

발간 등록 번호

11-B260003-000519-14

KOICA

지구촌 행복시대, KOICA가 함께합니다.

제28호 | 2016. 12월

개발과 이슈

영향평가(Impact Evaluation) 방법론 및 사례 : RCT(무작위 대조군 연구) 및 DID(이중차감법)을 중심으로

1. 영향평가(Impact Evaluation)의 목적
2. 영향평가(Impact Evaluation)의 정의
3. 영향평가를 위한 방법론 종류
4. 영향평가 윤리 기준
5. 타 원조기관의 영향평가 사례
 - 1) 실험설계(RCT) 사례 : 미국 MIT 산하 IPAL의 인도네시아 의료보험가입 요인 분석
 - 2) 준실험설계(DID) 사례: 일본 JICA의 식수개발사업 영향평가
6. 시사점 및 제언

파리 선언(Paris Declaration)에서 강조된 '개발 결과를 위한 성과관리' 이행과 정부 원조지출에 대한 성과 증명(Value for Money)에 대한 중요성이 커지고 있으며, 이에 따라 사업의 객관적인 성과제시와 공표가 중요해지고 있다.

이를 위해서는 과학적 방법론에 기반을 둔 영향평가, 신빙성 있는 대조군에 기반을 둔 평가 추정치가 필요하다. 과학적 방법론에 기반을 둔 영향평가를 시행하기 위해서는 1) 각 개별 사업 당 가장 적절한 영향평가 방법론이 무엇인지를 찾고 2) 사업기획단계에서부터의 준비가 가장 중요하다.

KOICA에서도 최근 과학적 방법론에 기반한 엄격한 영향평가가 시행되고 있으나, 아직 소수의 사업을 대상으로 하고 있으며, 사업의 현실적 상황 및 여건, 내용에 따라 다양한 방법론이 사용될 수 있음에도 불구하고, 충분히 사업기획 단계에서 고려가 되고 있지 못하고 있는 것이 현실이다.

따라서 본 연구는 대표적인 과학적인 영향평가 방법론을 알아보고, 방법론 별 특성을 살펴보고자 한다. 또한 대표적인 영향평가 방법론인 무작위 대조군 연구(이하 RCT, Randomized Control Trials)와 이중차감법(이하 DID, Difference in Difference)을 이용한 실제 사례를 살펴본다. 이를 통해 영향평가를 고려한 KOICA 사업계획시 필요한 시사점을 제시하고자 한다.

연구개발팀 이상미

1. 영향평가(Impact Evaluation)의 목적

영향평가는 프로젝트, 프로그램 등의 사업(intervention) 또는 정책의 파급효과에 대한 평가로서 일반적인 사업평가와 달리, 실험적 접근법에 근거를 두어 보다 엄격하게 평가 대상과 그 파급효과 사이의 인과 관계를 증명하고자 하는 평가를 의미한다.(World Bank, 2011) 특히 영향평가는 사업 설계, 수행, 평가 등 사업 초기부터 사업 주기의 전 과정에 원조효과성에 대해 고려하여, 방법론적 엄격성을 강조함으로써 “원조효과성을 어떻게 엄밀히 측정할까?”에 대한 답을 찾아가는데 도움을 주는 평가도구이다. 또한 정부의 정책결정 책임성, 투명성을 강화하는 증거기반 정책결정(Evidence-Based Policy Making) 도구로 활용된다.

[BOX 1] 평가가 ‘정책 지속가능성’에 활용된 사례

Case 멕시코 조건부 현금 지원정책 (Conditional Cash Transfer)

- 조건부 현금지원정책(Conditional Cash Transfer, CCT)는 일정소득 이하의 가난한 가구를 대상으로 하며, 자녀를 학교에 보내거나 건강검진을 정기적으로 받는 것과 같은 특정조건을 달성하는지 여부에 따라 현금을 지급하는 프로그램
- CCT 프로그램의 결과로 사업전후 학교등록율 0.7% 상승, 어린이 질병 유발율 23% 감소를 확인하였고, 멕시코 신정부 출범 시 동 정책을 지속적으로 유지하는 데에 객관적인 증거를 제공하는 계기가 됨. 아울러, 멕시코의 CCT 프로그램의 객관적 평가결과를 바탕으로 조건부 현금지급 프로그램은 전 세계에서 다양한 방식으로 확산 중

출처: World Bank(2011), KOICA 영향평가 학습 동아리 자료¹⁾

2. 영향평가(Impact Evaluation)의 정의

모니터링이 프로젝트·프로그램의 정보를 수집하고 분석하는 지속적인 과정이라면, 평가는 종료·진행 중인 프로젝트·프로그램 계획, 집행, 결과에 대한 목적성 있는 행위이다(표1 참조).

1) 평가실 라미혜 직원 작성

[표 1] 모니터링 및 평가의 차이

구분	모니터링(Monitoring)	평가(Evaluation)
주기	주기적/지속적	한정된 기간
범위	모든 프로그램	선정된 프로그램
자료	전체	표본 기반
정보의 깊이	사업이행 추적 및 관찰을 통한 자료	성과 및 영향, 과정(Why)에 집중
활용도	진행 중인 사업의 개선 및 관리	향후 타 사업의 질제고에 도움이 될 교훈 도출

출처 : World Bank(2011)

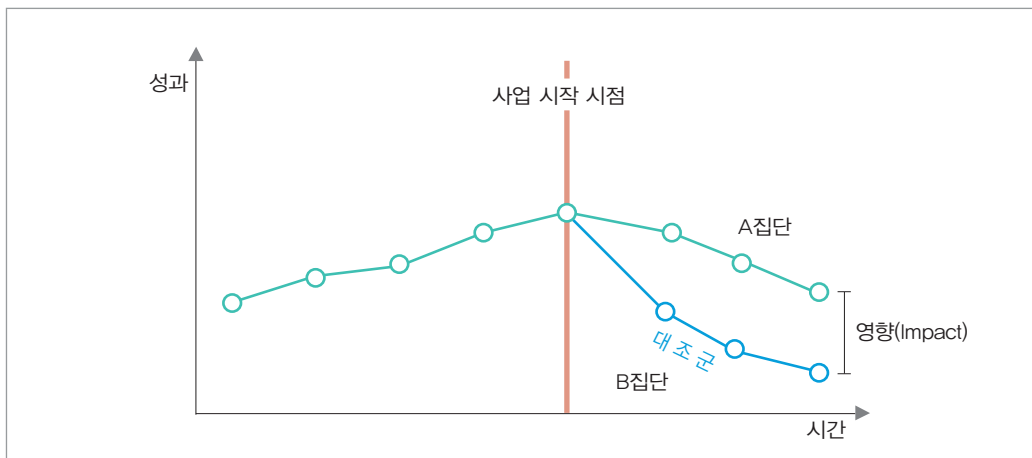
평가(evaluation)는 성과 및 영향, 과정(Why)을 밝혀내는 것이 관건이며, 이를 위해, 대표적으로 ① 사업 수행지역의 사업 전후(before/After)를 비교 방법, ② 사업 혜택을 받은 상태(실험군, Treatment group)와 혜택을 받지 않은 상태(비교군 혹은 대조군, Comparison or Control group)를 비교하는 방법이 사용된다.

가장 적용이 용이한 방법은 사업 전후 비교방법이나, 이는 한계를 가진다. 전후 비교만으로 우리사업의 성과를 과학적으로 측정하였다고 보기 힘들기 때문이다. 예를 들면, A라는 지역에 깨끗한 식수를 공급하기 위해 식수개발 원조사업을 5년 동안 진행한 경우, 사업의 성과 측정 방식을 설계한다고 가정하자. 사업 시작 전 대비, 사업 종료 후 대상지역의 수인성 질병 발생률이 10% 감소하였다면, 이 10%를 우리 사업으로 인한 영향이라고 단정할 수 있을까? 같은 지역에서 수행된 타 원조기관의 사업, 대상지의 날씨 변화 등 이 지역의 수인성 질병 발생률에 영향을 줄 수 있는 요인은 다양하게 존재하기 때문이다. 따라서 10%는 온전히 이 원조 사업의 영향이라고 보기는 힘들다.

따라서 원조 사업으로 인한 성과(영향)를 측정하기 위해서는 명확한 인과관계(cause-and-effect questions)에 기반을 둔 성과를 측정하는 접근이 필요하다. 이를 위해서는 가장 유사한 조건의 두 집단을 사업초기에 설정하고, 한 집단(실험군)에만 사업을 수행한 후, 사업을 수행하지 않은 집단(대조군)과 비교하여, ‘해당 사업(intervention)만으로 일어난 변화(영향)’를 분석하는 작업이 필요하다. 이를 영향평가(Impact Evaluation)라고 한다.

아래 [그림 1]의 A집단은 사업을 수행하는 대상 집단인 실험군이며, B집단은 A집단과 유사한 특성을 가지고 있으나, 사업이 수행되지 않은 집단이다. 사업이 종료된 후, A와 B집단 간의 차이가 곧 우리 사업의 영향(Impact)가 된다. 이때 B집단은 A집단의 counterfactual²⁾ 역할을 하게 된다.

[그림 1] 영향(Impact)의 정의



출처 : World Bank(2011), 저자 수정

3. 영향평가를 위한 방법론의 종류

영향평가 방법론은 크게 실험설계(experimental methods)와 준 실험설계(quasi-experimental methods)로 나눌 수 있다.³⁾

1) 실험설계(experimental methods)의 특성

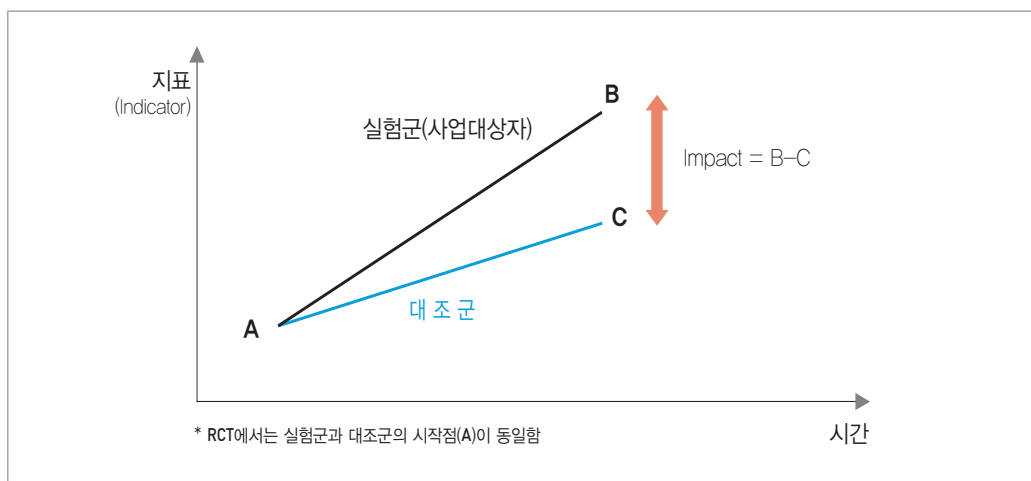
실험설계는 연구대상의 무작위로 추출(random selection)하고 배정(random assignment)

2) 결과에 대한 원인의 관계는 인과적인 연쇄(causal chain)에 의해 연결되어 있으며, 인과적 연쇄는 일종의 반사실 의존관계(counterfactual dependence)에 있다(안건훈 2005 : 134)

3) 흔히 사용되는 전-후 비교(before-after)는 non-experimental method에 속하며, 과학적 영향평가 방법론에 포함하지 않는다.

하여 동일한 실험군과 대조군을 구축하는 방식으로, 비교 대상(counterfactual)을 인위적으로 만든다. 이후 실험군만을 대상으로 사업을 시행하며, 사업이 종료되고 난 후, 대조군과의 비교를 통해 성과를 측정한다(그림 2 참조). RCT(Randomized Control Trials)는 대표적인 실험설계 방법이며, 사업의 영향을 평가하는데 있어 가장 엄격한 방식이며, 이상적인(gold standard)이라고 평가된다. 사업을 수행하는 대상지역과 (이론적으로) 동일한 대조군을 마련하여 그 변화를 측정하는 방식으로, 표본 선택 등으로 발생할 수 있는 바이어스를 무작위 추출을 통해 제거할 수 있다.

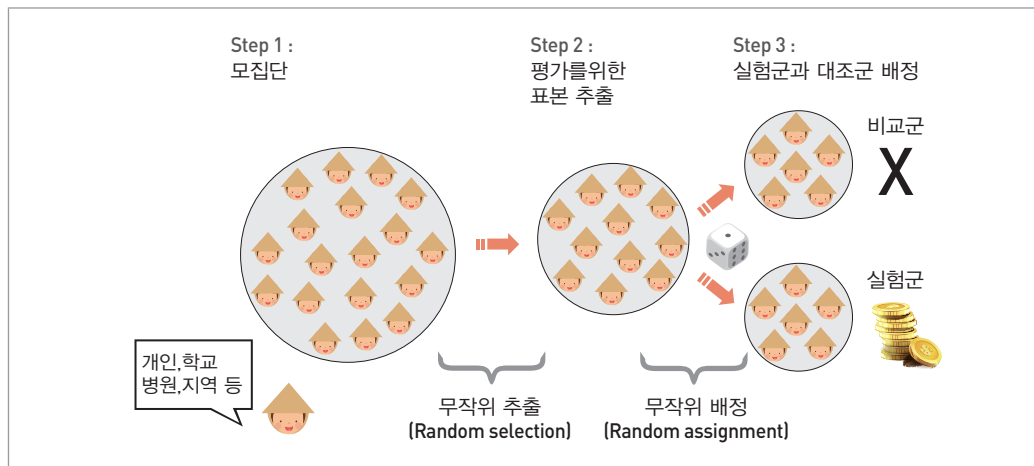
[그림 2] RCT의 영향(impact) 측정 방법



출처 : 저자작성

RCT에서는 특히 무작위 배정의 과정이 중요하며, 아래 [그림3]은 무작위 배정의 과정을 단계별로 보여준다. 첫 단계는 모집단에서 무작위로 표본을 추출(random selection)하는 단계이며, 이후 이를 실험군과 대조군으로 나누는 단계를 무작위배정(random assignment)라고 부른다.

[그림 3] 무작위 배정 단계



출처 : World Bank, 2011 : 57, 저자수정

RCT는 여러 장점을 가진다. ① 표본을 무작위추출을 함으로써, 평가하고자 하는 사업의 영향을 왜곡시킬 수 있는 바이어스(selection bias)를 제거할 수 있고, ② 눈에 보이지 않지만 사업의 성과에 영향을 줄 수 있는 요인들을 실험군과 대조군에 동등하게 나누어 흩뿌릴 수 있다. ③ 또한 사업을 시행하는 실험군을 선정하는 절차에 있어, (복권을 뽑듯이 무작위 추출로 선정이 되므로) 지역 주민의 의혹을 없앨 수 있으며, 비교군을 결정하는데 있어 시간을 절약할 수 있다(Michael, 2012). 예를 들어 A지역 지역주민 100명의 리스트가 있고, 이 중 50명을 뽑아 실험군에 배정한다면, 나머지 50명이 자동적으로 대조군이 되기 때문이다.⁴⁾

동시에 실험설계는 여러 가지 현실적인 단점도 가지고 있다. 첫째는 무작위추출을 시행함에 있어 상대적으로 시간과 비용, 노력이 많이 소요된다. 또한 윤리적인 문제가 발생할 수 있다. 예를 들어, 의약품 혹은 백신을 나누어 주는 사업이라 할 때, 대조군에는 제공하지 않는 상황이 윤리적으로 문제가 될 수 있다.⁵⁾

가장 큰 단점으로는 반드시 사업초기부터 설계가 되어야 한다는 점이다. 게다가 영향평가를 다수 수행하는 기관들은 사업 착수 전, 사전 테스트(pilot study)를 수행하고, 설계상에 문제

4) 반면, 준 실험설계 방법에서는 별도로 대조군을 선정하느라 시간을 소요하게 된다.

5) 이는 준 실험설계도 동일하게 가지는 단점이며, 실험군 집단에 대한 사업 종료 후 대조군에 동일한 사업을 진행하는 방법으로 보완할 수 있다.

가 없는지 등을 점검한 후 사업기획에 착수하기를 권고하고 있으며, 이를 감안하면, 실험설계를 통한 영향평가를 위해서는 실제 사업이 착수하기 이전 많은 노력이 요구된다.

실험설계의 또 다른 단점은 사업기간 동안 배정된 실험군과 대조군의 구성원 이탈(attrition)이 발생할 수 있다는 점이다. 가난한 계층, 교육을 덜 받은 계층, 인식개선·동기 부여가 덜 된 계층이 사업 참여 집단에서 이탈할 가능성이 높다고 알려져 있다(James C, 2006).

또한 평가할 프로그램을 선택할 때 ‘method-driven’이 될 가능성이 높으며, 무작위추출을 통해 선택에 의한 바이어스(selection bias) 통제가 가능하나, 이후의 타당도(validity)를 손상시킬 수 있는 것들에 대해서는 보장하지 않는다(Michael, 2012). 예를 들면, 사업이 수행되는 맥락에서의 특정한 변화를 찾아 보여주는 힘들며, 숨겨진 과정(black box)을 풀어내기는 힘들다.

실험설계는 보통 의약품 연구, 동물 행동연구, 심리학 연구 등에서 사용되는 방법으로 가장 이상적인 영향평가 방식이지만, 개발도상국을 대상으로 한 원조 사업 현장에서 사용하기에는 어려움이 많다. 실험실 수준의 통제를 위해서는 많은 사전 장치와, 위험관리 계획, 주기적인 모니터링 등의 많은 노력이 필요하다. 또한 앞에서 언급하였듯이, 실험설계는 상대적으로 비용이 많이 소요되는 경향이 있다.⁶⁾ 따라서 아래와 같은 상황과 같이, 필요한 때에 선별적으로 시행함이 필요하다.

Linda G(2016)는 아래와 같은 경우, 실험 설계(RCT) 적용을 추천하고 있다.

- 특정 사업(intervention)의 성과가 밝혀지지 않았거나 성과를 측정하여 홍보가 필요한 경우
- 향후 동일한 사업의 내용으로 사업 확대(scale-up)가 필요한 경우
- 혁신적인 기술 등을 모두에게 시도할 수 없기 때문에 테스트가 필요한 경우
- 평가 대상이 공간적으로 떨어져 있거나 일시적으로 고립되는 상황에 있는 경우
- 변화가 필요하나 해답은 잘 알지 못하는 경우
- 집단 중 일부는 대안에 대해 특별히 선호하는 바가 없고, 선정 과정에 있어 투명성이 요구되어서 추첨(lottery)방식이 필요한 경우

6) 다만 2차 자료(secondary data) 활용 등, 상황에 맞는 적절한 방법을 고안하여, 비용감소가 가능하다.

2) 준 실험설계(quasi-experimental methods)의 특성

준 실험설계(quasi-experimental design)⁷⁾ 방식은 무작위 배정에 의하여 실험군과 대조군을 만들 수 없을 때, 시행할 수 있는 방법이다. 실험설계와 동일한 점은 사업을 시행하는 실험군과 사업을 시행하지 않는 대조군을 정하고, 사업 종료 후 나타난 변화를 비교하는 방식이라는 점이다. 하지만 이때의 대조군은 무작위 추출 및 배정으로 선정되지 않은 대조군이라는 점이 실험설계와 차이가 있다. 이러한 이유로 준 실험설계에서의 대조군은 비교군(comparison group)이라고 흔히 부르며, 본 연구에서도 혼란을 막기 위해 비교군으로 칭하기로 한다.

준 실험설계는 실험설계에 비해 무작위 추출 및 배정으로 선정된 두 그룹의 비교가 아니라는 점에서 사업(intervention)의 순수한 성과(impact)를 측정하는 힘이 떨어진다는 단점이 있다. 하지만 동시에 현실적인 장점을 가지는데, 실험설계와는 달리, 반드시 사업 초기에 평가 설계를 하지 않아도 된다는 점이다⁸⁾—하지만 최소 사업초기의 기초조사(baseline survey)는 수행되어있어야 한다. 현실적인 제약사항으로 인해, 사업이 이미 시작된 상황이라 할지라도 준 실험설계 적용이 가능하다. 사업이 시작된 후일지라도, 타당한 방법으로 가능한 유사한 비교군⁹⁾(comparison group)을 만든다면, 영향 측정이 가능하기 때문이다. 단, 비교할 대상(counterfactual)을 얼마나 정교하게 구성하는지가 핵심이다.

따라서 평가자는 상황과 맥락(context), 사업의 특성에 따라 무엇을 비교할 대상(counterfactual)으로 가져갈 것인지 결정하게 되는데, 예를 들면, 해당 사업의 과거 데이터를 비교군으로 사용할 수 있다. 혹은 2차 자료(secondary data)를 이용하는 방법이 있으며, (사업이 이미 착수되었더라도) 사업지역과 유사한 지역을 찾아, 비교군으로 정하고 성과를 비교할 수 있다. 이는 특히 원조사업의 영향평가에 있어 크게 부각될 수 있는 장점이다. 개발도상국을 대상으로 하는 원조사업의 경우, 실험실과 같이 완벽한 통제가 더욱 쉽지 않고, 평가 설계 및 수행에 있어 많은 현실적인 제약이 동반되기 때문이다. 또한 준 실험 설계의 경우¹⁰⁾, 여러가지 방법들을 혼합하여 사용할 경우, 각자가 가지는 한계를 극복하고 더 강력한 방법론이 될 수 있다.

7) 유사실험설계라고도 부른다.

8) 하지만 준 실험설계 역시 가능한 사업 초기에 평가 설계에 착수하는 것이 좋다.

9) 유사하다는 의미는 사회경제적 요소, 기타 고려되어야 할 사항들이 통계적으로 유의미한 수준에서 거의 동일하다는 의미이다.

10) 예, DID와 방식과 matching을 혼합하는 경우

준 실험설계 방법 중 많이 사용되는 방식 중 하나는 이중 차감법(Difference in Difference, DID)이다. 사업 전후 비교(before&after) 방법은 사업 실시 전과 사업 실시 후의 변화만을 측정하는 방법임에 반해, DID는 사업 전후 시점과 사업시행유무(with program& without program)까지 포함하여 변화를 측정하는 방법이다. 즉 실험군과 비교군을 설정하고, 2개 그룹을 대상으로 사업시작전과 사업종료후의 변화량을 측정하여 영향을 찾아내는 방법이다(표 2 참조).

[표 2] DID의 성과 측정 방법(1)

구 분	Before (사업실시전)	After (성과를 측정하고자 하는 시점)
With program (실험군 : 사업시행)	A	B
Without program (비교군 : 사업시행없음)	C	D

※ 성과(Impact) = (B-A)-(D-C)

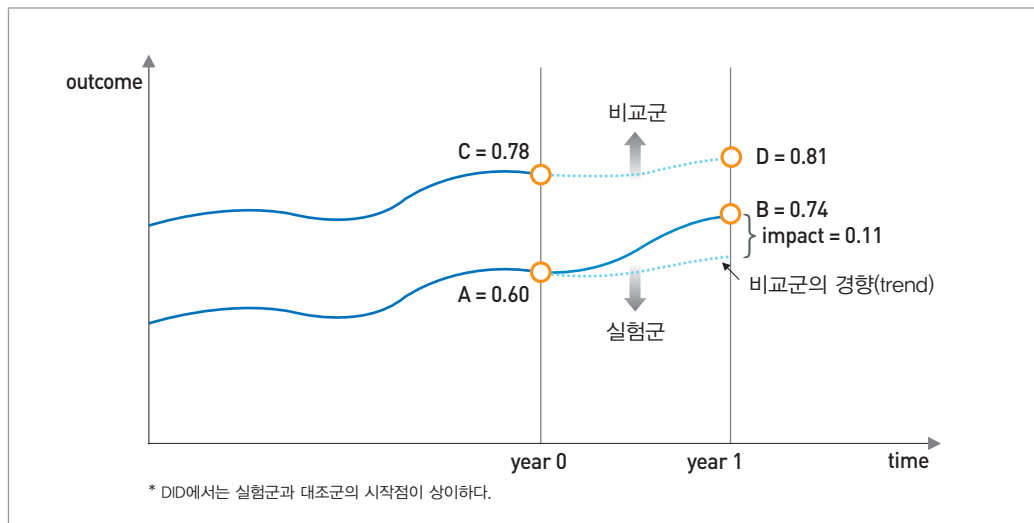
출처 : 저자작성

[표 2]를 좀 더 구체적으로 표현하면, [그림4]와 같다. 특정 지역에 식수개발사업을 수행하여 지역의 깨끗한 물 접근성을 향상시키는 사업이었다고 가정하기로 하자. 사업 시작 전(year 0)과 성과를 측정하고자 하는 시점¹¹⁾(year 1)을 기준으로 시점비교를 함과 동시에, 실험군과 비교군을 비교하는 방법으로 성과를 측정한다. 이때 중요한 전제는 ‘실험군 지역도 사업이 시행되지 않았다면 비교군과 동일하게 변화(equal trend)한다’는 것이다¹²⁾.

11) 사업의 성과를 측정하고자 하는 시점은, 사업수행 중간 시점(중간평가) 일수도 있으며, 종료이후 시점(사후평가)일 수도 있다.

12) equal trend의 타당도를 검증하는 방법이 여러 가지 있으며, 첫째는 사업 전에 실험군과 대조군의 수치(outcomes)를 조사해보는 것, 두 번째는 placebo test로, 실험군과 대조군 이외에 제 3의 지역을 정하고, 대조군과 제3의 지역을 다시 한 번 비교해 보는 방법, 셋째 사업시행 전 최소 1개의 pre-intervention 관측치를 두고 비교해 보는 방법 등이 있다(World Bank, 2011)

[그림 4] DID의 성과측정 방법(2)



출처 : World Bank, 2011 : 97

따라서 [그림 4]에서의 영향(impact)을 구하는 방법은 실험군에서 나타난 변화에서 사업을 수행하지 않았더라도 나타났을 변화(같은 기간 동안의 비교군의 변화량)를 빼주는 방식이다. 즉, $(B-A)-(D-C) = (0.74-0.60)-(0.81-0.78)=0.11$ 가 이 사업의 영향(impact)가 된다.

4. 영향평가 윤리 기준

과학적인 방법론을 기반으로 한 영향평가는 다양한 윤리기준 준수를 원칙으로 한다. 엄격한 방법론을 적용함에 있어, 여러 윤리적인 문제가 발생할 수 있기 때문이다. 따라서 사업 기획 단계에서 윤리적인 문제 발생을 막기 위한 장치를 고려하고, 사업 이행단계에서 꾸준한 모니터링이 필요하다.

World Bank(2011)은 과학적인 방법론을 기반으로 한 영향평가는 기본적으로 ‘해를 끼치지 않아야 한다(Do No Harm)’는 점을 강조하고 있다. 즉, 평가로 인한 직간접적인 피해가 없어야 하고, 사전에 내용을 평가 대상지 주민들에게 설명하는 등 잠재 위험에 대한 동의를 얻어야 한다는 의미이다. 또한 프로젝트 자체가 해로워서는 안되며¹³⁾, 투명하지 않거나 공정하지 않은 방식

으로 평가를 수행함으로써, 수혜를 못 받는 계층이 생기지 않도록 해야 함을 강조하고 있다.

SAGE(2016)는 데이터 수집, 처리 및 보관상에 실험대상자 보호 필요성을 강조하고 있으며, 이는 공개된 데이터 상의 익명성이 충분하지 않아 참가자에게 피해가 갈 수 있기 때문이다.

5. 타 원조기관의 영향평가 사례

앞에서 살펴 본, 실험설계 및 준 실험설계 방법론 중, 가장 대표적인 방법론인 RCT와 DID를 적용한 타 원조기관의 사례를 살펴보고자 한다.

1) 실험설계(RCT) 사례 :

JPAL의 인도네시아 의료보험가입요인 연구사업

(1) 사업 및 영향평가 개요

J-PAL(The Abdul Latif Jameel Poverty Action Lab)은 미국 MIT 내 RCT를 통한 개발협력정책 및 사업 평가, 연구 및 정책제언을 수행하는 기관이다. JPAL은 개발도상국에서 RCT를 통한 영향평가를 통해 해당 개발도상국 정부에 정책제언을 하거나, 특정 사업(intervention)의 효과를 RCT로 증명하여, 공유하는 활동을 펴고 있다. KOICA는 JPAL의 인도네시아 의료보험가입요인 분석 연구 사업을 지원(2015.10~2017.4)하고 있으며, 동 사례분석에서 RCT의 사례로 살펴보고자 한다.

인도네시아는 노동인구의 60%가 비공식 근로자(informal worker)이며 이중 60%가 의료보험을 보유하고 있지 못하다. 이에 인도네시아 정부는 2019년까지 재정적으로 지속가능한 범국민 의료보험을 마련하기 위해 노력 중이며, 이를 위해서는 의료보험 등록율과 보유율을 높이기 위한 효과적인 방안을 강구하여야 하는 과제를 안고 있다. JPAL은 RCT기법을 이용하여, 인도네시아 의료보험 가입 요인을 분석하고, 이를 통해 인도네시아 정부가 재정적 지속가능성이 확보된 의료보험 정책을 마련하는데 도움이 되고자 동 연구 사업을 추진하였다¹⁴⁾. 구체적으로는 아래의 질문(research question)의 답을 찾는 것이 연구 사업의 목적이라 볼 수 있다.

13) 세이프가드 등을 통해 사전에 위험요소를 고려하여 추진할 필요가 있다.

14) 한국 KOICA와 호주 정부의 자금지원으로 수행되고 있다.

- 인도네시아 국민¹⁵⁾의 의료보험 가입율을 높일 수 있는 요인은 무엇인가?

사업의 예산은 약 3억원이며, KOICA와 호주 정부가 공동으로 지원하고 있다. JPAL은 본격적으로 연구 사업을 시작하기 전에 시범연구사업(pilot study)를 시행하였다. 앞에서 언급한 바와 같이 RCT 수행에는 많은 노력과 시간과 비용이 소요되며, 실제 수행 중 발생할 수 있는 여러 위험요소들이 있으므로, 본 사업 착수 전, 동일한 내용이되 단기간·소규모로 진행하는 pilot study를 수행해 봄으로써, RCT 설계상의 문제점 등을 파악하여 본 사업에 반영할 수 있다(표3 참조).

[표 3] JPAL의 인도네시아 의료보험 가입요인분석 연구사업 요약

구 분	내 용
사업예산	약 3억원(KOICA와 호주정부 공동 자금지원)
사업대상지	인도네시아 Medan 및 Bandung지역 중 무작위 추출 지역
사업기간	2015.10-2017.4(1년 6개월) * Pilot Study는 2014년 수행
사업목적	RCT방식을 활용하여, 인도네시아 국민(특히, 자영업계층)의 의료보험 가입율을 높일 수 있는 요인 분석 * 연구질문 : 어떠한 종류의 요인이 빈곤층에 속하지 않지만, 비공식 근로자(자영업자 등)인 사람들이 의료보험을 등록하도록 만드는가?
사업수행방식	미국 MIT 교수진으로 구성된 JPAL소속 설계팀이 주로 RCT디자인 및 분석을 담당하며, 이후 JPAL 인도네시아 사무소에서 사업을 총괄 진행. 사업대상지 별 기초조사(설문지 조사 등) 등은 인도네시아 현지 기관에게 위탁용역하여 수행
특이사항	본격적으로 연구사업 시작 전, Pilot study를 JPAL 자금으로 수행하며, 도출된 결과를 본 연구사업에 반영

출처 : 저자작성

15) 빈곤층에 속하지 않고, 비공식근로자(자영업자 등)인 사람들을 대상으로 한다.

(2) 영향평가 설계

① 시범 연구사업(Pilot Study)

JPAL은 본 사업 대상지에 사업을 실시하기 전에, 2014년 11월, 인도네시아 Gresik, East Java 내 소지역인 Sidayu에 속한 마을 중, 무작위로 7개의 마을을 선정하여 시범사업(pilot study)을 진행하였다. 7개의 마을 중 4개의 마을에 '보조금 지원(바우처 제공)' 활동을 진행하고, 나머지 3개의 마을에서는 '의료보험 정보 제공'과 '가정방문을 통한 온라인 등록' 활동을 각기 진행하였다.

Pilot study 결과, 미미한 등록을 상수를 확인하였다. 가정방문을 통한 온라인 등록 활동을 통해 보험에 등록한 이는 한 명도 없었고, 4개 가구만이 보조금을 수령하였다. 낮은 등록율의 원인으로는 가족 내 모든 구성원을 가입시켜야 한다는 인식에 기인한 부담감과 지역 내 수도에 위치한 BPJS(인도네시아 사회보험 관리기관) 사무실까지 이동하여 등록을 진행해야 하는 번거로움, 은행 계좌와 같은 행정적 의무사항 등으로 파악되었다. 게다가, 정보 제공(캠페인 포함) 역시 마을주민 전반에 중요 메시지를 전달하는 데 있어 실패하였다. 위의 결과는 본 사업을 설계 하는데 있어 주요 고려사항 및 시사점으로 활용되었다.

② 본 연구사업(인도네시아 의료보험가입요인 분석 연구사업)

이후 JPAL은 본 사업에 착수하였으며, 인도네시아 Medan지역의 12개 마을, Bandung지역의 32개 마을을 무작위로 추출하고 아래와 같은 각기 서로 다른 활동을 수행하고, 해당 지역의 의료보험 가입율을 비교함으로써 의료보험 가입의 주요 요인을 파악하고자 하였다. 무작위 추출로 선정된 실험군에 대해서는 문헌조사와 pilot study를 통해 선정한 의료보험 가입에 영향을 주는 요인별로 변화를 주어 사업을 시행하였다. 보조금, 접수 장소, 의료보험 사용가능 대기 기간, BPJS의무화 정보 제공과 관련하여, 서로 다른 경우(variation)를 적용하였으며, 상세내용은 아래 [표4]와 같다.

[표 4] JPAL의 인도네시아 의료보험 가입요인분석 연구 요약

구 분	내 용
보조금 편차 (Subsidy variation)	보조금은 바우처로 제공하며, 보조금을 제공하지 않는 경우(no subsidy), 1+1 보조금 지원(a buy-one-get-one-free subsidy), 전액 지원으로 구분(바우처의 유효기간은 12개월임)
접수 장소 (Registration location)	온라인을 통한 현장 등록(onsite registration)과 BPJS 사무실에서의 등록(standard registration)으로 구분
의료보험 등록 후 이용가능할때까지의 대기 시간 (waiting period information variation)	의료보험 등록 후 2주 후에 병원 이용시 보험 적용을 받을 수 있다는 정보를 받은 경우와 받지 않은 경우로 구분
BPJS 정보 (Mandatory BPJS Information variation)	인도네시아 정부의 BPJS 가입을 해야 운전면허증, 여권 발급등의 행정서비스를 이용할 수 있다는 정책이 발효될 예정이라는 정보를 미리 공지한 경우와 그렇지 않은 경우로 구분

출처 : 저자작성

위의 서로 다른 경우의 수(variation)를 무작위 추출로 선정된 서로 다른 실험군(지역)에 적용하고, 각 해당 지역의 의료보험 가입률을 조사하는 방식으로 의료보험 가입 촉진 요인을 분석하였으며, 그 결과는 아래와 같다.

[표 5] 실험군 별 사업내용

	Subsidy Treatment				Registration Treatment		Waiting Period Treatment		Mandate Treatment	
	No-subsidy	Buy-1-get-1-free	Full subsidy	No-inpatient bonus	standard	Online	Standard	14-day	No Treatment	Mandate Info
Assigned Treatment	2785	1157	568	1149	2840	2819	2833	2826	2850	2809
Actual Enrollment	47	118	116	61	115	227	158	184	181	161
Actual Enrollment (%)	1.7%	10.2%	20.4%	5.3%	4.0%	81%	5.6%	6.5%	6.4%	5.7%

출처 : JPAL(2016)

(3) 소결

동 연구 사업은 2016년 12월 현재 진행 중인 사업으로 아직 최종 결과가 도출되기 전이므로, 현재 진행된 내용만 동 연구에 포함하였으며, 그럼에도 불구하고 영향평가를 고려한 원조 사업기획 시, 고려해야 할 몇 가지 시사점을 도출할 수 있었다. 첫째는 RCT에 드는 시간과 노력, 비용이 상대적으로 크기 때문에, 실패의 위험률을 줄이기 위해서는 pilot study가 필요하다는 점이다. 이를 통해, 평가 설계상의 문제점은 없는지, 이탈(attrition) 등의 위험(risk)을 관리할 수 있는 방법 등을 사전에 점검하고 본 사업의 사업기획 시 반영하여야 한다. 두 번째는 사업기획 단계에서 사업의 성과 혹은 영향을 점검하기 위한 명확한 질문(혹은 연구 질문)과 논리모형이 제시되어야 한다. 결국 사업의 성과를 측정할 때, 어떠한 지표를 사용할 것인지 등은 최초의 질문과 사업의 논리모형에 근거하는 것이므로, 사업기획 단계에서 이 두 가지를 결정하는 데에 충분한 시간과 노력을 투자하여야 한다.

2) 준 실험설계(Difference in Difference, DID) 사례 :

일본 JICA의 식수개발사업 영향평가¹⁶⁾

(1) 사업 및 영향평가 개요

본 사업은 일본 JICA가 2011.6월부터 2013.5월까지 약 70억의 예산으로 수행한 “잠비아 식수개발 사업(Rural Groundwater Development Project in Zambia)”사업이다. 사업구성요소는 1) 관우물(borehole)¹⁷⁾과 핸드펌프¹⁸⁾를 제공하는 하드웨어형 지원과 2) 시설유지 및 운영(정부기관과의 협의, 유지비용 납부 관리 등), 조직 관리를 위한 역량강화, 3)수처리, 손씻기, 화장실 사용등에 대한 인식개선 및 위생교육을 포함하는 소프트웨어형 지원으로 구성된다.

평가 종류는 사후평가로, 2014.4월에 평가를 수행하였다. 평가수행자는 JICA 내부 3명(평가부서)과 외부전문가 2명으로 구성되었으며, 평가 방법론은 준 실험설계 중, DID를 사용하였다. 평가 배경은 JICA는 무유상원조를 모두 수행하는 통합기관으로, 그 동안 무상원조에 대해서는 영향평가를 해오지 않았으며, 사업의 성과 측정에 있어 사업전후비교(before/after) 혹은 with program/without program 기준으로만 성과를 제시하였다. 하지만 이러한 방식은 인

16) 출처 : https://www.jica.go.jp/english/our_work/evaluation/tech_and_grant/impact/c8h0vm00006eh46d-att/zambia_en.pdf

17) 길이 30m 이상의 관을 삽입한 우물임

18) 1일에 30리터의 양으로 250명에게 안전한 식수를 제공할 수 있는 시설임

과관계(causal effect)를 측정할 수 없기 때문에, 영향평가가 필요하며, 무상원조(grant aid)로는 최초로 동 사업을 과학적인 방법론을 기반으로 평가하였다.

(2) 영향평가 설계

사업대상지는 Luapula지역의 3개 District(Milenge, Mwense, Nchelenge)를 대상¹⁹⁾으로 하며, 먼저 기초조사 시점에 Milenge District 중 사업이 수행된 마을 리스트에서 랜덤하게 14개 지역을 선정하고 이들을 실험군으로 정하였다. 그리고 사업대상지역과 유사한 특성을 가지면서, 사업(intervention)이 없었던 12개 지역을 비교군으로 정하였다.

같은 방식으로 Mwense District에서 19개 실험군과 17개 비교군을 선정하고, Nchelenge District에서는 17개 실험군과 15개 비교군을 선정하였다. 하지만 사업이 들어갔지만, 새로 설치된 관우물에서 물이 나오지 않는 등의 실패가 생기기도 했기 때문에, 당초 실험군이었던 곳이 비교군으로 바뀌고, 대신 비교군이었던 집단이 실험군으로 편입되는 등 변동이 불가피하였음을 밝히고 있다(JICA, 2014).

이에 따라, 최종적으로 1) Milenge에는 21개 사업대상지(실험군), 5개 비교군, 2) Mwense에서는 25개 사업대상지, 11개 비교군, 3)Nchelenge에는 18개 사업대상지, 14개 비교군으로 집계되었다. 아래 [표 6]은 각 District별 사업대상지 및 비교군의 수와 샘플 가구수를 표로 정리한 것이다

[표 6] District별 사업대상지 및 비교군의 수와 가구수

District	No. of Sample Sites			No. of Sample Households		
	All	Project	Control	All	Project	Control
Milenge	26	21	5	185	150	35
Mwense	36	25	11	236	156	80
Nchelenge	32	18	14	214	128	86
Total	94	64	30	635	434	201

출처 : JICA(2014)

19) Mause District는 기초조사 전에 사설자원사업 지원을 시작하여서, 대상 site에서 제외됨

동 사업의 영향평가 보고서에 따르면, 표본 가구들 중 이탈(attrition)이 일어난 현상을 명시하고 있고, 분석 결과 가족인원이 적은 가구일 경우 이탈이 심한 경향이 있다고 언급하고 있으며, 따라서 평가 방법론 설계 시, 적은 인원의 가족인 경우, 프로젝트의 인과관계(causal effect)를 설명하기 어려울 수 있다는 점도 제시하였다.

비교군 설정 시, 실험군(사업대상지)과 유사한 특성을 가진 비교군을 설정해야 하기 때문에, 실험군과 비교군이 유사특성을 가지고 있는지, 두 집단에 대한 인구사회학적 특성(socio-economic characteristic)을 아래 [표 7]과 같이 확인하였다. 실험설계 혹은 준 실험설계 모두 ‘실험군과 대조군 · 비교군이 유사한 특성을 가진 동일한 집단’이라는 점이 전제가 되며, 이를 확인하기 위하여, 주로 두 집단의 인구사회학적 특성이 유사한지를 조사하고, 통계적으로 유의미한지 확인하는 절차를 거친다.

[표 7] 실험군 - 비교군 간 특성 비교

	All (A)	Project (B)	Control (C)	Diff. (B)-(C)
Individual characteristics (no. of observations)	3303	2268	1035	
Female (=1)	0.509	0.513	0.500	0.012
Age	20.85	20.92	20.69	0.23
Age group [10-] (=1)	0.328	0.315	0.357	-0.041**
Age group [10+& 20-] (=1)	0.253	0.269	0.219	0.049***
Age group [20+& 30-] (=1)	0.138	0.135	0.146	-0.011
Age group [30+& 40-] (=1)	0.111	0.112	0.109	0.003
Age group [40+& 50-] (=1)	0.091	0.094	0.084	0.010
Age group [50+& 60-] (=1)	0.041	0.037	0.050	-0.013
Age group [60+] (=1)	0.037	0.037	0.035	0.003
Education level	3.875	3.977	3.650	0.327***
Household characteristics (no. of observations)	635	434	201	
female headed household (=1)	0.206	0.207	0.204	0.003
Age of the household head	43.06	43.43	42.25	1.18
Education level of male members	6.793	6.889	6.576	0.314
Education level of female members	5.170	5.368	4.741	0.628**

Household size	5.202	5.226	5.149	0.077
Ratio of dependents to household size	0.452	0.453	0.451	0.002
Monthly consumption per capita [ZMK]	174843	163219	199940	-36721
Village characteristics (no. of observations)	94	64	30	
Population	414	446	346	100

Note: t-test results are shown; * significant at 10%; ** significant at 5%; *** significant at 1%.

출처 : JICA(2014)

(3) 영향평가 결과

DID(Difference in Difference)방법은 준 실험설계의 대표적인 방법으로, 사업 전후(before/after), 사업이 수행된 경우와 수행되지 않은 경우(with program/without program)를 동시에 포함하여, 영향을 확인하는 방법이다. 따라서 앞의 과정을 통해 실험군과 유사한 비교군을 설정한 이후, 실험군 및 비교군에서 동일한 항목으로 대해 기초조사를 시행하게 된다. 이후, 사업(intervention)이 종료된 후, 결과값을 측정하고, 회귀분석(필요시 모델링을 기반으로 한 다중회귀분석)을 통해, 실험군 및 대조군의 경향을 측정하며, 이를 비교한다. 동 사업은 △가정에 저장된 물에서의 대장균 검출 여부 △ 설사 증상 발생률 △ 5세 이하 영아를 대상으로 한 설사 발생률, △ water source 접근성, △ 18세 이상 여성들 대상 물을 길러 오는데 걸리는 시간 등을 주요 지표로 두고 있으며, DID 방법으로 아래와 같이 성과를 측정하였다.

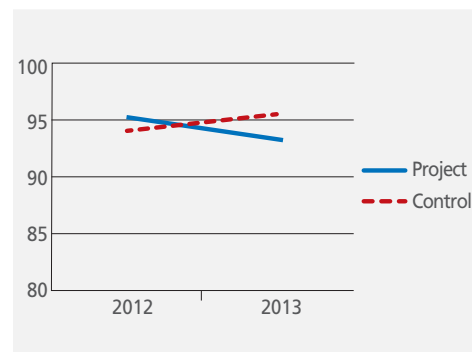
[표 8] 가정에 저장된 물에서의 대장균 검출 여부

Year	Project		Control		Diff. (B)-(D)
	(A)	(B)	(C)	(D)	
	Obs.	Mean	Obs.	Mean	
2012	434	95.2	201	94.0	1.1
2013	428	93.2	197	95.4	-2.2
Change		-1.9		1.4	-3.3

Note: Statistical test results are shown; * significant at 10%;

** significant at 5%; *** significant at 1%.

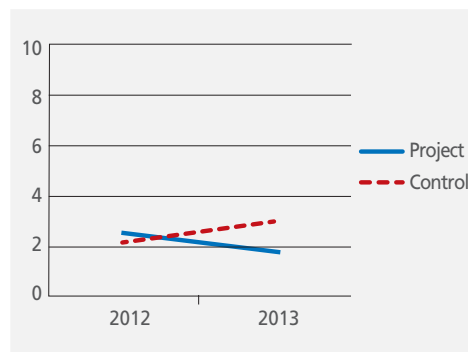
출처 : JICA(2014)



사업대상지에서의 대장균 검출은 사업 시작 후 1년이 지난 시기에, 1.9가 감소하였고, 비교군에서는 1.4가 증가하였다. 하지만, difference값인 1.1과 -2.2는 통계적으로 유의하지 않았다²⁰⁾. 동 사업의 영향 평가 보고서에서는 이 이유를 같은 시기, 우물의 관에서는 대장균이 검출되지 않았으나, 가정의 저장고에서는 대장균이 검출되었으며, 따라서 물이 이동하는 과정에서 혹은 저장고에서 오염되었을 가능성을 제시하였다. 반면, 설사증상 발생률은 [표 9]에서와 같이, 사업시행 후, 1.7%를 낮추었으며, 이는 10% 범위에서 통계적으로 유의미한 결과를 제시하고 있다.

[표 9] 설사증상 발생률

Year	Project		Control		Diff. (B)-(D)
	(A)	(B)	(C)	(D)	
	Obs.	Mean	Obs.	Mean	
2012	2268	2.5	1035	2.2	0.3
2013	2155	1.7	989	3.0	-1.4**
Change		-0.8		0.8	-1.7*



Note: Statistical test results are shown; * significant at 10%;
** significant at 5%; *** significant at 1%.

출처 : JICA(2014)

이 외에도, 5세 이하 영아를 대상으로 한 설사 발생률, water source 접근성, 18세 이상 여성들 대상 물을 길러 오는데 걸리는 시간 등에 대해 DID로 그 효과를 검증하였다. 결과는 설사증상 발생정도 감소는 유의미하였으며, 특히 5세 이하 영아를 대상으로 했을 때, 영향의 정도가 더 컸다. 또한 수자원 접근성 증가에도 유의미하게 영향을 주었다. 물 길러오는 시간 감소는 의미 있는 영향이 발견되지 않았다

20) 통계적으로 유의미한지의 여부는 해당 값의 * 유무로 확인할 수 있으며, 그림2와 그림3의 표 아래에 제시된 바와 같이, *는 10% 범위에서 유의, **인 경우, 5%, ***인 경우 1%범위에서 유의미하다는 의미이다.

(4) 소결

JICA는 동 사업 영향평가 보고서를 통해, 현실 세계에서의 시간, 예산 등의 제약으로 RCT로 평가를 진행하지 못했음을 한계로 지적하고 있다. 무작위 추출 방식으로 실험군과 비교군을 정하기 어려운 상황이었기 때문이었다. 두 번째로는 동 평가의 영역에서 벗어나는 부분이긴 하지만, “왜” intervention이 효과를 내지 못했는지에 대한 이유와 매커니즘을 밝혀내지 못했다는 점을 한계로 제시하고 있다. 더불어 영향평가를 통한 명확한 인과관계(causal effect)를 측정하는 것은 중요하고, 이때 적절한 평가 방법론이 사용되어야 한다는 점을 강조하고 있으며, 이러한 영향평가를 위해서는 충분한 자원(인력과 예산 등)이 필요하다는 점 또한 강조하고 있다. 마지막으로 이러한 엄격한 평가를 하기 위해서는 반드시 사업 초기부터 기획되어야 한다는 교훈을 제시하고 있다.

동 영향 평가는 원조시행기관이 해당 기관의 사업의 영향을 준실험 설계 방법론을 기반으로, 기존의 before/after 비교 측정방식에서 벗어나 사업의 성과를 엄격하게 측정한 사례이다. 동 사례에서 알 수 있듯이 RCT 적용이 힘든 경우(혹은 RCT 적용이 필요 없으나, 인과관계에 기인한 성과 측정이 필요한 경우), 준 실험설계를 통해 RCT에 준하는 인과관계에 기반한 성과를 측정할 수 있다. 실험군(사업대상지)과 유사한 비교군을 찾아, 동질성을 입증 한 다음, 변화량을 측정하는 방법이기 때문이다. 완벽한 실험설계 설정이 어려운 개발도상국 원조사업의 특성을 감안하면, 실험설계보다 현실적인 평가방법일 수 있다.

DID 등의 준 실험설계 방법은 RCT와 같이 반드시 사업초기에 설계가 되어야 하는 것은 아니지만, 가능한 사업초기부터 설계 하는 것이 필요하다. 동 사례의 경우처럼 사업초기부터 해당 사업지(District)에서 실험군과 비교군을 설정하고, 이후 사업이 종료된 후, 설계에 근거하여 성과 측정할 수 있도록 모니터링하고 관리하는 작업이 필요하다. JICA 사례에서도 알 수 있듯이, 표본의 이탈(attrition)은 매우 흔히 일어나는 상황으로, 이에 대한 고려도 필요하다.

또한 DID 방법 역시, RCT와 마찬가지로 ‘왜?’라는 질문에 대한 답을 제시할 수 있는 과정을 밝히는 데에는 한계가 있으며, 이를 위해서는 별도의 과정평가(process evaluation)이 필요하다.

6. 시사점 및 제언

실험설계와 준 실험설계는 각각의 뚜렷한 장·단점을 가지며, 따라서 평가(evaluation) 설계 시기, 활용·수집 가능한 자료의 수준, 발주자(client)의 의견 및 정치적 결정 등을 토대로 판단하여 적절한 방법론을 적용하여야 한다. 또한 RCT로 대표되는 실험설계가 이상적인 기준(gold standard)로 불리는 방법론 이긴 하지만, 현실적 한계 및 여건을 고려하면 준 실험설계(quasi-experimental design)를 택하고, 보완 할 수 있는 다양한 방법들을 활용하여 설명력을 높이는 방법도 좋은 대안이 될 수 있다.

국제원조사업의 경우, 평가 비용이 점점 높아지는 추세로, 상대적으로 원조에 들어가는 비용이 줄어들게 되므로, 이를 우려하는 목소리도 있다. 하지만, 장기적 관점에서 보았을 때, 평가를 기반으로 얻은 결과 및 교훈을 기반으로 증거기반 활동(evidence-based activity)을 하는 등의 환류체계가 구축되고, 향후 타 사업에 교훈을 적용하여 성과를 높일 수 있다면 사회적으로도 이익을 가져올 수 있는 활동이 될 수 있다.

따라서 평가자는 여러 가지 방법론의 장단점과 특성을 잘 이해하고 비용대비 효과가 높은 평가를 설계할 수 있어야 하며, 평가 결과가 어떻게 활용될 수 있을지, 어떠한 긍정적인 효과를 만들어낼지를 고민하고, 비판을 위한 비판이 아닌, 생산적인 결과가 도출될 수 있도록 고민하는 태도를 견지하는 것이 필요하다.

참고문헌

▶ 국내 문헌

강종려, 개발협력사업에서 영향평가 : KOICA 영향평가 설계 사례를 중심으로, KOICA 국제개발협력지, 2010

Linda G. Morra Imas와 3인, The Road to Results, World Bank, 2009, KOICA, 개발협력프로그램평가의 설계와 실행, 한울, 2016

안건훈, 인과성분석. 서울대학교출판부, 2005

▶ 국외 문헌

World Bank. *Impact Evaluation in Practice*, Washington, 2011

James C. MCDavid, *Program Evaluation and Performance Measurement*, SAGE, 2006

JICA, *Basic Research for Impact Evaluation Trial on Grant Aid: A Case from Rural Groundwater Development Project in Zambia*, 2014

JPAL, *Implementation and Preliminary Data Analysis Report*, 2016

Michael Bamberger, *RealWorld Evaluation(Working Under Budget, Time, Data, and Political Constraints)*, SAGE, 2012

Rachel Glennerster, *Running randomized evaluation*, Princeton, 2013



연구자료

연구개발 2016-25-212

『개발과 이슈』는 국제개발협력분야의 주요 현안에 대한 분석과 최근 연구 및 정책을 소개하여 국내외 국제개발협력에 대한 인식을 제고하고 관련 정책 수립과 사업 추진에 방향을 제시하기 위해 발간되는 비정기간행물입니다. 본 연구 보고서의 내용은 출처와 집필자 명시 하에 인용이 가능합니다. 본 연구의 내용은 한국국제협력단의 공식적인 견해와 일치하지 않을 수도 있습니다.

※ 본 자료는 KOICA ODA전자도서관(lib.koica.go.kr)을 통해서도 볼 수 있습니다.