The effects of food systems interventions on women's empowerment: Protocol for a rapid evidence assessment

Kishore Kumer Basak
International Initiative for Impact Evaluation (3ie)

Etienne Lwamba
3ie

Ingunn Storhaug
3ie

Pierre Marion
3ie and University of Sussex, Business School, Economics department

Ashiqun Nabi
3ie

Shannon Shishler
3ie

Anil Thota
3ie

Rapid Evidence Assessment Protocol

April 2024

International Initiative for Impact Evaluation

german cooperation
DEUTSCHE ZUSAMMENARBEIT

Implemented by

giz Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH

About 3ie

The International Initiative for Impact Evaluation (3ie) develops evidence on how to effectively transform the lives of the poor in low- and middle-income countries. Established in 2008, we offer comprehensive support and a diversity of approaches to achieve development goals by producing, synthesizing and promoting the uptake of impact evaluation evidence. We work closely with governments, foundations, NGOs, development institutions and research organizations to address their decision-making needs. With offices in Washington DC, New Delhi and London, and a global network of leading researchers, we offer deep expertise across our extensive menu of evaluation services.

About this report

This report provides a summary of the evidence available on the effects of food systems interventions on women's empowerment. It presents new insights into the findings of the living Food systems and nutrition (FSN) Evidence and Gap Map (E&GM) that was initially commissioned by the German Federal Ministry for Economic Cooperation and Development (BMZ) through Deutsche Gesellschaft für Internationale Zusammenarbeit's (GIZ) "Knowledge for Nutrition" Program. It also presents a protocol for a Rapid Evidence Assessment (REA) that aims to synthesize and appraise the evidence available on the effect of FSN interventions on women's empowerment. This report was made possible with generous support from GIZ.

 Please direct any comments or queries to the corresponding author, Etienne Lwamba, elwamba@3ieimpact.org.

Suggested citation: Basak, Kishore Kumer, Etienne Lawmba, Ingunn Storhaug, Pierre Marion, Ashiqun Nabi, Shannon Shisler, and Anil Thota, 2024. The effects of food systems interventions on women's empowerment: Protocol for a rapid evidence assessment. London: International Initiative for Impact Evaluation (3ie).

Abstract

In 2022, about 9% of the world's population faced chronic hunger. Despite a decline from 13% in 2001, recent indicators show alarming hunger trends, increasing from 8% in 2015. Women are disproportionately affected and constitute 60% of the world's food insecure population. Unequal access to food security is rooted in and reinforces restrictive gender norms and inequality.

This research aims to support GIZ’s work by conducting a Rapid Evidence Assessment (REA) examining the effects of Food Security and Nutrition (FSN) interventions on women’s empowerment outcomes in a sample of countries. This protocol details the methods and processes we will follow for this REA. In addition to the effects of FSN interventions, we will also assess whether they vary by context, intervention features, or other relevant moderators. We will also assess unintended consequences or adverse effects of these interventions. Finally, we will describe the evaluation strategies used by authors in this sector. Overall, the findings from this REA will provide a basis for decision-makers, funders, and practitioners to consult rigorous evidence for decision-making and program design on the nexus between FSN and women’s empowerment.

Contents

List of figures and tables

List of abbreviations

BMGF	Bill and Melinda Gates Foundation
BMZ	German Federal Ministry for Economic Cooperation and Development
BRAC	Bangladesh Rural Advancement Committee
CASP	Critical Appraisal Skills Program
CGIAR	Consultative Group for International Agricultural Research
E&GM	Evidence and Gap Map
FCAS	Fragile and Conflict-Affected Situations
FCDO	Foreign, Commonwealth and Development Office
FSN	Food Security and Nutrition
GBV	Gender-based Violence
GGPI	Gender Gap Performance Index
GHI	Global Hunger Index
GII	Gender Inequality Index
GIZ	Deutsche Gesellschaft für Internationale Zusammenarbeit
HKI	Helen Keller International
IFAD	International Fund for Agricultural Development
IFPRI	International Food Policy Research Institute
IGC	International Growth Centre
IMMANA	Innovative Methods and Metrics for Agriculture and Nutrition Actions
IPV	Intimate Partner Violence
ITS	Interrupted Time Series
IV	Instrumental Variable
L&MICs	Low- & Middle-Income Countries
PICOS	Population, Intervention, Comparison, Outcome, Study Design
PSNP	Productive Safety Net Program
QEX	Quantitative Data Extraction
REA	Rapid Evidence Assessment
Rob	Risk of Bias
SDG	Sustainable Development Goal
SE	Standard Error
SMS	Short Message Service
UNDP	United Nations Development Program
WHO	World Health Organization

1. Introduction

As per the Food and Agriculture Organization of the United Nations (FAO), a person is food insecure “when they lack regular access to enough safe and nutritious food for normal growth and development and an active and healthy life due to unavailability of food and/or lack of resource to obtain food” (FAO 2024). This definition directly relates to the concept of *hunger* that FAO defines as “an uncomfortable or painful sensation caused by insufficient consumption of dietary energy that becomes chronic when the person does not consume a sufficient amount of calories on a regular basis to live a normal, active and healthy life” (FAO 2024) and to the concept of *nutrition* that encompasses the provision of the necessary nutrient for health and growth.

In 2022, 9.2 percent of the global population faced chronic hunger, a 1 percent increase from 2019 (United Nations 2023). In 2005, about 793 million people experienced chronic hunger, decreasing to 598 million in 2010 and 589 million in 2015 but increasing again to 735 million in 2022 (FAO 2023a). In the global struggle against food insecurity, malnutrition, and hunger, women are more vulnerable than men: they often eat last and least, are more likely to live in extreme poverty, and are often more vulnerable to famine and food crises (WFP 2023). As a result, women represent 60% of the people who are food insecure in the world. Additionally, in nearly 2/3 of countries, women are more likely than men to report food insecurity (WFP 2023).

Deep-rooted gender norms and gender inequalities can both be causes and catalysts of gender disparities in food security and nutrition. Food insecurity, malnutrition, and hunger affect genders differently, both biologically and socially, and are highly influenced by gender norms. These gender disparities in food security, hunger, and nutrition are also related to aspects of health, livelihood, income, or social-political rights. As such, addressing food insecurity might contribute to women’s empowerment and gender equality more broadly. Similarly, women participate in all aspects of the food systems (farming, trade, marketing, care, etc.); thus, empowering women might contribute to improved food security (Feed the Future 2022).

In this iterative process, food systems and nutrition are interlinked with women’s empowerment and might lead to mutually positive outcomes (SPRING 2014). Recent work from Berretta and colleagues (2023) demonstrated gender transformative and women's empowerment interventions' positive and statistically significant effect on

nutrition-related outcomes, including food security, affordability, availability, and diet quality and adequacy. Through their challenge of gender norms and specific targeting of gender groups to address the root causes of gender inequalities, gender transformative interventions can contribute to improved food security. However, little is known on the other side of the nexus: the potential effect of food systems and nutrition interventions on women's empowerment.

Here, we present a protocol for a rapid evidence assessment (REA) to complement existing evidence synthesis by analyzing the effects of food systems and nutrition (FSN) interventions on women's empowerment outcomes. We aim to address the following research questions:

- What are the effects of food systems and nutrition interventions on outcomes related to women's empowerment?
- Are there unintended consequences, including adverse effects, of these interventions?
- Do effects vary by context, intervention features, or other moderators?
- What evaluation design strategies are used?

Our research will draw on the evidence available in 3ie's living FSN Evidence and Gap Map (E&GM) commissioned initially by GIZ/BMZ (Moore et al. 2021) up to its November 2023 update (Storhaug et al. 2023). This protocol outlines our methodology for a rapid evidence assessment (REA) that will synthesize the evidence identified on the effect of FSN interventions on women's empowerment outcomes.

Overall, the findings from this research can provide a basis for decision-makers, funders, and practitioners to consult rigorous evidence when designing and implementing development programming on the nexus between FSN and women's empowerment. Researchers and commissioners will also be able to use the findings to identify gaps in the FSN evidence base that can be meaningfully filled through new research on how FSN interventions can affect women's empowerment.

2. Background

2.1. The problem, condition, or issue

2.1.1. Hunger remains a critical global health challenge and a major barrier to sustainable development.

Urgent actions are needed to address food insecurity, hunger and malnutrition's root causes and consequences. Food insecurity, hunger, and malnutrition are multisectoral issues affecting economic development, health, education, equality, equity, and social development by generating losses of productivity, vulnerability to diseases, or social inequalities (United Nations 2024). Despite progress, significant portions of the global population, particularly in developing regions, face hurdles in obtaining a diverse and healthy diet (FAO et al. 2023c). According to a report released by five UN agencies, 735 million individuals worldwide experienced hunger, and 258 million people across 58 countries and territories faced acute food insecurity at crisis or worse levels (FAO et al. 2023; FSIN 2023). Although the Sustainable Development Goal (SDG) 2 targets ending hunger, achieving food security, improving nutrition, and promoting sustainable agriculture by 2030, indicators are not on track, and progress towards this goal has been severely hampered by the combined effects of the COVID-19 pandemic, conflict, climate change, and economic downturn (UNDESA 2023).

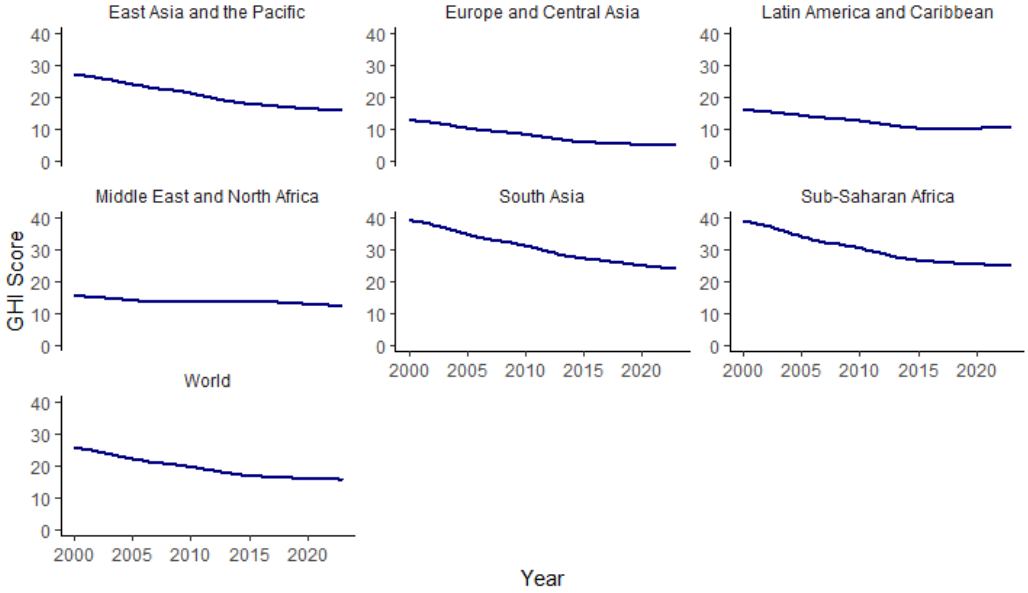
While global indicators show a slow improvement in addressing food insecurity, hunger, and malnutrition, we are witnessing a resurgence in recent years. According to the latest report by the Food and Agriculture Organization of the United Nations (FAO; FAO et al. 2023), the number of people affected by hunger in low- and middle-income countries (L&MICs) increased by 9.9 percent from 2019 to 2022. Likewise, moderate, or severe food insecurity in L&MICs rose from 25.4% in 2019 to 30.4% in 2022, affecting 2.37 billion people (FAO et al. 2023).

Indicators of food insecurity, hunger, and malnutrition show regional differences. Despite a moderate Global Hunger Index[1] (GHI) with a score of 18.3 for 2023, we observe regional differences and slow progress toward tackling global hunger. South Asia and Sub-Saharan Africa have had the highest GHI scores and the

[1] The GHI measures and tracks hunger globally as well as by region. It is calculated through four indicators: undernourishment, child stunting, child wasting, and child mortality. The lower the score, the better the situation. For more information: https://www.globalhungerindex.org/

most severe hunger situations in the 21st century (Figure 1). However, South Asia and Sub-Saharan Africa also improved their hunger situation most.

Figure 1. Trend in Global Hunger Index Score 2000-2023 by region



2.1.2. Gender inequalities: a vector of vulnerability to food crises

Women and girls are particularly vulnerable to food insecurity and hunger. According to FAO (2022), among 828 million people severely affected by hunger in 2021, 60 percent were women and girls. The report also shows that the gender gap in food insecurity continued to rise in 2021, as women were more likely to report moderate or severe food insecurity than men. The World Food Program highlighted similar trends in 2023 (WFP 2023): compared to men globally, 150 million more women are food insecure.

Gender inequality and restrictive gender norms partly explain women's higher vulnerability to food insecurity and hunger. This discrepancy is partly due to restrictive gender roles that assign women and men different responsibilities and opportunities in food systems, such as production, processing, trade, and consumption (Maraka 2021). These factors affect their ability to achieve food security and nutrition for themselves and their families. Women also face challenges and barriers to participating in and benefiting from food systems, such as lack of education, restricted access to resources, unequal economic opportunities, denied decision-making power, and limited control over time use (Oseni et al. 2015; Backiny-Yetna, McGee, and

Mcgee 2015; Kilic, Palacios-López, and Goldstein 2015). For example, in South Asia, only 13 percent of women own land, and in Africa, only 10 percent of women have access to credit (FAO 2023b). Over 90% of the world's female population (3.1 billion) lives in countries with low or moderate performance in achieving gender parity (UNDP 2023; WEF 2023).

2.1.3. Empowering women through food and nutrition

Food security and gender equality are interconnected concepts. Social norms, knowledge, skills, and the distribution of decision-making power within households affect both women's empowerment and nutrition outcomes. This nexus comprises three interconnected elements: how women might allocate their income for both food and non-food expenses and experience financial independence, the capacity of women to care for themselves and their families, and the energy expended by women in various roles and responsibilities (SPRING 2014). This interconnection between FSN and gender norms shows that transforming food systems to be more inclusive and empowering for women requires addressing the underlying gender norms and power dynamics that shape women's roles and opportunities in food systems.

Achieving women's empowerment can improve nutrition. Gender equality influences women's dietary intake, care practices, health-seeking behavior, and allocation of food within the household. As observed by Njuki and colleagues (2022, 1), "achieving gender equality and women's empowerment in food systems can result in greater food security, better nutrition, and more just, resilient, and sustainable food systems for all." Elevating women's agency and decision-making influence, both within households and in the governance of food systems, can contribute to their well-being, food security, and nutrition (Njuki et al. 2021). Additionally, this empowerment can strengthen their capacity to advocate for their rights and interests (Njuki et al. 2021; Berretta et al. 2023). However, despite the observed tradeoffs between dimensions of women's empowerment and FSN outcomes, the effect might vary according to context or socio-economic factors. In 2021, Quisumbing and colleagues analyzed the tradeoffs among women's empowerment dimensions and nutrition outcomes. Although they observed significant associations, they highlighted that other factors such as household wealth, age, country-level factors, or other underlying determinants should also be considered (Quisumbing et al. 2021b).

To complement the existing evidence, our research will explore whether addressing the root causes of food insecurity, hunger, and malnutrition can contribute to women's empowerment. Food systems are transforming due to dietary changes, production processes, and environmental factors. As presented in the FSN E&GM, researchers have analyzed the effects of these food systems interventions and transformations and their potential effect on control over resources, participation in decision-making, skills and knowledge, agency, and participation in community activities through primary research, but there is a gap of synthesis evidence (Storhaug et al. 2023). Our REA will synthesize this evidence to analyze the effect of FSN interventions on women's empowerment.

2.2. The interventions

Our review will analyze the effect of FSN interventions on women's empowerment outcomes. It will complement the analysis from Berretta and colleagues (2023), who analyzed the effect of women's empowerment interventions on food security and nutrition, by examining the effect of FSN interventions on women's empowerment outcomes. Through this complementary research, we will contribute to the body of synthesis evidence on the overall nexus between food systems and nutrition, and women's empowerment and gender equality.

2.2.1. Food systems and nutrition

Our scope includes a subset of the interventions in the FSN E&GM framework. The food systems interventions of interest for this REA fall into the food supply chain, food environment, and consumer behavior domains:

- The food supply chain includes activities that affect the process, from how the food is produced to the consumption and disposal of the waste (HLPE 2017). The FSN E&GM covers interventions focusing on improving the production system, storage and distribution, processing and packaging, and food loss and waste management.
- The food environment includes how consumers engage with the food system that provides and shapes dietary preferences, choices, and nutritional status based on physical, economic, political, and sociocultural contexts (HLPE 2017). The FSN E&GM comprises interventions focusing on availability and affordability, promotion and labeling, and quality and safety of food.

- The consumer behavior domain includes activities that shape individual preferences related to consumption, allocation of food within the household, food prices, and income available for food (HLPE 2017). The FSN E&GM includes interventions focusing on behavior change concerning the food system.

2.2.2. Women's empowerment and gender transformative approaches

Our definition of women's empowerment draws on Naila Kabeer's definition: "… a process by which women who have been denied the ability to make strategic life choices acquire such an ability" (Kabeer 1999, 437). In order to have the ability to make strategic life choices, women need resources, agency, and achievements (Richardson 2018):

- Resources include material, human and social, and institutional resources that create conditions leading to women's empowerment.
- Agency refers to being able to identify goals and having the ability to act upon them.
- Achievements refer to realizing those goals and the existence of an enabling environment, norms, and behavior that contribute to gender equality.

As per UNICEF's definition, Gender Transformative Approaches (GTA) are "concerned with redressing gender inequalities, removing structural barriers, such as unequal roles and rights and empowering disadvantaged population. They aim both to change overall structures than underpin gender inequality and to contribute to lasting change in individuals' lives" (Marcus et al. 2022, 2). Our REA will both analyze FSN interventions integrating GTA in their program design and FSN interventions not integrating GTA to compare their respective effect on women's empowerment outcomes.

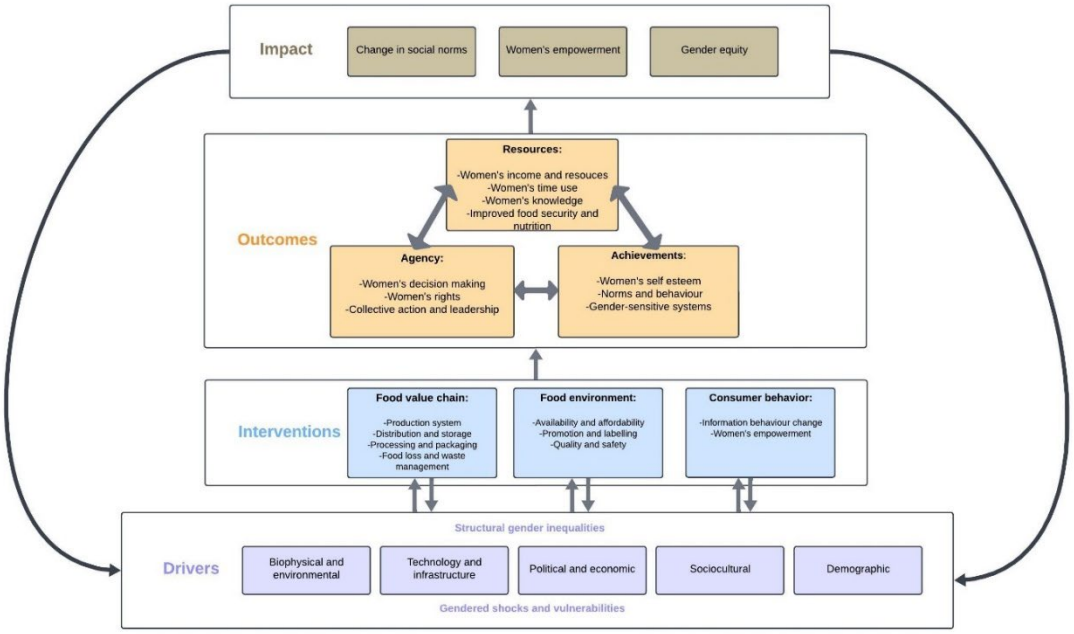
2.3. Expected theory of change

In order to understand how the food system affects women's empowerment, it is important to consider the role of gender in the food system. Njuki and colleagues (2022) adapted de Brauw and colleagues' (2019) food systems conceptual framework and applied a gender lens to it. Their model shows how women's empowerment can affect each part of the food system and how the relationship is

interlinked, as the food system can also affect outcomes of gender equity and women's empowerment. We have adapted the model but specified it to explore the link between food systems interventions and women's empowerment outcomes (Figure 2).

Food systems drivers **represent the variables that enable change or influence the functionality of the food systems** (HLPE 2017). These can be biophysical, environmental, technological, infrastructural, political, economic, sociocultural, and demographic. Some examples of specific drivers include population growth, climate change, and COVID-19 (HLPE 2017). Food systems conceptual frameworks can include gender as a part of the sociocultural drivers, considering how gender relationships and norms impact the food systems (Figure 2). This model reflects how every driver is affected by gender inequalities and considers how vulnerabilities and shocks have gender-based effects. Njuki and colleagues (2022) emphasize the importance of considering the role of gender in each of the drivers when analyzing their effect on the food system.

Figure 2: Theory of change



The *intervention* part of the model shows the three domains of the food system aligned with our published FSN E&GM: the food value chain, the food environment, and consumer behavior. Interventions within this pathway can operate on different scales, from household to global (Moore 2021).

These interventions can affect several *outcomes* for women. Those outcomes belong to the three pillars of women's empowerment (resources, agency, and achievements) and include knowledge, resources/income, time use, self-esteem, decision-making, as well as food and nutrition security outcomes as presented in the FSN E&GM. Improvements in these outcomes enable change for overall social norms, women's empowerment, and gender equity, leading back to addressing some of the drivers of structural gender inequality.

2.4. Rationale for the review

2.4.1. Review of existing literature

Despite the primary and synthesis literature on the effect of women's empowerment interventions on food security, little synthesis evidence exists on the effect of the food system and nutrition intervention on women's empowerment outcomes*.* The evidence mapped in the FSN E&GM shows that current synthesis evidence on the nexus between FSN and women's empowerment is mostly covered by reviews rated as low confidence. In 2022, Njuki and colleagues published a scoping review of the evidence on gender equality, women's empowerment, and food systems. They discovered more substantial evidence concerning the varying access women have to resources and the effect of women's empowerment and maternal education on nutrition or dietary diversity. However, they encountered limited evidence regarding gender considerations in food systems for urban women and in aquaculture value chains, as well as few effective strategies for involving men in the advancement of women's empowerment in food systems or issues related to migrations (Njuki et al. 2022).

Some synthesis evidence is available on the effect of women's empowerment interventions on FSN outcomes, although it is mostly rated as low confidence. The recent study by Berretta and colleagues (2023) aimed to address this gap and found that women's empowerment interventions improve nutrition-related outcomes, with the largest effects on food security, especially on aspects of affordability and availability. Diet quality and adequacy, anthropometrics, and micronutrient status also improve, but to a lesser extent. Well-being outcomes are not significantly affected.

However, no synthesis evidence is available on the effect of FSN interventions on women's empowerment outcomes. The FSN E&GM shows that the available

synthesis evidence either focused on the effect of specific interventions such as cash transfers (Evans and Popova 2014; Bastagli et al. 2019; Bastagli et al. 2016; Owusu-Addo, Renzaho, and Smith 2018), nutrition-sensitive agriculture (Ruel, Quisumbing, and Balagamwala 2018), or livestock and aquaculture (Blackmore, Lesorogol, and Iannotti 2018) on women's empowerment outcomes or was not rated as high confidence. Our review will aim to fill this gap of synthesis evidence on FSN interventions' effect on women's empowerment outcomes.

2.4.2. Relevance to policy and practice

Food security and gender equality are both among the core priorities and commitments of international development stakeholders. In 2020, donors spent USD 12 billion on food security and nutrition (Florizone and Smaller 2020). However, according to international observers, the commitment would need to increase by USD 14 to 19 billion annually until 2030 to end hunger (Oxfam 2022). On the other hand, financial commitment to addressing gender inequalities as a program's secondary objective has increased in the last ten years from USD 22.7 billion to 57.4 billion but remained relatively stable at USD 5 billion as a program's principal objective (OECD 2023).

Despite this parallel commitment to addressing food insecurity and gender inequality, relatively fewer programs address them as a nexus. An example of nexus focus includes BRAC's Targeting the Ultra Poor. Launched in 2002, the Program supported 1.77 million extremely poor households in Bangladesh through assistance on livelihoods, social safety nets, financial inclusion, and social integration to tackle food insecurity, irregular income, lack of skills and assets, and gender inequalities (BRAC 2016). In Africa, the Ethiopian Productive Safety Net Program (PSNP) supported rural populations in the fight against chronic food insecurity and resilience by providing economic opportunities and contributing to reducing gender inequalities (European Union 2014). Other examples include the UN Joint Program for Rural Women's Economic Empowerment: this multi-country program seeks to improve food security and nutrition for rural women and their households while increasing their income and economic autonomy, increasing their voice and agency, and strengthening gender-responsive legal frameworks, policies, and institutions (JP RWEE 2024).

Other development actors have taken action to support the growth of the evidence base on nutrition and women's empowerment. For example, the Innovative Methods and Metrics for Agriculture and Nutrition Actions (IMMANA) has supported the development of tools, capacity, and evidence in agriculture and food systems research, including on their link with women's empowerment (ANH Academy 2024). The International Food Policy Research Institute (IFPRI) is also working on building the evidence base through a portfolio of agricultural development projects under the Gender, Agriculture, and Asset Project (GAAP 2 2024). Despite its recent increase, more evidence is needed on the potential impact of FSN interventions on women's empowerment outcomes to inform the design of future programs and strengthen the consideration for this nexus.

2.4.3. Defining goals *and contribution*

The 3ie FSN E&GM provides us with the existing primary and secondary evidence on FSN and women's empowerment. 3ie has been maintaining its living FSN E&GM since January 2021 and has contributed to developing the evidence base on FSN interventions and their related outcomes. The map contains 2,537 articles, making it a valuable resource of evidence in the field. Because the E&GM has been conducted using a rigorous and systematic search and screening process, it can be leveraged as the basis for a Rapid Evidence Assessment (REA). REAs can support policy and program decisions by synthesizing the existing evidence on the effects of interventions on specific outcomes.

By building on the evidence from the FSN E&GM, our REA will present a unique synthesis of available evidence on the effects of interventions to support women's empowerment within the food system. It will also provide practical and policy-relevant implications for designing, implementing, and evaluating food systems interventions that are gender-responsive, gender-transformative, and nutrition-sensitive and that can enhance women's empowerment and nutrition outcomes. Finally, it will support the achievement of the Sustainable Development Goals (SDGs), especially SDG 2 (Zero Hunger), SDG 5 (Gender Equality), and SDG 17 (Partnerships for the Goals), by providing evidence on inclusive and sustainable food systems that empower women and improve nutrition for all.

3. Mapping of evidence on the effects of FSN interventions on women's empowerment

3.1. About the Food Systems and Nutrition E&GM

The FSN E&GM was commissioned by GIZ and was first published in 2021 (Moore et al. 2021). The E&GM provides an overview of the literature on food systems interventions on food security and nutrition outcomes in L&MICs. This is one of the largest E&GMs 3ie has undertaken, initially including 1,838 IEs and 178 SRs (Moore et al. 2021). Since then, it has become a living E&GM that has been updated quarterly for the past 2.5 years[2]. The map total as of January 2024 stands at 2,338 IEs and 199 SRs.

The original framework used the International Food Policy Research Institute's (IFPRI) extension of the High-Level Panel of Experts (HLPE) 2017 framework by De Brauw and colleagues that divides food systems interventions into three domains (de Brauw et al. 2019); the food value chain, food environment, and consumer behavior domain. Over the past year, the E&GM framework has been expanded to include any women's empowerment interventions and gender transformative interventions, not only women's empowerment interventions within the food systems. Cash transfer and social insurance programs are also being added to the online E&GM.

Our REA scope will be exclusively based on the primary and secondary evidence availability in the FSN E&GM. We will not undertake an additional search for evidence outside of the scope covered by the existing E&GM (more information about the scope of the FSN E&GM is available in Appendix 0). This excludes the additional search for qualitative evidence presented in the subsequent sections.

3.2. Evidence mapping on studies measuring women's empowerment outcomes

We identified 96 quantitative impact evaluations and five systematic reviews across 41 countries on the effects of food systems and nutrition interventions on women's empowerment outcomes (Table 1). This amounts to four percent of the quantitative impact evaluations and two percent of the systematic reviews identified across the FSN E&GM. We observed a significant increase in the number of studies

[2] Six updates have been undertaken since the initial map was published in 2021. More information about the FSN E&GM is available on the project page: https://www.3ieimpact.org/research/food-systems-and-nutrition-evidence-and-gap-map

between 2005 and 2023. About 60 percent of the studies analyzed the effect of interventions through experimental design.

Table 1. Distribution of quantitative impact evaluations and systematic reviews considering WE outcomes by intervention-outcome pairing

		Other empowerment outcome	Decision making	Control of resources	Gender transformative outcome	Ownership	Self-esteem	Grand Total
Production system	Education / information - other educational programs	11	14	7	4	5	3	**17**
	Education / information - Agricultural extension programs	12	10	5	5	4	2	**16**
	Livestock access	4	5	1	2	2	1	**9**
	Land markets & mgmt	4	5	1		3		**8**
	Education / information - Farmer field schools	3	1	5	2	1	2	**7**
	Other ag inputs	3	2	2	1	1	2	**7**
	Ag credit / savings	5	2	2	1	1	1	**6**
	Water access/management	3	4	3	2	1	2	**5**
	Improved seeds	3	3	1	1	1	1	**4**
	Market support	2	1	2			1	**3**
	Fertilizer access	1	2		1	1		**2**
	Other production system improvements	2	2	1				**2**
	Education / information - information/guidance	1		1				**2**
	Contract farming		1					**1**
	Provision of mechanical equipment	1	1					**1**
Empowerment	Women's empowerment efforts	25	23	8	9	6		**34**
	Gender transformative intervention	7	5	2	4	1	1	**10**
Behavior change communication	Classes	7	5	3	2	1	2	**11**
	Peer support/counselors	4	5	2	1	1		**9**
	Community meetings	5	4	3	3	2	1	**8**
	Healthy food social marketing campaigns		2	2	1	1		**3**
	Professional services (dietitians/nurses)	1	1	1	1	1	1	**2**
Availability and affordability	Direct provision of foods	8	4	2	1	1	1	**10**
	Provision or use of supplements	3						**3**
	Cash-for-food programs	1	2	1				**3**
Multi-component Program	Large multicomponent intervention		1	1	1	2		**2**
	MC: Peer support & community meetings	2	2	1				**2**
	MC: Classes & healthy food marketing		1					**1**
Distribution and storage	Storage/distribution. education	1	1					**2**
	Trade regulations		1					**1**
	On-farm storage	1	1			1		**1**
Processing and packaging	Fortification	2						**2**
	Processing/packaging education	1	1			1		**1**
Promotion and labeling	Food safety regulations	1						**1**
Food loss and waste management	Composting	1	1			1		**1**
Grand Total		**60**	**57**	**29**	**20**	**18**	**8**	**101**

4. Protocol for a rapid evidence assessment on the effects of FSN interventions on women's empowerment

In this section, we present our approach for an REA that will synthesize and appraise the evidence on the effects of FSN interventions on women's empowerment outcomes. We outline our research questions, the criteria determining the inclusion of studies in this review, and the methods we will use to describe and analyze this evidence.

4.1. Research Questions

Based on the findings of the FSM E&GM, we aim to synthesize and appraise the evidence identified on the effects of FSN interventions on women's empowerment outcomes. This research aims to help promote the wider use and understanding of evidence on what works to strengthen women's empowerment. To support these aims, we will address the following research questions:

- What are the effects of food systems and nutrition interventions on outcomes related to women's empowerment?
- Are there unintended consequences, including adverse effects, of these interventions?
- Do effects vary by context, intervention features, or other moderators?
- What evaluation design strategies are used?

4.2. Inclusion criteria and overview of the body of evidence

REA is a form of evidence synthesis that has been developed to address policy-relevant questions in less time and with fewer resources than what is typically required for full systematic reviews (Ganann, Ciliska, and Thomas 2010; Khangura et al. 2012; Collins et al. 2015; Barends, Rousseau, and Briner 2017; Snilstveit et al. 2018). There is no single definition of a rapid review, and recent analysis of study methods has highlighted the variation in rapid review methods (Hartling et al. 2015; Khangura et al. 2012; Tricco et al. 2015; Fenton Villar 2022). However, such approaches typically involve adjusting traditional systematic review methods and adopting one or more shortcuts to answer urgent questions more promptly (Schünemann and Moja 2015). The approach and methodology below are developed in line with other types of

rigorous evidence synthesis methodologies (Barends, Rousseau, and Briner 2017; Fenton Villar 2022).

4.2.1. Criteria for including and excluding studies

The inclusion criteria for the REA have been established in collaboration with GIZ and the Advisory Group and based on the existing scope and criteria of the FSN E&GM. Table 2 summarizes the type of participants, interventions, comparison, outcomes, and study designs (PICOS) considered in this REA, along with other inclusion criteria. Most of the criteria align with the E&GM described in the previous section, but we have applied some additional restrictions to limit the scope of the exercise to be feasible within the allocated time and budget.

The REA will primarily focus on GIZ's strategic intervention types where enough evidence is available for our analysis. We will exclude studies focusing on women's empowerment interventions (already covered in Berretta et al. 2023), ongoing studies, and studies on interventions measuring women's empowerment effects, which are covered by less than six studies in the existing FSN E&GM (as these will be areas where we will not be able to draw strong conclusions). Applying these criteria will reduce the body of evidence from the 101 identified studies to 60 that will be includable in the REA.

Table 2. Summary of criteria (PICOS) determining study eligibility for the REA

Criteria	Description
Participants	People of any age and gender residing in low- and middle-income countries (L&MICs). Excluding studies targeting participants with a clinical condition
Interventions[3]	Food systems and nutrition interventions within the following categories: - *Production system* - *Behavior change communication* - *Availability and affordability* Descriptions of included interventions are available in Appendix 1
Comparison	Business as usual, including pipeline and waitlist controls and alternative intervention

[3] We define interventions as an activity or a set of activities implemented in real-life settings by individuals or institutions, with the aim of creating a change for the people exposed to it. It then covers both internal or external, national or international, programs and policies implemented at the international, regional, national, or sub-national level. It is then used as a synonym to the following non-exhaustive list: treatment, initiative, program, project, policy, activity, etc.

Outcome	Measures of women's empowerment within the following categories: - *Resources* - *Agency* - *Achievements* Descriptions of included outcomes are available in Appendix 2
Study designs	We include quantitative impact evaluations, qualitative evaluations, and systematic reviews: - For quantitative impact evaluations, we include studies using an experimental or quasi-experimental design. - For qualitative evaluations, we include qualitative studies collecting primary data using mixed-method or qualitative methods, descriptive quantitative studies, and process evaluations focusing on interventions included in the quantitative impact evaluations. It is important to note that the original FSN E&GM did not include qualitative evaluations. Descriptions of included study designs are available in Appendix 3
Language	Studies in English Studies in English
Publication date	All studies were published from 20002000 onwards
Status of studies	We include completed quantitative impact evaluations and qualitative evaluations.
Publication status	We include studies published in any outlet, including peer-reviewed journals, working paper series, organizational reports, and unpublished author manuscripts.

4.3. Method for quantitative data extraction and analysis

4.3.1. Data extraction and coding procedures

We will extract the following data from each study, encompassing the study's context, methods, and findings, along with information about the cost and implementation of the intervention. (Provisional data extraction forms are provided in Appendix 4.)

- Descriptive data includes authors, publication date, status, and other information to characterize the study, including country, category of intervention and outcome, and intervention design.
- Methodological information on study design, analysis method, and type of comparison (if relevant).
- Quantitative data for outcome measures, including descriptions of outcome measures, sample sizes in each intervention and comparison group, the

outcome means, SDs, and test statistics (e.g., *t*-test, F test, *p*-values, 95 percent confidence intervals, if available).

- Qualitative data for the conditions of implementation and main barriers and facilitators to the implementation of the interventions.
- Cost data for the cost associated with the implementation of the intervention.

Descriptive data, methodological information, and cost data will be single coded by a trained reviewer and checked by another second reviewer. Two trained reviewers will independently code the quantitative data, and any disagreement will be resolved through discussion with a third reviewer (who must be a core team member). Before proceeding with independent data extraction, all coders will receive training on quantitative data extraction (QEX) and Risk of Bias assessment (RoB). Only coders meeting the minimum similarity threshold (inter-rater reliability of at least 85%) with the research team will be selected for independent data extraction. Additionally, all QEX and RoB will be reviewed for consistency by the research team, and a sample will be reviewed in detail by the research team to ensure the quality of the data extraction process.

4.3.2. Measures of treatment effects

An effect size (or treatment effect) expresses the direction and magnitude of the difference in outcomes between groups of observations, such as the difference in outcomes between observations in the intervention and comparison groups (Borenstein et al. 2009; Valentine, Aloe, and Lau 2015).

Effect sizes presented in empirical studies are rarely independent of the scale or unit of the outcome in the study, and the scale or unit of the outcome is generally not directly comparable across studies. To facilitate cross-study comparisons of the magnitudes of studies' effects in our analysis, we will extract data from each study to calculate standardized effect sizes (i.e., Cohen's d adjusted to Hedges *g*). We will choose the appropriate formulae for standardized effect size calculations about, and dependent upon, the data provided in the included studies and the outcome category (see Appendix 5 for the effect size formulae list) (Borenstein et al. 2009).

If different outcome categories exist under the same outcome construct, we will convert estimates to the most common standardized metric for comparability of

estimated effect sizes. We will use common transformations outlined by Borenstein and colleagues (2009) for converting between different measures of standardized effects. When studies provide multiple estimates for the same effect (e.g., using different model specifications), we will extract the authors' preferred specifications if they have identified one. Otherwise, we will select the most precise estimate (i.e., the estimate with the smallest standard error).

4.3.3. Criteria for determination of independent findings

Our analysis must accurately capture and account for co-dependencies between study estimates. This is because standard meta-analytic methods assume effect size estimates are independent and fail to qualitatively recognize that estimates derived from the same intervention or study can distort (inflate) our perceptions of the availability of evidence.

Dependent effect sizes can arise in several circumstances. For example, co-dependencies between estimates can arise when several publications stem from one study or several studies are based on the same data set. Some studies might have multiple treatment arms that are all compared to a single control group. Other studies may report outcome measurements from several time points or use multiple outcome measures to assess related outcome constructs. All such cases yield a set of statistically dependent effect size estimates (Borenstein, Hedges, and Rothstein 2009).

We will assess the extent to which relationships exist across the studies included in the review. We will avoid double-counting identical evidence by linking papers prior to data analysis and using the information provided in the studies, such as sample sizes, program characteristics, and key implementing and/or funding partners, to help support these assessments. Where we have several publications reporting on the exact same effect in the same underlying sample, one main study will be used for data extraction, and the linked studies will be stored to help any required search for further or missing information. To identify the main study, priority will be given to the most recent journal article, and, in the case of multiple reports/working papers, the most recent one will be selected.

We will extract effects reported across different interventions, outcomes, and subgroups within a study. We will address dependent effect sizes using data processing and selection techniques. We will utilize several criteria to select one effect estimate per outcome per study (further details of the criteria determining effect estimate selection are available in Appendix 6).

4.3.4. Unit of analysis issues

Unit of analysis errors can arise when the unit of allocation of treatment is different from the unit of analysis of effect size estimate, and this is not accounted for in the analysis (e.g., by clustering standard errors at the level of allocation). We will assess included studies for the prevalence of these issues and, where they exist, account for them by adjusting the reported standard errors (*SE*s) according to the following formula (Borenstein et al. 2009; Hedges 2009):

$$(d)' = (d) \cdot 1 + (m-1)c$$

Where *d* is the effect size, m is the average number of observations per cluster, and c is the intra-cluster correlation coefficient. If the included studies use robust Huber-White *SE*s to correct for clustering, we will calculate the *SE* of *d* by dividing *d* by the t-statistic on the coefficient of interest.

4.3.5. Dealing with missing data

In instances where there is missing or incomplete data, we will make every effort to contact study authors to obtain the required information. If we are unable to obtain the necessary data, we will report the characteristics of the study but state that it could not be included in the meta-analysis or reporting of effect sizes due to missing data. In line with recommendations on collating data in systematic reviews from study authors (see Mullan et al. 2009), we will report the number of studies for which authors were contacted, the information requested, any important details of the method of eliciting information, and the response of authors to the request.

4.3.6. Critical appraisals

We will assess the risk of bias in included studies using 3ie's risk of bias tool (see Appendix 7). This examines both the internal validity and statistical conclusion validity of experimental and quasi-experimental impact evaluation designs (Waddington et al.

2012). Two reviewers will undertake the risk of bias assessment independently. If there are disagreements, we will resolve them by discussion and the involvement of a third reviewer (who must be a member of the core team). We will compile a risk of bias assessment for each estimate we extract. This is to account for the fact that estimates for different outcomes in the same study may score differently in the assessment.

We will assess the risk of bias based on the following criteria by answering whether the estimate is *free from* each bias, with a response set of "Yes," "Probably Yes," "Probably No," "No," and "No Information" for each domain:

- Factors relating to baseline confounding and biases arising from differential selection into and out of the study (e.g., assignment mechanism).
- Factors relating to bias due to missing outcome data (e.g., assessment of attrition).
- Factors relating to biases due to spillovers, crossovers, and contamination.
- Factors relating to biases in outcome measurement (e.g., social desirability or courtesy bias, recall bias).
- Factors relating to biases in reporting of analysis.

We will report the results of the assessment for each of the assessed criteria for each estimate. In addition, we will use the results of the risk of bias assessments to produce an overall rating for each study as either "High risk of bias," "Some concerns," or "Low risk of bias," drawing on the decision rules in RoB2.0 (Sterne et al. 2019), rating studies as follows:

- "High risk of bias": if any of the bias domains were assessed as "No" or "Probably No."
- “Some concerns”: if one or several domains were assessed as “No Information," and none were "No" or "Probably No."
- "Low risk of bias": if all of the bias domains were assessed as "Yes" or "Probably Yes."

We will describe our analysis of the outcomes of our assessment of the reliability of included studies, and we also intend to explore whether there are systematic differences in estimated effects between primary studies with different risks of bias.

We will conduct a sensitivity analysis to assess the robustness of the results to the risk of bias associated with included studies.

4.3.7. Data synthesis

To synthesize the effects of interventions, we will combine a narrative synthesis of study findings with a cross-study meta-analyses of intervention effects. Our narrative synthesis will examine the range of intervention effects and the study settings. We will then include studies in the same meta-analysis when we identify two or more effect sizes using a similar outcome construct, the same intervention type, and where the type of comparison group is judged to be similar across the studies. If there are too few studies, or the included studies are considered too heterogeneous in terms of interventions or outcomes, we will present a narrative discussion of individual effect sizes alone (Wilson, Weisburd, and McClure 2011).

Because heterogeneity exists in theory due to the variety of interventions and contexts that could be included in the review, we will use inverse-variance weighted, random effects meta-analytic models to synthesize the effect estimates (Higgins et al. 2020). We will use the metafor package (Viechtbauer 2010) in R software to conduct the meta-analyses (R Core Team 2018).

4.3.8. Sub-group analysis and investigation of heterogeneity

In our analysis, we will examine and discuss the distribution of estimated effects across intervention and outcome types. We will also statistically assess heterogeneity by calculating the Q statistic, I^2, and τ^2 to provide an estimate of the amount of variability in the distribution of study effect sizes (Borenstein et al. 2009). We will complement this assessment with a graphical analysis using forest plots to illustrate the range of the standardized effects by intervention. This approach provides a summary effect estimate with studies weighted by the precision of the estimate using the inverse of the variance. Whenever feasible, we will conduct moderator analyses using meta-regression to investigate potential sources of heterogeneity.

Following the PROGRESS-PLUS approach (Oliver et al. 2017), we will assess moderators falling into three broad categories of extrinsic, methodological, and substantive characteristics. Examples of these categories include:

- Extrinsic characteristics: E.g., funder of the study (e.g., NGO vs private sector vs government investments), publication type, publication date.
- Methodological characteristics: E.g., study design, risk of bias, length of follow-up, categories of outcome measures.
- Substantive characteristics: E.g., participant characteristics (gender, age, socio-economic status, education), context (geographical setting; democratic setting), gender and/or hunger-related indicators (GHI, GII, etc.), intervention type, intervention features, type of implementing agency).

We intend to use random effects meta-regression to investigate the association between moderator variables and heterogeneity of treatment effects (Borenstein et al. 2009) and subgroup analyses to investigate heterogeneity by treatment subgroups (e.g., men and women, poor and non-poor, and so on). If these strategies are not possible (e.g., if we do not have a sufficient number of studies or data), we will discuss and explore the factors that may be driving the heterogeneity of results narratively by conducting cross-case comparisons (Miles and Huberman 1994).

4.3.9. Sensitivity analysis

We will conduct a sensitivity analysis to assess whether the results of the meta-analysis are sensitive to the removal of any single study. We will do this by removing studies from the meta-analysis one by one and assessing changes in results. We will also assess the sensitivity of our results to the inclusion of studies with a high risk of bias by removing these studies from the meta-analysis and comparing results to the main meta-analysis results. Furthermore, we will assess the sensitivity of our results to outliers. We will use studentized residuals to examine whether studies' estimated effects may be outliers (Viechtbauer and Cheung 2010), and studies with a studentized residual larger than the $100 \times (1 - 0.05/(2 \times k))$th percentile of a standard normal distribution will be considered potential outliers.

We also used both bivariate and multivariate (or partial) effects for calculating standardized effect sizes. A partial effect size is based on a regression coefficient measuring the treatment effect "holding all other variables constant" and is, therefore, measuring a different quantity to a bivariate relationship. Our standardized effect sizes are only strictly comparable in studies using a common model (Keef and Roberts 2004). However, only using bivariate effect sizes to calculate standardized effects

would not be suitable in this context due to the likely high risk of bias. This may cause quasi-experimental study designs that control selection bias (Waddington et al. 2014). We will use sensitivity analysis to examine systematic differences in partial and bivariate effects (omitting them from the analysis or controlling for these characteristics in a meta-regression).

Finally, not all multivariate models control for the same covariates, nor should models estimated for different study designs using data collected in different contexts necessarily do so. The risk of bias assessment evaluates likely specification errors, and the sensitivity analysis omitting high-risk studies (discussed above) should capture most of these issues. Otherwise, we assume the possible resulting multicollinearity issues are inconsequential (see H. Waddington et al. 2014).

4.3.10. Assessment of reporting biases

If meta-analysis is feasible, we will assess reporting biases in the literature using a rank correlation test (see Begg and Mazumdar 1994) when there are at least 10 studies contributing to the analysis. We will also use a regression test (Sterne and Egger 2005), using the standard error of the observed outcomes as a predictor, to check for funnel plot asymmetry.

4.4. Qualitative search and appraisal

4.4.1. Purpose of the qualitative analysis

For the review to be more beneficial for policymakers and practitioners, we will collect and analyze qualitative evidence to assess factors that enhance or hinder the effectiveness of interventions. We will investigate the context, configuration of different conditions of the intervention and the barriers and facilitators to intervention effectiveness (Thomas, O'Mara-Eves, and Brunton 2014). This will inform:

- how context, population characteristics and other factors impact the effectiveness of the included interventions
- to what extent effects vary by these factors (such as context, population characteristics, or other factors) and types of interventions
- the identification of potential unintended effects

4.4.2. Source and eligibility of qualitative evidence

Qualitative evidence is not included in the FSN E&GM. Thus, we will undertake an additional search for qualitative evidence. Qualitative evidence will be gathered from two sources:

- Descriptive and qualitative data from the included experimental and quasi-experimental studies from the included studies.
- A targeted search for additional papers on the interventions covered by the included impact evaluations to provide additional detail on these factors.

In order to be included, these papers must be related to the interventions in the included quantitative impact evaluations and also be one or more of the following types of studies:

- A *qualitative study* collecting primary data using qualitative or quantitative methods of data collection and analysis and reporting some information on all of the following: the research question, procedures for collecting data, procedures for analyzing data, and information on sampling and recruitment, including at least two sample characteristics.
- A *descriptive quantitative* study collecting primary data using quantitative methods of data collection and descriptive quantitative analysis and report some information on all of the following: the research question, procedures for collecting data, procedures for analyzing data, and information on sampling and recruitment, including at least two sample characteristics.
- A *process evaluation* assessing whether an intervention is being implemented as intended, what is felt to be working more or less well, and why. Process evaluations may include the collection of qualitative and quantitative data from different stakeholders to cover subjective issues, such as perceptions of intervention success, or more objective issues, such as how an intervention was operationalized. They might also be used to collect organizational information.
- A *project document* providing information about planned, ongoing, or completed interventions. Such documents may describe the background and design of an intervention or the resources available for a project. As such, these documents do not typically include much analysis of primary evidence, but they provide factual information about interventions. The purpose of including them

in our review is to ensure we will have sufficient information about the context and interventions in the included studies.

4.4.3. Search process

Search will begin by identifying the program names from included quantitative studies, followed by a citation search, and searching in the web browser and the implementer, funder or project websites and publications for linked qualitative evaluations. Relevant hits include, but are not limited to, reports, project documents, and web pages, as listed below. Key documents should fall within our inclusion criteria and PICOS:

- Project websites, including hosted by the implementer and donor.
- Qualitative research reports on the relevant project.
- Additional grey literature or reports.

4.4.4. Appraisal of qualitative evidence

We will assess the quality of included qualitative studies, process evaluations, and descriptive quantitative studies using a mixed-methods appraisal tool developed by Langer and colleagues (2016) and applied by Snilstveit and colleagues (2018). This tool builds on the Critical Appraisal Skills Program checklist (CASP 2011) and Pluye and colleagues' (2011) mixed-methods appraisal tool. Using our appraisal tool, we will make judgments on the adequacy of reporting, data collection, presentation, analysis, and conclusions drawn. We will assess the quality of the included qualitative studies and descriptive quantitative studies using six appraisal domains:

- The defensibility of the applied research design to answer the research question under investigation.
- The defensibility of the selected research sample and the process of selecting research participants.
- The rigor of the technical research conducted, including the transparency of reporting.
- The rigor of the applied analysis and credibility of the study's claims given the nature of the presented data.
- The consideration of the study's context (for qualitative studies only).
- The reflexivity of the reported research (for qualitative studies only).

We will not undertake a critical appraisal of the included project documents. They typically provide information about planned, ongoing, or completed programs, providing information about the design or resources available for a project, for instance. As such, these documents do not typically include much analysis of primary evidence, but they provide information about interventions. The purpose of including them in our review is to ensure we have sufficient information about the context and interventions included in our review.

For the rest of the qualitative studies, we will filter out studies of particularly low quality at this stage, using a fatal flaw approach following Dixon-Woods and colleagues (2005). Studies that do not meet either criterion of appraisal domains 1–4 above will be excluded from the synthesis. That is, they will be included in the review, and we will report on the studies' descriptive data, for example, applied intervention. However, no research findings will be extracted from these studies to feed into the review's synthesis. Each appraisal domain will be assessed on a scale of critical trustworthiness to low, medium, and high trustworthiness. We will allocate an overall critical appraisal judgment per study using a numerical threshold of the appraised quality domains.

We will, therefore, focus the appraisal on assessing the relevance of the documents against the interventions assessed in our review. Before extracting any data, we will ensure that the name of the intervention, the implementing agency, the context, and the timeline of the intervention described in the project document correspond to the intervention assessed in the impact evaluation included in our review. Finally, collecting data from a range of sources, especially if used for triangulation, can enhance confidence in the trustworthiness of the information included. If several sources are available, we will extract data from all sources for purposes of triangulation.

4.4.5. Data extraction and thematic coding

We will use computer-aided qualitative data analysis tools for thematic synthesis. Themes will closely follow these domains:

- *Context:* Any element related to external factors beyond the program's control that affects program impact. This can refer to, e.g., political factors such as type

of governance, societal factors such as norms, economic factors such as a recession, and cultural factors such as beliefs.

- *Intervention design:* Any variable that is related to the design and planning of the applied intervention. Design and planning of an intervention refers to the blueprint or schedule of the intervention and will typically outline what components the intervention consists of and in what sequence they will be applied. Examples of design variables refer to the size or type of cash transfer and outreach strategy, e.g., posters, reminders, and type of training.
- *Intervention implementation:* Any variable that is related to the implementation of the intervention in practice. This refers to variables that emerge while the intervention is applied and are usually not known in advance. Examples of implementation variables refer to the lack of attendance or uptake, payment difficulties, corruption, and elite capture.
- *Population characteristics:* Any variable that is related to the population targeted by the intervention or the population in which the effects are measured (in cases where these differ). This can refer to, e.g., the socio-economic status of the population, its educational status, and asset ownership. It is important not to confuse this with sample characteristics, where these variables might be reported to describe the composition of the study sample and only to look for data on how these characteristics might have influenced the program effects.
- *Influence on program effect:* Data that report on barriers and facilitators associated with the context, intervention design, implementation, and population characteristics that influence program effects.

4.4.6. Synthesis of qualitative information

The process of synthesis involves generating descriptive and analytic themes derived from the line-by-line coding for each of the four sections (population, design, implementation, and contextual interplay with effects).

The first data point of interest is the frequency reports. These are important to get an overall idea of the most common child codes and those related to each other, as well as those that are bounded by one study (unlikely to make it into the descriptive and thematic analysis) and themes across every four categories per intervention.

For each intervention, we will generate coding reports and compare and compile codes that can be merged together based on their thematic proximity.

We will assess the quality of included qualitative studies, process evaluations, and descriptive quantitative studies using a mixed-methods appraisal tool developed by Langer and colleagues (2017) and applied by Snilstveit and colleagues (2018). This tool is in Appendix 8.

4.4.7. Data presentation

A thematic synthesis will complement the meta-analysis conducted with the quantitative data. We will provide a narrative summary of the papers identified. This will include an overall description of the available literature and a general synthesis of findings. Key information from each study, such as intervention type, study design, country, outcomes, measurement type, effect sizes, and confidence rating, will be summarized in a table. Along with results from meta-analyses, we will narratively summarize qualitative information focusing on informing project design and implementation.

Acknowledgement

We are grateful to 3ie colleagues who provided inputs at various stages of the REA protocol and supported its administrative, communication and finance management. We are particularly grateful to the GIZ advisors and Julia Hoigt, for their support in the production of this review. We acknowledge and appreciate the contribution of the advisory group for this E&GM and REA protocol: Adam Trowbridge, Senior Evaluation Advisor at USAID; Agnes Quisumbing, Senior Research Fellow at International Food Policy Research Institute (IFPRI); Andrea Menefee, Deputy Director at CARE international; Bridget Fenn, Technical Associate at the Emergency Nutrition Network; Cheryl Doss, Professor at Tufts University; Geeta Sethi, Advisor at World Bank; Hazel Malapit, Senior Research Coordinator at International Food Policy Research Institute (IFPRI); Jessica Fanzo, Professor at Columbia University; Marie Ruel, Director at International Food Policy Research Institute (IFPRI); Michele Tarsilla, Chief of Humanitarian Evaluation at UNICEF; Sara Wuehler, Senior Technical Advisor at Nutrition International; Saul Morris, Director of Program Services at Global Alliance for Improved Nutrition; Thalia Sparling, Assistant Professor at London School of Hygiene and Tropical Medicine; Valentine Granet, Gender Advisor at Global Alliance for Improved Nutrition; Zlata Bruckauf, Senior Evaluation Specialist at UNICEF . We also acknowledge and appreciate the contribution and comments of the independent reviewers.

References

Ahmed, Akhter, John F. Hoddinott, and Shalini Roy. 2019. “Food Transfers, Cash Transfers, Behavior Change Communication and Child Nutrition: Evidence from Bangladesh.” 0 ed. Washington, DC: International Food Policy Research Institute. https://doi.org/10.2499/p15738coll2.133420.

Alaofè, Halimatou, Jennifer Burney, Rosamond Naylor, and Douglas Taren. 2019. “The Impact of a Solar Market Garden Programme on Dietary Diversity, Women’s Nutritional Status and Micronutrient Levels in Kalalé District of Northern Benin.” Public Health Nutrition 22 (14): 2670–81. https://doi.org/10.1017/S1368980019001599.

Alkire, Sabina, Ruth Meinzen-Dick, Amber Peterman, Agnes R Quisumbing, Greg Seymour, and Ana Vaz. 2013. “The Women’s Empowerment in Agriculture Index.” OPHI WORKING PAPER 58. https://www.ophi.org.uk/wp-content/uploads/ophi-wp-58.pdf.

ANH Academy. 2024. “IMMANA | ANH Academy.” 2024. https://www.anh-academy.org/immana.

Backiny-Yetna, Prospere, Kevin McGee, and Kevin Robert Mcgee. 2015. “Gender Differentials and Agricultural Productivity in Niger.” SSRN Scholarly Paper. Rochester, NY. https://papers.ssrn.com/abstract=2568933.

Bandiera, Oriana, Robin Burgess, Narayan Das, Selim Gulesci, Imran Rasul, and Munshi Sulaiman. 2017. “Labor Markets and Poverty in Village Economies*.” The Quarterly Journal of Economics 132 (2): 811–70. https://doi.org/10.1093/qje/qjx003.

Barbour, R. S. 1998. “Mixing Qualitative Methods: Quality Assurance or Qualitative Quagmire?” Qualitative Health Research 8 (3): 352–61. https://doi.org/10.1177/104973239800800306.

Barends, Eric, Denise M. Rousseau, and Rob B. Briner. 2017. “CEBMa Guideline for Rapid Evidence Assessments in Management and Organizations (Version 1.0).” Center for Evidence Based Management. https://cebma.org/assets/Uploads/CEBMa-REA-Guideline.pdf.

Bastagli, Francesca, Jessica Hagen-Zanker, Luke Harman, Valentina Barca, Georgina Sturge, and Tanja Schmidt. 2019. “The Impact of Cash Transfers: A Review of the Evidence from Low- and Middle-Income Countries.” Journal of Social Policy 48 (3): 569–94. https://doi.org/10.1017/S0047279418000715.

Bastagli, Francesca, Jessica Hagen-Zanker, Luke Harman, Valentina Barca, Georgina Sturge, Tanja Schmidt, and Luca Pellerano. 2016. Cash Transfers: What Does the Evidence Say? A Rigorous Review of Programme Impact and

of the Role of Design and Implementation Features. https://doi.org/10.13140/RG.2.2.29336.39687.

Begg, C. B., and M. Mazumdar. 1994. “Operating Characteristics of a Rank Correlation Test for Publication Bias.” Biometrics 50 (4): 1088–1101.

Berretta, Miriam, Meital Kupfer, Shannon Shisler, and Charlotte Lane. 2023. “Rapid Evidence Assessment on Women’s Empowerment Interventions within the Food System: A Meta-Analysis.” Agriculture & Food Security 12 (1): 13. https://doi.org/10.1186/s40066-023-00405-9.

Blackmore, Ivy, Carolyn Lesorogol, and Lora Iannotti. 2018. “Small Livestock and Aquaculture Programming Impacts on Household Livelihood Security: A Systematic Narrative Review.” Journal of Development Effectiveness 10 (2): 197–248. https://doi.org/10.1080/19439342.2018.1452777.

BMGF. 2019. “What Gets Measured Matters: A Methods Note for Measuring Women and Girls’ Empowerment.” Bill and Melinda Gates Foundation. https://www.gatesgenderequalitytoolbox.org/wp-content/uploads/BMGF_Methods-Note-Measuring-Empowerment-1.pdf.

Bold, Mara van den, Agnes R. Quisumbing, and Stuart Gillespie. 2013. “Women’s Empowerment and Nutrition: An Evidence Review.” 1294. IFPRI Discussion Paper. Washington, D.C.: International Food Policy Research Institute (IFPRI). https://ebrary.ifpri.org/digital/collection/p15738coll2/id/127840.

Bonuedi, Isaac, Lukas Kornher, and Nicolas Gerber. 2020. “Making Cash Crop Value Chains Nutrition-Sensitive: Evidence from a Quasi-Experiment in Rural Sierra Leone.” SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3603918.

Booth, Andrew. 2001. “Cochrane or Cock-Eyed? How Should We Conduct Systematic Reviews of Qualitative Research?” Proceedings of the Qualitative Evidence-Based Practice Conference, Taking a Critical Stance, May.

———. 2016. “Searching for Qualitative Research for Inclusion in Systematic Reviews: A Structured Methodological Review.” Systematic Reviews 5 (1): 74. https://doi.org/10.1186/s13643-016-0249-x.

Borenstein, M, LV Hedges, JPT Higgins, and HR Rothstein. 2009. “Multiple Comparisons within a Study.” In Introduction to Meta-Analysis, John Wiley&Sons, Ltd, 239-242. Ch.25. Chichester, U.K.

Borenstein, M, LV Hedges, and HR Rothstein. 2009. Introduction to Meta-Analysis. Chichester, UK: John Wiley & Sons, Ltd.

BRAC. 2016. “Targeting the Ultra Poor: Programme Brief 2016.” https://www.brac.net/images/index/tup/brac_TUP-briefNote-Jun17.pdf.

Braun, Joachim von, Kaosar Afsana, Louise O. Fresco, and Mohamed Hag Ali Hassan. 2023. “Food Systems: Seven Priorities to End Hunger and Protect the Planet.” In Science and Innovations for Food Systems Transformation, 3–9. Springer International Publishing Cham. https://library.oapen.org/bitstream/handle/20.500.12657/61280/978-3-031-15703-5.pdf?sequence=1#page=23.

Brauw, Alan de, Marrit Van den berg, Inge D. Brouwer, Harriette Snoek, Raffaele Vignola, Mequanint B. Melesse, Gaia Lochetti, et al. 2019. “Food System Innovations for Healthier Diets in Low and Middle-Income Countries.” IFPRI Discussion Paper 01816. https://books.google.co.uk/books?hl=en&lr=&id=CvqNDwAAQBAJ&oi=fnd&pg=PR2&ots=ru9sKc5Jy7&sig=rVXfxOzIeRyGlDvG09xLK60KHdc&redir_esc=y#v=onepage&q&f=false.

Britten, Nicky, Rona Campbell, Catherine Pope, Jenny Donovan, Myfanwy Morgan, and Roisin Pill. 2002. “Using Meta Ethnography to Synthesise Qualitative Research: A Worked Example.” Journal of Health Services Research & Policy 7 (4): 209–15. https://doi.org/10.1258/135581902320432732.

Brown, Elizabeth D., and Jeffery C. Tanner. 2019. Integrating Value for Money and Impact Evaluations: Issues, Institutions, and Opportunities. World Bank, Washington, DC. https://doi.org/10.1596/1813-9450-9041.

Campbell, Rona, Pandora Pound, Catherine Pope, Nicky Britten, Roisin Pill, Myfanwy Morgan, and Jenny Donovan. 2003. “Evaluating Meta-Ethnography: A Synthesis of Qualitative Research on Lay Experiences of Diabetes and Diabetes Care.” Social Science & Medicine (1982) 56 (4): 671–84. https://doi.org/10.1016/s0277-9536(02)00064-3.

Candy, Bridget, Michael King, Louise Jones, and Sandy Oliver. 2011. “Using Qualitative Synthesis to Explore Heterogeneity of Complex Interventions.” BMC Medical Research Methodology 11 (1): 124. https://doi.org/10.1186/1471-2288-11-124.

CASP. 2011. “Critical Appraisal Checklists.” CASP - Critical Appraisal Skills Programme. 2011. https://casp-uk.net/casp-tools-checklists/.

“Casp-Qualitative-Studies-Checklist-Fillable.Pdf.” 2018. 2018. https://casp-uk.net/checklists/casp-qualitative-studies-checklist-fillable.pdf.

Chaplin, Duncan D., Thomas D. Cook, Jelena Zurovac, Jared S. Coopersmith, Mariel M. Finucane, Lauren N. Vollmer, and Rebecca E. Morris. 2018. “The Internal and External Validity of the Regression Discontinuity Design: A Meta-Analysis

of 15 Within-Study Comparisons." Edited by Burt S. Barnow. Journal of Policy Analysis and Management 37 (2): 403–29. https://doi.org/10.1002/pam.22051.

Collins, Alexandra, Deborah Coughlin, James Miller, and Stuart Kirk. 2015. "The Production of Quick Scoping Reviews and Rapid Evidence Assessments A How to Guide."

De La Rue, Lisa, Joshua R. Polanin, Dorothy L. Espelage, and Terri D. Pigott. 2013. "PROTOCOL: School-Based Interventions to Reduce Dating and Sexual Violence: A Systematic Review." Campbell Systematic Reviews 9 (1): 1–43. https://doi.org/10.1002/CL2.106.

Dixon-Woods, M., and R. Fitzpatrick. 2001. "Qualitative Research in Systematic Reviews. Has Established a Place for Itself." BMJ (Clinical Research Ed.) 323 (7316): 765–66. https://doi.org/10.1136/bmj.323.7316.765.

Dixon-Woods, Mary, Shona Agarwal, David Jones, Bridget Young, and Alex Sutton. 2005. "Synthesising Qualitative and Quantitative Evidence: A Review of Possible Methods." Journal of Health Services Research & Policy 10 (1): 45–53. https://doi.org/10.1177/135581960501000110.

Ellis, Paul D. 2010. The Essential Guide to Effect Sizes: Statistical Power, Meta-Analysis, and the Interpretation of Research Results. 1st ed. Cambridge University Press. https://doi.org/10.1017/CBO9780511761676.

Estabrooks, Carole A., Peggy Anne Field, and Janice M. Morse. 1994. "Aggregating Qualitative Findings: An Approach to Theory Development." Qualitative Health Research 4 (4): 503–11. https://doi.org/10.1177/104973239400400410.

European Union. 2014. "Productive Safety Net Programme in Ethiopia." 2014. https://capacity4dev.europa.eu/projects/project_psnp_ethiopia_en.

Evans, David K, and Anna Popova. 2014. "Cash Transfers and Temptation Goods: A Review of Global Evidence."

Evans, David K. 2021. "Towards Improved and More Transparent Ethics in Randomised Controlled Trials in Development Social Science."

FAO, ed. 2022. Repurposing Food and Agricultural Policies to Make Healthy Diets More Affordable. The State of Food Security and Nutrition in the World 2022. Rome: FAO. https://doi.org/10.4060/cc0639en.

———, ed. 2023a. Urbanization, Agrifood Systems Transformation and Healthy Diets across the Rural-Urban Continuum. The State of Food Security and Nutrition in the World 2023. Rome: FAO. https://doi.org/10.4060/cc3017en.

———. 2023b. "The Status of Women in Agrifood Systems." FAO. https://doi.org/10.4060/cc5060en.

———. 2024. “Hunger.” Food and Agriculture Organization of the United Nations. 2024. http://www.fao.org/hunger/en/.

FAO, IFAD. 2020. Gender Transformative Approaches for Food Security, Improved Nutrition and Sustainable Agriculture: A Compendium of Fifteen Good Practices. Food and Agriculture Organization of the United Nations 2020.

FAO, IFAD, UNICEF, WFP, and WHO. 2023. The State of Food Security and Nutrition in the World 2023. FAO; IFAD; UNICEF; WFP; WHO; https://doi.org/10.4060/cc3017en.

Feed the future. 2022. “Empowering Women for Food Security.” Feed the Future. https://cg-281711fb-71ea-422c-b02c-ef79f539e9d2.s3.us-gov-west-1.amazonaws.com/uploads/2022/08/Womens-Empowerment-Fact-Sheet-1.pdf.

Fenton Villar, Paul. 2022. “Structured Literature Reviews: Building Transparency and Trust in Standards of Reporting Evidence.” World Bank, IEG Methods and Evaluation Capacity Development Working Paper Series, . https://ieg.worldbankgroup.org/sites/default/files/Data/Evaluation/files/SLR_2022.pdf.

Flemming, Kate. 2010. “Synthesis of Quantitative and Qualitative Research: An Example Using Critical Interpretive Synthesis.” Journal of Advanced Nursing 66 (1): 201–17. https://doi.org/10.1111/j.1365-2648.2009.05173.x.

Florizone, Richard, and Carin Smaller. 2020. “For an Extra USD 14 Billion per Year, We Can End Hunger, Double Incomes for Poorest Farmers, and Meet Climate Goals.” International Institute for Sustainable Development (blog). 2020. https://www.iisd.org/articles/insight/extra-usd-14-billion-year-we-can-end-hunger-double-incomes-poorest-farmers-and.

France, Emma F., Isabelle Uny, Nicola Ring, Ruth L. Turley, Margaret Maxwell, Edward A. S. Duncan, Ruth G. Jepson, Rachel J. Roberts, and Jane Noyes. 2019. “A Methodological Systematic Review of Meta-Ethnography Conduct to Articulate the Complex Analytical Phases.” BMC Medical Research Methodology 19 (1): 35. https://doi.org/10.1186/s12874-019-0670-7.

FSIN. (2023). *2023 Global Report on Food Crises*. https://www.fsinplatform.org/sites/default/files/resources/files/GRFC2023-compressed.pdf

GAAP 2. 2024. “Gender, Agriculture, and Assets Project – An IFPRI Info Site.” 2024. https://gaap.ifpri.info/.

Ganann, Rebecca, Donna Ciliska, and Helen Thomas. 2010. “Expediting Systematic Reviews: Methods and Implications of Rapid Reviews.” Implementation Science 5 (1): 56. https://doi.org/10.1186/1748-5908-5-56.

Glewwe, Paul, and Petra Todd. 2022. Impact Evaluation in International Development: Theory, Methods, and Practice. The World Bank. https://doi.org/10.1596/978-1-4648-1497-6.

Goulart, Céline M., Amber Purewal, Humaira Nakhuda, Anita Ampadu, Amanda Giancola, Jean-Luc Kortenaar, and Diego G. Bassani. 2021. “Tools for Measuring Gender Equality and Women’s Empowerment (GEWE) Indicators in Humanitarian Settings.” Conflict and Health 15 (1): 39. https://doi.org/10.1186/s13031-021-00373-6.

Grebmer, Klaus von, Jll Bernstein, Wendy Geza, and Ndlovu. 2023. “2023 Global Hunger Index: The Power of Youth in Shaping Food Systems.”

Hansen, Henrik, Ninja Ritter Klejnstrup, and Ole Winckler Andersen. 2013. “A Comparison of Model-Based and Design-Based Impact Evaluations of Interventions in Developing Countries.” American Journal of Evaluation 34 (3): 320–38. https://doi.org/10.1177/1098214013476915.

Haque, Md Ahshanul, Nuzhat Choudhury, S. M. Tanvir Ahmed, Fahmida Dil Farzana, Mohammad Ali, Farina Naz, Mohammad Jyoti Raihan, et al. 2022. “The Large-scale Community-based Programme ‘ Suchana’ Improved Maternal Healthcare Practices in North-eastern Bangladesh: Findings from a Cluster Randomized Pre-post Study.” Maternal & Child Nutrition 18 (1): e13258. https://doi.org/10.1111/mcn.13258.

Harden, Angela, James Thomas, Margaret Cargo, Janet Harris, Tomas Pantoja, Kate Flemming, Andrew Booth, Ruth Garside, Karin Hannes, and Jane Noyes. 2018. “Cochrane Qualitative and Implementation Methods Group Guidance Series-Paper 5: Methods for Integrating Qualitative and Implementation Evidence within Intervention Effectiveness Reviews.” Journal of Clinical Epidemiology 97 (May): 70–78. https://doi.org/10.1016/j.jclinepi.2017.11.029.

Hartling, Lisa, Jeanne-Marie Guise, Elisabeth Kato, Johanna Anderson, Naomi Aronson, Suzanne Belinson, Elise Berliner, et al. 2015. EPC Methods: An Exploration of Methods and Context for the Production of Rapid Reviews. AHRQ Comparative Effectiveness Reviews. Rockville (MD): Agency for Healthcare Research and Quality (US). http://www.ncbi.nlm.nih.gov/books/NBK274092/.

Heaton, Janet. 1998. “Secondary Analysis of Qualitative Data.” Social Research Update, no. 22. https://sru.soc.surrey.ac.uk/SRU22.html.

Heckert, Jessica, Deanna K. Olney, and Marie T. Ruel. 2019. “Is Women’s Empowerment a Pathway to Improving Child Nutrition Outcomes in a Nutrition-Sensitive Agriculture Program?: Evidence from a Randomized Controlled Trial in Burkina Faso.” Social Science & Medicine 233 (July): 93–102. https://doi.org/10.1016/j.socscimed.2019.05.016.

Hedges, LV. 2009. “Stochastically Dependent Effect Sizes.” In The Handbook of Research Synthesis and Meta-Analysis, Russell Sage Foundation, pp.281-298. Ch.13. H Cooper, LV Hedges and JC Valentine. New York.

Heidari-Beni, Motahar. 2019. “Early Life Nutrition and Non Communicable Disease.” In Primordial Prevention of Non Communicable Disease, edited by Roya Kelishadi, 1121:33–40. Advances in Experimental Medicine and Biology. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-10616-4_4.

Higgins, J, J Thomas, J Chandler, M Cumpston, P Li, M Page, and V Welch. 2020. Cochrane Handbook for Systematic Reviews of Interventions. Version 6.1 (updated September 2020). Cochrane. https://training.cochrane.org/handbook.

HLPE. 2017. “Nutrition and Food Systems. A Report by the High Level Panel of Experts on Food Security and Nutrition of the Committee on World Food Security.” Rome. https://www.fao.org/policy-support/tools-and-publications/resources-details/en/c/1155796/.

IFIAD. 2021. “Women’s Empowerment and Nutrition: Making the Connections.” Irish Forum for International Agricultural Development. https://www.ifiad.ie/wp-content/uploads/2021/06/Policy-Brief-on-Webinar-3.pdf.

International Initiative for Impact Evaluation (3ie). 2020. “Understanding the Current and Potential Role of Self-Help Group Federations.” 2020th ed. International Initiative for Impact Evaluation (3ie). https://doi.org/10.23846/B/WP/202008.

International Initiative for Impact Evaluation (3ie), Nick Moore, Charlotte Lane, International Initiative for Impact Evaluation (3ie), Ingunn Storhaug, International Initiative for Impact Evaluation (3ie), Amber Franich, et al. 2021. “The Effects of Food Systems Interventions on Food Security and Nutrition Outcomes in Low- and Middle-Income Countries.” 2021st ed. International Initiative for Impact Evaluation (3ie). https://doi.org/10.23846/EGM016.

Ishfaq, Sidra, Abedullah Anjum, Shahzad Kouser, Glenna Nightingale, and Ruth Jepson. 2022. “The Relationship between Women’s Empowerment and Household Food and Nutrition Security in Pakistan.” PLOS ONE 17 (10): e0275713. https://doi.org/10.1371/journal.pone.0275713.

Jensen, Louise A., and Marion N. Allen. 1996. “Meta-Synthesis of Qualitative Findings.” Qualitative Health Research 6 (4): 553–60. https://doi.org/10.1177/104973239600600407.

JP RWEE. 2024. “What We Do | Joint Programme on Accelerating Progress Towards Rural Women’s Economic Empowerment - JP RWEE.” 2024. https://www.jprwee.org/what-we-do.

Kabeer, Naila. 1999. “Resources, Agency, Achievements: Reflections on the Measurement of Women’s Empowerment.” Development and Change 30 (3): 435–64. https://doi.org/10.1111/1467-7660.00125.

Keef, Stephen P., and Leigh A. Roberts. 2004. “The Meta-Analysis of Partial Effect Sizes.” British Journal of Mathematical and Statistical Psychology 57 (1): 97–129. https://doi.org/10.1348/000711004849303.

Khangura, Sara, Kristin Konnyu, Rob Cushman, Jeremy Grimshaw, and David Moher. 2012. “Evidence Summaries: The Evolution of a Rapid Review Approach.” Systematic Reviews 1 (1): 10. https://doi.org/10.1186/2046-4053-1-10.

Kilic, Talip, Amparo Palacios-López, and Markus Goldstein. 2015. “Caught in a Productivity Trap: A Distributional Perspective on Gender Differences in Malawian Agriculture.” World Development 70 (June): 416–63. https://doi.org/10.1016/j.worlddev.2014.06.017.

Langer, Laurenz, Janice Tripney, and David Gough. 2016. “The Science of Using Science: Researching the Use of Research Evidence in Decision-Making.” EPPI-Centre, Social Science Research Unit, UCL Institute of Education, University College London. https://discovery.ucl.ac.uk/id/eprint/1493171/1/Science%20of%20Using%20Science%20Final%20Report%20April%202016.pdf.

Linden, Ariel. 2015. “Conducting Interrupted Time-Series Analysis for Single- and Multiple-Group Comparisons.” The Stata Journal: Promoting Communications on Statistics and Stata 15 (2): 480–500. https://doi.org/10.1177/1536867X1501500208.

Lombardini, Simone, Kimberly Bowman, and Rosa Garwood. 2017. “A ‘How to’ Guide to Measuring Women’s Empowerment: Sharing Experience from Oxfam’s Impact Evaluations.” Oxfam GB. https://oxfamilibrary.openrepository.com/handle/10546/620271.

L.Paterson, Barbara, Sally E.Thorne, Connie Canam, and Carol Jillings. 2001. Meta-Study of Qualitative Health Research. SAGE Publications, Inc. https://doi.org/10.4135/9781412985017.

Lufuke, Mosses, Yunli Bai, Shenggen Fan, and Xu Tian. 2022. “Women’s Empowerment, Food Security, and Nutrition Transition in Africa.” International Journal of Environmental Research and Public Health 20 (1): 254. https://doi.org/10.3390/ijerph20010254.

———. 2023. “Women’s Empowerment, Food Security, and Nutrition Transition in Africa.” International Journal of Environmental Research and Public Health 20 (1): 254. https://doi.org/10.3390/ijerph20010254.

Malapit, Hazel, Agnes Quisumbing, Ruth Meinzen-Dick, Greg Seymour, Elena M. Martinez, Jessica Heckert, Deborah Rubin, Ana Vaz, and Kathryn M. Yount. 2019. “Development of the Project-Level Women’s Empowerment in Agriculture Index (pro-WEAI).” World Development 122 (October): 675–92. https://doi.org/10.1016/j.worlddev.2019.06.018.

Malapit, Hazel Jean, Ruth Meinzen-Dick, Agnes Quisumbing, and Laura Zseleczky. 2020. “Women: Transforming Food Systems for Empowerment and Equity.” In 2020 Global Food Policy Report: Building Inclusive Food Systems, 36–45. International Food Policy Research Institute (IFPRI). https://EconPapers.repec.org/RePEc:fpr:ifpric:9780896293670_04.

Maraka, Joshua M. 2021. “Gender Equality, Women’s Empowerment and Food Systems.” CGIAR Gender Platform (blog). November 15, 2021. https://gender.cgiar.org/news/gender-equality-womens-empowerment-and-food-systems.

Marcus, Rachel, Fiona Samuels, Shoubo Jalal, and Helen Belachew. 2022. “Gender Transformative Programming.” Background Paper Series. UNICEF. https://www.unicef.org/lac/en/media/43146/file.

McDougall, C., J. Newton, F. Kruijssen, and A. Reggers. 2022. “Gender Integration and Intersectionality in Food Systems Research for Development: A Guidance Note.” https://digitalarchive.worldfishcenter.org/handle/20.500.12348/5286.

Miles, Matthew B., and A. Michael Huberman. 1994. Qualitative Data Analysis: An Expanded Sourcebook. SAGE.

Moore, Nick, Charlotte Lane, Ingunn Storhaug, Amber Franich, Heike Rolker, Josh Furgeson, Thalia Sparling, and Birte Snilstveit. 2021. “The Effects of Food Systems Interventions on Food Security and Nutrition Outcomes in Low- and Middle-Income Countries.” 2021st ed. International Initiative for Impact Evaluation (3ie). https://doi.org/10.23846/EGM016.

Mullan, Rebecca J., David N. Flynn, Bo Carlberg, Imad M. Tleyjeh, Celia C. Kamath, Matthew L. LaBella, Patricia J. Erwin, Gordon H. Guyatt, and Victor M. Montori. 2009. “Systematic Reviewers Commonly Contact Study Authors but Do so with

Limited Rigor." Journal of Clinical Epidemiology 62 (2): 138–42. https://doi.org/10.1016/j.jclinepi.2008.08.002.

Narayanan, Sudha, Erin Lentz, Marzia Fontana, Anuradha De, and Bharati Kulkarni. 2019. "Developing the Women's Empowerment in Nutrition Index in Two States of India." Food Policy 89 (December): 101780. https://doi.org/10.1016/j.foodpol.2019.101780.

Narayanan, Sudha, Erin Lentz, Marzia Fontana, and Bharati Kulkarni. 2022. "Rural Women's Empowerment in Nutrition: A Framework Linking Food, Health and Institutions." The Journal of Development Studies 58 (1): 1–18. https://doi.org/10.1080/00220388.2021.1961746.

Njuki, Jemimah, Sarah Eissler, Hazel J. Malapit, Ruth Suseela Meinzen-Dick, Elizabeth Bryan, and Agnes R. Quisumbing. 2021. "A Review of Evidence on Gender Equality, Women's Empowerment, and Food Systems." Washington, DC: International Food Policy Research Institute. https://doi.org/10.2499/p15738coll2.134469.

Njuki, Jemimah, Sarah Eissler, Hazel Malapit, Ruth Meinzen-Dick, Elizabeth Bryan, and Agnes Quisumbing. 2022. "A Review of Evidence on Gender Equality, Women's Empowerment, and Food Systems." Global Food Security 33 (June): 100622. https://doi.org/10.1016/j.gfs.2022.100622.

———. 2023. "A Review of Evidence on Gender Equality, Women's Empowerment, and Food Systems." In Science and Innovations for Food Systems Transformation, edited by Joachim Von Braun, Kaosar Afsana, Louise O. Fresco, and Mohamed Hag Ali Hassan, 165–89. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-15703-5_9.

OECD. 2009. "Gender Equality Indicators: What, Why, and How?" OECD. https://www.oecd.org/dac/gender-development/44952761.pdf.

———. 2023. "Development Finance for Gender Equality and Women's Empowerment - OECD." 2023. https://www.oecd.org/dac/financing-sustainable-development/development-finance-topics/development-finance-for-gender-equality-and-women-s-empowerment.htm.

Oliver, S, K Dickson, M Bangpan, and M Newman. 2017. Getting Started with a Review. An Introduction to Systematic Reviews. New York: SAGE Publications.

Oseni, Gbemisola, Paul Corral, Markus Goldstein, and Paul Winters. 2015. "Explaining Gender Differentials in Agricultural Production in Nigeria." Agricultural Economics 46 (3): 285–310. https://doi.org/10.1111/agec.12166.

Owusu-Addo, Ebenezer, Andre M N Renzaho, and Ben J Smith. 2018. "The Impact of Cash Transfers on Social Determinants of Health and Health Inequalities in

Sub-Saharan Africa: A Systematic Review." Health Policy and Planning 33 (5): 675–96. https://doi.org/10.1093/heapol/czy020.

Oxfam. 2022. "How Much Money Would It Take to End World Hunger?" 2022. https://www.oxfamamerica.org/explore/stories/how-much-money-would-it-take-to-end-world-hunger/.

Pearson, Alan. 2004. "Balancing the Evidence: Incorporating the Synthesis of Qualitative Data into Systematic Reviews." JBI Reports 2 (2): 45–64. https://doi.org/10.1111/j.1479-6988.2004.00008.x.

Pluye, P., E. Robert, M. Cargo, G. Bartlett, A. O'Cathain, F. Griffiths, F. Broadman, M.P. Gagnon, and M.C. Rousseau. 2011. "Proposal: A Mixed Methods Appraisal Tool for Systematic Mixed Studies Reviews." http://mixedmethodsappraisaltoolpublic.pbworks.com/. Http://www.webcitation.org/5tTRTc9yJ. WebCite®. http://mixedmethodsappraisaltoolpublic.pbworks.com/w/file/fetch/84371689/MMAT%202011%20criteria%20and%20tutorial%202011-06-29updated2014.08.21.pdf.

Polanin, Joshua R., and Birte Snilstveit. 2016. "Converting between Effect Sizes." Campbell Systematic Reviews 12 (1): 1–13. https://doi.org/10.4073/cmpn.2016.3.

Quisumbing, Agnes, Akhter Ahmed, John Hoddinott, Audrey Pereira, and Shalini Roy. 2021. 'Designing for Empowerment Impact in Agricultural Development Projects: Experimental Evidence from the Agriculture, Nutrition, and Gender Linkages (ANGeL) Project in Bangladesh'. *World Development* 146 (October): 105622. https://doi.org/10.1016/j.worlddev.2021.105622.

Quisumbing, Agnes R., Kathryn Sproule, Elena M. Martinez, and Hazel Malapit. 2021. "Do Tradeoffs among Dimensions of Women's Empowerment and Nutrition Outcomes Exist? Evidence from Six Countries in Africa and Asia." Food Policy 100 (April): 102001. https://doi.org/10.1016/j.foodpol.2020.102001.

Quisumbing, Agnes, Jessica Heckert, Simone Faas, Gayathri Ramani, Kalyani Raghunathan, and Hazel Malapit. 2022. "Women's Empowerment, Food Systems, and Nutrition." IFAD. https://www.ifad.org/en/web/knowledge/-/research-series-women-s-empowerment-food-systems-and-nutrition?p_l_back_url=%2Fen%2Fweb%2Fknowledge%2F%3Fmoen%3Dsearch%26page%3D75%26stent%3D10%26delta%3D125.

R Core Team. 2018. "R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria." R: A Language and Environment for Statistical Computing. 2018. https://www.r-project.org/.

Richardson, Robin A. 2018. “Measuring Women’s Empowerment: A Critical Review of Current Practices and Recommendations for Researchers.” Social Indicators Research 137 (2): 539–57. https://doi.org/10.1007/s11205-017-1622-4.

Rosenblad, Andreas. 2009. “Introduction to Meta-Analysis by Michael Borenstein, Larry V. Hedges, Julian P.T. Higgins, Hannah R. Rothstein.” International Statistical Review 77 (3): 478–79. https://doi.org/10.1111/j.1751-5823.2009.00095_15.x.

Ruben, Ruerd, Romina Cavatassi, Leslie Lipper, Eric Smaling, and Paul Winters. 2021. “Towards Food Systems Transformation—Five Paradigm Shifts for Healthy, Inclusive and Sustainable Food Systems.” Food Security 13 (6): 1423–30. https://doi.org/10.1007/s12571-021-01221-4.

Ruel, Marie T., Agnes R. Quisumbing, and Mysbah Balagamwala. 2018. “Nutrition-Sensitive Agriculture: What Have We Learned so Far?” Global Food Security 17 (June): 128–53. https://doi.org/10.1016/j.gfs.2018.01.002.

Salazar, Lina, Jossie Fahsbender, and Namho Kim. 2018. “Livestock Transfers, Food Security and Women’s Empowerment: Evidence from a Randomized Phased-in Program in Nicaragua.” IDB Publications, November. https://doi.org/10.18235/0001447.

Sánchez-Meca, Julio, Fulgencio Marín-Martínez, and Salvador Chacón-Moscoso. 2003. “Effect-Size Indices for Dichotomized Outcomes in Meta-Analysis.” Psychological Methods 8 (4): 448–67. https://doi.org/10.1037/1082-989X.8.4.448.

Sandelowski, M., S. Docherty, and C. Emden. 1997a. “Focus on Qualitative Methods. Qualitative Metasynthesis: Issues and Techniques.” Research in Nursing & Health 20 (4): 365–71. https://doi.org/10.1002/(sici)1098-240x(199708)20:4<365::aid-nur9>3.0.co;2-e.

———. 1997b. “Focus on Qualitative Methods. Qualitative Metasynthesis: Issues and Techniques.” Research in Nursing & Health 20 (4): 365–71. https://doi.org/10.1002/(sici)1098-240x(199708)20:4<365::aid-nur9>3.0.co;2-e.

Schünemann, Holger J, and Lorenzo Moja. 2015. “Reviews: Rapid! Rapid! Rapid! …and Systematic.” Systematic Reviews 4 (1): 4, 2046-4053-4–4. https://doi.org/10.1186/2046-4053-4-4.

Scriven, Michael. 2008. “A Summative Evaluation of RCT Methodology: & An Alternative Approach to Causal Research.” Journal of MultiDisciplinary Evaluation 5 (9): 11–24. https://doi.org/10.56645/jmde.v5i9.160.

Snilstveit, Birte, Jennifer Stevenson, Ian Shemilt, Mike Clarke, Emmanuel Jimenez, and James Thomas. 2018. “Timely, Efficient, and Living Systematic Reviews: Opportunities in International Development.” Centre for Excellence and Development Impact and Learning (CEDIL). https://doi.org/10.51744/CIP7.

Snilsveit, Birte, Jennifer Stevenson, Laurenz Langer, Natalie Tannous, Zafeer Ravat, Promise Nduku, Joshua Polanin, Ian Shemilt, John Eyers, and Paul J. Ferraro. 2019. “Incentives for Climate Mitigation in the Land Use Sector—the Effects of Payment for Environmental Services on Environmental and Socioeconomic Outcomes in Low- and Middle-Income Countries: A Mixed-Methods Systematic Review.” Campbell Systematic Reviews 15 (3): e1045. https://doi.org/10.1002/cl2.1045.

SPRING. 2014. “Understanding the Women’s Empowerment Pathway.” USAID/Strengthening Partnerships, Results, and Innovations in Nutrition Globally (SPRING) Project.

Sterne, Jonathan A. C., Jelena Savović, Matthew J. Page, Roy G. Elbers, Natalie S. Blencowe, Isabelle Boutron, Christopher J. Cates, et al. 2019. “RoB 2: A Revised Tool for Assessing Risk of Bias in Randomised Trials.” BMJ (Clinical Research Ed.) 366 (August): l4898. https://doi.org/10.1136/bmj.l4898.

Sterne, Jonathan A.C., and Matthias Egger. 2005. “Regression Methods to Detect Publication and Other Bias in Meta-Analysis.” In Publication Bias in Meta-Analysis, 99–110. John Wiley & Sons, Ltd. https://doi.org/10.1002/0470870168.ch6.

Storhaug, Ingunn, Charlotte Lane, Diana Cordova-Arauz, and Mark Engelbert. 2023. “Food Systems and Nutrition Evidence and Gap Map : Update #6 (May 2023 - July 2023).” 3ie. https://www.3ieimpact.org/sites/default/files/2023-11/Food-Systems-Nutrition-EGM-brief-nov23update-web.pdf.

Tashakkori, Abbas, and Charles Teddlie. 1998. Mixed Methodology: Combining Qualitative and Quantitative Approaches. Vol. 46. sage.

Thomas, James, Angela Harden, Ann Oakley, Sandy Oliver, Katy Sutcliffe, Rebecca Rees, Ginny Brunton, and Josephine Kavanagh. 2004. “Integrating Qualitative Research with Trials in Systematic Reviews.” BMJ : British Medical Journal 328 (7446): 1010–12.

Thomas, James, Alison O’Mara-Eves, and Ginny Brunton. 2014. “Using Qualitative Comparative Analysis (QCA) in Systematic Reviews of Complex Interventions: A Worked Example.” Systematic Reviews 3 (1): 67. https://doi.org/10.1186/2046-4053-3-67.

Tricco, Andrea C., Jesmin Antony, Wasifa Zarin, Lisa Strifler, Marco Ghassemi, John Ivory, Laure Perrier, Brian Hutton, David Moher, and Sharon E. Straus. 2015. “A Scoping Review of Rapid Review Methods.” BMC Medicine 13 (1): 224. https://doi.org/10.1186/s12916-015-0465-6.

Tschunkert, Kristina, and Caroline Delgado. 2022. “Food Systems in Conflict and Peacebuilding Settings: Ways Forward.” Stockholm International Peace Research Institute. https://doi.org/10.55163/SSKG9519.

UNDP. 2023a. “Gender Inequality Index.” Human Development Reports. United Nations. https://hdr.undp.org/data-center/thematic-composite-indices/gender-inequality-index.

———. 2023b. “Technical Note: Twin Indices on Women’s Empowerment and Gender Equality.” United Nations Development Programme. https://hdr.undp.org/sites/default/files/publications/additional-files/2023-07/paths_equal_2023_tn.pdf.

———. 2023c. “The Paths to Equal: Twin Indices on Women’s Empowerment and Gender Equality.” United Nations Development Programme. https://hdr.undp.org/system/files/documents/hdp-document/pathsequal2023pdf.pdf.

United Nations. 2023. “UN Stats: Zero Hunger.” 2023. https://unstats.un.org/sdgs/report/2023/goal-02/.

Valentine, Jeffrey C., Ariel M. Aloe, and Timothy S. Lau. 2015. “Life After NHST: How to Describe Your Data Without ‘p-Ing’ Everywhere.” Basic and Applied Social Psychology 37 (5): 260–73. https://doi.org/10.1080/01973533.2015.1060240.

Van Den Bold, Mara, Andrew Dillon, Deanna Olney, Marcellin Ouedraogo, Abdoulaye Pedehombga, and Agnes Quisumbing. 2015. “Can Integrated Agriculture-Nutrition Programmes Change Gender Norms on Land and Asset Ownership? Evidence from Burkina Faso.” The Journal of Development Studies 51 (9): 1155–74. https://doi.org/10.1080/00220388.2015.1036036.

Viechtbauer, Wolfgang. 2010. “Conducting Meta-Analyses in R with the Metafor Package.” Journal of Statistical Software 36 (August): 1–48. https://doi.org/10.18637/jss.v036.i03.

Viechtbauer, Wolfgang, and Mike W.-L. Cheung. 2010. “Outlier and Influence Diagnostics for Meta-Analysis.” Research Synthesis Methods 1 (2): 112–25. https://doi.org/10.1002/jrsm.11.

Villar, Paul Fenton, and Hugh Waddington. 2019. “Within Study Comparisons and Risk of Bias in International Development: Systematic Review and Critical

Appraisal." Campbell Systematic Reviews 15 (1–2): e1027. https://doi.org/10.1002/cl2.1027.

Waddington, Hugh Sharma, Paul Fenton Villar, and Jeffrey C. Valentine. 2023. "Can Non-Randomised Studies of Interventions Provide Unbiased Effect Estimates? A Systematic Review of Internal Replication Studies." Evaluation Review 47 (3): 563–93. https://doi.org/10.1177/0193841X221116721.

Waddington, Hugh, Birte Snilstveit, Jorge Hombrados, Martina Vojtkova, Daniel Phillips, Philip Davies, and Howard White. 2014. "Farmer Field Schools for Improving Farming Practices and Farmer Outcomes: A Systematic Review." Campbell Systematic Reviews 10 (1): i–335. https://doi.org/10.4073/CSR.2014.6.

Waddington, Hugh, Howard White, Birte Snilstveit, Jorge Garcia Hombrados, Martina Vojtkova, Philip Davies, Ami Bhavsar, et al. 2012. "How to Do a Good Systematic Review of Effects in International Development: A Tool Kit." Journal of Development Effectiveness 4 (3): 359–87. https://doi.org/10.1080/19439342.2012.711765.

WEF. 2023. "Global Gender Gap Report." World Economic Forum. https://www.weforum.org/publications/global-gender-gap-report-2023/.

WFP. 2022. "WFP and Gender." World Food Programme. https://docs.wfp.org/api/documents/WFP-0000138716/download/?_ga=2.155361586.806387667.1706181199-1803271924.1705654889.

———. 2023. "Women Are Hungrier." World Food Program USA. 2023. https://wfpusa.org/women-are-hungrier-infographic/.

Wilson, David B., David Weisburd, and David McClure. 2011. "Use of DNA Testing in Police Investigative Work for Increasing Offender Identification, Arrest, Conviction and Case Clearance." Campbell Systematic Reviews 7 (1): 1–53. https://doi.org/10.4073/csr.2011.7.

W.Noblit, George, and R. DwightHare. 1988. Meta-Ethnography. SAGE Publications, Inc. https://doi.org/10.4135/9781412985000.

World Bank. 2023a. "Classification of Fragile and Conflict-Affected Situations." World Bank. 2023. https://www.worldbank.org/en/topic/fragilityconflictviolence/brief/harmonized-list-of-fragile-situations.

———. 2023b. "World Bank Group Country Classifications by Income Level for FY24 (July 1, 2023- June 30, 2024)." June 30, 2023.

https://blogs.worldbank.org/opendata/new-world-bank-group-country-classifications-income-level-fy24.

World Vision. 2022. “Women’s Economic Empowerment Indicator, Guidance and Tools.” World Vision. https://www.wvi.org/sites/default/files/2022-03/WEE%20Indicator%20Guidance%20and%20Tools%20online%20version.pdf.

Yila, Jummai Othniel, and Almamy Sylla. 2019. “Women Empowerment in Addressing Food Security and Nutrition.” In Handbook of the Historiography of Biology, edited by Michael Dietrich, Mark Borrello, and Oren Harman, 1:1–11. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-69626-3_78-1.

Included studies in the Rapid Evidence Assessment (REA) scope

Adoho, Franck, Shubha Chakravarty, Dala T. Korkoyah Jr., Mattias Lundberg, and Afia Tasneem. 2014. ‘The Impact Of An Adolescent Girls Employment Program’. *World Bank Policy Research Working Paper*. Vol. 6832. http://documents.worldbank.org/curated/en/610391468299085610/The-impact-of-an-adolescent-girls-employment-program-the-EPAG-project-in-Liberia.

Ahmed, Akhter, Fiona Coleman, John F. Hoddinott, Purnima Menon, Shalini Roy, Aklima Parvin, Audrey Pereira, and Agnes Quisumbing. 2022. ‘Comparing Delivery Channels to Promote Nutrition-Sensitive Agriculture: A Cluster-Randomized Controlled Trial in Bangladesh’. *IFPRI - Discussion Papers*. https://doi.org/10.2499/p15738coll2.136473.

Ajefu, Joseph B., Nadia Singh, Ali Shayequazeenat, and Efobi Uchenna. 2022. ‘Women’s Inheritance Rights and Child Health Outcomes in India.’ *The Journal of Development Studies* 58 (4): 752–67. https://doi.org/10.1080/00220388.2021.2003333.

Aker, Jenny. 2013. ‘Cash Or Coupons? Testing The Impacts Of Cash Versus Vouchers In The Democratic Republic Of Congo’. *Other*. Vol. 320. https://www.cgdev.org/publication/cash-or-coupons-testing-impacts-cash-versus-vouchers-democratic-republic-congo-working.

Aker, Jenny, and Christopher Ksoll. 2016. ‘Can Mobile Phones Improve Agricultural Outcomes? Evidence From A Randomized Experiment In Niger’. *Food Policy* 60: 44–51. https://doi.org/10.1016/j.foodpol.2015.03.006.

Aker, Jenny C. 2017. ‘Comparing Cash And Voucher Transfers In A Humanitarian Context: Evidence From The Democratic Republic Of Congo’. *The African*

Journal of Food, Agriculture, Nutrition and Development (AJFAND) 31 (1): 44-70-4–70. https://doi.org/10.1093/wber/lhv055.

Alaofe, Halimatou, Jennifer Burney, Rosemond Naylor, and Douglas Taren. 2019. 'The Impact of A Solar Market Garden Program on Dietary Diversity, Women'S Nutritional Status and Micronutrient Levels in Kalalé District of Northern Benin'. *Public Health Nutrition* 22 (14): 2670–81. https://doi.org/10.1017/S1368980019001599.

Alemu, Sintayehu Hailu, Luuk Van Kempen, and Ruerd Ruben. 2018. 'Women Empowerment Through Self-Help Groups: The Bittersweet Fruits of Collective Apple Cultivation in Highland Ethiopia'. *Journal of Human Development and Capabilities* 19 (3): 308–30. https://doi.org/10.1080/19452829.2018.1454407.

Ali, Daniel A., Matthew Collin, Klaus Deininger, Stefan Dercon, Justin Sandefur, and Andrew Zeitlin. 2014. 'The Price Of Empowerment: Experimental Evidence On Land Titling In Tanzania'. *World Bank Policy Research Working Paper*. Vol. 6908. https://doi.org/10.1596/1813-9450-6908.

Ali, Daniel Ayalew, Klaus Deininger, and Markus JO -Not applicable Goldstein. n.d. 'Environmental And Gender Impacts Of Land Tenure Regularization In Africa: Pilot Evidence From Rwanda'.

Al-shami, Samer Ali, Abdullah Al Mamun, Nurulizwa Rashid, and Mohammed Al-shami. 2021. 'Microcredit Impact on Socio-Economic Development and Women Empowerment in Low-Income Countries: Evidence from Yemen'. *Sustainability* 13 (16). https://doi.org/10.3390/su13169326.

Anand, Paul, Swati Saxena, Rolando Gonzalez, and Hai-Anh Dang H. 2019. 'Can Women's Self-Help Groups Contribute To Sustainable Development? Evidence Of Capability Changes From Northern India'. *World Bank Policy Research Working Paper*. Vol. 9011. https://openknowledge.worldbank.org/bitstream/handle/10986/32418/WPS9011.pdf.

Anjomshoa, Hamideh, Moghaddameh Mirzaee, and Abedin Iranpour. 2018. 'The Application Of Social Cognitive Theory On Mothers' Feeding Practices For Children Aged 6 To 24 Months Old In Iran'. *International Journal of Pediatrics* 6 (7): 7983–97. https://doi.org/10.22038/ijp.2018.28326.2459.

Bahru, Bezawit Adugna, and Manfred Zeller. 2022. 'Gauging the Impact of Ethiopia's Productive Safety Net Program on Agriculture: Application of Targeted Maximum Likelihood Estimation Approach'. *Journal of Agricultural Economics* 73: 257–76. https://doi.org/10.1111/1477-9552.12452.

Baliki, Ghassan, Tilman BrÃ¼ck, Pepijn Schreinemachers, and Md. Nasir Uddin. 2019. 'Long-Term Behavioural Impact Of An Integrated Home Garden Intervention: Evidence From Bangladesh'. *Food Security* 11: 1217–30. https://doi.org/10.1007/s12571-019-00969-0.

Baliki, Ghassan, Pepijn Schreinemachers, Tilman Bruck, and Nasir Md. Uddin. 2022. 'Impacts of a Home Garden Intervention in Bangladesh after One, Three and Six Years'. *Agriculture & Food Security*, 11:48-1:48. https://doi.org/10.1186/s40066-022-00388-z.

Bandiera, Oriana, Niklas Buehren, Robin Burgess, Markus Goldstein, Selim Gulesci, Imran Rasul, and Munshi Sulaiman. 2012. 'Empowering Adolescent Girls: Evidence from a Randomized Control Trial in Uganda'. http://econ.lse.ac.uk/staff/rburgess/wp/ELA.pdf.

Bandiera, Oriana, Niklas Buehren, Markus Goldstein, Imran Rasul, and Andrea Smurra. 2018. 'The Economic Lives of Young Women in the Time of Ebola: Lessons from an Empowerment Program.' http://hdl.handle.net/10986/31219.

Banerjee, Abhijit, Dean Karlan, Robert Osei, Bram Thuysbaert, and Christopher Udry. 2017. 'Graduation From Ultra Poverty In Ghana'. *3ie Grantee Final Report*. Vol. Not applicable. http://www.3ieimpact.org/sites/default/files/2018-08/3ie%20Grantee%20Final%20Report%20OW2.206.pdf.

Bari, Abdul, Khan Ghulam, Bing He, and Yuichiro Yoshira. 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). https://doi.org/10.1371/journal.pone.0262031.

Baro, Mamadou, Lori Beaman, Micah Boyer, Sarah Custer, Tara Deubel, Megan Gash, James Greenberg, et al. 2013. 'Final Impact Evaluation Of The Saving For Change Program In Mali, 2009-2012'. *NORDA Reviews*. Vol. Not applicable. https://norad.no/en/toolspublications/publications/ngo-evaluations/2014/final-impact-evaluation-of-the-saving-for-change-program-in-mali-2009-2012/.

Barrett, Christopher B., Asad Islam, Mohammad Abdul Malek, Debayan Pakrashi, and Ummul Ruthbah. 2021. 'Experimental Evidence on Adoption and Impact of the System of Rice Intensification'. *American Journal of Agricultural Economics*, 1–29. https://doi.org/10.1111/ajae.12245.

Beath, Andrew, Fotini Christia, and Ruben Enikolopov. 2013. 'Randomized Impact Evaluation of Afghanistan's National Solidarity Program'. *The World Bank Group*. https://openknowledge.worldbank.org/handle/10986/16637.

Bernard, Tanguy, Melissa Hidrobo, Agnes Le Port, and Rahul Rawat. 2019. 'Nutrition-Based Incentives In Dairy Contract Farming In Northern Senegal'. *American Journal of Agricultural Economics* 101 (2): 404–35. https://doi.org/10.1093/ajae/aay036.

Bishop David JO -Not applicable. 2012. 'Effectiveness Review: Copperbelt Livelihoods Project, Zambia'.

Blacker, Lauren, Marianne V Santoso, Elias Mtinda, Rachel N Bezner Kerr, Sera L Young, Laura Harnett, Neema Kassim, and Clara JO -Not applicable Mollay. 2018. 'Evaluating the Impact of a Participatory Nutrition-Sensitive Agriculture Intervention on Women's Empowerment and Child's Diet in Singida, Tanzania'. https://www.cochranelibrary.com/central/doi/10.1002/central/CN-01557963/full.

Blackmore, Ivy, Carolyn Lesorogol, and Lora Iannotti. 2018. 'Small Livestock and Aquaculture Programming Impacts on Household Livelihood Security: A Systematic Narrative Review'. *Journal of Development Effectiveness* 10 (2): 197–248. https://doi.org/10.1080/19439342.2018.1452777.

Bold, Mara Van Den, Abdoulaye Pedehombga, Marcellin Ouedraogo, Agnes R. Quisumbing, Deanna Olney, and Andrew Dillon. 2015. 'Can Integrated Agriculture-Nutrition Programs Change Gender Norms On Land And Asset Ownership? Evidence From Burkina Faso'. *Journal of Development Studies* 51 (9): 1155–74. https://doi.org/10.1080/00220388.2015.1036036.

Bonuedi, Isaac, Lukas Kornher, and Nicolas Gerber. 2020. 'Making Cash Crop Value Chains Nutrition-Sensitive: Evidence from a Quasi-Experiment in Rural Sierra Leone.' *Other*. https://doi.org/10.2139/ssrn.3603918.

Bonuedi, Lukas Kornher, and Nicolas Gerber. 2021. 'Intervening in Cash Crop Value Chains for Improved Nutrition: Evidence from Rural Sierra Leone.' *The Journal of Development Studies* 58 (1): 38–54. https://doi.org/10.1080/00220388.2021.1945043.

Bryan, Elizabeth, and Dawit Mekonnen. 2023. 'Does Small-Scale Irrigation Provide a Pathway to Women's Empowerment? Lessons From Northern Ghana'. *Journal of Rural Studies* 97 (January 2023): 474–84. https://doi.org/10.1016/j.jrurstud.2022.12.035.

Caeyers, Bet, and Rob JO -Not applicable Fuller. 2015. 'Women's Empowerment in Rwanda: Evaluation of Women's Economic Leadership Through Horticulture Planting Material Business'. Vol. Series 2013-14. https://doi.org/10.21201/2015.550099.

Claudino JO -Not applicable. 2018. 'Preventing Weight-Related Problems among Adolescent Girls: A Cluster Randomized Trial Comparing the Brazilian New Moves' Program versus Observation.' https://pubmed.ncbi.nlm.nih.gov/28797704/.

Crookston, Benjamin T. 2021. 'Understanding Female and Male Empowerment in Burkina Faso Using the Project-Level Women's Empowerment in Agriculture Index (pro-WEAI): A Longitudinal Study'. *BMC Women's Health*. https://doi.org/10.1186/s12905-021-01371-9.

Daniel, Ali, Deininger Klaus, Goldstein Markus, and La Ferrara JO -Not applicable Eliana. 2015. 'Empowering Women Through Land Tenure Regularization: Evidence From The Impact Evaluation Of The National Program In Rwanda'. http://documents.worldbank.org/curated/en/241921467986301910/Empowering-women-through-land-tenure-regularization-evidence-from-the-impact-evaluation-of-the-national-program-in-Rwanda.

Das, Narayan, Rabeya Yasmin, Jinnat Ara, Md. Kamruzzaman, Peter Davis, Julia A. Behrman, Shalini Roy, and Agnes R. Quisumbing. 2013. 'How Do Intrahousehold Dynamics Change When Assets Are Transferred To Women? Evidence From BRAC's Challenging The Frontiers Of Poverty Reduction—Targeting The Ultra Poor Program In Bangladesh'. *International Food Policy Research Institute (IFPRI) Discussion Paper Series*. Vol. 1317. http://ebrary.ifpri.org/cdm/ref/collection/p15738coll2/id/127988.

Demont, Timothée. 2020. 'Coping With Shocks: The Impact of Self-Help Groups on Migration and Food Security'. https://halshs.archives-ouvertes.fr/halshs-02571730/.

Depenbusch, Lutz, Pepijn Schreinemachers, Ralph Roothaert, Sylvia Namazzi, Charles Onyango, Sophia Bongole, and James Mutebi. 2021. 'Impact of Home Garden Interventions in East Africa: Results of Three Randomized Controlled Trials'. *Food Policy* 104. https://doi.org/10.1016/j.foodpol.2021.102140.

Dohmwirth Carla, and Ziming Liu. 2020. 'Does Cooperative Membership Matter for Women's Empowerment? Evidence From South Indian Dairy Producers'. *Journal of Development Effectiveness* 12 (2): 133–50. https://doi.org/10.1080/19439342.2020.1758749.

Dyer, Julian, and Jeremy Shapiro. 2023. 'Pumps, Prosperity and Household Power: Experimental Evidence on Irrigation Pumps and Smallholder Farmers in Kenya'. *Journal of Development Economics* 163. https://doi.org/10.1016/j.jdeveco.2022.103034.

Evans, David K. 2021. “Towards Improved and More Transparent Ethics in Randomised Controlled Trials in Development Social Science.”

Gaddis, Isis, and Janneke Pieters. 2014. ‘The Gendered Labor Market Impacts Of Trade Liberalization: Evidence From Brazil’. *World Bank Policy Research Working Paper*. Vol. 7095. https://doi.org/10.1596/1813-9450-7095.

Glass, Nancy, Nancy A Perrin, Anjalee Kohli, Jacquelyn Campbell, and Mitima Mpanano Remy. 2017. ‘Randomised Controlled Trial of a Livestock Productive Asset Transfer Program to Improve Economic and Health Outcomes and Reduce Intimate Partner Violence in a Postconflict Setting’. *BMJ Global Health* 2 (1): e000165-000165. https://doi.org/10.1136/bmjgh-2016-000165.

Goldstein, Markus, Kenneth Houngbedji, Florence Kondylis, Michael O’Sullivan, and Harris Selod. 2015. ‘Formalizing Rural Land Rights In West Africa: Early Evidence From A Randomized Impact Evaluation In Benin’. *World Bank Policy Research Working Paper*. Vol. 7435. http://documents.worldbank.org/curated/en/947811468189268752/Formalizing-rural-land-rights-in-West-Africa-early-evidence-from-a-randomized-impact-evaluation-in-Benin.

Gram, Lu, Joanna Morrison, Naomi Saville, Shyam Sundar Yadav, Bhim Shrestha, Dharma Manandhar, Anthony Costello, and Jolene Skordis-Worrall. 2018. ‘Do Participatory Learning And Action Women’S Groups Alone Or Combined With Cash Or Food Transfers Expand Women’S Agency In Rural Nepal?’ *Journal of Development Studies* 55 (8). https://doi.org/10.1080/00220388.2018.1448069.

Greif, Adi, Dan Mattingly, Lauren Persha, Stephanie Fenner, Karol Boudreaux, Aidan Boyd, Cynthia Caron, and Heather Huntington. 2016. ‘Ethiopia Strengthening Land Tenure And Administration Program Endline Report’. *United States Agency for International Development (USAID) Working Paper Series*. Vol. Not applicable. https://www.land-links.org/wp-content/uploads/2016/09/USAID_Land_Tenure_ELTAP-ELAP_Impact_Evaluations_Endline_Report.pdf.

Grillos, Tara. 2018. ‘Women’s Participation In Environmental Decision-Making: Quasi-Experimental Evidence From Northern Kenya’. *Elsevier* 108 (2018): 115-130-15–130. https://doi.org/10.1016/j.worlddev.2018.03.017.

Haque, Mohammed A., Nuzhat Choudhury, Mohammad Jyoti Raihan, S. M. Tanvir Ahmed, Kazi Eliza Islam, Vanessa Self, Shahed Rahman, Lilly Schofield, Andrew Hall, and Tahmeed Ahmed. 2020. ‘Enhanced Women’s Decision-Making Power after the Suchana Intervention in North-Eastern Bangladesh: A Cluster Randomised Pre-Post Study’. *BMC Public Health*. https://doi.org/10.1186/s12889-020-08769-4.

Harris-Fry, Helen A, Kishwar Azad, Leila Younes, Abdul Kuddus, Sanjit Shaha, Tasmin Nahar, Munir Hossen, Anthony Costello, and Edward Fottrell. 2016. 'Formative Evaluation Of A Participatory Women'S Group Intervention To Improve Reproductive And Women'S Health Outcomes In Rural Bangladesh: A Controlled Before And After Study'. *Journal of Epidemiology and Community Health* 70 (7). https://doi.org/10.1136/jech-2015-205855.

Heckert, Jessica, Elena M. Martinez, Armande Sanou, Abdoulaye Pedehombga, Rasmané Ganaba, and Aulo Gelli. 2022. 'Can a Gender-Sensitive Integrated Poultry Value Chain and Nutrition Intervention Among the Rural Poor Increase Women's Empowerment in Burkina Faso?' *International Food Policy Research Institute (IFPRI) Discussion Paper Series*. Vol. 02144. https://doi.org/10.2499/p15738coll2.136428.

Heidari, Zeinab, Mahrokh Keshvari, and Shahnaz Kohan. 2016. 'Clinical Trial To Comparison The Effect Of Family-Centered Educational-Supportive Program On Mothers' Empowerment In Breast-Feeding'. *International Journal of Pediatrics*, 1445–51. https://doi.org/10.22038/ijp.2016.6556.

Hewett, Paul C., Karen Austrian, Erica Soler-Hampejsek, Jere R. Behrman, Fiammetta Bozzani, and Natalie A. JO -Not applicable Jackson-Hachonda. n.d. 'Cluster Randomized Evaluation of Adolescente Girls Empowerment Program (AGEP): Study Protocol'.

Hidrobo, Melissa, Amber Peterman, and Lori JO -Not applicable Heise. 2016. 'The Effect of Cash, Vouchers and Food Transfers on Intimate Partner Violence: Evidence from a Randomized Experiment in Northern Ecuador'. https://doi.org/10.1257/app.20150048.

Hughes, Karl. 2012. 'Effectiveness Review: Enhancing Access and Control to Sustainable Livelihood Assets, Philippines'. *Oxfam Effectiveness Review Series*. https://oxfamilibrary.openrepository.com/bitstream/10546/247232/1/er-livelihoods-philippines-effectiveness-review-081012-summ-report-en.pdf.

IFAD JO -Not applicable. 2018. 'Impact Assessment: Agricultural Sector Development Program–Livestock (ASDP-L) and Agriculture Service Support Program (ASSP)'. https://ideas.repec.org/p/ags/unadia/288463.html.

Independent Office of Evaluation of IFAD (IOE). 2015. 'Jharkhand Chhattisgarh Tribal Development Program'. *International Fund for Agricultural Development*. Vol. 3814-in. https://www.ifad.org/en/web/ioe/evaluation/asset/39823508.

Jaacks, Lindsay M., Lilia Bliznashka, Peter Craig, Michael Eddleston, Alfred Gathorne-Hardy, Ranjit Kumar, Sailesh Mohan, et al. 2023. 'Co-Benefits of Largescale Organic Farming on Human Health (Bloom): Protocol for a Cluster-Randomized

Controlled Evaluation of the Andhra Pradesh Community-Managed Natural Farming Program in India'. *PLoS One*. https://doi.org/10.1371/journal.pone.0281677.

Justine Briaux, Yves Martin-Prevel, Sophie Carles, Sonia Fortin, Yves Kameli, Laura Adubra, Renk Andre'a, et al. 2020. 'Evaluation of an Unconditional Cash Transfer Program Targeting Children's First-1,000-Days Linear Growth in Rural Togo: A Cluster-Randomized Controlled Trial'. *PLoS Medicine* 17 (11): e1003388-1003388. https://doi.org/10.1371/journal.pmed.1003388.

Kafle, Kashi, Hope Michelson, and Alex Winter-Nelson. 2016. 'Does She Have A Say? The Impact Of Livestock Transfer And Associated Training On Women'S Empowerment: Evidence From Zambia'. *Agricultural & Applied Economics Association*. https://doi.org/10.22004/ag.econ.236271.

Kafle, Kashi Ram. 2017. 'Impact Of Coordinated Asset Transfers And Asset Ownership On Poverty Reduction, Women'S Empowerment, And Child Education: Evidence From Zambia, And Tanzania'. *University of Illinois Urbana-Champaign*. https://www.ideals.illinois.edu/handle/2142/97678.

Kareem, Fatima Olanike, and Olayinka Idowu Kareem. 2021. 'Employment Responses to EU Food Safety Regulations: A Gendered Perspective'. *The European Journal of Development Research* 33: 1899-1929-899–1929. https://doi.org/10.1057/s41287-020-00330-8.

Karimli, Leyla, Bijetri Bose, and Njeri Kagotho. 2019. 'Integrated Graduation Program and Its Effect on Women and Household Economic Well-Being: Findings from a Randomised Controlled Trial in Burkina Faso'. *Journal of Development Studies* 56 (7). https://doi.org/10.1080/00220388.2019.1677887.

Kirkwood, Elizabeth K, Michael John Dibley, John Frederick Hoddinott, Tanvir Huda, Tracey Lea Laba, Tazeen Tahsina, Mohammad Mehedi Hasan, et al. 2021. 'Assessing the Impact of a Combined Nutrition Counselling and Cash Transfer Intervention on Women's Empowerment in Rural Bangladesh: A Randomised Control Trial Protocol'. *BMJ Open*. Vol. 11. https://doi.org/10.1136/bmjopen-2020-044263.

Kurdi, Sikandra, Yashodhan Ghorpade, and Hosam Ibrahim. 2019. 'The Cash For Nutrition Intervention In Yemen; Impact Evaluation Study'. https://www.ifpri.org/publication/cash-nutrition-intervention-yemen-impact-evaluation-study.

Lecoutere, Els JO -Not applicable. 2017. 'The Impact of Agricultural Co-Operatives on Women's Empowerment: Evidence from Uganda'. https://doi.org/10.1016/j.jcom.2017.03.001.

Lecoutere, Els, David J. Spielman, and Campenhout Van Bjorn. 2020. 'Women'S Empowerment, Agricultural Extension, And Digitalization Disentangling Information And Role-Model Effects In Rural Uganda'. https://www.ifpri.org/publication/womens-empowerment-agricultural-extension-and-digitalization-disentangling-information-0.

Leight, Jessica, Josue Awonon, Abdolulaye Pedehombga, Rasmané Ganaba, Elena M. Martinez, Jessica Heckert, and Aulo Gelli. 2020. 'The Impact of an Integrated Value Chain Intervention on Household Poultry Production in Burkina Faso Evidence from a Randomized Controlled Trial'. *Other*. Vol. 1919. https://doi.org/10.2499/p15738coll2.133686.

Lina, Salazar, Fahsbender Jossie, and Kim Namho. 2018. 'Livestock Transfers, Food Security And Women'S Empowerment: Evidence From A Randomized Phased-In Program In Nicaragua'. *IDB Working Paper Series*. https://doi.org/10.18235/0001447.

Lombardini, Simone. 2013. 'Livelihoods in Honduras: Evaluation of Strengthening Small-Scale Farmers' Agribusiness Capabilities'. *Oxfam GB*. https://policy-practice.oxfam.org/resources/livelihoods-in-honduras-evaluation-of-strengthening-small-scale-farmers-agribus-336555/.

Manda, Tiwari. 2021. 'Essays in Health and Development Economics'. *Other*. https://doi.org/10.17077/etd.005890.

Maredia, Mywish K, Murari Suvedi, Raul Pitoro, and Raju Ghimire. 2017. 'Impact Evaluation of the Feed the Future Cambodia Helping Address Rural Vulnerabilities and Ecosystem Stability (HARVEST) Project'. *Michigan State University, Department of Agricultural Economics*. https://ideas.repec.org/p/ags/midiwp/262393.html.

Martinez-Jaikel, Edward A. Frongillo, Christine E. Blake, and Maryah JO -Not applicable Fram. 2020. 'Reducing Both Food Insecurity and Excess Body Weight in Costa Rican Women: A Cluster Randomized Trial'. https://doi.org/10.1016/j.amepre.2019.11.021.

Mukherjee, Anirban, P.K. Singh, Shantanu Rakshit, and Satya JO -Not applicable Priya. 2019. 'Effectiveness of Poultry Based Farmers' Producer Organization and Its Impact on Livelihood Enhancement of Rural Women'. https://doi.org/10.56093/ijans.v89i10.95024.

Nikiema, Pouirkèta Rita. 2019. 'The Impact of School Feeding Programmes on Educational Outcomes: Evidence from Burkina Faso†'. *Journal of African Economies* 28 (3): 323–41. https://doi.org/10.1093/jae/ejy026.

Nisbett, Nicholas, Richard Longhurst, Inka Barnett, Francesca Feruglio, Jessica Gordon, John Hoddinott, Ferdous Jahan, et al. 2016. 'Impact Evaluation Of The DFID Program To Accelerate Improved Nutrition For The Extreme Poor In Bangladesh: Final Report'. *MQSUN Report*. Vol. Not applicable. http://www.ids.ac.uk/publication/mqsun-impact-evaluation-of-the-dfid-programme-to-accelerate-improved-nutrition-for-the-extreme-poor-in-bangladesh-final-report.

Ntakyo, Proscovia Renzaho, and Marrit van den Berg. 2022. 'The Unintended Side-Effects of a Major Development Strategy: Commercialization of Smallholder Production and Women Empowerment in Uganda'. *Journal of Development Studies* 0 (0). https://doi.org/10.1080/00220388.2022.2032671.

Osei, Robert Darko, Fred M. Dzanku, Isaac Osei-Akoto, Felix Asante, Louis S. Hodey, Pokuaa N. Adu, Kwabena Adu-Ababio, and Massa Coulibaly. 2018. 'Impact of Voice Reminders to Reinforce Harvest Aggregation Services Training for Farmers in Mali'. *3ie Series Report*. Vol. 90. http://3ieimpact.org/evidence-hub/publications/impact-evaluations/impact-voice-reminders-reinforce-harvest-aggregation.

Owusu-Addo, Ebenezer, Andre M. N. Renzaho, and Ben J. Smith. 2018. 'The Impact of Cash Transfers on Social Determinants of Health and Health Inequalities in Sub-Saharan Africa: A Systematic Review'. *Health Policy and Planning* 33 (5): 675–96. https://doi.org/10.1093/heapol/czy020.

Pamuk, Haki. 2021. 'Farmer Field Business Schools and Village Savings and Loan Associations for Promoting Climate-Smart Agriculture Practices: Evidence from Rural Tanzania'. *CGIAR Research Program on Climate Change, Agriculture and Food Security (CCAFS)*. Vol. 361. https://ccafs.cgiar.org/resources/publications/farmer-field-business-schools-and-village-savings-and-loan-associations.

Pandey, Vivek, Hari K. Nagarajan, and Deepak Kumar. 2021. 'Impact of Gendered Participation in Market-Linked Value-Chains on Economic Outcomes: Evidence from India'. *Food Policy* 104. https://doi.org/10.1016/j.foodpol.2021.102142.

Passarelli, Simone, Ramya Ambikapathi, Nilupa S. Gunaratna, Isabel Madzorera, Chelsey R. Canavan, Abdallah R Noor, Amare Worku, et al. 2020. 'A Chicken Production Intervention And Additional Nutrition Behavior Change Component Increased Child Growth In Ethiopia: A Cluster-Randomized Trial'. *Journal of Nutrition*, 1–12. https://doi.org/10.1093/jn/nxaa181.

Persha JO -Not applicable. 2018. 'PHASE II BASELINE AND PHASE 1 MIDLINE REPORT: IMPACT EVALUATION OF THE FEED THE FUTURE TANZANIA LAND TENURE ASSISTANCE ACTIVITY'. https://www.land-

links.org/document/lta-phase-ii-baseline-and-phase-1-midline-report-impact-evaluation-of-the-feed-the-future-tanzania-land-tenure-assistance-activity/.

Persha, Lauren, and Jacob Patterson-Stein. 2018. 'Impact Evaluation Of The Feed The Future Tanzania Land Tenure Assistance Activity'. *USAID*. https://www.land-links.org/document/lta-phase-ii-baseline-and-phase-1-midline-report-impact-evaluation-of-the-feed-the-future-tanzania-land-tenure-assistance-activity/.

Quisumbing, Agnes, Akhter Ahmed, John Hoddinott, Audrey Pereira, and Shalini Roy. 2021. 'Designing for Empowerment Impact in Agricultural Development Projects: Experimental Evidence from the Agriculture, Nutrition, and Gender Linkages (ANGeL) Project in Bangladesh'. *World Development* 146 (October): 105622. https://doi.org/10.1016/j.worlddev.2021.105622.

Raghunathan, Kalyani, S. Kannan, and A. JO -Not applicable Quisumbing. 2018. 'Women'S Self-Help Groups, Decision-Making, And Improved Agricultural Practices In India'. https://www.ifpri.org/publication/womens-self-help-groups-decision-making-and-improved-agricultural-practices-india.

Rath, Suchitra, Audrey Prost, Subhashree Samal, Hemanta Pradhan, K. Bhatia, Andrew Copas, S. Gagrai, et al. 2022. 'Effects of Community Youth Teams Facilitating Participatory Adolescent Groups, Youth Leadership Activities and Livelihood Promotion to Improve School Attendance, Dietary Diversity and Mental Health among Adolescent Girls in Rural Eastern India (JIAH Trial): A Cluster-Randomised Controlled Trial'. https://pubmed.ncbi.nlm.nih.gov/36618545/.

Riddle, Alison, Abby Ramage, Cynthia M. Kroeger, Zulfiqar A. Bhutta, Elizabeth Kristjansson, Monica Taljaard, Carol Vlassoff, et al. 2021. 'Protocol: The Effects of Empowerment-Based Nutrition Interventions on the Nutritional Status of Women of Reproductive Age in Low- and Middle-Income Countries'. https://onlinelibrary.wiley.com/doi/10.1002/cl2.1183.

Riddle, Alison Y., Abigail K. Ramage, Cynthia M. Kroeger, Vivian Welch, Carol Vlassoff, Zulfiqar Bhutta, Elizabeth Kristjansson, Monica Taljaard, and George A. Wells. 2017. 'The Effect Of Women's Empowerment Strategies On Adolescent Girls' Nutritional Status'. https://campbellcollaboration.org/library/effect-of-womens-empowerment-on-girls-nutritional-status.html.

Roy, Shalini, Jinnat Ara, Narayan Das, and Agnes R. Quisumbing. 2015. '"Flypaper Effects" In Transfers Targeted To Women: Evidence From BRAC's "Targeting The Ultra Poor" Program In Bangladesh'. *Journal of Development Economics* 117: 1–19. https://doi.org/10.1016/j.jdeveco.2015.06.004.

Rubio-Jovel, Karla. 2021. ‘Gender Empowerment in Agriculture Interventions: What Are We Still Missing? Evidence From a Randomized-Controlled Trial Among Coffee Producers in Honduras’. *Other* 5. https://doi.org/10.3389/fsufs.2021.695390.

Ruel, Marie, and Neha JO -Not applicable Kumar. 2019. ‘Targeting And Realigning Agriculture For Improved Nutrition (Train)’. https://ridie.3ieimpact.org/index.php?r=search/detailView&id=419.

Ruel, Marie T., Agnes R. Quisumbing, and Mysbah Balagamwala. 2018. ‘Nutrition-Sensitive Agriculture: What We Learned so Far?’ *Global Food Security* 17: 128–53. https://doi.org/10.1016/j.gfs.2018.01.002.

Santoso, Marianne V, Rachel N Bezner Kerr, Neema Kassim, Haikael Martin, Elias Mtinda, Peter Njau, Kelvin Mtei, John Hoddinott, and Sera L Young. 2021. ‘A Nutrition-Sensitive Agroecology Intervention in Rural Tanzania Increases Children’s Dietary Diversity and Household Food Security But Does Not Change Child Anthropometry: Results from a Cluster-Randomized Trial’. *The Journal of Nutrition* 151 (7): 2010–21. https://doi.org/10.1093/jn/nxab052.

Saville JO -Not applicable. 2013. ‘Community Intervention to Improve Growth among Children under Two in Rural India’. https://doi.org/10.1186/ISRCTN51505201.

Silvia Alonso. 2021. ‘MoreMilk Project: Milk Safety and Child Nutrition Impacts of a Training Scheme for Dairy Vendors in Eldoret, Kenya’. *Not Specified*. https://trialsearch.who.int/Trial2.aspx?TrialID=NCT05117892.

Simler, Kenneth R., Krista L. Jacobs, Joseph Mensah-Homiah, Josée Randriamamonjy, Doris Wiesmann, and Abizari Abdul-Razak. 2005. ‘Food-Based Approaches To Reducing Micronutrient Malnutrition: An Impact Evaluation Of The UNICEF ICBD Program In The Savelugu-Nanton District Of Northern Ghana’. *International Food Policy Research Institute (IFPRI) Discussion Paper Series*. Vol. Not applicable. https://www.3ieimpact.org/evidence-hub/impact-evaluation-repository/food-based-approaches-reducing-micronutrient-malnutrition.

Stark, Heather, Anteneh Omer, Aïssata Wereme N’Diaye, Amanda C. Sapp, Emily V. Moore, and Sarah L. JO -Not applicable McKune. 2019. ‘The Un Oeuf Study: Design, Methods and Baseline Data from a Cluster Randomised Controlled Trial to Increase Child Egg Consumption in Burkina Faso’. https://pubmed.ncbi.nlm.nih.gov/32770664/.

Tauseef, Salauddin. 2022. ‘The Importance of Nutrition Education in Achieving Food Security and Adequate Nutrition of the Poor: Experimental Evidence from

Bangladesh'. *Oxford Bulletin of Economics and Statistics* 84 (1): 0305–9049. https://doi.org/10.1111/obes.12465.

University Ghent. 2021. 'Biodiversity in the Diet in Vietnam'. *Other*. https://clinicaltrials.gov/ct2/show/NCT05144919.

Waid, Jillian L, Amanda S Wendt, Sheela S Sinharoy, Abdul Kader, and Sabine Gabrysch. 2022. 'Impact of a Homestead Food Production Program on Women's Empowerment: Pro-WEAI Results From the FAARM Trial in Bangladesh'. *World Development* 158. https://doi.org/10.1016/j.worlddev.2022.106001.

Warren, Andrea M., Edward A. Frongillo, Phuong H. Nguyen, and Purnima Menon. 2020. 'Nutrition Intervention Using Behavioral Change Communication Without Additional Material Inputs Increased Expenditures On Key Food Groups In Bangladesh'. *Journal of Nutrition* 150 (2020): 1284-1290-284–1290. https://doi.org/10.1093/jn/nxz339.

Zou, Wenbo, Travis Lybbert, Stephen Vosti, and Souheila Abbeddou. 2019. 'Early Childhood Nutrition, Parental Growth Perceptions And Educational Aspirations In Rural Burkina Faso'. *Journal of Development Studies* 56 (4): 798–816. https://doi.org/10.1080/00220388.2019.1605056.

Appendices

Appendix 0 – About the FSN E&GM

The initial literature search included 12 academic databases and was originally completed in September 2020. Since then, it has been repeated for the updates every four months from July 2021 until November 2023. Additionally, a grey literature search of 31 sector-specific databases has been searched twice, once during the original E&GM in September 2020 and once in January 2022. For each search, studies were uploaded to EPPI-Reviewer, de-duplicated, and screened independently in duplicate. We extracted interventions, outcomes, population, country, and methods data. All studies meeting the eligibility criteria established in the initial protocol (Moore et al. 2020) are published on the online map. We also publish a regular summary of the new evidence added and the overall evidence distribution.

About 75 percent of the impact evaluations (IEs) in the original E&GM used randomized designs. The updates have shifted to include more quasi-experimental studies (i.e. 44% of the studies identified from updates 1-6 are quasi-experimental). Since the original E&GM, we also observed a reduction in published systematic reviews and an increase in low-confidence systematic reviews. In the first report, 54% of the SRs were rated as low confidence and 19% as high confidence. Over 75% of the high-confidence SRs were published between 2015 and 2020. During the update period, 25 systematic reviews were identified, and 80% were low confidence.

We observed evidence clusters in the original E&GM For example, India, China, and Bangladesh gather higher numbers of studies. The most common interventions were *fortification*, *supplementation*, and *classes related to consumer behavior.* They have remained the most evaluated interventions, but the studies identified in the update period show a reduction in studies evaluating *supplementation* and *fortification* interventions. Overall, most interventions have been evaluated by a minimum of one IE, and interventions with 50 or more IEs have been synthesized by at least one SR. Outcome clusters have remained the same from the original report to the most recent update: *anthropometric*, *diet quality and adequacy,* and *micronutrient status*.

We also identified evidence gaps in the original E&GM. The gaps in intervention categories include *women's empowerment* interventions, *advertising* and *labeling* interventions, interventions to *support food packaging, and cold chain initiatives*. No

evaluations of *advertising regulations, food waste education,* or *food packaging* interventions were identified. We identified five gaps in outcomes, including *women's empowerment*, *economic, social, and political stability*, *food loss, environmental impacts of the food system,* and measures of *diet insufficiency*. Following the expansion of the scope, we also identified a gap in studies measuring *gender-transformative outcomes*.

Finally, we also observed synthesis gaps in *providing free or reduced-cost farm inputs to crop production, agricultural extension, agricultural information provision, agricultural insurance products,* and synthesis of outcomes related to *diet quality and adequacy measures*. By the sixth update, 85 new studies focused on those interventions and outcomes. Of these new studies, 24 were related to women's empowerment outcomes and 19 to women's empowerment interventions, most due to expanding the definition of women's empowerment. This means we have a much bigger evidence base for women's empowerment outcomes with 101 studies.

Appendix 1 – List of interventions

Category	Intervention	Description	Number of studies
	Education / information - other educational programs	Other educational programs supporting the adoption of new agricultural techniques. All mediums of education are included here so long as the information being exchanged is related to agricultural techniques / animal husbandry. Programs related to other educational topics (e.g. literacy) would not be included	17
	Education / information - Agricultural extension programs	Trained agents visit communities to teach current practices, organize cooperatives, and engage in other secondary activities	16
Production system	Livestock access	Activities supporting the management and daily work with livestock: fencings, animal housing, manual tools, protective equipment, antibiotics, or farm animals	9
Efforts to increase food production	Land markets & management	Systems to buy, sell, rent, or manage land related to agriculture	8
	Education / information - Farmer field schools	Farmer field schools bring together a group of farmers to learn agricultural techniques. They meet regularly during a production cycle, setting up experimentation and engaging in hands-on learning to improve skills and knowledge that will help adapt practices to their specific context. Demonstration farms may be used in farmer field schools or separately to show the use of certain agricultural techniques	7
	Other ag inputs	Provision of free or reduced-cost access to agricultural inputs excluding improved seed varieties, fertilisers, pesticides/herbicides, and livestock related ones	7
	Ag credit / savings	Creating or supporting agricultural credit and savings groups	6

Behavior change communication *Efforts to change behaviour through communication campaigns*	Classes	The use of a classroom structure to provide messages regarding healthy eating. This includes classrooms outside of school, online and ambiguous references to “nutrition education” or “education sessions”	11
	Peer support/counsellors	The use of peer support or counsellors to increase healthy eating. Includes home visits and other work by community health workers	9
	Community meetings	The use of community meetings to provide messages regarding healthy eating. Community meetings are public engagements for discussion and mobilization, not simply education. Education within established groups (such as women's self-help groups or microfinance groups) does not count as these are not open to the public	8
Availability and affordability *Efforts to improve food availability and affordability*	Direct provision of foods	State outlets that distribute food at free or reduced cost//Meals provided at free or reduced cost at school//Provision of food outside of state outlets and school meals. Often relates to the charitable distribution of food by religious or civil society groups	10

Appendix 2 – List of outcomes, descriptions, and corresponding indicators

Category	Outcome	Description	Indicators
Resources	Access to economic and livelihood resources	Measures of women's capacity to access, comprehend and effectively utilize basic services. It measures their access to essential activities throughout their lifespan, such as food, water, shelter, education, clothing	Access to basic services Access to institutions and public services Access to education and knowledge Access to income and finance services Access to information
	Ownership of land and assets*	Outcomes with regards to women's' ownership of land or assets.	Ownership of land Ownership of productive/non-productive assets
	Control over resources*	Measures of women's control of resources and the ability to make decisions on how the household resources are used without having to consult other members of the household.	Claim of resources Execution of resources Participation in decision making regarding land and assets Decision on use of income
	Time use*	Assessments of women's time use.	Distribution of time use of women's Time spent on specific activities (e.g. work, basic resources, education, other activities)
Agency	Decision-making*	Measures of women' participation in household decision making, may include financial decisions or intra-household food allocation.	Participation in decision-making at the household and community level Influence in decision-making
	Women's rights	Measures compliance to women's rights and ability to claim those rights and report misconduct.	Freedom of association Freedom of movement Control over body Rights awareness Women take action to claim their right Agreement with misconducts or good Behaviors

	Collective action and leadership	Measures women's participation in community activities through formal or informal networks to effectively claim their right and participated in decision-making.	Participation in community activities Women-led organizations Formation of organizations Actions to claim rights and influence decision-making
Achievements	Gender transformative outcomes*	Measure changes in gender norms, roles, responsibilities, and inequalities through fundamental shifts in societal attitudes, structures, and power dynamics.	Opinions on women's empowerment and gender equality Occurrence of IPV and GBV Resolution of cases of IPV and GPV Awareness of communities on women's rights Media coverage Attitudes towards women and gender norms
	Improved systems and policymaking	Measure actions from powerholders in favor of gender-equality and women's empowerment	Awareness of powerholders for gender equality Actions of powerholders for gender equality Women's representation in institutions Institutional change and gender transformative policymaking
	Self-esteem*	Any measure of self-esteem.	Self-confidence and self-esteem
	Other empowerment outcomes and indices*	Other measures or composite measures of the above.	Indices using multiple outcomes of the list to aggregate them into a gender equality and/or women's empowerment score

Notes: Outcomes with a “*” are labels included in the FSN E&GM. Other outcomes have been added to the framework for the purpose of this REA but were not part of the original FSN E&GM.

Appendix 3 – Study designs

Impact evaluations

We will include impact evaluation using experimental and quasi-experimental study designs to measure a change in outcomes that is attributable to an intervention. This includes studies that apply one of the following approaches:

1. Randomized evaluations with assignment at the individual, household, community or other cluster level, and quasi-randomized mechanisms using prospective methods of assignment such as alternation. This includes randomized trials where units are deliberately assigned to treatment and control groups for the purposes of research, and "natural experiments" where units are exposed to the treatment via some other random mechanism.
2. Non-randomized designs with either a known assignment variable(s) or a seemingly random assignment process:
    a. Regression discontinuity designs, where assignment is based on a threshold measured before intervention, and the study uses regression to model the assignment process.
    b. Natural experiments with clearly defined intervention and comparison groups which exploit apparently random natural variation in assignment (such as a lottery) or random errors in implementation, etc. Natural experiments that approximate randomized evaluations, regression discontinuity designs, or interrupted time series designs will be categorized as such.
3. Non-randomized studies with pre-intervention and post-intervention outcome data for both intervention and comparison groups, where data are individual level panel or pseudo-panels (repeated cross-sections), which use the following methods to control for confounding:
    a. Studies controlling for time-invariant unobservable confounding, including difference-in-differences, fixed-effects models, or models that contain a baseline measure of the dependent variable (e.g., an interaction term between time and intervention for pre-intervention and post-intervention observations).
    b. Studies assessing changes in trends in outcomes over a series of time points with a contemporaneous comparison (controlled interrupted time series, ITS), and with sufficient observations to establish a trend and control for effects on outcomes due to factors other than the intervention (such as seasonality).
4. Non-randomized studies that create a matched comparison group similar to the treated group on specific characteristics to control for observable confounding, including

statistical matching, exact covariate matching, coarsened-exact matching, and propensity score matching.

5. Studies that build a counterfactual through synthetic control approaches.
6. Non-randomized studies that control for confounding using instrumental variable (IV) approaches such as two-stage least squares procedures.

We will exclude before-after studies without a comparison group or cross-sectional studies that do not attempt to control for selection bias or confounding. Studies that only examine willingness-to-pay for goods, services, process, and business models will be excluded.

Experiments conducted in tightly-controlled settings, like those in a laboratory, lab-in-the-field studies, and studies that measure immediate reactions to a short-term exposure (i.e. studies where implementation and data collection is started and completed within a single day) will be excluded.

Qualitative evaluations

We will also include qualitative evaluations evaluating the impact of an intervention included in the quantitative impact evaluations and using the following designs:

1. Realist evaluations assume that projects and programs work under certain conditions and are heavily influenced by the way that different stakeholders respond to them. Authors must clearly state a theory tested through an intervention indicating how and for whom a program would work. They compare contexts, mechanisms, and outcomes within a program (not with a control). There is a strong emphasis on the social and historical context and comparison of those who benefited from the program and those who did not benefit (White and Phillips 2012). A realist evaluation is therefore not just designed to assess whether a development intervention worked or not. It is designed to address questions such as “What works (or doesn’t work)?”; “for whom (and to what extent)?”; “in which circumstances does it work?”; “How and why does it work?” (INTRAC 2017d).
2. Process tracing develops a set of (competing) hypotheses linking an intervention to an outcome including how these hypotheses could be (in)validated. Gather relevant evidence to determine which hypothesis most closely matches observed data. In its pure form, process tracing is based around a set of formal tests. These are designed to assess causation. They are applied to all the different possible explanations for how a particular change might have come about in order to confirm some and/or eliminate

others. Within the process tracing these different explanations are known as hypotheses (INTRAC 2017b).

3. Contribution analysis is a methodology used to identify the contribution a development intervention has made to a change or set of changes. The aim is to produce a credible, evidence-based narrative based on a theory of change that a reasonable person would be likely to agree with, rather than to produce conclusive proof. Contribution analysis can be used during a development intervention, at the end, or afterwards (INTRAC 2017a).
4. Contribution tracing is a participatory mixed-method (qual-quant) to establish the validity of contribution claims with explicit criteria to guide evaluators in data collection and Bayesian updating to quantify the level of confidence in a claim. Includes a contribution 'trial' with all stakeholders to establish what will prove/disprove the claim (HM Treasury 2020).
5. Qualitative impact assessment protocol (QuIP) studies serve to provide an independent reality check of a predetermined theory of change which helps stakeholders to assess, learn from, and demonstrate the social impact of their work. The QuIP gathers evidence of a project's impact through narrative causal statements collected directly from intended project beneficiaries. Respondents are asked to talk about the main changes in their lives over a pre-defined recall period and prompted to share what they perceive to be the main drivers of these changes, and to whom or what they attribute any change - which may well be from multiple sources (Remnant and Avard 2016).
6. General elimination methodology (GEM; Scriven 2008) builds upon his earlier Modus Operandi Method (1976) to provide an approach specifically geared towards substantiating causal claims. The methodology entails systematically identifying and then ruling out alternative causal explanations of observed results. It is based on the idea that for any event it is possible to draw up Lists of Possible Causes (LOPCs) or alternative hypothetical explanations for an outcome of interest. Each putative cause will have its own set of "footprints", or Modus Operandi (MO) – "a sequence of intermediate or concurrent events, a set of conditions or a chain of events that has to be present when the cause is effective (Scriven 2008)" (White and Phillips 2012, 38).
7. Qualitative comparative analysis (QCA) is a methodology that enables the analysis of multiple cases in complex situations. It can help explain why change happens in some cases but not others. QCA is designed for use with an intermediate number of cases, typically between 10 and 50. It can be used in situations where there are too few cases to apply conventional statistical analysis (INTRAC 2017c).

8. Outcome harvesting is designed to collect evidence of change (the ‘outcomes’) and then work backwards to assess whether or how an organization, program or project contributed to that change. Outcomes are defined as changes in the “Behavior writ large” (such as actions, relationships, policies, practices) of one or more social actors influenced by an intervention (Wilson-Grau 2015).

Appendix 4 – Provisional Quantitative data extraction for the REA

Variable label	Explanation
GENERAL INFORMATION	
Coder name	Record your name
Notes	Record any notes important for the team INDICATE IF IT IS LINKED STUDY HERE AND THE EPPI OF THE MAIN STUDY (for example "Linked to study 1111111").
PUBLICATION INFORMATION	
StudyID	This is the study ID - it should match the study ID from the Outcome Mapping Sheet (e.g., 946578). If EPPI is being used, this will be the EPPI ID
EstimateID	The estimate ID will provide a specific number for each effect size extracted and should include the original study number, underscore, then the unique ID number (e.g., 946578_1, 946578_2 and so on)
Study status	Select one of the following: 1) Completed; 2) Protocol; 3) Ongoing
Author name	Author last name For 1 author: leading author last name (e.g. Gomez) For 2 authors: both author last names with ampersand in between (e.g. Smith and Bahn) For 3 or more authors: leading author last name followed by et al. (e.g. Gupta et al.)
Year of publication	Year published (publication date, not preprint or first online publication dates)
Publication type	Select one of the following: i) Journal article; ii) Book or chapter; iii) Report or working paper; iv) Conference proceedings; v) Published protocol (select if a published protocol, registration, or pre-analysis plan; vi) Ongoing study (select if not a registered study or published protocol but, for example, a description of the study on an organizational or authors' personal webpage)
INTERVENTION INFORMATION	
Intervention sub-group	Choose one or more intervention sub-group code(s) for each corresponding effect size: Choose one or more intervention sub-group code(s) for each corresponding effect size: Production system: i) Education / information - other educational programs ii) Education / information - Agricultural extension programs iii) Livestock access iv) Land markets & management v) Education / information - Farmer field schools vi) Other ag inputs vii) Ag credit / savings Behavior change communication: i) Classes ii) Peer support/counsellors iii) Community meetings Availability and affordability: i) Direct provision of foods
Intervention name and abbreviation	Provide name and abbreviation of the intervention and its different components. Include details of sections of the manuscript and page numbers where authors describe name of the intervention. Please only report the intervention related to the estimate if a paper reports on more than one intervention or treatment arm. Elements here should coincide with the previous and next columns meaning that the name of the intervention taken from the study should match with the intervention subgroup. and intervention description If you feel that the description does not coincide with the previous and next columns, complete all columns as much as you can and leave a comment indicating if you think information across the 3 intervention columns here do not match. Intervention name and abbreviation (if any or put N/A if no name) appears here. Information should reflect the evaluated intervention. For example: "SHoMaP

Variable label	Explanation
	(Smallholder Horticulture Marketing Program). Name of the intervention: ___
Intervention description	Provide detailed description of the intervention and its different components. Include details of sections of the manuscript and page numbers where authors describe details of the intervention. Please only report the intervention related to the estimate if a paper reports on more than one intervention or treatment arm. Elements here should coincide with the 2 previous columns meaning that the description of the intervention taken from the study should match with the intervention subgroup and intervention name. If you feel that the description does not coincide with the two previous columns, complete all columns as much as you can and leave a comment indicating if you think information across the 3 columns here do not match. What is the intervention? ____ How did it work? ____ Where and when did it happen? ___ Any specification? ___ Also, reminder: any time you use verbatim, or you use data remember to put the page number, so it is easy to check and verify.
Country	Country of intervention
Non-staggered intervention	Have the treated observations been exposed to the intervention for the same amount of time? 1=Yes; 0=No
Year of the intervention	The earliest date (year) observations are exposed to the intervention.
Length of follow up	How many months have elapsed between the start of the intervention (earliest date observations are exposed to the intervention) and the date of the final outcome measurement. If less than one month, use decimals (e.g. one week would be .25, etc.).
Exposure to intervention	For how long are the observations exposed to the intervention (in months)? If less than one month, use decimals (e.g. one week would be .25, etc.). Note: If the intervention is active throughout the evaluation period, this value will be the same as the length of follow up. Answer here cannot be greater than in the length of the follow up. Minimum is 0.25 (cannot be 0).
METHOD INFORMATION	
Evaluation Design	Select one of the options below: 1. Experimental (defined as prospective randomized assignment, where randomisation is implemented by researchers (or by decision makers in the context of an evaluation study) 2. Quasi-experimental (including natural experiments and non-randomized studies)
Evaluation Method	● If Experimental, then select: i) Randomised controlled trial ● If Quasi-experiment or natural experiment, then select one of the following: i) Natural experiment in which exposure to treatment is random; ii) Regression Discontinuity Design (RDD); iii) Difference-in-Differences (DID) / Fixed effects estimation; iv) Instrumental variable (IV) estimation; v) Endogenous treatment-effects models (including endogenous switching regression, and other methods synonymous to the Heckman two step model); vi) Statistical matching (includes PSM or statistical weighting) vii) Interrupted time series (ITS); viii) Synthetic controls
Method description	Provide a brief description of the method applied and note if any methods have been combined.
Study population	Provide any details in the paper that describe how the study population was selected, covering: a) How is the population selected? what is the sampling strategy to recruit participants from that population into the study? b) What are the characteristics of study participants? Targeted population appears first in the cell. Explanations can follow afterwards Need to add all specified information in the study on the targeted population: subsistence farmers, commercial farmers, ... (look in data section).
Additional methods	Describe any additional methods used in analysis. If none, select not applicable.
ESTIMATE INFORMATION	

Variable label	Explanation
Analysis type for this effect size	Free text, what type of analysis was used (OLS regression, Probit regression, 2SLS, ANCOVA, etc.)
Treatment effect estimated	1=Intention to Treat (ITT), 2=Average Treatment Effect on the Treated (ATET), 3=Average Treatment Effect (ATE) 4 = Local Average Treatment Effect (LATE), 5=Other
Treatment effect estimated other	Provide details if other treatment effect estimated Also include the relevant matching algorithm here (ex: kernel, nearest neighbour, etc...)
Unit of analysis	What is the unit of analysis? UOA for this effect size: 1= Individual, 2= Household, 3= Group (e.g., community organization), 4= Village, 5 = Other, 6 = Not clear If OTHER, ALWAYS PLEASE SPECIFY with a comment in this cell. For example: 5 = Other, comment in this cell: "crop level" or 5 = Other, comment in this cell: "district level". In some cases, authors may use a different word for the choices 3= Group or 4= Village, but still use these 2 options. For example, if authors conduct analysis at the village leader, select 4. If authors conduct the analysis at the cooperative level, choose 3.
Covariate adjustment	Did the regression specification control for variables other than the treatment variable? 1= Yes; 0 = No. This includes matching variables.
Covariate adjustment description	List the control variables included in the specific specification related to the estimate (including any noted fixed effects).
Source	Note the page number, table number, column, and row you used to extract the estimate data [Open Answer]
OUTCOME DATASET	
Outcome sub-group	Choose an outcome sub-group code for each corresponding estimate: Resources i) Access to economic and livelihood resources ii) Ownership of land and assets iii) Control over resources iv) Time use Agency i) Decision-making ii) Women's rights iii) Collective action and leadership Achievements i) Gender transformative outcomes ii) Improved systems and policymaking iii) Self-esteem iv) other empowerment outcomes and indices
Outcome name	Record the outcome for the corresponding effect size. Use this open answer field to enter, in the author's own words, the name of the outcome. Need to be crop specific if specified. Crop

Variable label	Explanation
Outcome description	Record the outcome for the corresponding effect size. Use this open answer field to enter, in the author's own words, a description of the outcome. Be selective and concise with the excerpts being transcribed here as to ensure accurate and precise descriptions of the outcome. Include information about the unit of the outcome and how it has been measured. Include page numbers with every excerpt extracted. all units (ex: kg/hectare). level (per household, per capita,…) all crops or combination or single crop. Time period of measure (last 12 months, last season, last 7 days, last 24 hours) If measure standardized by standard deviation (nothing to specify if not) ALSO SPECIFY THE currency AND SPECIFY IF OUTCOME IS IN LOG. Don't need to specify type (binary, continuous, etc. as in the following column, also to check) Examples: "Maize yields kg/hectare in last harvest" "HDDS (out of 12 food groups over 7 days)" "Number of consumer durables (out of 20)" Need to be crop specific if specified. Crop
Outcome measurement	How was the data collected? 1=Self-reported, 2=Administrative data, 3=Satellite data
Outcome type	Record the type of outcome variable: 1=Continuous; 2=Discrete (including proportions); 3=Nominal (binary); 4=Ordinal (binary); 5=Nominal (non-binary); 6=Ordinal (non-binary); 7=Interval.
Levels or changes	0 = Unit is the level of outcome variable, 1 = Change in outcome variable
Reverse sign	Record no=0 if an increase is good, record yes=1 if a decrease is good and the sign needs to be reversed (i.e., decrease is good)
Outcome dataset	Record if data for this outcome comes from an identified dataset
TREATMENT VARIABLE INFORMATION	
Treatment	Record the treatment variable as written in the model (e.g., the variable name the author uses, such as "Marketing contract" or "Production contract"). This column enables to distinguish what treatment is evaluated here for this specific estimate. This is very important as many studies have multiple treatments. FOR EXAMPLE: 74713715 (Benali 2017). Authors evaluated the impact of CF offered by two export supply chain (ESC) actors HVESC and RESC. The first few estimates are from the combined effect of both actors ("ESC" in the article) and other are from HVESC specifically and others are from RESC. Here this information should appear in this column so ESC, HVESC or RESC.
Treatment type	Describe the types of treatment variable used: i) binary; ii) continuous; iii) categorical; iv) other
Comparison	1=No intervention (service delivery as usual), 2=Other intervention, 3=Pipeline (waitlist) control (still service delivery as usual) 4. Other
Describe comparison group	Describe the comparison group
Subgroup	Is this analysis of a subgroup or estimating heterogeneous effects? 0=no, 1=yes
Subgroup information	Describe the subgroup or variable interacted with the treatment variable (e.g., boys, girls), etc. If no subgroup or heterogeneity analysis, select not applicable
ESTIMATE DATA	
Mean treatment (Pre)	Outcome mean for the treatment group (pre-intervention)
SD treatment (Pre)	Outcome standard deviation for treatment group (pre-intervention)
Mean Control (Pre)	Outcome mean for the comparison group (pre-intervention)
SD Control (Pre)	Outcome standard deviation for control group (pre-intervention)
Mean treatment (Post)	Outcome mean for the treatment group (post-intervention)
SD treatment (Post)	Outcome standard deviation for treatment group (post-intervention)
Mean Control (Post)	Outcome mean for the comparison group (post-intervention)

Variable label	Explanation
SD Control (Post)	Outcome standard deviation for control group (post-intervention)
SD pooled (Pre)	Outcome standard deviation for pooled group (treatment and control) (pre-intervention)
SD pooled (Post)	Outcome standard deviation for pooled group (treatment and control) (post-intervention)
SD pooled (PP)	Outcome standard deviation for pooled group (treatment and control) (includes pre and post intervention data)
Mean difference	Overall mean difference (treatment - control)
SE difference	Standard error of the overall mean difference In the results table under the coefficients and in parentheses, authors do not always report SE but sometimes provides the t-stat or the p-values. Make sure you extract the correct information in the relevant column. SE in SE reg, T-stat in Tstat reg, etc.
Tstat difference	t-statistic of mean difference In the results table under the coefficients and in parentheses, authors do not always report SE but sometimes provides the t-stat or the p-values. Make sure you extract the correct information in the relevant column. SE in SE reg, T-stat in Tstat reg, etc.
Odds ratio	Odds ratio reported in the study
SE odds ratio	Odds ratio standard error reported in the study
Risk ratio	Risk ratio reported in study
SE Risk Ratio	Risk ratio standard error
Coeff reg	Report the regression coefficient of the treatment effect
SE reg	Report the associated standard error of the regression coefficient. In the results table under the coefficients and in parentheses, authors do not always report SE but sometimes provides the t-stat or the p-values. Make sure you extract the correct information in the relevant column. SE in SE reg, T-stat in Tstat reg, etc.
Tstat reg	Report the associated t statistic of the effect size (coefficient/SE) In the results table under the coefficients and in parentheses, authors do not always report SE but sometimes provides the t-stat or the p-values. Make sure you extract the correct information in the relevant column. SE in SE reg, T-stat in Tstat reg, etc. You can mentally calculate the t-statistic = coefficient estimate / standard error. DO NOT REPORT THE CALCULATE THE T-STAT IF IT NOT PROVIDED IN THE TEXT. If MENTALLY calculated t-stat or t-stat in the text is more than 10 or less than -10, simply check again the coef and se or t-stat. If still more than 10 or less than -10, put a comment indicating that you have checked this, and this is what is actually reported in the paper.
CI_LB reg	Report the associated Lower bound of the 95% Confidence interval of the effect size. If CI is reported for a different confidence level, indicate that in the notes section.
CI_UP reg	Report the associated Upper bound of the 95% Confidence interval of the effect size. If CI is reported for a different confidence level, indicate that in the notes section.
P value exact	Exact p value if given, if not, record as written in the manuscript (e.g., p < .001, or p > .05) In the results table under the coefficients and in parentheses, authors do not always report SE but sometimes provides the t-stat or the p-values. Make sure you extract the correct information in the relevant column. SE in SE reg, T-stat in Tstat reg, etc.
Interaction term 1 coeff	Required if requested to extract information from an interaction term (in addition to a single term) For estimates based on interaction term (Y=B1treat+B2treat*female), It is very

Variable label	Explanation
	important to extract both B1 (and associated SE, T-stat or p-value) in column Coeff reg and B2 (and associated SE, T-stat or p-value) in THIS column.
Interaction term 1 SE	Required if requested to extract information from an interaction term (in addition to a single term)
Interaction term 1 Tstat	Required if requested to extract information from an interaction term (in addition to a single term)
Interaction term 1 CI_LB	Required if requested to extract information from an interaction term (in addition to a single term)
Interaction term 1 CI_UP	Required if requested to extract information from an interaction term (in addition to a single term)
Interaction term 1 P value exact	Required if requested to extract information from an interaction term (in addition to a single term)
Clusters treatment	Number of clusters - treatment group Check thoroughly if authors cluster the standard errors for this estimate. Most often it is specified in the model or under the table (control + F and search for "cluster").
Clusters control	Number of clusters - control group Check thoroughly if authors cluster the standard errors for this estimate. Most often it is specified in the model or under the table (control + F and search for "cluster").
Clusters total	Number of clusters - total sample Check thoroughly if authors cluster the standard errors for this estimate. Most often it is specified in the model or under the table (control + F and search for "cluster").
N treatment	Sample size - treatment group DO NOT LEAVE EMPTY. IF THE ANSWER IS NOT CLEAR HERE, INDICATE THIS. Information on sample size is very important. Either 1) information on N treatment and N control or 2) information on N total is available.
N control	Sample size - control group DO NOT LEAVE EMPTY. IF THE ANSWER IS NOT CLEAR HERE, INDICATE THIS. Information on sample size is very important. Either 1) information on N treatment and N control or 2) information on N total is available.
N total	Sample size - total sample DO NOT LEAVE EMPTY. IF THE ANSWER IS NOT CLEAR HERE, INDICATE THIS. Information on sample size is very important. Either 1) information on N treatment and N control or 2) information on N total is available.
periods	Record how many time points (e.g. measurement points) there are in the analysis (e.g., cross sectional data is 1, panel data with 3 measurements is 3). DO NOT LEAVE EMPTY. IF THE ANSWER IS NOT CLEAR HERE, INDICATE THIS.
Does the sample size need to be corrected?	Often in panel data, models will report number of observations rather than number of participants. In this column you will indicate 1="Yes" if the sample size needs to be divided by the number of periods, and 0="No" if either it is cross-sectional data, or if the authors have already divided the number of observations by the number of panel assessments and thus no correction is necessary. DO NOT LEAVE EMPTY. IF THE ANSWER IS NOT CLEAR HERE, INDICATE THIS.
Source	Note the page number, table number, column, and row you used to extract the data
UNIT OF ANALYIS ERROR	
M: number of observations per cluster (unit of treatment allocation)	Example 1: Explanation column: Intervention (FFS) at the village level (p8 and p11) and analysis at the Household. There are approximately 2.5 villages per district and

Variable label	Explanation
c: intra-cluster correlation coefficient. Assumed at 0.05 for the moment.	the sample of 1986 households 108 district. m=1986/ (108*2.5) c=0.05
Explanation	Example 2: Explanation column: Intervention (land titles) at the household level (p759) and analysis at the Plots. p759: 325 farm households. In total, 1,678 plots were included in the survey. So, there are 3.6 plots per household on average. m=1678/325=5.16 c=0.05 Example 3: Explanation column: Intervention (soil and water conservation) at the village level (p27) and analysis at the Household. There were 1218 hh in 139 villages. m=1218/138=8.76 c=0.05
OTHER INFORMATION	
Other linked to previous columns	Provide any other relevant information from the study

Appendix 5 – Provisional quantitative data extraction for the REA

Continuous outcomes

For studies reporting regression results for continuous outcomes, we will standardize the effect sizes following the approach suggested by Keef and Roberts (2004). This includes dividing the regression coefficient (β) by the pooled standard deviation (SD) of the outcome.

$$SMD = \frac{\beta}{SD_{pooled}} \quad \text{(i)}$$

When using parsimonious regression specifications, this approach is analogous to Cohen's d (d), which is the difference in means between the treatment and control (or comparison) group divided by the pooled SD of the outcome (i.e. the standardized mean difference). Because Cohen's d can be biased in cases where sample sizes are small, in all cases we will simply adjust d using Hedges' method. This transformation adjusts Cohen's d to Hedges' g using the following formula (Ellis 2010):

$$g \cong d(1 - \frac{3}{4(n_T+n_C)-9}) \quad \text{(ii)}$$

Where n denotes the sample size of the treatment (nT) and control (nC) groups. If the intervention is expected to change the SD of the outcome variable, we will use the SD of the control group to compute d instead. If the study does not report the pooled SD but information about sample size is available for both the treatment and control groups, we will use regression coefficients and standard errors (SEs) or t statistic (t) to calculate the following:

$$d = t\sqrt{\frac{1}{n_T} + \frac{1}{n_C}} \quad \text{(iii)}$$

Alternatively, when only information on the total sample size (N) is available, we will use the following formula suggested by (Polanin and Snilstveit 2016):

$$d = \frac{2t}{\sqrt{N}} \quad \text{(iv)}$$

$$Var_d = \frac{4}{N} + \frac{d^2}{4N} \quad \text{(v)}$$

Here we will calculate the t-statistic (t) by dividing the coefficient by the SE. If the authors only report confidence intervals (CI) and no SE, we will calculate the SE from the confidence intervals:

$$SD = \sqrt{N}\ \frac{(upper\ CI - lower\ CI)}{I} \quad \text{(vi)}$$

where I is 3.29 for estimates using 90% confidence intervals, 3.92 for 95% confidence intervals, and 5.15 for 99% confidence intervals.

In cases in which significance levels are reported rather than t or a beta coefficient (b) with the associated SE, then we will impute t using a t-distribution table. If the precise probability value is not reported, then we will assume the following:

Prob > 0.1: t = 0.5

0.1 ≥ Prob > 0.05: t = 1.645

0.05 ≥ Prob > 0.01: t = 1.960

0.01 ≥ Prob: > 0.001: t = 2.576

Prob: ≥ 0.001: t = 3.291

In some cases, the studies we include in the review may not report a regression coefficient, but the group means ($\bar{X}$) and pooled SD for treatment and control group at follow up only (p + 1). Here we will calculate d using formulae provided in (Borenstein et al. 2009):

$$d = \frac{\bar{X}_{T_{p+1}} - \bar{X}_{C_{p+1}}}{SD_{p+1}} \quad \text{(vii)}$$

If the study does not report the pooled SD, it is possible to calculate it using the following formula:

$$SD_{p+1} = \sqrt{\frac{\left(n_{T_{p+1}} - 1\right) SD^2_{T_{p+1}} + \left(n_{C_{p+1}} - 1\right) SD^2_{C_{p+1}}}{n_{T_{p+1}} n_{C_{p+1}} - 2}} \quad \text{(viii)}$$

For studies reporting the difference in treatment and control group means and the pooled SDs at baseline (p) and follow up (p + 1):

$$d = \frac{\Delta\bar{X}_{T_{p+1}} - \Delta\bar{X}_{T_p}}{SD_p} \quad \text{(ix)}$$

Finally, for studies reporting mean differences between treatment and control group, standard error (SE) and sample size (n):

$$d = \frac{\Delta \bar{X}_{T_{p+1}}}{SE\sqrt{n}} \quad \text{(x)}$$

Outcomes measured as proportion of individuals

If outcomes are reported in proportions of individuals, we will calculate the Cox-transformed log odds ratio effect size (Sánchez-Meca, Marín-Martínez, and Chacón-Moscoso 2003):

$$d = LogOddsRatio\ \frac{\sqrt{3}}{\pi} \quad \text{(xi)}$$

Outcomes measured as proportion of events or days

If outcomes are reported based on proportions of events or days, we will use the standardized proportion difference effect size:

$$d = \frac{w_T - w_C}{SD(w)} \quad \text{(xii)}$$

Where w_T is the proportion in the treatment group and w_C the proportion in the comparison group, and the denominator is given by:

$$SD(w) = \sqrt{w1 - w} \quad \text{(xiii)}$$

Here w is the weighted average of w_T and w_C:

$$w = \frac{n_T w_T + n_C w_C}{n_T + n_C}$$

Appendix 6 – Criteria determining selection of effect estimates for data extraction

We will extract effects reported across different interventions, outcomes, and subgroups within a study. We will address dependent effect sizes using data processing and selection techniques. We will utilize several criteria to select one effect estimate per outcome per study:

- Where studies report effects from multiple model specifications, we will use the author's preferred model specification. Only if the preferred specification is unclear, we will use the most precise estimate of the treatment effect (measured by the one with the largest t-value). This reflects regression adjustments in designs, such as RCTs, IV, RDDs, are usually made on grounds of model efficiency but explorative specifications (e.g. including interaction terms, higher order terms, etc.) may not be the most efficient estimates (in fact they could be highly inefficient) where the added terms are not significant.
- Where studies report effects from multiple estimators, we will use the author's preferred specification. Only if the preferred specification is unclear, we will use the specification that appears most robust to falsification tests (e.g. according to sensitivity analysis for propensity score matching or placebo tests for difference-in-difference estimators).
- Where different studies report on the same program but use different samples (e.g., from different regions), we will include both estimates, treating them as independent samples, provided effect sizes are measured relative to separate control or comparison groups.
- Where studies report evidence according to subgroups of participants, we will record and report data on relevant subgroups separately.
- For studies with outcome measures at different time points, we will synthesize short- and long-term outcomes separately, following Rue and colleagues (2013).
- When studies include multiple outcome measures to assess related outcome constructs, we will follow our pre-specified preferred outcome order (described below) without reference to the results.
- When studies report multiple outcome subgroups for the same outcome construct, but do not present an effect for the full sample, we may calculate a

"synthetic effect size" using the sample-weighted average, and applying appropriate formulae to recalculate variances (Borenstein, Hedges, and Rothstein 2009).

- If studies include multiple treatment arms with only one control group and the treatments represent separate treatment constructs, we will calculate the effect size for treatment A versus control and treatment B versus control and include them in separate meta-analyses according to the intervention type.

This also includes criteria prioritizing specific outcome measures within included studies:

- Our analysis will prioritize synthesizing outcomes using composite or aggregate indicators. If a study does not report a composite measure, we will use the outcome that most closely relates to the intervention type and perform outcome mapping to identify the outcome in each study that appears most frequently across studies.
- Where an intervention targets or concerns one specific commodity, our analysis will use commodity specific outcomes for price, production, and intermediate outcomes. If the intervention targets and reports outcomes for more than one commodity, we will use the most frequent type of commodity specific outcome reported across studies. If a study reports outcomes for both arable and permanent crops, we will select one prioritized arable crop and one prioritized permanent crop to perform sensitivity analysis on the type of crop prioritized for this sample of studies. We will also distinguish between staple and cash crops, again taking the most frequent staple or cash crop reported across studies as the prioritized outcome.
- Specific preferences for measures of outcome constructs are also outlined in the table below. We will consider, where possible, creating combined and separate meta-analysis for different measures of the same outcome construct or control for groups of outcome measures in a meta-regression.

The number of possible ways any outcome construct can be measured is often large. The list of priority outcomes in the table below are unlikely to be entirely comprehensive given it only accounts for some common measures which are presented simultaneously in studies (which creates the need to choose a preferred outcome for the synthesis). We will consult subject experts, without reference to the

results of a study, to establish an order of preference should studies present multiple measurements of an outcome construct that is not already captured by the priority criteria established in this protocol.

Appendix 7 – Risk of bias assessment tool for REA

Questions	Explanations/Answers
EstimateID	The estimate ID will provide a specific number for each effect size extracted and should include the original study number, underscore, then the unique ID number (e.g., 946578_1, 946578_2 and so on)
Intervention description	Use a few words to describe the intervention. Should be same as in the QEX (column K).
Intervention sub-group	Should be same as in the QEX (column J).
	Production system: i) Education / information - other educational programs ii) Education / information - Agricultural extension programs iii) Livestock access iv) Land markets & management v) Education / information - Farmer field schools vi) Other ag inputs vii) Ag credit / savings Behavior change communication: i) Classes ii) Peer support/counsellors iii) Community meetings Availability and affordability: i) Direct provision of foods
Outcome sub-group	Should be same as in the QEX (column AD).
	Resources i) Access to economic and livelihood resources ii) Ownership of land and assets iii) Control over resources iv) Time use Agency i) Decision-making ii) Women's rights iii) Collective action and leadership Achievements i) Gender transformative outcomes ii) Improved systems and policymaking iii) Self-esteem iv) other empowerment outcomes and indices
Outcome description	Use a few words to describe the outcome. Should be same as in the QEX (column AE and AF).
Evaluation Method	Should be same as in the QEX (column R).
	1: Randomised controlled trial 2: Natural experiment 3: RDD 4: DiD & FE 5: IV 6: Endogenous treatment-effect models 7: Statistical matching 8: ITS (Interrupted-time series)
Evaluation Method description	Provide a short description of the evaluation method referencing the study (with the page number) (column S). For example: "In this study, we address the problem of selection on unobservable by combining PSM with the use of the double-difference (DD) estimator" pXXX.

Questions	Explanations/Answers
1. Unconfoundness and absence of selection bias	Q1. Was the allocation or identification mechanism able to control for selection bias? AND Was the evaluation method executed adequately to ensure comparability of groups throughout the study and prevent confounding? Yes Probably yes Probably No No Insufficient information
	Yes or **Probably Yes** if the study satisfies ALL of the following requirements (depending on the evaluation method) for this estimate (**No** or **Probably No** otherwise): If **Randomised controlled trial or a Natural experiment**: a) Centralized and independent treatment randomization at the start of the intervention is described (lottery, coin toss, random number generator). OR if randomization was done in sequence, authors provide detail on the exact settings and participants attending the lottery. OR if randomization was done in a special way, it is justified given the study setting (stratification, pairwise matching, unique random draw, multiple random draws etc.). b) A balance table is reported suggesting that allocation was random between all groups including subgroup receiving different treatment within control or treatment groups (if the comparison is relevant for this assessment). If **RDD**: a) Treatment assignment is made based on a pre-determined and independent discontinuity on a variable (assignment variable) AND the unit of analysis cannot manipulate the assignment variable. b) The mean of the covariates of the individuals immediately at both sides of the cut-off point are NOT statistically different OR authors control for the differences in the estimation model. c) Placebo tests are conducted to verify robustness. If **DiD & FE (also if combined with statistical matching)**: a) Authors use pre-treatment data to support the parallel trend assumption in the absence of the treatment OR in the absence of several rounds of pre-treatment data, authors control for differences in trends in the treatment and control groups. b) Authors control for relevant time-varying and constant differences between treatment and control groups observations OR placebo tests are conducted to verify robustness. c) If the intervention delivery is staggered, authors apply adjustments or use statistical matching. If **IV or Endogenous treatment-effect models**: a) Instrument(s) is(are) strongly correlated with the endogenous variable (indicate in the justification column, the relevance test results) OR authors use weak instrument and apply adjustments. b) Authors provide a convincing discussion that the instrument(s) is(are) exogenously generated (not correlated with the error term), for example, due to a "natural" experiment or random allocation. If **Statistical matching**: a) Authors use all relevant baseline and/or exogenous covariates (variables should not be affected by the treatment). b) Authors show based on diagnostics that the covariates are balanced after-matching across treatment and control groups' observations (with the exception of Kernel matching). c) Rosenbaum's test (or synonymous sensitivity analysis) is displayed and suggests that the estimate is not sensitive to the existence of hidden bias (bounds critical gamma cutoff value is >= 2). If **ITS**: a) Authors discuss and address non-stationarity (including seasonality). b) Authors control for other time-varying confounders such as other events OR authors use a valid control group or control outcome (unaffected by the

Questions	Explanations/Answers
	intervention). **Insufficient information**: insufficient information.
1. Justification	Q1. Answer justification (use references with page numbers).
2. Absence of non-random attrition AND significant missing data	Q2. Was the analysis conducted in the absence of non-random attrition and significant missing data? Yes Probably yes Probably No No Insufficient information
	Yes or **Probably Yes** if the study satisfies ALL of the following requirements for this estimate (**No** or **Probably No** otherwise): a) Attrition rate is less than 5% OR less than or equal to assumed in power calculations. b) Attrition is random (balance between attritors and non-attritors). OR Attrition is not random but authors apply convincing statistical techniques to identify and address the attrition bias (for example, only using observations in the analysis present in all rounds of the data). c) Analysis is conducted using most of the collected data (missing data is less than 10%) (for example, the data section reports that data was collected on 600 households and the number of observations used in the analysis is close to 600). **Insufficient information**: insufficient information.
2. Justification	Q2. Answer justification (use references with page numbers).
3. Absence of spillovers/crossovers and contamination	Q3. Was the study adequately protected against spillovers, crossovers, and contamination? Yes Probably yes Probably No No Insufficient information
	Yes or **Probably Yes** if the study satisfies ALL (a and b) of the following requirements for this estimate (**No** or **Probably No** otherwise): a) Treatment and control observations are sufficiently far away (geographically and/or socially) from one another and general equilibrium effects are unlikely so that the intervention is unlikely to spill-over to the control group observations. b) Treatment and control groups are isolated from other interventions which might affect the outcomes. Problems with crossovers and drop-outs are dealt with using intention to-treat analysis (ITT) or Local Average Treatment Effect (LATE) analysis. **Insufficient information**: insufficient information.
3. Justification	Q3. Answer justification (use references with page numbers).
4. Absence of outcome measurement bias	Q4. Was the outcome measured in the same way between study arms and outcome measurement was not affected by knowledge of the intervention? Yes Probably yes Probably No

Questions	Explanations/Answers
	No Insufficient information
	Yes or **Probably Yes** if the study satisfies the following requirement for this estimate (**No** or **Probably No** otherwise): a) Outcomes were not self-reported by participants (for example outcomes come from administrative records) OR outcomes are self-reported **but respondents are unlikely to be influenced by knowledge of administration of the intervention OR outcome assessors were blinded**. **IF RELEVANT:** **b) For self-reported outcomes: respondents in the intervention group are not more likely to have accurate answers due to recall bias;** **If relevant, discuss here how recall data bias may affect the outcome measurement of the intervention group. If it affects both treatment and control groups in a similar way, do not consider this as a risk of bias.** **Insufficient information**: insufficient information.
4. Justification	Q4. Answer justification (use references with page numbers).
5. Absence of reporting bias	Q5. Was the study free from selective analysis reporting? Yes Probably yes Probably No No Insufficient information
	Yes or **Probably Yes** if the study satisfies the following requirement for this estimate (**No** or **Probably No** otherwise): a) There is no evidence that outcomes were selectively reported (e.g. all relevant outcomes in the methods section are reported in the results section) OR There is only one possible way in which the outcome domain can be analyzed OR researchers have provided the reasons for any inconsistencies (not related to the nature of the results). **To answer yes here, the authors need to provide details of a pre-analysis plan and the included outcomes should be consistent with the ones discussed in the study.** **If the pre-analysis is available and it does not contain the extracted outcome you are considering (specific row here), unless author(s) provide an explanation in the included study, it should be considered as potential selective reporting (and therefore coded as Probably no or No).** **Insufficient information**: insufficient information (if no pre-analysis plan).
5. Justification	Q5. Answer justification (use references with page numbers).
6. Unit of analysis	Q6. Unit of analysis: Is unit of analysis in cluster allocation addressed in standard error calculation?
	Yes or **Probably Yes** if the study satisfies the following requirement for this estimate (**No** or **Probably No** otherwise): **Yes** if Unit of analysis (UoA) = Unit of randomization (UoR) OR if UoA ≠ UoR and standard errors are clustered at the UoR level OR data is collapsed to the UoR level **Insufficient information**: if not enough information is provided on the way the standard errors were calculated or what the unit of analysis is. "Not applicable" if it is not a cluster RCT.
6. Justification	Q6. Answer justification (use references with page numbers).

Questions	Explanations/Answers
7. Summary and implications	Q7. What are the key risks of bias identified, and what are their potential implications for interpreting the effects?
	Add a brief summary that focuses on the likely implications for interpreting the effects: are the effects likely to be suppressed due to the noted sources of bias? Exaggerated? Are there substantial quality issues or lack of clarity that should be considered? What should the reader keep in mind, to contextualize the findings?
Overall score	• "High risk of bias": if any of the bias domains were assessed as "No" or "Probably No". • "Some concerns": if one or several domains were assessed as "Insufficient Information" and none were "No" or "Probably No". • "Low risk of bias": if all of the bias domains were assessed as "Yes" or "Probably Yes".
Notes	Notes

Appendix 8 – Qualitative critical appraisal tool

Study type	Methodological appraisal criteria	Response		
		Yes	No	Comment
Screening questions: assessing 'fatal flaws' (Dixon-Woods 2005) *Configurative 'fatal flaws' based on Pawson (2003) TAPUS framework*	Configurative assessment: ✓ Study reports primary data and applied methods ✓ Study states clear research questions and objectives ✓ Study states clear research design, which is appropriate to address the stated research question and objectives (*Purposivity*) ✓ The findings of the study are based on collected data, which justify the knowledge claims (*Accuracy*)			
	Screening question based on abstract and/or superficial reading of full text: Further appraisal is not feasible or appropriate when the answer is 'No' to any of the above screening questions!			

Study type	Methodological appraisal criteria	Response		
		Yes	No	Comment / Confidence judgment
1. Qualitative and descriptive quantitative, and process evaluations	**I. RESEARCH IS DEFENSIBLE IN DESIGN (**providing a research strategy that addresses the question) Appraisal indicators: ✓ *Is the research design clearly specified and appropriate for aims and objectives of the research?* Consider whether			
	i. there is a discussion of the rationale for the study design			
	ii. the research question is clear, and suited to the inquiry			
	iii. there are convincing arguments for different features of the study design			
	iv. limitations of the research design and implications for the research evidence are discussed			
	Defensible / **Arguable** / **Critical** / **Not defensible**	*Worth to continue:*		
	II. RESEARCH FEATURES AN APPROPRIATE SAMPLE (following an adequate strategy for selection of participants)			

	Appraisal indicators: Consider whether						
	i. there is a description of study location and how/why it was chosen						
	ii. the researcher has explained how the participants were selected						
	iii. the selected participants were appropriate to collect rich and relevant data						
	iv. reasons are given why potential participants chose not take part in study						
	Appropriate sample	**Functional sample**	**Critical sample**	**Flawed sample**	*Worth to continue:*		
	III. RESEARCH IS RIGOROUS IN CONDUCT (Providing a systematic and transparent account of the research process) Appraisal indicators: Consider whether						
	i. researchers provide a clear account/description of the process by which data was collected (e.g. for interview method, is there an indication of how interviews were conducted? /Procedures for collection or recording of data?)						
	ii. researchers demonstrate that data collection targeted depth, detail and richness of information (e.g. interview/observation schedule)						
	iii. there is evidence of how descriptive analytical categories, classes, labels, etc. have been generated and used						
	iv. presentation of data distinguishes clearly between the data, the analytical frame used, and the interpretation						
	v. methods were modified during the study; and if so, has the researcher explained how and why?						
	Rigorous conduct	**Considerate conduct**	**Critical conduct**	**Flawed conduct**	*Worth to continue:*		
	IV. RESEARCH FINDINGS ARE CREDIBLE IN CLAIM/BASED ON DATA (Providing well-founded and plausible arguments based on the evidence generated) Appraisal indicators: Consider whether						
	i. there is a clear description of the form of the original data						
	ii. sufficient amount of data is presented to support interpretations and findings/conclusions						
	iii. the researchers explain how the data presented were selected from the original sample to feed into the analysis process (i.e. commentary and cited data relate; there is an analytical context to cited						

	data, not simply repeated description; is there an account of frequency of presented data?)					
	iv. there is a clear and transparent link between data, interpretation, and findings/conclusion					
	v. there is evidence (of attempts) to give attention to negative cases/outliers etc.					
	Credible claims	**Arguable claims**	**Doubtful claims**	**Not credible**	*If findings not credible, can data still be used?*	
	V. REASEARCH ATTENDS TO CONTEXTS (Describing the contexts and particulars of the study) Appraisal indicators: Consider whether					
	i. there is an adequate description of the contexts of data sources and how they are retained and portrayed?					
	ii. participants' perspectives/observations are placed in personal contexts					
	iii. appropriate consideration is given to how findings relate to the contexts (how findings are influenced by or influence the context)					
	iv. the study makes any claims (implicit or explicit) that infer generalization (if yes, comment on appropriateness)					
	Context central	**Context considered**	**Context mentioned**	**No context attention**		
	VI. RESEARCH IS REFLECTIVE (Assessing what factors might have shaped the form and output of research) Appraisal indicators: Consider whether					
	i. appropriate consideration is given to how findings relate to researchers' influence/own role during analysis and selection of data for presentation					
	ii. researchers have attempted to validate the credibility of findings (e.g. triangulation, respondent validation, more than one analyst)					
	iii. researchers explain their reaction to critical events that occurred during the study					
	iv. researchers discuss ideological perspectives/values/philosophies and their impact on the methodological or other substantive content of the research (implicit/explicit)					
	Reflection	**Consideration**	**Acknowledgement**	**Unreflective research**	*NB: Can override previous exclusion!*	
OVERALL CRITICAL APPRAISAL DECISION **Decision rule:**						

- a single critical appraisal judgement[4] in any of the 6 appraisal domains leads to a critical overall judgement.
- 2 or more high critical appraisal judgements in any of the 6 appraisal domains lead to an overall high risk of bias / low quality rating.
- 2 or more moderate critical appraisal judgements in any of the 6 appraisal domains lead to an overall moderate risk of bias / moderate quality rating.
- which means that for a study to be rated of low risk of bias / high quality at least 5 appraisal domains need be rated as of low critical appraisal.

HIGH QUALITY EMPIRICAL RESEARCH	**MODERATE QUALITY EMPIRICAL RESEARCH**	**LOW QUALITY EMPIRICAL RESEARCH**	**CRITICAL QUALITY EMPIRICAL RESEARCH**
(Study generates new evidence relevant to the review question and complies with all methodological criteria to ensure reliability and empirical grounding of the evidence).	(Study generates new evidence relevant to the review question and complies with reasonable methodological criteria to ensure reliability and empirical grounding of the evidence).	(Study generates new evidence relevant to the review question and complies with minimum methodological criteria to ensure reliability and empirical grounding of the evidence).	(The evidence generated by the study does not comply with minimum methodological criteria to ensure reliability and empirical grounding of the evidence).

Sources used in this section (in alphabetical order); Campbell et al (2003); CASP (2006); CRD (2009); Dixon-Woods et al (2004); Dixon-Woods et al (2006); Greenhalgh & Brown (2014); Harden et al (2004); Harden et al (2009); Harden & Gough (2012); Mays & Pope (1995); Pluye et al (2011); Spencer et al 2006; Thomas et al (2003); SCIE (2010).

Study type	**Methodological appraisal criteria**	**Response**		
		Yes	No	Comment /confidence judgment

[4] For the qualitative studies, we use a slightly different language to scale the critical appraisal assessments as compared to the quantitative studies. The far-right rating column always reflects a 'critical' appraisal judgement (i.e. 'unreflective research' above) with judgements moving further to the left on a scale from high to low critical appraisal.

2. Mixed-methods[2] *<u>Sequential explanatory design</u>* *The quantitative component is followed by the qualitative. The purpose is to explain quantitative results using qualitative findings. E.g., the quantitative results guide the selection of qualitative data sources and data collection, and the qualitative findings contribute to the interpretation of quantitative results.* *<u>Sequential exploratory design</u>* *the qualitative component is followed by the quantitative. The purpose is to explore, develop and test an instrument (or taxonomy), or a conceptual framework (or theoretical model). E.g., the qualitative findings inform the quantitative data collection, and the quantitative results allow a generalization of the qualitative findings.* *<u>Triangulation designs</u>* *the qualitative and quantitative components are concomitant. The purpose is to examine the same phenomenon by interpreting qualitative and quantitative results (bringing data analysis together at the interpretation stage), or by integrating qualitative and quantitative datasets (e.g., data on same cases), or by transforming data (e.g., quantization of qualitative data).* *<u>Embedded/convergent design</u>* *The qualitative and quantitative components are concomitant. The purpose is to support a qualitative study with a quantitative sub-study (measures), or to better understand a specific issue of a quantitative study using a qualitative sub-study, e.g., the efficacy or the implementation of an intervention based on the views of participants<u>.</u>*	**I. RESEARCH INTEGRATION/SYNTHESIS OF METHODS** (Assessing the value-added of the mixed-methods approach) Applied mixed-methods design: ○ Sequential explanatory design ○ Sequential explorative design ○ Triangulation design ○ Embedded design <u>Appraisal indicators:</u> Consider whether			
	i. the rationale for integrating qualitative and quantitative methods to answer the research question is explained *[DEFENSIBLE]*			
	ii. the mixed-methods research design is relevant to address the qualitative and quantitative research questions, or the qualitative and quantitative aspects of the mixed methods research question *[DEFENSIBLE]*			
	iii. there is evidence that data gathered by both research methods was brought together to inform new findings to answer the mixed-methods research question (e.g. form a complete picture, synthesize findings, configuration) *[CREDIBLE]*			
	iv. the approach to data integration is transparent and rigorous in considering all findings from both the qualitative and quantitative module (danger of cherry-picking) *[RIGOROUS]*			
	v. appropriate consideration is given to the limitations associated with this integration, e.g., the divergence of qualitative and quantitative data (or results)? *[REFLEXIVE]*			

For mixed-methods research studies, each component undergoes its individual critical appraisal first. Since qualitative studies are either included or excluded, no combined risk of bias assessment is facilitated, and the assigned risk of bias from the quantitative component similarly holds for the mixed-methods research.

The above appraisal indicators only refer to the applied mixed-methods design. If this design is not found to comply with each of the four mixed-methods appraisal criteria below, then the quantitative/qualitative components will individually be included in the review:

Mixed-methods critical appraisal:	Qualitative critical appraisal:	Quantitative critical appraisal:
1. Research is defensible in design 2. Research is rigorous in conduct 3. Research is credible in claim 4. Research is reflective	Include / Exclude	1. Low risk of bias 2. Risk of bias 3. High risk of bias 4. Critical risk of bias

Combined appraisal:
Include / Exclude mixed-methods findings judged with _____ risk of bias

Section based on Pluye et al (2011). Further sources consulted (in alphabetical order): Creswell & Clark (2007); Crow (2013); Long (2005); O'Cathain et al (2008); O'Cathain (2010); Pluye & Hong (2014); Sirriyeh et al (2011).