

제5장

현장조사 개선 방안

— 무응답 실태분석과 무응답 오차 평가 —

김서영 · 권순필

제1절 서론

1. 연구의 배경

인간의 생각과 행동을 증명하기 위한 도구로서 표본조사의 유일한 가치는 거대한 모집단을 편향(bias) 없이, 그리고 불확실성에 대해 측정 가능한 수준 내에서 설명 가능하다는 것이다. 이러한 추론을 위해 필요한 수학적 확률 이론이 거의 100여 년 전에 수립되었고(Pearson, 1903), 수학적 이론들은 1930년대 초반에 인간이 표본으로 사용되면서 확률적 이론에 근거한 조사통계 기틀을 마련하였다(Neyman, 1934). 그러나 실제 사회현상은 이론적 근거에 의해 정확하게 설명될 수 없는 한계에 부딪히는 경우가 발생하는 것도 사실이다. 그럼에도 불구하고 조사라는 방법은 현대사회의 필수적인 정보수집 도구로서 막강한 힘을 발휘하고 있다. 게다가 최근에 들어서면서 이들 조사의 힘(power of survey)은 확률표본을 통해서 집단 전체에 대한 정보를 관측하게 된다는 사실을 재인식하게 하였다. 즉, 원래 이론은 무응답이 없다는 것을 전제로 하고 있다. 그러나 불과 수십 년 전부터 선진국에서는 무응답자들이 이미 증가했던 것으로 알려졌다(de Leeuw & de Heer, 2001). 따라서 거대한 모

집단을 설명하기 위한 조사 능력을 제대로 발휘하기 위해서 조사환경과 특성에 따른 무응답의 잠재적 효과를 이해하는 것은 매우 중요하다 (Groves, 2006).

최근 연구들에 따르면 무응답률의 변화가 반드시 조사 추정치의 변화를 초래하는 것은 아니다(Lynn et al., 2002; Stoop, 2005; Keeter et al., 2000; Curtin et al, 2000). 그렇다 하더라도 대부분 조사 연구자들의 가장 공통적인 관심은 무응답률을 줄이는 것이다. 예를 들면, Alreck & Robert(1995)는 가능한 한 무응답을 줄이고 적절한 응답률을 유지하는 것이 중요하다고 하였다. Babbie(2007)는 사회연구 문헌 검토에 의해 최소한 50%의 응답률이 분석하기에 적정하고, 60%의 응답률은 좋은 편, 70% 응답률은 매우 좋다고 과감하게 언급하고 있다. 마지막으로 Singleton & Straits(2005)는 응답률에 신경을 쓰는 것은 매우 중요하며 면접조사의 경우 적어도 85%의 응답률이 적정하고, 70% 이하는 심각한 편향을 가져올 수 있다고 언급하고 있다.

이처럼 조사에 있어서 서로 상반되는 요구 사항들이 조사의 혼란을 야기하고 있다. 즉, 100%의 응답률을 요구하는 추정 패러다임, 줄어드는 응답률, 무응답률이 무응답 편향을 예측하지 못한다는 증거, 그리고 응답률을 최대화시키려는 노력들은 서로 다른 개념으로 해석되고 있으며 조사에 있어서 조사 실무자들에게 혼란을 야기할 수 있다. 따라서 이들 복잡한 요인들 간의 상호 조합은 실제 조사업무 관계자들 간에 발생하는 혼란을 줄이는 하나의 방법일 수 있다(Groves, 2006).

많은 조사에서 그렇듯이 조사에의 참여 여부는 조사를 권유받은 사람이 자유롭게 결정하도록 하고 있다. 그렇다면 조사 설계, 조사원의 태도, 표본 가구의 조사에 대한 사전 경험, 면접자와 가구의 상호관계, 그리고 조사에 대한 사회적 관심 중 어떤 요인들이 조사를 성공적으로 완성하는 데 작용할 것인가? 경우에 따라서는 모든 요인이 서로 유기적으로 관여할 수도 있고, 때로는 특정 요인만으로도 조사를 성공적으로 완성할 수 있다(Groves & Couper, 1998). 따라서 조사의 성격과 상황에 따라 어떤 요인에 더 집중적으로 노력을 투자할 것인지에 대한 전략을 세움으로써 조사원의 노력과 경제적 비용을 절감할 수 있다. 이는 현장조사에 투입되는 조사원의 노력과 노력의 대가로 자료의 품질은 얼마나

향상될 수 있는지에 대한 연구가 선행됨으로써 보다 정확하게 파악될 수 있는 부분이라 하겠다.

뿐만 아니라 자료의 품질은 무응답 오차와 직접적인 관련이 있다. 그리고 무응답 오차는 무응답률과 서로 상관되어 있다. 무응답이 발생하는 원인은 여러 가지(부재, 거부, 기타 등)가 있으며, 무응답 원인별로 무응답 오차에 기여하는 정도는 다르다. 무응답 원인별로 무응답 오차의 특성을 분석하는 것은 조사방법 측면에서는 가구에 대해서 차별화된 재접촉을 시도함으로써 무응답률을 줄이는 효과를 기대할 수 있고, 자료의 품질 측면에서는 전체 무응답에 더 많은 기여를 하는 무응답 층을 집중적으로 분석함으로써 자료 품질 향상에 기여할 수 있다는 데 그 의미가 있다. 이러한 관점에서 볼 때 지금이 조사에서 발생하는 무응답, 무응답률과 무응답오차에 대한 의미를 파악하고 관련된 제반사항들을 점검해 볼 만한 가장 적절한 시기라 여겨진다.

2. 연구목적 및 필요성

조사에 대한 부담은 가구를 대상으로 하든 사업체를 대상으로 하든 그 정도는 다르지만 항상 존재하기 마련이다. 특히, 가구대상조사의 경우 무응답 가구 설득에 대한 부담이 적지 않고, 2007년 연동표본제의 도입으로 가구의 응답부담은 다소 줄었다고 볼 수 있지만 조사원의 조사 부담은 더 증가했다는 것이 현장의 목소리이다. 물론 현장의 목소리에 너무 의존하다 보면 조사에 대한 일관된 정책을 수립하는 데 상당한 어려움이 따르게 될 것이다. 그러나 이들을 전적으로 무시한다면 보다 효율적인 조사방법 및 조사체계를 찾아내는 것이 쉽지 않을 것이다. 조사 설계자는 현장조사 현황을 충분히 파악하고 검토하고 이를 추후 조사 설계를 위한 정보로 사용할 수 있어야 할 것이다.

가구를 대상으로 하는 통계조사 중 경제활동인구조사(경활조사)는 기존 고정표본제에서 연동표본제로 전환되면서 매달 약 180조사구에서 900여 가구가 새로운 표본으로 교체되고 있다. 이에 따라 각 지방청(출장소)에서는 매월 신표본 가구 설득에 대한 노력이 고정표본제에 비해 가중되고 있는 것도 사실이다. 신표본의 경우, 기존 표본에 비해 응답이

불안정하고 조사원들의 노력 또한 이미 조사에 참여하고 있는 가구 방문에 비해 훨씬 소비적이라는 것이다. 따라서 연동표본제 도입으로 경활조사 실태본 가구에 대한 조사원의 노력과 이에 대한 가구의 응답실태를 분석함으로써 조사부담 정도를 측정하고, 이로부터 파생되는 문제점을 풀어나갈 필요가 있다.

한편 무응답 가구를 설득하려는 노력은 조사에 대한 응답률을 높이고자 하는 하나의 방법에 해당된다. 그렇다면 “응답률은 왜 높아야 하는가?”라는 원초적인 질문을 던질 수 있다. 일반적으로 조사를 통해 얻고자 하는 궁극적인 목적은 가능한 한 오차 없이 모집단에 대해 정확한 정보를 얻고자 하는 것이다. 그러나 조사환경이나 조사대상자들은 이를 뒷받침할 만큼 적극적이지 못하다. 어떤 조사의 경우, 특정 연령이나 학력 또는 특정 질병을 가진 그룹에서 응답을 더 많이 거부할 수 있다. 이는 조사에 따라 목적변수의 추정치에 대한 심각한 편향을 유발할 수 있는 핵심요인이 될 수 있다. 이러한 이유에서 무응답 원인별로 나타나는 응답가구의 특성을 분석하는 것은 무응답 오차의 특성을 파악하는 데 중요하며, 이로부터 무응답률 및 품질향상과 관련하여 무응답 원인별로 차별화된 대책을 강구할 필요가 있다.

3. 연구내용 및 기대효과

본 연구는 크게 두 가지 측면에서 접근한다. 첫째는 현장조사에서 나타나는 가구의 무응답 실태를 파악하기 위해 조사원이 직접 작성한 방문기록 자료를 분석하는 것이고, 둘째는 이러한 실태분석 결과를 기반으로 무응답 오차를 무응답 유형별로 평가하고자 한다. 현장조사 현황은 인력, 예산, 조사방법, 응답률 등을 포함하여 매우 다양한 측면에서 파악할 수 있다.

본 연구는 실태본 설득에 들어가는 조사자들의 노력정도와 가구의 반응 태도를 중심으로 응답실태를 파악하고, 관련 변수들 간의 연관성을 분석하는 데 초점을 맞추고자 한다. 조사원의 노력정도는 응답률과 응답전환율을 계산을 통해 파악하고, 응답전환 패턴이 거부이유별, 방문시기(시간, 요일)별 또는 가구의 반응에 따라 어떻게 다른지를 분석함으

로써 살펴보고자 한다. 이는 최초 대면시 가구의 거부 이유나 반응 정도 그리고 방문시기에 따라 응답전환 패턴이 다를 것이라는 기대에서 비롯된 것이다. 이처럼 초기 대면과 관련한 자료 분석은 조사원들의 응답전환을 위한 재방문 시기, 횟수 및 숙련된 조사원 투입 등과 같은 효과적인 대처 방안을 강구하는 데 중요한 정보가 될 것으로 기대한다.

또한 무응답 오차는 응답실태분석 결과에 기반하여 무응답유형별로 평가한다. 선행연구에 따르면 표본가구로서 응답가구와 무응답가구의 오차 간에는 서로 다른 현상을 보일 수 있다(Groves et al., 2001). 게다가 대부분 무응답 오차가 응답 거부와 부재에 의한 오차 중 응답 거부로 인한 오차가 클 것이라는 것이다. 이에 대해 Groves(2006)와 Heerwgh et al.(2007)은 무응답률 중 부재로 인한 무응답률이 거부에 의한 무응답률보다 상대적으로 더 낮은 반면에 오차의 경우에 있어서는 부재에 의한 오차가 더 큰 것으로 보고하고 있다. 무응답 오차는 응답자 자료에 기초한 구체적 주택유형 비율과 응답자와 무응답자를 모두 포함한 자료에 기초하여 추정치 비율을 구하고, 이 전체 추정치 비율과 응답자 그룹에서의 추정치 비율 간의 차이로 정의한다. 무응답 오차의 특성에 관한 연구는 가구의 응답패턴과 조사원들의 가구 방문 현황을 정확하게 파악함으로써 비로소 무응답 원인별 특성을 정확하게 평가할 수 있게 된다.

연구목적 달성을 위해 우선 경황조사 신표본 가구를 대상으로 응답 실태를 조사원들이 작성한 방문기록표를 통해 분석하고, 이를 기반으로 무응답 원인별 오차들의 특성을 결정하고자 한다. 첫째, 연동표본도입으로 인해 행하는 매월 신표본 설득에 대한 가구의 대응현황 및 조사원 노력에 대한 실태를 파악하고, 둘째, 무응답 오차를 무응답 원인별로 측정함으로써 오차 특성에 따라 자료품질향상에 필요한 차별화된 정책을 위한 기반을 제시하고자 한다. 이를 위해 국내 통계조사에서 무응답 특성은 추정치에 어떻게 반영되고 있고, 그것이 의미하는 것이 무엇인지에 대해 무응답 오차 유형별로 정확하게 진단할 필요가 있다. 이는 자료품질 향상과 관련하여 무응답률과 오차 제어에 필요한 중요한 절차가 될 것이다.

논문의 구성은 다음과 같다. 제2절에서 자료 설명과 응답실태 분석 결과를 제시한다. 제3절에서는 무응답오차 분석결과를 제시하고, 마지

막으로 제4절에서 연구에 대한 결과를 요약하고, 본 연구결과와 관련하여 향후 추진되어야 할 사항들에 대해서 포괄적으로 논의한다.

제2절 응답실태분석

1. 자료설명

연구 자료는 방문기록 자료와 표본명부 자료(통계청, 2007)를 사용하였다. 방문기록 자료는 조사원이 가구를 방문한 현황을 <부록 1>의 조사표에 직접 기입하는 방법이 사용되었다. 조사변수는 총 6개로 ① 최초 대면시 응답거부 여부, ② 응답거부 이유, ③ 화를 내는 정도, ④ 방문일시(요일), ⑤ 총 방문횟수(그 중 부재수), ⑥ 최종응답여부이다. 방문일시는 최초방문과 마지막 방문에 대해서만 방문시간과 요일을 기재하도록 하였다. 이는 조사원들의 적극적인 참여를 위해 가능한 한 조사표 작성 항목을 줄이기 위한 것이다. 방문가구는 전국 47개 지방청(출장소)의 2007년 10월 경황조사에 새롭게 포함되는 신표본 가구를 대상으로 하였다. 조사기간은 신표본 가구 설득에 들어가는 첫날부터 경황조사 본조사가 시작되기 바로 직전 주말까지로 정하고, 이 기간에 이루어진 가구의 설득과 방문현황을 기재하도록 하였다. 이런 점에서 방문기록 자료는 두 가지의 한계를 지니고 있다. 하나는 각 지방청(출장소) 또는 조사원마다 설득을 위한 작업 스케줄이 다르기 때문에 조사기간을 명확하게 정의할 수 없었다는 것과 다른 하나는 동일 가구를 여러 번 방문할 경우 처음과 마지막 방문에 대해서만 정보를 기록했기 때문에 그 사이에 이루어지는 방문에 대한 정확한 현황을 파악할 수 없다는 것이다. 따라서 본 연구는 방문횟수가 증가함에 따라 응답거부율, 전환율, 부재율의 변화 형태와 이들 간의 연관성 여부를 파악할 수 없다는 단점을 지니고 있다.

표본명부 자료는 무응답오차 계산에 필요한 핵심 변수 선정을 위해 사용되었다. 표본명부 자료는 경황조사 표본추출을 위해 사용되는 특성지표가 포함되어 있으며, 특히 목표 추정치(실업자수, 취업자수) 등에

영향을 주는 변수들이 포함되어 있다. 무응답 오차 측정을 위한 핵심변수로 고려할 만한 변수로는 주택유형(단독주택, 아파트, 연립주택, 다세대 주택, 기타 등)과 주택소유여부, 가구주산업, 가구주직업 등을 들 수 있다. 핵심 변수 선정에 대한 자세한 내용은 제3장 무응답오차 분석에서 설명하기로 한다. 조사 자료에 대한 집계현황은 다음과 같다.

무응답	조사대상 전체표본		이용가능 표본	
	n	%	n	%
응답	872	89.25	872	91.69
부재	25	2.56	25	2.63
거부	54	5.53	54	5.68
기타	4	0.41	-	-
이용불가	22	2.25	-	-
총합	977	100	951	100

경활조사 10월 신표본 구역인 02구역 가구를 대상으로 조사한 결과 977가구가 조사되었다. 977가구 중 기타(빈집)는 4가구로 응답이 되지 않았고, 22가구는 표본명부에 주어진 코드와 매치가 되지 않아 이용하지 않기로 하였다. 따라서 최종적으로 26가구를 제외한 951가구만을 유효 표본으로 간주하여 본 연구를 위한 분석 자료로 사용하였다.

2. 분석결과

본절에서는 수집된 951개 경활조사 신표본 가구의 응답 현황을 분석한 결과를 토대로, 조사원의 노력도와 응답가구의 반응을 중심으로 설명하였다. 조사원 노력도는 응답률과 무응답률, 방문횟수와 설득기간 및 방문시간(요일) 분석을 통해 측정하였다. 응답가구의 반응은 응답거부 이유와 거부할 때 보이는 감정표현(화를 내는 정도)의 정도 분석을 통해 파악하였다. 무응답 오차분석과 관련한 정보는 내용의 연결성을 유지하기 위해 3장에서 별도로 제시하기로 한다.

가. 조사원 노력도

조사원 노력도는 최종 응답률과 응답전환율을 통해 측정하였다. 응답전환율은 최초 대면에서 표본가구의 거부 이후 조사원의 노력 정도에 따라 변화되는 값으로, 최초 거부 이후 조사원들의 방문시간(요일)대의 변경 내용을 반영한다. 방문시간대 변경은 변경정도와 경향에 대해서 근무시간대·외 방문과 주말·주중으로 나누어 무응답 유형별로 살펴 보았다.

1) 응답률과 무응답률

<표 5-1>은 응답자와 무응답자 분포를 나타낸다. 여기서 응답자란 최초 대면시에 거부 없이 조사에 응답을 한 그룹이고, 응답전환자란 최초 대면시에 일단 거부를 한 후에 재방문시의 설득에 의해 조사에 응답한 자를 말한다. 무응답자는 부재, 거부, 기타(질병, 장애 등)의 이유로 응답을 하지 않은 자를 말한다. 본 연구는 조사대상을 신표본 가구로 제한하였고, 기타에 해당하는 무응답 유형이 극소수(4가구)로 파악되었기 때문에 무응답 유형은 부재와 거부 두 가지만을 고려하였다.

응답률 계산은 연구자에 따라 조사 참여자를 어떻게 정의하느냐에 따라 조금씩 다르게 적용될 수 있다. 여기서는 일반적인 적용 방법으로 조사에 참여한 유효 표본 가운데 조사에 응답한 가구의 비율로 정의하였다. 최종 응답률은 전체 951 유효표본 가구 중 91.7%로 신표본 가구가 경찰조사에 참여하겠다는 비율이 상당히 높게 나타났다. 이러한 결과는 경찰조사가 질문지가 비교적 간단하여 응답자 부담이 크지 않다는 점을 감안하면 그다지 놀라운 것은 아니다. 한편, 응답자 중 한 번도 거부하지 않고 조사에 참여하겠다고 응답한 최초 응답자 비율은 61.8% 정도로 최종 응답률 91.7%에 비하면 그리 높지는 않다. 그러나 29.9%의 응답자가 최초 거부로부터 최종 응답으로 전환되었고, 이는 조사원들의 재방문 등을 통한 설득 노력의 효과라 할 수 있다. 즉, 최초 대면에서 응답을 거부한 363가구 중 약 78.2%(284/363)가 응답 전환된 것으로, 전환 비율이 높게 나타났다. 한편, 무응답률은 전체 951가구 중 8.3%로 나타났다. 이 중 거부로 인한 무응답은 5.7%로 전체 무응답률의 68.7%, 부재

로 인한 무응답은 2.6%로 전체 무응답률의 31.3%를 각각 차지하였다. 예상대로 전체 무응답 중 거부로 인한 무응답이 부재로 인한 무응답보다 상대적으로 높게 조사되었고, 이들 부재로 인한 무응답 층은 조사원들의 더 많은 노력을 필요로 하는 잠재 요인 층으로 분류될 수 있다.

〈표 5-1〉 신표본 설득에 대한 응답현황

		N	%
응답자	최초응답자 ¹⁾	588	61.83
	응답전환자 ²⁾	284	29.86
	부분 합	872	91.70
무응답자	부재로 인한 무응답	25	2.63
	거부로 인한 무응답	54	5.68
	부분 합	79	8.30
총 합		951	100

주: 1) 조사에 대해 거부 없이 최초 대면시에 조사에 응답한 자.

2) 최소 한 번 이상 응답거부를 하였지만 최종적으로 조사에 응답한 자.

결국 전체 무응답률은 낮지만, 무응답 층을 유형별로 보면 거부로 인한 무응답이 부재에 의한 무응답률보다 상대적으로 높기 때문에, 이에 대해 조사 감독자(supervisor)는 최고의 조사 완성도를 갖는 유능한 조사원이 반드시 그 가구를 재조사할 수 있도록 하는 효율적인 대처를 해야 할 것이다.

2) 방문횟수, 부재횟수, 설득기간

<표 5-2>는 조사원들이 가구 설득에 투입하는 노력을 양적으로 파악하기 위한 것으로, 각 가구의 방문횟수와 부재횟수, 그리고 최종 응답여부가 결정되기까지 걸리는 기간을 나타낸다. 전체적으로 부재횟수를 포함하여 조사원들의 가구당 평균 방문횟수는 3회(3.09) 정도이고, 그 중 부재횟수는 2회(1.63) 정도인 것으로 나타났다. 그런데 무응답자의 경우

는 약 5회(부재자=4.68회, 거부자=4.92회)로 평균 3회보다 2회 정도 많은 것을 알 수 있다. 부재횟수의 경우도 무응답자에서 응답자보다 많고, 특히 최종 부재자의 경우는 조사원들이 평균 5번(4.68회)을 방문하고도 대면을 하지 못하는 것을 알 수 있다. 또한 설득기간은 약 8일 정도이고, 응답자(최초응답=약 6일, 응답전환=약 10일)보다는 무응답자(부재자와 거부자=약 15~16일)에서 최종 응답 협조를 받기까지 더 많은 기간이 소요되는 것으로 나타났다. 여기서 유의할 점은 응답전환자, 부재자, 거부자에서 나타나는 부재횟수의 수치적 의미는 각각 다를 수 있다는 것이다. 최초 응답자와 부재자에서 부재횟수는 조사원이 가구를 방문했을 때의 순수한 부재를 의미하고, 응답전환자와 거부자에서는 거부횟수와 부재횟수를 동시에 포함한다. 설득기간은 무응답자일수록 많이 소요되는데(부재자=15일, 거부자=16일), 이는 지방청마다 또는 조사원마다 조사 착수 일정이 다르다는 데서 비롯되는 부분도 있다.

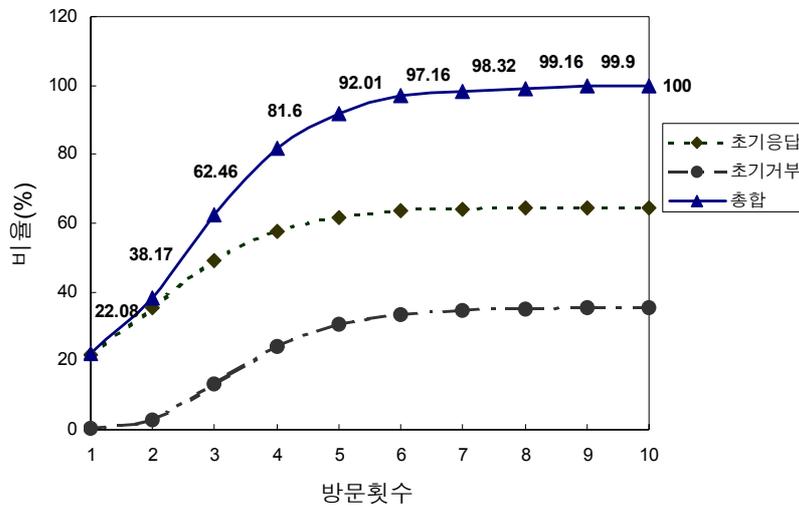
<표 5-2> 평균 방문횟수, 부재횟수, 설득기간

	전체	응답자		무응답자	
		최초응답자	응답전환자	부재자	거부자
방문횟수	3.09±1.68	2.44±1.47	3.96±1.37	4.68±1.41	4.92±1.54
부재횟수	1.63±1.52	1.36±1.42	1.73±1.24	4.56±1.47	2.83±1.78
설득기간(일)	8.23±12.07	6.36±11.18	10.11±12.02	14.60±12.26	15.68±15.8
총합 (명)	951	588	284	25	54

방문횟수는 조사원의 노력과 방문실태를 파악하기 위한 핵심내용으로 방문횟수별 응답률을 비교해 보기로 하자. [그림 5-1]은 방문횟수에 따른 응답자와 거부자의 누적비율을 나타낸다. [그림 5-1]에서 보는 바와 같이 4번째 방문까지는 방문가구의 비율이 급격하게 증가하는 반면에 5번째부터는 그 증가폭이 완만한 것을 알 수 있다. 즉, 4번째 방문으로 조사를 완성하는 가구의 비율은 약 82% 정도이다. 또한 단 한 번의 방문으로 조사가 완성되는 경우는 전체 951개 가구 중 약 22%이고, 전

체 약 62.5%가 평균 3번 이하의 방문으로 응답여부가 결정되는 것을 알 수 있다. 최초 대면에서의 응답자와 거부자의 누적분포를 보면 3번 이하에서는 최초 응답자의 비율이 빠르게 증가하는 반면, 3번과 5번 사이에서는 거부자의 곡선이 빠르게 증가하면서 거부하는 가구일수록 방문이 많이 이루어짐을 알 수 있다. 즉, 최초응답자에서 방문횟수는 부재횟수와 관련이 있고, 이는 결국 방문횟수가 증가하면 그만큼 부재자수 비율도 증가한다는 것을 예측할 수 있다. 그러나 본 연구 결과만으로는 연구의 한계 등을 고려하면 이 결과에 대한 정확한 결론을 내리기 어렵다. 따라서 보다 세부적인 조사계획에 따라 방문에 대한 매회 방문기록을 작성함으로써 재방문과 부재·거부간의 연관성을 분석할 필요가 있다. 이는 재방문 노력이 응답률 향상에 어떤 영향을 미치는지와 이로부터 파생되는 부재율과 거부율의 변화 패턴을 파악하는 데 도움이 될 것이다.

[그림 5-1] 방문횟수에 따른 누적 응답률



3) 시간대별 조사원의 가구 방문 현황

조사원들이 가구 방문시간대를 직접 결정하도록 하는 현재의 시스템은 현장 조사에 대한 부담을 가중시킬 뿐만 아니라 조사업무 전체적인

측면에서도 효율적이지 못하다. 왜냐하면 조사원 입장에서는 근무시간(18시) 이후의 방문에 대한 인센티브를 받을 수 없고, 조사관리자 측면에서는 조사 전반을 모니터링할 수 없기 때문이다. 더구나 현대 사회는 맞벌이, 노인·단독가구와 평일 취미생활을 즐기는 가구들이 증가하면서 일과시간(18시 이전)에 가구의 일원을 접촉하기가 쉽지 않다. 그러다 보니 조사원들은 목표 응답률을 달성하기 위해 근무시간 이후에 조사업무를 수행할 수밖에 없다.

<표 5-3>은 조사원들의 근무시간내·외 방문현황을 나타낸다. <표 5-3>에 의하면 조사원들은 표본가구의 약 87%(831가구)에 대해 근무시간 내에 최초 방문을 시도하지만, 마지막 방문에서는 근무시간 내의 방문비율이 63.1%로 감소하였다. 최초 방문에서 근무시간 내에 접촉을 시도했던 831가구 중 약 30.9%(257가구)가 마지막 방문에서는 근무시간 이후에 재접촉을 시도하는 것으로 나타났다.

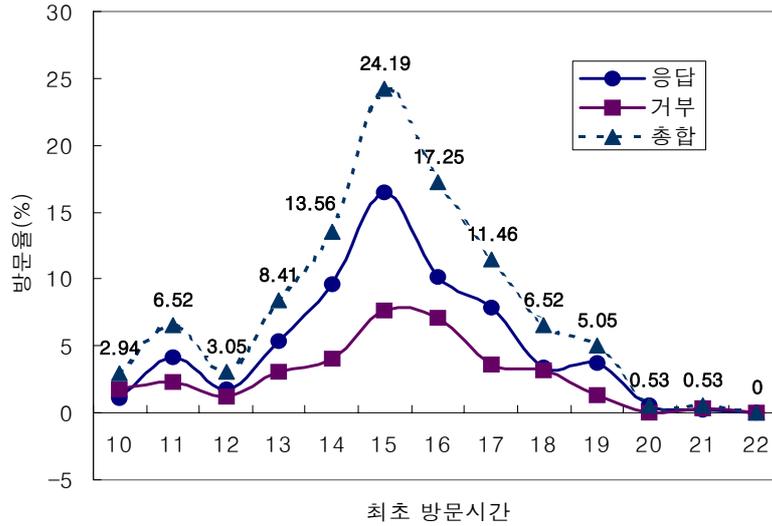
<표 5-3> 최초 방문시간과 마지막 방문시간대별 분포

		마지막 방문시간		총합
		18시 이전	18시 이후	
최초 방문시간	18시 이전	574(60.36)	257(27.02)	831(87.38)
	18시 이후	26(2.73)	94(9.88)	120(12.62)
총합		600(63.09)	351(36.91)	951(100)

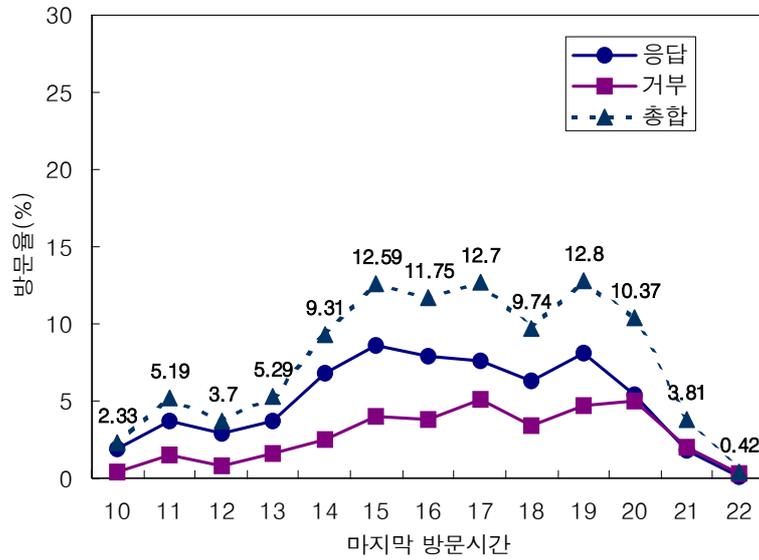
한편 응답자 유형별로 최초와 마지막 방문 시간대 변경이 어떻게 이루어지는지 살펴보자. [그림 5-2]와 [그림 5-3]은 응답 거부 여부에 따른 방문시간대별 응답자 분포를 나타낸다. [그림 5-2]에 의하면 최초 방문의 경우, 응답자와 응답 거부자 모두 14시에서 17시 사이에 가장 많은 접촉을 시도하는 것으로 나타났다.

그러나 [그림 5-3]의 마지막 방문시간대별 응답 분포를 보면, 전체적으로 재방문 시간이 18시 이후로 이동되는 양상을 보이며 응답자와 거

[그림 5-2] 최초 방문시간대별 방문비율



[그림 5-3] 마지막 방문시간대별 방문비율



부자 구별 없이 재방문 시간이 유사한 경향을 보인다. 이는 조사원들이 접촉되지 않은 가구의 무응답 유형별 성향을 고려하지 않고 일률적으로 근무시간 이후 재방문 시간을 결정한다고 볼 수 있다. 그러나 실제로 최초 응답자와 거부자의 재방문 이유를 보면, 최초 응답자는 부재로 인해 가구의 일원을 접촉하지 못해서 발생한 것이고, 거부자는 가구의 일원을 접촉하였으나 응답 거부로 인해 설득을 목적으로 재방문 하게 되는 경우가 대부분이다. 이러한 이유에서 보면 조사감독자는 조사원들에게 잠재적 응답자들과 더 많은 접촉을 위해서 주중의 다른 시각 또는 다른 날짜에 재방문 하도록 요구해야 한다. 그러나 현재 우리의 조사 시스템은 조사관리자가 없기 때문에 이러한 업무를 제어할 수 없을 뿐만 아니라, 현장 조사에 대한 철저한 관리를 하기가 어려운 실정이라 할 수 있다.

4) 요일별 조사원의 가구 방문 현황

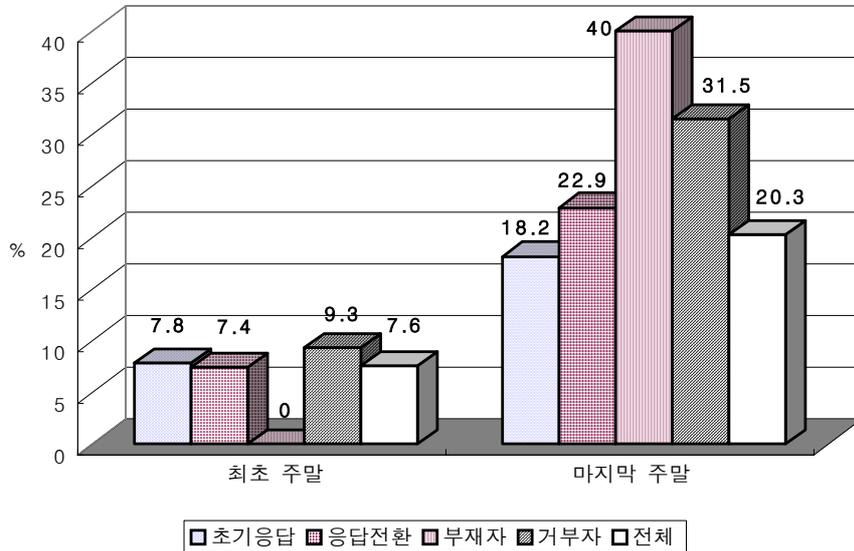
<표 5-4>는 조사원들의 노력도를 측정하기 위한 것으로 요일별 방문 현황을 나타낸다. 이는 최초 방문 이후에 조사원의 방문분포가 주말로 이동하는 현상을 파악하고자 하는 것이다. 따라서 세부적 요일별 방문 분포보다는 주중과 주말 방문현황으로 이분하여 그 현상을 파악하였다. 여기서 주말이란 '금요일 17시 이후부터 일요일'까지로 정의하였다. 주중방문의 경우, 최초 방문에서는 전체 대상자의 92.4%로 대부분 조사원들은 근무시간 이내에 최초 방문을 시도하는 것으로 나타났다. 그런데 이후 마지막 방문에서의 주중 방문 비율은 79.1%로 최초 방문에 비해 약 13.3%포인트가 주말방문으로 변경되었음을 알 수 있다. 한편, 전체 응답전환자 중 약 23%가 주말에 응답이 전환되었다. 응답유형별로 주말·주중 이동현황을 보면, 부재자의 경우는 부재가구 25가구 중 모두가 최초 방문은 주중에 접촉을 시도한 반면에 마지막 방문에는 25가구 중 40%(10가구)가 주말에 재방문이 시도되었지만, 여전히 모두 부재가구인 것으로 확인되었다. 최종적으로 응답을 거부한 경우, 조사자들은 최초에는 54가구 중 90.7%(49가구)를 주중에 방문을 시도했다가 마지막에는 54가구 중 31.5%(17가구)를 주말에 재방문을 시도한 것으로 나타났다.

〈표 5-4〉 최종 응답자의 주말·주중 방문 현황

		최초 방문 시기		마지막 방문 시기	
		주중	주말	주중	주말
응답자	최초응답자	542(56.99)	46(4.84)	481(50.58)	107(11.25)
	응답전환자	263(27.66)	21(2.21)	219(23.03)	65(6.83)
무응답자	부재자	25(2.63)	0(0.00)	15(1.58)	10(1.05)
	거부자	49(5.15)	5(0.53)	37(3.89)	17(1.79)
전체		879(92.43)	72(7.57)	752(79.07)	199(20.93)

최초방문 이후 조사원들의 주말 재방문이 증가한 것은 사실이다. [그림 5-4]는 최초 주말 방문과 마지막 주말 방문에 대한 분포를 나타낸다.

[그림 5-4] 최초와 마지막 주말방문 분포



[그림 5-4]에 따르면 최초응답자와 부재자는 부재가 재방문의 근본적인 이유이지만 주말 재방문 비율은 각각 18.2%와 40%로 재방문에 대한

변경정도는 다소 차이가 있다. 이는 최초응답자는 조사에 상당히 우호적인 그룹으로 굳이 주말 방문이 아니더라도 주중 근무시간 이후를 이용하여 성공적인 접촉의 가능성을 나타낸다고 볼 수 있다. 그러나 최종 부재자의 경우는 주말 재방문의 증가로 조사원의 부담은 상당히 증가하였지만, 재방문으로 인한 응답여부는 성공적이지 못한 것으로 보인다. 한편 거부로 인한 재방문인 경우 응답전환자 중 22.9%만이 주말 방문에서 응답전환에 성공하였고, 최종 거부자는 전체 거부자 중 31.5%가 주말에 재방문이 이루어졌지만 결국 응답전환에 실패한 것으로 나타났다. 여기서 한 가지 알 수 있는 것은 최초응답자와 응답전환자는 상대적으로 조사에 협조적인 응답자 층으로 이들 응답자 층의 특성을 잘 파악하여 주말 방문을 적극 활용한다면 응답률 향상에 크게 도움이 될 수 있다는 것이다. 그러나 거부자의 경우는 대체로 조사에 비협조적인 층으로 주중에 조사에 협조하지 않은 가구가 주말이나 근무시간 이후 재방문을 시도한다고 해도 다른 유형에 비해 상대적으로 성공률이 낮다는 것을 알 수 있다. 따라서 조사감독자는 조사기간 동안 응답유형 또는 무응답에 대한 이유를 모니터링함으로써 조사원들로 하여금 적절한 요일과 시간을 정하여 재방문을 하도록 권장해야 한다.

나. 가구의 반응

1) 무응답 유형별 거부 이유

<표 5-5>는 표본가구의 응답거부 이유에 대한 응답분포를 나타낸다. 가구를 대상으로 하는 조사에서 자주 거론되고 있는 응답 거부 이유 10가지를 선정하여 살펴보았다. 전체적으로 거부이유별 분포를 보면 ‘사생활 침해’가 25.9%로 가장 높고, 그 다음으로 ‘바쁘다’가 26.8%, ‘정부 통계 불신’이 16.7%, ‘무대응’ 등의 순으로 나타났다. 응답전환자의 경우도 순위별 거부이유는 전체 순위와 유사한 경향을 보였지만, 거부자는 ‘사생활 침해’, ‘무대응’과 ‘문전박대’ 순으로 전체경향과 다르게 나타났다. 즉, 최초로 거부한 그룹들의 거부이유에 따라 최종적으로 응답으로 전환되는 정도에는 차이가 있음을 알 수 있다.

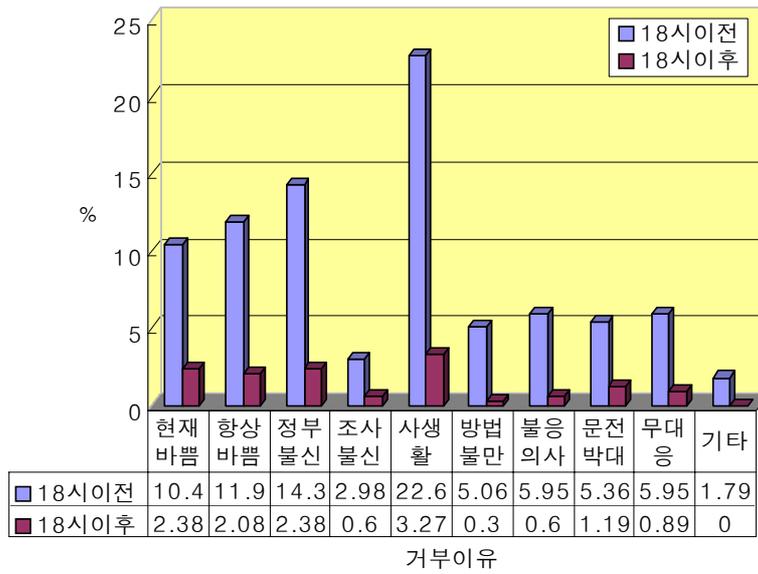
〈표 5-5〉 표본 가구의 응답거부 이유

	전체		응답전환자		거부자	
	N	%	N	%	N	%
1. 현재 바쁘다	43	12.80	40	14.18	3	5.56
2. 평상시 워낙 바쁘다	47	13.99	42	14.89	5	9.26
3. 정부통계 불신	56	16.67	48	17.02	8	14.81
4. 통계조사 불만	12	3.57	12	4.28	0	0.00
5. 사생활 침해	87	25.89	74	26.24	13	24.07
6. 조사방법에 대한 불만	18	5.36	16	5.67	2	3.70
7. 불응의사를 명확히 밝힘	22	6.55	17	6.03	5	9.26
8. 문전박대	22	6.55	15	5.32	7	12.96
9. 무대응(인기척 있음)	23	6.85	15	5.32	8	14.81
10. 기타	6	1.79	3	1.06	3	5.56
총 합	336	100	282	100	54	100

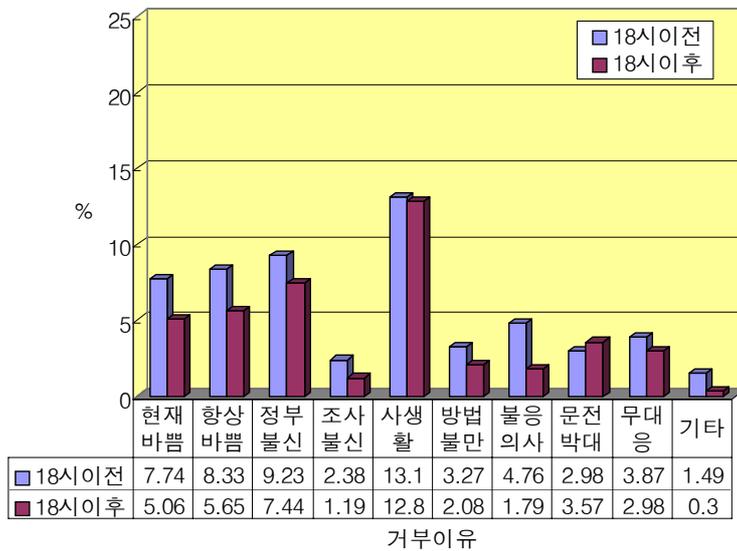
* 기타 이유: 간병차 장기출타, 합당한 금전보상요구(가계부와 연계), 명부작성만.

[그림 5-5]와 [그림 5-6]은 최초 응답 거부자들에 대해서 거부 이유별로 조사원들의 방문시간대별 이동분포를 나타낸 것이다. 거부이유별로 최초 방문에 비해 마지막 방문시에는 근무시간 이후 방문이 증가하였다. 눈에 띄는 현상은 ‘사생활 침해’라는 이유로 최초에 거부했던 가구는 전체의 25.9%였고, 이 중 3.3%만이 근무시간 이후에 이루어졌다. 그러나 마지막 방문에서는 근무시간 이후 방문이 약 12.8%로 최초 방문에 비해 9.5%포인트 증가하였다. 이는 <표 5-5>에서 언급한 바와 같이 ‘사생활 침해’로 거부한 87가구 중 74가구(85.1%)가 응답전환된 것과 무관하지 않다. 한편 ‘불응의사를 정확히 밝힘’이나 ‘문전박대’, ‘무대응’을 이유로 응답을 거부한 가구의 경우, 마지막 방문에서는 그 비율이 근무시간 이후로 많이 변경되었지만 <표 5-5>에 의하면 ‘바쁨’이나 ‘정부통계와 통계불신’, ‘사생활 침해’에 비해 상대적으로 응답전환이 낮은 경향을 보였다. Fisher의 정확검정(exact test)에 의하면 응답거부 이유에 따라 응답전환 여부는 유의수준 0.05하에서 통계적으로 유의한 차이($p=0.0078$)를 보였지만 방문시간과 응답전환 여부와는 유의한 차이를 보이지 않았다($p=0.1151$).

[그림 5-5] 최초 방문시간과 거부이유



[그림 5-6] 마지막 방문시간과 거부이유



이로부터 알 수 있는 사실은 ‘사생활 침해’ 또는 ‘바쁘다’는 이유로 응답을 거부하는 층은 처음 대면시에는 통계조사에 귀찮다는 반응을 보이지만, 결국은 조사에 대한 이해와 조사원 설득을 통해 조사에 협조하는 경향이 있다는 것이다. 그러나 ‘문전박대’와 ‘무대응’과 같은 반응을 보인 그룹은 통계조사에 대해 강하게 거부하는 층으로 아무리 시기를 달리하여 재방문과 같은 설득을 시도한다 하더라도 조사에 쉽게 협조하지 않는다는 것이다. 따라서 거부 이유에 따른 과학적이고 체계적인 설득 방안을 마련함으로써 원활한 조사를 진행할 수 있어야 한다. 최종 거부 층을 완전히 무시하는 것은 정해진 기간 내에 조사를 완성해야 되는 입장에서는 특정 층의 분포를 희생할 수밖에 없고, 이는 추정치에 대해 심각한 오류를 범할 소지가 높게 된다. 이런 이유에서 만약 무응답이 특정 그룹에서 높게 나타날 것으로 예측된다면 이들 층을 보다 세밀하게 모니터링할 필요가 있다(Statistics Canada, 2003).

2) 거부에 대한 반응 정도

<표 5-6>은 응답을 거부한 가구들의 거부할 때 나타내는 반응 분포이다. 표본가구에서 응답자가 조사에 협조하기를 거부할 때 보이는 반응을 ‘화를 내는 정도’에 따라 5점 척도로 측정하였다(매우 화냄=5점, 전혀 화를 내지 않음=1점). 우선 화를 내는 정도와 응답전환 여부(전환 vs. 거부)와의 관계를 Fisher의 정확성 검정(Fisher's exact test)에 비교한 결과 유의수준 0.05에서 통계적으로 유의한 차이를 보였다($p=0.0474$). 전체 거부자 중 ‘매우 화냄’ 또는 ‘약간 화냄’을 포함하여 어느 정도 화를 내는 거부자가 72% 정도로 대부분이 거부할 때 화를 내는 것으로 나타났다. 이를 응답전환자와 최종 거부자별로 살펴보면 응답전환자의 경우는 화를 내는 정도에 대한 분포가 전체 거부자의 반응정도와 비슷한 경향을 보였다. 구체적으로 살펴보면 ‘매우 화냄’ 그룹 중 약 28%는 최종적으로 응답 전환되었지만 나머지는 결국 거부한 것으로 나타났다.

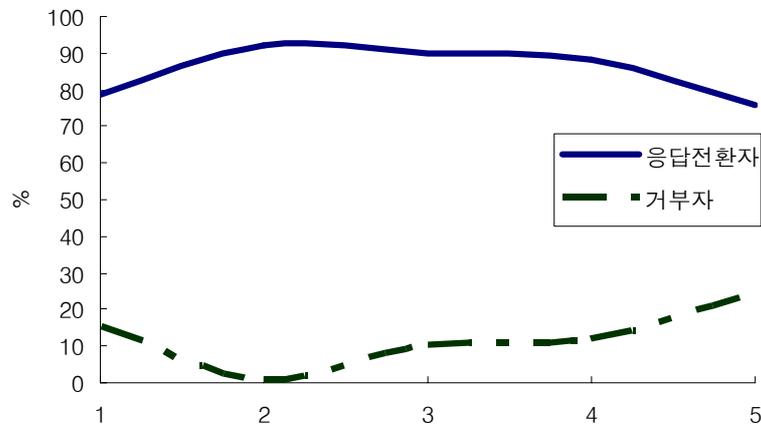
[그림 5-7]은 화를 내는 정도별로 응답전환 가능성을 나타낸다. 거부할 때 ‘별로 화를 내지 않음’ 그룹일수록 응답전환이 잘 되고, ‘매우 화를 냄’ 그룹일수록 전환되기 어려운 것으로 나타났다.

<표 5-6> 최초 대면시 거부자들의 반응

	전체 N (%)	응답전환자 N (%)	거부자 N (%)
매우 화냄	99(31.33)	75(27.99)	24(50.00)
약간 화냄	129(40.82)	114(42.54)	15(31.25)
보통	50(15.82)	45(16.79)	5(10.42)
별로 화를 내지 않음	25(7.91)	23(8.58)	2(4.17)
전혀 화를 내지 않음	13(4.11)	11(4.10)	2(4.17)
총합	316(100)	268(100)	48(100)

* Fisher's exact test, $\alpha = 0.05$; $p = 0.0474$.

[그림 5-7] 응답자의 화냄 정도에 따른 응답전환



* 5점 척도 - 1: 전혀 화를 내지 않음, 5: 매우 화를 냄.

이는 David & Melissa(2006)의 유럽 국가들의 가구를 대상으로 한 CPS(Current Population Survey) 조사에서 나타나는 가구의 거부 반응 정도와는 다르다. David & Melissa(2006)의 연구에서는 ‘매우 화를 냄’이나 ‘화를 내지 않음’ 그룹일수록 응답 전환이 쉽게 되는 반면에 거부할 때의 반응이 ‘그저 그렇다’와 같이 조사에 대해 무관심한 반응을 보일수록

오히려 응답 전환이 잘 되지 않는 것으로 제시하였다. 이처럼 유럽 국가의 조사에 대한 가구들의 반응과 우리의 경우가 다른 것은 각 국가간의 국민적 성향을 반영한 것으로 해석된다. 따라서 우리나라 응답자 성향이 외국과 다르면 통계조사에 대한 해외사례를 접목할 때도 이런 국민적 성향이 충분히 반영되어 한국형 조사방법 및 대처방안을 개발할 필요가 있다.

3) 응답전환 요인 분석

지금까지 현장 조사원들이 작성한 방문기록표를 근거로 조사원들의 방문실태와 조사원 노력, 가구의 반응 정도를 중심으로 살펴보았다. 이로부터 우리가 알 수 있었던 것은 조사원들의 노력 여하에 따라 응답률이 어느 정도까지는 향상될 수 있다는 것이다. 그렇다면 이 응답 전환에 영향을 주는 요인을 찾아보는 문제가 남아 있다. 지금까지는 간단한 이변량 분석(binary analysis)을 통해 응답거부이유, 방문시간, 방문요일, 설득기간 등의 영향력 제고를 위한 후보 변수들을 고려하였다. 그 결과 이 변수들은 응답전환 여부에 충분히 영향력 있는 변수로 판단되었고, 이들의 영향력 정도를 찾기 위해 구체적인 회귀모형에 의한 분석을 수행하였다.

종속변수는 최초 응답거부 이후에 응답 전환이 있었는지 여부를 나타내는 이진변수(binary variable)이다. 설명변수는 범주형 변수(응답거부이유, 방문시간, 요일)와 연속형 변수(설득기간)로 구성되었다. 본 연구는 적은 표본수를 고려한 효과적인 회귀모형 설정을 위해 원변수들을 몇 개의 범주형 변수들로 재구분하였고, 표본 수도 약간 축소(parsimonious)하였다. 응답 거부 이유는 기타를 포함하여 10개의 범주를 포함하는 변수로, 경험적 정보와 이변량 분석 결과를 이용하여 5개 그룹으로 재구성하였다 ; ‘바쁘다(현재 바쁘다, 항상 바쁘다)’, ‘통계불신(정부통계불신, 통계조사불만, 조사방법 불만)’, ‘사생활 침해’, ‘무대응(절대 안함, 문전박대, 무대응)’, ‘기타’로 구분하였다. 여기서 ‘기타’는 그 이유가 다양하고 명확하지 않을 뿐만 아니라 이에 해당되는 표본 수도 매우 적기 때문에(6가구) 분석에서 제외하였다. 그리고 나머지 4개 범주에 대해 3개의 가변수(dummy variable)를 생성하고, 기준 범주는 마지막 범주인

‘무대응’으로 설정하였다. 방문시간은 근무시간 전·후로 구분하여 18시 이전과 18시 이후 두 그룹으로, 방문요일은 주중·주말로 각각 두 범주를 갖는 변수로 재구성하였다. 마지막으로 방문기간은 구간변수로서 자료탐색 결과에 의해 곡선의 경향과 분산이 커지는 현상을 보였고 이를 극복하기 위해 곡선의 경향을 고려한 분산 안정화를 위해 제곱근(squared root) 변환을 시도하였다.

응답전환 요인	회귀계수	p-value	오즈비 추정치	
			오즈비	95% 신뢰구간
Intercept	1.609	0.0153	-	-
바쁘다: 바쁘다, 현재 바쁘다	1.336	0.0038	3.804	(1.540 9.400)
통계불신: 정부통계불신, 통계조사 불만, 조사방법 불만	0.984	0.0247	2.676	(1.133 6.318)
사생활 침해	0.775	0.0598	2.171	(0.968 4.866)
방문시간(18시 이전)	-0.0788	0.8128	0.924	(0.482 1.774)
방문요일(주중)	0.2253	0.2133	1.569	(0.772 3.192)
방문기간	-0.2049	0.0185	0.815	(0.687 0.966)
AIC = 279.14				
model $\chi^2 = 19.66$ (p=0.0046)				

최종 모형 선정은 AIC(Akaike's Information Criteria)통계량과 χ^2 통계량 기준에 의해 선정되었다. 모형에 대한 통계적 유의성 검증에는 유의수준 0.05를 사용하였다. 선택된 최종 모형에 의한 분석결과 ‘바쁘다’, ‘통계불신’은 ‘무대응’에 비해 응답 전환될 가능성이 높고 유의확률값은 각각 0.0038과 0.0247로 통계적으로 유의한 차이를 보였다. ‘사생활 침해’는 ‘무대응’에 비해서는 응답 전환이 되기 쉬웠지만 통계적으로 유의하지는 않았다(p=0.0598). 방문기간은 회귀계수값이 -0.2049로 방문기간이 늘어날수록 응답 전환이 어려워짐을 알 수 있고 통계적으로도 유의한 결과를 보였다(p=0.0185).

각 변수들의 응답 전환에 미치는 영향력 정도는 오즈비(odds ratio)를 이용하여 좀더 쉽게 이해할 수 있다. 오즈비에 의하면 ‘바쁘다’는 ‘무대

응’에 비해 약 4배 정도(오즈비=3.804) 응답 전환이 되기 쉽고, ‘통계불신’은 약 2.6배(오즈비= 2.676), ‘사생활 침해’는 약 2배 정도가 무대응에 비해 응답 전환이 쉬운 것으로 나타났다. 방문기간에 대한 오즈비는 1보다 작은 값(오즈비=0.5150)으로 방문기간이 1일씩 증가할 때 응답 전환 가능성은 0.8배 정도 감소한다는 것을 의미한다. 또한 방문시간은 통계적으로 유의하지는 않았지만, 18시 이전에 비해 18시 이후가 응답 전환 가능성이 약간 높고, 주말보다는 주중이 1.5배 정도 높은 것으로 나타났다. 이처럼 주말이 오히려 응답 전환이 낮은 것은 <표 5-4>에서 설명한 바와 같이 최종 부재자나 거부자는 확고한 무응답 층으로서 시간의 변동이 있다 하더라도 조사에 매우 비협조적이라는 것을 짐작할 수 있다.

4) 응답자와 응답전환자 비교

<표 5-7>은 2007년 경찰조사 표본에 대한 실업률을 최초응답가구와 응답 전환된 가구에 대해 계산한 것이다. 이 결과는 응답전환 효과를 증명하기 위한 하나의 방법으로 경찰조사의 목표 변수에 대한 추정치를 비교한 것이다. 응답 전환된 가구들의 추정치는 4.75%로 전체 응답가구들의 실업률 3.60%보다 상당히 높다. 또한 응답전환 가구들은 최초 응답가구들의 실업률 3.21%와는 큰 차이를 보이고 있다. 이 결과는 조사 환경 및 가구들의 응답 특성을 증명해 주는 결과로 실업자가 많은 가구일수록 응답을 거부할 가능성이 높다는 것을 알 수 있다. 따라서 조사 관리자는 조사품질과 응답률 향상을 고려하여 거부 가구들의 재방문에 대해 재접촉 방법 및 시기와 조사원 배치 등 제반 사항에 대한 신중한 검토를 할 필요가 있다.

<표 5-7> 최초 응답자와 응답전환자 실업률 비교

	최초응답자	응답전환자	전체 응답자
실업률 추정 (%)	3.21	4.75	3.60
취업자 수 (명)	755	341	1096
실업자 수 (명)	25	17	42
가구 수 (가구)	576	266	842

참고로 <표 5-7>의 분석결과는 10월 경황조사 표본조사에 의한 신표본 결과를 그대로 분석한 것으로, 설계 가중치 등을 전혀 고려하지 않은 실업률임을 밝혀둔다.

제3절 무응답 오차 분석

1. 무응답 오차에 대한 인식

지금까지 자료를 수집하는 데 들어간 조사원들의 노력과 가구의 반응에 대한 일반적인 현황과 관련한 탐색적 분석을 수행하였다. 조사원들의 노력은 최종적으로 무응답률을 줄이는 것과 관련된다. 그렇다면 무응답률을 줄이는 것이 자료의 품질향상과 관련이 있는가 하는 일반적인 의문을 가질 수 있다. 무응답률은 어떤 이유에서든(거부, 부재, 기타) 최종적으로 조사에 응답하지 않은 표본들의 비율을 말한다. 자료품질은 조사가 성공적으로 이루어진 표본에 기초한 추정치와 전체 표본에서의 추정치의 차이에 의해서 측정되었다. 참값에 가까운 추정치일수록 자료품질이 높다는 것을 의미한다. 자료품질에 대한 정의는 상대적으로 엄격하고, 보통 조사(survey)연구에서 정확한 정의를 내리기는 어렵다. 왜냐하면 참값은 일반적으로 알려져 있지 않고 만약 이 참값이 알려졌다면 조사 연구는 할 필요가 없기 때문이다. 그러나 경우에 따라 제한된 변수에 대해서 표본 내의 참값은 구할 수 있다. 예를 들면 투표행위에 대한 조사결과를 실제 선거결과에 연관시킬 수 있을 것이다(Voogt & Saris, 2005).

무응답이 실제로 조사결과에 어떤 영향을 줄 것인가를 자신 있게 설명하기는 어렵다. 확실한 것은 무응답은 조사단계에서의 노력으로 줄일 수 있다는 것, 그리고 무응답 오차가 발생하는 원인이 반드시 존재하는 것이다. 이에 대해 캐나다 통계청에서는 target 모집단의 핵심 부차 그룹에 대한 응답률을 계산하고 비교하라고 권장하고 있다(Canada statistics, 2007). 이들 부차 그룹은 연령, 성별, 지역 등이 될 수 있다. 이 방법은 무응답 편향 정도를 결정하는 데 직접적인 도움이 되는 것은 아니지만, 무응답 편향에 대한 존재 여부는 결정할 수 있다. 만약 무응답률이 부차

그룹에 걸쳐 매우 유사하다면, 무응답 편향이 존재하더라도 이 무응답은 자료에는 심각한 영향을 주지 않을 가능성이 높다는 것이다. 그러나 이러한 접근 방법이 갖는 문제점은 그런 부차 그룹 변수와는 별도로 무응답 편향에 대한 원인이 더 존재한다는 것이다. 즉, 응답 경향은 단지 부차 그룹 변수에 의해서만 영향을 받을 것 같지 않다는 것이다(Groves, 2006). 응답률을 비교하는 것은 무응답 편향 연구의 좋은 출발점이지만, 그렇다고 이것이 무응답을 설명하는 다른 방법을 대신해야 한다는 것은 아니다.

이러한 이유에서 무응답률과 함께 무응답 편향은 자료품질에 매우 중요한 역할을 한다고 볼 수 있다. 또한 무응답 편향은 무응답 원인에 따라 그 크기가 다르게 나타날 것이라는 것이 최근 연구들에서 보고되고 있다(Groves, 2006; Heerwgh et al., 2007). 무응답 원인은 가구의 어떤 특성에 기인된 것으로 이들 원인에 따른 무응답 편향을 측정하는 것은 자료 품질과 관련하여 매우 중요하다 하겠다. 본절에서는 무응답 원인별로 무응답 오차를 계산하고, 무응답률과 오차와의 관계를 살펴보고자 한다. 이러한 오차계산에 있어서 key 변수 선정은 어려운 문제이지만, 핵심 추정치와 관련된 변수를 선정하는 것이 일반적이다(Heerwgh et al., 2007).

2. key 변수 선정

표본의 부차 그룹들 간의 응답률 차이를 연구하는 데 있어서 응답률과 목표변수와 높은 상관성이 있을 것으로 판단되는 보조변수를 찾는 것은 중요하다. 부차 그룹간의 응답률이 통계적으로 보조변수 변수값들과 연관되어 있다는 것은 목표변수를 통해 구한 추정치가 무응답으로 인해 편향된다는 것을 의미할 수 있다(조사통계연구회, 2000).

본 연구는 매월 새롭게 표본대상이 되는 가구를 설득하는 과정에서 발생하는 일련의 상황들을 조사원들의 협조를 통해 조사하였다. 또한 조사원들의 부담을 덜어주기 위해 조사원들의 조사결과와 표본명부 자료를 연계하여 자료 분석에 활용하였다. 즉, 조사원들의 방문기록과 표본명부 자료의 코드를 연결시켜 연구 분석을 위한 새로운 자료를 구성하였다. 무응답 오차 분석은 실업자 수와 가장 연관성이 있는 핵심 변수가

필요하고, 이들 변수 선정에 있어 이용 가능한 특성들로 주택유형과 주택점유형태를 고려하였다. 이 변수들은 이미 모집단 자료(통계청, 2005)에 대한 연관성 분석을 통해 실업자수와 연관성이 있는 것으로서 현재 통계청에서 경활조사 표본 설계시에 분류지표로 사용되고 있는 변수들이다. 뿐만 아니라 주택유형은 가구조사의 응답 유형별 특성을 잘 반영할 수 있는 변수이기도 하다. 무응답 편향은 이들 변수를 사용하여 평가될 것이다. 즉, 주택유형에서는 주택, 아파트, 기타(연립, 다세대 등)가 차지하는 가구의 비율과 주택소유여부에서는 자가 vs 전/월세 등이 차지하는 가구의 비율을 비교하고자 한다.

특히, 주택유형은 부재율과 거부율에 영향을 미칠 수 있기 때문에 보다 넓은 의미에서 본다면 우리가 관심 있는 관심 모수가 될 수 있다(Groves & Couper, 1998). 따라서 이 변수의 추정치에 대한 오차를 연구하는 것은 흥미로운 것이고, 오차는 자료 품질에 대한 전체적인 설명이 가능할 뿐만 아니라 자료를 수집하는 데 들어가는 노력이 자료 품질과 어느 정도 연관성이 있는지에 대한 비전을 제시할 수 있을 것이다.

3. 무응답 오차 개념

표본설계에 의해 추출된 각 표본에 대해서 전체 표본평균(full sample mean) \bar{y}_n 와 응답자 평균 \bar{y}_r 와의 차이는 다음과 같다(Groves & Couper, 1998).

$$\bar{y}_n = \left(\frac{r}{n}\right)\bar{y}_r + \left(\frac{m}{n}\right)\bar{y}_m.$$

이것을 응답자 평균에 대해서 다시 정리하면,

$$\bar{y}_r = \bar{y}_n + \left(\frac{m}{n}\right) [\bar{y}_r - \bar{y}_m]. \quad (1)$$

이때, m 은 전체 표본 중 무응답자 수, r 은 응답자 수, $n(r+m)$ 은 전체 표본수를 나타낸다.

$$\begin{aligned} \text{응답자 평균} &= \text{전체 표본평균} + (\text{무응답률}) \\ &\quad \times (\text{응답자와 무응답자 간의 평균 차이}) \end{aligned}$$

결국, 전체 표본평균에 대한 응답자 평균의 편차는 무응답률, $\left(\frac{m}{n}\right)$ 과 (응답자와 무응답자 간의 평균 차)의 함수를 나타낸다. 이러한 단순 표현으로 볼 때, 동일한 표본 설계에 의해 추출된 전체 표본에 대해서 응답자 평균에 대한 기대값은 무엇을 의미하는가? 이 질문에 대한 대답이 곧 응답자 평균에서 편향의 특성을 결정한다. 여기서 편향은 통계량(statistic)의 기대값(estimate)과 모수(parameter)의 차이, 즉

$$bias = E(\text{표본 평균}) - \text{모집단 평균}$$

을 의미한다. 만약 전체 모집단으로부터 단순 임의추출에 의해 표본을 추출한다면 이는 표본 내의 응답자에 대해서만 관측 가능하게 된다. 이 때 얻어지는 응답자 평균 \bar{y}_r 과 이 응답자 평균을 사용하여 전체 모집단 평균을 추정한다면 다음과 같은 식의 결과를 얻게 된다. 이는 응답자 평균에 대한 편향으로 편향은 다음과 같은 근사식을 갖는다.

$$Bias(\bar{y}_r) \approx \left(\frac{M}{N}\right)(\bar{Y}_r - \bar{Y}_m), \quad var(\bar{y}_r) \approx \frac{1}{n_r} S_r^2$$

여기서, var 는 조사 결과 얻어진 응답자 평균에 대한 분산을 나타내고, S_r^2 은 전체 모집단에서 응답자들의 분산을 나타낸다. $Bias$ 를 다시 풀어 쓰면,

$$Bias(\text{응답자 평균}) = (\text{모집단에서의 무응답률}) \times (\text{모집단 내에서 응답자와 무응답자의 평균 차이})$$

와 같다. 여기서, M 은 모집단에서의 무응답자 수, N 은 모집단 전체 수를 나타낸다.

4. 무응답 오차와 오차의 구성

가. 주택유형(단독주택/아파트/기타)

<표 5-8>은 무응답 오차 계산을 위한 핵심변수인 주택유형별 분포를 나타낸다. 조사대상 가구 중 최종 수집된 951 표본 중에서 표본의 39.85%는 주택에, 44.79%는 아파트에, 그리고 나머지 15.35%는 기타유형(상가

주택, 다세대, 연립주택, 주거지의 거처)에 거주하는 것으로 나타났다. 응답한 표본 가구에 대해서만 주택유형을 살펴보면, 응답 가구 872가구 중 주택에 거주하는 비율은 40.71%(355가구), 아파트에 거주하는 비율은 44.27%(386가구)이고, 나머지 기타유형에 15.23%(131가구)가 거주하는 것으로 나타났다. 응답자의 응답유형별 주택유형에 대한 표본 분포는 전체 데이터의 주택유형과 유사한 분포를 보였다($\chi^2 = 5.1$, $df = 4$, $p = 0.2733$). 이는 표본설계로부터 오는 주택유형에 대한 편의는 거의 없다고 볼 수 있다.

〈표 5-8〉 무응답 원인별 주택유형 분포

응답 유형	주택		아파트		기타		전체	
	n	%	n	%	n	%	n	%
응답	355	93.2	386	90.61	131	89.73	872	91.69
부재	10	2.53	10	2.35	5	3.42	25	2.63
거부	14	3.54	30	7.04	10	6.85	54	5.68
	379	100	426	100	146	100	951	100

무응답은 그 원인에 따라 부재(noncontact), 거부(refusal)와 기타(질병, 이사 등)로 나눌 수 있다. 본 연구는 특정 달(month)의 신표본만을 대상으로 조사한 것이기 때문에 기타유형(빈집, 4가구)에 해당하는 표본이 매우 적었다. 따라서 무응답 유형에서 기타유형은 제외하고 부재와 거부만을 무응답 유형으로 고려하였다.

이제 구체적으로 무응답 오차를 계산해 보자. 만약 주택에 살고 있는 가구의 비율이 목표 추정치라면, 추정치에 대한 무응답 오차는 0.0086(0.86%)으로 응답자와 전체표본 평균과의 차이, $\bar{y}_r - \bar{y}_n = 0.4071 - 0.3985$ 에 의해 계산된다. 즉, 여기서 \bar{y}_r 는 응답자 중 주택 가구에 거주하는 비율, \bar{y}_n 는 전체 대상 표본가구 중 주택에 거주하는 가구의 비율(모수에 해당)을 나타낸다. 즉, Groves & Couper(1998)의 식 (1)을 다시 쓰면,

$$\bar{y}_r = \bar{y}_n + \left(\frac{m}{n}\right)(\bar{y}_r - \bar{y}_n),$$

이고, 이 식의 의미는 $\bar{y}_r = \bar{y}_n + nonresponse\ error$ 으로 결국 응답자 평균은 전체 표본평균과 무응답오차를 합으로 구성된다. 이로부터 무응답자의 평균은 $\bar{y}_m = 0.3036$ 이 된다. 이것은 단독주택에서 무응답자의 비율이 30.36%임을 의미한다. 이는 단독주택에 거주하는 응답자 비율이 40.71%라는 것과는 상당한 차이가 있다. 즉, 응답하지 않은 79(부재+거부=25+54) 표본가구 중 30.36%가 단독주택에 거주하는 것으로, 이는 단독주택거주 비율에 있어서 응답가구들과 무응답 가구들 간에 상당한 차이를 보인다는 것을 의미한다.

또한 이를 확장하여 무응답 그룹이 서로 다른 세부그룹으로 분할될 때 응답과 무응답 그룹 간에 발생하는 추가적인 차이를 볼 수 있다 (Mayer & Pratt Jr., 1966). 무응답 유형별로 오차 특성을 살펴보기 위해 식 (1)을 다음과 같이 확장할 수 있다(Groves & Couper, 1998:12).

$$\bar{y}_r = \bar{y}_n + \left(\frac{m_{ref}}{n}\right)[\bar{y}_r - \bar{y}_{m_{ref}}] + \left(\frac{m_{nc}}{n}\right)[\bar{y}_r - \bar{y}_{m_{nc}}] \quad (2)$$

여기서, m_{ref} , m_{nc} 는 각각 거부와 부재에 의해 발생한 무응답자 수를 나타낸다. 식 (2)에 의한 구체적인 계산은 다음과 같다.

$$0.4071 = 0.3985 + \left(\frac{54}{951}\right)[0.4071 - 0.2593] + \left(\frac{25}{951}\right)[0.4071 - 0.4000].$$

여기서 거부에 의한 오차는 0.0084(97.67%)로 무응답 오차의 대부분을 차지하고, 부재에 의한 오차는 0.0002(2.33%)로 전체 오차의 아주 적은 부분만을 차지한다는 것을 알 수 있다. 만약 추정치 대상이 단독주택 가구가 아니라 아파트나 기타주택유형에 대해서도 유사한 방법으로 계산할 수 있고, 이에 대한 결과도 단독주택에서와 마찬가지로 아파트 비율과 기타 비율을 추정할 경우 거부에 의한 오차가 각각 85.3%와 60.6%로 전체 무응답의 오차의 대부분을 차지한다. <표 5-9>는 주택유형에 따라 무응답 오차 유형별로 오차의 크기를 계산한 결과이다. <표 5-9>에서 오차 크기가 음수로 표현된 것은 응답자에서의 해당가구 유형 비율이 전체 표본이 차지하는 해당가구 비율보다 높기 때문에 나타나는 현상이다. 오차의 크기가 음수인 경우 오차 비율은 오차 크기의 절대값을 이용하여 계산하였다.

〈표 5-9〉 무응답 원인별 오차 크기

무응답오차 원인	단독주택		아파트		기타	
	오차크기	비율(%)	오차크기	비율(%)	오차크기	비율(%)
거부	0.0084	97.67	(-)0.0064	85.33	-0.002	60.6
부재	0.0002	2.33	0.0011	14.67	-0.0013	39.4
무응답 오차	0.0086	100	-0.0052	100	-0.0033	100

이 결과를 정리해 보자. 우선 단독주택이 목표 변수일 경우, 무응답 오차 0.0086이 의미하는 바는 단독가구 비율을 추정하는 데 있어서 약 0.86%는 잘못 추정될 수 있다는 것이다. 즉, 1,000가구당 약 10가구 정도는 단독주택이 아닌 다른 주택 유형으로 잘못 추정될 수 있음을 의미한다. 또한 무응답 오차는 다시 거부와 부재에 의한 오차로 나누어질 수 있고, 이 중 대부분이 거부로 인한 오차라는 것이다. 이 결과는 <표 5-1>에서 무응답 오차를 구성하는 무응답 비율 중 거부비율이 5.7%로 부재비율 2.6%보다 높다는 것과 관련하여 거부로 인한 무응답은 비율적인 측면이나 오차크기 측면에서 추정치의 편향을 가져올 수 있는 가능성이 상당히 잠재되어 있다. 이처럼 거부 원인이 무응답의 위험요인으로 다가오는 것은 주택유형에 있어서는 부재자가 거부자보다 응답자를 더 닮았기 때문으로 볼 수 있다.

나. 주택소유 여부(자가/전·월세 등)

앞에서는 주택유형을 추정치 변수(실업자 수)와 관련된 핵심변수로 고려하였지만, 여기서는 주택소유 여부를 핵심변수로 사용하여 무응답 오차 원인별로 계산해 보자. 주택소유 여부는 국민의 경제적 상황과 매우 상관된 변수로서 실업자와 연관성이 높은 핵심변수로 고려될 수 있다. 따라서 주택유형에서 계산했던 것과 동일한 방법으로 주택소유 여부 변수를 이용하여 무응답 유형에 따른 오차 크기를 측정하였다. 주택소유에 대한 가구의 분포는 <표 5-10>과 같다. 소유자는 응답비율이

93.7%로 무소유보다 높고, 무소유자의 경우는 부재보다 거부에 의한 무응답이 상대적으로 높게 나타났다.

〈표 5-10〉 주택소유 여부별 가구 분포

응답 유형	소유		무소유		전체	
	n	%	n	%	n	%
응답	595	93.7	277	87.66	872	91.69
부재	13	2.05	12	3.80	25	2.63
거부	27	4.25	27	8.54	54	5.68
전체	635	100	316	100	951	100

전체 표본 중 주택소유자는 68.23%이고 응답자 중 주택소유자는 66.77%로 조사되었다. 또한 주택소유자의 경우 따라서 주택 소유자 비율이 추정값이라면 각각 $\bar{y}_r = 0.6823$, $\bar{y}_n = 0.6677$ 이다. 식 (2)에 따르면 다음과 같다.

$$0.6823 = 0.6677 + \left(\frac{54}{951}\right)(0.6823 - 0.5000) + \left(\frac{25}{951}\right)(0.6823 - 0.5200).$$

여기서 거부에 의한 무응답 오차는 0.0104(70.75%), 부재에 의한 오차는 0.0043(29.25%)으로, 거부에 의한 오차가 전체 오차의 상당 부분을 차지하였다. 주택소유 여부 변수에 대해서도 주택유형과 마찬가지로 부재에 의한 오차보다는 거부에 의한 오차가 상대적으로 많은 비중을 차지하는 것으로 나타났다. 주택소유에 따른 오차는 표본을 설계할 때 소유자에 더 많은 표본(전체의 66.7%, 635가구)이 배정된 것과 무관하지 않을 수 있다.

지금까지의 무응답 오차 분석과 관련해서 선행 연구(Curtin, Presser, & Singer, 2000; Groves, Presser, & Dipko, 2004)는 약간 다른 결과를 제시하고 있다. 일반적으로 전체 무응답 비율 측면에서는 부재보다 거부에 의한 무응답 비율이 높은 반면, 무응답 오차 측면에서는 무응답 비율이 상대적으로 더 적은 부재에 의한 무응답 오차가 거부에 의한 무응답 오차보다 큰 것으로 보고되고 있다. 이들 연구는 무응답 오차에 대한 평

가 없이 거부율과 부재 비율만을 고려하여 거부에 의한 오차가 부재에 의한 오차보다 더 클 것이라는 잘못을 범할 수 있다고 말하고 있다. 즉, 무응답 오차는 결측률(missing rate)뿐만 아니라 결측에 의한 오차의 크기를 동시에 고려해야 한다는 것을 강조하고 있다.

그렇지만 본 연구 결과에서는 무응답 비율 측면에서는 거부율이 부재율보다 높고, 오차 역시 거부에 의한 오차가 부재에 의한 오차보다 훨씬 높게 측정되었다. 이는 거부율이 부재율보다 높지만, 부재에 의한 오차가 거부에 의한 오차보다 훨씬 크다는 선행 연구결과들과는 다른 양상을 보이는 것으로 나타났다. 물론 이 연구의 결과를 일반화시키기에는 자료나 조사방법 측면에서 여러 가지 제약이 따른다. 그러나 무응답 오차는 무응답률과 상관이 있고, 세부적으로는 무응답 유형별로 서로 다른 양상을 보인다는 것을 확인할 수 있다. 뿐만 아니라, 이 결과는 조사의 성격과 국내외 응답가구의 성향이 매우 다를 수 있기 때문에 주의할 필요가 있음을 깨닫게 한다.

제4절 결론 및 논의

1. 연구요약

본 연구의 목적은 경찰조사를 실시하는 가구를 대상으로 통계조사에서 발생하는 무응답 현황과 특성을 파악하고, 이를 기반으로 무응답 유형에 따른 무응답 오차를 평가하는 것이었다. 즉, 응답률과 조사원의 노력 정도 또는 응답률과 가구의 반응정도와의 관계를 살펴보고, 무응답의 유형을 심층 분석하고, 무응답 유형별로 오차를 측정함으로써 무응답 특성과 자료의 추정치의 관계를 살펴보고자 하였다. 무응답 오차를 평가할 때 key 변수값은 ‘주택유형’과 ‘주택소유’를 사용하였다.

조사원들은 응답률 향상을 위해 근무시간 변경을 포함한 여러 번의 재방문을 통해 불응가구 설득을 위한 노력을 하는 것으로 조사되었다. 조사결과 경찰조사 최종 응답률은 약 91.7%로 매우 높았고, 이 중 최초 응답 거부 이후 응답으로 전환된 가구의 비율이 상당히 높은 것으로 나

타났다. 이는 조사원들의 응답률 향상에 대한 노력의 효과라 할 수 있다. 한편, 일부 표본가구는 조사에 대해 거부감을 나타냈으며, 특히 응답 거부시 매우 화를 낸다거나 강하게 거부하는 가구들은 조사원들의 조사 부담을 가중시키는 하나의 요인으로 작용하였다. 표본 대표성 여부에 있어서 경찰조사 10월 신표본은 주택유형에 대해서는 단독과 아파트의 비중이 유사하였지만, 주택소유여부에 대해서는 주택 소유 가구의 비중이 다소 높고, 전월세를 포함한 주택 미소지자가 낮은 대표성을 갖는 것으로 나타났다. 연구의 결과는 다음과 같다.

첫째, 경찰조사의 표본가구는 상당히 조사에 협조적이었다. 최초 대면에서 응답에 협조하는 경우가 60% 이상이고, 조사원의 재방문 등의 설득노력으로 약 30% 이상이 응답 전환되었다.

둘째, 무응답의 유형은 부재보다 거부로 인한 발생이 더 많았다. 사용된 자료의 특성에 의하면 부재층은 거부층보다 응답층을 더 많은 것으로 파악되었다.

셋째, 조사원들은 응답률을 높이기 위해 여러 번의 재방문을 시도하고 있으며, 최초 대면에서 거부한 가구를 대상으로 근무시간외 또는 주말 시간을 이용하여 설득을 위한 재방문을 시도하는 경우도 많았다. 그러나 무응답 유형별 또는 가구의 거부 이유에 따라 재방문 시기 및 조사원 재배치 등과 같은 조치를 다르게 취한다는 근거는 찾을 수 없었다.

넷째, 표본가구는 주로 ‘사생활 침해’와 ‘바쁘다’를 이유로 응답을 거부하였지만, 결국은 무관심(‘무대응’, ‘문전박대’)에 비해 응답 전환이 되기 쉬웠다. 이는 회귀분석을 통해서 재확인되었으며, 응답 전환에 영향을 주는 주요 변수는 거부이유와 설득기간이었고, ‘무대응’이나 ‘문전박대’에 비해 ‘바쁘다’ 또는 ‘사생활 침해’라는 이유로 거부한 경우가 응답 전환이 되기 쉬웠다. 설득기간에 대해서는 설득기간이 길수록 응답 전환 가능성이 낮아진다는 것을 알 수 있었다. 이는 부재가구나 거부가구일수록 설득기간이 길어지기 때문에 발생하는 현상으로 볼 수 있다.

다섯째, 무응답 유형별로 무응답오차를 평가한 결과, 부재보다 거부로 인한 무응답 오차가 크게 나타났다. 즉, 부재자는 응답자와 매우 유사한 성격을 가졌으며, 이와 같은 부재자 그룹에서 응답을 얻지 못한다 하더라도 추정치에는 큰 영향을 주지 않을 것이라는 것이다. 그러나 전

체 무응답 오차의 대부분을 차지하는 거부 층의 경우, 설득을 통해 응답을 얻지 못하면 추정치에 엄청난 편향을 초래할 가능성이 높다.

2. 현장조사 개선을 위한 실천 사항

본절에서는 지금까지의 연구결과를 바탕으로 현장조사 개선을 위해 가장 우선으로 실천되어야 할 “Best Practice”를 거론하고자 한다. 특히 현장조사의 개선을 위해 자료수집 단계와 분석단계에서 검토되어야 할 문제들을 논의하고자 한다. 자료수집과 분석단계에서 이루어지는 일련의 과정들은 조사의 성패를 좌우하는 가장 중요한 단계이다. 이는 곧 자료의 품질과 직결되며 현장조사에 발견된 문제점을 보완하여 다시 조사 현장에 피드백하기 위해 필요한 중요한 단계이다. 그럼에도 불구하고 많은 조사에서 이들 과정에서의 섬세한 검토가 이루어지지 못하는 것이 아닐까 하는 우려가 없지 않다. 이 연구를 계기로 자료수집과 분석단계에 대한 정밀 진단이 왜 중요하고, 이 진단을 통해 무엇을 찾아내야 하는지에 대한 우리의 역할을 찾아낼 수 있을 것으로 기대한다. 여기에서 논의된 사항들은 캐나다 통계청의 *Best Practice in Public Opinion Research* (Canada Statistics, 2007)에 기반을 두었으며, 전체적인 검토를 통해 우리에게 우선적으로 필요하다고 생각되는 부분을 논의의 대상으로 삼았다.

자료수집단계에서 응답거부와 조사의 조기 종결을 방지하기 위한 실천은 다섯 가지 측면에서 살펴볼 수 있다.

① 유능한 조사원 활용을 검토해야 한다. 조사원은 표본이 조사에 잘 참여할 수 있도록 독려하고 자료를 정확하게 수집할 수 있도록 훈련되어 있어야 한다. 조사원들은 전문적이고 우호적이어야 하며, 이를 위해 전문적인 조사원 훈련 프로그램을 개발하고 이를 매뉴얼화해야 한다.

② 현장조사 감독자들이 자료수집과정에서 최소한의 조사에 대해 조사 전반의 과정을 모니터링하도록 해야 한다. 모니터링은 조사원이 최고의 작업을 수행하도록 독려함으로써 조사기간 동안의 조사의 질을 보장할 수 있을 것이다.

③ 자료수집 동안 무응답에 대한 이유를 모니터링해야 한다. 이는 단지

조사 목적과 예산 측면에서 연구를 유지하는 데만 도움이 되는 것은 아니다. 무응답에 대한 이유를 검토하는 것은 조사 감독자가 응답률을 최대한으로 높이기 위한 조정을 하는 데 도움이 될 수 있다. 만약 무응답률이 부재 때문에 높다면 감독자는 조사원들에게 응답자들과 더 많은 접촉을 하기 위해 주중의 다른 시각 또는 다른 날에 재방문을 하도록 요구해야 한다. 역으로 거부 때문에 무응답률이 높다면 조사감독자는 최고의 조사원만이 그 조사를 수행할 수 있도록 조정해야 한다.

④ 목표 모집단의 다른 부차 집단간의 무응답 수준을 모니터하도록 해야 한다. 이는 높은 응답률이 반드시 무응답 오차를 줄이는 것은 아니기 때문이다. 조사는 높은 응답률을 추구하지만 무응답자들이 실질적으로 응답자와 다르다면 그 결과들을 전체 목표 모집단으로 일반화시킬 수 없게 될 것이다. 따라서 보다 더 신중하게 핵심 특성을 갖는 표본에 대해 시간을 두고 적절한 비중으로 조사를 수행하도록 해야 할 것이다.

⑤ 거부전환을 시도해야 한다. 응답 거부 전환은 조사기관의 자료수집에 있어서의 핵심적인 측면이다. 초기에 조사를 꺼려 하는 사람들이 조사에 참여할 수 있도록 재방문 시기라든가 조사방법을 바꿈으로써 응답 전환을 시도해야 할 것이다.

자료수집과정에서 응답률 향상을 위한 노력은 모든 통계조사에서 끊임없이 진행되어 왔던 부분일 것이다. 이를 통해 실질적인 응답률 향상 뿐만 아니라 조사의 질도 상당히 높아진 것도 사실이다. 이처럼 응답률은 조사 설계·기획과 수집단계에서 잘 설명될 수 있다. 그렇다면 자료 분석 과정을 통해 조사 관리 기관은 무응답에 대한 설명과 무응답 편향에 대한 설명을 위해 최선의 절차를 따라야 한다. 이들 노력은 응답률 자체를 높이지는 못하지만 무응답을 보충하고 자료 품질 측면에서 신뢰도를 향상시키는 데 도움이 될 것이다. 자료분석 단계에서 가장 우선되어야 할 다섯 가지 실천 사항은 다음과 같다.

① 부차 그룹에 대해 응답률을 비교해야 한다. 이들 부차 그룹은 연령, 성별, 지역 등이 해당될 수 있다. 이 방법은 무응답 편향 정도를 결정하는 데 도움이 되는 것은 아니지만 편향의 존재 여부를 파악할 수 있다.

② 자료에 대하여 가중치를 부여해야 한다. 통계적 조정(statistical adjust)

또는 관측변수에 의한 가중치 부여는 무응답을 설명하기 위해 사용되는 가장 보편적인 방법이다(Groves et al., 2001). 무응답 편향이 거의 없다는 가정을 만족하는 경우, 자료에 가중치를 부여하는 것은 이용자에게 제공되는 결과가 모집단에 대한 대표성을 갖는다는 것을 보장하는 것이다. 그러나 가중치를 부여하는 것은 무응답 편향을 설명하는 데는 효과적이지 않다. 편향은 무응답이 응답과 무응답 간의 차이를 가져올 때 발생하고, 이러한 무응답 편향을 수정하기 위해 연구자들은 조사 동의에 영향을 줄 수 있는 요인이 무엇인지를 이해할 필요가 있다. 그러한 이해는 응답과 무응답 간의 체계적인 차이가 관측변수를 능가할 수 있기 때문에 중요하다(성별, 수입별, 지역별 등).

③ 응답자와 무응답자를 비교해야 한다. 가능하다면 응답자와 무응답자의 특성을 비교하기 위해 자료 수집 후에 무응답 분석을 수행해야 한다. 이러한 비교는 조사기관이 인구학적 또는 알려진 다른 변수들에 의해 무응답자와 응답자가 어떻게 다른지를 평가할 수 있을 것이다. 이 둘 사이에서 어떤 차이가 있다면 이는 무응답 편향이 존재함을 암시하는 것이다. 이를 위해 무응답자에 대한 정보를 파악할 수 있어야 하고, 이는 현장에서 조사원이 직접 가구의 방문기록을 작성하여 실제 조사 결과와 연계함으로써 실현될 수 있다.

④ 조사에 빨리 응답하는 자와 늦게 응답하는 자들의 특성을 비교해야 한다. 빠른 응답자와 늦은 응답자의 응답결과를 비교하는 근본 이유는 늦은 응답자일수록 거의 무응답자일 것이라는 가정에서 비롯된 것이다. 왜냐하면 만약 조사원이 이들을 접촉하기 위해서 추가 노력을 하지 않는다면 그들은 또한 무응답자가 되었을 것이기 때문이다. 만약 핵심 변수에 대해서 두 그룹간에 통계적으로 유의한 차이를 보이지 않는다면, 조사 자료는 보다 신뢰성을 가지고 일반화될 수 있다. 그러나 이 방법의 단점은 접근하기 어려운 응답자와 무응답자가 유사하다는 것을 전제로 한다는 것이다. 이 가정은 일반적으로 적용될 수 없기 때문에 때로는 무응답 편향의 존재에 관해 부정확한 정보를 제공할 수 있다는 점에 유의해야 한다.

⑤ 필요할 경우 무응답자 추적을 수행할 수 있어야 한다. 조사에서 우리가 무응답을 다루는 핵심은 무응답 편향을 인식함으로써 이 무응답

이 자료에 영향을 미치는 정도를 아는 것이다. 한 가지 방법은 무응답자와 응답자 간의 차이를 양적화하는 것이고, 이를 위해 현장조사를 완료하는 시점에서 무응답자들 중 부차 표본을 조사하는 것이다. 이 부차 그룹의 응답은 응답그룹과 비교될 수 있다. 그러나 단점은 많은 비용과 시간을 요한다는 것이다. 따라서 조사기관은 무응답 수준이 무응답 편향을 초래하는지를 우선 결정할 필요가 있다. 만약 무응답자와 응답자가 체계적으로 다르다는 타당한 근거가 있다면 많은 비용에도 불구하고 무응답자의 부차 표본 조사는 필요하고 가치가 있을 것이다.

3. 맺음말

현장조사 개선 문제는 여러 방면에서 접근할 수 있는 문제이지만, 현장조사 개선을 위한 궁극적인 목적은 조사 추정치를 전제로 하지 않을 수 없다. 게다가 조사원의 업무부담, 예산, 조사환경 등 복잡한 문제들이 산재되어 있지만 어느 하나 소홀히 할 수 있는 문제는 아니라고 생각된다. 무엇이 우선이 되어야 하는가를 결정하는 문제가 남아 있다.

통계조사기관으로서 자료의 품질은 생명과도 같다. 아무리 시대의 변화로 또는 국민의식의 변화로 조사환경이 날로 어려워진다고 하지만, 그렇다고 이러한 이유로 자료의 품질을 희생할 수는 없는 것이다. 이는 어느 누구에도 타당한 이유로 받아들여질 수 없는 부분이라 하겠다. 그렇다면 이 현실적인 문제를 어떻게 해결해 나갈 것인가가 우리의 숙제로 남는다. 이러한 문제는 경험적 또는 과학적 지식을 복합적으로 이용함으로써 해결될 수 있을 것으로 본다. 시대와 함께 변해 가는 조사현장의 상황을 자료를 통해 정확하게 진단하고 이로부터 정확한 처방전을 가지고 치료를 해 나가야 할 것이다. 조사에서 오차가 왜 발생하고, 왜 발생할 수밖에 없는지를 구분하여 정확하게 파악해야 한다. 왜 오차가 발생하는지를 알았다면 그 오차가 왜 발생할 수밖에 없었는지를 표본을 통해 정밀분석을 수행해야 할 것이다. 연구의 결과에서 살펴본 바와 같이, 오차의 원인에 따라 오차의 양적 정도를 파악함으로써 어떤 원인에 더 비중을 두고 조사에 임해야 하는지를 자료 품질과 연계하여 조사기획을 해 나가야 할 것이다.

물론 이는 현장의 몫이기 전에 조사기획 또는 관리 기관의 몫이 아닌가 싶다. 이러한 철저한 전략을 조사현장에 전달함으로써 조사기관으로부터 조사감독자를 통해 현장조사원들이 무엇을 우선적으로 해야 하는지를 계획하도록 해야 할 것이다. 실제 조사현장의 최전방에서 조사에 임하는 조사원들이 단지 응답률을 높이는 데만 집중하게 해서는 안 된다는 것이다. 왜 응답률을 높여야 하는지, 왜 응답전환 노력을 해야 하는지, 왜 거부자들을 설득해야 되는지, 그리고 어쩔 수 없는 무응답에 대해서는 왜 대체를 하게 되는지에 대해 자료의 품질과 관련하여 인지하도록 해야 한다. 이를 위해서 조사설계, 자료수집, 자료분석 등 각 단계에 대한 최적의 필요한 절차에 대한 매뉴얼이 우선적으로 작성되어야 하고, 이 매뉴얼을 근거로 조사원 훈련과 조사가 체계적으로 진행되어야 할 것이다.

그러나 현장조사 개선이 조사관리 기관의 노력만으로 해결될 수 있는 문제는 아니다. 현장에서 발생하는 문제는 현장에서 직접 조사를 경험한 자만이 가장 잘 알 수 있는 문제일 것이다. 즉, 현장 조사 개선은 조사원들의 협조가 없이는 절대 이루어질 수 없는 부분이라 생각된다. 아무리 훌륭한 조사 매뉴얼과 훈련 매뉴얼이 존재한다 하더라도 현장조사를 책임지고 있는 조사원들이 이들 지침에 따라 행동하지 않는다면 이런 매뉴얼들은 아무 쓸모가 없는 것이 되고 만다. 조사원들이 적극적으로 조사과정과 내용에 대한 전반적인 절차를 숙지하고 실천함으로써 현장조사 개선은 가능하리라 본다. 뿐만 아니라 조사원은 조사결과가 국가의 주요 정책 또는 현안 문제를 해결하는 데 필요한 중요 정보라는 점을 인지하고, 이를 위해 가장 솔직한 자세로 조사에 임할 수 있어야 한다. 결국 현장조사 개선을 위해 지금 현재 우리에게 가장 필요한 것은 현장조사 기관과 조사관리 기관의 상호이해와 적극적인 협조가 아닐까 한다.

한편, 조사를 관리하는 입장에서는 100% 응답률을 요구할 수는 없는 노릇이다. 더구나 100% 달성 노력이 단지 조사원 평가나 기관 평가를 목적으로 행해진다면 이것은 더더욱 위험한 행동이 될 것이다. 조사관리 기관이라면 최소한 점점 어려워지는 조사환경에서 최적의 응답률은 과연 100%여야만 하는지를 연구할 필요가 있다. 해외의 경우를 보면 우

리보다 먼저 조사환경이 열악해졌고, 조사를 바라보는 국민의 의식도 삭막해졌다 할 수 있다. 이들은 변해 가는 환경, 즉 인력, 예산, 시간 등을 고려하여 자료의 품질을 볼 때 100%가 최적의 응답률은 아니라는 관점에서 조사방법 개선을 위한 노력을 하고 있다. 물론 이러한 노력이 단지 경험에서만 행해질 것이 아니라 심층 분석 연구를 통해 과학적인 방법으로 이루어질 때 대외적으로도 설득력을 갖게 된다.

이러한 측면에서 본 연구의 결과만으로는 조사시간, 최적 응답률, 적정 방문횟수 등과 같은 문제를 직접적으로 논할 수는 없다. 이를 파악하기 위해서는 좀 더 대대적인 차원에서, 예를 들면 조사기관이 적극적인 의지를 가지고 진행되어야 할 과제라고 생각한다. 단지 소규모 단위의 연구만으로는 우리가 원하는 결과를 파악하기는 힘들 것이다.

마지막으로 이 연구를 계기로 보다 넓은 시각에서 현장조사 개선을 위한 노력이 계획적으로 이루어지길 바란다. 조사를 전적으로 기획하고 관리하는 입장에서 체계적인 연구 의지를 가지고 단지 해외의 방법 도입하는 차원이 아닌 우리의 환경에 맞는 최적의 조사 방법을 찾고 조사정비 체제를 구축할 수 있기를 기대해 본다.

참고문헌

- 조사통계연구회(2000), 「무응답 오차」, 자유아카데미.
- 통계청(2007), 「가구부문 표본개편 결과」, p.6.
- 통계청 (2005), 「인구주택총조사」.
- Alreck, Pamela L., Settle Robert(1995), *The Survey Research Handbook*, New York: McGraw-Hill.
- Babbie, E.(2007), *The Practice of Social Research*, 11th ed., Belmont, CA: Wadsworth.
- Canada Statistics(2007), *Best Practices in Public Opinion Research*.
- Curtin, R., S. Presser, and E. Singer(2000), “The effects of response rate changes on the index of consumer sentiment”, *Public Opinion Quarterly*, 64, pp.413-428.
- David, D. and H. Melissa(2006), “Finding Coverts: Who becomes a refusal conversion?”, *International Communications Research*, www.icrdurvey.com.
- De E. Leeuw and W. De Heer(2001), “Trends in households survey non-response: A longitudinal and international comparison”, in R. Groves, D. Dillman, J. Eltinge, and R. Little(eds.), *Survey Non-Response*, New York: Wiley, pp.41-54.
- Groves, R. M.(2006), “Nonresponse rates and nonresponse bias in household survey”, *Public Opinion Quarterly*, 70, pp.646-675.
- Groves, R. M. and M. P. Couper(1998), *Nonresponse in Household Interview Survey*, New York: Wiley.
- Groves, R. M., S. Presser, and S. Dipko(2004), “The role of topic interest in survey participation decisions”, *Public Opinion Quarterly*, 68, pp.2-31.

- Groves, R. M., W. Douglas, G. Liberty et al.(2001), “Common influences on noncontact nonresponse across household survey: theory and data”, Unpublished manuscript.
- Heerwgh, D., K. Abts, and G. Loosveldt(2007), “Minimizing survey refusal and noncontact rates: do our efforts pay off?”, *Survey Research Methods*, 1, pp.3-10.
- Keeter, S., C. Kenndey, M. Dimock, J. Best, and P. Craighill(2006), “Gauging the impact of growing nonresponse on estimates from a national rdd telephone survey”, *Public Opinion Quarterly*, 70, pp.759-779.
- Keeter, S., C. Miller, A. Kohut, R. M. Groves, and S. Presser(2000), “Consequences of reducing nonresponse in a telephone survey”, *Public Opinion Quarterly*, 64, pp.125-148.
- Lessler, Judith T., William D. Kalsbeek(1992), *Nonsampling Error in Surveys*, Wiley.
- Lynn, P., P. Clarke, J. Martin, and P. Sturgis(2002), “The effects of extended interviewer feeorts on nonresponse bias”, in R. Groves, D. Dillman, J. Eltinge, and R. Little(eds.), *Survey Nonresponse*, New York: Wiley, pp.135-147.
- Martin, B. and W. Karen(2007), “The impact of refusal conversion on survey response and error”, 2007 Joint Statistical Meetings, Salt Lake City, 2007.7.29~8.2.
- Mayer, C. S. and R. W. Pratt Jr.(1996), “A note on nonresponse in a mail survey”, *Public Opinion Quarterly*, 30, pp.637-646.
- Neyman, J.(1934), “On the two different aspects of the representative methods: The method of stratified sampling and the method of purposive selection”, *Journal of the Royal Statistical Society*, A. 97, pp.558-606.
- Pearson, K.(1903), “On the influence of antural of natural selection on the variability and correlations of organs”, *Philosophical Transactions of the Royal Society*, A. 200, pp.1-66.

- Rainer, S.(2004), "Different types of nonresponse, different causes, different effects and different remedies", European Science Foundation Seminar, [Measurement, Data collection and Data quality], 12-13 August 2004, Switzerland.
- Rainer, S. and T. Mark(2006), "The effect of the refusal avoidance training experiment on final disposition codes in the German", University of Konstanz, Germany.
- Singer, E.(2006), "Introduction-Nonresponse bias in household surveys", *Public Opinion Quarterly*, 70, pp.637-645.
- Singleton, R. and Bruce Straits(2005), *Approaches to Social Research*, 4th ed., New York: Oxford University Press.
- Stoop, I.(2005), "Nonresponse in sample survey", *The Hunt for the Last Respondent*, The Hague: Social and Cultural Planning Office.
- Ulrich, K.(2007), "Surveys from inside: An assessment of unit nonresponse bias with internal criteria", *Survey Research Methods*, 1, pp.55-67.
- Voogt, R. J. and W. E. Saris(2005), "Mixed mode design: Finding the balance between nonresponse bias and mode effects", *Journal of Official Statistics*, 21, pp.367-387.

< 부 록 >

1. 방문기록용 조사표

조 사 요 령

- (1) 응답거부 여부: 최초 방문시 응답거부가 있으면 ○, 아니면 ×
- (2) 응답거부 이유: 최초 방문 당시 응답자가 응답을 거부한 주된 이유(1개항만 선택) **단, 최초 방문이란 조사원이 응답자와 처음으로 대면했을 때를 말함**

바쁘다

- ① 현재 바쁘다(해주고 싶은데 시간이 없다, 방문한 시기(요일, 시간 등)가 적절치 않다)
- ② 평상시 워낙 바쁘다(해주고 싶지 않다)

귀찮다

- ③ 정부(통계청 포함)불신(뒤흠리는 조사인지 모르겠다. 조사해 가서 무엇에 쓰느냐)
- ④ 통계조사 불만(통계청 조사가 맞게 하고 있느냐, 대충 하는 거 아니냐)
- ⑤ 사생활 침해(왜 나한테 와서 조사하느냐)
- ⑥ 조사 방법에 대한 불만(달마다, 혹은 직접 찾아오는 것 불편하다)

화낸다

- ⑦ 정부(통계청 포함)불신(강력히 자신의 의견을 피력하는 경우)
- ⑧ 통계조사 불만(강력히 자신의 의견을 피력하는 경우)

무조건 불응(피한다)

- ⑨ 불응의사를 명확히 밝힘(설명도 듣고 이해도 하였으나 이유 여부를 막론하고 무조건 조사에 응하지 않겠다고 함)
- ⑩ 문전박대(대면은 했으나 설명도 필요없고 절대 응답해주지 않겠다고 강하게 거부)
- ⑪ 무대응(가구 내에 인기척이 느껴지지만 문을 열어주지 않고 대응하지 않는 경우)

기 타

- ⑫ 그 밖의 의견

- (3) 화냄 정도: 응답자가 어느 정도 화를 내며 응답을 거부하였는가? 단, 응답을 거부한 가구만 해당
- ① 매우 화냄
 - ② 약간 화냄
 - ③ 보통
 - ④ 별로 화내지 않음
 - ⑤ 전혀 화내지 않음
- (4) 방문일시: 기입 예) 9월 10일(월) 오후 3시 → 9월 10일 15시 (월)
단, 마지막 방문일시란? 최종 승낙 여부를 결정한 마지막 방문으로 본조사가 이루어지기 직전까지 기간에 활동한 것을 의미함. (즉, 본조사 기간의 응답결과는 포함하지 않음)
- (5) 총방문횟수: 기입 예) 총방문횟수 5회, 그 중 부재중 방문 2회이면 5(2).
단, 최종적으로 응답을 하지 않은 가구일 경우, (5)란에는 최종 방문시까지의 방문횟수와 그 중 부재중일 때 방문횟수를 위의 예와 같이 기입
- (6) 최종 응답 여부: 조사원의 최종 설득작업을 거친 후 응답을 했으면 ○, 아니면 ×
- (7) 응답 거부 이유가 ⑫인 경우 구체적 이유 작성

조 사 표

□ 최초 응답거부 이후의 설득과정에 대한 현황을 ‘작성요령’에 따라 다음 표에 기입하십시오.

조사구 번호	구역 및 거처 번호	가구 번호	최초 방문시			(4) 방문일시 월 일 시 (요일)	
			(1) 응답거부 여 부	(2) 응답거부 이 유	(3) 화냄 정도	최초	마지막
1234 5678	0102	0304	○	⑤	②	9월 17일 15시(월)	9월 21일 17시(금)

조사구 번호	(5) 총방문 횟수	(6) 최종응답 여부	(7) 응답거부 이유가 ‘㉔’인 경우 구체적인 거부 이유
1234 5678	3(1)	○	

※ 작성 예시 : 조사표 내의 이탤릭체 글씨처럼 작성