조사연구(2021. 8) 22권 3호: 1−25
DOI http://dx.doi.org/10.20997/SR.22.3.1

연구논문

비면접원조사는 면접원조사만큼 신뢰할 만한가: 코로나시대의 사회조사 한계 극복을 위한 연구*

조성겸** · 김성중*** · 강미나****

코로나사태와 조사환경의 변화로 인해 면접원조사의 실효성은 점차 줄어들고 있다. 이러한 상황에 대처하기 위해 다양한 조사방법을 동시에 사용하는 혼합조사를 과학적 사회조사에 도입할 필요성이 제기된다. 혼합조사의 도입에서 문제가되는 것은 상이한 조사방법을 동시에 사용함으로써 나타나는 모드효과이다. 이 연구에서는 모드효과 중 측정효과의 발생 여부와 발생할 경우 얼마나 심각한 수준인지를 실험을 통해 검증해 보았다. 실험을 위해 한 리서치전문회사가 보유한 온라인 패널을 대상으로 동일한 질문으로 구성된 전화조사와 모바일조사를 2개의 집단에 순서를 달리하여 실시하였다. 전화조사와 모바일조사가 동일한 결과를 산출하는지 검증하기 위해 동등성 검증을 실시하였다. 분석의 결과, 원자료를 사용한 비교에서는 양 조사간 오차의 크기에서 차이가 나타났다. 이상치의 영향을 제거한 절삭된 평균을 사용할 경우 양 조사방법 간 오차크기와 측정치 기대값의 실질적인 차이는 없는 것으로 나타났다. 본 연구의 결과는 비면접원조사가 측정오차를 증가시키거나 편향을 발생시키는 것은 사실이지만, 이는 충분히 관리 가능하므로 과학적 사회조사에서 혼합조사 도입이 가능하다는 것을 시사한다.

주제어: 혼합조사, 모드효과, 측정효과, 동등성 검증, 사회조사

* 이 연구는 충남대학교 학술연구비에 의해 지원되었음. This work was supported by research fund of Chungnam National University.
** 충남대학교 언론정보학과 교수(skcho99@gmail.com), 제1저자.
*** 충남대학교 자유전공학부 강사(sungjoongk@gmail.com), 교신저자
**** 국토연구원 선임연구위원(mikykang@krihs.re.kr), 제2저자.

Ⅰ. 문제 제기

코로나 바이러스19 사태는 한국뿐 아니라 지구촌 사회의 모든 영역에 영향을 미치고 있다. 전 세계인들은 일상생활의 전면적인 변화를 경험하였으며, 이러한 변화는 사회조사에서도 발생하고 있다. 특히 전통적인 면접원조사를 활용한 사회조사는 조사 수행 자체에 곤란을 겪고 있다. UN 경제사회국(Department of Economic and Social Affairs)이 2020년 5월 세계 각국의 국가통계청(National Statistical Offices)을 대상으로 코로나사태가 미친 영향에 대해 조사한 결과에 따르면, 조사에 응한 국가 중 96%의 국가에서 면대면조사에 영향이 있으며, 이 중 69%는 면대면조사를 전면 중단하였다고 보고하였다(UN 2020년 6월 5일). 2020년 12월에 이루어진 동일한 조사에 따르면 전체 응답국가의 57%가 면대면조사로 복귀할 수 있는 시점이 불투명하다고 대답하였다. 미국 내 지역별 코로나 바이러스 감염율과 2020 센서스 조사 응답률간의 관계를 분석한 연구(Bates & Zamadics 2021)에 따르면, 고령자나 소수인종 비율과 같은, 코로나바이러스 감염률에 영향을 미치는 요인을 통제한 이후에도 지역별 감염률이 높을수록 센서스 응답률이 낮아지는 경향이 있다.

한국의 경우 국가통계와 같이 과학적 표집을 이용하는 조사의 대부분은 면접원의 가정방문을 통해 이루어져 왔다. 2020년 7월 기준 국가승인 통계조사 534건 중 대면면접조사를 사용하는 조사가 423건(79.2%)에 이를 정도이다(이찬복 2020). 조사에 이용되는 표집틀이 인구센서스에 기반한 주소를 기반으로 하기 때문이다. 그러나 코로나 사태는 이러한 접근방법에 전면적인 전환을 요구하고 있다. 코로나 상황하에서 면접조사는 응답자와 면접원 모두에게 바이러스 감염 위험을 감수해야하는 행위이기 때문에 면접원들의 가구방문 조사는 많은 어려움을 겪고 있다. 이러한 어려움을 극복하기 위해 설문유치 또는 웹접속 등 면접원 없이 자기응답 방법에 의존하는 비율이 증가하는 추세이다.

코로나 사태 상황에서 비면접원조사의 현실적 필요성은 인정되지만 면접조사에 비해 오차가 증가할 수 있다는 우려가 제기된다. 비면접원조사와 면접원조사를 병행할 경우 두 조사모드를 사용하여 생산된 조사결과가 상호대체 가능한 수준의 등가성이 보장되어야 동시 사용이 가능하다. 나아가 국가승인통계를 위한 사회조사의

경우 과거 자료와의 연속성 문제가 남는다. 새롭게 시도되는 비면접원조사의 결과가 면접조사에 주로 의존해온 과거 자료와 유사한 수준의 신뢰성을 갖추지 못할 경우 과거 통계자료와 비교가 불가능해지는 문제를 야기한다. 결국 국가승인통계를 2021년을 기준으로 재설정하거나, 코로나 바이러스 사태가 완화될 때까지 전면 보류해야 하는 사태까지 가능할 수도 있다.

조사의 실무 현장에서는 면접원의 역할을 중시한다. 조사문항에 대한 설명, 응답입력의 정확성 점검 등 면접원이 측정오차의 감소에 큰 역할을 하고 있다는 것이다. 이런 이유에서 면접원조사의 측정오차가 낮을 것이라는 인식이 일반적이다. 백지선 외(2017)는 조사모드간 차이를 선택효과 즉 두 조사방법 간의 응답자 차이에서 발생하는 차이와 측정 자체에서 발생하는 차이로 구분하고 있다. 그들은 측정으로 한정해 본다면, 면접원조사는 면접원의 존재로 인해 사회적 바람직성 등과 같은 편향이 나타나는 반면, 비대면 조사는 그렇지 않을 것이라고 보았다.

모드효과의 유무와는 별개로 코로나 사태하에서 면접원의 가구방문을 통한 조사에만 전적으로 의존하는 것은 현실적으로 불가능하며 비면접원조사방법의 병행이 필요하다. 나아가 코로나 사태가 종식되더라도 사회조사에서 전화조사나 웹서베이 등의 비대면 방식에 대한 의존도는 높아질 가능성이 높다. 사생활과 안전 등의 이유로 가구방문조사를 꺼리는 경향이 증대하고 있기 때문이며, 비대면조사가 갖는 조사에 소요되는 시간, 비용 면에서 장점이 있기 때문이다. Dillman(2017)은 응답률 저하 경향에 대처하기 위한 대안으로써 혼합조사의 적극적 도입을 주장하였다.

과학적 엄밀성을 추구하는 사회조사에 혼합조사방법을 도입하려면 모드 차이가 조사결과에 미치는 영향이 있는지의 여부와, 영향이 있을 경우 어떠한 방식으로 어느 정도 규모로 나타나는지 파악할 필요가 있다. 지금까지의 모드효과 연구는 주로 모드의 선택효과에 집중되어온 반면, 측정과정에 미치는 효과에 대한 연구는 미미한 실정이다. 본 연구는 모드의 차이 중에서 측정효과에 초점을 두고 살펴보고자 한다. 면접원조사가 측정과정에 미치는 효과는 면접원이 면접과정에 미치는 영향으로부터 기인한다고 볼 수 있다. 면접원은 질문에 대한 이해도를 높일 수 있고, 응답의 일관성과 타당성을 일차 검증하는 역할을 한다. 또 면접원이 응답자를 대신하여 입력함으로써 입력오류 등을 방지할 수 있고, 중도탈락을 예방하는 역할도 할 수 있다. 하지만 응답과정에 부정적인 영향을 줄 수도 있다. 대표적으로 면접원의 존재를 의식하여 특정 주제에 대해 사회적으로 바람직한 방향으로 응답하는 편향이 존

재한다(Dillman & Christian 2003; Mavletova & Couper 2013; Burkill et al. 2016). 반면 비면접원조사는 사회적으로 바람직한 응답과 같은 편향성의 문제는 적으나 자기응답식(self-administered) 조사 고유의 문제점들이 존재한다.

이 연구는 코로나 상황에서 비면접원조사의 대폭 도입이 불가피하며, 조사환경의 변화로 인해 혼합조사방법의 도입이 필요하다는 인식에 기반하고 있다. 코로나 사태로 인해 국가승인통계에서조차 이미 비면접원조사를 병행하고 있는 실정이다. 즉, 이미 의도하지 않게 혼합조사가 이루어지고 있다. 상황적 요인에 의해 충분한 준비 없이 부분적으로 도입되는 혼합조사는 비표본오차와 표본의 편향 가능성만 높이는 결과를 초래할 우려가 있다. 본 논문은 혼합조사의 도입이 조사결과의 신뢰성과 과거자료와의 호환성에 어떠한 영향을 미칠 것인가를 가늠하기 위해 조사방법에 따라 결과에서 어떠한 차이가 발생하는가를 알아보고자 한다. 모드효과는 선택효과와 측정오차 두 가지 종류가 존재한다(백지선 외 2017). 이 연구는 이 중 면접원조사와 비면접원조사의 측정오차에서의 차이를 분석하고자 한다. 이 연구의 목적은 두 조사방법 간 품질의 우위를 판단하는 것이 아니라 혼합조사 도입시 우려되는 모드효과가 있는가 즉, 상이한 조사방법이 동등한 결과를 생산하는가를 검증하는 것이다.

Ⅱ. 기존 연구 검토와 연구문제

1. 혼합조사의 필요성과 문제점

혼합조사 도입의 필요성이 제기되는 이유는 무엇보다 외국에서 사회조사를 위해 주로 의존해온 전화조사의 응답률이 지속적으로 낮아지는 반면, 대안적인 방법인 웹서베이, 모바일조사 등은 표본의 대표성을 보장하기에는 아직 부족한 수준이기 때문이다(Dillman 2017). PEW research center(2012)에 따르면 1997년에서 2012년까지 15년간 RDD(random digit dialing) 방식 전화조사 응답률은 35%에서 9%로 하락했다. Dutwin & Lavrakas(2016)에 따르면 9개 조사기관의 평균 전화조사 응답률은 2008년 15.8%에서 2015년 9.3%로 추락했다. 무선전화 보급 확대와 이에 따른 유선전화 보급률 저하, 전화번호 식별 서비스 등의 보급은 전화조사 응답률을

더욱 떨어뜨리는 요인으로 작용하고 있다(Dillman 2017). 더 이상 단일조사방법으로는 충분한 응답률을 확보하는 것이 힘들어지면서 혼합조사 방식의 도입필요성이 제기되고 있다.

Dillman(2017)은 혼합조사의 도입이 늦어진 것은 첫째, 대면조사, 전화조사, 우편조사 등의 단일한 조사방법을 통해 충분한 응답률을 확보하는 것이 가능했으며, 둘째, 90년대 이전에는 기술적 한계로 상이한 방식의 조사를 동시에 진행하는 데 난관이 있었기 때문이며, 나아가 조사회사들이 일반적으로 하나의 조사방법에 특화되어 있던 제도적 특성 때문이라고 설명한다.

혼합조사 방식은 응답률과 응답자 대상(coverage)의 확대라는 장점이 존재하지만, 도입에서의 문제점도 많다. 혼합조사의 도입에서 가장 큰 우려는 이른바 모드효과의 존재 여부이다. 선택효과는 가중치 적용 등을 통해 보정이 가능하므로 큰 장애가 되지 않는다. 가장 큰 난점은 조사방법에 따른 측정에서의 차이이다(Dillman 2017). 전화조사의 대안으로 고려되는 웹서베이, 모바일조사 등은 비면접원조사이기 때문에 자기응답식 형식을 사용한다. 자기응답식 설문은 면접원의 존재가 가져다주는 여러 이점이 사라지는 결과를 낳으며, 나아가 청각적 자극과 시각적 자극에 응답자가 다르게 반응할 가능성도 있다(Dillman 2021). 면접원조사와 비면접원조사 간에 체계적인 측정 차이가 발생한다면 단일한 자료를 생산하기 위해 두 방법을 혼용하는 것은 실용적 이점에도 불구하고 고려하기 어렵다.

2. 모드효과에 관한 연구들

모드효과에 대한 기존 연구들은 주로 대면 인터뷰와 자기응답식 설문간의 차이를 중점적으로 살펴보았다. 기존 연구에서는 면접원의 존재를 의식한 사회적으로 바람직한 응답 경향과 민감한 주제에 대한 응답에서의 차이가 발견되었다. 많은 연구에서 대면 인터뷰가 자기응답식 설문에 비해 사회적으로 바람직한 응답을 유도하고, 바람직하지 않은 행위는 덜 보고하게 하는 효과가 있다는 결과가 보고되었다(Mavletova & Couper 2013).

사회적으로 바람직한 응답경향은 건강 여부를 묻는 단순한 질문에서도 나타난다. Hochstim(1967; Dillman & Christian 2005에서 재인용)의 연구에서는 대면 인터뷰에서는 43%가, 전화 서베이에서는 38%, 우편 서베이에서는 30%가 본인의 건강상

태를 ‘매우 양호’(excellent)하다고 대답하였다. Biemer(1992; Dillman & Christian 2005에서 재인용)가 65세 이상을 대상으로 한 조사에서는 자기응답형 설문의 경우 응답자의 15%가, 대면 인터뷰에서는 27%가 건강상태가 ‘매우 좋다’라고 응답하였다. Christensen et al.(2013)의 연구에서는 건강과 관련된 삶의 질, 건강관련 행위, 사회관계, 장기질환에 관한 지표에서 대면 인터뷰가 일관성 있게 점수가 높게 나타났다.

사회적으로 바람직한 응답경향은 사회에 대한 기여의향을 묻는 질문에서도 자주 발견된다. 공적 서비스에 대한 지불가치 조사에서는 대개 대면인터뷰에서 더 큰 금액에 동의하는 경향이 발견된다. Leggett(2003)의 연구에서는 대면 인터뷰 응답자가 자기응답식 설문에 응답한 이에 비해 국립공원에 대한 지불가치(willingness to pay)를 20% 더 높게 응답했다. 대면 인터뷰와 인터넷 설문을 비교한 Marta_Pedroso et al.(2007)의 연구에서는 대면 인터뷰에서 환경보존에 대한 지불가치가 더 크게 나타났다. Maguire(2009)의 연구에서는 가상적인 기부 질문에 전화응답자가 대면 인터뷰에 비해 적게 동의한 반면, 기부액수는 전화조사의 경우가 더 큰 경향을 보였다. 하지만 Covey et al.(2010)의 철도 안전에 대한 설문을 이용한 연구에서는 대면 인터뷰와 인터넷 설문 간 차이가 발견되지 않았다.

성생활, 혼외정사, 현재 성생활에 대한 만족도 등과 같은 민감한 질문의 경우에도 차이가 나타난다. Burkill et al.(2016)의 연구에서는 British National Survey of Sexual Attitudes and Lifestyles에 참여한 응답자 527명을 대상으로 웹서베이와 대면조사에서 민간한 질문에 대한 응답에서 차이가 있는지를 조사하였다. 분석 결과, 남성은 지난해 파트너 수에 대해 웹조사에 비해 대면조사에서 더 많이 대답하는 경향을 보였다. 여성의 경우에는 불법적인 약 복용, 현재 성생활에 대한 만족도, 캐주얼한 성관계에 대한 태도, 혼외정사에 대한 태도 등에서 조사모드에 따른 차이를 보였다. 연구자들은 이러한 차이에도 불구하고 두 설문방식 간 추정치에서 큰 차이가 나타나지는 않는다고 결론내리고 있다.

조사방법은 무응답비율에서도 다소 차이를 보인다. Christensen et al.(2013)의 연구에서는 자기응답식 설문이 대면면접조사에 비해 높은 무응답비율(37.9% vs. 23.7%)을 보인다고 보고하였다. 무응답비율에 영향을 미치는 변인은 결혼여부, 인종적 배경, 최종 학력이었으며, 이들 변인은 두 가지 모드에서 모두 무응답비율에 영향이 있는 것으로 나타났다.

설문방식과 측정오류(measurement error) 간의 관계에 있어서는 설문조사 방식에 따라 고유의 모드효과가 존재하는 것으로 나타났다. 대면조사의 경우 면접원의 인종과 성이 응답에 영향을 미치는 것으로 나타났으며(Huddy et al. 1997; Davis & Silver 2003), 전화조사의 경우 대면조사보다 그 정도는 덜 하지만 유사한 문제점들이 발견되었다(Hornik 1982).

스마트폰의 급속한 보급과 함께 PC 또는 태블릿을 이용한 웹서베이와 모바일 조사간의 차이점에 대한 연구도 활발하게 수행되었다. PEW research center의 조사(2015)에서는 모바일 조사가 웹서베이에 비해 낮은 응답률을 보였으나 모드별로 응답자의 인구통계학적 구성에서는 차이가 없는 것으로 나타났다. Mavletova & Coupe(2013)는 모바일 서베이와 PC를 이용한 웹서베이 간에 민감한 질문에 대한 응답률이 차이가 나는지에 대해 조사하였다. 이들은 스마트폰이 다른 사람이 접근하기 더 쉽기 때문에 웹서베이에 비해 사회적으로 바람직한 응답가능성이 높을 것이라고 가정하였다. 분석 결과 다섯 가지 민감한 질문 중 음주, 소득 등 2개의 문항에서 모드간 차이를 발견하였다. 음주와 소득 모두 모바일에서 적게 응답하는 경향을 보였다. 웹서베이와 모바일조사의 품질을 비교한 Mavletova(2013)의 연구에서는 모바일의 경우 응답률이 낮고, 주관식 응답의 길이가 더 짧은 것으로 나타났다.

Smyth, Olson, & Millar(2014)는 조사방법에 대한 선호도에 영향을 미치는 변인들을 분석하였다. 이들의 연구는 2개의 기존 설문조사 결과를 분석하여 면접원조사, 우편조사, 웹서베이 등 상이한 조사 모드의 선택에 응답의 편의성과 관련된 변인(친숙도, 매체접근성, 주의분산), 신체적, 인지적 부담과 관련된 변인(시각 vs 청각, 면접조사원의 존재), 규범과 관련된 변인(면접조사원 존재), 개인적 안전과 관련된 변인(설문조사의 정당성, 개인정보보호)등의 효과에 대해 분석하였다. 분석결과는 친숙성과 매체접근성이 모드 선호에 가장 큰 영향을 미치는 것으로 나타났다. 연령과 학력, 취직여부, 인터넷 사용능력 관련 변인, 인터넷 사용자 여부 등이 웹서베이를 선호에 영향을 미쳤다. 반면 고령자일수록 우편서베이를 웹서베이에 비해 선호하는 경향을 보였으나, 여성은 면접원조사에 비해 우편조사를 선호하는 경향을 보였다.

기존 연구결과에 따르면 모드 간 차이는 존재하지만 이러한 차이는 문항이나 응답자에 따라 달라지기 때문에 일반적인 규칙성을 찾기는 어렵다. 분명한 것은 모드 차이는 조사의 편향 또는 측정오차의 증대로 나타날 가능성은 언제나 있다는 점이다.

Ⅲ. 연구문제

본 연구의 목적은 모바일 조사와 전화조사 중 어느 조사방법이 더 정확한가를 판별하는 데 있지 않다. 우리는 과연 면접원의 존재 유무가 조사결과에 현저한 영향을 주는지, 즉 두 조사방법을 통해 획득한 데이터가 서로 동등하다고 판단할 수 있는지 여부를 분석해 보고자 한다.

조사 실무자들은 면접원이 개입하지 않을 경우 설문지 해석과 응답입력 등에서 오류가 증가할 것이라고 본다. 하지만 이러한 우려가 이론적, 실증적 증거에 기반하는 것은 아니다. 현재까지의 모드효과 연구들은 비면접원조사에서 면접조사에 비해 측정오차가 증가한다는 증거를 제시하고 있지 못하다. 실무자들의 우려는 그동안 조사품질관리가 전화 및 면접원조사를 중심으로 발달되어 왔기 때문에 자기응답식 설문에 적절한 조사방법이나 그에 따른 관리방법이 정교화되지 않았던 현실에서도 이유를 찾을 수 있다. 상이한 조사방법이 완벽하게 동일한 결과를 생산할 수는 없다. 비면접원조사가 면접원조사와 측정오차에서 차이가 있을 가능성은 크다. 관건은 '얼마나' 차이가 나는가이다. 두 조사방법을 사용해 수집한 자료의 측정오차의 차이가 사실상 동등한 수준이라고 판단할 수 있을 경우, 단일한 자료를 수집하기 위해 두 가지 조사방법을 사용하는 것이 타당하다고 판단할 수 있다. 반대로 두 조사방법의 결과가 동등하지 않다고 결론을 내릴 수 있을 수준으로 큰 측정오차가 존재한다면, 비면접원조사와 면접원조사의 동시 사용은 실용적 필요성에도 불구하고 선택가능한 대안이 되지 못한다. 만약 조사결과가 동일하지 않다면, 혼합조사가 아닌 면접원조사와 비면접원조사 중 조사결과의 품질이 더 좋은 것을 선택하여 단일모드조사를 실시해야 할 것이다. 이 경우에는 두 조사방법의 조사결과간 품질 차이를 검증하는 것이 필요하다.

연구문제: 비면접원조사와 면접원조사는 동등한 결과를 생산하는가?

앞서 지적한 대로 모드효과 중 선택효과는 혼합조사 방식 도입에 큰 장애가 되지

않는다. 관건은 측정오차이다. 이 연구에서는 모드효과 중 측정오차에 주목하여 비면접원조사와 면접원조사 결과의 동등성 여부를 검증하였다.

Ⅳ. 연구방법

1. 자료수집

연구의 타당도를 높이기 위해서는 연구결과에 영향을 줄 수 있는 내적/외적 요인을 통제할 필요가 있다. 그러나 조사모드에 관한 연구의 조사 자료 수집 자체가 시간과 비용을 소요하기 때문에 이상적인 연구설계를 하기가 쉽지 않다. 특히 일반인을 대상으로 모드를 달리하여 자료를 수집하는 것은 대단히 어렵다. 본 연구에서는 국토연구원의 <2020년도 일반가구 주거실태조사 시범조사>의 일부 문항을 모드를 달리하여 실시한 실험자료를 활용하여 연구문제에 대한 답을 구하였다.

실험은 한 리서치 전문회사가 보유한 온라인 패널을 대상으로 수행되었다. 실험은 면접원을 활용한 전화조사와 자기응답식 설문형태인 모바일조사 두 가지 방식으로 구성되었으며, 동일한 응답자들을 대상으로 2가지 조사방법을 사용하여 동일한 설문을 실시하였다. 면접원조사와 비면접원조사의 차이(또는 동등성) 검증을 위해서는 대면면접조사와 웹서베이 또는 ARS방식의 전화조사 등을 비교하는 것이 최선이지만, 코로나 상황을 고려하여 비대면조사를 원칙으로 하고 전화면접조사와 모바일조사를 비교하였다. 전화면접조사의 경우에도 조사의 처음부터 끝까지 면접원이 질문하고 응답을 기입하는 방식이기 때문에 대면조사와 실질적인 차이는 존재하지 않는다.

본 연구의 관심은 전화조사와 모바일조사의 결과가 동일한가에 있다. 그러므로 이상적인 실험설계는 동일한 응답자에게 동일한 질문을 모드만 달리하여 반복 질문하는 것이다. 그러나 이럴 경우 전화조사와 모바일조사 중 어느 방법이 먼저인가에 따른 순서효과가 있을 수 있기 때문에 순서를 각기 달리하는 방법을 사용했다. 응답자를 2개의 집단으로 구분하여, 집단 A에게는 모바일조사 후 전화조사를, 집단 B에게는 전화조사를 먼저 실시하고 2차로 모바일조사를 실시하였다. 집단 배정은 무

작위로 이루어졌다.

<표 1> 실험설계 (*N*=1차 6,054, 2차 4,138)

	1차 조사		2차 조사	
집단 A	모바일	1월 6일~19일	전화 면접	1월 27일~2월 18일
집단 B	전화면접	1월 6일~26일	모바일	2월 3일~19일

설문조사는 2021년 1월에서 2021년 2월 사이에 실시되었으며, 1차 조사와 2차 조사는 응답자에 따라 1~6주의 간격이 있었다. 참가자는 만 19세 이상 가구주 또는 가구주 배우자로 제한하였다. 응답자 모집은 온라인 패널에 가입된 이들에게 이메일을 통해 설문참가를 의뢰하는 형식으로 이루어졌으며, 목표한 인원에 도달할 때까지 모집이 이루어졌다. 1차 설문에 참여한 이는 6,054명이었으며 2차 설문에는 이 중 4,138명이 참가하였다. 집단 A의 경우 총 3,037명이 1차 조사에 응하였으며, 이 중 2,045명이 2차 조사에 응하였다. 집단 B의 경우 1차 조사에는 3,017명이, 2차 조사에는 2,093명이 참여하였다. 최종 분석에는 2차 조사까지 참여한 집단 A 응답자 2,045명과 집단 B 응답자 2,093명, 총 4,138명의 자료를 사용하였다.

2. 측정문항

실험에 사용된 설문은 100개가 넘는 문항을 포함하고 있으나 비교분석에는 설문문항 중 응답률이 높아 충분한 사례 수를 확보할 수 있는 주거환경 만족도, 소득, 주택가격, 생활비, 자산, 부채 등으로 한정하였다. 분석대상 중 만족도를 제외한 나머지 항목들은 자유응답식으로 응답하도록 되어 있었다. 통상적인 설문자료 분석에서는 응답신뢰성, 불성실한 응답 등에 대한 검토를 통해 문제가 있는 자료들을 삭제하는 작업을 거치지만, 우리 연구는 무응답, 부정확한 응답 등이 연구에서 비교하고자 하는 대상이므로 수집된 설문자료 그대로 분석에 사용하였다.

<표 2> 분석에 사용된 설문문항

유형	문항 내용	비고
폐쇄형	1. 주거환경에 대한 만족도Ⅰ(시설 접근용이성) - 상업시설 접근용이성 - 의료시설 접근용이성 - 공공기관 접근용이성 - 문화시설 접근용이성 - 도시공원 및 녹지 접근용이성 - 대중교통 접근용이성 - 교육환경 2. 주거환경 만족도Ⅱ(주거환경) - 주차시설 이용편의성 - 주변도로의 보행안전성 - 치안 및 범죄 등 방범 상태 - 자동차 경적 등 집주변의 소음정도 - 청소 및 쓰레기 처리상태 - 대기오염정도 - 이웃과의 관계	• 4점 척도: 매우만족/ 대체로 만족/ 약간 불만족/ 매우 불만족 • 14개 문항을 2개 요인으로 구분하여 각 요인별로 합산하여 분석 각 요인별로 7점-28점의 분포를 가짐
자유응답식	1. 가구원 전체의 월평균 정기 소득 2. 가구의 총자산 3. 가구의 부채 4. 가구의 월평균 생활비	• 만 원 단위로 응답

3. 분석방법

1) 동등성 검증(equivalence test)

사회과학연구에서는 일반적으로 두 집단의 특정 변수값의 차이가 유의수준 밖에 존재하는가 여부를 기준으로 두 집단의 차이가 유의미한지를 판단한다. 본 연구에서와 같이 2개 집단 간 비교의 경우 가장 보편적으로 사용되는 방법은 t-test 등의 방법을 이용해 집단 간 평균값 차이가 유의미한가를 검증하는 것이다. 그러나 본 연구의 목적은 비면접원조사결과가 면접원조사결과와 동일한 결과를 가져오는가를 알아보는 것이다. 통계치간에 차이가 발생할 확률을 산출하여 차이 유무를 검증하

는 방법으로는 이러한 동등성을 평가하기가 어렵다. 확률분포상 이론적으로는 유의미한 차이가 실용적으로는 허용가능한 차이일 수도 있다. 나아가 유의성 검증에서 두 집단 간에 차이가 있다는 가설이 기각되고 영가설(零假說: null hypothesis)이 채택된다고 해도, 그 결과는 차이가 없다는 것을 입증하는 것이 아니라 연구가설을 뒷받침할 증거가 불충분하다는 것만을 알려줄 뿐이다.

따라서 본 연구는 두 집단의 차이의 크기 즉 효과크기를 구한 다음 그것의 신뢰구간을 산출하여 두 집단 간 차이가 어느 정도 있는가를 검증하는, 즉 두 집단의 평균이 동등한지 여부를 검증하는 동등성 검증방법을 이용하였다. 동등성 검증은 집단 간 차이가 없다는 것을 보여주기 위한 통계방법이다. 동등성 검증은 약제학 분야에서 여러 종류의 처방약 간의 동등성 검증을 위해 오랜 기간 사용되어 왔으며, 최근에는 심리학 분야에도 도입되고 있다(Mara & Cribbie 2012). 또한 조사방법의 차이가 대표성에 영향을 주는지 분석하는 데도 이용되고 있다(Lie et al. 2010). 본 연구의 목적은 '조사방법 간 차이가 있다'는 것이 아니라 '차이가 없다'는 것을 입증하고는 것이기 때문에 동등성 검증방법이 적합하다. 동등성 검증은 집단 간 차이가 일정한 기준 내의 값일 경우 두 집단의 값은 동등한 것으로 판단한다. 동등성 검증을 위해서는 효과크기 및 효과크기의 신뢰구간을 추정할 필요가 있다. 집단간 효과크기는 그 차이를 표준편차로 표준화시킨 코헨의 d(Cohen's d)를 이용하였다. 코헨의 d와 신뢰구간은 JASP를 이용하였고, 절삭 데이터의 경우는 R의 Walrus 패키지와 UFS 패키지를 이용하였으며, d의 해석은 Sawilowsky(2009)를 참고하였다

연구문제에 대한 답은 2단계의 과정을 거쳐 구하였다. 먼저 조사모드에 따라 측정오차의 크기에 차이가 나는지 여부를 분석하였다. 측정오차의 크기 비교는 변량값의 비교를 통해 이루어졌다. 측정오차의 크기에서의 차이가 없다고 하더라도 조사방법에 따라 기댓값의 차이가 존재할 수 있다. 두 번째 단계에서는 전화조사와 모바일조사 간 기댓값에서 각기 다른 성격의 편향이 나타나는가를 검증해 보았다. 기댓값에서의 차이는 평균비교를 통해 수행되었다.

2) 절삭된(trimmed) 평균을 이용한 비교

비면접원조사, 특히 모바일 조사는 인터페이스의 불편함과 명백하게 잘못된 응답을 현장에서 정정해 줄 수 있는 면접원의 부재로 인해 입력오류나 응답자의 오해, 오독으로 인한 이상치가 발생할 가능성이 상대적으로 높다. 이러한 이상치는 조사 후 데이터 검증과정에서 수작업을 통해 제거하는 것이 가능하다. 그러나 이상치 처

리 원칙은 조사회사마다 상이하며 명확하게 합의된 원칙이 없으므로 자의적인 판단의 가능성이 있다. 보다 객관적인 원칙이 필요하다고 판단하여 이 논문에서는 상, 하위 5%의 값을 갖는 자료를 제외하고 산출하는 평균인 절삭평균을 적용하였다. 따라서 본 연구는 5% 절삭된 평균과 그에 기반한 조정된 분산(winsorized variance)[1]을 산출하여 검토하였다. 절삭된 평균과 분산은 아래 공식을 통해 산출하였다.

$$\overline{X_t} = \frac{1}{n-2g}\left(X_{g+1} + X_{g+2} + \ldots + X_{n-g}\right) \quad (1)$$

$\overline{X_1} \ldots \overline{X_n}$: 크기 순으로 정렬된 값
g : 절삭된 관찰값
$n-2g$: 절삭(trim) 후 남아 있는 관찰 값

$$\overline{X_w} = \frac{1}{n}\left([g+1]X_{g+1} + X_{g+2} + \ldots + [g+1]X_{n-g}\right)$$

$$SSD_w = [g+1]\left[X_{g+1} - \overline{X_w}\right]^2 + \left[X_{g+2} - \overline{X_w}\right]^2 + \ldots + [g+1]\left[X_{n-g} - \overline{X_w}\right]^2$$

$$S_w^2 = \frac{SSD_w}{(n-2g-1)} \quad (2)$$

Ⅴ. 분석결과

1. 조사모드에 따른 측정오차 크기 비교

조사과정에서 발생하는 측정오차는 변량을 증대시키는 결과를 낳는다. 따라서 두 방법 간의 측정오차의 차이는 변량 비교를 통해 검증이 가능하다. 통계치의 변량을 조사방법별로 보면 <표 3>과 같다.

---

1) 조정된 변량(winsorized variance)는 절삭된 평균과 유사하게 일정한 기준을 초과하는 상·하위 극단치의 값을 차상위 극단치(기준 내의 최대치)로 대체한 변량을 의미한다.

<표 3>에서 보듯이 전화조사에 비해 모바일 조사의 변량이 크게 나타났다. 두 조사는 동일한 사람이 응답했기 때문에 변량차이의 분석은 모간-피트만(Morgan-Pitman)의 비독립표본 변량분석방법을 적용해서 수행했다. 즉 두 종속표본의 변량이 동일하다는 다음 가설을 검증하기 위해 U(두 집단 값의 합)와 V(두 집단 값 간의 차이) 간의 상관계수를 구했는데, <표 4>에서 보는 바와 같이 모두 유의미한 것으로 나타났다.

<표 3> 조사방법 간 변량 비교

변수	조사방법	N	Std. Deviation
주건환경만족도 1	모바일	4,138	4.5
	전화	4,138	4.2
주건환경만족도 2	모바일	4,138	3.9
	전화	4,138	3.7
가구소득	모바일	3,970	54,276.3
	전화	4,104	370.0
가구생활비	모바일	3,986	23,873.3
	전화	4,117	141.3
가구자산	모바일	3,155	1,644,199.6
	전화	3,720	180,308.8
가구부채	모바일	2,365	370,331.2
	전화	2,316	17,446.1

모건-피트만 분석방법의 귀무가설은 두 종속표본의 변량이 동일하다는 것이다.

$$H_0 : \delta_1^2 = \delta_1^2 \tag{3}$$

영가설의 검증은 아래와 같은 절차를 밟는다.

$$U = X_1 + X_2$$

$$V = X_1 - X_2$$

$$H_0 : \rho_{\mu\nu} = 0$$

$$T_{uv} = r_{uv}\sqrt{(n-2)/\left(1-r_{uv}^2\right)} \tag{4}$$

<표 4> 모건-피트만 비독립표본 변량분석 결과

문항	rUV	t	유의도
주거환경만족도 1	.094	6.07	$p < .001$
주거환경만족도 2	.092	5.94	$p < .001$
가구소득	1.00	NA[2)]	NA
가구생활비	1.00	NA	NA
가구자산	.970	228.74	$p < .001$
가구부채	.250	11.84	$p < .001$

응답자료의 변량이 큰 것은 응답자들의 실수로 인한 잘못된 답변이 그대로 반영되어 있기 때문이다. 변량값으로 판단할 때 비면접원조사뿐 아니라 면접원조사에서도 이러한 오류는 존재하는 것으로 보인다. 비면접원조사의 경우 만원 단위로 응답하게 되어 있는 것을 오인한 탓이거나 입력과정에서 오타로 인한 오류가 발생했을 가능성이 있다. 이는 금액을 직접 입력하게 되어 있는 문항들에서 두 자리 이상의 값 차이가 발생하는 경우가 종종 있는 것을 통해 추측할 수 있다. 평균의 수십 배에 해당하는 이상치는 실제 사회조사에서는 내검 과정에서 관리가 가능하다. 문제는 이러한 실수에 의한 이상치를 제외하고도 차이가 있는지 여부일 것이다. 이상치의 영향을 축소시키기 위해 5% 조정된 분산값을 이용하여 방법 간 차이를 다시 살펴보았다.

2) 가구소득과 가구자산의 경우 U와 V의 상관관계값이 1.0으로 완벽한 상관관계를 가지므로 t값은 계산되지 않았다.

<표 5> 조사방법 간 조정된 변량 비교

문항	모바일조사	전화조사	rUV	t	유의도
주거환경만족도 1	4.4	4.1	.10	6.07	$p < .001$
주거환경만족도 2	3.8	3.5	.12	5.94	$p < .001$
가구소득	245.8	240.5	.03	2.07	$p = .019$
가구생활비	135.4	132.4	.09	5.95	$p < .001$
가구자산	49308.0	49928.4	-.02	-.980	$p = .836$
가구부채	12383.7	12539.2	.16	6.00	$p < .001$

<표 5>에서 나타나듯 조정된 변량값을 사용할 경우 두 방법간 차이가 현저하게 감소한다. 원래 자료와 비교할 때 응답상의 이상치를 5% 제거할 경우 면접원조사에 비해 비면접원조사의 변량이 현저하게 감소한 것을 볼 수 있다. 절삭된 평균을 사용할 경우 두 조사방법을 통해 생산된 자료가 동등하다는 것은 극단치의 영향이 크다는 것을 의미한다. 이러한 결과는 자료의 정제가 충실히 이루어진다면 측정오차를 조사방법 간에 실질적 차이가 없는 수준으로 감소시키는 것이 가능하다는 것을 시사한다.

2. 측정오차가 평균값에 미치는 영향

측정오차에서 차이가 없다는 것은 두 조사방법간 정밀성(precision)에서의 차이가 없다는 것을 의미한다. 그렇다면 조사방법에 따라 기댓값에서의 체계적 차이는 어떻게 되는가? 즉, 조사방법에 따라 각기 다른 편향이 나타날 가능성은 없는가? 기존 연구에 의하면 면접원의 존재는 민감한 질문에 대한 답변에 영향을 미친다고 한다. 이를 확인하기 위해 조사방법에 따른 기댓값에서의 차이를 동등성 검증을 통해 분석하였다. 동일한 응답자가 각각 모바일조사와 전화조사에 응답한 값을 비교하는 것이므로 대응표본 t-test를 이용하여 양 방법간 평균값을 비교하였고, 코헨의 d를 구하여 동등성 검증을 실시하였다.

<표 6> 조사방법 간 평균 비교

구분	t	df	p	Mean difference	SE	Cohen's d
주거환경만족도 1	-19.4	4137	$p < .001$	-1.0	0.05	-0.30
주거환경만족도 2	-20.9	4137	$p < .001$	-1.0	0.05	-0.33
가구소득(만 원)	1.6	3945	0.103	1414.3	866.7	0.03
가구생활비(만 원)	1.1	3973	0.257	429.5	379.3	0.02
가구자산(만 원)	1.1	2945	0.257	35702.9	31523.3	0.02
가구부채(만 원)	8.0	2050	$p < .001$	1995.8	248.0	0.18

<표 6>에서 나타나듯 동등성 검증 결과, 두 조사방법 간 측정오차의 크기 차이는 허용범위 내인 것으로 나타났다. 이는 직관적인 판단과는 배치되는 것인데, 4점 만점의 등간척도를 사용한 만족도 이외의 변인들에서는 가장 차이가 큰 자산의 경우 조사방법간 기댓값에서 35,702(만원)의 차이가 있음에도 이러한 차이는 허용범위 내인 것으로 나타났다. t값을 기준으로 판단해 보아도 부채를 제외한 소득, 생활비, 자산 모두 차이가 유의미하지 않은 것으로 나타났다. 이러한 결과는 이들 응답 값 차이의 표준편차가 컸기 때문인 것으로 판단된다. 즉, 이상치들의 존재가 결과에 영향을 크게 미친 것으로 추측된다. 이러한 결과는 통계적인 유의미성 여부와 무관하게 별다른 정보를 제공하지 못한다. 절삭된 평균을 이용하여 이상치의 영향을 최소화한 후 동등성 검증을 다시 실시한 결과는 <표 7>과 같다.

<표 7> 조사방법 간 절삭된 평균 비교

구분	t	df	p	Mean difference	SE	Cohen's d
주거환경만족도 1	(18.7)	3311	$p < .001$	(1.0)	0.1	0.16
주거환경만족도 2	(20.4)	3311	$p < .001$	(1.1)	0.1	0.19
가구소득(만원)	(9.2)	3157	$p < .001$	(19.8)	2.1	0.06
가구생활비(만원)	5.7	3179	$p < .001$	8.4	1.5	0.04
가구자산(만원)	0.0	2357	1.0	13.1	367.6	0.00
가구부채(만원)	9,5	1640	$p < .001$	1493.1	157.0	0.09

비교 결과 방법 간 차이가 현저하게 감소하였다. 비록 차이에 대한 t-test 결과는 통계적으로 유의미한 차이가 있는 것으로 나타나고 있으나 효과크기가 $d < 0.5$로 나타나고 있다. 즉, 동등성 검증 결과 두 조사방법 간 문항의 응답은 절삭된 평균을 사용할 경우 동등한 것으로 나타났다. 조사방법 간 문항 값을 비교해 볼 때 절삭된 평균값의 차이는 크지 않다. 예컨대 만족도의 경우 28점 만점으로 측정된 것인데 그 차이가 1점 정도에 불과했다. 소득의 경우에는 약 20만 원의 차이가 나타난다.

<표 8> 동등성 검증 결과

구분	Cohen's d	df	95% 신뢰구간
주거환경만족도 1	0.16	3311	.09~.23
주거환경만족도 2	0.19	3311	.12~.26
가구소득	0.06	3157	-.01~.13
가구생활비	0.04	3179	-.03~.11
가구자산	0.00	2357	-.08~.08
가구부채	0.09	1640	-.01~.19

보다 체계적으로 효과크기를 살펴보기 위해 95% 신뢰수준의 신뢰구간을 살펴보았다. 위 표에서 보듯이 자유응답식 질문의 경우 절삭된 평균을 사용할 경우 두 방법 간 차이는 거의 없다고 말할 수 있다. 폐쇄형 응답 문항인 만족도의 경우 자유응답식 문항들과 비교해서는 상대적으로 차이가 큰 편이지만 코헨의 d값으로 판단할 때 의미있는 차이가 아니다.

Ⅵ. 분석결과에 대한 논의 및 결론

본 연구는 코로나19 바이러스 사태로 인해 국내 사회조사가 주로 의존했던 면접원조사를 이용한 조사수행에 어려움을 겪고 있으므로 전화조사, 웹서베이, 모바일조사 등 면접원조사 외의 다양한 방법을 도입하여 혼합조사 방식을 채택하는 것이

불가피하다는 인식을 출발점으로 삼고 있다. 혼합조사의 필요성은 비단 코로나 사태뿐 아니라 조사환경이 점차 면접원조사의 어려움을 커지는 방향으로 변화하고 있고, 단일한 조사방식으로는 충분한 응답률을 확보하기 어렵다는 이유에서도 찾을 수 있다. 코로나 사태로 말미암아 면접원조사가 힘들어지자 국가통계를 위한 사회조사에서 유치기입, 웹조사가 부분적으로 도입되고 있는 상황이다. 문제는 사실상 혼합조사가 시행되고 있음에도 불구하고, 면접원조사를 보완하기 위한 임시방편의 성격이 강하기 때문에 충분한 사전준비가 부족하여 실제 조사가 원활하게 진행되는 효과는 적은 반면 비표본오차와 표본편차의 가능성만 높이고 있다는 것이다. 혼합조사의 도입으로 인한 모드효과의 가능성 여부에 대한 점검과 모드효과를 제거하거나 최소화하기 위한 치밀한 사전준비가 필요하다.

혼합조사의 도입에서 문제가 되는 것은 과연 상이한 조사방법이 서로 동등한 품질의 조사결과를 생산할 수 있는가 여부와 만약 조사결과에 차이가 있다면 이러한 차이가 자료품질에 얼마나 큰 영향을 미치는가이다. 즉, 모드효과 중 측정에 대한 효과가 존재하는가, 존재한다면 이는 실질적으로 자료의 호환성에 문제를 미칠 만큼 심각한가 여부이다. 본 연구에서는 전화조사와 모바일조사를 비교하는 실험을 통해 분석해 보았다.

이 연구의 목적은 조사모드 간 차이가 있음을 검증하는 것이 아니라, 양자 간의 차이가 실질적인 의미를 가질만한 수준인지, 즉 서로 동등한지 여부를 검증하는 것이었다. 그러므로 여기에서는 사회과학연구에서 일반적으로 사용되는 집단간 값 차이의 통계적 유의미성을 평가하는 것이 아니라, 집단 간 값 차이가 충분히 크지 않음을 검증하는 동등성 검증방법을 사용하였다.

분석의 결과, 두 조사방법을 통해 수집한 자료의 측정오차는 원자료값을 사용할 경우 동등하지 않은 것으로 나타났다. 특히 자유응답식 문항의 경우 측정오차의 차이가 상당히 크게 나타나는 것을 알 수 있다. 일반적으로 사회조사의 검수과정에서 명백한 실수에 의한 이상치는 제거되거나 대체되며, 나아가 모바일 조사의 경우 보다 정교한 디자인을 통해 실수로 인한 이상치 발생을 상당히 감소시키는 것이 가능하다. 이러한 점을 감안하여 절삭된 평균을 이용하여 동증성 검증을 실시할 경우, 두 조사방법간 차이는 없는 것으로 나타났다. 변량비교 결과, 모바일조사의 측정오차는 비교한 6개의 측정항목 모두에서 현저하게 큰 것으로 나타났지만, 이러한 차이는 절삭된 평균을 사용할 경우 실질적인 의미는 갖지 못하는 것으로 볼 수 있다. 비면접원조사는 면접조사에 비해 상대적으로 측정오차가 크지만 이상치를 제거할

경우 측정오차는 동등한 것으로 나타났다. 절삭된 평균을 사용할 경우 두 조사방법 간 측정오차가 동등한 것으로 나타난다는 사실은, 면접원조사와 비면접원조사 간에 측정오차가 존재하지만 설문설계와 내검과정을 통해 이러한 모드 간 차이를 용인가능한 범위 내로 축소하는 것이 가능할 것임을 시사한다. 측정오차가 동등하다는 것은 조사의 정밀성 수준이 동등하다는 것을 의미한다. 즉 조사의 오차 수준을 유지하기 위해 표본을 증가시킬 필요는 없다는 것이다. 비면접원조사에서 나타난 이상치의 원인으로 추측되는 단위 착오, 타이핑 실수로 인한 입력 오류 등은 더 정교화하게 설계된, 모바일조사에 최적화된 설문을 통해 대폭 줄일 수 있을 것이다. 나아가 비면접원조사의 반복적인 실시 과정에서 지속적으로 설계를 개선해나간다면 측정오차를 상당히 축소하는 것이 가능할 것이다. 이러한 결과는 비면접원조사와 면접원조사를 동시에 사용하는 혼합조사가 충분히 가능할 것임을 시사한다.

측정오차의 크기 차이와 독립적으로 조사방법에 따라 기댓값에서 차이가 발생할 수도 있다. 면접원의 존재가 민감한 질문에 대한 답변에 영향을 미치거나, 응답자들로 하여금 사회적으로 바람직한 응답을 보다 많이 하도록 할 수도 있다. 평균비교를 통한 검증 결과, 원자료를 이용한 비교는 양자 간의 차이가 허용범위 내인 것으로 나타나 조사방법 간에 동등한 것으로 나타났다. 그러나, 이러한 결과는 이상치의 영향인 것으로 판단되어 이상치의 영향을 최소화한 절삭된 평균을 이용한 검증을 다시 실시하였다. 분석결과, 조사방법간 기댓값 차이의 절대적인 크기도 현저히 감소하며 허용범위 내의 값을 갖는 것으로 나타났다. 즉, 조사방법에 따라 체계적인 응답 차이가 존재하지는 않는 것으로 드러났다.

결론적으로 측정오차 크기의 경우 원자료를 사용할 경우 두 조사방법이 동등한 결과를 생산하지 않으며, 이상치에 대한 처리가 이루어진 후에야 동등해진다. 응답의 편향성의 경우 두 조사방법간 차이는 없는 것을 나타났다.

이 연구는 여러 가지 한계를 갖는다. 무엇보다 분석에 사용한 자료가 이 연구를 위해 설계된 실험으로부터 획득한 것이 아니었기 때문에 본 논문의 연구문제에 대한 답을 구하기 위해 최적화된 자료가 아니라는 한계를 갖는다. 이로 인해 연구문제를 가용한 자료에서 해답을 구할 수 있는 것, 즉 면접원 유무에 따른 오차의 크기와 편향성으로 한정할 수밖에 없었다. 하지만 조사의 신뢰성을 결정하는 요인은 면접원 유무 외에도, 표본추출틀과 응답률 등 다양하게 존재한다. 측정오차외의 요인에 있어서 모드효과는 향후 연구에서 탐구해야 할 주제일 것이다.

이 연구는 면접원조사의 한 유형인 전화조사와 비면접원조사인 모바일조사를 비

교하였다. 즉, 매체의 차이가 있다. 전화조사와 모바일조사는 각각 청각적, 시각적 자극에 의존한다는 차이가 존재한다. 하지만 현실적으로 매체를 전화로 통일하여 면접원조사와 비면접원조사인 ARS방식을 사용하는 것은 후자의 응답률 및 중도탈락률이 높기 때문에 선택가능한 대안은 아니다. 동일한 설문이 1주~6주라는 기간을 두고 2차례에 걸쳐 질문되었기 때문에 응답자들이 설문에 익숙해졌을 가능성 또한 배제할 수는 없다. 하지만 이는 정확한 모드효과 검증를 위해서는 감수해야 할 한계일 수밖에 없다. 하지만 이러한 영향은 무선적으로 발생할 것이므로 연구결과에 의미있는 차이를 가져왔을 것이라고는 판단되지는 않는다.

분석의 결과, 전화조사와 모바일조사 간 가장 큰 차이를 보인 문항은 소득, 자산, 부채 등 응답자가 민감하게 반응할 가능성이 높은 것이었다. 이는 의도적인 설계의 결과가 아닌 실험에 사용된 개방형 설문의 다수가 위와 같은 사항에 대한 질문이었기 때문에 나온 결과이다. 그럼에도 불구하고, 면접원의 존재가 응답에 영향을 미쳤을 가능성을 완전히 배제할 수는 없다. 상이한 조사방법에 따라 응답률에서의 차이가 발생할 가능성도 있다. 그럼에도 불구하고 면접원 유무에 따른 차이가 크지 않았다는 것은 고무적인 결과이다. 본 연구는 일반인을 대상으로 한 무선표집이 아니라 조사전문회사가 보유한 온라인 패널을 대상으로 조사하였기 때문에 조사방법에 따른 응답률 차이를 분석하기에는 적절치 않았다. 본 연구는 상이한 조사방법만을 비교하였다.

한국의 과학적 사회조사가 면접원조사에 주로 의존해 왔었기 때문에 상대적으로 비면접원조사를 위한 설문설계, 품질관리 기법의 발달이 상대적으로 뒤처진 현실이다. 과학적 사회조사에 혼합조사를 도입하는 과정에서 새로운 조사방법 도입으로 인한 일시적 시행착오를 피하기란 어려울 것이다. 코로나 사태와 조사환경의 변화는 비면접원조사의 도입을 불가피한 선택으로 만들고 있다. 모드효과에 대한 근거 없이 막연한 불안감으로 인한 주저나 지나치게 방법상의 차이에 둔감한 것 모두가 정확한 조사자료를 얻는 데 도움이 되지 않는다. 이 연구의 결과는 현재 수준의 비면접원조사 기법을 사용한 사회조사도 면접원조사와 충분히 동등한 결과를 생산할 수 있다는 것을 보여준다. 관건은 새로운 조사방법을 도입할 때의 시행착오와 부작용을 최소화하는 실천적 방안을 마련하는 것이다.

한편, 원자료를 사용한 동등성 검증에서 조사방법 간에 측정오차와 기댓값의 차이가 존재하는 것으로 나타난 것은 무조건적인 동등성 가정은 위험하다는 것을 시사한다. 본 연구의 결과는 비면접원조사의 도입에 있어 긍정적인 면과 부정적인 함

의를 동시에 갖고 있다. 원자료를 이용한 동등성 검증 결과는, 응답자의 실수를 최소화할 수 있는 모바일 설문설계와 데이터 품질을 확보할 수 있는 철저한 검수과정이 없는, 준비되지 않은 비면접원조사의 도입은 측정오차의 증가와 응답에서의 편향 문제를 야기할 수 있다는 것을 보여준다. 다른 한편으로는 적절한 노력과 방법이 개발된다면 면접조사와 동등한 수준의 정확도와 정밀성을 가질 수 있다는 점을 시사한다. 본 연구의 결과를 단순하게 비대면 조사의 무조건적인 수용이나 배척으로 해석하는 것은 적절치 않다. 이 연구의 결과는 과학적 사회조사에 면접원조사와 비면접원조사의 병행을 통한 혼합조사의 도입이 충분히 가능하지만, 이는 반드시 응답자의 실수를 최소화하고 이상치에 대한 철저한 관리방안이 수립된 후에 이루어져야 한다는 것을 시사한다.

참고문헌

백지선·심규호·홍영택·장덕현. 2017. “혼합조사의 모드효과 추정: 선택효과와 측정효과의 분리방법.” ≪통계연구≫ 22(3): 1-24.

이찬복. 2020. “코로나19 시대의 사회조사 방법론.” ≪이슈와 동향≫ 34: 34-47.

Bates, N. and J. Zamadics. 2021. “COVID-19 Infection Rates and Propensity to Self-Respond in the 2020 U.S. Decennial Census.” *Survey Practice* 14(1). doi: https://doi.org/10.29115/SP-2021-0002

Burkill, S., A.Copas, M.P. Couper, S. Clifton, P. Prah, J. Datta, F. Conrad, K. Wellings, A.M. Johnson, and B. Erens. 2016. “Using the Web to Collect Data on Sensitive Behaviours: A Study Looking at Mode Effects on the British National Survey of Sexual Attitudes and Lifestyles.” *PLoS ONE* 11(2). doi:10.1371/journal.pone.0147983

Christensen, A.I., O. Ekholm, C. Glümer, and K. Juel. 2013. “Effect of Survey Mode on Response Patterns: Comparison of Face-to-face and Self-administered Modes in Health Surveys.” *European Journal of Public Health* 24(2): 327-332.

Davis, D.W. and B.D. Silver. 2003. “Stereotype Threat and Race of Inter-viewer Eects in a Survey on Political Knowledge.” *American Journal of Political Science* 47(1): 33-45.

Dillman, D.A. and L.M. Christian. 2003. “Survey Mode as Source of Instability in Responses across Surveys.” *Field Methods* 17(1): 30-52.

Dillman, D.A. 2017. “The Promise and Challenge of Pushing Respondents to the Web in Mixed-mode Surveys.” *Survey Methodology* 43(1): 3-30.

Dillman, D.A. 2021. Adapting Push-to-Web Survey Designs for Use in Different Countries and Situations. Webinar for: The World Association for Public Opinion Research (WAPOR). 2020년 3월 5일.

Dutwin, D. and P. Lavrakas. 2016. “Trends in Telephone Outcomes, 2008–2015.” *Survey Practice* 9(3).
doi: 10.29115/SP-2016-0017

Hornik, J. 1982. “Impact of Pre-call Request form and Gender Interaction on Response to a Mail Survey.” *Journal of Marketing Research* 19(1): 144-151.

Huddy, L., J. Billig, J. Bracciodieta, L. Hoeffler, P.J. Moynihan, and P. Pugliani. 1997. “The Effect of Interviewer Gender on the Survey Response.” *Political Behavior* 19(3): 197-220.

Konitzer, T., S. Eckmany, and D. Rothschildz. 2016. “Mobile as Survey Mode.” Proceedings of the Survey Research Methods Section, American Statistical Association.

Leggett, C.G. 2003. “Social Desirability Bias in Contingent Valuation Surveys Administered through In-person Interviews.” *Land Economics* 79(4): 561-575.

Liu, H. D. Cella, R.Gershon, J. Shen, L.S. Morales, W. Riley, and R.D. Hays. 2010. “Representativeness of the PROMIS Internet Panel.” *Journal of Clinical Epidemiology* 63(11): 1169-1178.

Maguire, K.B. 2009. “Does Mode Matter? A Comparison of Telephone, Mail, and In-person Treatments in Contingent Valuation Surveys.” *Journal of Environmental Management* 90(1): 3528-3533.

Mara, C.A. and A.A. Cribbie. 2012. “Paired-Samples Tests of Equivalence.” *Communications in Statistics-Simulation and Computation* 41(10): 1928-1943.
doi: http://dx.doi.org/10.1080/03610918.2011.626545

Marta-Pedroso, C., H. Freitas, and T. Domingos. 2007. “Testing for the Survey Mode Effect on Contingent Valuation Data Quality: A Case Study of Web Based Versus In-person Interviews.” *Ecological Economics* 62(3-4): 388-398.

Mavletova, A. and M.P. Couper. 2013. “Sensitive Topics in PC Web and Mobile Web Surveys: Is There a Dierence?” *Survey Research Methods* 7(3): 191-205.

Mavletova, A. and M.P. Couper. 2015. “A Meta-Analysis of Breakoff Rates in Mobile Web Surveys.” in Toninelli, D., R. Pinter, and P. de Pedraza(eds.). *Mobile Research*

Methods: Opportunities and Challenges of Mobile Research Methodologies. 81-98. London: Ubiquity Press.
doi: http://dx.doi.org/10.5334/bar.f.

Pew Research Center. 2012. *Assessing the Representativeness of Public Opinion Surveys*. http://www.people-press.org/2012/05/15/assessing-the-representativeness-of-public-opinion-surveys/

Sawilowsky, S.S. 2009. "New Effect Size Rules of Thumb." *Journal of Modern Applied Statistical Methods* 8(2): 597-599.

Smyth, J.D., K. Olson, and M.M. Millar. 2014. "Identifying Predictors of Survey Mode Preference." *Social Science Research* 48: 135-144.
doi: 10.1016/j.ssresearch.2014.06.002

Wilcox, R. 2015. "Comparing the Variances of Two Dependent Variables." *Journal of Statistical Distributions and Applications* 2(7).
doi: https://doi.org/10.1186/s40488-015-0030-z

UN Department of Economic and Social Affairs. 2021. "Monitoring the state of statistical operations under the COVID-19 Pandemic." 2020년 6월 5일.
https://unstats.un.org/unsd/covid19-response/covid19-nso-survey-report.pdf

UN Department of Economic and Social Affairs. 2021. "Monitoring the state of statistical operations under the COVID-19 Pandemic." 2020년 12월.
https://covid-19-response.unstatshub.org/survey/covid-19-nso-survey-report-3.pdf

<접수 2021.05.12; 수정 2021.06.15; 게재확정 2021.06.28>

Is Untact Survey as Reliable as Personal Interview Survey?: Overcoming the Obstacles to Scientific Social Survey in the COVID-19 Era

Sung Kyum Cho
(Chungnam National University)
Sungjoong Kim
(Chungnam National University)
Mina Kang
(Korea Research Institute for Human Settlements)

Personal interview survey is becoming increasingly harder to carry out due to changing survey environment and COVID-19 pandemic. Adoption of multimodal survey in scientific social surveys is necessary. Mode effect may be present when different survey methods are simultaneously used to construct a single data set. The current study investigated the presence and the magnitude of one type of mode effect, measurement effect, through an experiment. The experiment conducted telephone interview surveys and mobile web surveys, which used the same questionnaires, on two groups of volunteers, alternating the order of the two survey methods. Equivalence tests were carried out to investigate if telephone interview survey and mobile survey yield equivalent outcomes. The results showed significant differences between outcomes of the two survey methods when the raw data was used. When trimmed means were used controlling effects of the outliers, there was no meaningful difference in measurement errors and expected values between two survey methods. The findings imply that while the mobile web survey method may increase measurement error or yield biased values, it is possible to control them. With a careful planning and preparation, adoption of multimodal survey method in a scientific social survey is feasible.

Key words: multimodal survey method, mode effect, measurement effect, equivalence test, social survey