Department for Evaluation

International Initiative for Impact Evaluation

REPORT 3 2023

What do we know about the long-term outcomes of food systems interventions?

A rapid evidence assessment

The report is commissioned by the Department for Evaluation in Norad

The Department for Evaluation is responsible for conducting independent evaluations of activities funded over the ODA budget. The department is governed under a separate mandate and associated strategy issued by the Norwegian Ministry of Foreign Affairs and the Ministry of Climate and Environment. The department reports directly to the Secretary Generals of these two ministries.

Carried out by International Initiative for Impact Evaluation (3ie)

Written by Charlotte Lane (3ie), Pierre Marion (3ie), Suvarna Pande (3ie), Ingunn Storhaug (3ie), Diana Belen Cordova-Arauz (3ie), Paul Fenton Villar (3ie and University of East Anglia), and Birte Snilstveit (3ie).

December 2023

ISBN: 978-82-8369-179-5

International Initiative for Impact Evaluation

Coverphoto: **Ken Opprann** | Norad
Photo: **Sofi Lundin** | Panorama →

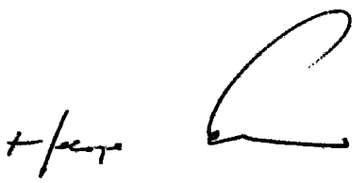
Foreword

The independent Department for Evaluation in Norad is tasked to contribute to evidence-based management of official Norwegian development assistance, through evaluative work. The department has identified a need for more evaluative knowledge about what works in the long run, in official development assistance.

Norway has launched a new food-security strategy in 2022 and the Norwegian agency for development cooperation (Norad) is managing a food security portfolio with an accompanying portfolio theory of change and knowledge plan which can support the implementation of this strategy.

To promote learning of what works in the long-run and ensure easy access to evaluative findings of long-term effects in food security, the Department for Evaluation has partnered with 3ie and funded work to assess and collate evidence on the extent to which food-security evaluations document long term effects and collate evidence from long-term evaluations.

We hope that the findings in this report, both on its own and together with the department's upcoming synthesis work, will provide useful inputs for the Norwegian aid administration in their work to achieve the Government's food security strategy.



Helge Østtveiten

Oslo, December 2023

Director, Department for Evaluation

About this report

This report presents the findings of a rapid evidence assessment (REA) on the long-term effects of food systems and nutrition interventions in low- and middle-income countries. Like systematic reviews, REAs use pre-specified systematic methods to search and screen studies for inclusion, extract and analyse data, and synthesise information around a framework. However, they often abridge one or more steps of the traditional systematic review process to accelerate the production of an output. To make this evidence assessment rapid, the search strategy employed in this work is limited to studies in the Food Systems and Nutrition Evidence Gap Map.

This report was funded by the Department for Evaluation of the Norwegian Agency for Development Cooperation (Norad). The terms of reference for this work are provided in Appendix 1. However, this report is the product of its authors, and responsibility for the accuracy of data included in this report rests with the authors. The findings, interpretations, and conclusions presented in this report do not necessarily reflect the views of the Department for Evaluation. Any errors and omissions are the sole responsibility of the authors. Please direct any comments or queries to Charlotte Lane at clane@3ieimpact.org.

About 3ie

The International Initiative for Impact Evaluation (3ie) promotes evidence-informed equitable, inclusive, and sustainable development. We support the generation and effective use of high-quality evidence to inform decision-making and improve the lives of people living in poverty in low- and middle-income countries. We provide guidance and support to produce, synthesise, and quality assure evidence of what works, for whom, how, why, and at what cost.

Suggested citation

Lane, C, Marion, P, Pande, S., Storhaug, I, CordovaArauz, DB, Fenton Villar, P., Snilstveit, B. 2023. Rapid Evidence Assessment: What facilitates success and failure of food systems interventions in the long-term. Oslo: Department for Evaluation in Norad and the International Initiative for Impact Evaluation (3ie).

Acknowledgements

Author contributions

Charlotte Lane conceived, designed and developed the main conceptual ideas and outline of this report, alongside Birte Snilstveit. Pierre Marion performed the analytic calculations and meta-analysis, with support from Paul Fenton Villar and Charlotte Lane. Suvarna Pande performed the qualitative analysis of findings, with support from Charlotte Lane. Ingunn Storhaug developed and wrote the introductory, theoretical and the methodological sections. She manages the Food Systems and Nutrition Evidence Gap Map which this work is based on. Diana Cordova-Arauz assisted with data extraction, designed the descriptive figures, and wrote the descriptive analysis of this report with support from Charlotte Lane and Ingunn Storhaug. Paul Fenton Villar took part in the data extraction process and supported the drafting of the report. All authors discussed the findings and contributed to the final manuscript.

Acknowledgements

The authors of this report are grateful for the insightful comments by Anette Wilhelmsen and Ida Kristine Lindkvist, senior advisers at the Norwegian Agency for Development Cooperation (Norad). We also want to thank Shannon Shisler for her technical review of this document.

Finally, we want to acknowledge the Germany's Federal Ministry for Economic Cooperation and Development (BMZ) through Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) through its "Knowledge for Nutrition" programme for funding the living Food Systems and Nutrition Evidence Gap Map, which this work is based on.

Executive summary

Background

The current global food system does not work for people, our planet, and prosperity. It is under threat from a range of factors, including COVID-19, climate change, and conflict. Therefore, food systems transformation is urgently needed. In order to succeed, the impacts of food systems interventions must be maintained in the long-term.

Short funding cycles limit the scope for long-term evaluation. Often, short-term impacts are assumed to reflect sustained change, but the dynamic nature of food systems means that, in reality, short- and long-term impacts may be different. There is an urgent need to know what works in the long-term to achieve food systems transformation, who it works for, and what it costs.

Objectives and questions

The overall objective of this study is to assess the time-perspective of impact evaluations of food systems interventions. Specifically, to assess whether long-term effects are evaluated, and investigate what facilitates long-term success and failure. In doing so we address the following research questions:

1. To what extent do impact evaluations of food systems interventions evaluate long-term effects (defined as 10 years or longer)?
2. To what extent do these impact evaluations document long-term environmental and climate effects?
3. What adverse effects are considered?
4. What are the effects of agricultural interventions on income, crop production, climate, and the environment in the long-term?
5. What facilitates long-term success and failure in agricultural interventions?

Methods

We conducted a rapid evidence assessment, relying on 3ie's Food Systems and Nutrition Evidence Gap Map (EGM; Moore et al., 2020). This map collects impact evaluations of food systems interventions conducted in low- and middle-income countries. Food systems interventions are conceptualised as relating to the food supply chain, the food environment, or consumer behaviour (HLPE, 2017; Brauw, 2019), and include interventions as disparate as farmer field schools, advertising regulations, and peer support networks. The Evidence Gap Map is periodically updated through systematic and comprehensive search and screening process, providing a unique source of evidence on the effects of food systems interventions.

All studies included in the Evidence Gap Map were screened to identify those measuring outcomes at least nine years after intervention initiation. These studies provided the basis for our descriptive analysis, addressing questions 1-3. To address questions 4 and 5 we conducted a more detailed assessment and analysis of a subset of studies, considering the effects of agricultural interventions on income, crop production, climate, and the environment.

Main findings

Only four per cent (n = 78) of the 2,019 impact evaluations considering food systems interventions evaluate long-term outcomes. However, the body of evidence on the long-term effects of food systems interventions has grown steadily over the last decade. Most evaluations of these interventions (53%) evaluate outcomes within a year of intervention initiation. The average length of follow-up is 2.3 years. The largest share of evaluations are in Sub-Saharan Africa (n = 31).

There is some variation in the extent to which different intervention types are evaluated in the long-term. Eight per cent of evaluations of food production interventions assess long-term outcomes; however, only one per cent of evaluations of interventions which provide messages to support behaviour change, such as nutrition classes or counselling, consider the long-term. Similarly, only two per cent of evaluations of food processing and packaging interventions take a long-term perspective.

Key policy relevant outcomes, like climate change and intervention cost-effectiveness, lack of evidence. Despite the growing evidence on the effect of the food system on climate change, very few evaluations consider the long-term effects of food systems interventions on climate or the environment (n = 4). Only five studies measuring long-term outcomes consider cost.

Limitations in the available evidence prevent us from drawing generalisable conclusions about effects. The 19 evaluations of agricultural interventions consider a range of outcomes and interventions, including agricultural education, soil and water conservation, the provision of agricultural inputs, land titling, credit, organic certification, and contract farming. Comparison between these variable interventions and outcomes is not theoretically meaningful. In addition, only seven included effect estimates are considered to have low risk of bias, causing serious concern about the quality of the underlying evidence. Effect estimates were subject to bias due to outcome measurement (73%) and confounding (66%).

Significant heterogeneity in the reported effects may reflect differences in interventions, outcomes, and contexts. The lowest reported effect is a 0.59 standard deviation reduction in coffee yield in Uganda, resulting from a Fairtrade-Organic certification program. The highest reported effect is a 0.47 standard deviation increase in the value of agricultural production in Nicaragua as a result of an agroforestry intervention. A watershed development program in India, an agricultural extension program in South Africa, and a national program providing reduced cost fertilizer and maize seeds in Zambia increased income in the long-term. On the other hand, two programs supporting fishponds in Bangladesh had no effect on income. Sustainable certification programs in Uganda and Colombia were found to decrease yield.

We identify a number of barriers and facilitators to the sustainability of outcomes. Authors of the included studies indicate that tailoring projects to local realities, leveraging dynamic contexts without inducing friction, and providing stable policies and engagement in the long-term may facilitate impact. The environment, family structure, and decisions about how resources are shared within a family can affect how outcomes are achieved. Implementing multi-faceted interventions which address the dynamic forces that affect agricultural outcomes, such as providing seeds to address availability issues and applicable farmer education to address knowledge gaps, may facilitate outcomes. Conversely, implementing complicated interventions which do not function smoothly, such as through complicated payment structures for farmers, may inhibit outcomes. Finally, stable policies and engagement may be necessary to allow time for outcomes to mature and be fully realized.

Most studies (16 of 19 studies reviewed) do not report adverse events, which could reflect: 1. the absence of such events, 2. insufficient evaluation, or 3. insufficient reporting. Evaluations in Colombia and Uganda found that sustainability certificates reduced agricultural production, possibly through insufficient fertiliser use. Another evaluation, in India, found that land titling could increase inequalities because benefits may be limited to upper caste individuals who already had significant land holdings.

Implications

The clearest implication of our study is that there is an urgent need for investment in the evidence base. However, the lack of evidence does not mean that interventions do not work in the long-term; the absence of evidence is not evidence of absence.

For policy-makers

There is insufficient evidence to serve decision-making needs. Policy-makers may consider how incentive structures could be updated to support long-term evaluations.

For program funders and implementers

Tailoring to local context, by accounting for the local environment, community dynamics, and other realities within program design, may benefit interventions. Leveraging dynamic context, by simultaneously addressing multiple factors affecting agricultural outcomes, may facilitate impact. Maintaining stable engagement and policies, including planning for project close-out, may provide the long time frame needed for some outcomes to be achieved.

For researchers and research funders

More research, including mixed-methods approaches and cost evidence, on the long-term effects of food systems interventions is needed generally. Statistical tests or adjustments to address confounding and a priori research design could improve the quality of the evidence base. Collecting data across the causal chain would provide a better understanding of how impacts develop over time.

Limitations of the method

The search strategy for this work is limited to studies already included in the Food Systems and Nutrition Evidence Gap Map. While we expect the map to be systematic, it is possible that some studies were omitted due to the broad nature of the map.

In addition, the evidence synthesis is limited to evaluations of agricultural interventions which have been evaluated in the long-term. The vast majority of evaluations do not take this time perspective, and we cannot make conclusions about the effectiveness of interventions that have not been evaluated. There could be an inherent bias in the types of interventions which receive long-term evaluations.

The high variation in the types of interventions included in the evidence synthesis makes generalization challenging. Therefore, we focus on presenting the results of individual studies and small clusters of similar evaluations. Results should be interpreted cautiously as most studies are subject to high risk of bias. Even findings from low risk of bias studies may not be applicable to other contexts. A larger, and higher quality, evidence base is needed to determine if food systems interventions work in the long-term.

Acronyms

Asian Development Bank	ADB
Germany's Federal Ministro for Economic Cooperation and Development	BMZ
Coronavirus disease 2019	COVID 19
Convention on Climate Change	COP
Danish International Development Agency	DANIDA
Department for International Development	DfID
Deutsche Gesellschaft für Internationale Zusammenarbeit's	GIZ
European Union	EUR
Expertise of the Expanding the Reach of Impact Evaluation	ERIE
Food and Agriculture Organization	FAO
Foreign, Commonwealth, and Development Office	FCDO
High Level Panel of Experts	HLPE
International Food Policy Research Institute	IFPRIR
International Fund for Agricultural Development	IFAD
International Initiative for Impact Evaluation	3ieR
Japan Society for the Promotion of Science	JSPS
Millennium Challenge Corporation	MCCR
Monitoring and evaluation	M&E
Norwegian Agency for Development Cooperation	Norad
National Institute of Health	NIH
Population, intervention, comparator, outcome and study design	PICOS
Rapid evidence assessment	REA
Swedish International Development Agency	SIDA
United States Agency for International Development	USAID
United Nations Framework Convention on Climate Change	UNFCCC

Contents

Introduction 11

The dynamics of food system interventions short- and long-term effects: Background 15

Methods 17

Criteria for including and excluding studies in this review 18

Selection of studies and data extraction 20

Analytical approach 20

State of the evidence on the long-term effects of food systems interventions on food security, nutrition, climate, and the environment 22

The volume of evidence on the long-term effects of food system interventions 23

The growth of evidence on the long-term effects of food system interventions 24

Intervention and outcome coverage 25

Geographic coverage 27

Methods of evaluations of the long-term effects of food system interventions 29

Funding 30

Synthesis of the evidence of the long-term effects of agricultural interventions on income, crop production, and environmental outcomes 34

Study quality and risk of bias 39

Effects by intervention type 40

Comparison of effects over time 43

Adverse events 44

Barriers and facilitators to long-term success 45

Cost evidence 46

Conclusion 47

Limitations in available studies 49

Implications 50

Strengths, limitations & future directions 51

References 52

List of appenndixes 63

Terms of References 64

Introduction

Today, an estimated two billion people do not have sufficient access to safe, nutritious food and water. A wide range of factors are contributing to these issues, including climate change, conflict, and COVID-19 (The Organization for Economic Cooperation and Development, 2021; Food and Agriculture Organization, 2022). While these factors continue to impact global food production, access, and affordability, they are also threatening progress towards global goals. For example, projections show that the global community is not on track to reach the nutrition targets established by the World Health Assembly or the Sustainable Development Goals (FAO, 2022; SUN, 2020; World Bank, 2021).

To get back on track, food systems interventions attempt to address vulnerabilities in our food systems that exacerbate food insecurity. These interventions promote sustainable food systems that deliver food and nutrition security for all, in ways that support economic development and protect the natural environment. They aim to build a future food system that works for people, our planet, and prosperity (Global Panel, 2020). Food systems interventions can target the food supply chain, the food environment, or consumer behaviour (HLPE, 2017; Brauw, 2019). They include interventions as disparate as farmer field schools, advertising regulations, and peer support networks. Growing interest in the implementation of food systems interventions is demonstrated by a series of recent high-profile events, including the Nutrition for Growth Summit, the United Nations Food Systems Summit, and the first ever Food Systems Pavilion at the 27th United Nations Climate Change Conference (also termed the Conference of the Parties of the UNFCCC or COP 27).

However, despite the increasing interest in food systems interventions, the long-term effects of these interventions are often unknown, both in theory and practice. While some argue that the short- and long-term food security and nutrition outcomes achieved by these interventions might be different (the effects may fade, they may increase, or only materialise with time), projects rarely evaluate the effects of interventions after a programme is complete or its initial funding is finished (ERIE, 2018).[1] In some cases, the expected long-term effects of interventions are debated and might even be controversial. For example, programs that advocate for the adoption of organic farming may inadvertently reduce farmer productivity over the long-term (Akoyi & Maertens, 2017).

In this study, we present a rapid evidence assessment (REA) on the long-term effects of food systems interventions on food security and nutrition outcomes in low- and middle-income countries. Our primary research objective is to identify and describe the available evidence on the effects of food systems interventions ten, or more, years after their inception. To achieve our objectives, we draw on the studies included in the living Food Systems and Nutrition Evidence Gap Map as of October 2022 (Moore et al., 2021). The Evidence Gap Map project systematically

1 Within the field of food systems and nutrition, the average period between the start of an intervention and the final evaluation measure is 2.3 years.

searches for and presents evidence on the effects of food system interventions. We provide novel insights on the state of the evidence identified by the Evidence Gap Map, presenting new information to answer the following research questions:

- To what extent do impact evaluations identified from the Food Systems and Nutrition Evidence Gap Map document long-term effects?
- To what extent do these impact evaluations document long-term environmental and climate effects?

To complement this, we conduct an evidence synthesis into the evidence available on the long-term effects of agricultural interventions on income, crop production, climate, and the environment. This aspect of the work aims to answer the following research questions:

- What are the effects of agricultural interventions on income, crop production, climate, and the environment in the long-term?
- What facilitates long-term success and failure in agricultural interventions?
- What adverse effects are considered by evaluations of these interventions?

The topic for additional synthesis was selected in consultation with the policy teams at the Norwegian Agency for Development Cooperation (Norad) and represents a topic of particular interest to its policy community working on food systems. It also represents a cluster of evidence identified in our initial mapping process. Our synthesis starts with a critical assessment of the quality of the evidence base. We then describe the range of effect sizes reported in different studies, compare effects over time where possible, present reported adverse events, and describe barriers and facilitators to interventions' long-term effects. We close by describing the available cost evidence.

From the 2,019 articles in the Food Systems and Nutrition Evidence Gap Map (Moore et al., 2021), we find 78 studies estimating the effects of food systems interventions on food security and nutrition outcomes in low- and middle-income countries at least ten years after their initiation. There appears to be some variation in the types of interventions evaluated in the long-term. Four per cent of evaluations overall consider long-term outcomes, but eight per cent of food production interventions, one per cent of behaviour change communication interventions, and two per cent of evaluations of food processing and packaging interventions are evaluated in the long-term. Only five studies identified report cost evidence on the long-term effects of food systems interventions.

Most long-term evaluations take place in Sub-Saharan Africa (n = 31), reflecting the longstanding interest in supporting food security in this chronically food insecure region. Ethiopia and India are the most studied countries with 11 evaluations each. However, because most studies do not report funding or implementing agencies, it is difficult to know the organisations driving this interest. Nonetheless, governments are the most commonly reported program implementers (n = 27, 35%), program funders (n = 21, 27%), and research funders (n = 36, 46%).

The most common approach for identifying causal impact is statistical matching (n = 36, 46%). This approach is often feasible when the decision to evaluate is made years after intervention initiation. Statistical matching can be conducted with endline only data. However, this can introduce bias and *a priori* statistical approaches may allow for stronger causal identification.

A total of 19 studies were included in the analysis of the effects of agriculture interventions on income, crop production, climate, and the environment. These studies consider 25 different interventions, including agricultural education (n = 5), agricultural education combined with provision of credit (n = 3), soil and water conservation (n = 4), water conservation combined with agricultural education (n = 1), sustainable certification (n = 3), land titling (n = 3), the provision of agricultural inputs (n = 1), the provision of agricultural inputs combined with agricultural education (n = 2), credit (n = 1) and contract farming (n = 2). Due to the variable nature of these interventions, it is difficult

to meaningfully combine these and develop general conclusions about effects. Individual interventions were found to be effective and can be referenced to understand how these work in local context.

The overall evidence base is relatively low quality, with 88 per cent of impact estimates included in our evidence synthesis rated as high risk of bias. Most evaluations lack the necessary statistical tests to validate the assumptions made by the statistical methods used and adjustments needed when the estimation assumptions are not met. Without requiring alternate evaluation approaches, studies could be strengthened by reporting sensitivity analysis for hidden bias in statistical matching and establishing pre-treatment parallel trends for difference-in-difference methods.

We find that, across six studies considering the effects of interventions that have an agricultural education component, there is a small, positive reported effect on production (SMD = 0.16; 95%CI: 0.09 to 0.22; $p < 0.001$). Six studies, five of which do not consider production, report that interventions with an agricultural education component do not have a statistically significant effect on income (SMD = -0.02; 95%CI: -0.13 to 0.09; $p = 0.77$). The one study that considers both outcomes reports no statistically significant change on either (Deschamps-Laporte, 2013). Two studies considering the effects of organic certification programs find that these interventions reduce crop production and have no effect on income (Akoyi & Maertens, 2017; Ibanez & Blackman, 2017). Among the five interventions which were evaluated at multiple time periods, two are reported to have no statistically significant effect at either time period; one had a positive effect early, which lost statistical significance in the follow-up; one had a positive and statistically significant effect early, which reduced but remained statistically significant two years later; and the last had statistically significant, positive effects which grew from one year to the next (Kumar & Quisumbing, 2011; Abdoulaye & Sanders, 2013; Funsani et al., 2016). Adverse events reported relate to insufficient fertilizer use (Akoyi & Maertens, 2017) and possible increases in inequality (Besley et al., 2015). Based on qualitative analysis of included studies, authors state that tailoring to local context, leveraging dynamic contexts without inducing friction, and providing stable policies and engagement may facilitate impact.

Despite the need to achieve long-term change, to our knowledge this is the first ever attempt to systematically collect the evidence about the long-term impacts of food systems interventions or to synthesise their impacts in the long-term. Our rapid evidence assessment therefore contributes to the literature by highlighting the size of the evidence on the long-term impacts of food systems and nutrition interventions. We demonstrate that this work is not only possible, but happening more often than expected. However, we also show that there is a need for continued research. Only four per cent of evaluations of food systems interventions take a long-term perspective. Although the body of evidence is modest in size, decision makers will likely find insufficient evidence to support specific decision-making needs due to sparce coverage of specific intervention-outcome combinations. Therefore, decision makers may wish to partner with researchers early in the design of interventions to support evaluability.

The studies identified can support policy-makers, funders, and implementers in understanding the long-term effects of these interventions. Individual studies can inform decision makers about the effectiveness of a particular intervention in a specific context. Understanding long-term effects can increase the reliability and validity of results and help decision makers reach cost-effective decisions, avoiding unintended consequences or investing in interventions that are ineffective in the long-term (U.S. Agency for International Development, 2021; David, 2023). We hope that this work will increase interest in what works in the long-term and support the consideration of potential long-term effects in project planning, monitoring, and evaluation.

The rest of the report is structured as follows: section 2 provides further explanation about why the effects of food system interventions may be different in the short- and long-term. Section 3 outlines our methods

and the inclusion criteria for our analysis. Section 4 presents the findings of our synopsis of the state of evidence on the long-term effects of food systems interventions on food security, nutrition, climate, and the environment. Section 5 provides the results of our synthesis of the long-term effects of agricultural interventions on income, crop production, climate, and the environment. Finally, Section 6 contains our concluding remarks, highlighting this study's key implications for policy and future research.

Photo: **Christopher Brandt** | Norad

1

The dynamics of food system interventions short- and long-term effects: Background

The processes between implementing interventions and achieving outcomes can be complex and theories of dynamic or complex relationships are difficult to accurately construct. To shape our thinking and convey our understanding of complex issues, we often simplify problems and the processes that determine them.

For example, centuries of economic research have presented simple theories about the functioning of markets, societies, and trade based on 'static models', with the dimension of time completely removed (Kay, 2006). However, the contexts in which food systems interventions are applied are not static. It is increasingly argued that the nutrition and food security outcomes achieved in the short- and long-term are not the same; they can be dynamic and change over time (Gerber et al., 2016; FAO, 2008).

For instance, in some circumstances, it is argued that an intervention that affects outcomes in the short-term could lose its effect in the long-term (Stern et al., 2012; Forss 2020; David, 2023). This phenomenon is known as fade-out or the abatement of effects (Bailey et al., 2020). This might be caused by a variety of issues, such as skill decay or other experiences which override early impacts (Green 2020; Schneider 2020; Blattman et al., 2014). Alternatively, in other circumstances, the effects of interventions, both positive and negative, may only begin to materialise in the medium- or long-term (Fross, 2020). For example, the potential for agricultural insurance to increase fertiliser and chemical use, thereby causing environmental damage, has been observed for decades (Horowitz et al., 1993; Chang et al., 2012). As another example, maternal supplements generally improve infants' anthropometric outcomes in the short-term but the full effects on children's cognitive development and health may only materialise later in life (Prado et al., 2017).

The long-term effects of interventions may occur through recursive processes (via feedback loops through which positive or negative effects build on themselves over time) or non-recursive processes ("domino effects" where improving one outcome in the medium- or long-term also improves another interacting factor; Hecht et al., 2019). It is also possible that the processes determining the short- and long-term effects of interventions vary by context. For example, a conservation agriculture intervention improved maize yields within a single season in a dry area of Malawi, but benefits were not measurable until the fifth season in a region characterised by high levels of rainfall (Ngwira et al., 2013). Thus, reliance on short-term data can result in over- or under-estimating the long-term effects of food system interventions.

However, few evaluations return to projects years after they end to measure these dynamic changes (Pollard and Lindkvist, 2020; USAID, 2021). The last outcome measurement tends to occur at the end of the project cycle and stakeholders move to the next program, making it difficult to prioritise long-term evaluations (ERIE, 2018). This focus on short-term outcomes may also be due to changes in funders programmatic priorities, unreliable funding streams, and funding being aligned with budget cycles, which are typically shorter than five years.

Recently though, there is a growing interest in long-term evaluations. USAID reflects the demand for long-term evaluations through its recently launched Expanding the Reach of Impact Evaluations (ERIE) initiative, which includes a guide for planning long-term impact evaluations (ERIE, 2018). The Norwegian Ministry of Foreign Affairs also focuses on the long-term effects of food systems in their strategy published in November 2022 (Norwegian Ministry of Foreign Affairs, 2022). They aim to reduce hunger by supporting sustainable ways of increasing food self-sufficiency in low- and middle-income countries. They seek to do this by promoting nutritious, locally produced food. This includes local food distribution, which may also create local jobs, increase income, and support food security.

2

Methods

Building on the growing interest in food systems, in this study, we present an overview of the existing evidence on the long-term effects of a broad set of food systems interventions on food security, nutrition, climate, and the environment in low- and middle-income countries. We relied on the living Food Systems and Nutrition Evidence Gap Map to identify studies evaluating the effects of food systems interventions at least a decade after their inception, and developed a descriptive analysis on the identified studies' key characteristics (e.g. interventions, outcomes, geographic location of interventions, study design, and funders).

Following this, we performed a synthesis of the evidence on the effects of agricultural interventions on income, crop production, climate, and the environment. We reviewed the quality of the evidence base and synthesised studies' reported estimates of the effects of agricultural interventions, including information on change over time where available. We then considered reported adverse effects, examined the barriers and facilitators impacting long-term outcomes, and summarised the available cost evidence. The protocol for this work was developed *a priori* and is provided in Appendix 2. Additional details on the methods are provided in Appendix 3.

Food Systems and Nutrition Evidence Gap Map and literature search

The Food Systems and Nutrition Evidence Gap Map was originally commissioned in 2020 by Germany's Federal Ministry for Economic Cooperation and Development (BMZ) through Deutsche Gesellschaft für Internationale Zusammenarbeit's (GIZ) "Knowledge for Nutrition" programme. The map supports the prioritisation of research within the food systems space and makes research more accessible to decision makers and researchers. In this capacity, it has been referenced in documents from the Millennium Challenge Corporation (Laborde et al., 2021); International Food Policy Research Institute (Njuki et al., 2021); Foreign, Commonwealth, and Development Office (Carter et al., 2021); and European Commission (European Commission, 2022). It has also been leveraged as a key database in the development of two rapid evidence assessments (Berretta et al., 2022; Kinzer, 2022) and a systematic review (Hammaker et al., 2022).

The methods underlying the Food Systems and Nutrition Evidence Gap Map have undergone two rounds of external peer review (Storhaug et al., 2022; Tree et al., 2022). The search strategy was developed in collaboration with an information search specialist. The original search for the map was conducted in 12 academic bibliographic databases and 31 sector-specific databases and websites in May 2020. The project has since become a living evidence gap map. The academic bibliographic database search has been updated four times. The last search available for this work was completed in July 2022. The grey literature search was last completed in January 2022. Studies identified through the regular updating of 3ie's Development Evidence Portal are added to the map on a rolling basis. At the time this REA began (October 2022) there were 2,074 impact evaluations included in the map (Moore et al., 2021).

Criteria for including and excluding studies in this review

For the state of the evidence review, we included all impact evaluations in the Food Systems and Nutrition Evidence Gap Map that present evidence of the effects of food system interventions 10 years, or more, after their inception (Table 1). This time period is somewhat arbitrary but has been adopted by researchers in the field of economics and those conducting randomized controlled trials (David, 2023; Bouguen et al., 2019). Furthermore, because most funding cycles are three to five years (ERIE, 2018), there was a consensus between the authors and the Department for Evaluation in Norad that outcomes achieved 10 years after the intervention started would be considered long-term effects in most cases.

The analysis for our evidence synthesis was restricted to the long-term effects of agricultural interventions on income, crop production, climate, and the environment. The selection of this topic was based on a concentration of evidence and interest from Norad's policy team. The Norwegian Ministry of Foreign Affairs' strategy prioritises increasing food security and reducing hunger through the local production of food with the expectation that this may improve income in the long-term (Norwegian Ministry of Foreign Affairs, 2022). Interventions eligible for the evidence synthesis relate to soil and water conservation; the provision of agricultural inputs; agricultural education; agricultural insurance; land titling and markets; sustainability farming certificates; agricultural credit and savings; and contract farming. Common agricultural inputs provided are seeds, fertilisers, and pesticides. Agricultural education generally takes the form of farmer field schools and extension services. Because an objective of this work is to specifically examine the extent of the evidence on climate and environmental outcomes, we considered interventions evaluating land-related outcomes as environmental outcomes for the purposes of this REA. These re-classified outcomes relate to pollution from fertilisers and cropping intensity. Table 1 presents the populations, interventions, comparators, outcomes, and study designs (PICOS) determining studies' inclusion in the evidence gap map, state of the evidence review, and evidence synthesis (additional detail in Appendix 3).

TABLE 1

Inclusion / exclusion criteria for the Food Systems and Nutrition Evidence Gap Map, state of the evidence review, and evidence synthesis

Criteria	Included	Excluded
Participants	All: Individuals in LMICs	All: Individuals in high-income countries
Interventions	Evidence Gap Map and state of the evidence review: Food systems interventions related to the production system, distribution and storage, processing and packaging, food loss and waste management, the availability and affordability of food, promotion and labeling, women's empowerment, and behavior change communication. Synthesis: Only studies considering agricultural interventions were considered for additional evidence synthesis.	All: All other
Comparison	All: Before-after, intervention-control, business as usual, alternate intervention.	All: No comparison
Outcome	Evidence Gap Map: Economic, agricultural, climate and environment, anthropometric, behavior change, bio-nutritional, developmental, diet quality and adequacy, food affordability and availability, food distribution, food safety, intrinsic motivators, micronutrient status, women's empowerment. State of the evidence review: Outcomes from the evidence gap map measured 10 years, or more, after the beginning of the intervention. Synthesis: Only studies considering crop production, income, climate, or the environment were selected for additional evidence synthesis.	All: All other
Study designs	Evidence Gap Map: Experimental and quasi-experimental impact evaluations, cost evidence, and systematic reviews State of the evidence and synthesis: Experimental and quasi-experimental impact evaluations, and cost evidence	All: Qualitative impact evaluations, descriptive or observational studies that do not assess effectiveness, modelling studies State of the evidence review and synthesis: Systematic reviews

Selection of studies and data extraction

Our review process started with re-screening all studies included in the Food Systems and Nutrition EGM for inclusion in this work. A single reviewer re-considered studies for their eligibility in the Food Systems and Nutrition EGM and recorded the time period over which each study measured outcomes. We excluded fifty-seven studies from the map during this process, most of which were linked to other studies which had already been included in the map.[2] As such, 2,019 studies were eligible for inclusion in this rapid evidence assessment.

Studies considering outcomes over 10 years, or more, were included for this review. Bibliographic, geographic, and other descriptive data was previously extracted. We extracted further qualitative and quantitative information from studies evaluating agricultural interventions, including effect sizes, barriers and facilitators, cost, and adverse effects. Quantitative data contained within included impact evaluations was extracted in duplicate and team members met to discuss and reconcile any differences. No additional search was conducted to identify qualitative studies. Rather, qualitative data within included impact evaluations was extracted by one team member. Impact evaluations of agricultural interventions included in the evidence synthesis were appraised, in duplicate, using a rapid critical appraisal tool (Appendix 2).

2 Linked studies are those that are published on the same intervention and present similar analysis. Often, this occurs when a working paper and a journal article are published on the same intervention, generally by the same authors. Only one publication among a set of linked studies is retained to avoid over-representing the evidence base. However, the identification and systematic removal of linked studies can be challenging and some linked studies were identified during this screening process. To maintain the systematic nature of the evidence base, the same decision rules applied by the Food Systems and Nutrition Evidence Gap Map were applied to identified linked studies and the evidence gap map was updated with these decisions.

Analytical approach

We present a descriptive analysis of the volume and characteristics of the studies we identify on the long-term effects of food systems interventions on food security, nutrition, climate, and the environment. Following this, we conduct a quantitative and qualitative evidence synthesis on the effects of agricultural interventions on income, crop production, and the environment in the long-term. No studies were identified that considered the effects of agricultural interventions on climate outcomes, so this could not be examined. In addition, the three studies identified on agricultural insurance were determined to be extreme outliers, likely due to the methods employed. They considered the same program, implemented in China, and calculated results at the province level, resulting in a sample size of 31. These studies are presented in Appendix 6, Appendix Table 1, and Appendix Table 2, but not discussed in the main report.

In our quantitative analysis, to account for variation in the units of outcome measures (ex. kilograms per hectare and kilograms per acre), we converted all reported and extracted effects into standard deviation changes in the outcome attributed to the intervention (Appendix 3). We use forest plots to illustrate the range of these standardized effects and present standardized, reported effects by intervention. When interventions could be meaningfully combined, we conducted a meta-analysis using inverse variance. This approach provides a summary effect estimate with studies weighted by the reported precision of the estimate (i.e. the variance). Additional sub-group and meta-regressions are presented in Appendix 6 to provide descriptive summary information across intervention types, regions, study methods, and more. However, these analyses often combine highly variable interventions, so should be interpreted cautiously and are presented largely for illustrative purposes.

In our qualitative analysis, to investigate how outcomes were achieved, we conducted a thematic analysis following the method by Thomas and Harden (2008). The synthesis progressed in three stages. First, one designated coder from the team read the full text of the studies included in the evidence synthesis and extracted insights into the following areas of inquiry: intervention description, sustainability, unintended consequences, adverse effects, hypothesized mechanism of action, and cost evidence. Insights were generally drawn from the author's own conclusions

about what worked well or did not, in the included interventions. Because none of the included evaluations were mixed-methods, qualitative data is not available to validate authors' conclusions. Insights were paraphrased and organised in Excel with each insight recorded in a unique row under the corresponding area of inquiry.

In the second stage, the single coder applied a deductive coding process and developed broad analytical themes. This was done by organising insights into similar clusters and then summarising the insights in each cluster. During this stage, the coder returned to the studies as needed. The purpose of this process was to identify common barriers to impact, facilitators of impact, causal mechanisms, and unintended consequences.

In the final stage, a second reviewer validated these themes, also returning to the studies to confirm conclusions and add nuance. The reviewer then arranged themes to present them coherently in the sections on adverse events, barriers and facilitators to long-term success, and cost evidence.

Qualitative findings may be specific to the study setting and may not translate as general conclusions. This does not however, take away from the thematic understanding of what worked and did not work for each intervention.

Photo: **Thomas Mukoya** | Reuters | NTB

3

State of the evidence on the long-term effects of food systems interventions on food security, nutrition, climate, and the environment

In this section, we present our descriptive analysis of the volume and characteristics of the studies we identify on the long-term effects of food systems interventions on food security, nutrition, climate, and the environment. First, we examine the overall volume and growth of evidence over time. We then consider the distribution of evaluations based on the interventions, outcomes, geography, methods, and funders reported.

The volume of evidence on the long-term effects of food system interventions

Overall, we find 78 impact evaluations examining the outcomes of food system interventions a 10, or more, years after their inception. This is four per cent of the 2,019 impact evaluations in the Food Systems and Nutrition Evidence Gap Map. On average, studies across the Food Systems and Nutrition Evidence Gap Map consider the outcomes of interventions 2.3 years after intervention initiation. Among the studies following interventions for 10 years, or more, the average study period is 14.5 years. The study that considers the longest period of time examines the outcomes of land rights reforms in India starting in the 1940s on wages in 2002 (approximately 62 years, Besley et al., 2015).

The number of identified impact evaluations generally declines as the length of the study period increases. Although we find 1,076 studies reporting on the outcomes of an intervention a year or less years after an intervention started, 533 studies report on outcomes between two and four years after intervention inception, and 208 studies consider between five and nine years.

TABLE 2

Number of impact evaluations according to the length of the study period

Study period	Number of impact evaluations	Percentage of total number of impact evaluations
One year or less	1,076	53%
2 to 4 years	533	30%
5 to 9 years	208	10%
10+ years	78	4%
10 to 14 years	55	3%
15 to 19 years	13	1%
20+ years	10	0.5%
Not applicable [1]	110	5%
Not indicated [2]	11	1%
Not available [3]	3	0%

This pattern persists when we examine the frequencies of studies reporting long-term effects. We find that 55 impact evaluations report on the outcomes of interventions 10 to 14 years after they started compared to 13 impact evaluations for the timeframe 15 to 19 years. Only 10 impact evaluations consider outcomes 20, or more, years after intervention initiation (Table 2).

Notes table 2: Study period refers to the number of years between the year of the intervention's initation and the year of the final outcome measurement. The total number of impact evaluations from the Food Systems and Nutrition Evidence Gap Map is 2,019.

[1] Study protocols and ongoing studies are coded as "not applicable" because the final outcome measurement has not l.

[2] Completed studies for which the number of years between intervention inception and the final outcome measure are not reported are coded as "not indicated".

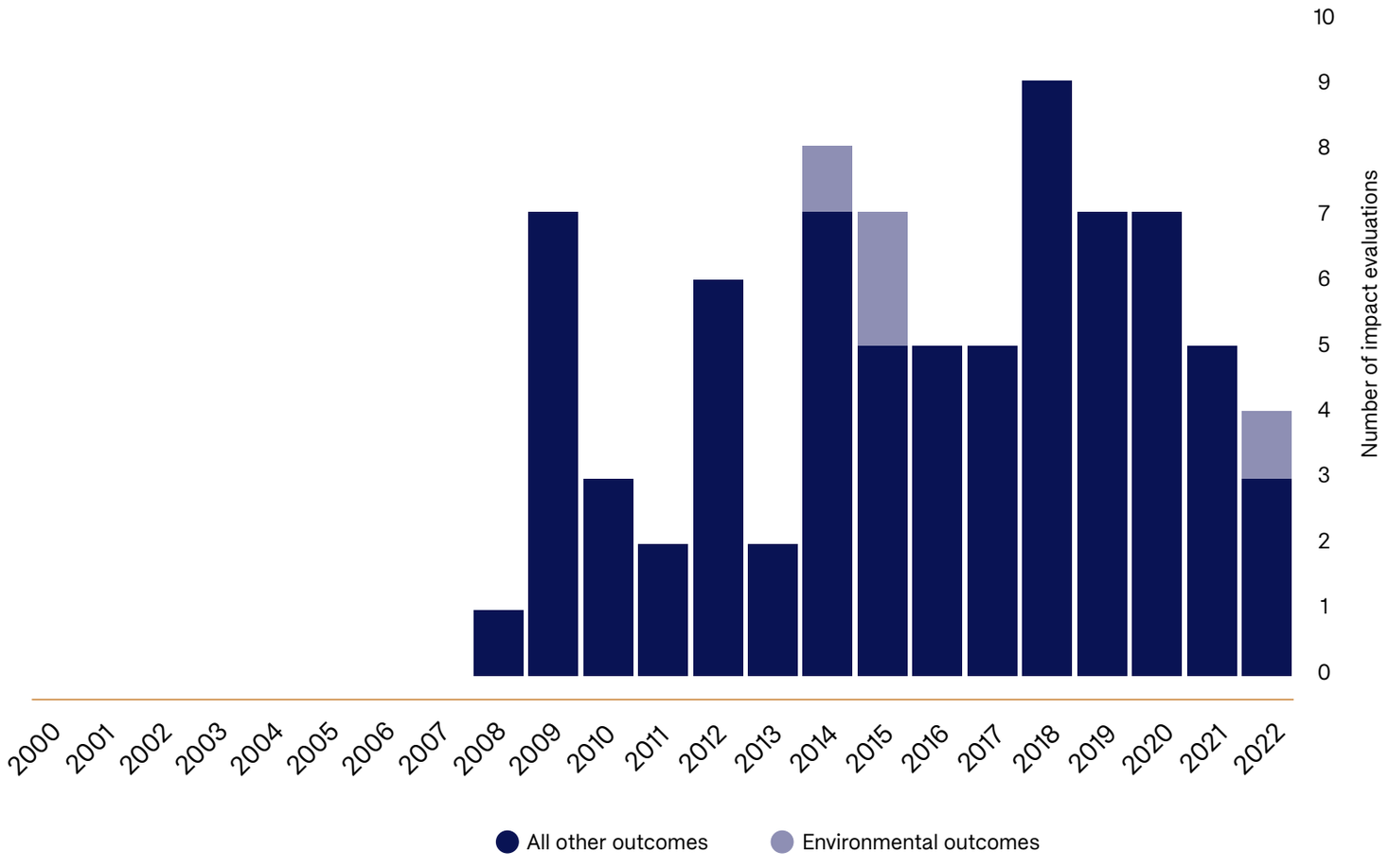
[3] Studies for which we were not able to retrieve the full text are coded as "not available". This happened for a small number of studies which were originally coded but PDFs were not saved and the study is no longer available.

The growth of evidence on the long-term effects of food system interventions

There has been a significant increase in the number of impact evaluations on the long-term effects of food systems interventions on food security, nutrition, and the environment in the last decade (Figure 1). Over three-quarters of impact evaluations (n = 65) following outcomes a decade, or more, after the inception of a food systems intervention have been published since 2012. This accounts for more than a three-fold increase in the total number of studies during the past 10 years (increasing from 13 studies before 2012 to 78 in 2022). The peak in the number of impact evaluations published is in 2018, with nine studies published; however, this is closely followed by eight studies published 2014. There is a noticeable decrease in the number of impact evaluations identified in 2022 (n = 4), but it is still too early to conclude that this represents a true trend in the data since this is a mid-year figure. All studies considering environmental or climate outcomes were published within the last decade.

FIGURE 1

Number of long-term impact evaluations of food system interventions by year



Notes: The values for 2022 represent a mid-year figure reflecting that the last update to the Food Systems and Nutrition Evidence Gap Map was completed in July 2022

Intervention and outcome coverage

Over half of the long-term evaluations identified consider interventions related to food production (n = 43, 55%; Table 3). Specifically, impact evaluations consider agricultural extension (n = 11; 13%, Appendix 3) and water access and management programs (n = 10, 13%). Evaluations of the direct provision of foods and the provision or use of supplements (n = 8, 10% each) are also common. Evaluations mostly consider agricultural (n = 32, 41%) and economic (n = 40, 51%; Table 3) outcomes. The next most commonly considered outcome is anthropometrics, with 15 studies (19%). The most common intervention-outcome groups evaluated are the effects food production interventions on economic (n = 29, 37%) and agricultural (n = 27, 35%) outcomes. Most often, these evaluations consider the effects of agricultural extension programs on economic outcomes (n = 9; 12%; Appendix 3).

Outcomes related to climate and the environment, including land-related outcomes, were only measured in four studies. Three of these were similar, considering soil contamination, fertiliser pollution, and cropping intensity. These studies considered land rights, agricultural extension programs, and agricultural insurance. One study considered the impact of a trade regulation intervention on non-food waste produced (Ibanez & Blackman, 2015).

Gaps in the literature base on the long-term effects of food systems interventions are generally similar to gaps in the literature on the effects of food systems interventions. These include evaluations of interventions related to food promotion and labelling, food quality and safety, and food loss and waste management. However, some intervention groups are more likely to be evaluated in the long-term than others.[3] Although the largest deviations from the expected four per cent of studies considering long-term evaluations are observed in intervention groups with few evaluations overall, some better studied intervention groups show modest deviations from the expected value. For example, eight per cent of evaluations of food production interventions (n total = 583) consider long-term outcomes, but only one per cent of behaviour change communication evaluations (n total = 555) and two per cent of evaluations of food processing and packaging interventions (n total = 297) consider long-term outcomes. Outcomes related to advertising and labelling, food loss, regulations, and time use are not measured in the long-term. When prioritizing research, careful consideration should be given to the theories of change and if long- and short-term impacts are expected to be different.

There are also difference in the types of interventions evaluated over time periods more or less than the mean among long-term evaluations (14.5 years).[4] However, this is likely driven by small sample sizes. For example, the only long-term evaluation on food safety considered a time period of 10 years. Of the three evaluations of women's empowerment interventions within the food system, two (66%) consider a 15-year time frame.

3 Chi-squared test based on the expectation that 4% of the studies in each intervention group would be long-term returned a p-value less than 0.00. This expectation is based on determination that 4% of the studies overall considered long-term outcomes.

4 Chi-squared test based on the expectation that 50% of the studies in each intervention group would consider outcomes longer than 14.4 years returned a p-value of 0.01. This expectation is based on the mean time period considered long-term evaluations.

TABLE 3

Frequency of long-term impact evaluations by intervention-outcome groups[1]

Outcomes \ Interventions	Behaviour change communication	Food loss and waste management	Food production	Food provision / price reduction	Food transport / storage	Processing and packaging	Quality and safety	Women's empowerment in the food system	Total unique count[3]
Economic	1	0	29	6	5	0	0	2	**40**
Agricultural	0	0	27	1	5	1	0	1	**32**
Climate / environment	0	0	3	0	1	0	0	0	**4**
Anthropometric	3	0	1	8	0	2	0	1	**15**
Bio nutritional	0	0	1	0	1	1	0	0	**1**
Behaviour change	0	0	7	2	1	0	0	0	**10**
Developmental	0	0	0	2	0	3	0	0	**5**
Diet quality / adequacy	0	0	4	2	1	0	1	0	**7**
Food affordability / availability	1	0	7	2	1	1	0	0	**9**
Food distribution	0	0	0	0	1	0	0	0	**1**
Food safety	0	0	0	0	0	0	1	0	**1**
Intrinsic motivators	2	0	1	0	0	0	0	0	**3**
Micronutrient status	0	0	1	2	0	2	0	0	**5**
Women's empowerment	0	1	4	1	0	0	0	2	**7**
Total unique count[2]	6	1	43	18	7	5	1	3	**78**

[1] Numbers reflect the number of unique evaluations considering that intervention-outcome combination. Colours represent the relative number of studies with light brown reflecting fewer studies and dark brown reflecting more studies.

[2] The total may be less than the sum of the column as a single study can be in several cells

[3] Total number of unique studies may be less than the sum of a row as a single study can be in several cells

Geographic coverage

The evaluated interventions were implemented in 27 countries across the world. Over a third of the evaluated interventions were implemented in Sub-Saharan Africa (n = 31; 40%), but smaller clusters of impact evaluations are also available on interventions in the Latin America and the Caribbean (n = 17, 22%), South Asia (n = 19, 24%), and East Asia and the Pacific (n = 8, 10%). Few studies take place in the Middle East and North Africa (n = 1, 1%) or Europe and Central Asia (n = 2, 3%). India and Ethiopia are the most common countries for long-term evaluations to take place in (n = 11 each, 14%). This is followed by China and Bangladesh (n = 6, 8%, each). No study took place in multiple countries.

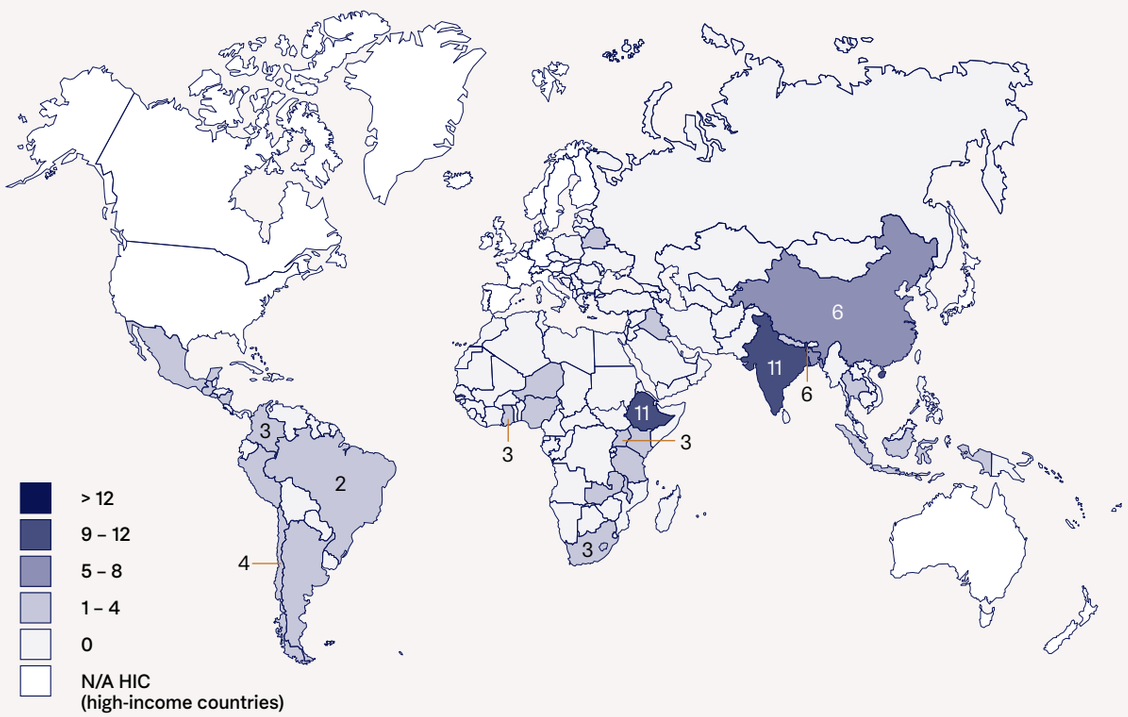
Although there are some differences in the types of interventions and outcomes considered in Sub-Saharan Africa and the rest of the world, it is not clear that these differences are meaningful (Appendix 5). No interventions related to food loss and waste management, food quality and safety, or women's empowerment interventions have been evaluated in Sub-Saharan Africa. No long-term evaluations in Sub-Saharan Africa considered outcomes related to climate and the environment, food safety, or micronutrient status. However, in all of these cases, few evaluations considered these topics generally, so it is not surprising that none occurred in Sub-Saharan Africa.

The distribution of studies geographically roughly matches the prevalence of food insecurity based on the Food Insecurity Index (Figure 2). However, population size also seems to play a role in dictating where long-term evaluations are conducted. For instance, while India is moderately food insecure (Food Insecurity Index = 3.96), it is tied for the most studies (n = 11). There are no evaluations in the Central African Republic, the Democratic Republic of the Congo, and Mozambique despite high food insecurity in these countries.

FIGURE 2

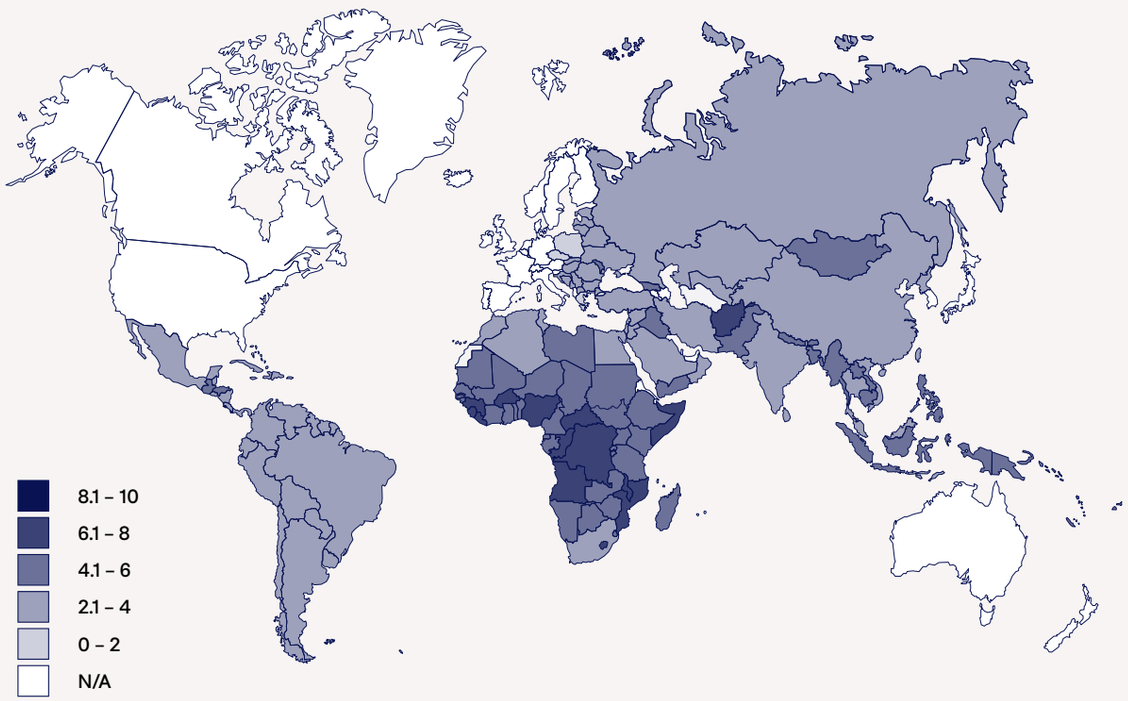
Correspondence between the number of long-term impact evaluations and the Food Insecurity Index, by country

Impact Evaluations by country



Notes: The map's value labels identify the number of studies from each country where three or more studies have been identified.

Food Insecurity Index



Notes: Higher number means higher food insecurity. The food insecurity score for each country is derived from the inverse of the Access to Food Index reported in the Global Food Security Q2 2022 report (Deep Knowledge Analytics, 2022).

Methods of evaluations of the long-term effects of food system interventions

Long-term impact evaluations generally employ quasi-experimental designs for causal inference (n = 57, 73%). The most common approach for evaluating food systems interventions in the long-term is statistical matching (n = 36; 46%; Table 4). Unlike many other study designs, such as randomised controlled trials (n = 21, 27%), statistical matching is feasible when the decision to evaluate is made years after intervention initiation, as seems to be common in the case of long-term evaluations. These evaluations are often conducted without baseline data, relying o nly on data collected at the time of the evaluation. Sometimes, statistical matching approaches are combined with other approaches, such as fixed effects estimation (n = 7, 10%) and instrumental variables (n = 4, 5%).

Other types of evaluation methods featuring in this literature include fixed effects (including difference-in-difference; n = 24, 31%), instrumental variables (n = 15, 19%) and regression discontinuity designs (n = 1, 1%). These designs are generally less common in the broader evaluation literature due to the data requirements often necessary to create a convincing application of these approaches. For example, fixed effects and difference-in-difference studies can be challenging as they require panel data, where data is collected over time for both the intervention and control groups. Meanwhile, identifying interventions with discontinuities or thresholds determining policy assignment can be challenging without in-depth knowledge of the implementation of policies. This is because not all policy intervention assignments are characterised by such thresholds or discontinuities. Instrumental variables similarly require in-depth knowledge of policies to ascertain whether exogenous factors determine assignment or uptake. Even with this detailed knowledge, it may not be possible to identify suitable instrumental variables that explain variation in a policy variable but have no independent effect on the outcome of interest (Glewwe & Todd, 2022).

The included studies offer some insight into the types of data that can be useful in developing instrumental variables and regression discontinuities. Manley and colleagues (2015) use the amount of monetary transfers households received as an instrument to determine the effect of a conditional cash transfer program. Geography is used as an instrumental variable in three studies. Akoyi and Maertens (2018) use distance to the nearest coffee export company's washing station as an instrument for farmers' adoption of private sustainability standards in Uganda. To determine the likelihood of engaging with agricultural cooperatives, Francesconi and Ruben (2012) use a dummy variable for urban and rural residency as well as a variable for enrolment in the local military air force. Location, rainfall, and prices are used as instrumental variables to support the estimation of

TABLE 4

Number of impact evaluations according to the length of the study period

Evaluation method	Number of studies	Percent of studies
Fixed effects (including difference-in-difference)	24	31%
Instrumental variables	15	19%
Randomised controlled trial	21	27%
Regression discontinuity design	1	1%
Statistical matching	36	46%

Notes: The total number of long-term evaluations is 78 (18 studies used several study designs).

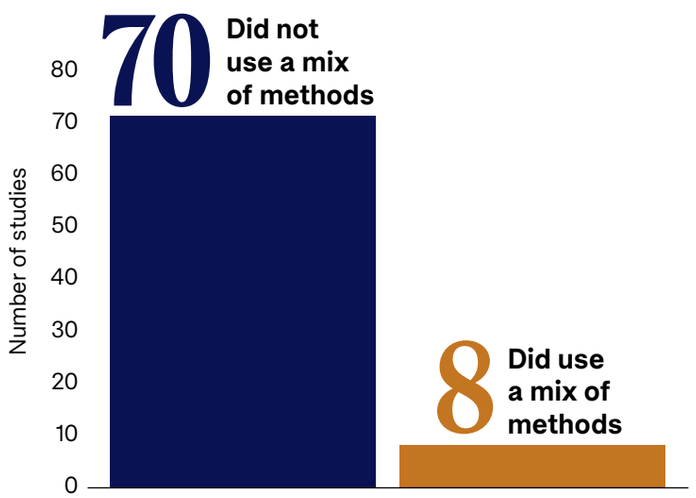
the effect of a safety net program on yield in Ethiopia (Filipski et al., 2017). Another study uses lagged values of fertile land holdings, adults in the household, and livestock to serve as an instrument in the identification of the impact of an agricultural extension program (Dercon et al., 2009).[5] One long-term evaluation uses a fuzzy regression discontinuity design to evaluate the impact of an unconditional cash and food assistance program in Bangladesh (Bari et al., 2022). The treatment is partially based on land ownership; those owning 0.5 acre or less are eligible for the program.

The idea of complementing rigorous quantitative impact evaluations with qualitative data has been gaining currency in recent years. Qualitative information can provide valuable insights to better understand and explain findings from quantitative analysis and they may also help validate findings (White, 2008). We find that approximately 10 per cent (n = 8; Figure 3) of impact evaluations on the long-term effects of food system interventions collect and analyse both qualitative and quantitative data. The qualitative portions of mixed methods research investigated implementation factors which might have effected outcomes, how local context effected outcomes, or the underlying theory of change. However, mixed methods studies remain a minority of all impact evaluations and future research would benefit from the richness of information gained from triangulating between quantitative and qualitative data.

5 We did not conduct an exhaustive review of all the studies using this approach. These examples are provided for illustrative purposes.

FIGURE 3

Frequency of mixed methods evaluations



Only five studies report cost evidence, accounting for 6 per cent of evaluations considering a time frame of 10 years or more. Three of these studies conduct cost-benefit analysis, one reports unit costs, and one calculates an internal rate of return. All of these evaluations consider agricultural interventions, such as land registration, the provision of seeds, or a soil and water conservation program. They focus on comparing the cost of interventions with changes in production or yield. No evaluations consider the cost of nutrition specific interventions and their corresponding health benefits, which are more challenging to quantify.

Funding

Approximately half of long-term evaluations do not provide information on the program implementing agency (n = 38, 49%, Table 5, page 33), the program funding agency (n = 51, 65%), or the research funding agency (n = 33, 42%). However, government agencies are the most commonly reported implementing agencies (n = 27, 35%). These are often the governments of the low- and middle-income countries where the intervention is implemented.

The second most commonly reported implementing agencies are non-profit organisations (n = 13, 17%), then international aid agencies (n = 3, 4%). No study reports charitable or private foundations as implementers. The governments of China, Ethiopia and India each implemented three of the evaluated interventions (Table 6, page 34). No other organization implemented more than one.

Government agencies are also the most commonly reported program funders (n = 21, 27%), followed by international aid agencies (n = 13, 17%). Non-profit organisations (n = 4, 5%), academic institutions (n= 2, 3%), international financial institutions (n = 2, 3%), charitable foundations (n = 1, 1%), and for-profit firms (n = 1, 1%) are less often reported as program funders. United States Agency for International Development (USAID), the Government of Ethiopia, and the

Government of India are the most program common funders (n = 4 each). Several other organisations were reported to have funded two programs each: International Fund for Agricultural Development (IFAD), Inter-American Development Bank (IDB), Government of Zambia, and Danish International Development Agency (DANIDA).

The most commonly reported research funders are also government agencies (n = 33, 42%) and international aid agencies (n = 19, 24%). Academic institutions (n = 8, 10%) and charitable or private foundations (n = 6, 8%) are more common as research funders than they are as program funders and implementers. Non-profit organisations (n = 10, 13%) and international financial institutions (n = 5, 6%) are also reported as funding some long-term evaluations. The most common funders reported are the National Institutes of Health (NIH; n = 5), Department for International Development (DfID; n = 5), followed by the World Bank Group (n = 4), the European Union (n = 3), Japan Society for the Promotion of Science (JSPS; n = 3) and Swedish International Development Agency (SIDA; n = 3).

Photo: **Gunnar Zachrisen** | Norad

TABLE 5

Number of studies by program implementing agency, program funding agency, and research funding agency types

	Implementing agency	Program funding agency	Research funding agency
Academic institution	1 (1%)	2 (3%)	8 (10%)
Charitable or private foundation	0 (0%)	1 (1%)	6 (8%)
For-profit firm	1 (1%)	1 (1%)	0 (0%)
Government agency	27 (35%)	21 (27%)	36 (46%)
International aid agency	3 (4%)	13 (17%)	19 (24%)
International financial institution	1 (1%)	4 (5%)	5 (6%)
Non-profit organisation	13 (17%)	4 (5%)	10 (13%)
Not specified	38 (49%)	51 (65%)	33 (42%)

Notes: The total number of long-term evaluations is 78. Where more than one agency type is reported, multi-coding is permitted (i.e. a study may include more than one listed funder). As a result, the percent values add to more than 100. However, if multiple agencies are reported for one study and they fall into the same agency type, these are counted once.

TABLE 6

Most commonly reported implementing agencies and funders

	Implementing agency	Program funding agency	Research funding agency
1.	Government of China Government of Ethiopia Government of India *(3)*	United States Agency for International Development (USAID) Government of India Government of Ethiopia *(4 each)*	Department for International Development (DfID) National Institutes of Health (NIH) *(5 each)*
2.	NA, all others only implemented a single intervention	International Fund For Agricultural Development (IFAD) Government of Zambia Danish International Development Agency (DANIDA) Inter-American Development Bank *(2 each)*	World Bank Group *(4)*
3.	NA, all others only implemented a single intervention	NA, all others only funded a single study	European Union Japan Society for the Promotion of Science (JSPS) Swedish International Development Agency (SIDA) *(3 each)*

Notes: Where more than one agency is reported by studies, multi-coding is permitted (i.e. a study may include more than one listed funder).

4

Synthesis of the evidence of the long-term effects of agricultural interventions on income, crop production, and environmental outcomes

In this section, we present the results of our analysis on the long-term effects of agricultural interventions on income, crop production, and environmental outcomes. Although climate outcomes were also eligible, no studies were identified which considered the effects of agricultural interventions on climate. After a brief overview of the findings, we describe the overall quality of the evidence base and present the effects of each type of agricultural intervention on all three outcomes, alongside a discussion of why outcomes may have been observed. Following this, we describe how these effects can develop over time, possible adverse effects, barriers and facilitators of impact, and cost evidence.

Our analysis is based on 19 studies reporting on the effects of 25 interventions, including agricultural education (n = 5), agricultural education combined with provision of credit (n = 3), soil and water conservation (n = 4), water conservation combined with agricultural education (n = 1), sustainable certification (n = 3), land titling (n = 3), the provision of agricultural inputs (n = 1), the provision of agricultural inputs combined with agricultural education (n = 2), credit (n = 1) and contract farming (n = 2).[6] The geographic coverage of these interventions is broad and consists of interventions implemented in sub-Saharan Africa (n = 12), Latin America and the Caribbean (n = 7), South Asia (n = 5), and East Asia (n = 1). A summary of the included studies is provided in Appendix Table 1 and the standardised effect estimates are reported in Appendix Table 2.

Due to the variation in intervention types (Figure 4) and the overall low quality of the evidence base (Figure 5), it is not possible to draw meaningful, generalized conclusions about the likely effects of these interventions. However, we note that this does not mean that the interventions do not work in the long-term: the absence of evidence is not evidence of absence. Many individual interventions are found to be effective, likely due to the specific activities implemented and local context. Rather than abandoning these interventions, more research is needed to understand their likely effects in the long-term.

The largest reported effect comes from a in Nicaragua which found that an educational agroforestry program improved the value of production by 0.47 standard deviations (95%CI: 0.26 to 0.69; p < 0. 001). In contrast, the most negative effect comes from a Fairtrade-Organic certification program in Uganda, which is reported to have reduced coffee yield by 0.59 standard deviations (95%CI: – 0.78 to – 0.41, p < 0.001).

We identify small, positive, and statistically significant reported effects on crop production from six evaluations of interventions that include agricultural education components (k = 6, SMD[7] = 0.16; 95%CI: 0.09 to 0.22, p < 0.001; Appendix 6). However, there is no statistically significant reported effect of interventions which include agricultural education components on income (k = 6; SMD = −0.02; 95% CI: −0.13 to 0.09; p = 0.77). The single study that considers both outcomes reports no effect on either (Deschamps-Laporte, 2013). Half of the studies included in these analyses include additional activities: the provision of agricultural inputs, water conservation activities, or credit.

6 Three studies on China's policy oriented agricultural insurance program were also identified. However, these conducted their evaluations on very small sample sizes, based on China's 31 provinces, and were identified as statistical outliers. They are included in all analyses in the appendix, but not presented in the main report (Niu et al., 2022; Lv, 2020; Li & Wang, 2022).

7 The standardized mean difference is a common statistical measure used in meta-analysis to account for variation in the units used to measure continuous outcomes. Because the outcomes measured by included studies are often different (ex. production per hectare and production per acre) effect estimates are transformed into standard deviation changes attributed to the intervention. This value therefore indicates that, when effect estimates are weighted by study precision, these interventions are reported to increase crop production by 0.16 standard deviations.

FIGURE 4A

Summary of estimates of the long-term effects of agricultural interventions on income and crop production

Study	Income received	Effect size with 95% CI	Weight (%)
Kumar & Quisumbing 2011 - Credit and agricultural education (improved vegetable varieties)		-0.28 [-0.59, 0.03]	5.65
Akoyi & Maertens 2017 - Certification (Fairtrade-Organic)		-0.15 [-0.33, 0.03]	7.68
Kumar & Quisumbing 2011 - Credit and agricultural education (long-term leases of fish ponds)		-0.14 [-0.44, 0.16]	5.83
Ibanez & Blackman 2015 - Certification		-0.10 [-0.36, 0.17]	6.26
Deschamps-Laporte 2013 - Agricultural education		-0.07 [-0.22, 0.08]	8.13
Kumar & Quisumbing 2011 - Credit and agricultural education (HH owned individual fish ponds)		-0.01 [-0.31, 0.29]	5.76
IFAD 2018 - Agricultural education		0.03 [-0.06, 0.12]	8.91
World Bank 2009 - Soil and water conservation		0.03 [-0.09, 0.15]	8.60
Abebe & Bekele 2014 - Soil and water conservation		0.03 [-0.31, 0.37]	5.28
Romero & Melo 2021 - Soil conservation		0.17 [0.09, 0.25]	9.02
Besley et al. 2016 - Land titling		0.17 [-0.06, 0.41]	6.87
Funsani et al. 2016 - Provision of agricultural inputs		0.20 [-0.03, 0.42]	6.99
Baiyegunhi et al. 2018 - Agricultural education		0.23 [0.00, 0.45]	6.96
Datta 2014 - Soil and water conservation		0.23 [-0.03, 0.48]	6.50

Notes: * reflects effect estimates that have a low risk of bias. A summary effect estimate is not presented as effects of such varying interventions cannot be meaningfully combined.

FIGURE 4B

Summary of estimates of the long-term effects of agricultural interventions on income and crop production

Study	Income received	Effect size with 95% CI	Weight (%)
Akoyi & Maertens 2017 - Certification (Fairtrade-Organic)		-0.59 [-0.78, -0.41]	6.64
Ibanez & Blackman 2015 - Certification		-0.27 [-0.54, -0.00]	4.81
World Bank 2009 - Soil and water conservation		-0.16 [-0.28, -0.04]	8.21
Chankrajang 2015 * - Land titling		-0.13 [-0.62, 0.35]	2.26
Ruml & Qaim 2020 - Contract farming (marketing contract)		0.00 [-0.17, 0.18]	6.74
Abebe & Bekele 2014 - Soil and water conservation		0.05 [-0.27, 0.37]	3.95
Gibbons et al. 2016 * - Water conservation and agricultural education		0.07 [0.05, 0.10]	9.70
Gibbons et al. 2016 * - Provision of agricultural inputs and agricultural education		0.09 [0.07, 0.12]	9.70
Deschamps-Laporte 2013 - Agricultural education		0.11 [-0.04, 0.27]	7.39
Ruml & Qaim 2020 - Contract farming (resources contract)		0.14 [-0.04, 0.32]	6.74
De los Santos-Montero & Bravo-Ureta 2017 - Agricultural education (forestry)		0.26 [0.10, 0.42]	7.20
Abdoulaye & Sanders 2013 - Provision of agricultural inputs and agricultural education		0.27 [0.12, 0.42]	7.46
Funsani et al. 2016 - Provision of agricultural inputs		0.37 [0.15, 0.60]	5.70
Melesse & Bulte 2015 - Land titling		0.40 [0.22, 0.58]	6.74
De los Santos-Montero & Bravo-Ureta 2017 - Agricultural education (agroforestry)		0.47 [0.26, 0.69]	5.97

Notes: * reflects effect estimates that have a low risk of bias. A summary effect estimate is not presented as effects of such varying interventions cannot be meaningfully combined.

There is suggestive evidence from three studies that effects of interventions change over time within the same context. An agricultural education program in Bangladesh (Kumar & Quisumbing, 2011) and an intervention in Niger providing improved sorghum technology packages, inorganic fertilizers, fungicide and agronomic recommendations (Abdoulaye & Sanders, 2013) are reported to have improved outcomes in the short-term, but effects reduced or lost statistical significance over time. Conversely, a national program in Zambia which provided reduced cost fertilizer and seeds had a modest impact on yields after 11 years and a larger impact the following year (Funsani et al., 2016).

Some authors report adverse events around insufficient fertilizer use (Akoyi & Maertens, 2017) and increased inequality (Besley et al., 2015). Both organic certification programs are reported to have reduced crop production, but had no statistically significant effect on income (Akoyi & Maertens, 2017; Ibanez & Blackman, 2015). Authors generally note that tailoring to local realities, leveraging dynamic contexts without including friction, and stable policies and engagement may facilitate impact.

Photo: **Khadija Farah** | NYT | NTB

Study quality and risk of bias

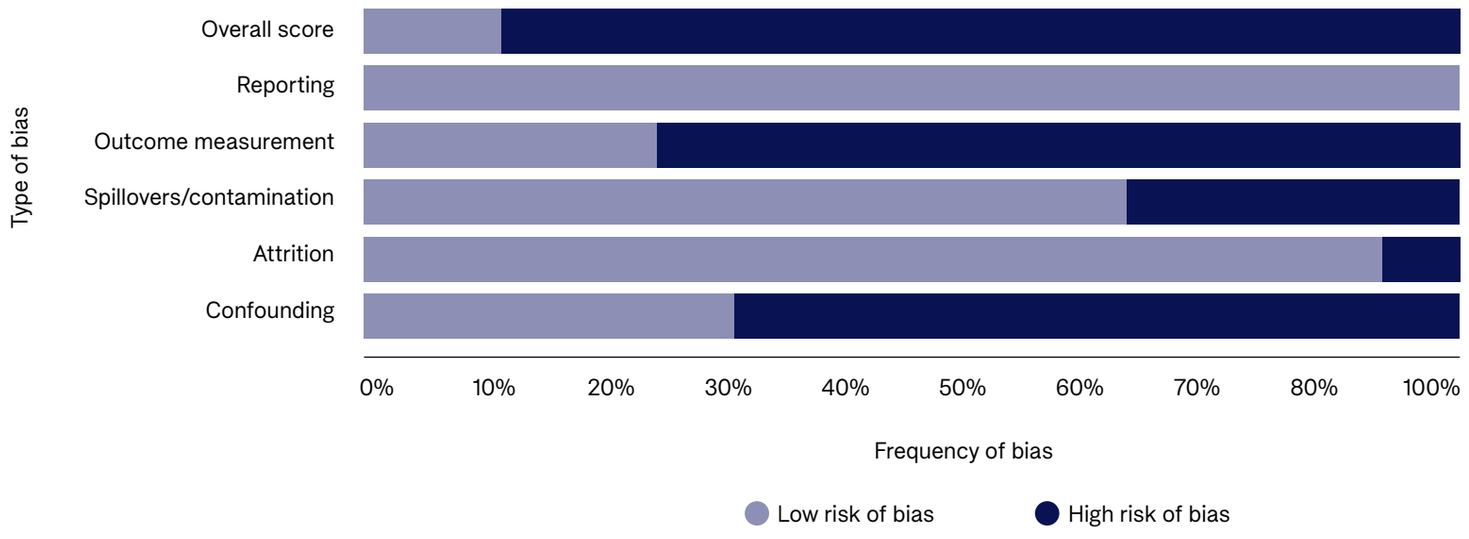
In the 19 studies in the evidence synthesis (25 interventions), we identified a total of 56 reported effects, including the short-term (n = 6)[8] estimates presented in Section V.3 and subgroup (n = 16) effects (Appendix Table 2). Among these 56 effects, seven are rated as low risk of bias with the remainder rated as high risk of bias (Figure 5). The most common sources of bias are related to outcome measurement (41 impact estimates rated as high risk of bias) and confounding (37 impact estimates rated as high risk of bias).

All studies use self-reported outcome measures, which can be subject to bias, especially for outcomes like income and crop production. Many studies do not address this source of bias. If respondents know that the survey is linked to the evaluated intervention, it is likely to influence their answers. These studies do not describe whether the data collection enumerators deliberately concealed information on the intervention to the respondents to limit this source of bias. By using secondary data that is not directly related to the project, data collection is made independent of project exposure and bias is avoided.

Many of these evaluations lack the statistical tests or adjustments to demonstrate that confounding factors were unlikely to bias estimates. Evaluations with statistical matching often do not conduct Rosenbaum bounds tests or analyses to show that the estimates are not sensitive to unobserved or hidden biases (26 out of the 31 impact estimates). Two of the studies using difference-in-difference do not establish parallel trends in outcomes of the treatment and control group before the intervention. These studies also do not control for time trends to account for differences over time in the treatment and control groups (4 out of 12 impact estimates).

FIGURE 5

Risk of bias of studies included in the evidence synthesis



Notes: Effect estimates (n = 56) are rated as low risk of bias or high risk of bias. In instances where bias was unclear or ambiguous, reviewers' judgment was applied based on the criteria outlined in Appendix Table A2.2.

8 Although the inclusion criteria restricted studies to those that considered long-term effects, short-term effects were extracted for studies that reported both. This includes the three estimates discussed in the section "Comparison of the short- and long-term effects" and two estimates in the study by Akoyi & Maertens (2017) which considered the effects of "Utz-Rainforest Alliance-4C" certification in the short-term. These estimates were not discussed above as the authors did not examine the long-term effects of this intervention preventing a comparison of the short- and long-term effects unlike in Kumar and Quisumbing, 2011. However, they are presented in the appendix for reference.

Effects by intervention type

In our analysis, we present the reported effects by intervention type: agricultural education, soil and water conservation, the provision of agricultural inputs, land titling, credit, organic certification, and contract farming. We focus on individual effects and explaining reported effects of specific interventions. In a limited number of instances where effects can be meaningfully combined, we present meta-analysis to develop average treatment effects weighted by study precision.

Agricultural education

We find four studies presenting estimates of the long-term effects of six interventions that have agricultural education components on crop production. Two of these interventions take place in Sub-Saharan Africa and four in Latin America. Three of the interventions included a non-educational component: the provision of agricultural inputs (k = 2) or water conservation. Studies measured maize, sorghum, and grape yield as well as the value of agricultural production per hectare. Reported effects are all positive, but not all statistically significant. The lowest reported effect is a 0.07 standard deviation (95%CI: 0.05 to 0.10,p < 0.001) increase in grape yield, and comes from an evaluation of an agricultural extension and irrigation program in Argentina (Gibbons et al., 2016). The largest reported effect is a 0.47 standard deviation (95%CI: 0.26 to 0.69; p < 0.001) increase in the value of agricultural production per hectare from an intervention in Nicaragua which supported the adoption of sustainable farming technologies and practices (De los Santos-Montero & Bravo-Ureta, 2017). The sustainable practices considered included terracing and integrated pest management. When intervention effects are weighted to account for differences in study precision,[9] the reported average effect is 0.16 standard deviations (95%CI: 0.09 to 0.22,p < 0.001).

We also find four studies presenting the estimated effects of six agricultural education programs on income. Three of these take place in Sub-Saharan Africa with the remaining taking place in Bangladesh. Three interventions, all from the same study, include the provision of credit alongside training in agricultural techniques (Kumar & Quisumbing, 2011). Studies measured total income (k = 5) and revenue. Reported effects on income are generally not statistically significant. The lowest effect comes from a program that provided credit and training on vegetable gardening in Bangladesh (g ˆ = -0.28; 95%CI: -0.59 to 0.03; p = 0.08; Kumar & Quisumbing, 2011). The highest effect comes from a farmer field school program in South Africa (ĝ= 0.23; 95%CI: 0.00 to 0.45; p = 0.05; Baiyegunhi et al., 2018). When interventions effects are weighted to account for study precision, reported effects are not statistically significant (SMD = -0.02; 95%CI: -0.13 to 0.09; p = 0.77).

The work by Gibbons and colleagues (2016) is the only evaluation of an intervention with an agricultural education component that we consider to have a low risk of bias. They find that two agricultural education interventions, which also provide agricultural inputs or water conservation efforts, have positive effects on yield (ĝ=0.07;95%CI:0.05 to 0.10,p<0.001 and ĝ=0.09;95%CI:0.07 to 0.12;p<0.001 respecitively). When these interventions are implemented jointly, they have a slightly larger effect than the effect of each individually (interaction term ĝ=0.03; 95%CI; 0.00 to 0.059; p = 0.05)

Deschamps-Laporte (2013) is the only study to consider the effects of interventions with agricultural education components on both production and income. An agricultural extension service in western Kenya is reported to have no effect on either outcome (g=0.11;95%CI: -0.04 to 0.27;p=0.14 and g= -0.07;95%CI: -0.22 to 0.08;p=0.36 respectively).

There is a strong behavioural component to these interventions which can be affected by dynamic contexts and develop over time. Mentoring and trust building initiatives may increase uptake of supported agricultural techniques, as could formal education and efficient land markets, particularly in the long

9 Inverse variance weighting was used to adjust for study precision. Inverse variance weights studies based on the variance of the calculated effect estimate. Studies which report smaller variance, often indicated through a smaller confidence interval, are given more weight in this analysis.

term (Deschamps-Laporte, 2013; Baiyenguhi et al., 2018). IFAD (2018) conclude that continued, long-term technical facilitation that accounts for local context may increase uptake of new production techniques. Kumar and Quisumbing (2011) highlight the effect of community dynamics on achieving long-term change. The vegetable garden intervention may have improved nutrition outcomes and resulted in the adoption of improved vegetable varieties by non-beneficiaries as it diffused over time. This was possible due to the low entry cost of the vegetable gardening techniques, but not seen in the fish pond interventions the authors also conducted due to the barriers to entry.

Soil and water conservation

We find three studies that look at the effects of soil and water conservation efforts on crop production. These take place in Niger, Ethiopia, and Argentina. One intervention included an educational component, is considered to have a low risk of bias, and is discussed above (Gibbons et al., 2016). Studies measured grape yield, teff yield, and the value of crop production. The World Bank (2009) report that natural resource management project in Niger, largely related to tree planning, reduced the value of crops produced (ĝ=-0.16;95%CI: -0.28 to-0.04;p=0.007). In contrast, a program in Argentina which provided irrigation infrastructure and financial incentives alongside education to support the competitiveness of small- and medium-scale producers is reported to have a positive effect on grape yield (ĝ=0.07;95%CI:0.05 to 0.10;p<0.001; Gibbons et al., 2016). The remaining study finds no statistically significant effect of a program providing educational and in-kind support for conserving and rehabilitating degraded agricultural lands on teff yields (ĝ=0.05; 95%CI: -0.27 to 0.37; p = 0.76; Abebe & Bekele, 2014).

We find four studies reporting the effects of four soil and water conservation efforts in India, Ethiopia, Chile, and Niger on total (k = 3) or crop income. Reported effects are all positive, but only one is statistically significant. The positive effect comes from an intervention in Chile which granted artisanal fisheries exclusive rights to manage, conserve, and exploit certain water areas (ĝ= 0.17;95%C:0.09 to 0.25;p<0.001; Romero & Melo, 2021). The smallest effect comes from the same World Bank (2009) study above that reports that the natural resource management project reduced the value of crops produced in Niger. Authors find the intervention had no effect on total income (ĝ=0.03;95%CI: -0.09 to 0.15,p=0.64), who report no statistically significant effect on teff yield above, also report no statistically significant effect on income (ĝ=0.03;95%CI: -0.31 to 0.37;p=0.86)

Datta (2014) is the only study to measure the effects of these efforts on environmental outcomes and find that the watershed development intervention, mostly constructing reservoirs in India, did not affect cropping intensity in the long-term (ĝ= 0.04;95% CI: -0.21 to 0.29;p=0.75). The intervention was also found to have no effect on crop income (ĝ= 0.23;95%CI: -0.03 to 0.48,p=0.08)

Several of these authors stress the effects that the dynamic contexts in which these interventions function can have on their outcomes (Gibbons et al., 2016; Datta, 2014; World Bank, 2009). They highlight the need for community and long-term engagement to allow for intervention effects to develop (Datta, 2014; World Bank, 2009; Abebe & Bekele, 2014). The World Bank (2009) note that part of the reason for the reduction in production associated with their activity could be that participants shifted to non-agricultural activities, which were also supported by the program. However, they also acknowledge that many of the activities supported by the program did not actually improve production.

Provision of agricultural inputs

Three studies report on the effects of providing agricultural inputs, two of which also included an educational component. Studies took place in Niger, Zambia, and Argentina. All three report positive effects on yield, either sorghum, maize, or grapes. The smallest effect comes from the low risk of bias study by Gibbons and colleagues (2016) discussed above. The largest effect comes from a national program in Zambia which reduced to cost of fertilizer and maize seeds for small scale farmers (ĝ=0.37;95%CI:0.15 to 0.60;p=0.001; Funsani et al., 2016). However, this

intervention had no statistically significant effect on income (ĝ=0.20;95%CI: -0.03 to 0.42;p=0.08). Although this intervention had the largest reported effect, authors posit that this intervention could have benefited from being implemented alongside an educational component, as was done in the other interventions.

Land titling

Three studies report on the effects of land titling interventions in India, Thailand, and Ethiopia. Results are highly variable, largely due to the specific local context in which the interventions were implemented. In Ethiopia, a land registration and certification program is reported to have increased the value added from agriculture (ĝ=0.40; 95%CI:0.22 to 0.59,p<0.001; Melesse & Bulte, 2015). Authors note that the state was the primary driver of land insecurity in Ethiopia. Therefore, formalization was effective at reducing insecurity. In contrast, in India, a land tenancy reform had no effect on wages (ĝ=0.17;95%CI: -0.06 to 0.41;p=0.14; Belsey et al., 2016) and may have actually increased inequality. Upper caste individuals, who generally had more land holdings before the intervention, may have benefited most from this program. The final evaluation, from Thailand, finds that partial property titling had no effect on the primary rice yield (ĝ=-0.13;95%CI: -0.62 to 0.35,p=0.59; Chankrajang, 2015) but increased secondary rice yield (ĝ=0.69;95%CI:0.18 to 1.21;p=0.009) without inducing negative environmental impacts (soil acidification ĝ=0.28;95%CI: -0.23 to 0.79;p=0.29) . This last study, by Chankrajang (2015), has a low risk of bias.

Credit

The provision of agricultural credit is considered by two evaluations evaluating four interventions. The microcredit component of a larger national intervention in Brazil which offered financing, infrastructure, and training, was reported to increase the value of crops produced in all five regions in which it was evaluated (Appendix Table 2; Maia et al., 2016). In contrast, the provision of credit alongside agricultural education on community fish ponds, individual fish ponds, or vegetable gardens, had no long-term effect on total income in Bangladesh (Kumar & Quisumbing, 2011). Although credit provision increased short-term per capita consumption expenditures, the effects abated in the long term.

Organic certification

Two organic certification programs, in Uganda (Akoyi & Maertens, 2017) and Colombia (Ibanez & Blackman, 2015), are reported to decrease production (SMD = -0.45; 95%CI: -0.76 to -0.14,p = 0.005) but have no statistically significant effect on income (MD = -0.13; 95%CI: -0.29 to 0.02; p = 0.08). The complex nature of the certification scheme and delays in payment may have posed a barrier to impact (Akoyi & Maertens, 2017). Inorganic fertilizers may actually be an important component of successful farming in these contexts.

Contract farming

One study evaluated the effects of two contract farming schemes: one in which a company provided resources in return for farmers selling to the company and another in which farmers agreed to an annual fixed price for goods which were regularly collected from the farms (Ruml & Qaim, 2020). Both types of contracts had, on average, no statistically significant effect on palm oil yield (ĝ=0.14; 95%CI: -0.04 to 0.32,p=0.12 and ĝ=0.00;95%CI: -0.17 to 0.18;p=0.96 respectively). However, the contract which provided resources to farmers was reported to increase yield for small (ĝ=0.75;95%CI:0.44 to 1.06;p<0.001) and medium (ĝ=1.36;95%CI:1.05 to 1.67;p<0.001) sized growers.

Comparison of effects over time

We attempt to compare the outcomes evaluated for the same intervention over time to establish commonalities in trajectories of change. Within-study comparisons, such as this, can help to mitigate the confounding factors that may bias comparisons of interventions across studies. For example, the comparison of two studies evaluating a similar intervention, one in the short-term and another in the long-term, may be confounded by differences in the contexts in which the two interventions were implemented (Fenton Villar & Waddington, 2019; Waddington et al., 2022). Comparing the outcomes of the same intervention in the short- and long-term minimises the risk of bias associated with these issues.

We identify only two studies, evaluating four interventions, reporting both the short- and long-term effects of agricultural interventions. An additional study was found that looks at long-term outcomes at different time points. Across these three studies, we see that reported effects follow highly variable patterns of change (Table 7). However, these reported estimates have a high risk of bias so findings should be interpreted with caution.

Kumar and Quisumbing (2011) study the effects of three interventions combining access to credit and

TABLE 7

Effects of agricultural interventions across time periods.

Study	Intervention	Outcome	Effect in first measurement (95% confidence interval) time period	Effect in later measurement (95% confidence interval) time period
Kumar and Quisumbing (2011)	Access to credit, agricultural trainings and leases to fishponds	Income	0.36 (0.06 to 0.65) 3 years	-0.14 (-0.44 to 0.16) 12 years
Kumar and Quisumbing (2011)	Access to credit and agricultural trainings to households with ownership of individual fishponds	Income	-0.34 (-0.65 to -0.04) 6 years	-0.01 (-0.31 to 0.29) 12 years
Kumar and Quisumbing (2011)	Access to credit and agricultural trainings for the adoption of improved vegetable varieties	Income	0.22 (-0.08 to 0.53) 2 years	0.28 (-0.59 to 0.03) 12 years
Abdoulaye and Sanders (2013)	Provision of sorghum technology packages and recommendations	Yield	0.56 (0.30 to 0.82) 9 years	0.36 (0.10 to 0.62) 10 years 0.36 (0.10 to 0.62) 11 years
Funsani and colleagues (2016)	Reduced cost fertilizer and maize seeds	Yield	0.28 (0.05 to 0.50) 11 years	0.36 (0.14 to 0.59) 12 years

agricultural education in Bangladesh. The effect of an intervention providing access to credit, training, and leases to fishponds is positive and statistically significant effects on income three years after the start of the intervention. However, these short-term effects dissipate and do not translate into significant, positive long-term effects. Conversely, effect estimates related to the intervention providing credit and agricultural training to households with ownership of individual fishponds indicate a negative and statistically significant effect after six years. Again, the effect dissipates and becomes very small and not statistically significant in the long-term. Finally, the reported effects of the intervention providing access to credit and agricultural training for the adoption of improved vegetable varieties are positive but not statistically significant after two years. In the long-term, the reported effect estimates turn negative but remained statistically insignificant. These findings suggest that short-term effects do not always translate into long-term impact.

Abdoulaye and Sanders (2013) is the second study with short- and long-term estimates. This study analysed the effects of an intervention in Niger which provided improved sorghum technology packages, inorganic fertilizers, fungicide and agronomic recommendations. Nine years after intervention inception, the estimated effect on sorghum yields was positive and statistically significant but magnitude (ĝ=0.56; 95%CI: 0.30 to 0.82,p<0.001, short-term effect). In the subsequent two years, the magnitude of the effect declined (ĝ=0.36; 95%CI: 0.10 to 0.62,p=0.006 , after 10 years and ĝ=0.36; 95%CI: 0.10 to 0.62,p=0.006 , after 11 years, long-term effects). Authors do not provide an explanation why the effects dissipated.

Funsani and colleagues (2016) do not report estimates in the short- and long-term but instead report several long-term estimates (more than 10 years after intervention inception). This study reported a modest and positive statistically significant effect of a national program in Zambia which reduced to cost of fertilizer and maize seeds for small scale farmers. They found a statistically significant positive effect on maize yields in 2013-2014 (11 years after intervention inception, ĝ=0.28; 95%CI: 0.05 to 0.50,p=0.01). In 2014-2015, the effect was slightly greater and still statistically significant (ĝ=0.36; 95%CI: 0.14 to 0.59,p=0.002).

Adverse events

Using both qualitative and quantitative approaches, we investigated if evaluated interventions have any adverse effects on our primary outcomes of interest or other relevant outcomes. However, not all studies measure adverse effects, and adverse effects are not always reported when identified. In addition, we did not conduct a comprehensive search to identify studies that were not included in the evidence synthesis but may have reported relevant adverse events. Adverse events are likely to be context specific, so implementers should consider if the concerns raised here are applicable to their specific situation.

The most apparent adverse event is related to insufficient fertilizer use. The two organic certification interventions were reported to reduce crop production but have no statistically significant effect on income (Akoyi & Maertens, 2017; Ibanez & Blackman, 2015). In Uganda, this was attributed to a decrease in the use of inorganic fertilisers for other crops intercropped with coffee (Akoyi & Maertens, 2017). Family labour was diverted towards cash crops, there was decreased coffee labour productivity, and reduced availability of other farm products. Although not stated by Ibanez and Blackman (2015), it seems likely that the certification program in Colombia would have resulted in similar changes in fertilizer use that could explain the decrease in production.

Another concern related to the potential for increasing inequality specific to a land titling program in India. Besley and colleagues (2015) conclude that the land titling program may have increased long-term inequality because individuals who did not belong to the upper castes generally had lower levels of pre-formalisation land holdings and may have been left out of benefitting from the program.

Barriers and facilitators to long-term success

Given the inconsistent impacts of agriculture interventions on income, crop production, and the environment, we focus our qualitative analysis on the barriers and facilitators to program success presented by study authors. While we discuss barriers and facilitators to impact for individual intervention types above, we present crosscutting themes in this section. Broadly, authors suggest three areas that could lead to improved long-term impact: tailoring to local realities, leveraging dynamic contexts without inducing friction, and providing stable policies and engagement.

Most authors acknowledge that the success or failure of the specific interventions in their evaluations does not necessarily imply that the intervention will work similarly in other contexts or if implemented again in the future. Nonetheless, the authors highlight the need to be aware of local realities at the design stage. The dynamic interaction between the targeted intervention and other ongoing activities may determine the direction of the effects. If the two work in tandem, benefits may be realized. But if the various actors counteract, outcomes may not be achieved. Continued engagement, including long-term stability, resourcing, and training, may be needed to ensure that outcomes are maintained.

Tailoring to local realities

Causal pathways can be moderated by local context, such as the environment, family structure, and decisions about how resources are shared within a household (IFAD, 2018; Maia et al., 2016; Romero & Melo, 2021; Kumar & Quisumbing, 2011). Accounting for these factors may facilitate impact. Tailoring allows interventions to account for farmer's decision-making processes and leverage these processes to achieve programme objectives (World Bank, 2016; IFAD, 2018). For example, developing and updating a focused curriculum and appreciating the agro–ecological contexts was believed to facilitate the success of an agricultural education program in Tanzania (IFAD, 2018). Going beyond technical facilitation to include mentoring and monitoring may have supported long-term adoption of recommended innovations.

Failure to internalise local context can result in interventions having negative, unintended consequences. For example, land reforms in India increased inequality in the long-term because benefits were largely restricted to upper castes (Besley et al., 2015). The reforms had no effect on wages. Melesse and Bulte (2015) note that the land formalization program in Ethiopia may not work well in contexts with strong customary land rights as this could result in the development of two parallel land tenure systems. The authors propose that such reforms, when enacted without sufficient institutional support or with a disregard for competing interests and local norms, may impede the "assurance" gains from such formalization. A Fairtrade-Organic certification in Uganda inadvertently lowered coffee production, possibly because inorganic fertilisers and pesticides were actually needed to maintain cultivation (Akoyi & Maertens, 2017). Other crop outputs suffered due to the diversion to organic production methods; therefore, future policy needs to account region-specific farm and intercropping patterns. Datta (2014) stresses the importance of engagement in the community and conclude that poorly conceived policies, detached from ground realities, can increase cost, vulnerability, and inequity in the long term.

Leveraging dynamic contexts without inducing friction

Due to the dynamic context in which agricultural interventions function, there is significant opportunity to leverage synergies between interventions and various actors. However, this same dynamic context can prove to be a barrier to impact when not properly addressed. A holistic understanding of possible countervailing forces that may hold back the expected benefits of interventions, especially when changing established production methods, may be needed for effective program design. Leveraging dynamic contexts without inducing friction can ease access and facilitate impact.

When it works, the simultaneous implementation of several interventions or a collaboration between

various stakeholders can facilitate impact (Gibbons et al., 2016). Evaluators of a watershed development initiative in India and agricultural education programs in Argentina stress the importance of building complementarities in program design, particularly when intervention seek to change production practices (Gibbons et al., 2016; Datta, 2014). In Argentina, the joint implementation of two agricultural interventions had a slightly larger impact than the implementation of each intervention individually, implying that synergies between programs may facilitate impact. Both Funsani and colleagues (2016) and Deschamps-Laporte (2013) indicate that accompanying seed and fertiliser provision programs with technical training on their use could facilitate impact in Zambia and Kenya.

However, dynamic contexts can induce friction (Akoyi & Maertens, 2017). For example, the private sustainability standards introduced in Uganda in 2006 intended to empower farmers through the power of Fairtrade (Akoyi & Maertens, 2017). The program was plagued by organic fertiliser shortages and delays in payments of certification checks. The payment structure, which was produce-pegged, worked against farmers. The scheme resulted in the diversion of family labour towards cash crops and reduced the availability of other farm products. It decreased coffee labour productivity and resulted in no meaningful change in income.

Stable policies and intervention engagement

Continuity or extended engagement can facilitate impact; erratic engagement or implementation can pose a barrier to impact. For instance, the long-term returns for irrigation projects may take time to develop, so continuing support may be needed after the formal end of programs (World Bank, 2009; Abebe, 2014). In Ethiopia, the state was the dominant driver of land tenure insecurity. So, when land rights were formalized, they became more secure, and production increased (Melesse & Bulte, 2015). Conversely, in India, changes in policies related to groundwater use were attributed to a reduction in the impact of a watershed development program (Datta, 2014). Generally, authors report that, trust, continued engagement with beneficiaries, awareness activities, and stable partnerships, including reliable long-term financial support, may facilitate outcomes (IFAD, 2018; Baiyegunhi et al., 2018; World Bank, 2009).

A common problem with large-scale interventions is what happens after the program is «over». The World Bank (2016) underscores the need for maintaining programme contact after it ends to sustain effects. Short-run early adopter effects may not be sustained long-term unless the intervention or agricultural approach is spread throughout the intervention community. The diffusion of intervention effects over time within a community can facilitate long-term impacts (Kumar & Quisumbing, 2011 and Besley et al., 2015). This necessitates understanding the role of mediating factors in sustaining the impacts. For example, Kumar and Quisumbing (2011) find that improved vegetable varieties spread more quickly through communities than fishponds because of their lower cost. This may explain some of the time-dependent effects of the intervention on nutritional indicators. Spillovers and variation in intrahousehold allocation also likely played a role.

Cost evidence

Little cost evidence is available from the included evaluations on the long-term effects of agricultural interventions (n = 3). The World Bank (2009) estimate an internal rate of return of 45 per cent from investing in a tree plantation program. There is a lower bound of 28 per cent if conservative estimates of fodder value are used. However, this estimate seems to be based on outcome measurements rather than impact measurements, which are reported not to be statistically significant or even negative. The land titling program in Ethiopia is estimated to have a one-time social cost of about 1 USD per plot and result in an annual gain of 75.40 USD in productivity (Melesse & Bulte, 2015). Although Funsani (2016) do not conduct a formal cost analysis, they indicate that the cost of the fertiliser subsidy program evaluated is high (40.2% of the government's annual agricultural budget) and may not be justified when considering the small effects of the program on yield and income. These results need to be considered cautiously as they only represent the evidence from a limited number of interventions and may not be generalisable. Future research should consider the use of more comparable cost metrics to allow for the identification of the most cost-effective interventions.

5

Conclusion

Although we find a larger evidence base than initially expected (n = 78), the vast majority (96%) of evaluations of food systems interventions consider outcomes within 10 years of intervention initiation. On average, outcomes are measures 2.3 years after interventions start, but 53 per cent of evaluations consider a time period of a year or less. Because few studies consider each specific intervention-outcome combination, the development of definitive conclusions of the long-term effects of food systems interventions challenging. Only five studies report cost evidence regarding the long-term effects of food systems interventions. As a result, there is considerable room for the evidence base to develop.

Photo: **Zohra Bensemra** | **Reuters** | **NTB**

There may be some variation in the types of interventions evaluated in the long-term. Four per cent of evaluations overall consider long-term outcomes, but eight per cent of food production interventions, one per cent of behaviour change communication interventions, and two per cent of evaluations of food processing and packaging interventions take a long-term perspective. The apparent focus on Sub-Saharan Africa (n = 31) is likely due to the long-standing interest in improving food systems in the chronically food insecure region. However, the inconsistent reporting of funding agencies makes it difficult to know which organisations are most interested in long-term evaluations. Government agencies seem to be the most common funders of long-term evaluations (n = 36). Because the decision to evaluate is often taken long after intervention initiation, many long-term evaluations rely on endline only data and use statistical matching approaches.

The most commonly evaluated intervention-outcome combinations are production systems interventions and agricultural or economic outcomes. Agricultural interventions which are subject to long-term evaluation are generally based on the theory that reducing environmental uncertainty and building resilience in the long-term will increase farmers' investment, access to land and markets, and engagement in global value chains. This is thought to improve measures of income, production, climate, and the environment. Despite the concentration of evidence, the variation in the types of agricultural interventions and outcomes considered results in few evaluations of each intervention-outcome combination and challenges in making generalizable conclusions about their effectiveness.

Nonetheless, some policy relevant findings are beginning to emerge and warrant further research. Individual, low risk of bias effect estimates can be referenced to understand effects in specific contexts (Gibbons et al., 2016; Chankrajang, 2015). Overall, the reported effects of agricultural interventions on income, crop production, and the environment are small or statistically insignificant in the long-term. However, results are highly variable across intervention types and regions. There is suggestive evidence from three studies that effects of an intervention can vary over time within the same context (Kumar & Quisumbing, 2011; Abdoulaye & Sanders, 2013; Funsani et al., 2016). So, short-term effect measurement should not be substituted for long-term evaluations. The effects of agricultural interventions on income and crop production seem to be facilitated by tailoring to local realities, leveraging dynamic contexts without inducing friction, and stable policies and engagement. Adverse events, specifically around fertiliser use and inequality, have been reported.

Few studies consider more than one of our included outcomes, and those that do report no meaningful link. There is weak evidence from two studies that income or production does not have to come at the cost of negative environmental effects. In Thailand, a partial land titling program had no effect on the share of land with acidic soil and slightly increased the secondary rice harvest (Chankrajang, 2015). There was no effect on the primary harvest. A watershed development project in India increased income without inducing a corresponding increase in cropping intensity, which often leads to land degradation (Datta et al., 2014). The studies that report on production and income generally found null effects for both (Funsani et al., 2016; Abebe & Bekele, 2014; Akoyi & Maertens, 2018; Deschamps-Laporte, 2013; World Bank, 2009; Ibanez & Blackman, 2015).

Limitations in available studies

The overall evidence base on the long-term effects of food systems interventions is limited. Due to the modest number of included studies and variation in interventions considered, there is insufficient evidence to reach conclusions about the likely effects of agricultural interventions on income, crop production, climate, and the environment in the long-term. The limited availability of evidence on the effects of food systems interventions on climate change may be due to methodological considerations and inclusion criteria. Evaluations of the effects of the food systems itself, as opposed to food systems interventions, would not be included in the Food Systems and Nutrition Evidence Gap Map, nor would evaluations of other types of interventions, such as efforts to reduce fossil fuel emissions by factories or cars.

The evidence that is available has meaningful risks of bias. However, many of the sources of bias of included studies could be readily addressed through additional statistical analyses, which would not necessitate completely changing the evaluation approach. Studies regularly lack sufficient reporting of common statistical and placebo tests and necessary statistical adjustments that might add confidence about the specific application of a study design. For example, studies using statistical matching could reduce risk of bias through the use of Rosenbaum bounds tests to determine if there is significant unobserved confounding. Studies using difference-in-difference methods could have controlled for time trends to account for differences in the treatment and control group over time. This adjustment would have increased the confidence in the assumption that treatment and control groups would have experience parallel trends in their outcomes in the absence of the intervention. Although self-reported data may be the only option for evaluators in some cases, blinding participants to the purpose of data collection or using secondary data that is not related to the program being evaluated can reduce biases in outcome measurement.

Another key limitation is that most studies in the evidence synthesis (n = 13) present the average treatment effect on the treated. In this context, the average treatment effect on the treated generally reflects the impact of the intervention on those who chose to adopt the agricultural practice supported by the intervention. When uptake is low, this does not always reflect the population-level average treatment effect (reported in 4 studies). The average treatment effect on the treated omits possible effects on people who chose not to adopt the intervention. However, a land certification program in Ethiopia reports similar production effects calculated through the average treatment effect on the treated and intention-to-treat analysis (Melesse & Bulte, 2015).

The lack of mixed methods and cost evidence research generally, and in our evidence synthesis in particular, reflects a need to gather qualitative information about structural impediments to change, details on cost structures, potential complementarities with other programs, and sustainability.

Implications

For policy-makers

A modestly sized but sparsely distributed body of research exists on the effects of food systems interventions on food security, nutrition, climate, and the environment 10 years, or more, after intervention initiation. The identified evaluations can be referenced to understand facilitators and barriers to impact in specific, local contexts.

- Because the food systems interventions evaluated in the long-term are so variable, policy makers may find **insufficient evidence on specific interventions or outcomes** of interest. The evidence becomes even more restricted when considering certain regions or vulnerable groups.

- Due to the highly variable nature of the evidence base, policy-makers should **avoid concluding that agricultural interventions do not work in the long-term** and maintain a state of equipoise. Intervention effects ranged widely and several interventions were found to be successful in specific local contexts.

The absence of cost evidence means that, even if interventions are found to be successful, it is not possible to determine if an intervention is worth the investment.

The **relationship between effects over time is variable,** so short-term evaluations cannot be used to make conclusions about long-term change.

Policy-makers and researchers may wish to partner early in the implementation of food systems interventions to **design interventions for evaluability** in the long-term. Often, the decision to evaluate is made years after intervention initiation. This can require the adoption of evaluation approaches which are less rigorous and less useful for decision making.

- **Incentive structures may need to be updated** to support long-term evaluations so that policy-makers have the information they need to make evidence-based decisions.

For program funders and implementers

Project design should consider the possibility of adverse events from agricultural interventions, such as inappropriate fertilsier use or increased inequality.

Design should anticipate project close-out and what happens when a program is "over" to limit short-term positive impacts abating with time.

The evidence indicates that interventions with *s***table policies and engagement and tailoring to local context while leveraging dynamic contexts** may facilitate impact. But, this finding needs to be validated through mixed-methods research.

Researchers and research funders

More mixed-methods work is needed to support the funding and design of contextually relevant interventions which have impact in the long-term. Because most food systems interventions require behaviour change to have long-term impact, mixed-methods research is needed to understand participant lived-experiences which underly mechanisms of change. Understanding of the local context and the mechanisms at play should lead to a more complete and precise theory of change which, in turn, can increase the likelihood of positive impacts and minimise the advent of adverse effects.

Although the effects of the food system on climate and the environment are now well established, additional research on **interventions within the food systems to improve climate and environmental outcomes** is needed.

Collecting data on indicators across the causal chain (environment, crop production, and income) at baseline and endline would provide valuable insights into how these interventions work in the long-term.

Cost evidence is needed to know if interventions represent an efficient use of limited resources.

Statistical tests or adjustments to address confounding and a **priori** research designs could strengthen the evidence base.

Strengths, limitations & future directions

Due to the rapid nature of this work, results should be interpreted more cautiously than those of a systematic review. Relying on the Food Systems and Nutrition Evidence Gap Map may result in some relevant studies being omitted from this evidence assessment. Although the evidence gap map is rigorous and broad, this broad nature means that some studies related to the long-term outcomes may have been missed. In particular, the Food Systems and Nutrition Evidence Gap Map only selects one study from any set of linked studies. In the original map (published 2021), the most recent study was selected. We expect this to be the one considering the longest time frame, but that may not always be the case. During the update period, we identified an additional 20 linked studies. In these cases, the originally included article was retained in the map.

Although we found a meaningful number of studies, only a small proportion of these could be synthesised. The studies presented are highly variable, making it difficult to draw broad, generalisable conclusions. Studies included in the evidence synthesis did not include qualitative research, so our barriers and facilitators analysis was largely restricted to the conclusions of the authors. Future work should expand the evidence base, allowing for the development of stronger conclusions about the effects of individual intervention types, and focusing on specific vulnerable groups. More mixed-methods research is needed to understand participant perceptions of these interventions as lived experiences are likely to affect the long-term impacts of food systems interventions. In addition, more cost evidence is needed to understand if these interventions are worth the investment.

This research project could be expanded through the additional synthesis of the studies that were not included in the evidence synthesis. In addition, a more detailed consideration of the methods used by the 78 long-term studies and the feasibility of applying these methods to other evaluations would be valuable in facilitating the future long-term evaluations ●

References

Abdoulaye, A. & Sanders, J. H. (2013). A Matching approach to analyze the impact of new agricultural technologies: productivity and technical efficiency in Niger.

Abebe, Y., & Bekele, A. (2014). The impact of soil and water conservation program on the income and productivity of farm households in Adama District, Ethiopia. Science, Technology and Arts Research Journal, 3(3), 198-203.

Akoyi, K. T., & Maertens, M. (2018). Walk the talk: private sustainability standards in the Ugandan coffee sector. The Journal of Development Studies, 54(10), 1792-1818.

Bailey, D. H., Duncan, G. J., Cunha, F., Foorman, B. R., & Yeager, D. S. (2020). Persistence and fade-out of educational-intervention effects: Mechanisms and potential solutions. Psychological Science in the Public Interest, 21(2), 55-97.

Baiyegunhi, L. J. S., Majokweni, Z. P., & Ferrer, S. R. D. (2019). Impact of outsourced agricultural extension program on smallholder farmers' net farm income in Msinga, KwaZulu-Natal, South Africa. Technology in Society, 57, 1-7.

Bari, M. A., Khan, G. D., He, B., & Yoshida, Y. (2022). The impact of unconditional cash and food assistance on contraceptive expenditure of rural households in Coastal Bangladesh: Evidence from fuzzy RDD. Plos one, 17(1), e0262031.

Berretta, M., Kupfer, M., Lane, C., & Shisler, S. (2022). Rapid Evidence Assessment on Women's Empowerment Interventions within the Food System: a meta-analysis. Available from: https://doi.org/10.21203/rs.3.rs-1754233/v1

Besley, T., Leight, J., Pande, R., & Rao, V. (2015). Long-run impacts of land regulation: Evidence from tenancy reform in India. Journal of Development Economics, 118, 72-87.

Blattman, C., Fiala, N., & Martinez, S. (2014). Generating skilled self-employment in developing countries: Experimental evidence from Uganda. The Quarterly Journal of Economics, 129(2), 697-752.

Borenstein, M, Hedges, LV, Higgins, JPT and Rothstein, HR, 2009a. Introduction to Meta-Analysis. Chichester, U.K: John Wiley & Sons, Ltd.

Bouguen, A., Huang, Y., Kremer, M., & Miguel, E. (2019). Using randomized controlled trials to estimate long-run impacts in development economics. Annual Review of Economics, 11, 523-561.

Brauw, Alan de. (2019). Endline Report On Impacts Associated With The Bangladesh Agricultural Value Chains Project. United States Agency for International Development (USAID) Working Paper Series. United States Agency for International Development (USAID) Working Paper Series. USAID. Available from: https://pdf.usaid.gov/pdf_docs/PA00TRHM.pdf

Campbell Collaboration, 2017. The Campbell Collaboration Online Library, Viewed 10 October 2017, https://campbellcollaboration.org/library.html

Carter, B., & Kelly, L. (2021). Social Inequalities and Famine and Severe Food Insecurity Risk.

Chang, H. H., & Mishra, A. K. (2012). Chemical usage in production agriculture: Do crop insurance and off-farm work play a part?. Journal of environmental management, 105, 76-82.,

Chankrajang, T. (2015). Partial land rights and agricultural outcomes: evidence from Thailand. Land Economics, 91(1), 126-148.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Erlbaum.

Datta, N. (2015). Evaluating impacts of watershed development program on agricultural productivity, income, and livelihood in bhalki watershed of Bardhaman District, West Bengal. World Development, 66, 443-456.

David, T. (2023, February 10). Long-run outcomes: Measuring program effectiveness over time (S. Kopper, Ed.). J-PAL. https://www.povertyactionlab.org/blog/2-10-23/long-run-outcomes-measuring-program-effectiveness-over-time

De los Santos-Montero, L. A., & Bravo-Ureta, B. E. (2017, November). Productivity effects and natural resource management: econometric evidence from POSAF-II in Nicaragua. In Natural Resources Forum (Vol. 41, No. 4, pp. 220-233). Oxford, UK: Blackwell Publishing Ltd.

Deep Knowledge Analytics. (2022). Food Security Index by Overall Score. Global Food Security Q2 2022. Available from: https://www.dka.global/global-food-security-q2-2022

Dercon, S., Gilligan, D. O., Hoddinott, J., & Woldehanna, T. (2009). The impact of agricultural extension and roads on poverty and consumption growth in fifteen Ethiopian villages. American Journal of Agricultural Economics, 91(4), 1007-1021.

Deschamps-Laporte, J. P. (2013). The impact of extension services on farming households in Western Kenya: A propensity score approach (No. 5/2013). Working Paper.

Ellis, P. D. (2010). The essential guide to effect sizes: Statistical power, meta-analysis, and the interpretation of research results. Cambridge university press.

ERIE. 2018. Guide for planning long-term impact evaluations (LTIEs). Available from: https://pdf.usaid.gov/pdf_docs/PA00T9HJ.pdf

European Commission. (2022). Supporting policy with scientific evidence. Available from: https://knowledge4policy.ec.europa.eu/online-resource/food-systems-nutrition-evidence-gap-map_en

Expanding the Reach of Impact Evaluation. (2018). Guide for planning long-term impact evaluations (LTIEs). Available from: https://pdf.usaid.gov/pdf_docs/PA00T9HJ.pdf

Fenton Villar, P. and Waddington, H. (2019) 'Within Study Comparisons and Risk of Bias in International Development: Systematic Review and Critical Appraisal'. Campbell Systematic Reviews. 15(1-2). p. e1027.

Filipski, M., Taylor, J. E., Abegaz, G. A., Ferede, T., Taffesse, A. S., & Diao, X. (2017). General equilibrium impact assessment of the Productive Safety Net Program in Ethiopia.

Food and Agriculture Organization (2008). An Introduction to the Basic Concepts of Food Security. Available from: https://www.fao.org/3/al936e/al936e00.pdf

Food and Agriculture Organization (2022). Tracking progress on food and agriculture-related SDG indicators 2022. Rome. Available from: https://doi.org/10.4060/cc1403en

Forss, K., Lindkvist, I., & McGillivray, M. (Eds.). (2020). Long term perspectives in evaluation: Increasing relevance and utility. Routledge.

Francesconi, G. N., & Ruben, R. (2012). The hidden impact of cooperative membership on quality management: a case study from the dairy belt of Addis Ababa. Journal of Entrepreneurial and Organizational diversity, 1(1), 85-103.

Funsani, W., Rickaille, M., Zhu, J., Tian, X., Chibomba, V., Avea, A. D., & Balezentis, T. (2016). FARMER INPUT SUPPORT PROGRAMME AND HOUSEHOLD INCOME: LESSONS FROM ZAMBIA'S SOUTHERN PROVINCE. Transformations in Business & Economics, 15.

Gerber, N., Kalkuhl, M., & Torero, M. (2016). Food and nutrition security: short and long term drivers–definition and indicators (No. 2201-2019-1518).

Gibbons, M. A., Maffioli, A., & Rossi, M. A. (2016). Money for Wine? Complementarities in the provision of private and public goods to wine producers (No. IDB-WP-689). IDB Working Paper Series.

Glewwe, P., & Todd, P. (2022). Impact Evaluation in International Development: Theory, Methods, and Practice. World Bank Publications.

Global Panel on Agriculture and Food Systems for Nutrition. (2020). Future food systems: for people, our planet, and prosperity.

Green, C. S. (2020). Interventions to do real-world good: Generalization and persistence. Psychological Science in the Public Interest, 21(2), 43-49.

Hammaker, J., Anda, D., Kozakiewicz, T., Bachina, V., Shisler, S., & Lane, C. (2022). Systematic review on Fiscal Policy Interventions in Nutrition. Frontiers in Nutrition, 2823. Available from: https://www.frontiersin.org/articles/10.3389/fnut.2022.967494/full?&utm_source=Email_to_authors_&utm_medium=Email&utm_content=T1_11.5e1_author&utm_campaign=Email_publication&field=&journalName=Frontiers_in_Nutrition&id=967494

Hecht, L., Bernard, M., Müller, N., Fabisch, G., Harder, A., & Luck-Sikorski, C. (2019). Efficacy of DiaLife, an education program for relatives of adult patients with diabetes–study protocol of a cluster randomized controlled trial. Trials, 20(1), 1-11.

High Level Panel of Experts on Food Security and Nutrition (HLPE). (2017). Nutrition and Food Systems: A Report by the High Level Panel of Experts on Food Security and Nutrition Report (HLPE)

Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. Child development perspectives, 2(3), 172-177.

Horowitz, J. K., & Lichtenberg, E. (1993). Insurance, moral hazard, and chemical use in agriculture. American journal of agricultural economics, 75(4), 926-935.

Ibanez, M., & Blackman, A. (2015). Environmental and economic impacts of growing certified organic coffee in Colombia. Environment for Development (February 2015).

International Fund for Agricultural Development (2018). IFAD IMPACT ASSESSMENT-Agricultural Sector Development Programme–Livestock (ASDP-L) and the Agriculture Service Support Programme (ASSP): Tanzania. Garbero, A., Chichaibelu, BB.

Kay, C. (2006). Rural poverty and development strategies in Latin America. Journal of agrarian change, 6(4), 455-508.

Kinzer, M. (2022). Land tenure formalization and regularization activities in low-income countries. Rapid evidence assessment brief. New Delhi: International Initiative for Impact Evaluation (3ie). Available from: https://3ieimpact.org/sites/default/files/2022-06/Land-tenure-formalization-REA-brief.pdf

Kumar, N., & Quisumbing, A. R. (2011). Access, adoption, and diffusion: understanding the long-term impacts of improved vegetable and fish technologies in Bangladesh. Journal of Development Effectiveness, 3(2), 193-219.

Laborde Debucquet, D., Parent, M., & Piñeiro, V. (2021). Prioritization of types of investments: Operational tools for MCC agricultural investments (Vol. 4). Intl Food Policy Res Inst.

Lane, Charlotte; Tree, Veronika; Storghaug, Ingunn; Engelbert, Mark. (2022). Food Systems and Nutrition Evidence Gap Map: Update #3 New Delhi: International Initiative for Impact Evaluation (3ie). Available from: https://www.3ieimpact.org/our-work/nutrition-and-food-security/food-systems-and-nutrition

Li, Y., & Wang, Z. (2022). Analysis on the effect of farmer income of policy-based agricultural insurance. Acta Agriculturae Scandinavica, Section B—Soil & Plant Science, 72(1), 386-400.

Lv, C. (2020). Empirical Analysis on the Effect of Agricultural Insurance on Production——Based on panel data of 31 provinces and cities in China from 2008 to 2018. In E3S Web of Conferences (Vol. 214, p. 01013). EDP Sciences.

Maia, G. A., Eusebio, G. S., & Silveira, R. L. (2016). Impact of microcredit on small-farm agricultural production: evidence from Brazil (No. 333-2016-14332).

Manley, J., Fernald, L., & Gertler, P. (2015). Wealthy, healthy and wise: does money compensate for being born into difficult conditions?. Applied Economics Letters, 22(2), 121-126.

Melesse, M. B., & Bulte, E. (2015). Does land registration and certification boost farm productivity? Evidence from Ethiopia. Agricultural Economics, 46(6), 757-768.

Moore, N, C Lane, I Storhaug, A Franich, H Rolker, J Furgeson, T Sparling, and B Snilstveit. 2021. "Food Systems and Nutrition Evidence Gap Map." (2021). Available from: https://developmentevidence.3ieimpact.org/egm/food-systems-and-nutrition-evidence-gap-map.

Moore, N, Lane, C, Engelbert, M, Rolker, H, Storhaug, I, Sparling, T and Snilstveit, B. (2020). Protocol: The effects of food systems interventions on food security and nutrition outcomes in low- and middle- income countries: an evidence gap map. New Delhi: International Initiative for Impact Evaluation (3ie).

Ngwira, A., Thierfelder, C., & Lambert, D. (2013). Conservation agriculture systems for Malawian smallholder farmers: Long-term effects on crop productivity, profitability and soil quality. Renewable Agriculture and Food Systems, 28(4), 350-363. doi:10.1017/S1742170512000257

Niu, Z., Yi, F., & Chen, C. (2022). Agricultural Insurance and Agricultural Fertilizer Non-Point Source Pollution: Evidence from China's Policy-Based Agricultural Insurance Pilot. Sustainability, 14(5), 2800.

Njuki, J., Eissler, S., Malapit, H., Meinzen-Dick, R., Bryan, E., & Quisumbing, A. (2021). A review of evidence on gender equality, women's empowerment, and food systems. Global Food Security, 33, 100622.

Norwegian Ministry of Foreign Affairs. (2022). Kraftsamlinig mot svolt- ein politikk for auka sjolvforsyning. Available from: https://www.regjeringen.no/no/dokumenter/matsikkerheit_strategi/id2948780/

Oya, C., Schaefer, F., Skalidou, D., McCosker, C., & Langer, L. (2017). Effects of certification schemes for agricultural production on socio-economic outcomes in low-and middle-income countries: a systematic review. Campbell Systematic Reviews, 13(1), 1-346.

Piza, C., Cravo, T. A., Taylor, L., Gonzalez, L., Musse, I., Furtado, I., ... & Abdelnour, S. (2016). The impact of business support services for small and medium enterprises on firm performance in low-and middle-income countries: A systematic review. Campbell Systematic Reviews, 12(1), 1-167.

Pollard, A. and I. Lindkvist (2020). 10 Why Take a Long View? Long Term Perspectives in Evaluation: Increasing Relevance and Utility. K. Forss, I. Lindkvist and M. McGillivray. New York, Routledge

Prado, E. L., Sebayang, S. K., Apriatni, M., Adawiyah, S. R., Hidayati, N., Islamiyah, A., ... & Shankar, A. H. (2017). Maternal multiple micronutrient supplementation and other biomedical and socioenvironmental influences on children's cognition at age 9–12 years in Indonesia: follow-up of the SUMMIT randomised trial. The Lancet Global Health, 5(2), e217-e228.

Romero, P., & Melo, O. (2021). Can a Territorial Use Right for Fisheries management make a difference for fishing communities?. Marine Policy, 124, 104359.

Schneider, B., & Bradford, L. (2020). What we are learning about fade-out of intervention effects: A commentary. Psychological Science in the Public Interest, 21(2), 50-54.

Snilstveit, B., Stevenson, J., Menon, R., Phillips, D., Gallagher, E., Geleen, M., ... & Jimenez, E. (2016). The impact of education programmes on learning and school participation in low-and middle-income countries.

Stern, E., Stame, N., Mayne, J., Forss, K., Davies, R., & Befani, B. (2012). Broadening the range of designs and methods for impact evaluations.

Storhaug, I. G., Lane, C., Moore, N., Engelbert, M., Sparling, T. M., & Snilstveit, B. (2022). Making the most of existing research: an evidence gap map of the effects of food systems interventions in low-income and middle-income countries. BMJ open, 12(6), e055062.

The Organization for Economic Cooperation and Development (2021). Food systems and the triple challenge. Agriculture Policy Brief. Available from: https://www.oecd.org/food-systems/understanding/triple-challenge/

Thomas, J, & Harden, A, 2008. Methods for the thematic synthesis of qualitative research in systematic reviews. BMC Medical Research Methodology, 8(1), 45.

Tree, V., Lane, C., 2022. Transforming nutrition. UN-Nutrition Journal, Volume 1. Rome, FAO. Available from: https://doi.org/10.4060/cc2805en

U.S. Agency for International Development. (2021). Expanding the reach of impact evaluation (ERIE). Available from: https://www.usaid.gov/PPL/MERLIN/ERIE

United Nations. (1997). Glossary of environment statistics, studies in methods. United Nations New York, NY.

Vaessen, J., Rivas, A., Duvendack, M., Jones, R. P., Leeuw, F., Van Gils, G., ... & Waddington, H. (2014). The effects of microcredit on women's control over household spending in developing countries: A systematic review and meta-analysis. Campbell systematic reviews, 10(1), 1-205.

Waddington, H. S., Fenton Villar, P. F., & Valentine, J. C. (2022). Can Non-Randomised Studies of Interventions Provide Unbiased Effect Estimates? A Systematic Review of Internal Replication Studies. Evaluation Review, 0193841X221116721.

Waddington, H., White, H., Snilstveit, B., Hombrados, J. G., Vojtkova, M., Davies, P., Bhavsar A., Eyers, J., Koehlmoos, T. P., Petticrew, M., Valentine, J. C. and Tugwell P., (2012). How to do a good systematic review of effects in international development: a tool kit, Journal of Development Effectiveness, 4(3), p. 359-387.

White, H. (2008) 'Of Probits and Participation: The Use of Mixed Methods in Quantitative Impact Evaluation'. IDS Bulletin. 39(1). pp. 98-109. Available from: https://doi.org/10.1111/j.1759-5436.2008.tb00436.x

Wilson, DB, Weisburd, D, McClure, D, 2011. Use of DNA testing in police investigative work for increasing offender identification, arrest, conviction and case clearance. Campbell Systematic Reviews 7, 1–53.

World Bank. (2009). Republic of Niger Impacts of Sustainable Land Management Programs on Land Management and Poverty in Niger. Environmental and Natural Resources Management Africa Region. World Bank Publications

World Bank (2016). "Agricultural Sector Risk Assessment: Methodological Guidance for Practitioners". Agriculture Global Practice. Discussion paper 10.

World Bank (2021). "Brief: Food Security and COVID-19." Available from: https://www.worldbank.org/en/topic/agriculture/brief/food-security-and-covid-19.

References for the evidence synthesis

Abdoulaye, A. & Sanders, J. H. (2013). A Matching approach to analyze the impact of new agricultural technologies: productivity and technical efficiency in Niger.

Abebe, Y., & Bekele, A. (2014). The impact of soil and water conservation program on the income and productivity of farm households in Adama District, Ethiopia. Science, Technology and Arts Research Journal, 3(3), 198-203.

Akoyi, K. T., & Maertens, M. (2018). Walk the talk: private sustainability standards in the Ugandan coffee sector. The Journal of Development Studies, 54(10), 1792-1818.

Baiyegunhi, L. J. S., Majokweni, Z. P., & Ferrer, S. R. D. (2019). Impact of outsourced agricultural extension program on smallholder farmers' net farm income in Msinga, KwaZulu-Natal, South Africa. Technology in Society, 57, 1-7.

Besley, T., Leight, J., Pande, R., & Rao, V. (2015). Long-run impacts of land regulation: Evidence from tenancy reform in India. Journal of Development Economics, 118, 72-87.

Chankrajang, T. (2015). Partial land rights and agricultural outcomes: evidence from Thailand. Land Economics, 91(1), 126-148.

Datta, N. (2014). Evaluating impacts of watershed development program on agricultural productivity, income, and livelihood in bhalki watershed of Bardhaman District, West Bengal. World Development, 66, 443-456.

De los Santos-Montero, L. A., & Bravo-Ureta, B. E. (2017, November). Productivity effects and natural resource management: econometric evidence from POSAF-II in Nicaragua. In Natural Resources Forum (Vol. 41, No. 4, pp. 220-233). Oxford, UK: Blackwell Publishing Ltd.

Deschamps-Laporte, J. P. (2013). The impact of extension services on farming households in Western Kenya: A propensity score approach (No. 5/2013). Working Paper.

Funsani, W., Rickaille, M., Zhu, J., Tian, X., Chibomba, V., Avea, A. D., & Balezentis, T. (2016). Farmer Input Support Programme and Household Income: Lessons from Zambia's Southern Province. Transformations in Business & Economics, 15.

Gibbons, M. A., Maffioli, A., & Rossi, M. A. (2016). Money for Wine? Complementarities in the provision of private and public goods to wine producers (No. IDB-WP-689). IDB Working Paper Series.

Ibanez, M., & Blackman, A. (2015). Environmental and economic impacts of growing certified organic coffee in Colombia. Environment for Development (February 2015).

International Fund for Agricultural Development (2018). IFAD IMPACT ASSESSMENT-Agricultural Sector Development Programme–Livestock (ASDP-L) and the Agriculture Service Support Programme (ASSP): Tanzania. Garbero, A., Chichaibelu, BB.

Kumar, N., & Quisumbing, A. R. (2011). Access, adoption, and diffusion: understanding the long-term impacts of improved vegetable and fish technologies in Bangladesh. Journal of Development Effectiveness, 3(2), 193-219.

Li, Y., & Wang, Z. (2021). Analysis on the effect of farmer income of policy-based agricultural insurance. Acta Agriculturae Scandinavica, Section B—Soil & Plant Science, 72(1), 386-400.

Lv, C. (2020). Empirical Analysis on the Effect of Agricultural Insurance on Production——Based on panel data of 31 provinces and cities in China from 2008 to 2018. In E3S Web of Conferences (Vol. 214, p. 01013). EDP Sciences.

Maia, G. A., Eusebio, G. S., & Silveira, R. L. (2016). Impact of microcredit on small-farm agricultural production: evidence from Brazil (No. 333-2016-14332).

Melesse, M. B., & Bulte, E. (2015). Does land registration and certification boost farm productivity? Evidence from Ethiopia. Agricultural Economics, 46(6), 757-768.

Niu, Z., Yi, F., & Chen, C. (2022). Agricultural Insurance and Agricultural Fertilizer Non-Point Source Pollution: Evidence from China's Policy-Based Agricultural Insurance Pilot. Sustainability, 14(5), 2800.

Romero, P., & Melo, O. (2021). Can a Territorial Use Right for Fisheries management make a difference for fishing communities?. Marine Policy, 124, 104359.

Ruml, A., & Qaim, M. (2020). Effects of marketing contracts and resource-providing contracts in the African small farm sector: Insights from oil palm production in Ghana. World Development, 136, 105110.

World Bank. (2009). Republic of Niger Impacts of Sustainable Land Management Programs on Land Management and Poverty in Niger. Environmental and Natural Resources Management Africa Region. World Bank Publications

References for the long-term impact evaluations not included in the evidence synthesis

Abebaw, D., Fentie, Y., & Kassa, B. (2010). The impact of a food security program on household food consumption in Northwestern Ethiopia: A matching estimator approach. Food Policy, 35(4), 286-293.

Agarwal, B. (2019). Does group farming empower rural women? Lessons from India's experiments. The Journal of Peasant Studies, 47(4), 841-872.

Ajefu, J. B., Singh, N., Ali, S., & Efobi, U. (2022). Women's Inheritance Rights and Child Health Outcomes in India. The Journal of Development Studies, 1-16.

Alderman, H., Hawkesworth, S., Lundberg, M., Tasneem, A., Mark, H., & Moore, S. E. (2014). Supplemental feeding during pregnancy compared with maternal supplementation during lactation does not affect schooling and cognitive development through late adolescence. The American journal of clinical nutrition, 99(1), 122-129.

Alem, Y., & Broussard, N. H. (2016). The impact of safety nets on technology adoption: A difference-in-differences analysis. Agricultural economics, 49(1), 13-24.

Amare, M., & Asfaw, S. (2012). Poverty reduction impact of food aid in rural Ethiopia. Journal of Development Effectiveness, 4(2), 235-256.

Anand, P., Saxena, S., Gonzales Martinez, R., & Dang, H. A. H. (2019). Can women's self-help groups contribute to sustainable development? Evidence of capability changes from Northern India. Journal of Human Development and Capabilities, 21(2), 137-160.

Asadullah, M. N., & Ara, J. (2016). Evaluating the long-run impact of an innovative anti-poverty programme: evidence using household panel data. Applied Economics, 48(2), 107-120.

Avitabile, C., Cunha, J. M., & Meilman Cohn, R. (2019). The medium term impacts of cash and in-kind food transfers on learning. Available at SSRN 3501896.

Bahta, Y., & Owusu-Sekyeer, E. (2018). Nexus between homestead food garden programme and land ownership in South Africa: Implication on the income of vegetable farmers.

Bari, M. A., Khan, G. D., He, B., & Yoshida, Y. (2022). The impact of unconditional cash and food assistance on contraceptive expenditure of rural households in Coastal Bangladesh: Evidence from fuzzy RDD. Plos one, 17(1), e0262031.

Behrman, J. R., Calderon, M. C., Preston, S. H., Hoddinott, J., Martorell, R., & Stein, A. D. (2009). Nutritional supplementation in girls influences the growth of their children: prospective study in Guatemala. The American journal of clinical nutrition, 90(5), 1372-1379.

Breza, E., CHANDRASEKHAR, A. G., & SHENOY, A. (2012). Skimming off the top: The unintended consequences of market expansion in the Indian dairy industry.

Carrillo, B., Iglesias, W. J., & Trujillo, J. C. (2015). Attainments and limitations of an early childhood programme in Colombia. Health Policy and Planning, 30(7), 906-916.

Chesterman, N. S., Entwistle, J., Chambers, M. C., Liu, H. C., Agrawal, A., & Brown, D. G. (2019). The effects of trainings in soil and water conservation on farming practices, livelihoods, and land-use intensity in the Ethiopian highlands. Land use policy, 87, 104051.

Dagunga, G., Amoakowaa, A., Ehiakpor, D. S., Mabe, F. N., & Danso-Abbeam, G. (2020). Interceding role of village saving groups on the welfare impact of agricultural technology adoption in the Upper East Region, Ghana. Scientific African, 8, e00433.

Dawang, C. (2018). Impact of Fishery Regulatory Innovation on Income and Nutrition of Smallholder Households in Plateau State, Nigeria.

Del Carpio, X. V., Loayza, N., & Datar, G. (2011). Is irrigation rehabilitation good for poor farmers? An impact evaluation of a non-experimental irrigation project in Peru. Journal of Agricultural Economics, 62(2), 449-473.

Dercon, S., Gilligan, D. O., Hoddinott, J., & Woldehanna, T. (2009). The impact of agricultural extension and roads on poverty and consumption growth in fifteen Ethiopian villages. American Journal of Agricultural Economics, 91(4), 1007-1021.

Durlak, J. A. (2009). How to select, calculate, and interpret effect sizes. Journal of pediatric psychology, 34(9), 917-928.

Elias, A., Nohmi, M., Kumi, Y., Akirai, S., Arega, A. (2014). The effect of agricultural extension service on the technical efficiency of teff (Eragrostistef) producers in Ethiopia. American Journal of Applied Sciences, 11(2), 223-239.

Filipski, M., Taylor, J. E., Abegaz, G. A., Ferede, T., Taffesse, A. S., & Diao, X. (2017). General equilibrium impact assessment of the Productive Safety Net Program in Ethiopia.

Francesconi, G. N., & Ruben, R. (2012). The hidden impact of cooperative membership on quality management: a case study from the dairy belt of Addis Ababa. Journal of Entrepreneurial and Organizational diversity, 1(1), 85-103.

Gahagan, S., Delker, E., Blanco, E., Burrows, R., & Lozoff, B. (2019). Randomized controlled trial of iron-fortified versus low-iron infant formula: developmental outcomes at 16 years. The Journal of pediatrics, 212, 124-130.

Gahagan, S., Yu, S., Kaciroti, N., Castillo, M., & Lozoff, B. (2009). Linear and ponderal growth trajectories in well-nourished, iron-sufficient infants are unimpaired by iron supplementation. The Journal of nutrition, 139(11), 2106-2112.

Garcia-Yi, J. (2014). Organic coffee certification in Peru as an alternative development-oriented drug control policy. International Journal of Development Issues, 13(1), 72-92.

He, S., Le, N. A., Ramirez-Zea, M., Martorell, R., Narayan, K. V., & Stein, A. D. (2020). Leptin partially mediates the association between early-life nutritional supplementation and long-term glycemic status among women in a Guatemalan longitudinal cohort. The American journal of clinical nutrition, 111(4), 804-813.

Jha, R., Imai, K. S., & Gaiha, R. (2009). Poverty, undernutrition and vulnerability in rural India: public works versus food subsidy. Chronic Poverty Research Centre Working Paper, (135).

Kinra, S., Sarma, K. R., Mendu, V. V. R., Ravikumar, R., Mohan, V., Wilkinson, I. B., ... & Ben-Shlomo, Y. (2008). Effect of integration of supplemental nutrition with public health programmes in pregnancy and early childhood on cardiovascular risk in rural Indian adolescents: long term follow-up of Hyderabad nutrition trial. Bmj, 337.

Kishore, A., & Chakrabarti, S. (2015). Is more inclusive more effective? The 'New Style'public distribution system in India. Food Policy, 55, 117-130.

Kuwornu, J. K., & Eric, S. O. (2012). Irrigation access and per capita consumption expenditure in farm households: Evidence from Ghana. Journal of Development and Agricultural Economics, 4(3), 78-92.

Liu, Y., Ma, W., Renwick, A., & Fu, X. (2018). The role of agricultural cooperatives in serving as a marketing channel: evidence from low-income regions of Sichuan province in China.

Lozoff, B., Castillo, M., Clark, K. M., & Smith, J. B. (2012). Iron-fortified vs low-iron infant formula: developmental outcome at 10 years. Archives of pediatrics & adolescent medicine, 166(3), 208-215.

Manea, R. E. (2020). School feeding programmes, education and food security in rural Malawi (No. 63-2020). Centre for International Environmental Studies, The Graduate Institute.

Manley, J., Fernald, L., & Gertler, P. (2014). Wealthy, healthy and wise: does money compensate for being born into difficult conditions?. Applied Economics Letters, 22(2), 121-126.

Mayda, A. M., & Steinberg, C. (2009). Do South-South trade agreements increase trade? Commodity-level evidence from COMESA. Canadian Journal of Economics/Revue canadienne d'économique, 42(4), 1361-1389.

Martin, R. M., Kramer, M. S., Patel, R., Rifas-Shiman, S. L., Thompson, J., Yang, S., ... & Oken, E. (2017). Effects of promoting long-term, exclusive breastfeeding on adolescent adiposity, blood pressure, and growth trajectories: a secondary analysis of a randomized clinical trial. JAMA pediatrics, 171(7), e170698-e170698.

Martínez, D., & Gori Maia, A. (2018). The impacts of cash transfers on subjective wellbeing and poverty: The case of Colombia. Journal of Family and Economic Issues, 39(4), 616-633.

Nandi, A., Behrman, J. R., Kinra, S., & Laxminarayan, R. (2018). Early-life nutrition is associated positively with schooling and labor market outcomes and negatively with marriage rates at age 20–25 years: evidence from the Andhra Pradesh children and parents study (APCAPS) in India. The Journal of nutrition, 148(1), 140-146.

Ngarava, S., Mushunje, A., & Chaminuka, P. (2018). Impact Of Livestock Development Programmes On Production & Risk: Case Of The Kaonafatso Ya Dikgomo (KyD) Programme.

Ntakyo, P. R., & Van Den Berg, M. (2022). The Unintended Side-Effects of a Major Development Strategy: Commercialization of Smallholder Production and Women Empowerment in Uganda. The Journal of Development Studies, 1-22.

Okwuokenye, G. F., & Okoedo-Okojie, D. U. (2014). Evaluation of extension agents commitment to the agricultural loans and inputs supply programme on special rice production in Delta State, Nigeria. Journal of Applied Sciences and Environmental Management, 18(2), 327-335.

Phadera, L., Sharma, D., & Wai-Poi, M. (2020). Iraq's Universal Public Distribution System: Utilization and Impacts During Displacement. World Bank Policy Research Working Paper, (9155).

Prado, E. L., Sebayang, S. K., Apriatni, M., Adawiyah, S. R., Hidayati, N., Islamiyah, A., ... & Shankar, A. H. (2017). Maternal multiple micronutrient supplementation and other biomedical and socioenvironmental influences on children's cognition at age 9–12 years in Indonesia: follow-up of the SUMMIT randomised trial. The Lancet Global Health, 5(2), e217-e228.

Prentice, A., Ward, K. A., Nigdikar, S., Hawkesworth, S., & Moore, S. E. (2019). Pregnancy supplementation of Gambian mothers with calcium carbonate alters mid-childhood IGF1 in a sex-specific manner. Bone, 120, 314-320.

Raghunathan, K. (2018). Women's self-help groups, decision-making, and improved agricultural practices in India.

Rand, J., & Tarp, F. (2009). Impact of an aquaculture extension project in Bangladesh. Journal of Development Effectiveness, 1(2), 130-146.

Rauniyar, G. (2012) Shallow Tubewell Irrigation in Nepal: Impacts of the Community Groundwater Irrigation Sector Project. Asian Development Bank.

Santos, I. S., Matijasevich, A., Assunção, M. C. F., Valle, N. C., Horta, B. L., Gonçalves, H. D., ... & Victora, C. G. (2015). Promotion of weight gain in early childhood does not increase metabolic risk in adolescents: a 15-year follow-up of a cluster-randomized controlled trial. The Journal of nutrition, 145(12), 2749-2755.

Shakya, T. R., Stewart, C. P., Christian, P., Katz, J., Schulze, K. J., wu, L. S. F., Leclerq, S. C., Khatry, S., & West, K. P. (2010). Maternal supplementation with vitamin A or β-carotene and cardiovascular risk factors among pre-adolescent children in rural Nepal. Journal of Developmental Origins of Health and Disease, 1(4), 262-270. https://doi.org/10.1017/S2040174410000255

Siddiqua, T. J., Roy, A. K., Akhtar, E., Haq, M. A., Wagatsuma, Y., Ekström, E. C., ... & Raqib, R. (2021). Prenatal nutrition supplementation and growth biomarkers in preadolescent Bangladeshi children: A birth cohort study. Maternal & child nutrition, 18(1), e13266.

Skröder, H., Kippler, M., Tofail, F., & Vahter, M. (2017). Early-life selenium status and cognitive function at 5 and 10 years of age in Bangladeshi children. Environmental health perspectives, 125(11), 117003.

Skugarevsky, O., Wade, K. H., Richmond, R. C., Martin, R. M., Tilling, K., Patel, R., ... & Kramer, M. S. (2014). Effects of promoting longer-term and exclusive breastfeeding on childhood eating attitudes: a cluster-randomized trial. International journal of epidemiology, 43(4), 1263-1271.

Wanyama, J. M., Nyambati, E. M., Mose, L. O., Mutoko, C. M., Wanyonyi, W. M., Wanjekeche, E., & Rono, S. C. (2010). Assessing impact of soil management technologies on smallholder farmers' livelihoods in North Western Kenya. African Journal of agricultural research, 5(21), 2899-2908.

Workneh, W. A., Takada, J., & Matsushita, S. (2020). The Impact of Using Small-Scale Irrigation Motor Pumps on Farmers' Household Incomes in Ethiopia: A Quasi-Experimental Approach. Sustainability, 12(19), 8142.

Yang, R., & Gao, Q. (2021). Water-Saving Irrigation Promotion and Food Security: A Study for China. Sustainability, 13(21), 12212.

Zhang, Z., Chen, Y. H., & Wu, L. H. (2021). Effects of Governmental Intervention on Foodborne Disease Events: Evidence from China. International Journal of Environmental Research and Public Health, 18(24), 13311.

List of appendixes

Appendix 1: Terms of Reference

Appendix 2: Protocol

Appendix 3: Methodological details

Appendix 4: Frequency of evaluation of specific interventions and outcome groups in the long-term

Appendix 5: Frequency of evaluation of intervention-outcome groups in the long-term in sub-Saharan Africa

Appendix 6: Meta-regression and subgroup analysis

Appendix 2 – 6 can be found as a separate document together with the report at norad.no/evaluation.

Appendix 1

Terms of References

Rapid evidence assessment

What facilitates success and failure of food systems interventions in the long-term

Background

Food security is a key priority area for the Norwegian government, which is also working on unifying its policies on climate and the environment and its development policy[1]. The Norwegian government plans to strengthen its food security support in development cooperation, of which food production is a key area. The Norwegian Agency for Development Cooperation (Norad) that manages a large part of this aid, practices results-based management which aims to use evidence to improve management, including knowledge from evaluations and research.

While there is a keen interest in evaluations and research, there is limited evidence of long-term effects of interventions. Pollard and Lindkvist (2020) argue that donor agencies do not sufficiently initiate and fund evaluations of long-term effects, partly because evaluations are often commissioned aligned with budget cycles, and there may be a limited demand for evidence after project funding have subsided. When funding cycles are typically shorter than 5 years, this means that we rarely find evaluations going back long after projects have ended. This is unfortunate as some effects may take longer to appear, and we may fail to learn what works in different contexts.

While there is some evidence that decentralized evaluation/project evaluations funded by donor agencies may have a shorter, rather than a longer- time perspective, less is known of the time-perspective of research in general and impact evaluations of food-security interventions in low- and middle-income countries in particular.

Evaluations with a short-time perspective may not capture that incremental change sparked by a project abates once the project has ended – or when transitional and transformational change is required to

1 See the Government's political platform (The Hurdal-platform).

achieve change but occurs only after a longer time[2]. Typically, transformational/transitional change may involve going from one state to another resulting in massive behavioral change. Unless data is collected after the desired change has occurred, we may erroneously conclude that a project was unsuccessful while if we had waited a little longer the project may have been regarded a massive success. Unless we pay particular attention to the long-term in project planning and evaluation, we risk myopic implementers that prefer incremental over transformational change if the results of projects targeting the first are measurable in the short-term, while the other may take a longer time.

The International Initiative for impact evaluation (3ie) published an Evidence Gap Map (EGM) in February 2022 on Food systems and nutrition evidence[3]. "*In the last few years, significant efforts have been made to improve food systems to facilitate better food security and nutrition outcomes. As a result, there is a vast amount of evidence on what works, but navigating the research is complicated, making it difficult for decision-makers to use the evidence.*" The evidence gap map provides an "*overview of the literature on food systems interventions. It reports on evidence from all key areas and intervention types within the food system and also identifies potential primary and synthesis evidence gaps.*"

While this Evidence Gap Map provides a comprehensive overview of topics covered and topics that are missing for more than 2000 impact evaluations and synthesis studies, 3ie has as of yet not checked gaps in terms of time-perspectives in these evaluations. This evidence gap-map and the literature already collected provides a unique opportunity to explore the time-perspective of studies of food security interventions.

To ensure we learn more about what works in the long-term, the Department for Evaluation wish to support 3ie's research, and the efforts to further strengthen the Evidence gap map on food systems and nutrition, by including the time-perspective of these evaluations in a Rapid Evidence Assessment (REA). More specifically the aim is to understand how outcome trajectories evolve over time, the extent to which these trajectories can be attributed to the intervention and synthesis evidence of what facilitates success in the long-term.

The suggested REA is part of a larger effort from the department for evaluation on food security where the department will a) map and synthesise evidence from process evaluations of climate adaptation in food production (out for tender now) b) analyse use of knowledge in the current food-security portfolio (planning stage) and c) investigate long-term effects of food-security aid. These terms of references are a part of c) with an aim to contribute with knowledge on what works in the long term.

Purpose

The purpose of the study is improved accessibility and knowledge of the time-perspective, whether long-term effects are covered, and what facilitates long-term success and failure, relying on the universe of studies identified by 3ie in their evidence gap-map on food systems and nutrition.

It is expected that this work will be of interest to a broad audience, including, agencies and governments working in the area of food security, researchers, and the general public. This includes the Norwegian aid-administration that is currently building a food-security portfolio. In particular we hope that the study will increase interest in what works in the longer term, and that potential long-term effects are considered in project planning, monitoring and evaluation (M&E).

Objectives

- Identify/map the time-perspective of evaluations included in the EGM (including average time from implementation to measurement of outcomes)

2 In the chapter 'From measuring impact to understanding change', Forss quotes Ackerman who distinguishes between three types of change: 1) Developmental, 2) Transitional and 3) Transformational. Developmental change is incremental, while transitional and transformational change is radical and involves a shift from one stage to another. The major difference between transitional change and transformational change is that transitional change is planned and involves shifting to a more desirable state, while transformational change is radical and can be unexpected.

3 See: https://www.3ieimpact.org/programme/nutrition-and-food-security/food-systems-and-nutrition - quotes are from this site.

- Identify whether outcome trajectories change over time (including identifying potential patterns, and whether these can be attributed to the intervention)
- For evaluations that has a long-perspective – explain what facilitates success/failure.
- For the same evaluations – synthesize any evidence related to long-term environmental and climate effects.

Scope

The scope are the more than 2000 impact evaluations and synthesis studies identified by 3ie for the EGM on food systems and nutrition.

In this rapid evidence assessment, long-term is defined as at least 10 years after implementation of an intervention. Evidence collected more than 5 years after the intervention, but less than 10 years is referred to as medium-term evidence.

Possible research questions

Time- focus/time-perspective of evaluations

1. To what extent do evaluations from the EGM on food systems document long-term effects? (For example average number of years from implementation to measurement of change in outcomes)

2. To what extent do these evaluations document long-term environmental and climate effects?

3. Do these evaluations consider adverse effects?

Synthesis of findings

4. How effective are food-security interventions in the long-term? (Can any patterns in impact trajectories be identified?)

5. What facilitates long-term success and failure?

Approach and methodology (based on 3ie's approach to rapid evidence assessment)

The suggested approach and methodology is 3ie's tried and tested approach to rapid evidence assessment developed in line with other types of evidence synthesis and described below.

Rapid Evidence Assessments (REA) are a type of evidence synthesis approach that has been developed to address policy relevant questions within a more limited time and resource context than what is typically available for full systematic reviews. There is no single definition of a rapid review and recent reviews of study methods have highlighted the variation in rapid review methods (Featherstone et al., 2015; Hartling et al., 2015; Khangura et al., 2012; Tricco et al., 2017). However, such approaches typically involve adjusting methods used in traditional systematic reviews and adopt one or more shortcuts to give more timely answers to urgent questions (Schünemann & Moja, 2015).

Common shortcuts include addressing more narrowly focused questions; limiting the number of sources consulted in the search; limiting the time frame and language of the search; using only one reviewer for screening, full text review, appraisal and data extraction; and presenting conclusions as a narrative summary rather than conducting a formal statistical synthesis (Ganann et al., 2010; Featherstone et al., 2015; Tricco et al., 2017). The reported production time for rapid reviews vary significantly, ranging from 5 minutes to 12 months, with most reviews typically taking between 1 and 6 months (Hartling et al., 2015; Featherstone et al., 2015; Tricco et al., 2017).

3ie's approach to REA uses explicit, systematic and transparent procedures to quickly respond to a specific policy question adopting the most rigorous

methods feasible within the time and resources available. The methods vary depending on the timeframe and budget.

Description of steps of the rapid evidence assessment:

Development of a study protocol describing the methods: The study protocol will outline the approach. The development of a study protocol a-priori approach ensures a systematic and transparent approach is followed. The protocol will be shared with stakeholders for additional inputs. It will include a description of the types of studies considered for inclusion, search and screening strategy, framework for data extraction, critical appraisal and analysis.

Search and screening: The main search and screening process was completed during the development of the original EGM, however the team will conduct some additional searches to identify qualitative studies related to the programmes evaluated in the included studies. The details of the search strategy and inclusion criteria will be determined at the protocol stage and will be informed by the scope of the literature identified. Screening of studies will be done by trained consultants with oversight from a Senior Evaluation Specialist.

Data extraction: The team will use the data extracted for the EGM and use a standardized data extraction tool for the quantitative data. In addition to relevant effect sizes, the data extraction template may include bibliographic, geographic and substantive data, including timelines for data collection, information on effect moderators, standardised methods information, and data on how studies address equity. The team may also extract the descriptive, methodological, and quantitative data from each included study using a standardised data extraction form.

Risk of bias (critical appraisal) of included studies: The team will assess the risk of bias of the eligible studies, drawing on the signalling questions in 3ie's risk of bias tool which covers both internal validity and statistical conclusion validity of experimental and quasi-experimental designs (Waddington et al. 2012), and the bias domains and extensions to Cochrane's ROBINS-I tool (Sterne et al., 2016). Qualitative information will be extracted from the additional qualitative studies and the quality of these studies will be assessed through the CASP tool (Critical Apraisal Skills Programme [CASP], 2008).

Data analysis (quantitative): Analysis will depend upon the number of studies ultimately included, but we aim to set a scope for the REA that includes no more than 20 studies in order to allow for a full and rigorous analysis, given the timeframe and resources available. The team will conduct a meta-analyses of studies that we assess to be sufficiently similar. Meta-analysis will be implemented when two or more effect sizes using a similar outcome construct and when the comparison group state is judged to be similar across the two, similar to the approach taken by (Wilson et al., 2011). Whenever feasible, the team will conduct moderator analyses to investigate sources of heterogeneity using random effects meta-regression. Following the PROGRESS-PLUS approach (Oliver et al., 2008), the team will assess moderators falling into three broad categories of extrinsic, methodological, and substantive characteristics to address inequity aspects. This moderator analysis will be used to determine the policy design features that affect impact.

Data analysis (qualitative): The team will perform a barriers/facilitators factors analysis to identify any implementing element which may or may not facilitate programme success, as well as any nuances about the context of each included study following the method by (Thomas & Harden 2008). Specific context-related information which can help to understand and explain the direction of the meta-analysis effects will be included to give an overall view of how those interventions work. The team will also discuss information on costs, and specifically address the sustainability of the outcomes in the medium- and long-term.

Reporting: Findings and implications for policy and research will be synthesised in a 40-page report using plain language to maximise the use and uptake of findings.

Deliverables

a. Inception report including a study protocol.

b. Draft report

c. Final report

Ethical consideration

All parts of the assignment shall adhere to recognised evaluation principles and the OECD Development Assistance Committee's quality standards for development evaluation, as well as relevant guidelines from the Department for Evaluation. It is expected that the assignment is carried out according to accepted research and evaluation ethics and it shall be undertaken with integrity and honesty and ensure inclusiveness of views. Ethical considerations and accompanying safeguards shall be documented throughout the review processes.

Organization of the rapid evidence assessment

The assignment will be led by 3ie's team leader and overseen by the Department for Evaluation in Norad. The team leader will report to the Department for Evaluation. The team leader shall be in charge of all deliveries and will report to the Department for Evaluation on the team's progress, including any problems that may jeopardise the assignment.
All decisions concerning the interpretation of this Terms of Reference, and all deliverables are subject to approval by the Department for Evaluation.

Quality assurance shall be provided by 3ie prior to submission of all deliverables.

Department for Evaluation

norad.no/evaluation