

2020년 중앙선거여론조사심의위원회 연구용역 결과보고서

# 지난 선거 투표 응답과 선거여론조사 편향의 관계연구

2020. 8

성균관대학교 산학협력단

본 연구보고서는 2020년도 중앙선거여론조사심의위원회 연구용역과제로서 연구내용은 중앙선거여론조사심의위원회 공식견해가 아님.

2020년 중앙선거여론조사심의위원회 연구용역 결과보고서

# 지난 선거 투표 응답과 선거여론조사 편향의 관계연구

---

2020. 8

본 연구보고서는 2020년도 중앙선거여론조사심의위원회 연구용역과제로서  
연구내용은 중앙선거여론조사심의위원회 공식견해가 아님.

**성균관대학교 산학협력단**

**과제명: 지난 선거 투표 응답과 선거여론조사 편향의 관계연구**

**사업추진단체: 성균관대학교 산학협력단**

**연구진:** 김지범 (성균관대학교 사회학과 교수)  
장덕현 (한국갤럽조사연구소 연구위원)  
김솔이 (성균관대학교 서베이리서치센터 연구원)



# CONTENTS

제1장 연구의 배경과 목적 .....	1
1.1 연구의 배경 .....	1
1.2 연구의 목적 .....	3
1.3 연구의 구성 .....	4
제2장 기존연구 검토 .....	5
2.1 전반적 조사오류 틀(Total Survey Error Framework) .....	5
2.2 투표 참여(Turnout) 연구 .....	8
2.3 투표 선택(Vote Choice) 연구 .....	10
제3장 연구 자료 및 특성 .....	13
3.1 자료 .....	13
3.2 자료 특성 .....	15
제4장 결과 .....	17
4.1 13대~18대 대선 투표율 .....	17
4.2 2017년 19대 대선 이후 지난 대선 후보 투표 .....	19
4.3 2012년 18대 대선 투표 후보 회상 .....	23
4.4 2016년 20대 총선 지역구 투표 후보 회상 .....	24
4.5 역대 대선 지난 대선 후보 투표 .....	28
제5장 선거여론조사 편향성 기준 및 제도개선 방안 .....	31
5.1 결과 요약 .....	31
5.2 편향성 기준 및 개선 방안 .....	32
참고문헌 .....	41

## 표목차

<표 1> 전반적 조사 오류틀로 본 선거조사 오류 - 대표성 오류와 측정 오류 .....	7
<표 2> 주요 해외 연구의 지난 선거 투표 회상과 실제투표 차이 비율 (%) .....	11
<표 3> 대선 투표 참여 여부 및 대선 후보 투표율 - 한국갤럽, 한국종합사회조사 (%) .....	18
<표 4> 2017년 19대 대통령 선거 투표 회상 - 21대 총선 출구조사 투표자 심층여론조사 (%) .....	19
<표 5> 2017년 19대 대통령 선거 투표 회상 - 한겨레신문-한국리서치, SBS-칸타코리아 결과 (%) .....	20
<표 6> 2017년 19대 대통령 선거 투표 회상과 실제 득표율 차이 - 한겨레신문-한국리서치, SBS-칸타코리아 결과 (%) .....	21
<표 7> 사전 선거여론조사와 출구조사의 대통령 선거 투표 회상 비교 (%) .....	22
<표 8> 2012년 18대 대통령 선거 투표 회상 - 리서치뷰 결과 (%) .....	23
<표 9> 2012년 18대 대통령 선거 투표 회상과 실제 득표율 차이 - 리서치뷰 결과 (%) .....	24
<표 10> 2016년 20대 국회의원 선거 지역구 투표 회상 - 2020년 21대 총선 선거여론조사 중앙일보-입소스 (%) .....	25
<표 11> 2016년 20대 국회의원 선거 지역구 투표 회상과 실제 득표율 차이 - 2020년 21대 총선 선거여론조사 중앙일보-입소스 (%) .....	26
<표 12> 사전 선거여론조사의 정확성 분석 - 2020년 21대 총선 선거여론조사 중앙일보-입소스 (%) .....	27
<표 13> 1987~2002년 대통령 선거 투표 회상 - 1992~2007년 한국갤럽 사전 선거여론조사 (%) .....	28
<표 14> 1997~2002년 대통령 선거 투표 회상 응답별 후보 지지도 - 한국갤럽 사전 선거여론조사 (%) .....	29
<표 15> 1997~2002년 대통령 선거 후보 지지도 정확성 비교 - 한국갤럽 사전 선거여론조사 (%) .....	30
<표 16> 표본 대표성 확보를 위한 절차, 과제 및 오차 .....	36
<표 17> 2016년 총선 선거여론조사 주요 조사 특성과 오차 간 상관관계 분석(Pearson's r) · .....	37

## 그림목차

[그림 1] 중앙일보 기사 .....	2
----------------------	---

## 제1장 연구의 배경과 목적

### 1.1 연구의 배경

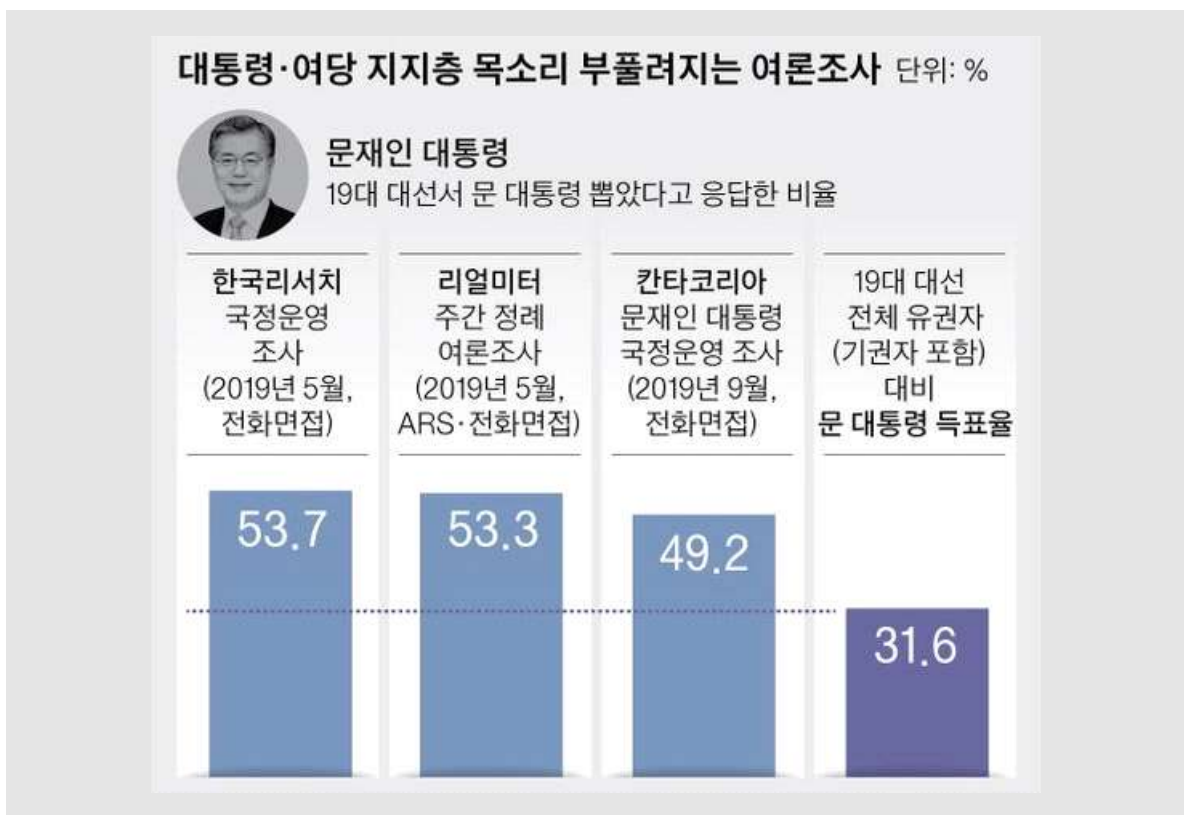
한국갤럽에서 최초로 1987년 대통령 당선자 예측, 1999년 출구조사가 도입된 후, 여론조사는 “여론조사 민주주의”(지병근 2010), “여론조사 공화국”(김준철 2015: 13), “여론전쟁 시대”(현경보 2019: 12)라고 말할 정도로 정당 후보자 선출 혹은 정책 결정에 심대한 영향을 끼치고 있다. 동시에, “여론조사공해”(홍영림 외 2012)라고 말할 정도로 여론조사결과가 미디어에 넘쳐나고 있다. 예로 중앙선거여론조사심의위원회 홈페이지에 등록된 19대 대선관련 조사는 1월 25일부터 5월 9일까지 293회인데, 이것은 18대 대선 관련 조사보다 100회 이상 증가한 것이다(윤희웅 2017). 반면, 조사환경은 점점 더 어려워지고 있고(김지범 외 2019), 조사결과에 대한 정치권의 신뢰와 불신이 이해관계에 따라 변하고 있다.

2020년 국회의원 총선거 이전 일부 언론 등에서 대통령 직무평가와 여당 지지율이 높은 것은 진정한 여론과 다르며, 선거여론조사에 편향에 있다는 주장이 제기되었다. 일부 언론은 선거이후 조사에서 물어본 ‘지난 선거에서의 투표 후보·정당’에 대한 응답 결과와 실제 투표결과와의 차이를 편향의 근거로 제시하였다. 현재의 여론조사가 정부 여당에 편향되어 있으며, 그 증거로 문 대통령 투표층이 실제보다 표본에 많이 포함된다는 ‘문 대통령 투표층 과대 표집’을 거론하였다.

대표적으로 2019년 11월 5일 중앙일보 ‘수상한 여론조사 응답자 절반이 문 투표층이었다’ 제하의 기사에서는 전체 유권자 중 2017년 문재인 후보 투표자는 31.6% (전체 유권자 기준, 득표율은 41.1%)인데 반해 주요 여론조사에서 문재인 투표자는 50%가 넘는 것이 편향의 증거라고 주장했다([그림 1]).

“여론조사를 둘러싼 논란 중 대표적인 것이 문 대통령 투표층의 ‘과대 표집’(특정 집단의 여론이 실제보다 부풀려 수집) 현상이다. (중략) 응답자 중 2017년 대선에서 문 대통령을 뽑았다는 응답자는 전체 응답자 1,000명 중 537명(53.7%)이었다. 하지만 실제로 2017년 대선에서 전체 유권자(기권자 포함) 대비 문 대통령의 득표율은 31.6%(4247만9710명 중 1342만3800표)였다.”

이 기사에서는 성별, 연령대별, 지역별 유권자의 인구 구성 비율뿐 아니라 ‘대선득표율에 따라 결과를 보정하는 여론조사는 드물다’고 하며 통계학자를 인용하여 ‘응답자의 정치성향을 고려하지 않은 조사는 왜곡된 결과로 이어질 수 있다’고 지적하여 대선득표율에 따라 할당하거나 보정하는 접근이 필요함을 말하고 있다.



[그림 1] 중앙일보 기사



또한 조선일보의 2020.3.27. ‘여론조사 표본은 입맛대로?’ 기사에서도 비슷한 논점으로 다음과 같이 여론조사의 신뢰성에 의문을 제기하고 있다:

“이 조사에서는 지난 대선 때 투표한 후보를 묻는 항목에 응답자 66%가 문재인 후보라고 답했다. 지난 대선에서 종로구의 투표율(77%)과 문 후보 득표율(42%)을 감안하면 여론조사 표본에 문 후보 지지자가 32%만 포함돼야 하지만, 이 조사에선 두 배 이상 많았다.”

이와 관련해서 선거여론조사 설문문항에 ‘최근 실시된 전국단위 선거에서 정당지지에 관한 사항이 반드시 포함되도록 하는 법안’(왜곡 여론조사 방지법)도 발의되기도 하였다.

언론과 정치권의 비판과 함께 관련 법안이 구체화됨에 따라, 지난 선거에서의 투표 응답이 선거여론조사의 편향을 나타낼 수 있는지에 대한 학술적이고 경험적인 연구의 필요성이 대두되었다. 여당이 압승한 21대 총선거 이후 이러한 비판은 과거보다 잦아 들었으나 여전히 일부 특정 정파, 지지층 사이에서는 재생산되고 있어, 향후 혼란을 최소화하기 위해 이 주제에 대한 깊이 있는 연구 필요성이 크다고 할 수 있다.

## 1.2 연구의 목적

본 연구는 한국에서 거의 관심의 대상이 아니었던 지난 선거에서의 투표 응답과 선거여론조사 편향과의 관계를 파악하려고 한다. 구체적으로, 첫째, 학술적 논의와 경험적 자료를 탐색하여 지난 선거 투표 참여(turnout) 정도와 투표 선택(vote choice) 응답과 선거여론조사 편향과의 관계를 규명하려고 시도한다. 둘째, 지난 선거 투표 응답 외에 선거여론조사의 편향을 파악하기 위한 기준과 방법을 논의하고자 한다.

연구 목적과 관련하여 두 가지 연구 질문에 대한 답을 찾으려고 한다. 첫째, 선거 이후 조사에서 물어본 지난 선거 투표 참여 와 투표 선택 응답이 실제 결과와 다른

현상의 이유는 무엇인가? 즉, 지난 선거 투표 후보 선택과 실제 선거결과 득표율의 차이는 표본 대표성의 문제를 가져오는 무응답오류인가, 아니면 기억 오류 및 사회적 바람직성 편향(social desirability bias)과 연관된 측정의 문제인가? 둘째, 선거 여론조사의 편향을 파악하기 위한 기준과 방법이 있는가? 선거여론조사의 편향성 근거로 참고할 만한 기준은 무엇인가? 편향을 방지하기 위한 조사 절차는 어떠해야 하는가?

### 1.3 연구의 구성

본 연구는 전체 5개의 장으로 구성되어있다. 연구 배경과 연구 목적을 설명한 제1장 다음으로는 제2장에서 전반적 조사오류 틀(total survey error framework)로 선거조사의 오류, 국내외 투표 참여와 투표 선택 연구를 기술한다. 제3장에서는 본 연구에서 이용한 다양한 자료에 대한 설명과 특성을 언급한다. 제4장은 3장에서 언급한 다양한 자료를 분석한 결과를 제시하는데, 지난 여러 번의 대선을 따로 나누어서 보여주기도 하고, 특정 연도의 대선을 여러 자료를 이용하여 비교하기도 한다. 그리고 마지막 제5장에서는 분석 결과를 요약하고 편향성 기준 및 개선 방안을 제시한다.

## 제2장 기존연구 검토

### 2.1 전반적 조사오류 틀(Total Survey Error Framework)

전반적 조사오류 틀은 조사에서 나타나는 다양한 종류의 오류를 체계적으로 파악하여 조사의 오류를 줄이려는 접근 방법으로 조사 설계부터 조사 품질 평가까지 모든 조사 단계에 유용한 지침을 제공한다(Groves 2004; Weisberg 2009). 조사 오류는 실제 값과 조사에서 측정된 값의 차이로 조사의 전 과정에서 발생한다. 조사의 과정은 조사의 설계, 실사, 자료화와 분석의 3단계로 요약할 수 있는데, 각 단계마다 오류가 발생할 수 있다.

조사 설계는 모집단을 정의하고 표본을 설계하며, 조사내용, 설문지를 구성하는 단계이다. 실사는 조사방법에 따라 조사대상자의 접촉과 참여 요청, 조사자의 질문에 응답자가 응답하는 과정이다. 자료화와 분석, 해석의 단계에는 자료 입력 및 정선작업, 통계표 집계, 결과 분석을 통해 보고서를 작성한다.

전반적 조사오류는 오류 발생 원인에 따라 크게는 표집오류(sampling error)와 비표집오류(non-sampling error)로 구성된다. 표집오류는 모집단이 아닌 표본을 사용한다는 점에서 발생한다. 비표집오류는 포함 오류, 무응답 오류, 측정 오류로 구분될 수 있다.

포함오류(coverage error)는 표본을 선정하는 표집틀의 정확성에 기인하며, 무응답 오류(nonresponse error)는 표본으로 선정된 응답자의 참여 여부 및 참여한 응답자의 항목 무응답과 연관되어 있다. 측정 오류(measurement error)는 측정하려는 것이 정확한 측정이 아닌 경우 발생하는데 응답자뿐만 아니라 조사원의 영향을 받을 수 있다.

표본조사의 특성상 표집오류 및 비표집오류의 발생은 불가피하다. 그러나 확률추출이라는 전제 하에서 전반적 조사오류 중 표집오류만 측정 가능하기 때문에 현재의 선거여론조사는 표집오류라는 제한적인 정보만 제공한다. 표집오류에 비해 비표집오류는 다양한 요인에 의해 발생하며, 다양한 오류 요인들이 조사의 설계, 실사, 분석의 조사 전 과정과 관련이 있다. 따라서 조사의 정확성 개선을 위해서는 비표집오류까지 포함한 전반적 조사 오류를 줄이기 위해 노력해야 한다.

조사에서 발생할 수 있는 오류는 표본 대표성의 차이를 발생시키는 것과 표본의 대표성에 영향을 주지 않지만 응답 결과에만 영향을 주는 것으로 구분할 수 있다. 먼저 조사의 설계과정에는 포함 오류, 표집 오류, 설문지 구성에 따른 오류가 발생할 수 있다. 설문지 구성은 문구나 맥락의 선택, 질문의 위치, 항목의 순서, 중간 항목이나 모름/응답거절 제시 등이 조사 응답에 영향을 준다. 조사과정에서 무응답 오류, 조사모드에 따른 오류, 조사원 효과, 응답자의 지식수준과 오해, 사회적 바람직성(social desirability), 만족화(satisficing)와 같은 심리적 기제에 의한 오류, 기억력에 따른 오류 등이 발생할 수 있다. 마지막으로, 자료화, 분석 단계에서는 단순한 코딩, 편칭, 정선작업, 프로그래밍 오류부터 해석의 오류까지 다양한 오류 가능성이 있다.

자료수집방법(면접조사, 전화조사, 우편조사, 인터넷 조사)은 표본과 측정결과에 모두 영향을 주는 자료수집방법효과(mode effect)를 가져온다. 즉 어떤 조사방법을 택하는가에 따라 응답자 표본이 달라지거나 조사방법에 따라 동일 응답자의 응답이 달라질 수 있다. 이런 현상이 한 가지 이상의 조사수집방법을 이용하는 혼합조사(mixed mode)에서는 선택효과와 측정효과로 구분된다.

<표 1> 전반적 조사 오류들로 본 선거조사 오류 - 대표성오류와 측정오류

전반적 조사 오류 틀	선거조사 오류 내용
포함오류	선거 유권자가 추출틀에 정확하게 포함되었는가?
표집오류	표집이 선거 유권자를 제대로 반영했는가?
무응답오류	비투표자는 조사에 참여하지 않는가? 당선자에게 투표하지 않은 사람은 조사에 참여하지 않는가?
측정오류	비투표자는 투표했다고 거짓말을 하는가? 당선자에게 투표하지 않은 사람이 당선자에게 투표했다고 응답하는가?

<표 1>에서 보듯이, 선거조사의 오류는 전반적 조사오류의 양대 축인 대표성 오류 (포함오류, 표집오류, 무응답오류)와 측정 오류로 파악할 수 있다. 사후선거조사에서 물어본 지난 대선 투표 참여와 투표후보 선택 응답이 실제 과거 선거결과와 차이가 나는 것은 한편으로는 포함률이나 무응답오류 때문에 조사된 표본 내에 여당 지지자가 과도하게 많은 것일 수도 있고, 다른 한편으로는 표본 구성의 문제가 아니라 과거 대선 투표 회상 응답에 사회적 바람직성, 기억력 등 심리적 과정이 연관되어 응답이 다르게 나타난 것일 수 있다.

만약 표본이 왜곡되었다면 현재의 선거여론조사의 표집틀과 조사방법과 조사대상자 접근방식에 근본적인 문제가 있으므로 대대적인 개선이 요구된다. 여당 지지자를 실제와 유사하게 축소하는 것이 편향을 보정하는 방법이 될 것이다. 반면, 과거 회상과 관련된 기억력 문제 혹은 사회적 바람직성에 따른 응답과 같은 단지 측정의 문제라면 이 질문의 특성을 이해하고 측정오류를 줄일 수 있는 문항을 개발하는 것이 필요하다고 볼 수 있다. 측정의 문제라면 이 질문 응답에 맞추어 여당 지지자를 축소하는 것은 현재의 정치 지형을 왜곡할 수 있다.

## 2.2 투표 참여(Turnout) 연구

### 2.2.1 해외 연구

투표율의 부정확성은 투표자와 비투표자의 특성 및 후보자 선택과 선거 결과 등의 경험적 연구에 잘못된 결론을 가져올 수 있기 때문에 선거 연구에서 중요한 주제로 다루어져 왔다(DeBell et al. 2018).

사후 선거 조사에서 틀리게 응답하는(misreporting) 경우는 투표했지만 투표하지 않았다고 대답하는 과소응답(under reporting)과 투표하지 않았으나 투표했다는 과대응답(over reporting)으로 구분되는데, 대부분의 틀린 보고는 과대응답이며, 투표하고 하지 않았다는 경우는 매우 적다(Silver, Anderson and Abramson 1986). 예로, 실제 투표율과의 차이는, 과소응답은 2%p 정도 내외이며(Belli et al. 2001), 과대응답은 13~17%p이다(Berent et al. 2016).

조사응답자의 투표 질문 응답과 실제 투표 등록 기록을 교차 검증하는 투표검증연구(vote validation studies)에서는 응답자의 투표 진위여부를 확인하여 과대응답에 기인한 사후선거조사 투표율과 실제 투표율의 차이를 투표격차(turnout gap)로 설명하려는 연구가 많이 진행되고 있다(미국: Sigelman 1982; Silver, Anderson and Abramson 1986; Belli et al. 2001; 네덜란드: Voogt and Saris 2003; 스위스: Sciarini and Goldberg 2017; 덴마크: Dahlgaard et al. 2019). 물론, 미국 연구에서 조사응답자를 투표 등록부와 연결할 때 연결 절차에 따른 오류(matching errors)의 문제가 지적되기도 한다(Berent et al. 2016).

응답자의 특성뿐만 아니라 과대응답을 가져오는 심리적인 요인에 대한 연구가 진행되었다. 과대응답에 대한 세 가지 동기 요인이 알려지고 있다. 첫째, 사회적 바람직성(social desirability)은 다른 사람들에게 긍정적인 모습을 보여주려는 의도 혹은 다른 사람으로부터 사회적 승인을 얻으려는 필요에 의해서 응답 범주를 선택하는 것을 의미한다(Ansolabehere and Hersh 2012). 둘째는 투표해야 하는 압력을 많이

받는 사람일수록 투표하지 않은 경우에 죄책감을 느끼게 되므로 거짓으로 투표했다고 말하는 경향이 높다(Bernstein et al. 2001). 셋째는 기억 망각과 혼동이다. 주요 연구에서는 선거일부터 시간이 지나 조사가 진행될수록 과대투표가 증가하는 것을 보여주고 있는데(Abelson et al. 1992; Belli et al. 2001; DeBell et al. 2018), 조사 투표 참여 응답이 이전 투표 참여 혹은 참여 의도와 실제 참여의 혼동의 결과일 수 있다.

과대응답이 측정오류의 문제뿐만 아니라 투표자가 사후선거조사에 더 참여하여 과잉대표(overrepresentation)되는 무응답오류에 기인한다고도 알려져 있다(Dahlgaard et al. 2019). Sciarini and Goldberg (2017)는 스위스의 공식적 투표 등록부를 표집틀로 이용하여 조사응답과 공식적 기록을 비교하였는데 조사에서 투표자의 과잉대표가 응답자가 과장응답한 것보다 투표격차를 더 설명한다는 점을 밝히고 있다.

### 2.2.2 국내 연구

국내 연구에서 조사된 투표 참여율이 실제 투표율보다 높게 나타나는 과대투표와 관련된 연구가 거의 없다(이현우, 전시홍 2010). 이준한(2005)의 연구에서는 1997년과 2002년 대선에서의 실제 투표율과 사회과학데이터센터의 선거 후 조사(선거일 다음 날부터 일주일 안에 조사 완료)의 설문 투표율을 성, 연령, 지역, 거주지에 따라 비교 분석하여 두 연도에서 성별에 따른 일관성 있는 차이는 없었고, 연령대 20~29세, 거주지로는 면지역에서 일관적으로 가장 높게 과대보고(overreporting)하는 경향이 있음을 밝히고 있다. 강신구(2017)는 2017년 대선 패널 자료를 이용하여 TV 토론이 지지후보에 대한 투표, 평가 등의 영향을 살펴본 연구에서 2차 패널 자료의 투표참여율이 98.5%로 실제투표율 77.2%와의 차이를 언급하면서 분석결과의 해석에 주의할 필요가 있음을 강조하고 있다.

이상의 논의를 통해 여론조사에서 투표 참여 응답이 과대 응답되는 경향이 있음을 설명했다.

## 2.3 투표 선택(Vote Choice) 연구

### 2.3.1 해외 연구

투표와 관련해 해외 주요 연구들은 시간에 따른 기억의 왜곡에 관심을 뒀으며, 이를 위해 반복적으로 동일한 조사 대상자의 응답을 확인할 수 있는 패널 데이터를 이용했다. 특히 투표 후보 정당 응답의 비일관성과 편향성을 확인하고, 그 원인을 규명하고자 했다(Himmelweit et al. 1978; Waldahl and Aardal 2000; Elsas et al. 2014; Durand 2015; Dassonneville and Hooghe 2017).

주요 해외 연구들의 지난 선거 투표 회상 응답과 실제 투표결과의 차이는 최소 8.7~28.9%p로, 회상 응답이 부정확하고 신뢰할 수 없음을 보고하고 있다. <표 2>에서 보면 주요 연구의 평균 차이는 20.1%p이며, 대체로 10~30%p 범위임을 알 수 있다.

더 중요한 점은 투표 후보/정당 회상 응답의 편향성으로, 주요 연구들은 회상 응답이 무작위적이지 않으며, 편향이 있음을 밝히고 있다. 편향의 방향성은 현재의 선호로 과거 선거 시점보다 현재 지지도가 높은 정당/후보에 응답하는 경향이 있었다.

비일관성 회상의 주요 요인은 유권자의 변동성, 망각, 인지적 일관성 유지 등이 지적되고 있다(Himmelweit et al. 1978; Waldahl and Aardal 2000; Elsas et al. 2014; Durand 2015; Dassonneville and Hooghe 2017).

이 중 유권자의 변동성(volatility)은 비일관성 회상의 가장 큰 원인으로 지적되고 있다. 한 정당, 정파에 꾸준히 투표한 유권자가 아닌 선거 때마다 다른 정당, 정파를 선택한(또는 투표와 기권을 반복한) 유권자에게 비일관성 응답이 많았다.



<표 2> 주요 해외 연구의 지난 선거 투표 회상과 실제투표 차이 비율 (%)

연구 논문	국가	회상연도	투표연도	회상연도 - 투표연도	투표 회상 - 실제 투표	비고
Weir (1975)	미국	1958	1956	2	8.7	
	미국	1960	1956	4	11.4	
Himmelweit et al. (1978)	영국	1974	1970	4	26	Himmelweit Longitudinal Study
	영국	1974	1966	8	28	
	영국	1970	1966	4	16	
	영국	1970	1964	6	24	
	영국	1970	1966	4	11	Butler and Strokes panel
	영국	1970	1964	6	18	
Waldahl and Aardal (2000)	노르웨이	1981	1977	4	18	
	노르웨이	1985	1981	4	16	
	노르웨이	1989	1985	4	20	
	노르웨이	1993	1989	4	21	
	노르웨이	1997	1993	4	27	
Dassonneville and Hooghe (2017)	벨기에	2014	2009	5	32.9	
	네덜란드	2003	2002	1	9.5	
	독일	2005	2002	3	28.9	
Durand (2015)	캐나다	2006-2011	2004	4	21	
	퀘벡	2008-2014	2007	5	25	
			평균	4.2	20.1	
			표준편차	1.5	7.2	

Waldahl and Aardal (2000)은 노르웨이 선거연구를 통해 투표 참여, 정당 투표 행동을 결합해 유권자를 안정적 투표자와 투표 정당 변동자, 지난 선거 비투표자, 이번 선거 비투표자, 안정적 비투표자 다섯 개 집단으로 구분하고, 1977년부터 1993년까지 이들 집단의 회상 오류를 분석한 결과 안정적 투표자의 오류 비율은 3~7%인 반면, 투표 변동자와 지난 선거 비투표자의 오류 비율이 34~44%, 57~75%에 달했다.

과거 투표 회상 응답을 신뢰할 수 없는 이유 중 하나는 기억력, 망각 때문이다. 선거 행동 자체가 어떤 유권자에게는 관심도가 적을 수 있으며, 또는 정치 상황에

따라 표심이 변했기 때문에 명확히 기억할 수 없는 사건일 수 있다. 투표 시점으로 부터 시간이 지날수록 응답 시점의 정치 태도와 일관성을 유지하려는 영향이 더 커지는 이유로(Elsas et al. 2014), 이는 시간 경과를 통한 망각이 다른 특성들과 결합해 회상 응답의 영향을 준다는 점을 보여준다.

회상 응답에는 사회적 바람직성이 영향을 줄 수 있다. 먼저 투표 참여가 바람직한 행동이므로 투표하지 않았으나 투표했다고 응답하며, 조사시점 선호도가 낮은 소수/포플리즘 정당 투표자는 응답을 회피할 수 있다(Himmelweit et al. 1978; Durand 2015). 현재 선호 정당/후보로 회상 응답이 편향되는 이유는 불명확한 기억을 현재의 선호에 일치시킴으로써 부조화를 해소하거나 자아 일관성을 유지하려는 심리 기제에 의한 것으로 설명된다(Waldahl and Aardal 2000; Elsas et al. 2014).

만약 표본 대표성이 낮은 근거가 지난 선거 회상 응답의 편향이라면 회상 응답을 실제와 유사하게 보정하는 것이 표본 대표성을 개선할 수 있는 방법이 될 것이다. 그러나 기존 연구에 따르면 지난 회상 응답을 통한 보정은 미미한 효과를 가질 뿐으로 실제 결과를 개선하지 못했다(Durand 2015).

### 2.3.2 국내 연구

국내에서 지난 선거 투표에 관한 연구는 찾아보기 힘들다(이계오, 장덕현 2011). 이계오와 장덕현(2011)은 2010년 지방선거에서 선거 전후 조사를 통해 사전 투표 의향과 사후 행동 응답의 차이를 비교하였다. 이 연구에서는 사전, 사후 응답의 일관성 있음을 발견했으나 사후 조사가 선거 직후 이루어져 기억 회상에 대한 연구로 적합하지 않고, 온라인 패널 대상의 조사였다는 한계가 있었다.

해외 주요 연구와 같이 실제 선거 투표와 회상 응답의 차이를 살피거나, 여러 시점에서 동일 응답자의 회상 일관성을 파악하기 위해서는 동일 응답자를 반복 조사한 패널 자료가 필수적이지만 국내에 이와 적합한 자료가 축적된 사례를 발견할 수 없었다.

## 제3장 연구 자료 및 특성

### 3.1 자료

본 연구에서는 공개되거나 공개되지 않은 중요한 자료를 많이 이용한다. 투표 검증 자료가 없기 때문에 연구 목적에 잘 부합하는 자료를 선정하는 것이 필수적이고 제시된 결과에 대한 해석의 오류를 줄일 수 있는 초석이다. 또한, 자료의 특성 및 보정에 대한 상세한 기술을 통해 제시하는 표에 대한 이해를 돕고자 한다.

#### 가. 1992~2007년 한국갤럽 대선 예측 선거여론조사

한국갤럽은 1987년 대통령 선거의 직선제 도입 이후 대선 사전여론조사를 실시했으며, 조사 결과에 기초해 득표율 예측치를 발표해왔다. 1992년, 1997년, 2002년, 2007년 4회는 지난 대선 투표 후보 회상 질문을 포함하고 있다.

#### 나. 한국종합사회조사(KGSS)

2003년부터 수행된 한국종합사회조사는 다단계 지역 확률 표집(multistage area probability sampling)방법에 따라 선정된 가구에서 한 명의 만 18세 이상 가구원을 대상으로 한 면접조사이다(김지범 외 2019). 2003년, 2008년에는 대통령 선거 투표 참여 여부만 물어 보았고, 2010년, 2013년, 그리고 2016년에는 참여 여부와 후보자에 대한 투표 질문을 물어 보았다. KGSS의 응답률은 2003년에 66%, 2008년에 60%, 2010년에 63%, 2013년에 52%, 2016년에 46%이다.

다. 2020년 21대 MBC-코리아리서치 총선 투표자 심층여론조사

4월 15일 출구조사 응답자 중 일부를 무작위 추출하여 총선 여론과 관련된 다양한 사항을 추가 질문하는 형식으로 이루어졌는데 3,142명의 투표자가 참여하였다. 설문 내용은 후보 선택 이유, 여야 심판론 공감도, 준연동형비례제·비례 위성 정당에 대한 의견, 대통령 직무평가 등이 있었으며, 2017년 19대 대선 투표 질문이 포함되었다. 출구조사는 다른 선거여론조사가 유권자 전체를 대상으로 조사하는 것과 다르게 투표자를 대상으로 한 유일한 조사이다. 총선 비례대표 투표 정당에 대한 심층여론조사 결과와 실제 득표율을 비교했을 때, 미래한국당에 대한 조사 결과는 34.2%(실제 득표율은 33.9%), 더불어민주당에 대한 조사결과 투표율은 33.3%(실제 득표율은 33.3%)로 조사결과의 정확성이 매우 높음이 확인되었다.

라. 2017년 이후 중앙선거여론조사심의위원회 홈페이지 등록된 선거여론조사 중 전국조사에서 2017년 19대 대선 투표 회상 응답을 포함한 조사

2017년 5월부터 2019년 5월까지 4회 조사한 한겨레신문-한국리서치와 2018년 2월부터 2019년 11월까지 5회 조사한 SBS-칸타코리아의 선거여론조사다. 9회 조사는 동일 의뢰처-조사기관일 때 대체로 동일한 조사 설계, 질문, 자료수집방법 등을 이용하여 비교 분석에 장점이 있다.

마. 2013년 이후 중앙선거여론조사심의위원회 홈페이지 등록된 선거여론조사 중 전국조사에서 2012년 18대 대선 투표 회상 응답을 포함한 조사

2016년 1월부터 2017년 4월까지의 리서치뷰의 조사는 자동응답 방식(ARS)의 조사로 무선전화 RDD 100%로 진행되었다.

바. 2020년 이후 중앙선거여론조사심의위원회 홈페이지 등록된 총선 선거구 여론조사 중 2016년 20대 총선 투표 회상 응답을 포함한 조사

2020년 21대 총선 선거여론조사 중 중앙일보-입소스가 실시한 10개 조사에 2016년 당시 지역구 투표 후보의 회상 질문이 포함되었다. 21대 총선 선거여론조사는 이 동통신 3사에서 제공하는 가상번호를 사용할 수 있었으며, 여론조사가 선거의 판세를 전반적으로 잘 반영한 것으로 평가된다.

### 3.2 자료 특성

자료들의 특성 및 보정 작업은 아래와 같다.

가. 선거예측을 위한 투표자 대상 출구조사는 투표자 모집단에 대해 절차상, 또는 실질적으로 대표성을 가지고 있으며, 현 시점 투표한 후보와 정당 응답이 포함되어 있기 때문에, 이를 통한 예측은 대체로 정확하다. 출구조사의 대표성을 고려할 때, 출구 조사의 결과와 과거 실제 선거 결과의 차이는 표본의 문제라기보다는 이 질문의 응답 편향, 기억의 측정 문제로 생각할 수 있다.

나. 지난 선거 후보 투표 응답과 다른 정치 지표의 추세도 고찰한다. 기존 연구들에 따르면 현재의 정치적 지지 성향에 지난 선거 투표의 기억을 일치시키려는 경향이 있으며, 이때 정치적 태도의 변동성이 영향을 주는 것으로 나타나고 있다. 현재 정치적 태도가 지난 선거 투표 회상 응답에 영향을 준다고 볼 수 있다. 따라서, 다른 정치지표와의 변화에 따라 지난 선거 투표 응답이 어떻게 변하는지를 파악할 것이다. 다만 이러한 접근은 회상 응답의 정확성을 파악할 수 없으므로 관계에 대한 결론을 도출할 수는 없다. 지난 선거 투표 응답이 다른 정치 지표의 변화에도 차이 나지 않는다면 이는 표본의 안정성을 나타내는 것으로 생각할 수 있다. 반대로 표본이 변화를 반영하지 못해 정치적 변동성을 과소 추정한다고 볼 수도 있다. 추세를 살피는 것만으로는 관계성에 대한 결론을 도출할 수 없다.

다. 지난 선거 후보 투표 응답을 실제 과거 선거 결과와 일치시키는 보정 후 사전 여론조사의 정확성이 개선되는가 혹은 보정 후의 선거여론조사 결과는 어떻게 변하는지도 파악한다. 보정이 정확성을 개선한다면 표본 자체가 편향되어 있으며, 과거 선거 투표에 대한 회상 응답을 신뢰할 수 있다는 의미로 볼 수 있다. 정확성이 개선되지 않는다면 표본 대표성 문제라기보다는 회상 응답을 신뢰할 수 없다는 근거가 될 수 있을 것이다.

보정의 방법은 지난 선거 투표 후보 질문에 대한 응답에 후보를 응답한 경우와 그렇지 않은 경우(투표 후보 말할 수 없다, 기권, 투표권 없음)를 구분하고, 지난 선거 투표 후보 응답자의 비율을 실제 후보별 득표율에 일치시켰다. 이외의 경우는 그 비율을 그대로 가져왔다.

예를 들어, 1997년에 1992년 대선 투표 후보 회상 응답에서 2,500명 응답자 중 투표 후보를 응답한 경우는 1,849명으로 74%, 이외의 경우는 651명으로 26%이다. 투표 후보 응답자 중 972명인 52.6%가 김영삼에 투표했다고 응답하였으나 실제 득표율은 42.0%로 776명이 기대된다. 따라서 이들의 경우 ‘기대 사례수/응답 사례수’로 가중을 적용하여 상대적 비중을 축소시켰다. 응답거절, 기권, 투표권 없음 등으로 후보를 응답하지 않은 경우는 기대 사례수와 응답 사례수가 동일하다고 가정했다. 이들의 경우 응답거절을 배분하기 어렵고, 5년이 지난 회상 시점에서 기권과 투표권 없음을 분리하기 어렵다.

이외 지난 대선 투표 회상 응답을 실제 득표율과 비교 시 응답거절, 기권, 투표권 없음은 제외하고 응답 비율을 재환산했다. 실제 득표율은 비투표자를 포함하지 않기 때문에 후보 득표율을 모두 합하면 100%가 되지만 회상 응답은 응답거절, 기권, 투표권 없음이 포함되기 때문에 후보별 응답의 합이 100%가 될 수 없다. 응답거절, 기권, 투표권 없음은 분리하거나 추정하기 어렵고, 일부 조사에서는 비율(또는, 사례수)을 확인할 수 없다.

제4장 결과

4.1 13대~18대 대선 투표율

<표 3>은 실제 대선 투표 참여율과 후보자 투표율 그리고 한국갤럽조사와 한국종합사회조사의 대선 투표 참여율과 후보자 투표율이다. 두 조사의 문항, 조사 방법 등의 다양한 차이를 고려할 때 단순 비교를 할 수 없다는 점을 염두에 두고 해석해야 한다. 실제 투표율과 두 조사의 투표율의 차이를 보면, 최소 10.7%p에서 최대 20.4%p 과대보고 되고 있다. 한국갤럽의 조사의 투표율은 모든 조사에서 90%가 넘었지만, 한국종합사회조사의 투표율은 2013년과 2016년 조사의 18대 투표에서만 거의 90%에 육박하고 있다. 17대 대선(2007년)에 대한 2008년과 2010년 KGSS는 문항의 차이가 없는데 거의 같은 투표율을 보이고 있다. 2016년의 조사도 유사하지만, 설문 문항이 바뀌었다. 2016년 조사에서 17대(2007년)과 18대(2012년)의 회상 투표율에서는 거의 10%p 차이가 있는데, 이런 차이가 실제 투표자 비율에 따른 차이를 반영한 것일 수도 있다. 즉, 실제 투표율이 낮은 연도에 대한 회상에서 조사의 투표자 비율도 낮아질 수 있다.

후보자에 대한 투표율을 보면, 2010년과 2013년의 한국종합사회조사 또한 대통령 당선자에 대해 과대보고 하고 있지만, 2016년 KGSS에서는 실제득표율과 거의 차이가 없다. 2013년 이후 한국종합조사와 한국갤럽조사의 후보자 투표율은 비슷하다.

<표 3> 대선 투표 참여 여부 및 대선 후보 투표율 - 한국갤럽, 한국종합사회조사 (%)

역대 대통령 선거	실제	92 G*	97 G	02 G	03 K**	07 G	08 K	10 K	13 K	13 G	14 G	15 G	16 K	
13대 (1987)	투표자 비율	89.2	-											
	노태우(민주정의당)	36.6	48.8											
	김영삼(통일민주당)	28.0	20.4											
	김대중(평화민주당)	27.0	26.4											
	김종필(신민주공화당)	8.1	4.4											
	기타후보	0.2	-											
14대 (1992)	투표자 비율	81.9		93.9										
	김영삼(민주자유당)	42.0		52.8										
	김대중(민주당)	33.8		30.2										
	정주영(통일국민당)	16.3		6.4										
	박찬종(신정당)	6.4		9.9										
	기타후보	1.5		0.7										
15대 (1997)	투표자 비율	80.7			91.4									
	이회창(한나라당)	38.7			37.8									
	김대중(새정치국민회의)	40.3			51.0									
	이인제(국민신당)	19.2			10.0									
	권영길(건설국민승리21)	1.2			0.9									
	기타후보	0.6			0.2									
16대 (2002)	투표자 비율	70.8				86.1	91.2							
	이회창(한나라당)	46.6				-	47.3							
	노무현(새천년민주당)	48.9				-	48.7							
	권영길(민주노동당)	3.9				-	3.3							
	기타후보	0.6				-	0.8							
17대 (2007)	투표자 비율	63.0						80.9	80.1				79.7	
	정동영(대통합민주신당)	26.1						-	21.9				22.6	
	이명박(한나라당)	48.7						-	58.8				57.4	
	권영길(민주노동당)	3.0						-	3.8				-	
	문국현(창조한국당)	5.8						-	5.9				-	
	이회창(자유선진당)	15.1						-	2.9				-	
	기타후보	1.3						-	6.7				20.0	
18대 (2012)	투표자 비율	75.8								88.9	93.0	92.1	91.5	88.5
	박근혜(새누리당)	51.6								56.0	59.1	58.9	58.5	51.2
	문재인(민주통합당)	48.0								43.2	39.8	39.4	38.8	42.8
	기타후보	0.4								0.8	1.2	1.6	2.7	6.0

주: \* 한국갤럽  
 \*\* 한국종합사회조사(KGSS)



## 4.2 2017년 19대 대선 이후 지난 대선 후보 투표

### 4.2.1 제21대 총선 출구조사

<표 4> 2017년 19대 대통령 선거 투표 회상 - 21대 총선 출구조사 투표자 심층여론조사 (%)

	문재인	홍준표	안철수	유승민	심상정	기타 후보	기권	투표권 없음	응답 거절
2017년 투표 회상	52.6	19.6	8.0	3.4	2.7	2.1	7.3	-	4.4
(A) 19대 대선 투표자 기준 재환산	59.5	22.2	9.0	3.7	3.1	2.4			
(B) 2017년 득표율	41.1	24.0	21.4	6.8	6.2	0.6			
(A-B)	+18.4	-1.9	-12.4	-2.9	-3.1	+1.8			

주: MBC-코리아리서치 출구조사, 전국 투표자 3,142명, 2020년 4월 15일

<표 4>는 MBC-코리아리서치가 발표한 심층여론조사 결과이다. <표 4>에서 후보자에 대한 (A)의 19대 대선 투표자 기준 재환산과 (B)의 2017년 득표율의 차이를 살펴보면, 문재인 후보에 대해서는 상당히 과대보고(18.4%p) 되었으며, 홍준표 후보에 대해서는 약간 과소보고되었고(1.9%p), 안철수 후보는 매우 과소보고(12.4%p)되었다. 투표자 대상의 출구조사에서 과거 투표 회상 응답은 실제 득표율과 크게 다를 수 있다.

### 4.2.2 2017년 19대 대선 이후 선거여론조사

<표 5> 2017년 19대 대통령 선거 투표 회상 - 한겨레신문-한국리서치, SBS-칸타코리아 결과 (%)

		2017년			2018년	2019년					평균
		한겨레-한국			SBS-칸타	한겨레-한국	SBS-칸타				
		5월 12-13일	8월 11-12일	10월 27-28일	2월 11-14일	5월 2-3일	5월 7-8일	8월 12일	9월 9일	11월 6일	
19대 대선 투표 회상 응답 - 기본	문재인	52.3	50.7	48.9	47.2	54.4	53.8	48.7	49.2	51.5	50.7
	홍준표	14.1	12.6	9.6	12.6	10.6	14.2	15.7	15.3	16.4	13.5
	안철수	14.8	12.6	10.9	9.2	7.6	7.9	8.1	10.5	8.3	10.0
	유승민	3.9	5.6	3.5	6.1	3.5	3.7	4.2	4.6	4.3	4.4
	심상정	5.8	4.3	5.5	4.7	4.6	2.3	2.8	2.9	3.1	4.0
	기타 후보	0.7	0.7			1.8	0.8	0.8	0.7	1.2	1.0
	기권/투표권 없음/응답거절	8.4	13.5	21.6	20.2	17.6	17.3	19.7	16.9	15.1	16.7
문재인 대통령 직무 평가	긍정 평가	77.8	78.6	76.4	67.4	51.6	49.3	56.2	45.1	47.6	61.1
	부정 평가	6.6	16.5	19.9	26.8	44.6	48.3	38.3	51.6	49.6	33.6
	모름/응답거절	15.6	4.9	3.7	5.8	3.7	2.4	5.5	3.3	2.8	5.3

<표 5>는 문재인 대통령 당선 후인 2017년 5월부터 2019년 11월까지 9개 선거여론조사 결과이다. 동일 의뢰처-조사기관일 때 대체로 동일한 설계와 질문, 자료수집 방법을 이용하기 때문에 유사한 조건에서 결과를 비교할 수 있다. 따라서, 해당 기간 중 과거 대선 투표 회상을 반복 조사한 한겨레신문-한국리서치, SBS-칸타코리아 결과를 중심으로 분석했다. 문재인 투표 회상 응답은 47.2~54.4%까지 7.2%p 범위 내에서 변동했다. 평균 50.7%를 중심으로  $\pm 3.7\%$ p 내이기 때문에 대략 1,000명 표본 조사의 최대 표본오차인  $\pm 3.1\%$ p보다 조금 높은 수준이었다. 반면, 홍준표 투표 회상 응답은 2017년 하반기까지 하락하다가 2019년부터 상승하였으며, 안철수 투표 회상 응답은 2017년 선거 직후보다 2019년 들어 더 하락하는 추세였다. 유승민, 심상정 투표 회상의 경우도 2017년 대비 2019년 점차 하락했다.

조사 시점에서의 대통령(후보자) 평가와 선거 투표 회상과의 관계가 있는지 살펴 보았다. 대통령 직무 긍정률은 45.1~78.6%까지 33.5%p 범위를 오르내렸다. 추세로 본다면 대통령 직무 긍정률은 당선 직후부터 2019년 5월까지 하락했고, 그해 8월 반등했으나 다시 하락하여, 전반적으로는 초기부터 하락하는 추세를 보였다. 반면 문재인 투표 회상 응답은 변화의 폭이 적을 뿐 아니라 추세도 명확하지 않았다. 즉, 유사 조사에서 대통령에 대한 직무 평가와 관계없이 회상 투표율에는 일관성 있는 관계가 없다고 판단된다.

<표 6>에서는 실제 득표율과 비교를 위하여, 기권/투표권 없음/응답거절자를 제외하고 재환산한 결과는 문재인 57.1~65.9%, 홍준표 12.2~19.6%, 안철수 9.2~16.2%였다. 문재인 투표는 평균 19.9%p 과대 회상된 반면 홍준표 -7.9%p, 안철수 -9.5%p 등 이외 후보들은 모두 과소 회상되는 경향이 있었다.

<표 6> 2017년 19대 대통령 선거 투표 회상과 실제 득표율 차이 - 한겨레신문-한국리서치, SBS-칸타코리아 결과 (%)

		2017년			2018년	2019년					평균
		한겨레-한국			SBS-칸타	한겨레-한국	SBS-칸타				
		5월 12-13일	8월 11-12일	10월 27-28일	2월 11-14일	5월 2-3일	5월 7-8일	8월 12일	9월 9일	11월 6일	
19대 대선 투표 회상 응답 - 투표자 기준 재환산	문재인	57.1	58.6	62.4	59.1	65.9	65.1	60.6	59.1	60.7	61.0
	홍준표	15.4	14.6	12.2	15.8	12.8	17.2	19.6	18.4	19.3	16.1
	안철수	16.2	14.6	13.9	11.5	9.2	9.6	10.1	12.6	9.8	11.9
	유승민	4.3	6.5	4.5	7.6	4.2	4.5	5.2	5.5	5.1	5.3
	심상정	6.3	5.0	7.0	5.9	5.6	2.8	3.5	3.5	3.7	4.8
	기타 후보	0.8	0.8	0.0	0.0	2.2	1.0	1.0	0.8	1.4	0.9
회상과 득표율 차이	문재인	+16.0	+17.5	+21.3	+18.1	+24.9	+24.0	+19.6	+18.1	+19.7	+19.9
	홍준표	-8.6	-9.5	-11.8	-8.2	-11.2	-6.9	-4.5	-5.6	-4.7	-7.9
	안철수	-5.3	-6.8	-7.5	-9.9	-12.2	-11.9	-11.3	-8.8	-11.6	-9.5
	유승민	-2.5	-0.3	-2.3	0.9	-2.5	-2.3	-1.5	-1.2	-1.7	-1.5
	심상정	+0.2	-1.2	+0.8	-0.3	-0.6	-3.4	-2.7	-2.7	-2.5	-1.4
	기타 후보	+0.2	+0.3	-0.5	-0.5	+1.6	+0.4	+0.4	+0.3	+0.9	+0.3

<표 7> 사전 선거여론조사와 출구조사의 대통령 선거 투표 회상 비교 (%)

	문재인	홍준표	안철수	유승민	심상정	기타 후보
(A) 사전여론조사 19대 대선 투표자 기준 재환산	60.7	19.3	9.8	5.1	3.7	1.4
(B) 출구조사 19대 대선 투표자 기준 재환산	59.5	22.2	9.0	3.7	3.1	2.4
(A-B)	+1.2	-2.9	+0.8	+1.4	+0.6	-1.0
(C) 2017년 득표율	41.1	24.0	21.4	6.8	6.2	0.6
(A-C)	+19.9	-4.7	-11.6	-1.7	-2.5	+0.9
(B-C)	+18.4	-1.9	-12.4	-2.9	-3.1	+1.8

주: (A) SBS-칸타코리아 선거여론조사(전화조사), 전국 유권자 1,009명, 2019년 11월 6일  
 (B) MBC-코리아리서치 출구조사, 전국 투표자 3,142명, 2020년 4월 15일

<표 7>은 출구조사의 심층조사에서 나타난 대통령 선거 투표 회상 응답과 앞서 언급한 사전 선거여론조사를 비교한 것이다. 사전 선거여론조사의 경우 2017년 이후 전반적으로 회상 응답의 변화는 크지 않았으나 투표자 회상 비율의 추세적 변화를 반영하기 위해 2019년 11월 SBS-칸타코리아 여론조사를 기준으로 삼았다. 두 조사에서 회상 응답의 비율 차이는 대체로 표본오차 범위(홍준표 2.9%p 차이) 안에 있다. 이러한 결과는 조사 시점 차이, 유권자 비율을 기준으로 하는 사전 선거여론조사와 투표자를 대상으로 하는 출구조사라는 차이점을 감안할 때 두 조사의 회상 응답은 차이가 거의 없다고 볼 수 있다.

### 4.3 2012년 18대 대선 투표 후보 회상

앞서, 2017년 대선 이후 2019년 11월까지의 <표 5>에서 지난 대선 투표 회상 응답 비율은 대통령 직무 평가의 변동에 비해 변화의 폭이 크지 않아 대체로 안정적이 었다. 정치적 상황이 급변할 때 지난 선거 회상 응답은 어떻게 변화하는지를 알아보 고자 한다. 2016년 11월 이후 2017년 5월 대선까지는 국정농단 및 대통령 탄핵, 대 선에 이르는 우리나라 정치가 급변하는 상황이었다.

<표 8>은 리서치뷰 정기조사로 2012년 대선 회상 투표율과 직무 평가율을 보여준 다. 해당 기간 중 중앙선거여론조사심위원회의 등록 자료를 탐색한 결과 지난 대선 투표 회상을 정기적으로 조사해 결과를 공개한 것은 이 조사가 유일했다. 대통령 직 무 긍정률은 4월까지 30%를 유지했고 9월 20%대로 하락했으며 태블릿PC 등 국정 농단 보도가 이어진 10월 31일에는 10%로 하락했다. 2016년 4월까지 박근혜 투표 가 문재인 투표보다 10%p 이상 많이 응답되었으나 2016년 12월 문재인 투표가 앞 서기 시작하여 2017년에는 문재인 투표가 10%p 가까이 많이 응답되었다.

<표 8> 2012년 18대 대통령 선거 투표 회상 - 리서치뷰 결과 (%)

		2016년							2017년	
		1월 15-17일	2월 13-15일	3월 25일	4월 6일	9월 28-29일	10월 31일	12월 8일	1월 30-31일	4월 4-6일
18대 대선 투표 회상 응답 - 기본	박근혜	48.1	49.7	52.6	49.8	44.5	40.3	35.3	35.9	36.3
	문재인	35.5	37.4	35.7	38.8	38.9	37.7	43.3	44.6	46.1
	기타 후보	5.5	5.7	3.6	3.4					
	기권/투표권 없음/응답거절	11.0	7.4	8.2	8.1	16.6	22.1	21.3	19.5	17.6
박근혜 대통령 직무 평가	긍정 평가	35.8	36.0	32.7	32.9	27.8	10.4			
	부정 평가	51.9	52.1	55.6	55.9	60.2	81.2			
	모름/응답거절	12.4	11.8	11.7	11.3	12.0	8.4			

<표 9>는 회상응답 재환산을 투표율과 실제 투표율의 차이를 보았는데, <표 8>과 유사한 변화를 보여주어, 회상 시점에서의 대통령 지지도 평가가 과거 투표행동의 회상에 영향을 준다는 점을 보여준다. 즉, 사회적 바람직성이 지난 대선 투표 회상에 영향을 미치는 점을 보여주었다고 생각할 수 있다.

<표 9> 2012년 18대 대통령 선거 투표 회상과 실제 득표율 차이 - 리서치뷰 결과 (%)

		2016년							2017년	
		1월 15-17일	2월 13-15일	3월 25일	4월 6일	9월 28-29일	10월 31일	12월 8일	1월 30-31일	4월 4-6일
18대 대선투표 회상 응답 - 투표자 기준 재환산	박근혜	54.0	53.6	57.3	54.2	53.3	51.7	44.9	44.6	44.1
	문재인	39.9	40.3	38.9	42.2	46.7	48.3	55.1	55.4	55.9
	기타후보	5.5	5.7	3.6	3.4	0.0	0.0	0.0	0.0	0.0
회상과 득표율 차이	박근혜	2.5	2.0	5.7	2.6	1.8	0.1	-6.6	-6.9	-7.5
	문재인	-8.2	-7.7	-9.2	-5.8	-1.4	0.3	7.1	7.4	7.9
	기타후보	5.0	5.2	3.1	2.9	-0.4	-0.4	-0.4	-0.4	-0.4

#### 4.4 2016년 20대 총선 지역구 투표 후보 회상

지금까지는 대통령 선거에서 투표 후보자에 대한 회상 응답을 살펴보았다. 대선과 달리 다른 선거 맥락에서도 동일한 회상 오류가 발생하는지, 대선 결과 해석이 다른 선거에서도 일반화할 수 있는 현상인지를 총선 지역구 투표의 회상 응답을 통해 살펴보고자 한다. 2020년 21대 총선 지역구 선거 여론조사는 최초로 휴대전화 가상번호를 이용해 조사를 실시했으며, 그동안 유선전화 중심의 조사의 한계로 인한 표본 추출틀 포함 오류를 해소하고 총선 여론을 대체로 정확하게 측정한 것으로 평가받고

있다. 특히 본 연구의 경우 총선을 앞두고 일부 언론 및 정치권에서 여론조사의 신뢰성을 문제 삼은 것으로부터 논의가 시작되었다. 따라서 이번 총선 선거여론조사에서 지난 선거 투표 후보 회상 응답에 어떠한 경향이 있었는지 살피는 것은 그러한 문제 제기에 대한 답으로서도 의미가 있다.

21대 총선 선거여론조사 중 중앙일보-입소스가 여론조사 공표 금지 직전 조사한 10개 관심 선거구 조사에서 지난 20대 투표한 후보에 대한 회상 응답이 포함되었으며(<표 10>), 이들을 중심으로 회상 응답과 실제 선거 결과를 비교 분석하고자 한다.

<표 10> 2016년 20대 국회의원 선거 지역구 투표 회상 - 2020년 21대 총선 선거여론조사 중앙일보-입소스 (%)

지역구		서울 강남구 을	서울 동작구 을	서울 광진구 을	서울 종로구	강원도 원주시 갑	대구 수성구 을	부산 남구 을	경남 김해시 을	세종시 을	경기 고양시 갑
조사일		4월 7-8일				4월 6-7일			4월 5-6일		
20대 총선 투표 회상 응답 - 기본	새누리당	17.1	32.3	14.6	19.5	21.1	13.5	17.4	12.5	10.4	12.4
	민주당	47.9	25.7	48.7	48.2	40.1	14.0	48.0	58.5	13.4	7.1
	기타 후보	0.4	6.1	2.5	2.3	1.8	36.4	1.6	0.4	35.3	49.6
	기권/투표권 없음/응답거절	34.6	36.0	34.3	30.0	37.1	36.0	33.0	28.6	40.9	31.2
21대 후보 지지도	민주당	44.8	53.6	48.1	58.4	47.6	24.5	46.9	46.8	50.2	24.1
	미래통합당	40.7	37.9	41.3	30.1	26.6	30.5	43.1	27.8	29.5	22.7
	주요 기타 후보						32.6 (홍준표)				40.1 (심상정)
	없다/모름/ 응답거절	9.9	5.9	8.6	8.6	15.0	11.9	9.8	12.7	18.6	12.5

<표 11> 2016년 20대 국회의원 선거 지역구 투표 회상과 실제 득표율 차이 - 2020년 21대 총선 선거여론조사 중앙일보-입소스 (%)

지역구		서울 강남구 을	서울 동작구 을	서울 광진구 을	서울 종로구	강원도 원주시 갑	대구 수성구 을	부산 남구 을	경남 김해시 을	세종시 을	경기 고양시 갑
조사일		4월 7-8일				4월 6-7일			4월 5-6일		
20대 총선 투표 회상 응답 - 투표자 기준 재환산	새누리당	26.1	50.4	22.2	27.9	33.5	21.1	26.0	17.5	17.6	17.9
	민주당	73.2	40.1	74.0	68.9	63.7	21.9	71.6	81.9	22.7	10.3
	기타 후보	0.6	9.5	3.8	3.3	2.9	57.0	2.4	0.6	59.7	71.8
20대 실제 득표율	새누리당	44.4	43.4	37.2	39.7	44.0	35.5	43.5	34.4	36.4	36.8
	민주당	51.5	31.5	48.5	52.6	43.9	17.7	48.1	62.4	10.6	8.7
	기타 후보	4.1	25.1	14.3	7.6	12.1	46.8	8.4	3.2	53.4	54.5
회상과 실제 차이	새누리당	-18.3	+7.0	-15.0	-11.9	-10.5	-14.3	-17.5	-16.9	-18.8	-18.9
	민주당	+21.8	+8.6	+25.5	+16.3	+19.8	+4.2	+23.5	+19.6	+12.1	+1.5
	기타 후보	-3.5	-15.6	-10.5	-4.3	-9.2	+10.1	-6.0	-2.6	+6.4	+17.3

<표 11>은 <표 10>을 20대 총선 회상 응답 결과와 실제 조사결과를 재환산해서 비교한 것으로 서울 동작구을 한 개 선거구를 제외하고 대부분 선거구에서 새누리당 후보 투표를 과소 회상한 것을 볼 수 있다. 동작구을 선거구를 제외하면 새누리당 후보를 -10.5~-18.9%p를 과소 회상했다. 동작구을의 경우도 국민의당 장진영 후보를 과소 회상한 것이 영향을 주었다고 볼 수 있다. 반면 민주당 후보 투표 회상은 실제 득표율 대비 +1.5~+25.5%p 높았다. 기타 정당, 무소속 유력 후보가 출마한 대구 수성구을, 세종시을, 경기 고양시갑을 제외하면 대부분 10%p 높았다. 기타 후보의 경우 유력 후보 출마 지역을 제외하면 대체로 실제보다 과소 회상했다.



<표 12> 사전 선거여론조사의 정확성 분석 - 2020년 21대 총선 선거여론조사  
중앙일보-입소스 (%)

지역구		서울 강남구 을	서울 동작구 을	서울 광진구 을	서울 종로구	강원도 원주시 갑	대구 수성구 을	부산 남구 을	경남 김해시 을	세종시 을	경기 고양시 갑
조사일		4월 7-8일				4월 6-7일			4월 5-6일		
후보별 지지도	민주당	44.8	53.6	48.1	58.4	47.6	24.5	46.9	46.8	50.2	24.1
	미래통합당	40.7	37.9	41.3	30.1	26.6	30.5	43.1	27.8	29.5	22.7
	A: 민주당-통합당	+4.1	+15.7	+6.8	+28.3	+21.0	-6.0	+3.8	+19.0	+20.7	+1.4
연령대별 투표율 반영*	민주당	43.7	51.8	47.1	56.3	47.3	22.4	44.5	45.4	48.5	23.2
	미래통합당	42.7	40.6	43.9	32.7	28.3	34.1	46.0	30.9	32.5	25.9
	B: 민주당-통합당	+0.9	+11.2	+3.1	+23.6	+18.9	-11.7	-1.5	+14.5	+16.0	-2.7
실제 득표율	민주당	46.4	52.2	50.4	58.4	48.6	25.1	50.5	49.7	58.0	27.4
	미래통합당	50.9	45.0	47.8	40.0	41.1	35.8	48.7	41.6	39.7	32.8
	C: 민주당-통합당	-4.5	+7.1	+2.6	+18.4	+7.4	-10.6	+1.8	+8.1	+18.3	-5.4
차이	A-C	+8.6	+8.6	+4.2	+9.9	+13.6	+4.6	+2.0	+10.9	+2.4	+6.8
	B-C	+5.4	+4.1	+0.5	+5.2	+11.5	-1.1	-3.3	+6.4	-2.3	+2.7

주: 연령대별 투표율을 한국갤럽 데일리오피니언의 사전 투표의향과 과거 투표 실현률을 반영하여 추정했으며 각 선거구의 연령대별 주민등록인구수\*예상 투표율로 산출함. 연령대별 예상 투표율은 20대 52.6%, 30대 52.4%, 40대 60.5%, 50대 72.2%, 60대 이상 76.7%임

<표 12>는 사전 선거여론조사의 정확성을 파악하기 위해 여당과 제1야당 후보의 지지도 차이와 실제 득표율 차이를 비교했다. 사전 선거여론조사는 선거일 7일 이전에 이루어진 조사로 전체 유권자 대상으로 연령대별 투표 참여율 차이가 반영되지 않았다. 따라서 연령대별 예상 투표율을 추정하였다. 예상 투표자 비중에 맞추어 연령대별 후보 지지도를 산출하고, 연령대별 투표율을 반영한 후보 지지도와 실제 득표율 차이를 비교하였다.

10개 지역구 중 당선자가 바뀐 경우는 서울 강남구을이었으며, 여야 간 차이로 산출한 오차 크기는 강원 원주시갑을 제외하고 모두  $\pm 10\%p$  이내였다. 표본크기(500명) 기준으로 볼 때 오차 범위 이내의 결과(최대 오차범위는 지지도간 차이이므로  $\pm 4.4\%p * 2 = \pm 8.8\%p$ )로 사전 선거여론조사가 대체로 정확했음을 알 수 있다.

### 4.5 역대 대선 지난 대선 후보 투표

<표 13> 1987~2002년 대통령 선거 투표 회상 - 1992~2007년 한국갤럽 사전 선거여론조사 (%)

구분	후보	소속정당	조사결과**	실제 득표율	차이
1992년 → 1987년 회상 응답*	노태우(당선)	민주정의당	48.8	36.6	+12.1
	김영삼	통일민주당	20.4	28.0	-7.6
	김대중	평화민주당	26.4	27.0	-0.6
	김종필	신민주공화당	4.4	8.1	-3.7
	기타후보			0.2	-0.2
1997년 → 1992년 회상 응답	김영삼(당선)	민주자유당	52.8	42.0	+10.8
	김대중	민주당	30.2	33.8	-3.6
	정주영	통일국민당	6.4	16.3	-10.0
	박찬중	신정당	9.9	6.4	+3.5
	기타후보		0.7	1.5	-0.8
2002년 → 1997년 회상 응답	이회창	한나라당	37.8	38.7	-0.9
	김대중(당선)	새정치국민회의	51.0	40.3	+10.7
	이인제	국민신당	10.0	19.2	-9.2
	권영길	건설국민승리21	0.9	1.2	-0.2
	기타후보		0.2	0.6	-0.4
2007년 → 2002년 회상 응답	이회창	한나라당	47.3	46.6	0.7
	노무현(당선)	새천년민주당	48.7	48.9	-0.2
	권영길	민주노동당	3.3	3.9	-0.6
	기타후보		0.8	0.6	0.1

주: \* 1997년 이후 유권자 기준에 따라 사후 가중처리한 결과, 반면 1992년 연령대별 투표율 예측을 반영한 결과임  
 \*\* 기권/투표권 없음/응답 저절을 제외하고 재환산함

<표 13>은 한국갤럽의 사전선거조사에서 과거 대선 후보 투표에 관한 질문에 대한 응답 결과이다. 우선, 회상 응답 결과와 실제 득표율을 비교하면 2002년을 제외하고 당선자 회상 응답이 실제 득표율보다 10%p 이상 많았다. 2002년 대선의 회상 응답은 실제 득표율과 1%p 이내의 차이를 보였는데, 예외적인 경우라 할 수 있다.

<표 14> 1997~2002년 대통령 선거 투표 회상 응답별 후보 지지도 - 한국갤럽 사전 선거여론조사 (%)

		사례수	1997년 후보 지지도				
			이회창	김대중	이인제	기타 후보	없다/모름/ 응답거절
전체		2,500	36.3	38.9	17.9	1.6	5.4
1992 대선 투표 후보	김영삼	972	53.5	18.7	22.9	0.7	4.2
	김대중	565	6.3	86.6	3.7	1.1	2.3
	정주영	117	48.8	21.2	22.6	2.6	4.8
	박찬중	181	46.4	18.0	24.7	5.7	5.2
	기타 후보	14	29.9	33.8	7.2	12.2	17.0
	말할 수 없다	200	34.8	26.7	12.0	0.2	26.3
	기권	130	35.3	37.1	20.7	4.2	2.7
	투표권 없었음	321	28.1	42.7	25.2	1.8	2.2
1997년 실제 득표율			38.7	40.3	19.2	1.8	
		사례수	2002년 후보 지지도				
			이회창	노무현	권영길	기타후보	없다/모름/ 응답거절
전체		2,055	42.1	45.5	4.2	0.6	7.5
1997년 대선 투표 후보	이회창	599	79.6	13.4	2.1	0.8	4.1
	김대중	809	20.6	68.3	3.6	0.5	7.0
	이인제	159	33.7	51.4	7.8	0.0	7.0
	권영길	15	6.8	59.0	29.4	0.0	4.8
	기타 후보	3	32.0	16.7		37.1	14.2
	말할 수 없다	121	43.2	20.7	4.9	0.9	30.3
	기권	160	38.9	45.6	7.1	0.0	8.4
	투표권 없었음	189	26.7	60.4	5.9	0.9	6.1
2002년 실제 득표율			46.6	48.9	3.9	0.6	

<표 14>는 1997년 사전선거조사에서 후보 지지도와 1992년 후보자 투표 회상 비율이다. 선거여론조사의 후보 지지도는 대통령 선거 직전의 조사로 실제 득표율 결과를 대체로 잘 반영하고 있다. 예로, 1992년 대선 투표 후보 회상 응답은 실제와 차이가 있었지만 1997년 후보 지지도 결과는 김대중, 이회창 후보의 지지도 차이가 2.6%p이고, 실제 득표율 차이인 1.6%p와 유사했다. 또한, 2002년의 경우도 노무현, 이회창 후보 지지도 차이는 3.4%p이고 득표율 차이는 2.3%p 차이로 거의 같았다.

<표 15> 1997~2002년 대통령 선거 후보 지지도 정확성 비교 - 한국갤럽 사전 선거여론조사 (%)

	이회창	김대중	이인제	기타 후보	없다/모름/ 응답거절	(이회창-김대중)
후보 지지도-단순	36.3	38.9	17.9	1.6	5.4	-2.6
후보 지지도- 지난 선거 투표자 비중 보정	34.8	40.8	17.3	1.7	5.4	-6.0
1997년 실제 득표율	38.7	40.3	19.2	1.8		-1.6

	이회창	노무현	권영길	기타후보	없다/모름/ 응답거절	(이회창-노무현)
후보 지지도-단순	42.1	45.5	4.2	0.6	7.5	-3.4
후보 지지도- 지난 선거 투표자 비중 보정	43.4	43.8	4.6	0.7	7.5	-0.4
2002년 실제 득표율	46.6	48.9	3.9	0.6		-2.3

<표 15>는 일부 주장대로 과거 투표 행동을 반영하는 것이 적절한가에 대해 지난 대선 투표 후보 회상 응답을 실제 득표율과 일치하도록 표본 비중을 보정하는 접근을 시도하였다. 보정을 거친 후 후보 지지도의 변화가 표본의 대표성을 향상시켰는지를 확인할 수 있다. 과거 대선 투표 후보 응답자의 비율을 실제와 일치하도록 응답자 비중을 보정하여 후보 지지도를 산출했다. 이 때 기권/투표권 없었음/응답거절의 경우는 응답자 비율을 그대로 적용했다.

과거 후보 투표율 기준 보정 결과 1997년 두 후보 간 지지도 차이는 2.6%p에서 6.0%p로 확대되었다. 이는 실제 차이인 1.6%p에서 오차가 확대된 결과이다. 2002년의 경우 두 후보 간 지지도 차이가 3.4%p에서 0.4%p로 줄었는데 보정 전후를 비교하면 보정 후 실제 결과인 2.3%p 대비 오차가 확대되었다.

요약하면, 선거여론조사의 후보 지지도는 현재 유권자의 정치적 선호를 정확히 반영하고 있으며, 과거 선거 투표의 회상 응답이 실제와 다른 것이 표본의 편향 증거가 될 수 없음을 확인할 수 있다.

## 제5장 선거여론조사 편향성 기준 및 제도개선 방안

### 5.1 결과 요약

여론조사에서 지난 선거 투표의 회상 응답에서 투표율은 최소 10%p 과대 보고 있으며, 투표자에 대한 투표율에서는 2002년 16대 예외가 있지만 대통령 당선자에게 투표했다는 비율이 당선자의 득표율보다 높았다. 21대 총선 출구조사에서 이뤄진 심층여론조사, 이외 대선과 총선 맥락의 다양한 선거여론조사에서 일부 예외는 있으나 전반적으로 지난 선거에 대한 회상 응답에는 편향이 있었다. 편향의 방향은 주로 당선자 또는 여당, 현재의 선호 정당을 향하고 있었으며, 편향의 크기는 거의 없는 것부터 20%p 이상까지 다양했다.

그러나 이러한 회상 응답의 편향은 여론조사 응답자 표본의 대표성 문제라기보다는 회상 응답의 오류, 측정의 문제라고 보여진다. 그 근거를 요약하면 다음과 같다.

첫째, 21대 총선 출구조사는 현재의 조사 역량이 총집결된 조사로, 전체 득표율 기준에서 예측 결과가 실제 결과와 거의 같았다. 이러한 출구조사의 심층조사에서도 당선자에 대한 과대보고되는 경향인 편승효과(bandwagon effect)가 있었다. 총선 출구조사의 심층조사와 사전여론조사의 지난 대선 투표 후보 응답 비율은 차이가 크지 않았다. 또한, 과거 대선 투표 후보 응답 기준에서 사전 선거여론조사는 출구조사 수준, 출구조사와 유사한 수준의 표본 대표성을 확보한 것으로 볼 수 있다. 물론, 출구조사는 투표자, 사전 선거여론조사는 유권자 전체를 모집단으로 하기 때문에 표본 특성이 다를 수 있지만, 지난 대선 투표 응답에서는 유사한 결과를 보이고 있다.

둘째, 21대 총선 지역구 선거여론조사는 투표일 7일 이전 조사라는 제한점이 있으나 선거 판세를 대체로 정확히 반영하였다. 이들 조사에서도 과거 총선 투표 후보의

회상은 여당 후보로의 과대 편향 오류가 있었다.

셋째, 역대 대선의 사전여론조사(한국갤럽)의 경우도 대체로 과거 대선 투표 후보에는 당선 후보로의 과대 편향 오류가 있었다. 그러나 지난 대선 투표 응답을 실제 득표율에 일치되도록 표본 크기를 보정했을 때 예상 득표율 차이가 개선되지 않거나 확대되었다. 지난 선거 투표 회상 응답을 통한 보정이 제한적으로 적용되어야 함을 보여준다. 비슷한 시기와 조건에서 이뤄지는 연속적인 추적조사에서 직전 조사 또는 직전 조사의 추세를 통해 파악한 지난 선거 투표 응답 비율에 일치시키는 접근이 가능할 것이다. 같은 시기와 조건이라면 회상 오류가 유사하게 발생할 것으로 추정할 수 있기 때문이다. 그러나 과거 선거결과와 회상 응답을 일치시키는 것은 회상 응답의 오류를 확인할 수 없기 때문에 오히려 표본의 왜곡을 불러올 가능성이 커진다.

지난 선거 투표 회상에 일관성이 없고 회상 응답에 오류가 있으며, 응답 결과를 전적으로 신뢰하여 표본 대표성이 낮은 근거로 삼거나 보정하는 것이 타당하지 않다는 결론은 앞서 논의한 해외 연구 사례와도 일관된 결과이다.

## 5.2 편향성 기준 및 개선 방안

### 5.2.1 편향성 기준

선거여론조사의 편향을 확인하는 방법은 (가) 실제 선거 결과와의 비교, (나) 응답자와 무응답자간 차이 없음을 검증·확인하는 방법이 있다. 이외에 모집단 정의의 차이, 시간 간격에 인한 영향이 있을 수 있다. 대표적으로 선거결과는 투표자들의 투표의 총합이지만 선거여론조사는 유권자 전체를 모집단으로 한다. 선거여론조사가 유권자 전체를 모집단으로 하기 때문에 유권자 중 투표자의 정당 또는 후보 지지 결과인 선거 결과와 모집단 특성 차이가 있으며, 정당 또는 후보를 말하지 않거나 선

택하지 못한 유보층 배분에서도 오차가 발생할 수 있다. 또한 투표 시점과 조사 시점은 조사마다 차이가 있다. 이러한 조건들은 선거여론조사가 정확하게 진행되었음에도 불구하고 실제 결과와 달라질 수 있음을 의미한다.

선거여론조사는 실제 선거결과를 통해 정확성을 확인할 수 있는 거의 유일한 조사로, 조사방법 연구자들은 선거여론조사를 통해 조사방법을 검증하고 개선해왔다. 2014년 선거여론조사심의위원회 출범 이후 등록된 자료 분석을 통해 사전 선거여론 조사와 실제 결과와의 차이에 대한 연구가 다양하게 진행되었다(김영원, 황다솜 2014; 장덕현, 조성겸 2015; 김영원, 윤지혜 2016; 장덕현 2016). 다만 이러한 연구들을 최근 선거여론조사에 일반화하기 위해서는 지역 단위 조사에서 휴대전화 가상번호를 사용할 수 없었던 점, 선거여론조사 공표 금지 기간으로 인한 시간 차이 등 상황적 요인을 감안해야 한다.

Groves and Peytcheva (2008)는 무응답 편향에 대한 메타 분석 연구에서 5가지 무응답 편향 추정치 산출 방법을 제시한 바 있다: (1) 표본추출틀 자료: 표본추출틀에 응답자와 무응답자에 대한 정보를 포함, (2) 보조 자료: 응답자, 무응답자에 연결 가능한 보조 자료 활용, (3) 스크리너 인터뷰: 향후 조사 진행을 위해 사전 스크리너 조사로 정보 수집, (4) 사후조사 연구: 무응답자에 대한 사후조사(follow-up) 진행, (5) 응답 의향 조사: 향후 조사 의향을 물어 의향자와 비의향자를 비교.

무응답 편향을 추정하기 위해서는 표본 추출틀 자료, 보조 자료와 같은 객관적 자료의 필요성이 매우 크다. 이러한 자료 확보가 어렵기 때문에 전화조사 무응답 편향에 대한 국내 연구는 거의 찾을 수 없으며 그마저도 기존 조사 응답자를 표본 추출틀로 활용한 연구였다(장덕현 외 2014). 반면, 스크리너 인터뷰, 사후 조사 연구, 응답 의향 조사를 통한 편향 추정은 편향을 확대하고 측정 오차가 혼입되는 등 정확성 문제가 있을 수 있다.

선거여론조사의 편향을 측정하기 어렵기 때문에 조사 간 차이에 대해서 업체나 전문가의 판단에 의존하지만 이러한 판단은 과학적인 근거에 의한 것이라기보다는 하

나의 의견에 불가한 경우가 많고, 자의적인 해석일 수 있다. 선거여론조사의 편향을 직접 측정할 수 없기 때문에 대안적인 지표들이 있으나 근거가 부족하거나 현재 우리나라 선거여론조사에 적합하지 않다. 예로, 편향의 중요한 지표로 간주되는 응답률이다. 응답률이 낮다면 무응답자의 비율이 높아지기 때문에 무응답 편향이 발생할 가능성이 높아진다고 볼 수 있다. 모든 조사 대상자가 응답하지 않는 이상 응답자와 무응답자의 특성이 달라질 가능성을 배제할 수 없다. 그러나 응답률이 상대적으로 낮더라도 응답자와 무응답자의 차이가 크지 않다면 표본 대표성이 크게 훼손되지 않을 수 있다. 즉, 낮은 응답률은 무응답 편향 발생의 필요조건일 뿐 충분조건이 될 수 없다(Groves and Peytcheva 2008).

따라서 무응답을 발생시키는 기제를 탐구할 필요가 있다. 크게 (1) 분리 원인 모형, 완전 임의 결측(MCAR: Missing Completely At Random), (2) 공통 원인 모형, 임의 결측(MAR: Missing At Random), (3) 조사변수 원인 모형, 비임의 결측(MNAR: Missing Not At Random)로 구분되는데, 조사변수 자체가 응답 결정의 요인이 되는 비임의 결측 조건(MNAR: Missing Not At Random)에서 무응답 편향 발생 가능성이 커진다. 선거여론조사에서 비임의 결측 조건은 지역, 성, 연령대와 상관 없이 특정 지지층이 더 응답하거나 덜 응답하는 경우, 또는 선거에 대한 관심층이 더 많이 응답하는 경우를 말한다.

응답률의 대안으로 편향을 나타내는 지표로 제시된 것이 대표성 지수(R-indicator)이다(Bethlehem, Cobben and Schouten 2008). 강한 대표성 가정은 모집단을 구성하는 모든 단위들이 표본이 될 확률( $p_k$ )이 동일함, 약한 가정은 모집단을 구성하는 각 층(stratum) 간 응답 확률( $p_h$ )이 동일함을 의미한다. 대표성 지수 산출을 위해서는 개별 단위들의 응답 확률을 알 수 없기 때문에 보조 정보로부터 응답 확률을 추정하며, 로지스틱, 프로빗 모형이나 CHAID 분류 분석 기법을 활용한다. 표본 단위에 대해 추정된 응답확률의 분산이 작을수록 대표성 지수는 높게 추정된다. 즉, 표본 단위의 응답확률이 동질적일수록 대표성이 높다고 볼 수 있다.



표본 단위의 추정 응답 확률  $\hat{\rho}_1, \hat{\rho}_2, \dots, \hat{\rho}_n$

평균 응답 확률  $\hat{\rho} = \frac{1}{N} \sum_{i=1}^n \frac{\hat{\rho}_i}{\pi_i}$ ,

대표성 지수(R-indicator)  $\hat{R}(\rho) = 1 - 2 \sqrt{\frac{1}{N-1} \sum_{i=1}^n \frac{(\hat{\rho}_i - \hat{\rho})^2}{\pi_i}}$

대표성 지수 산출, 응답 확률 추정을 위해서는 응답자와 무응답자 모두에게 측정 가능한 보조변수가 필요하지만 현재 선거여론조사에서 이러한 조건을 확보할 수 없다. 휴대전화 가상번호의 경우도 지역, 성, 연령대라는 제한된 정보를 활용할 수 있을 뿐이며, 주요 정치 지표에 대한 이들 변수의 설명력이 높다고 볼 수 없다.

이상 논의한 바와 같이 현재 선거여론조사에서 편향성을 나타내는 지표는 응답률이 거의 유일하지만 대표성을 나타내는 지수로서 충분하다고 볼 수 없다. 대안으로 활용될 수 있는 대표성 지수는 전체 표본에 대한 정보 부족으로 추정 불가능하다.

따라서 편향 축소를 위한 방법과 조건을 논의하고, 바람직한 절차적 기준을 제시하는 것이 현실적 대안이 될 것이다. 또한 응답자와 무응답자간 차이를 파악하고 개별 표본의 응답확률을 추정할 수 있는 보조변수를 확보하는 방안도 함께 모색되어야 한다.

### 5.2.2 편향 축소를 위한 방법

선거여론조사의 정확성 확보를 위해서는 모집단 대표성을 확보한 표본을 추출하여 적절한 질문과 진행을 통해 응답자 특성, 생각, 의견을 정확히 측정해야 한다. 표본 대표성 확보는 적합한 조사 방법 선택, 모집단 대부분을 포함하는 표본추출틀, 적절한 표본 추출, 충분한 접촉, 응답률 확보를 통한 응답자와 무응답자의 차이 축소 노력을 통해 이뤄져야 한다(<표 16>).

<표 16> 표본 대표성 확보를 위한 절차, 과제 및 오차

절차	과제	관련 오차
조사 모집단 정의	목적에 맞는 모집단 범위에 대한 명확한 정의	
조사방법 선택	조사방법의 특징과 적절성 평가	자료수집방법 효과
표본추출틀 확보	모집단 대부분을 포함한 명부	포함오차
표본추출	층화, 표본크기, 배분	표본오차
조사 진행	접촉률, 협조율 확보, 진행 관리	무응답 오차
모수 추정	적절한 가중치 산출	

조사방법의 선택, 응답률, 조사 시기, 가중배율 등 조사를 구성하는 주요 특성들이 정확성과의 관계성은 기존 연구 결과를 참고할 수 있다. 2014년 지방선거, 2016년 국회의원선거 선거여론조사의 정확성을 분석한 연구에서는(김영원, 황다솜 2014; 장덕현, 조성겸 2015; 김영원, 윤지혜 2016; 장덕현 2016) 전화면접조사가 ARS 대비 편향이 적고, 응답률이 높을수록, 가중 배율이 적을수록 편향이 축소되는 관계를 일관되게 밝히고 있다. 예를 들어, 2016년 총선 선거여론조사를 분석한 결과 선거일과 조사일 간격이 클수록, 가중배율이 클수록 오차가 증가하고, 응답률이 높을수록 오차가 감소하는 상관성이 있었다(<표 17>).

<표 17> 2016년 총선 선거여론조사 주요 조사 특성과 오차 간 상관관계 분석(Pearson's r)

	조사일(D-*)	응답률	최대가중배율	표본크기
오차	.100*	-.268**	.183**	.038
절대 오차	.125**	-.283**	.263**	.067

주: \* p <.05; \*\* p<.01

- 오차 = (사전지도도<sub>새누리</sub> - 사전지도도<sub>야당</sub>) - (득표율<sub>새누리</sub> - 득표율<sub>야당</sub>)
- 531건 사전 선거여론조사 분석: 3월 21일 이후 등록 688건 중 실제 선거결과 새누리당 후보가 1, 2위 내에 포함된 지역구 분석

이러한 결과에 따른다면, 동일한 조건 하에서 적합한 조사방법 선택, 조사 응답률 제고, 가중배율 축소가 조사의 정확성을 높일 수 있으며, 응답률과 가중배율이 편향성을 나타내는 지표로 활용될 수 있다고 판단된다. 다만 선거여론조사에서 휴대전화 가상번호가 활용되고, 접촉률 기준을 추가로 도입되었기 때문에 조사의 특성과 정확성과의 관계를 다시 검토하고 분석할 필요가 있다.

### 5.2.3 제도 개선 방안

이상의 논의를 바탕으로 제도 개선 방안을 제시한다.

가. 현재 중앙선거여론조사심의위원회에서는 선거여론조사기준의 제2장 일반기준 제4조 신뢰성과 객관성 확보를 위한 기준, 제5조 가중값 배율, 제6조 질문지의 작성, 제7조 조사대상자의 권리 보호를 언급하고 있다(<https://www.nesdc.go.kr/portal/content/view.do?menuNo=300010>). 선거여론조사의 일반기준에 덧붙여 품질(혹은 편향) 기준을 구체적으로 기술한다면 선거여론조사에 대한 불필요한 오해를 줄일 수 있다. 예로, 정치권, 언론 등에서 자주 언급하는 선거여론조사에 대한 비판 내용을 기준에 포함할 수 있다.

나. 현재 선거여론조사 편향에 대한 지표로 응답률과 가중값 배율이 언급되고 있다. 앞으로 표본 추출률과 외부자료 연계(응답자 및 무응답자에 대한 정보 보조자료) 등으로 편향 기준을 개선해야 할 것이다. 이를 위해서는 조사 절차 및 특성의 의미와 효과, 조사 정확성에 영향을 주는 요인 등에 대한 다양한 연구들이 이뤄져야 한다.

다. 해외에서는 조사 자료와 유권자 등록 및 투표 기록을 통한 연구가 많이 진행되었다. 한국에서도 조사 자료를 이용한 선거 연구 결과의 왜곡 현상과 투표과대보고 때문에 조사 자료의 신뢰성이 지속적으로 문제가 제기되고 있다. 선관위에서 유권자 개인이 투표를 했는지 안 했는지를 볼 수 있는 자료를 연구자에게 공개해 투표 검증 자료를 구축했으면 한다. 투표 검증 자료는 투표 참여에 있어서 편향이 어느 정도 되는지 여부를 정확하면서도 간단 명쾌하게 파악할 수 있게 한다. 조사 자료와 투표 참여 행정 자료를 연계할 때, 조사 자료를 선관위의 등록 자료와 연계하는 것 대신에 선관위의 등록 자료(지난 선거 투표 참여자 명단)를 표집틀로 이용하여 조사를 수행하여야, 무응답자에 대한 특성을 파악할 수 있다.

라. 행정 자료의 연계가 가능하지 않은 경우 투표 참여 혹은 후보자 투표에 대하여 과대보고를 낮출 수 있는 설문 문항 개발이 절실히 필요하다. 예로, 응답자의 사회적 바람직성 편향을 줄일 수 있는 설문의 질문 부분과 응답 범주 개발이다. 또한, 인지 면접 등을 통해 투표 문항의 문제에 대한 원인을 파악하여 문항을 새로이 만들어 낼 필요가 있다. 이렇게 개선된 문항이 실제로 과대보고의 문제를 줄일 수 있는지 확인하기 위해 문항 조사실험 연구를 수행한 후 선거여론조사에 이용할 수 있다.

마. 2016년 12월 8일 언론 5개 단체는 ‘선거여론조사보도준칙’([https://www.journalist.or.kr/news/section4.html?p\\_num=13](https://www.journalist.or.kr/news/section4.html?p_num=13))을 공동으로 제정하고 선포했다. 하지만, 신문 및 방송에서 선거여론조사 결과를 명백하게 잘못 해석하여 보도하는

경우가 많고, 이에 대하여 지속적으로 선거여론조사 결과 해석에 대하여 미디어와 일부 조사회사가 준수할 수 있는 방안의 필요성이 지적되고 있다(홍영림 외 2012; 윤희웅 2017). 선거여론조사는 넘쳐나지만 선거여론조사의 결과에 대한 기자들의 올바른 해석은 주변부로 밀려난 것 같다. 약 30년의 선거여론조사를 거쳤고, 이제 ‘선거여론조사보도준칙’이 제정된 지 거의 4년이 되어 가는 시점이다. 마지막으로 제안하고 싶은 것은 중앙선거여론조사심의위원회에서 기자들을 대상으로 한 선거여론조사 교육 프로그램 개발이다.



참고문헌

- 강신구. 2017. “제 19대 대통령선거와 TV토론회” 강원택 편. 《변화하는 한국유권자6》 pp. 257-299. 서울: 동아시아연구원.
- 김영원, 윤지혜. 2016. “2016년 국회의원선거 전화여론조사 정확성 분석” 《조사연구》 17(3): 109-137.
- 김영원, 황다솜. 2014. “2014년 지방선거 여론조사 전화조사방법에 따른 예측오차 및 편향” 《조사연구》 15(4): 1-32.
- 김준철. 2015. 《여론조사로 대통령 만들기-어떻게 할 것인가》 북앤피플.
- 김지범, 강정한, 김석호, 김창환, 박원호, 이윤석, 최성수, 최슬기, 김솔이. 2019. 《한국 종합사회조사 2003-2018》 서울: 성균관대학교 출판부.
- 김지범, 장덕현, 김솔이, 강정한. 2019. “조사동향: 여론조사에 대한 조사” 《조사연구》 20(1): 131-164.
- 윤희웅. 2017. “대선 여론조사와 여론조사 보도의 문제점” 《관훈저널》 143: 11-18.
- 이계오, 장덕현. 2011. “사전 태도 및 투표 의향과 실제 투표행동간 상관성 연구” 《조사연구》 12(1): 1-30.
- 이준한. 2005. “투표에 대한 오버리포팅(overreporting)” 《한국정당학회보》 4(1): 69-90.
- 이현우, 전시홍. 2010. “설문조사의 투표율 과장 현상에 대한 연구: 거짓응답자 추출을 중심으로” 《한국과 국제정치》 69: 37-69.
- 장덕현. 2016. “총선 여론조사 품질 어떻게 개선할 것인가?” 한국언론학회 춘계학술대회 발표 원고.
- 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 정밀성과 편향 비교” 《조사연구》 16(1): 129-153.
- 지병근. 2010. “서베이 민주주의(Survey Democracy)?: 6·2 지방선거 후보공천사례를 중심으로” 《한국정치연구》 19(3): 57-75.
- 현경보. 2019. 《여론전쟁》 상상.
- 홍영림, 김지연, 정한울, 허진재. 2012. “대선 여론과 여론조사보도” 《관훈저널》 125: 205-218.

- Abelson, Robert P., Elizabeth F. Loftus and Anthony G. Greenwald. 1992. "Attempts to improve the accuracy of self-reports of voting" in *Questions About Questions* pp. 138-53. edited by Judith M. Tanur. Russell Sage: New York.
- Ansolabehere, Stephen and Eitan Hersh. 2012. "Validation: What big data reveal about survey misreporting and the real electorate" *Political Analysis* 20(4): 437-459.
- Belli, Robert F., Michael W. Traugott and Matthew N. Beckmann. 2001. "What leads to voting overreports? Contrasts of overreporters to validated voters and admitted nonvoters in the American National Election Studies" *Journal of Official Statistics* 17(4): 479-498.
- Berent, Matthew K., Jon A. Krosnick and Arthur Lupia. 2016. "Measuring voter registration and turnout in surveys: Do official government records yield more accurate assessments?" *Public Opinion Quarterly* 80(3): 597-621.
- Bernstein, Robert, Anita Chadha and Robert Montjoy. 2001. "Overreporting voting: Why it happens and why it matters" *Public Opinion Quarterly* 65(1): 22-44.
- Bethlehem, Jelke., Fannie Cobben and Barry Schouten. 2008. "Indicators for the Representativeness of Survey Response" Proceedings of Statistics Canada Symposium 2008 Data Collection: Challenges, Achievements and New Directions.
- Dahlgaard, Jens Olav, Jonas Hedegaard Hansen, Kasper M. Hansen and Yosef Bhatti. 2019. "Bias in self-reported voting and how it distorts turnout models: Disentangling nonresponse bias and overreporting among Danish voters" *Political Analysis* 27(4): 590-598.
- Dassonneville, Ruth and Marc Hooghe. 2017. "The noise of the vote recall question: The validity of the vote recall question in panel studies in Belgium, Germany, and the Netherlands" *International Journal of Public Opinion Research* 29(2): 316-338.
- DeBell, Matthew, Jon A. Krosnick, Katie Gera, David S. Yeager and Michael P. McDonald. 2018. "The turnout gap in surveys: Explanations and solutions" *Sociological Methods & Research* 0049124118769085.



- Durand, Claire. 2015. "Should recall of previous votes be used to adjust estimates of voting intention?" *Survey Methods: Insights from the Field (SMIF)*.
- Groves, Robert M. 2004. *Survey Errors and Survey Costs*. Hoboken, NJ: John Wiley & Sons.
- Groves, Robert M. and Emilia Peytcheva. 2008. "The impact of nonresponse rates on nonresponse bias: Meta analysis" *Public Opinion Quarterly* 72(2): 167-189.
- Himmelweit, Hilde T., Marianne Jaeger Biberian and Janet Stockdale. 1978. "Memory for past vote: Implications of a study of bias in recall" *British Journal of Political Science* 8(3): 365-375.
- Sciarini, Pascal and Andreas C. Goldberg. 2017. "Lost on the way: Nonresponse and its influence on turnout bias in postelection surveys" *International Journal of Public Opinion Research* 29(2): 291-315.
- Sigelman, Lee. 1982, "The nonvoting voter in voting research" *American Journal of Political Science* 26(1): 47-56.
- Silver, Brian D., Barbara A. Anderson and Paul R. Abramson. 1986. "Who overreports voting?" *The American Political Science Review* 80(2): 613-624.
- Van Elsas, E. J., R. Lubbe, T. W. Van Der Meer and W. Van Der Brug. (2014). "Vote recall: A panel study on the mechanisms that explain vote recall inconsistency" *International Journal of Public Opinion Research* 26(1): 18-40.
- Voogt, Robert J. and Willem E. Saris. 2003. "To participate or not to participate: The link between survey participation, electoral participation, and political interest" *Political Analysis* 11(2): 164-179.
- Waldahl, R. and Aardal, B. (2000). "The accuracy of recalled previous voting: Evidence from Norwegian election study panels" *Scandinavian Political Studies* 23(4): 373-389.
- Weir, B. T. (1975). "The distortion of voter recall" *American journal of political science* 53-62.
- Weisberg, Herbert F. 2009. *The Total Survey Error Approach: A Guide to the New Science of Survey Research*. Chicago: University of Chicago Press.