The European Journal of Development Research (2023) 35:426–451
https://doi.org/10.1057/s41287-022-00569-3

SPECIAL ISSUE ARTICLE

Using Outcome Trajectory Evaluation to Assess HarvestPlus' Contribution to the Development of National Biofortification Breeding Programs

Boru Douthwaite[1] · **Nancy Johnson**[2] · **Amanda Wyatt**[3]

Accepted: 5 October 2022 / Published online: 5 November 2022

Abstract

Improving policies—broadly defined—is at the heart of the structural transformation agenda. This paper describes the use of a new evaluation method—outcome trajectory evaluation (OTE), based on both evaluation and policy process theory—to explore the influence of HarvestPlus, a large and complex research for development program focused on improving nutrition, on a specific policy outcome, namely the establishment of biofortification crop breeding programs in national agricultural research institutes in Bangladesh, India, and Rwanda. The findings support claims of significant HarvestPlus contributions while also raising issues that need to be monitored to ensure sustainability. The paper also discusses the pros and cons of the OTE approach in terms of methodological rigor and the accumulation of learning from one evaluation to the next.

Keywords Theory-based evaluation · Policy process evaluation · Middle-range theory · Biofortification

Résumé

L'amélioration des politiques, au sens large, est au cœur du programme de transformation structurelle. Cet article décrit l'utilisation d'une nouvelle méthode d'évaluation - l'évaluation de la trajectoire des résultats (outcome trajectory evaluation ou OTE en anglais), basée à la fois sur la théorie de l'évaluation et du processus politique - pour

✉ Boru Douthwaite
bdouthwaite@gmail.com

Nancy Johnson
N.Johnson-cp@cgiar.org

Amanda Wyatt
a.wyatt@cgiar.org

[1] Selkie Consulting Ltd, Westport, Ireland

[2] St Paul, MN, USA

[3] IFPRI, Washington, DC, USA

explorer l'influence de HarvestPlus, un vaste et complexe programme de recherche pour le développement axé sur l'amélioration de la nutrition, et orienté vers un résultat politique, à savoir la mise en place de programmes de biofortification et de sélection de cultures dans les instituts nationaux de recherche agricole au Bangladesh, en Inde et au Rwanda. Les résultats appuient les allégations selon lesquelles HarvestPlus permet des contributions importantes, tout en soulevant des problèmes qui doivent être surveillés pour assurer la durabilité. L'article discute également des avantages et des inconvénients de l'approche OTE en termes de rigueur méthodologique et d'accumulation d'apprentissage d'une évaluation à l'autre.

Introduction

While the key role that policy plays in sustainable development has long been recognized, rigorously documenting the influence of research on policy outcomes faces conceptual, empirical, and even political challenges (Renkow 2018; Slade et al. 2002). Addressing these challenges is increasingly urgent since improving policy is at the heart of the structural transformation agenda in international development. This paper describes the use of a new evaluation method—outcome trajectory evaluation (OTE)—to explore the influence of HarvestPlus, a large and complex research for development program focused on improving nutrition through agriculture, on a specific policy outcome, namely the establishment of biofortification crop breeding programs in national agricultural research institutes in Bangladesh, India, and Rwanda. By building on both evaluation and policy process theory, OTE seeks to improve the rigor of policy influence evaluation by ensuring that the evaluation covers all factors that are hypothesized to influence policy outcomes, not only those factors targeted by the program. By systematically considering all factors that potentially contribute to policy change, the approach reduces the risk of overstating program influence on an observed policy outcome.

The objectives of the paper are to describe a new approach to understanding and evaluating policy outcomes, provide an example of its use in a specific context, and reflect on some advantages of the approach and where it might be used that may be relevant to potential users. The paper is organized as follows. "Overview of the HarvestPlus Program and Its Policy Influence Efforts" section describes the HarvestPlus program and "Evaluation Approach" section presents the OTE approach. "Evaluation Findings and Results" and "How the Specific Strategies Worked to Influence Immediate Outcomes" sections present the general and specific findings of the evaluative review while "Discussion" section discusses their significance. "Conclusions" section concludes with recommendations for program implementers, evaluators, and for further research.

Overview of the HarvestPlus Program and Its Policy Influence Efforts

CGIAR and national agricultural research and extension systems (NARES) have collaborated on crop improvement programs for decades.[1] Starting in the late 1990s, CGIAR breeders began looking at the potential for breeding to increase the micronutrient concentration in staple crops (a.k.a. biofortification) in an effort to contribute to a reduction in micronutrient malnutrition, also known as hidden hunger. In 2003, the HarvestPlus program was established to work on biofortification at the CGIAR system level.[2] Once technical feasibility was established, the focus moved to breeding and testing varieties in practice and then to disseminating the varieties at scale and institutionalizing biofortification in national and international programs and policies. In 2016, four researchers behind biofortification were awarded the World Food Prize (World Food Prize Foundation).

From the beginning, HarvestPlus also invested systematically in understanding, estimating and tracking the impact of biofortified crop varieties on nutrition outcomes. The goal was to build an evidence base to convince not only the agricultural community but also the public health nutrition community that biofortification could be a cost-effective nutrition intervention (Bouis and Saltzman 2017; Johnson et al. 2017). Initial *ex ante* economic impact studies (Meenakshi et al. 2010) were followed by nutritional efficacy studies, by studies of factors affecting uptake by producers and consumers, by effectiveness studies and ultimately, by documenting the dissemination, adoption and consumption of biofortified varieties at scale (see for example Saltzman et al. 2017; HarvestPlus 2014, 2019). HarvestPlus' initial core research areas—crop breeding, nutrition, impact and policy, and reaching end users—were increasingly complemented by efforts designed to build capacity, establish partnerships, jump start dissemination, and engage with and influence policy.[3]

Consistent with Renkow (2018, p. 2), policy was broadly understood to include different types of policy-oriented outcomes to which CGIAR research contributes:

- *Changes in laws and regulations* governing economic incentives in agriculture or natural resource management—for example, agricultural, macro, trade, nutrition/health, and environmental policies;

[1] For the purposes of this paper, CGIAR and HarvestPlus are used interchangeably to describe the work with NARES on breeding programs for biofortified crops. HarvestPlus was established as a joint venture between two CGIAR Centers, the International Food Policy Research Institute (IFPRI) and the International Center for Tropical Agriculture (CIAT), in 2003. HarvestPlus is based at IFPRI and was part of the CGIAR Research Program (CRP) on Agriculture for Nutrition and Health (A4NH). HarvestPlus' crop research and breeding work draws on the expertise and resources of partner CGIAR Centers for breeding the respective biofortified crops which for the focus of an A4NH-commissioned evaluative review of HarvestPlus were the International Rice Research Institute (IRRI) (zinc rice for Bangladesh), the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), (iron pearl millet for India), and CIAT (iron beans for Rwanda).

[2] https://www.harvestplus.org/.

[3] For details on the African context, refer to the Special Issue of the African Journal of Food, Agriculture, Nutrition and Development devoted to Biofortification—https://www.ajfand.net/Volume17/No2/index.html#gsc.tab=0.

- *Creation of institutions*—for example, the formation of the Ethiopian Commodity Exchange or the agreement between India, Nepal, and Bangladesh to share rice varietal evaluation data among their respective countries to facilitate more rapid release and commercialization;
- *Changes in government investment priorities and budget allocations*—for example, increases in the share of budgets devoted to agricultural research associated within the Comprehensive Africa Agriculture Development Program (CAADP);
- *Innovations to the operations and management for government agencies and programs*—for example, monitoring and evaluation activities associated with operating social safety net programs like the Mexican PROGRESA conditional cash transfer program or the Ethiopian Productive Safety Net Program;
- *International treaties, declarations, or agreements among parties reached at major policy conferences*—for example, contributions of the International Food Policy Research Institute (IFPRI)'s trade policy research to the Doha Round of trade negotiations among the World Trade Organization membership or the substantial involvement of the CGIAR Research Program (CRP) on Climate Change, Agriculture, and Food Security (CCAFS) in crafting international climate treaties.

Over the course of the program, HarvestPlus sought to influence policy outcomes from all these categories, from influencing regional and national agricultural and nutrition policies (Baral and Birol 2020; HarvestPlus no date; Foley et al. 2021) to contributing to the establishment of a definition for biofortification under Codex Alimentarius (Bouis and Saltzman 2017; Saltzman et al. 2017) to mainstreaming breeding for nutrition in CGIAR (Rijsberman 2014; Baral and Birol 2020).

As mentioned earlier, the most recent phase of HarvestPlus's program was focused on scaling up biofortified varieties in target countries. To achieve this HarvestPlus worked with and supported breeders in NARES to develop, test, and release biofortified varieties. For biofortification breeding to become sustainable, these programs would need to become institutionalized meaning that they were no longer special projects funded by external donors but rather core parts of the national breeding strategies and programs. This is the outcome that the CRP on Agriculture for Nutrition and Health (A4NH), led by IFPRI, sought to assess when it commissioned an evaluative review of HarvestPlus' contribution to the development of national biofortification breeding programs in Bangladesh, India, and Rwanda (Douthwaite 2021).[4] Since HarvestPlus funding to the programs ended in 2018, it was possible to assess the status of those programs, 3 years later in 2021.

[4] The study was part of a series of evaluative studies and reviews to document lessons from both phases of the CRP (2012–2021). The studies were designed to inform future research and development efforts, mainly but not exclusively in CGIAR. Findings can inform both what the programs do and how they work. This study looked at lessons learned from HarvestPlus' work with NARES to develop and implement sustainable biofortification breeding programs. HarvestPlus was part of both phases of A4NH.

Table 1 Similarities between OTE and other approaches used to evaluate policy outcomes (Douthwaite et al., forthcoming)

Evaluation approach	Similarities to OTE
Process tracing (Collier 2011)	– Focus on unfolding events or situations over time to make causal inferences – The idea that causal inferences can be affirmed through building up a weight of evidence—in the form of various 'straw-in-the-wind' tests that individually do not affirm causal inference but collectively do – Use of criminal justice system analogies in explaining how the approach works
Outcome harvesting (Wilson-Grau 2019)	– The practice of 'back-casting' from an established outcome to understand what has contributed to it – Interviews with knowledgeable stakeholders to validate or repudiate causal claims
Outcome evidencing (Paz-Ybarnegaray and Douthwaite 2017)	– The use of outcome trajectories
Contribution analysis (Mayne 2012)	– Use of a contribution story, similar to the timeline used in OTE – The development and refinement of a theory of change (ToC) as a part of the analysis
Episode study (Carden 2009; Leksmono et al. 2006; Jones 2011)	– Back-casting from a well-defined policy change – Development of a historical narrative about what led to the policy change in question before assessing the relative role of research in that narrative

Evaluation Approach

Influencing policy is important for development, in particular sustainable development and structural transformation; however, documenting policy influence is hard. This is not because it is hard to know whether a policy change occurred, but rather because it is difficult to determine what factors led to the change. Quantitative methods for impact evaluation use statistical approaches to make causal claims; however, they require large sample sizes which are usually not possible in the case of policy influence. Thus, narrative and/or theory-based approaches are often recommended (White and Phillips 2012).

OTE adopts the outcome harvesting definition of an outcome: "A change in the behaviour, relationships, actions, activities, policies or practices of an individual, group, community, organization or institution" (Wilson-Grau 2019, p. 2). OTE shares similarities with four approaches used to evaluate policy outcomes (Table 1). However, OTE's main difference is in its use of theory. OTE, assumes, a priori, that significant policy-related outcomes, such as the establishment of a biofortification breeding program, are not single, one-off events, but rather are generated and sustained over time by an interacting and co-evolving system of actors, knowledge, technology and institutions. These systems are called *outcome trajectories* (Paz-Ybarnegaray and Douthwaite 2017) and are bounded and described by building and annotating a timeline of the events and processes thought to have contributed to the

outcome being evaluated. Outcome trajectories are generative mechanisms that are assumed to work as follows: Outcomes are understood to emerge in complex adaptive systems, through the interaction of actors, their strategies and decision-making, institutions, artifacts (i.e., technology), and knowledge (e.g., Axelrod and Cohen 1999; Douthwaite and Gummert 2010). This is more precise than the implicit definition of an 'episode' in the literature as simply the historical narrative constructed to explain a policy change (e.g., Leksmono et al. 2006; Carden 2009).

Understanding an outcome trajectory as a complex adaptive system has important implications for how OTE deals with the necessary and sufficient questions at the center of making causal claims in contribution analysis. The questions are: (1) Is the causal package sufficient? (2) Are the elements in the causal package necessary?. In OTE, the causal package is assumed to be the outcome trajectory from the outset.; The sufficiency question is "does the outcome trajectory add up to a credible explanation as to how the outcome in question emerged over time?" The assumption of the complex adaptive system dynamic driving the outcome trajectory, with outcomes emerging from interactions, makes it almost impossible to say whether any contributing strategy was necessary or not. What can be said is that strategies identified as contributing to an outcome trajectory are likely to have been necessary.

The other way OTE uses theory is to use a *middle-range theory* as the 'theory of the case' to help identify and describe an outcome trajectory and to avoid the risk of becoming bogged down in too much spurious detail, which is a known issue when working with case studies (Baxter and Jack 2008). Middle-range theories are positioned between 'grand' universal systems that describe all features in a stylized way and ad hoc explanations of singular cases (Pawson 2010, 2013). Middle-range theories apply to clusters of similar programs and can therefore help develop program theories of change (ToCs) that are comparable at cluster level, and so can aid cross-case learning and insight. A number of grand theories exist in the policy realm (Sabatier and Weible 2007) that have been simplified and described such that policy advocates and evaluators can choose which will best help their understanding and navigation of the policy processes in which they are involved (Stachowaik2013; Resnick et al. 2018). In the case of an evaluation of policy influence, starting with an existing policy process middle-range theory and adapting it to the specific program context can ensure that the evaluation is rooted in accepted understanding of how policy influence happens and that it considers the contributions of a program through this lens.

OTE does not explicitly seek out and discount rival explanations for the working of strategies included in the outcome trajectory. Rather OTE establishes whether the explanation is internally valid by understanding how the strategy is purported to have worked and whether this explanation is congruent with explanations of the timing and working of other interacting parts of the outcome trajectory. OTE seeks to establish external validity of an explanation by ensuring the way a strategy was supposed to work was congruent with the middle-range theory. OTE provides the opportunity for stakeholders involved to challenge and add to the middle-range theory during the course of an evaluation (see Step 4). Since the middle-range theory includes the main factors expected to account for rival explanations, it does in part account for rival expectations.

Implementation

Operationalizing OTE in an evaluation involves six steps which are described below for the case of the HarvestPlus study. Before describing the steps, we briefly describe the focus and scope of the evaluative review.

The main evaluation question addressed using OTE was how and to what extent did HarvestPlus contribute to the establishment and implementation of sustainable biofortification breeding programs in Bangladesh, India, and Rwanda.[5] The outcome trajectories to be described, modeled and evaluated are the establishment and implementation of three biofortification breeding programs, namely:

- Zinc rice in Bangladesh with the Bangladesh Rice Research Institute (BRRI);
- Iron pearl millet (IPM) in India with the All India Coordinated Research Project on Pearl Millet (AICRP-PM); and
- Iron beans in Rwanda with the Rwanda Agriculture and Animal Resources Development Board (RAB).

These 'crop × country' combinations were selected by HarvestPlus. The idea was to look at mineral nutrients (iron and zinc) in different crop × country contexts. In all cases, the programs were considered successful, and it was felt that sufficient information would be available on which to base the study. Details on how trajectories were constructed, including selection of interviewees, are described below in Step 2.

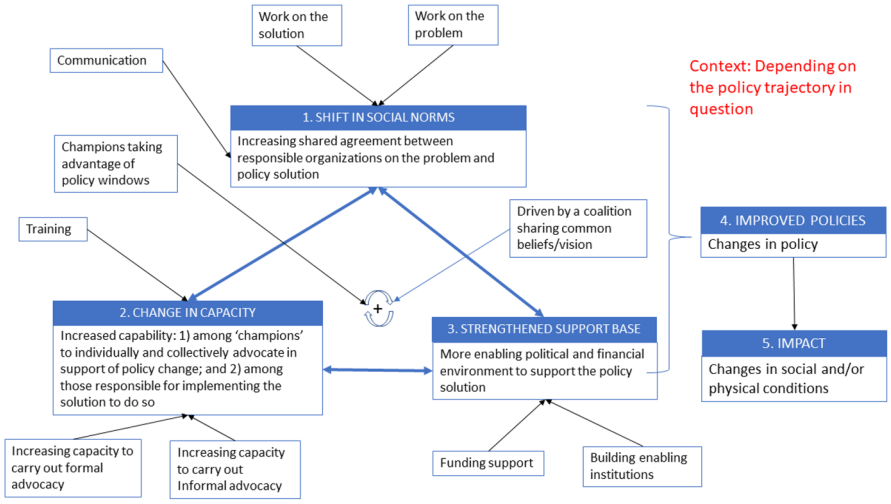
Step 0: Ensure That Outcome Trajectory Evaluation is Appropriate to Answer the Evaluation Questions

OTE is suited for carrying out after-the-fact 'ex post' evaluations in which it is possible to work backwards, i.e., backcast from a coherent set of achieved outcomes to establish contribution, such as the establishment of biofortification breeding programs. OTE is not suitable for before-the-fact 'ex ante' evaluations of likely outcomes that might result from a set of program activities. Nor is it suitable for evaluating a diffused set of outcomes, such as those that might result from efforts to build soft skills across a network of organizations.

OTE can be used when the following conditions are met:

- At least one clear and significant program outcome can be identified for which an outcome trajectory can be described;
- A middle-range theory exists to describe how similar types of outcomes emerge; and,
- Some documentation and actor recall of the events and processes in the outcome trajectory exists.

[5] The full evaluative review (Author et al. 2021) included other more forward-looking evaluation questions. While these built on the results obtained using OTE they were not addressed using the OTE method so are not reported here.



Source: AUTHOR et al. (forthcoming)

Fig. 1 A middle-range theory describing how CGIAR and partner interventions contribute to changes in policy-related outcomes. *Source* Author et al. (2020)

Step 1: Select Middle-Range Theory to Focus Description and Understanding of the Respective Outcome Trajectories

The first step in the evaluative review was to select a middle-range theory to provide a framework to help construct and make sense of the three outcome trajectories (crop × country breeding programs) and the contexts in which they operated. We chose one that was adapted and specified in a recent, similar evaluation conducted for another CRP on Roots, Tubers, and Bananas (Author et al., forthcoming) to describe how CGIAR contributed to four policy outcome trajectories (Fig. 1). One of the outcome trajectories in that evaluation was on the development of a Continental Declaration on biofortification to be adopted by the African Union Commission.

The selected middle-range theory is based on Kingdon's (1984) Policy Window theory, as interpreted by Stachowiak (2013). The theory was chosen in the first place from several middle-range theories described by Sabatier and Weible (2007) and others (e.g., Stachowiak 2013) because it has been assessed as widely applicable across a number of domains and was chosen by both the evaluation team and key stakeholders as the best fit.

Policy Window theory proposes that changes in policy-related outcomes occur during *windows of opportunity*, which help champions successfully connect two or more components of the policy process. The components are the way a *problem* is defined, the *policy solution* to the problem, and the *politics* surrounding the issue (Stachowiak 2013; Zahariadis 2008). Windows of opportunity are moments when progress can be made. They can be created by natural events such as pandemics, droughts, or earthquakes. They can also be changes in government, budget cycles,

or landmark meetings and summits held as part of ongoing national, regional, and global processes. Policy windows are often short in duration and may or may not be predictable.

The attraction of using a theory specified and adapted from a similar previous study was that it allowed us, the evaluation team comprised of the evaluator, plus the two evaluation managers, all co-authors of this paper, to learn from and build on it. We would be able to use the results to identify the different strategies that had proven useful to achieve outcomes—shifts in social norms, increased capacity and strengthened support base (see Fig. 1)—in different contexts for different types of policy-related outcomes (i.e., establishment and functioning of biofortified breeding programs in three countries). Our expectation was that this would make our modified version of Fig. 1 more broadly generalizable for future outcome evaluations of policy-oriented outcomes of CGIAR and research partners.

Step 2: Identify and Describe the Outcome Trajectory that Has Led to the Respective Biofortification Breeding Programs Being in Existence

The lead author ("the evaluator") developed timelines for the three respective outcome trajectories, based on interviews and reviews of available data, reports, and online publications (Author et al. 2021). Included in the data gathering was a detailed review of project documents carried out by HarvestPlus staff so as to identify:

- The investments made into the respective breeding programs over the study period, including but not limited to investments in technical capacity;
- The main outputs of the breeding programs in terms of the release of new biofortified varieties and provision of breeding material to other breeding programs; and
- Information relating to the three components of sustainability:
    - The competitiveness of new varieties released, measured in farmer and consumer preference and rates of adoption, e.g., adoption studies;
    - Political support in terms of awareness raising, extension and advocacy activities and outcomes associated with these, e.g., identification of policy documents that support biofortification; and
    - Technological progress, such as advances in ability to test for micronutrients, e.g., the arrival of a new X-ray fluorescent (XRF) machine.[6]

In developing the timelines, the evaluator wove together information from interviews, reports, and papers provided by HarvestPlus, and web-based keyword searches, to build a picture of the respective outcome trajectories. The evaluator interviewed a total of 18 people about the three cases. Interviewees represented

[6] The use of the XRF machine in breeding began in the early 2010s. The XRF machine accelerates the screening process for target minerals and is also used to confirm mineral levels in bred varieties.

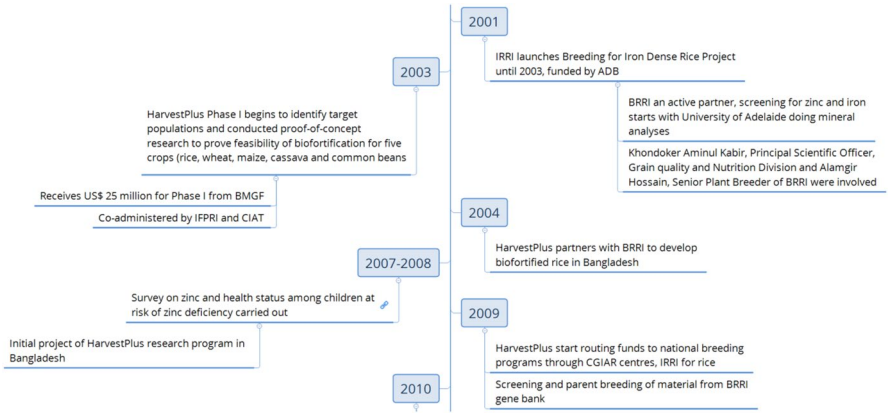


Fig. 2 Portion of the zinc rice breeding timeline in Bangladesh (based on Author 2021 p. 39)

organizations involved in the respective biofortification breeding programs and/or were individuals knowledgeable of the broader value chain context in which the breeding programs existed. Semi-structured interviews were recorded and detailed notes made of them upon listening to the interview a second time. The evaluator also carried out the analysis and wrote the report. There was no coding beyond grouping together answers to the same questions. Anonymized, detailed referencing was used to establish an audit trail such that facts and assertions made in the report could be checked back to their sources. Where pieces of data did not appear to fit, further clarification was sought and understanding adjusted.

Our approach was essentially a case study one in which understanding flowed from rich, thick picture descriptions of events gleaned from interviews in particular. We used the middle-range theory shown in Fig. 1 as the 'theory of the case' (Rule and John 2015) to help focus inquiry as we built annotated timelines of the three respective outcome trajectories. Specifically, we looked for manifestations of the three immediate outcomes—outcomes 1 to 3 shown in Fig. 1—together with the events and processes that may have contributed, which were recorded on the timeline, together with notes with regard to their significance (see for example Fig. 2).

Step 3: Validate the Outcome Trajectory Timelines with Key Stakeholders

The evaluator sent out the annotated outcome trajectory timelines for each case to the respective interviewees to validate, challenge, and add to the timelines. The annotated timelines were adapted based on this feedback. Response rate was above 50%.

Step 4: Identify Specific Strategies Used to Achieve the Immediate Outcomes, Adapt, and Validate Middle-Range Theory

Based on the annotated timelines, the rich, thick description from interviews, and document review, we identified the specific strategies used to achieve the general

strategies shown in Fig. 1. For example, the specific strategies related to 'communication' were as follows: (1) consumer marketing using print, radio, TV, and social media and (2) field demonstrations, field days, and shows. These two strategies were used in all three cases. The result is a complete list of specific strategies, and how they map—or not—to general strategies (see Table 3 in the next section).

Secondly, we adapted the generic descriptions of the immediate and intermediate outcomes to apply to the three cases. For example, we refined the description of immediate outcome 3 'strengthened support base' from 'more enabling political and financial environment to support the policy solution' to 'more enabling political and financial environments to support biofortification breeding programs alongside other efforts to disseminate and promote adoption.'

The evaluation team organized a workshop for interviewees to validate, challenge, and add to the specific strategies mapped onto the general strategies. This was done by showing the mapping to the workshop participants, general strategy by general strategy, and asking for their comments in terms of whether they thought the specific strategies had been applied, were described correctly, and whether any had been missed. This process was designed to help reduce confirmation bias on the part of the evaluation team.

Step 5: Use the Validated Timelines and Three-Case ToC to Answer the Evaluation Question

For each outcome trajectory, the evaluator used the conceptual framing, the case-specific timelines, the adapted ToC, notes from the in-depth interviews, and information from the document reviews to answer the evaluation question.

Step 6: Subject the Draft Report to Review for Fact and Inference Checking Before Finalizing

A first draft of the full evaluative review was provided to the evaluation managers to coordinate a review process to check facts and the legitimacy of inferences made. Comments and suggestions from a representative sub-set of interviewees were collated and considered. The changes made and not made were recorded and explained.

Evaluation Findings and Results

Overview of Outcomes and Other Key Characteristics of the Trajectories

Key information on HarvestPlus activities in each country, including outcomes achieved, is summarized in Table 2. In all three, the policy-oriented outcome sought was a national biofortification breeding program established in each country for each crop.

The level of investment by HarvestPlus in each trajectory fell sharply at the end of the program's third phase in 2018. For example, in Rwanda, HarvestPlus judged that it had met its objectives with respect to promoting and making iron

Table 2 Characteristics of the three biofortification breeding programs

	Zinc rice in Bangladesh	IPM in India	Iron beans in Rwanda
Policy-oriented outcome sought	Sustainable zinc rice breeding program at BRRI	IPM mainstreamed into the Indian Council of Agricultural Research (ICAR), (through AICRP-PM) and private sector breeding programs	Sustainable iron bean breeding program at RAB
Year trajectory began	2004	2004	2009
Key trajectory actors	More than 30 Host org: BRRI Research: CGIAR, Bangabandhu Sheikh Mujibur Rahman Agricultural University (BSMRAU), Bangladesh Institute of Nuclear Agriculture (BINA), BRRI	More than 40 Facilitating org: ICRISAT Host org: ICAR Research: ICRISAT, SAUs, 37 seed companies, farmer production organizations, nongovernmental organizations for product development and dissemination	More than 30 Host org: RAB Research: CGIAR, RAB
HarvestPlus investment	US$800,000 from 2010 to 2020	US$1,914,000 from 2009 to 2022	US$1.2 million from 2009 to 2019
Breeding outputs	9 High zinc rice varieties released by BRRI (6), BSMRAU (2), and BINA (1)	1 Open-pollinated variety (OPV) bred by ICRISAT; 8 hybrids produced by SAUs; 1 hybrid bred by private sector seed company	18 Iron bean varieties released by RAB
Future direction	BRRI to continue with dedicated high zinc (and iron) breeding program before transitioning to the mainstreaming of biofortification breeding in about 7 years (by 2028)	IPM breeding to take place under AICRP-PM umbrella screening varieties for notification and release, while helping to increase benchmark, and push for policies that create a demand for IPM	RAB to continue breeding iron bean varieties and to work to reduce the cost of good quality seed and make it more available, with private sector
Other outcomes	Zinc rice varieties reached 2,454,000 households directly from 2013 to 2020	In 2019, 240,000 farmers were growing IPM (largely Dhanashakti); in 2020, 65,000 farmers were growing IPM High iron (> 42 ppm) set as national benchmark for AICRP system	Biofortification successfully embedded into the country's food system: 15% of beans consumed are high in iron, 20% of beans grown are high in iron; 420,000 households growing iron beans (as of 2018)

From Foley et al. (2021, p. 121). Source of the data was "HarvestPlus Monitoring and Evaluation Team. HarvestPlus database. Published online 2019"

The drop from 2019 was attributed to COVID-19 restricting people's movement

bean seeds available. It reduced the overall investment to fund an ongoing liaison with the government and maintaining the XRF machine to test for iron levels in beans. The number of NARES bean breeders working in the country declined from 5 in 2018 to 2 in 2021. Nevertheless, breeding of iron bean varieties remains a priority for the national bean breeding program (Author et al. 2021). In India, separate biofortified hybrid trials for IPM were stopped in 2018 when the Indian Council of Agricultural Research (ICAR)'s AICRP-PM agreed in its annual meeting to screen all promising varieties and hybrids in national trials for iron and zinc content, set at 42 parts per million (ppm) and 32 ppm, respectively. HarvestPlus funding since 2018 has been used to fund the screening of about 8000 to 9000 lines a year, together with soil, at a cost of US$2.50 per sample (Author et al. 2021). HarvestPlus funding for IPM in India was much higher than for iron beans in Rwanda and zinc rice in Bangladesh. The difference is that funding was distributed among 30 organizations, including State Agricultural Universities (SAUs) and private sector seed companies.

Identifying Strategies and Adapting the Middle-Range Theory of Change

As described under Step 4, three data sources—the timelines, interviews, and document review—were used to adapt and specify the middle-range theory (Fig. 1). This was done by systematically looking across the data sources for evidence of the use of specific strategies that contributed to the general strategies shown in Fig. 1. This was a deductive process, driven to a large extent by the evaluator's understanding of how change happens, built on three decades of experience. To help reduce the risk of confirmation bias, the specific strategies identified were validated with the people interviewed in a workshop (Step 4) and modified accordingly.

It is important to note that outcome trajectory actors, including HarvestPlus, may not have understood that they were employing those strategies at the time. However, looking back, in the evaluator's view, and validated by key participants, the specific strategies do a plausible job of explaining how HarvestPlus, and other trajectory actors, contributed to the implementation of the respective general strategies.

We made some adaptations to the general strategies to better match the findings, as follows and as shown in Fig. 3:

- Under 'strengthened support base,' we added 'advocacy' as a strategy to create more enabling political and financial environments for the breeding, dissemination and adoption of biofortified crops.
- Under change in capacity, 'building formal capacity for advocacy' and 'building informal capacity for advocacy' are considered as one strategy: 'increasing and using capacity to carry out advocacy.' 'Training' is replaced with 'building technical capacity among breeders and value chain actors.'
- We reworded the strategy 'champions taking advantage of policy windows' to 'advocating for enabling policies.' We changed 'driven by coalitions sharing common vision' to 'holding events to generate policy windows.'

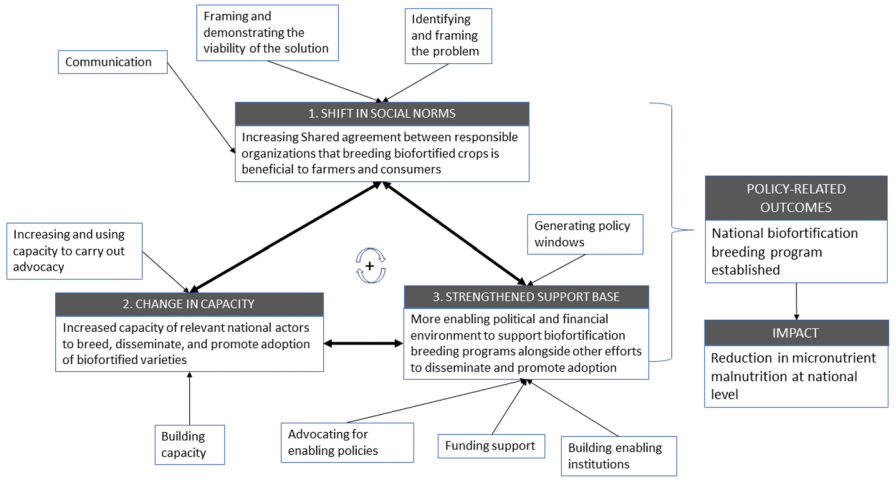


Fig. 3 Overview of how HarvestPlus contributed to the establishment and implementation of biofortification breeding programs in three countries, showing general strategies mapped onto immediate outcomes

Figure 3 is an adaptation of Fig. 1 to make it specific to the three cases in question.

The mapping of specific strategies onto general strategies onto immediate outcomes is shown in Table 3. The first column of the table shows the general strategies, organized by the immediate outcomes to which they contribute. The second column shows the specific strategies that map onto the general strategies. The third column indicates which cases used the specific strategies. The fourth provides links to summaries of the findings relating to the specific strategies, found later in the text. The full findings are provided in the evaluative review report (Author et al. 2021).

How the Specific Strategies Worked to Influence Immediate Outcomes

We consider in turn how the specific strategies, identified and validated in Step 4, worked to bring about the three immediate outcomes across the three cases. The order in which the specific strategies are presented depends on the general strategies they map onto, which in turn depends on the immediate outcomes the general strategies map onto as shown in Table 3. Figure 3 shows that the immediate outcomes influence each other and can happen at the same time. In the same way, it can be assumed that the general and specific strategies influence each other and may have happened simultaneously.

Summaries of findings with respect to specific strategies are given throughout the following text, to help provide an indication of where links between findings exist. More details of the individual findings can be found in Authoret al. (2021).Please confirm the section headings are correctly identified.

Table 3 General and specific strategies used to contribute to changes in the three immediate outcomes in the ToC, by country, with links to relevant findings

General strategies	Specific strategies used	Cases that used it	Relevant findings
For strengthened support base			
Funding support	Funding biofortification breeding in national programs	All	Finding 1
	Subsidizing the price of seed to allow greater farmer access	Bangladesh, Rwanda	Finding 2
Building and working within enabling institutions	Working with (India) or independently (Rwanda) of existing consortia/multistakeholder platforms sharing a similar objective	India, Rwanda	Finding 3
	Support to biofortification seed systems	All	Finding 4
Advocating for enabling standards and policies	Contribution to policy support for biofortification	All	Finding 5
Generating policy windows	Holding of conferences and crop meetings on biofortification	All	Finding 6
For changes in capacity			
Increasing and using capacity to carry out advocacy	Strengthening the capacity of biofortification champions	No evidence found	Finding 7
	Key trajectory actors using their professional links and expertise to engage in national and international meetings and conferences	All	Finding 8
Capacity development	Capacity development across the respective biofortified crop value chains	All	Finding 9
For shifting social norms			
Framing the problem	Identifying and framing micronutrient malnutrition as a global problem	N/A	Finding 10

Table 3 (continued)

General strategies	Specific strategies used	Cases that used it	Relevant findings
Framing and demonstrating the viability of the solution	Framing and demonstrating at a global level the viability of biofortification as a solution to the problem of micronutrient malnutrition	N/A	Finding 11
	Breeding and release of biofortified varieties in-country as proof of concept	All	Finding 12
	Carrying out commercialization assessments and market research for biofortified crops	All	Finding 13 Finding 14 Finding 15
	Carrying out trials to establish the efficacy of biofortified crops in producing positive health outcomes in the target country	All	Finding 16
	Carrying out adoption studies	All	Finding 17
Communication	Field demonstrations, field days, and shows	All	Finding 18
	Consumer marketing using print, radio, TV and social media	All	Finding 19

Findings Relevant to Immediate Outcome 3: Strengthened Support Base

Funding Biofortification Breeding in National Programs

Finding 1 HarvestPlus funded national partners to undertake biofortification breeding starting in 2009 in the three respective countries. Without this funding, it is highly unlikely that the biofortification breeding programs would have ever been set up.

Subsidizing the Price of Seed to Allow Greater Farmer Access

Finding 2 Once the biofortified varieties were approved for release, HarvestPlus subsidized the price of biofortified seed to allow greater farmer access in Rwanda and Bangladesh. While initially very successful in Rwanda, this led to a gap in the market when the indirect HarvestPlus subsidy was removed, and criticism that HarvestPlus should have worked to support value chain actors rather than become one itself. However, if HarvestPlus had been less interventionist, it may not have achieved such high adoption levels in such a short period of time. In Bangladesh, HarvestPlus also initially guaranteed a market for a portion of the private sector production and subsidized the price for any seed that the private sector marketed directly to consumers.

Working with or Independently of Existing Consortia/Multi-stakeholder Platforms Sharing a Similar Objective

Finding 3 HarvestPlus took different approaches to engaging with other trajectory actors depending on the size and governance structure of the country. In India, to reach the states growing pearl millet, HarvestPlus worked through the umbrella of ICAR's AICRP-PM, a consortium of 12 SAUs involved in breeding pearl millet. HarvestPlus also worked through the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT)-led Pearl Millet Hybrid Parents Research Consortium (PMHPRC) that included 30 seed companies. In Rwanda, HarvestPlus could have worked through the long-established International Center for Tropical Agriculture (CIAT)-led Pan-Africa Bean Research Alliance (PABRA) Network, but chose instead to work directly with RAB, taking a more interventionist approach to ensure rapid iron bean seed distribution and uptake in Rwanda.

Support to Biofortification Seed Systems

Finding 4 In Rwanda, HarvestPlus supported the development of a specification for iron beans to be used as part of iron bean seed certification. In India, the program was party to the decision by AICRP-PM to set the threshold of 42 ppm of iron and screen all new hybrid IPM candidate varieties to equal or exceed the level before promoting to the next level of testing and approving their release.

HarvestPlus supported the varietal release process in Bangladesh and Rwanda by helping to implement and fund necessary multi-locational trials and their analysis.

Contribution to Policy Support for Biofortification

Finding 5 Biofortification and the three case crops are supported in policy documents in all three countries. Perhaps not surprisingly, the direct contribution of HarvestPlus to the policy process appears to have depended on the size of the country. In India and Bangladesh—two large countries—the main policy conduits were the national lead organization for biofortification, that is, ICAR's AICRP-PM for India and BRRI for Bangladesh. In Rwanda, a much smaller country in which the HarvestPlus intervention was much more significant, HarvestPlus appears to have had more direct influence, together with RAB. Promotion of iron beans are supported more often and more specifically in Rwanda's policy documents than in the other two countries.

Holding a Global Conference on Biofortification and Crop Meetings

Finding 6 This review did not have the resources to identify specific policy windows that led to the policy support described under the previous finding. Policy windows for biofortification will likely have resulted from working collaboratively with country networks. In Rwanda, it is likely that they were generated by the Second Global Conference on Biofortification in 2014, which was held in Kigali. Respondents flagged the importance of crop meetings in influencing policy in all three countries.

Findings Relevant to Immediate Outcome 2: Change in Capacity

Strengthening the Capacity of Biofortification Champions

Finding 7 The review found no evidence of an overt strategy to build the capacity of champions in either of the three cases, as Policy Window theory might suggest.

Key Trajectory Actors Using Their Professional Links and Expertise to Engage in National and International Meetings and Conferences

Finding 8 Senior researchers working on the three respective breeding programs, and working for HarvestPlus globally, were supported by HarvestPlus to make the case for biofortification at conferences and in meeting with senior government officials and other key stakeholders, using and building their innate capacities in the process.

Capacity Development Across the Respective Biofortified Crop Value Chains

Finding 9 HarvestPlus facilitated training and capacity development in two main areas: for breeding and for value chain actors. Much of the capacity development

targeting end users (producers and consumers) happened as part of the communications strategy (see Findings 18 and 19).

Findings Relevant to Immediate Outcome 1: Shifting Social Norms

Identifying and Framing Micronutrient Malnutrition as a Global Problem

Finding 10 This general strategy did not apply to the three cases because by the time HarvestPlus began funding the three biofortification breeding programs, micronutrient malnutrition had already been established as a priority global health concern, in particular for women and children.

Framing and Demonstrating at a Global Level the Viability of Biofortification as a Solution to the Problem of Micronutrient Malnutrition

Finding 11 As with the previous finding, the general strategy did not apply to the three cases. Biofortification had been established globally as a solution to micronutrient malnutrition prior to 2009, in part through HarvestPlus running a 'gold standard' effectiveness study on orange-fleshed sweetpotato in Mozambique and Uganda and the resonance the results found, manifest in the Copenhagen Consensus ranking biofortification as the fifth best investment to tackle the world's most pressing development issues.

Breeding and Distributing Biofortified Crops in Country as Proof of Concept of Delivery at Scale

Finding 12 At country level, the most important step to establish biofortification as a viable solution to micronutrient malnutrition in the minds of trajectory actors was to breed and distribute biofortified crops and demonstrate that farmers and consumers would find them acceptable (more in Finding 17). Of the three cases, the first release of an approved biofortified variety was in Rwanda in 2010. One of the five varieties released, MAC-44, went on to become the most widely distributed high iron climbing bean variety in the region.

Carrying Out Commercialization Assessments and Market Research for Biofortified Crops

Finding 13 In India, HarvestPlus and the Global Alliance for Improved Nutrition (GAIN) commissioned the global consultancy firm Dalberg to carry out a commercialization assessment (Dalberg 2019). Although it was too late to influence the establishment of the breeding program, it was potentially relevant to its sustainability. The report estimated that by 2024, about 60% of on-farm consumption and 85% of rural and urban consumption will be of varieties with greater than 42 ppm of iron. The fact that all 17 of the pearl millet varieties released through government channels since 2017 have greater than 42 ppm of iron supports the estimate. HarvestPlus'

own global baseline to qualify as IPM is 47 ppm of iron and the breeding target is 77 ppm of iron. In India, the baseline and breeding targets are currently 42 ppm and 72 ppm of iron, respectively.

Finding 14 The barrier to commercialization identified in the Dalberg report, that no competitive biofortified alternatives exist for farmers using open-pollinated varieties (OPVs), would appear to be at odds with the launch and promotion of the Dhanashakti OPV by HarvestPlus as the first IPM which performed better than the variety it was bred from. Moreover, the adoption of IPM reported by HarvestPlus was largely of this variety. Adoption of nine hybrid varieties supported by ICRISAT/HarvestPlus was delayed by difficulties in licensing public sector-bred varieties to the private sector, a high-level issue that deserves further attention. The value of ICAR's relatively low iron threshold of 42 ppm may be to signal that the government seed system will set more and higher thresholds in 2 or 3 years and seed companies would do well to start breeding accordingly.

Finding 15 HarvestPlus took a much more proactive commercialization approach in Rwanda because of the challenges faced in distributing iron bean varieties to farmers. The program intervened directly in the seed value chain, becoming the largest seed distributer in the country for a period of time. The HarvestPlus Rwanda team experimented with a number of options to rapidly produce and distribute seed to farmers. Two of the most successful were swapping farmers' seed with iron bean seed and providing seed alongside agronomic training and the NARES.

HarvestPlus' approach was very successful in providing iron bean seed to farmers. By 2018, an estimated 420,000 farmers were growing iron beans with 15% of the population of Rwanda eating them. This took place in the context of a very low seed replacement rate among bean farmers. However, becoming a large institutional buyer and then leaving the market created, or at least led to the return of, a number of seed value chain issues, not least poor communication between institutional buyers and seed producers and the high cost of certified seed, unaffordable to most farmers.

Carrying Out Trials to Establish the Efficacy of Biofortified Crops in Producing Positive Health Outcomes in the Target Country

Finding 16 HarvestPlus supported at least four efficacy trials in the three cases, carried out either by Cornell University or the Swiss Federal Institute of Technology. Only two studies provided links to published results, both of which found significant positive health effects. Of the other two studies, no mention of their existence was found in the information HarvestPlus has made easily available online. In contrast, much mention is made by HarvestPlus of the two positive studies, to help frame biofortification as an effective solution to hidden hunger.

Carrying Out Adoption Studies

Finding 17 In each case HarvestPlus conducted adoption studies in part to document and understand the popularity of the respective biofortified crops. In all three cases, multiple studies were conducted in each country and adoption rates varied from one

study to the next. For example, in Bangladesh, published estimates in 2018–2019 of farmers who had ever grown zinc rice ranged from 250,000 to 500,000 to 1.5 million. In Rwanda, published papers and reports were more in agreement that about 15% of the population was eating iron beans in 2018, a level of adoption that is an order of magnitude higher than in Bangladesh or India.

Field Demonstrations, Field Days, and Shows

Finding 18 While field demonstrations, field days, and shows were important communication and awareness building strategies in all three cases, how they were run and by whom differed from country to country in terms of the role of the public versus the private sector, and the level of direct engagement by HarvestPlus.

Consumer Marketing Using Print, Radio, TV, and Social Media

Finding 19 HarvestPlus employed consumer marketing using print, radio, TV, and social media in all three cases. As with the previous finding, the specific strategies used varied from case to case. Bangladesh made most use of print media and television. India relied heavily on the private sector and on one company in particular. In Rwanda, much of the marketing happened in local markets, helped by celebrity endorsement.

Summary of How the Strategies Worked to Contribute to the Outcome

The findings allow for the following summary description as to how HarvestPlus contributed to setting up biofortification breeding programs in the three countries. The numbers refer to the three immediate outcomes involved—shift in norms (1); changes in capacity (2); and strengthened support base (3).

Biofortification was established as a viable solution to micronutrient malnutrition prior to and during HarvestPlus' first phase (1). During HarvestPlus's second phase that started in 2009, in each of the three cases, HarvestPlus provided funding support (3) and capacity development (2) to support national programs to work on biofortified breeding with CGIAR centers. As varieties were developed and released, HarvestPlus engaged with trajectory actors in the value chain, including seed systems, to build capacity (2) and strengthen the support base for biofortification at scale (3). HarvestPlus did this through capacity development (2), advocating for policies to support biofortification (3), securing funding to support biofortification (3), and building enabling institutions with other trajectory actors (3). These efforts at building institutions were complemented by efforts to build capacity (2) of key trajectory actors within their institutions to carry out and to advocate for biofortification (3). These actions led to changes in how trajectory actors viewed biofortification as a viable solution to micronutrient malnutrition (1). These widespread changes in awareness and perception influenced the enabling environment for the technology (3). The three trajectories were also driven forward by results-focused interventions by HarvestPlus backed by its main donors.

Discussion

The starting assumption that outcomes emerge from a system—an outcome trajectory—to which HarvestPlus contributed, proved useful. It helped broaden the scope of the review and in countering the tendency to assign too much causal power to the program being reviewed, rather than the system to which the program contributed.

OTE encourages the evaluator to look at the broad range of factors that might have influenced the outcome. This was certainly important in this evaluative review. In the original proposal for this study, the desired focus was on the role of capacity building in the NARES, what would have been one set of strategies relevant to only one immediate outcome that contributed to setting up the biofortification breeding programs. Instead, in the end, by using OTE, the evaluative review results unpacked specific strategies from HarvestPlus that were relevant to three immediate outcomes that influenced the outcome of interest.

Using an existing theory also identified some generalizable findings. For example, identifying and building the capacity of 'champions' is often considered an essential part of policy influence. However, HarvestPlus does not appear to have used that strategy to support the establishment of NARES-led biofortification breeding programs. A possible interpretation is that for types of policy influence that are more technical in nature, people who are already considered experts (e.g., crop breeders) can be more effective in this role and just need more venues in which to exert their existing influence. They may already have good 'soft skills' without having been trained in them, nor even knowing they possess them. Another possible interpretation is that the enabling environment was already primed or aware that micronutrient malnutrition was a problem and biofortification was a potential solution. Even if it wasn't explicitly employed for this policy change, HarvestPlus had engaged champions in complementary outcome trajectories that had a positive effect on this outcome trajectory (Author et al. 2020).

In a case like HarvestPlus, with a long history of careful monitoring, evaluation, learning, and impact assessment (MELIA) activities, information was available with which to assess the outcomes. The systematic way that the program went about building its research, advocacy and MELIA agenda over time resulted in the production of information that was relevant for this case. It is important to note that 'over time' refers not only the 2-, 3-, or 5-year time horizon of a specific project or grant but also to the 10, 20, or 30 years that it realistically takes for upstream agricultural research to lead to large-scale impact on the ground. Research for development (R4D) programs need to be honest with themselves and their stakeholders about how long research can take while at the same time recognizing the importance of incremental progress both in science and in maintaining support for the program. In addition, the amount of information available online made it easier for the evaluator to corroborate information given by interviewees, and vice versa.

OTE is for looking backwards, and is probably best to use in cases where the trajectory is long and somewhat complex. Program evaluators should invest in

OTE when it is needed to make the case and where there is an opportunity for lessons that can be generalized. For programs that have not been ongoing for multiple years or lack the thorough documentation and evidence generation modeled by HarvestPlus, OTE could be more challenging. OTE can be used outside the policy domain if suitable middle-range theories can be identified and agreed upon.

An advantage of OTE is that it allows for an accumulation of learning from one evaluation to the next, as called for by Pawson (2013). Just as we took an existing middle-range theory from an evaluation of a similar type of change, so future evaluations of similar policy-oriented outcomes can build on our specified middle-range theory. The existence of existing specified middle-range theories from which evaluators can select will make it easier for them to bring social science theory into the development of ToCs, alongside stakeholder theory.

One important difference to note between OTE and contribution analysis is in the way the former deals with the causal questions at the heart of the latter, namely: (1) is the causal package sufficient to produce the outcome in question? and (2) are the individual elements of the package necessary? OTE assumes, a priori, that the outcome trajectory is the causal package. All the actors' strategies identified in the outcome trajectory are likely to be necessary. For OTE, the two questions become: (1) has the evaluation identified and described the main causal actors, strategies and their interplay? Such that (2) does the outcome trajectory provides a credible explanation of how the outcome in question emerged over time?

Another important difference is how OTE and contribution analysis work with theory. Contribution analysis develops a bespoke theory of change to describe how a program is assumed to have contributed to outcomes. It considers alternative ToCs that might better explain causal links. OTE, on the other hand, selects an existing middle-range theory to specify and help guide the causal explanation of the outcome trajectory, or trajectories, to which the program has contributed. The reasons for choosing the middle-range theory need to be explained, in particular why one middle-range theory was chosen over others, if alternatives exist.

There are lessons from OTE for future studies focused on other policy-related outcomes. For example, both positive and less positive results should be reported equally from efficacy and adoption trials to allow for balanced decision-making with respect to supporting one solution to a policy issue over another.

OTE can be used to evaluate other types of outcomes other than policy-related ones, if a suitable middle-range theory can be identified and agreed upon by the evaluation team and the commissioners of the evaluation, at a minimum, and preferably by representatives of the evaluand and other primary intended users. A possible candidate middle-range theory, called the three-pathway model, is being used in evaluations carried out within the Irish agricultural research, development and extension system, Teagasc (2021, p 59). The model was also adopted by the European Commission (2017) in a set of guidelines for evaluating innovation in rural development programs in Europe, based on Author et al. (2017). More recently still, the model was the basis of a lesson-learning review of how CGIAR has achieved development outcomes, to inform its next phase (Author et al. 2021).

Conclusions

This paper describes the experiences of using OTE to assess the contribution of HarvestPlus to the establishment of NARES biofortification breeding programs. Use of the method resulted in a much broader and more detailed understanding of the program's contribution than would have been the case if the study had focused only on the most proximate program activities, namely building capacity of NARES breeders and programs. It generated a number of insightful and useful findings.

OTE is a combination of elements selected from five evaluation approaches that have been used to evaluate policy-related and other types of outcomes—process-tracing, outcome harvesting, outcome evidencing, contribution analysis, and episode studies. One difference is the assumption, a priori, that program outcomes emerge from one or more outcome trajectories, that is, a patterned and evolving set of interactions between actors, knowledge, technology, and institutions, over extended periods of time. A program evaluation needs to show if and how the program has contributed to the outcome trajectory or trajectories. A second difference is the use by OTE of middle-range theory, based on 'grand' social science theory, to delineate and understand the outcome trajectory that generated the set of outcomes being evaluated. Doing so allows for a greater use of published theory in evaluations as well as providing a framework to allow learning to accrue from one evaluation to the next.

OTE will be a good methodological choice when evaluating program contribution to significant, achieved sets of outcomes. Other evaluation approaches, e.g., outcome evidencing, will be more appropriate for formative evaluations in programs that have not yet contributed to a significant outcome and wish to identify winners early.

Funding HarvestPlus was part of the CGIAR Research Program (CRP) on Agriculture for Nutrition and Health (A4NH), which ended in 2021 and funded the evaluative review. A4NH was supported by the CGIAR Trust Fund contributors (https://www.cgiar.org/funders/).

References

Axelrod, R., and M.D. Cohen. 1999. *Harnessing Complexity*. New York: The Free Press.

Baral, A., and E. Birol. 2020. Catalyzing the Scale-Up of Crop Biofortification. A4NH Strategic Brief. Washington, DC: International Food Policy Research Institute. https://ebrary.ifpri.org/utils/getfile/collection/p15738coll2/id/133959/filename/134167.pdf. Accessed 20Jan 2022

Baxter, P., and S. Jack. 2008. Qualitative Case Study Methodology: Study Design and Implementation for Novice Researchers. *The Qualitative Report* 13: 544–559. https://doi.org/10.46743/2160-3715/2008.1573.

Bouis, H.E., and A. Saltzman. 2017. Improving Nutrition Through Biofortification: A Review of Evidence from HarvestPlus, 2003 Through 2016. *Global Food Security* 12: 49–58. https://doi.org/10.1016/j.gfs.2017.01.009.

Carden, F. 2009. Understanding Influence: The Episode Studies Approach. In *The Politics of Trade: The Role of Research in Trade Policy and Negotiation*, ed. D. Tussie, 273–289. Danvers: Republic of Letters Publishing.

Collier, D. 2011. Understanding Process Tracing. *PS: Political Science and Politics* 44 (4): 823–830.

Dalberg. 2019. *Commercialization Assessment: Pearl Millet in India*. Washington, DC: Dalberg. https://nutritionconnect.org/sites/default/files/2020-01/191213_India_Iron%20Pearl%20Millet_Report_vFINAL%20%28002%29.pdf.

Douthwaite, B. (2020). Mainstreaming of biofortification in the African Union: Evaluation of CGIAR contributions to a policy outcome trajectory. CGIAR Research Programs on Roots, Tubers and Bananas & Agriculture for Nutrition and Health: Lima, Peru. https://doi.org/10.4160/9789290605522

Douthwaite, B. (2021) Study on HarvestPlus' contribution to the development of national biofortification breeding programs. https://doi.org/10.2499/p15738coll2.134880.

Douthwaite, B., & Gummert, M. (2010). Learning selection revisited: How can agricultural researchers make a difference? Agricultural Systems, 103(5): https://doi.org/10.1016/j.agsy.2010.01.005.

Douthwaite, B., Mur, R., Audouin, S., Wopereis, M., Hellin, J., Saley Moussa, A., Bouyer, J. (2017). Agricultural research for development to intervene effectively in complex systems and the implications for research organizations (KIT Working Paper No. 12). Royal Tropical Institute. https://agritrop.cirad.fr/584543/

Douthwaite, B. and Child, K. (2021). How agricultural research for development achieves developmental outcomes: learning lessons to inform One CGIAR science and technology policy research. Colombo, Sri Lanka: International Water Management Institute (IWMI). CGIAR Research Program on Water, Land Ecos WLE Legacy Series 2https://doi.org/10.5337/2022.2

Douthwaite, B., Proietti, C., Polar, B. & Thiele, G. (Forthcoming). Outcome Trajectory Evaluation (OTE): An approach to tackle research-for-development's long-causal-chain problem. American Journal of Evaluation.

Foley, J.K., K.D. Michaux, B. Mudyahoto, L. Kyazike, B. Cherian, O. Kalejaiye, O. Ifeoma, P. Ilona, C. Reinberg, D. Mavindidze, and E. Boy. 2021. Scaling Up Delivery of Biofortified Staple Food Crops Globally: Paths to Nourishing Millions. *Food and Nutrition Bulletin* 42 (1): 116–132. https://doi.org/10.1177/0379572120982501.

HarvestPlus. 2014. *Biofortification Progress Briefs*. Washington, DC: HarvestPlus. https://www.harvestplus.org/sites/default/files/Biofortification_Progress_Briefs_August2014_WEB_0.pdf.

HarvestPlus. 2019. *Biofortification: The Evidence. A Summary of Research Informing Scaling Up of Biofortification to Improve Nutrition and Health Globally*. Washington, DC: HarvestPlus. https://www.harvestplus.org/sites/default/files/publications/Biofortification%20Evidence.pdf.

HarvestPlus. n.d. *Engagement*. HarvestPlus. https://www.harvestplus.org/what-we-do/engagement. Accessed 10 Jan 2022.

Johnson, N., D. Asare-Marfo, M. Zeller, and E. Birol. 2017. Building the Case for Biofortification: Measuring and Maximizing Impact in the HarvestPlus Program. *African Journal of Food, Agriculture and Nutrition Development* 17 (2): 12078–12091. https://www.ajol.info/index.php/ajfand/article/view/155135/144752.

Jones, H. 2011. *A Guide to Monitoring and Evaluating Policy Influence*. London: Overseas Development Institute.

Kingdon, J.W. 1984. *Agendas, Alternatives, and Public Policies*. Boston: Little, Brown.

Leksmono, C., J. Young, N. Hooton, H. Muriuki, and D.L. Romney. 2006. *Informal Traders Lock Horns with the Formal Milk Industry: The Role of Research in Pro-poor Dairy Policy Shift in Kenya*. London: Overseas Development Institute.

Mayne, J. 2012. Contribution Analysis: Coming of Age? *Evaluation* 18 (3): 270–280.

Meenakshi, J.V., N. Johnson, V. Manyong, H. De Groote, J. Javelosa, D. Yanggen, F. Naher, C. Gonzalez, J. Garcia, and E. Meng. 2010. How Cost-Effective is Biofortification in Combating Micronutrient Malnutrition? An Ex Ante Assessment. *World Development* 38 (1): 64–75. https://doi.org/10.1016/j.worlddev.2009.03.014.

Pawson, R. 2010. Middle Range Theory and Program Theory Evaluation: From Provenance to Practice. In *Mind the Gap: Perspectives on Policy Evaluation and the Social Sciences*, ed. J. Vaesen and F.L. Leeuw, 171–202. New York: Routledge.

Pawson, R. 2013. *The Science of Evaluation: A Realist Manifesto*. London: SAGE.

Paz-Ybarnegaray, R., and B. Douthwaite. 2017. Outcome Evidencing: A Method for Enabling and Evaluating Program Intervention in Complex Systems. *American Journal of Evaluation* 38 (2): 275–293. https://doi.org/10.1177/1098214016676573.

Renkow, M. 2018. *Outcomes of Policy-Oriented Research in the CGIAR*. SPIA Technical Note No. 4. Rome: CGIAR Independent Science and Partnership Council (ISPC) Secretariat. https://cas.cgiar.org/sites/default/files/pdf/ispc_technicalnote_outcomes_policy_research_cgiar_0.pdf.

Resnick, D., S. Haggblade, S. Babu, S.L. Hendriks, and D. Mather. 2018. The Kaleidoscope Model of Policy Change: Applications to Food Security Policy in Zambia. *World Development* 109: 101–120. https://doi.org/10.1016/j.worlddev.2018.04.004.

Rijsberman, F. 2014. CGIAR Commits to Mainstreaming Breeding for Mineral and Vitamin Traits. https://a4nh.cgiar.org/2014/04/02/getting-nutritious-foods-to-people-2/. Accessed 10 Jan 2022.

Rule, P., and V.M. John. 2015. A Necessary Dialogue: Theory in Case Study Research. *International Journal of Qualitative Methods* 14 (4): 1609406915611575. https://doi.org/10.1177/1609406915611575.

Sabatier, P.A., and C.M. Weible. 2007. The Advocacy Coalition Framework: Innovation and Clarification. In *Theories of the Policy Process*, 2nd ed., ed. P.A. Sabatier, 189–220. Boulder: Westview Press.

Saltzman, A., E. Birol, A. Oparinde, M.S. Andersson, D. Asare-Marfo, M.T. Diressie, C. Gonzalez, K. Lividini, M. Moursi, and M. Zeller. 2017. Availability, Production, and Consumption of Crops Biofortified by Plant Breeding: Current Evidence and Future Potential. *Annals of the New York Academy of Sciences* 1390 (1): 104–114. https://doi.org/10.1111/nyas.13314.

Slade, R., P. Hazell, F. Place, and M. Renkow. 2002. Evaluating the Impact of Policy Research: Evidence from the Evaluation of Rural Policy Research in Developing Countries. *Evaluation* 26 (4): 541–561. https://doi.org/10.1177/1356389020931881.

Stachowiak, S. 2013. *Pathways for Change: 10 Theories to Inform Advocacy and Policy Change Efforts*. Washington, DC: Center for Evaluation Innovation. https://www.evaluationinnovation.org/publication/pathways-for-change-10-theories-to-inform-advocacy-and-policy-change-efforts/.

Teagasc. 2021. *Teagasc Statement of Strategy 2021–2024*. https://www.teagasc.ie/media/website/publications/2021/Teagasc-Statement-of-Strategy.pdf. Accessed 9 May 2022.

White, H., and D. Phillips. 2012. *Addressing Attribution of Cause and Effect in Small N Impact Evaluations: Towards an Integrated Framework*. 3ie Working Paper 15. New Delhi: International Initiative for Impact Evaluation (3ie) c/o Global Development Network. https://www.3ieimpact.org/evidence-hub/publications/working-papers/addressing-attribution-cause-and-effect-small-n-impact.

Wilson-Grau, R. 2019. *Outcome Harvesting: Principles, Steps, and Evaluation Applications*. Charlotte: Information Age Publishing.

World Food Prize Foundation. *Biofortification Pioneers Win 2016 World Food Prize for Fight Against Malnutrition*. https://www.worldfoodprize.org/index.cfm/87428/40322/biofortification_pioneers_win_20_world_food_prize_for_fight_against_malnutrition. Accessed 5 Jan 2022.

Zahariadis, N. 2008. Ambiguity and Choice in European Public Policy. *Journal of European Public Policy* 15 (4): 514–530. https://doi.org/10.1080/135017608019.