

# IDS Bulletin

Transforming Development Knowledge

Volume 49 | Number 4 | September 2018

## THE MILLENNIUM VILLAGES: LESSONS ON EVALUATING INTEGRATED RURAL DEVELOPMENT

Editor **Chris Barnett**



<b>Notes on Contributors</b>	iii
<b>Foreword</b> Richard Longhurst	vii
<b>Introduction: Lessons from the Millennium Villages Evaluation; Where Next for Integrated Development?</b> Chris Barnett	1
<b>Integrated Development, Past and Present</b> Edoardo Masset	17
<b><u>The Cost-Effectiveness of Complex Projects: A Systematic Review of Methodologies</u></b> Edoardo Masset, Giulia Mascagni, Arnab Acharya, Eva-Maria Egger and Amrita Saha	33
<b>Assessing Value for Money in Integrated Development Programmes – The Case of a Millennium Villages Project in Northern Ghana</b> Arnab Acharya and Tom Hilton	53
<b>Abductive Reasoning to Explain Integrated Development: Lessons from the Multi-Method Evaluation of the Millennium Villages Project</b> Dee Jupp and Chris Barnett	67
<b>Can Immersion Research Add Value in Understanding Integrated Programme Interventions?</b> Dee Jupp, David Korboe and Tony Dogbe	83
<b>Learning About Integrated Development Using Longitudinal Mixed Methods Programme Evaluation</b> Emily Namey, Lisa C. Laumann and Annette N. Brown	97
<b>Applying Factorial Designs to Disentangle the Effects of Integrated Development</b> Holly M. Burke, Mario Chen and Annette N. Brown	115
<b>Glossary</b>	129

# The Cost-Effectiveness of Complex Projects: A Systematic Review of Methodologies\*

Edoardo Masset,<sup>1</sup> Giulia Mascagni,<sup>2</sup> Arnab Acharya,<sup>3</sup> Eva-Maria Egger<sup>4</sup> and Amrita Saha<sup>5</sup>

**Abstract** Most development interventions are complex, comprising several interacting activities affecting multiple outcomes. Impact evaluations of such interventions are widespread, but the literature offers little guidance on how to assess the cost-effectiveness of such integrated projects. We review the literature that conducts cost-effectiveness analyses of multiple interventions alongside impact evaluations in low- and middle-income countries. Only seven studies are identified in areas as diverse as de-worming, school support, conditional cash transfers, early childhood development, and social funds. We find that none of the applied approaches can be effectively employed in all instances, though each of them can be applied in some special cases. Furthermore, none of the studies reviewed addresses output synergies. Given the rising numbers of impact assessments in development practice and their importance for policy, research needs to develop sound methods to assess the cost-effectiveness of integrated interventions.

**Keywords:** systematic review, methodology, cost-effectiveness analysis, complexity, multisector integrated development, synergy, cost–consequence analysis, cost apportionment, cost–utility analysis, cost–benefit analysis.

## 1 Introduction

Thanks to efforts of organisations such as, among others, the International Initiative for Impact Evaluation (3ie), the Abdul Latif Jameel Poverty Action Lab (J-PAL), and the World Bank, a large number of impact evaluations have been conducted in low- and middle-income countries in recent years. The latest figures for these countries suggest that in the past 20 years, more than 4,000 impact evaluations were undertaken (Sabet and Brown 2018). Impact evaluations have estimated the impact of development interventions in areas such as agriculture, education, health, infrastructure, and governance. Most development interventions under evaluation are complex, however, being composed of

© 2018 The Authors. *IDS Bulletin* © Institute of Development Studies | DOI: 10.19088/1968-2018.160



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non Commercial No Derivatives 4.0 International licence (CC BY-NC-ND), which permits use and distribution in any medium, provided the original authors and source are credited, the work is not used for commercial purposes, and no modifications or adaptations are made. <https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>

The *IDS Bulletin* is published by Institute of Development Studies, Library Road, Brighton BN1 9RE, UK  
This article is part of *IDS Bulletin* Vol. 49 No. 4 September 2018 'The Millennium Villages: Lessons on Evaluating Integrated Rural Development'; the Introduction is also recommended reading.

several interacting activities affecting multiple outcomes. This integration across activities complicates the policymaker's task of using evaluation results to identify effective policies from those that are ineffective.

The policymaker's task to identify effective policies is further complicated by alternative policies aiming at the same goal. For example, what is the best way to reduce rural poverty: building roads, providing fertiliser subsidies, or conditional cash transfers? The analysis of impact alone is not sufficient here to determine effective policies. It is of importance to identify the costs of achieving the given impacts across policies. These choices can be informed by cost-effectiveness studies that consider the impact of the interventions in relation to their costs in a comparative way. Indeed, some impact evaluations collect data on project costs data and calculate cost-effectiveness ratios, the cost for obtaining one unit of the benefit outcome.

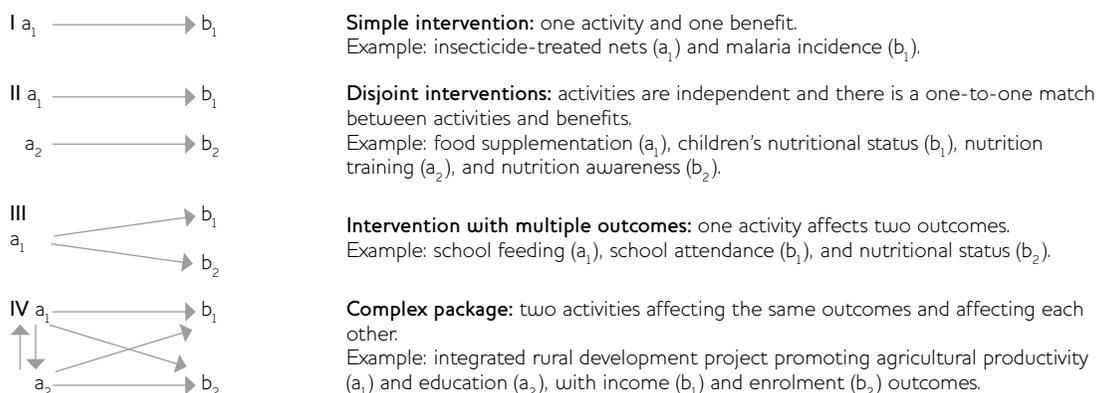
These cost-effectiveness analyses, however, face one major difficulty. Multiple activities in development projects also result in multiple outcomes, some of which are unintended. Multiple outcomes are not easily aggregated into a single index of effectiveness. It is not obvious that overarching welfare indices can be formulated and aggregated over different sectors. In relation to cost data, when there are multiple outcomes, we face the opposite problem: total project budgets cannot be easily disaggregated between the different project activities. As a result, it is often challenging to assign the cost of a project activity to its intended outcome. These are challenging methodological issues and reference books on cost-effectiveness analysis, such as Levin and McEwan (2001) and Drummond *et al.* (2005), offer little or no guidance as to how to assess the cost-effectiveness of complex interventions.

In this study, we review how researchers have conducted cost-effectiveness analyses of integrated development programmes within impact evaluations in low- and middle-income countries. The primary goal is to identify the best practices that are currently being used in order to assess the cost-effectiveness of complex interventions. A secondary goal is to map the existing cost-effectiveness literature of development programmes and identify gaps. Finally, we provide some recommendations for researchers conducting cost-effectiveness studies of complex interventions.

## **2 What are complex interventions?**

We define complex interventions as interventions consisting of one or multiple activities and producing multiple outcomes. Conversely, we define a simple intervention as an intervention consisting of one activity and one outcome, and no unintended outcomes. Figure 1 illustrates the different cases considered. In Figure 1,  $a_1$  and  $a_2$  represent intervention activities while  $b_1$  and  $b_2$  represent the benefits/outcomes of these activities. We employ a maximum of two activities and benefits in order to simplify, but the exposition can be easily generalised to more than two activities and benefits.

Figure 1 A typology of complex interventions



Source Authors' own.

Case I is a simple intervention. There is one activity ( $a_1$ ) producing a single benefit ( $b_1$ ). In the absence of unintended benefits, an incremental cost-effectiveness ratio can be calculated as the project budget (the cost of  $a_1$ ) over the impact on the outcome ( $b_1$ ), which represents the cost of obtaining a unit of the benefit for this particular intervention. For example, Bathia, Fox-Rushby and Mills (2004) compare the cost-effectiveness of two malaria control interventions: in-house residual spraying and insecticide-treated nets. The two interventions are randomly allocated at the community level and the prevalence of malaria before and after the intervention is assessed. Using project budget data, the authors calculate the average cost of averting one case of malaria for the two interventions. Insecticide-treated nets turn out to be more cost-effective by a large margin. Cost-effectiveness studies of simple interventions like this one do not present particular problems and will not be covered by our review. Case II is simply an extension of case I; it is presented to emphasise that many projects may have very separate distinct activities producing different outputs, and costs can easily be apportioned.

While cases I and II are simple, we believe they are also rare. Cases III and IV are complex and are much more common in international development. In case III, a single intervention achieves two outcomes; there is joint production through what can be thought of as a single activity. For example, school feeding programmes affect school attendance as well as the nutritional status of children. In this case, cost apportionment is not possible and cost-effectiveness ratios cannot be calculated to account for all project effects. In case IV, two activities affect the same two outcomes and also affect each other; there is joint production through multiple activities. For example, integrated rural development projects include, among others, activities promoting agricultural productivity and education. Both activities affect income and school attendance. The activities can also affect each other, for example, as more educated farmers are more likely to benefit from

Table 1 PICOTs<sup>6</sup> inclusion/exclusion criteria for selecting studies

	Included	Excluded
Population	Low-income, lower-middle-income, upper-middle-income	High-income, high-income OECD
Intervention	Complex programmes with multiple outcomes in different sectoral domains	Single-outcome programmes; programmes with multiple outcomes in the same domain; programmes with multiple outcomes evaluated only on one outcome
Comparison	Impact evaluations using experimental or quasi-experimental methods	Evaluations not employing experimental methods (RCTs) or quasi-experimental methods (regression discontinuity, matching methods, difference-in-differences)
Outcome	Multiple outcomes from different sectoral domains	Single outcomes; multiple outcomes within the same domain
Time	After 1999	Before 2000

Source Authors' own.

technical assistance. In this case, cost-effectiveness ratios can be calculated by cost apportionment but are difficult to interpret, as the cost-effectiveness ratio of each activity is not independent of costs incurred in the other activity.

### 3 Methodology: systematic review<sup>7</sup>

In this study, we conduct a systematic review of cost-effectiveness analyses of complex interventions in low- and middle-income countries. In our review, the term 'cost-effectiveness analysis' is equivalent to the term 'full economic evaluation' employed by Drummond *et al.* (2005: 9), a 'comparative analysis of alternative courses of action in terms of both their costs and consequences', which includes cost-effectiveness analysis, cost-utility analysis (CUA), and cost-benefit analysis (CBA).

#### 3.1 Search

We searched published and unpublished literature from the following databases: Medline, ERIC, the Social Sciences Citation Index, Econlit, IDEAS/RePEc, the J-PAL website, the World Bank website, the International Food Policy Research Institute (IFPRI) website, and the 3ie repository of impact evaluations. We included in the search all studies produced in English that included either in the title or the abstract the following terms: cost analysis; cost-effectiveness; cost-utility; or cost-benefit.

#### 3.2 Selection

We adopted the five inclusion and exclusion criteria listed in Table 1. First, studies of high-income countries were excluded. Second, we excluded simple programmes and only included complex multiple-outcomes interventions, as described in Figure 1. However, complex multiple outcomes interventions that were analysed only in relation to

one outcome were also excluded, as they do not offer any methodological insight. Third, we included only impact evaluations using experimental and quasi-experimental design: randomised control trials (RCTs), regression discontinuity designs (RDDs), matching methods, and difference-in-differences (DID) analyses. Fourth, we excluded studies looking at multiple outcomes that are different manifestations of the same latent construct, and we considered only multiple outcomes across different domains. For example, an intervention aiming at promoting women's empowerment and HIV prevention would be included, but an intervention aiming at improving school attendance and test scores would not be included. Lastly, we excluded all studies produced before 2000. This choice was made in the belief that few studies would be found before this date, given that the surge in impact evaluation studies in international development is a very recent phenomenon (Sabet and Brown 2018).

#### **4 Findings**

The systematic review resulted in only seven studies that fulfilled all criteria. This is surprising, as we find that an increasing number of impact evaluations (Sabet and Brown 2018) have not in turn produced comparable and credible CEA studies of complex projects. For such complex development projects, the limited results identify a critical knowledge gap where the policymaker's task of identifying effective policies remains constrained due to insufficient methods of identifying the costs of achieving given impacts across policies.

The detailed results of the search and selection process are summarised in Annexe 3. All the studies reviewed used effect sizes calculated using experiments (four randomised trials) or quasi-experiments (three propensity score-matching studies). The studies evaluated a wide range of development interventions: deworming (2); conditional cash transfers and food transfers (2); social funds (1); early childhood development (1); and preventative HIV school support (1). Costs were calculated using the 'ingredients approach' (all the components of the overall cost) in five cases and only two studies included 'social costs' (the cost of subsidising the deworming intervention and co-payments by the parents in the HIV preventative school support). Two studies also employed time discounting of benefits and costs (accounting whether a cost (or benefit) arises immediately or in the future).

##### **4.1 Cost-effectiveness methodologies of the selected studies**

To address the complexity of the interventions analysed, the studies took different approaches to assess the cost-effectiveness of the project under scrutiny. We review these approaches from the lens of our research question, i.e. whether and how the complexity challenge can be addressed with the applied method. To this end, we classified the cost-effectiveness approaches employed by the selected studies into four categories: cost-consequences analysis; cost apportionment; cost-utility analysis; and cost-benefit analysis. In what follows, we provide a brief description of each approach followed by a description and a critical appraisal of the studies employing this particular approach.

#### 4.1.1 Cost–consequence analysis

In this approach, all project costs ( $c_1, c_2, \dots, c_n$ ) and benefits ( $b_1, b_2, \dots, b_n$ ) relating to project activities ( $a_1, a_2, \dots, a_n$ ) are presented in a table for each of two or more alternative interventions (Drummond *et al.* 2005). This approach has a number of advantages (Mauskopf *et al.* 1998). First, it is extremely simple, and costs and benefits presented in this way are easily understood by policymakers. Second, it is transparent as no implicit trade-offs are imposed on decision makers. Third, it allows a more flexible decision-making process. Decision makers at different levels of operations have different goals and the results presented by cost–consequence analysis allow them to apply their own preferences in any specific context in which decisions are made. The main disadvantage of cost–consequence analysis is that it is more effective when there are few interventions and few outcomes. In these cases, the most cost-effective intervention can be identified by finding the ‘dominant’ project, which outperforms the other projects on all outcomes. For example, an education intervention may be unambiguously superior to another one by being more cost-effective in terms of both mathematical and reading achievements (Levin and McEwan 2001). However, such comparisons made by isolating dominant projects can only be made when comparing similar interventions across other complex projects.

The dominant choice is more difficult when an intervention is superior to another one on one domain but inferior on another domain; for example, if an education project is more cost-effective at improving mathematics test scores but less cost-effective at improving reading skills. In this case, as well as in the case when there are many projects and many outcomes, it is likely that decisions based on cost–consequence analysis are biased by visual inspection, vote counting, or other salient data. This approach limits an understanding of the potential connections between costs and respective benefits that may be different across interventions. By presenting the intervention costs and benefits as separate, it oversteps the problem of linking them. We are therefore somewhat sceptical that this method sufficiently accounts for the complexity of integrated projects.

Our review found three applications of this approach (Ahmed *et al.* 2009; Hidrobo *et al.* 2012; Miguel and Kremer 2004). Hidrobo *et al.* (2012) compare the cost-effectiveness of three different implementation modalities of a conditional cash transfer programme in Ecuador. Impact estimates are obtained through a randomised trial for the following outcomes: consumption, calories, household diet diversity score, dietary diversity score, and food consumption score. Costs of achieving a 15 per cent increase in each outcome are obtained for each intervention modality and compared. In the calculation of the cost-effectiveness ratio, the full cost of each intervention modality is divided by each outcome after removing project costs that are common across the three modalities. The results are presented in a table which shows that the food modality is dominated by the other two. The voucher

modality is superior to the cash modality for all but one outcome. This study is a good illustration of the cost–consequences approach but presents a number of limitations. First, the project had other two objectives in addition to increasing food security, namely empowering women and reducing tension between refugees and the host population. However, the cost-effectiveness analysis is entirely focused on the food security outcomes. Second, the outcomes considered are highly correlated and could be perceived by policymakers as attributes of the same construct so that the information provided may be excessive. Third, the authors did not report confidence intervals and the differences observed between cost-effectiveness ratios of difference interventions may lie between those intervals.

Ahmed *et al.* (2009) compare the cost-effectiveness of four different social transfers interventions to poor women in Bangladesh. Project effects are estimated using propensity score matching (PSM), and cost-effectiveness ratios of the four interventions are compared in two domains: poverty (cost of increasing per capita daily calories intake by 100Kcal; cost of increasing household monthly income by 100 Taka; and cost of reducing extreme poverty by 1 per cent) and empowerment (cost of increasing women’s participation in food decision-making; and the cost of increasing the percentage of women taking loans from non-governmental organisations (NGOs) by 1 per cent). Total project costs are used for all the outcomes. Two interventions clearly dominate the others in the income domain, while one intervention dominates all others in the empowerment domain. The dominant projects, however, are not the same across domains. The authors are careful not to draw conclusions from the analysis. This analysis suffers from the same problems highlighted in the previous study: cost-effectiveness is not measured for several other impacts of the intervention, the outcomes considered are highly correlated within each domain, and the point estimates do not include confidence intervals.

The study by Miguel and Kremer (2004) reports the use of four different cost-effectiveness approaches: a *health cost-effectiveness* approach, an *educational cost-effectiveness* approach, a *human capital investment* approach and an *externality* approach. Some of these approaches are applications of the cost–benefit analysis approach and will be discussed again below. However, the first three approaches together are an application of a cost–consequences approach. The authors estimate the impact of a randomised deworming intervention in rural areas of Kenya on school attendance, test scores (English, Mathematics, and Science), and parasitic worm infection. They calculate the cost per disability-adjusted life years (DALY) averted, the cost per additional year of schooling, and the increase in returns of education given the cost of treating a child. The cost per DALY is compared to a benchmark value for developing countries and the programme is found to be highly cost-effective. The cost per year of education is compared to cost-effectiveness ratios calculated from other interventions promoting primary education in Kenya, and it is found to dominate all alternative interventions.

Finally, the authors estimate a large increase in the net present value of wages. The combination of these results leads the authors to suggest that deworming is a highly cost-effective intervention in at least three different domains. This analysis has some limitations. First, the cost per treated child is obtained not from project data but from a similar project in Tanzania, which appears a bit arbitrary and *ad hoc*. Second, non-education interventions used as comparators are mostly hypothetical rather than real projects. Third, no confidence intervals for the cost-effectiveness ratios are reported. Finally, cost-effectiveness comparisons are not extended to test score outcomes for which the evaluation did not find a positive impact.

#### 4.1.2 Cost apportionment

In this approach, the costs of each project activity are calculated separately, and a separate cost-effectiveness ratio is calculated for the outcome of each activity (Dhaliwal *et al.* 2012). For example, in the case of an intervention providing cash transfers and health visits, the costs incurred in each activity are separately calculated and then divided, respectively, by the change in school attendance and morbidity. This approach is very appealing because of its conceptual and computational simplicity. However, as already noted, it can only be applied in the special case in which the different project activities are different interventions. A precondition for this approach to work is that there is a one-to-one mapping between costs ( $c_1, c_2, \dots, c_n$ ) and outcomes ( $b_1, b_2, \dots, b_n$ ) and there are no interactions between the activities. A further requirement of cost apportionment is that each activity must have only one outcome. Owing to these restrictions, this method does not address the challenges of assessing the cost-effectiveness of complex interventions satisfactorily.

We found that only one of the studies that were reviewed employed this approach, this (partial) exception of Abou-Ali *et al.* (2009). Abou-Ali *et al.* assessed the cost-effectiveness of a social funds intervention in Egypt which consisted of separate interventions in education, health, water and sanitation, roads, and microcredit. The costs of each intervention are separately calculated and cost-effectiveness ratios are calculated for each outcome. For example, the total cost of the social fund was 3 billion LE and the cost of the education component was 200 million LE. The authors use the latter figure to calculate the cost of having one less illiterate person. Similarly, they calculate the cost of saving one life under-five (using the total health costs), the cost of one less person with renal disease (using total water and sanitation costs), and the cost of creating one job (using total road costs). This study exposes some of the difficulties of this approach when it is applied to the assessment of integrated projects. First, cost figures for each intervention are crude and not available at any level of detail. Second, outcomes of each intervention are likely to be influenced by the other interventions. For example, the number of lives saved is affected by the health intervention but also by the road intervention so that attribution of intervention costs to a single outcome is rather arbitrary. Third, each intervention has several outcomes and it is not realistic to assign all cost to a single outcome.

#### 4.1.3 Cost–utility analysis

The cost–utility approach explicitly addresses multiple outcomes by aggregating utilities produced by the outcomes. The approach is based on the estimation of utility ( $U$ ) through a utility function specified for  $b$  outcomes ( $b_1, b_2, \dots, b_n$ ):

$$U=f(u_i(b_i))$$

The utility so obtained is then used to calculate a cost (C)–utility ratio:

$$CUR=C/U$$

The estimation of an overall utility of the intervention assumes knowledge of the utilities associated to each outcome and of the functional form used for their aggregation. This is not a simple task. Quality-adjusted life years (QALYs) and disability-adjusted life years (DALYs) used by health economists are applications of this approach. QALYs and DALYs aggregate all outcomes in terms of life years gained weighted by the quality of living under different levels of morbidity and disability. For example, in QALY, weights represent preferences over different health states, and are obtained through hypothetical lotteries conducted with experimental samples of subjects over different health states and are obtained through hypothetical lotteries conducted with experimental samples of subjects. The characteristics employed in developing the weights include such aspects of life as cognitive skills, physical strength, and emotional wellbeing, which are all crucial in the development of human capital. It is not obvious, however, how similar indices could be calculated in other sectors such as education or infrastructure or how an overall utility index could be calculated for all outcomes of all sectors. On the other hand, the problem of including other economic benefits and costs of health interventions has been acknowledged in health economics (Drummond *et al.* 2005). Health interventions can inflict costs to project beneficiaries; for example, by using their working time ( $c_e$ ), as well as benefits ( $b_e$ ), for example by increasing their productivity. These benefits and costs can be calculated and aggregated to project costs in the calculation of the cost–utility ratios, though this is rarely done:

$$CUR=(C+b_e-c_e)/U$$

We did not include cost-effectiveness analyses of health interventions using QALYs and DALYs in our review unless they explicitly tried to account for non-health outcomes of the intervention in this way.

We found only one cost–utility analysis that considered non-health benefits (Miller *et al.* 2013). Miller *et al.* assessed the cost-effectiveness of an HIV prevention intervention, which provided school support to orphan girls in Zimbabwe. They estimated the impact of the intervention through an RCT on three outcomes: early marriage, years of schooling, and health-related quality of life. In order to include the non-health outcomes in the cost–utility analysis, they estimated the returns to schooling (wages) resulting from increased years of education,

and the savings in medical costs resulting from the reduction in early marriage and therefore in HIV infection. They then subtracted these non-health benefits from project costs. The authors conclude that the intervention is highly cost-effective in comparison to the World Health Organization (WHO) benchmark. This approach is applicable to cases in which all but one of the outcomes can be expressed in monetary terms, and when the outcome that cannot be expressed in monetary terms is the relevant one. This study too, however, has some limitations. First, it is not clear that the evaluation included all the relevant outcomes of the intervention such as, for example, learning outcomes. Second, some of the preferences in relation to productivity gains and marriage choices might be already incorporated in the QALY weights, thus leading to a double counting of project effects. Finally, several of the cost calculations and estimations of benefits were rather *ad hoc* and arbitrary, which might be inevitable when trying to monetise all outcomes of an intervention.

#### 4.1.4 Cost–benefit analysis

Another approach to aggregate monetary and non-monetary benefits consists of expressing all non-monetary outcomes in monetary terms using opportunity costs and shadow prices. Cost–benefit analysis compares the streams of all project benefits ( $B = b_1 + b_2 + \dots + b_n$ ) and all project costs ( $C = c_1 + c_2 + \dots + c_n$ ), all expressed in monetary terms and discounted over time  $t$  at the rate  $r$ . One typical measure is the net benefit ratio ( $NBR$ ):

$$NBR = \sum_{t=1}^n \frac{B_t - C_t}{(1 + r)^{t-1}}$$

The  $NBR$  allows the economic evaluation of any project, not just in comparison to other projects, but also in absolute terms. It is able to tell whether a project is preferable to others and if a project is worthwhile regardless of other projects. The main limitation of this approach is that rarely all benefits can be expressed in monetary terms, unless the researchers are willing to make strong assumptions and several questionable imputations.

We found two cost–benefit analyses in our review (Baird *et al.* 2012; Bernal and Fernández 2013). Bernal and Fernández (2013) assessed the cost-effectiveness of an early childhood intervention in Colombia. They estimate the impact of the programme by length of beneficiary exposure using propensity score matching on the following outcomes: nutritional status, four indicators of socioemotional skills, and six indicators of cognitive development. The positive impact in each domain is translated into wage gains using the results found in the impact evaluation literature. The authors calculate the different values of the cost–benefit ratio depending on the child category considered and using different discount rates. Though the programme effects are, in general, modest, they found the programme to be cost-effective for children exposed for longer than 15 months. The limitations of this study are the following. First, the wage gains are estimated by applying parameters calculated by studies using data from different populations

from the one analysed and no sensitivity analysis is performed. Second, the different effects of the programme on wages are simply added to each other, thus assuming they are independent. However, if, for example, both nutritional and emotional improvements contribute to cognitive development directly and indirectly by affecting each other, this procedure results in double counting. Third, it is not obvious that the three selected indicators represent the whole impact of the intervention and other effects might be present. Finally, no allowance is made for the uncertainty of impact estimates due to sampling variation.

Baird *et al.* (2012) build on earlier work by Miguel and Kremer (2004) already discussed above. The authors revisit a sample of individuals ten years after the implementation of a randomised deworming intervention. They estimate the long-term impact of the intervention on education, employment, labour supply, and productivity, applying difference-in-differences analysis to the original project and control groups. They estimate the wage gains determined by an increase in hours worked. They propose assessing cost-effectiveness employing a *welfare* approach and a *social-planner* approach. In the first case, they compare wage gains to subsidy costs borne by the government and find that the gains largely exceed costs, and that government tax revenues generated by the programme largely compensate for the subsidy. Using the second approach, they calculate the internal rate of return of the project using discounted streams of earning gains and subsidy costs. The returns are shown to be four times the current interest rate, again showing the effectiveness of the project. This study, together with the earlier study by Miguel and Kremer (2004) is probably the best attempt to evaluate the cost-effectiveness of a complex intervention that we were able to find in our review. It does, however, have some limitations. First, no allowance is made for uncertainty of the results because of sampling variation. Second, cost estimates are based on a programme implemented in a different area and population.

## 5 Conclusions

Our review was able to find only seven studies assessing the cost-effectiveness of complex interventions. This is certainly a reflection of the limited use of cost-effectiveness analysis in the practice of impact evaluation. Despite the surge of impact evaluation studies in recent years, few cost-effectiveness analyses are conducted alongside impact evaluations. It should be emphasised, however, that the small number of studies found is also the result of the difficulty of identifying cost-effectiveness studies and of the restrictive selection criteria adopted. We identified studies by screening titles and abstracts, but cost-effectiveness analyses are often conducted as a secondary goal by impact evaluations and may not be reported in the title and abstract. In addition, we limited the review to impact evaluations using experimental and quasi-experimental designs, and did not consider cost-effectiveness analysis employing the results of other impact evaluations, of which there might be many. Finally, we defined complex interventions as interventions with multiple outcomes across sectors. This was done

mainly with the goal of excluding all cost–utility analyses using QALYs and DALYs produced by health economists, but which may have resulted in the exclusion of relevant studies in other sectors as well.

All the studies reviewed concluded that the interventions were cost-effective. These conclusions, however, are tempered by the methodological problems involved in assessing the cost-effectiveness of complex interventions. We identified the use of four different methodologies: cost–consequence analysis; cost apportionment; cost–utility analysis; and cost–benefit analysis. Each of these approaches may be employed effectively in specific cases, but none can be applied in all cases. Cost–consequence analysis is simple and easy to use, but it requires a cost-effectiveness comparison between few projects and few outcomes. Cost apportionment is a straightforward method of assessing cost-effectiveness, but it requires that each project component has a single outcome and that project components are independent. Cost–utility analysis has been applied very successfully in the health sector, but it is unclear whether utility indices like QALYs and DALYs can be developed in other sectors such as, for example, education and governance, and it is even less clear whether a single utility index can be formulated for all outcomes across all sectors. Finally, cost–benefit analysis effectively assesses the welfare impact of an intervention, but not all outcomes can be monetised unless we are willing to accept some peculiar assumptions.

In addition to the methodological difficulties outlined above, the studies reviewed shared a few other limitations. First, they rarely considered all the intended and unintended outcomes of the interventions. The choice of the outcomes often appeared to be motivated more by the availability of data rather than by a solid theory of how the interventions determine the outcomes. Second, none of the studies reported confidence intervals of the cost-effectiveness ratios. A meta-analysis of cost-effectiveness ratios of primary education interventions by McEwan (2014) shows how the inclusion of confidence intervals may considerably change the policy conclusions of a cost-effectiveness study. Finally, all studies suffered the practical difficulties of obtaining cost data and only two studies included social costs.

In summary, our review found few cost-effectiveness studies of complex interventions, no widely applicable methodologies, and a number of practical problems in measuring the costs and effects of the interventions. Much could be improved by conducting more cost-effectiveness studies along impact evaluations, and by exercising more care in the calculation and reporting of costs and outcomes. However, what appears to be more urgently needed is the discovery of methodologies able to aggregate outcomes and disaggregate costs, and a more systematic approach to the cost-effectiveness of complex interventions. We praise the studies included in this review for making the effort to assess cost-effectiveness across a multiplicity of outcomes. Most development interventions are complex, and the first wrong assumption made by many cost-effectiveness studies is that they are not.

## Annexe 1 Data extraction sheet

### Section 1 Study details

1.1 Authors	Author names
1.2 Title	Title of the study
1.3 Type of study	Indicate type of study and source, for example: 'Journal article, Journal of Development Effectiveness' or 'Working paper, IFPRI'
1.4 Brief description of programme/ intervention	A brief description of the project evaluated
1.5 Sector	Sectors of intervention such as education, health, etc.
1.6 Country (-ies) of implementation	Countries
1.7 Method used in the impact evaluation	Impact evaluation method: RCT, RDD, PSM, DID
1.8 Outcomes of the programme	Outcomes of the intervention

### Section 2 Cost data

2.1 Government or social	Specify which cost is considered: government (project budget); social (including shadow prices, costs to beneficiaries, costs to society, etc.)
2.2 Ingredients (1) or all budget (2)	Specify which costing method was used: ingredients (all the components of the overall cost), or all budget (the whole budget cost is included in the CEA)
2.3 If the ingredients approach is used, what cost categories are included?	List the cost categories used; for example, direct costs of the transfer, costs for delivering the transfer to the beneficiary, etc.
2.4 Quality assessment	An assessment of the quality of the cost data and whether enough information is given in the paper
2.5 Others	Any other relevant observation

### Section 3 Methods for cost-effectiveness analysis

3.1 Method category (codes 1 to 4)	(1) Total cost of the programme divided by the outcomes, separately for each of the outcomes. (2) The cost of the relevant component divided by the relevant outcome. (3) Welfare approach comparing costs and benefits (CUA or CBA). (4) Any other method.
3.2 Description of method	Brief description of the method used
3.3 Cost apportionment (YES/NO)	Whether cost apportionment was made or not
3.4 Treatment of multiple outcomes	Method to address cost-effectiveness of multiple outcomes
3.5 Discounting (YES/NO)	Whether discounting was used or not
3.6 Do authors flag a concern related to the treatment of multiple outcomes? (YES/NO)	Whether multiple outcomes issue is discussed
3.7 Do the authors mention synergies? (YES/NO)	Whether synergies between activities are discussed
3.8 Do the authors discuss or consider social costs (even if separate from CEA)? (YES/NO)	Whether social costs are considered
3.9 Quality assessment	Overall assessment of the cost data
3.10 Others	Any other relevant observation

Source Authors' own.

**Annexe 2 Study details**

**Table A1 Details of the studies included in the review**

Authors	Title	Intervention	Impact evaluation	Outcomes	Costs	Cost-effectiveness analysis	Conclusions
Ahmed <i>et al.</i> (2009)	Comparing Food and Cash Transfers to the Ultra-Poor in Bangladesh	A World Food Programme (WFP) social protection programme in Bangladesh. Four different interventions targeting poor women: (1) food rations; (2) food and cash; (3) food and cash as wage payments for public works; (4) cash as wages in road construction.	PSM	Food consumption; caloric intake; nutritional status; income; poverty; expenditure of former beneficiaries.	Ingredients approach (direct costs of the transfers and of delivering the transfers). No social costs. No cost apportionment.	The cost of each programme (including the transfer and the delivery cost) is divided by each outcome separately.	A ranking of the programmes is provided in relation to just one outcome (poverty reduction).
Hidrobo <i>et al.</i> (2012)	Cash, Food or Vouchers? Evidence from a Randomized Experiment in Northern Ecuador	A WFP social protection programme in Ecuador. Three different transfer modalities: (1) cash transfers; (2) food rations; (3) food vouchers.	RCT	Food expenditure; caloric intake; dietary diversity score, dietary diversity index, food consumption score.	Ingredients approach. After accounting for common costs, the differences in marginal costs are considered. Social costs considered but not included in CEA.	The cost of each programme is divided by each outcome separately.	Food is the most expensive method on all outcomes while vouchers are cheaper than cash transfers on four out of five outcomes.
Miguel and Kremer (2004)	Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities	A deworming and health education intervention by an NGO in a small district of rural Kenya.	RCT	Infection rates by four types of worms; school attendance; maths, English, and science test scores.	It uses the cost of US\$0.49 per pupil from a similar project in Tanzania because the programme is not running to scale and because it was difficult to remove the evaluation cost from the project cost. No social costs.	It calculates the cost per DALY for the worm infections; cost per additional years (or days) of schooling; the cost per education returns to schooling. All ratios use the same project cost figure.	The cost per DALY is compared to a benchmark while the other cost ratios are compared to those of other primary education interventions in Kenya. Deworming is found to dominate other projects.

Table A1 Details of the studies included in the review (cont.)

Authors	Title	Intervention	Impact evaluation	Outcomes	Costs	Cost-effectiveness analysis	Conclusions
Abou-Ali <i>et al.</i> (2009)	Evaluating the Impact of Egyptian Social Fund for Development Programs	The Egyptian social fund for development established in 1991. It includes (1) a community development programme (literacy classes, primary health care, small environmental projects); (2) a public works programme (potable water, sewage; and roads); (3) a microcredit programme (loans).	PSM	Illiteracy rate, employment rate, immunisation rate, under-five mortality rate, health expenditure, diarrhoea incidence, water-related morbidity, income and poverty.	It uses the total cost of the three programmes over the period 1991–2007. A crude apportionment of costs to project components is performed. No ingredients approach. No social costs.	For sectors, the cost of the relevant component is divided by the relevant outcome. For outcomes within sector, the total cost of the programme is divided by the outcomes, separately for each of the outcomes.	Most projects do not appear to be very cost-effective although to varying extents.
Miller <i>et al.</i> (2013)	Cost-Effectiveness of School Support for Orphan Girls to Prevent HIV Infection in Zimbabwe	School support provided to Grade 6 orphan adolescent girls in a province of rural Zimbabwe during 2010–14. Support consisted of a package of fees, uniforms, school supplies, and teacher training.	RCT	Three outcomes: school retention (metric is years of schooling); early marriage (metric is years unmarried); QALY (metric is a health-related self-assessment on a five-dimensional scale).	It uses the ingredients approach for 11 budget items. No social costs.	Cost per QALY. Multiple outcomes are accounted for by subtracting the monetarisation of other benefits from project costs.	The intervention is found to be highly cost-effective in relation to the WHO QALY threshold.
Baird <i>et al.</i> (2012)	Worms at Work: Long-Run Impacts of Child Health Gains	This is the same deworming intervention evaluated by Miguel and Kremer after a ten-year follow-up.	RCT, DID	Self-reported health, hours of work and meals per day.	It appears to use an ingredients approach. Social costs are included by considering the costs of subsidising the intervention.	They summarise the welfare benefits of the programmes in terms of wage gains. They compare wage gains to the cost of subsidy and calculate the internal rate or return of the project.	They find that the intervention is cost-effective.

Table A1 Details of the studies included in the review (cont.)

Authors	Title	Intervention	Impact evaluation	Outcomes	Costs	Cost-effectiveness analysis	Conclusions
Bernal and Fernández (2013)	Subsidized Child Care and Child Development in Colombia: Effects of Hogares Comunitarios de Bienestar as a Function of Timing and Length of Exposure	A home-based early child development programme for children under five, delivering childcare, supplemental nutrition, and psychosocial stimulation.	PSM	Cognitive abilities, nutritional status, and socio-psychological development.	Ingredients approach including parents' co-payments, but not other costs incurred by families.	All impacts are translated into wage increases using the results from other studies. A benefit-cost ratio is calculated.	The programme may increase future earnings though the benefit-cost ratio is modest.

Source Authors' own.

**Annexe 3 Data extraction and search results**

The studies finally selected for the review were independently analysed by three reviewers, covering questions regarding costing and cost-effectiveness methods (see Annexe 1). In particular, the methods employed to collect cost data were reviewed and the quality of such data assessed when possible. Further, the reviewers extracted the cost-effectiveness method applied and how it accounted for the presence of multiple outcomes (see Table A1 in Annexe 2).

The results of the search and selection process are illustrated in Figure 2. Search of the databases returned a total of 19,080 hits. Removal of duplicates and a first screening based on titles and abstract, which removed studies not relating to low- and middle-income countries and studies not conducting cost-effectiveness analysis, led to the selection of 2,235 studies. Further review of titles and abstract and a sequential application of the following selection criteria: (1) low- and middle-income country; (2) complex interventions; (3) cost-effectiveness analysis; (4) impact evaluations; and (5) analysis of multiple outcomes, led to the selection of 31 studies. Finally, an in-depth review of the 31 studies selected led to the exclusion of 24 other studies as closer inspection revealed they did not conform to the five selection criteria above. Only seven studies were selected for the final review.

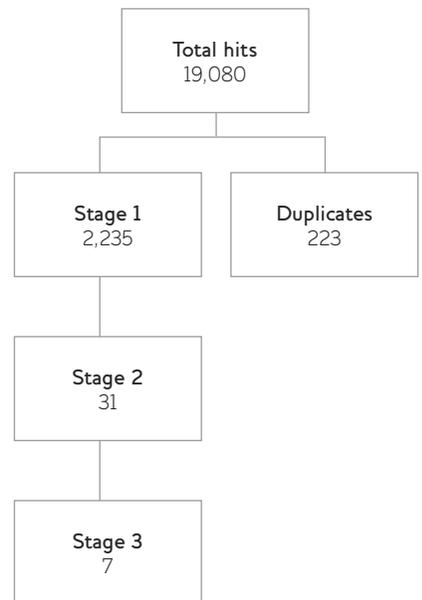
**Figure A1 Flow chart of the selection process**

Medline: 3,527; ERIC: 270; Web of Science: 9,056; Econlit: 432; Ideas: 4,434; J-Pal: 160; World Bank: 878; IFPRI: 200; 3ie: 123; Cochrane: 51; DFID: 97; Campbell: 247; EPPI: 199

**Stage 1:** By reviewing titles and abstracts we excluded: studies from high-income countries; non-cost-effectiveness studies; and duplicates

**Stage 2:** By reviewing titles and abstracts we sequentially selected: studies on low- and middle-income countries (1,920); complex programmes (992); cost-effectiveness analyses (70); impact evaluations (31)

**Stage 3:** In-depth review led to the exclusion of a further 24 studies



Source Authors' own.

### Notes

- \* This issue of the *IDS Bulletin* was prepared as part of the impact evaluation of the Millennium Villages Project in northern Ghana, 2012–17, funded by the UK Department for International Development (DFID) ([www.dfid.gov.uk](http://www.dfid.gov.uk)). The evaluation was carried out by Itad ([www.itad.com](http://www.itad.com)) in partnership with IDS ([www.ids.ac.uk](http://www.ids.ac.uk)) and PDA-Ghana ([www.pdaghana.com](http://www.pdaghana.com)). The contents are the responsibility of the evaluation team and named authors, and do not necessarily reflect the views of DFID or the UK Government.
- 1 Centre of Excellence for Development Impact and Learning (CEDIL) at the London School of Hygiene & Tropical Medicine.
  - 2 Institute of Development Studies, Brighton, UK.
  - 3 Honorary Assoc. Professor, London School of Hygiene & Tropical Medicine (2012–17).
  - 4 International Fund for Agricultural Development (IFAD), Rome, Italy.
  - 5 Institute of Development Studies, Brighton, UK.
  - 6 The PICO model is widely used in the synthesis of evidence as a strategy to formulating questions and organising a literature search. PICO stands for Population or Problem (what are the characteristics of the project population or the nature of the problem considered?), Intervention (what is the intervention?), Comparison (what is the counterfactual?), and Outcomes (what are the relevant outcomes?).
  - 7 This systematic review has not been registered with the Campbell Collaboration due to the difficulties of matching the criteria for a review of *methodologies* like this one, rather than the impact evaluation of *results*. However, we followed closely the Campbell Collaboration Systematic Review Guidelines. Due to the nature of our research question, we made no attempt to summarise the evidence in a quantitative way.

### References

- Abou-Ali, H.; El-Azony, H.; El-Laithy, H.; Haughton, J. and Khandker, S.R. (2009) *Evaluating the Impact of Egyptian Social Fund for Development Programs*, World Bank Policy Research Working Paper 4993, Washington DC: World Bank
- Ahmed, A.U.; Quisumbing, A.R.; Nasreen, M.; Hoddinott, J.F. and Bryan, E. (2009) *Comparing Food and Cash Transfers to the Ultra-Poor in Bangladesh*, IFPRI Research Monograph 163, Washington DC: International Food Policy Research Institute
- Baird, S.; Hicks, J.H.; Kremer, M. and Miguel, E. (2012) 'Worms at Work: Long-Run Impacts of Child Health Gains', unpublished paper
- Bathia, M.R.; Fox-Rushby, J. and Mills, A. (2004) 'Cost-Effectiveness of Malaria Control Interventions when Malaria Mortality is Low: Insecticide-Treated Nets Versus In-House Residual Spraying in India', *Social Science and Medicine* 59.3: 525–39
- Bernal, R. and Fernández, C. (2013) 'Subsidized Child Care and Child Development in Colombia: Effects of Hogares Comunitarios de Bienestar as a Function of Timing and Length of Exposure', *Social Science and Medicine* 97: 241–49

- Dhaliwal, I.; Duflo, E.; Glennerster, R. and Tulloch, C. (2012) *Comparative Cost-Effectiveness Analysis to Inform Policy in Developing Countries: A General Framework with Applications for Education*, Cambridge MA: Abdul Latif Jameel Poverty Action Lab (J-PAL), Massachusetts Institute of Technology (MIT)
- Drummond, M.F.; Sculpher, M.J.; Torrance, G.W.; O'Brien, B.J. and Stoddart, G.L. (2005) *Methods for the Economic Evaluation of Health Care Programmes*, 3rd ed., New York NY: Oxford University Press
- Hidrobo, M.; Hoddinott, J.; Peterman, A.; Margolies, A. and Moreira, V. (2012) *Cash, Food, or Vouchers? Evidence from a Randomized Experiment in Northern Ecuador*, IFPRI Discussion Paper 01234, Washington DC: International Food Policy Research Institute
- Levin, H.M. and McEwan, P.J. (2001) *Cost-Effectiveness Analysis*, Thousand Oaks CA: Sage Publications
- Mauskopf, J.; Paul, J.E.; Grant, D.M. and Stergachis, A. (1998) 'The Role of Cost-Consequence Analysis in Healthcare Decision-Making', *Pharmacoeconomics* 13.3: 277–88
- McEwan, P.J. (2014) 'Improving Learning in Primary Schools of Developing Countries: A Meta-Analysis of Randomised Experiments', unpublished paper
- Miguel, E. and Kremer, M. (2004) 'Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities', *Econometrica* 72.1: 159–217
- Miller, T.; Hallfors, D.; Cho, H.; Luseno, W. and Waehrer, G. (2013) 'Cost-Effectiveness of School Support for Orphan Girls to Prevent HIV Infection in Zimbabwe', *Prevention Science* 14.5: 503–12
- Sabet, S.M. and Brown, A.N. (2018) 'Is Impact Evaluation Still On the Rise? The New Trends in 2010–2015', *Journal of Development Effectiveness* 10.3: 291–304

This page is intentionally left blank