

## 연구논문

## 2008 총선 출구조사의 총조사오차 분석\*

A Total Survey Error Analysis of the Exit Polling for General Election 2008 in Korea

김영원\*\* · 곽은선\*\*\*

Young-won Kim · Eun-sun Kwak

본 논문에서는 2008년 18대 총선의 출구조사 자료를 이용하여 출구조사의 정확성을 평가할 수 있는 총조사오차 개념을 새로 정의하고, 출구조사에서 발생하는 총조사오차가 투표소 추출오차와 투표자 선정 및 응답 과정에서 발생하는 실사오차 중 어떤 것에 더 많은 영향을 받는지 분석했다. 또한 선거구별 무응답률이 총조사오차와 실사오차에 미치는 영향을 분석하였고, 중앙선거관리위원회의 투표율 분석자료를 이용하여 출구조사 표본의 대표성을 검증했다. 분석 결과 선거구 내 표본 투표소 추출 관련 오차보다는 표본투표소 내에서 투표자 선택 및 응답 과정에서 발생하는 오차가 더 컸던 것으로 나타났다. 또한 무응답률과 실사오차는 양의 상관관계를 갖는 것으로 나타났으며, 이는 특정한 지지성향을 갖는 사람들의 응답 거절로 인해 표본의 대표성이 떨어지고, 결과적으로 오차가 커지는 것으로 해석될 수 있다. 아울러 선관위와 출구조사 자료에 대한 카이제곱 검정을 통해 성/연령대별 구성비에 유의한 차이가 있다는 것을 확인할 수 있었다.

**주제어:** 출구조사, 무응답률, 총조사오차, 실사오차, 표본추출오차

In this study, we newly define the Total Survey Error(TSE) in exit poll and investigate the TSEs of the exit poll survey for the 18th general election of 2008 to analyse the cause of the exit poll prediction error. To explore the main cause and effect of the total survey error, the total survey error was divided by the sampling error which comes from sampling process of poll stations and the non-sampling error which comes from selecting voter and collecting responses from sampled voters in each electoral district. We consider the relationship between non-response rates and total survey error as well as non-sampling error. Also, we study the representativeness of the exit poll

\* 본 연구는 숙명여자대학교 2009학년도 교내연구비 지원에 의해 수행되었음.

\*\* 교신저자(corresponding author): 숙명여자대학교 통계학과 교수 김영원.

E-Mail : ywkim@sookmyung.ac.kr

\*\*\* 숙명여자대학교 통계학과 대학원생

sample by comparing the sex/age distribution of the exit poll data and the National Election Commission poll data.

**Key words** : exit poll, non-response rate, total survey error, non-sampling error, sampling error

## I. 서론

우리나라 방송사들의 선거예측조사는 1992년 대통령 선거부터 시작되어 최근 실시된 2010년 6.2 지방선거에 이르기까지 그 동안 수많은 시행착오를 거쳤으며, 표본설계나 조사 방법에 있어서 많은 발전이 이루어졌다. 특히 2010년 6.2지방선거의 출구조사 결과는 놀라운 예측 정확성을 보여줌으로써 많은 언론매체를 통해 찬사를 받기도 했다. 하지만 우리나라 선거예측조사의 수준은 아직 개선할 여지가 많으며, 특히 출구조사의 경우 응답과정에서 발생하는 비표본추출오차(non-sampling error)를 줄이기 위한 심층적인 연구가 필요하다는 지적이 많다.

총선의 경우 1996년 15대 총선 이후 네 차례에 걸쳐 선거예측조사가 이루어졌지만 아직 만족할 만한 수준의 결과를 얻지 못하고 있다. 1996년 15대 총선 예측조사에서는 전체 253개 선거구 중 39곳을 틀리는 참담한 결과를 보였고, 16대 총선에서는 전체 227개 지역구 가운데 21곳의 예측실패는 물론 제1당 예측에 실패하는 어이없는 결과가 나타났으며, 17대 총선에서는 전국 234개 지역구 중 19지역의 예측에 실패하는 결과가 나왔다(현경보 2005).

가장 최근인 2008년 총선에서도 만족할 만한 결과를 얻지 못했는데, KBS/MBC의 선거예측조사의 경우 243개 선거구 중에서 22개 선거구에서 예측을 실패하였다. 총선 예측조사의 경우 일반적으로 경합도가 높은 선거구에 대해서는 출구조사를 하고 그렇지 않은 선거구에 대해서는 전화조사를 통해 예측을 하게 된다. 2008년 18대 총선의 예측조사에서 나타난 특징적인 점들을 정리해 보면, 우선 과거 선거에 비해 초경합 지역이 늘어났고(1위~2위의 지지율 격차가 4% 이내인 초경합 지역이 17대 총선의 경우 38개이었던 반면, 18대 총선의 경우 51개로 증가했음), KBS/MBC 선거예측조사에서 예측에 실패한 22개 선거구 중 11개가 출구조사, 나머지 11개가 전화조사 대상 선거구였다는 특징이 있었다. 과거 선거예측조사의 경우 당선자를 잘못 예측한 선거구가 대부분 출구조사 지역이었던 점에 비해

2008년 총선 예측조사에서는 과거와는 다른 결과가 나온 것으로 보인다. 이런 현상은 18대 총선의 경우 주요 정당의 공천과정이 한 달 이상 늦어져 각 지역구별 경합도 판정에 어려움을 겪은 것이 원인이었다는 지적이 있다. 이는 전화조사와 출구조사를 병행하여 실시하는 우리나라의 총선 예측조사에서 경합도에 따른 선거구 분류과정이 보다 정교해져야 할 필요가 있다는 점을 시사하고 있다. 아울러 2008년 총선 예측조사에서도 과거와 마찬가지로 여당의 지지율이 실제보다 높게 예측되는 ‘여당 과대 계상’ 문제가 발생하였으며, 이로 인해 일부 선거구의 당선자 예측뿐만 아니라 정당별 의석수 예측에 있어서도 큰 오차가 발생하였다. 이런 오차는 응답자 선정 및 응답 과정에서 발생하는 것으로 일반적인 표본추출이론으로는 설명이 어렵다는 특징을 갖는다.<sup>1)</sup>

하지만 선거예측조사의 정확성에 대한 평가가 단순히 몇 개 지역구에서 당선자 예측에 실패했는가와 같은 인상적인 기준으로 이루어지는 것은 문제가 있다. 당선자 예측에 실패한 상당수의 선거구가 오차범위 내의 예측 실패였다는 점을 감안하면, 당선자를 잘못 예측한 선거구 수를 기준으로 선거예측조사의 정확성을 단순히 평가하는 것은 과학적인 접근으로 볼 수 없다(이준웅 2004). 아울러 선거예측조사의 정확성을 향상시키기 위해서는 경합도가 높은 선거구에 적용되는 출구조사의 조사방법을 개선하는 것이 핵심적인 과제로 볼 수 있다. 따라서 본 연구에서는 총선 예측조사에 사용되는 출구조사의 예측오차에 초점을 두었다. 특히 향후 출구조사의 정확성을 향상시키기 위해서는 앞에서 언급한 것과 같은 표본추출오차뿐만 아니라 ‘여당 과대 계상’ 등과 같은 다양한 오차의 성격과 발생 원인을 다각적으로 분석하여 출구조사 설계과정에 반영하는 것이 필요하다는 점을 고려해 연구를 수행했다는 점을 참고하기 바란다.

우리나라에서는 1999년 3.30재·보선에서 최초로 출구조사가 도입된 이후, 2002년 이후의 대통령 선거와 2000년 이후의 국회의원 총선거 및 2002년 이후의 지방선거에서 TV 방송사를 중심으로 출구조사가 수행되어 왔다. 2010년 지방선거를 비롯해 대통령 선거와 지방선거에서는 비교적 정확한 예측이 이루어졌지만, 앞에서 언급한 것처럼 총선에서는 아직 만족할 만한 결과를 얻지 못하고 있는 것이 현실이다.

출구조사와 관련된 국내 연구동향을 살펴보면, 류재복(2003)은 출구조사의 전반적인 개선방향을 정리하였고, 김정훈(2003)은 16대 대선 출구조사의 기획과 실행과정을 방법론 및 예측오차 중심으로 상세히 정리하였으며, 이준웅(2004)은 17대 총선 예측조사에서 발생한

1) 이 내용은 18대 총선 KBS/MBC 예측조사 사후 평가회의에서 미디어리서치(MBMR) 사회 여론조사본부 김지연 이사가 발표한 내용을 일부 인용한 것임.

예측오차와 편향을 조사거절 및 응답불성실 관점에서 구체적으로 평가하고 개선 방안에 대해 논의했다. 출구조사에서의 투표소 추출방법과 관련해 홍내리·허명희(2001)는 16대 총선 예측조사에서의 표본 투표소 수의 결정과 판단표본추출에 따른 선거예측조사의 문제점을 지적하고 이에 대한 개선방안을 제시하였으며, 조성검·김지연(2004)은 기존에 사용했던 '대표구 추출법'과 '계통추출법'의 효율성을 비교하였고, 김영원·엄윤희(2005)는 2004년 17대 총선 출구조사 자료를 이용하여 계통추출법을 근간으로 한 확률추출법의 효율성과 표본추출오차에 대해 논의하였다. 또한 김영원·김지현(2007)은 2006년 지방선거 출구조사에서 나타난 무응답 및 투표자의 성/연령대에 따른 투표성향과 출구조사 표본추출방법을 정리하고, 출구조사에서 발생하는 표본추출오차를 분석하는 동시에 표본의 대표성을 검증하고 있다. 한편 현경보(2005)는 17대 총선 예측조사에서 나타난 예측오차를 전반적으로 평가하는 동시에 전화조사와 출구조사의 정확성을 비교하였으며, 전체 출구조사 과정을 투표소 추출 과정과 실사 과정으로 구분하여 단계별로 발생한 오차를 분석해 표본추출오차보다는 비표본추출오차가 예측오차의 주된 요인이라는 점을 밝히고 있다. 현경보의 연구는 전체 오차를 투표소추출 관련 오차와 투표자 선정 및 응답 오차 등으로 구분해 분석한다는 관점에서 본 연구의 오차분석 방법과 유사하다고 볼 수 있다.

최근 출구조사 투표소추출 방법에 대한 연구가 주요 관심대상이 되고 있지만, 출구조사를 통한 전반적인 선거예측과정을 살펴보면 투표소 추출 방법뿐만 아니라, 조사원 관리 및 현장 실사 방법, 조사거절과 거짓 응답, 표본투표소에서의 투표자 선정방법, 무응답 및 가중치 적용방법 등 다양한 원인에 의해 예측오차가 발생함을 알 수 있다. 따라서 출구조사의 정확성 제고를 위해서는 출구조사의 특성을 정확히 이해하고 이를 기반으로 현실적인 한계를 감안한 출구조사 개선방안이 모색될 필요가 있다.

따라서 본 연구에서는 각 선거구에서 발생한 오차를 총조사오차(total survey error)로 정의하고, 각 선거구에서 발생한 총조사오차를 투표소 추출 관련 오차와 투표소 내 투표자 선택 및 응답오차로 나눠 분석함으로써, 출구조사의 정확성 제고를 위해 어떤 측면에 보다 집중적인 노력을 기울이는 것이 필요한지 검토하였다. 아울러 비표본추출오차를 나타내는 대표적인 지표로 볼 수 있는 무응답률과 출구조사에서 발생한 오차와의 상관관계를 분석해보았다. 또한 선택된 투표자가 출구조사 참여를 거절함으로써 표본의 대표성이 떨어지는 문제를 다루기 위해 중앙선거관리위원회(이하 '선관위')에서 제공한 선거구별 자료와 출구조사 자료의 성별·연령대별 분포를 비교·분석하였다. 18대 총선에서 KBS/MBC 선거예측조사는 미디어리서치(MBMR)와 코리아리서치(KRC) 2개 조사회사에서 담당하였다. 하

지만 출구조사 원자료 확보에 어려움이 있어 본 연구에서는 미디어리서치에서 조사를 수행한 출구조사 지역만을 기초로 분석이 이루어졌다는 점에 유의하기 바란다.<sup>2)</sup> 미디어리서치는 전국 246개 선거구 중 경합지역으로 분류된 46개 선거구에서 출구조사를 실시하였다.

## II. 출구조사 예측오차와 무응답률 현황

사후적으로 출구조사를 평가하는 대부분의 경우, 각 후보의 득표율 예측치와 실제 득표율의 차이로 정의될 수 있는 예측오차와 함께, 조사관리의 적절성 평가를 위해 무응답률을 우선적으로 살펴보게 된다. 출구조사에서의 무응답은 조사대상 투표자가 조사참여를 거절하거나 일부 조사항목에 응답하지 않을 때 발생하는데, 실제 출구조사에서는 대부분 조사거절에 의해 무응답이 발생한다. 일반적으로 응답자와 무응답자는 여러 가지 특성상 차이가 있는 경우가 많기 때문에 무응답률이 높아지면 조사의 정확도가 낮아지는 것으로 알려져 있다.

따라서 무응답률과 실제 출구조사에서의 예측오차 현황을 살펴보기 위해 18대 총선에서 미디어리서치가 출구조사를 실시한 46개 선거구별 무응답률과 예측오차를 정리하면 <표 1>과 같다. 한편 출구조사의 예측오차를 어떻게 정의하는 것이 적절한지에 대해서는 다양한 의견이 있을 수 있으며(Martin et al. 2005), 선거예측조사에서 오차는 다양한 척도로 설명될 수 있다. Martin et al.(2005)는 선거여론조사에서 그동안 사용되어 오던 다양한 형식의 정확성 척도(measure of poll accuracy)를 정리하고 새로운 척도를 제시하고 있다.

<표 1>의 예측오차는 다음과 같이 계산된 것이다.

$$\text{예측오차(prediction error)} = |(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)|$$

여기서  $\hat{p}_1$  과  $\hat{p}_2$  는 1위 후보자와 2위 후보자에 대한 득표율 예측값,  $p_1$  과  $p_2$  는 실제 득표율을 나타낸다. 제시된 예측오차는 1위와 2위 후보자의 득표율 예측치 차이와 실제 득표율 차이에 대한 절대 차이를 이용한 것으로 조성겸·김지연(2004)이 사용한 척도이며, 이런 척도는 Mosteller et al.(1949)이 제시한 다양한 척도 중 5번째 척도에 해당한다(Martin et al. 2005 재인용).

### <표 1> 46개 선거구별 예측오차와 무응답률 현황

(단위: %p, %)

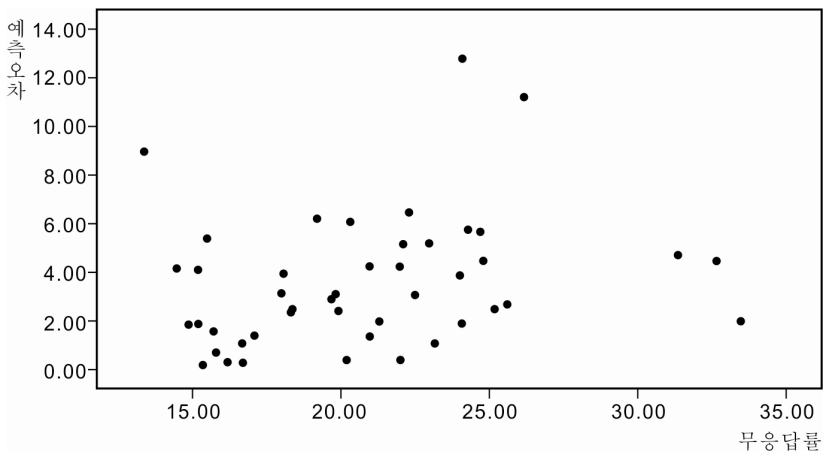
2) 본 연구가 가능하도록 2008년 18대 총선 KBS/MBC 공동예측조사의 출구조사와 선관위 개표 자료의 사용을 허락해 준 미디어리서치와 KBS 관계자분들께 감사드립니다.

선거구	예측오차	무응답률	선거구	예측오차	무응답률
종로구	0.3	20.2	고양덕양(갑)	4.1	14.5
성동(을)	5.1	22.1	고양일산동구	1.5	15.7
중랑(을)	1.9	21.3	남양주(갑)	4.0	15.2
성북(을)	2.8	19.7	시흥(을)	8.9	13.4
강북(을)	0.2	16.7	용인처인	2.4	25.2
도봉(을)	0.6	15.8	이천/여주	6.4	22.3
노원(을)	1.7	14.9	동해/삼척	1.0	23.2
마포(갑)	5.6	24.7	홍천/횡성	0.3	22.0
강서(갑)	1.8	15.2	청주홍덕(갑)	1.3	17.1
구로(갑)	3.1	18.0	청 원	11.2	26.2
금천구	6.0	20.3	증평/진천/괴산/음성	1.8	24.1
동작(을)	1.3	21.0	천안(을)	3.8	24.0
관악(을)	2.3	18.3	서산/태안	4.2	21.0
사하(갑)	3.9	18.1	부여/청양	1.9	33.5
수영구	6.2	19.2	군 산	4.4	24.8
달서(갑)	2.3	19.9	목 포	2.4	18.4
계양(갑)	0.1	15.4	안 동	4.4	32.7
대전동구	4.2	22.0	구미(을)	5.7	24.3
유성구	1.8	14.9	고령/성주/칠곡	12.8	24.1
성남수정	2.6	25.6	창원(을)	5.3	15.5
의정부(갑)	3.0	19.8	통영/고성	4.6	31.4
안양만안	0.2	16.2	제주시(갑)	1.0	16.7
안산상록(갑)	5.1	23.0	서귀포시	3.0	22.5

〈표 1〉을 보면 46개 분석대상 선거구의 평균 예측오차는 3.45%p 이고 고령/성주/칠곡, 청원, 시흥(을) 선거구의 경우 예측오차가 매우 큰 것으로 나타났다. 한편 46개 선거구에서 평균 무응답률은 20.65%인 것으로 나타났으며, 부여/청양(33.5%) 안동(32.7%), 통영/고성(31.4%)등의 지역에서 무응답률이 높은 것으로 나타났다.

선거여론조사를 포함해 일반적인 통계조사에서 조사의 품질을 평가하기 위해 가장 간편하게 사용되는 척도 중 하나는 무응답률이다. 이런 관점에서 선거구별 무응답률과 예측오차의 상관관계를 살펴보는 것이 의미가 있을 것이다. 무응답률과 예측오차의 관계를 보기 위해 산점도를 그려보면 <그림 1>과 같으며, 결코 무응답률과 예측오차가 강한 상관관계를 보인다고 말할 수 없다. 무응답률과 예측오차의 상관관계를 구해 보면  $0.253(p\text{-값}=0.09)$ 로 예상보다 높지 않다.

이런 결과는 예측오차만으로 출구조사의 정확성을 평가하기에는 한계가 있다는 것을 의미한다. 왜냐하면 출구조사는 여러 단계에 걸쳐 상당히 복잡한 과정을 통해 이루어지기 때문에 단순히 최종 예측결과의 차이만으로 산출되는 예측오차로 출구조사에서 단계별로 발생하는 오차를 종합적으로 설명하기에는 한계가 있다. 출구조사 과정에서 발생하는 오차를 단계별로 보면 우선 투표소를 추출하는 과정에서 생기는 표본추출오차, 투표소에서 투표자를 추출하는 과정에서의 표본추출오차, 조사거절이나 거짓 응답에 의한 오차, 조사원들이 실사규칙을 지키지 않음으로써 발생하는 오차 등 매우 다양한 오차가 누적되는 구조를 지니며, 이들 오차들이 서로 상쇄되는 방향으로 발생하는 경우에는 실제 단계별로는 많은 오차가 발생하지만 예측오차는 매우 작게 나타날 수도 있다. 따라서 본 연구에서는 출구조사에서 발생한 전체오차를 출구조사 관련 자료로 산출 가능한 ‘투표소추출오차’와 그밖의 오차(이하 ‘실사오차’)로 구분해 출구조사의 정확성을 평가해 보고자 한다.



<그림 1> 무응답률과 예측오차의 산점도

### Ⅲ. 출구조사의 총조사오차

#### 1. 총조사오차의 정의 및 오차의 분해

출구조사의 정확성을 판단할 때 일반적으로 예측오차의 관점에서 이야기를 하지만 출구조사 과정을 자세히 들여다보면 오차의 발생과정은 매우 복잡하다. 출구조사를 시행하기 위해서 각 선거구 내의 전체 투표소를 조사할 수 없으므로 이 중 일부 투표소를 추출하게 되고, 또한 선거 전에 추출된 투표소의 출구위치와 출구의 개수 등을 확인한다. 필요한 경우 투표소를 대체해 최종 표본투표소를 확정하고, 투표 당일 지정된 위치에서 조사원들이 출구조사를 실시하게 된다. 조사원들은 일정 간격으로 투표자 중 일부 투표자를 선정하여 조사를 실시하지만 조사대상자 중에는 조사거절자나 불성실 응답자 등이 발생한다.

이런 출구조사 과정을 고려해 본 연구에서는 전체 오차를 ‘총조사오차(total survey error)’로 정의하고 전체 투표소 중 일부 투표소를 추출하는 과정에서 발생하는 오차를 ‘투표소추출오차’, 나머지 추출된 투표소에서 투표자 중 일부 투표자를 조사함으로써 발생하는 오차와 실사를 진행하는 과정에서의 조사거절, 거짓응답 그리고 조사원의 조사규칙 불이행 등으로 발생하는 오차는 ‘실사오차’로 구분하기로 한다. 여기서 투표소추출오차는 출구조사에서 사용되는 2단계 집락표본추출에서 1차 집락추출에 따른 표본추출오차(sampling error)에 해당하고, 실사오차는 조사거절 등에 의한 다양한 비표본추출오차(non-sampling error)와 함께 집락(투표소) 내에서 이루어지는 2차 추출과정에서 발생하는 표본추출오차를 포함한 것으로 이해하면 된다. 여기서 이런 방식으로 총조사오차를 투표소추출오차와 실사오차로 구분하는 이유는 실제 방송사 출구조사 자료와 선관위 투표소별 개표 자료를 활용하면 총조사오차를 이와 같은 두 가지 오차로 구분해 산출하는 것이 가능하고, 개략적으로 보면 투표소추출오차는 표본추출오차, 실사오차는 비표본추출오차를 의미한다는 점을 고려한 것이다. 현경보(2005)의 연구에서도 이와 유사하게 출구조사의 전체 오차를 ‘1단계 오차’와 ‘2단계 오차’로 구분하고 있다.

특정 후보의 득표율에 대한 선거예측결과를 기준으로 하면, 출구조사에서 예측치를  $\hat{p}$ , 실제 득표율을  $p$ 로 나타내고, 표본으로 추출된 각 투표소별 개표결과를 이용해 계산한 후보자 득표율을  $\tilde{p}$ 로 표시하면 전체 오차는 다음과 같이 투표소추출오차와 투표소 실사오차로 구분할 수 있다.



$$\text{전체 오차}(\hat{p} - p) = \text{투표소추출오차}(\tilde{p} - p) + \text{실사오차}(\hat{p} - \tilde{p})$$

여기서 표본투표소 집합을  $s$ 로 표시하고, 개표 결과  $i$ 투표소에서 후보자의 실제 득표율을  $p_i$ , 실제 투표자 수를  $N_i$ 라고 하면 투표소별 개표결과로 계산한 후보자 득표율  $\tilde{p}$ 는 다음과 같이 계산한 것이다.

$$\tilde{p} = \frac{\sum_{i \in s} N_i p_i}{\sum_{i \in s} N_i}$$

결국  $\tilde{p}$ 는 표본으로 추출된 투표소에서 오차가 없이 완벽한 득표율 추정이 이루어지는 경우 얻어지는 후보자 득표율 추정값으로,  $(\tilde{p} - p)$ 는 순수하게 투표소추출에 따른 오차로 볼 수 있다.

실제 결과에서는 투표소추출오차인  $(\tilde{p} - p)$ 와 실사오차인  $(\hat{p} - \tilde{p})$ 가 서로 부호가 달라서 상쇄될 수 있기 때문에 본 연구에서는 투표소추출오차와 실사오차의 절대값을 취해 종합하는 방식으로 출구조사의 총조사오차를 정의하기로 한다. 아울러 선거예측조사의 정확성을 평가하고 특히 다른 선거들 간의 조사의 정확성 비교 평가를 위해서는 주요 관심대상인 1위 후보자와 2위 후보자에 대한 예측오차를 종합하는 것이 필요하다는 점을 고려해 (Martin et al. 2005), 본 연구에서는 1위 후보자와 2위 후보자의 투표소추출오차와 실사오차를 종합한 선거구별 총조사오차를 다음과 같이 정의한다.

$$\text{총조사오차(Total Survey Error)} = \{| \hat{p}_1 - \tilde{p}_1 | + | \hat{p}_2 - \tilde{p}_2 | \} + \{| \tilde{p}_1 - p_1 | + | \tilde{p}_2 - p_2 | \}$$

제시된 총조사오차(TSE)는 1위 후보자와 2위 후보자의 선거예측 결과와 개표결과의 차이를 단순히 계산하는 예측오차와는 달리 투표소추출오차와 실사오차를 각각 산출하고 각 오차의 절대값을 합해 2가지 요인에 의한 오차를 종합하는 개념의 척도이다. 이는 투표소추출오차와 실사오차가 서로 상쇄되는 방향으로 발생하여 우연히 예측오차가 적어지는 경우(실제 두 번 틀려서 예측오차가 적어지는 경우)에도 제시된 총조사오차를 사용하게 되면 이런 상쇄 작용을 제어할 수 있기 때문에 출구조사의 정확성을 보다 객관적으로 평가할 수 있다는 점을 고려해 새로 제안하는 척도이다. 참고로 제시된 총조사오차는 2명의 후보

자에 대한 오차를 더한 개념이기 때문에 일반적인 오차와 유사한 척도를 만들기 위해서는 산출된 값을 2로 나누어 사용할 수도 있을 것이다.

분석대상 46개 선거구의 총조사오차를 투표소추출오차와 실사오차로 구분해 정리해 보면 <표 2>와 같다. 본 연구에서 정의한 총조사오차와 예측오차의 관계는 어떤 수식으로 정확히 표현되지는 않지만, 개략적으로 예측오차가 작는데 총조사오차가 커지는 선거구의 경우는 투표소추출오차와 실사오차가 서로 상쇄되는 방향으로 오차가 발생하여 결국 두 번의 오류를 범해 예측치의 정확성이 높아진 경우로 해석할 수 있다. 예를 들어 유성구, 흥천/횡성, 증평/진천/괴산/음성, 부여/청양, 제주시(갑) 등의 경우 예측오차는 작지만 총조사오차가 매우 큰 지역들이다. 따라서 실제 출구조사의 정확성을 향상시키기 위해서는 단순히 예측오차만을 보고 조사의 정확성을 평가하는 것보다 가능한 투표소추출오차와 실사오차를 구분해 접근하는 자세가 필요할 것으로 보인다.

<표 2>에 제시된 46개 선거구의 총조사오차 평균은 6.31이고 선거구별로는 노원(을)의 1.7에서부터 부여/청양의 18.5까지 매우 다양하다. 그러나 대부분의 지역에서 총조사오차 중 투표소추출오차(A)에 비해 실사오차(B)가 차지하는 비중이 상대적으로 크다는 것을 알 수 있다. 46개 선거구의 총조사오차 중 실사오차가 차지하는 평균비율( $B/(A+B)$ )은 66.18%이다. 투표소추출오차가 작다는 것은 특정 선거구에서 표본투표소들이 전체투표소를 설명하는 데 있어서 표본의 대표성이 우수하다는 것을 의미하고, 실사오차가 크다는 것은 표본 투표소에서 출구조사 결과로 예측된 1위 후보자와, 2위 후보자 지지율과 실제 득표율과의 차이가 크다는 것을 의미한다. 실사오차가 커지는 이유로는 출구조사가 이루어지는 과정에서 실사 규칙이 잘 이루어지지 않았거나, 특히 특정 정당을 지지하는 투표자들의 조사 거절률이 상대적으로 높은 것이 주요인일 가능성이 높기 때문에, 결국 무응답률과 실사오차는 밀접한 관계를 가질 개연성이 높다.

아울러 총조사오차가 큰 선거구를 보면 대체적으로 실사오차가 큰 경우가 많으며, 18대 총선의 특수한 정치상황 때문에 자유선진당 또는 친박 무소속 후보 등과 관련된 경우가 대부분이라는 점에 유의할 필요가 있다. 특히 이런 지역은 많은 전문가들이 지적한 것과 같이 특정 정치성향을 가진 투표자들이 상대적으로 조사 참여를 거부하는 경우가 많고, 조사 참여자 중 거짓응답 가능성이 매우 높다고 예상되었다. 예를 들어 대전동구, 유성구, 천안(을), 부여/청양 등은 자유선진당 후보가 당선된 지역이고, 수영구, 안동, 고령/성주/칠곡 등은 무소속 후보가 당선된 지역이다.

〈표 2〉 선거구별 예측오차와 총조사오차

(단위: %p)

선거구	예측 오차	투표소 추출오차 (A)	실사 오차 (B)	총조사 오차	선거구	예측 오차	투표소 추출오차 (A)	실사 오차 (B)	총조사 오차
종로구	0.3	0.7	1.8	2.5	고양덕양(갑)	4.1	3.5	1.4	4.9
성동(을)	5.1	2.7	2.4	5.1	고양일산동구	1.5	1.0	2.1	3.1
중랑(을)	1.9	1.3	1.8	3.1	남양주(갑)	4.0	1.5	4.1	5.6
성북(을)	2.8	1.9	3.5	5.4	시흥(을)	8.9	5.4	3.5	8.9
강북(을)	0.2	1.3	2.5	3.8	용인처인	2.4	4.9	2.7	7.6
도봉(을)	0.6	1.0	1.4	2.4	이천/여주	6.4	2.9	3.5	6.4
노원(을)	1.7	0.1	1.6	1.7	동해/삼척	1.0	0.9	1.3	2.2
마포(갑)	5.6	1.7	3.9	5.6	홍천/횡성	0.3	1.3	4.0	5.3
강서(갑)	1.8	0.6	2.8	3.4	청주홍덕(갑)	1.3	1.3	4.4	5.7
구로(갑)	3.1	0.4	3.3	3.7	청 원	11.2	3.5	7.7	11.2
금천구	6.0	3.3	9.3	12.6	증평/진천/괴산/음성	1.8	3.4	5.2	8.6
동작(을)	1.3	1.4	0.5	1.9	천안(을)	3.8	3.6	7.4	11.0
관악(을)	2.3	1.1	1.6	2.7	서산/태안	4.2	3.2	2.2	5.4
사하(갑)	3.9	0.6	3.3	3.9	부여/청양	1.9	8.3	10.2	18.5
수영구	6.2	1.8	8	9.8	군 산	4.4	3.2	7.6	10.8
달서(갑)	2.3	2.0	4.3	6.3	목 포	2.4	1.6	4.4	6.0
계양(갑)	0.1	1.0	1.9	2.9	안 동	4.4	2.5	6.9	9.4
대전동구	4.2	1.6	5.8	7.4	구미(을)	5.7	2.7	3.0	5.7
유성구	1.8	1.9	6.3	8.2	고령/성주/칠곡	12.8	1.2	13.8	15.0
성남수정	2.6	1.8	4.4	6.2	창원(을)	5.3	1.5	3.8	5.3
의정부(갑)	3.0	0.9	2.1	3	통영/고성	4.6	2.0	4.2	6.2
안양만안	0.2	0.3	1.9	2.2	제주시(갑)	1.0	4.5	5.5	10.0
안산상록(갑)	5.1	1.0	4.3	5.3	서귀포시	3.0	3.4	5.0	8.4

〈표 2〉의 결과를 종합해 보면 대부분의 선거구에서 투표소추출오차가 크게 차이가 나지 않지만 실사오차는 다양한 원인에 의해서 선거구별로 차이가 크게 나고 있으며, 결국 이런 실사오차가 총조사오차를 결정하는 중요한 요인이 되고 있음을 알 수 있다. 따라서 총조사 오차와 실사오차와의 관계에 대해 좀더 자세히 살펴볼 필요가 있다.

## 2. 총조사오차와 실사오차의 관계

총조사오차와 실사오차의 관계를 살펴보기 위해 46개 선거구를 총조사오차의 크기로 정렬한 후 46개 선거구를 5개 그룹으로 균등하게 분할하고(중간인 그룹 3은 10개, 나머지 4개 그룹에는 9개씩 구분), 그룹별로 총조사오차와 투표소추출오차 및 실사오차의 평균을 정리한 결과는 〈표 3〉과 같다. 여기서 5개 그룹으로 구분한 특별한 이유는 없으며 연구자의 편의에 따라 구성한 것이고, 실제 3개 또는 6개 그룹으로 구분한 경우에도 실제 분석결과에 있어서는 큰 차이가 없음을 밝혀 둔다.

그룹 I 은 46개 선거구 중 총조사오차가 18.5로 가장 큰 부여/청양을 포함해 9개 선거구를 포함하고 있으며, 9개 선거구의 총조사오차의 평균은 12.03, 실사오차 평균은 8.49, 투표소추출오차 평균은 3.54인 것으로 나타났다. 그룹별 투표소추출오차와 실사오차를 보면 총조사오차 크기를 기준으로 구성된 그룹 순위에 따라 투표소추출오차와 실사오차가 모두 감소하는 것을 볼 수 있다.

〈표 3〉 총조사오차에 따른 그룹별 평균 오차

(단위: %p)

그룹	선거구	총조사 오차	실사 오차	투표소 추출오차
그룹 1	부여/청양, 고령/성주/칠곡, 금천구, 청원, 천안(을), 군산, 안동, 제주시(갑), 수영구	12.03	8.49	3.54
그룹 2	시흥(을), 증평/진천/괴산/음성, 서귀포시, 유성구, 용인처인, 대전동구, 이천/여주, 달서(갑), 성남수정	7.56	4.52	3.03
그룹 3	통영/고성, 목포, 청주흥덕(갑), 구미(을), 남양주(갑), 마포(갑), 서산/태안, 성북(을), 창원(을), 안산상록(갑)	5.62	3.78	1.84
그룹 4	홍천/횡성, 성동(을), 고양덕양(갑), 사하(갑), 강북(을), 구로(갑), 강서(갑), 중랑(을), 고양일산동구	4.03	2.62	1.41
그룹 5	의정부(갑), 계양(갑), 관악(을), 종로구, 도봉(을), 동해/삼척, 안양만안, 동작(을), 노원(을)	2.39	1.57	0.82

〈표 4〉 총조사오차에 따른 그룹별 분산분석

		제곱합	자유도	평균제곱	F	유의확률
투표소 추출오차	그룹간	46.233	4	11.558	7.590	0.0001
	그룹내	62.430	41	1.5226		
	합 계	108.663	45			
실사오차	그룹간	253.056	4	63.264	37.747	0.0001
	그룹내	68.716	41	1.676		
	합 계	321.772	45			

한편 총조사오차의 크기에 따라 분류된 그룹별로 투표소추출오차와 실사오차가 통계적으로 유의한 차이가 있는지 확인하기 위해 분산분석을 한 결과는 〈표 4〉와 같다. 그룹간의 차이를 보기 위해 등분산을 가정하지 않는 경우 사용되는 Tamhane T2와 Dunnett T3 등에 의한 다중비교 결과, 실사오차에 있어는 모든 그룹간에 유의적인 차이가 있는 것으로 나타났지만, 투표소추출오차의 경우 그룹 5만 다른 그룹들과 유의적인 차이가 있는 것으로 나타났다. 따라서 분산분석 결과 투표소오차나 실사오차 모두 그룹들간에 유의적인 차이가 있는 것으로 보이지만, 다중비교 결과를 참고로 하면 투표소추출오차에 비해 실사오차에서 그룹별로 더 유의적인 차이를 보이고 있다. 이런 결과는 실사오차가 투표소추출오차에 비해 총조사오차에 미치는 영향이 더 크다는 것을 시사한다고 볼 수 있다. 참고로 총조사오차와 실사오차의 상관계수는 .920이고 총조사오차와 투표소추출오차의 상관계수는 .739로, 총조사오차와 실사오차의 상관계수가 투표소추출오차에 비해 더 큰 것을 확인할 수 있다.

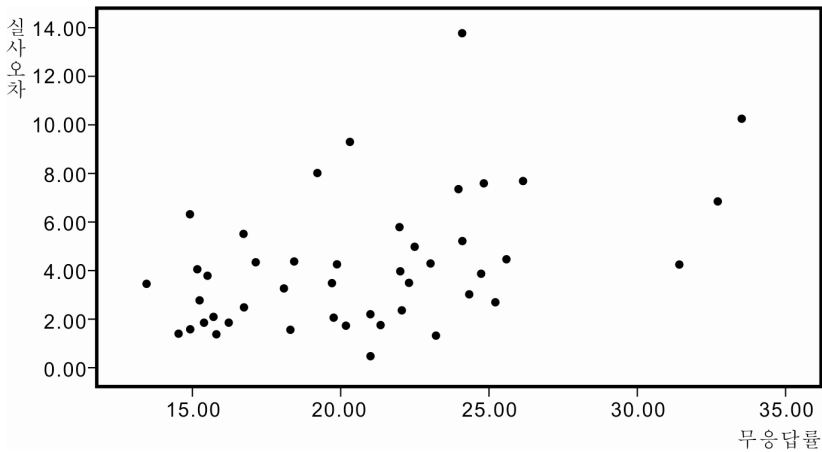
### 3. 무응답률과 실사오차

일반적으로 선거여론조사를 포함한 통계조사에서 무응답률이 높을수록 응답자 선정 등에 있어 편향이 발생할 가능성이 높고(허명희·홍내리 2001), 실제 출구조사에서는 대부분 조사거절에 의해 무응답이 발생한다. 일반적으로 동일한 성별과 연령대에 속하는 응답자와 무응답자의 경우에도 정치성향 등 여러 가지 특성상 차이가 있는 경우가 많기 때문에 무응

답률이 높아지면 특히 실사오차가 커지는 것으로 알려져 있다. 따라서 출구조사가 수행된 46개 선거구에서의 선거구별 무응답률과 실사오차의 관계를 살펴보는 것도 의미가 있을 것이다. 46개 선거구별 무응답률은 <표 1>과 같다. 무응답률이 30%가 넘는 선거구가 3개가 있으며, 평균 무응답률은 20.65%인 것으로 나타났다. 지역별로는 수도권에 비해 지방에서 무응답률이 다소 높은 것으로 나타났다.

무응답률과 실사오차의 관계를 확인해 보기 위해 산점도를 그려보면 <그림 2>와 같다. 그림에서 세로축은 실사오차, 가로축은 무응답률을 나타내며, 두 변수의 피어슨 상관계수는 0.44981( $p$ -값 = 0.0017)이다. II장의 <그림 1>에 제시된 무응답률과 예측오차와의 산점도와 비교해 보면 무응답률과 예측오차의 관계보다 무응답률과 실사오차의 관계가 강하며, 양의 상관관계를 갖고 있음을 확인할 수 있다.

무응답률과 실사오차의 관계를 살펴보기 위해 분석대상 46개 선거구를 무응답률의 크기에 따라 5개 그룹으로 구분하고 각 그룹별 실사오차와 투표소추출오차의 평균을 산출해 보면 <표 5>와 같다. 여기서 그룹 구성은 3% 간격으로 무응답률이 25% 이상, 22%~25% 이하, 19%~22 이하, 16%~19% 이하, 16% 이하인 5개 그룹으로 구분하고 그룹별로 실사오차에 있어서 유의적인 차이가 있는지를 검증해 보았다.



<그림 2> 무응답률과 실사오차의 산점도

〈표 5〉 무응답률에 따른 그룹별 평균오차

(단위: %p)

그룹	선거구	투표소 추출오차	실사 오차
그룹 A (6개)	부여/청양, 안동, 통영/고성, 청원, 성남수정, 용인처인	3.83	6.02
그룹 B (13개)	군산, 마포(갑), 구미(을), 고령/성주/칠곡, 증평/진천/괴산/음성, 천안(을), 동해/삼척, 안산상록(갑), 서귀포시, 이천_여주, 성동(을), 홍천/횡성, 대전동구	2.28	5.17
그룹 C (9개)	중랑(을), 서산/태안, 동작(을), 금천구, 종로구, 달서(갑), 의정부(갑), 성북(을), 수영구	1.83	3.72
그룹 D (8개)	목포, 관악(을), 사하(갑), 구로(갑), 청주흥덕(갑), 강북(을), 제주도(갑), 안양만안	1.39	3.36
그룹 E (10개)	도봉(을), 고양일산동구, 창원(을), 계양(갑), 남양주(갑), 강서(갑), 유성구, 노원(을), 고양덕양(갑), 시흥(을)	1.75	2.89

그룹별 투표소추출오차와 실사오차를 비교해 보면, 무응답률이 25% 이상인 선거구를 포함하는 그룹 A의 경우 실사오차 평균은 6.02, 투표소추출오차 평균은 3.83으로 나타났으며, 무응답률이 낮은(16% 이하) 그룹 E에서는 실사오차 평균은 2.89이고, 투표소추출오차 평균이 1.75로 나타났다. 무응답률을 기준으로 설정된 그룹 순서에 따라 실사오차의 경우 순차적으로 작아지는 것을 볼 수 있지만, 투표소추출오차의 경우 그룹 D와 그룹 E에서는 오히려 평균이 뒤집히는 현상이 나타난다. 이를 근거로 무응답률이 투표소추출오차에 비해 실사오차와 더 밀접한 관계가 있다고 할 수 있을 것이다.

한편, 실사오차가 그룹간에 유의적인 차이가 있는지 통계적으로 분석해 보기 위해 분산 분석을 해 본 결과는 〈표 6〉과 같다. p-값이 아주 크지는 않지만 실사오차가 무응답을 기준으로 구성된 그룹간에 유의적인 차이를 보인다는 것을 확인할 수 있다.

〈표 6〉 무응답률에 따른 그룹별 실사오차 분산분석

	제곱합	자유도	평균제곱	F	유의확률
그룹간	56.832	4	14.208	2.198	0.085
그룹내	264.939	41	6.461		
합계	321.772	45			

참고로 제시한 5개 그룹 가운데 그룹 D와 그룹 E는 46개 선거구의 평균 무응답률 20.65%보다 무응답률이 낮은 지역이다. 그런데 그룹 D와 그룹 E에 속한 지역들의 총조사 오차와 실사오차를 살펴보면 같은 그룹의 다른 지역과 비교하여 그 값이 눈에 띄게 큰 지역들이 있다. 예를 들어 그룹 E에 속해 있는 유성구의 경우 총조사오차는 8.2이고, 실사오차는 6.3으로 같은 그룹의 다른 지역에 비해 매우 큰 값을 갖는다. 이렇게 응답률은 높음에 반해 특히 실사오차가 크다는 것은 조사대상자가 조사에는 응하였지만 거짓응답을 한 사람의 비율이 상대적으로 컸다는 것으로 해석할 수 있다. 자신을 소수파라고 생각하는 유권자들이 자신이 주변의 다수 여론과 다르다고 생각하면 아예 조사참여를 거절하거나 거짓응답하는 경우가 생길 수 있는데 이런 현상을 흔히 침묵의 나선이론으로 설명한다.

아울러 이런 현상을 통해 무응답률이 높을수록 실사오차도 커지기 때문에 좀 더 정확한 예측을 위해서는 단순히 응답률을 높이기 위한 노력을 하면 된다고 생각할 수 있는데, 앞서서도 제시한 바와 같이 특정한 지지성향을 가진 사람들의 조사거절이 두드러진다면 응답은 하되 거짓으로 응답을 하는 사람이 많다면 단순히 응답률을 높여서는 실사오차를 낮출 수 없다는 점에 유의할 필요가 있다. 물론 이런 문제를 현장에서 해결하는 것은 쉽지 않을 것이다. 이론적으로는 각 선거구의 응답과정에서 발생하는 편향(bias)을 통계적인 모형을 이용해 추정하는 작업이 하나의 방안이 될 수는 있지만, 실제 선거마다 이런 편향이 발생하는 패턴이 달라지기 때문에 이런 모형을 개발하는 것 자체가 현실적으로 거의 불가능할 것으로 보인다.

〈표 7〉 무응답률에 따른 그룹별 평균오차 (3개 그룹의 경우)

(단위: %p)

그룹	선거구	투표소 추출오차	실사 오차
그룹 1	부여/청양, 안동, 통영/고성, 청원, 성남수정, 용인처인, 군산, 마포(갑), 구미(을), 고령/성주/칠곡, 증평/진천/괴산/음성, 천안(을), 동해/삼척, 안산상록(갑), 서귀포시	2.94	5.84
그룹 2	이천/여주, 성동(을), 대전동구, 홍천/횡성, 중랑(을), 서산/태안, 동작(을), 금천구, 종로구, 달서(갑), 의정부(갑), 성북(을), 수영구, 목포, 관악(을), 사하(갑)	1.77	3.66
그룹 3	구로(갑), 청주흥덕(갑), 제주시(갑), 강북(을), 안양만안, 도봉(을), 고양일산동구, 창원(을), 계양(갑), 남양주(갑), 강서(갑), 유성구, 노원(을), 고양덕양(갑), 시흥(을)	1.69	3.10



〈표 8〉 무응답률에 따른 그룹별 실사오차 분산분석 (3개 그룹의 경우)

	제곱합	자유도	평균제곱	F	유의확률
그룹간	63.2168	2	31.6084	5.256751	0.00907
그룹내	258.5554	43	6.012916		
합 계	321.7722	45			

한편 〈표 5〉에 제시된 그룹 구성 방법이 상당 부분 연구자의 편의에 의해 결정된 것이기 때문에 객관성에 문제를 제기할 수도 있을 것이다. 따라서 보다 일반적인 결과를 유추해 보기 위해 〈표 7〉과 같이 46개 선거구를 무응답 크기에 따라 상·중·하 3개 그룹으로 균등하게 분리하고 각 그룹별 투표소추출오차 평균과 실사오차 평균을 비교해 보았다. 아울러 그룹별 실사오차의 차이를 통계적으로 검증해 보기 위한 분산분석 결과는 〈표 8〉과 같다. 여기서도 다시 한 번 무응답률이 높아짐에 따라 실사오차가 증가하는 관계가 있다는 것을 확인할 수 있다.

#### IV. 표본의 대표성 검증

표본이 모집단을 잘 대표하는지 확인해 보는 것은 조사의 품질 평가 측면에서 매우 중요한 사안이다. 조사대상자가 거짓응답을 전혀 하지 않고 무응답이 없다는 가정 하에 관심변수에 영향을 주는 변수 관점에서 표본의 구성비가 모집단의 구성비와 같으면 가장 완벽한 조사가 이뤄질 수 있는 환경을 갖췄다고 할 수 있다. 출구조사에서는 실제 조사대상 투표자의 성별과 연령대가 파악 가능한 변수이고, 동시에 후보자 지지성향에 있어서도 성별과 연령이 많은 영향을 주는 중요한 변수라고 알려져 있다. 따라서 출구조사의 대표성을 검증하기 위해 선거구별로 출구조사의 성별 및 연령대별 분포와 실제 투표자 전체 모집단의 성별 및 연령대별 분포를 비교해 보는 것은 출구조사의 대표성을 사후적으로 검증해 본다는 관점에서 의미가 크다고 볼 수 있다.

출구조사를 할 때 표본투표소에서 조사원들은 출구조사 대상자를 계통추출에 의해 일정 간격을 두고 선택을 하지만 실제 조사대상자로 선정된 표본의 성별·연령대별 구성비와 모집단의 구성비는 차이가 있을 수 있다. 특히 실사규칙의 미준수 또는 조사원의 착오 등 실

사관리가 철저히 이루어지지 않는 경우 구성비에 있어서 차이는 더 커지게 되며, 결과적으로 이런 이유로 표본의 대표성에 문제가 발생하면 실사오차가 증가하는 주요한 원인이 될 수 있다.

〈표 9〉 무응답률이 높은 6개 선거구의 성별·연령대별 구성비 비교결과 (단위: %p)

구 분		부여/청양		안 동		통영/고성	
		선관위	출구조사	선관위	출구조사	선관위	출구조사
남	19~29세	5.36	2.98	6.55	3.68	4.61	3.87
	30대	5.07	6.37	7.76	6.53	5.46	8.12
	40대	10.58	12.6	13.38	9.81	11.5	12.6
	50대	10.91	11.52	9.72	11.24	11.41	12.49
	60대 이상	20.27	19.58	12.48	17.57	17.41	15.38
여	19~29세	2.29	2.78	4.25	3.73	3.19	4.55
	30대	5.36	7.11	9.77	7.9	4.58	7.41
	40대	9.42	10.7	13.42	10.3	9.18	10.53
	50대	9.31	8.74	8.36	10.4	10.62	7.94
	60대 이상	21.44	17.62	14.31	18.84	22.02	17.11
구 분		청 원		성남수정		용인처인	
		선관위	출구조사	선관위	출구조사	선관위	출구조사
남	19~29세	6.59	3.73	9.04	5.18	6.39	5.21
	30대	5.9	9.12	10.1	9.88	11.65	10.77
	40대	12.43	13.22	10.77	12.86	14.63	14.35
	50대	11.42	11.65	11.2	12.08	8.99	11.85
	60대 이상	14.59	15.75	12.35	11.64	11.99	13.1
여	19~29세	4.08	4.42	6.07	7.48	3.8	5.86
	30대	5.32	8.57	8.32	9.24	11.72	9.87
	40대	10.13	9.3	9.5	10.07	10.74	9.39
	50대	11.53	8.8	9.73	9.24	8.19	8.19
	60대 이상	18	15.43	12.91	12.32	11.89	11.42

<표 10> 무응답률이 낮은 6개 선거구의 성별·연령대별 구성비 비교결과

(단위: %p)

구 분		시흥(을)		고양덕양(갑)		노원(을)	
		선관위	출구조사	선관위	출구조사	선관위	출구조사
남	19~29세	8.13	4.79	7.09	4.76	6.86	5.44
	30대	14.07	15.47	11.6	11.45	8.15	10.42
	40대	15.13	19.79	15.76	17.58	13.34	16.22
	50대	9.28	9.11	7.53	8.06	9.25	10.97
	60대 이상	8.69	6.89	8.93	9.39	11.23	7.7
여	19~29세	5.92	4.63	6.56	6.03	6.18	7.58
	30대	12.62	14.47	13.97	12.46	9.16	11.84
	40대	10.11	12.63	13.76	13.18	12.5	14.83
	50대	6.62	4.89	5.85	6.91	9.42	8.07
	60대 이상	9.44	7.32	8.95	10.18	13.92	6.92
구 분		유성구		강서(갑)		남양주(갑)	
		선관위	출구조사	선관위	출구조사	선관위	출구조사
남	19~29세	7.3	5.18	8.49	5.79	7.42	4.5
	30대	11.69	13.6	10.85	9.92	9.89	10.67
	40대	15.43	17.04	9.48	13.06	12.88	14
	50대	9.06	10.2	11.1	12.12	10.18	10.4
	60대 이상	8.5	7.03	11.47	9.73	12.12	11.75
여	19~29세	5.01	6.79	7.36	7.37	5.48	5.34
	30대	12.65	14.07	9.77	10.95	11.01	10.85
	40대	13.48	13.67	9.64	12.21	11.04	12.09
	50대	7.61	6.82	11.66	10.8	8.62	8.56
	60대 이상	9.27	5.61	10.17	8.06	11.35	11.85

출구조사 표본과 모집단의 성별·연령대별 구성비를 비교하기 위해서는 각 선거구별로 투표자 모집단 전체의 구성비를 알아야 하지만 선관위에서 공개하는 것은 전체 모집단의 10% 정도에 해당하는 투표자 표본에 대한 분석자료이므로 이 자료를 이용하여 구성비를

비교하기로 한다. 2008년 18대 총선에서 투표자 모집단은 전국 13,246개 투표구에 선거인 37,796,035명이었는데 선관위는 이 중 1,379개 투표구에 선거인 3,901,220명(전체 선거인의 10.3%)을 표본으로 추출하여 분석한 결과를 제시하고 있다. 선관위에서 투표자 표본을 추출하는 과정은 중앙선거관리위원회(2008)를 참고하기 바란다.

본 연구의 분석대상이 되는 46개 선거구에 대한 출구조사와 선관위의 성별·연령대별 구성비 비교결과를 모두 여기에 제시하는 것은 어렵기 때문에 여기서는 선거구의 무응답률을 기준으로 무응답률이 높은 6개 선거구와 무응답률이 낮은 6개 선거구에 대한 비교 결과를 정리해 <표 9>와 <표 10>에 제시한다. 이들 12개 지역에 대한 분석결과를 제시한 이유는 일반적으로 무응답률이 높으면 표본의 대표성이 떨어지는 것으로 생각하기 때문에 무응답률이 높은 선거구와 무응답률이 낮은 선거구의 성별·연령대별 구성비가 출구조사와 선관위 자료에 있어서 어떤 차이가 있는지 독자들이 직접 확인해 볼 수 있도록 하기 위한 것이다.

아울러 출구조사와 선관위의 성별·연령대별 구성비가 통계적으로 유의한 차이가 있는지 확인하기 위해 각 선거구별로 카이제곱 검정을 수행해 보았다. 카이제곱 검정결과 모든 선거구에서 출구조사 표본의 구성비와 선관위 표본의 구성비가 유의적으로 차이가 있는 것으로 나타났다. 구체적인 분석결과는 생략하기로 한다. 이런 결과는 김영원·김지현(2007)의 연구에서 2006년 지방선거 출구조사 자료를 대상으로 한 분석에서 얻은 결과와 일치한다.

실제 이런 차이가 발생하는 원인을 살펴보면, 우선 선관위 표본에는 부재자 투표가 포함된 반면 출구조사 자료에는 제외되어 있기 때문에 특히 군 입대 등으로 부재자 투표가 많은 20대 남성의 구성비에 있어서 차이가 나는 것은 당연하다고 볼 수 있다. 아울러 선관위의 자료는 만 나이를 기준으로 하고 있으나 출구조사에서는 실제 응답자에 따라서 만 나이로 응답한 경우도 있지만 대부분 일반 나이로 응답한 경우가 많기 때문에 이 또한 차이가 발생하는 원인이 될 수 있다. 또한 조사거절자의 비율이 20% 수준에 이르는데 이들의 경우 조사원들이 조사대상자의 나이를 추측해 모의 투표용지에 기입하고 이를 근거로 출구조사의 성별·연령대별 구성비를 산출한 것이기 때문에 여기서도 일부 오류가 발생할 수 있다. 아울러 출구조사의 표본투표소 추출과정에서도 해당 투표소의 유권자 성별·연령대별 분포를 고려하지 않기 때문에 표본투표소의 성별·연령대별 분포가 전체 선거구의 성별·연령대별 분포를 그대로 따른다는 것을 보장할 수 없다.

따라서 <표 9>와 <표 10>에 제시된 선관위와 출구조사 자료의 비교결과는 부정확한 측면이 있다는 점을 감안해 참조하기 바란다. 하지만 일부 부정확한 측면이 있더라도 출구조사

의 대표성을 검증한다는 측면에서 나름대로 의미가 있는 정보로 활용될 수 있을 것이다. 비교 결과를 살펴보면 부재자 투표의 영향이 큰 20대 남성의 경우를 제외하고는 출구조사에서의 성별·연령대별 구성비가 선관위의 구성비에 비해 체계적으로 어떤 편향을 갖는 것으로 보이지는 않으며, 각 투표소별로 다른 경향을 보이고 있음을 볼 수 있다. 이는 현행 출구조사의 실사과정이 특정 연령층이나 성별을 과소 또는 과대 추출하는 체계적인 문제를 갖고 있는 것은 아니라고 긍정적으로 해석할 수도 있을 것이다. 이에 대한 보다 심층적인 분석은 향후에 고려될 필요가 있을 것이다.

## V. 요약 및 시사점

본 연구에서는 2008년 치러진 18대 국회의원 선거의 출구조사 자료와 KBS 개표자료를 이용하여 출구조사 예측결과의 오차 발생 원인에 대해 알아보았다. 출구조사의 정확도를 평가할 때 주로 사용하는 예측오차는 단계별로 발생하는 오차가 서로 상쇄되어 우연히 최종 예측오차가 작게 나올 수도 있다는 한계가 있다. 본 연구에서는 이를 극복할 수 있는 총조사오차 개념을 정의하고 전체 오차를 투표소추출오차와 실사오차로 구분하여 출구조사에서 발생하는 오차의 성격을 구체적으로 살펴보았다.

분석대상인 출구조사 자료 확보가 가능한 46개 선거구의 총조사오차의 크기는 매우 다양한데, 일부 선거구를 제외한 대부분의 선거구는 총조사오차에서 실사오차가 차지하는 비중이 매우 컸다. 이는 출구조사에서 발생하는 오차는 흔히 표본추출오차로 설명되는 투표소추출 과정에서 발생하는 오차보다는 투표소 내에서 투표자를 선정하는 과정이나 투표자로부터 응답을 얻는 과정에서 발생하는 실사오차가 더 크다는 것을 의미한다. 따라서 향후 우리나라 출구조사의 정확성 제고를 위해서는 실사과정에서 발생하는 다양한 오차를 줄이는 방안을 모색하기 위한 연구가 필요하다는 것을 시사한다.

대부분의 조사방법 연구자들은 무응답률과 선거예측조사의 정확성이 매우 높은 상관관계를 갖는다고 생각하고 있다. 이런 측면을 확인해 보기 위해서 무응답률과 총조사오차를 비교해 본 결과 무응답률이 높아질수록 총조사오차는 커지고 총조사오차 중 특히 실사오차가 무응답률과 강한 양의 상관관계가 있음을 확인할 수 있었다. 무응답이 조사자 선택 편향을 불러오며 잘못된 예측을 하게 되는 원인이 된다는 것을 간접적으로 확인할 수 있었다. 하지만 무응답률이 낮다고 해서 항상 실사오차가 작다고 볼 수는 없다는 점도 간과하면 안

될 것이다. 예를 들어 무응답률이 낮은 시흥(을)이나 유성구의 경우에도 총조사오차는 평균 이상인 것으로 나타나 무응답률이 출구조사의 정확성을 설명하는 절대적인 지표가 될 수 없다는 점도 유의해야 할 것이다.

한편 출구조사 표본의 대표성을 검증하기 위해 선관위에서 제공하는 성별·연령대별 자료를 이용하여 카이제곱 검정을 실시하였다. 그 결과 46개 전 선거구에서 출구조사 표본과 선관위 표본의 성별·연령대별 구성비에 차이가 있다는 결론을 얻었다. 물론 부재자가 출구조사 자료에서 제외되는 등 성별·연령대별 비교에 사용된 출구조사 자료의 신뢰성에 일부 문제가 있기 때문에 이런 차이에 너무 집착할 필요는 없을 것이다. 한편 전체적인 비교 결과를 종합해 보면 현재 출구조사의 실사과정이 어떤 체계적인 결함이 있어 특정 성별 또는 연령대에 대해 과대 또는 과소 표본추출하는 문제는 없다는 점을 확인할 수 있었다는 점이 시사하는 바가 더 클 수도 있다.

결국 본 연구 결과를 종합해 보면 18대 총선에서 치러진 출구조사는, 총조사오차의 개념으로 평가해 볼 때, 각 선거구내 투표소 추출의 문제보다는 투표소내 투표자 선정 및 실사과정에서 발생하는 문제가 더 컸다고 할 수 있다. 정확한 선거예측을 위해서 조사방법상의 모든 문제를 개선하는 것에는 한계가 있을 것이다. 본 연구에서 확인했듯이 출구조사는 투표소를 대표성 있게 추출하는 것 이외에도 출구조사를 실시하는 과정에서 발생하는 문제를 통제할 수 있어야 정확한 예측을 할 수 있다는 점을 보여주고 있다. 근본적으로 출구조사에 대한 사회인식이 긍정적으로 변화되어야 하겠지만 결국은 조사거절자와 거짓응답자들의 성향을 파악할 수 있는 모형이 개발되어 이를 출구조사 예측에 반영할 수 있어야 할 것이다. 또한 출구조사에 대한 철저한 사전준비로 실사과정에서 발생할 수 있는 문제를 최대한 예방하는 것도 출구조사의 정확성을 제고하는 데 큰 역할을 할 수 있을 것으로 생각된다.

## 참고문헌

- 김영원·김지현. 2007. “2006년 지방선거 출구조사 현황 및 예측오차.” 《조사연구》 8(1): 55-79.
- 김영원·엄윤희. 2005. “출구조사를 위한 투표소 확률추출 방법.” 《조사연구》 6(2): 1-32.
- 김정훈. 2003. “선거예측과 출구조사: 16대 대선을 중심으로.” 《조사연구》 4(2): 87-102.
- 류제복. 2003. “출구조사의 역사와 개선방향.” 《조사연구》 4(1): 31-48.
- 이준웅. 2004. “제17대 총선 예측조사의 문제: 조사거절자와 응답불성실자 편향을 중심으로.” 《언론정보연구》 41(1): 110-136.

- 조성겸·김지연. 2004. “출구조사의 투표소 표집방안 비교.” 《조사연구》 5(2): 3-29.
- 중앙선거관리위원회. 2008. “제18대 국회의원선거 투표율 분석. 보도자료.” (<http://www.nec.go.kr/>)
- 현경보. 2005. “선거 예측조사의 정확도와 그 영향요인 연구: 17대 총선 예측조사 결과의 사후 분석.” 《언론과학연구》 5(1): 301-336.
- 홍내리·허명희. 2001. “제16대 국회의원선거의 예측조사에 대한 사후적 검증.” 《조사연구》 2(1): 1-35.
- Martin, E.A., M.W. Traugott and C. Kennedy. 2005. “A Review and Proposal for a New Measure of Poll Accuracy.” *Public Opinion Quarterly* 69(3): 342-369.
- Mosteller, F., H. Hyman, P.J. McCarthy, E.S. Marks and D.B. Truman. 1949. *The Pre-Election Polls of 1948: Report to the Committee on Analysis of Pre-Election Polls and Forecasts*. New York: Social Science Research Council.

<접수 2010/10/11, 수정 2010/11/6, 게재확정 2010/11/12>