No.
29

HarvestPlus Working Paper | February 2018

Are We There Yet? The Rigorousness of Impact Evaluations of Nutrition-Sensitive Agricultural Interventions

Will Meister

HarvestPlus
Better Crops • Better Nutrition

HarvestPlus improves nutrition and public health by developing and promoting biofortified food crops that are rich in vitamins and minerals, and providing global leadership on biofortification evidence and technology. We work with diverse partners in more than 40 countries. HarvestPlus is part of the CGIAR Research Program on Agriculture for Nutrition and Health (A4NH). CGIAR is a global agriculture research partnership for a food secure future. Its science is carried out by its 15 research centers in collaboration with hundreds of partner organizations. The HarvestPlus program is coordinated by two of these centers, the International Center for Tropical Agriculture (CIAT) and the International Food Policy Research Institute (IFPRI).

CONTENTS

Are We There Yet? The Rigorousness of Impact Evaluations of Nutrition-Sensitive Agriculture Interventions

Will Meister[1]

Abstract

Several literature reviews have questioned the rigorousness of impact evaluations of nutrition-sensitive agriculture interventions. To understand how their rigorousness could be improved, this study seeks to establish a high-level overview of what rigorousness is, why it has not been realized, and how it could be improved through the use of randomized controlled trials (RCTs), long-term impacts and sustainability, and cost-benefit/cost-effectiveness analyses. To look further into these topics, the author interviewed a judgment sample of individuals from the agriculture-nutrition community. An understanding of rigorousness was derived from the interviews and literature, and was applied to biofortification studies connected to HarvestPlus. Issues related to funding, planning, time, knowledge, and intervention design were identified with regards to RCTs, long-term impacts and sustainability, and cost-benefit/cost-effectiveness analyses. Despite the arguments and difficulties with these three factors, the research community seems to be moving in the direction of greater rigor, as evidenced by some of the most recent research. Specifically, the impact evaluations of biofortification programs provide evidence that implementing more rigorous impact evaluations is possible. Given this movement, there is an opportunity to take a more standardized approach to impact evaluations of nutrition-sensitive agriculture interventions, using the weaknesses identified in the reviews to form the evaluation criteria.

1 Independent consultant. Email: will.meister@tum.de, Tel: (+49) 176-6134-5686

1. INTRODUCTION

Nutrition-Sensitive Agriculture Intervention

What defines a nutrition-sensitive agriculture intervention? Like other food-based approaches, a nutrition-sensitive agriculture intervention looks toward food to improve nutrition instead of toward fortification or supplementation. A nutrition-sensitive agriculture intervention differs from other intervention types, however, in its efforts to start at the root of the undernutrition problem—namely, if efforts are made to improve agriculture with nutrition in mind, then the impacts should reach those populations underserved by agriculture (Henson, Humphrey, and McClafferty 2013). Thus, nutrition-sensitive agriculture interventions are about bringing agriculture and nutrition closer together, because simply increasing agricultural production does not automatically equate to better nutrition. Rather, better nutrition depends on what is being produced and whether agriculture in the underserved areas can address the specific nutritional needs of the population in question (Henson, Humphrey, and McClafferty 2013). As Henson, Humphrey, and McClafferty (2013, 4) clearly summarized, "Agriculture should consider how the food it produces translates into good nutrition and better health."

However, it has been difficult to understand how agriculture interventions designed to improve nutrition can actually achieve this goal. Ruel (2001, 43, 50) concluded that some of the "most basic information needed to determine the usefulness of these strategies is simply not available" and "that evaluation designs must be improved. Habicht, Victoria, and Vaughan (1999; as cited in Ruel 2001) called for the focus to be placed on establishing plausibility, rather than causality, due to the complexity and difficulty of food-based approaches to micronutrient deficiencies. This call for a redirected focus illustrates the difficulty created by the novelty of nutrition-sensitive agriculture interventions. The long chain from the field to the table makes it difficult to causally link changes in agricultural production to improvements in nutrition (Global Nutrition CRSP 2012). This long chain means that many different stakeholders are involved in nutrition-sensitive agriculture interventions, requiring greater collaboration and coordination (Ruel 2001). In addition, the focus of these interventions is on enacting long-term change by altering both agricultural and eating practices, meaning changes to the way those targeted by the intervention live their lives (Ruel 2001). Thus, understanding the impacts requires a longer study period, which does not fit well into the modern obsession with instant gratification (Ruel 2001).

Despite the inherent challenges in nutrition-sensitive agriculture interventions, Ruel (2001, 67) summarized her review of the existing literature for food-based approaches as being in need of "the same

scientific rigor as other strategies" because of their potential to help in the fight against micronutrient deficiencies. A decade later, Webb and Kennedy (2012, 5) found that the "empirical evidence for impacts on nutrition ascribed to defined agricultural interventions is weak and mixed at best." They attributed this evidence gap to studies in this area having "significant methodological limitations."

Impact Evaluation

It is not only in the convergence of nutrition and agriculture that one finds criticism of the ability of these interventions to demonstrate impact; rather, impact evaluations of interventions as a whole have come under scrutiny. This scrutiny led to the emergence of 3ie (International Initiative for Impact Evaluation), the mission of which is to "increase development effectiveness through better use of evidence in developing countries" (3ie 2016). Leading to the development of this initiative was the identification of "the 'evaluation gap' ... [which] recognized the need for more rigorous ways to assess the impact of development interventions in order to give decision-makers the evidence they need" (3ie 2013). 3ie's intake of more than US$33 million in contributions and grants in 2015 provides strong evidence that impact evaluations in general are undergoing a period of improvement (3ie 2015). This improvement in the rigor of impact evaluations is partly based on the incorporation of cost-effectiveness analyses (3ie 2013). 3ie aims to help policy makers make decisions based on rigorous evidence that "links inputs to outcomes and impacts" (3ie, 2016; White 2009a, 7). As such, White (2009a, 7) identified "six key principles of a theory-based impact evaluation: 1) Map out the causal chain (programme theory); 2) Understand context; 3) Anticipate heterogeneity; 4) Rigorous evaluation of impact using a credible counterfactual; 5) Rigorous factual analysis; 6) Use mixed methods." Based on the evidence available, White (2009a, 18) called for "a greater volume of rigorous quantitative studies of what works in development," indicating the overall desire for improved impact evaluations in general.

As their popularity grows, impact evaluations are being used across an increasing variety of fields, as identified by Cameron, Mishra, and Brown (2016). Broadly speaking, the sectors for evaluation include health, nutrition and population, education, social protection, and agriculture, as well as economic policy, energy, finance, and transportation, among others. Within these sectors, a variety of different evaluation methods have been used. Notably, randomized controlled trials (RCTs) have been used the most in the health, nutrition, and population sector, at more than 80.0 percent of the identified impact evaluations, compared with the agriculture sector at 20.5 percent of the identified impact evaluations (Cameron, Mishra, and Brown 2016). Thus, one can see the important role impact evaluations have had on providing additional evidence for decision making.

Call for More Rigorous Evaluations

As summarized by Masset et al. (2011), methodological rigor has been affected by the failure to use RCTs and sufficient sample sizes, as well as conduct long-term studies. Ruel and Alderman (2013, 76) also called for the use of RCTs and the "assessment of cost and cost-effectiveness" in future research. The UK Department for International Development (DFID 2014) specifically highlighted the need for greater research on cost-effectiveness and sustainability in terms of both continuing practices and impacts. When researchers are able to address the issues of long-term costs, effectiveness, and sustainability, then those in charge of making policy decisions may be further awakened to the potential of agricultural interventions (Arimond et al. 2011, 70). The determinations made by Masset et al. (2011), for example, on how to improve the rigor of impact evaluations so that nutrition impacts are better understood, do not seem to fully explore the reasons for the failure of previous studies to achieve this level of rigor and how these issues can be ameliorated to improve future impact evaluations.

Randomized Controlled Trials

For some, RCTs are the strongest methods for impact evaluation and would be of the most benefit for researching the agriculture-nutrition link (Gertler et al. 2011, 50; Meeker and Haddad 2013, 11). The opportunities to use RCTs to examine causality has led some to declare them the "'gold standard' for program evaluation" (Bose 2010, 3; Habicht, Pelto, and Lapp 2009, 15). As explained by Meeker and Haddad (2013, 11), "The strongest methodology to determine attribution between agriculture and nutrition would be to conduct a randomized controlled trial." Similarly, according to Gertler et al. (2011, 50), "Randomized selection methods … represent the strongest methods for evaluating the impact of a program." The pursuit of understanding the link between interventional changes in agricultural practices and improvements in nutrition and health has led some researchers to call for an increase in "the use of randomized trials to determine the effectiveness of development assistance" (Winters, Salazar, and Maffioli 2010, 6).

In the criteria developed by Masset et al. (2011, 18) for rating the studies identified in their systematic review, a randomized experiment received a "high score" for the counterfactual analysis. However, the studies they reviewed were lackluster and failed to meet the threshold for a "high score," leading them to conclude that "rigorous counterfactual analysis has not been a major concern of this literature" (Masset et al. 2011, 29). The 23 papers selected for the review, as based on their search criteria, largely failed to achieve the level desired to be considered rigorous, with only three of those papers receiving a high score for their counterfactual analyses.

From their structural design alone, RCTs seem to provide certain benefits. The process of randomly selecting beneficiaries for an intervention is generally seen as being fairer and as allowing for greater transparency during the allocation of limited resources to a potential beneficiary population that is larger than the available resources (Gertler et al. 2011, 50). RCTs allow for a "valid solution" to the problem of having more willing participants than available resources by randomizing selection and creating control groups (Farley et al. 2012, 23). This approach allows for later phase-in of those not included in the earlier selection process (Gertler et al. 2011, 55). Within the target population, the use of randomized assignment to treatment and control groups helps ensure that the groups "will be similar not only in their observed characteristics but also in their unobserved characteristics" (Gertler et al. 2011, 52). This characteristic of randomized assignment also allows for comparability between treatment and control groups and helps limit bias (Gertler et al. 2011).

However, the use of RCTs in nutrition-sensitive agriculture interventions is not without controversy. The debate over the use of RCTs in development interventions stems from questions about how applicable RCTs are to this field (White 2009b, 16). As Cameron, Mishra, and Brown (2016) demonstrated, RCTs are not ubiquitous in agriculture-sector impact evaluations. In their examination of "current and planned research projects on agriculture for improved nutrition," Hawkes, Turner, and Waage (2012, 1, 30–31) looked at 151 research projects; of the projects that focused exclusively on an evaluation (28 projects), only four could be confirmed as RCTs.

Meeker and Haddad (2013, 11) clarified the following difficulties with implementing RCTs in this field: (1) "there is not much of a tradition of using these methods," (2) "there are concerns about cross-contamination of treatment and control," and (3) "the causal chains between agriculture and nutrition are long and complex." Despite these issues, they still encouraged "more experimentation with these methods in this domain" (Meeker and Haddad 2013, 11).

According to Scriven (2008, 12), the manifestation that RCTs have taken in development interventions is nowhere near how RCTs were designed to function, making their use in this area completely inapplicable. Evaluation methods, such as random treatment allocation and double-blind interventions, can not always be carried out in agriculture or food-based interventions. The results are studies that must show how likely it was that the intervention's outcomes can be attributed to the intervention itself and not to some other factor (Ruel 2001, 50). This limitation demands that evaluations consider the influence of confounding factors on the intervention's outcomes in order to rule out the effects of those factors (Ruel 2001, 50–51).

Other considerations, such as ethical and political constraints, also factor into this debate (White 2009b, 16). Furthermore, RCTs can be affected by a number of issues ranging from spillover effects to spontaneous adoption to self-selection (Farley et al. 2012, 23). This means that both the intervention and control groups could be influenced through various channels as participants move between the groups (Farley et al. 2012, 23). Attempting to conduct RCTs in an area as complex as agricultural interventions can also be "expensive and difficult to implement" (Masset et al. 2012, 4). Other researchers have echoed this idea; despite stating that "the optimal design to assess impact would be a randomized, controlled trial," Olney et al. (2009, 366) then qualified this statement by explaining that, as in many instances, limited resources prevented the application of this method to Helen Keller International's (HKI's) homestead food production (HFP) program in Cambodia.

A further criticism of RCTs is their lack of external validity. Devereux et al. (2013) and Deaton (2010) expressed these concerns when they emphasized the difficulty of applying the results of RCTs "outside the context in which they were obtained" (Deaton 2010, 449). Despite this criticism, however, in their list of recommendations to reorient nutrition evaluation research, Habicht, Pelto, and Lapp (2009, 52) recommended that RCTs would be best used for interventions "that have enough external validity (and define the conditions for this validity)" so as to increase the likelihood that results will be transferable.

For those interventions that fail to apply an experimental approach to an impact evaluation, self-selection bias becomes a relevant issue (Mareida 2009, 16). Self-selection bias is "bias [that] arises when individuals or groups receiving an intervention are self-selected or are assigned to the treatment group based on characteristics that may also affect their outcomes" (Mareida 2009, 16).

Long-Term Impacts

Evaluations of the long-term impact of food-based interventions are rare. Therefore, Girard et al. (2012, 218) and the Independent Evaluation Group (IEG 2011, x) called for a better understanding of how sustainable interventions truly are. Masset et al. (2011, 33) assumed that the paucity of long-term impacts stemmed from the shorter periods over which evaluations are conducted.

Cost-Effectiveness

The use of cost-benefit and cost-effectiveness analyses to disseminate findings does not occur as frequently as it should, according to White (2013, 46), who described cost-effectiveness analysis as "the best way to present results." IEG (2011, xii) also came to the conclusion that too few impact evaluations (IEs) were including cost-benefit analyses in evaluations, "even though IE methods and cost-benefit

analysis can complement each other." Hawkes, Turner, and Waage (2012, 30) also reported that "a very low number" of the projects they examined analyzed the cost-effectiveness of the intervention.

State of the Art

In autumn 2012, the Global Nutrition CRSP (Collaborative Research Support Program) organized a program with support from the Leverhulme Center for Integrative Research in Agriculture and Health (LCIRAH). They generated the following key questions:

- What are the actual linkages between agriculture and nutrition?
- What goals are we trying to achieve with the interventions—local improvements, progress to catch-up to a national mean, or the reaching of international standards/targets for outcomes and processes?
- What evidence do we need, at what level, rigor, and scale, to recognize causal or highly-plausible effects of complex interventions?
- What are the key metrics, and what essential data are needed by the different research communities to measure them? (Global Nutrition CRSP 2012, 1).

These questions shed light on the thinking of professionals in this field and indicate where the field was at the time of this study. Meeker and Haddad (2013) clearly summarized the state of the art in their paper: They explained that the pathways through which agriculture can improve nutrition have been identified, but the evidence supporting these pathways does not exist in sufficient quality and quantity to be able to link agriculture to improved nutrition. In addition, the reviews that have looked for this evidence have all come up somewhat empty-handed (Meeker and Haddad 2013).

This conclusion makes sense. If the studies analyzed were not aware of the criteria beforehand, how could they be able to fulfill it? The reviews were, however, an important element in raising questions of how to improve the rigorousness of these evaluations. Without having reviewed the quality of the work performed, how could improvements ever be suggested? For something to get better, there needs to be a baseline to measure it against; however, this also begs the question: What does *better* mean?

As an example, Cole et al. (2016) specifically addressed the weak points in nutrition-sensitive agriculture intervention research as identified by the systematic reviews performed to date. However, "agricultural economists, nutritionists, human health researchers, development specialists and evaluators hotly debate

what constitutes credible and actionable evidence to guide implementation and influence policy-making" (Cole et al. 2016, 12).

Focus of This Paper

In response to the aforementioned systematic reviews, a goal of this paper was to dissect the reasons the reviewed studies have failed to achieve the level of rigor needed to link agriculture to nutrition. However, it was important to start with a baseline of what rigor is and develop the reasoning from there, which is why the following questions were asked as part of this study:

- What is rigor in the context of impact evaluations of nutrition-sensitive agriculture interventions?
- Why has a level of rigorousness desired by the agriculture-nutrition community not yet been achieved with specific regard to RCTs, long-term impacts and sustainability, and cost-benefit/cost-effectiveness?
- How can the rigor desired by the agriculture-nutrition community be achieved?

At the heart of this study is the desire to determine whether the demands placed on improving the rigorousness of nutrition-sensitive agriculture interventions are achievable. As a nonexpert in this field, the author of this paper hopes to bring an outside perspective to an area that has undergone intense debate and criticism, with calls for increased rigor and responses questioning these calls.

This paper posits that the lack of rigor stems from fundamental limitations in how interventions can be structured. It posits that the nature of nutrition-sensitive agriculture interventions makes it difficult to achieve the desired level of rigor; however, with strong focus on improving the rigor of these interventions, a compromise will be found. Thus, studies released since the original research was conducted for this paper will be closer to addressing the issues identified by those in the field.

The next section provides an overview of how the research for this paper was conducted. The interview results are then summarized and discussed in comparison to the findings of other studies; they are also applied to studies on biofortification. The discussion focuses on what rigorousness means for impact evaluations. It also discusses why RCTs, long-term impacts and sustainability, and cost-benefit/cost-effectiveness analyses have not been frequently used and then explains how they could be used to increase rigor. Lastly, conclusions are given to add to the rigorousness of impact evaluations going forward.

2. SURVEY OF AGRICULTURE-NUTRITION EXPERTS

A judgment sample, with some elements of a convenience sample, was used to select participants using relevant literature, Internet searches, and snowball sampling. A list of 66 individuals was compiled. In addition, the Agriculture-Nutrition Community of Practice (Ag2Nut CoP) was contacted (https://knowledge-gateway.org/ag2nut), adding a potential 865 individuals that were members in February–March 2014.

A judgment sample was used instead of a random sample in order to target individuals with the expertise needed to answer the questions posed to them during the study. In this instance, individuals were selected for their specific expertise on the topic in a key informant sample. As explained by Marshall (1996, 523), "Qualitative researchers recognize that some informants are 'richer' than others and that these people are more likely to provide insight and understanding for the researcher."

After participants consented to take part in the study, their biographical data were collected. The study questions were directed toward understanding participants' thoughts on the following: the rigorousness of impact evaluations in their fields, RCTs, the paucity of long-term impacts, the links between agriculture and nutrition, intervention costs, frameworks, and metric standardization. To dive deeper into these topics, respondents were asked whether they would be willing to participate in a 15- to 30-minute follow-up Skype or telephone interview to recommend colleagues as potential participants and to make any general comments.

The interviewee-selection process focused on their years of experience, type of experience, and the level of engagement in a quantitative analysis, which has not been included in this paper. Availability during the study's time frame was also important.

Interviews with 12 individuals were conducted between March 26, 2014, and May 7, 2014. Self-selection bias could be present in the results. The small sample size was a result of the nature of qualitative research. A level of saturation was reached that seemed to suggest that continuing the interviews would not bring any additional findings. It should be noted that, in certain qualitative studies, researchers have been able to establish a consensus with as few as six participants (Mason 2010).

The interview consisted of 10 questions that served as a follow-up to topics covered in the questionnaire. A semistructured approach was used, and not all interviewees were asked each of the 10 questions. In general, the interview started with a question about the methodological rigorousness of impact

evaluations in this field. This question was typically followed by one about RCTs and rigor and the impact of RCTs on other methods. Interviewees were then typically asked about cost-effectiveness, research networks, minimum evaluation standards, guidance for implementers, and long-term impacts and sustainability.

The interviews were transcribed and analyzed first by openly coding them and then through categorization (Savin-Baden and Howell Major 2012). Word frequency clouds and tree maps were created to visually represent the data. The size of the codes in the tree maps is based on the number of references coded, whereas the color represents the number of coded items, with red representing more items coded and green representing fewer items coded.

3. SURVEY RESULTS AND DISCUSSION

What Is Rigorousness?

The criteria for rigorousness listed here serve as broad generalizations of the more common areas discussed by survey participants; this list does not pretend to be an all-inclusive definition of rigorousness in this field. Note that certain elements of the criteria were not without their objectors.

- Measures long-term impacts
- Conducts cost-benefit/cost-effectiveness analyses
- Uses a framework approach to guide research
- Examines nutrition impacts (using appropriate indicators)
- Uses sufficient sample sizes
- Is randomized
- Looks to understand causality
- Has clear and concise objectives
- Performs a contextual analysis

Figure 3.1 presents a tree map developed from the responses to the interview question about rigorousness.

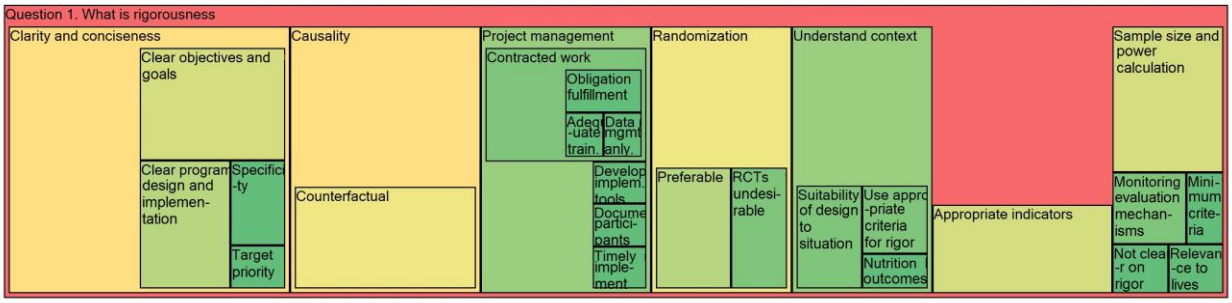

Figure 3.1 Tree map of Question 1: What is rigorousness? (source: Author's compilation)

Rigorousness of HarvestPlus Impact Evaluations

Using the criteria derived from this study for defining a rigorous evaluation, this paper asks how do select impact evaluations in Table 3.1 connected to HarvestPlus compare?

Table 3.1 Rigorousness of HarvestPlus impact evaluations (source: Author's compilation)

Study	Long-term impacts	Cost-effectiveness	Framework	Nutrition impacts	Sample size		Randomized	Causality	Clear objectives	Contextual analysis
					Intervention	Control				
Palmer et al. 2016: Provitamin A maize (PVAM)	6 months after baseline	Not included in study	Not specifically addressed	Vitamin A intake	*Cluster random sample of 542 children (4–8 years old), with 270 consuming white maize and 272 consuming PVAM	274 children not in intervention area	Cluster-randomized controlled study	Consumption of PVAM showed improvements in dark adaptation	Evaluate the impact of PVAM consumption on levels of impaired dark adaptation	Mkushi District, Zambia
Hotz, Loechl, de	2.5 years after	Second year of community	Not specifically	Vitamin A intake	*Cluster-sample	*129 children	Randomized controlled	Increase in OFSP consumptio	Evaluate scale-up and how	Focused on four districts in

Study	Long-term impacts	Cost-effectiveness	Framework	Nutrition impacts	Sample size		Randomized	Causality	Clear objectives	Contextual analysis
					Intervention	Control				
Brauw, et al. 2012: Orange-fleshed sweet potato (OFSP)	base-line	involvement by staff not cost-effective	addressed		*257 children (6–35 months) *263 women (mean 29.3 years in model 1; 28.4 years in model 2)	(6–35 months) *130 women (mean 29.1 years)	effectiveness study	n led to greater vitamin A intake	this affects OFSP consumption and vitamin A intake by targets	Zambézia Province, Mozambique
Hotz, Loechl, Lubowa et al. 2012: OFSP	2 years after base-line	Second year of community involvement by staff not cost-effective	Not specifically addressed	*Vitamin A intake *Vitamin A status	*Cluster-sample *166 children (6–35 months)	*105 children (6–35 months) *175 children	Randomized controlled effectiveness study	*Increase in OFSP consumption led to greater vitamin A intake	*Evaluate scale-up *Vitamin A intake	Focused on three districts in eastern and central Uganda

Study	Long-term impacts	Cost-effectiveness	Framework	Nutrition impacts	Sample size		Randomized	Causality	Clear objectives	Contextual analysis
					Intervention	Control				
					*306 children (3–5 years) *342 women (mean 32.8 years in IP group; 35.8 years in RP group)	(3–5 years) *213 women (mean 33.9 years)		*Children showed improved vitamin A status		
Low et al. 2007a, 2007b: OFSP	2 years after base-line	Less than US$0.01 to meet vitamin A	Program theory	*Vitamin A intake *Serum retinol	490 children (mean 35 months)	243 children (mean 35 months)	Quasi-experimental randomized	*Vitamin A intake improved *Serum retinol	*Formulate baseline *Identify intermediat	Focused on three districts in Zambézia Province,

Study	Long-term impacts	Cost-effectiveness	Framework	Nutrition impacts	Sample size		Randomized	Causality	Clear objectives	Contextual analysis
					Intervention	Control				
		needs of each child		concentration			controlled study	concentration improved	e outcomes *Impact on serum retinol concentration	Mozambique

The PATH Nutrition Embedding Evaluation Programme evaluated nutrition interventions using a "scoring system to assess strength of evidence" (PATH 2014, 6). When looking at biofortification interventions, PATH found that these interventions have focused mostly on measuring efficacy and have not examined scale-up except for with orange-fleshed sweet potato (PATH 2014, 8). The biofortification study that PATH scored (Hotz, Loechl, Lubowa, et al. 2012) received a "strength of evidence" score of "Limited," losing points for size, context, and consistency; however, the quality of the scored study was rated high, given that an experimental effectiveness design was used (PATH 2014, Annex 2).

DFID (2014, 44–45) reported that the eight biofortification studies it reviewed were "of moderate quality overall." Of particular note were the studies carried out by Hotz, Loechl, de Brauw et al. (2012) and Hotz, Loechl, Lubowa et al. (2012), which both used RCTs and included cost-effectiveness and sustainability analyses. Low et al. (2007a) appeared to base the intervention partially on the Sustainable Livelihoods Framework (S259).

According to the criteria outlined for an intervention to be rigorous, it appears that the aforementioned studies are strong candidates for being considered more rigorous. However, the use of RCTs would not equate to greater rigor for all. Regarding scale-up, the issue of external validity is also present in terms of how transferable the interventions are to other geographic areas.

For the issue of sustainability, the question of future plans can be raised in order to investigate the long-term impacts of these interventions—for example, "What is the impact 30 years from now?" However, the time frame over which HarvestPlus seeks to create impact must also be taken into consideration, because biofortified foods are seen as a segue to a more diverse diet.

In its cost-effectiveness analysis of orange-flesh sweet potatoes (OFSP) in Mozambique and Uganda, HarvestPlus compared two models of the intervention, as well as "a full range of subcomponents, activities, and messages" (HarvestPlus 2010, 451). By calculating costs for different activities within the intervention, this represents an important step toward increased rigor, in that HarvestPlus was then better able to identify which intervention activities were less cost-effective. Based on the recommendations of certain survey participants, HarvestPlus has achieved greater rigor through the application of this strategy.

HarvestPlus has also performed a number of ex ante cost-effectiveness analyses (HarvestPlus 2014, 47). These analyses seem to be important first steps to performing ex post analyses to evaluate the assumptions and projections made in the ex ante analyses. As a first step, the ex ante analyses serve to start the conversation on how to analyze the cost-effectiveness of biofortification interventions.

Based on the evidence to date, HarvestPlus (2016, 1) has been able to conclude that biofortification "can improve micronutrient status" cost-effectively. Randomized controlled efficacy trials have shown that biofortification can be an efficacious solution to the lack of micronutrients available in the diets of certain undernourished populations (HarvestPlus 2016). The cost-effectiveness of biofortification interventions has also been demonstrated through the OFSP intervention, which has been identified as being highly cost effective, according to the World Bank (HarvestPlus 2016, 4). Specifically looking at RCTs and cost-effectiveness, HarvestPlus has demonstrated that this level of rigor is achievable in nutrition-sensitive agriculture interventions.

Why Has Rigor Not Been Achieved?

Randomized Controlled Trials

As Figure 3.2 shows, those in favor of RCTs (eight interviewees in total) highlighted various reasons for their use, including fairness, transparency, credibility, and causality. Within this group, certain interviewees expressed a strong preference for RCTs over other methods; however, this preference was met with warnings from three interviewees that RCTs should only be used when appropriate, as was similarly expressed by Noonan and Kristjansson (2009).

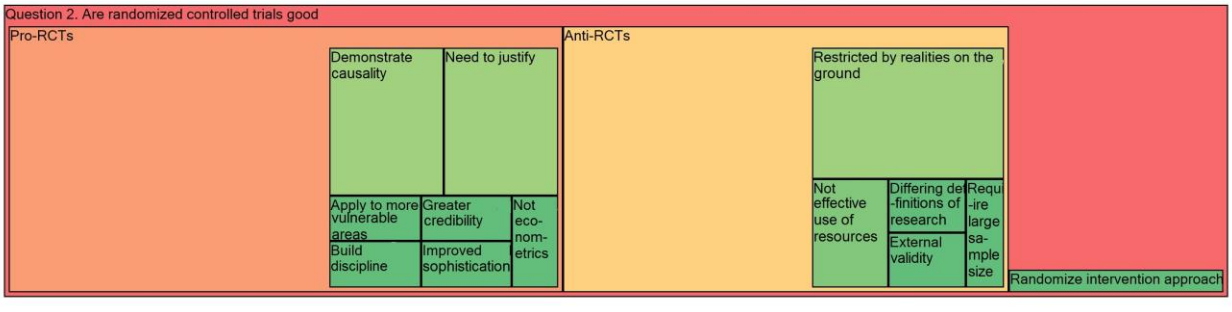

Figure 3.2 Tree map of Question 2: Are randomized controlled trials good? (source: Author's compilation)

As Figure 3.2 also shows, four interviewees argued against the use of RCTs; however, two of these interviewees also made arguments for their use, providing further evidence of the divided culture that RCTs seem to have created. Understanding the arguments against RCTs helps clarify why they were not used more frequently in previous studies. Overall, the feasibility of RCTs was questioned by Meeker and Haddad (2013, 11), based on the limited history of their use in agriculture evaluations and the length and complexity of "the causal chains between agriculture and nutrition."

Funding: Masset et al. (2012) highlighted the difficulties of implementing RCTs and their prohibitive costs. One interviewee also seemed to echo these concerns, citing budget limitations as a barrier to conducting RCTs (Figure 3.2). The cost of RCTs is also addressed by White (2013, 44), who emphasized that although primary data collection

is expensive, it is no more so than for other methods that also rely on data collection to the same degree as RCTs. Moreover, some quasi-experimental methods can be more expensive because not all of the data collected can be used (White 2013).

Operational rules of interventions: If RCTs are to become a priority in future research, then the way interventions are designed could be looked into with greater detail. Gertler et al. (2011, 144) countered a discussion of why randomization is so advantageous by focusing on its feasibility, specifically mentioning "conflict[s] with the operational rules of the program." Similarly, Roetman (2011, 3) explained that the use of RCTs "may go against the strategy and principles of many development agencies." Habicht, Pelto, and Lapp (2009, 5) stated that RCTs are "programmatically … difficult."

When asked if other methods serve as a barrier to the use of RCTs, interviewees expressed a level of uncertainty, as was apparent from the frequency of word usage displayed in Figure 3.3. In response to this question, interviewees used words such as *depends, change, objections,* and *dismiss*. However, some interviewees expressed strong viewpoints that RCTs have not influenced the use of other methods (Figure 3.4). These interviewees explained that the method selected was that which was most appropriate to the objectives and context of the evaluation.

According to Scriven (2008), an elitist group of RCT proponents had constructed a campaign to discredit the use of research methods other than RCTs. In response to this idea, two interviewees discussed the pushback they have witnessed stemming from the emphasis that RCTs have received (Figure 3.4). Overall, these interviewees seemed to indicate that placing too much emphasis on any one method was not the best decision and that the selection and acceptance of research methods should demonstrate greater tolerance.

Despite these objections, RCTs are being implemented, as evidenced by Olney et al. (2015, 1318), who specifically referenced the "weaknesses in this existing body of evidence." They used "a cluster-randomized controlled trial design to evaluate the impact of a carefully designed enhanced-homestead food production program implemented by an experienced nongovernmental organization ... on child anthropometry, Hb, anemia, and diarrhea" (Olney et al. 2015, 1318). Olney et al. (2015) gives an example of how the research community is responding to the findings of the systematic reviews that call for greater rigor.

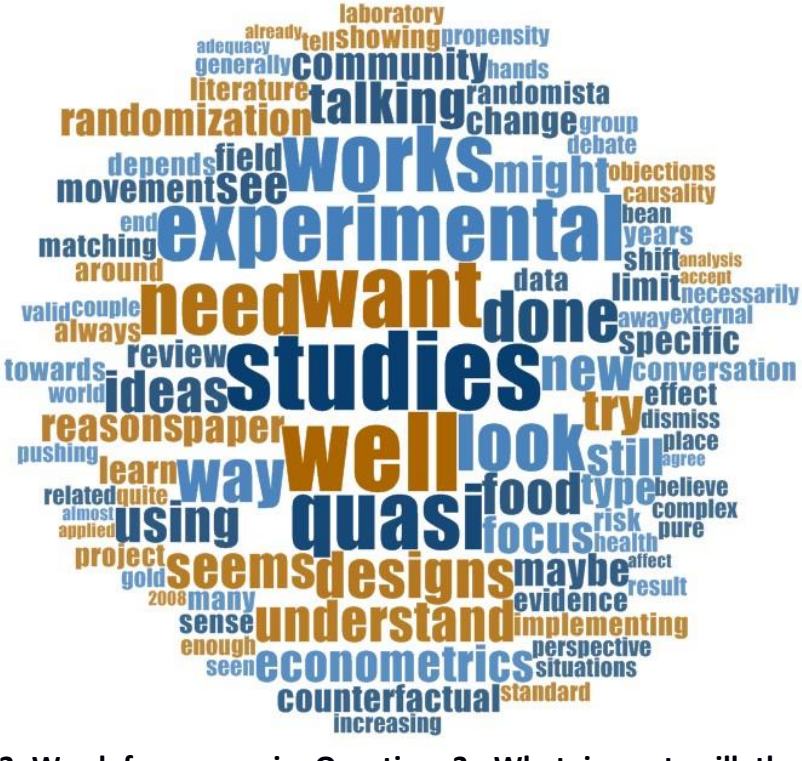

Figure 3.3 Word frequency in Question 3: What impact will the use of randomized controlled trials have on other research methods? (source: Author's compilation)

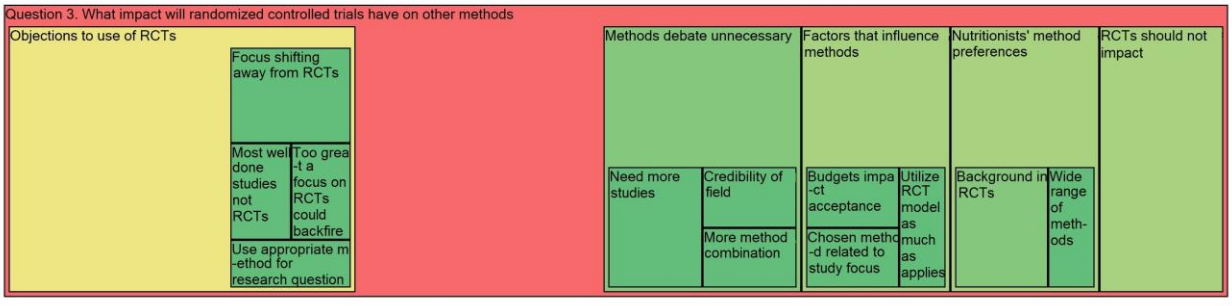

Figure 3.4 Tree map of Question 3: What impact will the use of randomized controlled trials have on other research methods? (source: Author's compilation)

Long-Term Impacts and Sustainability

Winters, Maffioli and Salazar (2011, 399) explained that evaluations typically "focus on impact indicators that can be measured in the period following the intervention, which often miss long-term effects and issues of project sustainability." They followed this statement by calling for a longer time frame for evaluating agriculture interventions. Similarly, Masset et al. (2011, 33) found that evaluations tended to favor the measurement of "short term indicators of hunger (wasting and underweight) … [than] long-term indicators (stunting)"; they attributed this to the short-term focus usually taken by evaluations.

Furthermore, time has to pass for researchers to be able to measure impacts, as pointed out by another interviewee (Figure 3.5). In contrast, Gertler et al. (2011, 160) explained that long-term impacts may not always be detectable because impacts could diminish over time—a concern that was also discussed by an interviewee (Figure 3.5), who thought that causality would be difficult to show over a longer time frame. This concern is understandable; however, one would hope to see sustainable impacts from the intervention. Of course, one could argue that the difficulty in understanding such long-

term impacts is the unknown of what may have happened between the end of the intervention and the time of the evaluation of the long-term impacts.

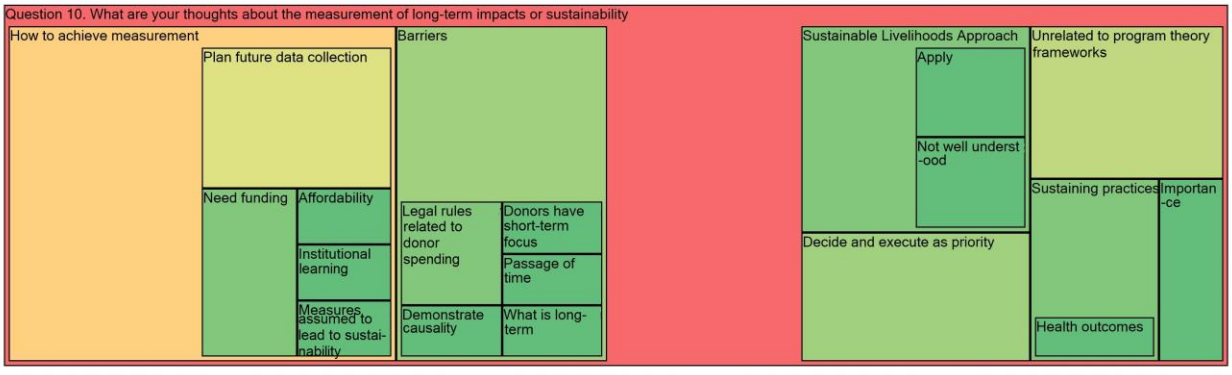

Figure 3.5 Tree map of Question 10: What are your thoughts about the measurement of long-term impacts/sustainability? (source: Author's compilation)

Some interviewees mentioned lack of funding as a concern, indicating the lack of financial support provided for understanding the long-term impacts that interventions may have (Figure 3.5). Looking to HKI's HFP program in Bangladesh as an example of a long-term impact and sustainability study, such a program "has had a variety of bilateral and multilateral funding sources," speaking to the complexity of funding such an intervention and the evaluations thereof (Iannotti, Cunningham, and Ruel 2009, 1). If several different funding sources have to be acquired in order to fund these types of interventions, the difficulties in finding funding for a long-term impact study become apparent.

When asked about long-term impacts, one interviewee discussed how donors can influence the length of the intervention and evaluation based on their expectations (Figure 3.5). One could argue that a connection between donors' expectations and a lack of funding may exist—if donors have expectations that evaluations will be shorter term, then fewer resources may be provided. This could explain why respondents

tended to agree more with these two factors (donor's expectations and lack of funding). However, with donor agencies showing a growing interest in impact evaluations in order to get "better evidence on the effective use of development resources," there is a possibility that donors will begin to expect impact evaluations that look at long-term impacts due to the potential to understand how their invested resources have paid off (Gertler et al. 2011, 226).

Some interviewees emphasized the importance of planning when discussing how to improve the understanding of the long-term impacts and the sustainability of interventions (Figure 3.5). Planning encompasses many aspects of what is needed to successfully conduct a long-term study, including funding and adequate documentation. The discussion also focused on incorporating planning into the overall design of an intervention and evaluation plan to ensure that it is made a priority. Winters, Salazar and Maffioli (2010, 57) discussed the importance of planning, specifically with reference to "providing guidance for future project leaders, government officials or evaluators," who may later be responsible for the intervention and who would need to know what was planned. Planning to evaluate the long-term impacts of an intervention before the intervention is implemented likely will result in a much better chance that such impacts will then be evaluated (Gertler et al. 2011).

Cost-Benefit/Cost-Effectiveness Analyses

When interviewees were asked about examples of cost-effectiveness studies, it was difficult for them to provide examples. However, as Figure 3.6 shows, a few examples were provided.

In trying to understand why cost-benefit analysis is not an integral part of an impact evaluation, IEG (2011, 12) simply stated that the answer is unclear. However, interviewees did provide various examples of limitations that they thought have affected the use of cost-effectiveness analyses (Table 3.2). Two respondents

questioned the application of this method to this field of study, citing the need to understand which benefits to measure and how to measure them, as well as the many assumptions involved. Another respondent who commented on this theme stated that cost-benefit analysis is not needed at all.

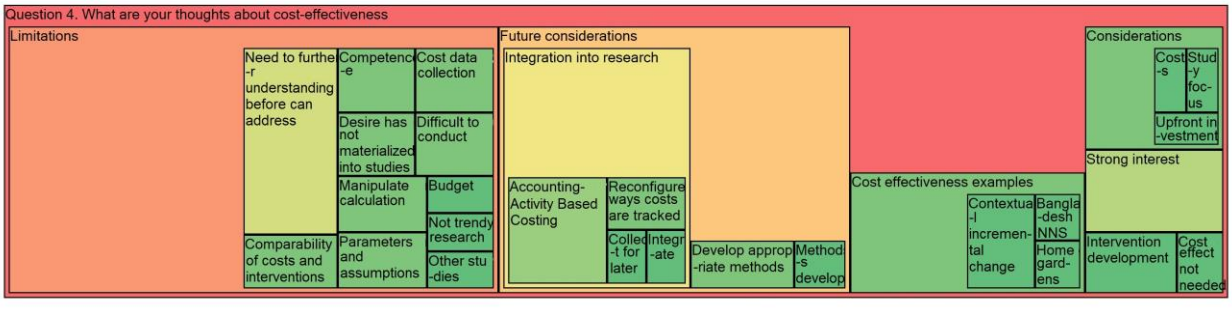

Figure 3.6 Tree map of Question 4: What are your thoughts about cost-effectiveness? (source: Author's compilation)

Table 3.2 Limitations of cost-effectiveness analysis in nutrition-sensitive agriculture interventions (source: Author's compilation)

Limitation	Number of interviewees
Budget (external consultant)	1
Comparability of costs and interventions	2
Competence	2
Collecting data on costs	2
Desire has not materialized into studies	2
Difficult to conduct	2
Manipulatable calculation	2
Need for further understanding before can address	5
Not trendy research topic	1
Other studies needed for comparison	1
Parameters and assumptions	2

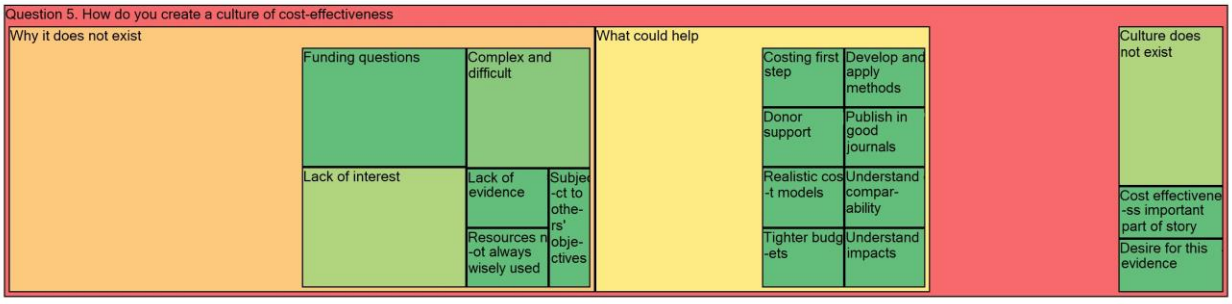

Figure 3.7 Tree map of Question 5: How do you create a culture of cost-effectiveness? (source: Author's compilation)

One interviewee referred to the importance of incorporating the collection of data on costs into the intervention design, so that adequate plans are made to ensure its collection (Figure 3.6). Moreover, this interviewee discussed how, prior to implementation, decisions should be made regarding how costs will be assigned—whether through activity-based costing or some other method—due to the problems associated with retrospective cost allocation.

Two interviewees brought up lack of knowledge or understanding as a barrier to conducting these types of analyses (Figure 3.6). Furthermore, a common theme among interviewees was the need to further the understanding of not only cost-effectiveness but also nutrition-sensitive agriculture interventions themselves. This theme was also discussed by Webb (2013, 11), who wrote of the need to understand the impacts of the interventions first and foremost. As explained by two interviewees (Figure 3.7), the complexity of nutrition-sensitive agriculture interventions has left some organizations with a lack of knowledge or understanding of how to conduct cost-effectiveness analyses. Within its assessment of agriculture impact evaluations, IEG (2011, xii–xiii) hypothesized that the lack of cost measurement could stem from evaluators not having the skills necessary to "measur[e] project costs along with impacts." Based on these findings, this appears to be an area of extreme importance if the use of cost-

benefit/cost-effectiveness analyses in nutrition-sensitive agriculture interventions is to be increased.

4. HOW TO ACHIEVE DESIRED RIGOR

One approach to achieving the desired level of rigor could be the Sustainable Livelihoods Approach, which was not mentioned by any interviewees when they were asked about long-term impacts and sustainability. This lack of response could be an indication of the impact this approach has had. When asked specifically about this approach, however, one interviewee thought the agriculture-nutrition community showed a lack of understanding of it (Figure 3.5). However, as Knutsson (2006, 90, 92) explained, "Important donor institutions … [have] adopted the Sustainable Livelihoods Approach (SLA) as [the] basis for their development programs and practices," which in large part could be due to it being one possible way "to handle complex problems with a more integrated and holistic perspective." It should be noted that Knutsson (2006, 96) also stated that more work needs to be carried out to expand the use of this approach.

To measure sustainability, one interviewee suggested that interventions should measure outcomes that have been hypothesized to lead to enhanced sustainability (Figure 3.5); the frequency with which this is tested could be higher, and furthermore, it has not been randomly tested. The advantage of this approach is that it allows one to measure outcomes in the present that are hypothesized to lead to impacts in the future; however, without verification of these impacts in the future, support for this suggestion is hard to muster.

According to one interviewee, identifying *what* is measured is critical; the potential to measure long-term impacts would be better served by examining how the health of beneficiaries has been affected rather than by understanding whether beneficiaries continued to practice the measures they were instructed to follow (Figure 3.5). Another

interviewee referred to a study that looked at sustainable practices when citing how long-term impacts have been measured. That study was by Bushamuka et al. (2005) and argued for a similar suggestion to look at outcomes that are hypothesized to lead to impacts. The study by Bushamuka et al. (2005) did just that; thus, the next era of evidence will need to go a step further to examine the hypothesized impacts. In a field designed to treat the underlying causes of undernutrition by addressing the root of the problem, sustainability and understanding the long-term impacts of interventions are critical elements (Herforth and Ballard 2016). It is systematic change at work, and these changes take time, which means that impact evaluations will also have to reflect this.

As the word cloud in Figure 3.8 shows, the creation of a culture of cost-effectiveness reveals such words as *incentives, donors, desirable, years, time,* and *funds* (when looking beyond more frequently used words). The terms *researchers, community,* and *trying* were also mentioned by interviewees with some frequency. These words appear to provide insight into the factors needed to create such a culture within the research community.

Figure 3.8 Word frequency in Question 5: How do you create a culture of cost-effectiveness? (source: Author's compilation)

The repeated calls for cost-effectiveness studies (Hawkes, Turner, and Waage 2012; Masset et al. 2012; Ruel 2001; White 2013) appear to have had an impact. As mentioned during an interview, methods are being developed to study this aspect of interventions (Figure 3.6). This interviewee placed an emphasis on ensuring that methods developed to analyze cost-benefit and cost-effectiveness would be able to fully comprehend nutrition-sensitive agriculture interventions by looking at the entire package of outcomes, rather than simply looking at one or two specific outcomes. Similarly, Mareida (2009, 22) discussed the use of cost-benefit analyses and the difficulties associated with the traditional cost-benefit analysis framework when applied to agriculture interventions, with their "people-related impacts." Expanding on this idea, Mareida (2009, 22) called for the development of new approaches to cost-benefit analysis, such as the "extended cost-benefit analysis." The social return on investment

(SROI) could be another possibility because it looks at the "economic, social and environmental costs and revenues of a project" (Roetman 2011, 12).

Moreover, the idea of collecting cost data now for future use when methods or resources may become available to conduct such studies, was discussed by one interviewee (Figure 3.6) as a possible way to approach this knowledge gap. This approach would mean doing the legwork now to make analysis possible in the future. Activity-based costing and the use of cost centers were specifically cited (Figure 3.6) as methods to be used now to record cost data for future use in analysis, as well as for budgeting. Ultimately, publishing is a key component; as one interviewee stated, and without this incentive, researchers have little motivation to look into this analysis (Figure 3.6).

As found by Herforth and Ballard (2016), the evidence gaps identified in the highlighted systematic reviews to date are now being filled by the next round of studies. Key performance indicators in nutrition-sensitive agriculture research have evolved to better establish the connection between nutrition and agriculture (Herforth and Ballard 2016). By examining the evidence to date, Herforth and Ballard (2016), as with the systematic reviews highlighted in the literature, have taken a critical eye to the research; as such, they asked: Does what we are doing make sense?

5. CONCLUSIONS

This study looks at the agriculture-nutrition community's thinking regarding the level of rigorousness in impact evaluations of nutrition-sensitive agriculture interventions. Members of this community were interviewed to gain an understanding of what rigorousness is, why it has not been achieved, and how it can be achieved for impact evaluations of nutrition-sensitive agriculture interventions. With a specific focus on RCTs, long-term impacts and sustainability, and cost-benefit/cost-effectiveness

analyses, the impression garnered from this study is that it will be no easy feat to improve the rigor in impact evaluations of nutrition-sensitive agriculture interventions. However, the growing interest and continued development of research in this field seem to equate to strong assets for improvements in rigor.

The interview results indicate that RCTs are a viable option for impact evaluations of nutrition-sensitive agriculture interventions—at the least, they are an option that should be further pursued. Despite the objections to RCTs, there seems to be a strong push to better understand the potential of this method to establish causality. In certain cases (such as biofortification), RCTs have been shown to be implementable. With regards to long-term impacts, it is a matter of matching expectations to reality and waiting for impacts to become longer term. Nutrition-sensitive agriculture interventions seek to treat one of the underlying causes of undernutrition, and doing so takes time. Moreover, treating these underlying causes also seems to be a matter of gaining an understanding of what nutrition-sensitive agriculture interventions are capable of achieving, as explained by Herforth and Ballard (2016).

Similar to RCTs, cost-effectiveness is another area that, though being questioned, seems to be a viable option to help understand how effectively nutrition-sensitive agriculture interventions use the resources made available to them. Again, certain interventions have demonstrated the applicability of cost-effectiveness to nutrition-sensitive agriculture interventions. What seems to be a recurring theme throughout these areas of rigor is funding. For example, Cole et al. (2016) worked with a funder who shared their research goals, which seems to be an important element of achieving greater rigor.

The key to understanding whether rigor will be achieved is to first understand how researchers in the field have responded to the likes of the reviews of Masset et al. (2011). This entailed listening to the criticism and taking action in response to the calls

for greater rigor. As shown in Cole et al. (2016), the specific weaknesses identified in the systematic reviews to date (Masset et al. 2012; Ruel and Alderman 2013; Webb 2013; Webb and Kennedy 2012) were taken into consideration when designing the Mama SASHA program (Sweet potato Action for Security and Health in Africa), which sought to increase the production and consumption of biofortified OFSP. By pursuing the specific areas of RCTs, long-term impacts and sustainability, and cost-benefit/cost-effectiveness analysis, one could ask whether it is important to develop a set of standardized criteria to help organizations create interventions that focus their efforts on these identified points of weakness. Although the idea of standardization is somewhat controversial because each study is unique, the use of specific criteria to understand the intervention design could help bring the research on agriculture-nutrition interventions closer to that desired in the systematic reviews performed to date.

As an example, Haddad (2013, 41) wrote that "a new systematic review of agricultural interventions that actually seek to improve nutrition status found only 23 evaluations between 1990 and 2010 that were able to establish a credible counterfactual." He continued: "So half of all agricultural interventions *with the express purpose of improving nutrition* ... have a zero impact on nutrition status" (Haddad 2013, 41). The idea of singling out specific studies presents an interesting opportunity to develop a certain set of standard criteria to rate interventions. The idea would not be to force each intervention to perform according to these criteria; rather, it could be to help identify how this intervention compares to the criteria set forth in order to increase the rigor of impact evaluations in agriculture-nutrition interventions.

Such an approach could also help simplify a rating system. As explained by Wilczek (2013), a complex problem should be made understandable by a computer; essentially, it should be a string of 0s and 1s, which, in the case of the intervention criteria, would

be a series of yes's and no's—as in, "Did the intervention use an RCT to establish causality?" Along similar lines, Scriven (2010, i) proposed the use of a checklist to improve evaluations, "inspired by the technique that aeronautical engineers developed to reduce plane crashes due to pilot error, and that civil engineers use to reduce mistakes in the construction of large buildings." The idea of standardization was further encouraged by the World Bank (2014, 33), which recommended the "establish[ment of] a new common vision globally for agriculture's role in improving nutrition, with measurable outcomes and targets."

The argument against such an approach is that not all interventions can fulfill such criteria based on their individual design needs; this approach would automatically result in them receiving a lower score despite the quality research that would be performed on them. And the criteria would, of course, be only one set of standardized criteria. Other sets of standardized criteria could also be developed to refute the claims of these standardized criteria. In examining the indicators used in nutrition-sensitive agriculture interventions, Herforth and Ballard (2016, 6) warned against the use of "standard indicators"; although these "facilitate comparison of results across studies, it is also important that indicators be selected for plausible and realistic outcomes for each project's objectives, activities, and statistical power." Further support for this argument comes from the World Bank (2014), which placed an emphasis on asking the right kinds of questions for nutrition-sensitive agriculture interventions. An example of this is the Food and Agriculture Organization's use of different categorical terminology for "'nutrition-sensitive agriculture' (that increases access to nutritious food and ensures no negative effects on other causes of malnutrition) and 'nutrition-specific agriculture' (that enhances nutrient content of food, such as biofortification or zinc fertilizers)" (World Bank, 35). Understanding this differentiation can also shed light on why biofortification impact evaluations have achieved favorable ratings and are able to fulfill the criteria set out in the systematic reviews.

The fact that the research community is responding to calls for improved rigor calls into question the author's hypothesis that the complexity of nutrition-sensitive agriculture interventions fundamentally limits them from achieving this rigor. Interventions are using an RCT approach, are beginning to look at their long-term impacts, and are seeking to understand the cost-effectiveness of such interventions. However, as Herforth and Ballard (2016) found, if nutrition-sensitive agriculture interventions are to have the impacts they desire (such as to reduce stunting), then they need to be large and long, which may not be the most effective use of limited resources. Moreover, Pinstrup-Andersen (2013; as cited in World Bank 2014, 35), in response to the conclusion that nutrition-sensitive agriculture interventions need to ask the right kinds of questions, stated that "few of these kinds of questions can be answered by randomized controlled trials; many would require population-level analyses of food and dietary trends, related to food and agricultural policies and investments." Thus, the nutrition-sensitive agriculture interventions performed prior to the influential reviews, such as Masset et al. (2011), appear to have been an attempt to ask a different set of questions that were not the most appropriate for the answers desired. With the interest in nutrition-sensitive agriculture interventions, there appears to be a shift toward asking a new set of questions more related to the achievable goals of such interventions.

REFERENCES

3ie (International Initiative for Impact Evaluation). 2013. *From Influence to Impact: 3ie Strategy 2014–2016.* New Delhi.

———. 2015. "Federal Form 990." Washington, DC.

———. 2016. "Strategy & Vision." Accessed October 30, 2016. http://www.3ieimpact.org/en/about/strategy-vision/.

Arimond, M., C. Hawkes, M. T. Ruel, Z. Sifri, P. R. Berti, J. L. Leroy, J. W. Low, L. R. Brown, and E. A. Frongillo. 2011. "Agricultural Interventions and Nutrition: Lessons from the Past and New Evidence." In *Combating Micronutrient Deficiencies,* edited by B. Thompson and L. Amoroso, 41–75. Rome: Food and Agricultural Organisation of the United Nations; Oxfordshire, UK: CAB International.

Bose, R. 2010. *A Checklist for the Reporting of Randomized Control Trials of Social and Economic Policy Interventions in Developing Countries: CEDE Version 1.0.* 3ie Working Paper. New Delhi: International Initiative for Impact Evaluation.

Bushamuka, V. N., S. de Pee, A. Talukder, L. Kiess, D. Panagides, A. Taher, and M. Bloem. 2005. "Impact of a Homestead Gardening Program on Household Food Security and Empowerment of Women in Bangladesh." *Food and Nutrition Bulletin* 26 (1): 17–25.

Cameron, D. B., A. Mishra, and A. N. Brown. 2016. "The Growth of Impact Evaluation for International Development: How Much Have We Learned?" *Journal of Development Effectiveness* 8 (1): 1–21.

Cole, D. C., C. Levin, C. Loechl, G. Thiele, F. Grant, A. Webb Girard, K. Sindi, and J. Low. 2016. "Planning an Integrated Agriculture and Health Program and Designing Its Evaluation: Experience from Western Kenya." *Evaluation and Program Planning* 56: 11–22.

Deaton, A. 2010. "Instruments, Randomization, and Learning About Development." *Journal of Economic Literature* 48 (2): 424–455.

Devereux, S., K. Roelen, C. Bene, D. Chopra, J. Leavy, and J. A. McGregor. 2013. *Evaluating Outside the Box: An Alternative Framework for Analysing Social Protection Programmes.* IDS Working Paper 431. Brighton, UK: Institute of Development Studies.

DFID (UK Department for International Development). 2014. *Can Agriculture Interventions Promote Nutrition?* Agriculture and Nutrition Evidence Paper. London.

Farley, K., S. Lucas, J. Molyneaux, and K. Penn. 2012. *Impact Evaluations of Agriculture Projects.* Millennium Challenge Corporation: Principles into Practice. Washington, DC: Millennium Challenge Corporation.

Gertler, P., S. Martinez, P. Premand, L. Rawlings, and C. Vermeersch. 2011. *Impact Evaluation in Practice.* Washington, DC: World Bank.

Girard, A. W., J. L. Self, C. McAuliffe, and O. Olude. 2012. "The Effects of Household Food Production Strategies on the Health and Nutrition Outcomes of Women and Young Children: A Systematic Review." *Paediatric and Perinatal Epidemiology* 26: 205–222.

Global Nutrition CRSP (Collaborative Research Support Program). 2012. *Integrating Agriculture and Nutrition Actions to Improve Maternal and Child Nutrition: Metrics for Coordinated Research.* Nutrition CRSP Research Briefing Paper No. 5. Boston: Global Nutrition CRSP and Tufts University.

Habicht, J. P., G. H. Pelto, and J. Lapp. 2009. *Methodologies to Evaluate the Impact of Large Scale Nutrition Programs.* Doing Impact Evaluation Paper No. 13. Washington, DC: World Bank.

Habicht, J. P., C. G. Victoria, and J. P. Vaughan. 1999. "Evaluation Designs for Adequacy, Plausibility and Probability of Public Health Programme Performance and Impact." *International Journal of Epidemiology* 28: 10–18.

Haddad, L. 2013. "From Nutrition Plus to Nutrition Driven: How to Realize the Elusive Potential of Agriculture for Nutrition?" *Food and Nutrition Bulletin* 34 (1): 39–44.

HarvestPlus. 2010. *Reaching and Engaging End Users (REU) with Orange Fleshed Sweet Potato (OFSP) in East and Southern Africa.* Final Report submitted to the Bill and Melinda Gates Foundation. Washington, DC.

———. 2014. *Biofortification Progress Briefs.* Washington, DC.

———. 2016. *Biofortification: The Evidence*. Washington, DC.

Hawkes, C., R. Turner, and J. Waage. 2012. *Current and Planned Research on Agriculture for improved Nutrition: A Mapping and a Gap Analysis.* London: Department for International Development.

Henson, S., J. Humphrey, and B. McClafferty. 2013. *Linking Nutritious Agriculture by Design: A Tool for Program Planning.* GAIN-IDS Discussion Paper. Abuja, Nigeria: Global Alliance for Improved Nutrition.

Herforth, A., and T. J. Ballard. 2016. “Nutrition Indicators in Agriculture Projects: Current Measurement, Priorities, and Gaps.” *Global Food Security* 10: 1–10.

Hotz, C., C. Loechl, A. de Brauw, P. Eozenou, D. Gilligan, M. Moursi, M. Munhaua, P. van Jaarsveld, A. Carriquiry, and J. V. Meenakshi. 2012. “A Large-Scale Intervention to Introduce Orange Sweet Potato in Rural Mozambique Increases Vitamin A Intakes Among Children and Women.” *British Journal of Nutrition* 108 (1): 163–176.

Hotz, C., C. Loechl, A. Lubowa, J. K. Tumwine, G. Ndeezi, A. Nandutu Masawi, R. Baingana, A. Carriquiry, A. de Brauw, J. V. Meenakshi, and D. O. Gilligan. 2012. “Introduction of β-Carotene-Rich Orange Sweet Potato in Rural Uganda Resulted in Increased Vitamin A Intakes Among Children and Women and Improved Vitamin A Status Among Children.” *Journal of Nutrition* 142 (10): 1871–1880.

Iannotti, L., K. Cunningham, and M. Ruel. 2009. *Improving Diet Quality and Micronutrient Nutrition: Homestead Food Production in Bangladesh.* IFPRI Discussion Paper 00928. Washington, DC: International Food Policy Research Institute.

IEG (Independent Evaluation Group). 2011. *Impact Evaluations in Agriculture: An Assessment of the Evidence.* Washington, DC: World Bank.

Knutsson, P. 2006. “The Sustainable Livelihoods Approach: A Framework for Knowledge Integration Assessment.” *Human Ecology Review* 13 (1): 90–99.

Low, J. W., M. Arimond, N. Osman, B. Cunguara, F. Zano, and D. Tschirley. 2007a. “Ensuring the Supply of and Creating Demands for a Biofortified Crop with a Visible Trait: Lessons Learned from the Introduction of Orange-Fleshed Sweet Potato in Drought-Prone Areas of Mozambique.” *Food and Nutrition Bulletin* 28 (2): S258–S270.

———. 2007b. “A Food-Based Approach Introducing Orange-Fleshed Sweet Potatoes Increased Vitamin A Intake and Serum Retinol Concentrations in Young Children in Rural Mozambique.” *Journal of Nutrition* 137 (5): 1320–1327.

Mareida, M. K. 2009. *Improving the Proof: Evolution of and Emerging Trends in Impact Assessment Methods and Approaches in Agricultural Development.* IFPRI Discussion Paper 929. Washington, DC: International Food Policy Research Institute.

Marshall, M. N. 1996. “Sampling for Qualitative Research.” *Family Practice* 13: 522–525.

Mason, M. 2010. “Sample Size and Saturation in PhD Studies Using Qualitative Interviews.” *Forum Qualitative Sozialforschung* 11 (3): art. 8. http://nbn-resolving.de/urn:nbn:de:0114-fqs100387.

Masset, E., L. Haddad, A. Cornelius, and J. Isaza-Castro. 2011. *A Systematic Review of Agricultural Interventions That Aim to Improve Nutritional Status of Children.* London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

———. 2012. “Effectiveness of Agricultural Interventions That Aim to Improve Nutritional Status of Children: Systematic Review.” *The BMJ* 344 (1): d8222.

Meeker, J., and L. Haddad. 2013. *A State of the Art Review of Agriculture-Nutrition Linkages.* AgriDiet Briefing, Brighton, UK: Institute of Development Studies.

Noonan, E., and E. Kristjansson. 2009. “The Difficult Path Toward Better Evidence and Better Decisions.” In *Better Evidence for a Better World,* edited by E. Noonan and M.W. Lindsey, 3–6. 3ie Working Paper No. 2. New Delhi: International Initiative for Impact Evaluation.

Olney, D. K., A. Pedehombga, M. T. Ruel, and A. Dillon. 2015. “A 2-Year Integrated Agriculture and Nutrition and Health Behavior Change Communication Program Targeted to Women in Burkina Faso Reduces Anemia, Wasting, and Diarrhea in Children 3–12.9 Months of Age at Baseline: A Cluster/Randomized Controlled Trial.” *Journal of Nutrition* 145: 1317–1324.

Olney, D. K., A. Talukder, L. L. Iannotti, M. T. Ruel, and V. Quinn. 2009. “Assessing Impact and Impact Pathways of a Homestead Food Production Program on Household and Child Nutrition in Cambodia.” *Food and Nutrition Bulletin* 30 (4): 355–369.

Palmer, A. C., K. Healy, M. A. Barffour, W. Siamusantu, J. Chileshe, K. J. Schulze, K. P. West Jr., and A. B. Labrique. 2016. “Provitamin A Carotenoid/Biofortified Maize Consumption Increases Pupillary Responsiveness Among Zambian Children in a Randomized Controlled Trial.” *Journal of Nutrition* 146: 2551–2558.

PATH. 2014. *The Evidence for Maternal and Child Nutrition Interventions: Gap Analysis*. Washington, DC: PATH.

Pinstrup-Andersen, P. 2013. “Nutrition-Sensitive Food Systems: From Rhetoric to Action.” *The Lancet* 382 (9890): 9–10.

Roetman, E. 2011. *A Can of Worms? Implications of Rigorous Impact Evaluations for Development Agencies.* 3ie Working Paper No. 11. New Delhi: International Initiative for Impact Evaluation.

Ruel, M. T. 2001. *Can Food-Based Strategies Help Reduce Vitamin A and Iron Deficiencies? A Review of Recent Evidence.* Food Policy Review No. 5. Washington, DC: International Food Policy Research Institute.

Ruel, M. T., and H. Alderman. 2013. “Nutrition-Sensitive Interventions and Programmes: How Can They Help to Accelerate Progress in Improving Maternal and Child Nutrition?” *The Lancet* 382 (9891): 66–81.

Savin-Baden, M., and C. Howell Major. 2012. *Practical Introduction to Qualitative Research: Methodology Matters.* London: Routledge Chapman & Hall.

Scriven, M. 2008. “A Summative Evaluation of RCT Methodology and an Alternative Approach to Causal Research.” *Journal of MultiDisciplinary Evaluation* 5 (9): 11–24.

———. 2010. “Rethinking Evaluation Methodology.” *Journal of MultiDisciplinary Evaluation* 6 (13): i–ii.

Webb, P. 2013. *Impact Pathways from Agricultural Research to Improved Nutrition and Health: Literature Analysis and Research Priorities.* Rome: Food and Agriculture Organization of the United Nations; Geneva: World Health Organization.

Webb, P., and E. Kennedy. 2012. *Impacts of Agriculture on Nutrition: Nature of the Evidence and Research Gaps.* Nutrition CRSP Research Briefing Paper 4. Boston: Global Nutrition CRSP and Tufts University.

White, H. 2009a. *Theory-Based Impact Evaluation: Principles and Practice.* New Delhi: International Initiative for Impact Evaluation.

———. 2009b. “We All Agree We Need Better Evidence. But What Is It and Will It Be Used?” In *Better Evidence for a Better World,* edited by M. Lipsey and E. Noonan, 15–18. 3ie Working Paper No. 2. New Delhi: International Initiative for Impact Evaluation.

———. 2013. “An Introduction to the Use of Randomised Control Trials to Evaluate Development Interventions.” *Journal of Development Effectiveness* 5 (1): 30–49.

Wilczek, F. 2013. “Simplicity.” In *This Explains Everything: Deep, Beautiful, and Elegant Theories of How the World Works,* edited by J. Brockman, 35–37. New York: HarperCollins.

Winters, P., A. Maffioli, and L. Salazar. 2011. “Introduction to the Special Feature: Evaluating the Impact of Agricultural Projects in Developing Countries.” *Journal of Agricultural Economics* 62 (2): 393–402.

Winters, P., L. Salazar, and A. Maffioli. 2010. *Designing Impact Evaluations for Agricultural Projects.* Impact-Evaluation Guidelines Technical Notes No. IDB-TN-198 Washington, DC: Inter-American Development Bank.

World Bank. 2014. *Learning from World Bank History: Agriculture and Food-Based Approaches for Addressing Malnutrition.* Agriculture and Environmental Services Discussion Paper, Washington, DC.

ACKNOWLEDGMENTS

The author would like to thank all those who took part in the survey for taking time out of their busy schedules to share their opinions. And to univ. Prof. Mag. Dr. Klaus Salhofer and dr. Ekin Birol for their help and guidance throughout this process.