

연구보고서(수시) 2016-02

ODA 평가체계 연구

- 보건복지분야 사업평가의 시사점



김현경 · 이현주 · 전진아 · 이소영 · 박복영 · 김부열 · 송민영

【책임연구자】

김현경 한국보건사회연구원 부연구위원

【주요 저서】

시간제 일자리 확산이 소득불평등과 빈곤에 미치는 영향
한국보건사회연구원, 2015(공저)

보건복지분야 ODA 활성화 방안 연구
한국보건사회연구원, 2015(공저)

【공동연구진】

이현주 한국보건사회연구원 연구위원

전진아 한국보건사회연구원 부연구위원

이소영 한국보건사회연구원 부연구위원

박복영 경희대학교 국제대학원 교수

김부열 KDI 국제정책대학원 조교수

송민영 한국보건사회연구원 전문연구원

연구보고서(수시) 2016-02

ODA 평가체계 연구

- 보건복지분야 사업평가의 시사점

발 행 일 2016년 12월 31일

저 자 김 현 경

발 행 인 김 상 호

발 행 처 한국보건사회연구원

주 소 [30147]세종특별자치시 시청대로 370
세종국책연구단지 사회정책동(1~5층)

전 화 대표전화: 044)287-8000

홈페이지 <http://www.kihasa.re.kr>

등 록 1994년 7월 1일 (제8-142호)

인 쇄 처 고려씨앤피

정 가 7,000원

© 한국보건사회연구원 2016
ISBN 978-89-6827-345-2 93330

국제협력사업이 효과 있게 진행되어야 함은 이론의 여지가 없다. 그럼에도 원조가 수원국의 실정과 정책목표에 부합하도록 진행되어 결과적으로 좋은 성과를 내었는가를 평가하는 작업에 실질적으로 관심을 갖기 시작한 것은 비교적 최근이다.

한국 또한 2010년 경제협력개발기구 개발원조위원회(OECD DAC) 가입과 더불어 원조의 규모가 급증했음에도 원조사업에 대한 모니터링과 평가는 예산 규모와 빈도, 환류 차원에서 충분하게 이뤄지고 있지 못하다. 이러한 평가의 미비함은 국제협력사업이 활발하게 이뤄지고 있는 보건분야에서도 별반 다르지 않다.

한국보건사회연구원은 보건 및 사회정책 분야의 공공정책을 수립하고 평가해 온 오랜 경험을 바탕으로 해당 분야의 국제협력사업 평가의 발전 방향에 대해 고민하고 개선점을 제안할 필요성을 느낀다. 따라서 발전된 평가체계와 보건 및 사회정책 분야 평가사례를 통해 한국의 평가 관련 쟁점과 발전 방향을 찾아보고자 했다.

본 연구는 우리 연구원의 김현경 부연구위원의 책임하에 이현주 연구위원, 전진아 부연구위원, 이소영 부연구위원, 송민영 전문연구원, 그리고 경희대학교 박복영 교수, KDI 정책대학원 김부열 교수가 참여하여 수행되었다. 모든 연구진의 노고에 감사드리며, 아울러 보고서 작성과 관련하여 여러 가지 중요한 조언을 아끼지 않은 자문위원분들과 원내의 검독위원들께 감사의 말씀을 전한다.

마지막으로, 본 연구가 개별 국제협력사업과 각 분야 정책의 효과뿐만

아니라 수원국의 보건과 복지를 포함하는 전체 사회보장의 향상을 위한
평가 개선에 기여할 수 있기를 기대한다.

2016년 11월

한국보건사회연구원 원장

김 상 호

목 차

Abstract	1
요 약	3
제1장 서 론	7
제1절 연구의 필요성	9
제2절 연구의 목적	10
제3절 주요 연구내용	11
제4절 연구방법	12
제2장 국제개발협력사업 평가의 목적과 방법	15
제1절 개발사업 평가의 목적과 방법	17
제2절 영향평가	30
제3장 주요 공여국 평가체계와 사례: 보건의료 및 사회정책분야를 중심으로	67
제1절 세계보건기구(WHO)	69
제2절 영국	85
제3절 일본	108
제4장 한국의 공적개발원조(ODA) 사업평가	121
제1절 한국 ODA 평가체계와 현황	123
제2절 보건의료분야 평가사례	137

제5장 한국의 개발협력사업 평가 개선방안 157

제1절 한국의 개발협력사업 평가체계 개선방안 159

제2절 보건복지분야 기획과 성과관리 개선방안 163

참고문헌 177

부록 183

부록 1. 영향평가의 방법들 183

부록 2. 지속가능한 개발목표(SDGs)의 보건복지분야 평가목표와 지표 192

표 목차

〈표 2-1〉 모니터링과 평가의 차이점	18
〈표 2-2〉 논리모형의 주요 구성요소	24
〈표 2-3〉 사업평가와 영향평가 비교	31
〈표 2-4〉 샘플 대표성 평가	50
〈표 2-5〉 여학생 장학금 프로그램 연구설계	51
〈표 2-6〉 말라위 남학생 포경수술사업 연구설계	54
〈표 2-7〉 준비조사를 통한 설문조사요원 교육	56
〈표 2-8〉 기초선조사 변수에서의 밸런스	61
〈표 2-9〉 종료조사 비응답 여부 밸런스	62
〈표 2-10〉 학교 출석 ITT 분석	64
〈표 2-11〉 사전적 생각에 따른 Heterogeneous 효과	66
〈표 3-1〉 캄보디아 MDG 보건지표	78
〈표 3-2〉 캄보디아 보건 증진에의 저해 요소	79
〈표 3-3〉 캄보디아 원조 부합성(alignment) 지수 (2005년)	82
〈표 3-4〉 원조영향평가위원회(ICAI)의 성공지표	93
〈표 3-5〉 파키스탄의 선거 문화 개선 사업 평가 : 투표율 분석 결과 요약	102
〈표 3-6〉 원조영향평가위원회(ICAI) 세 가지 유형의 평가	104
〈표 3-7〉 JICA의 평가 프레임워크: 평가 기준별 평가 결과 등급 분류 체계	110
〈표 3-8〉 몽골의 사회 분야 지원 프로그램 중 사회보장 영역 정책 매트릭스	115
〈표 3-9〉 사회보장 영역 사후평가를 위한 평가 프레임워크	117
〈표 4-1〉 평가 관련 국가개발협력위원회 활동 현황 및 결과	123
〈표 4-2〉 KOICA, EDCF 평가 건수·예산·인력 추이	128
〈표 4-3〉 평가 조직에 대한 논의	129
〈표 4-4〉 메타·국별·분야별 평가 현황	132
〈표 4-5〉 종합적인 평가계획 흐름 및 시행 여부	136
〈표 4-6〉 KOICA 성과관리 조사 결과 요약	140

〈표 4-7〉 KOICA 과테말라 치말테낭고주 모자보건 개선사업 내용	142
〈표 4-8〉 OECD DAC 5대 평가기준과 범분야 이슈에 따른 평가지표	143
〈표 4-9〉 KOFIH 라오스 통합모자보건 증진사업 규모	148
〈표 4-10〉 KOFIH 라오스 통합모자보건 증진사업 평가범위	150
〈표 4-11〉 KOFIH 라오스 통합모자보건 증진사업 평가결과	152
〈표 5-1〉 JICA의 PDCA 평가체계	165
〈표 5-2〉 DFID 생식 및 모자보건 프로그램 성과 지표	172

부표 목차

〈부표 2-1〉 Tier에 의한 SDGs 지표 분류	193
〈부표 2-2〉 SDGs 보건 분야의 세부목표와 지표	194
〈부표 2-3〉 SDGs 사회정책 분야의 목표 및 지표	196

그림 목차

[그림 2-1] 성과달성을 위한 논리모형	24
[그림 2-2] 가정과 구성요소 간의 수평적 논리 관계	25
[그림 2-3] 프로그램 성과에 영향을 미치는 잠재적 환경요인	27
[그림 2-4] 사업전후 비교평가와 오류	36
[그림 2-5] 사업유무 비교평가와 오류	38
[그림 2-6] 코이카 미얀마 현장연구 샘플 숫자 및 검정력 계산 예제	48
[그림 2-7] 데이터 이중입력방식	59
[그림 3-1] 사례: 캄보디아의 보건 ODA 흐름(2000~2010년)	72
[그림 3-2] 캄보디아의 보건 영역별 ODA 분배 비중	73
[그림 3-3] 수원국별·공여기관별 보건 개발사업 건수와 금액	74
[그림 3-4] IHP+의 모니터링 및 평가 프레임워크	76
[그림 3-5] 캄보디아 정부 보건분야 공공지출과 원조추정액	80
[그림 3-6] 캄보디아와 기여국의 보건분야 사업 우선순위 차이	81
[그림 3-7] 원조사업의 단계와 VfM의 3E 평가기준	89
[그림 3-8] 영국의 공적개발협력 및 그 평가 관련 조직의 개요	90
[그림 3-9] 보건 분야 실용적 연구(Operational Research, OR) 역량 강화 프로그램의 논리 구조도	100
[그림 3-10] JICA의 평가 체계	111
[그림 3-11] JICA 조직도 (2016. 10. 1.)	113
[그림 4-1] KOFIH 라오스 통합모자보건 증진사업 평가주체 간 관계도	150
[그림 5-1] JICA의 평가 환류체계	162
[그림 5-2] 평가절차	166
[그림 5-3] 사례: 우즈베키스탄 전염성 질환 관리 역량강화 사업의 논리적 구조도	169
[그림 5-4] 사례: 우즈베키스탄 전염성 질환 관리 역량강화 사업의 PDM	170



부도 목차

[부도 1-1] 랜덤화 조건에서 사업의 영향 184

[부도 1-2] 이중차이모형(DD)을 통한 사업의 영향 측정 187

Abstract <<

Study on the ODA evaluation systems

This study aims to review the ODA evaluation systems and to explore ways to improve current monitoring and evaluation systems of ODA for health and social protection. We look into issues of monitoring and evaluation of ODA for health and social protection.

We also review types of evaluation that have been applied to measure the efficiency and efficacy of ODAs. ODA monitoring and evaluation varies by contents and types of ODAs. It includes project evaluation, program evaluation, policy evaluation, sectoral evaluation, thematic evaluation, and country assistance evaluation. This study, especially, focused on the impact evaluation which is recently highlighted in the ODA monitoring and evaluation. Impact evaluation is to put value on whether the ODA actually impacts and/or benefits the recipient's life well-being not to figure out whether the ODA meets the planned activities.

In addition, this study explores what and how a multilateral organization(i.e., WHO) and bilateral donors(i.e., England and Japan) monitor and evaluate ODA efficiency and effectiveness. We also analyze the current ODA monitoring and evaluation system in South Korea with a focus on child and maternal

2 ODA 평가체계 연구 - 보건복지분야 사업평가의 시사점

health.

Findings of this study highlights the importance of strengthening the capacity for monitoring and evaluation of ODA for health and social protections. This study recommends a more rigorous and objective evaluation scheme. We also need to improve existing feedback system and transparency of monitoring and evaluation.

* key word: ODA, Efficiency, Effectiveness, Monitoring and Evaluation

1. 연구의 필요성과 목적

한국은 2010년 경제협력개발기구 개발원조위원회(OECD DAC) 가입과 더불어 원조의 규모가 급증했음에도 원조사업에 대한 모니터링과 평가는 예산 규모와 빈도, 환류 차원에서 충분하게 이뤄지고 있지 못하다. 이러한 평가의 미비함은 국제협력사업이 활발하게 이뤄지고 있는 보건 분야에서도 별반 다르지 않다.

한국보건사회연구원은 보건 및 사회정책 분야의 공공정책을 수립하고 평가해 온 오랜 경험을 바탕으로 해당 분야의 국제협력사업 평가의 발전 방향에 대해 고민하고 개선점을 제안할 필요성을 느낀다. 따라서 발전된 평가체계와 보건 및 사회정책 분야 평가사례를 통해 한국의 평가 관련 쟁점과 발전 방향을 찾아보고자 했다.

평가의 목적과 방법론, 최근 엄밀한 정량적인 방법으로 주목받고 있는 영향평가, 그리고 원조사업을 시행하는 국제기구(예: WHO)와 영국과 일본의 평가개요와 보건복지분야 사례를 통해 한국의 평가체제와 방식에 있어서 개선되어야 할 방향과 이 가운데서도 보건복지분야에 주목해서 짚어볼 만한 개선방안을 제시하고자 했다. 요약에서는 보건복지분야에 있어서 개선방안에 대해 정리한다.

2. 보건복지분야 기획과 성과관리 개선방안

보건복지분야의 평가가 내실화되고 사업의 성과를 높이기 위해서 다음

과 같은 과제가 요구된다.

첫째, **보건복지분야 평가의 종합적 기획과 다변화가 요구된다.** 현재 평가가 모니터링과 개별 프로젝트의 종료평가에 머물러 있는 경향이 있는데, 일본의 도식화된 평가 단계와 같이 단계별로 평가를 다양화할 필요가 있다. 특히 사업기획 단계에서 사업형성 및 사전 타당성 조사를 강화함으로써 수원국의 주체성을 발휘하도록 하고 수원국의 국가정책이나 수요에 맞는 사업을 기획할 필요가 있다. 사업형성의 질을 제고하는 것은 사업 전체의 성과를 좌우한다는 측면에서 사전기획은 매우 중요하며, 여기에 예산과 시간, 인력을 투자할 필요가 있다. 그리고 현재와 같이 형식적으로 이뤄지는 중간평가를 내실화하여 중간단계에서 장애요소를 제거함으로써 성과를 제고할 것이 요구된다.

이러한 평가의 다변화와 함께 이루어져야 하는 것이 사업의 다변화이다. 사전기획과 타당성 조사 등을 통해 사업이 개별적인 프로젝트가 아니라 국가 전체 혹은 분야별, 프로그램별 접근이 요구되는 경우에는 이에 맞는 다양한 형태의 사업이 기획 및 시행될 필요가 있다. 이는 특히 보건 분야 개발목표 달성을 위해서는 빈곤, 복지전달체계와 같은 통합적인 사회안전망 설계가 요구된다는 합의의 확산에 비추어 보건복지분야에 있어서 더욱 요구되는 지점이다.

둘째, **논리모형 접근법과 정량적 평가의 확대**가 요구된다. 2014년 메타평가에 따르면, 한국의 원조사업 평가에 있어서 사업의 논리적 연관성을 설명하는 PDM의 논리적 상관관계가 취약하여 성과측정의 실효성이 부족하다. 지나치게 추상적이거나 광범위하고 논리적 연결고리가 약한 접근을 지양하고 논리모형을 내실화할 필요가 있으며, 이는 또한 사전조사의 내실화와 사전평가의 확대를 필요로 한다.

그리고 평가의 엄밀성을 높이기 위해 보건 및 사회정책 관련 지표를 사

용해 정량적 평가를 확대하는 것이 바람직하다. 특히 저개발국의 경우 객관적인 데이터 수집이 객관적이고 정량적인 평가의 걸림돌로 작용하는 것이 현실이지만 이는 점차 개선되고 있고 기초선 조사가 강화되면 그 현실성은 더욱 높아진다. 예를 들어, 모자보건 분야의 개발목표에 제시된 계량화된 지표와 WHO의 모자보건 관련 지표 등을 활용하고, 기초선 조사, 종료 후 조사를 강화함으로써 보다 엄밀한 평가를 실현할 수 있을 것이다.

이러한 엄밀한 평가를 확장하여 **장기적 영향 또는 파급효과에 대한 평가(impact evaluation)**를 강화할 수 있다. 장기적인 파급효과의 중요성이 커지면서 이를 측정하기 위한 방법론적인 발전도 상당하다. 사업 전후의 상당기간을 계획하고 추적해야 하므로 예산과 시간, 인력이 투입되어야 하는 방법인 만큼 종합적인 원조사업 평가의 강화와 더불어 영향평가 역시 강화될 수 있을 것으로 기대한다.

마지막으로, 사업수행 및 평가 시 **수원국의 문화 및 젠더 규범을 반영한 보건의료적 접근이 요구된다**. 예를 들어, 모자보건 관련 의료서비스 이용 행태는 국가 및 지역사회의 문화와 젠더 규범에 많은 영향을 받기 때문에, 병원 건립이나 의료인력 파견, 현지 전문의료인 양성 등 물리적인 차원과 더불어 수원국의 문화적·젠더적 접근이 병행되어야 하는 분야이다. 이는 비단 모자보건사업에 국한되지 않으며, 많은 사업이 국가의 사회문화적 배경의 영향을 받기 때문에 사업기획 및 평가에 있어서 **다학제적 접근**과 이를 위한 다양한 분야 전문가의 참여가 필수적이라 할 수 있다.

*주요 용어: 공적개발원조, ODA, 개발협력사업, 모니터링, 성과평가, PDM, 영향평가, 무작위통제실험

제 1 장

서론

제1절 연구의 필요성

제2절 연구의 목적

제3절 주요 연구내용

제4절 연구방법

제1절 연구의 필요성

- 원조효과성 제고가 국제개발협력의 핵심과제로 대두되면서 국제사회가 원조의 성과관리를 더욱 강화하고 있고, 우리나라 또한 체계적 성과관리를 위한 방향을 모색하는 중임.
- 원조효과 제고에 관한 파리선언(2005)을 시작으로 효과성 제고가 원조의 핵심과제로 대두되기 시작했음.
- 우리나라 역시 「제2차 국제개발협력 기본계획(2016~2020)」에서 ODA 평가 및 환류시스템 강화 전략을 추진하고 있으며, 효과성 제고를 위한 유·무상 원조사업 통합평가를 위한 조직 마련 등 체계적 성과관리를 위한 방향을 모색하는 중임.
 - 우리나라 보건부문 ODA 사업은 비중이 큰 반면 평가가 매우 제한적으로 이루어져 왔고 주로 프로젝트평가에 그치고 있으며 분야별·국별 성과를 확인하기 힘들기 때문에, 선진공여국 및 국제기구의 보건분야 ODA 사업평가의 체계와 방식을 연구하여 보건분야 성과관리를 체계화할 필요가 있음.
- 수원국 중심의 사업을 통해 원조효과성을 제고하기 위해서는 수원국의 정책, 현지 실정, 수원국의 국제협력수요와 개발협력사업의 현황 등 수원국에 대한 이해가 필요함. ODA 평가체제와 다양한 평가 결과에 대한 연구는 평가 틀에 대한 시사점을 제공할 뿐만 아니라

수원국에 대한 이해를 돕고, 국제협력수요를 파악하는 데도 유용함.

- 수원국에 대한 연구는 사전적 평가라는 단계를 통해 사업의 효과성 제고에 기여하는바, 다양한 평가사례는 또한 수원국에 대한 이해를 돕고, 개발협력수요 파악과 사업방향 설정에 도움을 줄 수 있음.
- 그리고 하드웨어 지원 중심으로 이루어졌던 우리나라 보건분야 ODA 사업이 교육 및 훈련, 제도 설계, 종합적인 보건 및 사회복지 프로그램 개발과 같은 소프트웨어 중심의 ODA 사업 강화로 이어지기 위해서는 이와 같은 ODA 사업의 사례와 평가결과를 검토할 필요가 있음.
- 한국의 보건복지분야 사업기획과 평가의 현황을 이해하고, 선진 공여국에서 해당 분야 사업을 설계하고 평가하는 사례를 통해 시사점을 도출함으로써 한국 평가와 최종적인 성과의 개선방향이 평가조직과 예산과 같은 전체 사업구조의 개선인지, 평가 방법론이나 사업기획 접근법의 개선인지 등에 대한 다각적인 고민을 해 볼 수 있을 것으로 보임.

제2절 연구의 목적

- 다양한 국제협력사업 평가모형과 보건복지분야 ODA 사업평가 사례들을 검토함으로써 우리나라 ODA 사업 평가체계의 방향을 모색하고 시사점을 얻고자 함.
- 현재 원조효과성 제고를 위한 국제사회의 논의 동향과 이를 달성하기 위한 구체적인 성과관리 방식, 관련 쟁점에 대해 연구.

- 선진공여국의 보건의료 및 사회정책 분야 ODA 사업의 다양한 평가사례를 통해 사전, 중간, 사후 단계별 평가의 역할과 효과적인 평가종류에 대한 함의를 도출.
- 영향평가 등 발전된 평가기법 학습과 이를 적용한 사례 검토를 통해 사업별 평가에 적합한 방법론 연구.
- 한국의 성과관리 현황에 기반하여 우리나라 성과관리 및 사업평가 개선방향 모색.

제3절 주요 연구내용

- 국제사회의 논의 동향과 성과관리 현황
 - 국제개발협력과 성과관리 동향
 - 국제개발협력 성과관리를 둘러싼 쟁점
- 성과평가 방법
 - 정책·전략, 국별, 분야별, 형태별·프로젝트·프로그램 평가 등 평가 방법
 - 영향평가(Impact Evaluation) 방법론과 사업 설계
- 주요 공여국 평가체계와 사례: 보건의료 및 사회정책 분야를 중심으로
 - 다자간 협력사업 성과관리모델과 평가사례
 - WHO 등 국제기구의 국제협력사업 평가체계와 사례 검토.

- 선진공여국 평가사례와 함의: 영국, 일본 등
 - 영국, 일본과 같이 평가 및 성과관리 노력이 활발한 양자 공여국의 평가체계와 사례로부터 시사점 도출.

□ 한국의 ODA 사업평가

- 평가현황
 - 평가체계와 조직
 - 평가종류별(정책·전략, 국별, 분야별, 형태별·프로젝트·프로그램 평가), 평가단계별(사전, 중간, 종료, 사후 평가) 평가 현황
 - 메타평가
- 보건의료분야 평가
 - 보건분야 평가
 - 프로젝트평가: 모자보건사업을 중심으로

□ 한국의 성과관리체계 방향 제안

- 한국 평가체계 개선방안
- 보건복지분야 사업의 기획과 사업평가 개선을 위한 방안

제4절 연구방법

□ 선행연구 및 문헌 고찰

- 개발원조평가에 대한 국제환경 변화와 쟁점에 대한 문헌 고찰.
- 선진공여국과 국제기구의 평가보고서 검토.

- 우리나라 협력사업 시행기관 보고서와 국무총리실 평가소위원회 메타평가 검토.
 - 다양한 영향평가 방법론을 사용한 학술문헌 고찰.
- 국제개발협력사업 성과관리 전문가 인터뷰 및 간담회
- 우리나라 ODA 성과평가 전문가와 학계 전문가 인터뷰와 자문회의를 통한 의견 수렴.
 - 보건, 농업, 아동 등 다양한 분야 평가 및 개발협력사업 전문가와의 정례적인 포럼과 자문회의를 통해 의견 수렴.

제 2 장

국제개발협력사업 평가의 목적과 방법

제1절 개발사업 평가의 목적과 방법

제2절 영향평가

2

국제개발협력사업 평가의 목적과 방법

제1절 개발사업 평가의 목적과 방법

1. 평가의 목적과 기준

가. 평가의 정의, 목적 및 유형

- 개발 프로그램이나 정책은 한 사회에 대한 개입을 통해서 사회적 개선을 목표로 하는데, 평가(evaluation)는 이런 목표의 달성 여부 및 달성 정도를 확인하는 절차임.
- OECD DAC(1991)에 따르면 “평가는 활동, 정책 혹은 프로그램의 가치나 의의를 확인하는 과정으로서, 예정되거나 진행 중이거나 혹은 완료된 개입에 대해 평가가 이루어질 수 있으며 최대한 체계적이고 객관적으로 이루어져야 함.”
- 평가는 크게 ‘과정평가’와 ‘성과평가’로 구분할 수 있음.
 - 과정평가(formative evaluation)는 프로그램, 정책 혹은 프로젝트가 집행되는 방식을 조사하는 것으로, 실제 진행이 계획된 작업논리(operational logic)에 따라 이루어지는지와 그 결과를 평가하는 것임.
 - 성과평가(summative evaluation)는 개입이 종료된 후에 이루어지는 평가로 예상된 결과가 달성되었는지를 평가하는데, 영향평가(outcome or impact evaluation)라고 부르기도 함.

- 중간평가(midterm evaluation)는 과정평가에 속하며, 종료 평가 및 사후평가는 성과평가에 속한다고 할 수 있음.
- 과정평가는 프로젝트나 프로그램의 집행과 개선에 초점을 맞추는 반면, 성과평가는 결과에 초점을 맞추어 특정 프로젝트나 프로그램의 지속, 확대, 혹은 종료 등을 결정하는 데 활용됨.

□ 모니터링(monitring)은 개입이 진행되는 과정에서 개입의 집행 정도와 그 결과를 측정 및 관리하는 것이라면, 평가는 결과를 특정 개입과 연결하여 분석하는 것임.

○ 모니터링과 평가는 성과에 관한 자료를 수집한다는 점에서는 동일하지만 그 방식과 실시 시점에는 차이가 있음(표 2-1 참조).

〈표 2-1〉 모니터링과 평가의 차이점

모니터링	평가
지속적이고 연속적인 작업	주기적이며 일정 기간 내의 작업
내부 활동	내부 평가, 외부 평가, 참여적 평가
관리자의 책임	평가자, 스태프, 관리자의 공동책임
성과향상을 위한 지속적 피드백	주기적 피드백

자료: Morra Imas, L. G. & Rist, R. C(2009). p.17.

□ 평가의 궁극적 목적은 의사결정자에게 정보를 제공하여 프로젝트나 프로그램에 대해 올바른 결정을 할 수 있도록 돕는 것이지만, 구체적으로는 다음 네 가지의 목적이 있음.

○ 윤리적 목적(ethical purpose): 프로젝트나 프로그램의 집행경과 및 그 성과를 정치지도자와 시민에게 보고하려는 목적

- 정보제공을 통해 개발사업에 대한 공여국 내 책임성(accountability)을 담보함으로써 민주주의의 원리를 구현하기 위한 것이라고 할 수 있음.

- 관리적 목적(managerial purpose): 물적, 인적 자원을 다양한 프로그램 사이에 더욱 합리적으로 배분함으로써 프로그램의 관리와 성과를 향상하려는 목적
- 의사결정 목적(decisional purpose): 프로젝트, 프로그램 혹은 정책의 지속, 종료, 수정 등에 관한 결정에 활용하려는 목적
- 교육 및 동기유발 목적(educational and motivational purpose): 원조집행기관과 수원기관에게 개입의 과정과 그 성과를 이해시킴으로써, 더 나은 성과를 낼 수 있도록 유도하려는 목적

※ OECD DAC(1991)은 원조사업 평가의 목적을 “① 교훈의 피드백을 통한 향후 사업의 개선, ② 대중에 대한 정보제공 등 책임성의 기반 마련”으로 요약하고 있음.

- 프로젝트나 프로그램에 대한 평가는 개발을 위한 개입과 관련된 아래와 같은 주요한 질문에 대한 답을 찾는 데 도움이 될 수 있음.

- 개입의 성과 혹은 효과가 무엇인가?
- 개입이 계획된 대로 이루어지고 있는가?
- 개입이 성과를 나타내는 데 지역별로 차이가 있는가?
- 누가 개입의 수혜자인가?

- 평가의 대상에 따라 프로젝트평가, 프로그램평가, 정책평가, 조직평가, 분야별 평가, 주제별 평가, 국가지원평가 등이 있음.

- 프로젝트평가(project evaluation): 한 지역 혹은 다수 지역에서 진행된 단일 프로젝트에 대한 평가.
- 프로그램평가(program evaluation): 공통의 목적을 가진 복수의 활동 및 프로젝트가 포함된 개입, 즉 프로그램에 대한 평가.
- 정책평가(policy evaluation): 개발을 위해 어느 조직이 도입한 기준, 가이드라인, 규칙 등에 대한 평가.
- 분야별 평가(sectoral evaluation): 교육, 보건, 농업 등과 같은 특정 영역 전반에 이루어진 개입에 대한 평가.
- 주제별 평가(thematic evaluation): 양성평등, 글로벌 파트너십 등과 같은 특정 이슈에 대한 평가로 범분야(cross-cutting) 이슈를 다루는 경우가 많음.
- 국가별 평가(country assistance evaluation): 특정 국가에 대해 이루어진 다양한 개입의 성과를 국가 단위로 평가.

나. 평가의 원칙과 기준

- 평가의 책임성 및 비교 가능성을 높이고 신뢰도와 질을 향상하기 위해서는 평가의 원칙과 기준의 마련이 필요함.
- OECD DAC(1991)은 개발사업 평가에 관한 원칙을 제시한 바 있는데, 대부분의 공여기관들이 평가 시 이 원칙의 준수를 위해 노력하고 있음.
- 원조기관은 명확한 평가 가이드라인과 방법을 정립하고 있어야 하며, 평가의 역할과 임무를 명확히 해야 함.

- 평가는 중립적이어야 하며 개발원조의 결정, 집행 및 관리 과정과는 독립적으로 이루어져야 함.
 - 평가과정은 최대한 공개되어 그 결과가 광범하게 활용될 수 있어야 함.
 - 평가결과는 정책결정자 및 운영진에 대한 피드백을 통해 반드시 활용되어야 함.
 - 평가과정에서 수원국과 공여국 간의 파트너십이 중요하며, 이를 통해 수원국의 제도구축을 지원하고 행정부담을 최소화해야 함.
 - 평가는 원조사업 기획 단계부터 중요하게 고려되어야 하며, 프로젝트나 프로그램의 명확한 목표 설정이 평가의 핵심 전제 조건임.
- OECD DAC(1991)은 또한 개발사업의 평가를 위한 5대 기준을 제시하고 있는데, 대부분의 사업 평가에 이 기준이 적용되고 있음.
- 적절성(relevance): 개발을 위한 개입의 목적이 수원자의 요구, 수원국의 필요, 국제사회의 우선순위, 파트너 및 개발기관의 정책 등과 부합하는지를 평가.
 - 효과성(effectiveness): 원조 활동이 계획된 목표를 달성한 정도를 평가.
 - 효율성(efficiency): 산출물(outputs)을 투입물(inputs)과 비교하여 평가하는 것으로, 가능한 한 작은 비용 혹은 투입으로 성과가 달성되었는지를 평가함.
 - 일반적으로 동일한 산출물을 얻을 수 있는 다른 대안(alternative)과 비교함으로써 비용 효율성의 달성 여부를 평가함.

- 파급효과(impact): 개발을 위한 개입이 직접적 혹은 간접적으로 유발한 긍정적 및 부정적 변화를 종합적으로 평가하는 것으로, 의도된 변화와 의도되지 않은 변화가 모두 평가의 대상이 됨.
 - 개발사업의 핵심 효과뿐만 아니라 지역의 사회, 경제, 환경 및 기타 개발지표에 미친 효과를 모두 포괄함.
- 지속 가능성(sustainability): 사업의 성과가 시간이 경과된 후에도 계속 유지될 수 있는지 여부 및 이와 관련된 위험 요인을 평가함.
 - 특히 공여국의 지원이 중단된 후에도 사업의 성과가 유지될 수 있는지가 평가의 주안점이며, 환경적·제도적·금융적 측면에서 지속 가능성을 평가함.

2. 성과관리와 평가

- 1980년대 후반 이후 선진국에서 공공부문의 성과 제고에 대한 압력이 강해진 결과, 성과를 중심으로 정책이나 공공부문의 개입을 관리하는 성과기반관리(RBM: results-based management) 시스템이 도입되었음.
- 성과기반관리는 그 후 개도국 개발을 위한 개입 활동에도 도입되었는데, 새천년개발목표(MDGs: Millennium Development Goals)과 고채무빈곤국(HIPC: Heavily Indebted Poor Countries) 이니셔티브 등에 적용됨.
- 성과기반관리는 개입을 통해 달성하고자 하는 결과(outcomes)와 그 결과를 보여주는 지표(indicators), 그리고 개입을 통해 달성하고자 하는 지표의 목표치(targets)를 설정한 후 목표 달성을 위해 노

력하는 관리 방식임.

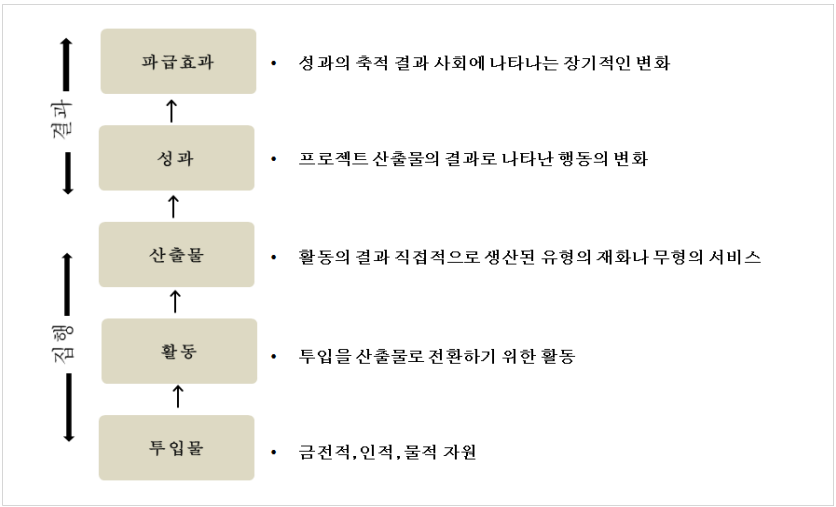
- 성과기반관리를 위해서는 성과 및 지표에 관한 정보를 관리자 및 대중에게 적기에 제공함으로써, 문제점 파악과 개선방안 도출을 돕는 것이 매우 중요함.
- 이런 정보는 모니터링과 평가시스템을 통해 획득할 수 있는데, 성과기반관리를 위한 모니터링과 평가를 ‘성과기반 모니터링 및 평가(results-based monitoring and evaluation)’라고 함.
 - 전통적 모니터링 및 평가(M&E)는 투입물(예산, 인력, 자원), 활동(프로젝트, 프로그램), 산출물(생산된 재화와 서비스)에만 초점을 맞춤.
 - 반면, 성과기반 M&E는 전통적 M&E에 성과(outcomes)나 파급효과(impacts)와 같은 더욱 포괄적인 결과에 대한 M&E를 추가함.
- 성과기반관리 혹은 M&E를 위해서는 ‘논리모형 접근법(logical framework approach)’이 폭넓게 이용되고 있음.
 - ‘논리모형(logical framework)’은 어느 조직이나 노력이 성과를 달성할 수 있는 논리적 연결을 제시하고 이에 필요한 기본 가정(underlying assumptions)을 정리한 것임.
 - 논리모형은 성과연쇄모형(results chain models), 성과모형(outcome models) 등으로 불리기도 함.
 - 논리모형은 일반적으로 <표 2-2>에 정리된 5개 주요 요소로 구성되며 [그림 2-1]과 같이 이들 사이의 인과관계를 모형화한 것임.

〈표 2-2〉 논리모형의 주요 구성요소

구성요소	내용
투입물 (inputs)	프로젝트, 프로그램 또는 정책에 투입되는 자원 (자금, 인력, 장비 등)
활동 (activities)	목표를 달성하기 위해 하는 행위
산출물 (outputs)	활동의 결과 직접적으로 생산된 유형의 재화나 무형의 서비스
성과 (outcomes)	프로젝트 산출물의 결과로 나타난 행동의 변화 (예: 금연, 모기장 사용)
파급효과 (impacts)	성과의 축적 결과 나타나는 장기적인 변화로 개입의 궁극적 목표와 유사

자료: Morra Imas and Rist(2009). p.109.

〔그림 2-1〕 성과달성을 위한 논리모형



자료: Morra Imas and Rist(2009). p.111.

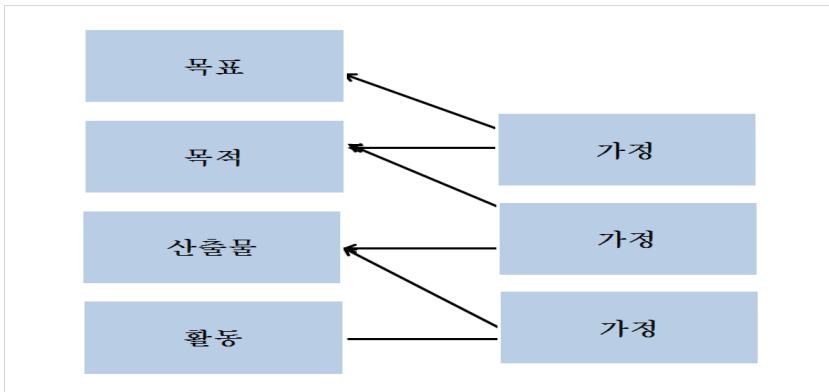
- 논리모형은 프로그램 기획의 논리, 즉 구성요소 사이의 인과적 연결을 명확히 하고, 각 단계별 성과지표를 설정하기 위해 사용함.

- 각 구성요소의 결과 혹은 목표는 직접 측정될 수 없으며, 해당 결과와 관련된 정보를 제공하는 다양한 지표인 성과지표(performance indicators)를 통해서만 관찰됨.

○ 논리모형에서 구성요소 간의 수직적 인과관계와 더불어 중요한 요소 중 하나는 가정(assumptions)인데, 이는 앞 단계의 요소가 이후 단계에서 예상된 결과를 낳기 위한 조건을 의미함.

- 예를 들어 활동(activities) 수준의 가정이란, 활동이 그 이후 단계의 성과로 이어지기 위해 해결되어야 하는 예측 가능한 위험(risks)을 의미함.
- 따라서 가정은 한 단계 위의 목표로 이어지기 위해 해결되어야 하는 또는 회피되어야 하는 위험들을 의미함.
- 구성요소 간의 인과관계를 ‘수직적 논리’라고 한다면, 가정과 구성요소 간의 관계는 ‘수평적 논리’라고 할 수 있으며 [그림 2-2]와 같이 도식화할 수 있음.

[그림 2-2] 가정과 구성요소 간의 수평적 논리 관계



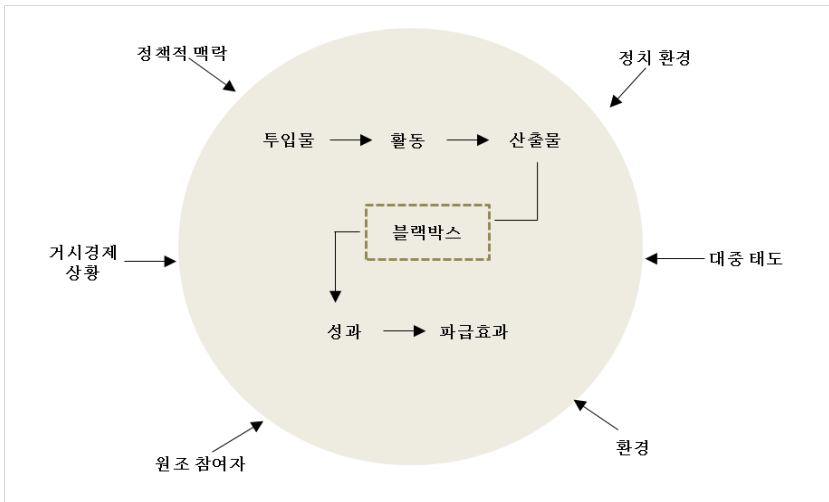
자료: KOICA(2009). p.28을 이용하여 재구성.

- 특정 프로젝트나 프로그램의 논리모형을 간략한 표로 작성한 것을 프로젝트기획매트릭스(PDM: project design matrix)라고 하는데, PDM은 사업의 모니터링 및 평가에 지속적으로 활용됨.
 - PDM에는 크게 네 가지 항목이 포함되는데, 각 구성요소(활동, 산출물, 성과, 파급효과)별로 주요내용(narrative summary), 객관적 성과지표, 지표 수집방법, 가정을 간략히 서술하고 있음.¹⁾
- 구성요소 간의 인과관계와 더불어, ‘논리모형’에서 중요한 것은 이런 인과관계가 실현될 수 있는 가정(assumptions)을 명확히 하는 것임.
 - 개입이 이루어지는 주위의 환경(정치적, 정책적, 경제적 환경 등)을 고려하여 가정이 현실성 있게 설정되어야 함.
 - 사업수행(투입물, 활동, 산출물)의 결과(성과, 파급효과)로의 연결 여부를 결정하는 블랙박스(black box)는 결국 이런 환경에 의해 결정됨.
 - 사업수행과 결과 간의 인과관계 및 결과에 영향을 미칠 수 있는 잠재적 환경요인은 [그림 2-3]과 같이 정리할 수 있음.
- 평가단계에서 평가자는 사업수행 단계에서 이미 변화의 이론이 설정된 경우에 기존 이론을 정교화하고 확인하는 작업을 해야 함.
 - 따라서 평가는 기존의 PDM 혹은 평가를 위해 정교화된 평가용 PDM(ePDM)에 기반하여 이루어져야 하며, 수직적 논리의 타당성 및 가정의 실현 여부가 성과에 미친 영향을 분석하게 됨.

1) PDM의 내용 및 작성 방법에 대한 자세한 설명은 KOICA(2009)를 참고.

- 논리모형이 없는 경우 평가자는 논리모형을 구축하여 프로그램 관리자와 함께 타당성을 확인한 후 이를 기반으로 평가를 진행하게 됨.

[그림 2-3] 프로그램 성과에 영향을 미치는 잠재적 환경요인



자료: Morra Imas and Rist(2009). p.153.

- ‘이론기반 평가(theory-based evaluation)’는 논리모형접근법과 매우 유사하며 때로는 동일한 것으로 이해되는데, 다만 프로젝트나 프로그램의 작동 원리에 대해 더욱 치밀한 이해를 바탕으로 하고 있음.
- 단순히 선형의 인과관계를 가정하는 것이 아니라, 성과에 영향을 미칠 수 있는 중요한 요인들 간의 인과관계 및 상호작용을 치밀하게 설계함.
- 이런 관계를 도식화한 모형을 ‘변화 이론(theory of change)’ 또

는 ‘프로그램 이론(program theory)’이라고 함.

- ‘변화 이론’은 “사회 변화 노력을 기획하고 평가하기 위한 혁신적 도구로서 사회 변화 노력의 장기 목표를 달성하기 위해 필요한 구성물들(building blocks)의 청사진”이라고 할 수 있음(ActKnowledge and Aspen Institute, 2003).

○ 이런 이론의 설계는 복잡한 활동을 통해 예상되는 다양한 결과를 구조적으로 이해하고, 많은 요인 중 결정적으로 중요한 요인들을 확인하는데 도움이 될 수 있음.

○ 모니터링 단계에서는 문제를 조기에 시정하고 예상하지 못한 부작용에 대응할 수 있도록 하며, 평가단계에서는 심층 분석이 필요한 요인을 선정하고 예상된 성과와 파급효과를 확인하는 데 활용됨.

- 하지만 활동이 다양할 경우 모형이 과도하게 복잡해지며 지나치게 많은 요인과 가정을 나열하는 문제점도 있음.
- 어떤 요인이 중요한 요인인지에 대해 이해관계자들 사이에 견해가 다를 경우 이를 조정하기가 쉽지 않을 수 있음.

3. 원조효과와 평가의 중요성

□ 2000년 이후 개발원조가 빈곤 감축에 미치는 효과의 제고 필요성이 대두되면서 공여국들이 원조의 성과관리에 역점을 두기 시작함.

○ 원조의 효과가 불확실하고 원조가 오히려 부패를 조장한다는 비판이 제기되자, 국제사회는 원조 확대(more aid)보다는 더욱 효과적인 원조(better aid)에 주력함.

- 원조의 효과 제고를 위한 국제사회의 노력은 2005년의 ‘원조효과 제고를 위한 파리선언’, 2008년의 ‘아크라 행동 계획’, 2011년의 ‘효과적 개발협력을 위한 부산 파트너십’ 등을 통해 확인되었음.
 - 원조의 효과 제고는 원조 집행 과정의 관리나 투입을 중심으로 한 평가가 아니라, 성과 중심의 관리 혹은 결과 중심의 평가가 이루어져야 함을 의미함.
 - 공여국 내에서는 공공행정의 개혁이나 재정지출 효율화의 한 부분으로 대외원조의 성과를 납세자 혹은 그 대표인 의회에 객관적 지표를 이용해 설명하라는 책임성(accountability) 요구가 강화되었음.
- 이런 국내외적 요구에 대응하여 원조사업의 평가를 더욱 강화하고 평가의 객관성과 엄밀성을 강화하려는 노력이 대부분의 공여국에서 나타났음.
- 다른 정책과 마찬가지로 대외원조의 경우에도 투입한 자원 대비 성과를 객관적 지표로 측정하고, 그 성과가 입증될 수 있는 사업에만 재원이 투입되어야 한다는 증거기반접근법(evidence-based approach)이 확산됨.
 - ※ 증거기반접근법이란 공공정책에서 기존 정책이나 시험적 정책에 대한 평가를 통해 확인된 정책 성과에 대한 증거를 이용함으로써, 정책의 도입, 지속, 확대, 중단 등을 결정하고자 하는 접근법임.
 - 전통적으로 원조사업의 경우 효과가 광범위하고 포괄적이며 성과 달성에 오랜 시간이 소요된다는 이유 등으로, 정량평가가 어렵고 실용적이지 않다는 인식이 지배적이었음.

- 하지만 성과관리와 효과 제고의 요구, 비효율성과 낭비의 우려 등이 제기되면서 원조사업의 경우에도 엄밀한 평가를 통해 성과에 대한 객관적이고 계량적인 증거를 도출하려는 노력이 급격히 증가함.
 - 원조효과를 정량적으로 엄밀히 측정하려는 영향평가의 도입이 그 대표적 노력이라고 할 수 있음.
 - 뉘플로(Duflo, E.) 등이 이끌고 있는 매사추세츠공과대학(MIT)의 Poverty Action Lab 등 다양한 개발연구기관들이 RCT 등의 방법을 통한 원조사업의 효과 측정과 그를 활용한 효과가 큰 원조사업의 개발을 시도하고 있음.

제2절 영향평가

1. 영향평가의 목적과 유형

가. 사업평가와 영향평가

- 사업평가(operational evaluation)는 사업이 계획한 성과의 달성 여부를 평가하는 것이라면, 영향평가(impact evaluation)은 수혜자의 삶의 질 변화가 개발사업에 기인한 것인지를 엄밀히 분석하는 것임.
- 사업평가는 M&E 프레임워크에 따라 사업의 목적, 목표, 지표를 기초로 회고적 평가를 하며, 주로 수혜자나 기타 이해관계자와의 인터뷰나 설문조사 등의 방법이 사용됨.
- 사업평가에서는 수혜자의 삶의 질에 나타난 변화를 평가하기는 하지만, 통계 수집의 어려움 등으로 이 변화가 특정 사업의 효과

인지를 정확히 확인하기 어려운 경우가 대부분임.

- 영향평가는 사업효과의 확인 가능성을 판단하고, 측정된 효과 중 어느 정도가 사업의 결과이고 어느 정도가 사업 외 다른 요인의 결과인지를 엄밀히 분석하는 것임.
- 사업평가와 영향평가는 배타적이라기보다 상호 보완적인 성격이 강함.
 - 사업평가에 사용되는 양식은 더욱 엄밀한 영향평가를 위해 유용한 자료가 될 수 있음.
 - 특히 통계자료의 이면에 있는 사회적, 제도적 맥락을 이해하고 사업이 목표로 하는 방향과 예상한 성과 채널 및 중간단계의 성과를 이해하는 데 유용함.
 - 또한 영향평가의 결과 사업이 기대한 성과를 달성하지 못한 경우 그 잠재적 원인을 추론하는 데 사업평가 결과가 유용한 정보를 제공할 수 있음.

〈표 2-3〉 사업평가와 영향평가 비교

	사업평가	영향평가
목적	계획된 성과 달성 여부 평가	수혜자의 삶의 질 변화 중 사업의 효과만을 측정
평가의 초점	투입부터 파급효과까지의 전 단계를 평가	주로 성과나 파급효과의 후반부 단계를 평가
평가방법	인터뷰, 설문조사, 현지조사 등	엄밀한 통계분석 방법
평가대상	모든 사업에 대해 실시	혁신적 혹은 대규모 사업 등에 대해 선택적으로 실시

자료: 저자 작성.

□ 프로젝트나 프로그램과 같은 개발사업의 궁극적인 목적이 목표로 하는 변화를 유도하는 것이라면, 그런 개입의 결과 실제로 어느 정도의 변화가 나타났는지를 확인하는 영향평가야말로 평가의 최종 단계라고 할 수 있음.

○ 개발사업에서 투입물보다는 성과를 중요시하고, 객관적 성과에 기반하여 사업 실시 및 계속 여부를 결정하는 증거기반 정책결정(evidence-based policy making)이 강조되면서 영향평가의 중요성이 한층 강조됨.

- “정책결정자나 시민사회가 공공프로그램 결과의 책무성을 요구하는 상황에서, 영향평가는 성과 및 목표 달성 여부에 대한 강건하고 신뢰할 수 있는 증거를 제공”할 수 있음(Gertler et al., 2011, p.4.).

○ 영향평가의 핵심은 목표 집단에 나타난 변화 중 프로젝트나 프로그램에 기인한(attributed) 효과만을 정확히 추출하는 것임.

- 영향평가의 핵심적 과제는 개입(프로젝트나 프로그램)과 결과 사이의 인과관계(causal relationship)를 정확히 식별하고 측정하는 것임.

- 이 인과관계의 식별 및 측정에는 다양한 이론적, 통계적 방법이 사용되는데 이를 영향평가 방법(impact evaluation methodology)이라고 함.

○ 양질의 영향평가를 위해서는 상당 규모의 기술적·재정적 자원이 필요하기 때문에, 일반적으로 아래의 특성을 가진 사업들에 대해 실시됨.

- 새롭고 혁신적(innovative)인 접근법을 사용한 사업

- 향후 유사 사업이 반복적(replicable)으로 진행될 수 있는 사업
- 사업 규모, 수혜자 수, 포괄 범위 등에서 전략적 성격(strategically relevant)을 가진 대표 사업
- 프로그램의 효과가 거의 검증되지 않은(untested) 사업
- 주요 정책결정에 대해 영향력 있는(influential) 결과가 예상되는 사업

나. 영향평가의 유형

- 영향평가에는 정성적 평가와 정량적 평가가 있지만, 대부분의 영향평가는 정량적 방식으로 이루어짐.
 - 정성적 평가에서는 사업이 없었을 경우의 대안적(alternative) 결과를 가정하기 어려우므로 사업의 성과 역시 정확히 측정하기 어려움.
 - 하지만 사업의 상세 내용, 주요 이해관계자, 사회·문화적 및 제도적 맥락 등에 관한 정성적 정보는 올바른 정량적 평가를 위해 중요함.
 - 사업 특성에 따라 사업효과에 대한 종합적 이해를 위해 정성적 평가와 정량적 평가를 결합한 혼합 접근법(mixed-methods approach)도 사용됨.
- 정량적 영향평가에는 사후(ex post)영향평가와 사전(ex ante)영향평가의 2종류가 있음.
 - 사전영향평가는 사업 대상지의 현재 조건하에서 계획된 사업이

미래에 어떤 효과를 발생하게 할 수 있는지를 추정하는 평가임.

- 일반적으로 시뮬레이션(simulation) 방법이 사용되는데, 주요 주체들 간 및 외부환경과의 관계를 모형화한 구조 모형(structural models)을 설정하고 주요 변수의 예측치에 대한 가정을 근거로 사업의 예상 효과를 평가함.

○ 사후영향평가는 사업이 진행된 이후 그 사업으로 인해 수혜자에게 나타난 변화를 측정하는 것임.

- 개발사업과 같은 외부의 개입(intervention) 혹은 처치(treatment)가 있었을 경우 대상 집단 즉 수혜자에게 나타난 변화를, 처치가 없을 경우와 비교하는 처치효과모형(treatment effect model)이 주로 사용됨.
- 이 모형은 사업이 수혜자에게 영향을 미치는 구체적 메커니즘을 설정하지 않을 수 있다는 점에서 사전영향평가 방법과 차이가 있음.
- 또한 수혜그룹과 비수혜그룹에 관한 다양한 통계자료의 수집이 필요하고, 사업 시작 단계부터 평가를 고려한 준비가 필요하므로 사전평가에 비해 많은 비용과 시간이 소요됨.

2. 영향평가 방법론

□ 사업 혹은 개입에 귀속시킬 수 있는 효과만을 측정하는 영향평가에서 가장 큰 문제는 사업을 실시하지 않았을 때의 가상적 결과(counterfactual outcome)를 측정할 수 없다는 것임.

○ 성과지표를 Y로 하는 특정 개발 프로그램 T의 순수한 효과는 다

음과 같이 표현될 수 있음.

$$\alpha = (Y|T=1) - (Y|T=0)$$

- 즉 프로그램 T가 실시되었을 때(T=1)의 결과(Y)와 실시되지 않았을 때(T=0)의 결과의 차이로 측정됨.
- 하지만 동일한 개인 혹은 그룹에 대해 동일한 시점에서 프로그램이 실시되었을 때의 결과와 실시되지 않았을 때의 결과를 동시에 측정하는 것은 불가능함.
- 프로그램에 참가한 그룹을 대상으로 결과를 측정할 경우 ($Y|T=1$)은 측정 가능한 실제 결과(actual outcome)이지만, ($Y|T=0$)은 가상적 결과로 측정이 불가능함.

○ 가상적 결과를 대체할 수 있는 적절한 비교그룹(comparison group)을 찾는 방법이 결국 영향분석 방법을 결정한다고 할 수 있음.

□ 사업의 성과를 측정하기 위해 사업평가에서 흔히 사용하는 사업전후비교(before-and-after comparison)나 사업유무비교(with-and-without comparison)는 사업의 효과 α 를 정확히 측정하지 못함.

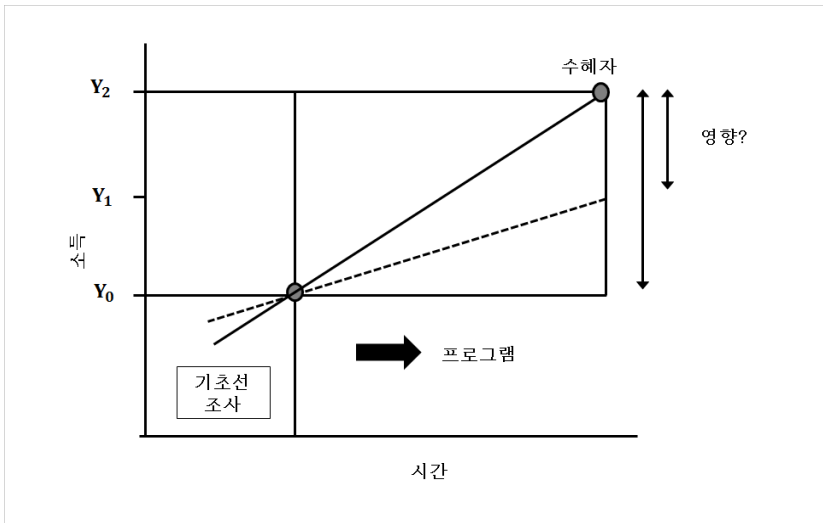
○ 수혜자 그룹을 대상으로 특정 사업 실시 전과 후의 성과지표를 측정하여 이 차이를 사업의 효과로 평가하고자 하는 방법이 전후비교방법임.

- 하지만 사업이 진행되는 기간 동안 사업 실시 외에도 다양한 다른 요인들이 수혜자에게 영향을 미쳤을 수 있기 때문에 단순히 전후의 결과 차이를 모두 해당 사업에 귀속시킬 수 없음.
- 예를 들어 소액금융(micro-finance)사업을 통해 농촌 가구

의 소득 증가를 목표로 할 경우, 사업이 진행되는 동안 가구원의 취업 상황, 기후, 농작물 가격, 전반적 경기변동 등 다른 요인들도 가구 소득에 영향을 미칠 수 있어 소액금융사업 후 소득 변화를 모두 동 사업의 효과로 판정할 수는 없음.

- [그림 2-4]에서 전후비교방법은 사업 시작 전후의 성과지표(소득) 차이 ($Y_2 - Y_0$)를 사업의 성과로 측정하지만, 사업 외 다른 요인들에 기인한 소득 증가가 ($Y_1 - Y_0$)라면 이 측정치는 과대평가된 것임.
- 따라서 정확한 평가를 위해서는 사업 기간에 발생한 다른 요인들의 효과를 제거할 수 있을 때만 사업의 순수한 효과를 정확히 측정할 수 있음.

[그림 2-4] 사업전후 비교평가와 오류



자료: Khandker et al.(2010). p.24.

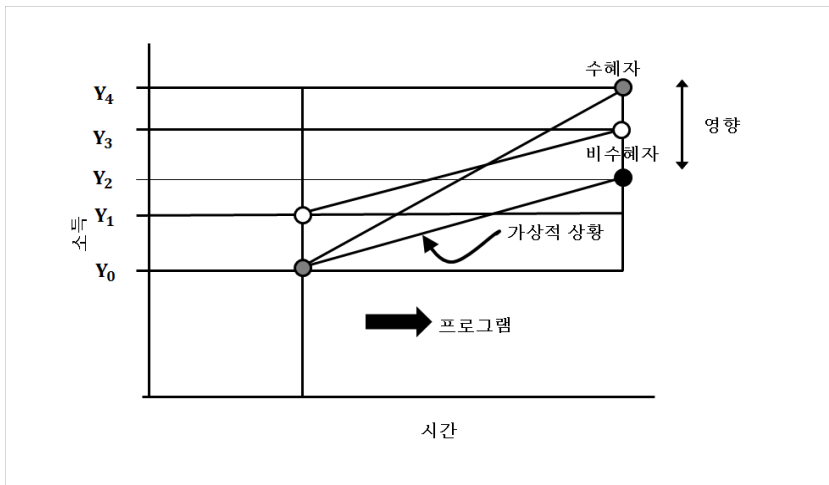
○ 사업종료 시점에 사업 수혜집단(treatment group)과 비수혜집단(control group)의 성과지표를 비교하여 그 차이를 사업의 효과로 측정하고자 하는 방법이 유무비교방법임.

- 하지만 이 방법의 경우 수혜집단과 비수혜집단이 동일한 집단 혹은 동일한 확률적 특성을 가진 집단이 아니라는 데서 문제가 발생함.
- 대부분의 개발사업에서 수혜집단은 임의추출(random) 방식으로 선정되는 것이 아니라, 일정한 특성을 가진 집단으로 구성되기 때문에 소위 선택편의(selection bias)가 발생함.
 - 유의적 대상 선발(purposive program placement): 개발 사업은 빈곤층 혹은 빈곤 지역과 같이 특정한 성격을 지닌 집단을 대상으로 이루어지기 때문에 사업 시행자의 대상자 선정 과정에서 편의(bias)가 발생함.
 - 자기선택 편의(self-selection bias): 사업 참여 여부를 잠재적 수혜 대상자들이 스스로 결정한다고 하더라도 특정 성향을 가진 사람들이 주로 사업에 참여할 가능성이 있어 편의가 발생함.
- [그림 2-5]에서 유무비교방법은 수혜집단과 비수혜집단의 성과지표(소득)의 차이인 ($Y_4 - Y_3$)를 사업효과로 측정하지만, 시작 단계부터 수혜집단의 소득은 비수혜집단에 비해 ($Y_1 - Y_0$) 만큼 낮은 상태였기 때문에 이 측정치는 정확하지 않음.
 - 예를 들어, 소액금융사업 참가 가구를 선발할 때 사업수행기관이 소득이 낮은 가구를 선정하거나 소득수준이 상대적으로 낮은 사람들이 주로 이 사업에 지원했다면 시작부터 수혜

그룹의 평균소득은 비수혜그룹에 비해 낮을 것임.

- 만약 사업이 없었을 경우 수혜집단의 소득이 비수혜집단과 같은 폭으로 증가했을 것으로 가정할 경우 사업의 실제 효과는 $(Y_4 - Y_2)$ 가 될 것이므로, 유무비교방법은 사업 효과를 과소 측정하게 됨.

[그림 2-5] 사업유무 비교평가와 오류



자료: Khandker et al.(2010). p.23.

□ 요약하면 특정 프로그램 또는 정책의 영향을 평가할 때 인과관계²⁾를 입증하기 위해서는 특정 프로그램에 참여했던 사람들이 만약 프로그램에 참여하지 않았을 경우 얻게 될 가상의 결과나 상황 (Counterfactual)이 필요한데 Rubin(1974)은 이를 가상결과모델

2) 과거 실증분석은 단순 상관관계 분석에 머무는 경우가 많았고, 체계적이지 못하고 작위적인 측면이 강했음. 이에 대해 Leamer(1985)는 다양한 분석모형, 가정들을 적용하여도 일관된 결론이 도출되는지를 살펴보는 민감도 분석(Sensitivity Analysis)을 대안으로 제시하였지만, 이 역시 엄밀한 인과관계를 입증하는 데 한계가 있음.

(Potential Outcome)로 설명.

- 프로그램 또는 정책의 영향을 평가할 때 단순 전후 비교로는 인과관계를 입증하기 어려운데, 그 이유는 프로그램 기간 중 발생할 수 있는 다양한 요소들이 우리가 평가하고자 하는 변수에 영향을 미칠 수 있기 때문임.
- 또한 해당 프로그램에 참여한 그룹과 참여하지 않은 그룹을 비교하여 그 차이를 정책의 영향으로 측정하더라도 일반적으로 프로그램에 참여한 사람과 참여하지 않은 사람은 서로 다른 사람이기 때문에 이들의 단순 비교는 선택편의(Selection Bias) 문제를 일으킴.
- 결국 가상적 결과의 측정이 불가능한 상황에서 영향평가의 정확성은 수혜집단과 가장 유사한 비교집단을 선정함으로써 선택 편의를 최소화하는 데 달려있다고 할 수 있음.
- 수혜집단과 비수혜집단의 성과지표(예: 소득)의 평균값은 모두 측정 가능하며 그 차이는 다음과 같이 표현할 수 있음.

$$D = E(Y_i(1)|T_i = 1) - E(Y_i(0)|T_i = 0)$$

- $Y_i(1)$ 는 가구 i 가 사업에 수혜자로 참여했을 때의 소득을 나타내고, $Y_i(0)$ 는 사업에 참여하지 않았을 때의 소득을 의미함.
- $E(Y_i(1)|T_i = 1)$ 는 수혜집단($T_i = 1$)의 수혜시 소득($Y_i(1)$)의 평균을, $E(Y_i(0)|T_i = 0)$ 은 비수혜집단($T_i = 0$)의 비수혜시 소득($Y_i(0)$)의 평균을 의미함.
- 앞의 설명한 바와 같이 이 차이는 사업의 정확한 효과가 아닌데, 이를 교정하기 위해 수혜집단이 사업에 참여하지 않았을 때의 가상적 결과 즉 $E(Y_i(0)|T_i = 1)$ 을 이용하면 위 식은 아

래와 같이 표현됨.

$$\begin{aligned} D &= E(Y_i(1)|T_i = 1) - E(Y_i(0)|T_i = 1) \\ &\quad + E(Y_i(0)|T_i = 1) - E(Y_i(0)|T_i = 0) \\ &= E(Y_i(1) - Y_i(0)|T_i = 1) \\ &\quad + [E(Y_i(0)|T_i = 1) - E(Y_i(0)|T_i = 0)] \end{aligned}$$

- 위 식의 앞부분 즉 $E(Y_i(1) - Y_i(0)|T_i = 1)$ 이 순수한 사업효과 또는 처치효과(treatment effect)임.
- 반면 뒷부분 즉 $[E(Y_i(0)|T_i = 1) - E(Y_i(0)|T_i = 0)]$ 은 사업이 없었을 경우에 두 집단 간의 가상적 결과 차이를 나타내므로 선택편의(selection bias)를 나타냄.
- 결국 뒷부분 즉 선택편의를 제거 혹은 최소화할 수 있는 현실적 방법을 찾는 것이 영향평가의 핵심이라고 할 수 있음.

□ 1990년대 후반 이후 실증경제학 분야에서는 도구변수(IV: Instrumental Variables), 무작위통제실험(RCT: Randomized Controlled Trial)과 같은 연구설계가 매우 중요하다는 입장(Angrist and Pischke, 2010)이 대두되었고, 엄밀한 인과관계를 입증하기 위한 연구 디자인 중 무작위통제실험 방법이 특히 주목받게 되었음 (Duflo et al., 2007).

○ 무작위통제실험 방식은 의학 분야에서 이미 오랜 기간 동안 가장 높은 수준의 증거를 도출해내는 연구방법론으로 인정받고 있었고, 이러한 연구방법론이 1990년대 후반 사회과학분야에 크게 확산

3. 무작위통제실험(RCT)의 설계와 분석

□ 무작위통제연구는 프로그램 또는 정책의 인과적 영향을 입증하는

아주 강력한 연구방법론으로 학계 및 많은 연구기관에서 이러한 방법론을 적용한 연구를 진행

- 본고에서는 가상결과모델에서 발생하는 선택편의(selection bias)를 효과적으로 제거할 수 있는 무작위통제연구의 기본적인 내용을 실제 연구 사례를 통해 간략하게 살펴보고자 함.
- 무작위통제연구에 대한 보다 깊고 포괄적인 논의를 검토하려면, 사회과학분야 무작위통제연구에서 주도적인 역할을 하고 있는 J-PAL 연구소의 실제 현장에서 축적된 연구 경험을 기반으로 작성된 Glennerster and Takavarasha(2013)를 참고할 것

가. 무작위통제연구 설계

- 본 절에서는 무작위통제연구를 설계함에 있어 고려해야 할 중요한 단계들을 간략하게 살펴보고자 함.

1) 연구주제 선정³⁾

- 다음과 같은 주제들은 무작위통제연구 설계가 필요⁴⁾

3) 연구주제를 선정하기 위한 좋은 출발점은 기존 연구문헌 등을 검토하는 것. 기존 연구문헌 검토를 통해 기존 연구에서 밝혀낸 점과 추가 연구가 더 필요한 부분들을 파악할 수 있음. 무작위통제연구 관련 문헌들을 검토할 경우, 다음과 같은 기관들에서 제공하는 연구문헌을 살펴보면 유용함.

- Jameel Poverty Action Lab (J-PAL) at MIT
- Innovations for Poverty Action (IPA)
- International Initiative for Impact Evaluation (3ie)
- Development Impact Evaluation Initiative (DIME) in World Bank.

또한 경제학의 경우 Handbook of Economics Series, Annual Review of Economics, Journal of Economic Perspectives, 및 Journal of Economic Literature와 같은 학술지에서 유용한 문헌검토 가능.

4) 연구주제의 성격에 따라 무작위통제연구가 꼭 필요한 경우도 있고 그렇지 않은 경우도 있음.

- 프로그램 또는 정책의 효과가 있었는지 여부
 - 프로그램의 어떠한 요소가 가장 큰 영향을 미쳤는지 여부
 - 두 개의 다른 프로그램 중 어떤 프로그램이 더 좋은 결과를 가져오는지 여부
 - 두 개의 다른 프로그램이 동시에 적용되었을 때 시너지가 생기는지 여부
 - 한 곳에서 도출된 결과가 다른 맥락에서도 재현될 수 있는지 여부
- 무작위통제연구가 필요한 연구주제를 선정하고 난 후 해야 할 일은 이러한 연구 주제를 바탕으로 설문조사 문항을 개발하는 것임.
- 연구가설과 이를 설명할 메커니즘 등이 설문조사 문항에 잘 반영되어야 함.
 - 최근에는 학술지 공식 사이트에 연구논문과 함께 데이터 원자료, 분석코드, 설문조사 문항 등을 모두 공개하는 경우가 많아, 해당 논문에 첨부된 설문조사 문항을 참고하여 설문조사 문항을 개발할 수 있음.
 - 또한 Demographic and Health Survey(DHS)와 같이 해당 국가에서 정기적으로 진행되는 설문조사 문항을 검토하게 되면, 해당 국가 맥락에 맞는 문항을 쉽게 개발할 수 있음.
 - 일례로 코이카에서 추진하고 있는 미얀마 새마을운동사업의 경우, 단기적으로 새마을운동이 마을 사회적 자본 형성에 미친 영향을 평가할 계획을 갖고 있는데, 이와 관련된 설문조사 문항을 개발하기 위해 DHS 설문조사 문항과 세계은행에서 사회적 자본 측정을 위해 개발한 Integrated Questionnaire for

the Measurement of Social Capital(Grootaert et al., 2004)을 이용하여 설문조사 문항을 개발하였음.

2) 현지 협력 기관 선정

□ 무작위통제연구를 통한 영향평가는 대부분 저개발국가에서 시행되기 때문에, 연구진이 현지 기관과 협력하는 것이 필수적임.

○ 현지 파트너 기관이 현장연구를 진행하는 데 필요한 법적·실무적인 토대를 제공해 주게 되며, 현지 파트너 기관을 통해 해당 필드 영향평가 연구를 수행하는 데 따른 사업 승인, 연구윤리 승인 등을 받게 됨.

- Kim et al.(2016)의 경우 아프리카 말라위에서 에이즈예방사업을 위한 중고등학교 남학생 포경수술사업을 무작위통제연구로 진행하였는데, 현지 말라위 대양누가병원과 협력 관계를 맺고 연구를 추진.

- 남학생 포경수술사업에 대한 승인 및 무작위통제연구에 대한 연구윤리심사 모두 말라위 대양누가병원 명의로 취득하였고, 연구사업에 필요한 현지 인력 역시 대양누가병원을 통해 모집·선발함.

○ 현지 기관과 파트너십을 형성할 때 중요한 점은 무작위통제연구가 왜 중요한지에 대한 상호 이해가 형성되어야 한다는 것임.

- 비정부기구(NGO) 등과 같은 현지 기관들은 연구 중심의 사업에 대한 반감을 가지고 있는 경우가 있으며, 특히 무작위통제연구 방식에 대해 사업 대상을 위하는 것이 아니라 연구를 위한

수단으로 사업을 한다고 생각하여 반감을 가지는 경우도 있음.
따라서 엄밀한 인과관계를 밝히기 위한 무작위통제연구의 중요
성에 대한 상호 이해가 현장 영향평가 연구의 선행 조건임.

3) 사업 및 연구윤리 승인

□ 무작위통제연구를 현장에서 실행하기 전, 사업에 대한 현지 정부의
승인과 연구윤리심사(IRB: Institutional Review Board)를 받아
야 함.

- Kim et al.(2016)의 경우, 현지 파트너 기관인 대양누가병원을
통해 말라위 정부(보건부)의 남학생 포경수술사업 승인을
받았음.
- 말라위 정부의 공식적 사업 승인을 취득한 후, 비공식적인
지역 추장들의 사업 승인도 함께 받았음. 저개발국가 농촌
환경에서 진행되는 현장연구의 경우 공식적인 사업 승인도
중요하지만, 비공식적인 traditional authorities의 승인을
받는 것 역시 사업 진행에 필요한 부분임.

○ 연구윤리심사의 경우 현장연구를 진행하는 현지 국가에서, 또한
연구진이 속해 있는 연구기관에서 각각 연구윤리심사를 받는 것
이 일반적임.

- Kim et al.(2016)의 경우 보건부 산하 National Health
Science Research Committee에서 IRB 승인을 받았고, 후
에 연구진이 속해있는 Columbia University에서 IRB 심의
를 받았음.
- 보건사업의 경우 보통 해당국 정부의 보건부 산하에 IRB 심의

위원회가 존재하는 경우가 많으며, 국가에 따라 지방정부 및 중앙정부 모두에서 각각의 IRB를 받아야 하는 경우도 있음.

4) 사전분석계획(Pre-Analysis Plan) 등록

□ 자료 선별(Data mining) 및 좋은 결과만 보고하는 관행(Cherry picking)을 피하기 위해서 무작위통제연구에 대한 사전분석계획(Pre-Analysis Plan: PAP)을 공개 등록하는 것이 확산.

○ 무작위통제연구에서 다양한 결과변수들을 분석하게 되면, 이 중에서 일부는 확률상 실제 프로그램의 효과와는 상관없이 결과변수들이 통계적으로 유의미하게 나올 수 있으며, 특정 소그룹에 대한 heterogeneous한 프로그램 효과를 분석하는 경우에도 다양한 소그룹(성별, 나이, 교육수준 등)별로 분석을 하다 보면 확률상 통계적으로 유의미한 heterogeneous 효과를 우연히 찾아낼 수도 있음. 여기서 좋은 결과나 선별된 자료만 보고될 수 있음.

- 사전분석계획을 미리 등록하게 되면 결과변수를 측정하는 방식이 매우 다양하거나, 실증분석 모형(specification)을 다양하게 적용해 볼 수 있는 경우, 또는 사전적으로 해당 프로그램이 특정한 소그룹에 대한 이질적인 효과가 있을 것으로 예상될 때, 자료선별 및 좋은 결과만 보고한다는(Cherry Picking) 비판을 예방할 수 있음.

○ 의료, 보건, 의학 분야 연구에서는 이러한 사전분석계획 등록이 오래전부터 일반화되어 있었지만, 최근에는 사회과학분야에서도 이러한 사전분석계획 등록이 연구의 엄밀성과 객관성 확보 차원에서 더 중요해지고 있음.⁵⁾

○ Finkelstein et al.(2010)은 경제학 분야에서 등록된 초기 PAP 중 하나로, 오레곤 주에서 저소득층 대상으로 Medicaid 건강보험을 무작위로 제공한 오레곤 메디케이드 실험에 대해 사전분석 계획을 정리하였음.

- 본 사전분석계획은 1)무작위통제연구 설계 방식 및 데이터 출처, 2) 회귀분석 및 분석 프레임, 3) 무작위배정 검증(verification) 및 기초선 설문조사 결과, 4) 주요 분석 및 보조 분석, 5) 추가 분석, 6) 해석 및 한계 등의 내용을 포함.

○ 하지만 일부 연구자들은 프로그램의 영향이 어떤 방향으로 나타날지 사전적으로 예측할 수 없는 경우가 많기 때문에, 사전에 분석계획을 PAP를 통해 고정하게 되면 연구 분석에 대한 유연성이 떨어질 수 있다는 점을 비판.

○ 가장 일반적으로 사용되는 PAP 작성 시기는 기초선 설문조사 데이터를 수집하고, 프로그램 사업을 진행한 후, 종료 설문조사 데이터를 수집한 상황에서 최종 분석을 시작하기 전임.

- 이상적으로는 기초선 설문조사 데이터를 수집하기 전, 프로그램을 시작하기 전에 하는 것이 좋으나, 정보 부족 등의 이유로 구체적인 PAP를 작성하기가 어려울 수 있음.
- Finkelstein et al.(2010) 역시 종료 설문조사 분석을 시작하기 전에 PAP를 작성.

5) 2013년 전미경제학회(AEA)에서 사회과학분야 무작위통제연구의 사전분석계획(PAP) 등록 사이트를 개설. www.socialscienceregistry.org에서 다양한 연구들의 PAP를 다운로드 받을 수 있음.

5) 검정력 계산(Power Calculation)

□ 현장에서의 무작위통제연구는 많은 예산이 소요되기 때문에 적절한 샘플 숫자를 사전에 계산하는 것이 중요하므로, 샘플 특성에서 모집단 모수를 추정할 때 검정력(statistical power)은 중요한 역할을 함.⁶⁾

○ 검정력은 유의수준, minimum detectable effect(MDE) size, 결과변수의 variance, 실험군과 통제군 간의 비율, 및 샘플 숫자에 의해 영향을 받음. 이러한 관계 속에서 특정 수준의 검정력을 확보하기 위해 필요한 적절한 샘플 숫자를 계산할 수 있음.

- Stata와 같은 대부분의 통계 프로그램에서 검정력을 계산할 수 있는 기능이 있지만, Optimal Design(OD)과 같은 검정력 계산 전문 소프트웨어를 사용하는 것이 편리함.⁷⁾ OD 프로그램의 경우 MDE와 같은 검정력 결정 요인 등이 변화할 때 필요한 샘플 숫자가 어떻게 변화되는지를 그래프를 통해 시각적으로 보여줌.

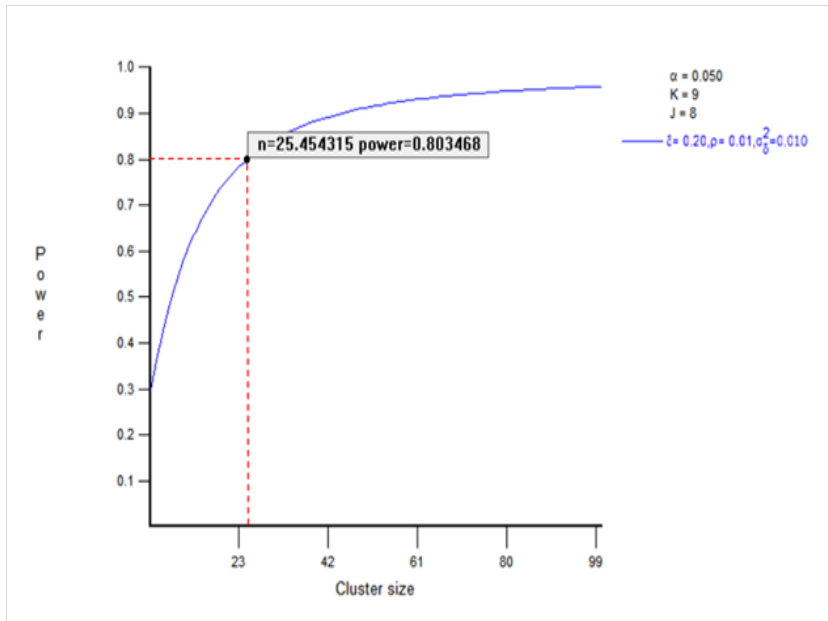
• [그림 2-6]은 OD 프로그램을 사용해서 샘플 숫자를 계산한 예임. 한국국제협력단(KOICA)에서 미얀마 9개 주에서(K, the total number of stratum=9), 주별 최소 8개 마을을 선정한다고 했을 때(J, the number of clusters per stratum=8) 80% 검정력 수준을 달성하기 위한 마을 내 설문조사 대상 가구 수를 계산해 보면 유의수준은 5%($\alpha =$

6) 통계추론에서 false positive 확률을 α , false negative 확률을 κ 로 설정하면, 유의수준(significance level)은 α 이며, 검정력(power)은 $1-\kappa$ 임.

7) Optimal Design 프로그램은 다음 링크에서 무료로 다운로드받을 수 있음.
<http://hlmssoft.net/od/>

0.05), 표준화된 MDE는 0.2 표준편차($\delta = 0.2$)로 설정하게 되면 마을당 26개 가구를 조사하게 될 경우 80%의 검정력을 확보할 수 있게 됨.⁸⁾

[그림 2-6] 코이카 미얀마 현장연구 샘플 숫자 및 검정력 계산 예제



자료: Optimal Design 이용 저자 작성.

6) 무작위 추출법(Random Sampling)

□ 무작위통제연구를 수행하기 전에, 우선 모집단을 대표할 수 있는 샘플을 확보하는 것이 중요함.⁹⁾

8) 보다 자세한 검정력 분석과 관련된 내용은 Glennerster and Takavarsha(2013) 제6장을 참고할 것.

9) 일례로 1936년 미국 대통령 선거 여론조사에서 당시 공화당 후보였던 Landon이 낙승

○ 모집단 센서스 데이터가 가용한 경우, 센서스 데이터에서 샘플을 무작위로 추출하고, 모집단 모수와 샘플 통계량을 비교하여 추출된 샘플의 대표성을 확인할 수 있음.

- 코이카 미얀마 새마을사업은 총 100개 미얀마 농촌 마을을 대상으로 진행되고 있는데, 미얀마 정부의 협조를 받아서 9개 주 사업 대상지 100개 마을에 거주하고 있는 18,456가구 전수 데이터를 수집한 후 18,456가구 중에서 5,515가구를 무작위로 추출.

• <표 2-4>에서 모집단 모수 평균과 샘플 통계량 평균을 비교하였는데 가구 수, 직업, 경작지 면적, 가구 자산 등의 변수에서 통계적으로 유의미한 차이를 발견하지 못했음. 이는 5,515가구 샘플이 전체 18,456가구를 대표함을 보여줌.

○ 현장에서 무작위통제연구를 장기간에 걸쳐 진행할 경우, 연구 대상 지역의 인구 센서스 시스템을 구축할 수 있으면 이상적임.

- 가나 나브롱고 보건연구센터(NHRC: Navrongo Health Research Center)의 경우 1980년대 비타민 A의 효능을 입증하는 보건연구 (VAST: Vitamin A Supplementation Trials)를 진행한 이후, 사업 대상지 약 15만 가구를 대상으로 매 3개월마다 정기 인구 센서스 시스템을 구축하여 지금까지 운영해 오고 있음.¹⁰⁾

할 것으로 예상하였으나, 그 결과는 반대로 루스벨트 대통령이 낙승함. 당시 여론조사가 정반대의 결과를 예측한 이유는 당시 여론조사기관이 발행하던 주간잡지 구독자, 차량 및 전화 소유자를 대상으로 샘플을 추출했기 때문임. 당시 차량 및 전화를 소유하고 있었던 사람들은 평균보다 부유한 계층이었고, 공화당 지지자가 과대 추출되었기 때문에 정반대의 예측 결과가 도출되었음.

10) 나브롱고 보건연구센터: <http://www.navrongo-hrc.org/>

〈표 2-4〉 샘플 대표성 평가

	가구 수			직업	토지	가구 자산			
	총인원	여성 숫자	단독 가구	비농부	경작면적 (acre)	가축	농기계	자동차 및 오토바이	재봉틀
모집단 (N = 18,456가구)									
평균	4.01	2.07	0.04	0.50	2.90	12.78	0.19	0.01	0.002
(표준 편차)	(1.67)	(1.12)	(0.20)	(0.50)	(5.85)	(107.56)	(0.55)	(0.14)	(0.05)
샘플 (n = 5,515가구)									
평균	3.98	2.06	0.04	0.49	2.94	12.49	0.19	0.01	0.002
(표준 편차)	(1.67)	(1.12)	(0.20)	(0.50)	(5.85)	(107.56)	(0.55)	(0.14)	(0.04)
t	-1.40	-0.90	-0.75	-0.93	0.41	-0.20	0.54	0.42	-0.53
(p- value)	(0.16)	(0.37)	(0.45)	(0.35)	(0.68)	(0.84)	(0.59)	(0.67)	(0.60)

자료: 저자 작성.

7) 무작위배정(Random Assignment)

□ Rubin(1974)의 가상결과모델에서 무작위배정을 통해 실험군과 비교할 적절한 가상의 상황(counterfactual)을 구축할 수 있음을 살펴봐왔음. 무작위배정은 3가지 방식으로 적용될 수 있음.

○ 다음의 세 가지 무작위배정방식들(프로그램 제공, 프로그램 제공 순서, 프로그램 참여 격려)을 통해, 현장연구를 진행할 시 거의 모든 상황에서 무작위통제연구를 진행할 수 있는 가능성을 찾을 수 있음.

□ **프로그램 제공 무작위배정**: 첫 번째 방식은 프로그램 또는 정책의 혜택을 무작위로 배정하는 방식으로 가장 일반적으로 사용되는 방식임.

- Kim(2014)은 말라위에서 에이즈예방사업의 일환으로 33개

중고등학교 여학생들에게 1년간 장학금 지급 프로그램을 운영. <표 2-5>는 총 124개 학급 중에서 무작위로 62개 학급을 선발하여 장학금 프로그램을 제공한 연구설계를 보여줌.

- 사업 예산 제약으로 인해 33개 학교 124개 학급 전체 학생들에게 장학금을 제공할 수 없어, 전체 학생 중 일부를 장학금 수혜자로 선정.
- 많은 경우, 프로그램 관련자(stakeholder)들이 제비뽑기(무작위 선정) 방식을 선호하는 경우가 많은데, 무작위 선정 방식이 가장 공정한 선택 방식이라 생각하기 때문임. 장학금 수혜자 선정 방식을 놓고 참여 학교 33개 교장 선생님들과 장시간 토론하였는데, 좀 더 가난한 학생들에게 혜택을 주자는 의견이 처음에는 강하게 제시되었지만, 단일 학교 내에서도, 서로 다른 학교 간에 누가 더 가난한지 객관적으로 측정하기 어렵기 때문에 공정하게 무작위로 선정하는 것으로 결론을 내렸음.¹¹⁾

<표 2-5> 여학생 장학금 프로그램 연구설계

그룹	배정	학급 수	학생 수
G1	실험군 (Treatment)	62	2,102
G2	통제군 (Control)	62	1,895
전체		124	3,997

주: 무작위배정 단위는 학급임.

자료: Kim(2014). p.8.

11) 무작위배정을 위한 난수는 엑셀 프로그램에서 생성하였고, 33개 교장선생님들 입회하에 지역교육감(Division Education Officer)이 무작위배정 버튼을 눌렀고, 이에 따라 생성된 무작위배정 결과에 대해 33개 교장 선생님들의 동의 서명을 받았음. 이와 같이 무작위배정방식을 연구진이 아닌 프로그램 주요 당사자들이 결정·실행하게 하는 것은 향후 무작위통제연구 현장연구를 진행하는 데 매우 큰 근거가 될 수 있음.

□ **프로그램 제공 순서 무작위배정**: 두 번째 방식은 프로그램은 모두에게 제공하되, 프로그램 제공 순서를 무작위로 배정하는 방식임.

- Miguel and Kremer(2004)는 케냐의 75개 초등학교를 대상으로 구충제 보급 사업을 진행했는데, 예산 및 행정 제약으로 인해 75개 학교를 3그룹으로 나누어서 첫 번째 그룹인 25개 학교는 1998년에 구충제를 보급받았고, 1999년에는 첫 번째 그룹 25개 학교와 두 번째 그룹 25개 학교가 구충제를 보급받았고, 2001년에 최종적으로 세 번째 그룹인 25개 학교가 구충제 프로그램에 참여함. 따라서 1998년에는 첫 번째 그룹 25개 학교가 실험군으로 배정되었고, 나머지 50개 학교가 통제군이었으며, 1999년에는 첫 번째 그룹과 두 번째 그룹 50개 학교가 실험군, 세 번째 그룹 25개 학교가 통제군의 역할을 함.
- Kim et al.(2016)은 말라위에서 에이즈예방사업의 일환으로 중고등학교 남학생 포경수술사업을 진행하였는데, 협력 병원 에서 하루 최대 10명까지만 포경수술이 가능하며, 포경수술을 받는 일자를 무작위로 배정하는 두 번째 방식을 사용.

○ 이러한 방식은 프로그램 제공을 동일한 시간에 함께 제공하기 어려운 경우나 예산 제약으로 모두에게 프로그램을 제공하기 어려운 경우 사용.

○ 두 번째 방식은 연구사업에 참여하는 대상자 모두에게 궁극적으로는 프로그램을 제공한다는 점에서 첫 번째 무작위배정방식에 비해 윤리적 비판을 덜 받지만, 궁극적으로는 모두가 프로그램 혜택을 받게 되어 통제군이 없어지게 되므로 프로그램의 장기 효과를 분석할 수 없는 약점을 가짐.

□ **프로그램 참여 격려 무작위배정**: 어떤 프로그램이 모두에게 제공되고 있는 상황에서, 해당 프로그램의 참여율이 낮은 경우, 참여율을 높이기 위한 다양한 인센티브를 고안하여 무작위로 선정된 일부 그룹에게만 제공.

- 예를 들어, 에티오피아에서는 기본적인 가족계획 서비스가 지역 보건소에서 모두에게 무료로 제공되고 있지만 보건소까지 갈 때 소요되는 교통비 부담 등으로 가족계획 서비스 이용률은 높지 않은 편인데, 이 경우 무작위로 교통비 보조금을 제공하여 실험군의 가족계획 서비스 이용율을 높여볼 수 있음.

□ 무작위배정방식이 결정된 후, 고려해야 할 사항은 무작위배정단위임.

○ 무작위배정단위는 프로그램이 진행되는 단위에서 이루어지는데, 보통 개인 수준이나 마을 단위에서 무작위배정이 이루어짐.

○ 하지만 실험군과 통제군 사이에 파급효과(spillover)가 발생할 경우, 무작위배정단위를 파급효과가 발생할 수 있는 단위로 조정하는 것이 필요함.

- Miguel and Kremer(2004)의 경우, 회충 감염 특징상 학생 수준에서는 구충제 사업의 긍정적인 파급효과가 구충제 사업에 참여하지 않은 통제군 학생에게 영향을 미칠 수 있기 때문에, 학생 개인 수준이 아닌 학교 단위로 무작위배정단위를 설정.

○ 이러한 파급효과를 측정하는 것이 주요 연구 주제라면 개인 수준과 그룹 수준에서 2단계 무작위배정방식을 적용해 볼 수 있음.

- <표 2-6>은 Kim et al.(2016)에서 적용한 2단계 무작위배정 방식을 보여줌. 포경수술을 받으러 가는 결정은 학급 내 친구

들 사이에서 파급효과가 발생할 수 있기 때문에, 1단계에서는 우선 학급 수준에서 무작위배정을 하였고, 2단계에서는 50% 프로그램 제공 학급에 대하여 개인 수준에서 학급의 절반 남학생을 무작위로 실험군에 배정. 이 경우 Group 3(본인은 포경수술 프로그램을 받지 못했으나 같은 학급 나머지 절반의 친구들은 프로그램 제안을 받은 경우)과 Group 4(본인도 포경수술 프로그램 제안을 받지 못했고, 나머지 학급 친구들도 받지 못한 경우)의 특성을 비교함으로써 파급효과를 측정할 수 있음.

〈표 2-6〉 말라위 남학생 포경수술사업 연구설계

	그룹	배정	학급 수	학생 수
100% 프로그램 제공	G1	실험군	41	1,293
50% 프로그램 제공	G2	실험군	41	679
	G3	통제군		679
0% 프로그램 제공	G4	통제군	42	1,323
총계			124	3,974

주: 무작위배정은 두 단계로 이루어짐. 우선 학급 단위에서 100% 프로그램 제공학급, 50% 프로그램 제공학급, 0% 프로그램 제공학급을 선정한 후, 50% 프로그램 제공 학급 중에서 학생 개인 수준에서 절반을 무작위로 선정하여 실험군으로 배정.

자료: Kim et al.(2016). p.5.

○ 또한 무작위배정은 그룹 내에서 감정적 문제를 발생시킬 수 있음.

- 예를 들어 같은 학급에서 공부하고 있는데 일부 학생은 장학금을 받고 다른 학생들은 장학금 수혜를 받지 못하는 경우, 통제군에 무작위로 배정된 학생들이 강하게 불만을 표시할 수도 있고, 그로 인해 통제군 학생들의 사업종료 설문조사 참여율이 낮아질 수도 있음.
- 이 경우 무작위배정단위를 학급으로 올리게 되면 이러한 불만

들을 어느 정도 통제할 수 있음. 최소한 같은 학급, 같은 마을에 사는 사람들은 모두 똑같이 프로그램 혜택을 받거나 받지 못하게 되기 때문임.

- Kim(2014)의 경우 여학생 장학금 프로그램을 운영하면서 학생 개인 수준에서 무작위배정을 하지 않고, 학급 수준에서 무작위배정을 하였는데, 학급 내 학생들의 불만, 사기 저하, 차별감 등을 고려하여 배정단위를 학생 개인이 아닌 학급으로 상향 조정.

○ 하지만 무작위배정단위를 개인 수준에서 그룹 수준으로 올리게 되면 파급효과 문제나 그룹 내 불만을 줄일 수 있지만, 검정력이 줄어들게 되는 단점이 있으므로, 무작위배정단위를 결정할 때는 사전에 검정력 계산을 함께 진행하는 것이 필요함.

8) 설문조사

□ 설문조사는 현장 무작위통제연구에서 가장 중요한 과정 중 하나로, 프로그램을 진행하기 전 기초선조사를 실시하고, 프로그램이 끝나 후 사업종료 조사를 수행하는 것이 일반적임.

□ 특히 기초선조사는 현장연구에 있어서 가장 중요한 기초이므로 기초선조사를 신중하게 계획하는 것이 중요함.

○ 설문조사요원들을 고용하여 훈련하는 것이 기초선조사 준비의 첫 단계인데, 현장연구에서 설문조사 경험이 있는 조사요원들을 선발하는 것이 좋음.

- Kim et al.(2016)의 경우 기초선조사 설문조사요원들을 선발

하기 위해 말라위 통계청에 연락하여 사업 대상지인 릴롱웨 지역에서 직전 말라위 DHS 2016 설문조사에 조사요원으로 참여한 사람들의 명단을 받아서 해당 명단에서 설문조사요원들을 선발.

- 조사요원들을 고용한 후에 집중적인 훈련이 필요한데, 우선 설문조사 문항들에 대해 조사요원들이 완벽하게 이해할 수 있도록 하는 교육이 필요하며, 설문조사 수행 전체 과정에 대해서도 충분한 숙지가 필요함.
- 기초선조사를 실시하기 전에 준비 설문조사(Pilot Survey)를 몇 차례 실시하여 설문조사 문항 및 설문조사 진행방식의 문제점들을 점검해볼 수 있고, 동시에 조사요원들이 기초선조사를 실시할 수 있는 준비가 되었는지 확인.

〈표 2-7〉 준비조사를 통한 설문조사요원 교육

	설문조사 소요시간	전체 실수	Skip 패턴	Missing Value	특이값	기타	데이터 입력
1차 준비조사	1시간 49분	25.93	2.17	14.53	0.33	2.77	6.13
2차 준비조사	1시간 35분	1.65	0.06	1.06	0.00	0.16	0.37

주: 총 변수 숫자는 1차 준비조사는 873개, 2차 준비조사는 880개였음. 2차 준비조사 변수 숫자는 1차 준비조사 이후 설문지 수정 등을 통해 다소 변경되었음. 전체 실수는 다섯가지 유형으로 분류됨. 첫째 skip 패턴을 실수한 경우, 둘째 변수 값을 missing한 경우, 셋째, 특이값은 해당변수 평균값에서 4 표준편차 이상의 차이를 나타낸 경우, 마지막으로 데이터 숫자를 잘못 입력한 경우임.

자료: 저자 작성.

- 코이카 미얀마 새마을운동사업 기초선조사를 앞두고 두 차례 준비조사를 실시하였는데, 첫 번째 준비조사에서는 설문조사요원들이 평균 26개의 실수를 하였지만, 1차 준비조사 후에

조사요원 재교육을 실시하고 2차 준비조사를 실시하였을 때에는 조사요원들의 실수가 크게 감소(표 2-7 참조). 2차 준비조사에서 조사요원들의 정확도 및 속도가 크게 향상된 것을 확인한 후에 기초선조사를 실시하였음.

□ 기초선조사 이후 프로그램 진행이 끝난 후 실시하는 종료조사에서 중요한 이슈 중 하나는 종료조사 비응답(Attrition)임.

○ 종료조사에서 조사하지 못했던 비응답자를 대상으로 추적 설문조사를 실행하는 것이 일반적임.

- 만약 비응답 비율이 실험군과 통제군이 다를 경우 무작위통제 연구의 내적 정합성을 위협할 수 있기 때문에, 기초선조사를 진행하는 시점에서 1년 뒤 또는 몇 년 뒤 종료조사 시 해당 대상자와 연락할 수 있는 다양한 추적 정보를 물어보는 것이 중요함.
- Kim(2014)은 사업종료 설문조사를 실시하였는데, 기초선 설문조사 대비 68.4%(2,733명)의 학생을 종료 설문조사에서 조사할 수 있었고, 나머지 31.6%(1,264명)는 결석, 전학, 중퇴 등의 이유로 종료 설문조사를 실시하지 못했음. 종료조사 비응답자 1,264명 중 15%(187명)를 무작위로 추출하여 비응답 전체에 대한 대표 샘플을 구축한 후, 이들 대상으로 추적 설문조사를 실시(Thomas et al., 2001; Thomas et al., 2012). 187명의 추적조사 대상자 학생 중 128명의 학생을 추적하는데 성공하여 가정방문 설문조사를 진행하였고, 이를 통해 결과적으로 90.4%의 유효설문조사율(Effective Survey Rate: ESR)을 달성.¹²⁾

□ 설문조사 실시 후, 조사된 결과를 데이터로 입력하는 해야 하는데, 이러한 데이터 입력 과정이 무작위통제연구 분석하기 전 마지막 단계.

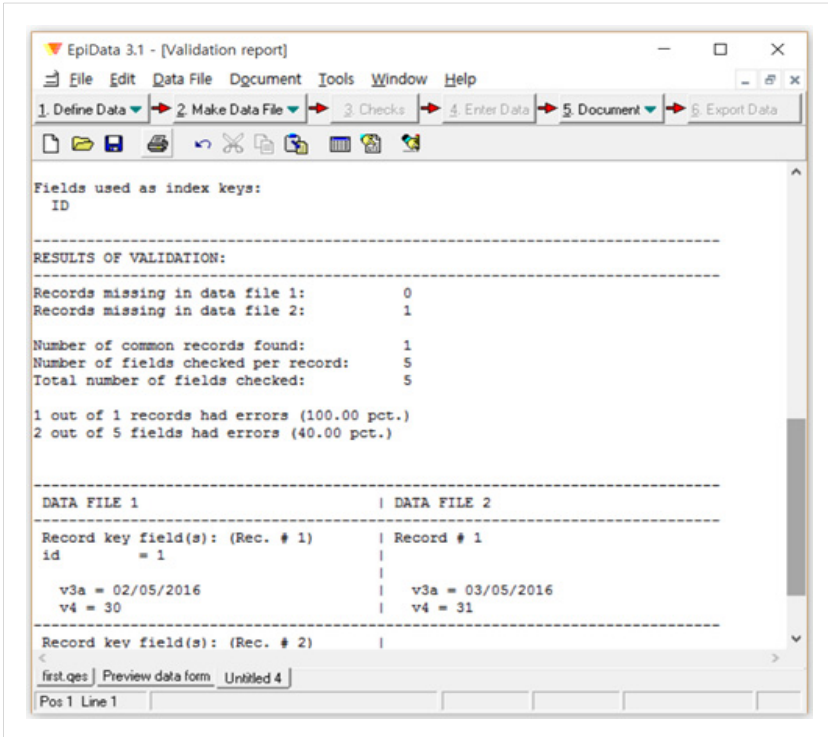
○ 데이터 입력 시 매우 많은 실수가 발생하기 때문에, 단일입력방식 보다는 이중입력방식(Double Data Entry)을 선호.

- 이중입력방식은 동일한 데이터를 서로 다른 데이터 입력 요원이 각각 입력한 후에 입력값의 차이를 비교하여 오류를 수정하는 방식으로, 단일입력방식에 비해 데이터 입력의 정확성을 크게 높여줄 수 있음. 하지만 이중입력방식의 경우에 비용과 시간이 크게 늘어나게 됨.
- [그림 2-7]은 5개 변수를 이중입력방식으로 입력한 후, 이 중 2개 변수(v3a 및 v4)에서 이중입력값이 다른 경우를 보여줌. 이러한 차이를 확인하게 되면, 데이터 입력 요원이 설문조사 원자료를 확인해서 최종적으로 정확한 값을 확정.
- 데이터 입력을 위해 특별히 설계된 EpiData와 같은 소프트웨어를 이용하는 것이 좋음.¹³⁾ EpiData 소프트웨어는 각 변수별 적합한 입력값 범위 등을 사전에 프로그램할 수 있어서, 적합한 범위를 넘어서는 값을 실수로 입력한 경우, 자동으로 데이터 입력 에러를 알려줌. 또한 EpiData는 이중입력방식 기능도 제공.

12) 유효설문조사율(Effective Survey Rate: ESR)은 종료조사 성공율과 추적조사 성공율의 합수임. 유효설문조사율 = 종료설문조사율 + (1-종료설문조사율)×추적설문조사 성공률로 계산함. Kim(2014)의 경우 종료조사 비응답 학생에 대한 15% 샘플 추출을 진행했으므로 추적조사 학생에 대한 가중치 6.67을 부여. 말라위 여학생 장학금 프로그램의 유효설문조사율을 90.4%(68.4% + 31.6%×69.6%)임. 보통 유효설문조사율을 90% 이상으로 달성하는 것을 바람직하게 생각하며 Thomas et al.(2012)의 경우 95% 유효설문조사율을 달성하였음.

13) EpiData 소프트웨어는 <http://epidata.dk/>에서 무료로 다운로드받을 수 있음.

[그림 2-7] 데이터 이중입력방식



자료: EpiData 이용 저자 작성.

- 최근 들어 종이 설문조사 방식 대신 모바일 설문조사 방식이 많이 활용됨.
 - 모바일 설문조사 방식의 가장 큰 장점은 설문조사와 동시에 데이터 입력이 되기 때문에 추가적인 데이터 입력 과정이 필요 없다는 점임.
 - CommCare라는 모바일 설문조사 플랫폼이 가장 많이 사용됨.¹⁴⁾
 - 이호철 등(2016)에서는 에티오피아 모자보건사업 설문조사

를 CommCare에서 어떻게 구축하였는지 단계별로 자세히 소개함.

나. 무작위통제연구 분석

1) 기초선조사 밸런스

□ 무작위통제연구 분석을 시작할 때 가장 먼저 확인하는 것은 실험군과 통제군 간의 기초선조사 변수들의 밸런스가 맞는지를 살펴보는 것임.

○ 무작위배정을 하게 되면 실험군과 통제군 사이에서 다양한 기초선조사 변수들의 특성이 평균적으로 비슷해짐.

- Kim(2014)은 말라위 여학생 장학금 프로그램 연구에서 회귀 분석을 통해 실험군과 통제군 간의 기초선조사 변수 밸런스를 확인.

- 〈표 2-8〉 열 (2)를 보면 다양한 기초선조사 변수상에서 아버지 고등교육 여부 변수를 제외하고는 실험군과 통제군 간의 차이가 거의 없음을 볼 수 있음. 또한 F-test 값이 통계적으로 유의미하지 않아($p = 0.325$), 모든 기초선조사 변수에서 두 그룹 간의 차이가 없다는 귀무가설을 기각할 수 없음. 이는 실험군과 통제군이 그룹 평균적으로 다양한 기초선조사 변수상 다르지 않음을 보여줌.

14) CommCare는 5가지 서비스(Scope, Launch, Boost, Growth, and Scale)를 제공하고 있다. 이 중 가장 기초적인 Scope 서비스는 무료로 사용할 수 있음.
<https://www.commcarehq.org/home/>

〈표 2-8〉 기초선조사 변수에서의 밸런스

종속변수	평균 (표준편차) (1)	여학생 장학금 프로그램 실험군 배정 여부 (2)
나이 (세)	16.16 (1.856)	0.003 (0.011)
고아 여부	0.054 (0.227)	-0.037 (0.032)
아버지 고등교육 여부	0.198 (0.399)	0.047** (0.022)
어머니 고등교육 여부	0.082 (0.274)	-0.024 (0.030)
아버지 직업	0.256 (0.436)	-0.031 (0.023)
어머니 직업	0.106 (0.307)	-0.017 (0.027)
가구 자산 (0-16)	7.59 (3.455)	0.003 (0.007)
학교 유형	0.245 (0.430)	0.086 (0.102)
p-value of joint F-test		0.325
관측수		3,993
R-squared		0.055

주: 1) 고아 여부는 두 부모 모두 사망시 =1. 부모 고등교육 여부는 2년제 또는 4년제 대학교육시 =1. 부모 직업은 전문직, 정부공무원일 경우 =1. 가구 자산은 16개 자산 문항에서 소유하고 있는 자산 개수. 학교 유형은 Community Day Secondary School이 아닌 Conventional Secondary School에 재학할 경우 =1.

2) 두 번째 열은 여학생 장학금 프로그램의 무작위배정 밸런스를 나타냄. 강건(Robust) 표준 오차는 학급 레벨에서 clustering 하였음. 괄호 안은 표준오차.

3) *** p<0.01, ** p<0.05, * p<0.1

자료: Kim(2014). p.10.

2) 종료선조사 Attrition 밸런스

□ 실험군, 통제군 간 기초선조사 밸런스를 확인한 후, 다음으로 해야 할 분석은 종료조사 비응답(attrition) 여부의 밸런스 테스트임.

- 만약 두 그룹 간의 비응답 여부가 다르게 된다면, 이 역시 무작위 통제연구의 내적 정합성에 위협이 됨.
- Kim(2014)은 회귀분석을 통해 여학생 장학금 프로그램의 비응답 여부 밸런스를 확인.

〈표 2-9〉 종료조사 비응답 여부 밸런스

종속변수	종료설문조사 또는 추적설문조사에 응답한 경우 = 1			
	Treatment (1)	Adjusted (2)	Main effect (3)	Interaction (4)
여학생 장학금 프로그램 배정	0.040* (0.022)	0.038* (0.022)		0.068 (0.207)
나이		-0.015** (0.006)	-0.012 (0.008)	-0.006 (0.011)
고아 여부		-0.081** (0.039)	-0.071 (0.057)	-0.025 (0.079)
아버지 고등교육		-0.003 (0.025)	-0.009 (0.038)	0.001 (0.051)
어머니 고등교육		-0.070** (0.034)	-0.069 (0.050)	-0.000 (0.069)
아버지 직업		-0.024 (0.017)	-0.041 (0.027)	0.035 (0.034)
어머니 직업		-0.028 (0.027)	-0.004 (0.038)	-0.049 (0.053)
가구 자산 (0-16)		-0.001 (0.003)	-0.005 (0.004)	0.006 (0.007)
학교 유형		0.049** (0.023)	-0.008 (0.035)	0.102** (0.048)
관측수	3,997	3,993	3,993	
R-squared	0.014	0.024	0.027	

주: 1) 열 (3)과 열 (4)는 하나의 회귀분석 결과임. 행 (3)은 main effects, 열(4)는 장학금 프로그램 배정 변수와 다양한 기초선 변수 간의 interaction 변수임.
2) ***p<0.01, **p<0.05, *p<0.10
자료: Kim(2014). p.12.

- 〈표 2-9〉 열 (2)를 보면 다양한 기초선 변수를 통제한 이후에도 장학금을 받았던 실험군 여학생들이 3.8%p 더 종료 설

문조사에 참여하였고 이로 인해 attrition 편이가 발생하였음. 이러한 경우, attrition 편이가 기초선 특성과 연관되어 있는지를 교차항 분석으로 열 (3), 열 (4)에서 살펴보았는데, attrition 편이가 기초선 특성과 연관되어 있지는 않았음 (Thomas et al., 2001).

3) Intention-to-treat(ITT) 분석

□ Intention-to-treat(ITT) 분석은 무작위로 배정된 실험군 전체 평균과 통제군 전체 평균을 비교하는 방식임.

○ 기초선 밸런스와 종료조사 무응답(attrition) 밸런스를 확인한 후, 결과변수에 대한 프로그램 또는 정책의 영향을 분석할 때 가장 기본적으로 사용되는 분석 방식임.

○ ITT 분석은 실험군 내에서 실제로 프로그램에 참여한 사람들의 영향을 평가하지는 않는다. 왜냐하면 실험군에 배정되었더라도, 모두가 프로그램에 참여하는 것은 아니기 때문이다. 실제로 프로그램에 참여한 사람들만의 영향을 평가하는 방식은 Treatment of treated(TOT) 분석이라 한다. 하지만 TOT 분석의 경우, 실제로 프로그램에 참여한 사람들의 영향만을 평가하기 때문에 ITT보다 더 강한 결과를 도출할 수 있으나, 실험군에 배정되어 실제적으로 프로그램에 참여한 사람들의 특성이 실험군에 배정되어 프로그램에 참여하지 않은 사람들과 체계적으로 다르다면, 결과적으로 실험군 내 프로그램 참여자의 평균 특성과 통제군 집단의 평균 특성 간의 사전 기초선 밸런스가 깨지게 되어 TOT 분석 결과에 편이가 발생하게 된다. 따라서 결과의 정도는 좀 작더라도

편의가 발생하지 않는 ITT 분석을 더 선호함.

- ITT 분석을 할 때 보통 통제변수 벡터를 추가하지 않은 상황에서 결과변수를 우선 추정하고, 그 후 통제변수 벡터를 추가한 상황에서 다시 한번 결과변수를 추정.
- 만약 무작위배정으로 실험군과 통제군 간의 기초선 변수 밸런스가 완벽하게 맞는다면 통제변수를 추가하는 것이 결과변수 추정치에 아무런 영향을 미치지 못하지만 이러한 분석 방식을 통해 결과변수 추정치가 강건(robust)하다는 것을 보여줄 수 있고, 동시에 통제변수들이 결과변수 추정치의 정확도(precision)를 높여주는 효과도 기대할 수 있음.
- Kim(2014)은 여학생 장학금 프로그램이 학교 출석에 미치는 영향을 ITT로 분석.

〈표 2-10〉 학교 출석 ITT 분석

종속변수	학기 당 결석일수			
	(1)	(2)	(3)	(4)
여학생 장학금 프로그램	-1.707*** (0.345)	-1.645*** (0.275)	-2.187*** (0.437)	-2.187*** (0.420)
통제군 평균	3.794		0.672	
기초선 통제변수	x	o	x	o
관측수	2,715	2,704	2,700	2,689
R-squared	0.027	0.047	0.027	0.042

주: 1) 열 (1), (2)의 경우 종속변수는 종료 설문조사 데이터를 이용함. 열(3), (4)의 경우 종속변수를 종료 설문조사 및 기초선 설문조사의 차이를 이용함. 통제변수로는 나이, 고아 여부, 부모 고등교육 여부, 부모 직업, 가구자산, 학교 유형이 포함.

2) *** p<0.01, ** p<0.05, * p<0.10

자료: Kim(2014). p.16.

- <표 2-10> 열 (1), (3)은 기초선 통제변수를 포함하지 않고 ITT 추정치를 계산하였고, 열 (2), (4)는 통제변수를 포함하고 계산한 결과, 결과변수인 학기 당 결석 일수 계수 추정치는 거의 변화가 없고(실험군 여학생이 평균적으로 학기당 1.7일 또는 1.6일 더 적게 결석함), 통제변수를 추가하였을 때 표준오차가 0.345에서 0.275로 감소하여 정확도(precision)가 향상되는 것을 볼 수 있음.

4) Heterogeneity 분석

- 프로그램 영향에 대한 ITT 분석을 한 후, 특정 소그룹에 대해서 프로그램이 이질적인 사업효과(heterogeneous treatment effect)를 가질 수 있다고 생각되는 경우, heterogeneity 분석을 할 수 있음.
- 실험군 무작위배정 더미 변수와 상이한 사업효과가 발생할 수 있는 소그룹 간의 교차(interaction) 변수를 생성하여 회귀분석을 진행.
- Heterogeneity 분석을 하기 위해 해당 소그룹만 남기고 나머지 샘플을 삭제하는 소집단(sub-group) 분석을 할 수도 있으나, 이 경우에 샘플 숫자가 작아져 검정력을 확보하기 어려울 수 있으므로 소집단 분석보다는 전체 샘플을 다 이용하는 교차항(interaction terms)을 이용한 heterogeneity 분석을 보통 사용.
- Kim et al.(2016)의 경우 포경수술에 대한 사전적 생각이 포경수술을 받는 데 영향을 끼칠 수 있다고 생각하고 사전적 생각에 대한 heterogeneity 분석을 진행.

- <표 2-11> 열 (3)을 보면 사전에 포경수술이 매우 고통스러울 것이라고 생각하는 학생이 그렇지 않은 학생에 비해 3.8%p 더 적게 포경수술을 받는 것으로 분석되었지만 사전에 포경수술의 에이즈 예방 효과 등을 알고 있거나, 포경수술이 종교적으로 무슬림만 받아야 한다고 생각하는 경우에는 유의미하게 다른(heterogenous) 효과가 발견되지 않았음.

<표 2-11> 사전적 생각에 따른 Heterogeneous 효과

종속변수	포경수술 받았는지 여부			
	(1)	(2)	(3)	(4)
포경수술 실험군 배정	0.146*** (0.021)	0.144*** (0.023)	0.162*** (0.023)	0.151*** (0.023)
포경수술의 에이즈예방효과를 알고 있는 경우		0.007 (0.012)		
실험군 배정 x 포경수술의 에이즈 예방 효과를 알고 있는 경우		0.003 (0.023)		
포경수술이 매우 고통스럽다고 생각함			-0.023** (0.010)	
실험군 배정 x 포경수술이 매우 고통스럽다고 생각함			-0.038* (0.021)	
포경수술은 무슬림만 받는다고 생각				-0.014 (0.016)
실험군 배정 x 포경수술은 무슬림만 받는다고 생각				-0.032 (0.029)
관측수	3,952	3,949	3,945	3,942

주: *** p<0.01, ** p<0.05, * p<0.10
자료: Kim et al.(2016). p.14.

제 3 장

주요 공여국 평가체계와 사례: 보건의료 및 사회정책분야를 중심으로

제1절 세계보건기구(WHO)

제2절 영국

제3절 일본

3

주요 공여국 평가체계와 사례: << 보건의료 및 사회정책분야를 중심으로

제1절 세계보건기구(WHO)

1. 개요

□ WHO 원조 평가의 특성

- WHO는 2005년 파리선언(Paris Declaration)을 바탕으로 보건 원조의 효과성을 분석하는 프레임워크를 제시하고 있으며 보건원조의 모니터링과 평가의 중요성을 강조하였음.
 - 파리선언에서는 국가 우선순위와 원조 자금의 부합성(alignment), 국가 조달 체계(Procurement system) 활용, 원조의 예측 가능성, 일반적 배열과 절차(common arrangements and procedures)의 활용, 결과 중심의 프레임워크에 대한 고려를 강조하였음.
- WHO의 모니터링 및 평가는 개별 사업에 대한 평가라기보다 각 국가를 대상으로 한 원조사업의 규모를 모니터링하거나 국가별 새천년개발목표(MDGs: Millenium Development Goals) 달성 정도와 같이 보건원조사업의 영향(impact)을 평가하는 차원에서 이루어지고 있음.
 - 평가를 위한 구체적인 틀이나 지침을 제시하기보다 평가를 위해 고려해야 하는 요소나 평가 원칙 등을 제시하고 있음.
 - WHO는 수원국별 보건 ODA의 목적과 규모 등 보건 관련

ODA 흐름을 파악하고 있으며, 이에 대한 보고서를 주기적으로 발간하고 있음.

- 최근까지 WHO는 보건 관련 MDG를 중심으로 각 국가의 보건 관련 MDG 이행과 달성 정도를 파악하는 것에 초점을 두었으며, 이행과 달성에 영향을 주는 주요 요소로서 보건원조를 모니터링하고 평가하였음.
- MDG와 최근의 지속 가능한 개발목표(SDGs: Sustainable Development Goals)의 이행과 달성 정도를 파악하는 것은 결국 보건 사업의 성과와 영향을 평가하는 것을 의미함.

2. WHO의 원조 평가사례

□ ‘From whom to whom’ 보고서(WHO, 2012)에 나타난 보건 ODA 흐름¹⁵⁾

○ 수원국별 보건 ODA의 목적과 규모(volume)를 파악하고 보건 관련 사업의 재정적 불균등에 대한 정보를 제공하기 위해 구성되었음.

- 보건 ODA의 투명성을 증진하고, 보건 자원들에 대한 전반적 검토(overview)를 통해 보건계획 수립에 근거를 제공함.
- 기여국이 기존 보건 ODA 자원, 규모, 목적을 수원국별로 파악할 수 있도록 정보를 제공하여 원조 계획을 더욱 효과적으로 수립할 수 있도록 하는 데 기여함.
- 수원국 보건사업의 주요 기여국 및 기구, 계획된 약정(commit-

15) 이 부분은 WHO(2012), From whom to whom. 보고서의 내용을 요약 정리한 것임.

ments)과 실행을 위한 자금 등 외부 펀딩에 대한 정보를 제공함으로써 수원국 차원에서는 국가 개발 계획 수립 시 자원과 예산 분배를 위한 정보를 정리하는 데 효과적임.

○ 2010년에 500,000명의 인구를 가진 119개 WHO 회원국을 포함. 이 보고서에 담긴 자료들은 2009~2010년 전 세계적으로 분배된 모든 보건 관련 ODA의 75%를 망라함.

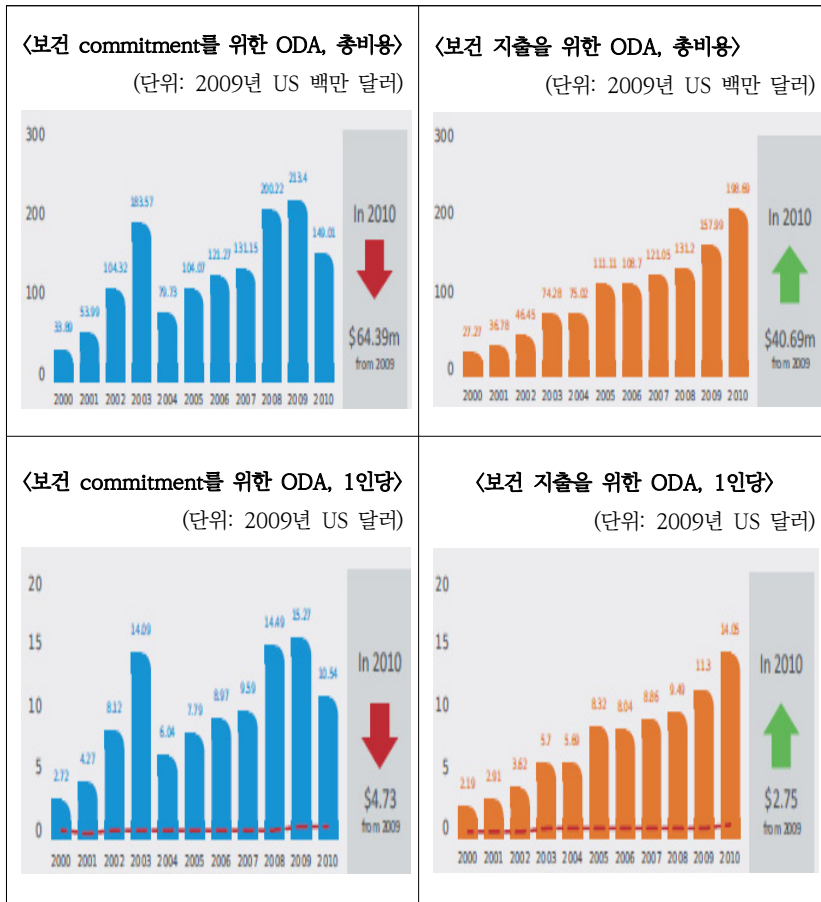
- OECD의 Creditor Reporting System database를 활용함.
- 주요 결과: 2000년에 비해 보건 관련 ODA의 실제 자금 이전 (transfer)은 증가하였음. 이는 MDG6인 HIV(인간면역결핍 바이러스)/AIDS(후천성면역결핍증), 결핵, 말라리아 등 질병 퇴치를 위한 자금 증가와 관련이 높음.
- 2009~2010년 78개 수원국에서 단일(single) 보건 ODA 지출(disbursement)에서 가장 많이 나타난 목적이 MDG6였음.
- 47개 국가에서 MDG6를 위한 ODA가 보건 관련 지출된 총 ODA의 50% 이상을 차지함.

○ 평가 내용

- 기본적으로 119개 회원국별로 정리가 되어있으며, 회원국의 기본 인구 데이터, ODA 흐름, 보건 지출 등에 대한 일반적인 정보와 더불어 총 ODA 중 보건 ODA가 차지하는 비중, 1인당 보건 ODA(health ODA per Capita), 동일 WHO 지역 내 1인당 평균 ODA 약정과 지출 현황과 변화 추이가 제시됨.
- 국가별로 내용이 제시되어 있음. 캄보디아의 사례를 들어 살펴보면, 캄보디아의 보건분야 약정(commitment)을 위한

ODA 총비용은 2000년 대비 증가하였으나 2008년과 2009년 대비 감소하였음. 보건 지출을 위한 ODA 총비용은 2000년 이후 지속적으로 증가하고 있는 추세임.

[그림 3-1] 사례: 캄보디아의 보건 ODA 흐름(2000~2010년)



자료: WHO(2012). From whom to whom (2nd). p.49.

- 또한 보건 정책 및 행정 관리, MDG6(HIV/AIDS, 결핵, 말라리아 등 질병 퇴치), 재생산 건강 및 가족계획, 기타 보건 목적으로 구분하여 영역별 ODA 분배 비중이 제시되어 있음.
- 2009~2010년 양자 및 다자기구에서 각 회원국이 얼마의 보건 ODA를 받았는지에 대한 건수와 금액이 제시되어 있음. 또한 가장 많은 기여를 한 상위 5개 국가 및 다자기구의 기여 비중을 제시하고 있음.

[그림 3-2] 캄보디아의 보건 영역별 ODA 분배 비중



자료: WHO(2012). From whom to whom (2nd). p.49.

[그림 3-3] 수원국별·공여기관별 보건 개발사업 건수와 금액

HOW MUCH DONORS DISBURSED IN 2009-2010					
Bilateral	Number of Disbursements	Total (in million euro/US dollars)	Multilateral	Number of Disbursements	Total (in million euro/US dollars)
Australia	73	22.03	ARDF	-	-
Austria	1	0.00	AFESD	-	-
Belgium	15	4.30	AsDB Special Fund	1	0.63
Canada	47	0.70	EU Institutions	22	7.63
Denmark	5	1.44	GAVI	8	8.31
Finland	3	0.29	Global Fund	22	107.06
France	16	5.63	IDA	28	18.68
Germany	43	14.37	IDB Special Fund	-	-
Greece	-	-	ORD	-	-
Ireland	-	-	UNAIDS	2	1.56
Italy	4	0.36	UNDP	2	0.05
Japan	42	20.56	UNFPA	65	11.09
Korea	61	1.78	UNICEF	59	4.49
Kuwait	-	-	UNPBF	-	-
Luxembourg	6	0.55	UNRWA	-	-
Netherlands	-	-	WFP	4	4.33
New Zealand	1	0.02	TOTAL	213	163.84
Norway	4	0.57	ACER/AFRO: AFRO - WHO African Region; AMRO - WHO Region of the Americas; EMRO - WHO Eastern Mediterranean Region; EURO - WHO European Region; WDOs - Millennium Development Goals; SEARO - WHO South-East Asia Region; WPRO-WHO Western Pacific Region; ARDF: African Development Fund; AFESD: Arab Fund for Economic and Social Development; AsDB Special Fund - Asian Development Bank; EU Institutions - European Commission, European Community; GAVI - Global Alliance for Vaccines and Immunization; Global Fund - Global Fund for HIV/AIDS, Tuberculosis and Malaria; IDA-International Development Association (World Bank); IOR SpF-Fund - Inter American Development Bank, Special Fund; OFID - OPIC Fund for International Development; UN - United Nations; UNAIDS - Joint United Nations Programme on HIV/AIDS; UNDP - United Nations Development Programme; UNRWA - United Nations Relief and Works Agency; UNICEF - United Nations Children's Fund; UNPBF - United Nations Peacekeeping Fund; UNRWA - United Nations Relief and Works Agency; WFP - World Food Programme.		
Portugal	-	-			
Spain	2	0.03			
Sweden	24	0.32			
Switzerland	4	5.50			
United Arab Emirates	-	-			
United Kingdom	17	33.32			
United States of America	324	81.08			
TOTAL	692	192.83			

자료: WHO(2012). From whom to whom (2nd). p.50.

□ 모니터링 및 평가

○ 각국의 보건사업과 정책을 모니터링하고 평가하기 위해 WHO에서 지금까지 사용한 주요 지표는 보건 관련 MDG였음. 최근 보건 관련 SDG가 제시됨에 따라 큰 목표는 구성되었지만 이를 측정하고, 모니터링하고 보건사업을 평가할 수 있는 지표들은 구성되고 있는 중임.

- 특히 모자보건과 관련하여 MDG4(아동 사망 감소), MDG5(모성 건강 향상)에 따라 5세 미만 아동의 사망률, 아동의 홍역 백신 적용 범위(coverage), 모성사망비, 숙련된 보건인력이 참여한 출산 경험율, 산전 관리율, 피임 실천율 등이 지표로 선정되어 목표치가 제시되었으며, 매년 MDG 달성 정도가 모니

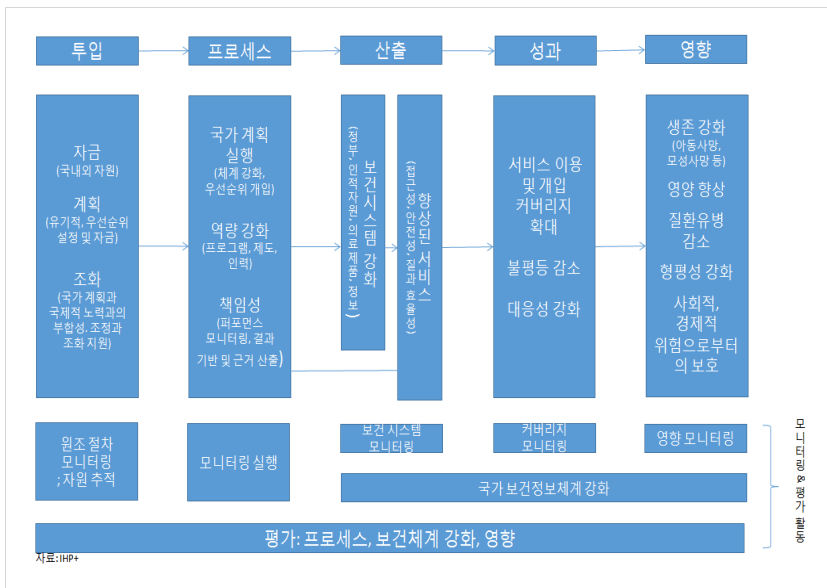
터링되고 평가되었음.

- WHO는 모니터링 및 평가의 주요 원칙으로 공평성(Impartiality), 독립성(Independence), 효용성(Utility), 질(Quality), 투명성(Transparency) 5가지를 제시함(WHO, 2013).
 - WHO는 평가 문화와 조직적 학습(Organizational Learning) 강화를 강조하면서 참여적 접근(Participatory approach), 성, 형평성 및 인권에 대한 다영역적 통합적인 전략을 강조함.
 - WHO는 본부 및 지역 사무소, 국가 사무소 등 조직 전반에 대한 평가를 수행하며, 내부 평가 및 외부 평가를 고루 활용하고 있음. 최근 들어 WHO와 다른 조직 및 기관이 참여하는 공동 평가(Joint Evaluation)가 강화되는 추세임.
 - WHO의 평가 유형을 살펴보면, 주제별 평가(Thematic evaluation), 프로그램 평가(Programmatic evaluation), 특정 사무소에 대한 평가(Office-specific evaluation)가 있음.
 - 프로그램 평가가 단일 사업, 프로그램에 대한 평가라고 한다면 주제별 평가는 프로젝트평가의 개념과 유사함. 영역에 대한 효과성, 지속 가능성, 적용 가능성 등에 대한 심도 있는 평가를 진행하는 것이 주제별 평가임.
 - WHO의 일반적인 평가기준은 타당성(Relevance), 효율성(Efficiency), 효과성(Effectiveness), 지속 가능성(Sustainability), 영향(Impact)이며, 그 외에도 효용성(Utility), 형평성(Equity), 유기성(Coherence), 시너지(Synergy), 부가성(Additionality), 중량(Deadweight), 대체성(Displacement)을 제시하고 있음.

○ 보건분야 이니셔티브들을 모니터링하고 평가하기 위한 전략적 프레임워크

- 모니터링 및 평가의 전략적 프레임워크는 International Health Partnership(IHP+) 워킹그룹에 의해 디자인된 것임.

[그림 3-4] IHP+의 모니터링 및 평가 프레임워크



자료: WHO(2008). Effective Aid Better Health. p.27 Figure 3 수정 번역.

- 집합적 활동(Collective Action), 국가 프로세스와의 부합성 (Alignment with country processes), 국가 참여와 독립성 간의 균형(Balance between country participation and independence), 조화된 접근(Harmonized approaches), 역량강화와 보건 정보 강화(Capacity building and health information strengthening), 적절한 펀딩(Adequate funding)

의 6가지 원칙을 내세움.

- 재정, 서비스 전달, 의약품, 보건 통계 및 거버넌스 등의 영역에서
의 진전 사항을 보여줄 수 있는 보건체계 지표들을 포함함.

□ 캄보디아 보건 사업에 대한 평가¹⁶⁾

- 보건 관련 MDG 달성에 대한 고위급 포럼 이후 캄보디아 국 가
사례연구를 진행함(WHO and Ministry of Health Kingdom
of Cambodia, 2007).
- 캄보디아 국가 사례연구를 통해 MDG 달성을 저해하는 인적
자원, 제도적 구조에 대한 고려와 더불어 재정적·행정적인 절
차에 대한 평가를 수행하였음.
 - 캄보디아의 건강수준과 보건 관련 MDG 달성 여부 및 가능
성, MDG 달성 저해 요인 및 향후 개선방안에 대한 연구를
진행하였음.
- WHO and Ministry of Health Kingdom of Cambodia
(2007)는 캄보디아 정부의 보건 관련 예산 규모보다 원조 자
금의 규모를 파악하고 재정 규모를 키우거나, 보건 서비스 분
야의 규모를 키우는 것 등의 시나리오를 구축하여 각 시나리오
가 캄보디아의 MDG 달성에 어떤 영향을 미치는지를 분석하
였음.

○ 캄보디아의 건강수준

- WHO and Ministry of Health Kingdom of Cambodia

16) 이 부분은 'WHO and Ministry of Health Kingdom of Cambodia(2007). Scaling up for Better Health in Cambodia. WHO'의 내용을 바탕으로 작성하였음.

(2007)는 캄보디아의 건강수준은 소득수준이 증가하고 보건 의료 비용이 감소하고, 보건의료 지출이 증가하면서 향상되고 있다고 보고하였음. 캄보디아의 영아 및 아동 사망률이 감소하고 출산율이 낮아졌으며, 산전 관리율 및 HIV/AIDS 유병률 역시 감소하여 MDG 달성을 목전에 두고 있다고 평가하였음. 하지만 모성사망, 피임 실패, 숙련된 보건인력이 참여한 출산 경험, 결핵과 말라리아 퇴치는 지속적 노력이 필요하다고 지적함.

〈표 3-1〉 캄보디아 MDG 보건지표

지표	단위	2000	2005 목표	2005 현황	2015 목표	On track
MDG4. 아동사망감소						
신생아 사망	출생아 10만 명당	95	75	66	50	V
5세 미만 아동 사망	출생아 10만 명당	124	105	83	65	V
MDG5 모성건강증진						
모성사망비	출생아 10만 명당	437	343	472	140	
출산율	아동 수	4.0	3.8	3.4	3.0	V
근대화된 피임실패	%	19	30	27	60	
숙련된 보건인력에 의한 출산	%	32	60	44	80	
산전관리율	%	30.5	60	60.2	90	V
MDG6 HIV/AIDS, 말라리아와 기타 질환						
15~49세 성인의 HIV 유병률	10만 명당 비율	30 1/	2.3	0.6	1.8	V
결핵사망률	인구수	90 1/	68	N.A.	32	
공공보건체계에 보고된 말라리아 사망률	%	0.4	0.3	0.36	0.1	
공공보건체계에 보고된 Dengue 사망률	%		1	0.74	0.3	V

자료: WHO and Ministry of Health Kingdom of Cambodia(2007). Scaling up for Better Health in Cambodia. p.5 Table 1.

○ 캄보디아의 MDG 달성 저해 요소

- WHO and Ministry of Health Kingdom of Cambodia (2007)는 캄보디아의 MDG 달성 저해 요소 중 하나로 보건 분야 협력사업이 분절화되어 있고, 캄보디아 국가의 보건 우선순위에 밀접하게 부합하지 않는다는 점을 언급하고 있음. 특히 원조 자금의 효과성, 효율성은 매우 심각한 것으로 평가되었으며, 원조 자금의 규모와 포괄범위(coverage)는 중간 수준인 것으로 평가되었음.

〈표 3-2〉 캄보디아 보건 증진에의 저해 요소

	양/커버리지 저해 정도	효율성/효과성 저해 정도
국내 파이낸싱	중간 (높은 본인부담 지출)	심각 (공공 재정 관리, 낮은 위기관리)
외부 파이낸싱	중간	심각 (부조화와 부합성 낮음)
국내 인적 자원	중간 (지역 행정)	높음 (낮은 공공임금과 공공/민간 기술 수준)
외부 인적 자원	낮음 (광범위한 기술지원)	중상 (중기 목표의 부족)
국내 제도	중간	중상 수준 (복잡한 체계와 변화하는 제도들)
외부 제도	중간 (분절화되어 있음)	중간

자료: WHO and Ministry of Health Kingdom of Cambodia(2007). Scaling up for Better Health in Cambodia. p.33 Table 12.

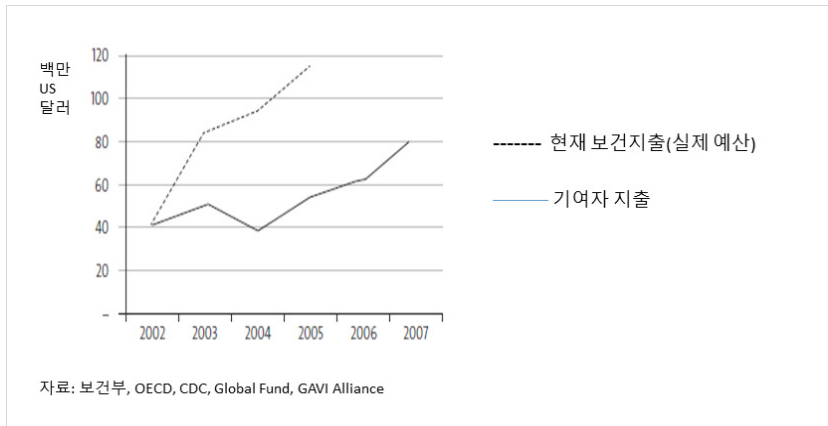
○ 캄보디아 보건 관련 원조 자금 흐름

- 2002년 대비 2007년 보건 관련 지출과 원조 자금은 증가하였음(WHO and Ministry of Health Kingdom of Cambodia, 2007). 그동안 미국, 일본, 영국, 비영리 기구(The Global

Fund to Fight AIDS, Tuberculosis and Malaria)들로부터 원조를 받았음.

- 이 연구는 캄보디아의 보건 관련 정부예산보다 원조 자금이 더욱 많다는 점을 언급하면서 원조기구나 국가와 캄보디아 정부 간 긴밀한 협력이 필요하다고 언급함.

[그림 3-5] 캄보디아 정부 보건분야 공공지출과 원조추정액



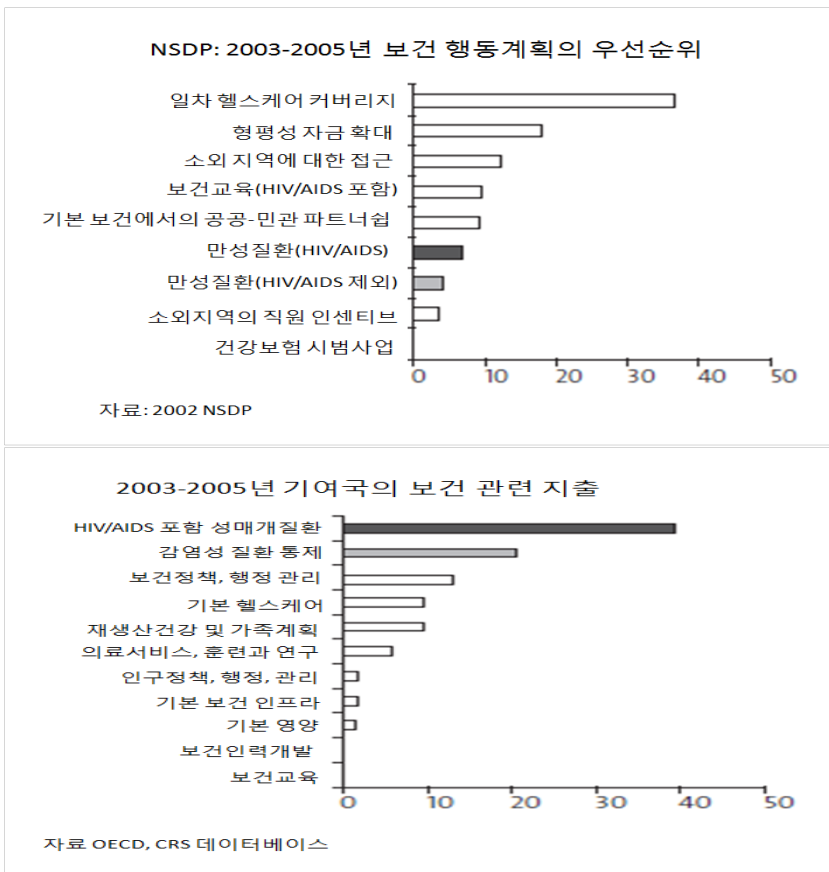
자료: WHO and Ministry of Health Kingdom of Cambodia(2007). Scaling up for better health in Cambodia. p.19 Figure 5.

○ 캄보디아 보건 관련 원조의 문제

- 자원의 배치 및 적절한 활용은 원조의 효과성을 증가시키지만 현재의 보건원조는 구조적으로 분절화되어 있고 유연적이지 않아 MDG 달성을 저해하는 장애물로 작용한다고 언급함 (WHO and Ministry of Health Kingdom of Cambodia, 2007).
- 특히 특정 질환에 대한 지정된 자금(earmarked finance)이 많아지는 것은 실제 국가가 필요로 하고 그동안 소외된 영역에

대한 접근을 하는 데 제한적으로 작용한다고 강조함(WHO and Ministry of Health Kingdom of Cambodia, 2007). 구체적으로 다음의 그림과 같이 캄보디아의 보건 우선순위는 보편적 의료보장인 데 반해 원조 자금은 주로 성매개질환 관리에 집중되어 있음을 알 수 있음.

[그림 3-6] 캄보디아와 기여국의 보건분야 사업 우선순위 차이



자료: WHO and Ministry of Health Kingdom of Cambodia(2007). Scaling up for better health in Cambodia. p.21 Figure 6.

- 또한 원조 기구 및 국가들 간 조화로운 관리 및 실행 등이 이루어지지 않고 있다고 언급하면서, 캄보디아는 다른 지역에 비해 원조기구 간 조화도가 상대적으로 낮은 수치로 나타난 것을 지적함(WHO and Ministry of Health Kingdom of Cambodia, 2007).
- 캄보디아와 같이 원조 자금이 정부 보건 관련 예산보다 많은 상황에서 원조기구와 국가와 수원국 간 조화가 이루어지지 않으면 중장기 재정 및 자원 동원 계획, 인적 자원 양성 등 보건 영역과 직접적으로 관련 있는 정보들이 분절적으로 존재할 수밖에 없고, 이 때문에 정보를 적시에 수집하는 데 시간과 비용이 많이 소모된다고 강조함(WHO and Ministry of Health Kingdom of Cambodia, 2007).

〈표 3-3〉 캄보디아 원조 부합성(alignment) 지수(2005년)

지표	캄보디아	조사된 31개 국가의 평균
국가 재정관리체계의 신뢰도[Country Policy and Institutional Assessment(CPIA) 평가]	낮음	-
원조지출 대비 예산에 보고된 원조 비율(%)	79	90
국가 프로그램을 통해 조정된 기술 지원의 비중(%)	36	43
국가 체계를 통한 원조(%)		
예산집행	17	43
재정보고	9	34
감사	3	34
조달	6	37
지출 예정된 원조 자금 대비 실제 지출된 것으로 보고된 원조 자금의 비율(%)	69	65

자료: OECD(2006). Progress Report on the 2006 Survey on Monitoring the Paris Declaration의 내용을 재구성한 WHO and Ministry of Health Kingdom of Cambodia(2007). Scaling Up for Better Health in Cambodia. p.20 Table 8.

○ 캄보디아 보건 관련 원조의 개선방안

- WHO and Ministry of Health Kingdom of Cambodia (2007)는 캄보디아 국가 사례연구를 통해 캄보디아 원조에 있어 캄보디아 국가 정책과의 조화와 부합성(alignment)을 강화해야 한다고 강조하면서, 이를 위한 원조기구들 간 원조의 조화를 강화하기 위한 노력이 이루어지고 있다며 ‘Sector-Wide Management’를 소개하기도 함.
- 점진적으로 원조 자금은 실제 보건 정책이 수행되는 영역에 제공될 수 있도록 해야 한다고 제안함. 이는 결국 각 국가와의 협의를 통해 국가의 보건 우선순위에 따라 원조가 이루어져야 하며, 이는 결국 각국의 보건 증진에 기여할 수 있다고 보고 있음.

3. WHO의 원조 평가의 함의

□ 그동안 이루어진 WHO의 ODA 모니터링 및 평가는 개별 사업에 대한 평가라기보다 각 국가를 대상으로 한 원조사업의 규모를 모니터링하거나 MDG 달성 정도와 같이 영향을 평가하는 차원에서 이루어지고 있음.

- WHO에서 하는 평가는 프로그램 평가에서부터 주제 평가, 특정 사무소에 대한 평가까지 다양하나, 최근까지 WHO는 보건 MDG를 중심으로 각 국가의 보건 관련 MDG 이행 정도와 달성 정도를 파악하는 것에 초점을 두었음. 이는 보건 사업의 성과와 영향을 평가하는 것을 의미하며, 일반적으로 주제 평가나 특정 사무소에 대한 평가로 이루어졌음.

- WHO는 모니터링과 평가를 위해 고려해야 하는 원칙들과 기준들을 제시하고 있음. 구체적인 평가 도구를 제시하지는 않지만, 모니터링과 평가를 하는 데 내부 및 외부 평가자가 고려해야 할 요소들과 질문들을 제시함.
- 최근 WHO는 SDG 이행 및 달성 정도를 파악하기 위하여 Global Health Observatory를 설치하여 194개 회원국의 보건 관련 1000여 개의 지표들에 대한 데이터를 구성하고 공개하고 있음.
- 국제적인 보건 관련 지표들의 현황과 변화 추이를 살펴볼 수 있으며, ODA 사업의 효율성, 효과성, 영향을 파악할 수 있는 근거 자료를 생산하고 이를 공유하고 있음.
- ODA와 관련하여 수원국 리스트와 금액, 원조국 리스트와 금액들을 연도별로(2000~2010) 제시하고 있으며, 목적별 ODA 자금 규모(예: 보건 정책, MDG6)도 제시되어 있음.¹⁷⁾
- Global Health Observatory는 다양한 보건 관련 지표들을 보유하고 있어 보건 관련 ODA 사업의 성과와 영향을 평가하기 위한 지표 선정에 활용할 수 있음. 또한 WHO에서 지속적으로 구축하는 보건 통계는 실제 ODA 사업 평가에 활용할 수 있을 것으로 기대함.

17) 관련 구체적인 내용은 WHO 홈페이지 참조
(<http://apps.who.int/gho/data/node.wrapper.HEALTHSYSTEMS00>).

제2절 영국

1. 영국 공적개발원조(ODA)와 평가개요

□ 영국은 국제개발협력에서 모범적인 국가 중 하나이며 평가에서도 이러한 점이 드러나는 사례 국가임.

○ 영국은 ODA 규모 및 ODA/GNI가 매우 높은 국가이며 ODA의 운영에서 중앙 집중적인 형태를 취하고 있는 국가임.

○ 한편 영국 공적개발원조의 특징은 자국의 이익만을 우선으로 하지 않고 세계의 빈곤 감소를 공적개발원조의 주요 원칙으로 수용하고 있다는 점임(김현경 등, 2015, p.100).

- OECD DAC에서 영국의 공적원조를 모범 사례로 인정한 바 있음.
- 영국의 공적개발원조는 1990년대 중반에 그 목적이 크게 변화.
 - 1997년 노동당 정부는 공적개발원조의 목적으로 세계 빈곤 퇴치를 명시하고 백서를 발간(신상협, 2011, pp.179-180; 김현경 등, 2015, p.101에서 재인용).
 - 2002년 국제개발기본법(International Development Act)을 제정. 동법에는 영국의 공적개발원조는 모두 비구속성 원조(untied)로 추진한다는 것, 국제개발부에 빈곤 완화라는 의무 수행의 핵심 목표를 부과한다는 것이 핵심 내용으로 포함됨.
 - 영국은 2001년 이후 조건부(tied) 원조를 모두 폐지하여 공적개발원조의 100%가 비구속성 원조임. 이는 북유럽 일부

국가와 유사한 특성임. 다자간 협력의 중요성을 강조하고 이를 증대하는 중임(김현경 등, 2015, p.101)

- 공적 협력 사업의 효과를 높이고 투명성을 확보하기 위하여 정확하고 객관적인 평가와 이 결과를 반영하는 구조를 강조.
- 영국은 국별 성과관리와 국별 평가의 결과를 예산편성과 연동하는 근거기반(evidence-based) 접근을 DAC 공여국 중에서 가장 선진적으로 추진한다는 평가를 받고 있는 국가임(정지선, 오탉현, 2013, p.70).
- 영국은 영국 공적협력사업의 목적을 명시한 관련 법인 국제개발법(2002)을 제정한 후에, 국제협력사업에 대한 보고와 투명성을 강조한 법인 국제개발법[International Development (Reporting and Transparency) Act 2006]을 제정한 바 있음. 18)

□ 영국의 대외원조 정책과 사업수행을 담당하고 있는 국제개발부(DFID: Department for International Development)는 최근 평가를 개선하기 위한 다양한 노력을 기울이고 있음.

○ 영국의 사례는 향후 우리나라가 지향해야 할 원조사업 평가체계와 방법에 유용한 시사점을 제공할 것으로 판단됨.

○ 영국의 DFID는 평가뿐만 아니라 국제개발과 관련된 다양한 영역에서 국제적 표준(global standard)을 선도하고 있음.

18) International Development (Reporting 2 and Transparency) Act 2006 (c. 31)을 참조.

2. 평가의 원칙과 체계

- DFID는 2013년 평가정책의 방향과 원칙을 정리한 “국제개발협력 평가정책(International Development Evaluation Policy)”을 발표했는데, 이는 원조사업에서 평가를 얼마나 중요하게 인식하고 있는지를 명확히 보여줌(DFID, 2013a).
- 개발평가의 중요성에 대한 인식의 제고, 평가의 설계·수행·활용과 관련하여 명확하고 일관된 기준의 제시, 평가의 엄밀성과 윤리성 제고 등이 동 정책 수립의 배경임.

가. Value for Money(VfM) 원칙 적용

- 원조사업의 성과관리 및 평가의 중요성에 대한 DFID의 인식을 가장 명확히 보여주는 것은 원조사업에 대해서도 'Value for Money (VfM)'의 원칙을 도입한 것임.
- VfM은 영국 공공분야 전반에서 지출되는 공적자금이 효율적으로 사용되는지를 확인하기 위한 평가의 원칙으로 1980년대부터 도입이 시작되었음.
- 영국 감사원(National Audit Office)에 따르면 VfM은 “목표한 성과를 달성하기 위해 자원을 최적의 방법으로 사용하는 것”을 의미함.¹⁹⁾

19) 자세한 설명은 다음을 참조할 것. <https://www.nao.org.uk/successful-commissioning/general-principles/value-for-money/assessing-value-for-money/> 2016. 9. 20. 인출.

○ VfM을 평가하는 데는 흔히 3E라고 불리는 세 가지 기준이 사용됨.

- 경제성(economy): 투입된 자원의 비용을 최소화하였는가?
(spending less)
- 효율성(efficiency): 생산에 사용된 자원(투입)과 생산된 산출물(재화 혹은 용역)의 관계(spending well)
- 효과성(effectiveness): 공공지출(산출물)이 목표로 한 결과와 실제 결과 간의 관계(spending wisely)

□ DFID는 원조사업에도 VfM을 적용하기 위한 노력을 꾸준히 해 왔으며, 2011년에는 이를 정리한 “DFID의 VfM 접근법”을 발표하였음.

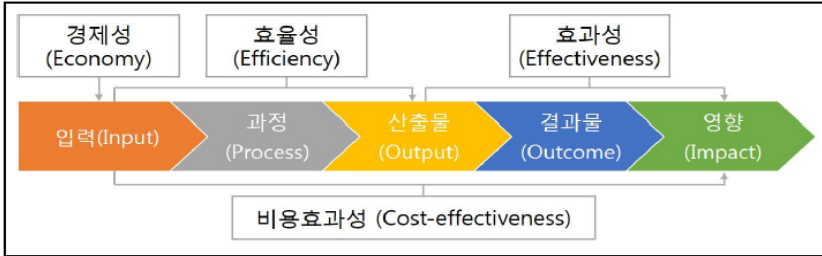
○ DFID의 사업에서 VfM은 “빈민의 삶의 질 향상을 위한 최적의 자원 이용 혹은 재정투입의 영향력 극대화”를 의미하며, 이것이 사업의 모니터링 및 평가에 적용되는 기본 원칙임(DFID, 2011, p.3).

○ VfM을 위해서는 원조의 영향력 극대화를 위해 목표하는 결과(산출물과 성과)와 비용이 명확해야 하며, 증거기반을 강화하고 성과 달성을 위한 기본 가정을 분명히 해야 함.

○ 원조사업의 평가에서도 [그림 3-7]과 같이 각 단계에서 VfM의 3E원칙이 적용됨.

- 경제성: 양질의 투입물을 적절한 가격에 구매하였는가?
- 효율성: 투입이 산출물로 적절히 변환되었는가?
- 효과성: 사업의 산출물이 빈곤이 감소하는 결과물로 이어졌는가?
- 비용효과성(cost-effectiveness): 위 3E의 결과 투입이 빈곤 감소에 얼마나 영향을 미쳤는가?

[그림 3-7] 원조사업의 단계와 VfM의 3E 평가기준



자료: DFID(2011). p.4.

○ DFID는 VfM 원칙을 실현하기 위해 사업의 기획·집행·평가 등의 단계에서 유의해야 할 점을 제시하고 있는데, 이런 점들은 결국 영향평가의 중요성을 강조한 것이라고 할 수 있음.

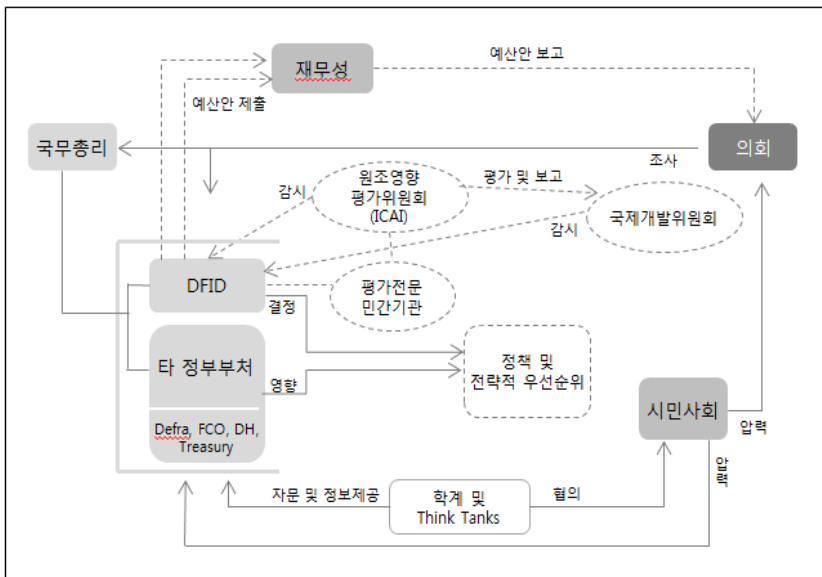
- VfM은 무조건 조달을 최저가로 하라는 것이 아니라 원가의 구성을 이해하고 저렴한 비용으로 적절한 질을 확보하라는 의미임.
- 개입, 즉 사업이 효과를 발휘할 수 있는 조건 혹은 가정을 근거에 기반하여 명확히 설정하여야 함.
- 성과지표는 측정하기 용이한 지표가 아니라 목표한 결과를 가장 잘 드러내는 지표가 되도록 해야 함.
- 나타난 성과가 DFID의 사업에서 비롯되었다는 점, 즉 인과의 귀속(attribution)을 분명히 해야 함.
- 그리고 지원 범위와 범분야 이슈 고려(coverage), 파트너십(coordination), 일관성(coherence), 귀속(attribution) 등이 추가적인 평가기준으로 고려됨(정지선, 오택현, 2013, p.83).
- DFID뿐만 아니라 시민사회, 다자기구 등 DFID의 자금의

사용되는 모든 주체를 포함한 원조에 VfM 원칙이 강화되어야 함.

나. 평가조직

- 영국의 공적개발원조에 대한 평가는 영국 공적개발원조의 주관 기관(DFID)의 평가 팀과 독립평가기관이 수행.

[그림 3-8] 영국의 공적개발협력 및 그 평가 관련 조직의 개요



자료: Donor Tracker(2012). p.4; 정지선, 오태현(2013). p.72 재인용된 내용을 수정하여 재구성.

- 정부 기관으로 공적개발원조를 수행하는 부처의 평가 팀(Evaluation Department)
- DFID는 사업별 보고서뿐 아니라 주제별 보고서, 그리고 연간 보

고서 등을 작성하고, 연간보고서를 보면 자산 형성, 빈곤 완화, 교육, 보건 등 주제별 성과를 정리(DFID, 2016).

○ 실질적인 평가는 ITAD²⁰⁾ 등과 같은 외부의 평가 및 컨설팅을 전문으로 하는 기관에 용역을 주는 독립평가의 형태로 수행되는 경향이 강함.

○ 한편 DFID의 평가보고에서는 민간 평가전문기관과 현지의 평가팀의 역할이 동시에 강조되고 있음.

- 예를 들어 베트남에 대한 평가 중 하나인 ‘Evaluation of the Long-term Development Cooperation Between the UK and Vietnam’은 조사평가업체인 Landell Mills Limited (LML)와 베트남 현지 조사기관인 Mekong Development Research Institute (MDRI)이 공동 수행.

- 평가진은 평가 전문 컨설턴트와 주제별·분야별 컨설턴트로 구성되며, 영국 컨설턴트와 현지 컨설턴트의 협력하에 진행(정지선, 오테현, 2013, p.81).

○ DFID의 평가 팀은 현지 팀과 협력하면서 평가과정을 지원.

- 예를 들어 한 개도국 전문가 교육 프로그램에 대한 평가를 위해 BCURE evaluation team과 BCURE implementing teams을 운영. 후자는 현지 팀으로 평가 팀이 해당국을 방문할 때 그리고 1단계 평가과정에서 참여하여 지원하는 역할을

20) 1984년 설립된 민간 조직으로 국제 개발에 대한 모니터링과 평가로 그 효과를 높이는 것을 목적으로 하며, 영국 DFID 외 여러 국가의 국제개발협력 주체들, 국제기구들에게 컨설팅을 제공함. 전 세계 현지 사무국을 운영 중인데, 2016년 현재 아프리카와 아시아 지역을 포함하여 9개 현지 사무국을 운영. <http://itad.com/about-us/>에서 2016. 9. 2. 인출.

수행하는데, DFID 평가 팀은 평가결과를 실행에 반영하는 과정(Evidence into Action)에서 이 팀에 행정적 지원과 유용한 자문을 제공(Vogel and Punton, 2016).

〈참고〉 에티오피아의 2003~08년 지원에 대한 국별 평가의 과정

- 평가과정을 보면 1주간의 사전점검출장(inception visit) 이후 3주간의 본격적인 현지 조사를 실시. 현지조사 기간 동안 개도국 정부관계자, DFID 과거·현재 담당자 등 주요 이해관계자에 대한 인터뷰를 실시하고, DFID 내 관련 문서 등 문헌검토를 병행.
- 사전점검 출장에서는 주요 이해관계자를 파악하고 평가의 범위를 확정하는 한편 표준 평가매트릭스를 해당 국가에 맞게 조정하고 현지 출장과 관련한 중요 사항을 결정.
- 본격적인 현지조사 출장은 사전점검하고 1개월 이후에 실시되며 대개 3주의 기간이 소요. DFID의 평가부서에서는 DFID의 인트라넷 및 인터넷에서 추출한 해당 수원국에 대한 전략문서, 프로젝트·프로그램 관련 정보 문서 및 파일을 평가단에 제공.
- 평가연구진은 다양한 형태로 구성되지만 해당 국가에 대한 DFID 지원내용 및 전략적 방향, 주요 쟁점 등을 고려하여 분야별, 주제별 전문가가 참여. 에티오피아 국별 평가는 평가단이 총 6인으로 구성. 개발 전문 컨설팅업체 ITAD의 평가 전문 컨설턴트가 PM 역할을 수행하였고, 그밖에 인도적 지원, 공공재정 관리 및 예산지원, 경제개발 및 취약성, 거버넌스 및 시민사회, 교육 및 물 등 DFID의 에티오피아 중점 지원분야·주제별 전문가가 공동평가진으로 참여(정지선, 오택현, 2013, pp.85-86).

□ 영국의 원조평가에서 또 하나 주목할 만한 점은 평가의 독립성 강화를 위해 ‘원조영향평가위원회(ICAI: Independent Commission for Aid Impact)’라는 독립기관을 2011년 설립한 것임.²¹⁾

- 원조사업 평가의 독립성과 객관성 제고를 위해 설립하였으며, DFID와 독립된 기관이며 평가의 결과는 의회에 직접 보고하도록 되어 있음.

21) <http://icai.independent.gov.uk/>에서 2016. 9. 2. 인출.

- 이 조직은 독립성을 유지하기 위하여 다양한 이해관계자 사이의 갈등에 대응하는 정책을 두고 있음.

○ ICAI는 원조예산이 수혜자에 미치는 영향과 납세자를 위한 VfM을 평가하는 데 중점을 두고 있음.

- DFID의 평가부서(EvD: evaluation department)는 과정평가, 성과평가, 영향평가를 수행하는 반면, ICAI는 메타평가 및 시행·수행기관에 대한 감사 기능에 초점을 맞추는 등 평가의 책무성에 중점을 둠.
- 따라서 ICAI는 주제별 조사와 정기적 조사, DFID의 원조관계 변화(ICAI, 2016a) 등과 같은 종합적이고 포괄적인 평가를 수행하기도 함.
- 원조영향평가위원회의 성공지표(ICAI, 2016b, p.8)를 보면 영국의 원조가 제 가치를 창출하도록 하는 것이 가장 근본적인 목적에 해당됨을 알 수 있음.

〈표 3-4〉 원조영향평가위원회(ICAI)의 성공지표

목표	핵심 성과 지표	측정 기준
파급효과(impact) 긍정적인 변화를 통해 투자가가치 창출	ICAI의 제안이 수용된 비율	매년 자체 추후조사
효과적인 책임성 (effective accountability) 영국 원조의 효과성에 대한 신뢰할 만한 독립적인 증거를 생산함으로써 정부가 책임성을 확보할 수 있도록 의회를 지원	ICAI 전반적 활동 및 1) 영국 원조의 도전과제와의 연관성, 2) 생산한 증거기반의 질, 3) 수행과 평가의 독립성에 대한 위원회의 만족도	위원회 위원과 내부 지원으로부터의 피드백
	매년 6~8편의 평가보고서 발간 (보고서는 방법론적 접근법에 대한 설명을 포함해야 함)	출판된 보고서 수

목표	핵심 성과 지표	측정 기준
효과적인 학습(effective learning) 정부 및 기타 개발 이해집단과 함께 학습하고 토론하는 장을 마련하는 데 효과적으로 기여	파급효과 극대화를 위한 이익집단과의 관계 확대 및 증진	6개월마다 열리는 행사의 수, 2개월마다 측정된 ICAI 트위터 계정 팔로워 수
	평가가 조회된 수	각 평가별 웹사이트 조회 수
	평가가 방송 및 SNS 매체에 언급된 수	평가 공개 후 두 달간 언론에서 언급된 수, 2개월마다 측정된 트윗에 언급된 수
효율성(efficiency) 올바른 거버넌스를 통해 효율적으로 사업 수행	개발사업의 역동적인 맥락을 반영하도록 생산물의 범위 확장	차별화된 스타일의 평가 생산
	전체 예산 중 평가 관리	현 위원회의 임기 4년간 ICAI의 전체 예산 준수

자료: <http://icai.independent.gov.uk/>에서 2016. 9. 2. 인출.

○ ICAI는 책임성(accountability), 환류와 학습(feedback and learning), 분석적 엄격성(analytical rigour), 접근성(accessibility), 투명성(transparency)을 핵심 가치로 하고 있음.²²⁾

- 책임성: IDC와 DFID를 통한 영국 원조지출의 효과를 높이기 위하여 객관성, 중립성 중시.
- 환류와 학습: 사업에 대한 검토 과정으로 이어지는 효과적인 환류시스템 구축.
- 분석적 엄격성: 근거의 사용과 분석에서 엄격함을 유지.
- 접근성: 명백한 목적을 가지고, 독자, 청중을 고려하면서 평가.
- 투명성: 자료와 근거의 사용과 제시 등 조사에서 투명성을 견지.

○ 국제개발위원회의 해당 부서는 원조영향평가위원회의 제안에 대

22) <http://icai.independent.gov.uk/>에서 2016.9.2. 인출.

한 응답 및 반영계획 등을 반드시 제시해야 함.

- 보고서의 결론에서는 각 평가기준별 등급 제시와 함께 개선을 위한 권고안을 명시하고, 각 권고안별로 DFID 등 담당 부처 및 기관은 개선방안을 제시해야 함.
- ICAI는 각각의 개선방안을 검토하여 의견을 제시하는 한편, 이행상황을 모니터링함.

○ 국제개발위원회(IDC)가 원조영향평가위원회의 계획에 자문과 승인을 담당하고 있음.

다. 평가의 내용

□ 평가내용의 구성에서 영국의 국제 원조의 목표가 주요하게 고려됨.

○ 영국 국제원조의 전략적 목표는 크게 네 가지임(DFID, 2016, p.9).

- 세계평화, 안전, 거버넌스의 증진
- 회복 능력과 위기 대응 능력 함양
- 세계 풍요 증진
- 극단적 빈곤에 대한 대응과 세계의 가장 취약한 집단에 대한 원조

□ 국제개발협력에 대한 규정을 담은 관련 법에는 국제개발협력과 관련하여 평가하고 보고할 내용을 담고 있음.

- 국제개발법[International Development (Reporting and Transparency) Act 2006]은 의회에 대한 보고 의무 및 통계의 보고, 그리고 국제개발협력에 대한 지출 수준, 목표, 투명성 확보에 대한 내용을 담고 있음.

- 투명성 확보와 관련하여 미래의 원조 할당, 목표의 유지, 원조 수원국과 수혜국의 동의, 부패 방지 등 이상적 관리 등에 대한 내용이 포괄되어 있음.
- 동 법에 따르면 연간보고서에 포함되어야 하는 내용(section 2)은 영국의 양자 간, 다자간 원조액, 위의 원조에 포함된 부채 경감액, 공적개발원조액 등임.
- 동 법에서는 해당 내용이 지역별, 국가별, 그리고 영역별로 파악될 수 있어야 한다고 적시하면서 필요한 기타 집계 방식도 추가 제시하고 있음.
- 한편 매년 평가보고서를 발간. 예를 들어 2015년의 DFID 보고서에서는 2014/5년 DFID는 930만 명에게 현금 지원을 하였고, 2,850만 명의 영유아와 임산부, 수유모에게 영양 지원을 하였고 350만 명에게 식품을 지원한 것으로 보고하였음(DFID, 2015; 김현경 등, 2015, p.103에서 재인용).
 - 대부분 극빈국은 원조 대상으로 하여 식품 영양, 건강 등으로 사업 내용이 집중되는 경향을 보임.

라. 평가결과 공유와 환류

- 영국의 공적개발원조에 대한 평가의 가장 큰 특징 중 하나는 평가결과를 공유하고 이를 원조계획과 활동에 반영하는 환류시스템을 가졌다는 점임.
- 평가결과를 공유하는 다양한 기제를 운영.
 - 공식적으로 정부 내에서 주체들의 활동에 대한 평가결과를 의회

에 보고하고 이를 공개.

- 실제 ICAI 홈페이지에서 수원국별, 시기별, 사업 영역별 평가결과를 확인할 수 있음.²³⁾
- 영국 정부의 통계자료 공개 홈페이지에서 국제개발협력의 각종 통계자료를 엑셀 상태로 다운받을 수 있음.²⁴⁾

□ 평가결과의 반영 기제도 다양

- 원조 주관부처의 평가보고서의 마지막에는 경영진 의견이라는 별도의 장을 두어 독립평가자가 제시한 평가결과 및 제언에 대한 DFID와 같은 주관부처 측의 입장과 개선 노력을 항목별로 구체적으로 명시(정지선, 오태현, 2013, p.86).
 - 일례로 2016년 6월 DFID는 DFID Management Response를 ICAI로 제출. 물과 위생(WASH: Water, Sanitation and Hygiene) 사업에 대한 ICAI의 영향 조사에 대응한 것으로 2016년 5월에 제출된 보고서의 의견에 대한 대응.
- 평가결과는 각 사업의 예산 결정에 반영
 - 예산안에 대한 검토 및 국가별, 지역별, 부서별 예산편성 단계에서 각각의 현지 사무소는 향후 추진 전략과 함께 중점 분야별 목표 달성치에 대한 성과(results offer)를 예산 요구의 근거로 함께 제시(정지선, 오태현, 2013, p.78).

23) ICAI 홈페이지 참조. 그리고 다음 사이트에 국가별 영역별 평가결과물을 정리하여 볼 수 있도록 공개. <https://www.gov.uk/government/collections/evaluation-reports>

24) 관련 사이트 주소는 다음을 참조.

<https://www.gov.uk/government/statistics/statistics-on-international-development-2015>

○ 경쟁적인 성과기반 자원배분(competitive and performance based allocation)방식에 따라 국별 개발사업 추진에 대한 비용 대비효과(VfM)를 제고.

- 국별 성과관리 차원에서 현지 사무소에 성과관리 자문관 (results advisor)을 두고 매년 3월과 9월 연 2회에 걸쳐 국가 협력전략에 포함되어 있는 중점 분야별 목표 달성에 대한 이행 성과를 모니터링하여 본부의 성과관리위원회에 보고. 성과관리위원회에서는 달성 여부를 측정하고 목표치 조정의 필요성에 대해 논의하며 매해 연례보고서에서 국별 성과를 공개. 국가별로 목표 달성 성과가 우수한 국가, 미흡한 국가가 정량적으로 파악되어 다음 해 우선순위별 예산을 받기 위한 근거 자료로 활용(정지선, 오테현, 2013, p.79).

□ 한편 DFID는 평가가 사업 수행단위에서 사업 사이클의 일부임을 인식하고 또 그렇게 운영하도록 하는 소위 ‘평가의 내재화(embedding evaluation)’ 노력을 2010년부터 꾸준히 추진하고 있음.²⁵⁾

○ 새로운 프로그램을 기획할 때는 기존의 평가결과나 증거를 일상적으로 활용해야 하며, 평가를 프로그램 사이클의 일부로 받아들여야 함.

○ 사업성과의 증거가 불확실한 사업이나 혁신적인 사업의 경우 사업 팀은 신뢰할 수 있는 평가계획을 수립해야 함.

○ 더욱더 많은 사업에 대해 영향평가를 포함한 양질의 평가를 외부에 위탁하도록 함.

25) 자세한 내용은 DFID(2013b) 참고.

- 결론적으로 예산지출 효율성 강화 압력하에서 객관적이고 엄밀한 원조평가 요구가 증대되었는데, 영국은 이에 대응하여 평가의 원칙 제시, 중요성 인식 제고, DFID 내 평가체계 개선, 독립기관 설립, 구체적 평가 매뉴얼 작성 등 다양한 노력을 기울이고 있음.²⁶⁾

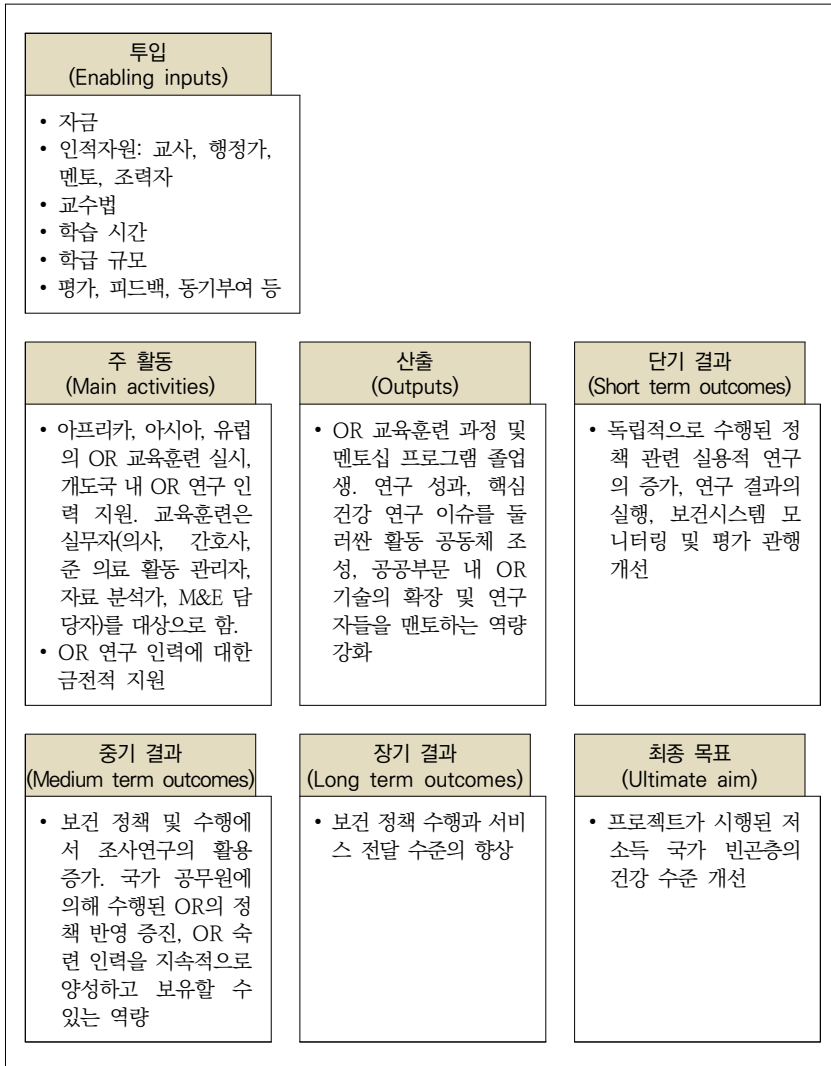
3. 평가사례

- 사례 1: 2015년 DFID의 ‘개도국의 실용적 조사(Operational Research)역량 함양 프로그램’에 대한 평가(Melin et al., 2015).
 - 개도국 보건 전문가들이 객관적 근거를 갖춘 조사를 수행하고 이를 정책, 그리고 전략과 행동에 반영하는 능력(Operational Research Capacity)을 함양하는 일종의 교육 프로그램에 대한 평가.
 - 해당 프로그램은 여러 국가에서 진행된 프로그램으로 이 프로그램에 대한 평가는 영역에 대한 평가로 이해할 수 있음.
 - 인도 등에서 진행된 해당 코스 11개 사업에 대한 종합 평가.
 - 양적·질적 연구방법을 활용하여 효과성, 효율성, 적절성, 그리고 지속 가능성(relevance and sustainability of the project)을 중심으로 평가.
 - [그림 3-9]는 이러한 평가지표를 전체 평가의 구조에서 이해할 수 있는 논리구조도임.
 - 다음과 같은 분석방법을 병행 활용(Melin et al., 2015, p.13).
 - 면담을 통한 질적 평가(Qualitative assessment through

26) DFID의 평가체계 개선에 관한 더욱 자세한 설명은 KOICA(2013), pp.48-70을 참고.

interviews)

[그림 3-9] 보건 분야 실용적 연구(Operational Research, OR) 역량 강화 프로그램의 논리 구조도



자료: Melin et al.(2015). p.10.

- 참여자의 순수효과에 대한 자가평가(Self-assessment of net effect by participants via the electronic questionnaire)
- 통제집단과의 비교를 통한 이중차이비교(Difference-in-difference analysis through comparison with a control group)

□ **사례 2:** 파키스탄의 선거문화 개선 사업 평가(Balagamwala, Gazdar, 2014)

- 파키스탄의 선거문화를 변화시키기 위한 두 개의 프로젝트(SERP: Supporting Electoral Reforms in Pakistan와 STAEP: Supporting Transparency, Accountability and Electoral Processes)에 대한 영향평가
 - 외부 민간 평가조직이 해당 평가를 수행.
- 주요 평가 지표는 2013년 선거의 ‘투표율’과 ‘민주적인 절차의 강화’를 중심으로 구성.
- 비교할 통제적인 상황에 대한 정보를 입수하기 어려운 것이 이 평가의 약점. 그래서 분석에서 기여분석방법(Contribution Analysis)을 활용하고, 양적·질적 방법을 동원하여 이 개입이 어떻게 성과에 기여하였는지를 면밀하게 규명.
- 아래의 표는 주요 평가 지표 중 하나인 투표율에 대한 분석으로 지표의 변화와 각 요인의 기여를 분석하고 있으며 분석은 정책수준, 환경수준, 지역, 그리고 관련된 맥락에 대한 것임.
- 성과에 대하여 정책수준과 지역수준 등 다양한 수준에서 그 변화를 정리. 이후 프로그램에서 고려할 점들을 정리하여 보고서에 담

고 있음.

○ 해당 보고서에는 본론의 뒷부분에 해당 국가의 위험수준에 대한 종합평가를 담고 있는데 예를 들어 이동, 치안, 홍수 등 여러 영역에서 해당 국가가 어느 수준으로 위험한지를 간략하게 평가하고 정리하였음. 이는 참여하는 요원 및 이 프로젝트와 관련하여 영향을 받을 수 있는 제3자의 안전 등을 위한 조치임.

○ DFID는 이 보고서에 대하여 의견을 제출함. 이는 규정에 따른 것으로 평가보고서에 대응하여 원조사업을 담당한 주체들이 향후 반영계획을 답으로 전달하여야 함.²⁷⁾

〈표 3-5〉 파키스탄의 선거 문화 개선 사업 평가: 투표율 분석 결과 요약

구분		양적 분석	정책 수준	선거구 수준	지역사회 기반	맥락 분석
지표	투표율 증가	강하게 지지 - 지역 간, 지역 내 편차	강하게 지지			
	등록 개선	증거 혼재 - 선거구 간 편차	강하게 지지	강하게 지지	증거 혼재	
연결고리	선거 관리 역량 증진이 투표 참여 증가로 이어짐	의견 조사 결과 강하게 지지	강하게 지지 - 공급측 요소가 더 효과적; 관찰이 검표에 대한 인식 강화로 이어짐	약하게 지지 - 지역 요인이 선거 관리 및 투표율에 영향을 미침	약한 증거 - 폭력이 감소한다면, 지역 요인이 더욱 중요해 질 것	약한 증거 - 민주적 전환, 정치적 협아가 더욱 중요
	선택권 확대가 투표율 증가로 이어짐	일부 증거 존재	강하게 지지		제한적 지지 - 정당 및 후보자가 투표자 동원	안정적인 선거를 위해서는 정당 사이의 협력도 중요

27) 해당 문건의 제목은 'DFID Management Response to Impact Evaluation of DFID's Electoral Programmes in Pakistan'임. 동 자료는 해당 보고서가 있는 사이트에 연결되어 공개되고 있음.

구분	양적 분석	정책 수준	선거구 수준	지역사회 기반	맥락 분석
사업 성과	ECP의 사전 선거 이미지에 대한 기여	의견 혼재 - SERP가 ECP 이미지 향상으로 이어졌으나 ECP 역량을 강화 하지는 못함	증거 혼재	증거 혼재 - 지역 요인이 선거 관리 경험을 형성함	숙고할 만한 의견 - 정치 엘리트들 위한 신호 장치
	검표 준비환경에 대한 기여		강하게 지지		숙고할 만한 의견 - 투표자를 동원하는 정치 엘리트들 위한 신호 장치
	투표자 직접 동원에 대한 기여	약하게 지지 - CRG 영향에 명확한 차이 없음.	의견 혼재 - 지배적인 개입 요인 없음.	약한 증거 - 프로그램의 제한적 가시성. 지배적인 개입 요인 없음	

자료: Balagamwala & Gazdar(2014). p.27.

□ 사례 3: ICAI의 연간보고서

○ 해당 보고서는 2015/6년 활동에 대한 보고서임(ICAI, 2016b).

- ICAI는 중기 평가계획을 수립하여 운영되는데 2016/7년 계획을 보면 조사를 세 개로 구분하여 운영.
- 단기조사(두개의 조사를 예정), 탄력적인 대응 조사(이 주제는 해당 시기 후반 결정, 하나의 조사를 예정), 그리고 해당 시기를 전반적으로 조사하는 연간조사가 그 세 가지 조사임.
- 동 보고서는 ICAI의 평가계획 개요, 평가의 틀 그리고 이전 평가의 검토, 해당 기간에 대한 평가로 구성됨.
- ICAI는 2015년 이후 물과 위생 그리고 여성에 대한 폭력을 새로운 주제로 설정(ICAI, 2016b, p.1).

○ ICAI의 평가(reviews)는 세 가지 유형으로 구분.

- 영향(impact)에 대한 평가, 수행(performance)에 대한 평가 그리고 학습(learning)에 대한 평가임.

〈표 3-6〉 원조영향평가위원회(ICAI) 세 가지 유형의 평가

영향(Impact)	수행(Performance)	학습(Learning)
• 정부가 달성했다고 주장하는 것들의 기저를 이루는 요소를 평가함. • 타당성과 중요성을 모두 평가하되, 전자는 근거에 기반하여, 후자는 수혜자 관점에서 살펴봄. • 영국 정부가 그 결과를 얼마나 잘 측정했는지, 정부 프로그램과 포트폴리오가 향후 더욱 영향력을 갖도록 결과자료를 이용하는 지 등을 살펴봄.	• 원조의 효과성과 투자 가치를 면밀하게 살펴봄. • 특히 책임성에 초점을 맞춤 • 정부 원조가 얼마나 잘 관리되고 전달되었는지, 투자 가치를 성취하고 있는지, 원조 프로그램을 관리하는 데 적절한 시스템과 역량을 보유하고 있는지 등을 살펴봄.	• 원조 프로그램에서 새롭게 부상하고 있는 영역 • 영국 정부가 무엇이 효과적인지에 대한 근거기반을 얼마나 잘 생산해 내고 믿을 만한 프로그램을 개발하는 데 그 지식을 어떻게 사용하는지에 관해 살펴봄 • 독립적인 정밀 조사가 진행되는 동안 지속적인 학습에 기여하도록 ICAI는 정부와 보다 활발하게 상호작용함.

자료: ICAI(2016b). p.2.

○ 매년 조사 유형에 따른 조사 대상을 선정하고 조사를 수행

- 첫해에는 DFID의 주요 사업 영역 중 하나인 ‘여성에 대한 폭력 대응(programming on violence against women and girls)’과 ‘국제조세에 관한 활동(work on international tax)’에 대한 학습이 잘 이루어졌는지 평가. 영향에 대한 조사는 물과 위생(water and sanitation) 그리고 현금 지원(cash transfers)에 대하여 수행. 그 이유는 후자 사업의 경우 사업의 성숙으로 결과에 대한 질 좋은 자료를 갖추었기 때문임.
- 성과에 대한 조사는 DFID의 사업 중 갈등 지역의 수탁자 리스크 관리 사업(management of fiduciary risk in conflict-

affected areas)과 교육에서 주변화된 소녀들에 대한 지원 (support to marginalised girls in education)에 대하여 수행.

- 이 보고서는 해당 팀의 첫해 보고서에 대한 정리와 보고 외에 지난 4년간의 보고에 대한 간략한 정리와 처리 내용을 담고 있음.
- 그리고 해당 기간의 영국 원조지출의 변화를 정리.

4. 합의

□ 평가의 강조

- 근거를 기반으로 하는 정책기획의 추세와 맞물려 더욱 중요해짐.
- 최근 ODA의 운영을 참고할 때 국가별 종합평가, 사업 영역별 종합 평가 등 다양한 측면의 평가를 종합하고 향후 계획에 반영할 필요성이 커지고 있음.

□ 평가의 종합적 기획과 다변화, 그리고 다시 종합

- 평가의 대상과 단위가 변화하는 추세를 반영하여, 과거에는 프로젝트평가를 통한 단기적 성과에 대한 평가가 주류를 이루었다면, 최근에는 주제·부문별, 정책·전략 평가, 프로그램 평가, 국별 평가를 통한 중장기적 개발성과와 전략적 방향성 파악에 중점을 두는 추세.
- 개별 사업에 대한 평가보다는 개발성과와 조직 효율성과 같이 전략적인 측면으로 평가의 초점이 이동하는 경향이 강함(정지선, 오태현, 2013, pp.41-42).

- 평가에 대한 평가(evaluation of evaluation)인 메타평가의 경우 주로 기존에 실시한 평가에 대한 종합 검토에서 시작되었으나 최근에는 주제, 영역, 원조 수단별로 대상을 세분화하는 추세(정지선, 오탉현, 2013, p.43).
- 한편 평가는 다양한 국면에서 시행하되 그 결과의 종합이 더욱 크게 활용되는 현상도 공존.
- 여러 평가를 종합하여 재평가하면서 국가별, 그리고 영역별 사업의 개선과 기획에 빠르게 반영할 수 있도록 하는 보다 현실적인 평가를 선호.
- 다양한 축으로 분석결과와 정리, 그리고 결과의 집합이 가능하여야 정책설계에 유용.

□ 영향(impact)평가 등 평가에서 결과에 주목.

- 사업의 영향을 명확하게 파악하기 위한 분석방법을 적극 개발하고 활용.
- 분석방법에서 양적 방법 외 평가의 정확성을 위하여 필요한 질적 분석 방법, 기여분석 등 다양한 접근을 개방적으로 수용.
- 평가의 설계에서 영향을 정확하게 평가할 수 있도록 적절한 분석 방법을 사전에 설계하되 그 방법의 선택에서 보다 적극적이고 개방적일 필요가 있다는 점을 보여줌.

□ 평가의 독립성을 확보하기 위한 기제 마련

- 국제개발협력에 다양한 이해관계자가 참여하는 과정이므로 평가의 독립성 확보는 중요한 과제.

- 영국은 평가조직의 독립성을 확보하기 위한 조직구조, 그리고 조직운영의 다양한 정책 고려가 돋보임.
 - 예를 들어 독립적인 평가조직의 예산 지원 부처와 기획 및 운영에 대한 감독 부서가 상이함.
- 평가의 독립성을 보장하고 평가자체에 대한 평가가 가능하도록 하는 기제의 설계와 검토가 필요함.

□ 평가의 반영기제 강화

- 국제원조사업에 대한 평가는 반드시 그 반영을 가능하게 하는 기제를 갖추고 있음. 예를 들어 해당 사업주체의 응답과 모니터링이 이러한 경향을 반영.
- 한편 평가결과의 확산과 공유를 통하여 이러한 지식이 관련 사업 및 관계자에게 폭넓게 영향을 주도록 하여 평가의 효과를 극대화하는 상황임.
- 예산에도 반영하는 방식으로 실제 지출의 효과성을 높이는 방안으로 활용.

□ 현장성 고려

- 평가에서 국제원조사업이 진행되는 현장 팀의 역할을 강화하고 현장과 사업에 맞추어 평가의 지표와 방식을 원용할 수 있도록 하여 현장성을 강화.
- 이러한 일련의 개선 과정은 실제 국제원조사업의 효과성을 높이는 데 기여할 것으로 추정.

□ 평가주체의 역량 중시

- 영국은 개발원조를 수행하는 조직의 인력이 전문성을 갖추도록 지속적으로 지원.
 - 영국 정부는 2015/6년 DFID 인력의 상당한 비율인 약 18% 수준을 교육 대상에 포함하여 전문성 제고 교육을 실시.
 - 프로그램 운영, 위험관리 등에 대한 1~3일의 교육과정을 제공.
 - 이후 설문 결과 이로써 참여 인력의 응답은 역량강화를 경험하였다고 긍정적으로 응답(DFID, 2016, p.6.)
- 또한 현장 전문가의 참여가 강조되면서 현지 전문가의 역량을 높이고 영국 국제개발원조의 목적과 방식에 대한 공유 기회를 확대.

제3절 일본

- 이 절에서는 일본국제협력단(JICA: Japan International Cooperation Agency)의 평가체계와 사례를 소개함(JICA, 2016a; JICA, 2016b).
- 1995년 해외경제협력기금(OECF: Overseas Economic Cooperation Fund)과 일본수출입은행(Export-Import Bank of Japan)이 일본국제협력은행(JBIC: Japan Bank for International Cooperation)으로 통합되었고, 2008년에는 JICA와 JBIC를 통합하여 New JICA가 탄생함.
- 현재 New JICA는 기술협력, 양허성 차관, 무상원조를 포함하여 일본의 모든 ODA 사업을 실질적으로 담당하고 있음.

1. 평가개요

□ 평가 방향

○ PDCA(Plan-Do-Check-Action) Cycle

- JICA의 평가활동은 프로젝트의 진행 단계에 따라 다음의 네 가지 절차로 구성됨: 사업계획 단계에서 실시하는 사전조사(Plan), 사업진행 단계에서 실시하는 모니터링(Do), 사업종료 후 실시하는 사후조사(Check), 평가결과를 실제 사업 및 향후 사업계획에 반영하는 피드백(Action).

○ 평가 방법 및 기준의 일관성

- 기술협력(Technical Cooperation), 양허성 차관(ODA Loans), 무상원조(Grant Aid) 각각의 특성을 고려하되 일관성 있는 철학과 프레임워크에 기초하여 평가함.
- 평가 프레임워크는 다음과 같은 특징을 지님: 1) 프로젝트 단계에서의 평가는 PDCA 사이클을 따름, 2) 평가기준은 OECD DAC가 제시한 5가지 기준을 따름, 3) 평가결과는 JICA 자체 등급 분류 체계(Rating system)에 맞춰 외부에 공개함(표 3-7 참조).

○ 주제별 평가를 통한 영역 간 종합적인 평가

- 각 주제별 사업에 공통적으로 적용할 수 있는 교훈을 제공하고 평가양식을 개발함.

○ 객관성과 투명성

- 외부 전문가로 하여금 사후평가를 진행하도록 하고 평가자문 위원회에 제3자를 포함하여 평가결과의 객관성을 증진하고,

평가결과를 JICA 웹사이트에 공개하여 투명성을 보장함.

○ 결과의 활용 및 평가의 질 증진

- PDCA 사이클 중 Action 단계의 피드백을 활용하여 Plan과 Do 단계의 질 증진을 도모함.

〈표 3-7〉 JICA의 평가 프레임워크: 평가 기준별 평가 결과 등급 분류 체계

평가 기준		등급 분류 기준		
		높음(high)	중간(fair)	낮음(low)
적절성	• 원조 타당성 (수혜국의 개발 정책과의 관련성, 일본의 ODA 정책 및 JICA의 개발 전략과의 관련성) • 개발 수요와의 관련성	관련성 높음.	일부 관련됨.	심각한 문제 있음.
효과성/영향력	• 목표 기간 내 예상했던 결과 성취도	80% 이상 달성	50~80% 달성	50% 미만 달성
	• 긍정적인 간접 효과 및 부정적 효과의 상태	- 간접 효과 발생 - 부정적 효과 없음	- 간접 효과에 다소 문제 있음 - 부정적 효과 일부 발생	- 간접 효과에 문제 있음 - 중대한 부정적 효과 발생
효율성	• 계획안과 실제 결과 비교(투입, 기간, 비용 등)	효과적 (100% 혹은 조금 못 미침)	다소 비효율 (100~150%)	비효율 (150% 이상)
지속 가능성	• 재정적 지속가능성(사업 수행 및 관리를 위한 예산 확보 가능성)	확실	다소 문제 있으나 개선 가능성 있음	불충분

자료: JICA(2016b). p.10.

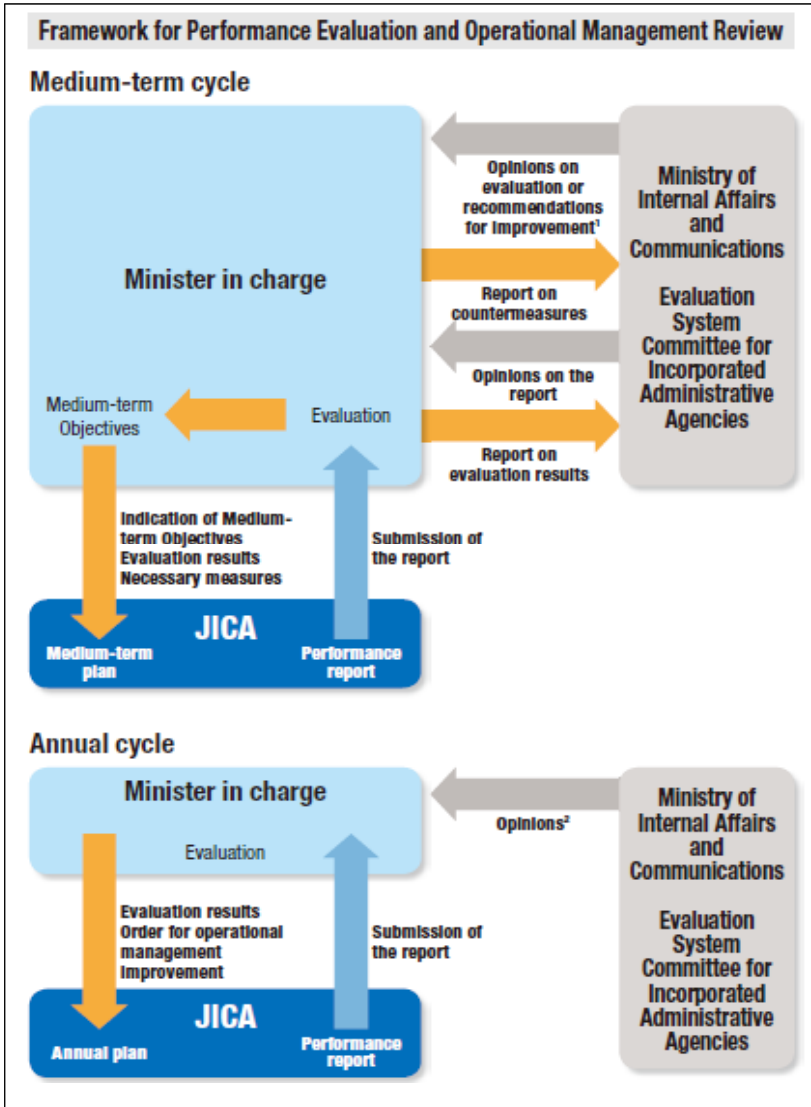
□ 평가체계

○ 단기목표기반 관리 및 성과평가시스템

- JICA는 5년 단위 중기 계획을 수립하여 일본 총무성(Ministry of Internal Affairs and Communications) 장관의 승인을

받은 후, 그에 맞춰 1년 계획을 수립하여 정부에 제출함.

[그림 3-10] JICA의 평가 체계



자료: JICA(2016a). p.150.

- JICA는 매년 정부에 성과 보고서를 제출하고, 이에 근거하여 정부는 사업 계속 여부를 판단함.
- 동시에 총무성 산하에 평가위원회(the Evaluation System Committee for Incorporated Administrative Agencies)를 외부기관으로 설치하여, 중기 목표와 평가결과에 대한 별도의 의견 제출하도록 함.

○ 중기 및 단기 효과를 성취하기 위한 노력

- 국별, 지역별, 이슈별, 분야별 분석에 바탕을 둔 전략적 추진 계획 마련.
- JICA 내부적으로 외부 전문가가 참여하는 내부 자문위원회를 구성하여 중기 및 단기 계획 달성을 위해 사업 진행 상황을 꾸준히 모니터링함.
- 정부의 평가결과, 총무성 산하 평가위원회의 의견, 내부 모니터링 결과를 사업의 질 제고와 관리 최적화를 위해 피드백함.
- 국제사회와 사업 경험 및 평가결과를 공유함(예: 2015년 제3차 재난 위험 감소를 위한 UN 세계총회에서 샌다이의 경험 공유).

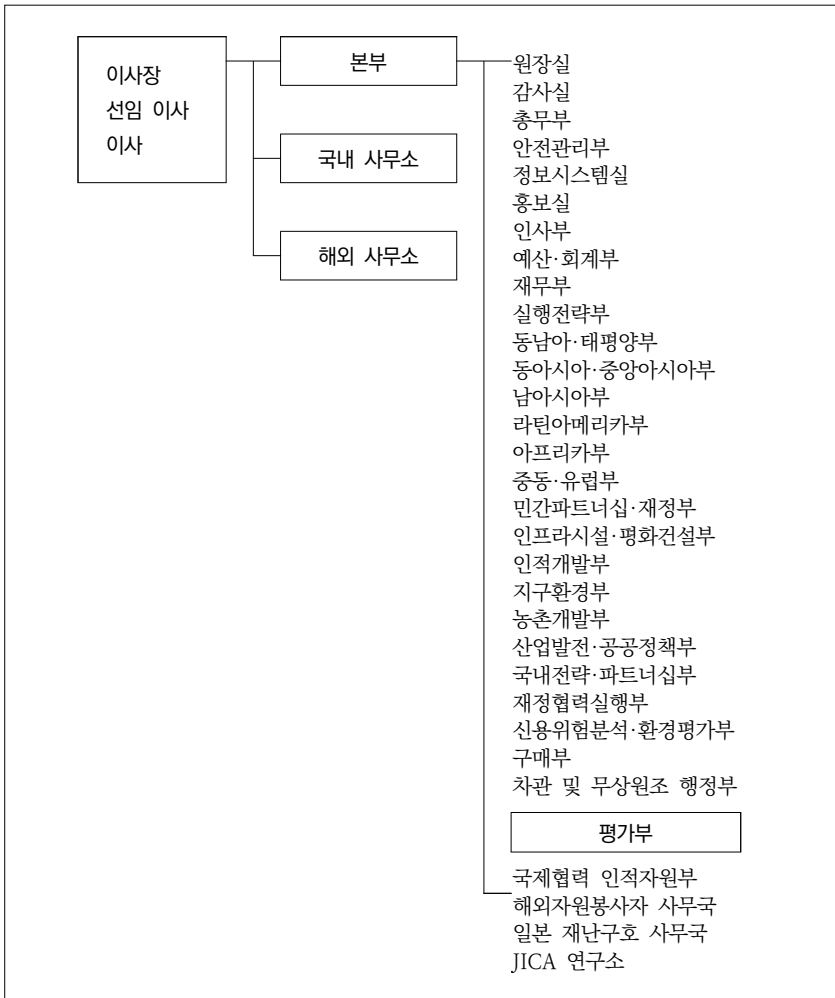
□ 평가보고서

- 사전평가, 중간점검, 사후평가, 주제별 평가, 영향평가로 나누어 2000년 이후 작성된 영문 보고서 요약본을 JICA 홈페이지에 모두 공개함.
- 또한 2000년 이후 매년 연간평가보고서(Annual Evaluation Report)를 발간하여, 주요 평가결과를 소개하고 종합적인 시사점을 도출함.

□ 평가조직

○ JICA의 평가부서는 headquarter 내 독립 부서의 하나로 포함되어 있음(그림 3-11 참조).

[그림 3-11] JICA 조직도(2016. 10. 1.)



자료: <http://www.jica.go.jp/english/about/organization/index.html>에서 2016. 10. 11. 인출.

2. 평가사례

□ 사례1: 몽골의 사회 분야 지원 프로그램 자금원조(JICA, 2014).

○ 배경

- 2008년 세계 경제 위기로 광물 가격이 급락하자 2009년 1분기 외환 보유고는 2007년 대비 절반 수준으로 떨어지고 재정 적자는 GDP 대비 15% 수준까지 치솟음. 이에 몽골 정부는 국제통화기금(IMF: International Monetary Fund)에 원조를 요청하게 됨.
- IMF의 대기성차관협정(Stand-by Arrangement)에 근거한 단기 유동성 자금을 지원받기 위해 몽골 정부는 금융시스템 안정성 회복, 사회보장시스템의 재정 부담 완화, 중기 재정의 안정성 확보를 약속함. 이에 부응하여 세계은행(WB: World Bank), 아시아개발은행(ADB: Asian Development Bank), 일본이 몽골에 대해 재정 지원 의사를 밝힘.

○ 프로그램 개요

- 몽골 사회 분야 지원 프로그램의 목표는 단기적으로 국가 재정 파산을 막고 중장기적으로 공공 재정 관리 역량을 강화하는 것이었음.
 - 구체적으로 경제위기 시에도 빈곤층의 보호에 차질이 발생하지 않도록 수급자 선정 과정을 개선하고, 기초 사회 서비스를 보장 및 확대하며, 사회 분야 필수 지출 규모를 유지하고자 함.
- 자금 지원은 JICA, ADB, WB, IMF 공동으로 두 차례에 걸쳐

이뤄짐.

- 1단계: 2009년 6월부터 8월까지 총 28억 9천만 엔을 연 이율 0.8%, 30년 상환 조건으로 지원함.
 - 2012년 3월부터 7월까지 총 15억 5천만 엔을 연 이율 0.8%, 20년 상환 조건으로 지원함.
- 프로그램 중 사회보장 분야의 정책 개선 영역은 복지, 보건, 교육, 도시개발 등 총 네 개를 포함했는데, 각 정책 개선 영역에 대한 시기별 정책 수단은 정책 매트릭스로 요약됨.

〈표 3-8〉 몽골의 사회 분야 지원 프로그램 중 사회보장 영역 정책 매트릭스

분야	정책 개혁 영역	1단계 정책 수단	2단계 정책 수단
복지	효율성 증진 및 빈곤층 보호를 위한 사회 복지 개혁: 1) 대리자 산조사를 통한 선별 증진, 2) 아동수당 등 사회복지프로그램 통합, 3) 재정 안정성 확보	복지고용부(Ministry of Social Welfare and Labor)가 개혁안 준비 완료: 1) 제도적 변화, 2) 전환 전략을 포함한 실행계획, 3) 예상되는 어려움에 대한 대비책, 4) 이해당사자의 책임성, 5) 재정 등에 관한 내용 포함	• 내각이 의회에 새로운 사회복지프로그램안을 제출함. • 의회가 2010년 예산안을 승인함.
	국가건강보험제도 아래서 의약품할인제도 확대를 통해 병원 입원률 감소	• 기존 할인제도에 대한 비판적 평가 • 제도 개선방안 마련을 위한 전담반 신설 • 할인제도 확대 지침 마련	• 평가결과를 근거로 의약품할인제도 개혁안 승인 • 새로운 제도의 핵심 영역에 대해 시범사업 실시
	대리자산조사를 통한 선별적 국가건강보험료 보조금 지원제도 실시	• 대상자 선별을 위한 총리실 산하 조직 신설	• 선별적 건강보험 지원금 제도 실행계획안 마련 • 건강보험법 개정안 초안 마련
보건	갑작스러운 공급부족에 대비해 필수약품 확보를 위한 응급의료체계 구축	보건부(Ministry of Health)가 전담반 설치	보건부가 재원 마련, 지침안 실행 가능성 확보

분야	정책 개혁 영역	1단계 정책 수단	2단계 정책 수단
	가구의 무기영양소 보호범위 확대를 통한 아동 영양결핍 예방	- 보건부가 전담반 설치 - 보건부가 선별 방식을 포함한 확대안 마련	보건부가 가구 무기영양소 보호범위 확대안 승인
	빈곤층을 위한 의료카드 프로그램을 통한 의료비 현금지출 감소	보건부와 복지노동부 합동 대책반 설치	- 보건부가 대리자산조사에 근거한 의료카드 도입안 승인(대리자산조사가 완료될 때까지는 과도기 전략에 기반) - 보건부가 실행계획안 마련
교육	중등교육과정 교과서 보조금 지원 대상자 선별작업 개선	교육문화과학부(Ministry of Education, Culture and Science)가 대리자산 조사의 실행 가능성 평가를 위한 전담반 설치	교육문화보건부가 전담반의 작업 결과를 고려하여 선별된 지원대상자에 대한 보조금 승인
	아동 발달을 돕는 새로운 교수법 확대 보급을 통한 교사훈련시스템 서비스 지원 증진(게르 지역 포함)	교육문화보건부가 실행계획 초안 준비	교육문화보건부가 실행계획안 완성
	새로운 교육 기준에 맞춰 교사 평가 시스템 개편	교육문화보건부가 새로운 교수법에 대한 평가시스템 결과를 평가하는 전담반 설치	교육문화보건부가 새로운 평가 시스템 승인
도시	기초 서비스 보장 및 생활환경 개선을 위해 울란바토르 마스터플랜 개정 및 우선순위 지정(게르 지역 포함)	도로교통건설부(Ministry of Roads, Transportation, Construction and Urban Development)와 울란바토르 시청이 환경 및 주거 전문가로 구성된 전담반 설치	의회가 마스터플랜 개정안 자금 조달을 위한 2010년 회계연도 예산 승인
	게르 지역 생활 조건 개선을 위한 역량 강화 및 지침 개발	울란바토르 시청이 전담반 구성	게르 지역 개발 및 환경 정책, 사회안전망 정책 개발을 위해 울란바토르 시청이 역량 강화 교재 및 지침을 개발

자료: JICA(2014). pp.5-6.

○ 사후평가 개요

- 외부 평가자에 의해 2014년 7월부터 2015년 11월까지 진행.
- 사후평가를 진행하기 위해서는 산출(output), 결과(outcome), 영향력(impact)을 고려한 별도의 평가 프레임을 작성할 필요가 있었음.
- 정책 매트릭스는 단기 정책 수단을 한눈에 파악할 수 있도록 해 주지만, 언제까지, 무엇을, 어떻게 수행할지에 관한 프로그램의 구체적인 내용이나 궁극적 목표에 대해서는 제시하지 않을 뿐만 아니라 정책 개혁의 완성에 필요한 정책 수단의 중장기 효과를 고려하지 못함.

〈표 3-9〉 사회보장 영역 사후평가를 위한 평가 프레임워크

프레임워크	목표	지표
산출	국가의 재정적 파산을 피함	
결과	정책 수단의 성취	* 새로운 교수법과 관련된 교사용 지침서 수: 효과성(effectiveness) 지표로 활용될 수 있으나 데이터 수집이 제한적
영향력1 (중기)	정책 개혁 영역의 성취, 평가 시점에 맞춰 실행 및 결과 지표(Operation and Effect Indicator) 선정	• 대리자산조사(proxy means test) 방식을 사용한 가구조사를 수행한 행정 구역 수 • 의약품 할인 제도에 대한 정부 보조금 • 국가 건강 보험에 대한 정부 보조금 • 영양결핍 아동 비율 • 빈곤인구 중 의료카드 보급률 • 중등 교육 과정에서 무료 교과서를 배급 받은 학생 비율 • 교사 훈련 프로그램 참석률
영향력2 (중장기)	경제위기 동안 빈곤층 보호, 기초 생활보호 대상자 선정 과정의 적절성 향상, 빈곤층의 생활 환경 개선	• 빈곤율 • 빈곤인구의 월 소비액 • 게르(ger) 지역 개발 참여 지역 기관 수
영향력3 (장기)	공공 지출 관리 향상, 사회 발전 증진	

자료: JICA(2014). p.10의 Table 2와 3을 바탕으로 필자가 재구성.

- 몽골 사회 분야 지원 프로그램은 일본이 일반적으로 수행하는 ODA 프로젝트와 달리 정책 수단 실시를 조건으로 조건 성취 시 자금을 지원함. 따라서 자금 지원 전 정책 조건 성취 여부를 프로그램 효과로, 자금 지원 후 개혁 진척을 프로그램의 영향력으로 간주하고 평가를 진행함. 특히 후자에 관하여, 평가 시점에서 개혁의 효과나 진행 정도를 명확하게 파악할 수 없었기 때문에 현재 진행 중인 정책 개혁만을 프로그램의 영향력으로 간주함.
- 효율성(efficiency)은 다음과 같은 이유에서 평가하지 않음.
 - 투입은 몽골 정부의 일반예산, 산출은 사회보장 분야의 개혁에 해당하므로 투입과 산출이 호응하지 않고, 산출을 계량화하기 힘들.
 - 예산 지원 프로그램의 경우, 프로그램 기간과 비용에 관해 계획과 실재는 항상 일치하기 때문에 평가의 의미가 없음.

○ 사후평가 결과

- 적절성(relevance): 높음(high)
 - 프로그램 목표가 몽골의 자금 수요 및 개발 수요뿐만 아니라 몽골의 개발 계획과 일본의 ODA 정책에도 부합함.
 - IMF 프로그램하에서 다른 국제 공여기관과 함께 지원한 덕분에 규모의 경제와 효율성을 달성함.
 - 몽골 정부는 자금 지원 조건으로 요구된 사항에 대해 정책 매트릭스를 통해 미리 대응하였으며 정책 매트릭스의 내용 또한 적절하였음.

- 효과성(effectiveness): 높음(high)
 - 정책 매트릭스로 제시한 정책 수단이 모두 시행되었으며 JICA에 의한 확인 절차와 평가도 적절하게 진행됨.
- 영향력(impact): 보통(fair)
 - 일부 개혁 분야에서 예상했던 것만큼의 진척이 이뤄지지 않았음.
 - 또한 재정 상황 악화에 대한 우려 속에 계획했던 효과가 나타날 것인지 여부에 대해 지켜볼 필요가 있음: 개혁을 위한 시스템은 갖추었으나 심각한 재정 상황 악화로 인해 전 분야에서 개혁이 지연되었음.
 - 또한 몽골 정부의 정책 수행 역량에 대해서도 우려가 제기됨.

○ 시사점

- 자금원조 ODA의 경우 목표 달성에 시간이 걸리기 때문에 프로그램의 영향력을 단기에 파악하기 어려움.
- 따라서 프로그램의 영향력, 즉 효과의 지속 가능성을 파악하기 위해서는 수원국 정부와 공여기관 사이의 지속적인 교류가 필요함.

3. 합의

□ 오랜 국제개발협력 역사를 통해 일본은 상당히 체계적인 평가체계를 갖추게 된 것으로 보임. JICA의 평가체계의 장점은 다음과 같이 요약됨.

□ 체계성

- 영국, 독일 등 선진공여국과 마찬가지로 일본은 프로젝트의 계획부터 평가단계에 이르기까지 목표, 활동, 투입, 외부 조건 간 논리적 상관관계를 중시하는 Project Design Matrix(PDM)을 채택하고 있음.
- 구체적으로 평가는 PDCA 사이클에 따라 프로젝트의 진행 단계에 맞춰 체계적으로 실시함.
- 환류 역시 단기(1년 단위)와 중장기(5년 단위)로 나누어 체계적으로 진행됨.

□ 구체성

- 몽골에 대한 자금원조사업 사후평가 사례에서 볼 수 있듯이, 프로젝트의 성격에 따라 일반적인 평가 프레임워크를 적용할 수 없는 경우에는 별도의 평가 프레임워크를 구성하여 적용함.

□ 투명성

- 2000년 이후 작성된 JICA의 모든 평가결과가 웹사이트에 공개됨.
- 모든 평가보고서가 영문으로도 요약되어 있어 누구나 쉽게 열람할 수 있음.

제 4 장

한국의 공적개발원조(ODA) 사업평가

제1절 한국 ODA 평가체계와 현황

제2절 보건의료분야 평가사례

4

한국의 공적개발원조(ODA) << 사업평가

제1절 한국 ODA 평가체계와 현황

〈표 4-1〉 평가 관련 국가개발협력위원회 활동 현황 및 결과

구 분		주요문서	핵심내용	주요 의결 사항
평가 시스템 구축	국무 회의 (‘05. 11)	- 국제개발협력 개선방안	- 평가 및 사후관리 시스템 정비	- 독립적 외부평가 방식 도입 - 기존 시행기관의 평가 담당 조직 및 평가 기능을 강화 - 평가 규정을 제정
	국 개 위 2차	- ODA 중기전략 (2008~2010) - 2007년도 ODA 추진계획	- 평가시스템 구축 - 국개위 중심으로 평가 및 모니터링 강화	- ‘통합평가 가이드라인(지침)’ 제정 추진 - 국개위 중심으로 종합평가 실시 - 원조기관 내 평가예산 확보 및 전문인력 보강 - 선진 공여국 및 국제개발원조기구 등과 공동 평가를 실시하여 평가 경험 및 지식을 축적
	3차	- 2008년도 ODA 추진계획	- 평가시스템 구축	- ‘통합평가지침’ 및 ‘통합평가시행 매뉴얼’ 마련 - 위 평가지침에 따른 정책/사업 평가 실시 - 국개위 중심으로 민간참여확대, 공동평가 실시 - 평가 결과는 공개
	5차	- 2009년도 ODA 추진계획	- 통합평가시스템 구축	- ‘통합평가지침’ 및 ‘통합평가매뉴얼’(09. 12. 제정) - 통합평가지침에 따른 시범평가 실시 - 외부기관평가 실시(EDCF 베트남 국별 종 합평가 실시)
평가 시스템 선진화	7차	- 국제개발협력 선진화 방안	- 통합평가시스템 구축 - 환류 시스템 제도화 - 평가결과 공개	- 총리실 평가소위 중심의 유무상 통합평가 방식으로 개편하여 객관성 확보: 평가소위 에서 민관합동으로 유무상 연계영역을 직 접 평가하고, 기관별 자체평가결과를 심의 - 평가결과와 환류시스템을 제도화하며 평가 결과는 공개 - NGO, 컨설턴트 등 민간 전문인력을 평가 에 활용

124 ODA 평가체계 연구 - 보건복지분야 사업평가의 시사점

구 분	주요문서	핵심내용	주요 의결 사항
8차	- 분야별 국제개발 협력 기본계획 (2011~2015)	- 자체평가 역량 강화 - 환류시스템 제도화	- '통합평가 추진방안'에 따라 자체평가 역량을 강화 - 독립된 평가조직을 설치하고 외부전문가, 협력대상국 관계자 등의 참여를 통해 평가 결과의 객관성과 신뢰성을 제고 - 프로젝트 평가뿐만 아니라 정책평가, 국별평가, 주제별 평가 등으로 평가대상을 다변화 - 평가결과는 대국민 공개, 다음 연도 사업 계획 수립에 반영
9차	- 2011년 ODA 시행계획	- ODA 평가 강화	- 자체평가역량강화: 수원국과의 공동평가 방식으로 녹색성장 관련 사업에 대한 프로젝트 평가, 인도네시아 국별 평가 실시
11차	- 2012년 ODA 시행계획	- 통합평가추진 - 범분야(cross-cutting issue) 평가 도입 - 평가 내실화 및 효율화	- 소위평가와 자체평가 결과를 공유 및 환류 - 범분야 이슈에 대한 범부처 가이드라인 마련을 위하여 범분야 평가 예시 사례 도출 추진: 여성, 환경 분야 - 국별지원전략(CPS) 체제와 연계하여 전략 성과평가: 성과지표 설정 및 사실자료에 기반한 평가체제안 마련, 현지 ODA 협의체의 평가기능 강화 - 메타평가 추진: 자체평가 일부를 소위에서 심층평가하여 평가의 질을 측정
14차	- 2013년 ODA 시행계획	- 메타평가 실시 기반 구축 추진 - '통합평가지침', '통합평가매뉴얼' 개정(13년 하반기 중)	- 일정 규모 이상의 사업에 대해 자체평가 실행을 의무화 - 평가 전문성 강화를 위하여 '평가전문가 pool'을 구축, 평가 워크숍 개최 - '자체평가 가이드라인'을 작성하여 평가품질 표준화 - 중요과제는 평가소위에서 통합평가 실시 - 사업분야별 메타평가 도입 등 메타평가를 본격 실시 - 성과 평가를 예산, 인사 등 인센티브와 연계하는 방안 검토
17차	- 2014년 시행계획	- 평가협업 강화 - 자체평가 강화	- EDCF, KOICA 간 평가실무협의회, 공동평가 확대 - 무상협력 시행기관 간 평가기법 및 노하우 공유: ODA 평가 워크숍 개최, KOICA 평가실무교육 강화 - 각 기관의 자체평가계획 수립 및 제출을 의무화 - 자체평가 환류계획에 대한 이행상황 점검 - '12~13년 실시된 자체평가에 대한 메타평가를 실시하고 개선방안 마련

구 분		주요문서	핵심내용	주요 의결 사항
평가 시스템 강화	20차	- 2015년 시행계획	- 자체평가 강화	- 평가대상을 다변화: 프로젝트 평가 위주에서 주제별, 분야별, 국별 평가로 확장 - KOICA ODA 교육원에 정규평가교육과정을 개설하여 시행기관의 자체평가 역량을 제고
	22차	- 제2차 국제개발협력 기본계획 (2016~2020)	ODA 평가 및 환류 시스템 강화 - 외부평가 강화 - 독립평가부서 구축 - 예산 및 인력 증가	- 각 시행기관의 평가 예산을 점진적으로 확대 - ODA 사업전담기관의 평가 담당 직위의 전문직위제 실시 - 대규모 사업에 대한 외부평가 실시로 재정 책임성 및 투명성 확보 - 국무조정실 내 평가 전담 부서를 설치하여 각 부처 평가업무를 관리 감독 및 지원 - 사업 심사 시 당해 사업의 기존 실적 평가 결과와 연계하여 환류 강화
	23차	- 2016년 시행계획	- 자체평가 내실화	- 시행기관의 대규모 사업에 대한 외부평가 실시를 의무화 - 자체평가 결과 및 개선 사항에 대한 이행 상황 지속 점검 - 평가교육을 개편하여 각 시행부처 평가 담당자의 역량 강화를 지원 - '14~'15년간 실시된 자체평가 결과에 대한 메타평가를 실시
	26차	- 2017년 시행계획	- 자체평가 강화 - 통합평가 내실화 - 평가체계 개편	- 대규모 사업에 대한 외부평가 실시를 의무화 - 메타평가 결과를 바탕으로 개선방안 마련 - 통합평가시스템 출범('10) 이후의 성과 및 한계를 분석하고 개선안 마련

자료: ODA KOREA 국제개발협력위원회 활동 현황 및 결과 http://www.odakorea.go.kr/hz.bltn2.ODACmtRsltSIPL.do?brd_seq=24&bltn_div=oda#script에서 2016. 10. 1. 인출.

1. 평가지침

□ (통합평가지침·매뉴얼) 평가소위원회는 통합평가시스템을 구축하기 위해서 「통합평가지침」과 「통합평가매뉴얼」을 제정(2009년).

○ 동 지침과 매뉴얼은 2010년 7월에 발표된 「국제개발협력기본법」과 부합하도록 다듬어져 2010년 10월 개정.

- 현재 사용하고 있는 최종본은 2014년에 2차로 개정한 것인데, ODA 시행기관의 자체평가계획 수립과 자체평가 결과 보고 의무를 명문화한 「국제개발협력기본법」 개정(2013년 7월) 사항과 평가소위원회의 2012년 메타평가 결과를 반영하기 위한 개정.
 - 2014년에 개정된 주요 내용은 다음과 같음: ① 용역제안서의 목차 표준 및 점검 목록을 신설, ② 평가보고서의 목차 표준 및 점검 목록을 신설, ③ 평가결과반영 이행상황의 반기별 점검 및 그 결과를 평가소위에 보고하도록 의무화(국제개발협력위원회, 2013a, p.73).
- **(메타평가 가이드라인)** 제16차 평가소위원회는 2014년도 메타평가를 실시하여 그 결과로 ‘국제개발협력 메타평가 가이드라인(안)’을 마련
- 동 가이드라인은 메타평가의 정의와 목적, 시행체제, 모형 및 방법론, 결과 활용을 제시함(국제개발협력위원회, 2014a, pp.25-36).
- **(시행기관 가이드라인)** 주 시행기관인 EDCF와 KOICA는 자체적으로 평가 가이드라인을 마련
- 대외경제협력기금(EDCF)은 「사후평가보고서 작성 가이드라인(2011)」, 「평가매뉴얼(2011)」, 「공동평가지침서(2012)」를 사용.
 - 한국국제협력단(KOICA)은 「사업관리자를 위한 개발협력사업평가 가이드라인(2008)」, 「DAC 평가 시리즈-공동평가 업무관리 가이드라인(2014)」, 「평가업무수행 길라잡이(2014)」를 사용.

2. 평가조직

- **(통합평가 체계)** 통합평가 체계에 대한 논의는 2008년 OECD DAC의 특별검토 수검 이후로 지속. DAC은 특별검토에서 한국 ODA 평가의 객관성과 전문성 확보를 위하여 한국정부의 '유·무상 통합평가 체계 구축'과 '독립적 평가문화 확립'을 권고, 이에 따라 정부는 통합평가체제 구축을 결정하고 당해 8월에 평가소위원회를 구성(국제개발협력위원회, 2009, pp.5-6).
- 2010년 7월 발효된 「국제개발협력기본법」 및 「국제개발협력시행령」에서 ODA 평가와 평가소위원회에 대해 규정함으로써 통합평가체제의 법적 근거를 마련.
- 이후 정부는 「국제개발협력 통합평가 추진방안」, 「ODA 선진화 방안」에서 총리실 평가소위 중심의 유·무상 통합평가 방식으로 개편할 것을 명시함으로써(국제개발협력위원회, 2010b, p.22, p.88; 국제개발협력위원회, 2010a, p.34.) 평가소위평가와 자체평가가 병행되는 방식의 통합평가체제가 자리 잡게 되었음.
- 각 시행기관이 소관 사업에 대해 자체평가를 실시한 후 그 결과를 평가소위가 심의하고, 동시에 평가소위원회는 정책 및 전략·국별·분야별·주제별·형태별·프로젝트/프로그램별로 대표성 있는 사업을 선정하여 심사하고 소위평가를 실시.
- **(시행기관의 평가조직)** 제1차 분야별 기본계획(2011년~2015년)과 제2차 분야별 기본계획(2016년~2020년)은 평가 예산과 평가 인력의 확대, 평가역량강화를 주요 개선과제로 제시하고 있으나, 개선 성과는 뚜렷하지 않은 실정

○ 기획재정부는 대외경제협력기금(EDCF) 내에 평가팀을, 외교통상부는 한국국제협력단(KOICA) 평가실, 보건복지부는 한국국제보건의료재단(KOFIH) 평가협력부로 평가담당 부서를 설치하여 평가지침과 방법 등 체계적인 평가시스템을 구축하여 운영.

- 그 외의 다수의 사업수행기관도 수행사업에 대한 결과보고서나 전문가 용역 등의 평가를 실시하고 있지만, 평가 담당부서나 평가지침 등의 평가시스템 마련은 미흡한 실정(국제개발협력위원회, 2011, p.43).

○ 정부의 평가제도 점검결과, ODA 예산은 지속적으로 확대되어 2010년 1조 3천억 원에서 2014년 2조 3천억 원으로 크게 증가하였지만, 해당 기간 중 원조 시행기관의 평가예산과 인력의 변화는 크지 않은 형편임(표1-1 참고)(국제개발협력위원회, 2015a, p.25).

〈표 4-2〉 KOICA, EDCF 평가 건수·예산·인력 추이

(단위: 건, 백만 원)

구분	KOICA				EDCF			
	2011	2012	2013	2014	2011	2012	2013	2014
평가 건수	18	18	24	31	8	9	10	10
평가 예산*	631	694	926	908	278	280	461	463
평가 건당 비용	35	38	38	29	34	31	46	46
담당 인력	7	7	7	7	4	5	6	7

주: *ODA 평가용역비 기준.

자료: 국제개발협력위원회, 2015a. 제18차 평가소위원회 보고안건(제18-2호), 국제개발협력 통합 평가제도 개선방안. p.25.

〈표 4-3〉 평가 조직에 대한 논의

시 기	문 서	주요 내용	시행 여부
2005	국제개발협력 개선방안	기존 시행기관의 평가담당 조직 및 평가기능을 강화	미흡
2007	국제개발협력 중기전략 (2008~2010)	원조기관 내 평가 예산을 확보하고 전문인력을 보강	미흡
2010	국제개발협력 선진화 방안	총리실 평가소위 중심의 유·무상 통합평가 방식으로 개편하여 객관성을 확보	시행
2010	분야별 국제개발협력 기본계획 (2011~2015)	독립된 평가조직을 설치	시행
2013	2014년 시행계획	EDCF, KOICA 간 평가실무협의회, 공동평가 확대	시행
2014	2015년 시행계획	KOICA ODA 교육원에 정규평가교육과정을 개설하여 시행기관의 자체평가 역량을 제고	시행
2015	통합평가제도 개선방안	- 각 시행기관은 평가 예산 증가율을 ODA 예산 증가율과 동일한 수준으로 확보 - 평가 담당 직원의 교육 참여 권장 및 ODA 사업 전담기관의 평가 담당 직원의 전문직위제 실시 - 국무조정실 내 평가 전담 부서 설치 추진 - 장기적으로 ODA 평가 전담 기구 신설·지정 검토 - 국무조정실 내 평가 Helpdesk를 구축하여 각 부처 평가사업에 대해 자문·지원	제2차 기본 계획에 반영
2015	제2차 국제개발협력 기본계획 (2016~2020)	- 외부평가 강화, 독립평가부서 구축, 예산 및 인력 증가 - 각 시행기관의 평가예산을 점진적으로 확대 - ODA 사업전담기관의 평가 담당 직위의 전문 직위제 실시 - 국무조정실 내 평가 전담 부서를 설치하여 각 부처 평가업무를 관리 감독 및 지원	추진 중
2015	2016년 시행계획	- 자체평가 내실화 - 평가교육을 개편하여 각 시행부처 평가 담당자의 역량 강화를 지원	추진 중

자료: ODA KOREA 홈페이지, www.odakorea.go.kr에서 2016. 10. 1. 인출.

3. 메타평가

- 메타평가란 ODA 시행기관이 실시한 자체평가의 일부를 평가소위원회에서 심층 분석하여 평가의 질을 측정하는 것을 의미하며, 이를 통해 자체평가의 품질을 향상시키고 평가결과의 활용을 제고하고자 함.
- 제14차 국제개발협력위원회는 메타평가를 본격적으로 실시하기 위한 기반을 구축해 나가기로 결의(2012년 12월)하고, 평가소위원회의 과제로 2년마다 자체평가에 대한 메타평가를 실시할 것을 의결
- 평가소위원회는 시범 평가로서 2012년 ‘자체평가에 대한 메타평가’를 처음으로 실시
 - 2010~2011년에 원조 시행기관이 실시한 39개 자체평가 중에서 16개를 표본 추출하여, 해당 기관의 평가시스템, 평가과정, 평가보고서를 메타평가의 대상으로 삼았음.
 - 평가 방식은 외부 전문기관(한국개발전략연구소: KDS)에 평가 용역을 위탁한 후 그 결과를 토대로 소위평가를 실시.
 - 메타평가의 분석틀을 마련하여 평가 항목표를 구성하였으며, 평가 결과로부터 개선방안을 도출(국제개발협력위원회, 2012a, pp.1-2).
- 2014에 실시한 두 번째 메타평가는 대외경제정책연구원에 평가 용역을 위탁한 후 그 용역결과를 토대로 평가소위원회가 평가를 실시
 - 평가대상은 2012~2013년에 실시된 자체평가 69건 중 표본 추출한 29건.

- 방법론으로는 자체평가 단계를 평가계획, 평가과정, 평가결과로 구분하여 시행기관의 단계별 평가품질을 검토하되, 메타평가 분석틀을 마련하여 평가품질을 등급화.
- 동 메타평가 실시를 통해 평가소위원회는 메타평가 개선방안 및 메타평가 가이드라인(안)을 마련(국제개발협력위원회, 2014a, pp.25-36).

4. 국별·분야별 평가

□ 국별 평가는 총 11개국에 대해 16건이 실시

- 2010년 통합평가체제가 출범한 이후로 평가소위원회가 실시한 ‘국별 통합평가’가 2건이며, EDCF와 KOICA가 각자 실시한 국별 평가가 14건임.

□ 분야별 평가는 11개 분야에 대해 21건이 실시됨.

- 평가소위원회가 실시한 ‘분야별 통합평가’는 5건, EDCF는 5건, KOICA는 11건임.
- 상기 기관 이외의 ODA 시행기관들은 프로젝트평가에 치중하고 있어서 국별 평가나 분야별 평가 실적이 없음.

〈표 4-4〉 메타·국별·분야별 평가 현황

분 류	기 관	제 목	연 도	용역 기관
메타평가	평가 소위원회	자체평가에 대한 메타평가	2012	한국개발 전략연구소
		ODA 자체평가(2012~2013년)에 대한 메타평가	2014	대외경제 정책연구원
국별 평가	평가 소위원회	캄보디아 ODA 지원 종합 평가	2010	직접평가
		미얀마 국별 종합 평가	2015	서울대학교 국제개발협력 평가센터
	한국수출 입은행 (EDCF)	EDCF 베트남 국별지원 종합평가보고서	2010	대외경제정책 연구원
		인도네시아 국가지원전략 및 지원사업 평가	2011	한국개발전략 연구소
		방글라데시 국별평가 보고서	2012	대외경제정책 연구원
		필리핀 국별평가 보고서	2013	글로벌개발 협력건설팅
	한국국제 협력단 (KOICA)	과테말라 지원사업 국별평가 보고서	2008	한국개발전략 연구소
		탄자니아 지원사업 국별평가 보고서	2008	대외경제정책 연구원
		베트남 국별지원전략 및 지원사업 평가결과보고서	2009	한경대학교 개도국기술 협력연구소
		인도네시아 국가지원전략 및 지원사업 종합평가	2009	한국개발전략 연구소
		우즈베키스탄 국별지원전략 및 지원사업 종합평가	2010	산업연구원
		이집트 국별지원전략 및 지원사업 종합평가	2010	한국개발전략 연구소
		필리핀 국별지원전략 및 지원사업 종합평가	2010	우송대학교 국제교육연구소
		CPS 성과관리를 위한 국별 평가방안 연구-시범국: 파라과이	2012	(사)글로벌 발전연구원
분야별 평가	평가 소위원회	ICT 분야 ODA 통합평가	2011	산업연구원
		농림수산 분야 ODA 사업 통합 평가	2012	한국농촌경제 연구원
		보건 분야 ODA 통합 평가	2013	서울대학교병원

분 류	기 관	제 목	연 도	용역 기관
		수자원 분야 ODA 통합평가	2014	한국수자원공사연구원
		교육 분야 ODA 통합평가	2014	APEC국제교육협력원·상명대학교
	EDCF	상·하수도 분야 평가보고서	2011	한국수자원공사
		ICT 분야 평가보고서	2012	산업연구원
		교통 분야 종합평가	2013	한국교통연구원·서영엔지니어링
		에너지 분야 종합평가	2014	한국개발정책학회
		전자정부 분야 종합평가	2015	국민대학교 산학협력단
	KOICA	베트남 병원건축사업 사후평가보고서	1998	평가보고서 미공개로 알 수 없음 (평가보고서 공개시점 이후부터 분야별 평가 시행)
		중남미 보건의료지원사업 사후평가보고서	1999	
		잠업개발 지원사업 사후평가보고서	2000	
		중남미 식수공급 지원사업 사후평가보고서	2000	
		IT사업 수행체제 평가 보고서	2002	
		메콩강위원회(MRC) 지원사업 평가보고서	2002	
		직업훈련 프로젝트사업 외부위탁 평가결과보고서	2003	
		거버넌스 관점에서 본 한국 국제개발협력사업	2005	
		농업분야 협력사업 평가	2005	
		한국국제협력단 보건의료 협력사업 평가	2005	
		교육분야 지원사업 분야별 평가보고서	2008	한국개발전략연구소

주: 상기 평가 현황은 공개된 평가보고서만을 대상으로 작성하였음. 현재 각 기관의 웹 사이트상에 KOICA는 2006년 이후의 평가보고서를, EDCF는 2011년 이후 평가보고서만을 공개하고 있음.

자료: ODA KOREA(www.odakorea.go.kr), 한국수출입은행(www.edcfkorea.go.kr), 한국국제협력단(www.koica.go.kr)에서 2016. 10. 1. 인출.

5. 종합적인 평가계획 흐름 및 시행 여부

□ 정부는 중기계획, 선진화 방안, 분야별 기본계획 등 ODA 관련 주요 문서에서 평가의 독립성 확보, 환류 강화, 평가역량강화를 평가와 관련한 주요 개선 과제로 제시해 왔고, 매년 수립되는 시행계획도 세 가지 개선 과제의 이행을 지속적으로 언급하고 있음. 시행 성과를 살펴보면, 독립성 확보를 위한 통합평가체계 마련과 환류 강화 면에서 제도적인 개선 성과가 뚜렷하지만, 원조시행기관의 평가역량강화 부문은 성과가 미흡한 것으로 보임.

○ **(평가 독립성 확보)** 평가의 독립성 확보는 「국제개발협력 개선방안(2005)」 이후로 현재까지 지속적으로 논의되고 있는 중점 과제 중 하나임. 이전에는 원조사업을 시행하는 기관이 각기 자체적으로 평가를 수행하여 평가의 질과 객관성이 미흡하던 상황에서 정부는 평가소위 중심의 통합평가체제를 수립하고, 통합평가지침과 매뉴얼 등을 마련(2009년)하여, 평가의 독립성을 높일 수 있는 기반을 마련.

- OECD DAC은 2012년 동료 평가 결과로, 평가소위원회의 독립성과 절차를 더욱 강화하여 평가기능을 높일 것을 권고하였음(국제개발협력위원회, 2013b, p.9).

○ **(환류 강화)** 평가결과의 환류(Feedback)는 「국제개발협력 선진화 방안」과제 1, 2차 「분야별 기본계획」에서 ODA의 주요 개선 과제로 언급, 이에 따라 정부는 평가결과의 환류 강화를 골자로 하는 「국제개발협력 통합평가지침」을 개정(2014년 1월 13일)하여, 각 시행기관으로 하여금 평가결과 반영계획을 수립하고 반기별로 그 이행상황을 점검하게 하고 있음.²⁸⁾

- 정부는 이행상황에 대한 점검을 통해 평가결과로 도출되는 대부분의 정책제언이 추상적이며 차기 사업에 적용하기가 어려울 뿐 아니라 연구용역이 지연되어 환류계획 수립이 늦어진 경우 등이 발견되고 있다고 지적한 바 있음(국제개발협력위원회, 2012b, p.7).

○ **(평가역량강화)** 정부는 평가역량과 객관성을 높이기 위해 수원국·국제기구·관계기관 간 공동평가를 지속적으로 추진하는 한편²⁹⁾ ODA 담당자의 평가 관련 이해도를 높이고 시행기관의 자체평가 수행 능력을 향상하기 위해서 ‘ODA 평가교육 프로그램’을 마련해서 실시하고 있음.

- 2014년 7월 국무조정실 주도로 KOICA 교육원과 협의하여 시범교육과정을 마련하였는데, 국무조정실은 평가교육 참가자의 조사와 선발, 사후관리, 향후 교육과정 조정과 기획 등 전반적인 평가교육 과정을 주관하고, KOICA 교육원은 평가교육 프로그램을 개발하고 평가교육 과정을 시행(국제개발협력위원회, 2014b, pp.10-12).
- 2015년부터는 KOICA 교육원에 ODA 평가 관련 정규과정을 신설하여 연 2회 실시하기로 의결(국제개발협력위원회, 2015b, p.44).

28) 「국제개발협력 통합평가지침」제11조(평가결과반영계획의 수립), 제13조(평가결과의 반영), 2014년 1월 13일 신설.

29) 국제개발협력위원회(2012. 1. 31). 제9차 평가소위원회 의결안건(제9-1호), 2011년 ODA 자체평가결과, p.7: (수원국과의 공동평가 사례) KOICA-콜롬비아 ‘중미 직업훈련 사업’ 공동평가, EDCF-캄보디아 ‘직업훈련원 건립사업’ 공동평가, (관계기관 공동평가 수행 사례) 농식품부-KOICA ‘캄보디아 농촌분야 사업’ 공동평가.

〈표 4-5〉 종합적인 평가계획 흐름 및 시행 여부

시 기	문 서	주요 내용	시행 여부
2005	국제개발협력 개선방안	독립적 외부평가 방식 도입	실시: 통합평가 체계 마련
		기존 시행기관의 평가담당 조직 및 평가기능 강화	미흡
		평가 규정을 제정	실시: 국제개발 협력법 제정
2007	ODA 중기전략 (2008~2010)	‘통합평가 가이드라인(지침)’ 제정 추진	실시: 통합평가 지침·매뉴얼 제정
		국개위 중심으로 종합평가 실시	실시
		원조기관 내 평가예산 확보 및 전문인력 보강	미흡
		선진 공여국 및 국제개발원조기구 등과 공동 평가를 실시하여 평가 경험 및 지식을 축적	실시
2010	국제개발협력 선진화 방안	총리실 평가소위 중심의 유·무상 통합평가 방 식으로 개편	실시
		평가결과와 환류시스템을 제도화하고 평가결 과는 공개	실시
2010	분야별 ODA 기본계획 (2011~2015)	‘통합평가 추진방안’에 따라 자체평가 역량을 강화	미흡
		독립된 평가조직을 설치하고, 외부전문가·협 력대상국 관계자 등의 참여를 통해 평가결과 의 객관성과 신뢰성을 제고	실시
		프로젝트 평가뿐만 아니라 정책평가, 국별평 가, 주제별평가 등으로 평가대상을 다변화	실시
		평가결과는 대국민 공개, 다음연도 사업계획 수립에 반영	실시
2015	제2차 ODA 기본계획 (2016~2020)	각 시행기관의 평가 예산을 점진적으로 확대	추진 중
		ODA 사업전담기관의 평가 담당 직위의 전문 직위제 실시	
		대규모 사업에 대한 외부평가 실시로 재정책 입성 및 투명성 확보	
		국무조정실 내 평가 전담 부서를 설치하여 각 부처의 평가업무를 관리·감독하고 지원	
		사업 심사 시 당해 사업의 기존 실적 평가결과 와 연계하여 환류를 강화	

자료: ODA KOREA 홈페이지 www.odakorea.go.kr에서 2016. 10. 1. 인출.

제2절 보건의료분야 평가사례

1. 부문별 평가: 보건

□ 분야별 ODA 기본계획의 평가 다변화에 따라 2013년 ‘보건분야 ODA 종합평가 연구’(오주환 등, 2013) 수행.

○ (평가목적) 보건 ODA 전반과 개별 보건사업에 대한 평가를 통해 보건분야 사업현황을 평가하고, 효과를 제고하기 위해 부처 간 조정, 프로그램 보완, ODA 일관성 및 효율성을 제고하는 대안을 도출하고자 함(오주환 등, 2013, p.1).

○ (평가대상) 기획재정부 및 EDCF, 외교부 및 KOICA, 보건복지부 및 KOFIH의 보건분야 ODA 전략과 체계, 사업내용(오주환 등, 2013, p.2)

- 대상기관의 정관 및 법, 조직과 조직구조, 인원과 인력관리 및 사업추진 체계, 모니터링 및 평가, 국내외 파트너십
- 대상기관이 추진한 보건분야 사업의 적절성, 효과성, 효율성, 영향력, 지속 가능성

□ 결과 및 6가지 제안사항(오주환 등, 2013, pp.169-182)

- 보건의료 원조기관 간 협업 수행체계 수립 필요
- 기관 간 협력(결과)에 대한 인센티브 제공
- 기관 간 협력(과정)에 대한 권고안 제공
- 기관 간 협력과정에서의 분업에 대한 권고안 제공
- 각 기관의 자료 통합 게시
- 각 기관의 보건의료분야 원조효과성 강화를 위한 원조 실행 상황

모니터링 정례화 및 사업종료 평가 독립성 강화

□ 보건분야별 평가가 보건분야 ODA 사업평가에 주는 시사점

○ 국내 보건분야 ODA 추진체계를 분석한 결과, 분야와 무관하게 자체 평가, 모니터링 강화 등의 성과관리와 사업평가에 대한 요구가 있고, 보건분야의 유·무상 연계 강화 등의 조치가 평가를 통해 원조효과성 제고로 이어지지 못하는 한계가 지적되고 있음.

- 분석결과, 1) 국제개발협력위원회 중심의 통합적 추진체계가 공고화되고 관계기관 간 연계·협업 환경이 개선되었고 2) 평가 대상기관의 체계적 사업 수행 역량이 개선되었으나 인적 역량강화 및 자체 평가 등의 과제가 있고 3) ODA 모니터링 시스템 활용 개선이 필요하며 4) 그간 보건분야 ODA 관계기관 간의 분업 논의를 재고할 필요성이 있다고 판단(오주환 등, 2013, p.13).
- 보건분야 유·무상 연계, 기관 간 협업 등 조정의 성과가 결과 관리로 나타나지 않는 한계를 극복할 필요가 있음(오주환 등, 2013, p.22).

○ 대상기관 자체평가 개선 필요성(오주환 등, 2013, p.25)

- 자체 평가의 기틀이 마련됨.
 - 2013년 11월 현재 EDCF와 KOICA 자체 평가매뉴얼 보유, KOFIH는 제작 중임.
 - 평가의 중요성에 대한 인식 공유
- 평가의 독립성, 자체평가 미비, 평가결과의 환류기제 미비
 - EDCF는 평가업무를 사업부서에서 관할하여 객관성·독립성

논란이 제기될 여지가 많음.

- 향후 일정 규모 이상의 ODA에 관해서는 자체 평가 의무화 필요가 있음.

□ 첫 번째 시행된 보건분야별 평가로 기관별 보건의료 ODA 현황을 파악하고, OECD·DAC 5대 평가원칙에 따라 종합적으로 평가해 보는 의의가 있으나, 초점이 보건의료 원조기관 간 협업에만 모아져 있어 어떤 내용의 사업이나 프로그램이 수행되어야 할지에 대한 비전을 제시하지 못하는 한계가 있음.

- 보건분야 ODA 사업은 실행 주체 간 협력 및 조정이 여전히 더욱 강화될 필요가 있음.
- 하지만 단기 프로젝트형 사업의 비중이 커 장기적·종합적인 기획이 어려운 점과 사업의 다양성이 미비한 점이 보건분야의 주요한 문제점으로 지적되고 있음(김현경 등, 2015, pp.163-164).
- 보건분야 성과평가가 미비하고 성과관리 인력 부족 역시 꾸준히 지적되고 있음(김현경 등, 2015, pp.165-166).

□ KOICA의 보건의료분야 성과관리 가이드라인 마련을 위해 분석한 KOICA의 성과관리 제반여건에 대한 분석내용은 KOICA(2013b)에 요약되어 있음.

- 메타평가의 진단과 평가결과와 유사하여 이로부터 개선방안을 도출하기 용이함.

〈표 4-6〉 KOICA 성과관리 조사 결과 요약

구 분	항 목		주요 내용
성과 관리 체계 및 제반 환경	성과관리 관련 역사적 흐름		KOICA 내에서 성과관리에 대한 논의는 매우 최근에 시작 성과관리에 대한 기본적인 정책적 방향이나 시행계획은 마련되어 있을지라도 실제 그것이 KOICA의 내부 프로젝 트 관리에 적용되어 내재화되기에는 충분한 시간이 필요 할 것임
	성과관리 관련 제반환경		한국 정부차원의 분산된 정책적 움직임(정부 차원의 국제 개발협력 개선방안 및 CPS 마련, 한국형 ODA 추진전략 마련, KOICA 차원의 신진화 계획 및 국가협력전략 마련 등)으로 인해, 실질적으로 각 세부사업을 어떤 차원의 전략 에 어떻게 일치(Aligning)시킬 것인지에 대한 문제가 존재
	성과관리 관련 교육훈련 제공을 통한 인식 제고		ODA 평가실 주도로 제공된 교육훈련 기회를 통해 성과관 리의 기본개념, PDM 작성법, 평가 등에 대한 기본적 지식 축적이 이루어졌으나, 실제 성과관리에 대한 개념과 현업 에의 적용도 측면에서는 단내 직원들 간 개인차가 상당히 존재
전략 수준 성과 관리	보건의료분야 전략 수립 및 성과관리 현황		2011년 <KOICA 보건의료 전략(2011-2015)>을 마련하였으 나 KOICA의 보건의료분야 정책적 체계 수립 노력은 비교 적 광장히 최근에 시작되었기 때문에, 현재까지 구체적인 정책적 지침과 전략적 방향성을 제시하기에는 한계 존재
	국별 전략 수립 및 성과관리 현황		일부 기 수립된 전략들(예 : 일부 범정부 CPS, KOICA의 분 야별 전략)이 존재함에도 불구하고, 전략-사업 연계 부족 문제는 계속하여 존재 사업발굴과정에서의 요청주의 및 정치/외교적 요인이 작 용으로 인해 국별 프로그래밍 시행에 한계 존재
사업 수준 성과 관리	기획 단계	수원국 보건 현황 및 전략 검토	사업 발굴 및 기획 시 수원국의 보건 전략을 검토한다는 답변은 많았던 반면, 보건 전략의 R/F는 검토하고 있지 않다는 답변이 많았음
		아국 보건전략 검토	현재 KOICA 실무자들이 아국의 전략을 검토하여 사업과 일치하려는 노력을 하고 있으나 전략 간 연계가 부족하여 혼란의 가능성이 있음
		사전조사(타당 성 조사) 및 기준치 자료 수집	사업대상지 지역수준의 기본 정보 및 기준치 자료 수 집 과정이 적극적으로 이루어지고 있지는 못한 것으로 나타남

구 분	항 목	주요 내용
	PDM 작성	PDM 작성 시 성과 목표 설정 단계에서 투입(Input)과 산출물(Output)에 대한 규정은 이루어지고 있으나, 상위 성과인 산출 수준이나 목표 수준의 성과 명제 도출은 충분히 이루어지지 못하고 있음 지표 층위간의 논리적 연관성을 검토하는 것이 중요하나, 현재 KOICA의 사업 타당성 검토 시 이러한 부분이 충분히 점검되지 못하고 있음 확인수단 설정 부분에서는 보건의료 전문가에 대한 의존도가 상당히 높으며, 위험 및 가정 서술 시에도 명확한 인식이 공유되어 있지 못한 것으로 나타남
		M&E 계획 수립
		PDM 및 M&E 계획 공유
	모니터링 단계	일련의 M&E 계획 수립과정이 현재 KOICA내에서 활발히 이루어지지 못하고 있으며, 특히 지표 별 목표치와 이정표를 설정하는 것에 어려움이 큰 것으로 조사됨
		모니터링
		PDM 업데이트
	평가 및 환류 단계	모니터링 결과 취합 및 보고
		평가성 사정 및 평가 TOR 작성
		평가 결과 환류 체계

자료: KOICA(2013b). pp.248-249.

2. 프로젝트평가: 모자보건

가. 과테말라 모자보건(KOICA)³⁰⁾

□ 평가배경 및 목적

- KOICA의 과테말라 모성 및 아동 사망률 감소를 위한 ‘과테말라 치말테낭고 주 모자보건 개선 사업’을 통해 목표 달성 정도를 측정하고 결과에 영향을 미치는 요인에 대한 분석을 통해 향후 모자보건사업을 위한 개선방안을 도출하고자 평가함.

□ 사업 내용 및 사업 내역

- 사업기간은 2007-2009년
- 사업비는 300만 달러(3년간)

〈표 4-7〉 KOICA 과테말라 치말테낭고주 모자보건 개선사업 내용

구분	내 용	
한국	- 모자보건센터 건립	- 치말테낭고주 국립병원 내 모자보건센터 신축
	- 모자보건소 건립	- 치말테낭고주 파춘군에 모자보건소 신축
	- 기자재 지원	- 모자보건센터와 모자보건소 필요 기자재 지원
	- 의료인력교육	- 보건행정가, 의사, 간호사 등 의료인력에 대한 교육프로그램 지원
	- 기타	- 조사 및 평가비
과테말라	- 건축부지 제공 - 기자재 설치를 위한 공간 확보 - 기자재 통관 및 운송 지원 - 파견 전문가 업무수행지원(사무공간, 자료 제공 등) - 프로젝트 수행 관련 행정 및 인력지원 등	

자료: KOICA(2013a). p.10에서 발췌.

30) KOICA(2013a). 과테말라 치말테낭고주 모자보건개선사업 사후평가 보고서(2013.12). 를 요약·정리한 것임.

□ 사업 목표

- 개발목표(상위목표): 모자보건분야 MDGs 달성 기여
- 사업목표(하위목표): 안전한 출산환경 조성, 산전관리 서비스 확대

□ 평가 방법

- OECD DAC 5대 평가기준에 따른 평가 지표 범분야 이슈(성 주류화) 평가 지표에 따른 평가매트릭스

〈표 4-8〉 OECD DAC 5대 평가기준과 범분야 이슈에 따른 평가지표

평가기준 Criteria	세부질문 Sub-Questions	평가지표 Performance Indicators	자료출처 Sources of Information	정보수집방법 Information Collection
적절성	• 동 사업은 과테말라의 국가발전전략 및 분야별 개발 우선순위에 부합하는가?	• 과테말라 국가발전 전략과의 일치성 • 과테말라 보건사회부 정책과의 일치성	• 과테말라 국가개발계획 • 과테말라 보건사회부 정책 • 정부관계자	• 문헌조사 • 인터뷰
	• 한국의 ODA 정책 및 과테말라 지원전략, KOICA 보건분야 지원정책에 부합하는가?	• 한국의 과테말라 지원정책과의 일치성 • KOICA 보건부문 원조정책과의 부합성	• 대 과테말라 지원정책 • KOICA 보건분야 중기계획 • KOICA 통계	• 문헌조사
	• 사업의 목표가 MDGs 달성과 연관이 있는가?	• 과테말라 사업목표 및 MDGs 목표 3.4.5 관련 항목 비교분석	• 과테말라 MDGs 지표	• 문헌조사
	• 타 공여기관의 모자보건 개선사업과 조화를 이루는가?	• 타 공여기관의 모자보건 지원 사업 현황	• 타 공여기관의 모자보건 지원 사업 계획서 및 보고서	• 사례 비교조사
	• 사업계획 시 수원국 및 수혜자의 의견 반영을 위한 적절한 절차가 있었는가?	• 과테말라의 지역별 모자보건 수준 • 모자보건시설 운영예산의 적절성 • 모자보건시설 활용여부 및 활용 비율	• 모자보건현황 분석자료 • 사전조사보고서 • 실시현의결과 보고서 • 중간평가보고서 • 과테말라 정부관계자	• 문헌조사 • 인터뷰 • 설문조사
	• 사업대상지역 선정과정에서 수원국 정부의 의 충분한 협의가 있었는가?	• 대상지역 협의과정	• 사전조사보고서 • 실시현의결과 보고서 • 중간평가 보고서 • 과테말라 정부관계자	• 문헌조사 • 인터뷰 • 설문조사
효율성	• 개발된 산전관리 서비스는 산모 수요에 부합하는가?	• 산전관리 프로그램과 모자보건 현황과의 부합성	• 과테말라 보건의료현황 • 중간평가 보고서	• 문헌조사 • 인터뷰 • 설문조사
	• 모든 인적, 물적자원 (인력, 자금, 시간)은 시의 적절하게 투입되어 최소비용으로 최대효과를 산출하도록 사용되었는가?	• 인력투입의 시의 적절성 • 자금의 시의적절성 • 시간의 적절성	• 사전조사보고서 • 실시현의 결과보고서 • 중간평가보고서 • 사업종료보고서 • PMC 및 관계자	• 문헌조사 • 인터뷰
	• KOICA의 유사사업과 비교할 때, 산출물 대비 투입이 효율적이었는가?	• 센터 및 보건소 신축비용 비교 • 기자재 지원 비용 비교	• KOICA의 모자보건 사업문서	• 사례 비교 조사
	• 이해관련자인 효율적인 조정 및 의사소통 채널은 수립되었는가?	• KOICA, 과테말라 보건사회부 및 PMC 협의기구	• 사업종료보고서 • 후원원 운영 기본계획안 • PMC 및 관계자	• 문헌조사 • 인터뷰
	• 효율성을 해치는 구조적 요인들은 무엇인가?	• 보건사회부의 모자보건 사업운영 시 역할 • 보건사업에 대한 리더십	• 과테말라 의료운영 시스템 • 의료운영 및 역량강화 조직 및 관리(거버넌스) • 이해관계자	• 문헌조사 • 인터뷰

평가기준 Criteria	세부질문 Sub-Questions	평가지표 Performance Indicators	자료출처 Sources of Information	정보수집방법 Information Collection
	• 계획 대비 사업의 산출물(output)이 일치하는가? - 모자보건시설 신속시 계획된 수용시설 및 기자재가 모두 설치 및 활용되었는가? - 의료인력(간호사, 산파)의 역량강화 프로그램이 제공되었는가? - 임신부 산전관리 사업 및 교육자료가 개발되었는가?	• 계획 대비 모자보건시설 및 기자재 설치 및 활용 정도 • 모자보건센터 및 보건소 운영 기본계획 수립 • 의료인력 역량강화사업 및 교육자료 품질 • 임신부 관리사업 및 교육자료 품질	• 모자보건시설 검수결과표 • 모자보건시설 운영기본계획안 • 의료인력 역량강화 커리큘럼 및 교재 • 초청연수결과 보고서 • 전문가 파견 보고서	• 문헌조사 • 직접조사
효과성	• 모자보건센터 및 보건소 운영 효과를 높이기 위한 방안은 있는가?	• 훈련된 운영 기본계획안	• 사업종료보고서 • 종료평가보고서 • 훈련된 운영 기본계획안	• 문헌조사
	• 사업을 통해 계획된 단기사업목표(initial outcomes)대로 출산환경이 사업시행 이전보다 개선되었는가? • 분만시 모자보건센터 및 보건소 활용이 증가하였는가? • 훈련받은 모자보건 의료인력이 증가하였는가? • 임신부 대상 수술 시스템이 구축되었는가?	• 모자보건센터 및 보건소 내 분만인용품 • 훈련받은 모자보건 의료인력의 증가율 • 훈련받은 모자보건 의료인력에 의한 출산 증가율 • 임신부 응급수술 시스템 구축 • 제공된 기자재의 질적 수준 및 활용도	• 모자보건시설 운영 현황 자료 • 교육받은 의료인력의 활동현황 • 커리큘럼 및 교재 • 교육비용 운용대장 • 초청연수결과보고서 • 전문가 파견 보고서 • 병원관계자	• 문헌조사 • 직접조사 • 인터뷰
	• 과테말라의 모자보건환경이 개선되었는가?	• 모성 및 아동 사망률 개선정도 • 교육받은 의료인에 의한 출산 증가율	• 과테말라 보건 의료 현황 • 정부관계자 • 지역주민	• 문헌조사 • 설문조사 • 인터뷰
영향력(파급효과)	• 과테말라 내 모자보건개선에 대한 인식 이 개선되었는가?	• 현지 임신부 및 가임기 여성의 산전관리 프로그램 등록률 • 교육받은 산파의 수	• 과테말라 보건 의료 현황 • 지역주민	• 설문조사 • 인터뷰
	• 관련 분야 내 보건 의료 사업이 확대 되었는가?	• 확대 된 개발 된 보건 의료 프로그램의 현황 • 지원된 모자보건센터의 이용률 및 이용 목적	• 과테말라 보건 개발 사업 현황 보고서	• 직접관찰 및 인터뷰 (관계자)
지속가능	• 과테말라 정부는 지속적인 정책적, 제도적 지원이 가능한가? - 수원국이 정책 및 제도적으로 모자보건개선에 대한 지원전략을 수립하고 있는가? • 중앙-지방간 모자보건 의료체계의 연계가 잘 이루어졌는가?	• 과테말라의 모자보건개선 중장기계획 수립 및 진행 유무 • 중앙-지방간 모자보건의료체계 협력 현황 및 향후 계획	• 사전조사보고서 • 실시결과결과보고서 • 사업종료보고서 등 • 정부관계자	• 문헌조사 • 인터뷰
	• 모자보건시설이 효과적인 운영 조직을 갖추고 있는가?	• 모자보건시설 조직도 및 업무분장	• 모자보건시설 조직도 및 운영계획서	• 문헌조사
성	• 모자보건시설이 관련시설을 운용 및 유지할 기술적 역량을 보유하고 있는가?	• 의료인력의 훈련시설장비 운용 및 보수 기술	• 기자재 관리대장 조직도 및 운영계획	• 문헌조사
	• 모자보건사업의 결과가 유지 및 확산될 수 있는가? - 산전관리시스템이 과테말라 사회문화 조건에 조화되는가?	• 과테말라 정부의 교육 훈련 재정 시스템 • 교육비 책정 현황 • 인력 투입 현황 • 교육받은 산파에 의한 수해자 증가율	• 연도별 모자보건 관련 운영 예산 및 지출 현황 • 과테말라 보건 개발 사업 현황 보고서 • 병원관계자	• 문헌조사 • 설문조사 • 인터뷰
기타 (성 주류화)	• 성 주류화에 대한 고려가 사업의 기획단계에서부터 포함되었는가? • 사업이 수혜자의 젠더관계 및 양성평등에 미친 영향이 있는가? • 과테말라 여성 보건인력의 사회적 역량강화가 이루어 졌는가?	• 성 주류화에 대한 사업 전략의 유무 • 여성의 사회적 참여 및 위치 신장 • 모자보건사업에 대한 남성 및 여성의 인식 및 반응 변화	• 사전 타당성 조사 보고서 • 종료 평가 보고서 • 과테말라 보건 개발 사업 현황 보고서 • 지역 여성주민	• 문헌조사 • 설문조사 • 인터뷰

자료: KOICA(2013a). pp. 87-90.

□ 평가의 한계 및 제약 요인

- 모성사망률, 임산부 등록률, 시설 분만을 등 통계자료들을 수집하기 어려움. 더 나아가 정부기관 및 병원의 데이터 관리 시스템이 제대로 구축되지 않아 입수한 지표의 표시방법이 일관적이지 않음.
- WHO에서 사용하는 시설 분만, 숙련된 인력에 의한 분만, 비숙련 인력에 의한 분만, 가정분만이라는 분만방법에서의 구분이 과테말라에서는 그대로 적용이 안 됨. 즉, 시설에서의 분만이 곧 숙련된 인력에 의한 분만을 의미.
- 과테말라 내 산파는 기존의 조산사와 구분됨. 즉, 하는 일은 같으나 산파는 비의료 전문 인력임.
- 계획했던 건축 관련 회사와 연락이 불가하여 면담이 이루어지지 않음.
- 설문 관련 협조가 원활하게 이루어지지 않음.

□ 평가의 문제점과 개선방안

- 평가에서 활용된 방법은 대체로 심층 면접 및 설문조사로 대체로 정성적인 평가임.
- 객관적 자료수집의 어려움이 있으나 정량적 지표를 산출하려는 노력이 필요할 것으로 보임.

나. 라오스 모자보건(KOFIH)³¹⁾

□ 라오스 통합모자보건 증진사업 1차 종료평가 배경 및 목적

- 라오스 모성 및 아동 사망률 감소를 위한 ‘라오스 국가 모자보건 전략 2009~2015’ 지원을 목적으로 하는 1차 ‘라오스 통합모자보건 증진사업’이 종료됨에 따라, 주요 성과를 정리하고 이를 기반으로 지역보건의료체계의 개선방안을 도출하기 위해 평가함.

□ 1차 ‘라오스 통합모자보건 증진사업’의 개요

○ 라오스의 국가 통합모자보건 전략 2009~2015 수립 배경

- 1991~2003년 호주 세이브더칠드런(SCA: Save the Children Australia)에서, 사야보리(Sayaboury Province) 지방 보건부와 함께 보건의료시설 접근성 향상, 모성사망률과 이환율 향상을 위해 모자보건 서비스를 활용하였고, 포괄적 일차 보건의료 역량강화 프로그램을 수행함.
- 프로그램 수행 결과 보건의료시설에 대한 접근성과 이용률이 향상되고, 신생아 사망률, 5세 이하 아동 사망률에 있어서 향상됨.
- SCA의 지역보건 프로그램은 모자보건사업의 표준 모델로 2006년에는 세계보건기구(WHO)에서 2009년에는 유니세프 연차 보고서에서 사례연구로 다루어짐.
- 이를 통해 사업이 확대되어 라오스의 모자보건 전략을 통한 지

31) 이명선 등(2014). 라오스 통합모자보건 증진사업 1차 종료평가 결과보고서. 한국국제보건의료재단(KOFIH) 발간자료를 요약·정리한 것임.

역보건의료체계 강화를 위해 벤치마킹 사례로 인식됨.

○ 라오스 국가 모자보건 전략 2009~2015 추진과정

- 2007년 라오스 보건부는 모자보건 관련 목표 달성을 위해 범 분야적 조율기구(sector wide coordination mechanism) 역할을 하는 정부기관, UN기구, 양자 간 공여주체, 비정부조직인 모자보건 기술 실무 그룹을 설비하고 WHO와 해외 원조기구 등 협력조직(UNICEF, UNFPA, WB, ADB, JICA, BTC, Lux-Dev 등)과 함께 정책을 입안하고 실행함.
- 2009년 라오스 국가 모자보건 전략을 세우고 지역보건의료체계 역량강화를 위해 모자보건 영역을 국가 중점과제로 추진함.

○ 한국국제보건의료재단과의 협력

- WHO 라오스 사무소 모자보건 팀이 2010년 한국에 방문하여 보건복지부와 한국국제보건의료재단과 함께 라오스 모자보건 증진사업의 필요성과 추진을 위한 협력방안에 관해 협의.
- 해당 사업은 기존 원조의 한계인 수직적 접근방식에서 벗어나 한국의 모자보건사업 핵심 요소인 지역사회에 기반한 통합적 접근을 활용하는 ‘한국형 보건의료 원조 모델’로 발전할 가능성이 있다고 평가하여 한국과 라오스 양국 보건부 간 MDG4, MDG5 등의 달성을 위한 포괄적 양자 협력사업 협약을 체결하고 WHO·WPRO와 사업 추진에 관한 협약을 2010년 5월 체결함.

○ 라오스 통합모자보건 증진사업의 사업 기간, 사업 대상, 사업비

- 2010~2012년
- 사업지역은 라오스 동북부 2개 도의 18개 군

- 대상자는 여성 121,000명, 5세 미만 아동 62,000명
- 사업비는 2010~2012년간 50억 3100만 원(연간 약 15억 규모)

〈표 4-9〉 KOFIH 라오스 통합모자보건 증진사업 규모

(단위: 원, %)

구분	2010	2011	2012
KOFIH 실행	1.9억(7.5)	3.4억(12.5)	6.2억(37.5)
WHO 실행	13.7억(92.5)	11.1억(87.5)	10.2억(62.5)
계(전체 실행)	15.6억(100)	14.5억(100)	16.4억(100)

자료: 이명선 등(2014). p.53.

○ 사업 목표

- 개발 목표(상위 목표)는 ‘중앙정부, 지방정부, 보건소, 지역사회까지의 거버넌스, 전달체계 구축 및 단계별 역량강화를 통한 라오스 모자보건 국가전략 이행 지원(이명선 등, p.53)’임.

○ 사업 내용

- 라오스 국가 모자보건 전략을 지원
 - 국가전략 이행을 위한 정부의 관리 능력을 강화하기 위해 보건소에 대하여 관리 감독.
 - 모자보건 전략 서비스 질 관리를 위해 조산사 양성 교육 프로그램 지원 등을 통한 서비스 제공자 인력 지원, 서비스 제공을 위한 필수의료기기 및 시설 지원, 보건소 직원을 통한 방문 모자보건 서비스 제공(백신 접종 및 산전 진찰), 보건소를 통한 무료분만 서비스 제공.

- 마을 주민의 인식을 제고하기 위해 보건교육 자료를 제작, 배포.
- 보건소의 서비스 질을 개선하고 인식을 제고하기 위해 보건소 시설 지원.
- 국가 모자보건 환경 개선을 위해 수원국 인력을 활용하여 기생충 퇴치 활동 수행.
- 현지 업무를 지원하기 위한 전문가와 행정 인력을 파견하고 사업 대상 지역에 모니터링을 포함하는 국내 기술 지원.

○ 사업 수행 특성

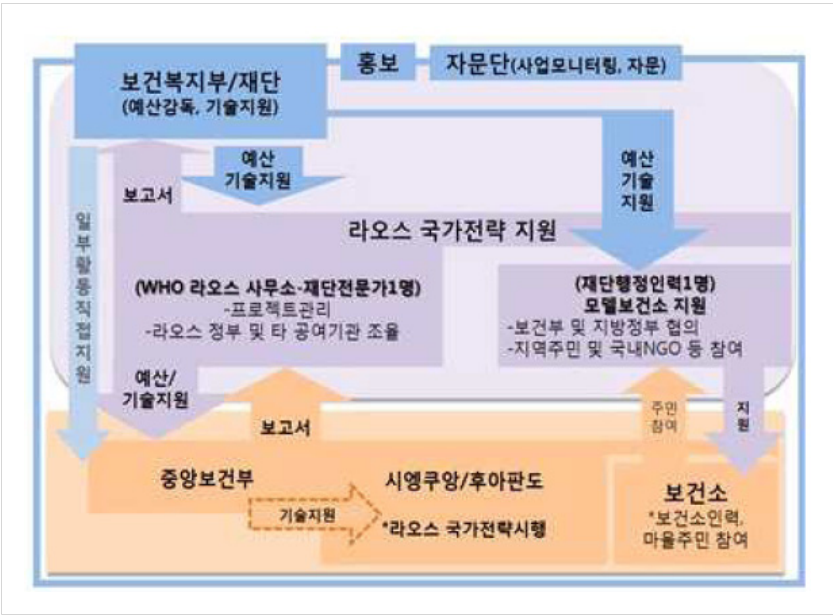
- 통합모자보건 증진사업의 활동이 단계별로 투입됨.
 - 개발 및 평가 워크숍을 우선 진행한 후 서비스 핵심 훈련으로 이행됨.
 - 시범사업 실시 후 사업을 확대하며 사업에 대한 평가 실시.
- 라오스 국가 모자보건 전략과 일치.
 - 사업의 내용이 라오스 국가 전략에 맞춘 보건의료체계 강화에 해당됨.

□ 1차 ‘라오스 통합모자보건 증진사업’ 평가의 개요

○ 평가주체

- 사업집행기관(KOFIH), 협력 사업수행기관(WHO), 수원기관(라오스 보건부)을 제외한 제3자인 대학 산학협력단에서 용역으로 평가 수행함.

[그림 4-1] KOFIH 라오스 통합모자보건 증진사업 평가주체 간 관계도



자료: 이명선 등(2014). p.59.

○ 평가 범위

〈표 4-10〉 KOFIH 라오스 통합모자보건 증진사업 평가범위

구분	세부 내용
라오스 통합모자보건체계 개선 1차 사업 종료에 따른 성과 평가	- 2010~12년 추진된 라오스 통합모자보건체계 개선사업의 추진배경 및 과정에 대한 기록 정리 - 사업평가 모형개발 및 모형에 따른 성과측정
2차 사업 추진방안에 대한 제언	- 지역단위 보건의료체계 강화사업을 위한 프레임워크 개발 - 지역보건의료체계 지표 측정을 통한 기본선 수립 등 현황 파악 - 모자보건사업 확대를 통한 지역보건의료체계 강화 방안 도출 - 보건의료체계 강화를 위한 지역의 내부 자원 및 동기 요소 파악

자료: 이명선 등(2014). p.14.

○ 평가 기간

- 2013년 9월 25일~2014년 2월 24일

○ 평가 지표: PDM 재구성에 따른 평가 지표

- 영향 지표(사업지역 모성사망률 및 영유아 사망률 감소라는 목표 지표): 모성사망률, 영유아 사망률, 5세 이하 아동 사망률
- 산출 지표
 - 기획과 관리: 자체 모자보건 마이크로플랜이 있는 보건소 비율(자체 모자보건 마이크로플랜이 있는 보건소 수/사업 지역에 있는 보건소 수)
 - 정보 시스템: 정보수집에 대한 훈련을 받은 마을 보건요원 비율(정보수집에 대한 훈련을 받은 마을 보건요원 수/사업 지역에 있는 마을 수)
 - 기술 훈련: 주요 모자보건 의료 서비스에 대한 훈련을 받은 보건소 직원 비율(모자보건 의료 서비스 훈련을 받은 스태프가 한 명이라도 근무하는 보건소 수/사업 지역에 있는 보건소 수)
 - 장비: 모자보건 전략 실시를 위한 필수 장비를 갖춘 보건소 비율(모자보건 전략 실시를 위한 필수 장비(분만실, 분만침대)를 갖춘 보건소 수/사업지의 보건소 수)
 - 지역사회 참여: 지역사회 참여를 위한 주요 자료 보급률(마을 보건요원이 VHV 매뉴얼, VHV 키트, 핵심 메시지 카드, 카운슬링 매뉴얼을 받은 마을의 수/사업지의 지역 마을의 수)
 - 지역사회 참여: 모자보건 서비스 트레이닝을 받은 마을 보건요원 비율(1명 이상의 마을 보건요원이 모자보건 서비스 트레이닝을 받은 마을의 수/사업지의 지역 마을의 수)

- 지역사회 참여: 모자보건 활동에 참여한 마을 리더 비율(1년에 적어도 1번 이상 마을 리더가 모자보건 활동에 참여한 마을의 수/사업지의 지역 마을의 수)
- 지역사회 참여: 모자보건 프로그램에 참여한 마을의 비율(마을 사람들이 보건소 세부 계획(마이크로플랜) 수립에 참여한 마을의 수/사업지의 지역 마을의 수)

□ 평가결과

○ PDM 재구성에 따른 연차별 평가결과

〈표 4-11〉 KOFIH 라오스 통합모자보건 증진사업 평가결과

	내용	지표	1차 연도	2차 연도	3차 연도
영향	사업 지역 내 모성·영유아 사망률 감소를 통한 MDGs 4, 5번 달성	모성사망률(100,000명당): 260(2015년 목표)	-	-	357
		5세 이하 영·아동사망률(1,000명당): 55(2015년 목표)	-	-	73
		영아사망률(1,000명당): 45(2015년 목표)	-	-	68
산출	기획과 관리	자책 모자보건 마이크로 플랜이 있는 보건소 비율	100%	100%	100%
	정보 시스템	정보수집에 대한 훈련을 받은 마을 보건요원 비율	100%	100%	100%
	기술 훈련	주요 모자보건 의료 서비스에 대한 훈련을 받은 보건소 직원 비율	77%	82%	85.9%
	장비	모자보건 전략 실시를 위한 필수 장비를 갖춘 보건소 비율	100%	53%	63.9%
	지역사회 참여	지역사회 참여를 위한 주요 자료 보급률	100%	100%	100%
		모자보건 서비스 트레이닝을 받은 마을 보건요원 비율	100%	100%	100%
		모자보건 활동에 참여한 마을 리더 비율	100%	100%	100%
		모자보건 프로그램에 참여한 마을의 비율	100%	100%	100%

자료: 이명선 등(2014). pp.63-64.

○ 기초선 조사(Baseline survey)에 따른 평가결과

- 가임기 여성의 특성

- 학력에 있어서 캄보디아 지역은 중등교육 과정이 36.7%이고, 시엥코 지역은 미취학이 53.6%로 캄보디아 지역 여성의 학력이 더 높음.
- 산전관리를 받은 여성은 두 지역 모두 100%로 나타났고, 혈압·소변·혈액검사 중 혈압검사를 가장 많이 받은 것으로 나타남.
- 분만장소는 캄보디아 지역은 다른 사람 집 63.2%, 의사의 도움 68.4%로 나타났고, 시엥코 지역은 병원 55%, 간호사·조산사 41.2%, 의사의 도움 35.3% 순으로 나타나 의료기관에서의 출산은 사업 시작이 빠른 시엥코 지역에서 높게 나타남.
- 산후 수진은 산전 수진보다 낮게 나타남.
- 두 지역 모두 피임을 위한 방법으로 피임약·콘돔이 가장 많이 사용되고 있으며, 가족계획 방법에 대한 지식을 얻는 장소는 정부기관병원(90%)인 것으로 나타남.

- 바우처 이용률 및 보건프로그램 욕구

- 바우처 이용률은 사업 시작이 빠르고 학력 수준이 높은 시엥코의 여성이 60.4%, 캄보디아 여성이 20.0%로 나타남.
- 프로그램에 대한 요구도는, 모자보건 프로그램 증진에 대하여 77% 이상, 응급의료 후송체계 16.7% 순으로 나타남.
- 5세 이하 아동의 모자보건 서비스 이용
- 아동의 교육 프로그램 참여의 경우 캄보디아 지역은 71.4%인데 비해 시엥코 지역은 16.7%로 낮게 나타남.

- 아동의 질병관리에 있어 감기증상 여부, 질병으로 인한 진찰 및 치료 여부 및 질병치료를 위한 투약 여부에 있어서 대체로 씬느아 지역이 높게 나타남.
- 아동의 예방접종 카드 소지 여부는 사업 실시 기간이 긴 시엥코 지역이 다소 높게 나타난 반면, 국가 홍역, 국가 면역의 날, 지방 보건의 날 캠페인 참여 여부는 사업 실시 기간이 긴 시엥코 지역이 훨씬 낮게 나타남.
- 보건의료기관의 상태
 - 준비성 또는 가용성에 대한 조사 결과 대체로 씬느아 지역이 우수함.

○ 평가의 어려운 점

- 연구에 소요되는 시간 측면: 라오스의 행정 절차가 경직되고 복잡하여 연구 수행에 시간 소요가 필요함.
- 언어적 측면: 라오스는 라오어를 사용하기 때문에 영어를 함께 사용하는 사람이 많지 않아 의사소통에 불편함이 있는데, 특히 설문조사 시 언어적 문제 발생.
- 인력의 측면: 역량 있는 보건 연구자들의 수가 부족하여 연구를 함께 수행하는 데 있어서 어려움이 있음.
- 인프라 측면: 도로 및 교통이 불편하여 사업지역 방문 시 시간 소요
- 자료수집의 측면: 라오스 내의 성과지표 관련 자료(HMIS, VHV 자료 등) 수집이 어려움. 결과 지표인 모성사망률, 영유아 사망률에 대한 신뢰할 만한 자료가 존재하지 않아 자료를 획득하면서 정부에 보고되는 사례들이 발생하면서 사망자가

늘어나기도 함.

○ 평가 지표의 개선방안

- 향후 통합모자보건사업을 평가할 경우, 각기 다른 평가 틀을 통해 통합모자보건사업의 요소를 평가하기보다는 통합모자보건사업 평가 틀로 단일화하여 평가할 것을 제언함.
- 사전기획을 통해 자료수집의 어려움 극복.
- 라오스 국가의 행정 절차의 어려움과 언어소통의 불편함 등 수원국에 대해 충분히 이해하고 상황을 숙지할 수 있는 상황까지 고려하여 적절한 사업기획과 사전조사를 위한 시간이 충분히 확보될 필요가 있음.
- 이 사업과 같이 다년간, 몇 단계에 걸쳐 진행되는 경우 중간 및 사후평가결과를 반영하여 이후 사업의 성과제고로 이어질 수 있도록 환류체계를 더욱 강화할 필요가 있음.

제 5 장

한국의 개발협력사업 평가 개선방안

제1절 한국의 개발협력사업 평가체계 개선방안

제2절 보건복지분야 기획과 성과관리 개선방안

5

한국의 개발협력사업 평가 << 개선방안

제1절 한국의 개발협력사업 평가체계 개선방안

□ 독립성과 전문성을 갖춘 평가조직과 평가역량강화

- 원조사업의 효과성을 제고하기 위해서는 인력의 평가 전문성이 강화되어야 하고, 독립적·중립적·객관적인 평가가 이루어져야 함. 이를 위한 제도적 노력 또한 강구되어야 함(김현경 등, 2015, p.57).
 - (독립성 부족) 평가 담당 부서가 사업실시 부서와 독립되지 않은 경우도 있어 평가의 독립성 보장이 어려움.
 - (전문성 부족) 2-3년마다 순환근무를 함에 따라 담당 직원의 전문성이 축적되지 못함.
- 2008년 국제개발협력위원회 산하 평가소위원회를 구성하고, 메타평가를 실시하는 등 통합적인 평가체계 마련에 있어 성과가 있으나, 원조 시행기관의 평가예산과 평가역량의 변화는 크지 않고, 평가 독립성을 확보할 수 있는 체계나 조직을 갖추지 못한 상황임.
 - 2008년 OECD DAC이 특별 검토에서 한국 ODA 평가의 객관성과 전문성 확보를 위해 유무상 통합평가체제 구축과 독립적 평가 문화 확립을 권고한 후, 정부는 통합평가체제 구축을 결정하고 당해 8월 평가소위원회를 구성.
 - 제1차 분야별 기본계획(2011년~2015년)과 제2차 분야별 기본계획(2016~2020년)은 평가예산과 평가인력의 확대, 평가

역량강화를 주요 개선 과제로 제시하고 있으나, 개선 성과는 뚜렷하지 않은 실정임.

- ODA 예산은 지속적으로 확대되어 2010년 1조 3천억 원에서 2014년 2조 3천억 원으로 크게 증가하였지만, 해당 기간 중 원조시행기관의 평가예산과 인력의 변화는 크지 않은 형편임.

○ 영국의 경우 영국 ODA 주관 기관인 DFID의 자체평가 팀은 외부의 평가 및 컨설팅을 전문으로 기관에 용역을 주는 형태로 독립성과 전문성을 제고하는 한편, 2011년 원조영향평가위원회라는 독립기관을 설립하여 평가의 독립성을 강화하고자 함.

- 일본의 경우 외부 전문가로 하여금 JICA 사업의 사후평가를 진행하도록 하고 평가자문위원회에 제3자를 포함하여 평가결과의 객관성을 증진하고자 함.

○ 시행기관 평가부서와 사업부서를 분리하여 독립성을 꾀하고, 외부의 평가수행기관의 평가 전문성을 강화하는 한편, 실제 평가수행 팀에 ODA 평가 전문가 및 분야·주제별 전문가 참여를 확대³²⁾해야 함.

- EDCF의 경우 평가부서가 사업부서에 소속되어 있어 보고체계의 독립성이 부족함(권율 등, 2014).
- KOICA의 경우 JICA와 같이 독립적인 평가부서를 운용.
- 시행기관이 평가를 외부의 평가수행기관에 의뢰하여 사업주체와 평가주체를 분리하는 형태로 독립성을 제고하고자 하지만, 이와 같이 진행되는 평가 건수가 적을 뿐 아니라 평가수행기관의 전문성도 낮음.

32) DFID의 2003~2008년 에티오피아 국별 평가의 사례가 좋은 예.

- 평가 팀 구성에 분야·주제별 전문가, 평가 전문가의 참여가 미흡(권을 등, 2014).
- 설문조사 응답자의 68%가량이 해당 평가 이전에 수행한 ODA 평가 경험이 없거나, 1~2건으로 전문성이 부족함.
- 현지 전문가 참여 강화, 현지 팀의 평가 전문성 제고.
- DFID는 현지 팀과 협력하면서 평가과정을 지원함.

○ 평가역량을 강화하기 위한 프로그램 내실화가 요구됨.

- 정부는 ODA 담당자의 평가 관련 이해도를 높이고 시행기관
의 자체평가 수행 능력을 향상하기 위해서 ‘ODA 평가교육 프
로그램’을 마련해서 실시하고 있음.

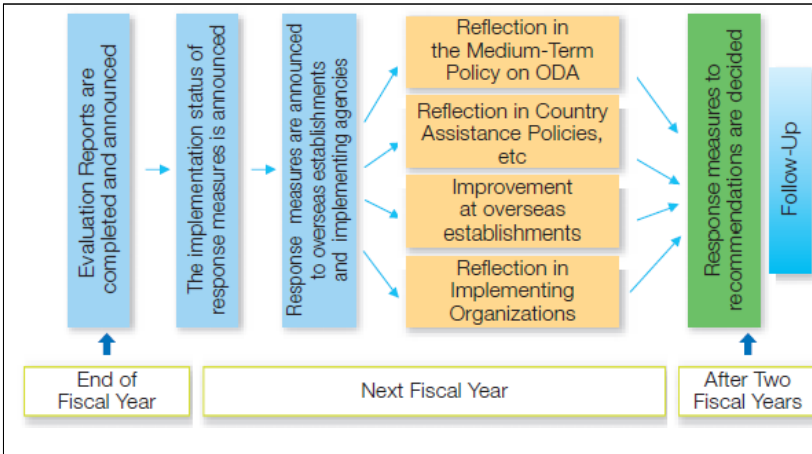
□ 평가의 환류시스템 제도화와 평가결과 공개

○ 평가결과의 환류(Feedback)는 「국제개발협력 선진화 방안」과제 1, 2차 「분야별 기본계획」에서 ODA의 주요 개선 과제로 언급, 이에 따라 정부는 평가결과의 환류 강화를 골자로 하는 「국제개발협력 통합평가지침」을 개정(14년 1월 13일)하여, 각 시행기관으로 하여금 평가결과 반영계획을 수립하고 반기별로 그 이행상황을 점검하게 하고 있음.

- 정부는 이행상황에 대한 점검을 통해, 평가결과로 도출되는 대부분의 정책제언이 추상적이어 차기 사업에 적용하기가 어려울 뿐 아니라 연구용역이 지연되어 환류계획 수립이 늦어진 경우 등이 발견되고 있다고 지적한 바 있음.
- 또한 평가결과가 예산에 반영되지 않아 실효성있는 환류가 되지 못함.
- KOICA의 평가결과는 자체 웹사이트와 OECD DReC에 공

개되고 있으나, EDCF의 경우 보고서가 잘 공개되지 않는 경향이 있음. 이는 평가부서의 독립성 미비와 연관되어 있음.

[그림 5-1] JICA의 평가 환류체계



자료: KOICA(2013b). p.196.

○ 영국과 일본의 다양한 환류 기제를 참고

- 영국은 독립 평가자가 제시한 평가결과 및 제언에 대해 DFID와 같은 시행기관의 답변과 개선 노력을 명시토록 하고 평가결과를 각 사업의 예·결정에도 반영. 이러한 경쟁적인 성과기반 자원배분 방식이 개발협력사업에 있어서 비용대비효과(VfM)를 제고할 수 있음.
- 일본은 PDCA cycle의 마지막 단계에서 평가결과를 실제 사업 및 향후 사업계획에 반영하는 피드백(Action) 단계를 두어, 평가결과가 사이클 처음의 Plan과 Do 단계에 반영될 수 있도록 함.

□ 평가의 강조

- 마지막으로 모니터링과 성과관리의 중요성에 대한 인식이 높아지고 있으나, 여전히 ‘사전조사-모니터링-종료평가-환류’로 이어지는 구조가 원활하지 않음.
 - 이는 원조예산 대비 낮은 평가예산과 평가인력 부족으로 이어짐.
 - 또한 사전 타당성 조사 및 사전 형성 조사라는 핵심적인 사업 기획 단계의 부재를 통해 전체적인 사업 부실로 이어질 수 있음.
 - 2010년 OECD DAC 가입을 계기로 평가체계에 대한 논의를 이어 오고 있는바, 내실 있는 평가시스템 확립이 절실함.

제2절 보건복지분야 기획과 성과관리 개선방안

□ 평가의 종합적 기획과 다변화

- 현재 모니터링과 개별 프로젝트의 종료평가에 머물러 있는 경향이 있는데, 평가를 다양한 국면에서 시행하여 그 결과를 종합적으로 파악하고, 사업을 기획함으로써 원조효과성 제고에 기여할 수 있음.
 - 수원국의 보건 및 사회정책의 목표와 주민의 요구(needs)분석 결과를 반영한 사업, 평가결과를 활용한 근거 중심의 사업이 기획되어 ‘적절’한 사업이 기획될 수 있도록 사업기획 단계를 강화할 필요가 있음.
 - 기초설계연구(Basic Design Study)와 기초선 조사(Baseline survey)에 충분한 시간과 인력 등의 자원 투입 요구

○ 사업기획 단계의 사업형성 및 사전 타당성 조사를 강화하는 것이
필요함(김현경 등, 2015, pp.56-57).

- 평가의 중요성에 대한 인식 제고를 통해 사업형성 및 사전조사
에 투입되는 예산 및 시간을 확대함으로써 사업형성의 질을 제
고해야 함.
- 특히 사전 타당성 조사의 객관성과 엄밀성을 제고하는 노력이
필요한데, 이를 위해서는 유능한 전문가와 컨설팅업체의 활용
이 중요함.
- 사업의 사전기획은 하나의 절차가 아니라 사업의 전체 효과를
좌우한다는 인식의 전환이 필요함.
- 사업기획 단계의 강화는 수원국의 주체성을 발휘하여, 수원국
이 필요로 하는 ‘적절’한 사업을 기획할 수 있는 가능성을 높임.

○ 중간평가 내실화 강화

- 한국 개발협력사업의 단기성으로 인해 중간평가가 형식적으
로 이뤄지는 경향이 있음.
- 중간평가는 실제 사업 추진과정에서 장애요소가 되는 부분, 사
전기획 조사에서 충분히 고려하지 못한 부분 등을 수원국과 공
여국이 논의하여, 최종 지향하는 성과들이 생산되도록 하는 브
리지적인 성격이어야 함. 실제로는 착수 시의 계획과 거의 다
르지 않은 내용으로 중간사업계획을 세우는 경우가 많음.

〈표 5-1〉 JICA의 PDCA 평가체계

PCDA Cycle	Plan	Do		Check		Action
평가 종류	사전평가 (Ex-ante Evaluation)	중기검토 (Mid-term Review)	종료평가 (Terminal Evaluation)	사후평가 (Ex-post Evaluation)	사후 모니터링 (Ex-post Monitoring)	피드백 (Feed-back)
지원 양식	Technical Cooperation Loans Grants	Technical Cooperation		Technical Cooperation Loans Grants	Loans	평가 결과의 반영 및 유사 프로젝트 의 기획과 실행에 대한 제언 제공
시기	프로젝트 실행 전	프로젝트 실행의 중간단계	프로젝트 완료 6개월 전부터	프로젝트 완료 3년 후부터	프로젝트 완료 7년 후부터	
타겟	모든 프로젝트	4년 이상 진행되는 프로젝트	모든 프로젝트	예산 2,000만 엔 이상의 프로젝트	사후평가가 효과성과 지속가능성 을 담보할 시	
평가 원칙	내부평가	내부평가 (수원국과의 합동평가)		외부/내부	-	
평가 방법	DAC 5대 평가기준에 기반한 니즈(needs), 기대되는 결과물 (expected outcomes), 프로젝트 계획에 대한 확인	DAC 5대 평가기준에 기반한 프로젝트 결과물 평가	DAC 5대 평가기준에 기반한 프로젝트 결과물의 종합적인 달성도 평가	DAC 5대 평가기준에 기반한 평가	-	

(출처 : JICA, 2013 : 50-51) (재구성)

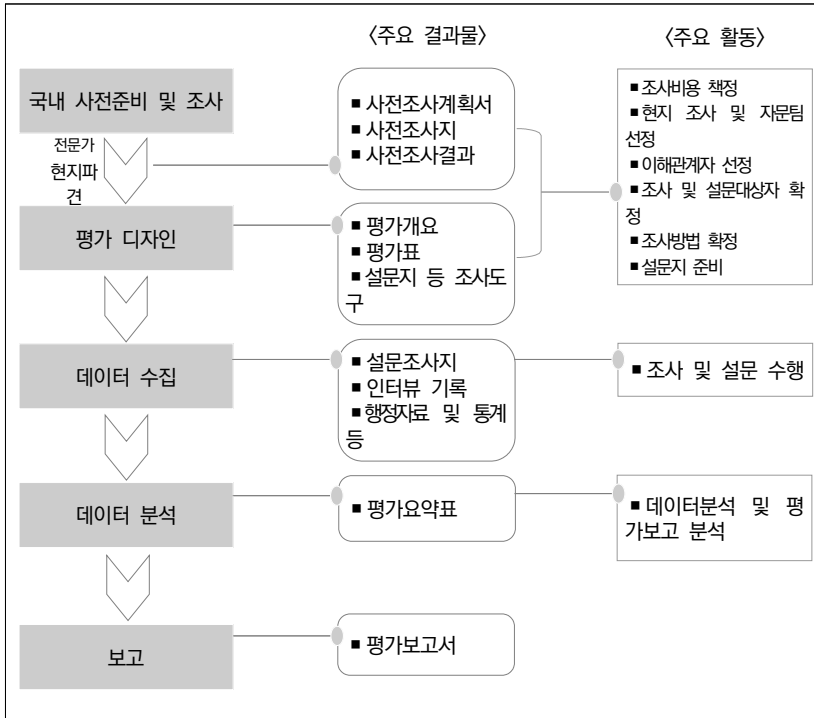
자료: KOICA(2013b). p.193.

○ 일본과 같이 사전평가의 체계화와 지원이 요구됨.

- 일본은 PDCA 사이클을 따라 사전평가, 중간평가를 다음과 같은 체계로 실행함.
- 한국은 사후평가를 위주로 하여 아직은 사전기획 및 중간평가에 충분한 시간과 재원을 투입하지 못하고 있으나 점차 확대되는 경향.
- 다음과 같은 평가절차와 각 절차에서 도출되어야 하는 결과물

과 활동을 확인하는 것이 바람직함.

[그림 5-2] 평가절차



자료: 국제개발협력위원회(2014), pp.9-25; 김현경 등(2015), p.164 재인용.

- 국별·분야별·프로그램별 평가로의 다변화를 통한 평가의 질적 향상
 - 한국의 평가소위원회가 실시한 ‘분야별 통합평가’는 5건, EDCF는 5건, KOICA는 11건이며, 상기 기관 이외의 ODA 시행기관들은 프로젝트평가에 치중하고 있어서 국별 평가나 분야별 평가 실적이 없음.
 - 영국은 평가의 대상과 단위가 변화하는 추세를 반영하여, 과거

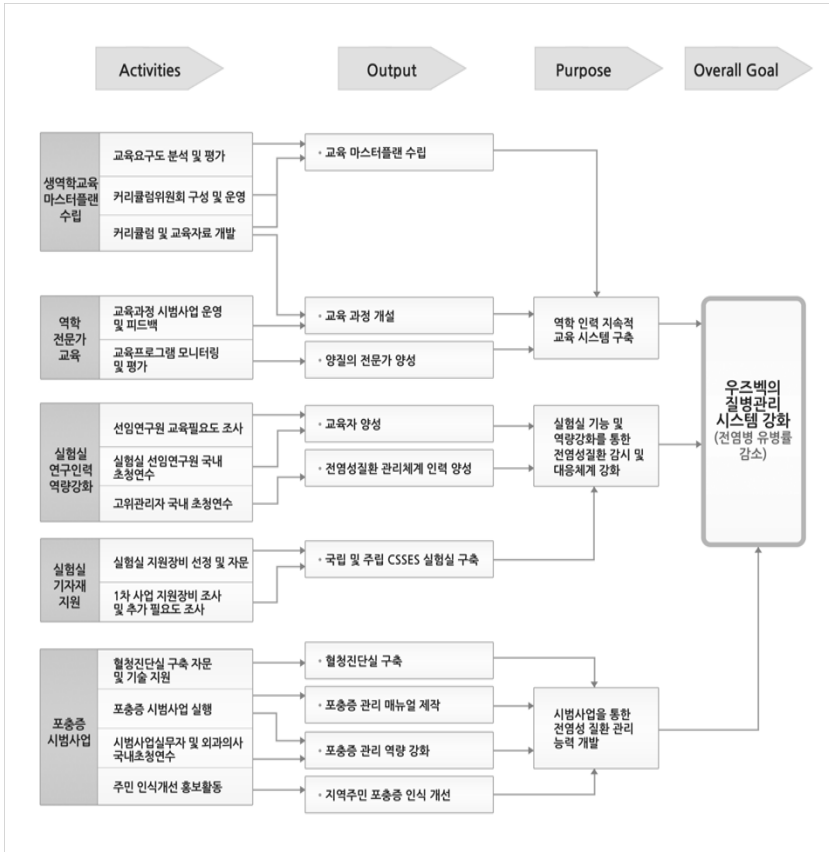
에는 프로젝트평가를 통한 단기적 성과에 대한 평가가 주류를 이루었다면, 최근에는 주제·부문별 평가, 정책·전략 평가, 프로그램 평가, 국별 평가를 통한 중장기적 개발성과와 전략적 방향성 파악에 중점을 두는 추세임.

- WHO에서 하는 평가는 프로그램 평가에서부터 주제 평가, 특정 사무소에 대한 평가까지 다양하나, 최근까지 WHO는 보건 MDG를 중심으로 각 국가의 보건 관련 MDG 이행 정도와 달성 정도를 파악하는 것에 초점을 두었음. 이는 보건 사업의 성과, 영향(impact)을 평가함으로써 각국의 원조수요와 공여 흐름과의 일치성 정도, 국별 보건 정책과의 조화성 등을 종합적으로 평가하고 개선방향을 찾을 수 있다는 강점이 있음.
- 또한 WHO의 보건분야 평가 틀은 보건분야의 이니셔티브를 모니터링하고 평가하기 위한 것으로 재정, 서비스 전달, 의약품, 보건 통계 및 거버넌스 등의 진전 사항을 점검할 수 있는 보건체계 지표들을 포함하므로 여타 분야와는 다른 분야별 특성을 고려할 수 있을 것으로 기대됨.
- JICA의 몽골 사회분야 지원 프로그램에 대한 평가는 프로그램 정책 매트릭스에서 알 수 있는 바와 같이 복지, 보건, 교육, 도시를 아우르는 범분야적 특성이 있고, 특히 최근 사회안전망(Social Protection) 구축과 같은 포괄적인 프로그램은 정책 평가나 프로그램 평가와 같은 평가 방식을 필요로 할 것임.
- 세계은행의 베트남 사회안전망 강화 프로젝트(김현경 등, 2015, pp.122-124)나 국제기구(ILO, UNICEF 등)의 캄보디아 사회안전망 구축 프로젝트 등이 이러한 사례에 해당함.

□ 논리모형 접근법 강화

- 메타평가 결과 사업의 논리적 연관성을 설명하는 PDM의 논리적 상관관계가 취약하여 성과측정의 실효성이 부족함(권율 등, 2014).
 - 전반적으로 평가대상사업 PDM의 구성요소 간 논리적 상관관계가 부족함.
 - 사업초기단계에 성과관리 미흡으로 체계적이고 활용 가능한 PDM이 제대로 구축되지 못해 자체평가의 성과측정과 평가 품질을 저해하는 요인으로 작용함.
 - 특히 프로젝트 사후평가는 사업 상위목표로 연계되는 폭넓은 연결고리에 대한 다면적인 검토가 필요하나 개별사업에 한정하여 평가대상 및 범위에 대한 분석 미흡함.
 - PDM상 최종 사업목표가 지나치게 추상적이거나 광범위하고 논리적 연결고리가 부족한 경우가 많음.
- 따라서 시행기관은 PDM 수립을 위한 사전조사를 내실화하고, 기관별 상황에 따라 사전평가제도 도입의 가능성 및 적절성 검토 필요
 - 프로젝트 단위 평가에서 벗어나 통합적, 전략적 평가를 점진적으로 확대하는 방안에 대한 고려도 요구됨.
 - 평가수행기관은 평가대상 및 범위에 대한 종합적 분석을 강화하여 사업개선을 위한 교훈 및 제언 도출에 중점을 두어야 함.
- 영국과 일본 사례의 PDM 참조
 - 다음은 우즈베키스탄 전염성 질환 관리 역량강화 사업의 논리적 구조도와 PDM 사례.

[그림 5-3] 사례: 우즈베키스탄 전염성 질환 관리 역량강화 사업의 논리적 구조도



자료: '국제개발협력사업 평가'(강성욱) 발표자료.

[그림 5-4] 사례: 우즈베키스탄 전염성 질환 관리 역량강화 사업의 PDM

PDM			
Project Name: 우즈베키스탄 전염성 질환 관리 역량강화 사업		Duration: 45개월(2013~2017)	Ver. No: 1.1
프로젝트 요약 (Narrative Summary)	지표 (Performance Indicators)	자료 입출수단 (Means of Verification)	가정 및 위험요소 (Important Assumptions)
Overall Goal 식중독·수역·매개 전염성 질환 관리의 효율화 및 현대화를 위해 현 우즈베키스탄 내 질병관리 시스템을 강화	RCI에 합의된 식중·수역·매개 미생물 및 기생충 주요 전염병 유행률	보건부 통계자료 Demographic Health Survey	1. 우즈베키스탄 보건부의 정책일관성이 유지된다 2. TIMPE의 지속적인 운영이 보장된다
Project Purpose 1. 역학인력 지속적 교육 시스템 구축 2. 실험실 기능 및 역량강화를 통한 전염성 질환 감시 및 대응체계 강화 3. 시범사업들을 통한 전염성 질환 관리 능력 개발	1. 보건부 필수 보수교육 과정으로의 선정여부 2-1 검체분석 소요 시간 2-2 감시보고의 적시성 및 일관성 3-1 포충증 유행률	1. 보건부 공식서신 2. CSSES 실험실 통계 연감 3. Baseline 및 Endline Survey	1. TIMPE의 지속적 운영이 보장된다 2. 전염성질환 관리를 위한 행정·법률·의학적·적극적인 개선의지가 지속된다 3. 사업지역에 내전이나 전쟁으로 인한 사회불안이 일어나지 않는다
Outputs 1-1 역학 및 위생 전문가 교육을 위한 마스터플랜 수립 1-2 교육과정 개설 1-3 양질의 역학 및 위생 전문가 양성 2-1 국립 및 주립 CSSES 실험실 구축 2-2 교육자 (Master Trainer) 양성 2-3 보건부 및 CSSES 전염성 질환 관리체계 인력 양성 3-1 실험실 진단실 구축 3-2 포충증 관리 매뉴얼 제작 3-3 포충증 관리 역량강화 3-4 지역주민 포충증 인식개선	1-1 역학 및 위생 전문가 교육 마스터플랜 수립 여부 1-2 역학 마스터플랜의 지점에 따라 수행된 연간 교육과정 실시 횟수 1-3 연간 역학 및 위생 전문가 교육과정 수료자 수 1-4 교육생 중 최종 수행평가 80점 이상인 비율 2-1 기지재 설치 여부, 필수 진단 항목 완비 여부 2-2 미생물 및 기생충 실험실 선임연구원 최종 수행평가 80점 이상인 비율 2-3 실험실 고위관리자 최종 수행평가 80점 이상인 비율 3-1 사마르칸트 이사에프 기생충학연구소 및 4개주 실험실 진단실 구축 여부 3-2 기지재 설치 여부, 필수 진단 항목 완비 여부 3-3 포충증 관리 매뉴얼 제작 여부 3-4 포충증 치료환자 수 3-5 조기 포충증 발병률 3-6 포충증 양성률 비율 3-7 포충증 관련 지역사회 인식도	1. 사업결과보고서 2. TIMPE 통계연감 1. 기지재 지원 결과보고서 2. post-training tests 1. 기지재 지원 결과보고서 2. 사업결과보고서 3. 포충증 관리 시범사업 결과 보고서	1. 역학전문가 교육자료가 현실 적용 가능하게 마련된다 2. 우즈베키스탄 보건부가 TIMPE 교육 과정을 제도화한다 3. 교육받은 연수생이 본래의 위치에 서 근무한다 4. 포충증관리 매뉴얼이 우즈베키스탄 정부로부터 채택된다
Activities <마스터플랜 수립> 1-1 앞선 전염병 유행 및 위기관리 업무 능력 분석 1-2 역학 및 위생전문가 교육요구도 분석 및 평가 1-3 커리큘럼위원회 구성 및 운영 (워크숍 및 세미나 개최) 1-4 역학 전문가 교육커리큘럼 및 교육 자료 개발	Inputs [KOICA] 1. 총예산 : \$ 약 600만 2. 인력	External Factors 1. 역량있는 기지재 업체가 입찰에 참여한다 2. 주요 이해관계자들이 사업실현에 적극 참여한다 3. 시범사업 주민들이 적절하게 협력한다	
<역학교육> 1-5 역학 전문가 교육과정 시범사업 운영 및 피드백 1-6 우즈베키스탄 측 교육 프로그램 운영을 위한 현지강사진 확보 (ToT 전문가 섭외) 1-7 현지 CSSES 역학 및 위생전문가 교육프로그램 운영 1-8 수료자 활동보고서 대회 중 2회 실시 1-9 교육프로그램 모니터링 및 평가 2-1 국립 CSSES 및 주립 CSSES 실험실 지원 장비 및 선임연구원 교육 필요도 조사 2-2 1차 사업 지원 장비 조사 및 추가 필요도 조사 2-3 국립 CSSES, 주립 CSSES 실험실 구축을 위한 지원 장비 리스트 확정 및 기지재 제안요청서 지원 2-4 초청연수 커리큘럼 및 교재 개발 2-5 미생물 (4명) 및 기생충 (4명) 실험실 선임 연구원 국내 초청연수 2회 2-6 미생물 (4명) 및 기생충 (4명) 실험실 연수과정 수료자에 의한 동물 교육과정 수행 모니터링 및 현지 교육과정 지원 2-7 보건부 및 CSSES 고위관리자 6명 국내초청연수 실시 (2명) 2-8 사업대상지에 대한 현지업체를 통한 안정적인 유지보수 지원 3-1 사마르칸트 이사에프 기생충학연구소 실험실진단실 구축 지원 및 기술지원 3-2 4개주 CSSES 실험실진단실 구축 지원 및 기술지원 3-3 시범사업관리위원회 구성 및 운영 3-4 시범사업 대상 마을 선정 3-5 시범사업마을 Baseline Survey 수행 3-6 시범사업 실무자(포충증) 10명 국내 2주간 초청연수 3-7 포충증 관리 워크숍 4회 국내 4주간 초청연수 3-8 시범사업 실무자(포충증) 10명 역학조사팀 구성 및 시범사업 운영 기술지원 (모니터링 및 평가) - 주민 실험조사 및 초음파검사 - 개 구충 사업 - 양 포충 감열을 조사 3-9 지역주민 포충증 인식개선 홍보 활동 3-10 시범사업마을 Endline Survey 수행	- PM 1명 - 분야별 전문가 - 추가 실무자 파견 3. 기간: 계약 후 45개월 [우즈베키스탄] 1. 사업관리 인력 지원 2. 사무실 및 관련인력 제공, 기타 행정지원 등	4. 시범사업을 방해하는 외부적 사건이 일어나지 않는다	
		Precondition 1. 사업수행계획에 따라 분야별 전문가가 적시에 파견된다 2. 우즈베키스탄 보건부와 CSSES 등 거버넌스가 작동된다	

자료: '국제개발협력사업 평가'(강성욱) 발표자료.

□ 평가의 객관성과 엄밀성을 높이기 위해 개발협력목표(MDGs, SDGs)의 보건 및 사회정책 관련 지표를 사용해 **정량적 평가 확대**.

○ 데이터 객관성 부족의 문제가 있음(권율 등, 2014).

- 정량적 목표치, 측정지표가 필요한 효과성 분석에서 대부분 정성적 접근과 제한적 수준의 설문조사에 의존(권율 등, 2014).
- 2가지 모자보건 평가사례 역시 동일한 한계점을 가지고 있음.

○ WHO는 각국의 보건사업과 정책을 모니터링하고 평가하기 위해 보건 관련 MDG 주요 지표 사용. 최근 보건 관련 SDG가 제시됨에 따라 큰 목표는 구성되었지만, 이를 측정하고 모니터링하고 보건사업을 평가할 수 있는 지표들은 구성되고 있는 중임.

- 특히 모자보건과 관련하여 MDG4(아동 사망 감소), MDG5(모성 건강 향상)에 따라 5세 미만 아동의 사망률, 아동의 홍역 백신 coverage, 모성사망비, 숙련된 보건인력이 참여한 출산 경험율, 산전 관리율, 피임 실천율 등이 지표로 선정되어 목표치가 제시되었으며, 매년 MDG 달성 정도가 모니터링되고 평가되었음.
- 향후 구성 중인 SDG 지표(부록)에 따른 정량적이고 객관적인 평가가 요구됨.

○ 한국의 모자보건사업 평가사례의 한계로 각각 지적된 정량적 지표를 이용한 객관적 평가의 한계는 WHO의 모자보건 관련 지표, KOICA 보건의료 지원 전략 2011~2015 전략 성과 핵심 지표 가운데 관련 지표를 이용해 기초선조사, 종료후 조사가 엄밀하게 이루어질 때 극복할 수 있음.

- 이는 사업형성 및 타당성 조사 강화의 필요성으로도 이어짐.

〈표 5-2〉 DFID 생식 및 모자보건 프로그램 성과 지표

구분	내용
핵심 성과 지표	• 임신 중과 출산 중 구조된 여성과 신생아 수(모성사망률과 신생아사망률에서 모델 추출) • 최신 가족계획 수단을 사용하는 여성의 수(피임보급률 통해 계산) • 원치않은 임신의 방지 건수 • 숙련된 조산사를 통한 분만 수와 비율
강력 추천 지표	• 임신 중 산전 진찰을 최소 4번 이상 받은 여성의 비율 • 인구 10,000명당 의료 인력 수 • 인구 500,000명당 응급 산과 및 신생아 진료가 가능한 의료 시설 수
*추천 지표	• 소녀 및 청소년기 여성 대비 정상 출산 수와 비율 • 15-49세의 임신하지 않은 여성 중 체질량지수가 정상 범위 밖인 여성의 비율 • 제왕절개를 통한 정산 분만 비율 • 6개월 동안 모유 수유한 신생아 비율
*개발 중 지표	• 18세 이하 여성의 첫 번째 출산 비율 • 권한강화(Empowerment)를 위한 지표 • 최소 1회의 산후 진찰을 받은 여성과 신생아 비율 • 만족스러운 서비스 경험을 보고한 여성의 비율

(*주 : 상기의 표에 기재된 추천 지표와 개발 중 지표는 전체 지표 중 일부를 발췌한 것임.)
(출처 : DFID, 2010b : 48)

자료: KOICA(2013b). 보건의료분야 ODA 사업 성과관리 가이드라인 연구결과 보고서. p.75.

○ 최근 WHO는 SDG 이행 및 달성 정도를 파악하기 위하여 Global Health Observatory를 설치하여 194개 회원국의 보건 관련 1000여 개의 지표들에 대한 데이터를 구성하고 공개하고 있으므로, 보건분야 ODA 사업평가의 엄밀성과 객관성 제고를 위해 이와 같은 보건 통계를 활용할 수 있을 것으로 기대함.

□ 영향평가 확산을 통한 장기적 영향 또는 파급효과에 대한 평가 강화.

○ 선택편의를 줄이는 다양한 방식 가운데 무작위통제연구를 활용한 영향평가 확대.

- WHO는 전체적인 자금 흐름 등을 추적하는 국제개발협력 결

과 보고서를 내는 것과 더불어 <참고>와 같이 개별 보건 프로그램에 대해서 무작위통제방식을 통해 정책의 성과를 평가함.

<참고> WHO의 우유의 불소효과(milk fluoridation) 프로그램의 임상 유효성 측정 (Petersen et al., 2015, Long term evaluation of the clinical effectiveness of community milk fluoridation in Bulgaria, Community Dental Health 32, pp. 199-203.)

- 기초연구설계: 그룹 간 final 횡단면 비교를 통한 parallel arm 5년 코호트 조사
- 참가자: 8곳의 불가리아의 도시(cities/towns)에 살고 있는 3살 아동 - 2004년을 기점으로 모집된 표본(1,782명)을 랜덤 샘플링하여 추출. 5년 후인 2009년 8살이 된 하위표본(약 30%)을 무작위로 선출하여 454명을 대상으로 후속 실험 처치를 하였고, 276명은 통제집단에 배치. 횡단면 비교 조사를 위해, 2004년 통제 도시에서 살면서 충치가 있는 284명의 3살 아동이었으며, 2009년 8살이 된 276명을 대상으로 진행
- Treatment: 6개의 처치 집단들에서 매일 학교에서 0.5mg 불소가 첨가된 100ml 또는 200ml의 우유나 요구르트를 제공받은 1,498명의 아이들을 대상으로 실시하였으며, 추가로 180명은 불소가 첨가되지 않은 우유를 제공받음. 2개의 통제 집단들에서 284명은 불소가 첨가되지 않은 우유를 제공받음.
- 주요결과측정: 유치와 영구치에서 충치 경험
- 결과: 유치의 경우, 불소가 첨가된 우유를 마신 그룹이 불소가 첨가되지 않은 우유를 마신 그룹에 비해 충치증가가 낮음. 각각 충치 감소는 46%($p<0.001$) 30%($p<0.01$)로 나타남. 영구치의 경우 충치 감소가 61%와 53%($p<0.001$). 8살 아동의 횡단면 조사는 시간이 지나면서 충치경험에 있어서 중요한 변화를 보여주는데, 통제집단의 경우 DMFS가 11% 감소된 데 반해 불소가 첨가된 우유를 마신 아동의 경우 43%가 감소(2004년과 2009년). 불소가 첨가된 우유를 마신 아동들은 DMFS는 68% 감소하였으나 통제집단은 3% 증가.
- 결론: 불가리아에서 불소가 첨가되지 않은 우유를 마신 아동의 경우와 비교하였을 때, 매일 배달되는 불소가 첨가된 우유가 상당히 충치진행을 낮추었다. 우유의 불소효과를 알아본 전국적 실험은 공중보건 제도가 아이들의 충치에 효과적으로 대응할 수 있다는 점을 시사.

- 임의추출, 통제집단 설정의 어려움, 대상의 장기적 추적의 어려움, 평가예산 미비 등으로 인해 이러한 평가를 하기 어려운 측면이 있으나 보다 엄밀하게 사업의 성과를 평가할 수 있는 방법이므로 확대 및 강화할 것을 제안.

- 보건분야에서 통제집단을 필요한 의료 행위에서 배제하는 듯한 설계가 있어 윤리적인 측면의 문제 제기가 있으나, 순위를 정하는 방식 등으로 이와 같은 문제를 피할 수 있는 다양한 평가 방식 설계가 가능할 것으로 보임.

○ 다른 평가지표보다 장기적인 관점에서 사업의 최종적인 목표 달성을 평가한다는 측면에서 원조의 효과성 제고라는 성과관리의 목표에 가장 부합하는 평가 잣대임.

□ 수원국의 문화 및 젠더 규범을 반영한 보건의료적 접근³³⁾과 이를 반영한 평가

○ 모자보건분야는 단지 병원 건립이나 의료인력 파견 및 현지 전문 의료인 양성 등으로 단기간의 눈에 보이는 일정 수준(모성사망률 감소, 영아사망률 감소 등)의 성과를 얻을 수 있는 분야 중 하나. 그만큼 의료자원의 부재와 부족에 의한 모자보건의 취약성은 수원국이 지닌 명확한 일차적 문제점이기 때문임.

○ 하지만 모자보건과 관련된 의료서비스 이용 행태는 특히나 해당 국가 및 지역사회의 문화와 젠더 규범에 많은 영향을 받음. 모성 사망률, 영아사망률, 인공임신중절을 및 피임 실천율 등은 병원 건립이나 의료인력의 전문화 및 의료서비스 제공 등에 의해 획기적 성과를 일정 기간, 일정 수준까지 올릴 수는 있지만, (대다수의 국가들에서) 그 이상의 성과로 이어지지 않는 것은 다른 부분에서 해결되지 않은 문제들이 있기 때문임.

○ 저개발국가일수록 여성의 학력과 소득 수준 등 사회경제적 지위

33) 김동식(한국여성정책연구원)의 자문을 토대로 함.

가 낮고, 사회문화적으로 여성으로서 여성다운 행위를 요구하는 젠더 역할과 젠더 규범이 강하기 때문에 부부 및 연인 관계에서 성별 간의 권력(의사결정)구조는 여성에게 상대적으로 취약함. 따라서 모자보건분야를 비롯한 보건의료적 개입이 필요한 분야의 경우, 수원국의 문화적·젠더적 접근이 병행되어야 함. 그렇지 않으면 일정 수준 이상의 성과(효과성)를 얻을 수 없으며, 결국 지속 가능하지 않고, 효율성 달성도 낮을 수밖에 없음.

○ 이러한 접근에 대한 강조는 문화적 및 젠더적 접근이 모자보건분야에 중요하다는 점도 있지만, 이보다는 수원국이 직면한 여러 사정 및 상황을 사전에 진단하고, 그에 맞는 사업 설계와 추진 방법의 적절성 평가, 그에 맞는 맞춤형 성과 및 단계별로 지속 가능성 설계가 이루어져야 한다는 포괄적인 의미도 있음. 상당수의 수원국들은 이들의 사회적, 경제적, 문화적 여건에 따라 더블 사이클이나 혹은 또 다른 형태의 접근이 필요함.

○ 한국의 모자보건사업 평가에서도 4장의 사례와 같이 이를 반영한 성과지표를 포함하는 경향이 늘어나고 있음.

□ 보건복지분야 사업기획과 평가에 있어서 다양한 분야의 전문가로 구성된 **다학제적 접근의 필요성**³⁴⁾

○ 보건의료 관련 ODA 사업의 경우, 그동안 전문가 영역으로 의료인들이 대다수 수행하였음. 그러나 현장에서 의료시스템이 들어갈 때는 실제 의료보다는 현장에 대한 이해가 우선되어야 함. 이를테면, 수원국의 문화적 측면, 젠더적 규범 및 환경적 요소가 해

34) 김동식(한국여성정책연구원)의 자문을 토대로 함.

당 보건의료에 대한 수용성이 높은지, 아니면 그 원인을 파악하여, 그에 맞게 보건의료적 접근방안을 달리 모색하거나, 수원국 및 현장 관계자·전문가와 전반적인 부분에 대해 논의하여 대체방안을 마련하는 것이 필요함. 이를 위해서는 의료인에 의한 의료적 논의와 방안 마련은 한계가 분명 있음.

- 사업 착수 이전에는 수원국의 해당 사업지역에 대한 이해도가 있는 문화인류학자가, 그리고 보건의료적 개입에서의 계층 간, 성별 간, 연령 간 차별이 존재한다면 불평등 학자 등이 의료인과 함께 논의 테이블에 있을 필요가 있음.
- 실제 모자보건 ODA 사업을 추진한 두 국내기관의 사업내용에서 한 사업은 의료인이 주로 투입된 경우였고, 다른 하나는 다양한 분야의 전문가가 투입된 경우였는데, 실제 모자보건사업을 통한 임신부 및 해당 지역사회의 모자보건 역량강화는 후자의 전문가 구성이 더 월등하였음.
- 실제 보건분야 사업은 해당 사업과 연관된 사회정책적 배경에 대한 검토와 지원대상에 대한 사회적 이해, 복지 전달체계 구축 등과 함께 이루어질 필요가 있음.

○ 다양한 분야의 다학제적 접근은 수원국의 니즈(needs)에 대한 올바른 이해와 사업과정에서의 예측하지 못한 문제 파악 및 개선, 그리고 이를 통한 내실 있는 성과물 생산을 위해 모든 사업의 기본이 되어야 할 것임.

참고문헌 <<

- 국제개발협력위원회. (2015a). 제18차 국제개발협력 평가소위원회 보고안건(제18-2호), 국제개발협력 통합평가제도 개선방안.
- 국제개발협력위원회. (2015b). 제17차 평가소위원회, 2015년 통합평가계획(안) 의결안건 (제17-2호).
- 국제개발협력위원회. (2014a). 제16차 평가소위원회 의결안건(제16-1호), ODA 시행기관의 자체평가에 대한 메타평가 결과(안).
- 국제개발협력위원회. (2014b). 제15차 평가소위원회, ODA 평가 시범교육과정 실시 결과(보고안건 제15-2호).
- 국제개발협력위원회. (2013a). 제13차 평가소위원회 의결안건(제13-4호), 국제개발협력 평가관련 규정 제·개정(안).
- 국제개발협력위원회. (2013b). 제15차 국제개발협력위원회 의결안건(제15-1호), OECD/DAC 개발협력 정책·집행 평가 결과 및 활용방안(안).
- 국제개발협력위원회. (2012a). 제11차 평가소위원회 의결안건(제11-1호), ODA 시행기관의 자체평가에 대한 메타평가 결과.
- 국제개발협력위원회. (2012b). 제9차 평가소위원회 의결안건(제9-1호), 2011년 ODA 자체평가 결과.
- 국제개발협력위원회. (2011). 제8차 평가소위원회 의결안건(제8-2호), 국내 주요기관의 공적개발원조 체제 평가결과.
- 국제개발협력위원회. (2010a). 제7차 국제개발협력위원회 의결안건(제7-5호), 국제개발협력 통합평가 추진방안.
- 국제개발협력위원회. (2010b). 제7차 국제개발협력위원회 의결안건(제7-1호), 국제개발협력 선진화 방안.
- 국제개발협력위원회. (2009). 제5차 국제개발협력위원회(안건 2), ODA 선진화 추진계획(안).
- 권율 등. (2014). ODA 시행기관의 자체평가에 대한 메타평가. (2014.12.2.) 대외경제정책연구원.

- 김지현. (2016a). SDGs지표 체계와 모니터링: 불평등성 완화 논의를 중심으로. 국제개발협력지. KOICA, 2016. No. 2.
- 김지현. (2016b), SDGs 지표 확정과 의의, 제31회 국제개발협력포럼, 2016. 8,29.
- 김현경 등. (2015). 보건복지분야 ODA 활성화 방안 연구. 한국보건사회연구원.
- 대외경제협력기금, 한국수출입은행. (2012). EDCF 공동평가 지침서.
- 대외경제협력기금, 한국수출입은행. (2011a). 사후평가보고서 작성 가이드라인.
- 대외경제협력기금, 한국수출입은행. (2011b). EDCF 평가매뉴얼.
- 오주환 등. (2013). 보건 분야 ODA 종합평가 연구. 서울대학교병원.
- 이명선 등. (2014). 라오스 통합모자보건 증진사업 1차 종료평가 보고서. 한국국제보건의료재단 발간자료.
- 이호철, 김도형, 김부열, 남은우. (2016). M-Survey를 이용한 모자보건 설문조사 효과성 분석. 국제지역연구, 25(1), pp.171-198.
- 정지선, 오태현. (2013). ODA 국별 성과관리체계 및 평가방법에 관한 연구. 대외경제정책연구원
- 한국국제협력단. (KOICA), OECD. (2014a). 개발원조위원회(DAC) 평가 시리즈-공동평가 업무관리 가이드라인.
- 한국국제협력단. (KOICA). (2014b). 평가업무수행 길라잡이.
- 한국국제협력단. (KOICA). (2013a). 과테말라 치말테낭고주 모자보건개선사업 사후평가 보고서(2013.12), 한국국제협력단.
- 한국국제협력단. (KOICA). (2013b). 보건의료분야 ODA사업 성과관리 가이드라인 연구결과 보고서, 한국국제협력단.
- 한국국제협력단. (KOICA). (2008). 사업관리자를 위한 개발협력사업평가 가이드라인.
- 한국국제협력단. (KOICA). (2009). 프로젝트 기획, 모니터링 및 평가 방법론, 한국국제협력단.

- on Community Change (<http://www.theoryofchange.org/>).
- Angrist, J., Pischke, J-S. (2010). "The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con out of Econometrics," *Journal of Economic Perspectives*, 24(2), pp.3-30.
- Balagamwala, M., Gazdar, H. (2014). *Impact Evaluation of DFIDs Electoral Programmes in Pakistan*. September 2014. Collective for Social Science Research
- DFID.(2011), "DFID's Approach to Value for Money (VfM)", *DFID*, July 2011.
- DFID.(2013a). *International Development Evaluation Policy*. May 2013.
- DFID (2013b), *Annual Evaluation Report*, *DFID*, July 2013.
- DFID(2016), *Annual Report and Accounts 2015-16*.
- Duflo, E., Glennerster, R., Kremer, M. (2007). Using Randomization in Development Economics Research: A Toolkit. *Handbook of Development Economics*, 4, pp.3895-3962.
- Finkelstein, A., Taubman, S., Allen, H., Gruber, J., Newhouse, J., Wright, B., Baicker, K., & the Oregon Health Study Group. (2010). *The Short-run Impact of Extending Public Health Insurance to Low Income Adults: Evidence from the First Year of the Oregon Medicaid Experiment - Analysis Plan*. December 1, 2010.
- Gertler, P. J., Martinez, S., Premand, P., Rawlings, L. B., & Vermeersch, C. M. (2011). *Impact Evaluation in Practice*. World Bank Publications.
- Glennerster, R., Takavarasha. K. (2013). *Running Randomized Evaluations: a practical guide*. Princeton and Oxford: Princeton University Press.

- Grootaert, C., Narayan, D. Jones, V. N., Woolcock, M. (2004). *Measuring Social Capital: An Integrated Questionnaire*. World Bank Working Paper No. 18.
- ICAI. (2016a). *When aid relationships change: DFID's approach to managing exit and transition in its development partnerships*. November 2016.
- ICAI. (2016b). *ICAI Annual Report 2015 to 2016: Report to Parliament's International Development Committee*. June 2016.
- ICAI. (2016c). *Corporate Strategy 2015/16-2018/19*. April 2016.
- JICA (2016a). *JICA 2015 Annual Report*.
- JICA (2016b). *JICA 2015 Annual Evaluation Report*.
- Khandker, S. R., Koolwal, G. B., Samad, H. A. (2010). *Handbook on Impact Evaluation: Quantitative Methods and Practices*. The World Bank.
- Kim, B., H. Kim, and C. Pop-Eleches. (2016). *Peer Effects in the Demand for Male Circumcision: Evidence from Secondary Schools in Malawi*. Paper presented at the GRIPS Seminar Series in Economics, National Graduate Institute for Policy Studies, Tokyo, Japan.
- Kim, B. (2014). *The impact of Girls' Education Support Program on Human Capital Development: Evidence from a Randomized Evaluation in Malawian Secondary Schools*. Paper presented at the 2015 Annual Meeting of the Population Association of America, San Diego, CA, USA.
- KOFID, KCOC, INDI LAB, KOICA. (2016). 알기 쉬운 지속가능발전목표 SDGs. 국제개발협력시민사회포럼(KoFID).
- Leamer, E. (1985). Sensitivity Analyses Would Help, *American Economic Review*, 75(3), pp.308-313.
- Melin, G., Allinson, R., Grudin, M., Mostert, B., Ploeg, M., Sadesk, F.,

- Varnai, P., Warbis, M. & Zegel, S. (2015). *Evaluation of the Developing Operational Research Capacity in the Health Sector Project*.
- Miguel, E., Kremer, M. (2004). Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities. *Econometrica*, 72(1), pp.159-217.
- Morra Imas, L. G. & Rist, R. C. (2009). *The Road to Results: Designing and Conducting Effective Development Evaluations*. World Bank Publications.
- OECD DAC (1991). *Principles for evaluation of development assistance*.
- Petersen et al. (2015). "Long term evaluation of the clinical effectiveness of community milk fluoridation in Bulgaria," *Community Dental Health* 32, pp.199-203.
- Rubin, D. (1974). Estimating Causal Effects of Treatments in Randomized and Nonrandomized studies, *Journal of Educational Psychology*, 66(5), pp.688-701.
- Thomas, D., Frankenberg, E., Smith, J. P. (2001). Lost but Not Forgotten. *Journal of Human Resources*, 36 (3), pp.556-592.
- Thomas, D., Witoelar, F., Frankenberg, E., Sikoki, B., Strauss, J., Sumantri, C., Suriastini, W. (2012). Cutting the costs of attrition: Results from the Indonesia Family Life Survey. *Journal of Development Economics* 98(1), pp.108-123.
- UN Economic and Social Council (2016). *Report of the Inter-Agency and Expert Group on Sustainable Development Goal Indicators*.
- Vogel, I. and Punton, M. (2016). *Building Capacity to Use Research Evidence (BCURE) Evaluation: Stage 1 Synthesis Report*. ITAD
- WHO(2008). *Effective Aid Better Health. Report prepared for the*

Accra High Level Forum on Aid Effectiveness 2-4 September 2008. WHO.

WHO(2012). *From Whom to Whom? Official Development Assistance for Health.* 2nd Edition 2000-2010. WHO.

WHO and Ministry of Health Kingdom of Cambodia(2007). *Scaling Up for Better Health in Cambodia.* WHO.

JICA (2016c). 2016년 10월 1일 현재 조직도.

(<http://www.jica.go.jp/english/about/organization/index.html>
에서 2016. 10. 11. 인출)

JICA(2014), Ex-post Evaluation of Japanese ODA Loan “Social Sector Support Program (I), (II).

(http://www2.jica.go.jp/en/evaluation/index.php?anken=&country1=Mongolia&area2=&country2=&area3=&country3=&field1=&field2=&field3=&start_from=&start_to=&search=Search에서 2016. 9. 1. 인출)

부록 1. 영향평가의 방법들³⁵⁾

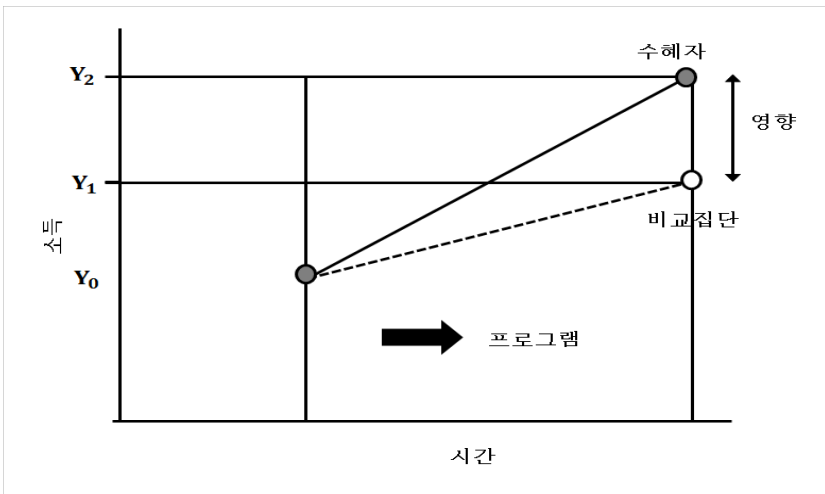
- 랜덤화(randomization) 방법 (또는 무작위통제실험)
 - 수혜집단과 비수혜집단이 사업실시 이전에 확률적으로 동일한 특성을 지닌 집단이 되도록 만들면 선택편의를 쉽게 제거할 수 있음.
 - 모집단에서 수혜자와 비수혜자를 무작위 혹은 랜덤하게 선발하면, 수혜집단과 비수혜집단은 확률적으로 동일한 집단이 되는데, 이처럼 수혜자의 무작위 선정을 통해 사업 영향을 평가하는 방법을 랜덤화라고 함.
 - 그 후 무작위로 선정된 두 그룹 중 한 그룹에만 프로그램의 혜택을 제공하게 되면, 프로그램 혜택 여부를 제외한 다른 모든 특성들이 두 그룹 간에 같아지게 되어 개인 수준에서는 관찰할 수 없었던 가상결과(Counterfactual)를 그룹 수준에서는 구성할 수 있게 됨.
 - 무작위로 배정된 프로그램 혜택 여부와 결과는 아무 상관이 없기 때문에 선택편의가 완전히 사라짐.
 - 따라서 실험군과 대조군의 평균 결과 값 차이를 측정함으로써 프로그램의 영향을 엄밀하게 평가할 수 있게 됨.
- ※ 랜덤 실험법은 신약의 효능검사와 같이 의학계에서 폭넓게 사용되었는데 그후 사회정책이나 개발정책의 효과를 확인하기 위한 목적

35) 영향평가 방법론에 대한 더욱 상세한 설명과 사례에 대해서는 Khandker et al.(2010)와 Gertler et al.(2011)을 참고.

으로도 다양하게 활용되고 있으며 현장실험(field experiment) 혹은 무작위대조군시험(RCT, randomized controlled trial)이라고도 함.

- [부도 1-1]과 같이 표본이 충분히 크다면 무작위 추출된 수혜그룹과 비수혜그룹의 초기 평균소득은 동일하며, 사업진행 과정에서 다른 요인의 영향도 확률적으로 동일하게 받으므로 사업의 효과는 두 집단간의 소득격차인 ($Y_2 - Y_1$)이 됨.

[부도 1-1] 랜덤화 조건에서 사업의 영향



자료: Khandker et al.(2010). p.34.

○ 랜덤화는 이론적으로는 사업의 영향을 정확히 측정할 수 있는 방법이지만 실제 사업현장에서 적용하는 데는 다음과 같은 어려움이 있음.

- 특정 목적을 지닌 사업이나 정책을 실시할 때 수혜대상자를 무

작위로 선발하는 것은 윤리적, 정치적 문제를 야기함.

- 전국 단위의 대규모 사업에서는 특정 지역에서의 시범사업과 다른 결과가 나타날 수 있음.
- 무작위로 선발된 수혜자가 사업 참여를 거부할 경우 완전한 랜덤화가 이루어지지 않을 수 있음.
- 사업의 효과가 수혜자에게 국한되지 않고 비수혜자도 간접적인 파급영향을 받을 수 있음.

□ 성향점수매칭(PSM, propensity score matching)방법

○ 성향점수매칭 방법은 수혜그룹에 포함될 확률모형을 근거로 비교집단을 구성하고 수혜집단과 비교집단 간 성과 차이를 통해 사업영향을 평가하는 방법임.

- 대규모의 비수혜집단 중에서 수혜집단의 각 개인과 관측 가능한(observed) 특성의 측면에서 가장 유사한 개인들로 구성된 집단을 비교집단으로 구성하면, 사업이전의 수혜집단과 비교집단이 확률적으로 유사한 집단이 되어 선택편의를 줄일 수 있다는 원리를 이용함.
- 수혜집단의 무작위선정이 어려운 상황에서는 PSM이 차선의 방법으로 이용되고 있음.

○ PSM은 다음과 같은 순서로 진행됨.

- ① 참여확률모형추정: 수혜자와 비수혜자 집단을 합쳐 관측가능한 특성들을 변수로 하여 수혜집단 포함 확률모형(model of probability of participation)을 추정하는데 이때 probit (혹은 logit)이 사용됨.

- ② 공통영역 확정: 추정된 모형을 이용해 수혜자들과 비수혜자들의 참여확률 즉 성향점수(propensity score)를 계산하여, 두 집단의 성향점수가 겹치는 공통영역(region of common support)을 확인하여 이 영역 외의 개인들은 표본에서 제외함.
- ③ 수혜자와 비수혜자의 매칭: 표본 내 수혜자 각 개인과 성향점수가 가장 유사한 비수혜자들을 매칭하는 방식으로 비교집단을 구성함.
- ④ 수혜집단과 비교집단의 성과비교: 사업이 종료된 시점에서 수혜집단과 비교집단 간 평균성과의 차이를 사업의 효과로 평가.

○ PSM은 수혜대상자를 무작위로 선정하지 않고, 기초선조사가 필요하지 않다는 장점을 갖지만 다음의 조건들이 충족될 때만 정확한 결과를 얻을 수 있음.

- 관측된 특성만으로 사업 참여확률 혹은 성향점수를 추정하므로, 관측되지 않는 특성들은 참여확률에 영향을 미치지 않다는 가정을 충족해야 함.
- 개별 수혜자와 유사한 특성을 가진 비수혜자를 발견할 수 있을 정도로 충분히 큰 비수혜자 표본이 확보되어야 함.

□ 이중차이모형(DD, double difference Model)

- 이중차이모형(DD or difference in difference)은 수혜집단과 비수혜집단에 대해 사업전후 성과지표의 차이를 측정하고, 이 차이를 두 집단 간에 비교함으로써 사업의 효과를 측정하는 방법임.
- 랜덤화나 PSM은 사업종료 후 수혜집단과 비교집단의 성과지표 차이만을 고려한다는 의미에서 단일차이모형이라고 한다

면, DD는 사업 후 두 집단 간의 차이를 사업이전 기준선과 비교한다는 점에서 차이를 이중으로 계산하는 모형임.

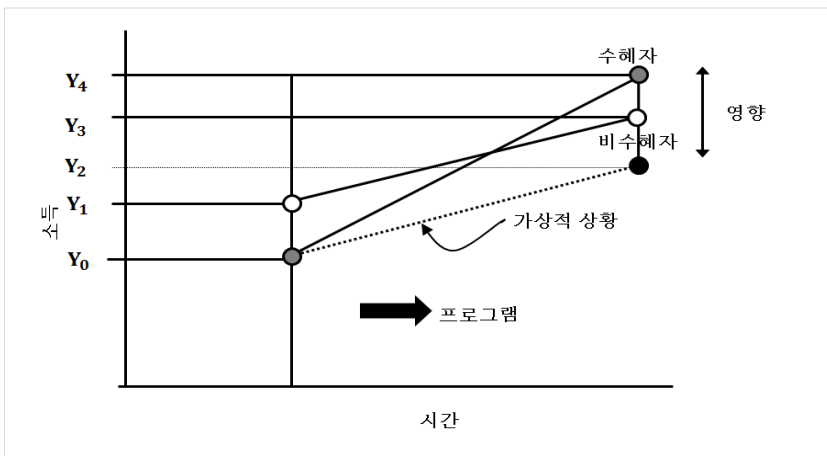
- 따라서 DD모형은 사업 후뿐만 아니라 사업 전의 값도 이용하기 때문에 패널데이터(panel data)를 이용하는 모형임.

○ 이중차이모형에서 사업성과는 다음 식과 같이 측정됨.

$$DD = E(Y_1^T - Y_0^T | T_1 = 1) - E(Y_1^C - Y_0^C | T_1 = 0)$$

- Y_t^T , Y_t^C 는 t 시점에서의 수혜자와 비수혜자의 성과를 나타내며, $t=0, 1$ 는 각각 사업 시작 전과 사업 종료 후의 시점을 의미함.
- $T_1=1, 0$ 은 각각 사업 실시 지역과 비실시 지역을 의미함.
- [그림 2-7]에서 사업의 효과는 수혜집단의 사업전후 성과차이인 $(Y_4 - Y_0)$ 와 비수혜집단의 사업전후 성과차이인 $(Y_3 - Y_1)$ 의 차이, 즉 $DD = (Y_4 - Y_0) - (Y_3 - Y_1)$ 임.

[부도 1-2] 이중차이모형(DD)을 통한 사업의 영향 측정



자료: Khandker et al.(2010). p.75.

○ DD모형은 수혜집단과 비수혜집단의 성과차이에 영향을 미치는 요인들이 시간에 따라 변하지 않는다(time-invariant)는 가정에 기초하고 있음.

- 즉, [부도 1-2]에서 사업 시작 전 두 집단의 성과지표 차이 ($Y_1 - Y_0$)은 사업 종료 이후에도 그대로 유지되는 것 ($(Y_3 - Y_2) = (Y_1 - Y_0)$)으로 가정하고 있음.
- 이는 곧 수혜집단이 사업에 참여하지 않았다고 가정했을 $t=0$ 과 $t=1$ 시점에서의 가상적(counterfactual) 지표차이 ($Y_2 - Y_0$)가 비교집단인 비수혜집단의 차이 ($Y_3 - Y_1$)와 동일하다고 가정함을 의미함.
- 이런 가정 하에서 $DD = (Y_4 - Y_0) - (Y_3 - Y_1)$ 는 측정하고자 하는 이론적 효과 ($Y_4 - Y_2$)와 동일하게 됨.
- DD는 비교집단 구성을 위해 수혜자의 무작위선정이나 관찰 가능한 요인만 사업 참여 확률에 영향을 미친다는 PSM의 가정이 필요하지 않다는 장점이 있지만, 두 집단 간의 차이가 불변한다는 가정을 필요로 함.
 - DD모형을 일반화한 것이 고정효과 패널 회귀분석(fixed effect panel regression)이라고 할 수 있음.

□ 도구변수 추정법(instrumental variable estimation)

○ 도구변수 추정법은 수혜집단을 선정이 무작위적으로 이루어지지 않아 내생성(endogeneity) 혹은 선택편의가 발생할 때, 수혜 여부를 설명할 수 있는 외생변수 즉 도구변수를 이용해 사업효과를 추정하는 방법임.

- 도구변수 추정법은 회귀분석 방법 중의 하나인데, 수혜그룹과 비수혜그룹의 성과차이를 추정하기 위한 아래 식을 이용해 설명할 수 있음.

$$Y_i = \alpha X_i + \beta T_i + \epsilon_i$$

- 여기서 Y_i 는 개인 i 의 성과지표, X_i 는 성과에 영향을 미칠 수 있는 통제변수의 벡터, T_i 는 사업 수혜 여부를 나타내는 더미 변수, ϵ_i 는 오차항
- 특정 조건 하의 수혜자 선정 혹은 자기선택(self-selection)으로 인한 관측되지 않는 개인의 이질성(unobserved individual heterogeneity)이 있는 경우 수혜 여부 즉 T_i 는 오차항 ϵ_i 와 독립적이지 않게 됨.
- $\text{cov}(T, \epsilon) \neq 0$ 이면 두 집단 간의 평균성과의 차이, 즉 사업효과 β 의 추정치인 β 는 일치추정량이 되지 않음.
- T 와 ϵ 의 상관관계를 제거하기 위해서는 T 와는 밀접한 관계를 갖지만 ϵ 와는 독립적인 외생변수, 즉 도구변수가 필요한데 도구변수 Z 는 다음과 같은 성격을 가져야 함: $\text{cov}(Z, T) \neq 0$, $\text{cov}(Z, \epsilon) = 0$.
- 도구변수 Z 를 이용해 2단계추정(2SLS) 방법으로 사업효과 β 의 불편추정량을 구할 수 있음.
 - 1단계로 T 를 도구변수 Z 와 통제변수 X 에 대해 회귀분석하여 T 의 예측값($\hat{T}$)을 구함.
 - 2단계로 영향분석 추정에서 T 대신 1단계에서 얻은 T 의 예측값($\hat{T}$)을 이용하여 추정함.

○ 도구변수 추정법은 DD와 달리 선택편의가 시간불변이라는 가정이 필요하지 않지만, 적절한 도구변수를 찾기 어려운 경우가 많음.

- 도구변수 Z 가 수혜처리 T 의 상관관계가 약할 경우(weak instrument) 사업영향 추정치에 심각한 편의가 발생할 수 있음.

□ 회귀단절모형 (RD, regression discontinuity)

○ 비실험적인 상황(수혜자의 무작위 선정이 아닌 경우)에서 일정한 기준에 따라 사업의 수혜자격(eligibility) 여부를 결정하는 경우가 있는데, 이때 탈락경계선(cut-off threshold)에 근접한 수혜자와 비수혜자를 비교하여 사업성과를 측정하는 방법이 회귀단절모형임.

- 경계선에 매우 근접한 수혜자그룹과 비수혜자그룹은 사전적으로 큰 차이가 없는 거의 동질적인 집단이라고 볼 수 있어, 사업 후 이 두 집단의 성과를 비교하는 방식으로 성과를 측정할 수 있음.
- 예를 들어 성적 50점미만 학생들에게만 보충교육을 실시한 후 그 성과를 측정하기 위해, 성적 50점 이상~51점 미만의 비참여 학생과 성적 50점 미만~49점 이상의 참여 학생의 보충교육 후 성적을 비교할 수 있음.
- 이 방법에서는 수혜 경계선이 T 의 일종의 외생적 도구변수 역할을 하는 것이라고 할 수 있음.

○ 회귀단절모형은 추정이 비교적 간단하기 때문에 쉽게 활용할 수 있다는 장점이 있지만 단점도 있음.

- RD의 추정치는 불연속점에서는 국지적으로 불편(unbiased)

추정량임.

- 다양한 사회정책에 대해 단일한 자격기준이 적용될 경우 동일한 비교집단에 대해 다양한 정책의 효과를 확인할 수 있는 장점이 있음.
- 랜덤화와 달리 수혜자격이 있는 대상자를 영향평가를 위해 사업 수혜에서 배제할 필요가 없음.
- 하지만 이 추정치는 불연속점 주변에서의 효과만을 의미할 뿐, 모집단 전체의 효과로 일반화할 수는 없음.
- 또한 경계선에 인접한 집단만을 비교하기 때문에 표본을 충분히 확보하지 못할 위험이 있음.

부록 2. 지속가능한 개발목표(SDGs)의 보건복지분야 평가목표와 지표

가. SDGs의 평가 및 모니터링 체계

- SDGs의 달성을 감시하는 역할을 맡은 유엔기구인 고위급정치포럼(HLPF: high-level political forum on sustainable development)임.
- 고위급정치포럼은 정기적으로 SDGs 이행 상황을 평가하며, 평가는 모든 나라에게 보고를 요청하는 자발적 형태로 이루어짐.
- 보고서는 각 국가에서 장관급 지위를 가진 사람이 주도하여 준비해야 하며 해당 국가 내의 관계자들이 절차에 참여할 수 있도록 보장해야 함. 각국 정부는 보고서를 유엔에 제출하여 고위급정치포럼의 검토를 받게 됨(KOFID, KCOC, INDI LAB, KOICA, 2016, p.22).
- SDGs는 17개의 목표, 169개의 세부목표 그리고 230개의 지표로 이루어짐.
- 지표는 유엔에 소속된 ‘지속가능 개발목표 지표 개발을 위한 관계기관 및 전문가 그룹(Inter-agency and Expert Group on SDGs, IAEG-SDGs)’이 제출한 보고서를 2016년 3월 UN 통계위원회(UN Statistical Commission, UNSC)가 승인하는 절차에 따라 수립
- IAEG-SDGs는 230개의 지표를 가용성에 따라 크게 3개의 Tier로 분류

- Tier I은 국제적으로 합의된 통계방법론이 수립되어 있으며 데이터를 쉽게 얻을 수 있는 지표를 의미하고, Tier II는 방법론은 수립되어 있지만 데이터를 얻기가 쉽지 않은 지표, Tier III은 방법론이 아직 개발되지 않은 지표를 의미함(김지현, 2016a, pp.5-6).
- 어느 Tier에도 속하지 않거나 복수의 Tier에 속하는 지표는 다중/미정으로 분류한다(부표 2-1 참조).

〈부표 2-1〉 Tier에 의한 SDGs 지표 분류

목표	Tier I	Tier II	Tier III	다중/미정
1. 빈곤	25%	33%	33%	8%
2. 기아	50%	14%	36%	0%
3. 보건	58%	27%	12%	4%
4. 교육	30%	40%	20%	10%
5. 양성평등	21%	29%	43%	7%
6. 물	55%	0%	45%	0%
7. 에너지	67%	0%	33%	0%
8. 일자리	59%	35%	6%	0%
9. 산업	67%	8%	25%	0%
10. 불평등	36%	0%	45%	18%
11. 도시	13%	47%	40%	0%
12. SCP	8%	23%	62%	8%
13. 기후변화	0%	29%	57%	14%
14. 해양생태계	20%	0%	80%	0%
15. 육상생태계	43%	7%	36%	14%
16. 거버넌스	39%	30%	26%	4%
17. 글로벌 파트너십	41%	21%	32%	6%

자료: 김지현(2016a). p.6 〈표 1〉 재인용.

나. SDGs의 보건 및 사회정책분야 목표 및 지표

□ SDGs의 보건분야 목표는 ‘3. 전 연령층의 모든 사람을 위한 건강한 삶 보장과 웰빙’이며, 세부목표로 3.1~3.9와 3.a~3.c가 수립

○ 세부목표 별로 달성도를 측정하기 위해 26가지 지표 마련 (부표 2-2 참조)

〈부표 2-2〉 SDGs 보건 분야의 세부목표와 지표

세부목표	지표
3.1 2030년까지 전 세계 산모사망률을 10만명당 70명 미만으로 낮춘다.	3.1.1 산모사망률
	3.1.2 출산 시 의료인력 참여 비율
3.2 2030년까지 모든 국가들이 출생 인구 1,000명당 적어도 신생아 사망률은 12명, 5세 미만 사망률을 25명까지 낮추는 것을 목표로 하여, 신생아·영유아·5세 미만 아동의 예방가능한 사망을 근절한다.	3.2.1 5세 미만 사망률
	3.2.2 신생아 사망률
3.3 2030년까지 AIDS, 결핵, 말라리아, 소외열대질환(NTD)과 같은 전염병을 근절하며, 간염·수인성 질환·기타 감염성 질환을 퇴치한다.	3.3.1 비감염 인구 1,000명당 HIV 신규 감염자 수
	3.3.2 인구 1,000명당 결핵 발병 건수
	3.3.3 인구 1,000명당 말라리아 발병 건수
	3.3.4 인구 10만 명당 B형 간염 발병 건수
	3.3.5 소외열대질환(NTD) 관련 대응이 필요한 사람 수
3.4 2030년까지 비전염성 질환으로 인한 조산으로 인한 사망률을 3분의 1 수준으로 줄이고, 정신건강과 웰빙을 증진한다.	3.4.1 심혈관계질환, 암, 당뇨병, 만성호흡기 질환으로 인한 사망률
	3.4.2 자살사망률
3.5 마약류, 알코올을 포함한 약물 오남용의 예방과 치료를 강화한다.	3.5.1 약물 사용 장애를 극복하기 위한 치료의 보장(약물치료, 심리적 치료, 재활 및 요양치료를 포함)

세부목표	지표
	3.5.2 알코올의 오용(매해 15세 이상의 성인이 1인당 사용하는 리터단위 알코올 사용량)
3.6 2020년까지 세계적으로 도로교통사고로 인한 사망 및 상해를 절반으로 줄인다.	3.6.1 도로 교통사고로 인한 injuries
3.7 2030년까지 가족계획, 정보와 교육, 국가전략 및 계획에 출산보건을 통합하는 것을 포함하여 여성 건강 및 출산보건 서비스에 대한 보편적인 접근을 보장한다.	3.7.1 가족계획을 위해 현대적 피임법을 사용할 수 있는 가임기 여성(15-49세)의 비율 3.7.2 청소년 출산율
3.8 재무위험관리, 양질의 필수 보건서비스에 대한 접근, 양질의 안전하고 효과적이며 적절한 가격의 필수 의약품 및 백신에 대한 접근을 보장함으로써 모두를 위한 보편적 의료보장(UHC)을 달성한다.	3.8.1 필수 보건서비스의 보장(임신, 출산, 신생아·아동의 건강, 전염성 질환, 비전염성 질환에 대한 추적치료와 서비스 역량, 접근성을 제공하는 필수서비스의 평균적인 보장) 3.8.2 인구 1,000명당 의료보험 또는 공중보건시스템의 혜택을 받는 사람의 수
3.9 2030년까지 유해한 화학물질이나 공기, 수질, 토지 오염으로 인한 사망 및 질병을 대폭 줄인다.	3.9.1 가정과 주변의 공기오염으로 인한 사망률 3.9.2 안전하지 않은 물과 위생시설, 위생결여로 인해 발생한 사망률(안전하지 않은 식수위생 서비스에 대한 노출) 3.9.3 비의도적 중독에 의한 사망률
3.a 모든 국가에서 적절하게 세계보건기구 <담배규제기본협약>의 이행을 강화한다	3.a.1 15세 이상 인구 중 연령 표준 흡연율
3.b 개발도상국에 주로 영향을 미치는 전염성 및 비전염성 질병에 대한 백신 및 의약품의 연구개발을 지원하고, 공중보건을 보호하고, 특히 모든 사람에게 의약품에 대한 접근을 보장하기 위해 무역 관련 지적재산권 협정의 모드 조항을 활용할 수 있는 개발도상국의 권리를 확인하는 <TRIPS 협정과 공중보건에 관한 도하선언>에 따라, 적정가격의 필수 의약품과 백신에 대한 접근을 제공한다	3.b.1 적정가격의 의약품 및 백신을 지속적으로 이용할 수 있는 인구의 비율 3.b.2 의료연구와 기초보건분야에 대한 ODA 지원 규모(순지출 기준)

세부목표	지표
3.c 개발도상국, 특히 최빈국과 군서도서개도국에서의 보건재원과 보건인력의 채용·개발·훈련·보유를 대폭 확대한다	3.c.1 보건인력의 밀도와 분포도
3.d 모든 국가, 특히 개발도상국에서 국내 및 국제적 보건위험에 대한 조기 경보, 위험경감과 관리를 위한 역량을 강화한다.	3.d.1 국제보건규칙(IHR)의 역량과 보건비상사태에 대한 준비

주: 세부목표의 국문 번역은 KOFID 등(2016) 부록 2. 참조, 지표 번역은 저자 작성.
 자료: UN Economic and Social Council(2016).

〈부표 2-3〉 SDGs 사회정책 분야의 목표 및 지표

목표	세부목표	지표
1. 모든 곳에서 모든 형태의 빈곤 종식	1.3. 국가별로 최저 생계 보장 등을 포함하여 모두를 위한 적절한 사회보장시스템 및 조치를 이행하고, 2030년까지 빈곤층과 취약계층에 대한 실질적 보장을 달성한다.	1.3.1 성별 및 아동, 실업자, 노인, 장애인, 임산부·신생아, 산재피해자, 빈곤계층과 취약계층별로 세분화되는 사회보호 안전망·시스템에 의해 보장되는 인구 비율
	1.4 2030년까지 모든 남성과 여성, 특히 빈곤층과 취약계층이 경제적 자원에 대한 권리와 더불어 기초 공공 서비스, 토지 및 기타 유형의 자산·유산·천연자원·적정기술, 소액금융을 포함한 금융서비스에 대한 옹호권과 통제권에 대한 접근에 동등한 권리를 가질 것을 보장한다.	1.4.1 기초서비스 접근성이 있는 가구에 살고 있는 인구 비율
	1.5 2030년까지 빈곤층 및 취약계층의 복원력을 구축하고 기후 관련 재해와 경제적, 사회적, 환경적 충격 및 재난에 대한 노출과 취약성을 경감한다.	1.5.1 인구 10만 명당 재해로 인한 사망자, 실종자, 피해자 수
		1.5.2 재해로 발생된 직접적인 경제적 손실(GDP)
		1.5.3 국가나 지역수준에서 재해위험 경감전략을 마련한 국가의 수

목표	세부목표	지표
8. 모두를 위한 지속적·포용적·지속 가능한 경제성장, 생산적인 완전고용과 양질의 일자리 증진	8.5 2030년까지 장애인 및 청년을 포함하여 모든 여성, 남성을 위한 생산적 완전고용과 양질의 일자리 창출 및 동일가치노동에 대한 동일임금을 달성한다.	8.5.1 성별, 직업별, 연령별, 장애 여부에 따른 시간당 평균임금
		8.5.2 성별, 연령별, 장애 여부별 실업률
10. 국내 및 국가 간 불평등 감소	10.2 2030년까지 연령, 성별, 장애 여부, 인종, 민족, 출신, 종교, 경제 및 기타 지위와 관계없이 모든 사람의 사회적·경제적·정치적 포용을 증진하고 확대한다.	10.2.1 성별, 연령별, 장애 여부 별 중위소득의 50% 이하의 삶을 사는 인구의 비율
	10.3 차별적인 법규, 정책, 관례를 철폐하고 이와 관련한 적절한 법, 정책, 활동을 증진하는 등의 노력을 포함하여 평등한 기회를 보장하고 결과의 불평등을 감소한다.	10.3.1 국제인권법에서 금지하고 있는 차별을 기준으로 지난 12개월 동안 개인적으로 차별받거나 괴롭힘을 당했다고 신고한(reporting) 인구의 비율
	10.4 재정정책, 임금정책, 사회보호 정책과 같은 정책을 도입하고 점진적으로 더 높은 수준의 평등을 달성한다.	10.4.1 GDP 중 임금과 사회보호이전(social protection transfer)으로 구성된 노동의 몫

주: 국문 번역은 KOFID 등(2016) 부록 2. 김지현(2016b) 참고.

자료: UN Economic and Social Council(2016).

간행물회원제 안내

▶ 회원에 대한 특전

- 본 연구원이 발행하는 판매용 보고서는 물론 「보건복지포럼」, 「보건사회연구」도 무료로 받아보실 수 있으며 일반 서점에서 구입할 수 없는 비매용 간행물은 실비로 제공합니다.
- 가입기간 중 회비가 인상되는 경우라도 추가 부담이 없습니다.

▶ 회원종류

- 전체간행물회원 : 120,000원
- 보건분야 간행물회원 : 75,000원
- 사회분야 간행물회원 : 75,000원
- 정기간행물회원 : 35,000원

▶ 가입방법

- 홈페이지(www.kihasa.re.kr) - 발간자료 - 간행물구독안내

▶ 문의처

- (30147) 세종특별자치시 시청대로 370 세종국책연구단지 사회정책동 1~5F
간행물 담당자 (Tel: 044-287-8157)

KIHASA 도서 판매처

- | | |
|---|---|
| ■ 한국경제서적(총판) 737-7498 | ■ 교보문고(광화문점) 1544-1900 |
| ■ 영풍문고(종로점) 399-5600 | ■ 서울문고(종로점) 2198-2307 |
| ■ Yes24 http://www.yes24.com | ■ 알라딘 http://www.aladdin.co.kr |