

Rapid Evidence Assessment on Women's Empowerment Interventions within the Food System: a meta-analysis

Miriam Berretta (✉ mberretta@3ieimpact.org)

International Initiative for Impact Evaluation (3ie)

Meital Kupfer

3ie

Charlotte Lane

3ie

Shannon Shisler

3ie

Research Article

Keywords: women's empowerment, review, food system, meta-analysis

Posted Date: June 13th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1754233/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Rapid Evidence Assessment on Women's Empowerment Interventions within the Food System: a meta-analysis¹

Miriam Berretta, MSc, The International Initiative for Impact Evaluation, Synthesis and reviews office (SRO), London office

Meital Kupfer, MA, The International Initiative for Impact Evaluation, Synthesis and reviews office (SRO), London office

Shannon Shisler, PhD, The International Initiative for Impact Evaluation, Synthesis and reviews office (SRO), London office

Charlotte Lane, PhD, The International Initiative for Impact Evaluation, Washington DC office

¹ Women's empowerment interventions assessment

Contents

Abstract	5
1 Background	7
1.1 How are women's empowerment interventions expected to improve nutrition?	9
2 Objectives and research questions	10
3 Methodology	11
3.2 Search strategy.....	15
3.3 Data extraction.....	16
4 Results	17
4.1 Characteristics of the included studies	17
4.2 What are the effects of women's empowerment interventions on nutritional and food security outcomes?.....	24
4.4 Implications for implementation and sustainability.....	33
5 Discussion	39
5.1 Summary and application of findings.....	39
5.2 Gaps and limitations in available studies.....	40
5.3 Implications for policy and practice	41
5.4 Implications for researchers.....	42
5.4 Strengths, limitations & future directions.....	43
6 Declarations	44
7 References	45
8 Appendices	49
Appendix 1: Rapid Evidence Assessment on Women's Empowerment in Food Systems Interventions – Protocol	49
1 Background	50
2 Research questions.....	52
3 Methodology.....	52
4. Protocol references.....	62
Appendix 2: Data extraction tool	66
Quantitative data extraction tool	67
Appendix 3: Critical appraisal tools	74

3.1 Appraisal of Risk of Bias for Impact Evaluations using RCT Designs	74
3.2 Appraisal of Risk of Bias for Impact Evaluations using Quasi-experimental designs	99
3.3 Qualitative analysis tool	128
Appendix 4: Additional meta-analysis results	141
Detailed results for food security	141
Detailed results for food affordability/availability.....	141
Detailed results for diet quality and adequacy	141
Detailed results for anthropometric measures	142
Detailed results for well-being outcomes	142
Appendix 5: Detailed risk of bias.....	142

List of abbreviations

EGM	Evidence Gap Map
GTA	Gender transformative approaches
QED	Quasi-experimental design
RCT	Randomized control trial

List of Figures

Figure 1: THEORY OF CHANGE, FROM NJUKI ET AL. (2021)	10
Figure 2: RISK OF BIAS OF THE INCLUDED RANDOMISED CONTROL TRIALS	23
Figure 3: FOREST PLOT SHOWING THE EFFECT OF EMPOWERMENT INTERVENTIONS ON FOOD SECURITY OUTCOMES	27
Figure 4: FOREST PLOT SHOWING THE EFFECT OF EMPOWERMENT INTERVENTIONS ON FOOD AFFORDABILITY/AVAILABILITY OUTCOMES.....	29
Figure 5: FOREST PLOT SHOWING THE EFFECT OF EMPOWERMENT INTERVENTIONS ON DIET QUALITY AND ADEQUACY	30
Figure 6: FOREST PLOT SHOWING THE EFFECT OF EMPOWERMENT INTERVENTIONS ON WEIGHT RELATIVE TO HEIGHT	31
Figure 7: FOREST PLOT SHOWING THE EFFECT OF EMPOWERMENT INTERVENTIONS ON WELLBEING.....	33

List of Tables

Table 1: PICOS.....	12
Table 2: Included outcomes and indicator extracted for evidence synthesis	13
Table 3: Summary of included studies	19
Table 4: Meta-analytical results	25
Table 5: Risk of bias in experimental studies.....	143
Table 6: Risk of bias in quasi-experimental studies.....	144
Table 7: Effect estimates from included studies in REAT	145

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.

The objective was to rapidly synthesise 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 grey literature sources performed for the regular maintenance of the living Food System and Nutrition Evidence Gap Map. Impact evaluations and systematic reviews of impact evaluations that considered the women's empowerment interventions in food systems and food security and nutrition outcomes were eligible for inclusion. We conducted an additional search for supplementary, contextual, 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, and well-being effects were smaller. Due to the limited number of included

studies for anthropometric and well-being outcomes, these effects are inconclusive. 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, in particular combining a livelihoods component with behavioural change communication.

Researchers should consider issues with inconsistent data and reporting, particularly relating to seasonal changes, social norms and time taken between rounds of data collection. Future studies on gender-transformative approaches should carefully consider contextual norms and avoiding stereotyping women into pre-decided roles which may perpetuate social norms.

Keywords: women's empowerment, review, food system, meta-analysis

1 Background

Most research on women within food systems focuses on their roles as caregivers and cooks (Moore et al. 2021). However, women are key actors within food systems, serving as producers, wage workers, traders, processors and consumers. Nonetheless, women face negative, differential access to affordable, nutritious foods relative to men. Often living in more vulnerable conditions than men due to societal norms, 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 (WHO 2014). As a result, women's empowerment interventions represent a key opportunity to improve nutrition-related outcomes. 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 (Njuki et al. 2021).

We considered interventions that attempt to empower women and/or girls to access, participate in and take control of components of the food system, for example by improving decision-making on household expenditures. 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](#) (Section 3.1.2).

Gender transformative approaches (GTA) acknowledge the equal role that all genders have in women's empowerment and thus target men as agents of change to transform structural barriers and social norms (Cole et al. 2015). While many women's empowerment interventions include

GTA approaches, women's empowerment and GTA differ mainly in the following aspects (adapted from Wong et al., 2019):

- ❖ 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.

Most of the interventions included in the analysis focus on the economic and social empowerment of women with the goal of providing them the means and ability to affect dietary decisions; (Cheung et al. 2021; Thompson and Clement 2019). The interventions we looked at primarily relate to behaviour 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.

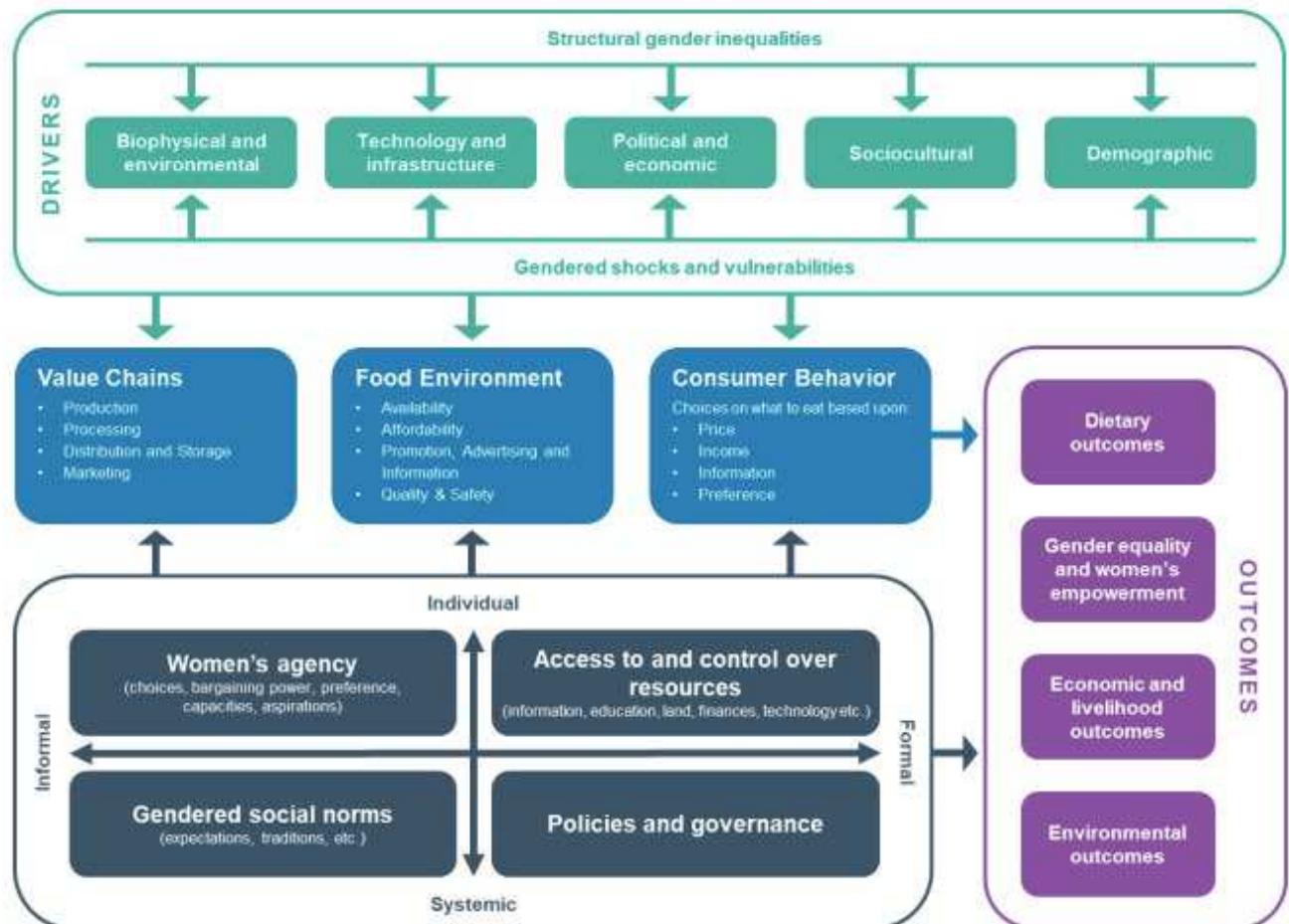
This rapid evidence assessment is expected to support policy makers, experts, and stakeholders in making evidence-informed decisions regarding food systems and nutrition interventions by synthesising available evidence on women's empowerment interventions in food systems. Women's empowerment has been highlighted as a critical, crosscutting theme for food systems transformation (UN 2021). However, many food systems and nutrition interventions are criticised as disempowering because they can entrench stereotypes by targeting women and girls explicitly in these roles of caregivers or cooks. Stakeholders can use this work to better 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.

1.1 How are women's empowerment interventions expected to improve nutrition?

Our theory of change is based on the pathways developed by Njuki and colleagues (2021). 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. To transform food systems, changes in gender equality are needed at individual and systemic as well as formal and informal levels.

Improvements in women's empowerment are expected to facilitate women's interactions with the food system and improve nutrition outcomes of women and their communities directly and indirectly. By allowing 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. This is because they allow women to make the choices that are most likely to benefit them while addressing the broader social and cultural context.

Figure 1: THEORY OF CHANGE, FROM NJUKI ET AL. (2021)



Source: Adapted from de Brauw et al. (2019).

2 Objectives and research questions

The objective of this work was to rapidly synthesise 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. We also sought to understand factors affecting implementation and sustainability, including equity. We specified the following research questions *a priori*:

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?

3 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. However, due to time and resource limitations several steps were shortened (Barends et al. 2017). The main shortcut was that we did not conduct a new search for impact evaluations, but rather relied on an existing evidence gap map (EGM) of food systems and nutrition interventions in low- and middle-income countries (Moore et al., 2021). The search for the EGM was extensive and systematic, covering 12 academic databases and 13 grey literature sources. The EGM is a 'living' EGM; while the original search was complete in May 2020, the search is continuously updated with new studies added to the EGM through January 2022. To date, we have screened over 160,000 articles for inclusion in the EGM. It currently contains 2,647 studies.

For the REA, we select a sub-sample of studies from the EGM – those focusing on women's empowerment. While we are confident the EGM search is both systematic and extensive, the broad scope of the EGM means that it may not have the sensitivity of a more targeted search which would be expected in a systematic review. For the EGM, single screening 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. The protocol for the REA was developed *a priori* and is provided in Appendix 3. Below we summarise the methodology.

3.1 Criteria for including and excluding studies in the review

We adopted the same inclusion criteria as Moore et al. (2021), as described in Table 1 and summarised below, with the intervention criteria revised as per the scope of the REA.

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	No comparator
	An alternate intervention	
Outcome(s)	Food security	All else
	Food affordability and availability	
	Diet quality and adequacy	
	Anthropometrics	
	Iron, zinc, vitamin A, and iodine status	
Study designs	Measures of well-being	Efficacy trials Before-after with no control group Cross-sectional studies, etc.
	Experimental, quasi-experimental, systematic reviews and cost evidence	

3.1.1 Types of study participants

We included impact evaluations in low- and middle- income countries as defined by the [World Bank Country and Lending Groups classification](#) in the first year of the intervention. Studies in which a country held high-income status for only one year before reverting to low- or middle-income status were also included. However, this did not apply to any of the studies included. If the year of intervention was not available, we used the publication year.

3.1.2 Types of interventions

We 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.

3.1.3 Types of outcome measures

We considered studies measuring 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 (Table 2). These outcomes could be measured using a variety of indicators. We pre-specified preferred and alternate outcome. Composite measures were always preferred over disaggregated ones.

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
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 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.

3.1.4 Types of comparators

Alternate intervention or business as usual comparators, including pipeline and waitlist controls, were included as valid comparators. Studies with no comparator were excluded.

3.1.5 Types of study design

We included studies adopting the following study designs:

- Randomised controlled trials
- 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
- High or medium quality systematic reviews that include a quantitative or narrative synthesis

Ex-post cost-effectiveness analyses and qualitative studies associated with an included impact evaluations were included.

3.1.6 Date, language, and form of publication

The evidence gap map which forms the basis of this work was restricted to English language studies published after 2000. Additional information on inclusion/exclusion criteria can be found in the protocol in Appendix 1.

3.2 Search strategy

We did not perform any new searches for impact evaluations for this rapid evidence assessment. However, 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 programme or intervention, if available, as well as the country the intervention took place in. Eligible qualitative study designs were (Snilstveit 2012):

- *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 analysing 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 of the following: the research question, procedures for collecting data, procedures for analysing 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 operationalised. They might also be used to collect organisational 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 (Noyes et al. 2020).

3.3 Data extraction

Data extraction templates were modified from 3ie's standard coding protocol for systematic reviews (Appendix 2). Two team members extracted bibliographic, geographic information, methods, and substantive data. Substantive data were related to interventions, 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 were extracted by a single reviewer.

3.3.3 Critical appraisal of quantitative and qualitative studies

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. We assessed the quality of nine included qualitative studies using a mixed-methods appraisal tool developed by Langer and colleagues (2017) and applied in Snistveit and colleagues (2019) (Appendix 3.3).

3.4 Data presentation

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 are summarised in tables. Results from meta-analyses and associated forest plots are presented in the section on the findings. Qualitative information is summarised in a section on implications for implementation and sustainability.

3.5 Analytical approach

We conducted four meta-analyses to provide summary effect estimates on the four outcomes for which we had sufficient data. We chose the appropriate formulae for effect size calculations in reference to, and dependent upon, the data provided in included studies. We conducted random effects meta-analyses when we identified two or more studies that we assessed to be sufficiently similar. We assessed heterogeneity by calculating the Q statistic, I^2 , and r^2 to provide an estimate of the amount of variability in the distribution of the true effect sizes (Borenstein et al. 2009). We were unable to explore heterogeneity using moderator analyses due to the small number of included studies.

Qualitative data were synthesised 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. The meta-analysis conducted with the quantitative data has thus been complemented by a thematic synthesis utilizing the extracted qualitative data.

4 Results

4.1 Characteristics of the included studies

We included ten studies retrieved through the last systematic search done for the Food Systems and Nutrition Evidence Gap Map, conducted in January 2022 (Table 3). One low quality systematic review was identified and excluded from analysis. Four of the ten 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 programme, an agricultural training programme, and two fully independent evaluations of Targeting-Ultra-Poor programme (TUP) with a time gap of eight years

and somewhat different intervention designs. More information on study characteristics can be found in Supplemental Table 1 (online).

Table 3: Summary of included studies

Country	Study design	Intervention	Authors' interpretation of effects
Bangladesh	Randomised controlled trial	The Transfer Modality Research Initiative (TMRI) provided: Cash or food transfers, with or without nutrition behaviour 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.
Bangladesh	Randomised controlled trial	The Targeting the Ultra-Poor programme 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.
Tanzania	Randomised controlled trial	The Homestead Food Production Programme provided: (a) agricultural training for women and inputs to promote homestead food production (b) nutrition and public health counselling for women to improve diet and health-related behaviours.	Household dietary diversity score increased, but there was no effect on food security.

Sierra Leone	Quasi-experimental design	<p>The Pro-Resilience Action (PROACT) project provided:</p> <p>(a) the LANN was a participatory community-based intervention involving nutrition education, behavioural 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.</p> <p>(b) a cash crop, income-oriented intervention aimed at enhancing economic access to nutritious foods. It included a nutrition programme directed at improving nutrition knowledge and stimulating nutrition-sensitive spending and allocation of other household resources.</p>	LANN, combined with the cash-crop intervention, improved dietary diversity and food consumption among women and children. LANN alone did not have any effect.
India	Quasi-experimental design	<p>The District Poverty Initiative in India:</p> <p>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 mobilisation.</p>	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.
Bangladesh	Quasi-experimental design	<p>The Targeting the Ultra-Poor (TUP) programme provided:</p> <p>(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</p> <p>(b) vitamin A supplements for children under five</p>	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.
Bangladesh	Randomised controlled trial	<p>The Suchana project provided:</p> <p>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.</p>	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.

Burkina Faso	Randomised controlled trial	The Enhanced Homestead Food Production (E-HFP) programme provided: (a) agricultural assets (b) behaviour change communication on agricultural activities, optimal infant and young child feeding, health, hygiene, and care practices.	The E-HFP programme reduced wasting but had null effects on hemoglobin levels among children.
Ghana	Quasi-experimental design	The Enhancing Child Nutrition through Animal Source Food Management (ENAM) programme provided: (a) microcredit loans (b) weekly nutrition, technical, and entrepreneurship training on viable income-generation activities.	The programme 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.
Uganda	Quasi-experimental design	This large-scale agricultural extension programme 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 programme reduced meals skipped, worries about insufficient food, and limited variety of food among smallholder women. It increased per capita food consumption.

Randomised 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. Ultimately, 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 behaviour change communication ($n = 3$; Tables 3, Table 6 in Appendix 6 and Supplemental table 1 online). Behaviour 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 contextualised 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 randomised controlled trials except Blakstad and colleagues (2021) have an overall rating of ‘some concerns’, mainly due to reporting bias, performance bias, and selection bias (Figure 6; Appendix 5). Deininger and Liu (2009) also encountered issues related to deviation from the intended interventions and the unit of analysis did not correspond to the unit of randomisation.

Two quasi-experimental studies were rated as having a low risk of bias (Figure 7; Pan et al. 2015; Marquis et al. 2015), one study as having ‘some concerns’ (Emran et al. 2009), and one as having a high risk of bias (Bonuedi et al. 2020). The major sources of bias were related to reporting bias, spill-over, cross-over and contamination, performance bias, and confounding.

Figure 2: RISK OF BIAS OF THE INCLUDED RANDOMISED CONTROL TRIALS

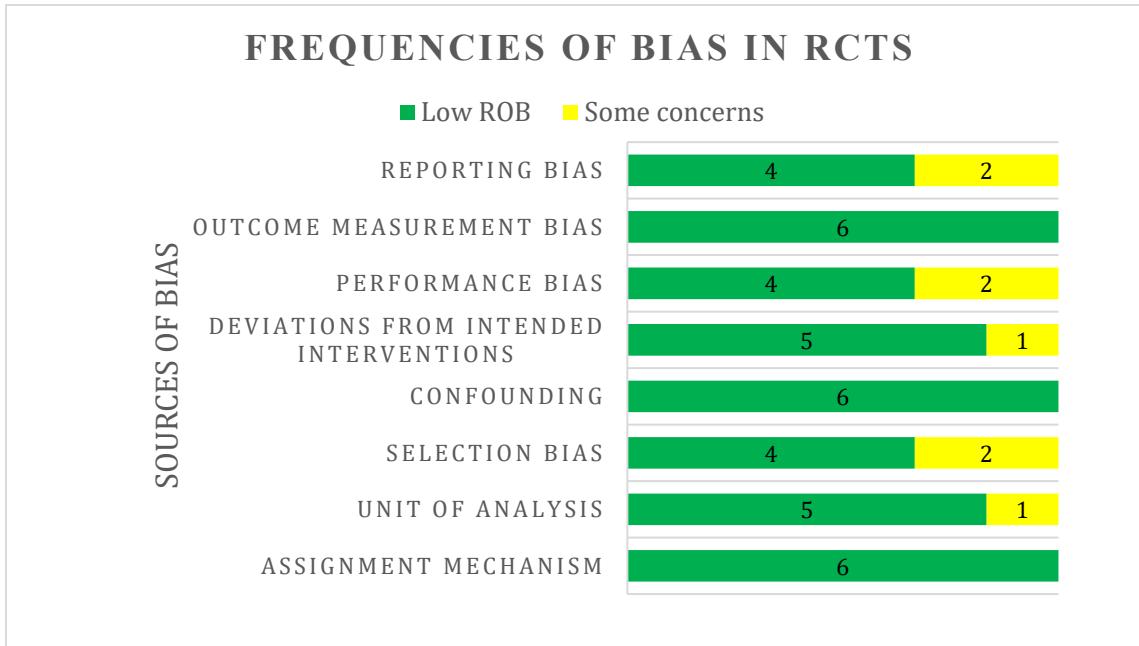
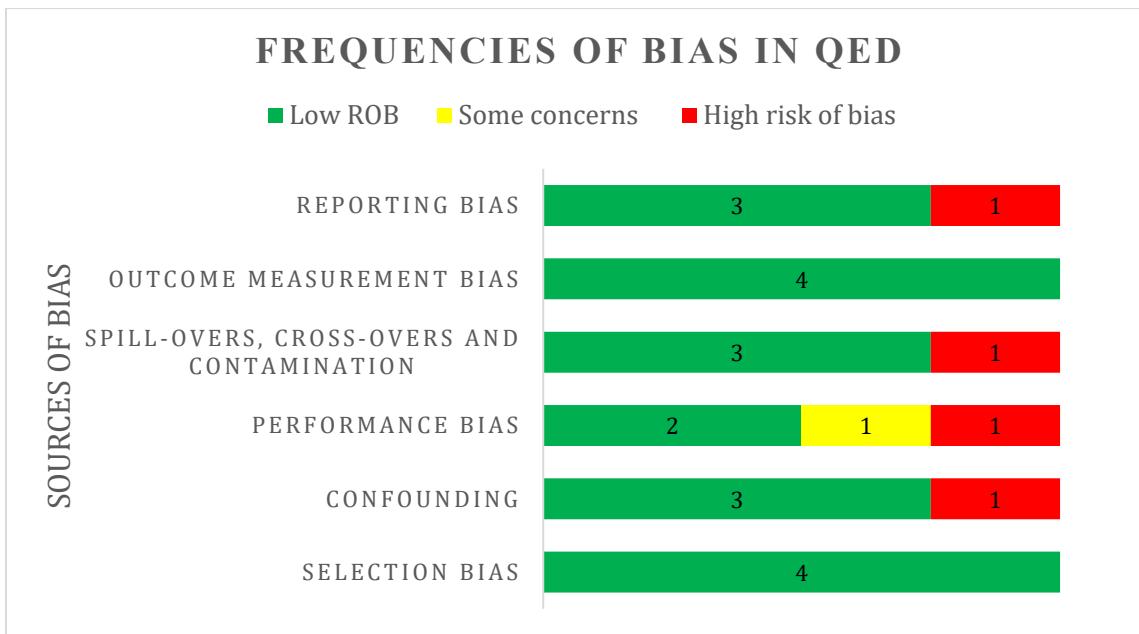


FIGURE 7: RISK OF BIAS OF THE INCLUDED QUASI-EXPERIMENTAL INCLUDED STUDIES



4.2 What are the effects of women's empowerment interventions on nutritional and food security outcomes?

Standardised effects are reported in Table 6 in Appendix 6. The meta-analysis results are reported in Table 4.

Table 4: Meta-analytical results

Outcomes	# 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 (Q and I^2)	Range of effects
Food security	4 (12,545)	0.24** [0.00; 0.47]	9.5% [0; 18.1%]	111.16***, 97.3%	0.07 to 0.67
Food affordability/availability	6 (12,972)	0.23*** [0.09; 0.38]	9.1% [3.6%; 14.8%]	187.27***, 91.99%	-0.11 to 0.49
Diet quality and adequacy	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 relative to length	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	2 (10,100)	0.08* [0.01; 0.15]	3.2% [0.4%; 6%]	2.9*, 65.6%	-0.11 to 0.04

Notes: *** indicates a p-value < 0.01; ** p-value < 0.05; * p-value < 0.1

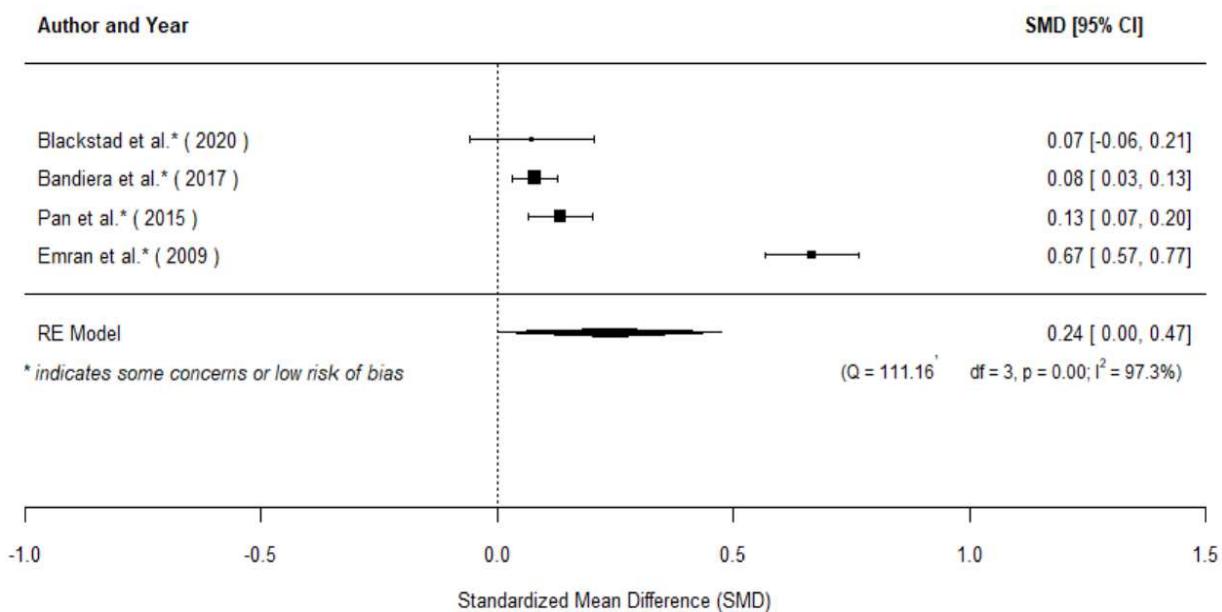
Effect of women's empowerment interventions on food security outcomes is promising

Our analysis of the effects of women's empowerment interventions suggest they improved food security outcomes overall ($\hat{\mu} = 0.24$ [95% CI: 0.00 to 0.47], $p = 0.05$, Figure 2). 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 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 (Blakstad et al. 2020; Bandiera et al. 2017; Emran et al. 2009; Pan et al. 2015). All studies provided training or education, mostly related to agriculture. Three also provided some form of asset transfer (Bandiera et al. 2017; Emran et al. 2009; Pan et al. 2015).

Two studies were assessed as having some concerns related to risk of bias (Bandiera et al. 2017; Emran et al. 2009) and two were assessed as low risk of bias (Blakstad et al. 2020; Pan et al. 2015).

Figure 3: FOREST PLOT SHOWING THE EFFECT OF EMPOWERMENT INTERVENTIONS ON FOOD SECURITY OUTCOMES



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

Our analysis of the effects of women's empowerment interventions suggest they improved the availability and affordability of food ($\hat{\mu} = 0.23$ [95% CI: 0.09 to 0.38] $p < 0.01, 3$). 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) (Ahmed et al. 2019; Bonuedi et al. 2020; Deininger and Liu 2009; Emran et al. 2009; Pan et al. 2015). We included two estimates for Ahmed and colleagues as the results were reported for

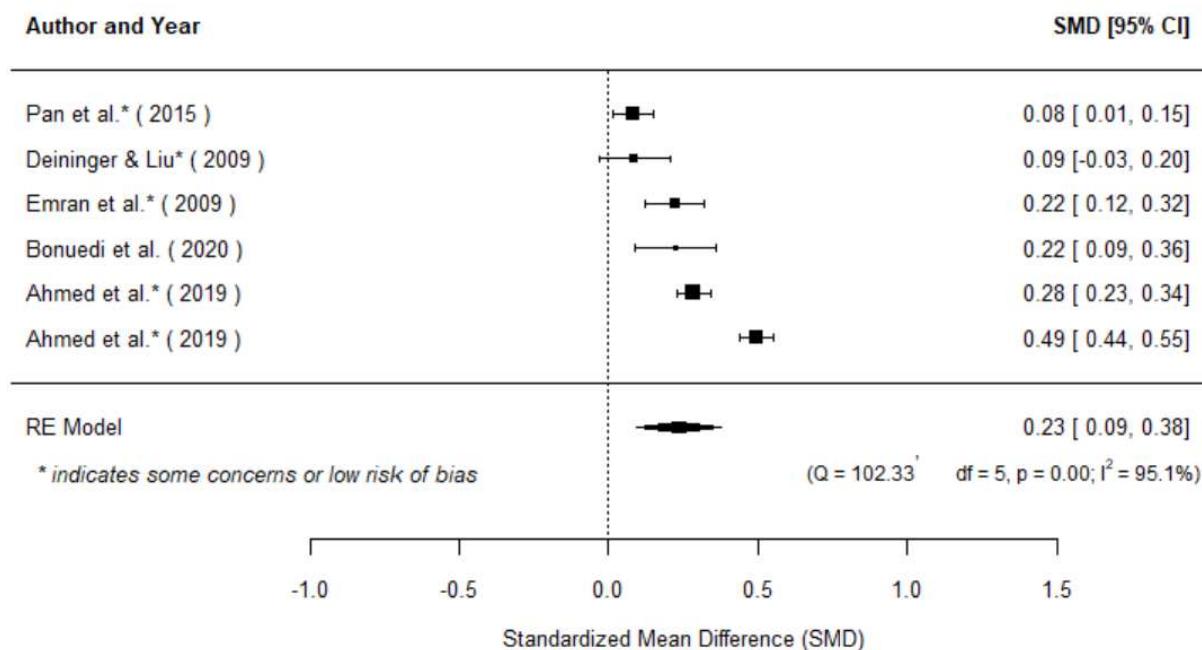
independent samples from the North and South of Bangladesh, without an overall estimate for all the areas.

All studies but Deininger and Liu (2009) included assets transfer such as cash, cash crops (Ahmed et al. 2019, Bonuedi et al. 2020), or livestock, seeds, or vitamin A supplements (Emran et al. 2009; Pan et al. 2015). All studies, except Ahmed and colleagues (2019), included trainings or education on nutrition (Bonuedi et al. 2020), or agriculture (Emran et al. 2009; Pan et al. 2015), or enterprise/accountability (Deininger and Liu 2009). Two studies also included a behaviour change communication component (Ahmed et al. 2019, Bonuedi et al. 2020).

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 and colleagues was assessed as having a high risk of bias, the remaining studies have either some concerns (Ahmed et al. 2019; Deininger and Liu 2009; Emran et al. 2009) or low risk of bias (Pan et al. 2015).

Figure 4: FOREST PLOT SHOWING THE EFFECT OF EMPOWERMENT INTERVENTIONS ON FOOD AFFORDABILITY/AVAILABILITY OUTCOMES



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

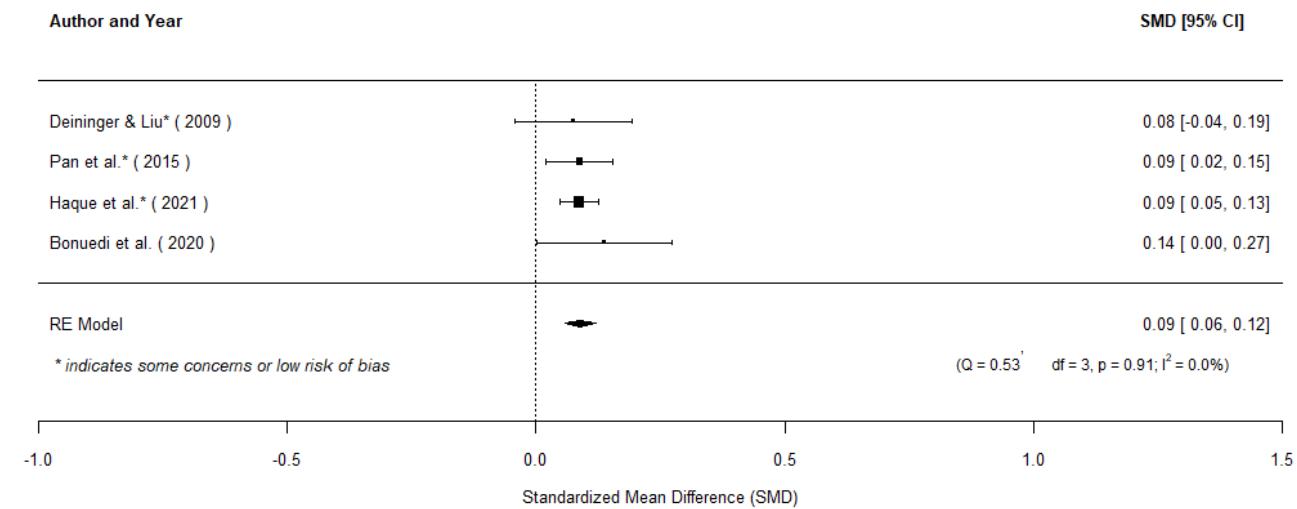
Our analysis of the effects of women's empowerment interventions suggest they improved diet quality and adequacy ($\hat{\mu} = 0.09$ [95% CI: 0.06 to 0.12], $p < 0.01$, Figure 4). Women receiving these interventions had a 53.6% chance of having diet quality and adequacy scores above the mean in the control group. The variations among the range of effects was 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 (Bonuedi et al. 2020; Haque et al. 2021; Deininger and Liu 2009; Pan et al. 2015). All four studies employed training/education interventions focused on

agriculture (Bonuedi et al. 2020; Haque et al. 2021; Pan et al. 2015) or enterprise/accountability (Deininger and Liu 2009). Two studies also transferred assets (Bonuedi et al. 2020; Pan et al. 2015), and one included a behavioural change communication component (Bonuedi et al. 2020).

One study was scored as low risk of bias (Pan et al. 2015), two were scored as having some concerns (Deininger and Liu 2009; Haque et al. 2021), and one was rated as high risk of bias (Bonuedi et al. 2020).

Figure 5: FOREST PLOT SHOWING THE EFFECT OF EMPOWERMENT INTERVENTIONS ON DIET QUALITY AND ADEQUACY



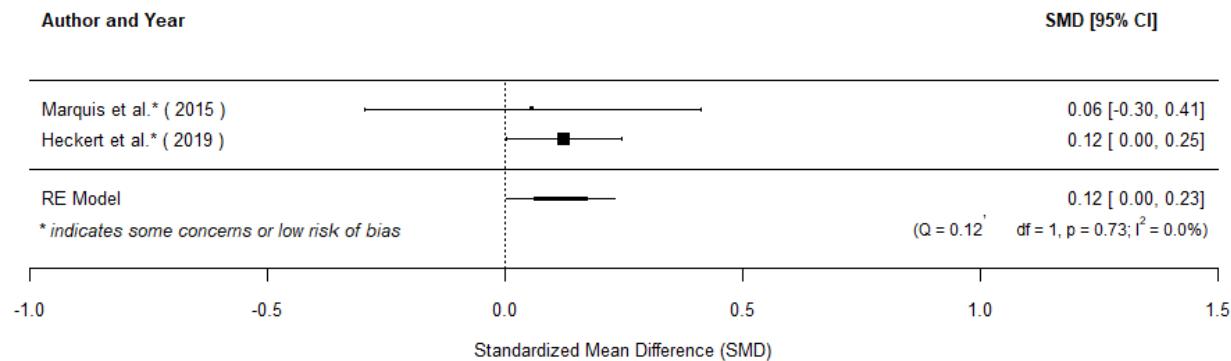
*Effect of women's empowerment interventions on anthropometrics is promising
but there is a lack of evidence*

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: 0.00 to 0.23], Figure 4). 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 (Heckert et al. 2019; Marquis et al. 2015). Both studies transferred agricultural (Heckert et al. 2019) or financial assets (Marquis et al. 2015). The Heckert and colleagues' study also included a behavioural change communication strategy, while Marquis and colleagues included entrepreneurship training.

Marquis and colleagues (2015) also report a decrease in weight-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 and colleagues (2015) had low risk of bias.

Figure 6: FOREST PLOT SHOWING THE EFFECT OF EMPOWERMENT INTERVENTIONS ON WEIGHT RELATIVE TO HEIGHT



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 and colleagues 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 and colleagues evaluated an agricultural education and behaviour change communication strategy, but they found no effect on haemoglobin levels ($g = -0.10$ [95% CI: -0.03 to 0.23]). Both studies were rated as having some concerns about bias.

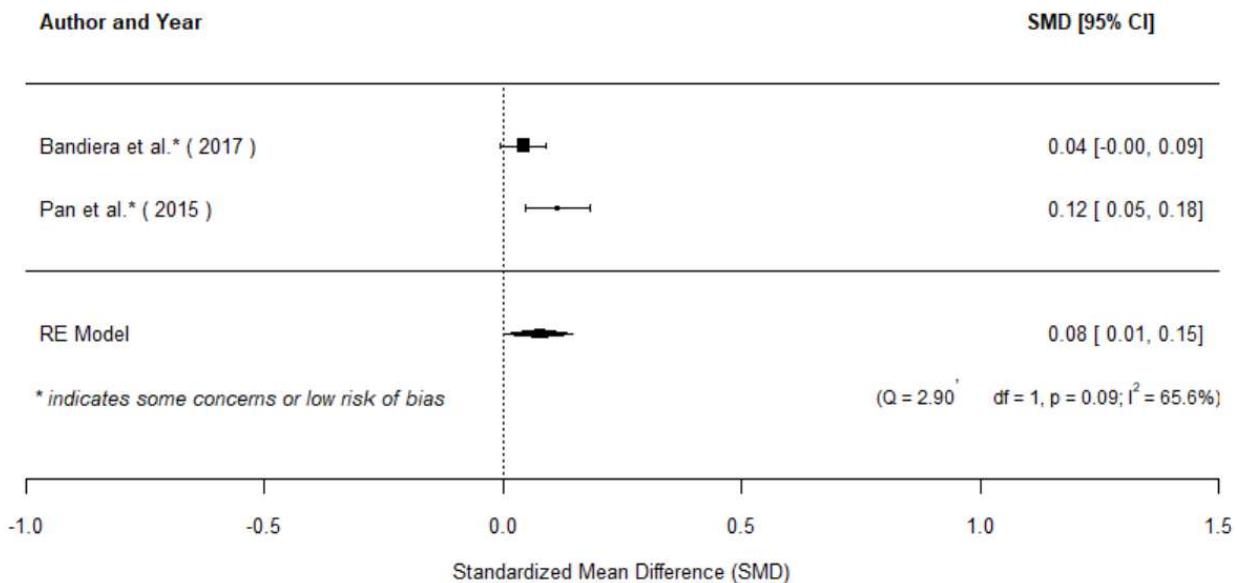
Effect of women's empowerment interventions on mental well-being outcomes is small but also there is a lack of evidence

Our analysis of the effects of women's empowerment interventions shows minimal effect on mental health outcomes ($\hat{\mu} = 0.08$ [95% CI: 0.01 to 0.14], $p = 0.088$, Figure 5). Women receiving these interventions had a 53.2% chance of having wellbeing scores above the mean in the control group.

Bandiera and colleagues (2017) reported a mental health index constructed based on self-reported happiness and mental anxiety, while Pan and colleagues (2015) 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 and colleagues' (2015) paper was assessed as having a low risk of bias, while Bandiera and colleagues' (2017) paper was assessed as having some concerns related to performance bias.

Figure 7: FOREST PLOT SHOWING THE EFFECT OF EMPOWERMENT INTERVENTIONS ON WELLBEING



4.4 Implications for implementation and sustainability

4.4.1 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) programme 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 (Heckert et al. 2019). The accompanying behaviour change communication intervention allowed mothers to better communicate with men to improve familial support and adopt positive nutrition behaviours, such as improved feeding practices. Similarly, the Suchana programme in Bangladesh resulted in improvements in women's empowerment and

maternal healthcare practices by using a gender-transformative approach (Haque et al. 2021). Women became more confident to discuss issues around food and management of household resources with their partners (Bonuedi et al. 2020). Self-help group participation improved social awareness and leadership skills. Women mobilised to protest child marriage and violence against women in their communities (Kabeer and Datta 2020). The Targeting-Ultra-Poor programme (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 (Pan et al. 2015). Similarly, because of the TUP programme, the labour market choices of household members aside from the targeted woman also shifted (Bandiera et al. 2017). However, women themselves did not have increased labour participation. Women in the programme spent most of their time at home and were generally not employed outside of the home (Roy et al. 2019). 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 labour and time intensive, which resulted in high attrition (Blakstad et al. 2020). This additional labour was an increased burden on women and took away from their time to acquire and prepare food for their families (Bonuedi et al. 2020). 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.

4.4.2 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 and colleagues' theory of change. In highly patriarchal societies, such as Sierra Leone, deeply entrenched social and cultural norms marginalise women, restrict their decision-making and exclude them from accessing or controlling household resources (Bonuedi et al. 2020). Single-focus interventions that only targeted nutrition or value-chain inputs without behaviour change communication related to social norms were not able to fully realise potential impacts because entrenched norms were significant barriers to long-lasting change (Pan et al. 2015). 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 (*Ibid*). 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 (Roy et al. 2015). 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 (Roy et al. 2015). 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 (Kabeer and Datta 2020).

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 neighbours who also grew crops at home had higher dietary diversity (Blakstad et al. 2020). Participants who were close to markets were able to access, trade and procure food and related items easier than those who were farther away (Bandiera et al. 2017). 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 (Kabeer and Datta 2020). 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 (Bonuedi et al. 2020). Complementary programme arms can reinforce each other in achieving desired results and reduce implementation costs to achieve the same objectives (Bonuedi et al. 2020). The asset-based component of the PROACT programme in Sierra Leone had little effect. However, when combined with a behaviour 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. Behaviour change communication components of the TMRI programme 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 programme (Roy et al. 2018).

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 (Kieran et al. 2018). 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 (Roy et al. 2015). 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 (Kabeer and Dutta 2020).

Unfortunately, the long time needed to change social norms was a barrier to these interventions achieving impact in the short period over 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. 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. Behaviour change communication can be slow to expand women's empowerment and households' social status and networks (Ahmed et al. 2019). 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 (Pan et al. 2015). 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 (Deininger and Liu 2009). Because the study utilised 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 (Heckert et al. 2019; Blakstad et al. 2020).

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 (Ahmed et al. 2019). 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 (Deininger and Liu 2009). 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 programme targeted microcredit groups, and experienced significant attrition among those who were not benefiting from the loan programme (Marquis et al. 2015). This may not have been observed if the intervention targeted women directly and did not work through the microcredit group.

4.4.4 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 (Deininger and Liu 2009). The TUP programme in Bangladesh also showed that average benefits, including increased household welfare, were 3.21 times larger than costs. Big push programmes, like the TUP, required large investment. However, in this case, it resulted in cost-effective and sustainable change in household welfare, including nutrition (Bandiera et al. 2017).

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 behaviour change component was added to the asset transfer (Bonuedi et al. 2020). Similarly, when added to an asset transfer program, the TMRI women's empowerment behaviour change communication component cost \$50 per beneficiary per year, which is a relatively low cost compared to stand alone behaviour change communication interventions (Ahmed et al. 2019). 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 (Pan et al. 2015). However, when calculating costs, the additional cost of such labour should not be ignored, especially because these costs are often born by the women that these interventions are trying to help (Blakstad et al. 2020).

5 Discussion

5.1 Summary and application of findings

Overall, our analyses suggest women's empowerment interventions can improve nutrition outcomes, contributing to overall better families' health. 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 nutrition outcomes when gender and social norms were considered. However, often, entrenched norms and roles were not acknowledged in these interventions (Kieran et al. 2018). 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.

With ten included studies, the evidence base is small, which can reduce generalizability. 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.

5.2 Gaps and limitations in available studies

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. While we were able to run a four meta-analysis, interpretation of the results is limited due to the low number of studies.

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 important 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.

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. These limitations could result in findings being somewhat unreliable.

5.3 Implications for policy and practice

- Women's empowerment interventions are promising for improving food security, affordability, and diet diversity.
- 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 sensitising men to women's pursuits of work outside of the home (Hagan et al. 2012).
- GTA require cultural and social adaptation to local contexts through strengthened local partnerships and capacities while taking into account 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 policies' and interventions' full potential.

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 (Roy et al. 2015).

- 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 behavioural change communication and advocacy, may be more effective than interventions focusing on just livelihoods or behavioural change.

5.4 Implications for researchers

- The evidence base is relatively limited; therefore, more research is needed across a range of geographies and intervention sub-types. Cost data will also be needed to determine if these impacts are cost-effective.
- 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 ensure that data collection is representative of different seasons and contextual changes, to avoid incomplete or insufficient data (Marquis et al. 2015; Haque et al. 2021; Blakstad et al. 2020).

5.4 Strengths, limitations & future directions

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 Food Systems and Nutrition Evidence Gap Map, of which the last search was run in 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. Finally, due to time and resource constraints the protocol and report has not been subject to external peer review.

6 Declarations

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.

Competing interests

The authors declare that they have no competing interests.

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" programme.

Authors' contribution

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 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.

Acknowledgements

Not applicable.

7 References

Included studies

Ahmed A, Hoddinott J, Roy S. Food transfers, cash transfers, behavior change communication and child nutrition. *Intl Food Policy Res Inst.* 2019;1868.

Bandiera O, Burgess R, Das N, Gulesci S, Rasul I, Sulaiman M. Labor markets and poverty in village economies. *Q J Econ* [Internet]. 2017;132(2):811–70. Available from: <http://dx.doi.org/10.1093/qje/qjx003>

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* [Internet]. 2021;17(2):e13096. Available from: <http://dx.doi.org/10.1111/mcn.13096>

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* [Internet]. 2020; Available from: <http://dx.doi.org/10.2139/ssrn.3603918>

Deininger K, Liu Y. Economic and social impacts of self-help groups in India. World Bank Group; 2009.

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.

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. *Maternal & Child Nutrition*. 2022;18(1).

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* [Internet]. 2019;233:93–102. Available from: <http://dx.doi.org/10.1016/j.socscimed.2019.05.016>

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* [Internet]. 2015;145(2):335–43. Available from: <http://dx.doi.org/10.3945/jn.114.194498>

Pan Y, Smith SC, Sulaiman M. Agricultural extension and technology adoption for food security: Evidence from Uganda. *Am J Agric Econ* [Internet]. 2018;100(4):1012–31. Available from: <http://dx.doi.org/10.1093/ajae/aay012>

Additional Qualitative Studies

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 BRACCs challenging the frontiers of poverty reduction targeting the ultra poor program in Bangladesh. *SSRN Electron J* [Internet]. 2013; Available from: <http://dx.doi.org/10.2139/ssrn.2405712>

Hagan LL, University of Ghana, Aryeetey R, Colecraft EK, Marquis GS, Nti AC, et al. Microfinance with education in rural Ghana: Men's perception of household level impact. *Afr J Food Agric Nutr Dev* [Internet]. 2012;12(49):5776–88. Available from: <http://dx.doi.org/10.18697/ajfand.49.enam7>

Huda K, Kaur S. 'It was as if we were drowning': shocks, stresses and safety nets in India. *Gend Dev* [Internet]. 2011;19(2):213–27. Available from: <http://dx.doi.org/10.1080/13552074.2011.592632>

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;

Kieran C, Gray B, Gash M. Understanding Gender Norms in Rural Burkina Faso: A Qualitative Assessment. 2018.

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.

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. Journal of Development Economics. 2015;117:1–19.

Roy S, Hidrobo M, Hoddinott J, Ahmed A. Transfers, behavior change communication, and intimate partner violence: Postprogram evidence from rural Bangladesh. Rev Econ Stat [Internet]. 2019;101(5):865–77. Available from: http://dx.doi.org/10.1162/rest_a_00791

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 [Internet]. 2015;51(9):1155–74. Available from: <http://dx.doi.org/10.1080/00220388.2015.1036036>

Other References

Barens, E., Rousseau, D. M. & Briner, R. B. CEBMa Guideline for Rapid Evidence Assessments in Management and Organizations. Amsterdam. 2017. Available from: <https://www.cebma.org/wp-content/uploads/CEBMA-REA-Guideline.pdf>

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.

Cheung J, Gursel D, Kirchner MJ, Scheyer V. Practicing Feminist Foreign Policy in the Everyday: A Toolkit. Germany; 2021.

Herforth A, Harris J. Understanding and Applying Primary Pathways and Principles. Brief #1. Improving Nutrition Through Agriculture Technical Brief Series. In: USAID/Strengthening Partnerships, Results, and Innovations in Nutrition Globally (SPRING) Project. Arlington, VA; 2014.

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. URL: https://bonndoc.ulb.uni-bonn.de/xmlui/bitstream/handle/20.500.11811/9132/fss_briefs_review_evidence_gender_equality.pdf?sequence=3&isAllowed=y

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.

Savitri AI, Idris NS, Indawati W, Saldi SRF, Amelia D, Baharuddin M, et al. Breastfeeding Attitude and Volume Optimization (BRAVO) trial: study protocol for a randomized controlled trial. *Trials* [Internet]. 2016;17(1). Available from: <http://dx.doi.org/10.1186/s13063-016-1397-y>

WHO. Understanding the Women's Empowerment Pathway. Brief #4. Improving Nutrition through Agriculture Technical Brief Series. Arlington, VA: USAID/Strengthening Partnerships, Results, and Innovations in Nutrition Globally (SPRING) Project. 2014.

Thompson L. Defining Feminist Foreign Policy. International Center for Research on Women. 2019;1–7. United Nations Food Systems Summit, Chapter 2 Key Inputs from Summit Workstreams ACTION TRACKS. 2021. <https://foodsystems.community/food-systems-summit-compendium/action-tracks/> [Web page accessed 27/01/2022]

Wong, F, Vos A, Pyburn R, Newton J. Implementing gender transformative approaches in agriculture. A Discussion Paper for the European Commission. 2019.

8 Appendices

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

Contents

1	Background	7
1.1	The problem, condition or issue	Error! Bookmark not defined.
1.2	The interventions	Error! Bookmark not defined.
1.3	Expected Theories of Change	9
1.4	Rationale for the review	Error! Bookmark not defined.
2	Research questions	10
3	Methodology	11
3.1	Criteria for including and excluding studies in the review (PICOS)	12
3.2	Search strategy	15
3.3	Selection of studies	16
3.5	Data presentation	17
3.6	Limitations	17
4.	References	Error! Bookmark not defined.

1 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 have the ability to take care of themselves and their families (WHO, 2014). 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 (Njuki et al, 2021).

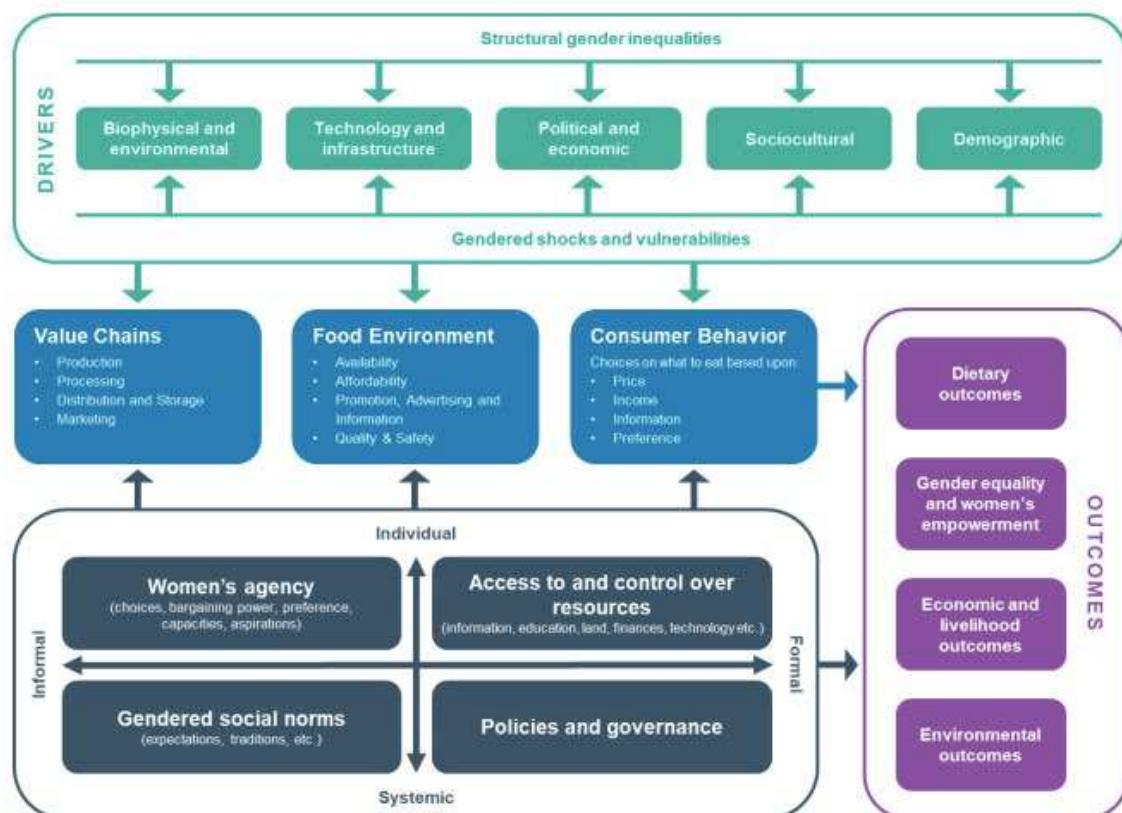
The interventions

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.

1.3 Expected Theories of Change

Our theory of change is based on the pathways developed by Njuki et al (2021) 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.

FIGURE 1. THEORY OF CHANGE, FROM NJUKI ET AL (2021)



Source: Adapted from de Brauw et al. (2019).

1.4 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 (UN, 2021), key decisionmakers 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.

2 Research questions

- 8 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?
- 9 Are there any unintended consequences of such interventions?
- 10 Do effects vary by context, approach to empowerment, or other moderators?

3 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 (Barends et al. 2017).

3.1 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	All else

	empowerment and giving women the capabilities to make decisions on the purchase and consumption of a healthy diet	
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 impacts 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. 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. 2019	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. 2017	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. 2020	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. 2021	<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 and Liu 2009	The intervention is self-help groups for women living in poverty in India.
Emran et al. 2009	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. 2019	The intervention is the Enhanced Homestead Food Production (E-HFP) program, a nutrition- and gender-sensitive agriculture training program.
Marquis et al. 2015	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. 2021	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. 2015	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	<p>Preferred outcome: measures of content in blood/tissue (ex. haemoglobin levels)</p> <p>Secondary outcome: intake in weight (grams, micrograms, etc.)</p> <p>Tertiary outcome: intake in percentage relative to recommended intake</p> <p>Other measures will be considered</p>
Well-being	<p>Preferred outcome: perceived well-being</p> <p>Secondary outcome: anxiety</p>

Types of Comparators

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

3.1.4 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.

3.1.5 Date, language, and form of publication

All proceeding restrictions are from the EGM.

- Date: 2000
- Language: English

3.2 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' EGM², 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

² <https://gapmaps.3ieimpact.org/evidence-maps/food-systems-and-nutrition-evidence-gap-map>

availability, accessibility, and affordability and nutritional status. We will conduct additional targeted searches to identify qualitative studies and process evaluations of the included interventions.

3.3 Selection of studies

3.3.1 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.

3.3.2 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.

3.3.3 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.

3.3.4 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 (Snilstveit 2012):

- *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 analysing data, and information on sampling and recruitment, including at least two sample characteristics.
- *A descriptive quantitative study* collecting primary data using quantitative methods of data collection and descriptive quantitative analysis and report some information on all of the following: the research question, procedures for collecting data, procedures for analysing 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 operationalised. They might also be used to collect organisational 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 (Noyes et al. 2020).

We will assess the quality of included qualitative studies, process evaluations, and descriptive quantitative studies using a mixed-methods appraisal tool developed by Langer and colleagues (2017) and applied in Snistveit and colleagues (2019). 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.

3.4 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 r^2 to provide an estimate of the amount of variability in the distribution of the true effect sizes (Borenstein et al., 2009). 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.

3.5 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.

3.6 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.

4. Protocol references

Ahmed, A., Hoddinott, J. and Roy, S., 2019. *Food transfers, cash transfers, behavior change communication and child nutrition: Evidence from Bangladesh* (Vol. 1868). Intl Food Policy Res Inst.

Bandiera, O., Burgess, R., Das, N., Gulesci, S., Rasul, I. and Sulaiman, M., 2017. Labor markets and poverty in village economies. *The Quarterly Journal of Economics*, 132(2), pp.811-870.

Barends, E., Rousseau, D. M. & Briner, R. B. 2017. CEBMa Guideline for Rapid Evidence Assessments in Management and Organizations. Amsterdam. Available from: <https://www.cebma.org/wp-content/uploads/CEBMA-REA-Guideline.pdf>

Bonuedi, I., Kornher, L. and Gerber, N., 2020. Making cash crop value chains nutrition-sensitive: Evidence from a quasi-experiment in rural Sierra Leone. Available at SSRN 3603918.

Choudhury, N., Raihan, M.J., Ahmed, S.T., Islam, K.E., Self, V., Rahman, S., Schofield, L., Hall, A. and Ahmed, T., 2020. The evaluation of Suchana, a large-scale development program to prevent chronic undernutrition in north-eastern Bangladesh. *BMC Public Health*, 20, pp.1-9.

Deininger, K. and Liu, Y., 2009. Economic and social impacts of self-help groups in India. World Bank Group.

Emran, M.S., Robano, V. and Smith, S.C., 2009. Assessing the frontiers of ultra-poverty reduction: Evidence from CFPR/TUP, an innovative program in Bangladesh. *TUP, An Innovative Program in Bangladesh (March 5, 2009)*.

Heckert, J., Olney, D.K. and Ruel, M.T., 2019. Is women's empowerment a pathway to improving child nutrition outcomes in a nutrition-sensitive agriculture program?: Evidence from a randomized controlled trial in Burkina Faso. *Social Science & Medicine*, 233, pp.93-102.

Herforth, Anna, and Jody Harris. 2014. Understanding and Applying Primary Pathways and Principles. Brief #1. Improving Nutrition Through Agriculture Technical Brief Series. Arlington, VA: USAID/Strengthening Partnerships, Results, and Innovations in Nutrition Globally (SPRING) Project

Marquis, G.S., Colecraft, E.K., Sakyi-Dawson, O., Lartey, A., Ahunu, B.K., Birks, K.A., Butler, L.M., Reddy, M.B., Jensen, H.H. and Huff-Lonergan, E., 2015. An integrated microcredit, entrepreneurial training, and nutrition education intervention is associated with better growth among preschool-aged children in rural Ghana. *The Journal of nutrition*, 145(2), pp.335-343.

Mosha, D., Canavan, C.R., Bellows, A.L., Blakstad, M.M., Noor, R.A., Masanja, H., Kinabo, J. and Fawzi, W., 2018. The impact of integrated nutrition-sensitive interventions on nutrition and health of children and women in rural Tanzania: study protocol for a cluster-randomized controlled trial. *BMC nutrition*, 4(1), pp.1-8.

Njuki, J., Eissler, S., Malapit, H., Meinzen-Dick, R., Bryan, E., Quisumbing, A. 2021. A review of evidence on gender equality, women's empowerment, and food systems: Food Systems Summit Brief Prepared by Research Partners of the Scientific Group for the Food Systems Summit. URL: https://bonndoc.ulb.uni-bonn.de/xmlui/bitstream/handle/20.500.11811/9132/fss_briefs_review_evidence_gender_equalit_y.pdf?sequence=3&isAllowed=y

Pan, Y., Smith, S.C. and Sulaiman, M., 2018. Agricultural extension and technology adoption for food security: Evidence from Uganda. *American Journal of Agricultural Economics*, 100(4), pp.1012-1031.

Savitri, A.I., Idris, N.S., Indawati, W., Saldi, S.R.F., Amelia, D., Baharuddin, M., Sastroasmoro, S., Grobbee, D.E. and Uiterwaal, C.S., 2016. Breastfeeding Attitude and Volume Optimization (BRAVO) trial: study protocol for a randomized controlled trial. *Trials*, 17(1), pp.1-10.

WHO. 2014. Understanding the Women's Empowerment Pathway. Brief #4. Improving Nutrition through Agriculture Technical Brief Series. Arlington, VA: USAID/Strengthening Partnerships, Results, and Innovations in Nutrition Globally (SPRING) Project.

United Nations Food Systems Summit 2021. Chapter 2 Key Inputs from Summit Workstreams ACTION TRACKS. <https://foodsystems.community/food-systems-summit-compendium/action-tracks/> [Web page accessed 27/01/2022]

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)
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 .25, etc.)
Post-intervention or change	0 = Post-intervention, 1 = Change from baseline

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 organisation), 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
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_coeff	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 < .001, or p > .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
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
CI_I	THIS IS FOR SENIOR QUANT LEAD TO FILL OUT
CI_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
cilow_3sf	THIS IS FOR SENIOR QUANT LEAD TO FILL OUT
cihigh_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

3.1 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= Randomised 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	-	

		<p>= Heckman selection model</p> <p>5 = Fixed effects regression</p> <p>6 = Covariate adjusted estimation</p> <p>7 = Propensity weighted regression</p> <p>8 = Comparison of means</p> <p>= Other (please state)</p>		
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	<p>Provide any details in the paper that describe how the study population was selected, covering:</p> <p>a) How is the population selected? what is the sampling strategy to recruit participants from that population into the study?</p> <p>b) What are the characteristics of that study participants?</p> <p>Was this a pilot programme aimed at being scaled up?</p> <p>d) Were there specific factors of success or failure in the implementation?</p>	
General	Type of comparison group	1=No intervention (Service delivery as usual)	Indicate type of comparison group	

		2=Other intervention 3=Pipeline (waitlist) control (still service delivery as usual)		
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.	
1: Assignment mechanism - Assessment	Assignment mechanism: Was the allocation or identification	1= Yes, 2 = Probably Yes, 3 = Probably No, 4	a) The authors describe a random component in sequence	Score "Yes" if all criterion a), b), c) and d) are satisfied.

	mechanism random or as good as random?	= No, 8 = Unclear	<p>generation/ randomisation 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</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</p>
--	--	-------------------	---	--

			<p>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>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	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 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>Score "Not applicable" if it is not a cluster RCT.</p> <p>Score "No" otherwise.</p>	
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	<p>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.</p> <p>Score "Probably yes" if attrition falls into the green zone AND if survey</p>	

		<p>respondents were randomly sampled.</p> <p>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.</p> <p>Score "Probably no" if there is attrition which is likely to be related to the intervention OR there is some indication that the survey respondents were purposely sampled in a</p>	
--	--	--	--

			<p>way that might have led the sampling to be different between treatment and control groups, or attrition falls into the yellow zone.</p> <p>Score "No" if attrition falls into the red zone.</p>	
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	1= Yes, 2 = Probably Yes, 3 = Probably No, 4	a) Baseline characteristics are similar in magnitude;	Score "Yes" if criterion a) and b) are satisfied;

	analysis executed adequately to ensure comparability of groups throughout the study and prevent confounding	= No, 8 = Unclear	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)	<p>Score "Probably yes" if a) is not satisfied but b) is satisfied and imbalances are small in magnitude OR if only a) is satisfied.</p> <p>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"</p> <p>Score "Probably no" if a) and b) are not satisfied and the magnitude of imbalances are small</p> <p>Score "No" if a) and b) are not satisfied and the magnitude of imbalances are large and covariates are clear determinant of the outcomes.</p>
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).	
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</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>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 behaviours; AND the survey process did not reveal information to the control</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>
--	--	--	--

			<p>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>	
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 behaviours ('survey effects').</p>	
6. Performance bias - Assessment	<p>Performance bias: 1= Yes, 2 = Probably 3 = Probably Yes, 4 = No, 5 = Unclear</p>	<p>Was the process of monitoring individuals unlikely to introduce</p>	<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</p>	<p>Score "Yes" if either criterion a) or b) are satisfied; 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>motivation bias among participants?</p>	<p>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</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). 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' behaviours.</p> <p>Score "No" if neither criterion a) or b) are satisfied.</p>
--	--	--	---

			the context of a retrospective (ex post) evaluation.	
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	a) Outcome assessors are blinded, or the outcome measures are not likely to be biased by their judgement. b) For self-reported outcomes: respondents in the intervention group are not	Score "Yes" if criterion a), b), c) and d) are satisfied: 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 outcomes.

		<p>more likely to have accurate answers due to recall bias;</p> <p>c) For self-reported outcomes:</p> <p>respondents do not have incentives to over/under report something related to their performance or actions,</p> <p>OR researchers put in place mechanisms to reduce the risk of reporting bias</p> <p>(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 the future) OR authors</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>
--	--	--	---

			<p>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>	
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	<p>a) A pre-analysis plan or trial protocol is published and referred to or the trial was preregistered, or the outcomes were preregistered;</p> <p>b) Authors report results corresponding to the outcomes announced in the method section (there is no outcome reporting bias);</p> <p>c) Authors report results of unadjusted analysis and intention to treat (ITT) estimation, alongside any adjusted and treatment-on-the</p>	<p>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);</p> <p>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.</p>

			<p>treated/complier average-causal effects analysis.)</p> <p>d) Authors use the appropriate analysis method (use baseline data when available) and different treatment arms are differentiated in the analysis</p> <p>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)</p>	
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).	
--	--	--	---	--

3.2 Appraisal of Risk of Bias for Impact Evaluations using Quasi-experimental designs

Risk of bias assessment tool (QED)

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: randomised or as-if randomised 2= Natural experiment: regression discontinuity (RD) 3= CBA (non-randomised assignment with treatment and contemporaneous comparison group, baseline and end line data collection) – individual repeated		
---------	---	--	--	--

		<p>measurement 4= CBA pseudo panel (repeated measurement for groups but different individuals)</p> <p>5= Interrupted time series (with or without contemporaneous control group)</p> <p>6= Panel data, but no baseline (pre-test)</p> <p>7 = Comparison group with end line data only</p>		
General	Methods used for analysis:	1 = Statistical matching (PSM,	-	

	Which methods are used to control for selection bias and confounding?	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	
--	--	--	--

		<p>7 = Propensity weighted regression</p> <p>8 = Comparison of means</p> <p>= Other (please state)</p>	
General	Study population	<p>Open answer</p> <p>Provide any details in the paper that describe how the study population was selected, covering:</p> <p>a) How is the population selected? what is the sampling strategy to recruit participants from that population into the study?</p> <p>b) What are the characteristics of that study participants?</p> <p>c) Was this a pilot program aimed at being scaled up?</p>	

			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	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

		<p>(Regression discontinuity design) and blinded to participants or;</p> <p>b) if not blinded, individuals reasonably cannot affect the assignment variable in response to knowledge of the participation decision rule;</p> <p>c) and the sample size immediately at both sides of the cutoff point is sufficiently large to equate groups on average.</p>	<p>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.</p> <p>Score "Unclear" if it is unclear whether participants can affect it in response to knowledge of the allocation mechanism.</p> <p>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.</p> <p>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</p>
--	--	---	---

				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.
1: Selection bias - Justification	For assignment based nonrandomised programme 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>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>**Accounting for and matching on all relevant characteristics is usually only feasible when the programme allocation rule is known and there are no errors of targeting. It is unlikely that studies not based on randomisation 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 taken into account 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</p>	Score “No” if relevant characteristics are omitted from the analysis.
--	--	--	---

			taken into account through the use of 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	
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 have an effect on 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	2 - Group equivalence:	1= Yes, 2 = Probably Yes,	

- Assessment	was the method of analysis executed adequately to ensure comparability of groups throughout the study and prevent confounding?	3 = Probably No, 4 = No, 8 = Unclear		
2: Confounding - Justification	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-</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>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 programme.</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: Confounding - Justification	For non-randomised 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 programme**;</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>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 programme – or even whether the non-participants were individuals that refused to participate in the programme, as opposed to individuals that were not given the opportunity to participate in the programme – can help in the assessment of whether the covariates accounted for in the regression capture all the relevant characteristics</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>
--	--	---	--

			that explain differences between treatment and comparison groups.	
2: Confounding - Justification	For statistical matching studies including propensity scores (PSM) and covariate matching** **Matching strategies are sometimes complemented with difference-indifference only uses in the	Open answer	<p>a) Matching is either on baseline characteristics or time-invariant characteristics which cannot be affected by participation in the programme; 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 programme</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>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</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 yield the same results and authors take into account the use of control observations multiple times against the same treatment in their standard error calculation.</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>
--	---	--

	estimation methods. This combination approach is superior since it			
2: Confounding - Justification	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 programme;</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</p> <p>(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>

		<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.</p> <p>However, it plays a different role in the different methods of analysis. While in</p>	
--	--	--	--

			<p>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>	
2: Confounding - Justification	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 identifying instruments are individually significant ($p \leq 0.01$); for Heckman</p>	<p>Score "Yes, if a, b, c, d (if relevant) is addressed." 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. Score "UNCLEAR" if relevant confounders are</p>

		<p>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 overidentifying 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 randomisation 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</p>	<p>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>
--	--	---	--

			programme (for example, weather, infrastructure, community fixed effects, and so forth) through multivariable analysis.	
3: Performance bias - Assessment	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 (randomised 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>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). 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>

			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.	Score "Probably no" if there was imbalance in the frequency of monitoring in intervention groups, which might have influenced participants' behaviours. Score "No" if both criterion a) and b) are not satisfied.
3: 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).	
4: Spillovers, crossovers and contamination: - Assessment	4 - Spillovers, crossovers and contamination: was the study adequately protected against	1= Yes, 2 = Probably Yes, 3 = Probably No, 4 = No, 8 = Unclear	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	Score "Yes" if criterion a), b), c) and d) are satisfied; 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>spillovers, crossovers and contamination?</p>	<p>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 behaviours; AND the survey process did not reveal</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>
--	--	--	--

			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)	
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,	a) Outcome assessors are blinded or the outcome measures are not likely to be biased by their judgement.	Score "Yes" if criterion a), b), c) and d) are satisfied: 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

		8 = Unclear	<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 the 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>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 outcomes.</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>
--	--	-------------	---	--

			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.	
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	1= Yes, 2 = Probably Yes, 3 = Probably No, 4 = No, 8 = Unclear	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	Score "Yes" if a), b), c) and d) are satisfied OR if a) is not met and it is a retrospective NRS. Score "Probably Yes" if authors combined methods and reported relevant tests (d) only for one method OR if all the criteria are met except for a) and it is a prospective NRS

	analysis reporting?	<p>attribution given the data available);</p> <p>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) ;</p> <p>d) Requirements for specific methods of analysis:</p> <ul style="list-style-type: none"> - 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 	<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>
--	---------------------	--	---

			<p>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> <ul style="list-style-type: none"> - 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). 	
6: Reporting bias - Justification	Analysis reporting bias - 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>	

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?	

3.3 Qualitative analysis tool

Study type	Methodological appraisal criteria	Response		
		Yes	No	Comment
<p><i>Screening questions: assessing 'fatal flaws' (Dixon-Woods 2005)</i></p> <p><i>Configurative 'fatal flaws' based on</i></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>) 			

<i>Pawson (2003) TAPUS framework</i>				
	<p><i>Screening question based on abstract and/or superficial reading of full text:</i></p> <p><i>Further appraisal is not feasible or appropriate when the answer is ‘No’ to any of the above screening questions!</i></p>			
Study type	Methodological appraisal criteria	Response		
1. Qualitative and descriptive quantitative, and process evaluations	<p>I. RESEARCH IS DEFENSIBLE IN DESIGN (providing a research strategy that addresses the question)</p> <p><u>Appraisal indicators:</u></p> <p>✓ Is the research design clearly specified and appropriate for aims and objectives of the research?</p>	Ye s	No	Comment / Confidence judgment

	Consider whether						
	<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>						
	Defensible	Arguable	Critical	Not defensible	<i>Worth to continue:</i>		
	II.RESEARCH FEATURES AN APPROPRIATE SAMPLE (following an adequate strategy for selection of participants)						
	<u>Appraisal indicators:</u>						
	Consider whether						
	<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>						

	Appropriate sample	Functional sample	Critical sample	Flawed sample	Worth to continue:
III. RESEARCH IS RIGOROUS IN CONDUCT (Providing a systematic and transparent account of the research process)					
<u>Appraisal indicators:</u>					
Consider whether					
<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>						
Rigorous conduct	Considerate conduct	Critical conduct	Flawed conduct		<i>Worth to continue:</i>		
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							
<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>							
<i>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?)</i>							

	<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>						
	Credible claims	Arguable claims	Doubtful claims	Not credible	<i>If findings not credible, can data still be used?</i>		
	<p>V. REASEARCH ATTENDS TO CONTEXTS</p> <p>(Describing the contexts and particulars of the study)</p> <p><u>Appraisal indicators:</u></p> <p>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>						

	<i>iv. the study makes any claims (implicit or explicit) that infer generalisation (if yes, comment on appropriateness)</i>						
Context central	Context considered	Context mentioned	No context attention				
VI. RESEARCH IS REFLECTIVE							
<i>(Assessing what factors might have shaped the form and output of research)</i>							
<u>Appraisal indicators:</u> Consider whether							
<i>i. appropriate consideration is given to how findings relate to researchers' influence/own role during analysis and selection of data for presentation</i>							
<i>ii. researchers have attempted to validate the credibility of findings (e.g. triangulation, respondent validation, more than one analyst)</i>							
<i>iii. researchers explain their reaction to critical events that occurred during the study</i>							
<i>iv. researchers discuss ideological perspectives/values/philosophies and their impact on the methodological or other substantive content of the research (implicit/explicit)</i>							

	Reflection	Consideration	Acknowledgement	Unreflective research	<i>NB: Can override previous exclusion!</i>
OVERALL CRITICAL APPRAISAL DECISION					
Decision rule:					
- a single critical appraisal judgement ³ in any of the 6 appraisal domains leads to a critical overall judgement.					
- 2 or more high critical appraisal judgements in any of the 6 appraisal domains lead to an overall high risk of bias / low quality rating.					
- 2 or more moderate critical appraisal judgements in any of the 6 appraisal domains lead to an overall moderate risk of bias / moderate quality rating.					
- which means that for a study to be rated of low risk of bias / high quality at least 5 appraisal domains need be rated as of low critical appraisal.					
HIGH QUALITY EMPIRICAL RESEARCH	MODERATE QUALITY EMPIRICAL RESEARCH		LOW QUALITY EMPIRICAL RESEARCH		CRITICAL QUALITY EMPIRICAL RESEARCH

³ For the qualitative studies, we use a slightly different language to scale the critical appraisal assessments as compared to the quantitative studies. The far right rating column always reflects a ‘critical’ appraisal judgement (i.e. ‘unreflective research’ above) with judgements moving further to the left on a scale from high to low critical appraisal.

(Study generates new evidence relevant to the review question and complies with all methodological criteria to ensure reliability and empirical grounding of the evidence).	(Study generates new evidence relevant to the review question and complies with reasonable methodological criteria to ensure reliability and empirical grounding of the evidence).	(Study generates new evidence relevant to the review question and complies with minimum methodological criteria to ensure reliability and empirical grounding of the evidence).	(The evidence generated by the study does not comply with minimum methodological criteria to ensure reliability and empirical grounding of the evidence).
Sources used in this section (in alphabetical order); Campbell et al (2003); CASP (2006); CRD (2009); Dixon-Woods et al (2004); Dixon-Woods et al (2006); Greenhalgh & Brown (2014); Harden et al (2004); Harden et al (2009); Harden & Gough (2012); Mays & Pope (1995); Pluye et al (2011); Spencer et al 2006; Thomas et al (2003); SCIE (2010).			
Study type	Methodological appraisal criteria	Response	
		Yes	No
			Comment /confidence judgment

<p>2. Mixed-methods²</p> <p><u>Sequential explanatory design</u></p> <p><i>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.</i></p> <p><u>Sequential exploratory design</u></p> <p><i>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.</i></p> <p><u>Triangulation designs</u></p> <p><i>the qualitative and quantitative components are</i></p>	<p>I. RESEARCH INTEGRATION/SYNTHESIS OF METHODS</p> <p>(Assessing the value-added of the mixed-methods approach)</p> <p>Applied mixed-methods design:</p> <ul style="list-style-type: none"> ○ Sequential explanatory design ○ Sequential explorative design ○ Triangulation design ○ Embedded design <p><u>Appraisal indicators:</u></p> <p>Consider whether</p> <p><i>i. the rationale for integrating qualitative and quantitative methods to answer the research question is explained [DEFENSIBLE]</i></p>			
--	--	--	--	--

<p>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><i>ii. the mixed-methods research design is relevant to address the qualitative and quantitative research questions, or the qualitative and quantitative aspects of the mixed methods research question</i> [DEFENSIBLE]</p>		
<p><u>Embedded/convergent design</u></p> <p>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>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, synthesise findings, configuration)</i> [CREDIBLE]</p>		
	<p><i>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)</i> [RIGOROUS]</p>		
	<p><i>v. appropriate consideration is given to the limitations associated with this integration, e.g., the divergence of qualitative and quantitative data (or results)?</i> [REFLEXIVE]</p>		

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 2.Research is rigorous in conduct 3.Research is credible in claim 4.Research is reflective	Include / Exclude	1.Low risk of bias 2.Risk of bias 3.High risk of bias 4.Critical risk of bias

Combined appraisal:

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

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

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 (Bandiera et al. 2017) 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

The nine additional qualitative studies were assessed. Five (Kabeer and Data 2020; Das et al. 2013; Kieran et al. 2018; Roy et al. 2015; Roy et al. 2019) were found to be high-quality, with the remaining four (Huda and Kaur 2011; Hagan et al. 2012; van den Bold et al. 2015; Only et al.

2016) marked as medium quality according to the assessment tool. The main 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.

Table 5: Risk of bias in experimental studies

Author	Overall									
	(Year)	Score	Assignment mechanism	Unit of analysis	Selection bias	Confounding	Deviations from intended	Performance bias	Outcome measurement	Reporting bias
Heckert (2019)	Some concer ns	Some concer Risk	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	Low Risk	Some conce rns
Deining er (2009)	Some concer ns	Some concer Risk	Some concer ns	Some concer ns	Some concer ns	Some concer ns	Some concer ns	Some concer ns	Some concer ns	Some concer ns
Haque (2021)	Some concer ns	Some concer Risk	Low Risk	Low Risk	Concer ns	Low Risk	Low Risk	Low Risk	Low Risk	Some conce rns

Blakstad (2020)	Low ROB	Low Risk	Low Risk	Low Risk					
Bandiera (2017)	Some concerns ns	Low Risk	Some concerns ns	Low Risk	Low Risk				
Ahmed (2019)	Some concerns ns	Low Risk	Some concerns ns	Low Risk	Low Risk				

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	Low risk
Marquis (2015)	Low ROB	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk
Emran (2009)	Some concerns	Low risk	Low risk	Some concerns	Low risk	Low risk	Low risk
Bonuedi (2020)	High ROB	Low risk	High risk	High risk	High risk	Low risk	High risk

Appendix 6: Effect estimates from included studies

Table 7: Effect estimates from included studies in REA†

First author	Year	Country	Intervention type	Evaluation / synthesis method	Outcome	Standardized effect estimate (Confidence Interval)	Sample size
<i>Food security</i>							
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	0.07 (0.03; 0.12)	6,732

					eat two meals a day*		
Emran	2009	Bangladesh	Training / education Asset transfer	Difference-in- difference and statistical matching	Meals twice a day	0.6 (0.5; 0.7)	1,569
					Food availability: Sufficient food to meet the household's needs*	0.66 (0.56; 0.76)	1,569
Blackstad	2020	Tanzania	Training / education	Randomized control trial	Household food insecurity access scale*	0.07 (-0.05; 0.2)	876
Pan	2015	Uganda	Training / education Asset transfer	Regression discontinuity	Skip meals*	-0.13 (-0.2; -0.06)	3368
<i>Food affordability and availability</i>							
Ahmed	2019	Bangladesh	Asset transfer (Cash)	Randomized control trial	Per capita monthly food	0.13 (0.08; 0.19)	5000

				consumption - North		
				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	0.11 (0.06; 0.17)	5000

				consumption - North		
				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
				Food consumption score -South	0.12 (0.07; 0.18)	5000
		Assets transfer (Cash and food)		Per capita monthly food	0.11 (0.06; 0.17)	5000

				consumption - North		
				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
		Behaviour change		Per capita monthly food	0.32 (0.27; 0.38)	5000

			communication Asset transfer (Cash)		consumption - North		
					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	Behaviour change	Statistical matching	Total food consumption	-0.04 (-0.18, 0.09)	836

		communication Training / education Assets transfer		expenditure in the 12 months preceding the survey (Food production and market purchases) (LOG)- Household*		
				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

Deininger	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
Emran	2009	Bangladesh	Assets transfer Training / education	Difference-in- difference and statistical matching	Grain stocks (kg)*	0.22 (0.12; 0.32)	1569

Pan	2015	Uganda	Training / education Assets transfer	Regression discontinuity	Per capita food consumption*	0.08 (0.01; 0.15)	3368
<i>Diet quality and adequacy</i>							
Bonuedi	2020	Sierra Leone	Behaviour 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
			Behaviour change communication Training / education Assets transfer		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
Deninger	2009	India	Training / education	Difference-in- difference and statistical matching	Protein intake p.c. (g/day) in the total population*	0.08 (-0.04, 0.19)	1099.5
					Protein intake p.c. (g/day) among the poor of the poor	0.32 (0.05, 0.60)	202
					Protein intake p.c. (g/day)	0.66 (0.30, 1.02)	121.5

					among the poor		
					Protein intake p.c. (g/day) among the non-poor	0.20 (-0.24, 0.64)	78.5
Pan	2015	Uganda	Training / education Asset transfer	Regression discontinuity	Variety of foods consumed*	0.09 (0.02, 0.15)	3368
<i>Anthropometrics</i>							
Heckert	2019	Burkina Faso	Behaviour 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	-0.42 (-0.77, -0.06)	121.6
					Height-for-age z-score	0.40 (0.04,0.75)	121.6

<i>Micronutrient status</i>							
Haque	2021	Bangladesh	Training / education	Randomized control trial	Consumption of at least 100 IFA tablets during pregnancy	0.25 (0.21, 0.28)	10722
					Received vitamin a capsule after last delivery	0.20 (0.16, 0.24)	10722
Heckert	2019	Burkina Faso	Behaviour change communication Asset transfer	Randomized control trial	Change in hemoglobin (g/dL)	0.10 (-0.02, 0.23)	1035
<i>Wellbeing outcomes</i>							
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.