

## 부패와 사회복지정책에 관한 선행 실험연구를 통해 살펴본 실험연구 시 주의해야 할 다섯 가지 사항들

장한일\*

박일주\*\*

---

우리는 한국 사회과학계에서의 실험연구 활성화에 기여하고자 다음의 두 가지를 이 논문에서 수행한다. 첫째, 한국 연구자들이 관심가질 것이라 생각되는 부패와 사회복지 정책에 관한 실험 논문 10편을 소개한다. 둘째, 소개된 논문들을 활용하여, 실험연구 시 연구자들이 주의해야 할 다섯 가지 사항들에 대해서 논의한다. 이 사항들은 (1) 관찰대상의 평균적 특질들이 실험조건들 사이에서 차이가 없음을 경험적으로 확인하는 균형검사의 중요성, (2) 관찰대상에 대해서 의도된 실험적 조작이 이루어지지 못할 때 발생하는 비순응 문제, (3) 실험적 조작이 의도된 관찰대상이 아닌 다른 대상으로 확산되는 일명 확산효과의 문제, (4) 관찰대상이 연구자가 의도한 방식으로 실험적 조작을 받아들이는지 확인하는 조작검사의 중요성, 그리고 (5) 실험적 조작에 의해 영향 받았을 수 있는 변수가 통제변수로 사용될 때의 위험성 등이다.

주제어: 실험연구, 균형검사, 비순응 문제, 확산효과, 조작검사

---

\* 교신저자 겸 제1저자, 뉴욕대학교(New York University)에서 정치학 박사학위를 취득하고, 현재 국민대학교 정치외교학과 교수로 있다. 주요 관심분야는 선거, 집단갈등, 이민 등이다 (hanilchang@kookmin.ac.kr).

\*\* 공동저자, 고려대학교에서 행정학 박사학위를 취득하고, 현재 한국과학기술기획평가원 부연구위원으로 재직중이다. 주요 관심분야는 창업정책, 혁신정책, 정책분석 등이다(ijpark@kistep.re.kr).

## 1. 들어가는 말

다양한 사회과학 연구 목적 중 하나인 이론 구축을 위하여 연구자들은 가설을 수립하고 이 가설을 정성적(qualitative) 방법이나 정량적(quantitative) 방법을 통하여 검증한다. 이때 정량적 방법은 관찰대상이 통제조건과 조작조건에 무작위로 배정되었는지에 따라서 관찰(observational) 방법과 실험(experimental) 방법으로 구분된다. 전자의 방법에서는 관찰대상이 어느 조건에 배정될 것인가가 그 대상의 내생적 특질에 의해서 영향을 받을 수 있지만, 후자의 방법에서는 이러한 가능성이 무작위적 배정으로 인하여 배제된다. 예를 들어 어떤 연구자가 시민들의 집회 참가 경험이 그들의 정치 이데올로기에 영향을 끼치는지에 대해서 연구한다고 가정해보자. 이때 관찰 방법은 시민들의 내생적 특질(예: 정치 효능감, 연령, 거주지)의 영향력에 따라 나누어진 두 부류의 시민들 - 집회 불참자와 집회 참가자 - 의 정치 이데올로기를 비교하는 방식이다. 이와 달리 실험 방법은 집회 참가 여부가 시민 개개인의 내생적 특질들과 상관없이 무작위로 결정된 상황에서 두 부류의 시민들의 정치 이데올로기를 비교하는 방식이다.

이 두 연구 방법 중 실험 방법을 활용한 연구가 최근 빠른 속도로 증가하고 있는 현상은 주로 다음과 같은 두 가지 이유 때문이다.<sup>1)</sup> 첫째, 실험 방법은 독립변수의 영향력을 측정할 때 비편향성(unbiasedness)을 확보하기에 관찰 방법 대비 용이하다. 관찰 방법의 경우 관찰대상의 내생적 특질들이 통제집단과 조작집단에서 발견되는 종속변수 상의 차이에 영향을 줄 수 있는데, 이때 연구자가 영향력을 발휘하는 모든 특질들을 완벽하게 인지하고 그 영향력의 정도를 측정하는 것은 거의 불가능에 가깝다. 그런데 실험 방법은 충분한 수의 관찰대상을 그것들의 내생적 특질들과 무관하게 통제조건과 조작조건에 무작위로 배정하고, 이 과정에서 두 조건에 배정된 관찰대상들의 내생적 특질들을 평균적으로 동일하게 만든다. 달리 표현하면, 실험 방법은 관찰대상의 내생적 특질들이 가질 수 있는 영향력을 실험적 조작 이전 실험조건 수준에서 통제한다.

둘째, 실험 방법은 인과관계를 주장할 때 관찰 방법보다 용이하다. 먼저 관찰 방법

1) 물론 실험 방법의 한계 역시 존재한다. 예를 들어, 연구자에 의해서 인위적으로 만들어진 실험적 조작은 종종 현실의 모사에 불과하고, 이 경우 그 조작이 모사하고자 했던 현실 속 사건의 실제적 효과를 측정하는 데에는 한계적일 수밖에 없다. 그리고 이러한 한계점들의 존재로 인해서 여전히 관찰 연구 방법은 실험 방법과 상호 보완적으로 광범위하게 활용되고 있다.

의 경우 통제조건 혹은 조작조건에 이미 배정되어 있는 관찰대상들을 대상으로 종속변수의 변수값에 관한 정보를 확보하는 연구 방법이므로, 관찰대상들이 특정 조건에 배정되었던 시점이 종속변수 상의 변수값이 결정되었던 시점보다 선행했는지 아닌지를 객관적으로 파악하는 것이 종종 어렵다. 위에서 언급한 예를 활용하여 이야기하자면, 설사 집회 참여자들과 비참여자들 사이에서 정치 이데올로기적 차이가 관찰된다고 하더라도 집회 참여 유무가 정치 이데올로기적 차이에 선행했는지 아닌지를 구분하기 쉽지 않다. 이에 반해서 실험 방법의 경우, 무작위 배정을 통해서 평균적으로 동일한 특질을 갖게 된 통제조건과 조작조건의 관찰 대상들 중 조작조건의 관찰 대상에게만 실험적 조작이 가해진 후 통제조건과 조작조건 사이에서 발생하는 종속변수상의 차이가 발생했는지를 확인하는 방식이므로, 실험적 조작이 그 차이의 발생에 선행한다는 점이 상대적으로 뚜렷하다. 다시 한 번 위에서 사용한 예를 생각해 보면, 실험연구자는 관찰대상들 일부를 집회에 참가시킨 이후 모든 관찰 대상자들의 정치 이데올로기를 측정할 것이므로, 집회 참여자들과 비참여자들 사이에서 관찰된 정치 이데올로기의 차이는 집회 참여 유무의 결과물이라고 주장하는 데 어려움이 없다.

해외에서는 실험방법이 광범위한 연구주제들에 적용되어 연구되고 있으나, 한국 사회과학계에서는 그렇지 못하다. 이러한 상황에서 이 논문은 한국에서의 실험연구 활성화를 위하여 해외에서의 주요한 연구 성과물들을 한국 사회과학 연구자들에게 소개하고자 한다. 다양한 주제들 중 지면 제약 때문에 현재 한국 사회과학자들이 관심을 가질 것으로 생각되는 두 가지 연구 주제 - 부패와 사회복지정책 - 만을 다루며, 우리는 다음과 같은 기준으로 논문들을 선택한다. 첫째, 사회과학 각 분야 최상위급 저널에서 출판된 논문들을 우선했다. 둘째, 우리는 각 연구주제 안에서 실험적 조작, 종속변수, 실험이 시행된 국가, 그리고 실험 방식 측면에서 최대한의 다양성을 추구한다. 따라서 이 글은 일반적으로 한 개의 독립변수나 종속변수에 초점을 맞춰 선행연구들을 검토하는 논문들과는 구별된다.

먼저 부패와 관련하여 다음 장에서 소개될 논문들은 (1) 마을 도로 건설 사업과 관련하여 감사율을 높이는 방식과 주민 참여 감시 방식이 건설 사업비 착복에 끼치는 영향에 관해서 인도네시아에서 시행된 실험(Olken, 2007), (2) 뇌물의 양과 뇌물을 제안 받은 사람의 임금이 뇌물 수수 여부에 끼치는 영향을 알아보기 위하여 부르키나 파소(Burkina Faso)에서 시행된 실험(Armantier and Boly, 2011), (3) 교통경찰들이 어떤 운전자를 대상으로 뇌물을 요구하는지에 관해서 말라위(Malawi)에서 시행된 실험(Robinson and Seim, 2018), (4) 현직 정치인의 부패 관련 정보가 유권자들의 투표 행위에 끼치는 영향을 알기 위하여 멕시코에서 시행된 실험(Chong et al., 2015), 그

리고 (5) 뇌물제공자에게는 법적 책임을 묻지 않고 오직 수수자에게만 그 책임을 묻는 방식이 뇌물 거래에 끼치는 영향에 대해서 인도에서 시행된 실험(Abbink et al., 2014)에 관한 논문들이다.

다음으로, 사회복지 정책과 관련하여 우리는 (1) 의료보험 미가입 빈곤층에게 의료 보험을 제공할 때 발생하는 응급실 사용의 변화를 분석한 미국 오레곤(Oregon) 주에서의 실험(Taubman et al., 2014), (2) 전일제(full-time) 고용과 관련된 재정적 동인을 복지 수당을 받는 사람들에게 제공했을 때 그들의 고용 상태에 끼치는 영향에 대해서 캐나다에서 실시된 실험(Michalopoulos et al., 2005), (3) 빈곤층 여성들의 경제활동 지원이 이들의 빈곤에 끼치는 영향에 관해서 방글라데시에서 시행된 실험(Bandiera et al., 2017), (4) 빈곤층 경제적 지원이 아동들의 교육에 끼치는 영향에 대한 멕시코에서의 실험(Dubois et al., 2012), 그리고 (5) 빈곤층 경제적 지원이 단체장의 재선에 끼치는 영향에 관한 필리핀에서의 실험(Labonne, 2013)에 관한 논문들을 소개한다.

이후 우리는 5장에서 연구자들이 실험을 계획할 때 주의해야 할 많은 사항들 중 다섯 가지 사항 - 균형검사(balance test), 비순응(noncompliance) 문제, 확산효과(spillover effect), 조작검사(manipulation check), 실험적 조작에 의해서 영향받았을 수 있는 변수를 통제변수로 활용할 때 발생하는 문제 - 에 대해서 깊이 있게 논의할 것이다. 이 과정에서 우리는 각 주의사항에 대한 독자들의 이해를 돕기 위하여 앞서 언급한 10편의 논문들을 예로 사용할 것이다. 우리가 논의할 주의사항 각각은 이미 다른 연구자들에 의해서 깊이 있게 논의된 바 있으나, 우리의 논문처럼 특정한 주제의 논문들 몇 편을 활용하여 복수의 주의사항들을 함께 설명하는 방식의 논문은 출판된 바 없는 것으로 알고 있다. 우리는 이러한 방식이 실험연구에 익숙하지 않은 연구자들에게 각각의 주의사항을 설명하는 데 있어 효과적일 것이라고 기대한다. 그러나 언급된 10편의 논문만을 활용하기에 '실험자 효과(experimenter effects)'와 같이 이 논문들과 관련성이 적은 사항들의 경우 논의하지 않는다는 한계가 있음을 밝혀둔다.

## II. 다양한 실험 방식들

앞서 언급한 10편의 논문을 소개하기에 앞서, 실험연구에 익숙하지 않은 독자들을 위하여 이번 장에서는 몇 가지 실험연구 방식들을 간략히 소개한다.<sup>2)</sup> 먼저, 여론조사 실험(survey experiment)은 데이터를 모으기 위해서 관찰 방법이 흔히 사용하는 여

론조사와 흡사하다. 다만 여론조사실험의 경우, 여론조사 문항에 대한 응답과 무관하게 무작위로 선별된 일부 참가자들에게만 실험적 조작(예: 문항, 선택지, 사진, 음성 등)을 노출시킨다는 점에서 통상적인 여론조사와 다르다. 그럼에도 불구하고 여론조사실험은 다른 실험 방식들에 비해 실험적 조작을 위하여 실험 설계상 요구되는 추가 노력이 크지 않다는 점 때문에 최근 온라인 여론조사 방식과 결합하여 많은 연구자들에 의해서 활발하게 활용되고 있다.

두 번째, 실험실 실험(laboratory experiment)은 연구자가 통제조건과 조작조건을 실험실 안에 만들고 - 컴퓨터 소프트웨어를 이용하여 실험실 내의 컴퓨터 상에서 가상적으로 만들거나 혹은 컴퓨터 소프트웨어의 사용 없이 실제로 만들고 - 실험 참가자들을 그 조건들 속으로 무작위로 배정하는 방식이다. 이 실험 방식을 통해서 연구자는 실험적 상황들에 대한 높은 통제력을 발휘할 수 있고 이로 인해서 높은 내적 타당성(internal validity)을 기대할 수 있다. 그러나 실험실이 일반적으로 대학교 내부에 위치한다는 점 때문에 대학생들이 실험 참가자로 모집되는 경우가 많으며, 이러한 경우 일반 시민들의 선호나 행위 등을 설명하는 데 한계가 있다.<sup>3)</sup>

다음으로 현장 실험(field experiment)의 경우, 실험적 조작이 실험실이 아니라 참가자들의 일상적 생활공간에서 발생하고 이로 인해서 참가자들이 종종 자신이 실험연구의 관찰대상이라는 점을 인식하지 못할 수도 있다는 점에서 실험실 실험과 구분된다. 그리고 현장 실험의 현장성은 이 실험 방법의 장단점에도 영향을 끼친다. 먼저, 연구자는 관찰대상의 생활공간(즉, 실험환경)을 완벽히 통제할 수 없고 이로 인해서 종종 내적 타당성을 확보하는 데 어려움을 겪는다. 그러나 표본 대표성의 문제와 관련하여서는 실험실 실험에 비해서 상대적으로 대표성을 확보하기 용이한데, 이는 실험실 실험이 갖고 있는 참가자들의 자기 선택(self-selection)의 문제가 덜 하거나 부재하기 때문이다.<sup>4)</sup>

- 
- 2) 좀 더 자세한 설명은 Morton and William(2010)과 안도경(2014)을 참고하길 바란다.
- 3) 대학생 표본과 일반 시민 표본 사이의 차이가 실재하는지에 관한 연구들이 이미 존재하며, 그 결과는 혼재되어 있다. 예를 들어 Cappelen et al.(2015)는 두 개의 대학생 표본과 국가 수준에서의 대표성을 가진 성인 표본을 대상으로 게임(dictator game)과 신뢰 게임(trust game)을 실시했고, 이를 통하여 성인 표본의 선택이 대학생 표본의 선택과 근본적으로 차이가 있음을 보인다. 이에 반해서 Kam(2005)은 정보 인식과 관련된 두 개의 변수들에 대한 반응이 학생 표본과 American National Election Study의 성인 표본들 사이에서 어떻게 다른지를 비교했고, 반응의 분산 양태가 두 표본 사이에서 매우 유사함을 보고한다. 그리고 Falk et al.(2013)의 경우 신뢰 게임에서의 선택 방식과 관련하여서는 학생 표본과 국가 수준에서의 대표성을 가진 성인 표본 사이에 큰 차이가 없는데 반해서, 성인 표본이 더 많은 보상적 선택을 한다는 것을 발견했다. 자세한 논의는 Druckman and Kam(2011)과 Falk and Heckman(2009)을 보길 바란다.

이밖에도 실험실 실험과 현장 실험을 혼합한 ‘현장 속의 실험실 실험(laboratory in the field experiment)’이라는 방식이 활용되곤 한다. 이는 연구자들이 대학 외부에 실험실을 설치함으로써 대학생뿐만 아니라 다양한 배경을 가진 사람들을 피실험자로 모집할 수 있다는 장점뿐만 아니라, 실험이 실험실 내부에서 이루어진다는 점 때문에 연구자가 실험 환경에 대한 높은 통제력을 발휘할 수 있기 때문이다. 그리고 이상의 실험 방식들과 달리, 관찰대상이 자연발생적으로 만들어진 통제조건과 조작조건에 무작위로 배정되는 상황을 ‘발굴’해서 연구하는 자연 실험(natural experiment)도 존재한다.<sup>5)</sup>

### III. 부패에 관한 실험연구들

이번 장과 다음 장에서는 부패와 사회복지 정책에 관한 실험연구 다섯 편씩을 소개한다. 선택된 논문들을 소개할 때 연구가 이루어진 시·공간적 배경이 무엇인지, 어떠한 실험적 조작이 가해졌는지, 그리고 그 실험적 결과의 효과가 무엇인지에 초점을 둔다.

#### 1. Olken(2007)의 현장 실험

일반적으로 부패는 직접적으로 관찰·측정하기 어렵기 때문에 무엇이 부패를 조장하는지 그리고 어떻게 부패를 줄일 수 있는지 등에 대한 과학적 연구가 어렵다.<sup>6)</sup> 그렇지만 실험연구는 부패를 측정할 수 있는 방법을 실험 계획 단계에서부터 준비함으로써 이러한 한계를 극복하곤 하며 우리가 첫 번째로 소개할 Olken(2007)의 현장 실험연구가 바로 이 경우에 해당한다.

이 실험연구가 시행될 2003년도 당시 인도네시아 정부는 World Bank로부터 지원

---

4) 실험실 실험 관련 ‘자기 선택’ 문제에 관해서는 Cleave et al.(2013)과 Harrison et al.(2009)을 보길 바란다.

5) 종종 연구자들은 통제조건과 조작조건으로 관찰대상이 무작위로 배정되었는지는 상관없이, 연구 대상의 예외성, 혹은 특정 시점을 전후로 급격한 변화를 강조하기 위하여 자연실험이라는 표현을 사용한다(예: Fortin et al., 2004; Powell et al., 2017). 현장 실험과 자연 실험의 차이에 대한 자세한 논의는 Gerber and Green(2008)을 보길 바란다.

6) 이러한 이유 때문에 Fisman and Wei(2004)와 Bandiera et al.(2009)의 경우 간접적인 자료를 통해서 부패에 대해서 연구했다.

받은 돈을 이용하여 해마다 대략 15,000여 촌락들의 소규모 경제활동 및 사회간접자본개발을 지원하는 'Kecamatan (subdistrict) Development Project'를 시행하고 있었다. 그런데 문제는 인도네시아 정부가 자신들이 지급한 지원금이 실제로 그 목적에 맞게 잘 사용되고 있는지를 실효성 있게 감독할 방안을 갖고 있지 못했다는 점이며, 이로 인해서 촌장이나 촌락의 집행부가 지원금을 착복할 가능성이 존재했었다.

이러한 상황 속에서 Olken(2007)은 동 자바(East Java)와 중앙 자바(Central Java)지역에서 도로사업 지원금을 받게 된 608개의 촌락들을 대상으로 현장 실험을 집행했다. 이들 촌락들은 6개의 실험 조건들 가운데 하나에 무작위로 배정되었는데, 이 6개의 실험 조건은 지원금 사용에 관한 정부 감사를 받을 확률이 4%인지 100%인지에 따라서 3개씩의 실험 조건으로 1차적으로 구분되었다.<sup>7)</sup> 이 3개의 실험 조건들은 다시 지원금 사용 내역이 정기적으로 보고되는 촌락회의에 촌락민들이 초대되지 않는 조건, 초대되는 조건, 그리고 초대된 촌락민들이 추가적으로 자기 의견을 익명으로 피력할 수 있는 조건으로 구분되었다.

부패의 정도를 측정하기 위하여 Olken(2007)은 각별한 노력을 기울였다. 먼저 그는 각 건설 사업에 사용된 건축자재의 양과 구매가격 그리고 건설 노동자들의 임금을 자체적으로 고용한 조사팀과 함께 직접 조사한 후 이에 기반하여 총사업비용을 추정했다. 그 후 부패의 정도를 보고된 최종 경비 대비 이 경비와 Olken(2007)의 조사팀이 자체적으로 계산한 비용 사이의 차이의 비율로 측정했다.<sup>8)</sup>

3개의 실험적 조작 - 정부 감사, 촌락민들의 촌락회의 참여, 촌락민들의 의견 개진 - 의 부패감소 효과 여부는 다음과 같이 드러났다. 첫째, 정부 감사의 확률이 4%에서 100%로 증가함에 따라 부패의 정도는 27.7%에서 19.2%로 감소했으며, 이 변화는 통계적으로 유의미했다. 둘째, 촌락민들이 사업지원금 감시에 직접 참여하는 것은 착복된 공사비 추정에 사용된 3개의 항목 중 노동 임금 착복 감소에만 영향을 끼칠 뿐, 다른 항목들에 대해서나 3개 항목의 총량에 관해서는 유의미한 효과가 없었다.

7) Olken(2007)에 따르면 대부분의 경우에서 정부 감사는 단순히 절차적인 것으로서 부패행위 자체를 입증하는데 사용하기에는 부족했다.

8) Olken(2007)과 마찬가지로 Bertrand et al.(2007) 역시 부패의 정도를 추론적으로 측정하고 있다.

## 2. Armantier and Boly(2011)의 현장 실험

우리가 소개할 두 번째 연구는 제안되는 뇌물의 양과 뇌물을 제안받은 사람의 임금 이 뇌물 수수 여부에 끼치는 영향을 알아보기 위하여 부르키나 파소(Burkina Faso)에서 실시한 연구이다. 실험은 상급 교육기관 입학 을 위한 국가시험 기간 동안에 진행되었으며, 이 기간 동안 많은 공립·사립학교들이 외부에서 채점자들을 모집해서 시험 답안지를 채점한다는 점을 활용했다. 좀 더 구체적으로 말하자면, 연구자들은 자체적으로 준비한 거짓 시험 답안지를 채점할 사람들을 국가시험 기간 동안 모집했고, 모집된 247명에게 그들이 피실험자라는 점을 알려주지 않은 채 이들을 뇌물 수수 기회에 노출했다. 연구자들은 실험이 끝난 후 피실험자들에게 뇌물 제공은 실험의 일부였다는 점을 통보했고, 모든 피실험자들은 그들이 실험에서 내린 결정들을 데이터화하는데 동의했다.<sup>9)</sup>

참가자들이 실험을 위해서 마련된 곳에 도착하면, 그들은 6개의 실험조건 가운데 하나로 무작위로 배정되었다. 먼저, 통제조건에서는 각 참가자들은 20개의 답안지를 채점한 후 실제 아르바이트를 할 때 받게 되는 양과 유사한 양의 일당을 받았으며, 11 번째 답안지와 관련하여 이 일당의 1/5에 해당하는 액수를 뇌물로 제안받았다. 실험 디자인상 뇌물 수수는 발각될 위험이 없었으며 뇌물을 수수한 참가자가 뇌물 공여자의 답안을 어떻게 채점하든 그 참가자의 일당에 영향을 주지 않았다. 두 번째 실험조건은 뇌물 액수가 통제조건 의 두 배가 된다는 점을 제외하고는 통제조건과 아무런 차이가 없었으며, 세 번째 실험조건은 통제조건에서 제공되는 일당의 140%를 제공한다는 사실을 참가자가 실험 장소에 도달했을 때 통보한다는 점을 제외하고는 통제조건과 아무런 차이가 없었다. 네 번째와 다섯 번째 실험조건 의 참가자들은 통제조건에서와 달리 자신이 채점한 결과의 정확도를 평가받았고 잘못 채점된 경우의 수에 따라서 일당이 삭감되었다. 두 조건의 차이는 네 번째 조건에서는 채점한 20개의 답안지 가운데 한 개의 답안지만 평가되었지만, 다섯 번째 조건에서는 5개의 답안지가 평가되고 그중 부정확도가 가장 높은 답안지를 기준으로 일당 삭감이 이루어졌다는 점이다. 마지막 실험조건은 부차적인 조건으로서, 통제조건에서의 뇌물 수수가 갖는 기본 효과를 파악하기 위하여 뇌물 수수 기회만 없을 뿐 통제조건과 동일했다.

논문은 다음과 같은 결과를 보고한다. 첫째, 뇌물 액수의 증가는 뇌물수수 빈도를

9) 당시에 이 연구가 해당 기관의 Institutional Review Boards(IRBs)로부터 연구승인을 받았는지 불분명하다. 그리고 사전 동의 없이 참가자를 비윤리적인 상황에 처하게 했다는 점에서 비윤리적인 연구라고 비판받을 여지가 충분히 있다.

25% 증가시킨 데 반해서 일당의 증가는 그 빈도를 23% 감소시켰으며, 이러한 변화량은 통계적으로 유의미했다. 그리고 두 종류의 정확도 측정 방식들 모두 뇌물수수 빈도를 낮추긴 했으나, 한 개의 답안지 채점만 평가되는 경우의 뇌물수수 감소 효과가 27%로 더 컸고 이 효과만이 통계적으로 유의미했다.<sup>10)</sup> 뇌물은 채점자의 채점과 관련해서는 다음과 같은 영향력을 끼쳤다. 첫째, 뇌물을 수수한 채점자들이 그렇지 않은 채점자들에 비해서 그들이 뇌물 공여자로 알고 있는 사람의 답안지를 합격시킬 확률이 32% 정도 높았다. 둘째, 뇌물 액수와 일당이 증가했을 때 뇌물 수수자들이 뇌물 공여자로 인식하고 있는 사람의 답안지를 합격시킬 확률은 각각 24%와 25% 증가했다. 그러나 뇌물을 거부한 사람에 대해서는 뇌물 액수와 일당의 증가가 통계적으로 유의미한 변화를 만들지 못했다. 셋째, 채점의 정확도를 확인하는 두 가지 방식은 뇌물 수수자와 거부자 모두에게 통계적으로 유의미한 영향력을 끼치지 못했다.

### 3. Robinson and Seim(2018)의 현장 실험

다음으로 소개할 연구는 부패한 행정 관리들이 어떤 사람을 대상으로 뇌물을 요구하는지에 관해 말라위(Malawi)에서 실제 교통경찰들을 대상으로 시행된 현장 실험 연구다. 이 연구를 위해서 연구자들은 6명의 말라위 사람을 연구 협조자로 고용한 후 이들을 20시간 정도 훈련시켰다. 그리고 이들을 매 4-5시간씩 5일에 걸쳐서 경찰검문소를 지나며 운전을 하게 했고, 그 와중에 경찰들이 뇌물을 요구하는지에 관해서 기록을 남기도록 요구했다.

경찰들이 뇌물을 요구할 수 있는 상황을 만들기 위해서 연구자들은 연구 협조자와 다른 운전자의 안전을 위협하지 않는 범위 안에서 연구 협조자가 도로교통법을 어기고 있는 것 “처럼” 보이도록 만들었다. 일반적으로 말라위 운전자들은 자신의 차가 책임보험에 가입했음을 표시하기 위해서 원거리에서도 눈에 잘 띄는 스티커 - disc라고 현지에서 불림 - 를 자동차 앞유리에 붙이고 다닌다. 그런데 말라위에서 자동차 책임보험 가입은 법적 의무사항이기 때문에, 경찰들은 스티커를 붙이지 않은 자동차를 발견할 경우 정차시켜 검문을 하고 만약 이때 운전자가 책임보험 가입을 증명하지 못할 경우 벌금을 부과한다. 이러한 상황에서 연구자들은 실험에 사용하는 모든 자동차들을 책임보험에 가입했음에도 불구하고 스티커를 자동차 앞유리에 붙이지 않았고, 자동차

10) 5개의 답안지 채점이 평가되는 경우의 효과가 더 적고 통계적으로 유의미하지 않은 결과에 대해서 저자들은 “더 강한” 감시가 뇌물수수 거부에 대한 내적 동인을 감소시켰기 때문이라고 해석한다.

사고 발생의 경우와 같이 꼭 필요한 상황이 아니라면 자동차 안쪽에 따로 숨겨둔 스티커를 경찰에 보여주지 말라고 연구 협조자에게 지시했다.

연구자들은 두 개의 실험적 조작을 교차시킴으로써 만들어지는 4개의 실험조건에 맞춰서 연구 협조자들의 자동차, 의상, 행동 지침들을 바꿨다. 첫 번째 실험적 조작은 연구 협조자들이 비싼 자동차(Audi나 Mercedes), 정장, 시계, 선글라스, 향수, 그리고 스마트폰을 사용하게 함으로써 그들의 사회경제적 지위가 높음을 드러내는 것이었다. 이러한 조작이 가해지지 않는 실험조건에서의 연구 협조자들은 값싼 자동차(Toyota Tercel), 청바지 차림의 일상복, 티셔츠, 그리고 플라스틱 샌들 혹은 맨발을 통해서 그들의 사회경제적 지위가 낮음을 드러냈다. 두 번째 실험적 조작은 연구 협조자의 정치적 영향력을 드러내는 것으로서, 협조자들은 말라위 정치인들의 전형적인 이미지를 따라 검문 받는 중 전화통화하거나 잘난 척했으며 말라위에서 통상적으로 상위 당 간부에게만 허용되는 정당 배지나 단추 역시 착용했다. 두 번째 실험적 조작이 가해지지 않은 실험 조건에서의 연구 협력자들은 정당 배지나 단추를 착용하지 않았을 뿐만 아니라 경찰들에게 아첨하듯이 행동했다.

연구자들은 총 333개의 경찰 반응을 수집했으며 이를 분석한 결과는 다음과 같다. 첫째, 비싼 자동차의 경우 47%가 검문을 받았지만 그렇지 않은 자동차의 경우 58%가 검문을 받았고 이 차이는 통계적으로 유의미했다.<sup>11)</sup> 둘째, 일단 검문을 받게 되었을 때, 경제사회적 지위는 뇌물 요구 빈도에 통계적으로 유의미한 영향을 주지 못했다(낮은 지위 89%, 높은 지위 92%). 셋째, 정치적 영향력은 뇌물 요구 빈도를 낮추는 것으로 드러났다(영향력 없음 94%, 영향력 있음 85%). 끝으로, 뇌물 요구 빈도를 낮추는 정치적 영향력의 효과는 오직 낮은 경제사회적 지위를 연기한 연구협조자들에게서만 두드러졌다.

#### 4. Chong et al.(2014)의 현장 실험

정치인들의 부패는 정치체제와 상관 없이 항상 존재하는 위협이지만, 민주주의 정치체제를 가진 국가의 시민들은 적어도 이론적으로는 선거를 통해서 부패한 정치인을 축출할 수 있다. Chong et al.(2014)은 실제로도 시민들이 부패한 정치인을 선거에서 축출하기 위하여 투표하는지 알아보기 위하여 2009년 멕시코 지방 선거 1주일 전 현

11) Robinson and Seim(2018)에 따르면 이 결과만으로는 경찰들이 차를 검문할 때 경제사회적 지위와 정치적 영향력 중 어느 것을 더 고려하는지 판별하기 어렵다. 왜냐하면, 실험디자인 측면에서 정치적 영향력이 자동차를 통해서 드러나지 않음에도 불구하고 말라위에서 경제사회적 지위와 정치적 영향력이 밀접하게 연관되어 있기 때문이다.

장 실험을 실시했다.

연구자들은 먼저 3개의 조작조건 각각에 배정할 150개의 선거구를 무작위로 선택했으며, 각 조작조건에 따라 서로 다른 유인물을 시민들에게 배부했다. 첫 번째 유형의 조작조건에 배정된 선거구에 거주하는 시민들이 받은 유인물에는 자기 지역의 시장들이 '착복'한 사회간접자본자금의 비율에 관한 정보가 담겨 있는데 반해서, 다른 두 유형의 조작조건에 배정된 선거구에 거주하는 시민들이 받은 유인물에는 '착복하지 않고 실제 사용된' 혹은 '가난한 지역'에서 사용된 사회간접자본자금의 비율에 관한 정보가 담겨 있었다.<sup>12)13)</sup> 통제조건에 배정된 1,910개의 선거구에 거주하는 시민들은 아무런 유인물을 받지 않았다.

연구자들은 선거구별 실제 선거 결과를 이용하여 실험적 조작의 효과를 확인했다. 먼저, 현직 시장의 부패에 관한 정보는 투표참여를 1.3%p 감소시켰고, 여당과 야당 후보에 대한 지지도를 각각 0.43%p와 0.86%p씩 감소시켰다. 그리고 이러한 부정적 효과들은 현직 시장의 부패 정도가 높을 때 더 강하게 나타났다. 다음으로 현직 시장의 부패에 관한 정보는 여당에 대한 정당일체감을 0.07%p 줄였으나, 야당에 대한 정당일체감에 대해서는 통계적으로 유의미한 수준의 변화를 만들어내지 못했다.

## 5. Abbink et al.(2014)의 실험실 실험

부패를 현장에서 직접 관찰·측정하기 어렵다는 한계를 극복하기 위하여 많은 연구자들이 실험실 실험을 활용했으며,<sup>14)</sup> 우리가 이번 장에서 마지막으로 소개할 Abbink et al.(2014)은 이러한 흐름 속에서 생산된 연구 중 하나이다. 이들은 뇌물 거래의 책임을 뇌물 수수자에게만 묻는 방식이 뇌물 거래 보고와 액수에 영향을 주는지 알아보기 위하여 인도의 6개 대학 - 두 개의 도시로부터 3개씩 - 을 선택했고 각 대학으로부터 60명의 학부 학생과 석사과정 학생들을 모집했다. 그 다음 각 대학에서 두 개의

12) 실험디자인 상 두 번째와 세 번째 실험조건은 플라시보(placebo) 테스트를 위한 것이다. 만약 첫 번째 실험조건에 배정된 선거구에 거주하는 시민들이 아무런 유인물을 받지 않은 통제조건에 배정된 선거구에 거주하는 시민들과 구별되는 투표 행위를 할 때, 이 차이의 원인이 부패에 관한 정보 때문인지 아니면 예상 밖의 유인물을 받았기 때문인지 구분하기 어렵다. 이 둘을 분리해내기 위해서 시민들이 부패에 관한 정보가 아닌 다른 정보가 담긴 유인물을 제공받는 실험조건이 필요하다.

13) 따라서 각 실험조건에 배정된 시민들은 자기 선거구의 시장이 누구인지에 따라서 서로 다른 정보를 받았다.

14) 부패에 관한 또 다른 실험실 연구를 찾는다면, Abbink et al.(2002), Barr and Serra (2009), 그리고 Cameron et al.(2009)를 보길 바란다.

실험 모임(모임 당 30명의 학생들이 참여)을 열었고, 이 모임들을 4개의 실험 조건 중 서로 다른 두 개의 실험 조건에 배정했다.

이때 통제조건에 배정된 학생들은 공무원과 시민 역할 둘 중 하나를 무작위로 맡게 되고, 공무원이 먼저 뇌물을 요구할 것인지 아닌지를 결정한 후 시민은 (1) '뇌물 공여 거부', (2) '뇌물 공여', (3) '뇌물 공여 후 보고' 중 하나를 선택한다. 만약 시민이 (2)나 (3)을 선택하면, 뇌물거래가 각각 5%와 40%의 확률로 발각되고 발각 시 공무원과 시민 모두 일정한 금전적 손해를 받게 된다. 조작조건 1은 뇌물거래 발각 시 오직 공무원만 금전적 손해를 입게 된다는 점을 제외하고는 통제조건과 동일하다. 조작조건 2는 시민이 (3)을 선택했음에도 불구하고 뇌물거래가 발각되지 않은 공무원이 그 시민에 대해서 금전적인 보복을 가할 것인지 말 것인지를 선택할 수 있다는 점을 제외하고는 조작조건 1과 동일하다. 조작조건 2에서 (그리고 조작조건 1에서도 마찬가지로) 시민이 (2)나 (3)을 선택한 후 뇌물거래가 발각되면 시민은 뇌물로 사용한 돈을 돌려받았다. 그러나 조작조건 3에서는 뇌물거래가 발각된다고 하더라도 시민에게 그가 뇌물로 사용한 돈을 돌려주지 않았다.

분석 결과 뇌물 거래의 책임을 공무원에게만 부과했을 때 뇌물 공여 거부를 선택한 시민들의 비율은 크게 변하지 않는데 반해 뇌물 공여 후 보고를 선택한 시민들의 빈도는 증가했다. 뇌물 공여 후 보고의 비율이 통제조건에서 25%였는데 조작조건 1에서는 59%로 증가했다. 그리고 조작조건 2와 3에서 이 비율은 각각 42%와 35%로 감소했다는 점은 조작조건 1에서의 큰 증가가 ① 공무원의 보복에 대해서 걱정할 필요가 없다는 점과 ② 뇌물거래 발각 시 뇌물로 사용한 돈을 돌려받을 수 있다는 생각에서 부분적으로 비롯됨을 의미한다. 다음으로 4개의 실험 조건들 중 조작조건 1에서 공무원이 요구한 뇌물액수의 평균값이 감소하긴 했으나 이는 공무원의 뇌물 요구 빈도 자체가 감소했기 때문이며, 전체적으로 4개의 실험 조건 사이에서 뇌물액수의 평균값은 통계적으로 유의미한 차이가 나지 않았다.

## IV. 사회복지 정책에 관한 실험들

### 1. Taubman et al.(2014)의 현장 실험

사회복지 정책 분야와 관련하여 우리가 소개할 첫 번째 논문은 미국 오레곤(Oregon) 주의 의료보조제도 'Medicaid'의 효과에 관한 논문이다. 2008년 오레곤 주

는 저소득층 성인들에게 Medicaid 프로그램을 제공하기 위하여, 3만 명을 추첨 (lottery)하여 통보한 후 지원서를 제출한 이들에 한해 Medicaid의 혜택을 받을 수 있도록 해주었다. Taubman et al.(2014)은 이 과정에서 새롭게 혜택을 받게 된 저소득 성인의 응급실 이용 횟수에 어떠한 변화가 발생하는지를 주 정부와 병원으로부터 얻은 데이터를 이용하여 확인한다. 뿐만 아니라 이들은 추첨 뒤 약 1년 후에 실시한 우편 설문조사와 2년 후 실시한 대면 인터뷰를 통해서 확보한 데이터들을 행정 데이터와 비교 활용한다.

추첨 선정이 통지된 2008년 3월부터 2009년 9월까지 약 18개월간의 응급실 사용 결과와 관련하여 행정 데이터를 통해 발견한 결과는 다음과 같다. 첫째, 이 기간 동안 Medicaid의 혜택을 받은 사람들의 응급실 방문 확률이 대략 7%p 높았으며(비수혜자 34.5%, 수혜자 41.5%) 이 차이는 통계적으로 유의미했다. 뿐만 아니라 Medicaid 프로그램은 방문 횟수 역시 증가시켰는데 대략 0.41회 정도 더 잦은 방문을 했던 것으로 나타난다. 이는 프로그램 혜택을 받지 못하는 사람이 1.02회 방문이라는 점을 감안하면 41%의 증가에 해당하며, 이러한 증가는 통계적으로 유의미했다. 그런데 설문조사를 통해서 확인한 결과는 Medicaid 프로그램이 응급실 사용과 관련하여 어떠한 유의미한 변화도 만들어내지 못한 것으로 나왔으며, Taubman et al.(2014)은 이러한 차이가 데이터의 시간적 길이에 있어서의 차이나 설문조사 특질 상 응답자들의 자기 선택(self-selection) 문제 등으로부터 비롯된 것으로 추측한다.

## 2. Michalopoulos et al.(2005)의 현장 실험

기존의 복지제도들이 수혜자들의 노동 동인을 약화시킨다는 문제점이 제기되었고 이 문제점을 해소할 수 있는 대안적 방안들이 고민되던 중, 'Self-Sufficiency Project(SSP)'라는 프로그램이 1990년 캐나다 브리티시 컬럼비아 주에서 기획되었다. 이 프로그램의 핵심은 수급자가 주당 30시간(논문에서의 전일제 고용 기준) 이상 일할 경우 SSP가 목표 소득으로 설정한 금액으로부터 그 수급자의 노동임금을 뺀 나머지의 절반에 해당하는 보조금을 3년 동안 지급하는 것이었다. 이 프로그램 하에서 SSP 보조금 수급자들은 수급자격을 유지하기 위하여 주당 30시간 이상씩 꾸준히 일하게 될 것으로 기대되었는데, 왜냐하면 기존 복지혜택에 전적으로 의존하거나 혹은 순전히 노동임금에 의존해 생계를 유지하는 경우에 비해서 수급자의 소득이 더 높아지기 때문이다.

주 정부는 SSP 프로그램이 실제로도 이러한 기대에 부합하게 작동하는지 알아보기

위하여 그 프로그램을 시범적으로 실시했다. 이 과정에서 역내 거주자 중 지난 6개월 동안 기존의 복지 혜택을 받지 못했던 19세 이상의 편부모 3,316명이 선별되었고, 이들은 이후 보조금 수급 지원자가 될 수 있는 기회가 제공된 조작조건과 그렇지 않은 통제조건 둘 중 하나에 무작위로 배정되었다.

실험조건 배정 12개월 그리고 30개월 후 실시된 설문조사를 통해서 Michalopoulos et al.(2005)은 다음과 같은 결과를 발견한다. 첫째, SSP 보조금은 (노동 시간과 무관하게) 일하는 사람의 수와 주당 30시간 이상 일하는 사람의 수를 12.1%p와 12.5%p 증가시켰다. 둘째, SSP 보조금은 노동 시간과 임금에도 영향을 끼쳤는데, 통제조건에 배정된 사람들에 비해서 조작조건에 배정된 사람들의 평균 노동 시간과 평균 임금이 각각 20시간과 \$242만큼 더 많았다. 셋째, 수급자들의 증가된 소득은 증가된 세금 납부로 이어졌고 이로 인해서 SSP 보조금을 위해서 소요된 순비용은 0에 가까워졌다. 끝으로, SSP 보조금은 수급자 개인뿐만 아니라 가정의 소득 증대와 빈곤 수준 감소로 이어졌다.

### 3. Bandiera et al.(2017)의 현장 실험

세 번째로 소개할 논문은 방글라데시 ‘Targeting the Ultra-Poor (TUP)’ 프로그램의 장·단기 효과를 측정하는 논문이다. BRAC이라는 NGO 단체에 의해서 시행된 이 프로그램은 최빈층 여성이 가축 및 기술을 보유함으로써 고소득을 올릴 수 있도록 국가 혹은 민간의 보조를 받지 않는 최빈층 여성에게 일회성으로 \$560 상당의 가축 및 사육 관련 교육 훈련을 지원했다. 프로그램 시행을 위해서 13개 최빈곤 구역이 1차적으로 선별되었고, 각 구역 내에서 무작위로 선별된 두 개의 군(subdistrict)이 통제조건과 조작조건에 무작위로 배정되었다. 이때, 각 군별로 있던 BRAC 지사의 8km 반경 안에 있는 모든 촌락은 한꺼번에 같은 실험 조건으로 포함되었으며, 조작조건에 배정된 촌락들은 2007년에 프로그램 혜택을 받았고 통제조건에 배정된 촌락들은 2011년 이후 그 혜택을 받았다.

Bandiera et al.(2017)은 프로그램의 장·단기 효과를 측정하기 위하여 통제조건과 조작조건에 할당된 1,309여 촌락의 6,700여개 빈곤 가구를 포함하여 총 21,000여 가구를 대상으로 7년간 4회에 걸쳐 설문조사를 실시했다. 확보된 데이터를 분석한 결과 연구자들은 최빈곤층 여성이 빈곤을 탈출하는 데 있어 TUP 프로그램이 장·단기적으로 효과적이었음을 밝힌다. 좀 더 구체적으로 말하자면, 통제조건에서는 가축이 좀 더 부유한 계층의 여성들만의 전유물에 가까웠다. 그런데 조작조건의 최빈곤층 여성들

은 가축 사육을 시작했을 뿐만 아니라, 이는 그들의 재산 증가와 빈곤 감소로 이어졌다.

#### 4. Dubois et al.(2012)의 현장 실험

1998년 멕시코 정부는 저소득층 교육, 보건, 영양 상태를 개선하기 위하여 일정한 조건을 충족시키는 가구에 현금을 제공(conditional cash transfer)하는 'Progesa'라는 프로그램을 시행했다. 이때 지원자에게 요구된 조건들 중 하나는 학령기 아동이 월 3회 이상의 결석 없이 85% 이상의 출석률을 유지하는 것이었다. 프로그램 시행 초기에 506개의 마을이 선정되었고 이들 중 무작위로 선택된 320개 마을(조작조건)에는 프로그램 시작 시기인 1998년 5월부터 프로그램이 실시되었다. 그리고 나머지 186개 마을(통제조건)에는 프로그램 시작 3년 후부터 프로그램 혜택이 제공되었다. Dubois et al.(2012)는 혜택 수급 시작 시기가 3년 차이난다는 점을 활용하여 학령기 아동의 학교 등록 여부 및 성적의 변화가 발생했는지를 다양한 행정 데이터를 활용하여 분석했다.

그 결과는 다음과 같다. 첫째, 1년간 프로그램 혜택을 받은 가구의 아동의 초등학교 성적이 그렇지 않은 가구의 아동보다 6% 정도 더 높았다. 둘째, 중등학교 학생들에 대한 프로그램 효과는 초등학교의 것과 달리 남녀 학생들 각각 프로그램 경험 시 성적의 평균 한계효과가 -22%와 -17%로 나타났다. 셋째, 학교 등록률의 경우 비수혜 가구 대비 수혜 가구의 아동 학교 등록률이 학년을 막론하고 평균 3.5% 높았다.

#### 5. Labonne(2013)의 현장 실험

마지막으로 우리가 소개할 논문은 멕시코의 Progesa 프로그램과 유사한 필리핀의 'Pantawid Pamilyang Philipino Program (4Ps)'가 단체장의 재선에 어떤 영향을 끼쳤는지를 분석한 논문이다. 이 프로그램은 0-14세 아동 및 임신부가 있는 가정 중 1인당 소득이 해당 지역 빈곤선 이하인 가구를 대상으로 시행되었다. 이때 대상 가구는 아동 수와 상관없이 1년간 월 PHP 500씩의 건강 보조금을 받았고, 만약 학령기 자녀가 있을 시 최대 3명까지 10개월간 자녀 당 월 PHP 300을 교육 보조금으로 받았다. 이에 따라 학령기 3명의 자녀가 있는 빈곤선 이하의 가구의 경우 연간 PHP 15,000을 지급받았다.

Labonne(2013)에 따르면 4Ps의 시행 지역(municipality)의 선정에 있어서 시장들의 영향력은 없었음에도 불구하고, 시장들은 자신의 재선 승리를 위하여 자기 지역에서 4Ps 시행에 자신들이 기여했음을 주장한다. 이러한 상황에서 Labonne(2013)은

4Ps가 현직 시장에 대한 투표율을 증가시키는지 분석한다. 연구자가 분석한 데이터에는 투표 정보를 알 수 없는 2개 지역을 제외한 총 17개 지역이 포함되어 있었으며, 이 중 9개의 지역에서는 각 지역 내 모든 촌락에서 4Ps가 시행되었다. 그리고 나머지 8개 지역의 경우 각 지역 내 촌락 중 무작위로 선정된 일부만이 4Ps 보조금을 받는 조작조건으로, 그리고 나머지 촌락은 보조금을 받지 못하는 통제조건으로 배정되었다.

2010년 5월 지방선거 득표율을 분석한 결과 Labonne(2013)은 다음과 같은 것들을 발견한다. 첫째, 정치적 경쟁이 큰 지역에서는 4Ps가 모든 촌락에 시행되었을 경우 전체 촌락의 반 정도에 시행되었을 경우보다 현직 단체장에 대한 지지율이 26%p 높은 것으로 나타났다. 둘째, 정치적 경쟁이 크지 않은 지역에서는 프로그램 수혜 여부가 현직 단체장 지지율 변화에 통계적으로 유의미한 영향을 끼치지 못하는 것으로 나타났다.

## V. 실험 연구 시 주의사항

지금까지 우리는 부패와 사회복지 정책에 관한 10편의 연구 논문을 소개했다. 글의 서두에서도 밝힌 바와 같이 우리는 이 연구들이 한국에서 복제시행(replication)되거나 한국적 상황에 맞게 변형되어 적용될 수 있다고 본다. 이번 장은 위의 논문들을 비판적으로 검토하면서, 그 과정에서 한국 연구자들이 주의해야 할 사항들에 대해서 논의한다.

### 1. 균형검사 시행

실험연구는 관찰대상을 실험조건에 무작위로 배정함으로써 관찰대상의 내생적 특질들이 종속변수에 끼칠 수 있는 영향력을 실험조건 수준에서 통제하려 하지만, 이 기대가 항상 실현되는 것은 아니다. 따라서 실험조건에 배정된 관찰대상의 평균적 특질들이 실제로도 차이가 없다는 점을 경험적으로 확인하는 ‘균형검사’는 실험연구에 있어서 중요한 위치를 차지하며 대략 2가지 방식으로 이루어진다. 첫 번째 방식은, Olken(2007)이 논문에서 사용한 것처럼 어떤 특정 실험조건에 대한 배정이 관찰대상의 내생적 특질들과 관련되어 있거나 영향을 끼칠 수 있을 만한 변수들과 무관하게 이루어졌는지를 회귀분석 모델을 이용하여 검정하는 방식이다. 그리고 두 번째 방식은 Chong et al.(2015), Taubman et al.(2014), Michalopoulos et al.(2005),

Labonne(2013), Bandiera et al.(2017) 등이 사용한 방식으로 관찰대상의 내생적 특질들과 관련되어 있거나 영향을 끼칠 수 있을 만한 변수 각각에 대해 각 실험조건에 배정된 관찰대상들이 갖는 평균값이 실험조건들 사이에서 차이가 없음을 확인하는 방식이다.

어떤 변수들을 회귀분석 모델에 포함시킬 것인지 혹은 어떤 변수에 대해서 실험조건 수준에서 비교할 것인지는 위에 기술된 2가지 방식 중 어떤 방식을 사용하는지와는 무관하게 무엇이 관찰대상인지와 어떤 맥락인지에 따라 달라진다. 예를 들어, 촌락이 관찰대상이었던 Olken(2007)의 연구에서는 촌락의 인구 수, 이슬람 사원의 수, 빈곤 가구 수, 촌장의 교육 수준, 촌장의 나이, 촌장직 급여 등의 변수들이 활용되었고, 시민이 관찰대상이었던 Taubman et al.(2014)의 연구에서는 시민의 출생년도, 나이, 언어, 전화번호를 제공했는지 아닌지, 주거 동네의 중위 가구 수입 등이 활용되었다. 그리고 빈곤 가구의 여성이 관찰대상이었던 Bandiera et al.(2017)의 경우 여성이 축산활동에 투자하는 시간, 농업에 투자하는 시간, 가사에 투자하는 시간, 가구 재산, 가구 저축 등이 활용되었다.

균형검사에 사용할 변수들 중 실험적 조작에 영향을 받을 수 있는 변수에 대한 데이터는 그 조작이 관찰대상에 가해지기 전에 수집되어야 한다. 예를 들어, Taubman et al.(2014)의 경우 균형검사에 사용된 변수에 관한 데이터가 언제 수집되었는지 논문에서 분명히 밝히지 않고 있는데, 만약 실험적 조작 이후에 수집되었다면 보고된 결과와는 달리 주거 동네의 중위 가구 수입 변수의 통제조건과 조작조건 상의 통계적 균형은 달성되지 않았을 수도 있다. 좀 더 구체적으로 말하자면, Medicaid 프로그램에 선택된 시민은 그 프로그램의 혜택을 통해서 절약한 돈을 시장경제 활동을 통해서 증식시켰을 수도 있으며, 만약 이것이 실제로 발생했다면 Taubman et al.(2014)이 보고한 결과는 주거 동네의 중위 가구 수입 변수와 관련하여 실험적 조작 이전에는 조작조건에서의 평균값이 통제조건에서의 평균값보다 더 낮았을 수 있음을 의미한다.

덧붙여, 균형검사의 중요성은 관찰대상의 수가 감소할수록 증가한다. 이는 관찰대상의 수가 감소할수록 관찰대상의 평균적 특질들이 실험조건들 사이에서 차이가 날 가능성이 증가하기 때문이다. 물론 이 가능성은 이론적 확률일 뿐, 실제로 많은 수의 관찰대상을 각 실험 조건에 배정했다고 하더라도 불균형이 발생할 수 있다. 예를 들어 Michalopoulos et al.(2005)의 경우 두 개의 실험조건 각각에 1,422개와 1,430개의 관찰대상들을 배정했지만 참가자들의 성별과 결혼 경험의 유무 등 몇 개의 변수에 대해서 통제조건과 조작조건에서 통계적 차이가 발생했다.

그런데 앞서 소개된 논문들 가운데에서 Armantier and Body(2011)의 경우 각 실

험조건에 배정된 관찰대상의 개수가 평균 41.17개에 불과함에도 균형검사의 결과를 보고하지 않았다. 실험적 조작의 효과를 회귀모형을 통해서 분석하는 과정에서 채점자들에게 대한 인구학적 변수들(예: 연령, 성별, 종교)이 통제변수로 포함되어 있다는 점을 감안하면 이 변수들을 활용하여 균형검사를 충분히 실시할 수 있었을 것으로 추정되며 그럼에도 불구하고 균형검사의 결과를 보고하지 않았다는 점은 Armantier and Body(2011)가 자신들이 발견한 실험조건들 간의 큰 불균형을 숨겼을 지 모른다고 의심하게 만든다. 그리고 Abbink et al.(2014)의 경우 각 실험조건에 90개의 관찰대상을 배정하고 있는데, 이들 역시 균형검사의 결과를 보고하지 않는다. 이들의 실험 디자인상 각 대학의 학생들이 오직 2개의 실험조건 가운데 하나에만 무작위로 배정되었다는 점 그리고 각 실험조건에 배정된 대학은 총 6개 중 3개씩에 불과했다는 점을 감안한다면, 각 실험조건에 90개의 관찰대상을 배정했다고 하더라도 그것이 실험조건 간의 균형을 보장했을 것이라고 기대하기는 어렵다. 따라서 Abbink et al.(2014) 역시 실험조건들 간의 불균형성을 의도적으로 숨기고 있다고 생각할 수 있는 여지가 남아 있다.

끝으로, 만약 실험적 조작에 의해 영향받지 않는 한 개 이상의 변수와 관련하여 관찰대상의 평균적 특질들이 실험조건들 사이에서 차이가 있는 것으로 드러날 경우, 실험적 조작의 효과에 대한 판정은 실험조건 수준에서의 분석이 아닌 개별 관찰대상 수준에서의 분석을 통해서 이루어져야 한다. 왜냐하면, 불균형을 발생시킨 바로 그 변수가 통제조건과 조작조건 사이에서 측정되는 종속변수상의 평균적 차이에 영향을 끼칠 수도 있기 때문이다. 개별 관찰대상 수준에서의 판정은 일반적인 회귀분석방법을 활용하면 되며 이때 불균형을 일으킨 변수를 포함하여 균형이 확인된 변수들도 함께 통제변수화 한다. 그리고 통계모델에 실험적 조작에 관한 변수를 독립변수로 포함시키고 이 변수의 계수가 통계적으로 유의미한지 확인한다.

## 2. 비순응 문제 주의

실험연구를 계획할 때 결정해야하는 것들 중 하나는 몇 개의 실험적 조작을 할 것인지 혹은 몇 개의 실험조건을 만들 것인지에 관한 결정이다. 이 어려움은 일반적으로 실험연구비용의 제약 때문에 발생하는 것으로서, 결정의 기준은 각 실험조건에 배정된 관찰대상의 개수를 충분히 확보하는 데 있어야 한다. 관찰대상 수의 증가는 통계적 검증력(statistical power)을 증가시킨다는 점에서 각각의 실험조건에 가능한 많은 관찰대상을 배정하면 좋으며 최소 30개를 배정하는 것이 통상적이다.

그런데 문제를 복잡하게 하는 것은 관찰대상에 대한 실험적 조작이 항상 성공하는 것이 아니라는 점이다. 달리 표현하면, 연구자가 관찰대상을 조작조건에 배정했다고 하더라도 그 관찰대상에 실험적 조작이 작동하지 않을 수도 있으며 - 이 문제는 실험연구 문헌에서 ‘비순응 문제’로 지칭된다<sup>15)</sup> - 이 가능성은 실험적 조작의 효과와 그 효과의 통계적 유의미성에 영향을 끼칠 수 있다. 예를 들어 Chong et al.(2014)이 유인물을 배부하는 과정에서 다양한 이유 때문에 유인물의 내용이 시민들에게 전달이 되지 않을 수도 있다. Chong et al.(2014) 역시 이 문제를 고려한 상태에서 통계분석을 해도 결과는 크게 바뀌지 않음을 보인다. 그러나 그들이 고려하지 않았던 점(혹은 그들의 논문에서 논의하지 않았던 가능성)은 어떤 시민이 유인물을 받았다고 하더라도 만약 그것을 곧바로 쓰레기통에 버렸다면 그 유인물에 담긴 정보에 노출되지 않았을 수도 있다는 점이다. 그리고 이러한 가능성의 존재는 만약 모든 시민들이 유인물에 담긴 정보에 제대로 노출되었다면 Chong et al.(2014)이 보고하고 있는 실험적 조작의 효과보다 더 큰 효과가 발생했을 수도 있다는 점을 의미한다.

비순응 문제가 심각할 것으로 예상될 경우 측정된 실험적 조작의 효과는 연구자가 관찰대상에 실험적 조작을 가하려 했던 “의도”의 효과(intention-to-treat effect 혹은 ITT effect)로 일반적으로 해석되어야 한다. 강제성을 부과할 수 없는 정책의 효과를 사전적으로 측정하기 위한 실험의 경우처럼 ITT 효과를 측정하는 것만으로도 큰 의미를 갖는 연구 상황도 분명히 존재한다. 그러나 중요한 점은 ITT 효과로 사후적으로 해석하던 안하던 상관없이 비순응의 문제는 실험적 조작의 효과(혹은 의도의 효과)의 크기에 영향을 준다는 것이다.

따라서, 실험연구자는 비순응 문제를 염두에 두면서 관찰대상의 수와 실험적 조작의 개수를 결정해야 한다. Chong et al.(2014)의 경우에서처럼 의도된 실험적 조작의 실패가 다양한 상황에서 발생할 수 있다는 점, 그리고 일반적으로 각 상황에서 어느 정도 그 실패가 발생할 것인지에 대해서 미리 알기 어렵다는 점에서 관찰대상의 수와 실험적 조작의 개수를 결정하는 것은 어느 정도는 연구자의 주관적인 판단에 좌우될 수밖에 없다. 그럼에도 불구하고 선행연구를 살펴보거나 혹은 시범연구(pilot study)를 직접 시행해보는 것은 성공적으로 조작된 관찰대상의 비율을 끌어올릴 수 있는 실험 디자인을 찾아내는 데 도움이 될 것이며, 이는 곧 적은 수의 관찰대상만으로도 실험적 조작의 효과를 통계적으로 판별하는 데 도움이 될 것이다.

15) 비순응 문제에 대한 자세한 논의는 Horiuchi et al.(2007)을 보길 바란다.

### 3. 확산효과 주의

연구자는 조작조건에 배정된 관찰대상에 실험적 조작이 실제로 발생했는지의 여부 뿐만 아니라, 통제조건에 배정된 관찰대상이 의도와 달리 실험적 조작에 노출되었을 가능성과 복수의 조작조건이 존재할 때 관찰대상이 배정된 조작조건과 무관한 실험적 조작에 노출될 가능성에 대해서도 주의를 기울여야 한다. 그 이유는 이러한 가능성이 실제화될 때 연구자가 의도한 실험적 조작의 효과는 상쇄되고 연구자는 그 효과에 대한 잘못된 판단을 내릴 수 있기 때문이다.

실험적 조작이 의도된 관찰대상이 아닌 다른 대상으로 확산되는 현상은 문헌에서 ‘확산효과’라고 지칭되는데<sup>16)</sup>, 이것이 무엇인지에 대해서 좀 더 구체적으로 알아보기 위하여 Chong et al.(2014)의 연구를 다시 살펴보자. 이 연구의 연구자들은 A라는 시민에게 시장이 얼마만큼의 사회간접자본자금을 착복했는지에 관한 유인물을 제공하려 의도했고, ‘다행스럽게도’ A는 그 유인물을 주의 깊게 읽었다고 가정해보자. 이 경우 비순응의 문제는 발생하지 않았으나 만약 A가 그 유인물을 통해서 알게 된 정보를 통제조건에 배정된 B에게 혹은 실제 사용된 사회간접자본에 관한 유인물만을 받기로 되어 있는 C에게 알려주었다고 생각해보자. 이런 상황이 빈번할수록 누설된 정보로 인해서 각 실험조건에 배정된 시민들 사이의 행동 간 차이는 약화될 것이다. 그리고 이러한 약화는 A와 같은 사람이 없었을 때에 비해서 시장의 사회간접자본 착복 정보가 갖는 선거 효과를 발견할 가능성을 낮추게 될 것이다. Olken(2007)의 실험적 조작 역시 확산효과와 가능성을 갖고 있었다. 예를 들어, 어떤 촌락에서 정부 감사를 받게 되었을 때 이 소식이 근처의 다른 촌락으로 전해지게 될 경우, 그 촌락의 촌장은 자기 촌락의 사업에 대해서도 감사가 있을 것이라고 생각하면서 원래는 착복하려 했던 공사비를 착복하지 않기로 결정할 수도 있다. 그리고 이 가능성은 감사받을 확률을 증가시키는 실험적 조작의 효과를 약화시킬 수도 있다. 또 다른 예로 Bandiera et al.(2017)이 활용한 TUP 프로그램을 생각해보 수 있다. 이 연구에서 어떤 가구의 여성이 그 프로그램의 혜택을 받는 조작조건에 배정되었을 때 그녀의 사회/경제적 활동이 통제 조건에 배정된 다른 가구의 여성의 사회/경제적 활동에 영향을 줄 수도 있었기 때문이다.

확산효과는 일반적으로 서로 다른 실험조건에 배정된 관찰대상들 간의 교류를 완전

16) 확산효과가 실제로 발생했는지를 어떻게 확인할 것인지, 어떤 실험 디자인을 통해서 확산효과를 약화시킬 수 있는지, 그리고 확산효과가 발생했을 가능성이 있을 경우 어떤 식으로 통제분석을 할 것인지에 관한 자세한 논의는 Sinclair et al.(2012)를 보라.

히 차단할 수 없을 때 발생하므로 상대적으로 실험 환경에 대한 통제력이 약한 현장실험을 진행하려는 연구자들에게 큰 고민일 수밖에 없다.<sup>17)</sup> 이러한 상황에서 확산효과의 가능성을 해소하거나 혹은 약화시키기 위하여 연구자들이 일반적으로 사용하는 방식은 관찰대상을 무작위로 배정할 때 개별적인 배정이 아니라 군집적으로 배정(clustered randomization)하는 것이다. 예를 들어 Olken(2007)의 경우 촌락들을 실험조건에 배정할 때, 어느 조건에 배정할 것인지를 각 촌락마다 독립적으로 결정하지 않고 군(subdistrict) 단위에서 결정했으며 같은 군에 속한 촌락들은 통째로 같은 실험조건에 배정했다. 군집배정방식의 기본 아이디어는 물리적 거리의 증가가 확산효과의 발생 가능성을 억누를 것이라는 사전적 기대이다. 따라서 실제로 군집배정방식이 얼마나 효과적이었는지는 Bandiera et al.(2017)나 다른 연구가 했던 것처럼 경험적으로 확인할 필요가 있다.

그런데 우리가 소개한 현장실험연구들 모두가 군집배정방식을 활용한 것은 아니다. 예를 들어 Dubois et al.(2012)와 Labonne(2013)의 연구에서 활용된 조건부 현금 지원 프로그램에서 수급 가구의 사회경제적 활동이 비수급 가구에 영향을 미칠 수 있음에도 불구하고,<sup>18)</sup> 수급 가구를 결정하는 과정에서 군집배정방식이 활용되지 않았다. 이는 관찰대상의 무작위적 배정 과정이 연구자가 아닌 정부에 의해서 이루어졌기 때문으로 추정된다. 확산효과의 가능성에 대한 사후적 논의조차도 결여되어 있어 구체적인 판단은 어렵지만, 수급자가 군집배정방식으로 선택되었다면 그때 발견되는 조건부 현금 지원 프로그램의 효과는 Dubois et al.(2012)와 Labonne(2013)가 보고하는 효과보다 더 컸을지도 모른다.

#### 4. 조작검사 시행

우리는 실험을 디자인하는 연구자들이 비순응의 문제에 대해서 주의를 기울일 것을 앞서 이야기한 바 있다. 그런데, 설사 비순응하는 관찰대상이 없는 경우에도 문제는 남아있다. 그것은 관찰대상이 연구자가 의도한 것과 다른 방식으로 실험적 조작을 받아들일 수도 있다는 점이다.

예를 들어, Robinson and Seim(2018)은 사회경제적 지위와 정치적 영향력의 차이를 실험적으로 조작하기 위해서 자동차의 종류와 운전자의 외향과 행동양식에 변화

17) 확산효과는 다른 실험 방식에서도 발생할 수 있다. 예를 들어, 여론조사실험에서의 확산효과에 대해서는 Transue et al.(2009)를 보라.

18) 이에 관해서는 Angelucci and De Giorgi(2009)를 참고하길 바란다.

를 주었는데, 그들의 관찰대상이었던 김문소의 교통경찰들이 자동차의 종류나 운전자의 외향과 행동양식을 통해서 운전자의 사회경제적 지위와 정치적 영향력을 실제로 추측했는지 그리고 그 추측이 그들의 의도와 실제로 일치했는지는 다른 문제이다. 좀 더 구체적으로 말하자면, 만약 Robinson and Seim(2018)이 말하는 것처럼 말라위(Malawi)에서 사회경제적 지위와 정치적 영향력이 밀접하게 연결되어 있다면, 사회경제적 지위에 관한 실험적 조작과 정치적 영향력에 관한 실험적 조작 둘 중 하나만 관찰대상에게 가해진다고 하더라도 관찰대상이 다른 하나에 관해서도 함께 조작되는 현상이 발생할 수도 있다. 즉, 낮은 사회경제적 지위와 높은 정치적 영향력의 실험조건에 배정된 교통경찰들이 운전자의 높은 정치적 영향력을 표상하는 것들을 보고 그 운전자가 높은 사회경제적 지위를 갖고 있는 것으로 인식했을 수도 있다. 같은 방식으로 높은 사회경제적 지위와 낮은 정치적 영향력의 실험조건에 배정된 교통경찰들이 운전자의 높은 사회경제적 지위를 표상하는 것들을 보고 그 운전자가 높은 정치적 영향력을 갖고 있는 것으로 인식했을 수도 있다.

실험적 조작이 연구자의 의도대로 관찰대상에게 받아들여지는가의 문제는 종속변수와 관련하여 통제조건과 조작조건에서 관찰된 차이가 만약 실험적 조작이 연구자의 의도대로 관찰대상에게 받아들여졌을 경우 발견될 수 있는 차이와 다를 수 있다는 점에서 중요하다. 예를 들어, Robinson and Seim(2018)은 (1) 낮은 사회경제적 지위와 낮은 정치적 영향력의 조건에 배정된 교통경찰관에 비해서 (2) 낮은 사회경제적 지위와 높은 정치적 영향력의 조건에 배정된 교통경찰관들의 뇌물 요구 빈도가 더 낮다는 것을 발견하고 이를 높은 정치적 영향력의 효과로 해석했다. 그러나 만약 위의 단락에서 언급한 첫 번째 가능성이 실제로 발생했다면, 연구자들이 비교하고 있는 저 두 실험조건은 교통경찰관들의 '인식'의 관점에서는 각각 (1) 낮은 사회경제적 지위와 낮은 정치적 영향력의 조건과 (2) 높은 사회경제적 지위와 높은 정치적 영향력의 조건에 해당할 것이며, 따라서 그들이 관찰한 차이의 일부가 높은 사회경제적 지위에 의해서 비롯되었을 가능성을 완전히 배제할 수 없게 된다. 뿐만 아니라 연구자들은 뇌물청구 억제에 관한 높은 정치적 영향력의 효과는 높은 사회경제적 지위의 운전자들과 관련해서는 관찰되지 않았다고 보고하고 있는데<sup>19)</sup>, 이는 위의 단락에서 언급한 두 번째 가능성 때문에 높은 사회경제적 지위를 가진 운전자들과 관련하여 정치적 영향력의 차이가 제대로 인식되지 않았기 때문일 수도 있다. 달리 표현하면, 교통경찰관들의 '인식'의 관

19) 즉, 높은 사회경제적 지위와 낮은 정치적 영향력의 조건에 배정된 교통경찰관들과 높은 사회경제적 지위와 높은 정치적 영향력의 조건에 배정된 경찰관들 사이에서 뇌물 요구 빈도와 관련해서 아무런 차이를 발견하지 못했다는 것이다.

점에서는 두 조건 모두 높은 사회경제적 지위와 높은 정치적 영향력의 조건에 해당했었을 수도 있다. 이는 연구자들이 논문에서 보고하고 있는 결과들은 만약 사회적 지위와 정치적 지위에 관한 실험적 조작이 제대로 이루어졌을 경우 발견할 수 있는 결과들과는 다를 수 있음을 시사한다.

연구자가 의도한 방식대로 실험적 조작이 관찰대상에게 전달되었는지를 그 조작이 마무리된 이후 관찰대상을 대상으로 확인하는 검사를 ‘조작검사’라고 하며 일반적으로 설문조사형태의 문항을 활용한다. Robinson and Seim(2018)의 경우 관찰대상이 교통경찰이었고 또 연구주체가 뇌물이었다는 점 때문에 이들을 대상으로 직접적인 조작 검사를 시행하는 것은 어려울 수도 있으나, 이 경우 실험 지역에 포함되지 않은 지역에서 검문 중인 교통경찰에게나 혹은 교통경찰이 아닌 다른 유형의 경찰에게 실험적 조작에 대한 인식을 물어보는 등 제 3자를 대상으로 조작검사를 시행하고 이를 통해서 조작의 성공 여부 판단의 객관성을 확보할 수 있다. 그런데 우리가 소개한 논문 10편 중 Robinson and Seim(2018)를 포함하여 그 어떤 논문도 조작검사의 결과를 보고하지 않았다. 이는 Robinson and Seim(2018)의 연구를 제외하고는 실험적 조작의 특성상 연구자의 의도와 다른 방식으로 관찰대상에게 전달될 가능성이 비교적 낮았기 때문일 수도 있다. 그러나 이에 대한 판단은 주관적일 수 있으므로, 가능한 실험 디자인 단계부터 조작검사를 준비하는 것이 바람직하다.<sup>20)</sup>

## 5. 실험적 조작에 의해서 영향받았을 수 있는 변수를 통제변수로 사용하지 말 것

실험결과를 데이터화한 후 실험적 조작의 효과를 회귀분석을 통해 검정할 때 연구자들은 어떤 변수를 통제변수로 사용할 것인가에 대해 고민하게 된다. 이때 일반적으로는 관찰적 연구에서 사용되는 방식을 그대로 적용하면 되지만, 실험적 조작에 의해서 영향을 받았을 수 있는 변수는 통제변수로 사용하지 말아야한다. 왜냐하면 그 변수를 추가하면 회귀분석의 결과가 관찰 불가능한 혹은 측정 불가능한 요인들에 의해 왜곡될 수 있기 때문이다(King and Zeng 2006; Montgomery et al. 2018).

예를 들어, Armantier and Boly(2011)는 자신들의 실험적 조작을 통계적으로 검증하기 위하여 회귀모형을 사용하고 있는데 이때 사용된 통제변수들 중에는 채점자가

20) 조작 검사를 통과하지 못한 관찰대상을 데이터에서 누락시키는 것이 일반적이었으나 이것이 통계적 편향을 불러일으킬 수 있음을 Aronow et al.(2019)와 Montgomery et al.(2018)은 지적하고 있다.

채점하는데 걸린 총 시간에 관한 변수도 포함되어 있다. 문제는 채점자에 대한 뇌물 공여가 총 20개의 답안지 중 11번째 답안지를 채점할 때 이루어진다는 점이다. 이 뇌물 공여의 시점에서 채점자가 뇌물을 수수했는지의 여부는 그 채점자가 이후 남아 있는 답안지를 채점하는데 걸리는 시간에 영향을 끼칠 수 있으며, 이때 뇌물수수 여부는 그 채점자가 어떤 실험조건에 배정되었는지에 좌우된다. 따라서 11번째 답안지부터 20번째 답안지까지 채점하는데 걸리는 시간은 실험조건에 따라서 달라지며, 이는 곧 총채점시간 자체가 실험조건에 따라서 달라짐을 의미한다. 즉, 1번부터 10번까지의 답안지를 채점하는데 소요되는 시간의 관점에서 봤을 때 채점자들이 실험조건에 무작위로 배정되었다고 하더라도 총채점시간의 관점에서 봤을 때에는 채점자들이 실험조건에 무작위적이지 않는 방식으로 배정되었다는 것임을 의미한다. 결과적으로, 총채점시간이 통계모델에 포함될 경우, 총채점시간과 관련된 관찰 불가능하고 측정 불가능한 요인들이 통계결과를 왜곡할 수 있게 된다.

## VI. 나가며

사회복지정책의 효과에 대한 최근 논쟁 속에서 그 효과를 경험적으로 검정해보기 위한 방법으로 실험연구에 대한 관심이 한국 사회과학계에서 급속히 증가했다. 국내에서 실험연구가 활성화되는 것에 일조하기 위하여 우리는 이 논문에서 두 가지를 수행했다. 첫째, 국내 사회과학계 연구자들이 많은 관심을 갖고 있을 것으로 생각되는 부패 및 사회복지 정책에 대한 실험 논문 10편을 소개했다. 둘째, 실험연구를 할 때 주의해야 할 다섯 가지 사항에 대해서 논의하였다. 그것들은 실험조건에 배정된 관찰대상의 평균적 특질들이 실제로도 차이가 없다는 점을 경험적으로 확인하는 ‘균형검사’의 중요성, 실험적 조작이 실제로는 의도한 대로 이루어지지 않는 ‘비순응 문제’의 발생 가능성을 염두에 두어 관찰대상의 수와 실험적 조작의 개수를 결정하는 것의 중요성, 실험적 조작이 의도된 관찰대상이 아닌 다른 대상으로 확산되는 일명 ‘확산효과’ 통제 의 필요성, 연구자가 의도한 방식대로 실험적 조작이 관찰대상에게 전달되었는지를 그 조작이 마무리된 이후 관찰대상을 대상으로 확인하는 ‘조작검사’의 중요성, 그리고 실험적 조작에 의해서 영향받았을 수 있는 변수는 통제변수로 사용하지 말아야한다는 것 등이다.

서론에서 밝힌 바와 같이, 이러한 주의사항들은 이미 다른 실험연구 개론서나 개별 논문 등에서 깊이 있게 논의된 바 있다. 그러나 대부분의 경우 이러한 논의들은 한두

개의 사항들에 초점을 맞추고 있거나 혹은 지나치게 이론적인 논의에 치우쳐서 실험연구에 익숙하지 않은 연구자들이 손쉽게 이해하기 어려운 부분들이 존재한다. 이와 달리 우리는 부패와 사회복지 정책에 관한 실험 논문 10편을 활용하여 복수의 주의사항들을 함께 논의했는데, 이 방식이 실험연구에 익숙하지 않은 연구자들에게 그 주의사항들을 전달함에 있어서 좀 더 효율적이었길 기대한다.

끝으로 두 가지를 이야기하면서 이 글을 마무리하고자 한다. 첫째, 관찰 대상에게 인위적인 실험적 조작을 가하는 실험연구의 경우 그 과정에서 발생할 수 있는 윤리적 문제들에 대해서 연구자가 각별한 주의를 기울여야 한다. 특히 Armantier and Boly(2011)의 경우처럼 참가자들의 사전 동의 없이 뇌물수수라는 비윤리적인 상황에 그들을 밀어 넣는 연구는 분명히 지양되어야 할 것이다. 둘째, 실험 방법은 인과관계 수립과 관찰/측정 불가능한 변수 통제 문제에 있어서의 용이성 때문에 그 연구 결과물이 갖는 학문적 파급력이 큰 편이지만 그만큼 실험 디자인에 있어서 높은 엄밀성을 요구받는다. 달리 표현하면, 단순히 실험적 조작을 관찰대상에 무작위로 할당하는 것만이 실험연구 디자인의 전부라 아니라는 것이며, 연구자는 실험적 조작의 효과를 왜곡할 수 있는 다양한 가능성들과 대처 방안에 대해서 디자인 단계에서부터 고민해야 한다. 물론 연구자들은 종종 Michalopoulos et al.(2005)처럼 정부나 사회단체가 디자인하고 이미 시행된 실험의 결과에 관한 데이터만을 분석할 수 있는 경우도 있다. 최선은 연구자들이 정부나 사회단체가 실험을 준비하는 단계부터 개입하여 실험적 조작의 효과를 과학적으로 검증해낼 수 있는 디자인을 자문하는 것이겠지만 이는 다양한 이유로 어렵다. 이 경우 연구자의 역할은 종속변수와 관련하여 통제조건과 조작조건을 단순 비교하는 것을 넘어서, 관찰된 차이가 실험적 조작이 아닌 다른 원인에 의해서 왜곡되었을 가능성을 다양한 추가적 자료들을 활용하여 검토하는 것이 되어야할 것이다.

## ■ 참고문헌

- Abbink, Klaus, Bernd Irlenbusch, & Elke Renner. 2002. "An experimental bribery game." *Journal of Law, Economics, and Organization*, 18(2): 428~454.
- Abbink, Klaus, Utteeyo Dasgupta, Lata Gangadharan, & Tarun Jain. 2014. "Letting the briber go free: an experiment on mitigating harassment bribes." *Journal of Public Economics*, 111: 17~28.
- Anduiza, Eva, Aina Gallego, & Jordi Muñoz. 2013. "Turning a blind eye:

- Experimental evidence of partisan bias in attitudes toward corruption.” *Comparative Political Studies*, 46(12): 1664~1692.
- Angelucci, Manuela, & Giacomo De Giorgi. 2009. “Indirect effects of an aid program: how do cash transfers affect ineligibles’ consumption?.” *American Economic Review*, 99(1): 486~508.
- Armantier, Olivier, & Amadou Boly. 2011. “A controlled field experiment on corruption.” *European Economic Review*, 55(8): 1072~1082.
- Aronow, Peter M., Jonathon Baron, & Lauren Pinson. 2019. “A note on dropping experimental subjects who fail a manipulation check.” *Political Analysis*, 27(4): 572~589.
- Bandiera, Oriana, Andrea Prat, & Tommaso Valletti. 2009. “Active and passive waste in government spending: Evidence from a policy experiment.” *American Economic Review*, 99(4): 1278~1308.
- Bandiera, Oriana, Robin Burgess, Narayan Das, Selim Gulesci, Imran Rasul, & Munshi Sulaiman. 2017. “Labor markets and poverty in village economies”. *The Quarterly Journal of Economics*, 132(2): 811~870.
- Banerjee, Abhijit, Esther Duflo, Nathanael Goldberg, Dean Karlan, Robert Osei, William Parienté, Jeremy Shapiro, Bram Thuysbaert, & Christopher Udry. 2015. “A multifaceted program causes lasting progress for the very poor: Evidence from six countries.” *Science*, 348(6236): 1260799.
- Barr, Abigail, & Danila Serra. 2009. “The effects of externalities and framing on bribery in a petty corruption experiment.” *Experimental Economics*, 12(4): 488~503.
- Bertrand, Marianne, Simeon Djankov, Rema Hanna, & Sendhil Mullainathan. 2007. “Obtaining a driver’s license in India: an experimental approach to studying corruption.” *The Quarterly Journal of Economics*, 122(4): 1639~1676.
- Cappelen, Alexander. W., Knut Nygaard, Erik Ø. Sørensen, & Bertil Tungodden. 2015. “Social preferences in the lab: A comparison of students and a representative population.” *The Scandinavian Journal of Economics*, 117(4): 1306~1326.
- Cameron, Lisa, Ananish Chaudhuri, Nisvan Erkal, & Lata Gangadharan. 2009.

- “Propensities to engage in and punish corrupt behavior: Experimental evidence from Australia, India, Indonesia and Singapore.” *Journal of Public Economics*, 93(7-8): 843~851.
- Chong, Alberto, Ana L. De La O, Dean Karlan, & Leonard Wantchekon. 2014. “Does corruption information inspire the fight or quash the hope? A field experiment in Mexico on voter turnout, choice, and party identification.” *The Journal of Politics*, 77(1): 55~71.
- Cleave, Blair L., Nikos Nikiforakis, & Robert Slonim. 2013. “Is there selection bias in laboratory experiments? The case of social and risk preferences.” *Experimental Economics*, 16(3): 372~382.
- Druckman, James N., & Cindy D. Kam. 2011. “Students as experimental participants.” *Cambridge Handbook of Experimental Political Science*, 41~57.
- Dubois, Pierre, Alain De Janvry, & Elisabeth Sadoulet. 2012. “Effects on school enrollment and performance of a conditional cash transfer program in Mexico.” *Journal of Labor Economics*, 30(3): 555~589.
- Falk, Armin, & James J. Heckman. 2009. “Lab experiments are a major source of knowledge in the social sciences.” *Science*, 326(5952): 535~538.
- Falk, Armin, Stephan Meier, & Christian Zehnder. 2013. “Do lab experiments misrepresent social preferences? The case of self-selected student samples.” *Journal of the European Economic Association*, 11(4): 839~852.
- Fisman, Raymond, & Shang-Jin Wei. 2004. “Tax rates and tax evasion: Evidence from “Missing imports” in China.” *Journal of Political Economy*, 112(2): 471~496.
- Fortin, Bernard, Guy Lacroix, & Simon Drolet, 2004. “Welfare benefits and the duration of welfare spells: evidence from a natural experiment in Canada.” *Journal of Public Economics*, 88(7-8): 1495~1520.
- Gerber, Alan S., & Donald P. Green. 2011. “Field experiments and natural experiments.” In Goodin, R. E (eds). *The Oxford handbook of political science*, Oxford Press.
- Harrison, Glenn W., Morten I. Lau, & E. Elisabet Rutström. 2009. “Risk

- attitudes, randomization to treatment, and self-selection into experiments.” *Journal of Economic Behavior & Organization*, 70(3): 498~507.
- Horiuchi, Yusaku, Kosuke Imai, & Naoko Taniguchi. 2007. “Designing and analyzing randomized experiments: Application to a Japanese election survey experiment.” *American Journal of Political Science*, 51(3): 669~687.
- Kam, Cindy D. 2005. “Who toes the party line? Cues, values, and individual differences.” *Political Behavior*, 27(2): 163~182.
- King, Gary, & Langche Zeng. 2006. “The Dangers of Extreme Counterfactuals.” *Political Analysis*, 14(2): 131-159.
- Labonne, Julien. 2013. “The local electoral impacts of conditional cash transfers: Evidence from a field experiment.” *Journal of Development Economics*, 104: 73~88.
- Michalopoulos, Charles, Philip K. Robins, & David Card. 2005. “When financial work incentives pay for themselves: evidence from a randomized social experiment for welfare recipients.” *Journal of Public Economics*, 89(1): 5~29.
- Montgomery, Jacob M., Brendan Nyhan, & Michelle Torres. 2018. “How conditioning on posttreatment variables can ruin your experiment and what to do about it.” *American Journal of Political Science*, 62(3): 760~775.
- Morton, Rebecca B., & Kenneth C. Williams. 2010. *Experimental political science and the study of causality: From nature to the lab*. Cambridge University Press.
- Muralidharan, Karthik, Paul Niehaus, & Sandip Sukhtankar. 2016. “Building state capacity: Evidence from biometric smartcards in India.” *American Economic Review*, 106(10): 2895~2929.
- Olken, Benjamin A. 2007. “Monitoring corruption: evidence from a field experiment in Indonesia.” *Journal of Political Economy*, 115(2): 200~249.
- Powell, Benjamin, Jeff R. Clark, & Alex Nowrasteh. 2017. “Does mass immigration destroy institutions? 1990s Israel as a natural experiment.”

*Journal of Economic Behavior & Organization*, 141: 83~95.

Robinson, Amanda Lea & Brigitte Seim. 2018. "Who is targeted in corruption? Disentangling the effects of wealth and power on exposure to bribery." *Quarterly Journal of Political Science*, 13(3): 313~331.

Sinclair, Betsy, Margaret McConnell, & Donald P. Green. 2012. "Detecting spillover effects: Design and analysis of multilevel experiments." *American Journal of Political Science*, 56(4): 1055~1069.

Taubman, Sarah L., Heidi L. Allen, Bill J. Wright, Katherine Baicker, & Amy N. Finkelstein. 2014. "Medicaid increases emergency-department use: evidence from Oregon's Health Insurance Experiment." *Science*, 343(6168): 263~268.

Transue, John E., Daniel J. Lee, & John H. Aldrich. 2009. "Treatment spillover effects across survey experiments." *Political Analysis*, 17(2): 143~161.

Zizzo, Daniel John. 2010. "Experimenter demand effects in economic experiments." *Experimental Economics*, 13(1): 75~98.

안도경. 2014. "정치학과 실험." 《평화연구》, 22(2): 326-359.

## Five lessons for Experimenters from a Review of Experimental Studies on Corruption and Social Welfare

Han Il Chang & Iljoo Park

To promote experimental studies among social scientists in South Korea, we introduce 10 articles that experimentally study corruption and social welfare, of which the social scientists are expected to be interested. Utilizing the articles, we then discuss five issues to which the social scientists have to pay attention while conducting experimental studies. The issues are (1) the significance of balance tests to verify no differences in the average attributes of observational units between experimental conditions, (2) noncompliance issues that emerge as experimental manipulation on observational units fails, (3) spillover effects that concern the unintended exposure of observational units to experimental manipulation, (4) the importance of manipulation checks to verify that observational units took experimental manipulation in the manner intended by the experimenters, and (5) the inappropriateness of using as a control variable a factor that is likely to have been affected by experimental manipulation.

※ Keywords: experiment, balance test, noncompliance, spillover effect, manipulation check