterly* 73(5): 845-865.
- Newport, Frank. 2004. *Polling Matters: Why Leaders Must Listen to the Wisdom of the People*. New York: Warner Books, Inc.
- Panagopoulos, Costas. 2009. “Polls and Elections: Preelection Poll Accuracy in the 2008 General Elections.” *Presidential Studies Quarterly* 39(4): 896-907.
- Statistics Canada. 2009. “Statistics Canada Quality Guidelines.” Statistics Canada..
- Traugott, Michael W. 2005. “The Accuracy of the National Preelection Polls in the 2004 Presidential Election.” *Public Opinion Quarterly* 69(5): 642-654.
- United Nations Economic Commission for Europe (UNECE). 2003. “Definition of Quality in Statistics.” in *Handbook of Statistical Organization*, Third Edition: The Operation and Organization of a Statistical Agency. United Nations.
- Verba, Sidney. 1996. “The Citizen as Respondent: Sample Surveys and American Democracy.” *American Political Science Review* 90(1): 1-7.

## **Survey Quality and Representativeness in Election Polls: A Comparative Analysis of the 21th and 22nd Legislative Elections in South Korea**

In-Kyun Kim  
(Yonsei University)  
Seung-Joo Lee  
(Yonsei University)  
Jong-Ho Im  
(Yonsei University)

This study examines the challenges in current election polling practices in South Korea and explores potential improvements. In South Korean election survey, data collection primarily relies on quota sampling, which introduces sampling errors and distorts representativeness. The nature of quota sampling also complicates efforts to ensure verifiability and reproducibility, hindering improvement. Strict regulations by the National Election Commission and limited survey budgets further constrain available methodologies, while low response rates intensify representativeness distortions. To address these issues, this study employs simulation analyses to compare quota sampling with the scientific approach of probability sampling. Simulation Analysis 1 investigates how differing response rates affect quota sampling versus probability sampling, while Simulation Analysis 2 assesses representativeness across quota sampling, probability sampling, and re-contact probability sampling under similar response rate conditions. Additionally, survey data from the 21st legislative election held on April 8, 2020, and the 22nd legislative election on April 10, 2024, were analyzed with a focus on predictive accuracy and bias. Findings indicate that quota sampling produces greater representativeness distortion than probability sampling, that low response rates exacerbate this distortion, and that quota sampling lacks reproducibility.

Key words: election poll, survey quality, reponse rate, the 21st and 22nd legislative election