RESEARCH ARTICLE

From impact evaluation to decision-analysis: assessing the extent and quality of evidence on 'value for money' in health impact evaluations in low- and middle-income countries

[version 1; peer review: 1 approved, 1 approved with reservations]

Noemi Kreif ¹, Andrew J. Mirelman², James Love-Koh¹, Sangjun Kim¹, Rodrigo Moreno-Serra¹, Paul Revill¹, Mark Sculpher ¹, Marc Suhrcke¹

¹Centre for Health Economics, University of York, UK, York, UK

²Health Systems Governance and Financing, World Health Organization, Geneva, Switzerland

v1 First published: 07 Jan 2021, 5:1
<https://doi.org/10.12688/gatesopenres.13198.1>

Latest published: 07 Jan 2021, 5:1
<https://doi.org/10.12688/gatesopenres.13198.1>

Abstract

Background: Health impact evaluations (HIEs) are currently the main way of assessing policy changes in low-and middle-income countries (LMICs). However, evidence on effectiveness alone cannot reliably inform decisions over the allocation of limited resources. Health economic evaluation provides a suitable framework for 'value for money' assessments.

Methods: In this article we explore to what extent economic evaluations have been conducted alongside published health impact evaluations, then we assess the quality of these, using criteria from an economic evaluation reference case developed for use in LMICs.

Results: Among the 2419 HIEs stored in the International Initiative for Impact Evaluations (3ie) database, and among the 8155 studies identified by the Ovid Medline database search, only 70 studies included an economic evaluation. When measured against the quality assessment criteria, study quality showed great variation. Many studies did not fulfil the basic requirements for economic evaluation, such as stating the perspective of the budget holder, using generic health measures that can be compared across diseases, or suitably reflecting uncertainty.

Conclusions: Greater effort should be directed towards bringing the fields of impact evaluation and economic evaluation together to better inform resource allocation decisions in global health.

Open Peer Review

Approval Status

	1	2
version 1
07 Jan 2021	view	view

1. **Laura Anselmi**, The University of Manchester, Manchester, UK
2. **Ciaran N. Kohli-Lynch** , Northwestern University, Chicago, USA
University of Glasgow, Glasgow, UK

Any reports and responses or comments on the article can be found at the end of the article.

Keywords

impact evaluation, economic evaluation, cost-effectiveness, global health, low- and middle-income countries

Corresponding author: Noemi Kreif (noemi.kreif@york.ac.uk)

Author roles: **Kreif N:** Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Writing – Original Draft Preparation, Writing – Review & Editing; **Mirelman AJ:** Conceptualization, Data Curation, Formal Analysis, Methodology, Writing – Original Draft Preparation, Writing – Review & Editing; **Love-Koh J:** Conceptualization, Formal Analysis, Investigation, Writing – Original Draft Preparation, Writing – Review & Editing; **Kim S:** Conceptualization, Formal Analysis, Investigation; **Moreno-Serra R:** Conceptualization, Methodology, Writing – Original Draft Preparation, Writing – Review & Editing; **Revill P:** Conceptualization, Methodology, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing; **Sculpher M:** Conceptualization, Methodology, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing; **Suhrcke M:** Conceptualization, Methodology, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This research was jointly funded by the Bill and Melinda Gates Foundation (OPP1087363) and the National Institute for Health Research (NIHR) (16/137/90) using UK aid from the UK Government to support global health research. The views expressed in this publication are those of the author(s) and not necessarily those of the NIHR or the UK Department of Health and Social Care. *The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

Copyright: © 2021 Kreif N *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Kreif N, Mirelman AJ, Love-Koh J *et al.* **From impact evaluation to decision-analysis: assessing the extent and quality of evidence on ‘value for money’ in health impact evaluations in low- and middle-income countries [version 1; peer review: 1 approved, 1 approved with reservations]** Gates Open Research 2021, 5:1 <https://doi.org/10.12688/gatesopenres.13198.1>

First published: 07 Jan 2021, 5:1 <https://doi.org/10.12688/gatesopenres.13198.1>

Introduction

The demand for better evidence on “what works” in health policy in low- and middle-income countries (LMICs) has soared in recent years (Shi *et al.*, 2015). The growing demand has at least partly been met by a considerable growth in *impact evaluations* (Cameron *et al.*, 2016; Gertler *et al.*, 2016), studies that seek to causally attribute changes in an outcome to a given intervention, using either an experimental or quasi-experimental design (3ie Impact Evaluation Glossary, 2012). Not only has the quantity of such studies increased, but quality has also improved overall, as impact evaluations have adopted increasingly rigorous evaluation methodologies (McEwan, 2012), providing policy makers with relevant, credible evidence on intervention effectiveness.

However, effectiveness evidence alone is not enough to determine whether to invest in a programme, in the (ubiquitous) situation of limited resources. Decision-makers also need to assess the costs and opportunity costs associated with different policy choices. Although there is widespread agreement that health economic evaluation methods can suitably inform resource allocation decisions, there is evidence that these tools are under-utilised (Evans, 2016) and often imperfectly applied (McEwan, 2012) in the impact evaluation field. This is in contrast to the increasingly routine use of economic evaluation, often applying novel and sophisticated methods, to inform decisions on the adoption and reimbursement of drugs, devices and other health care interventions, as e.g. in health technology assessment (HTA) (Drummond *et al.*, 2015).

In this paper we assess the extent to which ‘value for money’ has been considered in published health-related impact evaluations in LMICs. To our knowledge, this is the first study to assess this aspect in a comprehensive and rigorous manner, focusing on health-related interventions. For those studies with some kind of economic evaluation, we examine the quality of the economic assessment. We use quality criteria derived from the International Decision Support Initiative (iDSI) Reference Case for Economic Evaluation. This provides a set of principles and flexible methodological specifications intended to support the planning, conduct, and reporting of economic evaluations of health interventions in LMICs. (iDSI, 2016; Wilkinson *et al.*, 2016). On the basis of these quality assessments, we provide suggestions for future research and highlight how economic evaluation can be better conducted alongside impact evaluation.

Methods

Literature review

Data sources, search and screening criteria. We conducted a purposive literature review using two strategies to identify relevant published impact evaluation and accompanying economic evaluation studies.

Firstly, we used the repository of published impact evaluations from the International Initiative for Impact Evaluations (3ie). The 3ie database aims to stock all published¹ impact evaluations

of development interventions. 3ie first conducted a systematic literature search in 2014, covering more than 35 databases of published studies, including journal articles, book chapters, reports, or working papers. This database was last fully updated in 2016². 3ie’s inclusion criteria require that a study has been conducted in a LMIC country (classified by World Bank criteria in the year of publication); examines the effectiveness of a specific development intervention; and uses an approved experimental (i.e. randomised controlled trials (RCTs), cluster RCTs), or quasi-experimental econometric strategy (i.e. differences-in-differences (DiD), propensity score matching (PSM), instrumental variables (IV), regression discontinuity design (RDD), or other methods including multivariate matching and regression approaches). It covers interventions in many sectors (including agriculture and rural development, economic policy, energy, disaster management and others).

For health, we searched the “health, nutrition and population” segment of the repository, between the years 2000 and 2016. This includes RCTs which go beyond laboratory trials and examine interventions in real world settings, but it excludes those trials which only address the biomedical efficacy of a drug or treatment. To identify studies that have conducted any economic evaluation, we searched for articles with the keywords: “economic evaluation” OR “cost-effectiveness” OR “cost-benefit” OR “cost-utility” OR “cost”.

Secondly, we conducted a literature search replicating the search strategy used in the 3ie repository and using the same terms to identify economic evaluations. This was in case the 3ie search, which focused on identifying impact evaluations, missed any accompanying economic evaluations studies. We searched the Ovid MEDLINE database, which has an implicit focus on the health field, extending the search period until 2017 (see *Extended data* (Kreif, 2020) for full details). The inclusion/exclusion criteria are summarised in Table 1. We only included studies that used health outcome measures. Health outcomes were interpreted broadly, including health benefits such as height gained, weight loss, life years gained, and intermediate outcomes such as the accuracy of a malaria testing kit, patient attendance rate, etc. This led to the exclusion of, for instance, the study by Alatas *et al.* (2012) which compared different mechanisms for the targeting of poor households for social welfare programs, with outcome measures including per capita consumption of beneficiaries, but excluding health outcomes.

The resulting articles were screened by reviewing abstracts and full texts. Screening was conducted by one researcher (SK) for the 3ie search and was split between three researchers (NK, AM, JLK) for the Ovid Medline search. We retained only those studies that conducted: cost-benefit analysis (CBAs), cost-effectiveness analysis (CEAs) (including cost-utility analysis), cost-consequences analysis (CCA), or cost-minimisation analysis (CMAs); following the categories of economic evaluations outlined in Drummond *et al.* (2015). Descriptions of CEAs and CBAs can be found. For the purposes of this review we

¹ As a journal article, working paper, or report (by e.g. 3ie, World Bank)

² 3ie aims to update the database every 2 years.

Table 1. Exclusion and exclusion criteria employed in the study.

Criteria	Include	Exclude
Location/Setting	Low- and Middle-Income Country (World Bank Income Classification)	High-Income Country (World Bank Income Classification)
Impact evaluation criterion (1): microdata used	Treatment effect estimated within study through analysis of microdata for all comparator groups	Treatment effect estimated externally (using a published estimate) or through meta-analysis; treatment effect simulated (microdata only for one arm of RCT or observational study)
Impact evaluation criterion (2): study design	Experimental (<i>pragmatic / field / real-world RCTs</i>), and quasi-experimental (<i>DiD, IV, PSM, RDD</i>) designs	RCTs for biomedical efficacy, other study types (i.e. uncontrolled <i>before and after</i> , <i>uncontrolled observational study</i>), and
Type of economic evaluation	Cost-effectiveness, cost-benefit, cost-consequence, cost minimisation Return on investment	Other frameworks (<i>MCDA</i>) or no economic evaluation framework
Outcomes	Direct health outcomes (<i>mortality, quality of life, self-assessed health</i>), intermediate health outcomes (<i>utilisation, behaviour change</i>); Monetary valuations of health (CBA, ROI, i.e. VSL, earnings, health-related productivity costs)	Outcomes not directly linkable to health (<i>consumption, educational attainment</i>)*
Costs	Estimation of partial/full economic costs and/or opportunity costs	No consideration of costs or of intervention costs only
Language	English language articles only	Non-English language (based on 3ie inclusion criteria)

*While there is ample research that links items such as consumption and education to health, we define 'direct' linkage as relating to proximal determinants of health such as access, utilisation and health behaviour.

classified any papers that use measures of health as their primary outcomes as CEAs and those that value outcomes monetarily, usually based on some notion of the 'value' of varied health effects, as CBAs. CMAs estimate outcomes in various ways but their central findings and recommendations are based only on minimizing costs. We included any studies where incremental costs and effects were combined either in ratios, net benefits or were presented side-by-side in a cost-consequences style³. We excluded studies that only presented immediate direct intervention costs (e.g. unit price of a diagnostic test kit) without assessing full delivery costs (e.g. human resource costs) and/or future related costs (e.g. of treatment).

Quality assessment

We assessed the quality of economic evaluation studies using criteria adjusted from 11 principles of the iDSI Reference Case (Wilkinson *et al.*, 2016). We used the 11 Reference Case principles to construct a set of criteria for the quality assessment of the retrieved studies. We excluded one principle on 'evidence' – that "an economic evaluation should consider all available evidence relevant to the decision problem" – as this was not deemed applicable for impact evaluations that generally rely on a single study. Other iDSI principles were merged to create criteria to better fit impact evaluations. The final set of criteria is shown in Table 2, including (1) the transparent statement of

the decision problem, and the inclusion of appropriate comparators; (2) details of the analysis, including the measure of health outcomes, time horizon and discounting; (3) the perspective of the study; (4) costs; (5) heterogeneity; (6) uncertainty; (7) constraints; and (8) equity.

We recorded a range of descriptive information from each study: year and journal of publication, country location of the intervention, type of health intervention, impact evaluation method, the type of economic evaluation and the health outcomes measured. For each study, at least two reviewers conducted the assessment and any discrepancies were resolved through discussion with the wider team.

Results

Literature search results

The 3ie impact evaluation repository contained 2,419 impact evaluations in the Health Nutrition and Population sector, published between 2000 and 2016. Of these, more than half employed an experimental design (n=1,313), while the studies with non-experimental design applied various econometric methods, including DiD (n=166), PSM (n=135), IV (n=71), RDD (n=20) and other approaches⁴ (n=715). Of the entire set of Health, Nutrition and Population studies, we found only 117 mentioned costs and, following screening of abstracts and full

³ <https://www.nice.org.uk/Media/Default/guidance/LGB10-Briefing-20150126.pdf>

⁴ Including regression and multivariate matching methods, but in many cases labelled as "not specified" by 3ie.

Table 2. Reference case principles and assessment criteria.

Principle	iDSI Reference Case Summary	Assessment criteria
(1) Decision problem		
<i>Transparency</i>	An economic evaluation should be communicated clearly and transparently to enable the decision-maker(s) to interpret the methods and results.	The decision problem should be fully and accurately described.
<i>Comparators</i>	The comparator(s) against which costs and effects are measured should accurately reflect the decision problem.	The intervention(s) that are currently offered to the population should be the base case comparator.
(2) Analysis		
<i>Measure of health outcome</i>	The measure of health outcome should be appropriate to the decision problem, should capture positive and negative effects on length of life and quality of life, and should be generalizable across disease states.	Disability-Adjusted Life Years (DALYs) averted or other generic measures that capture length and quality of life should be used (e.g. the QALY).
<i>Time horizon</i>	The time horizon used in an economic evaluation should be of sufficient length to capture all costs and effects relevant to the decision problem	Lifetime time horizon should be used. Shorter time horizons can be used where it is shown that all costs and effects that are relevant to the decision problem have been captured.
<i>Discounting</i>	Appropriate discount rate should be used to discount costs and effects to present value.	An annual discount rate for costs and effects should be used. When time horizon is greater than 30 years, the impact of lower discount rates should be explored.
(3) Perspective	Non-health effects and costs associated with the health interventions that don't accrue to the health budget should be identified where relevant to the decision problem. All costs and effects should be disaggregated, either by sector of the economy or to whom they accrue.	<ol style="list-style-type: none"> 1. The perspective of the study should be described. 2. Analysis should reflect direct health costs and health outcomes. 3. Additional analysis should adopt a disaggregated societal perspective, to include non-health effects and costs that fall outside the health budget.
(4) Costs	All differences between the intervention and the comparator in expected resource use and costs of delivery to the target population(s) should be incorporated into the evaluation.	<ol style="list-style-type: none"> 1. Costs of all resources relevant to the decision problem should be considered. 2. Cost implications of a rollout of the program to the population should be considered. 3. Out of pocket payments are considered.
(5) Uncertainty	The uncertainty associated with an economic evaluation should be appropriately characterised.	<p>The economic evaluation should explore uncertainty:</p> <ol style="list-style-type: none"> 1. In the structure of the analysis in the economic evaluation 2. Due to source of parameters, and/or precision in the estimation of parameters of the economic evaluation (e.g. one sensitivity analysis, probabilistic sensitivity analysis).
(6) Heterogeneity	The cost and effects of the intervention on sub-populations within the decision problem should be explored and the implications appropriately characterised.	Heterogeneity in cost-effectiveness should be explored in population subgroups, where subgroup formation is justified by the evidence base regarding differences in relative costs and effects, and the influence on absolute effects.
(7) Constraints	The impact of implementing the intervention on the health budget and on other constraints should be identified clearly and separately.	Budget impact analysis should be performed that provides an estimate of the implications of implementing the intervention on various budgets, or an empirical CEA threshold.
(8) Equity	An economic evaluation should explore the equity implications of implementing the intervention.	Equity implications should be considered at all stages of the economic evaluation, including design, analysis and reporting.

texts, only 42 of these conducted an economic evaluation. In our Ovid MEDLINE search, 380 studies were identified. Studies were excluded at the full text review stage typically due to failing to qualify as an impact evaluation, either because it did not estimate the effectiveness of the intervention (31 excluded studies) or did not employ an experimental or quasi-experimental method (18 excluded studies). See [Figure 1](#) for a flow chart of the search and screening process.

The remaining core body of 70 studies (see [Table 4a](#) and [Table 4b](#)) were published in journals from a variety of fields, including medical and public health, health policy, and economics journals and in impact evaluation reports by 3ie journals. Publications covered many geographical areas; with a majority focused on Sub-Saharan Africa and several on Central and South America and East Asia. The interventions ranged from health interventions, including disease prevention and treatment (e.g. intermittent preventive treatment [IPT] for malaria or a school-based HIV education programme), health services (e.g. health facility improvement), health promotion (e.g. commitment devices for smoking cessation), to non-health interventions with potential health-improving impacts (e.g. a national social fund for development).

The majority (n=56) of studies employed randomised designs and, of the non-experimental studies, four conducted a PSM analysis as a primary statistical approach, four used DiD designs, one performed IV estimation, one employed an interrupted time series design, while four further studies used other controlled

observation designs (e.g. covariate matching). Economic evaluations were mostly CEAs (n=54), with a minority of CBAs (n=11), CMAs (n=3) and CCAs (n=2).

Results of the quality assessment

[Table 3](#) summarises the results of the quality assessment. Here we discuss how the published studies fared against each principle and highlight some overarching patterns.

Decision problem. Transparent description of the decision problem requires describing the need for a policy decision, accurate description of the comparators and target population and stating the evaluation perspective. We found 50 studies (71%) clearly described the decision problem, as e.g. [Alfonso et al. \(2015\)](#) who framed their decision problem as the comparison of a voucher scheme combined with obstetrical quality improvements to the status quo. The nature of the review, where we conditioned on first conducting an impact evaluation, meant all included studies considered a comparator; most often (n=55) this was 'do nothing' or the status quo.

The nature of the review, where we conditioned on first conducting an impact evaluation, meant that all included studies had considered a comparator, and most often (n=55) this comparator was the 'do nothing' or status quo. An example of a study that did not consider the status quo was [Duflo et al. \(2007\)](#). While in the impact evaluation they compared five different types of education interventions including the current national programme (current practice) for HIV/AIDS prevention, in the

Figure 1. Flowchart of search strategy and results.

Table 3. Summary results of the quality assessment.

Criterion	Met		Not met	
Decision problem				
Clearly stated	50	(71%)	20	(29%)
Appropriate comparators	55	(79%)	15	(21%)
Analysis				
Health measure (QALY, DALY or similar)	20	(29%)	50	(71%)
Time horizon	23	(33%)	47	(67%)
Discounting	29	(41%)	41	(59%)
Perspective				
Stated	36	(51%)	34	(49%)
Health sector perspective considered	49	(70%)	21	(30%)
Societal perspective considered	26	(37%)	44	(63%)
Costs				
Relevant costs	49	(70%)	21	(30%)
Scale-up costs	17	(24%)	53	(76%)
OOP expenditure included	15	(21%)	55	(79%)
Heterogeneity				
in cost-effectiveness explored	6	(9%)	64	(91%)
Uncertainty				
Structural	6	(9%)	64	(91%)
Parameter	30	(43%)	40	(57%)
Constraints				
Budget impact analysis	3	(4%)	67	(96%)
Equity considerations	0	(0%)	70	(100%)

economic evaluation they compared the incremental cost per teen pregnancy averted only among the comparator interventions (not considering the current practice as a comparator).

Analysis

Measure of health outcome. The assessment criteria highlighted the importance of using generic health measures such as DALYs or QALYs, incorporating both morbidity and mortality consequences of an intervention and to facilitate cross-program comparisons. However, only 20 studies used a generic health measure; 18 using DALYs and two using QALYs. The remaining studies used a wide range of outcome measures, including narrower health indicators (e.g. life years saved, depression free days, malaria detected), and behavioural change (quitting smoking, number of new adopters attributed to a campaign). The studies that conducted CBA measured health effects and

then converted them to monetised benefits (e.g. wage benefit due to effective treatment, value of statistical life saved).

Time horizon. The assessment criteria considered the lifetime horizon as a gold standard and required a clear justification for a shorter time horizon. The majority of studies (n=47) did not follow this requirement and used either a shorter time horizon without any justification or failed to explicitly state the time horizon at all. Examples include studies that maintained the time horizon of an RCT and stuck to measured, intermediate outcomes, for example accuracy of a diagnosis, as reported by Bualombai *et al.* (2003). An example of a study that considered a longer time horizon is the evaluation of a pre-school intervention (Behrman *et al.*, 2004), where the short term impact of the intervention, a gain in height, was translated into wages gained later in life via a cost-benefit analysis.

Discounting. Less than half of the studies (n=29) applied a discount rate in their economic evaluation (or justified the use of zero discount rate), and typically applied the same rate to discount costs and effects. Many studies use discount rates that come from recommended guidelines for economic evaluation, usually at the level of 3% or 5%. For instance, Nonvignon *et al.* (2012) used a discount rate of 3%, then varied the rate in sensitivity analysis.

Perspective. Around half of studies (n=36) clearly stated the perspective used. For example, Alfonso *et al.* (2015) outline how they conducted the evaluation both from the societal and medical sector perspectives. We found that, for the majority of the studies (n=49), the health services perspective was considered in the primary analysis; although there were other examples, such as Shepard *et al.* (2003) that took a donor perspective (USAID). Another 26 studies went further, however, and adopted a wider societal perspective, either by considering costs beyond the health budget, or some non-health benefits of the intervention (e.g. opportunity cost of waiting time, travel fees, wage loss, out of pocket payments, and education benefits); although these choices were rarely justified. CBA that valued life years saved using the value of statistical life (VSL), or lifetime earnings, also implicitly considered benefits beyond the health care sector. An example that incorporates some of these approaches is the study by Abou-Ali *et al.* (2010), evaluating the Egyptian Social Fund for Development, a complex nationwide inter-sectoral policy initiative including interventions in education, health, sanitation and microcredit.

Costs. The assessment criteria required that all resource implications relevant to the decision problem were counted and costed, and the implications of a potential rollout or scale-up of a program had been considered. Overall, 49 studies were deemed to have incorporated all relevant costs, although we had to rely on what was reported, which was challenging and may thus likely be an overestimate. A minority of the studies (n=17) considered a potential scale up of a programme, as e.g. Jan *et al.* (2010): a microfinance intervention to address the partner violence problem in South Africa is assessed based initially on the implementation costs of the programme, with exploration of

Table 4a. Articles included for quality assessment: 31e search.

Authors	Year	Journal, issue	Title
Abou-Ali, Hala, <i>et al.</i>	2010	Journal of Development Effectiveness, 2.4:521–555	Evaluating the impact of Egyptian social fund for development programmes
Alfonso, Y. Natalia, <i>et al.</i>	2013	Health Policy and Planning, 30.1: 88–99.	Cost-effectiveness analysis of a voucher scheme combined with obstetrical quality improvements: quasi experimental results from Uganda
Aracena, Marcela, <i>et al.</i>	2009	Journal of Health Psychology, 14.7 : 878–887.	A cost-effectiveness evaluation of a home visit program for adolescent mothers.
Araya R, <i>et al.</i>	2006	American journal of psychiatry, 163(8):1379–87.	Cost-effectiveness of a primary care treatment program for depression in low-income women in Santiago, Chile
Ashraf, N. <i>et al.</i>	2013	3ie Impact Evaluation Report, 9	"No margin, no mission." Evaluating the role of incentives in the distribution of public goods in Zambia
Banerjee, A. <i>et al.</i>	2010	British Medical Journal, 340 c2220.	Improving immunisation coverage in rural India: clustered randomised controlled evaluation of immunisation campaigns with and without incentives
Barasa, EW. <i>et al.</i>	2012	PLoS medicine 9.6 e1001238.	A multifaceted intervention to improve the quality of care of children in district hospitals in Kenya: a cost-effectiveness analysis.
Barham, T.	2011	Journal of Development Economics , 94.1 74–85.	A healthier start: the effect of conditional cash transfers on neonatal and infant mortality in rural Mexico.
Behrman JR, <i>et al.</i>	2004	The Review of Economics and Statistics. Feb;86(1):108–32.	Evaluating preschool programs when length of exposure to the program varies: A nonparametric approach.
Bhatia MR, <i>et al.</i>	2004	Social Science & Medicine, 31;59(3):525–39.	Cost-effectiveness of malaria control interventions when malaria mortality is low: insecticide-treated nets versus in-house residual spraying in India. 2004 Aug
Bojang, Kalifa A., <i>et al.</i>	2011	PLoS medicine 8.2 e1000409.	Two strategies for the delivery of IPTc in an area of seasonal malaria transmission in The Gambia: a randomised controlled trial.
Briceño, Bertha, and Claire Chase. 7.4: 423–434.	2015	Journal of Development Effectiveness	Cost-efficiency of rural sanitation promotion: activity-based costing and experimental evidence from Tanzania
Bualombai, P., <i>et al.</i>	2003	In: Briggs, A., K. Claxton, and M. Sculpher, Decision Modelling for Health Economic Evaluation. 2006, Oxford: Oxford University Press.	Determining cost-effectiveness and cost component of three malaria diagnostic models being used in remote non-microscope areas
Cerda, R., <i>et al.</i>	2011	The International Journal of Tuberculosis and Lung Disease , 15.3 363–368.	Health care utilization and costs of a support program for patients living with the human immunodeficiency virus and tuberculosis in Peru.
Cohen, J and Dupas P.	2010	The Quarterly Journal of Economics 1–45.	Free distribution or cost-sharing? Evidence from a randomized malaria prevention experiment
Colvin M, <i>et al.</i>	2006	Sexually Transmitted Infections, 1;82(4):290–4	Effectiveness and cost effectiveness of syndromic sexually transmitted infection packages in South African primary care: cluster randomised trial. 2006 Aug
Davis, Wendy A., <i>et al.</i>	2011	Bulletin of the World Health Organization 89.3: 211–220	Cost-effectiveness of artemisinin combination therapy for uncomplicated malaria in children: data from Papua New Guinea.
Duflo E, <i>et al.</i>	2007	Background paper to the 2007 World Development Report	Education and HIV/AIDS prevention: evidence from a randomized evaluation in Western Kenya.
Fairall, L, <i>et al.</i>	2010	Tropical Medicine & International Health 15.3 277–286	Cost-effectiveness of educational outreach to primary care nurses to increase tuberculosis case detection and improve respiratory care: economic evaluation alongside a randomised trial

Authors	Year	Journal, issue	Title
Freudenberg S, <i>et al.</i>	2004	World Journal of Surgery, 28(4):421–4	Fishing Line Suture: Cost-saving Alternative for Atraumatic Intracutaneous Skin Closure—Randomized Clinical Trial in Rwanda.
Giné, X, <i>et al.</i>	2010	American Economic Journal: Applied Economics, 2(4): 213–35.	Put Your Money Where Your Butt Is: A Commitment Contract for Smoking Cessation.
Jan, S, <i>et al.</i>	2010	Health policy and planning 26.5: 366–372.	Economic evaluation of a combined microfinance and gender training intervention for the prevention of intimate partner violence in rural South Africa
José Díaz, J, and Jaramillo M.	2009	Journal of Development Effectiveness 1.4 (2009): 387–412.	Evaluating interventions to reduce maternal mortality: evidence from Peru's PARSalud programme.
Kincaid DL, Do MP.	2006	Journal of Health Communication, Feb 1; 11(52):69–90.	Multivariate causal attribution and cost-effectiveness of a national mass media campaign in the Philippines.
Leong K, <i>et al.</i>	2006	Family Practice, 17;23(6):699–705.	The use of text messaging to improve attendance in primary care: a randomized controlled trial.
Manandhar DS, <i>et al.</i>	2004	The Lancet. 17;364(9438):970–9.	Effect of a participatory intervention with women's groups on birth outcomes in Nepal: cluster-randomised controlled trial.
Michaels-Igbokwe, C., <i>et al.</i>	2016	BMC Public Health , 16.1: 196	Cost and cost-effectiveness analysis of a community mobilisation intervention to reduce intimate partner violence in Kampala, Uganda.
Miguel E, Kremer M.	2004	Econometrica, 1;72(1):159–217.	Worms: identifying impacts on education and health in the presence of treatment externalities.
Muñoz, K. <i>et al.</i>	2016	3 <i>ie. Impact Evaluation Report</i> 26.	Validation of hearing screening procedures in Ecuadorian school.
Nizalova, OY, and Vyshnya M.	2010	Health economics 19.S1: 107–125	Evaluation of the impact of the Mother and Infant Health Project in Ukraine.
Nonvignon, J, <i>et al.</i>	2012	Tropical Medicine & International Health 17.8: 951–957.	Is home management of fevers a cost-effective way of reducing under-five mortality in Africa? The case of a rural Ghanaian District
Patel AB, <i>et al.</i>	2003	Cost Effectiveness and Resource Allocation. Aug 29;1(1):7.	Economic evaluation of zinc and copper use in treating acute diarrhea in children: A randomized controlled trial.
Pereira, SM., <i>et al.</i>	2012	The Lancet infectious diseases 12.4:300–306.	Effectiveness and cost-effectiveness of first BCG vaccination against tuberculosis in school-age children without previous tuberculin test (BCG-REVAC trial): a cluster-randomised trial.
Ramachandran A, <i>et al.</i>	2007	Diabetes Care 1;30(10):2548–52.	Cost-effectiveness of the interventions in the primary prevention of diabetes among Asian Indians.
Sedlmayr, R, <i>et al.</i>	2013	Malaria Journal, 12.: 102.	Health impact and cost-effectiveness of a private sector bed net distribution: experimental evidence from Zambia
Simwaka BN, <i>et al.</i>	2009	Journal of Development effectiveness, 1(4):492–506.	Retrospective analysis of a school-based malaria treatment programme demonstrates a positive impact on health and education outcomes in Mangochi district, Malawi.
Solon, Orville, <i>et al.</i>	2009	Health Policy 92.1:89–95.	An evaluation of the cost-effectiveness of policy navigators to improve access to care for the poor in the Philippines
Subramanian S, <i>et al.</i>	2009	Bulletin of the World Health Organization.;87(3):200–6.	Cost-effectiveness of oral cancer screening: results from a cluster randomized controlled trial in India.
Thirumurthy H, <i>et al.</i>	2008	Journal of Human Resources. Jul 1; 43(3):511–52.	The economic impact of AIDS treatment labor supply in Western Kenya.
Tripathy, P, <i>et al.</i>	2010	The Lancet 375.9721: 1182–1192.	Effect of a participatory intervention with women's groups on birth outcomes and maternal depression in Jharkhand and Orissa, India: a cluster-randomised controlled trial.
Tun-Lin, W., <i>et al.</i>	2009	Tropical Medicine & International Health 14.9 1143–1153	Reducing costs and operational constraints of dengue vector control by targeting productive breeding places: a multi-country non-inferiority cluster randomized trial

Table 4b. Articles included for quality assessment: Ovid Medline search.

Anchala, R., <i>et al.</i> ,	2015	Journal of the American Heart Association, 4(1): p. e001213.	Evaluation of effectiveness and cost-effectiveness of a clinical decision support system in managing hypertension in resource constrained primary health care settings: results from a cluster randomized trial.
Ar'kov, O.Y., <i>et al.</i>	2011	Applied Health Economics & Health Policy, 9(2): p. 89–99.	Influenza vaccination in healthy working adults in Russia: observational study of effectiveness and return on investment for the employer.
Bergmann, J.N., <i>et al.</i>	2017	AIDS & Behavior, 21(3): p. 703–711.	Outcomes and Cost-Effectiveness of Integrating HIV and Nutrition Service Delivery: Pilots in Malawi and Mozambique.
Bigna, J.J., <i>et al.</i>	2014	The Lancet Infectious Diseases, 14(7): p. 600–8.	Effect of mobile phone reminders on follow-up medical care of children exposed to or infected with HIV in Cameroon (MORE CARE): a multicentre, single-blind, factorial, randomised controlled trial.
Bustamante Valles, K., <i>et al.</i>	2016	Journal of Neuroengineering & Rehabilitation, 13(1): p. 83.	Technology-assisted stroke rehabilitation in Mexico: a pilot randomized trial comparing traditional therapy to circuit training in a Robot/technology-assisted therapy gym.
Buttorff, C., <i>et al.</i>	2012	Bulletin of the World Health Organization, 90(11): p. 813–21.	Economic evaluation of a task-shifting intervention for common mental disorders in India.
Dangour, A.D., <i>et al.</i>	2011	PLoS Medicine / Public Library of Science, 8(4): p. e1001023.	Effect of a nutrition supplement and physical activity program on pneumonia and walking capacity in Chilean older people: a factorial cluster randomized trial.
Ezenduka, C., <i>et al.</i>	2012	PLoS Neglected Tropical Diseases, 6(9): p. e1818.	Cost-effectiveness analysis of three leprosy case detection methods in Northern Nigeria.
Fottrell, E., <i>et al.</i>	2013	JAMA Pediatrics, 167(9): p. 816–25.	The effect of increased coverage of participatory women's groups on neonatal mortality in Bangladesh: A cluster randomized trial.
Gebremedhin, S., <i>et al.</i>	2016	BMC Public Health, 16: p. 457	The effectiveness bundling of zinc with Oral Rehydration Salts (ORS) for improving adherence to acute watery diarrhea treatment in Ethiopia: cluster randomised controlled trial.
Gulluoglu, B.M., <i>et al.</i>	2013	Annals of Surgery, 257(1): p. 37–43.	Efficacy of prophylactic antibiotic administration for breast cancer surgery in overweight or obese patients: a randomized controlled trial.
Habyarimana, J. and W. Jac	2015	Proceedings of the National Academy of Sciences of the United States of America, 112(34): p. E4661–70.	Results of a large-scale randomized behavior change intervention on road safety in Kenya.
Hanson, K., <i>et al.</i>	2003	Bulletin of the World Health Organization, 81(4): p. 269–76.	Cost-effectiveness of social marketing of insecticide-treated nets for malaria control in the United Republic of Tanzania.
McBain, R.K., <i>et al.</i>	2016	Health Policy & Planning, 31(4): p. 415–24.	Costs and cost-effectiveness of a mental health intervention for war-affected young persons: decision analysis based on a randomized controlled trial.
Mendoza-Cano, O., <i>et al.</i>	2017	International Journal of Environmental Research & Public Health, 14(8): p. 08.	Cost-Effectiveness of the Strategies to Reduce the Incidence of Dengue in Colima, Mexico.
Meng, L., <i>et al.</i>	2013	PLoS ONE, 8(10): p. e77971.	The costs and cost-effectiveness of a school-based comprehensive intervention study on childhood obesity in China.
Nadkarni, A., <i>et al.</i>	2017	Lancet, 389(10065): p. 186–195.	Counselling for Alcohol Problems (CAP), a lay counsellor-delivered brief psychological treatment for harmful drinking in men, in primary care in India: a randomised controlled trial.

Obireli-Neto, P.R., <i>et al.</i>	2015	Journal of Managed Care & Specialty Pharmacy, 21(1): p. 66–75.	Economic evaluation of a pharmaceutical care program for elderly diabetic and hypertensive patients in primary health care: a 36-month randomized controlled clinical trial.
Obure, C.D., <i>et al.</i>	2017	Sexually Transmitted Infections, 93(7): p. 482–486.	A comparative analysis of costs of single and dual rapid HIV and syphilis diagnostics: results from a randomised controlled trial in Colombia.
Patel, V., <i>et al.</i>	2017	Lancet, 389(10065): p. 176–185.	The Healthy Activity Program (HAP), a lay counsellor-delivered brief psychological treatment for severe depression, in primary care in India: a randomised controlled trial.
Pereira, S.M., <i>et al.</i>	2012	The Lancet Infectious Diseases, 12(4): p. 300–6.	Effectiveness and cost-effectiveness of first BCG vaccination against tuberculosis in school-age children without previous tuberculin test (BCG-REVAC trial): a cluster-randomised trial.
Rashid, R.M., <i>et al.</i>	2014	Asian Pacific Journal of Cancer Prevention: Apjcp, 15(13): p. 5143–7.	Cost effective analysis of recall methods for cervical cancer screening in Selangor--results from a prospective randomized controlled trial.
Shepard, D.S., <i>et al.</i>	2003	Studies in Family Planning. 34(2): p. 117–26.	Cost-effectiveness of USAID's regional program for family planning in West Africa.
Stanback, J., <i>et al.</i>	2007	International Journal for Quality in Health Care, 19(2): p. 68–73.	Improving adherence to family planning guidelines in Kenya: an experiment.
Stevenson, M., <i>et al.</i>	2008	Injury Prevention, 14(5): p. 284–9.	Reducing the burden of road traffic injury: translating high-income country interventions to middle-income and low-income countries.
Sun, X., <i>et al.</i>	2015	Trials, 16: p. 496.	The cost-effectiveness analysis of JinQi jiangtang tablets for the treatment on prediabetes: a randomized, double-blind, placebo-controlled, multicenter design.
Sweat, M., <i>et al.</i>	2000	Lancet, 356(9224): p. 113–21.	Cost-effectiveness of voluntary HIV-1 counselling and testing in reducing sexual transmission of HIV-1 in Kenya and Tanzania.
Weobong, B., <i>et al.</i>	2017	PLoS Medicine / Public Library of Science, 14(9): p. e1002385.	Sustained effectiveness and cost-effectiveness of the Healthy Activity Programme, a brief psychological treatment for depression delivered by lay counselors in primary care: 12-month follow-up of a randomised controlled trial.

likely programme scale up costs. By indicating the decrease in the per capita cost due to economies of scale, the study concluded that the cost per DALY averted would be lower when scaled up beyond the pilot programme.

Heterogeneity. Only six studies considered heterogeneity in the cost-effectiveness estimates. One of these, [Subramanian *et al.* \(2009\)](#), implemented a CEA of visual screening for oral cancer detection and showed that cost-effectiveness was better for high risk individuals rather than the overall population of interest, due to higher per-case health gains, even though the cost-per-case was also higher. Typically, evaluations tended to explore heterogeneity in the impact evaluation estimates, for example by conducting subgroup analysis, or estimating other forms of treatment effect heterogeneity, but did not take this forward to the economic evaluation component of the study. For example, [Barham \(2011\)](#) estimates heterogeneous impacts of the PROGRESA conditional cash transfer program on infant mortality, by pre-intervention municipality characteristics including access to piped water or proportion of illiteracy. However, when conducting their CBA the authors only employed the overall impact estimate.

Uncertainty. The iDSI Reference Case recommends that economic evaluations systematically explore all sources of uncertainty, including choices in the structure of an analysis and precision of estimated parameters. A relevant source of structural uncertainty in impact evaluations is the choice and specification of the econometric method used to adjust for confounding. We found that while authors often carefully explored the implications of these methodological choices on their impact estimates, this approach was rarely ($n=6$) taken forward to the economic evaluation stage. For example, [Barham \(2011\)](#) reported impact estimates from a wide range of model specifications, including different definitions of the intervention variable, and different sets of control variables. They found qualitatively similar results across specifications and chose to use the lowest of the estimates in the cost-benefit analysis. [Abou-Ali *et al.* \(2010\)](#) took a somewhat different approach: they use three different statistical approaches to obtain impact estimates: regression, nearest neighbour matching and kernel matching; and they reported separate cost-benefit estimates based on each of the impact estimation method. However, for the economic evaluation they only made use of those estimates that were found statistically significant. This latter choice omits uncertainty attributable to the degree of precision in the estimation of the impact parameter – a pattern that we find in many of the included studies: less than half of the studies ($n=30$) took into account any kind of parameter uncertainty, with only few studies reporting probabilistic sensitivity analysis. Characterising uncertainty due to assumptions on the parameters was more common, as in [Alfonso *et al.* \(2015\)](#) and [Michaels-Igbokwe *et al.* \(2016\)](#), who presented tornado diagrams to investigate the sensitivity of the cost-effectiveness parameter by varying assumptions on relevant parameters.

Constraints. Our criteria related to constraints focused primarily on budget impact due to its importance for health financing, although this is not the only constraint that could be considered.

Only three studies performed an explicit budget impact analysis, among which [Barasa *et al.* \(2012\)](#), in their evaluation of a hospital improvement intervention in Kenya, estimate the budget attributable to the scale up of the intervention, and compare this to the annual health budget of the country. Several papers attempted to give an estimate of the budget impact by providing the cost per a given administrative unit. For example, an analysis by [Simwaka *et al.* \(2009\)](#) provided the cost per student in a school-based malaria programme.

Opportunity costs that fall on health budgets are sometimes reflected in economic analysis through the cost-effectiveness threshold ([Drummond *et al.*, 2015](#)), although the use of an explicit cost-effectiveness threshold such as the UK one is rare in LMICs, due in part to the lack of empirical evidence on their opportunity costs. Several of those studies that estimated cost-effectiveness in terms of DALYs compared the resulting ICERs to thresholds derived from the GDP per capita of a given country (following a traditional and by now largely disowned recommendation by the World Health Organization; [Bertram *et al.*, 2016](#)) rather than a threshold reflecting, at least in part, an appropriate measure of opportunity cost.

Equity. While several impact evaluation studies have implicitly touched upon the equity principle, by either evaluating a programme specifically designed for a deprived population (e.g. [Behrman *et al.*, 2004](#)) or by conducting subgroup analysis by levels of deprivation (e.g. [Barham, 2011](#)), only one study formally incorporated equity in the impact evaluation. [Abou-Ali *et al.* \(2010\)](#) examined the distribution of resources of the Egyptian Social Fund by sector, using a Lorenz Curve analysis. They found that more funds for education and wastewater programmes were allocated disproportionately in favour of the relatively wealthy income group. In contrast, other programmes (e.g. portable water, health and micro credit) allocated relatively more funds for the poor income group. However, this analysis did not extend to assessing the distributional impacts also in relation to cost-effectiveness.

Discussion Summary

The main finding from our study is that too few published impact evaluations include a full economic evaluation, and those that do, have economic evaluations of variable quality. When searching within the 3ie database, we found that among the 2,419 published impact evaluations in the “Health, Nutrition and Population” category, only a small fraction, i.e. $n=42$ (2%), had made an attempt to conduct an economic evaluation. Based on the complementary Ovid MEDLINE search, a further 28 studies passed the criteria of having both an impact evaluation and an economic evaluation component, resulting in a total of 70 studies. The quality of economic evaluations, when assessed against a set of criteria derived from the iDSI reference case, was found to vary greatly).

Explanation of our findings

This literature review shows two distinct types of impact evaluations with economic evaluation components that have major differences in their design, and which may explain some

of our findings. First, there are those typically published in economics journals – most of which were identified through the 3ie database – which positioned the economic evaluation as a relatively small part of the overall work. In these studies, typically using non-experimental designs (e.g. Barham, 2011; Behrman *et al.*, 2004; Cohen & Dupas, 2010; Giné *et al.*, 2010; Miguel & Kremer, 2004; Nizalova & Vyshnya), the impact evaluation was generally conducted using highly sophisticated econometric methods, addressing the heterogeneity of programme impacts and the sensitivity of the results to choices of the econometric specification. However, the same level of sophistication was not normally applied to the accompanying economic evaluation, which often relied on just one point-estimate and did not systematically consider uncertainty or the impact of different methodological choices on the estimates. Analyses of costs were typically ‘back-of-the-envelope’ and were not sufficient to provide decision makers with a reliable picture of program cost-effectiveness. The effects of different assumptions regarding the input parameters and the resulting cost-effectiveness estimates were likewise not assessed. While the lack of availability of detailed cost information and long-term outcomes may in part explain the choices above, even with such constraints, it should be possible to provide a decision maker with a fuller picture on the expected cost-effectiveness of such programs. A helpful example is a study by Cohen & Dupas (2010), published in the Quarterly Journal of Economics, evaluating different levels of cost-sharing when distributing insecticide treated bed-nets for malaria prevention. The authors showed their results in a cross tabulation where their key assumptions were varied. Often, these studies defined their research question as testing an economic or behavioural hypothesis (e.g. How do people respond to incentives? Does early education work?), and not as a decision problem for resource allocation. Hence, the assessment criteria requiring a definition of the decision problem as well as the comparators often fell short in these studies.

The second type of studies can be best described as “within-trial” economic evaluations (Sculpher *et al.*, 2006). In these, the impact evaluations and economic evaluations were typically conducted as part of an evaluation of an RCT, often with detailed information on costs. While these studies were often precise in defining the decision problem, the comparators, and even the perspective of the study, they fell short with respect to other important assessment criteria. Notably, the studies rarely looked beyond the trial: short time horizons were used, the outcomes of interest were typically intermediate outcomes measured in the trial instead of generic health measures (e.g. DALYs), and sensitivity analysis was rarely comprehensive.

Future research

One objective of this work was to generate a discussion of methodological gaps that could guide future research on how to better combine economic evaluations and impact evaluations. The need and value for the fields coming together has also been noted elsewhere. McEwan (2012), for example, outlines the main requirements for CEA and CBA in the health and education sectors; Dhaliwal *et al.* (2013), propose a framework to synthesize published impact estimates to compare the cost-effectiveness of a range of interventions in the

educations sector; and Evans & Popova (2014) assess how uncertainty can be reflected and suggest the use of probabilistic sensitivity analysis in impact evaluation. In the UK, researchers have also noted the emphasis on effectiveness studies for evaluating pay-for-performance and have called for the accompanying assessment of cost-effectiveness (Meacock *et al.*, 2014; Meacock, 2019).

There have been methods advancements in both the fields of impact evaluation and economic evaluation that could lead to cross-fertilization and merging of approaches. An important concern is quantifying generalisability (or ‘external validity’), in which analyses undertaken for one setting can inform assessments in another (Dhaliwal *et al.*, 2013; Drummond *et al.*, 2005; McEwan, 2012; Sculpher *et al.*, 2004; Vivaldi, 2016). One potential avenue to increase generalisability in impact evaluations is the careful definition of the parameter of interest (e.g. average treatment effects, average treatment effect on the treated, local average treatment effects), with the linked challenge of estimating heterogeneous treatment effects (Wager & Athey, 2017). Other advancements may be in the incorporation of spill-over effects and externalities (Miguel & Kremer, 2004), econometric modelling of the relationship between intermediate and long term outcomes (Athey *et al.*, 2016), and the use of machine learning to estimate treatment effects (Wager & Athey, 2017).

Our review has focused on single studies for which causal estimates have been obtained, but the way impact and economic evaluations are combined to meet the needs of policy-making is likely to involve the fuller use of decision-modelling (Briggs *et al.*, 2006). This can enable the synthesis of multiple forms of evidence, including effect estimates from multiple studies (Welton *et al.*, 2012); extrapolation of treatment effects and costs beyond a study’s follow-up period, as well as from intermediate effects to outcomes; and exploration of the consequences of heterogeneity, uncertainty, as well as changes in key parameters such as prices or effect sizes.

Decision-analytic modelling is now a much more common vehicle for economic evaluation than single-study cost-effectiveness analysis, but is still rarely used as an extension to impact evaluation studies. One reason could be that the interventions assessed in impact evaluations often have the characteristics of what have been described as “complex” interventions, consisting of one or multiple activities producing multiple outcomes (Masset *et al.*, 2018). Moreover, when evaluated at the level of jurisdictions, effects for many of these interventions may be dynamic (with externalities). The development of methods to address these issues, and how they can be modelled at the level of whole systems, is likely to be a research priority in the coming years. Future research could also better reflect the distributional aspects of policies on different socioeconomic groups (Cookson *et al.*, 2017; Welch *et al.*, 2017) and inform cost-effectiveness assessments costs and benefits fall across multiple sectors (Claxton *et al.*, 2010; Remme *et al.*, 2017).

Limitations

This review has been a first step in examining more closely the linkage (or lack thereof) between the literatures of impact

evaluation and economic evaluation. There are several limitations: First, as the literature search was for single studies, it inherently excluded studies that synthesised evidence from many sources (e.g. meta-analyses, systematic reviews). Second, while the inclusion criteria of the 3ie database and of our OVID MEDLINE search provided some quality assessment for the econometric methodology applied, an in-depth assessment of the applied econometrics methods was beyond the scope of this study.

Conclusion

The fields of impact evaluation and economic evaluation have largely developed separately; each to a high level of methodological sophistication. More efforts should now be directed towards bringing the two fields together, with a view to better informing resource allocation decisions in global health. Research funders, as well as national and international policy institutions, can play important roles in supporting the generation of new methods research to achieve this aim.

Data availability

Underlying data

All data underlying the results are available as part of the article and no additional source data are required.

Extended data

York Research Database: Data appendices to the paper “From impact evaluation to decision-analysis: assessing the extent and quality of evidence on ‘value for money’ in health impact evaluations in low- and middle-income countries”. <https://doi.org/10.15124/cd49d13f-0553-46e7-ab13-c726fc5c97e5> (Kreif, 2020).

Datadepo.zip contains the following extended data:

- 3ie_QA.csv. (Quality assessment of articles identified in 3ie.)
- Ovid_allscreen.csv. (All articles identified in the initial screen of Ovid.)
- Ovid_fulltext_excludes.csv. (Reasons for exclusion after full text review in Ovid search)
- Ovid_fulltext_screen.csv. (Articles that required full-text screening.)
- Ovid_QA.csv. (Quality assessment of articles identified in Ovid.)
- Ovid_search_terms.txt. (Search terms used in Ovid.)

Data are available under the terms of the [Creative Commons Attribution 4.0 International license](#) (CC-BY 4.0).

References

- Alatas V, Banerjee A, Hanna R, *et al.*: **Targeting the poor: evidence from a field experiment in Indonesia.** *Am Econ Rev.* 2012; **102**(4): 1206-40.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Athey S, Chetty R, Imbens G, *et al.*: **Estimating Treatment Effects using Multiple Surrogates: The Role of the Surrogate Score and the Surrogate Index.** *arXiv preprint.* arXiv: 1603.09326, 2016.
[Reference Source](#)
- Bertram MY, Lauer JA, De Joncheere K, *et al.*: **Cost-effectiveness thresholds: pros and cons.** *Bull World Health Organ.* 2016; **94**(12): 925.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Briggs A, Sculpher M, Claxton K: **Decision modelling for health economic evaluation.** Oxford. University Press. 2006.
[Reference Source](#)
- Cameron DB, Mishra A, Brown AN: **The growth of impact evaluation for international development: how much have we learned?** *J Dev Effect.* 2015; **8**(1): 1-21.
[Publisher Full Text](#)
- Claxton K, Palmer S, Sculpher M, *et al.*: **Appropriate perspectives for health care decisions.** *CHE Research Paper.* 2010; **54**.
[Reference Source](#)
- Cookson R, Mirelman AJ, Griffin S, *et al.*: **Using cost-effectiveness analysis to address health equity concerns.** *Value Health.* 2017; **20**(2): 206-12.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Dhaliwal I, Duflo E, Glennerster R, *et al.*: **Comparative cost-effectiveness analysis to inform policy in developing countries: a general framework with applications for education.** *Education Policy in Developing Countries.* 2013; **17**: 285-338.
[Reference Source](#)
- Drummond M, Manca A, Sculpher M: **Increasing the generalizability of economic evaluations: recommendations for the design, analysis, and reporting of studies.** *Int J Technol Assess Health Care.* 2005; **21**(2): 165-71.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Drummond MF, Sculpher MJ, Claxton K, *et al.*: **Methods for the economic evaluation of health care programmes.** Oxford University Press. 2015.
[Reference Source](#)
- Evans DK: **Why don't economists do cost analysis in their impact evaluations?** *The World Bank. Development Impact Blog.* 2016; Accessed 27 Nov. 17.
[Reference Source](#)
- Evans DK, Popova A: **Cost-effectiveness measurement in development: accounting for local costs and noisy impacts.** *World Bank Policy Research Working Paper No. 7027.* 2014.
[Publisher Full Text](#)
- Gertler PJ, Martinez S, Premand P, *et al.*: **Impact evaluation in practice.** World Bank Publications. 2016.
[Reference Source](#)
- iDSI **Reference Case for Economic Evaluation.** 2016.
[Reference Source](#)
- Kreif N: **Data appendices to the paper “From impact evaluation to decision-analysis: assessing the extent and quality of evidence on ‘value for money’ in health impact evaluations in low- and middle-income countries”.** York Research Database. 2020.
<http://www.doi.org/10.15124/cd49d13f-0553-46e7-ab13-c726fc5c97e5>
- Masset E, Mascagni G, Acharya A, *et al.*: **The Cost-Effectiveness of Complex Projects: A Systematic Review of Methodologies.** *iDSI Bulletin.* Brighton: IDS 2018.
[Publisher Full Text](#)
- McEwan PJ: **Cost-effectiveness analysis of education and health interventions in developing countries.** *J Dev Effect.* 2012; **4**(2): 189-213.
[Publisher Full Text](#)
- Meacock R: **Methods for the economic evaluation of changes to the organisation and delivery of health services: principal challenges and recommendations.** *Health Econ Policy Law.* 2019; **14**(1): 119-134.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Meacock R, Kristensen SR, Sutton M: **The cost-effectiveness of using financial incentives to improve provider quality: A framework and application.** *Health Econ.* 2014; **23**(1): 1-3.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Remme M, Martinez-Alvarez M, Vassall A: **Cost-Effectiveness Thresholds in Global Health: Taking a Multisectoral Perspective.** *Value Health.* 2017; **20**(4): 699-704.
[PubMed Abstract](#) | [Publisher Full Text](#)

Sculpher MJ, Claxton K, Drummond M, *et al.*: **Whither trial-based economic evaluation for health care decision making?** *Health econ.* 2006; **15**(7): 677-87. [PubMed Abstract](#) | [Publisher Full Text](#)

Sculpher MJ, Pang FS, Manca A, *et al.*: **Generalisability in economic evaluation studies in healthcare: a review and case studies.** *Health Technol Assess.* 2004; **8**(49): iii-iv, 1-192. iii-iv. [PubMed Abstract](#) | [Publisher Full Text](#)

Shi Y, Zhang L, Rozelle S: **When will we ever learn... to change policy: current state of impact evaluation.** *J Dev Effect.* 2015; **7**(4): 402-22. [Reference Source](#)

Vivalt E: **How much can we generalize from impact evaluations?** Working paper. 2016. [Reference Source](#)

Wager S, Athey S: **Estimation and inference of heterogeneous treatment**

effects using random forests. *J Am Stat Assoc.* 2017; **113**: 523, 1228-1242.

[Publisher Full Text](#)

Welch VA, Norheim OF, Jull J, *et al.*: **CONSORT-Equity 2017 Extension and elaboration for better reporting of health equity in randomised trials.** *BMJ.* 2017; **359**: j5085. [PubMed Abstract](#) | [Publisher Full Text](#)

Welton N, Sutton AJ, Cooper NJ, *et al.*: **Evidence Synthesis for Decision Making in Healthcare.** Chichester: Wiley. 2012. [Publisher Full Text](#)

Wilkinson T, Sculpher MJ, Claxton K, *et al.*: **The international decision support initiative reference case for economic evaluation: an aid to thought.** *Value Health.* 2016; **19**(8): 92. [PubMed Abstract](#) | [Publisher Full Text](#)

Open Peer Review

Current Peer Review Status: ? ✓

Version 1

Reviewer Report 22 February 2021

<https://doi.org/10.21956/gatesopenres.14406.r30266>

© 2021 Kohli-Lynch C. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Ciaran N. Kohli-Lynch

¹ Center for Health Services & Outcomes Research, Northwestern University, Chicago, IL, USA

² Health Economics and Health Technology Assessment, University of Glasgow, Glasgow, UK

This is a well-written and well-conducted study that provides a useful assessment of health economic impact evaluations in low- and middle-income countries. I include only minor suggestions for clarifications in the text.

Abstract:

- The abstract gives a concise summary of the background, methods, results, and conclusions of the paper.

Introduction:

- The introduction contains a good overview of the growing role of impact evaluations of interventions affecting health in low- and middle-income countries. The need for economic evaluation alongside these impact evaluations is made clear.
- Later in the paper it is stated that an objective of this study was to “generate a discussion of methodological gaps that could guide future research on how to better combine economic evaluations and impact evaluations.” A short discussion on the distinct need to restrict this analysis to low- and middle-income countries would be informative. Has there been a relatively slower uptick in impact evaluations and accompanying economic evaluations in LMICs? Does data sparsity make conducting these analyses more difficult? More generally, the introduction would be improved by an explicit statement of the objectives for the study.

Methods:

- The methods are well-described and are appropriate for the research question. The condensed iDSI criteria used to assess quality are reasonable. Some clarification on two reasons for study exclusion would be helpful.
- First, only the “Health, Nutrition and Population” segment of the 3ie database was reviewed.

Could studies that are included in other segments of the database evaluate interventions which do not primarily affect health but have caused health-related externalities?

- Second, the study excluded “trials which only address the biomedical efficacy of a drug or treatment.” It is later stated, “Health outcomes were interpreted broadly, including health benefits such as height gained, weight loss,” A definition and/or example of *biomedical efficacy* would be informative. It is not immediately clear why some clinical outcomes (e.g. weight loss) are evaluated but others are omitted when both may be intermediaries of disease-related outcomes and health-related quality of life.

Results:

- The results section is well-written and contains useful, relevant information. It would provide useful clarification to state that Duflo *et al.* used average cost-effectiveness ratios in their study of HIV/AIDS prevention.

Discussion:

- The discussion provides a useful summary of the results and explanation of the findings. The need for future research is highlighted, namely development of methods for synthesising results from disparate impact evaluations, increasing the generalisability of impact evaluations, and assessing the multi-sectoral impact of interventions.

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

Not applicable

Are all the source data underlying the results available to ensure full reproducibility?

No source data required

Are the conclusions drawn adequately supported by the results?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: health economics, cardiovascular disease prevention

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 19 February 2021

<https://doi.org/10.21956/gatesopenres.14406.r30263>

© 2021 Anselmi L. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Laura Anselmi

Health Organisation, Policy and Economics (HOPE) Group, The University of Manchester, Manchester, UK

The study systematically identifies health impact evaluations implemented in low and middle-income countries and reviews the approach to the assessment of value for money and the standard to which it is conducted. Only 70 of the studies reviewed has made an attempt to conduct an economic evaluation, with a great variability in quality.

The study is relevant and well conducted in all aspects. I have only some minor comments.

1. I believe that the point that there is little evidence of economic evaluation of health policies, rather than specific health interventions, applies to all settings, not LMICs specifically. I guess this is what the authors are hinting at when mentioning studies evaluating pay-for-performance and having called for the accompanying assessment of cost-effectiveness.
2. The opening statement of the abstract "Health impact evaluations (HIEs) are currently the main way of assessing policy changes in low-and middle-income countries (LMICs)." may be misleading, as for most non-clinical policies the evaluation concerns health care access and utilisation rather health. It is explained later that health is defined to comprise health care use and health behaviours, but it may be worth clarifying this difference since the beginning. I think this is particularly important because evidence on the health impact is still lacking for many policies.
3. The introduction to the relevance of the review is based on important references. Given the fast increase in the volume of impact evaluations, it would be valuable to add some more recent ones, if available. An example, but perhaps not the best, may be Thondoo *et al.* (2019¹).
4. Studies are selected based on the 3ie database, which aims to stock all published impact evaluations of development interventions, and through an additional search. I wonder if there may be specific health policy evaluations which may not be "classed" as development and that would not be included. If so, I would suggested highlighting this point.
5. There is an increased tendency to evaluate the health impact of non-health policies. See for example Mindell *et al.* (2008²). I wonder if by searching specifically the "health, nutrition and population" segment of the repository some evaluations not concerned with the health sector may have been missed out, if they were of relevance.
6. When reporting the results of the literature search I thought it could be informative, if not

too cumbersome, to have also the number of studies for the type of interventions, the measure of health used and the type of data used. I believe the authors could draw from here also important implications for further research, perhaps emphasizing (but I may be having the wrong expectation here) the lack of direct health measures and the underutilisation of administrative data routinely collected. Perhaps some of the recommendations could then point to the need of improving routine data to facilitate these evaluations.

7. Does “medical sector perspective” refer to health care sector perspective more broadly?
8. It is unclear if the authors argue for the need of more studies evaluating value for money, or for the need of all studies evaluating health impact to also include an economic evaluation.
9. It is stated only later in the paper that one objective was to generate a discussion of the methodological gaps that could guide future research on how to better combine economic evaluations and impact evaluations. It would be worth highlighting this at the beginning.
10. There is a mention of the need for increased generalisability across settings. It is not clear what settings mean here, whether across countries, or care settings or perhaps type of intervention. I also wonder if some of the suggestions made point at understanding generalisability by understanding the mechanisms to impact.
11. A final point is about methodological improvements in further research. Is there a need for validations of health measures in specific settings?

References

1. Thondoo M, Rojas-Rueda D, Gupta J, de Vries D, et al.: Systematic Literature Review of Health Impact Assessments in Low and Middle-Income Countries. *International Journal of Environmental Research and Public Health*. 2019; **16** (11). [Publisher Full Text](#)
2. Mindell JS, Boltong A, Forde I: A review of health impact assessment frameworks. *Public Health*. 2008; **122** (11): 1177-87 [PubMed Abstract](#) | [Publisher Full Text](#)

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

Yes

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Health economics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.