American Society for Nutrition
Excellence in Nutrition Research and Practice

JN THE JOURNAL OF NUTRITION

journal homepage: https://jn.nutrition.org/

Perspectives

# Food Systems Interventions for Nutrition: Lessons from 6 Program Evaluations in Africa and South Asia☆

Lynnette M Neufeld [1,*], Stella Nordhagen [2], Jef L Leroy [3], Noora-Lisa Aberman [2], Inka Barnett [4], Eric Djimeu Wouabe [5], Amy Webb Girard [6], Wendy Gonzalez [2], Carol E Levin [7], Mduduzi NN Mbuya [2], Eduardo Nakasone [8], Christina Nyhus Dhillon [2], Dave Prescott [9], Matt Smith [10], David Tschirley [8]

[1] *Food and Nutrition Division, Food and Agriculture Organization of the United Nations (FAO), Rome, Italy;* [2] *Knowledge Leadership Team, Global Alliance for Improved Nutrition (GAIN);* [3] *Nutrition, Diets and Health Unit, International Food Policy Research Institute, Washington, DC, United States;* [4] *Institute of Development Studies (IDS), University of Sussex, Brighton, United Kingdom;* [5] *Results for Development (R4D), Washington, DC, United States;* [6] *Nutrition and Health Sciences Program, Hubert Department of Global Health, Rollins School of Public Health, Emory University, Atlanta, GA, United States;* [7] *Department of Global Health, University of Washington, Seattle, WA, United States;* [8] *Department of Agricultural, Food and Resource Economics, Michigan State University, East Lansing, MI, United States;* [9] *The Partnering Initiative, Oxford, United Kingdom;* [10] *Mokoro Ltd, Oxford, United Kingdom*

A B S T R A C T

Although there is growing global momentum behind food systems strategies to improve planetary and human health—including nutrition—there is limited evidence of what types of food systems interventions work. Evaluating these types of interventions is challenging due to their complex and dynamic nature and lack of fit with standard evaluation methods. In this article, we draw on a portfolio of 6 evaluations of food systems interventions in Africa and South Asia that were intended to improve nutrition. We identify key methodological challenges and formulate recommendations to improve the quality of such studies. We highlight 5 challenges: a lack of evidence base to justify the intervention, the dynamic and multifaceted nature of the interventions, addressing attribution, collecting or accessing accurate and timely data, and defining and measuring appropriate outcomes. In addition to more specific guidance, we identify 6 cross-cutting recommendations, including a need to use multiple and diverse methods and flexible designs. We also note that these evaluation challenges present opportunities to develop new methods and highlight several specific needs in this space.

*Keywords*: methods, project impact, theory of change, complex interventions, systems interventions

## Introduction

The past few years have seen a flurry of publications, presentations, and calls to action related to food system transformation for human, animal, and planetary health [1–5]. These messages coalesced around the 2021 United Nations Food Systems Summit, resulting in consolidation of evidence on a variety of food systems-related topics [6], new coalitions, and new commitments and investments [7]. This included the recognition of the importance of a healthy diet to prevent all forms of malnutrition [8]. To support access to healthy diets, development donors have funded, and implementing partners have designed and implemented, a range of different food systems for nutrition interventions in recent years.

There is limited evidence, however, on how to make healthy diets accessible to all: the evidence base on effective actions to "transform food systems for nutrition" is tenuous and in urgent need of expansion. As a result, although strong evidence is

*Abbreviations:* DALY, disability-adjusted life year; FE, food environment; GAIN, Global Alliance for Improved Nutrition; LMIC, low- and middle-income country; MSME, micro, small- and medium-sized enterprise; RCT, randomized controlled trial; SUN, Scaling Up Nutrition.

☆ The views expressed in this publication are those of the author(s) and do not necessarily reflect the views of FAO.

* Corresponding author. *E-mail address:* lynnette.neufeld@fao.org (L.M. Neufeld).

https://doi.org/10.1016/j.tjnut.2024.04.005
Received 11 January 2024; Received in revised form 27 March 2024; Accepted 2 April 2024; Available online xxxx

building in some areas (e.g., taxation on sugar-sweetened beverages [9,10]), overall guidance to inform effective actions is lacking.

Much of food system transformation happens through large, long-term, and external processes such as urbanization, globalization, and income growth. However, many governments and organizations are aiming to foster faster progress toward desired areas of transformation through intentional, specific policies and interventions. Evaluating these types of interventions aiming to transform food systems is difficult. First, because of the systems nature of these interventions, there is a need to be comprehensive, considering the inherent complexity and unpredictability of the system itself. There is a risk that evaluations simplify the systems complexity to instead focus narrowly on individual intervention components to align with conventional evaluation methods. In the process, they may inadvertently disregard the interconnectedness between system actors and components—which may be affected by the intervention and can affect other parts of the system. A more holistic evaluation may be required to fully understand what changes might have been triggered (or not) and why. Second, and related, the standard nutrition evaluation toolbox does not easily apply to these types of interventions. For example, randomized controlled trials (RCTs) are often infeasible, making it difficult to construct a valid counterfactual.

There are thus critical gaps in our understanding of what constitutes potentially impactful food systems actions for nutrition and how to assess this potential. In this article, we contribute to filling the latter gap, with a focus on the contexts of low- and middle-income countries (LMICs) in Asia and Africa. We do so by drawing on a portfolio of evaluations of food systems interventions to identify key methodological challenges and formulate recommendations to improve the quality of such studies. We take advantage of 6 evaluations examining diverse interventions using varied methods to learn and draw recommendations that can be applied more broadly.

The evaluations were commissioned by the Global Alliance for Improved Nutrition (GAIN) from 2017 to 2021 and focused on food systems interventions designed and implemented by GAIN. In late 2021, the evaluation teams and invited experts convened to review and share experiences. Through this process, 5 main challenges were identified: lack of an evidence base to justify the intervention; the dynamic and multifaceted nature of the interventions; addressing attribution; obtaining accurate, timely data; and defining and measuring appropriate outcomes. Here we summarize these challenges, illustrate examples of how they were addressed in these cases, and highlight where challenges remain. For each challenge, we include recommendations for how to address it. We also reflect on what lessons and tools nutrition can borrow from other fields in which complex evaluations are common.

## The Evaluation Portfolio

The 6 programs evaluated were selected for evaluation by GAIN based on having a clear theory of change with potentially measurable exposures and outcomes as well as sufficient intensity to potentially lead to measurable changes across the theory of change (see Box 1 for key terms used in this article) [11–16]. They focused on generating demand for nutritious foods, strengthening the supply of nutritious foods, supporting subcomponents of the supply chain (such as traditional markets), and combinations of these. Programs were implemented in LMICs in Asia and Africa; some evaluations covered only one country or subnational area, whereas others covered multiple settings. One of the programs had a specific focus on children and one on adolescents; 4 worked through businesses (either as producers of food or enablers of employee nutrition). Supplemental Table 1 provides a detailed overview of the 6 interventions, and details of the accompanying evaluations are provided in Supplemental Table 2; a brief summary is provided below. Some results of the evaluation have been published [17–20], and others are forthcoming.

The Scaling up Nutrition (SUN) Business Network supports micro, small, and medium-sized food enterprises (MSMEs) in LMICs through networking, training, and advocacy to create an enabling environment with the eventual goal of increasing accessibility of nutritious food for consumers. The evaluation aimed to assess the relevance, internal consistency, and feasibility of the program's theory of change as well as progress in achieving change. A theory-based mixed-methods evaluation

**Box 1**
Definitions of key evaluation-related terms used in this article.

- **Attribution** – Ascription of a causal link between a given intervention and observed changes [11].
- **Contribution analysis** – A method for determining whether and how an intervention contributed to an observed result, based on analyzing the underlying theory of change (its assumptions and evidence to support changes along it) as well as other influencing factors [11,12].
- **Evaluation** – The systematic assessment of an ongoing or completed project or program, including its design, implementation, and/or results. The assessment can be both qualitative and quantitative. The aim is to determine one or all of: relevance and fulfillment of objectives, efficiency, effectiveness, impact, and sustainability [11].
- **Food system** – "all the elements (environment, people, inputs, processes, infrastructures, institutions, etc.) and activities that relate to the production, processing, distribution, preparation and consumption of food, and the outputs of these activities, including socioeconomic and environmental outcomes" [13].
- **Food systems for nutrition intervention** – A set of actions deliberately seeking to improve nutrition outcomes through changes to the food system, particularly when intervening across multiple areas of the food system or relying on interactions within the food system.
- **Most significant change** – A qualitative monitoring and evaluation methodology to assess the most significant change in the lives of stakeholders in the project as opposed to seeking impact solely on predetermined outcomes [14].
- **Quality improvement** – a systematic approach to more efficiently attain intended outcomes by tracking outcome measures, implementing change ideas one at a time, and scaling up when ideas improve outcomes [15].
- **Theory of change** – The mechanism through which an intervention is expected to achieve change for individuals and groups (compared to what the situation would otherwise have been), including assumptions about causal links between activities or inputs, intermediate outputs or outcomes, and eventual impact [11,16].

was conducted, which included an extensive review of relevant documentation and in-depth interviews with >100 project collaborators and others working on topics of business support and nutrition. Desk reviews were conducted of networks in Sri Lanka, Bangladesh, Kenya, Malawi, Tanzania, Nigeria, and Pakistan, and 3 in-country case studies were conducted of advanced networks in Zambia, Indonesia, and Mozambique.

Supporting Frontline Food System Workers was implemented during the COVID-19 pandemic. It delivered emergency grants to employers to enable them to provide nutritious foods to vulnerable food system front-line workers in Bangladesh, India, Pakistan, and Kenya. The summative evaluation sought to document and analyze the program's relevance and effectiveness, identify implementation challenges, and formulate lessons learned and best practices for future implementation of workforce nutrition and emergency nutrition programs. The analysis also considered how gender-related issues were addressed in the process of awarding grants and how that translated into impacts for women and other workers in situations of vulnerability.

The Bhalo Khabo Bhalo Thakbo ("Eat Well, Live Well") Campaign in Bangladesh aimed to trigger adolescent-led social actions to promote individual behavior change (healthier snack food choices among adolescents) and broader societal change (improved availability of healthier snacks). The campaign's evaluation used a theory-based mixed-methods approach with contribution analysis as the overarching methodological framework. The objectives were to elucidate credible causal claims about the campaign's contribution to change, verify the core assumptions that underpinned the campaign logic, and identify implementation bottlenecks and opportunities.

The Marketplace for Nutritious Foods program in Kenya aimed to increase the supply of nutritious foods in affordable, desirable, and convenient forms by providing MSMEs with grant funding and technical assistance and bringing them together for networking and group training via a community of practice. The evaluation broke the complex program into smaller components and focused evaluation activities on shorter segments of the theory of change. This strategy allowed the evaluators to determine the impact of individual program components (e.g., technical assistance, informal networking) on different segments of food supply chains, from producers to consumers. This was done through 7 separate studies, including a field experiment, choice and trust experiments, participant surveys, and qualitative interviews.

The Nigeria Egg Demand Creation Campaign sought to create demand for eggs for feeding children aged 6 to 59 mo through diverse activities such as television and radio spots; point-of-purchase advertisements; and neighborhood, compound, and market shows. The activities were designed to emotionally engage caregivers, increase their confidence in choosing eggs, improve their knowledge about the benefits of eggs, and subsequently prompt them to choose eggs at the point of purchase. It was evaluated using a quasi-experimental design, with intervention and comparison arms.

The Building Resilience of Nutritious Food MSMEs in the Food System program sought to ensure business continuity of food system MSMEs during and after the COVID-19 pandemic. This was done by providing emergency grants and training to cope with immediate financial and management challenges as well as technical assistance to support post-COVID-19 trade and build resilience to future uncertainties. It was evaluated through case studies applying a theory-based design and contribution analysis.

## Overview of Evaluation Challenges and Opportunities

The evaluations encountered many challenges, related to the nature of the interventions and, in some instances, the implementation context. We focus on challenges that are more unique to food systems-focused actions and those that most affected our ability to effectively evaluate the programs in question.

### Challenge 1: lack of an evidence base to justify the intervention approach

The focus on "food systems for nutrition" interventions is relatively new. Many such interventions are designed based on theoretical or intuitive understandings of how food systems plausibly work but without a strong evidence base to justify the approach. This contrasts with more traditional nutrition interventions, such as promotion of exclusive breastfeeding through social behavior change communication or vitamin A supplementation campaigns, which build on strong evidence bases regarding the underlying biological pathways, the particular behavior that needs to be changed, and the actions used for doing so. The weak evidence base for food system interventions presents 2 challenges for evaluation. First, it creates the need for a "leap of faith" when formulating the theory of change (Box 1) [11–16] needed to design the evaluation. Due to evidence gaps, a theory of change for such interventions will also likely contain more and stronger assumptions than one for a more tested type of intervention. Second, prior research cannot be relied on to interpret unexpected or contradictory results.

#### *Recommendations for good evaluation practice*

- Evaluators should undertake a high-quality "design evaluation" prior to program implementation. In a design evaluation, external actors not involved with developing the program undertake a detailed review of its approach before it is implemented. Design evaluations include a deep examination of the theory of change to identify and, as feasible, test critical assumptions within it. If it is unlikely that the program will be effective as designed, it is modified. Where assumptions cannot be verified a priori, they become focuses of the evaluation itself. Design evaluations have been used in some social policy evaluations, such as previously in Mexico [21].
- When participating in a design evaluation or responding to a request for proposals, evaluators should point out when a program has a poorly developed theory of change, shows logic gaps, contradicts existing evidence, or features unreasonable assumptions. Ideally, a design evaluation provides the opportunity to address such issues before implementation. If they are not addressed, then potential evaluators of the implemented program should be cautious about embarking on evaluations, as doing so may be an inappropriate use of resources and a violation of ethical principles for human subjects research.

- Evaluators should include the intervention-implementing team in the evaluation design from the outset, with a focus on strengthening the theory of change [22]. The evaluators should aim to understand how the implementing team thinks the intervention will have an effect and what the critical assumptions behind that logic are; this information can be used to understand the steps in the theory of change, which outcomes to assess, and which assumptions to verify. Although this early and deep collaboration may present additional challenges in terms of timing and mutual understanding, ultimately it will enhance evaluators' understanding of the thinking behind the intervention, implementers' understanding of existing evidence on the topic, and the potential for uptake of evaluation results and recommendations. Developing close collaboration with implementers can also facilitate timely access for evaluators to the program's monitoring data, which can be used to confirm or disprove certain program assumptions, and ensure evaluators are quickly informed about changes to program implementation and why they were made, enabling them to promptly adjust methods and tools, as required.

## Challenge 2: the dynamic and multifaceted nature of interventions

Many food systems interventions necessarily require adjustments over the course of implementation. This is first due to an incomplete understanding of the underlying mechanisms, which necessitates regular course-correction in response to initial implementation experiences and emerging information; these course corrections are intended to increase the likelihood of impact. Second, many food systems interventions involve private-sector food companies. In their pursuit of profit maximization, these companies may make rapid changes—in response either to the intervention itself or to other market considerations. For example, the Marketplace for Nutritious Foods intervention had to contend with companies unexpectedly shifting their focus areas and working with multiple agents in their supply chains, which made their products difficult to trace for evaluation purposes (see Challenge 4, below).

Fluidity is particularly the case for interventions that are participatory, highly innovative, iterative, or locally led or codesigned—as was the case for several of the interventions included in this portfolio. For example, although the Bhalo Khabo Bhalo Thakbo campaign in Bangladesh had well-defined entry activities for the campaign, follow-on activities were later developed together with the adolescents using a participatory and iterative method involving various food system actors (e.g., corner shop owners, street food vendors, and school canteens). This meant that the evaluation had to adapt to these emerging and changing activities while ensuring transparency in any methodological changes made and adhering to a rigorous approach that would deliver high-quality evidence.

External factors may also necessitate adaptation of both the intervention and the evaluation. As several of the interventions (and all of the evaluations) in this portfolio were implemented during the COVID-19 pandemic, numerous related adaptations were required. In the Bhalo Khabo Bhalo Thakbo program, all school-based face-to-face campaign activities had to be redesigned and moved to virtual channels due to prolonged school closures in Bangladesh. This required a significant change in the evaluation design and data collection methods. Although adaptability is necessary for any type of real-world evaluation, the need is particularly acute for food systems interventions due to the interrelated nature of the food system. As a consequence, changes in one part of the system may have implications for another part.

Finally, food systems interventions often include multiple actors from different backgrounds and with different interests and incentives. This may present challenges at the design, implementation, and evaluation stages. In particular, it can be difficult to coordinate interrelated actions from different actors—such as businesses and governments—in line with evaluation timelines. For example, in cases where a program aims to shift business action through policy advocacy but policy is lagging behind, action can stall. This was true for SUN Business Network's advocacy work with food manufacturers to eliminate trans-fat in major food products in Pakistan and Nigeria, where businesses had no incentive to transition production until government could properly enforce policies.

### *Recommendations for good evaluation practice*

- To account for this complexity and dynamism, evaluators can use a *theory of transformation,* instead of a traditional theory of change, to guide evaluation design. This concept emerged in relation to systems interventions (e.g., [23,24]) and recognizes that transformation is more multidimensional, cross-sectoral, and longer-term than a traditional project, and thus, it requires not one but multiple theories of change operating at different levels that, when combined, explain systems-level transformation.
- Evaluators can apply complexity-aware or "realist" evaluation methods from other fields, such as systems mapping and visualization (e.g., social network analysis, community mapping), system dynamics modeling, and narrative methods (e.g., Most Significant Change Technique) [24–28]. These may need to be refined to be specifically applicable to food systems interventions.
- Evaluators should expect adaptations—and, indeed, embrace and normalize them as critical to the success of interventions in the dynamic food system context. If an intervention changes, the evaluation design will likely need to follow. However, not all types of evaluation adaptations are acceptable: evaluation changes should not be guided by the desire to increase the likelihood of finding impact, such as choosing easier-to-accomplish outcomes or those that show early signs of change. In the interest of rigor, interpretability, and replicability, change must be carefully and transparently documented, including why, how, when, and by whom adaptations are made. This documentation should include updating the theory of change to reflect these adaptations, as well as capturing any changes to underlying assumptions that either contributed to the change or occurred as a result of it. Where feasible, changes should be reflected in updates to publicly registered protocols.
- Adaptation can be facilitated by applying systems evaluation principles, such as a design that evolves with implementation strategy, real-time feedback on immediate outcomes, and embracing "failures" as inevitable opportunities for adaptation [12,29,30].

- Evaluators should retain as much flexibility as possible in methods and data collection timing—and in the budgeting and staff allocation needed to support this; evaluation funders should revise their expectations to align to this.
- In some cases, evaluators can use formal methods to track the process of change itself. For example, in GAIN's workforce nutrition interventions, quality improvement methods were used to track how workers perceived small adaptations to meals over time [31]. Other methods include rapid-cycle improvement [32,33], community quality improvement [34], and complexity-aware monitoring [35,36].
- Evaluators should consider and explain whether adaptability has implications for scaling the intervention. Subsequent interventions will rarely follow the same path of adaptations, but in some cases, it may be that impact was partially dependent on following that path. When scaling, it is thus important to reflect on which elements of the intervention and its adaptative process were critical to success and which were not.

## Challenge 3: addressing attribution

Well-designed evaluations allow attribution of changes in outcomes to the program or intervention, i.e., they rule out the possibility that the observed differences are due to non–program-related factors. The fundamental challenge to impact evaluation is thus to identify what would have happened in the absence of the program: the counterfactual. The counterfactual is typically constructed by finding a comparison group that is similar to the group receiving the program in all relevant characteristics, except for receiving the program [37–39].

When interventions are reasonably self-contained, individuals or groups of individuals can be assigned to either receiving (the "treatment") or not receiving the intervention (the "control"). If the assignment is done randomly, one can assume that differences in outcomes between the treatment and control group must be due to the program. The control group thus provides a valid counterfactual for the treatment group exposed to the program [40].

Many food systems interventions are intrinsically challenging to manipulate experimentally. Evaluating these programs thus may require different evaluation methods. Quasi-experimental and complexity-aware methods can be used to evaluate demand-side interventions, such as social marketing to promote the consumption of healthy diets or the distribution of vouchers or cash to purchase nutritious foods. These types of interventions may be amenable to randomization at the recipient level, or, if that is not possible, a phased roll-out can be used to compare those who already receive the program benefits with those who are yet to receive them. If randomization is not possible but program benefits are limited to those falling below a certain threshold (e.g., a wealth level), a regression-discontinuity design can be used in which households just below and just above the cutoff are compared to estimate the program's impact [38].

Attributing the impact of supply-side interventions is more challenging. They are often implemented across an entire region or country (e.g., taxes and front-of-pack labeling), and consumers might only be reached indirectly (e.g., as in the Marketplace for Nutritious Foods intervention). Supply-side interventions are similar to natural experiments, which are broadly defined as experiments in which the exposure is not being manipulated by the researcher. Confounding in the evaluation of natural experiments can be minimized when there is abrupt change in exposure across the whole population and when the expected impacts are large [41]. Unfortunately, for many supply-side interventions, these conditions are unlikely to hold.

Other strategies to increase the internal validity are to use multiple pre- and postmeasures to control for secular changes, accurately measure multiple potential confounders, and apply combinations of methods to address different sources of bias [41]. Although not part of this portfolio, the evaluation of the sugar-sweetened beverage tax in Mexico used the change in trends over time to estimate impact [42], and a similar strategy was used in Chile to assess the impact of front-of-pack labeling [29,43].

In some cases, it may be possible to embed RCTs within other types of complex evaluation. For example, the Marketplace for Nutritious Foods evaluation could not address the overall research questions (i.e., the impact of technical and financial assistance on MSMEs' ability to produce nutritious products) through experimental methods. They instead used quasi-experimental and qualitative methods for this purpose, applied to participating/nonparticipating firms. A cluster RCT was used to examine one specific assumption in the overall theory of change—whether the availability of nutritious foods improved the quality of diets—by saturating randomly selected market segments with a program-supported nutritious food product.

### *Recommendations for good evaluation practice*

- Evaluators can use a combination of different evaluation designs, measuring key outcomes along the theory of change, with mixed quantitative and qualitative methods. This can help to minimize confounding and bias and strengthen the case for attribution.
- Evaluators should adhere to key principles such as using valid instruments for measurement, controlling for known confounders, ensuring statistical power, and maintaining transparency by registering the research protocol.
- Because causal attribution is often problematic in these interventions, evaluators should take particular care to formulate clear and concise research questions that are directly linked to the intervention theory of change.
- Evaluators should encourage discussion among those involved in program implementation and uptake of evaluation results, including donors, to challenge and redefine our collective mental model of *attribution* and to include innovative, rigorous methods [30]. These can include:
  - o Applying contribution analysis to assess the credible *contribution* of the food system intervention to the desired outcomes/impacts rather than seeking statistical *attribution*, which may be impossible to determine [12,44].
  - o Breaking a complex theory of change into components and evaluating those via nested evaluations or focusing only on theory of change segments known to contribute to or constrain achievement of the target outcomes.
  - o Using complexity-aware methods, as mentioned in the prior section.

## Challenge 4: collecting or obtaining access to accurate, timely data

Good evaluations rely on accurate and timely data, and collecting or otherwise accessing data is a key challenge in many evaluations. In evaluations of complex food systems with

multiple actors across various sectors, access to relevant high-quality data can be particularly difficult for several reasons.

First, evaluating food systems interventions often requires obtaining data from private companies. Assessing profitability is central to determining whether market-based interventions work from both business and nutrition perspectives. Sometimes standard surveys can be used to collect such data, but some companies may be reticent to share such data—particularly in an evaluative context in which "negative" results may have financial implications. Although this challenge may be greater with large firms, MSMEs—which represent the bulk of food systems actors in many of the contexts in which these evaluations are conducted [45]—may not even track such data or may do so with insufficient quality needed for evaluation. Private-sector actors may also have limited time or willingness to participate in research activities. To address this last challenge, the Marketplace for Nutritious Foods evaluation used a streamlined questionnaire for MSME managers, collected via email with enumerators' support over the phone where needed; this meant that managers could complete the questionnaire when convenient and could access relevant firm financial documents when doing so.

Second, evaluations of food systems interventions are often trying to understand how food travels through that system—for example, to determine who ultimately consumes the products produced or sold by a company and how it reaches them. This requires "tracking" a food from producers and distributors to markets and ultimately to households, and even members within households. The exact path foods follow may not be predictable in advance—a challenge for data collection. For interventions that work with unbranded foods (e.g., eggs or vegetables), it is even more difficult to collect these data, as products are typically untraceable along the supply chain. In addition, foods of interest may not be visually identifiable—as is the case for many biofortified crops or unpackaged and repackaged fortified products. For commodities that cannot be easily traced or identified, difficulties are not limited to quantifying supply volumes. Assessing changes in consumer perceptions and demand for those foods and how they contribute to diets is also challenging.

Third, the complex nature of food systems interventions means that theories of change and routes from activities to impact are often long, complex, and not always predictable—e.g., when will training and technical assistance lead to a firm installing new equipment? When will that lead to changes in production? When will those changes percolate down to altered consumer behavior? This makes it difficult to plan when to collect data. It also creates challenges with obtaining evaluation funding, as funding cycles are often shorter than the many years needed to observe systems-level changes.

Finally, in addition to impact-related data, it is highly valuable to collect data related to intervention costs and to compare these to the benefits. However, collecting data on intervention costs for food systems interventions entails an additional set of challenges (Box 2) [46–51].

### *Recommendations for good evaluation practice*

- To help incentivize firms' sharing of data, evaluators can:
  - o Carefully explain how data will be protected to ensure that no one beyond the core research team has access to non-anonymized data on individual firms.
  - o Use transparency agreements to facilitate sharing among multiple firms that would benefit from access to one another's (aggregate) data.
  - o Make data collection as convenient and rapid as possible for participants, thus increasing their willingness to participate. This could include the use of online forms, phone interviews, and scheduling interviews outside of business hours.
  - o Focus only on those indicators that are most crucial for the central research questions—which is also in line with best practices for research ethics.
  - o Enable participating companies to benefit from the evaluation data by providing them with customized analysis focused on their direct business needs, such as estimates of their marketing reach by channel. This could increase their incentive to provide high-quality data, though it may require redesigning research questions or indicators to ensure they are of business relevance.
- For study participants who may not normally keep comprehensive records (e.g., MSME financial records), evaluators and implementers could train them to do this. However, these data will need to be interpreted with caution and be considered part of the provided intervention.
- To properly time data collection, evaluators must collaborate closely with implementers; although such collaboration is a hallmark of most evaluations, it may need to be more regular for a dynamic food systems intervention than in a traditional evaluation in which all data collection and implementation schedules can be planned in advance.
- For (bio)fortified foods in which the (bio)fortification is not associated with a visible trait, evaluators can consider testing to verify whether they are indeed (bio)fortified. For most unbranded foods, however, practical methods to track them throughout the supply chain are still lacking.

## Challenge 5: defining and measuring appropriate outcomes

Food systems interventions for nutrition also pose evaluation challenges when it comes to defining and measuring appropriate outcomes. These interventions may purposefully have multiple key endpoints. For example, the SUN Business Network aims to promote business viability and to increase the availability, accessibility, and affordability of nutritious foods in the market. It thus recognizes that both are critical for businesses to contribute to improved nutrition. The complexity of food systems interventions and the interlinked nature of food systems' feedback loops also raise the likelihood of unintended positive and negative consequences, which should be tracked.

For interventions that are iterative or participatory, perspectives about priority outcomes may vary vastly over time and across different individuals and organizations involved. As such, it can be difficult to reconcile varying perspectives of what defines program success. With regard to dietary outcomes, for example, there are various aspects to what constitutes "nutritious" or "healthy" and a variety of tools and approaches to their measurement [52]. Understanding of these concepts and terminology varies among business owners, policymakers, and consumers. Another outcome area that has increased in prominence in recent years is the food environment, which comes with its own set of challenges from an evaluation perspective (Box 3)

**Box 2**
Challenges and opportunities for capturing costs and benefits of food systems interventions.

Understanding the effectiveness of food systems interventions for nutrition is essential for deciding what to invest in to shape future food system transformation—but it is not sufficient. Ensuring the efficient use of scarce resources requires comparing estimates of effectiveness with information on costs.

Economic evaluations, such as cost-benefit and cost-effectiveness analysis, can be used to inform efficient resource allocation in agriculture and health, respectively. For example, cost-benefit analysis can capture the monetary benefits associated with increased productivity and incomes of an agriculture investment to increase crop yields. On the health side, cost-effectiveness analysis can capture the comparative costs and effects of investments that seek to address a single health or nutrition outcome, such as vitamin A status through supplementation or the duration of acute malnutrition episodes through better treatment. However, when considering food systems interventions, a more holistic view is needed. To properly assess such interventions, economic evaluations need to consider a wider range of benefits, not just a single outcome, as is typical in most cost-effectiveness analyses. Benefits fall into 3 categories: *1*) those that can be easily monetized, like changes in productivity, income, profits, and the value of home consumption; *2*) health and nutrition benefits that can quantified using disability-adjusted life years (DALYs) (e.g., anemia and severe acute malnutrition) or associated monetary values; and *3*) potential benefits that cannot be monetized or quantified using DALYs, such as changes in knowledge and practices, enhanced dietary diversity, improved food security, reduced vulnerability, and greater women's empowerment. The final category is the hardest to measure and incorporate in economic evaluation, but accounting for such benefits can provide a more complete picture of an intervention's effects and help to appropriately value the holistic approach taken by food systems interventions.

It is challenging to quantify and value the full range of benefits along the supply chain or theory of change. However, omission of these benefits will underestimate the return on investment for food systems interventions. Even where benefits can be measured and valued, current economic evaluations are not consistently including them, including for food system investments. A recent review found that 60% of cost studies of nutrition interventions captured only a single benefit, though many of the interventions could have had impacts on other outcomes, and <10% of these studies included some of the more hard-to-measure/-value benefits, such as dietary diversity [46].

A challenge unique to food system interventions, especially those involving the private sector, is measuring and attributing the intervention's impact on producer or supplier quantities, input costs, market prices, revenues, and net profit and linking these to consumer benefits. Given the heterogeneity in how businesses invest, benefits to both producers and consumers may accrue in different time periods and extend beyond the intervention period. Capturing the effect on market prices of foods and linking these prices to consumer diets is challenging, which further limits capturing the full range of benefits. For example, the Marketplace for Nutritious Foods intervention worked with businesses that varied in terms of revenue, number of employees, products, infrastructure, and net worth. The investments made possible by the intervention may have represented a small fraction of total upfront investments or running costs, and their incremental benefit in terms of increased quantity or quality of product may have been small. Linking the intervention's investments to profits was challenging. Thus, it was difficult to accurately estimate average net costs, and the resulting estimates do not necessarily represent net costs and benefits at scale or over time. Another issue related to food systems investments is how to avoid to double counting costs and benefits that may simultaneously improve both production and consumption at the market and household levels.

To do better, we must develop standard methods to generate evidence on diverse types of costs and benefits across a range of outcomes, both specific to nutrition and not (as done through the work of the Strengthening Economic Evaluation of Multisectoral Strategies for Nutrition project [47–51]).

[13,53–64]. Reaching an agreement on what is being evaluated can thus pose critical-to-surmount challenges.

The need to assess multiple outcomes leads to another challenge. Some in the evaluation community have argued that it is necessary to limit the number of evaluated outcomes to reduce the risk of false positive findings, i.e., to avoid multiple comparison bias. Limiting the number of outcomes or statistically correcting for multiple hypothesis testing is not needed when all outcomes are declared and registered in advance, one measure is used per outcome, and evaluators report on all outcomes irrespective of whether an intervention effect was found [40]. At the same time, the need to collect, analyze, and report on multiple outcomes does add to evaluation cost and complexity—as well as to the time burden placed on survey respondents and others who provide data. How to interpret results when assessing many outcomes that may change in opposite directions also poses interpretation challenges.

As noted above, many food systems interventions have very long, often complex theories of change, often aligning to the general framework given in Figure 1 [65]. The potential effect of factors exogenous to the intervention increases with longer causal paths, resulting in weakening intervention impacts from one logical step in the theory of change to the next. As such, null results may be found on outcomes far down that long causal pathway, though meaningful impacts might exist at a more proximal point. It is thus essential to properly identify outcomes that are responsive to the intervention and to define both meaningful and detectable levels of change. Related to and given the long causal chain, the response might take longer to appear than the timeframe of the evaluation. This is compounded by issues such as loss to follow-up, spillovers, and influence of confounders, which tend to become more severe over time. Finally, determining the adequate effect magnitude to constitute success is also challenging, given the vast scale of the food system in comparison to the likely impact of an intervention—as well as the previously mentioned factors that may reduce detectable effect size.

### *Recommendations for good evaluation practice*

- Beyond focusing on impact on intervention endpoints, evaluators should include multiple intermediate outcomes.
- From the outset, evaluators and implementers should explicitly define what constitutes success and how it can be measured, clearly anchoring the specific intervention to its contribution in the broader food system framework.
- Through a thorough design evaluation, evaluators should try to identify potential unintended consequences a priori and

**Box 3**
The Food Environment.

Many food system interventions target the food environment (FE), defined as "the interface that mediates people's food acquisition and consumption within the wider food system" [53]. Key characteristics of the FE that affect food consumption include food availability, affordability, vendor and product properties, and messaging [54]. Interventions aimed at improving diet quality may thus seek to influence, for example, the availability of a particular nutritious food within a given FE. The FE is conceptually attractive, as it offers a space for impact that is more tractable and specific than the full food supply chain or food system, but also at a higher level of scale than the individual or household, suggesting greater efficiency and scalability of interventions. The importance of modifying the FE to improve nutrition has been widely recognized in the research literature and by major international bodies [13,55–59].

From an evaluation perspective, however, the FE poses certain challenges. First, capturing the FE of any given individual is complex. A person typically passes through several different geographic spaces in a day or week, including where they live, work, attend school, or socialize; within each space, they may source food from multiple different outlets [60], the characteristics of which may also change throughout the day (e.g., open-air markets during the day and street food sellers at night). People across the globe increasingly use online food delivery platforms [61], which do not have a physical FE presence. Mapping the bounds of this space for any one individual is thus complex. Second, it may not be feasible to aggregate these individual-level FEs into a salient population-level FE for the population of interest. Even among those living within a given household or neighborhood, differences in factors such as job type and school location can result in large differences in the nature of the FE individuals are exposed to. The aggregate FE across all neighborhood residents may thus extend across multiple neighborhoods. Third, the bounds of FEs are far from static: they may change by season (e.g., school year compared with summer) as well as over time due to factors such as changes to transportation corridors, new residential developments, or zoning laws—a challenge exacerbated by the typically long duration of FE interventions. Jointly, these issues make it difficult to pinpoint a real-world area that has salience as an FE, is in line with the lived experience of the target population, *and* is well-defined enough to serve as a sampling frame, especially when reused over time. The challenge is compounded by several other factors. Most evaluation designs would require many FEs (i.e., as units of observation or clusters) to facilitate statistical analysis. FE attributes such as convenience and desirability cannot be easily measured in large-scale surveys. Finally, the different dimensions of the FE cannot be easily summarized in a simple compound metric, which complicates statistical analysis.

Additionally, within the bounds of a FE, its characteristics are constantly in flux, particularly for outlets such as mobile street vendors [62]; food availability and prices change regularly across seasons or even hours. This creates a high level of background "noise" against which to try to detect any intervention effect. At the same time, the potential effect of most FE-focused interventions is fairly small once spread across all target FEs. For example, supporting an MSME through the Marketplace for Nutritious Foods program to supply a nutritious food product may result in a few hundred additional servings of nutritious food produced per week, but that could entail only a dozen additional servings within any given FE, once spread across all the firm's sales areas. Detecting this "signal" of FE change, over the pre-existing "noise," requires either a very large change or a very large sample, which may be infeasible given the constraints on implementation of both interventions and evaluations. Evaluating FE-focused interventions thus comes with considerable practical challenges. Additional challenges relate to the specific metrics to use [63,64].

The FE remains useful as a construct to guide thinking about what drives food choices and which levers are the most potentially impactful. However, it may be an evaluable unit only in certain, specific contexts (e.g., interventions specifically focusing on a workplace or other controlled setting) or with a narrow focus (e.g., restricting the research question to just examine out-of-home purchases made in and around a school). Evaluators should thus be cautious when targeting FE changes as key outcome indicators.



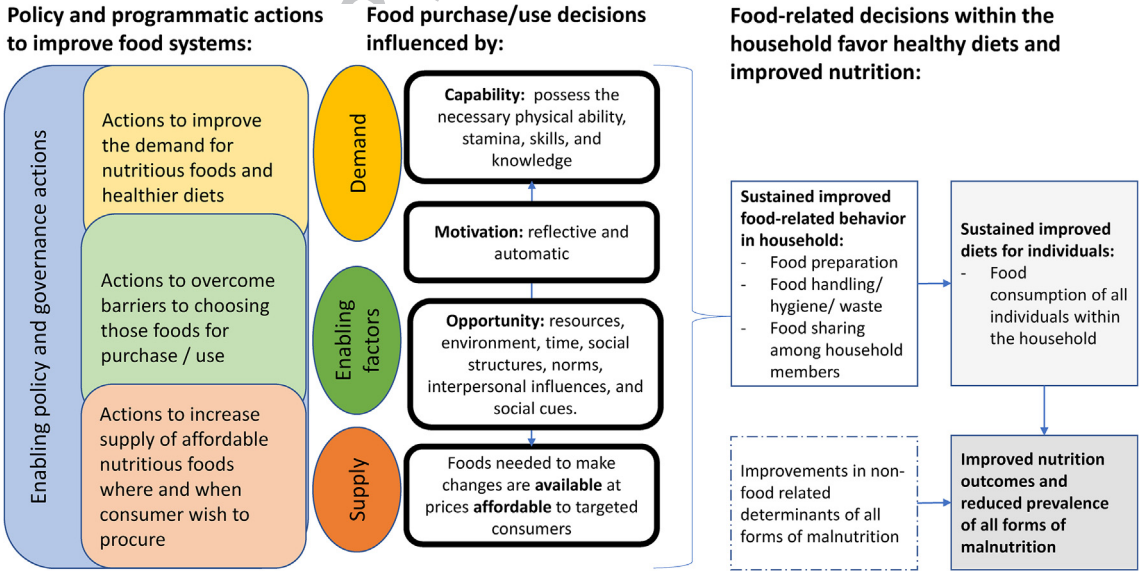


**FIGURE 1.** Food systems actions for dietary improvement and nutrition outcomes (elaboration by the authors, drawing on [65]). Dashed box indicates nutrition outcome-related factors that fall outside of the scope of food systems related policies and programs.

put methods in place to detect those changes. During implementation, they should engage in regular dialogue with implementers to identify unforeseen unintended consequences and be ready to adapt methods to take these into account.

- Evaluators, working with implementers, should consider making explicit the temporal dimension of a theory of change: i.e., mapping out the timeframe in which impacts on each key outcome can be expected to manifest. This will make clearer which effects are beyond the temporal scope of the intervention and thus may be excluded or deprioritized for evaluation. It may also facilitate the planning of longer-term evaluations, including justifying requests for long-term funding.
- Where statistical analysis is used, evaluators must select outcomes for which the study can be adequately powered. This may entail focusing on proxies or determinants rather than the ultimate outcomes themselves [66].
- Evaluators should reset expectations around what meaningful nutrition outcomes are. These should move beyond biological measures of malnutrition and include critical outcomes such as dietary intake. It is important to choose outcomes that align with the theory of change and are amenable to change through the examined food systems interventions—and during the timeframe used for the intervention. For example, food access is known to be meaningful for nutrition and feasible to change within a typical funding cycle/evaluation window, even if increased access may not be directly associated with biological manifestations of malnutrition.
- Although direct measurement may need to focus on what is achievable in the short term, evaluators may be able to model or simulate aspirational long-term potential outcomes to provide indicative values; these can function as "reality checks" by putting feasible bounds on what *might* be achieved in the long run.

## Conclusion

The current recognition of the need to transform food systems for human and planetary health is unprecedented. It creates a major opportunity for improving nutrition in tandem with achieving other societal goals. However, it is not a forgone conclusion that food systems will transform in ways that favor healthy diets and nutrition outcomes. Ensuring that food systems transformation truly leads to sustainable healthy diets for all requires a much stronger evidence base. The nutrition community has for decades benefited from robust evidence on the impact of nutrition-focused interventions [67–69]. We are now moving into an era in which we need to build a similar evidence base that will guide food systems interventions.

Based on the experiences highlighted in this article, we have 6 cross-cutting recommendations for designing better evaluations of these food systems interventions. First, evaluations should always be guided by a strong theory of change, developed in close collaboration with the implementing team and subject to critical reflection from both parties. Second, due to the inherent complexity of food systems interventions, a range of outcomes should be assessed to understand the impact at multiple stages of what is often a long theory of change, gauge trade-offs and synergies, and consider unintended consequences. Third, both to assess this range of outcomes and to increase the plausibility of any attribution claims, a combination of methods should be used, with triangulation across them. Whenever possible, this should include qualitative as well as quantitative methods. Fourth, evaluators should reach beyond the methods typically used in nutrition evaluations and consider those from other fields, including complexity-aware methods such as contribution analysis and system dynamics modeling. When assessing quality of evidence, methods should be evaluated on their own merits, considering fit for purpose and the rigor of application, and not automatically seen as inferior to a statistical "gold standard," which may not be feasible to attain. Fifth, adaptive and flexible evaluation designs are needed to accommodate changes in interventions as they are implemented; the risks this entails, such as more selective reporting, should be recognized and dealt with transparently. Finally, research questions, methods, hypotheses, analysis, and results, and how these may change over time, must be documented transparently. This recommendation underpins several of the above recommendations: transparency is essential when justifying which methods were chosen and how researchers triangulated across their potentially divergent results, as well as why and when changes to evaluation design were made.

We have also revealed some areas where future research is needed. These include development of methods that can better trace food as it moves through the supply chain and metrics focusing on the more understudied aspects of food systems, such as environmental sustainability, resilience of food businesses, and equity in food access. Evaluators should also work to harmonize evaluation frameworks and indicators for diverse food system interventions to make it easier to build a comparable evidence base across evaluations. Such harmonization may require nutrition researchers to grapple with some of the points of contention within the community, such as classification systems for food processing and metrics for assessing the healthfulness of diets.

At the same time, there is considerable scope for the use of diverse methods that have been used to evaluate complex interventions in other fields, such as in education [70]. For example, the Marketplace for Nutritious Foods evaluation discussed here drew heavily on methods used in economics [71], such as laboratory and choice experiments. From this experience, we feel there is potential for wider application of economic methods. Several examples of such evaluations are underway or have been published recently. For example, Cooper et al. [72] are using system dynamics modeling to evaluate a market intervention to improve nutrient-dense food access as part of efforts to address stunting in India, Indonesia, and Senegal. Other novel methods permit the quantification of short-term impacts on agriculture, for example, as a result of COVID-19 [73]. Importantly, many of the system dynamics methods use some type of formalized community or collaborator engagement to build and/or refine the systems models that underpin the evaluation (for example, [74]).

Process-wise, it would be helpful to develop approaches for designing evaluations of complex systems interventions and train researchers on their use. This should include better guidance on how to triangulate across mixed methods within a complex food systems context while lessening risk of bias. It

would also be useful to elaborate best practices for evaluators and implementers to work together throughout a changing program, while maintaining evaluator independence. To create an enabling environment for transparent evaluation of complex food system interventions, there is a need to adjust protocol registries to encourage documenting change over time and to educate evaluation funders on the benefits of making funding more flexible and longer-term. Finally, there is a need for renewed discussion of what constitutes robust scientific evidence in these kinds of interventions and how a variety of evaluation methods beyond RCTs can be better integrated into guidelines [75].

Overall, our collective experience across the portfolio of interventions showcased here underlines the importance of comprehensive evaluations that use a diversity of methods, encompasses a range of outcomes at various points in the causal chain and an inclusive approach working closely with different types of collaborators, and centers the experiences and meets the needs of those with greatest knowledge of their local food system. When doing so, it is essential to not be bound to rigid disciplinary norms about appropriate methods. Instead, we must seek evidence-generating methods that, although rigorous, best fit a given set of constraints and needs and provide actionable results—i.e., *good enough* evidence to inform real-world decisions.

The methodological challenges outlined here should not stop us from evaluating. It is essential to do what we can to generate credible evidence about what works to guide food system transformation to achieve societal goals. Challenges are equally opportunities for innovation of new and better methods. Similarly, we should not let methodological issues constrain the ambition of program design and implementation. Although the global policy and advocacy narrative currently focuses on "food systems transformation," so far, the existing evidence illustrates merely how to tweak around the edges of the system. Better evaluation methodologies are urgently needed to support the more ambitious transformation goal.

## Author contributions

Q3

The authors' responsibilities were as follows – LMN: conceived of the manuscript and wrote the first draft; SN: led the development of subsequent drafts; JLL, N-LA, IB, EDW, AWG, WG, CEL, MNNM, EN, CND, DP, MS, DT: contributed to the writing process; all authors: contributed to the evaluations and/or workshop on which the article draws; and all authors: read and approved the final manuscript.

## Conflict of interest

JLL reports financial support was provided by International Food Policy Research Institute and is an Editorial Board Member for *The Journal of Nutrition* and played no role in the Journal's evaluation of the manuscript All other authors report no conflicts of interest.

Q4

## Funding

This work was part of the Making Markets Work to Improve the Consumption of Safe and Nutritious Foods program, which was supported by the Bill & Melinda Gates Foundation, Federal Ministry for Economic Cooperation and Development (Germany), International Development Research Centre (Canada), Irish Aid, Ministry of Foreign Affairs of the Netherlands, and the Swiss Agency for Development and Cooperation. Jef Leroy's time was partially supported by the CGIAR Research Initiative on Sustainable Healthy Diets through Food Systems Transformation (SHiFT) which is funded through contributions to the CGIAR Trust Fund: https://www.cgiar.org/funders/.

Q5

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.tjnut.2024.04.005.

## References

Q6

[1] I.D. Brouwer, M.J. van Liere, A. de Brauw, P. Dominguez-Salas, A. Herforth, G. Kennedy, et al., Reverse thinking: taking a healthy diet perspective towards food systems transformations, Food Secur 13 (2021) 1497–1523, https://doi.org/10.1007/s12571-021-01204-5.

[2] A. Lartey, J. Meerman, R. Wijesinha-Bettoni, Why food system transformation is essential and how nutrition scientists can contribute, Ann. Nutr. Metab. 72 (3) (2018) 193–201, https://doi.org/10.1159/000487605.

[3] T.A. Reardon, D. Tschirley, B. Minten, S. Haggblade, L.S. Liverpool-Tasie, M. Dolislager, et al. [Internet], Transformation of African Agrifood Systems in the New Era of Rapid Urbanization and the Emergence of a Middle Class (2015) [cited 25 October, 2020]. Available from: https://ebrary.ifpri.org/digital/collection/p15738coll2/id/130005.

[4] M. Zurek, A. Hebinck, O. Selomane, Climate change and the urgency to transform food systems, Science 376 (6600) (2022) 1416–1421, https://doi.org/10.1126/science.abo2364.

[5] P. Webb, T.G. Benton, J. Beddington, D. Flynn, N.M. Kelly, S.M. Thomas, The urgency of food system transformation is now irrefutable, Nat. Food 1 (10) (2020) 584–585, https://doi.org/10.1038/s43016-020-00161-0.

[6] [Internet]J. von Braun, K. Afsana, L.O. Fresco, M.H.A. Hassan (Eds.), Science and Innovations for Food Systems Transformation, Springer International Publishing, Cham, 2023 [cited 3 April, 2023]. Available from: https://link.springer.com/10.1007/978-3-031-15703-5.

[7] Global Nutrition Report, Nutrition Accountability Framework Commitment Tracker [Internet], Global Nutrition Report (2022) [cited 11 May, 2023]. Available from: https://globalnutritionreport.org/resources/naf/tracker/.

[8] L.M. Neufeld, S. Hendriks, M. Hugas, Healthy diet: a definition for the United Nations Food Systems Summit 2021, in: J. von Braun, K. Afsana, L.O. Fresco, M.H.A. Hassan (Eds.), Science and Innovations for Food Systems Transformation, Springer International Publishing, Cham, 2023, pp. 21–30. Available from: https://link.springer.com/10.1007/978-3-031-15703-5_3.

[9] T. Andreyeva, K. Marple, S. Marinello, T.E. Moore, L.M. Powell, Outcomes following taxation of sugar-sweetened beverages: a systematic review and meta-analysis, JAMA Netw. Open 5 (6) (2022) e2215276, https://doi.org/10.1001/jamanetworkopen.2022.15276.

[10] A.M. Teng, A.C. Jones, A. Mizdrak, L. Signal, M. Genç, N. Wilson, Impact of sugar-sweetened beverage taxes on purchases and dietary intake: systematic review and meta-analysis, Obes. Rev. 20 (9) (2019) 1187–1204, https://doi.org/10.1111/obr.12868.

[11] Organisation for Economic Co-operation and Development, Development Assistance Committee, Glossary of Key Terms in Evaluation and Results-Based Management [Internet], 2nd Edition, OECD-DAC, Paris, 2022. Available from: https://one.oecd.org/document/DCD/DAC/EV(2022)2/en/pdf.

Q7

[12] M. Marra, R. Schwartz, K. Forss (Eds.), Evaluating the Complex: Attribution, Contribution and Beyond, 1st ed., Routledge, New Brunswick, NJ, 2011.

[13] High Level Panel of Experts on Food Security and Nutrition (HLPE), Nutrition and food systems [Internet], Food and Agriculture Organization of the United Nations, Rome, 2017. Available from: https://www.fao.org/policy-support/tools-and-publications/resources-details/en/c/1155796/.

[14] R. Davies, J. Dart [Internet], The 'Most Significant Change' (MSC) Technique: A Guide to Its Use (2005), https://doi.org/10.13140/RG.2.1.4305.3606 [cited 19 March, 2024]. Available from:.
[15] P.B. Batalden, F. Davidoff, What is "quality improvement" and how can it transform healthcare? Qual. Saf. Health Care. 16 (1) (2007) 2–3, https://doi.org/10.1136/qshc.2006.022046.
[16] S.C. Funnell, P.J. Rogers, Purposeful program theory: Effective use of theories of change and logic models, John Wiley & Sons, Hoboken, 2011.
[17] L.M. Larson, E.A. Frongillo, B.E. Kase, L.M. Neufeld, W. Gonzalez, I. Erhabor, et al., Effectiveness of the Eggs Make Kids demand-creation campaign at improving household availability of eggs and egg consumption by young children in Nigeria: a quasi-experimental study, Matern. Child Nutr. 19 (1) (2023) e13447, https://doi.org/10.1111/mcn.13447.
[18] M. Smith, D. Prescott, T. Kirkbride, R. de Mel, L. Loveday, L. Raphael, et al., SUN Business Network Evaluation 2019, Mokoro Ltd, Oxford, UK, 2019.
[19] M.K. Maredia, M. Porter, E. Nakasone, D.L. Ortega, V. Caputo, Does increasing the availability of a nutritious food produced by a small- and medium-sized enterprise increase its consumption? Evidence from a field experiment in Kenya, Appl. Eco. Perspect. Policy (2023), https://doi.org/10.1002/aepp.13402.
[20] D.L. Ortega, V. Caputo, M. Maredia, Effectiveness of nutritional information in increasing demand for enriched yoghurt among poor households in Kenya: implications for agribusiness marketing strategies, Agribusiness 39 (1) (2022) 285–290, https://doi.org/10.1002/agr.21784.
[21] CONEVAL. Evaluación de la política social: evaluación de diseño [Internet]. Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL) [cited 12 June, 2023]. Available from: https://www.coneval.org.mx/Evaluacion/MDE/Paginas/Evaluaciones_Diseno.aspx.
[22] I. Barnett, W. Gonzalez, M. Bipul, D. Chowdhury, E.W. Djimeu, A.K. Deo, et al., Improving adolescents' food choices: learnings from the Bhalo Khabo Bhalo Thakbo ("Eat Well, Live Well") campaign in Bangladesh, Field Exchange 66 (2021) 23.
[23] M.Q. Patton, The Global Alliance Formally Adopts a Theory of Transformation [Internet], Global Alliance for the Future of Food. (2020). Available from: https://futureoffood.org/insights/theory-of-transformation/.
[24] K. Skivington, L. Matthews, S.A. Simpson, P. Craig, J. Baird, J.M. Blazeby, et al., A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance, BMJ 374 (2021) n2061, https://doi.org/10.1136/bmj.n2061.
[25] B. Douthwaite, J. Mayne, C. McDougall, R. Paz-Ybarnegaray, Evaluating complex interventions: a theory-driven realist-informed approach, Evaluation 23 (3) (2017) 294–311, https://doi.org/10.1177/1356389017714382.
[26] R.G. Whitaker, N. Sperber, M. Baumgartner, A. Thiem, D. Cragun, L. Damschroder, et al., Coincidence analysis: a new method for causal inference in implementation science, Implement. Sci. 15 (1) (2020) 108, https://doi.org/10.1186/s13012-020-01070-3.
[27] International Institute for Environment and Development (IIED), Clearing the fog: new tools for improving the credibility of impact claims [Internet], IIED, London, 2016. Available from: https://www.iied.org/sites/default/files/pdfs/migrate/17359IIED.pdf.
[28] B. Lee, P. Sommerville, S. Farley, R. Mayega, Spaces MERL: Systems and complexity white paper [Internet], United States Agency for International Development (USAID), Washington, DC, 2016. Available from: https://linclocal.org/wp-content/uploads/2017/02/pa00m7qz.pdf.
[29] L.S. Taillie, M. Reyes, M.A. Colchero, B. Popkin, C. Corvalán, An evaluation of Chile's Law of Food Labeling and Advertising on sugar-sweetened beverage purchases from 2015 to 2017: a before-and-after study, PLOS Med 17 (2) (2020) e1003015, https://doi.org/10.1371/journal.pmed.1003015.
[30] E. Stern, N. Stame, J. Mayne, K. Forss, R. Davies, B. Befani, Broadening the range of designs and methods for impact evaluations [Internet]. DFID Working Paper 38, Department for International Development, London, 2012. Available from, https://www.oecd.org/derec/50399683.pdf.
[31] C. Green, C. Nyhus Dhillon, J. Meerman, G. Stone, Use of Quality Improvement Methods for Business-Led Workforce Nutrition Programmes [Internet]. Working Paper #34, Global Alliance for Improved Nutrition (GAIN), Geneva, 2023. Available from: https://www.gainhealth.org/sites/default/files/publications/documents/GAIN-Working-Paper-Series-34-Use-of-quality-improvement-methods-for-business-led-workforce-nutrition-programmes-pdf.pdf.
[32] M. Skillman, C. Cross-Barnet, R. Friedman Singer, C. Rotondo, S. Ruiz, A. Moiduddin, A framework for rigorous qualitative research as a component of mixed method rapid-cycle evaluation, Qual. Health Res. 29 (2) (2019) 279–289, https://doi.org/10.1177/1049732318795675.
[33] M. Gold, D. Helms, S. Guterman, Identifying, monitoring, and assessing promising innovations: using evaluation to support rapid-cycle change, Issue Brief, Commonw. Fund). 12 (2011) 1–12.
[34] S. Tesfaye, D. Barry, A.G. Gobezayehu, A.H. Frew, K.E. Stover, H. Tessema, et al., Improving coverage of postnatal care in rural Ethiopia using a community-based, collaborative quality improvement approach, J. Midwifery Womens Health. 59 (Suppl 1) (2014) S55–S64, https://doi.org/10.1111/jmwh.12168.
[35] L. Wilson, A Guide to Complexity-Aware Monitoring Approaches for MOMENTUM Projects [Internet], United States Agency for International Development (USAID), Washington, DC, 2020. Available from: https://usaidmomentum.org/resource/a-guide-to-complexity-aware-monitoring-approaches-for-momentum-projects/.
[36] L. Golding, J. Petraglia, Complexity Matters: Aligning the Monitoring and Evaluation of Social and Behavior Change with the Realities of Implementation, The CORE Group, Washington, DC, 2020.
[37] H. White, An introduction to the use of randomised control trials to evaluate development interventions, J. Dev. Effect. 5 (1) (2013) 30–49, https://doi.org/10.1080/19439342.2013.764652.
[38] P.J. Gertler, S. Martinez, P. Premand, L.B. Rawlings, C.M.J. Vermeersch, Impact evaluation in practice, World Bank, Washington, DC, 2011, https://doi.org/10.1596/978-0-8213-8541-8.
[39] S. Khandker, G.B. Koolwal, H. Samad, Handbook on impact evaluation: quantitative methods and practices, World Bank, Washington, DC, 2010, https://doi.org/10.1596/978-0-8213-8028-4.
[40] J.L. Leroy, E.A. Frongillo, B.E. Kase, S. Alonso, M. Chen, I. Dohoo, et al., Strengthening causal inference from randomised controlled trials of complex interventions, BMJ Glob. Health 7 (6) (2022) e008597, https://doi.org/10.1136/bmjgh-2022-008597.
[41] P. Craig, C. Cooper, D. Gunnell, S. Haw, K. Lawson, S. Macintyre, et al., Using natural experiments to evaluate population health interventions: new Medical Research Council guidance, J. Epidemiol. Community Health. 66 (12) (2012) 1182–1186, https://doi.org/10.1136/jech-2011-200375.
[42] M.A. Colchero, B.M. Popkin, J.A. Rivera, S.W. Ng, Beverage purchases from stores in Mexico under the excise tax on sugar sweetened beverages: observational study, BMJ 352 (2016) h6704, https://doi.org/10.1136/bmj.h6704.
[43] L.S. Taillie, M. Bercholz, B. Popkin, M. Reyes, M.A. Colchero, C. Corvalán, Changes in food purchases after the Chilean policies on food labelling, marketing, and sales in schools: a before and after study, Lancet Planet. Health 5 (8) (2021) e526–e533, https://doi.org/10.1016/s2542-5196(21)00172-8.
[44] T. Delahais, J. Toulemonde, Making rigorous causal claims in a real-life context: has research contributed to sustainable forest management? Evaluation 23 (4) (2017) 370–388, https://doi.org/10.1177/1356389017733211.
[45] K.M. Demmler, The Role of Small and Medium-Sized Enterprises in Nutritious Food Supply Chains in Africa [Internet], Global Alliance for Improved Nutrition (GAIN) (2020). Available from: https://www.gainhealth.org/sites/default/files/publications/documents/gain-working-paper-series-2-the-role-of-small-and-medium-sized-enterprises-in-nutritious-food-supply-chains-in-africa.pdf.
[46] J. Wun, C. Kemp, C. Puett, D. Bushnell, J. Crocker, C. Levin, Measurement of benefits in economic evaluations of nutrition interventions in low- and middle-income countries: a systematic review, Matern. Child Nutr. 18 (2) (2022) e13323, https://doi.org/10.1111/mcn.13323.
[47] G. Thai, A. Margolies, A. Gelli, N. Sultana, E. Choo, N. Kumar, et al., The economic costs of a multisectoral nutrition programme implemented through a credit platform in Bangladesh, Matern, Child Nutr. 18 (2) (2022) e13323, https://doi.org/10.1111/mcn.13323.
[48] A. Gelli, C.G. Kemp, A. Margolies, A. Twalibu, M. Katandu, C. Levin, Economic evaluation of an early childhood development center–based agriculture and nutrition intervention in Malawi, Food Secur 14 (1) (2022) 67–80, https://doi.org/10.1007/s12571-021-01203-6.
[49] A. Margolies, A. Gelli, R. Daryanani, A. Twalibu, C. Levin, When communities pull their weight: the economic costs of an integrated agriculture and nutrition home-grown preschool meal intervention in

Malawi, Food Nutr. Bull. 42 (1) (2021) 3–22, https://doi.org/10.1177/0379572120986693.
[50] C. Levin, W.A. Masters, A. Gelli, H. Harris-Fry, S. Kadiyala, S. Kalamatianou, et al., Economic Evaluation of Multisectoral Actions for Health and Nutrition [Internet], Agriculture, Nutrition and Health Academy Working Group of Economic Evaluations. Innovative Methods and Metrics for Agriculture and Nutrition Actions programme, London, 2019. Available from, https://www.advancingnutrition.org/resources/economic-evaluation-multisectoral-actions-health-and-nutrition.
[51] C. Levin, L. Tokos Harp, J. Crocker, J. Wun, C. Kemp, E. Choo, et al., Strengthening Economic Evaluation for Multisectoral Strategies (SEEMS) – Nutrition: Common Approach Guidance Document [Internet], University of Washington, 2023. Available from: https://r4d.org/resources/seems-nutrition-common-approach-guidance-document/.
[52] WHO, FAO, UNICEF, Rockefeller Foundation, Healthy Diets Metrics: Technical Expert Meeting on Harmonizing and Mainstreaming Measurement of Healthy Diets Globally, Rockefeller Foundation, Bellagio, Italy, 2022.
[53] C. Turner, A. Aggarwal, H. Walls, A. Herforth, A. Drewnowski, J. Coates, et al., Concepts and critical perspectives for food environment research: a global framework with implications for action in low- and middle-income countries, Glob. Food Sec. 18 (2018) 93–101, https://doi.org/10.1016/j.gfs.2018.08.003.
[54] J. Fanzo, L. Haddad, R. McLaren, Q. Marshall, C. Davis, A. Herforth, et al., The Food Systems Dashboard is a new tool to inform better food policy, Nat. Food. 1 (2020) 243–246, https://doi.org/10.1038/s43016-020-0077-y.
[55] Global Panel, Improving Nutrition Through Enhanced Food Environments, Policy Brief No. 7, Global Panel on Agriculture and Food Systems for Nutrition, 2017. London, UK.
[56] Food and Agriculture Organization of the United Nations, Influencing food environments for healthy diets, FAO, Rome, 2016.
[57] A. Drewnowski, E.C. Monterrosa, S. de Pee, E.A. Frongillo, S. Vandevijvere, Shaping physical, economic, and policy components of the food environment to create sustainable healthy diets, Food Nutr. Bull 41 (2_suppl) (2020) 74S–86S, https://doi.org/10.1177/0379572120945904.
[58] C. Turner, S. Kalamatianou, A. Drewnowski, B. Kulkarni, S. Kinra, S. Kadiyala, Food environment research in low- and middle-income countries: a systematic scoping review, Adv. Nutr. 11 (2) (2020) 387–397, https://doi.org/10.1093/advances/nmz031.
[59] S. Westbury, I. Ghosh, H.M. Jones, D. Mensah, F. Samuel, A. Irache, et al., The influence of the urban food environment on diet, nutrition and health outcomes in low-income and middle-income countries: a systematic review, BMJ Glob. Health 6 (10) (2021) e006358, https://doi.org/10.1136/bmjgh-2021-006358.
[60] C. Hannah, J. Davies, R. Green, A. Zimmer, P. Anderson, J. Battersby, et al., Persistence of open-air markets in the food systems of Africa's secondary cities, Cities 124 (2022) 103608, https://doi.org/10.1016/j.cities.2022.103608.
[61] C. Li, M. Mirosa, P. Bremer, Review of online food delivery platforms and their impacts on sustainability, Sustainability 12 (14) (2020) 5528, https://doi.org/10.3390/su12145528.
[62] S. Giroux, J. Blekking, K. Waldman, D. Resnick, D. Fobi, Informal vendors and food systems planning in an emerging African city, Food Policy 103 (2021) 101997, https://doi.org/10.1016/j.foodpol.2020.101997.
[63] S. Ahmed, G. Kennedy, J. Crum, C. Vogliano, S. McClung, C. Anderson, Suitability of data-collection methods, tools, and metrics for evaluating market food environments in low- and middle-income countries, Foods 10 (11) (2021) 2728, https://doi.org/10.3390/foods10112728.
[64] A. Herforth, S. Ahmed, The food environment, its effects on dietary consumption, and potential for measurement within agriculture-nutrition interventions, Food Secur 7 (2015) 505–520, https://doi.org/10.1007/s12571-015-0455-8.
[65] S. Michie, M.M. van Stralen, R. West, The behaviour change wheel: a new method for characterising and designing behaviour change interventions, Implement. Sci. 6 (2011) 42, https://doi.org/10.1186/1748-5908-6-42.
[66] R. Glennerster, K. Takavarasha, Running randomized evaluations: a practical guide [Internet], Princeton University Press, 2013 [cited 8 June, 2023]. Available from: http://www.jstor.org/stable/10.2307/j.ctt4cgd52.
[67] Z.A. Bhutta, J.K. Das, A. Rizvi, M.F. Gaffey, N. Walker, S. Horton, et al., Evidence-based interventions for improvement of maternal and child nutrition: what can be done and at what cost? Lancet 382 (9890) (2013) 452–477, https://doi.org/10.1016/s0140-6736(13)60996-4.
[68] Z.A. Bhutta, T. Ahmed, R.E. Black, S. Cousens, K. Dewey, E. Giugliani, et al., What works? Interventions for maternal and child undernutrition and survival, Lancet 371 (9610) (2008) 417–440, https://doi.org/10.1016/s0140-6736(07)61693-6.
[69] D. Hargreaves, E. Mates, P. Menon, H. Alderman, D. Devakumar, W. Fawzi, et al., Strategies and interventions for healthy adolescent growth, nutrition, and development, Lancet 399 (10320) (2022) 198–210, https://doi.org/10.1016/s0140-6736(21)01593-2.
[70] K. Kainz, A. Metz, N. Yazejian, Tools for evaluating the implementation of complex education interventions, Am. J. Eval. 42 (3) (2021) 399–414, https://doi.org/10.1177/1098214020958490.
[71] M.K. Maredia, E. Nakasone, M. Porter, S. Nordhagen, V. Caputo, E.W. Djimeu, et al., Using novel multi-method evaluation approaches to understand complex food system interventions: insights from a supply chain intervention intended to improve nutrition. Q8
[72] G.S. Cooper, H. Davies-Kershaw, P. Dominguez-Salas, U. Fahmida, B. Faye, E. Ferguson, et al., Investigating market-based opportunities for the provision of nutritious and safe diets to prevent childhood stunting: a UKRI-GCRF action against stunting hub protocol paper, BMJ Paediatr, Open 8 (Suppl 1) (2024) e001671, https://doi.org/10.1136/bmjpo-2022-001671.
[73] I. Haqiqi, M. Bahalou Horeh, Assessment of COVID-19 impacts on U.S. counties using the immediate impact model of local agricultural production (IMLAP), Agric. Syst. 190 (2021) 103132, https://doi.org/10.1016/j.agsy.2021.103132.
[74] S. Gerritsen, S. Harré, D. Rees, A. Renker-Darby, A.E. Bartos, W.E. Waterlander, et al., Community group model building as a method for engaging participants and mobilising action in public health, Int. J. Environ. Res. Public Health. 17 (10) (2020) 3457, https://doi.org/10.3390/ijerph17103457.
[75] R.J. Stoltzfus, How can the scientific community support the generation of the evidence needed to improve the quality of guidelines for micronutrient interventions? Adv. Nutr. 5 (1) (2014) 40–45, https://doi.org/10.3945/an.113.004721.