

REVIEW

Open Access



Rapid evidence assessment on women's empowerment interventions within the food system: a meta-analysis

Miriam Berretta^{1*} , Meital Kupfer¹, Shannon Shisler^{1,2} and Charlotte Lane²

Abstract

Background Women's empowerment interventions represent a key opportunity to improve nutrition-related outcomes. Still, cross-contextual evidence on the factors that cause poorer nutrition outcomes for women and girls and how women's empowerment can improve nutrition outcomes is scant. We rapidly synthesized the available evidence regarding the impacts of interventions that attempt to empower women and/or girls to access, participate in and take control of components of the food system.

Methodology We considered outcomes related to food security; food affordability and availability; dietary quality and adequacy; anthropometrics; iron, zinc, vitamin A, and iodine status; and measures of wellbeing. We also sought to understand factors affecting implementation and sustainability, including equity. We conducted a rapid evidence assessment, based on the systematic literature search of key academic databases and gray literature sources from the regular maintenance of the living Food System and Nutrition Evidence Gap Map. We included impact evaluations and systematic reviews of impact evaluations that considered the women's empowerment interventions in food systems and food security and nutrition outcomes. We conducted an additional search for supplementary, qualitative data related to included studies.

Conclusion Overall, women's empowerment interventions improve nutrition-related outcomes, with the largest effects on food security and food affordability and availability. Diet quality and adequacy, anthropometrics, effects were smaller, and we found no effects on wellbeing. Insights from the qualitative evidence suggest that women's empowerment interventions best influenced nutritional outcomes when addressing characteristics of gender-transformative approaches, such as considering gender and social norms. Policy-makers should consider improving women's social capital so they can better control and decide how to feed their families. Qualitative evidence suggests that multi-component interventions seem to be more sustainable than single-focus interventions, combining a livelihoods component with behavioral change communication. Researchers should consider issues with inconsistent data and reporting, particularly relating to seasonal changes, social norms, and time between rounds of data collection. Future studies on gender-transformative approaches should carefully consider contextual norms and avoid stereotyping women into pre-decided roles, which may perpetuate social norms.

Keywords Women's empowerment, Review, Food system, Meta-analysis

*Correspondence:

Miriam Berretta

mberretta@3ieimpact.org

Full list of author information is available at the end of the article



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Introduction

Most research on women within food systems focuses on their roles as caregivers and cooks [1]. However, women are key actors within food systems, serving as producers, processors, distributors, vendors, and consumers. Often living in more vulnerable conditions than men due to societal norms, women face negative, differential access to affordable, nutritious foods relative to men. Gendered food systems interact with gender equality and equity at individual and systemic (community) levels, as well as in formal (traditions and economic roles) and informal (household norms) ways, also referred to as the four quadrants of change (Fig. 1). To achieve food systems transformation, women will need to have adequate agency and control over resources. Social norms, policies, and governance structures must be fair and equitable to allow women access to food and livelihood opportunities. However, many food systems and nutrition interventions are criticized as disempowering because they can entrench stereotypes by targeting women and girls explicitly in the roles of caregivers or cooks.

Improvements in women’s empowerment are expected to facilitate women’s interactions with the food system

and improve the nutrition of women and their communities directly and indirectly. Women can improve their own and their children’s nutritional status when they have the socio-economic power and social capital to make decisions on food and non-food expenditures and the ability to take care of themselves and their families [3]. By giving women more control and self-determination, women’s empowerment interventions are expected to have larger impacts than similar interventions that do not incorporate an empowerment approach. Women’s empowerment interventions may allow women to make the choices that are most likely to benefit them while addressing the broader social and cultural context. As a result, women’s empowerment interventions represent a key opportunity to improve nutrition-related outcomes, and women’s empowerment has been highlighted as a critical, crosscutting theme for food systems transformation [4]. However, cross-contextual evidence on the factors that cause poorer nutrition outcomes for women and how women’s empowerment can improve nutritional outcomes is still scant [2].

Gender-transformative approaches (GTA) acknowledge the equal role that all genders have in women’s

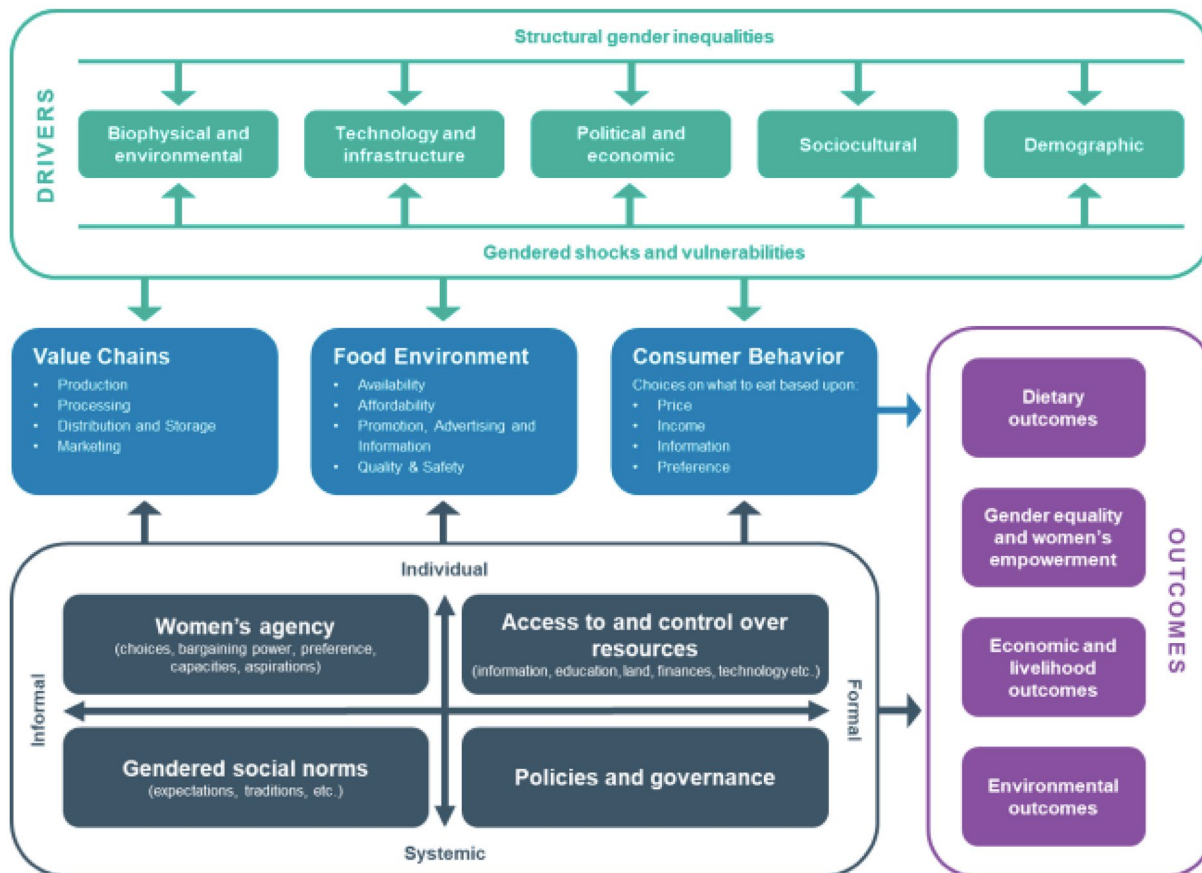


Fig. 1 Theory of change, from Njuki et al. [2]

empowerment and thus target men as agents of change to transform structural barriers and social norms [5]. While many women's empowerment interventions include GTA approaches, women's empowerment and GTA differ mainly in the following aspects (adapted from [6]):

- Approaches to women's empowerment often focus only on women. GTA, on the other hand, aim to address broader social contexts and avoid essentializing men and women.
- A central element of GTA is intersectionality, i.e., considering the interconnections between different social identities, such as gender, race, ethnicity, or geographic location.

For our purposes, women's empowerment interventions within the food system are defined as "efforts targeted at increasing women's abilities to make decisions regarding the purchase and consumption of healthy foods" based on 3ie's Food Systems and Nutrition Evidence Gap Map [1]. Moore et al. [1] determined that, as of January 2022, there were 21 evaluations of the impacts of interventions that target women's abilities to make decision regarding the purchase of healthy foods, for example by improving decision-making on household expenditures. However, these studies had not been synthesized to determine average treatment effects and key contextual factors driving to impact. In this rapid evidence assessment, we focus on 10 of those studies which looked at specific outcomes related to food security, food affordability and availability, diet quality and adequacy, anthropometrics, iron, zinc, vitamin A, iodine status, and measures of well-being.

This rapid evidence assessment provides a novel synthesis of the available evidence on the impacts of interventions to support women's empowerment within the food system, contributing to the literature base on both women's empowerment and food systems. It is expected to support policymakers, experts, and stakeholders in making evidence-informed decisions regarding the implementation and design of such interventions. Stakeholders can use this work to understand how to better integrate gender-transformative approaches as one characteristic of feminist development policies, to improve nutritional outcomes in the project and study design process while acknowledging and moving past the use of stereotypes.

In this rapid assessment, we run a meta-analysis and a barriers and facilitators analysis of interventions on the economic and social empowerment of women with the goal of providing them the means and ability to affect dietary decisions; [7, 8]. As a result, we focus on food

environment and dietary measures, a subset of the factors presented in Fig. 1. Measures of wellbeing are also considered due to their direct link with women's empowerment. The interventions we identified primarily relate to behavior change communication, skills training, and asset transfers. Interventions were often complex and integrated other components, such as microcredits, self-help groups, and provision of vitamins supplements. They often targeted men as well as women, making them gender-transformative.

Objectives and research questions

The objective of this work was to rapidly synthesize the available evidence regarding the impacts of interventions that attempt to empower women and/or girls to access, participate in and take control of components of the food system. Outcomes considered are limited to measures of the food environment and diet. This fills the synthesis gap identified by Moore et al. [1]. We also sought to understand factors affecting implementation and sustainability, including equity. We specified the following research questions a priori (Appendix 1):

1. What are the effects of women's empowerment interventions within the food system on food availability, accessibility, and affordability, of healthy diets or nutritional status?
2. Are there any unintended consequences of such interventions?
3. Do effects vary by context, approach to empowerment, or other moderators?

Methodology

To respond to these research questions, we conducted a rapid evidence assessment (REA). As far as possible this REA is based on the rigorous methodologies adopted in a systematic review [9]. However, due to time and resource limitations, the search and screening process and the data extraction process were shortened [10]. These abbreviated steps allowed for the *rapid* nature of this rapid evidence assessment. The protocol for the REA was developed a priori in February 2021 and is provided in Appendix 1.

Search and screening based on the EGM by Moore et al. [1]

We did not conduct a new search for impact evaluations, but relied on an existing, open-source evidence gap map (EGM) by Moore et al. [1]. The EGM includes all impact evaluations and systematic reviews of impact evaluations of interventions within the food system which measure outcomes related to food security and nutrition in low- and middle-income countries (Appendix 7). Because the

Table 1 PICOS

Criteria	Included	Excluded
Participants	People of any age and gender residing in low- and middle-income countries	High-income countries
Intervention(s)	Interventions aimed at increasing women's empowerment and giving women the capabilities to make decisions on the purchase and consumption of a healthy diet	All else
Comparison	Business as usual, including pipeline and waitlist controls An alternate intervention	No comparator
Outcome(s)	Food security Food affordability and availability Diet quality and adequacy Anthropometrics Iron, zinc, vitamin A, and iodine status Measures of well-being	All else
Study designs	Experimental, quasi-experimental, systematic reviews and cost evidence	Efficacy trials Before-after with no control group Cross-sectional studies, etc.

search conducted by Moore et al. [1] was not specifically focused on women's empowerment, rather it included women's empowerment among a variety of other topics, it is possible that some articles may have been missed. However, there is no reason to believe that there would have been any systematic bias in the types of articles that were omitted or that this would have meaningfully affected results.

- The search by Moore et al. [1] was extensive and systematic, covering 12 academic databases and 13 gray literature sources (Appendix 7). Single screening with safety first was used at both title and abstract and full text stages. A machine learning classifier was applied to automatically exclude studies with a low probability of inclusion. Although the original search was complete in May 2020, the search is continuously updated with studies added to the EGM through January 2022 considered for this REA. As of January 2022, over 160,000 articles were screened for inclusion in the EGM and 2,647 studies were included Appendix 7.

Because this REA is based on the search by Moore et al. [1], the same criteria for eligible populations, comparators, and study designs employed by Moore et al. [1] were used for this REA. Moore et al. [1] included interventions which targeted women's empowerment within food systems. Women's empowerment interventions which functioned outside the food system, such as those related to economic empowerment outside of the food system, were not included. From the 21 studies on women's empowerment interventions included in their EGM, we selected the ten studies evaluating outcomes related to the food

environment (food security and food affordability and availability), diet (diet quality and adequacy, anthropometrics, and micronutrient status), or well-being. Table 1 presents the population, interventions, comparisons, outcomes, and study designs (PICOS), modified from Moore et al. [1], employed by this REA.

Although we did not perform any new searches for impact evaluations for this rapid evidence assessment, we conducted a targeted search in Google Scholar looking for the qualitative papers related to included studies to allow us to investigate how impacts were achieved. The search included the name of the program or intervention, if available, as well as the country the intervention took place in. Eligible qualitative study designs were [11]:

- A *qualitative study* collecting primary data using mixed methods or quantitative methods of data collection and analysis and reporting some information on all 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 reporting some information on all 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 and what is felt to be working well and why. Process evaluations may include the collection of qualitative and quanti-

Table 2 Included outcomes and indicator extracted for evidence synthesis

Outcome	Indicators*
Food security	Preferred outcomes: food security indexes and composite scores Secondary outcome: skipped meals Tertiary outcome: reports of insufficient food
Food affordability and availability	Preferred outcome: per capita food consumption in monetary units Secondary outcome: per capita food consumption in weight Other measures, such as the cost of a food basket, will be considered if these are not available
Diet quality and adequacy	Preferred outcomes: composite diet scores such as the nutrient rich food index Secondary outcome: dietary diversity and other food variety measures Tertiary outcome: intake of specific foods
Anthropometrics	Preferred outcomes: body mass index, weight for length, length for age, weight for age Other measures, such as MUAC and ponderal index, will be considered if these are not available
Micronutrient (iron, zinc, vitamin A, iodine) status	Preferred outcome: measures of content in blood/tissue (ex. hemoglobin levels) Secondary outcome: intake in weight (grams, micrograms, etc.) Tertiary outcome: intake in percentage relative to recommended intake Other measures will be considered
Well-being	Preferred outcome: perceived well-being Secondary outcome: anxiety

*Indicators are listed by preference based on a priori specification. Such a priori specification reduces bias by preventing subjective reporting of outcomes by the team conducting the Rapid Evidence Assessment. Most indicators were ultimately not found in the studies

tative 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.

While the identification of qualitative evidence was limited to studies linked to the included impact evaluations, the process of data extraction, critical appraisal, and evidence synthesis was independent.

Data extraction

Data extraction templates were modified from 3ie's standard coding protocol for systematic reviews, reflecting another shortened step for the purposes of making this assessment rapid (Appendix 2). The primary modification to the tool was a restriction on the number and type of outcomes considered. The outcomes considered were broad and could be measured using a variety of indicators. To restrict the number of outcomes extracted, we specified preferred and secondary indicators of interest a priori (Table 2). This limited the analysis to be conducted to only the specified outcomes. Composite measures were always preferred over disaggregated ones. If multiple analyses were presented considering the same outcome (ex. Univariate analysis and a regression with control variables), the data from the model preferred by the author was extracted. If no preferred model was

specified, the model with the most control variables was used.

Two team members extracted bibliographic, geographic information, methods, and substantive data. Substantive data were related to interventions, selected outcomes, population (including gender/age disaggregation, when available), and effect sizes. Discrepancies were reconciled through a discussion between the two team members. Qualitative information on barriers and facilitators to implementation, sustainability and equity implications, and other considerations for practitioners was extracted by a single reviewer.

Included quantitative impact evaluations were appraised by two independent team members using a critical appraisal tool (Appendix 3). Qualitative studies linked to included impact evaluations were critically appraised by a single reviewer using a mixed methods appraisal tool developed by CASP [12] and applied in Snilstveit et al. [11] (Appendix 3).

Synthesis approach

We provide a narrative summary of the papers identified. This includes an overall description of the 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 is summarized in tables. Results from meta-analyses and associated forest plots are presented in the section on the

FREQUENCIES OF BIAS IN RCTS

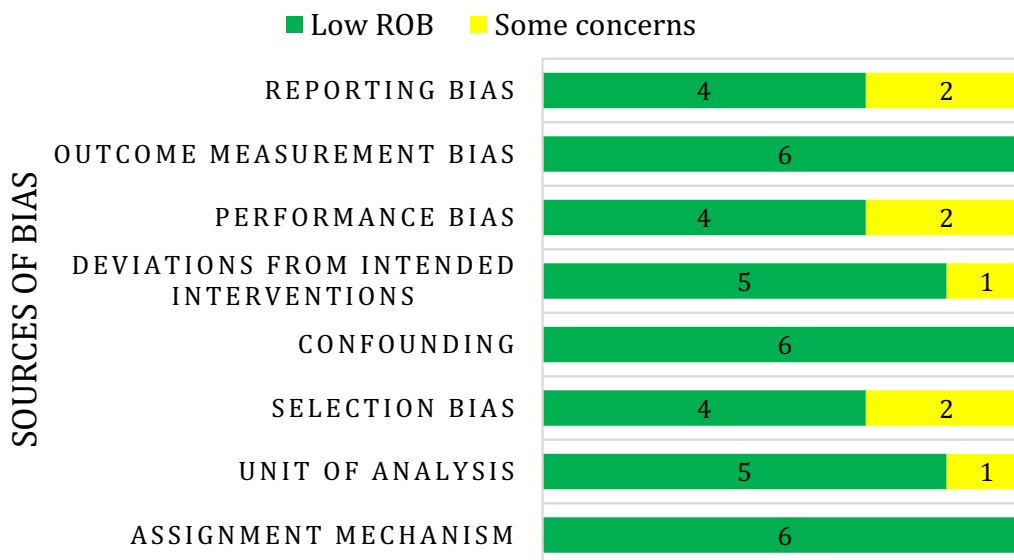


Fig. 2 Risk of Bias of the included randomized control trials

findings. Qualitative information is summarized in a section on implications for implementation and sustainability.

Meta-analysis

In addition to presenting individual effect estimates for all six outcomes, we conducted five meta-analyses to provide summary effect estimates on the five outcomes for which we had sufficient data. This meta-analysis provides additional value relative to presenting the individual effect estimates by presenting a summary effect estimate. Meta-analyzed effects have the benefit of being supported by a broader (Figs. 2 and 3), potentially more generalizable evidence base than individual point estimates. Previous works have statistically synthesized similar evidence, for instance, on food security and food affordability and availability [13, 14], anthropometrics measures [14, 16, 17] micronutrients status [18–20], diet quality and adequacy [21, 22],

Because only ten studies were included, meta-analysis was conducted at the outcome (column 1, Table 2), not the indicator level (column 2, Table 2). However, due to variations in the indicators used and their interpretation, we also present the standardized effect estimates for each study in each forest plot (Figs. 4, 5, 6, 7 and 8) and Appendix 6. The decision to conduct meta-analysis was made on a case-by-case basis after considering if the indicators adequately captured the same underlying concept

[23]. We also summarize the findings of each study, including narratively reporting on individual effects, in Table 3. For all outcomes except micronutrient status, the metrics were determined to be sufficiently similar to warrant a joint analysis in addition to the presentation of individual effects.

To compare the effect sizes, we converted all of them to a single metric, Cohen’s *d*. We then converted all Cohen’s *d* to Hedges *g* to correct for small sample sizes. We chose the appropriate formulae for effect size calculations in reference to, and dependent upon, the data provided in included studies. For example, for studies reporting means (*X*) and pooled standard deviation (*SD*) for treatment (*T*) and control or comparison (*C*) at follow-up only, we used the following formula:

$$d = \frac{X_{Tp+1} - X_{Cp+1}}{SD}$$

If the study did not report the pooled standard deviation, it is possible to calculate it using the following formula:

$$SD_{p+1} = \sqrt{\frac{(n_{Tp+1} - 1)SD_{Tp+1}^2 + (n_{Cp+1} - 1)SD_{Cp+1}^2}{n_{Tp+1} + n_{Cp+1} - 2}}$$

where the intervention was expected to change the standard deviation of the outcome variable, we used the standard deviation of the control group only: For studies

FREQUENCIES OF BIAS IN QED

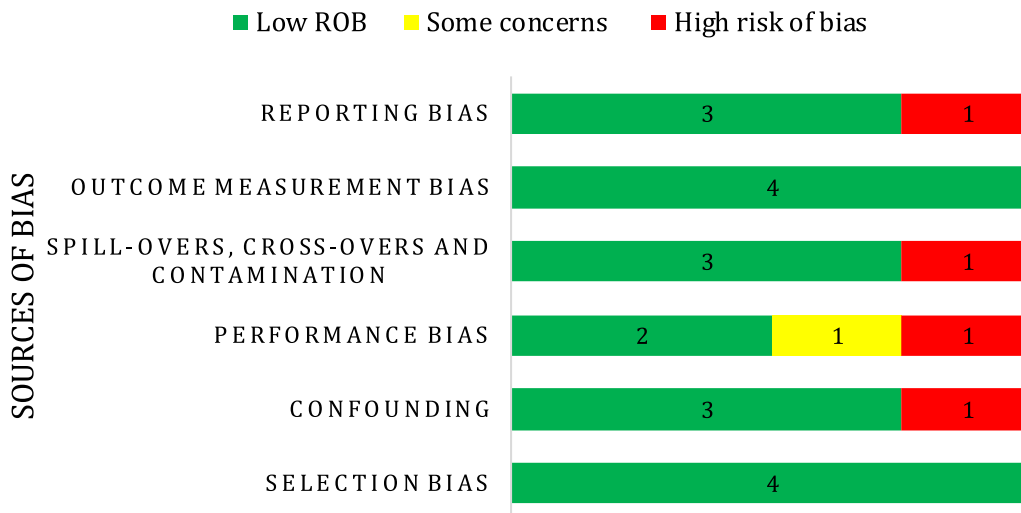


Fig. 3 Risk of bias of the included quasi-experimental included studies

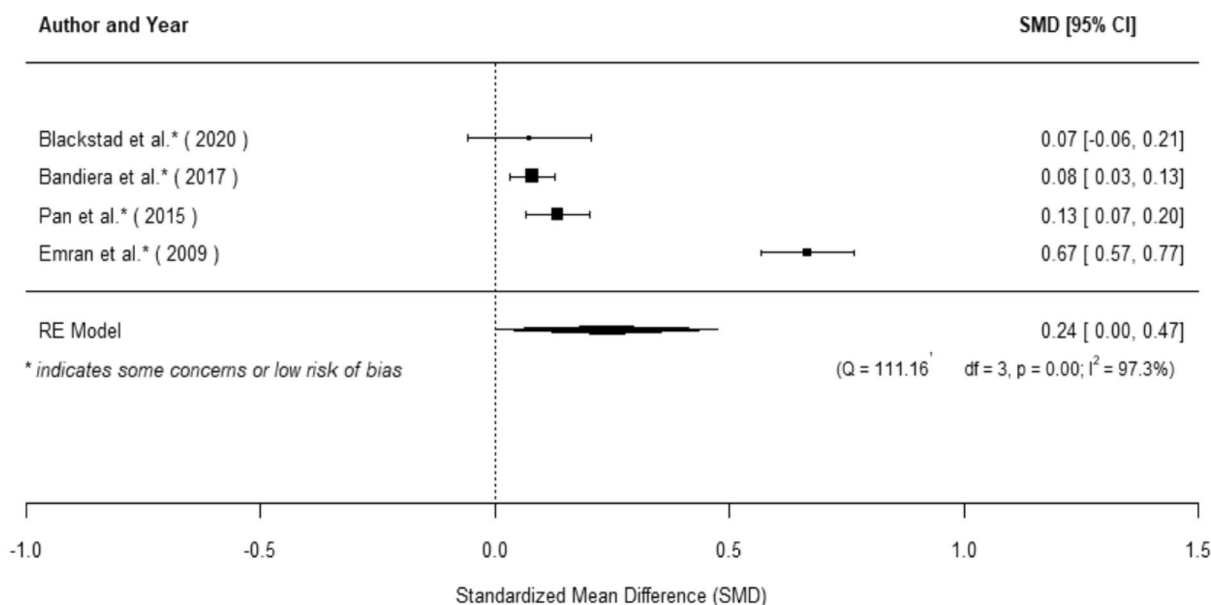


Fig. 4 Forest plot showing the effect of empowerment interventions on food security outcomes

reporting means (\bar{X}) and standard deviations (SD) for treatment and control or comparison groups at baseline (p) and follow-up ($p + 1$):

$d = \frac{\Delta X_{p+1} - \Delta X_p}{SD_{p+1}}$ For studies reporting mean differences (ΔX) between treatment and control and standard deviation (SD) at follow-up ($p + 1$):

$d = \frac{\Delta X_{p+1}}{SD_{p+1}} = \frac{\Delta X_{Tp+1} - \Delta X_{Cp+1}}{SD_{p+1}}$ For studies reporting mean differences between treatment and control, standard error (SE) and sample size (n):

$$d = \frac{\Delta X_{p+1}}{SD \sqrt{n}}$$

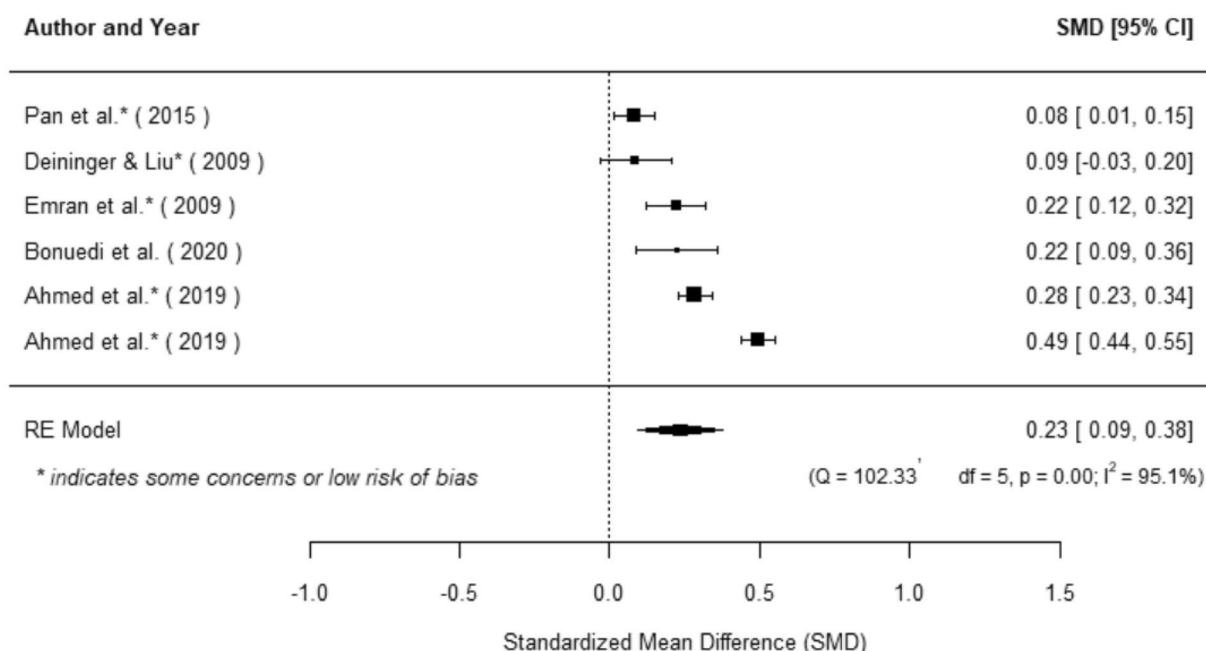


Fig. 5 Forest plot showing the effect of empowerment interventions on food affordability/availability outcomes

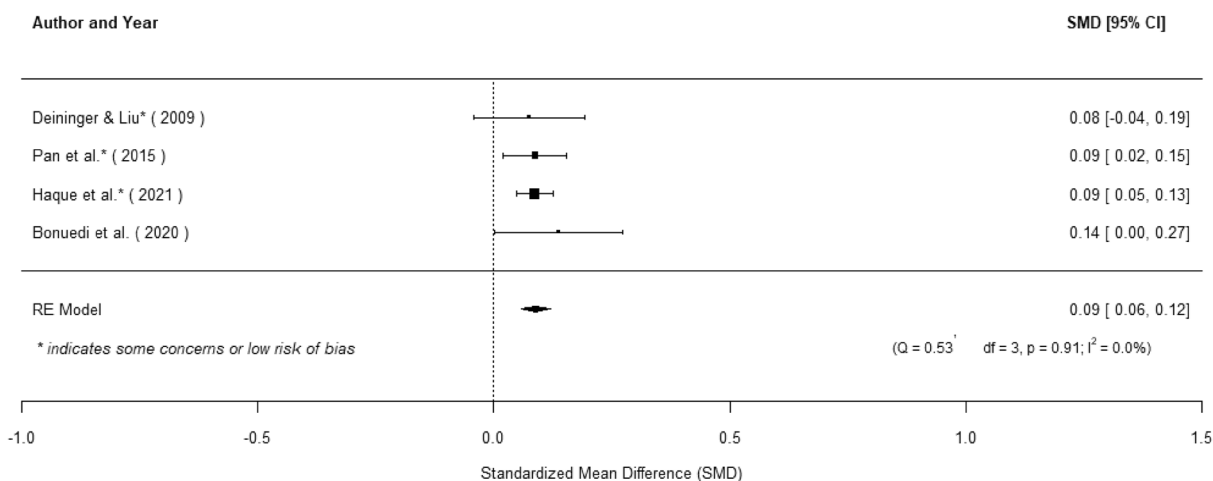


Fig. 6 Forest plot showing the effect of empowerment interventions on diet quality and adequacy

For studies reporting regression results, we followed the approach suggested by Keef and Roberts (2004) using the regression coefficient and the pooled standard deviation of the outcome. Where the pooled standard deviation of the outcome was not unavailable, we used the regression coefficients and standard errors or t-statistics to do the following, where sample size information is available in each group:

$$d = \sqrt{\frac{1}{n_T} + \frac{1}{n_C}}$$

where n denotes the sample size of treatment group and control. We used the following where total sample size information (N) is available only (as suggested in Polanin [34]):

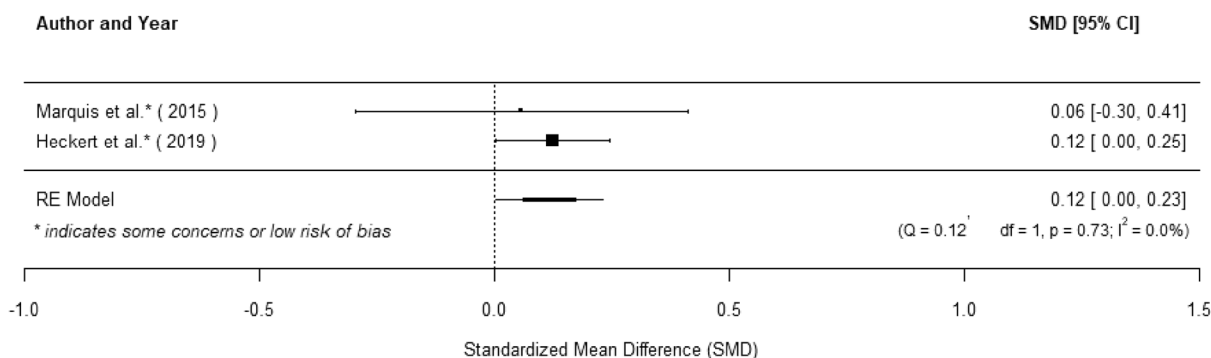


Fig. 7 Forest plot showing the effect of empowerment interventions on weight relative to height

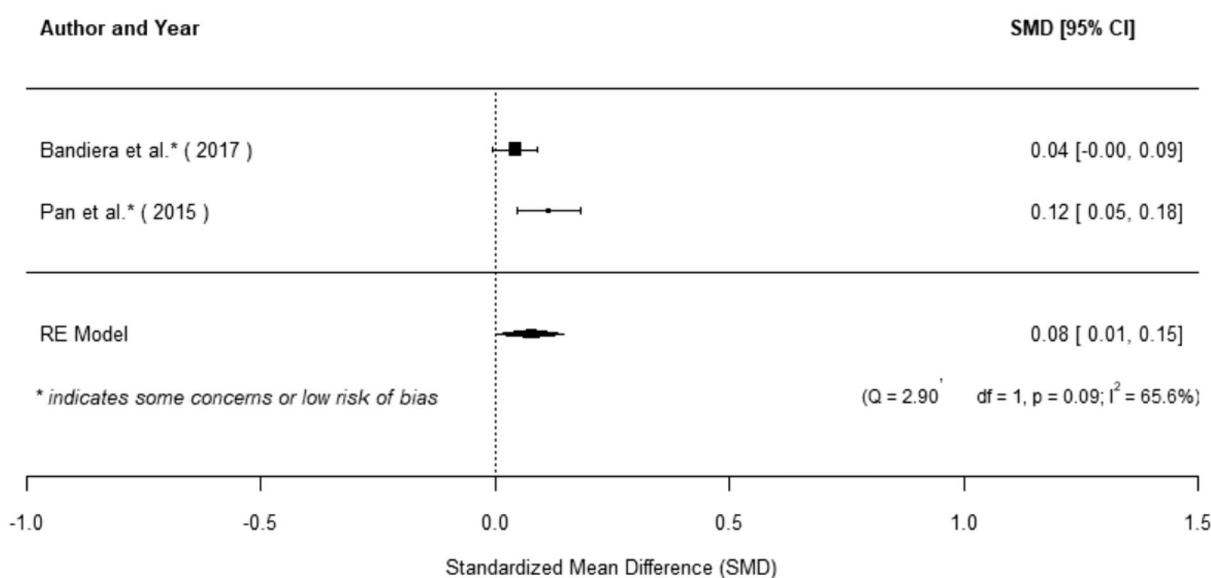


Fig. 8 Forest plot showing the effect of empowerment interventions on wellbeing

$$d = \frac{2t}{\sqrt{N}} \text{Var}_d = \frac{4}{N} + \frac{d^2}{4N}$$

When necessary, we calculated the t statistic (*t*) by dividing the coefficient by the standard error. If the authors only report confidence intervals and no standard error, we calculated the standard error from the confidence intervals using the following:

$$SD = \sqrt{N} \times (\text{upper limit} - \text{lower limit}) / 3.92$$

If the study did not report the standard error, but did report *t*, we extracted and used this as reported by the authors. If an exact p value was reported but no standard error or *t*, we used the following Excel function to determine the t-value.

$$= T.INV.2T (\text{exact } p \text{ value}, (n - 1))$$

where outcomes were reported in proportions of individuals, we calculated the Cox-transformed log odds ratio effect size [35]:

$$d = \text{Log Odds Ratio} \times \frac{\sqrt{3}}{\pi}$$

where OR is the odds ratio calculated from the two-by-two frequency table.

We fitted a random effects meta-analyses model when we identified two or more studies that we assessed to be sufficiently similar. We assessed heterogeneity using the DerSimonian–Laird estimator by calculating the Q statistic, *I*², and *τ*² to provide an estimate of the amount of variability in the distribution of the true effect sizes [23].

Table 3 Summary of included studies

Study reference and country	Study design	Intervention	Authors' interpretation of effects
Ahmed et al. [24] [Bangladesh]	Randomized controlled trial	The Transfer Modality Research Initiative (TMRI) provided: Cash or food transfers, with or without nutrition behavior change communication (BCC) for rural women living in poverty	All four interventions (Cash transfer; food transfer, cash + food, cash + BCC) increased monthly per capita food consumption, daily per capita intake caloric, and food consumption score. The effects are slightly higher for cash + BCC, particularly from the food consumption score
Bandiera et al. [25] [Bangladesh]	Randomized controlled trial	The Targeting the Ultra-Poor program provided: (a) livestock assets and skills transfers for the poorest women. Women were offered a menu of assets to support income generating activities. Assets included livestock and goods for small-scale retail operations, tree nurseries and vegetable growing (b) each asset was offered with a package of complementary training and support	Food security among women improved, but there was no effect on mental health status
Blakstad et al. [26] [Tanzania]	Randomized controlled trial	The Homestead Food Production Program provided: (a) agricultural training for women and inputs to promote home-stead food production (b) nutrition and public health counseling for women to improve diet and health-related behaviors	Household dietary diversity score increased, but there was no effect on food security
Bonuedi et al. [27] [Sierra Leone]	Quasi-experimental design	The Pro-Resilience Action (PROACT) project provided: (a) the LANN was a participatory community-based intervention involving nutrition education, behavioral change communication and awareness creation on the benefits of consuming diverse diets, proper child feeding and water, sanitation, and hygiene (WASH) practices, and sustainable agriculture and natural resource management in rural areas (b) a cash crop, income-oriented intervention aimed at enhancing economic access to nutritious foods. It included a nutrition program directed at improving nutrition knowledge and stimulating nutrition-sensitive spending and allocation of other household resources	LANN, combined with the cash crop intervention, improved dietary diversity and food consumption among women and children. LANN alone did not have any effect
Deninger et al. [28] [India]	Quasi-experimental design	The District Poverty Initiative in India: Supported new self-help groups for women living in poverty in India by training leaders and accountants from new self-help groups in basic management and accounting. The SHGs combined savings generation and micro-lending with social mobilization	The creation of SHGs had mixed effects on food consumption (RS/year), energy intake per capita (kcal/day), and protein intake p.c. (g/day) among poor, non-poor, and poorest of the poor. The three outcomes improved among the poor. Energy intake increased also for the poorest of the poor, but the other two outcomes were not significant for them. None of the outcomes improved among the non-poor
Emran et al. [29] [Bangladesh]	Quasi-experimental design	The Targeting the Ultra-Poor (TUP) program provided: (a) health, education, and training for poor women, including trainings in livestock and poultry rearing; fruit, vegetable, and herb cultivation; operation of tree nurseries; and village vending (b) vitamin A supplements for children under five	The probability of having two meals a day, the probability of having sufficient food to meet the household's needs, and grain stock increased. The highest impacts were reported on the first two outcomes

Table 3 (continued)

Study reference and country	Study design	Intervention	Authors' interpretation of effects
Haque et al. [30] [Bangladesh]	Randomized controlled trial	The Suchana project provided: Training on agriculture, aquaculture, and market development, including challenging the gender barriers to agriculture, health and nutrition practices among the beneficiary women, husbands, and other household members	The Suchana project increased food consumption during pregnancy, the consumption of vitamin A capsules after last delivery, and the consumption of at least 100 IFA tablets during pregnancy. Greater impacts were reported for the first two outcomes
Heckert et al. [31] [Burkina Faso]	Randomized controlled trial	The Enhanced Homestead Food Production (E-HFP) program provided: (a) agricultural assets (b) behavior change communication on agricultural activities, optimal infant and young child feeding, health, hygiene, and care practices	The E-HFP program reduced wasting but had null effects on hemoglobin levels among children
Marquis et al. [32] [Ghana]	Quasi-experimental design	The Enhancing Child Nutrition through Animal Source Food Management (ENAM) program provided: (a) microcredit loans (b) weekly nutrition, technical, and entrepreneurship training on viable income-generation activities	The program had a positive effect on height-for-age z score, a negative effect on weight-for-age z score, and a null effect on BMI-for-age z score of preschool-aged children
Pan et al. [33] [Uganda]	Quasi-experimental design	This large-scale agricultural extension program for smallholder women farmers provided: (a) training through model farmers (b) easier access to and affordability of seeds sold through farmers serving as community agriculture promoters	The program reduced meals skipped, worries about insufficient food, and limited variety of food among smallholder women. It increased per capita food consumption

We were unable to explore heterogeneity using moderator analyses due to the small number of included studies.

Qualitative synthesis

The meta-analysis conducted with the quantitative data has been complemented by a thematic synthesis utilizing the extracted qualitative data. Qualitative data were synthesized thematically by a single team member and reviewed by two other team members. Themes considered related to non-nutrition impacts, barriers and facilitators to impact, and cost evidence.

Results

Characteristics of the included studies

We included ten studies retrieved through the systematic search done for the Food Systems and Nutrition Evidence Gap Map, conducted in January 2022 (Table 3). An additional, low-quality systematic review was identified and excluded from analysis. Four of the ten included studies were implemented in Bangladesh, while the remaining studies were in Burkina Faso, Ghana, India, Sierra Leone, Tanzania, and Uganda. The four studies in Bangladesh represent unique evaluations of a cash transfer program, an agricultural training program, and two fully independent evaluations of Targeting-Ultra-Poor program (TUP) with a time gap of eight years and somewhat different intervention designs. More information on study characteristics can be found in Additional file 1: Table S1.

Randomized controlled trials ($n=4$) and difference-in-difference were the most common designs ($n=4$). Half of the studies using difference-in-difference also used statistical matching ($n=2$). One study used statistical matching alone and one used regression discontinuity to identify counterfactuals. Nine additional qualitative papers associated with seven interventions were also identified and included.

Almost all studies provided training ($n=8$). Some also provided asset transfers ($n=6$) and behavior change communication ($n=3$; Tables 3, 6 in Appendix 6, and Additional file 1: Table S1). Behavior change communication interventions generally communicated messages about women's empowerment and women's roles within their communities. Often, they targeted men, making them gender-transformative. Training and educational interventions focused on agriculture and/or nutrition, but some also considered entrepreneurship and water, sanitation, and hygiene. Asset transfers were largely related to cash or agricultural inputs, including livestock.

Food affordability and availability outcomes were the most common ($n=5$). Diet quality and adequacy and food security outcomes were also common ($n=4$ each).

Anthropometric measures, micronutrient status, and well-being outcomes were less common ($n=2$ each).

We found nine qualitative reports related to seven interventions. Additional qualitative information was not found for the remaining interventions. The qualitative components of the main studies and additional studies were minimal and primarily focused on contextual information from the researchers. Many of the qualitative studies used focus group discussions or key informant interviews to better understand participants' lived realities. Qualitative data contextualized results of empowerment interventions and food and nutrition security based on the differing intervention locations and intersecting social, cultural and gender norms that influence the impacts on nutrition and other key outcomes.

All the randomized controlled trials except Blakstad et al. [26] have an overall rating of 'some concerns', mainly due to reporting bias, performance bias, and selection bias (Fig. 7; Appendix 5). Deiningner and Liu [28] also encountered issues related to deviation from the intended interventions and the unit of analysis did not correspond to the unit of randomization.

Two quasi-experimental studies were rated as having a low risk of bias (Fig. 8; [32, 33]), one study as having 'some concerns' [29], and one as having a high risk of bias [27]. The major sources of bias were related to reporting bias, spill-over, cross-over and contamination, performance bias, and confounding.

What are the effects of women's empowerment interventions on food environment, diet, and well-being outcomes?

Standardized effects are reported in Table 7 in Appendix 6, calculated as outlined in the Methodology section. The meta-analysis results of the random effects model are reported in Table 4. We could not run a meta-analysis on micronutrient status because the two studies looking at it measured different underlying concepts which could not be meaningfully combined.

Effect of women's empowerment interventions on food security outcomes is promising

Our analysis of the effects of women's empowerment interventions suggests they improved food security outcomes overall ($\hat{\mu} = 0.24$ [95% CI: 0.001 to 0.47], $p = 0.048$, Fig. 4). Women receiving these interventions had a 59.5% chance of having food security scores above the mean in the control group. There was significant variation in the size of the effect, ranging from 0.07 in Tanzania, to 0.67 in Bangladesh.

We included four studies which reported the following indicators: food security index (whether the household

Table 4 Meta-analytical results

Outcomes	Specific outcomes indicators	# of included effects (total number of beneficiaries)	Overall effect size [95% CI]	Estimated percentile change compared to control group [95% CI]	Heterogeneity of overall effect (I ²)	Range of effects
Food security	Food security index—whether HH had surplus food or deficit; enough food to eat, and could afford to eat two meals a day;	4 (12,545)	0.24* [0.00; 0.47]	9.5% [0; 18.1%]	111.16***, 97.3%	0.07 to 0.67
	Food availability; sufficient food to meet the household's needs; Household food insecurity access scale; Skip meals					
Food affordability/availability	Food consumption score; Total food consumption expenditure in the 12 months preceding the survey (Food production and market purchases) (LOG)-Household;	6 (12,972)	0.23** [0.09; 0.38]	9.1% [3.6%; 14.8%]	187.27***, 91.99%	− 0.11 to 0.49
	Food consumption (RS/year); Grain stocks (kg); Per capita food consumption					
Diet quality and adequacy	Household dietary diversity; Additional food consumed during pregnancy; Protein intake p.c. (g/day) in the total population; Variety of foods consumed	4 (16,025.5)	0.09** [0.06; 0.12]	3.6% [2.4%; 4.8%]	0.53***, 0%	0.076 to 0.14
	Weight-for-length z score; BMI-for-age z score	2 (1156.6)	0.12* [0.00; 0.23]	4.8% [0; 9.1%]	0.12, 0%	0.06 to 0.12
Well-being outcomes	Mental health index; Worry about insufficient food	2 (10,100)	0.08 [0.01; 0.15]	3.2% [0.4%; 6%]	2.9*, 65.6%	− 0.11 to 0.04

* is $p < 0.05$, ** is $p < 0.01$ and *** is $p < 0.001$. For more information on the results and the studies please see Additional file 1: Table S1

had surplus food or deficit, enough food to eat, and could afford to eat two meals a day), household food insecurity assessment scale (HFIAS), skipped meals, and food available to meet a household's needs of two meals a day [25, 26, 29, 33]. All studies provided training or education, mostly related to agriculture. Three also provided some form of asset transfer [25, 29, 33].

Two studies were assessed as having some concerns related to risk of bias [25, 29] and two were assessed as low risk of bias [26, 33].

Effect of women's empowerment interventions on food affordability and availability outcomes is promising

Our analysis of the effects of women's empowerment interventions suggests they improved the availability and affordability of food ($\hat{\mu} = 0.23$ [95% CI: 0.09 to 0.38] $p < 0.01$, Fig. 5). Women receiving these interventions had a 59.1% chance of having food affordability and availability scores above the mean in the control group. There was significant variation in the size of the effect, ranging from 0.08 in Uganda, to 0.49 in Bangladesh.

Food affordability and availability was measured in five included studies, per capita food consumption, food consumption per capita (Rs/year), total food consumption expenditure (food production and market purchases in the 12 months preceding the survey), and grain stock (kg) [24, 26, 28, 29, 33]. We included two estimates for Ahmed et al. as the results were reported for independent samples from the North and South of Bangladesh, without an overall estimate for all the areas.

All studies but Deininger and Liu [28] included assets transfer, such as cash, cash crops [24, 27], or livestock, seeds, or vitamin A supplements [29, 33]. All studies, except Ahmed et al. [24] included trainings or education on nutrition [27], or agriculture [29, 33], or enterprise/accountability [28]. Two studies also included a behavior change communication component [24, 27].

Ahmed and colleagues also reported increases in monthly food consumption per capita in both northern and southern regions of their intervention area (North areas: $g = 0.32$ [95% CI: 0.27 to 0.38]; South areas: $g = 0.22$ [95% CI: 0.16 to 0.27]) and per capita daily intake caloric (North areas: $g = 0.22$ [95% CI: 0.17 to 0.28]; South areas: $g = 0.09$ [95% CI: 0.043 to 0.15]). Three other intervention arms (provision of food, cash, or food plus cash) were also evaluated. However, we were not able to include them in the meta-analysis as they were not comparable to the other studies. All three reported similar impacts.

Only Bonuedi et al. were assessed as having a high risk of bias, the remaining studies have either some concerns [24, 28, 29] or low risk of bias [33].

Effect of women's empowerment interventions on diet quality and adequacy outcomes is promising

Our analysis of the effects of women's empowerment interventions suggests they improved diet quality and adequacy ($\hat{\mu} = 0.09$ [95% CI: 0.06 to 0.12], $p < 0.01$, Fig. 6). Women receiving these interventions had a 53.6% chance of having diet quality and adequacy scores above the mean in the control group. The variations among the range of effects were not as high as for other outcomes, ranging from 0.08 in India to 0.14 in Sierra Leone.

Four studies reported impacts related to diet quality and adequacy, such as dietary diversity and amount of food or protein consumed [27, 28, 30, 33]. All four studies employed training/education interventions focused on agriculture [27, 30, 33] or enterprise/accountability [28]. Two studies also transferred assets [27, 33], and one included a behavioral change communication component [27].

One study was scored as low risk of bias [33], two were scored as having some concerns [28, 30], and one was rated as high risk of bias [27].

Effect of women's empowerment interventions on anthropometrics is promising but there is a lack of evidence

Our analysis of the effects of women's empowerment interventions suggests they improved measures of weight relative to height ($\hat{\mu} = 0.12$ [95% CI: 0.002 to 0.23], $p = 0.046$ Fig. 7). Children of women receiving these interventions had a 54.8% chance of having anthropometrics scores above the mean in the control group.

Two studies reported impacts on anthropometric measures of children based on WHO z-scores [31, 32]. Both studies transferred agricultural [32] or financial assets [32]. The Heckert and colleagues' study also included a behavioral change communication strategy, while Marquis and colleagues included entrepreneurship training. Marquis et al. [32] also report a decrease in weigh-for-age ($g = -0.42$ [95% CI: -0.77 to -0.06]) and an increase in height-for-age ($g = 0.40$ [95% CI: 0.04 to 0.75]). Heckert and colleagues were scored as having some concerns about bias while Marquis et al. [32] had low risk of bias.

Effect of women's empowerment interventions on micronutrient status is promising but there is a lack of evidence

Two studies considered the effects of women's empowerment interventions on micronutrient status, but these could not be meaningfully combined in a meta-analysis because they measured different underlying concepts.

Haque et al. found that Suchana's gender-transformative approach, which encompassed a portfolio of agriculture and entrepreneurship trainings, increased the consumption of iron, folic acid tablets ($g = 0.25$ [95% CI: 0.21 to 0.28]). Heckert et al. evaluated an agricultural education and behavior change communication strategy, but they found no effect on hemoglobin levels ($g = -0.10$ [95% CI: -0.03 to 0.23]). Both studies were rated as having some concerns about bias.

Effects of women's empowerment interventions on mental well-being outcomes is not significant and there is a lack of evidence

Our analysis of the effects of women's empowerment interventions shows no effect on mental health outcomes ($\hat{\mu} = 0.08$ [95% CI: 0.01 to 0.14], $p = 0.088$, Fig. 8). Bandiera et al. [25] reported a mental health index constructed based on self-reported happiness and mental anxiety, while Pan et al. [33] measured the level of worries regarding insufficient food. Both studies evaluated assets transfer interventions, such as livestock, seeds, vegetables growing, and specific trainings which accompanied to the transfers. Pan et al. [33] paper was assessed as having a low risk of bias, while Bandiera et al. [25] paper was assessed as having some concerns related to performance bias.

Implications

Implications for non-nutrition outcomes

Authors of many of these studies concluded that the interventions accomplished their goals of supporting women's empowerment, often by introducing gender-transformative approaches which challenged traditional social norms. The Enhanced Homestead Food Production (E-HFP) program in Burkina Faso included a gender-transformative approach in which it improved men's perceptions of women as farm managers and increased respect and communication in agri-business activities [31]. The accompanying behavior change communication intervention allowed mothers to better communicate with men to improve familial support and adopt positive nutrition behaviors, such as improved feeding practices. Similarly, the Suchana program in Bangladesh resulted in improvements in women's empowerment and maternal healthcare practices using a gender-transformative approach [30]. Women became more confident to discuss issues around food and management of household resources with their partners [27]. Self-help group participation improved social awareness and leadership skills. Women mobilized to protest child marriage and violence against women in their communities [37]. The Targeting-Ultra-Poor program (TUP) in Bangladesh increased saving and borrowing opportunities for women. These

interventions allowed women to accumulate savings and spend more judiciously, rather than consistently responding to immediate needs.

Two interventions which combined training with improved accessibility of agricultural assets increased opportunities for paid work. The agricultural intervention in Uganda resulted in an increase in work for wages and freed up off-farm work times for the entire household, including women [33]. Similarly, because of the TUP program, the labor market choices of household members aside from the targeted woman also shifted [25]. However, women themselves did not have increased labor participation. Women in the program spent most of their time at home and were generally not employed outside of the home [38]. In fact, women reported that they preferred to stay at home due to low pay and social stigma in workplaces.

Similarly, two interventions focusing on household farming for improved nutritional outcomes were labor and time intensive, which resulted in high attrition [26]. This additional labor was an increased burden on women and took away from their time to acquire and prepare food for their families [27]. When data collection coincided with harvest months in Sierra Leone, women's involvement in the farming activities increased their time constraints and adversely affected caregiving practices.

Barriers and facilitators

Restrictive social norms preventing women from being able to take advantage of the interventions as intended was a common barrier. Structural gender barriers act as a driver of inequality in the household and community, as specified in Njuki et al. theory of change (Fig. 1). In highly patriarchal societies, such as Sierra Leone, deeply entrenched social and cultural norms marginalize women, restrict their decision-making and exclude them from accessing or controlling household resources [27]. Single-focus interventions that only targeted nutrition or value chain inputs without behavior change communication related to social norms were not able to fully realize potential impacts because entrenched norms were significant barriers to long-lasting change [33]. Even if women were given the tools to work outside the home or own assets, they were often blocked from leveraging these tools by norms that dictate how women can act and work [33]. Gender-transformative approaches address this social barrier by including men to ensure that the full impacts of interventions can be leveraged and realized as intended.

In the TUP program, asset transfers that were intended for women members of households were controlled by men due to social norms [39]. Social norms delineated what type of assets women were allowed to own. Larger

livestock, like cattle, were automatically perceived to belong to men because they were higher in value and traded more often. Their sale required an adult male's consent, which restricted women's ability to own and manage them. Restrictions almost always came from jealous or violent husbands. When the TUP transferred small livestock such as poultry, that women more often owned, it was easily controlled by women [39]. Religious norms also played a role in restricting women's public movements. Care responsibilities were reinforced by conservative social norms for women in Bangladesh, where women were demarcated as primary caregivers in the home [37].

In some contexts, community and men's support also facilitated improvements in outcomes, demonstrating the importance of gender-transformative approaches that actively challenge gender norms and power inequities between genders. In the Homestead Food Production intervention in Tanzania, women who lived near neighbors who also grew crops at home had higher dietary diversity [26]. Participants who were close to markets were able to access, trade and procure food and related items easier than those who were farther away [25]. If husbands and other men in the household or community were more receptive to change, then progress was more visible with women in the TUP [37]. If a husband was more open to his wife engaging in out-of-house activities, livelihood strategies were more successful.

Multi-component interventions may leverage synergistic effects to have greater impacts than the individual components would have [27]. Complementary program arms can reinforce each other in achieving desired results and reduce implementation costs to achieve the same objectives [27]. The asset-based component of the PROACT program in Sierra Leone had little effect. However, when combined with a behavior change communication component, it increased women's decision-making power, shifting women's roles in the household, and expanding women's ability to work outside the house. Behavior change communication components of the TMRI program in Bangladesh combined with the incentive of asset transfers allowed women's sustained participation and achieved an overall improvement in household indicators over the course of the program [38].

Interventions which do not address equity can be less successful and re-enforced social norms. Often, entrenched norms and roles were not acknowledged within included interventions [40]. Failure to address these norms may have resulted in some interventions being unsuccessful. This was seen in the Bangladesh asset transfer program which did not address norms around livestock ownership and resulted in men gaining control

over some of the transferred assets [39]. Interventions which took place at the home and approached women as caregivers and providers may have further perpetuated the stereotype of women within these roles [37].

Unfortunately, the long time needed to change social norms was a barrier to these interventions achieving impact in the short period in which they were evaluated. The theory of change from women's empowerment interventions to improved nutrition outcomes assumes a change in social norms, which requires a significant amount of time (Fig. 1). Change within the food system is a dynamic process which often depends on other changes outside the scope of these interventions. Moreover, change processes are not straightforward and can be accompanied by setbacks, sometimes occurring parallel to positive effects. Behavior change communication can be slow to expand women's empowerment and households' social status and networks [24]. Impacts often become apparent in the long-term when foundational improvements consolidate and are dependent on internal and external factors. Food and nutrition security and women's empowerment may need to be achieved in stages, according to different resources and opportunities [33]. For example, in India, the District Poverty Initiative fostered group formation and supported more mature groups, which could have significant economic benefits in the long term [28]. Because the study utilized data from three and six years after group formation, the research implies there may have been impacts on capital endowments and economic effects on individuals and the group itself. Authors of evaluations that occurred within 12 months of the interventions' end indicated that a more comprehensive understanding of women's empowerment and nutritional outcomes would require longer-term and more frequent data collection [26, 31].

Specific characteristics of the target group can affect impacts and may explain heterogeneity in results. Household decisions regarding assets and nutrition were shaped by local ecological and economic conditions [24]. In India, target groups that were the poorest saw the largest asset accumulation and empowerment improvements. This resulted in the poorest benefitting both socially and economically [28]. Interventions which leverage existing groups may experience high attrition if the groups themselves experience attrition. For example, the Enhancing Child Nutrition through Animal Source Food Management program targeted microcredit groups, and experienced significant attrition among those who were not benefitting from the loan program [32]. This may not have been observed if the intervention targeted women directly and did not work through the microcredit group.

Cost information

Cost reporting was low ($n=3$). When studies reported cost data, either through cost per participant or cost benefit analysis, the benefits generally outweighed the costs. The District Poverty Initiative in India found that net present value of benefits from the project were approximately \$1,690 million, significantly more than the project cost of \$110 million. Even if benefits only lasted for one year the estimated benefits still significantly exceed project costs, with a benefit–cost ratio of 1.5 to 1 [28]. The TUP program in Bangladesh also showed that average benefits, including increased household welfare, were 3.21 times larger than costs. Big push programs, like the TUP, required large investment. However, in this case, it resulted in cost-effective and sustainable change in household welfare, including nutrition [37].

Multi-component interventions can be cost-effective because they combine complementary initiatives, such as interventions targeting nutrition and social norms. This was seen in PROACT where impacts were only achieved once a behavior change component was added to the asset transfer [27]. Similarly, when added to an asset transfer program, the TMRI women's empowerment behavior change communication component costs \$50 per beneficiary per year, which is a relatively low cost compared to stand-alone behavior change communication interventions [24]. Low-cost additional activities can have greater impact than expected, especially when integrated with other components. The training of model farmers in Uganda improved cultivation methods at relatively low cost when compared with the cost of inputs, such as a high-yield and drought-resistant seeds. Both training and the provision of inputs improved women's efficiency in household gardens [33]. However, when calculating costs, the additional cost of such labor should not be ignored, especially because these costs are often born by the women that these interventions are trying to help [26].

Discussion

Overall, our analyses suggest women's empowerment interventions can improve measures of the food environment and diet. We find significant and positive effects on food security (0.24 [95%CI: 0.00 to 0.47], $n=4$), food affordability and availability ($\hat{\mu} = 0.023$ [95% CI: 0.06 to 0.38], $n=6$), and diet quality and adequacy ($\hat{\mu} = 0.09$ [95% CI: 0.06 to 0.12], $n=4$). With two studies considering outcomes related to weight-for-length ($\hat{\mu} = 0.12$ [95% CI: 0.00 to 0.23]) and wellbeing ($\hat{\mu} = 0.08$ [95% CI: 0.01 to 0.15]) each, the evidence is too limited to draw conclusions. Although impacts on diet quality and adequacy, anthropometrics, and well-being were positive, they were smaller than impacts on more proximate outcomes, such

as food security and food affordability and availability. Impacts seem to reduce along the causal chain. Some of the more final outcomes, such as anthropometric and well-being measures, can take years to meaningfully change. As such, modest early effects may imply longer-term change.

Insights from the qualitative evidence suggest that women's empowerment interventions best influenced food environment and diet outcomes when gender and social norms were considered. However, often, entrenched norms and roles were not acknowledged in these interventions [40]. When community, and especially male support, was found, it may have facilitated impact. Including gender-transformative approaches in women's empowerment interventions may be essential to challenge and overcome existing social norms which often prevent the achievement of intended impacts. Such transformative approaches may be necessary to allow women to fully benefit from ongoing interventions. Restrictive social norms may prevent women from taking full advantage of the interventions and reduce potential impacts.

Although women's empowerment interventions are promising approaches for improving measures of the food environment and diet, interventions may need to move beyond women's empowerment interventions include GTA and gain the buy-in of men and the community. This can result in increased power of women in household decision-making while also sensitizing men to women's pursuits of work outside of the home [41]. GTA require cultural and social adaptation to local contexts through strengthened local partnerships and capacities while considering intersectionality, e.g., by considering different interconnections between gender, socioeconomic class, and caste divisions. GTA and intersectionality, both characteristics of feminist development policy, are crucial to progress on gender equality and leverage the full potential of policies and interventions. Similarly, interventions should attempt to improve women's social capital so they can better control and decide how to acquire and prepare food for their families [39]. Focusing on the duration of interventions is also important. Long-term interventions may be needed to account for slow processes, such as changing social norms. Multi-component interventions, which combine a livelihoods component (asset transfer or financial services) with behavioral change communication and advocacy, may be more effective than interventions focusing on just livelihoods or behavioral change.

With ten included studies, the evidence base is small, which can reduce generalizability. Variation in the measures considered in the meta-analysis may drive heterogeneity in results. However, the overall quality of the

evidence is fair with most of the studies ($n=6$) rated as having ‘some concerns’ regarding bias. Three studies were assessed as having ‘low risk of bias.’ Given the low number of studies available and potential biases, the results should be interpreted with some caution.

Although the evidence was generally of high quality, we had some concerns related to reporting, performance, and selection bias of the randomized controlled trials. Within the quasi-experimental studies, we found issues related to reporting bias, spill-over, cross-over and contamination, performance bias, and confounding. Some authors reported issues with incomplete or low-quality data, for instance, incomplete children’s health or vaccination records. Moreover, some children aged out during the evaluation period making the data inconsistent. Other studies did not collect data across seasons, an essential element when collecting data on agriculture outcomes, which can act differently across seasons. Short interventions and short data collection periods might also prevent impacts from being identified. These limitations could result in findings being somewhat unreliable.

Strengths, limitations & future directions

The interventions considered in this analysis were multifaceted, often considering two or three components: behavior change communication, training, and asset transfers. As such, it is not possible to determine which of these approaches is most effective. Future work can isolate the effects of these different pathways, as done by Bonuedi et al. [27], to determine which of these components is most effective.

The meta-analyses presented here combine disparate indicators of broad concepts. The combined analysis of these different indicators is justified because they measure the same underlying concept. However, the variation in indicator used by each study may explain the heterogeneity in results. For example, the analysis on food security combines a food security index, household food insecurity assessment scale, number skipped meals, and indicator of whether food is available to meet a household’s needs of two meals a day. The framing of food attributes as positive versus negative can affect attitudes toward food [42], so framing questions around food security and insecurity may produce different results. As such, individual effect estimates should also be considered and are reported within each forest plot and in Appendix 6. Summaries of the effects identified by each study are provided in Table 3. Future work should move toward standardizing measurement to allow for better comparability. Some of such efforts already exist, but should be further supported to allow for stronger synthesis [43, 44].

Given the limited evidence base, more research is needed in this field broadly. All the studies were

implemented in Sub-Saharan Africa or South Asia, leaving evidence gaps in Central America, South America, and Central Asia. Most studies were implemented in contexts that were particularly patriarchal and restrictive for women, meaning that results in more egalitarian societies may be different. Although we were able to run a five meta-analysis, interpretation of the results is limited due to the low number of studies and variation in the indicators synthesized. Cost data will also be needed to determine if these impacts are cost-effective. To determine the sustainability of impacts over time, future studies should have longer intervention periods to ensure accurate capture of perceived impacts. Qualitative data can add rich depth to quantitative findings by adding context, experiences and meaning to the lived experiences of project participants. Mixed-methods studies should focus on identifying impacts and then using qualitative research to interrogate how these impacts were achieved. Studies in places with caste divisions, such as India or Bangladesh, could have benefited from a disaggregation in the experiences and outcomes of women and households from different castes. Future studies should try to avoid outcome measurement bias, reporting bias, spill-over, cross-over and contamination, performance bias, confounding, and selection bias. Future studies should also ensure that data collection is representative of different seasons and contextual changes, to avoid incomplete or insufficient data [26, 30, 32].

Due to the rapid nature of this work, results should be interpreted with caution. The studies included in this review are those found through the systematic search for the EGM produced by Moore et al. [1] as of January 2022. It is possible that a more sensitive and targeted search strategy would identify additional studies. Moreover, the REA is limited in the scope of interventions included. Only those which take place within the food system are considered; interventions functioning outside of the food system may influence nutrition outcomes but have not been considered.

Appendices

Appendix 1: Rapid Evidence Assessment on Women’s Empowerment in Food Systems Interventions – Protocol

Background

The problem, condition, or issue

Women are key actors within food systems, serving as producers, wage workers, traders, processors, and consumers. Women also face differential outcomes related to accessing and affording nutritious foods or a healthy diet. Some evidence shows that women—often living in more

vulnerable conditions than men due to societal norms—can improve their own and their children’s nutritional status when they have socio-economic power to make decisions on food and non-food expenditures (especially accessing resources) and can take care of themselves and their families [3]. As a result, women’s empowerment interventions represent a key opportunity to improve nutrition-related outcomes. There is substantial agreement about pathways to improve women’s empowerment in food systems. However, cross-contextual evidence on the factors that cause poorer nutrition outcomes for women, and how women’s empowerment can improve nutritional outcomes is still scant [2].

The interventions

We will include interventions that integrate activities to empower women and/or girls to access, participate and take control in components of the food system, for example improving decision-making on household expenditures. We have extracted relevant papers from the Food Systems and Nutrition evidence gap map that have any intervention component relating to women’s empowerment.

Expected theories of change

Our theory of change is based on the pathways developed by Njuki et al. [2] to presume that women’s empowerment can lead to improved nutrition with a variety of other influencing factors. Gendered food systems interact with gender equality and inequality in a four-dimensional space: individual, systemic, formal, and informal.

Rationale for the review

This rapid evidence assessment is expected to inform decisions regarding gender and women’s empowerment in nutrition and food systems interventions. Given that women’s empowerment has been highlighted as a critical, crosscutting theme for the transformation of the food system [4], key decision-makers have indicated interest in this area. Researchers can use this work to better understand how to intertwine gender-sensitive or -transformative interventions for improved nutritional outcomes.

Research questions

1. What are the effects of women’s empowerment interventions within the food system on the availability, accessibility, and affordability of healthy diets or nutritional status?
2. Are there any unintended consequences of such interventions?
3. Do effects vary by context, approach to empowerment, or other moderators?

Methodology

To respond to these research questions, we will conduct a rapid evidence assessment, based on a systematic literature search of key academic databases. Literature will be screened for quality and summarized visually and in a narrative format. A rapid evidence assessment is based upon the rigorous methodology adopted in a systematic review; however, many steps are shortened [10].

Criteria for including and excluding studies in the review (PICOS)

Criteria	Included	Excluded
Participants	People of any age and gender residing in low- and middle-income countries (L&MICs)	High-income countries
Intervention(s)	Interventions aimed at increasing women’s empowerment and giving women the capabilities to make decisions on the purchase and consumption of a healthy diet	All else
Comparison	Business as usual, including pipeline and waitlist controls An alternate intervention	No comparator
Outcome(s)	Food affordability, accessibility, and availability Iron, zinc, vitamin A, and iodine status Anthropometric measures Diet quality and adequacy Measures of well-being	All else
Study designs	Experimental, quasi-experimental, systematic reviews and cost evidence	Efficacy trials, before-after with no control group, cross-sectional studies and so on

Types of study participants

Only studies which consider populations in low- and middle-income countries (as defined using the World Bank Country and Lending Groups classification in first year of intervention or if not available then Publication year) will be considered. The exception to this is if a country held high-income status for only one year before reverting to L&MIC status. These will be included even if the intervention began in the high-income year. As of the writing of this protocol, this applies to Argentina (2014, 2017), Venezuela (2014), Mauritius (2019), and Romania (2019). If the study is conducted in a high-income

country but measures impact on people, firms, or institutions in an L&MIC, it can be included. For example, we would not exclude a study that measures impact of New Zealand's immigration visa lottery on residents of Tonga.

Types of interventions

Eligible interventions were identified during the development of the Food Systems and Nutrition Evidence Gap Map [1]. The map defined women's empowerment interventions as "efforts targeted at increasing women's abilities to make decisions regarding the purchase and consumption of healthy foods." After completing the search, we found that these interventions were primarily related to agriculture skills training, asset transfers, microcredit, and behavior change.

Citation	Intervention
Ahmed et al. [24]	The intervention consists of two treatment arms: cash or food transfers, with or without nutrition behavior change communication (BCC), to women living in poverty in rural Bangladesh
Bandiera et al. [25]	The intervention is a nationwide asset transfer "plus" program in Bangladesh. The intervention transfers livestock assets and skills to the poorest women
Bonuedi et al. [27]	The intervention is two-pronged: (1) cash crop and (2) nutrition components. (1) Included farmer field schools (FFS), productive inputs, and value chain linkages. (2) Included gender-sensitive nutrition behavior change and awareness creation
Choudhury et al. [45]	<i>Suchana</i> improves nutrition service delivery, nutrition governance, and the knowledge of women and girls regarding gender norms and gender-based violence that can impact mother and child nutrition
Deininger et al. [28]	The intervention is self-help groups for women living in poverty in India
Emran et al. [29]	This is an asset transfer "plus" intervention, bundling asset transfers with capacity building (health, education, and training) for poor women with the goal of helping them graduate to the standard micro-credit program of BRAC
Heckert et al. [31]	The intervention is the Enhanced Homestead Food Production (E-HFP) program, a nutrition- and gender-sensitive agriculture training program
Marquis et al. [32]	This is a microcredit "plus" intervention that provides microcredit loans and weekly sessions of nutrition and entrepreneurship education for 179 women with children 2–5 years of age
Mosha et al. [26]	The agricultural training and provision of inputs intervention includes the provision of small agricultural inputs to women, garden training support, and nutrition and health counselling to improve food security
Pan et al. [33]	A large-scale agricultural extension program for smallholder women farmers to improve food security in Uganda

Types of outcome measures

The table below outlines outcome indicators that will be extracted. These outcomes can be measured using a variety of indicators. We have indicated the preferred outcomes and alternate outcomes which could be used if preferred outcomes are not reported. Composite measures will always be preferred over disaggregated ones.

Outcome	Indicators
Food security	Preferred outcomes: food security indexes and composite scores Secondary outcome: skipped meals Tertiary outcome: reports of insufficient food
Food affordability	Preferred outcome: per capita food consumption in monetary units Secondary outcome: per capita food consumption in weight Other measures, such as cost of a food basket, will be considered if these are not available
Food availability/accessibility	Preferred outcomes: food assets, production (community gardens,) and stores Other measures, such as distance and accessibility to markets
Diet quality and adequacy	Preferred outcomes: composite diet scores such as the nutrient rich food index Secondary outcome: dietary diversity and other food variety measures Tertiary outcome: intake of specific foods
Anthropometrics	Preferred outcomes: body mass index, weight for length, length for age, weight for age Other measures, such as MUAC and ponderal index, will be considered if these are not available
Iron, zinc, vitamin A, and iodine status	Preferred outcome: measures of content in blood/tissue (ex. hemoglobin levels) Secondary outcome: intake in weight (grams, micrograms, etc.) Tertiary outcome: intake in percent-age relative to recommended intake Other measures will be considered
Well-being	Preferred outcome: perceived well-being Secondary outcome: anxiety

Types of comparators

- Business as usual, including pipeline and waitlist controls
- An alternate intervention
- Studies with no comparator are excluded

Types of study design

Experimental, quasi-experimental, systematic review, and cost evidence will be considered. The following study designs will be included.

- Randomized controlled trial
- Regression discontinuity design
- Controlled before-and-after studies, including
 - Propensity-weighted multiple regression
 - Instrumental variable
 - Fixed effects models
 - Difference-in-differences (and any mathematical equivalents)
 - Matching techniques
- Interrupted time series
- Systematic reviews that include a quantitative or narrative synthesis

Ex-post cost-effectiveness analyses will be included, provided that they are associated with an included impact evaluation.

Date, language, and form of publication

All proceeding restrictions are from the EGM.

- Date: 2000
- Language: English

Search strategy

We will not perform any new searches for this REA. Instead, we will look at the ten studies of women's empowerment interventions identified in the Food Systems and Nutrition 'living' EGM,¹ updated every four months (last update December 2021). We specifically searched for interventions using women's empowerment within the food system implemented in low- and middle-income countries. This EGM was developed through a systematic search and screening process equal to that of a systematic review. However, because interventions had to function within the food system to be included, many women's empowerment interventions, such as those related to self-help groups broadly, were not included. Ultimately, the EGM includes ten evaluations of women's empowerment interventions which considered outcomes related to food availability, accessibility, and affordability and nutritional status. We will conduct additional

targeted searches to identify qualitative studies and process evaluations of the included interventions.

Selection of studies

Screening Because we are utilizing the results of the Food systems EGM, there is no search and screening process to select the studies. Rather, within the FSN EGM, we selected ten studies that have women's empowerment interventions associated with the relevant outcomes.

Data extraction and coding procedures Data extraction templates will be modified from 3ie's repository coding protocol and the coding protocols typically used for systematic reviews (Appendix 2). This includes bibliographic, geographic information and substantive data, as well as standardized methods information. In addition, two members of the team will extract data independently on interventions, outcomes, population (including gender/age disaggregation, when available), and effect sizes corresponding to the outcomes indicated above, and any discrepancies will be reconciled. On interventions, outcomes, population (including gender/age disaggregation, when available), and effect sizes corresponding to the outcomes indicated above, and any discrepancies will be reconciled. Qualitative information on barriers and facilitators to implementation, sustainability and equity implications, and other considerations for practitioners will also be extracted.

Critical appraisal All the included quantitative impact evaluations will be appraised by two independent members of the team using a critical appraisal tool (Appendix 1.1 and 1.2). Qualitative studies linked to included impact evaluations will also be critically appraised.

Qualitative search and appraisal In addition to qualitative evidence from the included studies to assess factors that determine or hinder the effectiveness of interventions using a combination of qualitative synthesis, we will conduct a basic search on the programs in each of the ten papers, looking for the following relevant papers [11]:

- *A qualitative study* collecting primary data using mixed- methods 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

¹ <https://gapmaps.3ieimpact.org/evidence-maps/food-systems-and-nutrition-evidence-gap-map>.

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 and what is felt to be working 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.

While the identification of qualitative evidence is limited to studies linked to the included impact evaluations, the process of data extraction, critical appraisal, and evidence synthesis is independent.

We will assess the quality of included qualitative studies, process evaluations, and descriptive quantitative studies using a mixed methods appraisal tool developed by CASP [12] and applied in Snilstveit et al. [46]. This tool is in Appendix 1.3. The meta-analysis conducted with the quantitative data will thus be complemented by a thematic synthesis utilizing the extracted qualitative data.

Analytical approach for quantitative data If sufficient data is available, we will conduct meta-analysis to provide summary effect estimates. We will choose the appropriate formulae for effect size calculations in reference to, and dependent upon, the data provided in included studies. We will conduct random effects meta-analyses when we identify two or more studies that we assess to be sufficiently similar. We will 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 the true effect sizes [23]. We will explore heterogeneity through the use of moderator analyses if the data allow. We will also test for the presence of publication bias if at least 10 studies are included in the analysis.

Data presentation

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. Results from meta-analyses and their associated forest plots will be presented when the data is sufficient. Qualitative information will be summarized narratively in a practitioner’s brief to support project design and implementation. An updated theory of change will be developed based on the combination of qualitative and quantitative data.

Limitations

Due to the rapid nature of this work, results should be interpreted more cautiously than those of a systematic review. Relying on the existing Food Systems and Nutrition EGM may result in some relevant studies being omitted from this evidence assessment. The small number of studies which are expected to be retrieved through this REA may restrict the possibility of using meta-analysis and our ability to draw generalizable conclusions.

Appendix 2: Data extraction tool

Variable group	Variable Label
Publication info	Record type
	Record Title
	Record authors
	Publication year
	URL link
Intervention and implementation considerations	Intervention
	Intervention details
	Unintended consequences
	Barriers and facilitators to implementation
Evaluation considerations	Study design
	Covariates
	Outcomes
Sustainability and financial considerations	Sustainability comments
	Cost effectiveness comments
Other	Other
	Confidence rating (srr only)

Quantitative data extraction tool

Variable level	Explanation
Study ID (DEP)	This is the study ID from DEP (e.g., 17347)
Study ID (EPPI)	This is the study ID from EPPI reviewer. It should match the study ID from the Outcome Mapping Sheet (e.g., 41504196)
Estimate ID	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., SC-SR1_1, SC-SR1_2 and so on)
Evaluation design	0 = Experimental Design (e.g., RCT), 1 = Quasi-Experimental Design
How counterfactual is chosen	Free text (e.g., random control trial, propensity score matching, etc.)—Multiple codes are ok
Analysis type for this effect size	Free text, what type of analysis was used (Regression, 2SLS, ANCOVA, etc.)- Multiple codes are ok
Estimate type	Type of data for this effect size: 1 = Continuous—means and SDs, 2 = Continuous—mean difference and SD, 3 = Dichotomous outcome—proportions, 4 = Regression data
Comparison	1 = No intervention (service delivery as usual), 2 = Other intervention, 3 = Pipeline (waitlist) control (still service delivery as usual)
Describe comparison group	Free text, describe the comparison group
Country	Select the countries in which the study was conducted (drop down menu). There is a multi-country option for situations when there are more than 15 countries, and no disaggregated effects provided for each country
Subgroup	Is this analysis of a subgroup? 0 = no, 1 = yes
If yes to subgroup, describe	Free text, describe the subgroup if applicable (e.g., boys, girls). If no subgroup, type N/A
Source	Note the page number, table number, column, and row you used to extract the data
Treatment effect	1 = Intention to Treat (ITT), 2 = Average Treatment Effect on the Treated (ATET), 3 = Average Treatment Effect (ATE) 4 = Local Average Treatment Effect (LATE)

Variable level	Explanation
Intervention codes	
Intervention description	Use this open answer field to enter, in the author's own words, a description of the intervention, up to a paragraph or so; more detail information will be preferred. Be selective and concise with the excerpts being transcribed here as to ensure accurate and precise descriptions of the intervention. Include page numbers with every excerpt extracted. Do this for each Treatment arm
Intervention	Record the intervention for the corresponding effect size
Exposure to intervention (in months)	How long is the intervention exposure itself?
Evaluation period (in months)	The total number of months elapsed between the end of an intervention and the point at which an outcome measure is taken post intervention, or as a follow-up measurement. If less than one month, use decimals (e.g., measurement immediately after the intervention end would be coded as 0, one week would be 0.25, etc.)
Post-intervention or change from baseline?	0 = Post-intervention, 1 = Change from baseline
Outcome Codes	
Outcome description	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 page numbers with every excerpt extracted. Do this for each outcome
Outcome	Record the outcome for the corresponding effect size
Effect Size Data Extraction	
Reverse Sign (i.e., decrease is good)	Record no if an increase is good, record yes if a decrease is good and the sign needs to be reversed
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
Mean_t	Outcome mean for the treatment group
Sd_t	Outcome standard deviation for treatment group
Mean_c	Outcome mean for the comparison group

Variable level	Explanation
Sd_c	Outcome standard deviation for control group
Mean_overall_diff	Overall mean difference (treatment—control)
Diff_se	Standard error of the overall mean difference
Diff_t	t statistic of mean difference
Odds ratio	Odds ratio reported in the study
OR_se	Odds ratio standard error reported in the study
Risk ratio	Risk ratio reported in study
RR_se	Risk ratio standard error
Reg_coef	Report the regression coefficient of the treatment effect
Reg_SE	Report the associated standard error of the regression coefficient
Reg_t	Report the associated t statistic of the effect size (coefficient/SE)
Exact p value	Exact p value if given, if not, record as written in the manuscript (e.g., $p < 0.001$, or $p > 0.05$)
Clust_t	Number of clusters—treatment group
Clust_c	Number of clusters—control group
Clust_T	Number of clusters—total sample
n_t	Sample size—treatment group
n_c	Sample size—control group
n_T	Sample size—total sample
Periods (1 if cross-sectional)	Record how many periods of evaluation there are (e.g., cross section is 1, panel data with 3 measurements is 3)
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 "Yes" if the sample size needs to be divided by the number of periods, and "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
Treatment variable	Record the treatment variable as written in the model (e.g., the variable name the author uses, such as ("Intervention x Time"))
Dataset	Record if data comes from an identified dataset
Coder	Record your name
Notes	Record any notes important for the team
n_T_revised	THIS IS FOR SENIOR QUANT LEAD TO FILL OUT
sp	THIS IS FOR SENIOR QUANT LEAD TO FILL OUT
d	THIS IS FOR SENIOR QUANT LEAD TO FILL OUT

Variable level	Explanation
g	THIS IS FOR SENIOR QUANT LEAD TO FILL OUT
Var(d)	THIS IS FOR SENIOR QUANT LEAD TO FILL OUT
se(d)	THIS IS FOR SENIOR QUANT LEAD TO FILL OUT
Cl_l	THIS IS FOR SENIOR QUANT LEAD TO FILL OUT
Cl_u	THIS IS FOR SENIOR QUANT LEAD TO FILL OUT
Remove	THIS IS FOR PROJECT MANAGER TO FILL OUT
Formula Used	THIS IS FOR SENIOR QUANT LEAD TO FILL OUT
g_1	THIS IS FOR SENIOR QUANT LEAD TO FILL OUT
g_rev	THIS IS FOR SENIOR QUANT LEAD TO FILL OUT
g	THIS IS FOR SENIOR QUANT LEAD TO FILL OUT
vi	THIS IS FOR SENIOR QUANT LEAD TO FILL OUT
wi	THIS IS FOR SENIOR QUANT LEAD TO FILL OUT
ywi	THIS IS FOR SENIOR QUANT LEAD TO FILL OUT
95ci_lower	THIS IS FOR SENIOR QUANT LEAD TO FILL OUT
95ci_upper	THIS IS FOR SENIOR QUANT LEAD TO FILL OUT
ci_low_3sf	THIS IS FOR SENIOR QUANT LEAD TO FILL OUT
ci_high_3sf	THIS IS FOR SENIOR QUANT LEAD TO FILL OUT
ci	THIS IS FOR SENIOR QUANT LEAD TO FILL OUT
wb_g	THIS IS FOR SENIOR QUANT LEAD TO FILL OUT
Checked	THIS IS FOR EFFECT SIZE RELIABILITY CHECKER TO FILL OUT
ROB Category	THIS IS FOR SENIOR QUANT LEAD OR PM TO FILL OUT

Appendix 3: Critical appraisal tools
Appraisal of risk of bias for impact evaluations using RCT designs

The following table provides a provisional tool to guide the risk of bias assessment for quantitative impact evaluations.

Provisional risk of bias assessment tool (RCT)

General	ID	EPPI ID	
General	Study first author	Open answer	
General	Time taken to complete assessment	Minutes	
General	Design type: What type of study design is used?	1 = Randomized controlled trial (RCT) (random assignment to households/individuals) or quasi-RCT 2 = Cluster-RCT (quasiRCT)	–
General	Methods used for analysis: Which methods are used to control for selection bias and confounding?	1 = Statistical matching (PSM, CEM, covariate matching) 2 = Difference-in-differences (DID) estimation methods 3 = IV-regression (2stage least squares or bivariate probit) 4 = Heckman selection model 5 = Fixed effects regression 6 = Covariate adjusted estimation 7 = Propensity-weighted regression 8 = Comparison of means= Other (please state)	–
General	Design and analysis method description	Open answer	Briefly describe the study design and analysis method undertaken by the authors
General	Study population	Open answer	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 that study participants? Was this a pilot program aimed at being scaled up? d) Were there specific factors of success or failure in the implementation?
General	Type of comparison group	1 = No intervention (Service delivery as usual) 2 = Other intervention 3 = Pipeline (waitlist) control (still service delivery as usual)	Indicate type of comparison group
General	Type of comparison group (if other)	Open answer	
General	Ethical clearance	Open answer	Provide any details of ethical research clearances granted. Report unclear if this information is not available
General	Study registration	Open answer	Provide any details of study registration, including registry IDs, etc.

General	ID	EPPI ID		
1: Assignment mechanism—Assessment	Assignment mechanism: Was the allocation or identification mechanism random or as good as random?	1 = Yes, 2 = Probably Yes, 3 = Probably No, 4 = No, 8 = Unclear	<p>a) The authors describe a random component in sequence generation/ randomization method (e.g lottery, coin toss, random number generator) and assignment is performed for all units at the start of the study centrally or using a method concealed from participants and intervention delivery</p> <p>b) If public lottery is used for the sequence generation, authors provide detail on the exact settings and participants attending the lottery</p> <p>c) If a special randomization procedure is used to ensure balance, it is well described and justified given the study setting (stratification, pairwise matching, unique random draw, multiple random draws, etc.)</p> <p>d) 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)</p>	<p>Score "Yes" if all criterion a), b), c) and d) are satisfied</p> <p>Score "Probably Yes" if only criterion a) and b) are not satisfied OR if only criteria c) is not satisfied</p> <p>Score "Unclear" if d) is not satisfied because no balance table is reported</p> <p>Score "Probably No" if d) is not satisfied because there is no balance table reported and there is evidence suggesting a problem in the randomization, such as baseline coefficients in a diff-in-diff regression table are very different or sample size is too small for the procedure used (using stratification when there are less than two units for each intervention and control group in each strata can lead to imbalance)</p> <p>Score "No" if d) is not satisfied because there are large imbalances concerning a large number of variables, providing evidence that the assignment was not random. If this is scored as no, use the NRS tool</p>
1: Assignment mechanism—Justification	Assignment justification	Open answer	Justification for coding decision (Include a brief summary of justification for rating, mentioning your response to all sub-questions, cite relevant pages)	
2: Unit of analysis—Assessment	Unit of analysis: Is unit of analysis in cluster allocation addressed in standard error calculation?	1 = Yes 2 = No 3 = Not reported/unclear 4 = Not applicable	<p>Score "Yes" if UoA = UoR OR if UoA ≠ UoR and standard errors are clustered at the UoR level OR data is collapsed to the UoR level</p> <p>Score "Not reported/unclear" if not enough information is provided on the way the standard errors were calculated or what the unit of analysis is</p> <p>Score "Not applicable" if it is not a cluster RCT</p> <p>Score "No" otherwise</p>	

General	ID	EPPI ID	
3: Selection bias-Assessment	Selection bias Was any differential selection into or out of the study (attrition bias) adequately resolved?	1 = Yes, 2 = Probably Yes, 3 = Probably No, 4 = No, 8 = Unclear	Score "Yes" if there is no attrition or attrition falls into the green zone and the study establishes that attrition is randomly distributed (e.g., by presenting balance by key characteristics across groups) AND if survey respondents were randomly sampled Score "Probably yes" if attrition falls into the green zone AND if survey respondents were randomly sampled Score "Unclear" if there is an attrition problem but no information provided on the relationship between attrition and treatment status, OR if there is not enough information on how the population surveyed was sampled Score "Probably no" if there is attrition which is likely to be related to the intervention OR is some indication that the survey respondents were purposely sampled in a way that might have led the sampling to be different between treatment and control groups, or attrition falls into the yellow zone Score "No" if attrition falls into the red zone
3: Selection bias-Justification	Selection bias justification	Open answer	Justification for coding decision (Include a brief summary of justification for rating, mentioning your response to all sub-questions, cite relevant pages)
4: Confounding- Assessment	Confounding and group equivalence: Was the method of analysis executed adequately to ensure comparability of groups throughout the study and prevent confounding	1 = Yes, 2 = Probably Yes, 3 = Probably No, 4 = No, 8 = Unclear	a) Baseline characteristics are similar in magnitude; b) Unbalanced covariates at the individual and cluster level are controlled in adjusted analysis; c) Adjustments to the randomization were taken into account in the analysis (stratum fixed effects, pairwise matching variables)? (Bruhn and McKenzie 2009) Score "Yes" if criterion a) and b) are satisfied; Score "Probably yes" if a) is not satisfied but b) is satisfied and imbalances are small in magnitude OR if only a) is satisfied Score "Unclear" if no balance table is provided or if imbalances are controlled for but they are very large in magnitude and assignment mechanism is not coded as "Yes" or "Probably yes" Score "Probably no" if a) and b) are not satisfied and the magnitude of imbalances are small Score "No" if a) and b) are not satisfied and the magnitude of imbalances are large, and covariates are clear determinant of the outcomes
4: Confounding-Justification	Confounding justification	Open answer	Justification for coding decision (Include a brief summary of justification for rating, mentioning your response to all sub-questions, cite relevant pages)

General	ID	EPPI ID		
5: Deviations from intended interventions—Assessment	Deviations from intended interventions: Spillovers, crossovers, and contamination: was the study adequately protected against spillovers, crossovers, and contamination?	1 = Yes, 2 = Probably Yes, 3 = Probably No, 4 = No, 8 = Unclear	<p>a) There were no implementation issues that might have led the control participants to receive the treatment (implementer's mistake)</p> <p>b) The intervention is unlikely to spillover to comparisons (e.g., participants and non-participants are geographically and/or socially separated from one another and general equilibrium effects are not likely) or the potential effects of spill overs were measured (e.g., variation in the % of unit within a cluster receiving the treatment)</p> <p>There is no risk of contamination by external programs: the treatment and comparisons are isolated from other interventions which might explain changes in outcomes</p> <p>d) There is nothing in the surveys that might have given the control participants an idea of what the other group might receive OR they did but there is no risk that this has changed their behaviors; AND the survey process did not reveal information to the control group that they did not have before (e.g., the study aims to measure increase in take up of a service or product that participants might not know about) Authors might put something in place in the design of the study that allows to control for that survey effect (e.g., a pure control with no monitoring except baseline end line)</p>	<p>Score "Yes" if criterion a), b), c) and d) are satisfied;</p> <p>Score "Probably yes" if there is no obvious problem but there is no information reported on potential risks related to spillovers, contamination, or survey effects in the control group OR if there were issues with spillovers but they were controlled for or measured</p> <p>Score "Unclear" if spillovers, crossovers, survey effects and/or contamination are not addressed clearly</p> <p>Score "Probably no" if any of the criterion a), b), c) or d) are not satisfied but the scale of the issue is not clear</p> <p>Score "No" if any of the criterion a), b), c) or d) are not satisfied and happened at a large scale in the study</p>
5: Deviations from intended interventions—Justification	Deviations justification	Open answer	<p>Justification for coding decision (Include a brief summary of justification for rating, mentioning your response to all sub-questions, cite relevant pages)</p> <p>For example, intervention groups are geographically separated, authors use intention to treat estimation or instrumental variables to account for non-adherence, and survey questions are not likely to expose individuals in the control group to information about desirable behaviors ('survey effects')</p>	

General	ID	EPPI ID		
6. Performance bias -Assessment	Performance bias: Was the process of monitoring individuals unlikely to introduce motivation bias among participants?	1 = Yes, 2 = Probably Yes, 3 = Probably No, 4 = No, 8 = Unclear	<p>a) The authors state explicitly that the process of monitoring the intervention and outcome measurement is blinded and conducted in the same frequency for treatment and control groups, or argue convincingly why it is not likely that being monitored could affect the performance of participants in treatment and comparison groups in different ways (such as resulting in Hawthorne or John Henry effects)</p> <p>b) The outcome is based on data collected in the context of a survey, and not associated with a particular intervention trial, or data are collected from administrative records or in the context of a retrospective (ex post) evaluation</p>	<p>Score "Yes" if either criterion a) or b) are satisfied;</p> <p>Score "Probably yes" if the study is based on data collected during a trial and there is no obvious issue with the monitoring processes, but authors do not mention potential risks</p> <p>Score "Unclear" if it is not clear whether the authors use an appropriate method to prevent Hawthorne and John Henry Effects (e.g., blinding of outcomes and, or enumerators, other methods to ensure consistent monitoring across groups)</p> <p>Hawthorne effects may result where participants know that they are being observed and John Henry Effects may result from participant knowledge of being compared</p> <p>Score "Probably no" if there was imbalance in the frequency of monitoring in intervention groups, which might have influenced participants' behaviors</p> <p>Score "No" if neither criterion a) or b) are satisfied</p>
6. Performance bias-Justification	Performance bias justification	Open answer	Justification for coding decision (Include a brief summary of justification for rating, mentioning your response to all sub-questions, cite relevant pages)	
7. Outcome measurement bias - Assessment	Outcome measurement bias: Was the study free from biases in outcome measurement?	1 = Yes, 2 = Probably Yes, 3 = Probably No, 4 = No, 8 = Unclear	<p>a) Outcome assessors are blinded, or the outcome measures are not likely to be biased by their judgment</p> <p>b) For self-reported outcomes: respondents in the intervention group are not more likely to have accurate answers due to recall bias;</p> <p>c) For self-reported outcomes: respondents do not have incentives to over/under report something related to their performance or actions, OR researchers put in place mechanisms to reduce the risk of reporting bias (researchers not strongly involved in the implementation of the program and it is clear that their answers to the survey will not affect what they receive in future) OR authors have measured the risks of bias through falsification tests or measuring the effect on placebo outcomes in cases where there was a risk of reporting bias</p> <p>d) Timing issue: the data collection period did not differ between intervention and comparison group; the baseline data is not likely to be affected by the beginning of the intervention or affects a small percentage of the study participants</p>	<p>Score "Yes" if criterion a), b), c) and d) are satisfied;</p> <p>Score "Probably yes" if there is a small risk related to any of a), b), c) or d) and there is no more information provided to justify the absence of bias OR if there was a high risk of bias, but authors have either controlled it in their design or measured it with a placebo outcome</p> <p>Score "Unclear" if it there is a high risk related to any of a), b), c) or d) and there is no more information provided to justify the absence of bias</p> <p>Score "Probably no" if there are high risk related to a), b), c) or d) and it is clear that authors were not able to control for this bias</p> <p>Score "No" if there is evidence of bias</p>

General	ID	EPPI ID		
7. Outcome measurement bias-Justification	Outcome measurement justification	Open answer	Justification for coding decision (Include a brief summary of justification for rating, mentioning your response to all sub-questions, cite relevant pages)	
8. Reporting bias-Assessment	Analysis reporting: Was the study free from selective analysis reporting?	1 = Yes, 2 = Probably Yes, 3 = Probably No, 4 = No, 8 = Unclear	a) A pre-analysis plan or trial protocol is published and referred to or the trial was preregistered, or the outcomes were preregistered; b) Authors report results corresponding to the outcomes announced in the method section (there is no outcome reporting bias); c) Authors report results of unadjusted analysis and intention to treat (ITT) estimation, alongside any adjusted and treatment-on-the-treated/complier average-causal effects analysis.) d) Authors use the appropriate analysis method (use baseline data when available), and different treatment arms are differentiated in the analysis e) Authors have reported all the analysis which could help understand the results and no other bias is assessed as unclear due to the lack of an important analysis (e.g., a balance table or a subgroup analysis)	Score "Yes" if all the criterion a), b), c), d), and e) are satisfied; Score "Probably yes" if all the conditions are met except a), or if all the conditions are met but there is some element missing that could have helped understand the results better (e); Score "Unclear" if there is not enough information to determine that there is an analysis missing; Score "Probably no" if any of the criterion b), c) or d) are not satisfied; Score "No" if any of the criterion b), c) or d) are not satisfied and there is evidence that the analysis results would be different because large imbalances were not controlled for, compliance was very low and ITT estimation was not reported or different treatment arms were pooled
8. Reporting bias-Justification	Analysis reporting justification	Open answer	Justification for coding decision (Include a brief summary of justification for rating, mentioning your response to all sub-questions, cite relevant pages)	
9. Other bias-Assessment	Other risks of bias Is the study free from other sources of bias?	1 = Yes, 4 = No		
9. Other bias-Justification	Other bias justification	Open answer	Justification for coding decision	
10. Blinding-observers-Assessment	Blinding of participants?	1 = Yes 2 = No 8 = unclear 9 = N/A	If there is no information, code NO. If there is information but it is ambiguous, code UNCLEAR	
10. Blinding—observers—Assessment	Blinding of outcome assessors?	1 = Yes 2 = No 8 = unclear 9 = N/A	If there is no information, code NO. If there is information but it is ambiguous, code UNCLEAR	
10. Blinding-analysts-Assessment	Blinding of data analysts?	1 = Yes 2 = No 8 = unclear 9 = N/A	If there is no information, code NO. If there is information but it is ambiguous, code UNCLEAR	
10. Blinding-method(s)	Method(s) used to blind	Open answer (including describe method of placebo control) No 9 = N/A	Describe method(s) used to blind	
11. External validity-Assessment	External validity	Open answer	a) What do authors say about external validity?	Include all information that can help assess the external validity of the results

Summary of justification for rating, mentioning your response to all sub-questions, cite relevant pages).

Appraisal of risk of bias for impact evaluations using quasi-experimental designs

Risk of bias assessment tool (QED)

Code	Question	Coding	Criteria	Decision-rules
General	ID	EPPI ID		
General	Time taken to complete assessment	Minutes		
General	Study first author	Open answer		
General	Outcomes assessed	Open answer		
General	Study design: What type of study design is used?	1 = Natural experiment: randomized or as-if randomized 2 = Natural experiment: regression discontinuity (RD) 3 = CBA (non-randomized assignment with treatment and contemporaneous comparison group, baseline, and end line data collection) – individual repeated measurement 4 = CBA pseudo panel (repeated measurement for groups but different individuals) 5 = Interrupted time series (with or without contemporaneous control group) 6 = Panel data, but no baseline (pre-test) 7 = Comparison group with end line data only		
General	Methods used for analysis: Which methods are used to control for selection bias and confounding?	1 = Statistical matching (PSM, CEM, covariate matching) 2 = Difference-in-differences (DID) estimation methods 3 = IV-regression (2-stage least squares or bivariate probit) 4 = Heckman selection model 5 = Fixed effects regression 6 = Covariate adjusted estimation 7 = Propensity-weighted regression 8 = Comparison of means = Other (please state)	–	

Code	Question	Coding	Criteria	Decision-rules
General	Study population	Open answer	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 that study participants? c) Was this a pilot program aimed at being scaled up? d) Were there specific factors of success or failure in the implementation?	
General	Ethical clearance	Open answer	Provide any details of ethical research clearances granted. Report unclear if this information is not available	
1: Selection bias-Assessment	1—Mechanism of assignment: was the allocation or identification mechanism able to control for selection bias?	1 = Yes, 2 = Probably Yes, 3 = Probably No, 4 = No, 8 = Unclear		
1: Selection bias-Justification	For regression discontinuity designs	Open answer	a) Allocation is made based on a predetermined discontinuity on a continuous variable (Regression discontinuity design) and blinded to participants or; b) if not blinded, individuals reasonably cannot affect the assignment variable in response to knowledge of the participation decision rule; c) and the sample size immediately at both sides of the cutoff point is sufficiently large to equate groups on average	Score "Yes" if criteria a), b), c) are all satisfied Score "Probably Yes" if there are minor differences in between both sides of the cut-off point but authors convincingly argue that the differences are unlikely to affect the outcome, OR individuals are not blinded and there are low risk of them affecting the assignment, but the authors do not mention it Score "Unclear" if it is unclear whether participants can affect it in response to knowledge of the allocation mechanism Score "Probably No" if there are differences between individuals on both sides of the cut-off point, and there are doubts that the differences are due to individuals altering the assignment OR the participants are blinded but there is evidence that the decisions that determined the discontinuity is based on differences between the two groups or differences in time Score "No" if the sample size is not sufficient OR there is evidence that participants altered the assignment variable prior to assignment. If the research has serious concerns with the validity of the assignment process or the group equivalence completely fails, we recommend assessing risk of bias of the study using the relevant questions for the appropriate methods of analysis (cross-sectional regressions, difference-in-difference, etc.) rather than the RDDs questions

Code	Question	Coding	Criteria	Decision-rules
1: Selection bias-Justification	For assignment-based nonrandomised program placement and self-selection (studies using a matching strategy or regression analysis, excluding IV)	Open answer	<p>a) Participants and non-participants are either matched based on all relevant characteristics explaining participation and outcomes, or;</p> <p>b) all relevant characteristics are accounted for,** and the data set used contains relevant variables that are measured in a relevant way (i.e., they were not collected for a different purpose initially and therefore are good proxy for some characteristics)</p> <p>**Accounting for and matching on all relevant characteristics is usually only feasible when the program allocation rule is known and there are no errors of targeting. It is unlikely that studies not based on randomization or regression discontinuity can score "YES" on this criterion. There are different ways in which covariates can be taken into account. Differences across groups in observable characteristics can be considered as covariates in the framework of a regression analysis or can be assessed by testing equality of means between groups. Differences in unobservable characteristics can be taken into account using instrumental variables (see also question 1.d) or proxy variables in the framework of a regression analysis, or using a fixed effects or difference-in-differences model if the only characteristics which are unobserved are time-invariant</p>	<p>Score "Yes" if a) or b) and c) are satisfied</p> <p>Score "Probably yes" if a) or b) are addressed for but there is some doubt related to c), OR authors combined statistical matching and difference-in-difference to cope with unobservable differences, OR they only did statistical matching and there were clear rules for selection into the program (no self-selection)</p> <p>Score "Unclear" if · it is not clear whether all relevant characteristics (only relevant time-varying characteristics in the case of panel data regressions) are controlled</p> <p>Score "Probably no" if only a statistical matching was done and there was self-selection into the program</p> <p>Score "No" if relevant characteristics are omitted from the analysis</p>
1: Selection bias-Justification	For identification based on an instrumental variable (IV estimation)	Open answer	<p>Score "Yes" if an appropriate instrumental variable is used which is exogenously generated: for example, due to a 'natural' experiment or random allocation</p> <p>Score "Probably yes" if there is less evidence (no balance table showing differences between the intervention and comparison group)</p> <p>Score "Unclear" if the exogeneity of the instrument is unclear (both externally as well as why the variable should not enter by itself in the outcome equation)</p> <p>Score "Probably no" if there is evidence that enrolment in the program is correlated with a variable that might also influence outcome and on the instrumental variable</p> <p>Score "No" if it is clear that the instrument is not exogenous and affect the outcome through other channels than the program</p>	
2: Confounding-Assessment	2—Group equivalence: was the method of analysis executed adequately to ensure comparability of groups throughout the study and prevent confounding?	1 = Yes, 2 = Probably Yes, 3 = Probably No, 4 = No, 8 = Unclear		

Code	Question	Coding	Criteria	Decision-rules
2: Con- found- ing-Justi- fication	For regression discontinuity design	Open answer	<p>a) The interval for selection of treatment and control group is reasonably small OR authors have weighted the matches on their distance to the cutoff point; and</p> <p>b) the mean of the covariates of the individuals immediately at both sides of the cut-off point (selected sample of participants and non-participants) are overall not statistically different based on t-test or ANOVA for equality of means;</p> <p>c) Significant differences in covariates of the individuals have been controlled in multivariate analysis; and for cluster assignment, authors control for external cluster-level factors that might confound the impact of the program</p>	<p>Score "Yes if criterion a), b), c) and d) are addressed</p> <p>Score "Probably yes" if b) is not addressed but c) is addressed and differences in means are not large</p> <p>Score "Unclear" if insufficient details are provided on controls; or if insufficient details are provided on cluster controls</p> <p>Score "Probably no" if b) is not addressed (absence of a difference test or balance table) and there are doubt regarding the continuity on both sides of the cut-off point (a)</p> <p>Score "No" otherwise</p>
2: Con- found- ing-Justi- fication	For non-randomized trials using difference-in-differences methods of analysis	Open answer	<p>a) The authors use a difference-in-differences (or fixed effects) multivariate estimation method;</p> <p>b) the authors control for a comprehensive set of individual time-varying characteristics, and for cluster assignment, authors control for external cluster-level factors that might confound the impact of the program**;</p> <p>c) and the attrition rate is sufficiently low and similar in treatment and control, or the study assesses that dropouts are random draws from the sample (for example, by examining correlation with determinants of outcomes, in both treatment and comparison groups);</p> <p>**Knowing allocation rules for the program – or even whether the non-participants were individuals that refused to participate in the program, as opposed to individuals that were not given the opportunity to participate in the program – can help in the assessment of whether the covariates accounted for in the regression capture all the relevant characteristics that explain differences between treatment and comparison groups</p>	<p>Score "Yes, if a, b, c, d (if relevant) is addressed and baseline imbalances between groups were relatively low OR the method was combined by a statistical matching</p> <p>Score "Probably yes" if all possible variables are controlled for and the selection into the program was done according to clear rules, but baseline imbalances between groups were very large</p> <p>Score "Unclear" if insufficient details are provided; or if insufficient details are provided on cluster controls</p> <p>Score "Probably no" if some time-varying characteristics are not controlled for and the program was self-selected by the intervention groups</p> <p>Score "No" if any of the criterion is not addressed</p>

Code	Question	Coding	Criteria	Decision-rules
2: Con- found- ing-Justi- fication	For statistical matching studies including propensity scores (PSM) and covariate matching** **Matching strategies are sometimes complemented with difference-indifference only uses in the estimation the common support region of the sample size, reducing the likelihood of existence of time variant unobservable differences across groups affecting outcome of interest and removing biases arising from time-invariant unobservable characteristics, regression estimation methods. This combination approach is superior since it	Open answer	<p>a) Matching is either on baseline characteristics or time-invariant characteristics which cannot be affected by participation in the program; and the variables used to match are relevant (for example, demographic and socio-economic factors) to explain both participation and the outcome (so that there can be no evident differences across groups in variables that might explain outcomes); and, for cluster assignment, authors control for external cluster-level factors that might confound the impact of the program</p> <p>b) in addition, for PSM Rosenbaum's test suggests the results are not sensitive to the existence of hidden bias; and,</p> <p>c) with the exception of Kernel matching, the means of the individual covariates are equated for treatment and comparison groups after matching;</p> <p>d) different matching methods including varying sample sizes yields the same results and authors consider the use of control observations multiple times against the same treatment in their standard error calculation</p>	<p>Score "Yes, if a, b, c, and d (if relevant) are addressed</p> <p>Score "Probably yes" if the selection into the program was done according to clear rules, which are used for the matching but there are slight imbalances remaining after matching</p> <p>Score "Unclear" if relevant variables are not included in the matching equation, or if matching is based on characteristics collected at end line; or if insufficient details are provided on cluster controls</p> <p>Score "Probably no" if the program was self-selected by the intervention groups or participants OR if the selection into the program was done according to clear rules but there is no baseline data available to match the participants or groups on</p> <p>Score "No" if matching was done based on variables that are likely to be affected by the program or any other scenario that affect a), b) c) or d)</p>
2: Con- found- ing-Justi- fication	For regression-based studies using cross-sectional data (excluding IV)	Open answer	<p>a) The study controls for relevant confounders that may be correlated with both participation and explain outcomes (for example, demographic and socio-economic factors at individual and community level) using multivariate methods with appropriate proxies for unobservable covariates, and, for cluster assignment, authors control particularly for external cluster-level factors that might confound the impact of the program;</p> <p>b) and a Hausman test with an appropriate instrument suggests there is no evidence of endogeneity**;</p> <p>c) and none of the covariate controls can be affected by participation;</p> <p>d) and either, only those observations in the region of common support for participants and non-participants in terms of covariates are used, or the distributions of covariates are balanced for the entire sample population across groups;</p> <p>**The Hausman test explores endogeneity in the framework of regression by comparing whether the OLS and the IV approaches yield significantly different estimations. However, it plays a different role in the different methods of analysis. While in the OLS regression framework the Hausman test mainly explores endogeneity and therefore is related with the validity of the method, in IV approaches it explores whether the author has chosen the best available strategy for addressing causal attribution (since in the absence of endogeneity OLS yields more precise estimators) and therefore is more related with analysis reporting bias</p>	<p>Score "Yes if a, b, c and d are addressed</p> <p>Score "Probably yes" if all criteria are addressed but authors did not report the Hausman test (b)</p> <p>Score "Unclear" if relevant confounders are controlled but appropriate proxy variables or statistical tests are not reported; or if insufficient details are provided on cluster controls</p> <p>Score "Probably no" if any of the criterion other than b) is not addressed</p> <p>Score "No" if none of the criterion are addressed</p>

Code	Question	Coding	Criteria	Decision-rules
2: Con- found- ing-Justi- fication	For identification based on an instrumental variable (IV estimation)	Open answer	<p>a) The instrumenting equation is significant at the level of $F \geq 10$ (or if an F test is not reported, the authors report and assess whether the R-squared (goodness of fit) of the participation equation is sufficient for appropriate identification); b) the identified instruments are individually significant ($p \leq 0.01$); for Heckman models, the identifiers are reported and significant ($p \leq 0.05$);</p> <p>c) where at least two instruments are used, the authors report on an over-identifying test ($p \leq 0.05$ is required to reject the null hypothesis); and none of the covariate controls can be affected by participation and the study, and authors convincingly assesses qualitatively why the instrument only affects the outcome via participation. If the instrument is the random assignment of the treatment, the reviewer should also assess the quality and success of the randomization procedure in part a)</p> <p>d) and, for cluster assignment, authors particularly control for external cluster-level factors that might confound the impact of the program (for example, weather, infrastructure, community fixed effects, and so forth) through multivariable analysis</p>	<p>Score "Yes, if a, b, c, d (if relevant) is addressed</p> <p>Score "Probably yes" if one of the tests required for criterion a) or b) is not reported but the other is, and the rest of the criterion are addressed, and the instrument is convincing</p> <p>Score "UNCLEAR" if relevant confounders are controlled for but appropriate statistical tests are not reported; or if insufficient details are provided on cluster controls</p> <p>Score "Probably no" if exogeneity of the instrument is not convincing and appropriate tests are not reported</p> <p>Score "No" otherwise if any of the tests required for criterion a), b) or c) are reported and not satisfied</p>
3: Perform- ance bias- Assess- ment	3—Performance bias: was the process of being observed free from motivation bias?	1 = Yes, 2 = Probably Yes, 3 = Probably No, 4 = No, 8 = Unclear	<p>a) For data collected in the context of a particular intervention trial (randomized or nonrandomised assignment), the authors state explicitly that the process of monitoring the intervention and outcome measurement is blinded, or argue convincingly why it is not likely that being monitored could affect the performance of participants in treatment and comparison groups in different ways (such as resulting in Hawthorne or John Henry effects)</p> <p>b) The study is based on data collected in the context of a survey, and not associated with a particular intervention trial, or data are collected from administrative records or in the context of a retrospective (ex post) evaluation</p>	<p>Score "Yes" if either criterion a) or b) are satisfied;</p> <p>Score "Probably yes" if the study is based on survey data collected during a trial and there is no obvious issue with the monitoring processes, but authors do not mention potential risks</p> <p>Score "Unclear" if it is not clear whether the authors use an appropriate method to prevent Hawthorne and John Henry Effects (e.g., blinding of outcomes and, or enumerators, other methods to ensure consistent monitoring across groups)</p> <p>Hawthorne effects may result where participants know that they are being observed and John Henry Effects may result from participant knowledge of being compared</p> <p>Score "Probably no" if there was imbalance in the frequency of monitoring in intervention groups, which might have influenced participants' behaviors</p> <p>Score "No" if both criterion a) and b) are not satisfied</p>
3: Perform- ance bias-Justi- fication	Performance bias-Justification	Open answer	Justification for coding decision (Include a brief summary of justification for rating, mentioning your response to all sub-questions, cite relevant pages)	

Code	Question	Coding	Criteria	Decision-rules
4: Spillovers, crossovers, and contamination-Assessment	4—Spillovers, crossovers, and contamination: was the study adequately protected against spillovers, crossovers, and contamination?	1 = Yes, 2 = Probably Yes, 3 = Probably No, 4 = No, 8 = Unclear	<p>a) There were no implementation issues that might have led the control participants to receive the treatment (implementer's mistake) The intervention is unlikely to spillover to comparisons (e.g., participants and non-participants are geographically and/or socially separated from one another and general equilibrium effects are not likely) or the potential effects of spill overs were measured (e.g., variation in the % of unit within a cluster receiving the treatment)</p> <p>c) There is no risk of contamination by external programs: the treatment and comparisons are isolated from other interventions which might explain changes in outcomes</p> <p>b) There is nothing in the surveys that might have given the control participants an idea of what the other group might receive OR they did but there is no risk that this has changed their behaviors; AND the survey process did not reveal information to the control group that they did not have before (e.g., the study aims to measure increase in take up of a service or product that participants might not know about) Authors might put something in place in the design of the study that allows to control for that survey effect (e.g., a pure control with no monitoring except baseline end line)</p>	<p>Score "Yes" if criterion a), b), c) and d) are satisfied;</p> <p>Score "Probably yes" if there is no obvious problem but there is no information reported on potential risks related to spill overs, contamination, or survey effects in the control group OR if there were issues with spillovers but they were controlled for or measured</p> <p>Score "Unclear" if spillovers, crossovers, survey effects and/or contamination are not addressed clearly</p> <p>Score "Probably no" if any of the criterion a), b), c) or d) are not satisfied but the scale of the issue is not clear</p> <p>Score "No" if any of the criterion a), b), c) or d) are not satisfied and happened at a large scale in the study</p>
4: Spillovers, crossovers, and contamination-Justification	Spillovers, crossovers, and contamination-Justification	Open answer	Justification for coding decision (Include a brief summary of justification for rating, mentioning your response to all sub-questions, cite relevant pages)	
5: Outcome measurement bias-Assessment	5—Outcome measurement bias	1 = Yes, 2 = Probably Yes, 3 = Probably No, 4 = No, 8 = Unclear	<p>a) Outcome assessors are blinded, or the outcome measures are not likely to be biased by their judgment</p> <p>b) For self-reported outcomes: respondents in the intervention group are not more likely to have accurate answers due to recall bias;</p> <p>c) For self-reported outcomes: respondents do not have incentives to over/under report something related to their performance or actions, OR researchers put in place mechanisms to reduce the risk of reporting bias (researchers not strongly involved in the implementation of the program and it is clear that their answers to the survey will not affect what they receive in future) OR authors have measured the risks of bias through falsification tests or measuring the effect on placebo outcomes in cases where there was a risk of reporting bias</p> <p>d) Timing issue: the data collection period did not differ between intervention and comparison group; the baseline data is not likely to be affected by the beginning of the intervention or affects a small percentage of the study participants</p>	<p>Score "Yes" if criterion a), b), c) and d) are satisfied;</p> <p>Score "Probably yes" if there is a small risk related to any of a), b), c) or d) and there is no more information provided to justify the absence of bias OR if there was a high risk of bias, but authors have either controlled it in their design or measured it with a placebo outcome</p> <p>Score "Unclear" if it there is a high risk related to any of a), b), c) or d) and there is no more information provided to justify the absence of bias</p> <p>Score "Probably no" if there are high risk related to a), b), c) or d) and it is clear that authors were not able to control for this bias</p> <p>Score "No" if there is evidence of bias</p>

Code	Question	Coding	Criteria	Decision-rules
5: Outcome measurement bias-Justification	Outcome measurement bias-Justification	Open answer	Justification for coding decision (Include a brief summary of justification for rating, mentioning your response to all sub-questions, cite relevant pages)	
6: Reporting bias-Assessment	6—Selective analysis reporting: was the study free from selective analysis reporting?	1 = Yes, 2 = Probably Yes, 3 = Probably No, 4 = No, 8 = Unclear	<p>a) a pre-analysis plan is published, especially for prospective NRS, but it should also be for retrospective studies b) authors use 'common' methods of estimation (i.e., credible analysis method to deal with attribution given the data available); c) There is no evidence that outcomes were selectively reported (e.g., results for all relevant outcomes in the methods section are reported in the results section); d) Requirements for specific methods of analysis:</p> <p>- For PSM and covariate matching: (a) Where over 10% of participants fail to be matched, sensitivity analysis is used to re-estimate results using different matching methods (Kernel Matching techniques); (b) For matching with replacement, no single observation in the control group is matched with a large number of observations in the treatment group.—For IV (including Heckman) models, (a) The authors test and report the results of a Hausman test for exogeneity ($p \leq 0.05$ is required to reject the null hypothesis of exogeneity); (b) the coefficient of the selectivity correction term (Rho) is significantly different from zero ($P < 0.05$) (Heckman approach)</p> <p>- For studies using multivariate regression analysis, authors conduct appropriate specification tests (e.g., testing robustness of results to the inclusion of additional variables, or (very rare) reporting results of multicollinearity test, etc.)</p>	<p>Score "Yes" if a), b), c) and d) are satisfied OR if a) is not met and it is a retrospective NRS</p> <p>Score "Probably Yes" if authors combined methods and reported relevant tests (d) is only for one method OR if all the criteria are met except for a) and it is a prospective NRS</p> <p>Score "Unclear" if intended outcomes not specified in the paper OR if any of the requirements for d) are not reported</p> <p>Score "Probably No" if b) is addressed, but authors did not present results for all outcomes announced in the method section OR did not meet requirement d) although reported</p> <p>Score "No" if authors use uncommon or less rigorous estimation methods such as failure to conduct multivariate analysis for outcomes equations OR if some important outcomes are subsequently omitted from the results or the significance and magnitude of important outcomes was not assessed</p>
6: Reporting bias-Justification	Analysis reporting bias—Justification	Open answer	Justification for coding decision (Include a brief summary of justification for rating, mentioning your response to all sub-questions, cite relevant pages)	
7: Other bias-Assessment	7—Other risks of bias: Is the study free from other sources of bias?	1 = Yes, 4 = No	Score "Yes" if the reported results do not suggest any other sources of bias. Score "No" if other potential threats to validity are present, and note these here (e.g., coherence of results, survey instruments used are not reported)	
7: Other bias-Justification	Other risks of bias-Justification	Open answer	Justification for coding decision (Include a brief summary of justification for rating, mentioning your response to all sub-questions, cite relevant pages)	
8: External validity	8—External validity	Open answer	Open answer- what do authors say about external validity if anything?	

Qualitative analysis tool

Study type	Methodological appraisal criteria	Response		
		Yes	No	Comment
<p>Screening questions: assessing 'fatal flaws' (Dixon-Woods 2005) Configurative 'fatal flaws' based on Pawson (2003) TAPUS framework</p>	<p><u>Configurative assessment:</u></p> <ul style="list-style-type: none"> • 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 (<i>Purposivity</i>) • The findings of the study are based on collected data, which justify the knowledge claims (<i>Accuracy</i>) <p>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!</p> <p>I. RESEARCH IS DEFENSIBLE IN DESIGN (providing a research strategy that addresses the question) <u>Appraisal indicators:</u> Bullet <i>Is the research design clearly specified and appropriate for aims and objectives of the research?</i> Consider whether</p> <ul style="list-style-type: none"> <i>i. there is a discussion of the rationale for the study design</i> <i>ii. the research question is clear, and suited to the inquiry</i> <i>iii. there are convincing arguments for different features of the study design</i> <i>iv. limitations of the research design and implications for the research evidence are discussed</i> <p>Defensi- Arguable Critical ble</p> <p>II. RESEARCH FEATURES AN APPROPRIATE SAMPLE (following an adequate strategy for selection of participants) <u>Appraisal indicators:</u> Consider whether</p> <ul style="list-style-type: none"> <i>i. there is a description of study location and how/why it was chosen</i> <i>ii. the researcher has explained how the participants were selected</i> <i>iii. the selected participants were appropriate to collect rich and relevant data</i> <i>iv. reasons are given why potential participants chose not take part in study</i> <p>Appropriate sample Functional sample Critical sample Flawed sample</p> <p>III. RESEARCH IS RIGOROUS IN CONDUCT (Providing a systematic and transparent account of the research process) <u>Appraisal indicators:</u> Consider whether</p> <ul style="list-style-type: none"> <i>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?)</i> <i>ii. researchers demonstrate that data collection targeted depth, detail, and richness of information (e.g., interview/observation schedule)</i> <i>iii. there is evidence of how descriptive analytical categories, classes, labels, etc. have been generated and used</i> <i>iv. presentation of data distinguishes clearly between the data, the analytical frame used, and the interpretation</i> <i>v. methods were modified during the study; and if so, has the researcher explained how and why?</i> <p>Rigorous Considerate conduct Critical Flawed conduct conduct conduct</p> <p>IV. RESEARCH FINDINGS ARE CREDIBLE IN CLAIM/BASED ON DATA (Providing well-founded and plausible arguments based on the evidence generated) <u>Appraisal indicators:</u> Consider whether</p> <ul style="list-style-type: none"> <i>i. there is a clear description of the form of the original data</i> <i>ii. sufficient amount of data is presented to support interpretations and findings/conclusions</i> 			<p>Not defensible</p> <p>Worth to continue:</p> <p>Worth to continue:</p> <p>Worth to continue:</p>

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

Study type	Methodological appraisal criteria	Response		
		Yes	No	Comment
OVERALL CRITICAL APPRAISAL DECISION				
Decision rule:				
– a single critical appraisal judgment ² in any of the 6 appraisal domains leads to a critical overall judgment				
– 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 (study generates new evidence relevant to the review question and complies with all methodological criteria to ensure reliability and empirical grounding of the evidence)	Moderate-quality Empirical research (study generates new evidence relevant to the review question and complies with reasonable methodological criteria to ensure reliability and empirical grounding of the evidence)	Low-quality Empirical research (study generates new evidence relevant to the review question and complies with minimum methodological criteria to ensure reliability and empirical grounding of the evidence)		Critical quality Empirical research (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. [9]; CASP (2006); CRD (2009); Dixon-Woods et al. (2004); Dixon-Woods et al. (2006); Greenhalgh and Brown (2014); Harden et al. (2004); Harden et al. (2009); Harden and Gough (2012); Mays and 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
<p>2. <i>Mixed-methods</i>²</p> <p><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</p> <p><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</p> <p><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)</p> <p><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</p>	<p>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</p> <p>i. the rationale for integrating qualitative and quantitative methods to answer the research question is explained [DEFENSIBLE]</p> <p>ii. 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]</p> <p>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]</p> <p>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]</p> <p>v appropriate consideration is given to the limitations associated with this integration, e.g., the divergence of qualitative and quantitative data (or results)? [REFLEXIVE]</p>			

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

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:

<u>Mixed-methods critical appraisal:</u>	<u>Qualitative critical appraisal:</u>	<u>Quantitative critical appraisal:</u>
1. Research is defensible in design	Include/Exclude	1. Low risk of bias
2. Research is rigorous in conduct		2. Risk of bias
3. Research is credible in claim		3. High risk of bias
4. Research is reflective		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 and Clark (2007); Crow (2013); Long (2005); O’Cathain et al. (2008); O’Cathain (2010); Pluye and Hong (2014); Sirriyeh et al. (2011)

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 judgment (i.e., ‘unreflective research’ above) with judgements moving further to the left on a scale from high to low critical appraisal

Appendix 4: Additional meta-analysis results

Detailed results for food security

A total of $k = 4$ studies were included in the analysis. The observed outcomes ranged from 0.07 to 0.67, with the majority of estimates being positive (100%). The estimated average outcome based on the random effects model was $\hat{\mu} = 0.24$ (95% CI: 0.00 to 0.47). Therefore, the average outcome differed significantly from zero ($z = 1.97, p = 0.05$). According to the Q-test, the true outcomes appear to be heterogeneous ($Q(3) = 111.16, p < 0.01, \hat{\tau}^2 = 0.06, I^2 = 97.30\%$).

An examination of the studentized residuals revealed that one study [25] had a value larger than ± 2.50 and may be a potential outlier in the context of this model.

Detailed results for food affordability/availability

We included a total of $k = 6$ studies were included in the analysis. The observed outcomes ranged from 0.08 to 0.49, with the majority of estimates being positive (100%). The estimated average outcome based on the random effects model was $\hat{\mu} = 0.23$ (95% CI: 0.09 to 0.38). Therefore, the average outcome differed significantly from zero ($z = 3.19, p < 0.01$). According to the Q-test, the true outcomes appear to be heterogeneous ($Q(15) = 187.27, p < 0.01, \hat{\tau}^2 = 0.02, I^2 = 91.99\%$).

An examination of the studentized residuals revealed that one study (Ahmed et al. 2019 had a value larger than ± 2.96 and may be a potential outlier in the context of this model.

Detailed results for diet quality and adequacy

We included a total of $k = 4$ studies in the analysis. The observed outcomes ranged from 0.08 to 0.14. The estimated average outcome based on the random effects model was $\hat{\mu} = 0.09$ (95% CI: 0.06 to 0.12). Therefore, the average outcome differed significantly from zero ($z = 5.64, p < 0.01$). According to the Q-test, there was no significant amount of heterogeneity in the true outcomes ($Q(3) = 0.53, p = 0.91, \hat{\tau}^2 = 0.00, I^2 = 0.00\%$).

An examination of the studentized residuals revealed that none of the studies had a value larger than ± 2.50 and hence there was no indication of outliers in the context of this model.

Detailed results for anthropometric measures

We included a $k = 2$ studies in the analysis. The estimated average outcome based on the random effects model was $\hat{\mu} = 0.12$ (95% CI: 0.00 to 0.23). Therefore, the average outcome did not differ significantly from zero ($z = 1.99, p = 0.05$). According to the Q-test, there was no significant amount of heterogeneity in the true outcomes ($Q(1) = 0.12, p = 0.73, \hat{\tau}^2 = 0.00, I^2 = 0.00\%$). Given the small number of studies, this result should be interpreted with caution.

Detailed results for well-being outcomes

We included a $k = 2$ studies in the analysis. The estimated average outcome based on the random effects model was $\hat{\mu} = 0.08$ (95% CI: 0.01 to 0.15). Therefore, the

average outcome did not differ significantly from zero ($z = 2.11, p = 0.034$). According to the Q-test, there was significant amount of heterogeneity in the true outcomes ($Q(1) = 2.90, p = 0.08, \hat{\tau}^2 = 0.00, I^2 = 65.57\%$). Given the small number of studies, this result should be interpreted with caution.

Appendix 5: Detailed risk of bias

See Tables 5 and 6

The nine additional qualitative studies were assessed. Five [37, 38, 39, 40, 48] were found to be high quality, with the remaining four [41, 49, 50, 51] marked as medium quality according to the assessment tool. The main

Table 5 Risk of bias in experimental studies

Author (Year)	Overall Score	Assignment mechanism Unit of analysis	Selection bias	Confounding	Deviations from intended Performance bias	Outcome measurement Reporting bias	
Heckert (2019)	Some concern	Low	Low	Low	Low	Low	Some concern
	ns	Risk	Risk	Risk	Risk	Risk	ns
Deining er (2009)	Some concern	Low	Some concern	Some concern	Low	Some concern	Low
	ns	Risk	ns	ns	Risk	ns	Risk
Haque (2021)	Some concern	Low	Low	Some concern	Low	Low	Low
	ns	Risk	Risk	ns	Risk	Risk	Risk
Blakstad (2020)	Low	Low	Low	Low	Low	Low	Low
	ROB	Risk	Risk	Risk	Risk	Risk	Risk
Bandiera (2017)	Some concern	Low	Low	Low	Low	Low	Some concern
	ns	Risk	Risk	Risk	Risk	Risk	ns
Ahmed (2019)	Some concern	Low	Low	Low	Low	Low	Some concern
	ns	Risk	Risk	Risk	Risk	Risk	ns

Table 6 Risk of bias in quasi-experimental studies

Author (Year)	Overall Score	Selection bias	Confounding	Performance bias	Spillovers, crossovers, and Outcome	measurement bias Reporting bias
Pan (2015)	Low ROB	Low risk	Low risk	Low risk	Low risk	Low risk
Marquis (2015)	Low ROB	Low risk	Low risk	Low risk	Low risk	Low risk
Emran (2009)	Some concerns	Low risk	Low risk	Some concerns	Low risk	Low risk
Bonuedi (2020)	High ROB	Low risk	High risk	High risk	High risk	High risk

factor differentiating high and medium quality qualitative studies was the level of rigor and detail provided in the methods. Triangulating data by interviewing different population groups in a given community allowed for different perspectives, making qualitative studies more rigorous. Sometimes the male head of household was interviewed along with the woman beneficiary, as well as other community members, which can affect the information reported. Studies were high quality if they triangulated data, used ethical methods (i.e., did not add additional burden onto women’s time) and added rich contextual layers to quantitative findings in other studies or the same study.

Appendix 6: Effect estimates from included studies

See Table 7

Appendix 7: Food system EGM framework and search strategy

See Table 8

The complete Food system EGM framework can be found at this link: <https://www.3ieimpact.org/sites/default/files/2021-01/EGM16-Online-appendix-A-Additional-methods-detail.pdf>

Website searched

Below is the list of databases and organizational websites searched in the FSN EGM. This online Appendix provides more detailed information about the search strategy: <https://www.3ieimpact.org/sites/default/files/2021-01/EGM16-Online-appendix-B-Search-strategy.pdf>

Academic databases

We conducted electronic searches of the following databases of published sources:

- MEDLINE
- EMBASE
- Cochrane Controlled Trials Register (CENTRAL)
- CINAHL
- CAB Global Health
- CAB Abstracts
- Agricola
- PsychINFO
- Africa-Wide Information
- Academic Search Complete
- Scopus
- Campbell Library

Table 7 Effect estimates from included studies in REAT

First author	Year	Country	Intervention type	Evaluation/synthesis method	Outcome	Standardized effect estimate (Confidence Interval)	Sample size
Food security							
Bandiera	2017	Bangladesh	Training/education Asset transfer	Randomized control trial and difference-in-difference	Food security index—whether HH had surplus food or deficit, enough food to eat, and could afford to eat two meals a day*	0.07 (0.03;0.12)	6732
Emran	2009	Bangladesh	Training/education Asset transfer	Difference-in-difference and statistical matching	Meals twice a day	0.6 (0.5; 0.7)	1569
Blackstad	2020	Tanzania	Training/education	Randomized control trial	Food availability: Sufficient food to meet the household's needs*	0.66 (0.56;0.76)	1569
Pan	2015	Uganda	Training/education Asset transfer	Regression discontinuity	Household food insecurity access scale*	0.07 (− 0.05; 0.2)	876
Food affordability and availability					Skip meals*	− 0.13 (− 0.2; − 0.06)	3368
Ahmed	2019	Bangladesh	Asset transfer (Cash)	Randomized control trial	Per capita monthly food consumption—North	0.13 (0.08;0.19)	5000
					Per capita monthly food consumption—South	0.08 (0.03;0.14)	5000
					Per capita daily intake caloric—North	0.07 (0.02;0.13)	5000
					Per capita daily intake caloric—South	0.02 (− 0.03;0.08)	5000
					Food consumption score—North	0.17 (0.12;0.23)	5000
					Food consumption score—South	0.07 (0.02;0.13)	5000
			Asset transfer (Food)		Per capita monthly food consumption—North	0.11 (0.06;0.17)	5000
					Per capita monthly food consumption—South	0.07 (0.02;0.13)	5000
					Per capita daily intake caloric—North	0.13 (0.08;0.19)	5000
					Per capita daily intake caloric—South	0.01 (− 0.05; 0.06)	5000
					Food consumption score—North	0.24 (0.18;0.29)	5000

Table 7 (continued)

First author	Year	Country	Intervention type	Evaluation/synthesis method	Outcome	Standardized effect estimate (Confidence Interval)	Sample size
			Assets transfer (Cash and food)		Food consumption score—South	0.12 (0.07; 0.18)	5000
					Per capita monthly food consumption—North	0.11 (0.06; 0.17)	5000
					Per capita monthly food consumption—South	0.11 (0.06; 0.17)	5000
					Per capita daily intake caloric—North	0.07 (0.02; 0.13)	5000
					Per capita daily intake caloric—South	0.03 (− 0.03; 0.09)	5000
					Food consumption score—North	0.16 (0.11; 0.22)	5000
					Food consumption score—South	0.11 (0.06; 0.17)	5000
			Behavior change communication		Per capita monthly food consumption—North	0.32 (0.27; 0.38)	5000
			Asset transfer (Cash)		Per capita monthly food consumption—South	0.22 (0.17; 0.28)	5000
					Per capita daily intake caloric—North	0.22 (0.17; 0.28)	5000
					Per capita daily intake caloric—South	0.10 0.04; 0.15)	5000
					Food consumption score—North*	0.49 (0.44; 0.55)	5000
					Food consumption score—South*	0.28 (0.23; 0.34)	5000
Bonuedi	2020	Sierra Leone	Behavior change communication Training/education Assets transfer	Statistical matching	Total food consumption expenditure in the 12 months preceding the survey (Food production and market purchases) (LOG)-Household*	− 0.04 (− 0.18; 0.09)	836
					Total food consumption expenditure in the 12 months preceding the survey (Food production and market purchases) (LOG)-Household*	0.22 (0.09; 0.36)	836

Table 7 (continued)

First author	Year	Country	Intervention type	Evaluation/synthesis method	Outcome	Standardized effect estimate (Confidence Interval)	Sample size
Deinger	2009	India	Training/education	Difference-in-difference and statistical matching	Food consumption (RS/year)—All groups*	0.09 (− 0.03; 0.2)	2199
					Energy intake p.c. (kcal/day)—All groups	0.02 (− 0.09; 0.14)	2199
					Food consumption (RS/year)-POP (Poorest of the poor)	0.19 (− 0.08; 0.47)	404
					Food consumption (RS/year)-Poor	0.42 (0.06; 0.77)	243
					Food consumption (RS/year)-Non-poor	− 0.11 (− 0.54; 0.33)	157
					Energy intake p.c. (kcal/day)-POP (Poorest of the poor)	0.36 (0.09; 0.64)	404
					Energy intake p.c. (kcal/day)-Poor	0.59 (0.23; 0.95)	243
					Energy intake p.c. (kcal/day)-Non-poor	− 0.08 (− 0.52; 0.36)	157
					Grain stocks (kg)*	0.22 (0.12; 0.32)	1569
					Per capita food consumption*	0.08 (0.01; 0.15)	3368
Diet quality and adequacy	2020	Sierra Leone	Behavior change communication Training/education	Propensity score matching	Household dietary diversity*	0.14 (0.00, 0.27)	836
					Women's dietary diversity	0.10 (− 0.05, 0.26)	636
					Children's dietary diversity	− 0.05 (− 0.21, 0.11)	575
					Household dietary diversity	0.23 (0.09, 0.26)	836
					Women's dietary diversity	0.31 (0.15, 0.46)	636
					Children's dietary diversity	0.29 (0.12, 0.45)	575
Haque	2021	Bangladesh	Training/education	Randomized control trial	Additional food consumed during pregnancy*	0.09 (0.05, 0.13)	10722
					Protein intake p.c. (g/day) in the total population*	0.08 (− 0.04, 0.19)	1099.5
Deninger	2009	India	Training/education	Difference-in-difference and statistical matching	Protein intake p.c. (g/day) among the poor of the poor	0.32 (0.05, 0.60)	202

Table 7 (continued)

First author	Year	Country	Intervention type	Evaluation/synthesis method	Outcome	Standardized effect estimate (Confidence Interval)	Sample size
Pan	2015	Uganda	Training/education Asset transfer	Regression discontinuity	Protein intake p.c. (g/day) among the poor Protein intake p.c. (g/day) among the non-poor Variety of foods consumed*	0.66 (0.30, 1.02) 0.20 (- 0.24, 0.64) 0.09 (0.02, 0.15)	121.5 78.5 3368
Anthropometrics							
Heckert	2019	Burkina Faso	Behavior change communication Asset transfer	Randomized control trial	Weight-for-length z score*	0.12 (0.00, 0.25)	1035
Marquis	2015	Ghana	Training/education Asset transfer	Difference-in-difference	BMI-for-age z score*	0.06 (- 0.30, 0.41)	121.6
					Weight-for-age z-score Height-for-age z score	- 0.42 (- 0.77, - 0.06) 0.40 (0.04, 0.75)	121.6 121.6
Micronutrient status							
Haque	2021	Bangladesh	Training/education	Randomized control trial	Consumption of at least 100 IFA tablets during pregnancy Received vitamin a capsule after last delivery Change in hemoglobin (g/dL)	0.25 (0.21, 0.28) 0.20 (0.16, 0.24) 0.10 (- 0.02, 0.23)	10722 10722 1035
Heckert	2019	Burkina Faso	Behavior change communication Asset transfer	Randomized control trial			
Wellbeing outcomes							
Bandiera	2017	Bangladesh	Training/education Asset transfer	Randomized control trial and difference-in-difference	Mental health index	0.04 (- 0.00, 0.09)	6732
Pan	2015	Uganda	Training/education Asset transfer	Regression discontinuity	Worry about insufficient food	- 0.11 (- 0.18, - 0.04)	3368

*Indicates estimates that were used in meta-analysis. Only one outcome per study per analysis was included to maintain independence of observations. Outcomes were selected based on comparability with other studies

†Some studies appear multiple times because they report data related to multiple outcomes

Table 8 PICOS summary of criteria for the inclusion and exclusion of studies

Criteria	Inclusion criteria	Exclusion criteria
Population	<p>Program participants that were located in a L&MIC in the first year of implementation³</p> <p>Impact evaluations with at least one effect size for an L&MIC country population</p> <p>Studies focused on the prevention of clinical conditions</p>	<p>Studies focused on niche populations, such as athletes or the military</p> <p>Efficacy studies, unless they were completed in a sufficiently real-world setting</p> <p>Studies targeting participants with a clinical condition</p> <p>Studies focused on high-income country migrant populations in L&MICs and vice versa</p>
Intervention	<p>Interventions that directly intervene on an aspect of the food system within its three primary domains: the food supply chain, the food environment and consumer behavior</p> <p>Studies evaluating multiple food systems interventions</p>	<p>Interventions not in the food system or interventions targeting drivers of the food system without an explicit food system focus</p> <p>Unconditional cash transfer programs</p> <p>Interventions focused on the financing of a food systems intervention</p>
Comparisons	<p>Appropriate comparisons included: business as usual, an alternative treatment, no treatment or an early-versus-late comparison (where those that took part in earlier years are compared to those that took part in later years)</p>	<p>Studies that did not justify and make use of an appropriate comparison group</p>

Source: 3ie 2020

The cutoff at the year 2000 was made arbitrarily to make the volume of search results more manageable

Gray literature sites searched

To identify relevant gray literature, we searched the following databases (some of which contain a mixture of published and gray literature):

- Google Scholar
- EconLit
- ENN-Network
- IDEAS/RePEc
- Innovative Methods and Metrics for Agriculture and Nutrition Actions grantee database
- WHO Global Index Medicus
- Gray Literature Report
- Social Science Research Network (SSRN)
- Eldis
- Epistemonikos
- 3ie Development Evidence Portal
- Registry of International Development Impact Evaluations (RIDIE)
- Oxfam Policy & Practice

Below is a list of organizational websites we manually searched for additional related studies.

- [AgEcon Search \(University of Minnesota\)](#)
- [Innovations for Poverty Action](#)

- [Abdul Latif Jameel Poverty Action Lab](#)
- [Global Development Network](#)
- [World Bank Development Impact Evaluation \(DIME\) and Impact Evaluation Policy Papers](#)
- [Inter-American Development Bank](#)
- [Center for Global Development](#)
- [Center for Effective Global Action \(CEGA\)](#)
- [Department for International Development Research for Development \(R4D\)](#)
- [USAID](#)
- [International Food Policy Research Institute](#)
- [CIGAR](#)
- [Food and Agriculture Organization of the United Nations \(FAO\)](#)
- [High Level Panel of Experts on Food Security and Nutrition](#)
- [World Food Programme](#)
- [Action Against Hunger](#)
- [UNICEF](#)
- [United Nations Evaluation Group](#)
- [Asian Development Bank](#)
- [World Agroforestry Centre \(ICRAF\)](#)
- [International Livestock Research Institute \(ILRI\)](#)
- [Nutrition International](#)

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40066-023-00405-9>.

Additional file 1. It contains additional information about the included studies in terms of Intervention type, Detailed intervention, Evaluation method, Hypotheses mechanisms of action, Impacts Barriers and facilitator to impact, implementation, and evaluation, Equity consideration, Sources of bias, Risk of bias, Effectiveness, and Conclusions.

Acknowledgements

Not applicable.

Author contributions

MB contributed to extract the effect sizes and analyze them through the meta-analysis. She was a major contributor in writing the manuscript. MK searched the additional qualitative studies, extracted the data and analyzed them. They were a major contributor in writing the manuscript. CL contributed to extract the effect sizes and analyze them through the meta-analysis. She was a major contributor in reviewing and writing the manuscript, as well as in ensuring its overall quality. SS ensured the meta-analysis were conducted following the highest standard and corrected any mistakes. She was a major contributor in reviewing the manuscript, as well as in ensuring its overall quality. All authors read and approved the final manuscript.

Funding

This project has been commissioned and funded by Germany's Federal Ministry for Economic Cooperation and Development (BMZ) through Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) through its "Knowledge for Nutrition" program.

Availability of data and materials

All the data used to support finding in this study derive from the included study, please see the full list in the Reference section.

Declarations

Ethics approval and consent to participate

This is not applicable to this manuscript as we did not use primary data.

Consent for publication

This is not applicable as we did not include any individual person's data in any form.

Competing interests

The authors declare that they have no competing interests.

Author details

¹The International Initiative for Impact Evaluation, Synthesis, and Reviews Office (SRO), London, UK. ²The International Initiative for Impact Evaluation, Washington, DC, USA.

Received: 20 June 2022 Accepted: 10 January 2023

References

- Moore N, Lane C, Storhaug I, Franich A, Rolker H, et al. The effects of food systems interventions on food security and nutrition outcomes in low- and middle-income countries. *International Initiative for Impact Evaluation* (3ie); 2021.
- Njuki J, Eissler S, Malapit H, Meinzen-Dick R, Bryan E, Quisumbing A. A review of evidence on gender equality, women's empowerment, and food systems: Food Systems Summit Brief Prepared by Research Partners of the Scientific Group for the Food Systems Summit. 2021. https://bonndoc.ulb.uni-bonn.de/xmlui/bitstream/handle/20.500.11811/9132/fss_briefs_review_evidence_gender_equality.pdf?sequence=3&isAllowed=y
- WHO. Understanding the women's empowerment pathway. Brief #4. Improving nutrition through agriculture technical brief series. Arlington: USAID/Strengthening Partnerships, Results, and Innovations in Nutrition Globally (SPRING) Project. 2014.
- United Nations Food Systems Summit. Chapter 2 key inputs from summit workstreams action tracks. 2021. <https://foodsystems.community/food-systems-summit-compendium/action-tracks/>. Accessed 27 Jan 2022.
- Cole SM, Kantor P, Sarapura S, Rajaratnam S. Gender-transformative approaches to address inequalities in food, nutrition and economic outcomes in aquatic agricultural systems. 2015.
- Wong, F, Vos A, Pyburn R, Newton J. Implementing gender transformative approaches in agriculture. A Discussion Paper for the European Commission. 2019.
- Cheung J, Gursel D, Kirchner MJ, Scheyer V. Practicing feminist foreign policy in the everyday: a toolkit. Germany; 2021.
- Thompson L. Defining feminist foreign policy. Washington: International Center for Research on Women; 2019. p. 1–7.
- Campbell Collaboration. (2017). Campbell systematic reviews: Policies and guidelines.
- Barends, E., Rousseau, D. M. & Briner, R. B. CEBMa Guideline for Rapid Evidence Assessments in Management and Organizations. Amsterdam. 2017. <https://www.cebma.org/wp-content/uploads/CEBMA-REA-Guideline.pdf>
- Snijlsteit B. Systematic reviews: from 'bare bones' reviews to policy relevance. *J Dev Effect*. 2012;4(3):388–408. <https://doi.org/10.1080/19439342.2012.709875>.
- CASP (2018). Qualitative Checklist. [online] Available at: https://casp-uk.net/images/checklist/documents/CASP-Qualitative-Studies-Checklist/CASP-Qualitative-Checklist-2018_fillable_form.pdf Accessed: 1st March 2022.
- Korth M, Stewart R, Langer L, Madinga N, Rebelo Da Silva N, Zaranyika H, van Rooyen C, de Wet T. What are the impacts of urban agriculture programs on food security in low and middle-income countries: a systematic review. *Environ Evid*. 2014;3(1):1–10.
- Stewart R, Langer L, Da Silva NR, Muchiri E, Zaranyika H, Erasmus Y, Randall N, Rafferty S, Korth M, Madinga N, de Wet T. The effects of training, innovation and new technology on African smallholder farmers' economic outcomes and food security: a systematic review. *Campbell Syst Rev*. 2015;11(1):1–224.
- Goudet SM, Bogin BA, Madise NJ, Griffiths PL. Nutritional interventions for preventing stunting in children (birth to 59 months) living in urban slums in low-and middle-income countries (LMIC). *Cochrane Database Syst Rev*. 2019. <https://doi.org/10.1002/14651858.CD011695.pub2>.
- Sguassero Y, de Onis M, Bonotti AM, Carroli G. Community-based supplementary feeding for promoting the growth of children under five years of age in low and middle income countries. *Cochrane Database Syst Rev*. 2012;2012(6):CD005039. <https://doi.org/10.1002/14651858.CD005039.pub3>.
- Visser ME, Schoonees A, Ezekiel CN, Randall NP, Naude CE. Agricultural and nutritional education interventions for reducing aflatoxin exposure to improve infant and child growth in low-and middle-income countries. *Cochrane Database Syst Rev*. 2020. <https://doi.org/10.1002/14651858.CD013376.pub2>.
- Shah D, Sachdev HS, Gera T, De-Regil LM, Peña-Rosas JP. Fortification of staple foods with zinc for improving zinc status and other health outcomes in the general population. *Cochrane Database Syst Rev*. 2016. <https://doi.org/10.1002/14651858.CD010697.pub2>.
- Suchdev PS, Peña-Rosas JP, De-Regil LM. Multiple micronutrient powders for home (point-of-use) fortification of foods in pregnant women. *Cochrane Database Syst Rev*. 2015. <https://doi.org/10.1002/14651858.CD011158.pub2>.
- Suchdev PS, Jefferds MED, Ota E, da Silva LK, De-Regil LM. Home fortification of foods with multiple micronutrient powders for health and nutrition in children under two years of age. *Cochrane Database Syst Rev*. 2020;2(2):CD008959. <https://doi.org/10.1002/14651858.CD008959.pub3>.
- Gera T, Sachdev HS, Boy E. Effect of iron-fortified foods on hematologic and biological outcomes: systematic review of randomized controlled trials. *Am J Clin Nutr*. 2012;96(2):309–24. <https://doi.org/10.3945/ajcn.111.031500>.

22. Ota E, Hori H, Mori R, Tobe-Gai R, Farrar D. Antenatal dietary education and supplementation to increase energy and protein intake. *Cochrane Database Syst Rev*. 2015. <https://doi.org/10.1002/14651858.CD000032.pub3>.
23. Borenstein M, Hedges LV, Higgins JP, Rothstein HR. *Introduction to meta-analysis*. New York: John Wiley & Sons; 2021.
24. Ahmed A, Hoddinott J, Roy S. Food transfers, cash transfers, behavior change communication and child nutrition. *Intl Food Policy Res Inst*. 2019;1868.
25. Bandiera O, Burgess R, Das N, Gulesci S, Rasul I, Sulaiman M. Labor markets and poverty in village economies. *Q J Econ*. 2017;132(2):811–70. <https://doi.org/10.1093/qje/qjx003>.
26. Blakstad MM, Mosha D, Bellows AL, Canavan CR, Chen JT, Mlalama K, et al. Home gardening improves dietary diversity, a cluster-randomized controlled trial among Tanzanian women. *Matern Child Nutr*. 2021;17(2):e13096. <https://doi.org/10.1111/mcn.13096>.
27. Bonuedi I, Kornher L, Gerber N. Making cash crop value chains nutrition-sensitive: evidence from a quasi-experiment in rural Sierra Leone. *SSRN Electron J*. 2020. <https://doi.org/10.2139/ssrn.3603918>.
28. Deininger K, Liu Y. Economic and social impacts of self-help groups in India. World Bank Group; 2009.
29. Emran MS, Robano V, Smith SC. Assessing the frontiers of ultra-poverty reduction: Evidence from CFPR/TUP, an innovative program in Bangladesh. TUP, An Innovative Program in Bangladesh. 2009.
30. Haque MA, Choudhury N, Ahmed ST, Farzana FD, Ali M, Naz F, et al. The large-scale community-based programme 'Suchana' improved maternal healthcare practices in north-eastern Bangladesh: findings from a cluster randomized pre-post study. *Matern Child Nutr*. 2022. <https://doi.org/10.1111/mcn.13258>.
31. Heckert J, Olney DK, Ruel MT. 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. *Soc Sci Med*. 2019;233:93–102. <https://doi.org/10.1016/j.socscimed.2019.05.016>.
32. Marquis GS, Colecraft EK, Sakyi-Dawson O, Lartey A, Ahunu BK, Birks KA, et al. An integrated microcredit, entrepreneurial training, and nutrition education intervention is associated with better growth among preschool-aged children in rural Ghana. *J Nutr*. 2015;145(2):335–43. <https://doi.org/10.3945/jn.114.194498>.
33. Pan Y, Smith SC, Sulaiman M. Agricultural extension and technology adoption for food security: evidence from Uganda. *Am J Agric Econ*. 2018;100(4):1012–31. <https://doi.org/10.1093/ajae/aay012>.
34. Polanin, J.R. and Snilstveit, B., 2016. Converting between effect sizes. *Campbell Systematic Reviews*, 12(1), pp.1-13.
35. Sánchez-Meca, J., Marín-Martínez, F., & Chacón-Moscoso, S. (2003). Effect-size indices for dichotomized outcomes in meta-analysis. *Psychological methods*, 8(4), 448.
37. Kabeer N, Datta S. Randomized control trials and qualitative impacts: what do they tell us about the immediate and long-term assessments of productive safety nets for women in extreme poverty in West Bengal? (No. 19–199). Working Paper Series. 2020
38. Roy S, Hidrobo M, Hoddinott J, Ahmed A. Transfers, behavior change communication, and intimate partner violence: postprogram evidence from rural Bangladesh. *Rev Econ Stat*. 2019;101(5):865–77. https://doi.org/10.1162/rest_a_00791.
39. Roy S, Ara J, Das N, Quisumbing AR. Flypaper effects" in transfers targeted to women: evidence from BRAC's "Targeting the Ultra Poor" program in Bangladesh. *J Dev Econ*. 2015;117:1–19.
40. Kieran C, Gray B, Gash M. Understanding gender norms in rural Burkina Faso: a qualitative assessment. 2018.
41. Hagan LL, Aryeetey R, Colecraft EK, Marquis GS, Nti AC, University of Ghana, et al. Microfinance with education in rural Ghana: Men's perception of household level impact. *Afr J Food Agric Nutr Dev*. 2012;12(49):5776–88. <https://doi.org/10.18697/ajfand.49.enam7>.
42. Dolgoplova I, Li B, Pirhonen H, Roosen J. The effect of attribute framing on consumers' attitudes and intentions toward food: a Meta-analysis. *Bio-based Appl Econ J*. 2021;10:253–64.
43. Data4Diets' (2021), International Dietary Data Expansion Project. <https://index.nutrition.tufts.edu/data4diets>. Accessed 4 Sep 2022.
44. DQQ Tools & Data' (2021), Global Diet Quality Project. <https://www.globaldietquality.org/dqq>. Accessed 11 Apr 2022.
45. Choudhury N, Raihan MJ, Ahmed ST, Islam KE, Self V, Rahman S, Schofield L, Hall A, Ahmed T. The evaluation of Suchana, a large-scale development program to prevent chronic undernutrition in north-eastern Bangladesh. *BMC Public Health*. 2020;20:1–9.
46. Snilstveit, B, Stevenson, J, Langer, L, da Silva, N, Rabat, Z, Nduku, P, Polanin, J, Shemilt, I, Evers, J, Ferraro, PJ, Incentives for climate mitigation in the land use sector – the effects of payment for environmental services (PES) on environmental and socio-economic outcomes in low- and middle-income countries: a mixed-method systematic review 3ie Systematic Review 44. London: International Initiative for Impact Evaluation (3ie). 2019. <https://doi.org/10.23846/SR00044>
47. Critical Appraisal Skills Programme. (2006) 10 questions to help you make sense of qualitative research. Public Health Resource Unit. Retrieved from <http://www.biomedcentral.com/content/supplementary/2046-4053-3-139-S8.pdf>
48. Das N, Yasmin R, Ara J, Kamruzzaman M, Davis P, Behrman J, et al. How do intrahousehold dynamics change when assets are transferred to women? Evidence from BRACs challenging the frontiers of poverty reduction targeting the ultra-poor program in Bangladesh. *SSRN Electron J*. 2013. <https://doi.org/10.2139/ssrn.2405712>.
49. Huda K, Kaur S. 'It was as if we were drowning': shocks, stresses and safety nets in India. *Gend Dev*. 2011;19(2):213–27. <https://doi.org/10.1080/13552074.2011.592632>.
50. Olney DK, Dillon A, Ruel MT, Nielsen J. Lessons learned from the evaluation of Helen Keller International's enhanced homestead food production program. *Achieving a nutrition revolution for Africa: The road to healthier diets and optimal nutrition*. 2016;67–81.
51. van den Bold M, Dillon A, Olney D, Ouedraogo M, Pedehombga A, Quisumbing A. Can integrated agriculture-nutrition programmes change gender norms on land and asset ownership? Evidence from Burkina Faso. *J Dev Stud*. 2015;51(9):1155–74. <https://doi.org/10.1080/00220388.2015.1036036>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.