Department for Evaluation

International Initiative for Impact Evaluation

APPENDIXES 2-6

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

A rapid evidence assessment

Appendixes 2-6

Contents

Appendix 2: Protocol

A2.1. Background

The global community is not on track to reach the nutrition targets established by the World Health Assembly and the Sustainable Development Goals (FAO 2022; SUN 2020; World Bank 2021). To course correct, we need to ensure that the impacts achieved by ongoing nutrition and food security interventions are sustained over the long-term. However, long-term impacts of nutrition and food security interventions are often unknown as few evaluations return to projects long after they ended (Pollard & Lindkvist 2020; USAID 2021). Evaluations tend to focus on short-term outcomes due to changes in funders programmatic priorities, unreliable funding streams, and funding being aligned with budget cycles, typically shorter than 5 years.

Nutrition and food security outcomes achieved in the short- and long-term are often different. Incremental change achieved during a project may abate after the project ends. Or, when transitional and transformational change is required, outcomes can take a long time to develop.[1] For example, women's empowerment interventions improve food security, affordability, and availability in the short term (Berretta et al. 2022). However, their effects on longer term outcomes, like diet quality and adequacy or well-being have not yet been established. Reliance on short-term data can therefore result in over- or under-estimating long-term impacts, depending on the causal mechanisms underlying change. Long-term evaluations are needed to establish long-term impact.

Most funders and implementers are interested in achieving sustained change. Understanding long-term effects increases the reliability and validity of the results and can help funders make cost-effective decisions (USAID 2021). Unless we pay particular attention to long-term impacts in project planning and evaluation, we risk myopic implementers that prefer incremental over transformational change if the results of projects targeting the first are measurable in the short-term and the other takes a longer time to develop.

Therefore, there is growing interest in quantifying long-term impacts. USAID has highlighted the need for long-term evaluation by launching its Expanding the Reach of Impact Evaluations (ERIE) initiative, which includes a guide for planning long-term impact evaluations (ERIE, 2018). The Norwegian aid-administration is currently building a food-security portfolio and wishes to understand how to measure and achieve long-term change.

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

A2.2. Research questions

Choice of topic

Based on consultation with the Department for Evaluation at NORAD, 3ie will engage in a rapid evidence assessment to describe the available evidence base on the long-term effects of food security and nutrition interventions. We hope this work will increase interest in what works in the long-term and support the consideration of potential long-term effects in project planning, monitoring, and evaluation (M&E). It can help in portfolio prioritization, highlighting future areas for intervention and research based investments.

Time- focus/time-perspective of evaluations:

1. To what extent do impact evaluations from the Food Systems and Nutrition EGM document long-term effects?
2. To what extent do these impact evaluations document long-term environmental[2] and climate effects?
3. Do these impact evaluations consider adverse effects?

Synthesis of findings:

4. How effective are food-security interventions in the long-term?
5. What facilitates long-term success and failure?

A2.3. Method

Rapid Evidence Assessments (REA) are a type of evidence synthesis approach that have been developed to address policy relevant questions within a more limited time and resource context than what is typically available for full systematic reviews. There is no single definition of a rapid review and recent reviews of study methods have highlighted the variation in rapid review methods (Featherstone et al., 201; Hartling et al., 2015; Khangura et al., 2012; Tricco et al., 2017). However, such approaches typically involve adjusting methods used in traditional systematic reviews and adopt one or more shortcuts to give more timely answers to urgent questions (Schünemann & Moja, 2015).

The approach and methodology below is 3ie's standard approach to rapid evidence assessment developed in line with other types of evidence synthesis. It is based on the rigorous methodology of Barends and colleagues (2017). Other references related rapid evidence assessment methodology are provided in the References section.

A2.3.1 Criteria for including and excluding studies in the review (PICOS)

Only studies included within the Food Systems and Nutrition Evidence Gap Map (EGM) will be considered for this rapid evidence assessment (Table A1.1). Evidence synthesis will be limited to studies considering outcomes more than 10 years after the beginning of

[2] Environmental impacts will be considered broadly and could include a variety of measures such as soil degradation and water quality.

intervention implementation. If fewer than 10 studies are identified to be eligible for inclusion based on this criteria, we will consider reducing this time period, potentially to five years.

Table A1.1: PICOS adapted from the Food Systems and Nutrition EGM for the purposes of this rapid evidence assessment.

Criteria	Included	Excluded
Participants	Individuals in L&MICs	Individuals in high-income countries
Interventions	Food systems interventions related to the production system, distribution and storage, processing and packaging, food loss and waste management, the availability and affordability of food, promotion and labeling, women's empowerment, and behavior change communication.	All other
Comparison	Before-after Intervention-control Business as usual Alternate intervention	If there is no comparison
Outcome	Long-term food security and nutrition outcomes (at least 10 years after the beginning of the intervention)	All other
Study designs	Experimental and quasi-experimental impact evaluations, systematic review, and cost evidence	Qualitative impact evaluations, Descriptive or observational studies that do not assess effectiveness Modelling studies

Notes: Additional details provided in the appendix to the Food Systems and Nutrition EGM (Moore et al. 2021).

Types of study participants

Because only studies within the Food Systems and Nutrition EGM are eligible for inclusion in this rapid evidence assessments, the inclusion criteria for participants in the EGM restricts the inclusion of participants in this work. The Food Systems and Nutrition EGM includes studies with the following populations:

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) are included in the Food Systems and Nutrition EGM. The exception to this is if a country held high-income status for only one year before reverting to LMIC status. These are 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 impacts on people, firms, or institutions in an LMIC, it can be included. For example, we do not exclude a study that measures impacts of New Zealand's immigration visa lottery on residents of Tonga.

Types of interventions

This REA will include any of the interventions in the Food Systems and Nutrition EGM.

Types of outcome measures

This REA will include any outcome in the Food Systems and Nutrition EGM measured at least 10 years after the beginning of an intervention. To specifically examine the extent of the evidence on climate and environmental outcomes, we re-consider interventions evaluating land-related outcomes and code some of these as environmental outcomes for the purposes of this REA. These re-classified outcomes relate to use of chemical fertilizers and cropping intensity.

Types of Comparators

The Food Systems and Nutrition EGM and this REA include studies with the following comparators:

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

Types of study design

Experimental, quasi-experimental, systematic review, and cost evidence are included in the Food Systems and Nutrition EGM. The following study designs are 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 is included, provided that they are associated with an included impact evaluation.

This REA will consider all experimental, quasi-experimental, and cost-effectiveness studies in the Food Systems and Nutrition EGM. Systematic reviews will not be considered in this REA. Systematic reviews are unlikely to report results separately for outcomes more than 10 years after the intervention. Although it is possible that a few consider long-term outcomes, they may choose a different threshold. Given the low likelihood of relevance, we have chosen to exclude these.

Date, language, and form of publication

Other inclusion criteria from the Food Systems and Nutrition EGM which will affect this work include restrictions based on the following:

- Date: 2000
- Language: English
- Publication forms: only completed studies, not protocols

A2.3.2 Search strategy

No new search will be run for this REA. Instead, the results of the living Food Systems and Nutrition EGM, will be relied on as the basis of the search strategy. The EGM was first published in January 2021 and is currently updated with new studies every four months. Studies included through the search conducted in July 2022 will be reviewed for this REA.

A2.3.3 Selection of studies and data extraction

Screening

Study screening will adopt a rapid synthesis approach. A single reviewer will record the time period over which each study in the Food Systems and Nutrition EGM measures impact. A subset of these will be checked by a second reviewer for quality assurance. The studies which consider outcomes over 10 years will be selected for additional data extraction.

Data extraction and coding procedures

Bibliographic, geographic, and other descriptive data has already been extracted from all studies included in the Food Systems and Nutrition EGM. Additional data related to effect sizes, moderators, barriers and facilitators of impact, sustainability and equity implications, and other considerations for practitioners will also be extracted (Appendix 1). Moderators to be considered include extrinsic, methodological, and substantive characteristics.

Critical appraisal

Included impact evaluations will be appraised using a rapid critical appraisal tool (Appendix 2). Appraisal will be conducted by a single coder and reviewed by another team member.

A2.3.4 Analytical approach

Quantitative analysis

A meta-analysis will be conducted if two or more effect sizes considering sufficiently similar interventions, outcomes, and populations are identified. We will choose the appropriate formulae for effect size calculations in reference to, and dependent upon, the data provided in included studies. We will assess heterogeneity by calculating the Q statistic, I^2, and τ^2 to provide an estimate of the amount of variability in the distribution of the true effect sizes (Borenstein et al., 2009). We will explore heterogeneity through the use of moderator analyses if the data allow. Moderators to be considered will include extrinsic, methodological, and substantive characteristics, if data are available. We will also test for the presence of publication bias if at least 10 studies are included in the analysis. If sufficiently similar studies are not identified, effect sizes will be presented individually and interpreted with caution.

To ensure the independence of effects within outcome categories, only one effect size will be extracted from each study for each outcome category. If a study considers multiple indicators for the same outcome category, the most biologically relevant outcome will be extracted. If multiple model specifications are tested and presented, the author's preferred model will be used. Additional detail on model and indicator prioritization will be established once study designs and outcome categories of included studies are known, but before data extraction begins.

Qualitative analysis

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

A2.3.5 Data presentation

We will provide a narrative summary of the papers considering outcomes over 10 years or longer. The summary 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. A meta-analysis and forest plot will be presented when the data is sufficient. Qualitative information will be summarized narratively to support project design and implementation. Below, we outline how information will be presented to respond to each research question (Table A2.2).

Table A2.2: Data presentation for each research question.

Research question	Presentation
1. To what extent do impact evaluations from the Food Systems and Nutrition EGM document long-term effects?	Plot showing the distribution of impact evaluations measuring outcomes each year after the start of an intervention.
2. To what extent do these impact evaluations document long-term environmental and climate effects?	Plot showing the distribution of impact evaluations measuring *environmental* or *climate* effects each year after the start of an intervention. Depending on the extent of this evidence base, these outcomes may be presented in a disaggregated manner.
3. Do these evaluations consider adverse effects?	Description of adverse effects mentioned in studies considering long-term outcomes. Adverse effects may be discussed qualitatively or quantitatively by authors.
4. How effective are food-security interventions in the long-term?	Meta-analysis if data allows. If data is insufficient, individual effect sizes will be presented. If articles report short- and long-term effects, these *may* both be presented, depending on resource limitations.
5. What facilitates long-term success and failure?	Narrative summary of qualitative evidence.

A2.3.6 Limitations

Due to the rapid nature of this work, results should be interpreted more cautiously than those of a systematic review. Relying on the Food Systems and Nutrition EGM may result in some relevant studies being omitted from this evidence assessment. Although the EGM is rigorous and broad, this broad nature means that some studies related to the long-term outcomes may have been missed. 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.

This study will consider the sustained effects of interventions after they start, regardless of the period during which they occurred. Interventions that took place for a short period of time and those that are still ongoing will all be included, potentially adding heterogeneity to

results. This will be considered in analysis, presentation, and discussion. Sub-groups will be created as theoretically justified.

A2.4 References

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

Berretta, M., Kupfer, M., Lane, C., Shisler, S. 2022. "Rapid evidence assessment on women's empowerment interventions within the food system: a meta-analysis." Submitted for review at Agriculture & Food Security. Available from: https://www.researchsquare.com/article/rs-1754233/v1

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

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

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

USAID. 2021. Expanding the reach of impact evaluation (ERIE). Available from: https://www.usaid.gov/PPL/MERLIN/ERIE

A2.4.1 Additional references on methods for rapid evidence assessments

Thomas, J., Newman, M. and Oliver S. 2013. Rapid evidence assessments of research to inform social policy: taking stock and moving forward. Evidence & Policy, 9(1): 5-27.

Haby, M.M., Chapman, E., Clark, R., Barreto, J., Reveiz, L. and Lavis, J.N. 2016. What are the best methodologies for rapid reviews of the research evidence for evidence-informed decision making in health policy and practice: a rapid review. Health research policy and systems, 14(1): 1-12.

Ganann R., Ciliska D. and Thomas H. 2010. Expediting systematic reviews: methods and implications of rapid reviews. Implementation Science, 5(1): 1-10.

Hunter, J., Arentz, S., Goldenberg, J., Yang, G., Beardsley, J., Lee, M.S. and Myers, S.P. 2020. Choose your shortcuts wisely: COVID-19 rapid reviews of traditional, complementary and integrative medicine. Integrative Medicine Research, 9(3): 100484.

Harker, J. and Kleijnen, J. 2012. What is a rapid review? A methodological exploration of rapid reviews in Health Technology Assessments. International Journal of Evidence-Based Healthcare, 10: 397–410.

Hartling, L., Guise, J.M., Kato, E., Anderson, J., Aronson, N., Belinson, S., Berliner, E., Dryden, D., Featherstone, R., Foisy, M. and Mitchell, M. 2015. EPC Methods: An Exploration of Methods and Context for the Production of Rapid Reviews. Research White Paper.

Tricco, A.C., Antony, J., Zarin, W., Strifler, L., Ghassemi, M., Ivory, J., Perrier, L., Hutton, B., Moher, D. and Straus, S.E. 2015. A scoping review of rapid review methods. BMC medicine, 13(1): 1-15.

Varker, T., Forbes, D., Dell, L., Weston, A., Merlin, T., Hodson, S. and O'Donnell, M. 2015. Rapid evidence assessment: increasing the transparency of an emerging methodology. Journal of evaluation in clinical practice, 21(6): 1199-1204.

Watt, A., Cameron, A., Sturm, L., Lathlean, T., Babidge, W., Blamey, S., Facey, K., Hailey, D., Norderhaug, I. and Maddern, G. 2008. Rapid reviews versus full systematic reviews: an inventory of current methods and practice in health technology assessment. International Journal of Technology Assessment in Health Care, 24(2): 133-9.

Appendix 1 of protocol: Provisional data extraction

Table A1.1: Provisional data extraction fields.	
VARIABLE LABEL	**EXPLANATION**
Study ID	This is the study ID - it should match the study ID from the Outcome Mapping Sheet (e.g., 946578)
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., 946578_1, 946578_2 and so on)
STUDY DESCRIPTIVE INFORMATION	
Author	Author last name For 1 author: leading author last name (e.g. Gomez) For 2 authors: both author last names with ampersand in between (e.g. Smith & Bahn) For 3 or more authors: leading author last name followed by et al. (e.g. Gupta et al.)
Year	Year published
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 - dichotomous outcome, 5 = Regression data - continuous outcome

Comparison	1=No intervention (service delivery as usual), 2=Other intervention, 3=Pipeline (wait-list) control (still service delivery as usual)
Describe Comparison Group	Free text, describe the comparison group
Subgroup	Is this analysis of a subgroup? 0=no, 1=yes
If yes to subgroup, describe	Free text, describe the subgoup 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 description	Provide detailed description of the intervention such that a reader could easily understand what happened. Avoid copying text directly from the article as it is likely to be verbose. Summarize in your own words but include page numbers for quick reference. If more than two or more interventions are being evaluated, please provide descriptions for each intervention arm under separate rows, e.g. description of cash transfer (in all rows where estimate id's evaluate the cash transfer), description of cash transfer + community mobilization (in all rows where estimate id's evaluate the multicomponent intervention).
Intervention code	Dropdown menu with intervention codes
Exposure to intervention (in months)	How long is the intervention exposure itself? If time series is used, indicate the lenght of the period covering data points when the intervention was going on.
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 from baseline?	0 = Post-intervention, 1 = Change from baseline

OUTCOMES	
Outcome description	Record the outcome for the corresponding effect size. Use this open answer field to enter, in the author's own words, a description of the outcome. Be selective and concise with the excerpts being transcribed here as to ensure accurate and precise descriptions of the outcome. To the extent possible, be sure to include numbers, units, population, and comparators. Include page numbers with every excerpt extracted.
Outcome codes	Dropdown menu with outcome codes
Dataset	Record if data for this outcome comes from an identified dataset
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

diff_p-value	p-value 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)
reg_CI_LB	Report the associated Lower bound of the 95% Confidence interval of the effect size. If CI is reported for a different confidence level, indicate that in the notes section.
reg_CI_UP	Report the associated Upper bound of the 95% Confidence interval of the effect size. If CI is reported for a different confidence level, indicate that in the notes section.
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 1="Yes" if the sample size needs to be divided by the number of periods, and 0="No" if either it is cross-sectional data, or if the authors have already divided the number of observations by the number of panel assessments and thus no correction is necessary.
Treatment Variable	Record the treatment variable as written in the model (e.g., the variable name the author uses, such as ("Intervention x Time")
CODING RECORDS	
dataset	Record if data comes from an identified dataset
coder	Record your name
Notes	Record any notes important for the team
DO NOT USE THIS SECTION	
n_T_revised	THIS IS FOR SENIOR QUANT LEAD TO FILL OUT
sp	THIS IS FOR SENIOR QUANT LEAD TO FILL OUT
d	CODING RECORDS
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_l	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
yi_1	THIS IS FOR SENIOR QUANT LEAD TO FILL OUT
yi_rev	THIS IS FOR SENIOR QUANT LEAD TO FILL OUT
yi	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_yi	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
QUALITATIVE DATA	
Collaborating organization	Implementing partner, funding agency (of the research and/or implementation of the intervention), and/or university affiliation
Hypothesized mechanism of action	Describe the study's theory of change (described by authors) or path from cause to effect on stated outcomes
Unintended consequences	Describe any unintended consequences. These could be quantified or observed. They can include unexpected patterns in outputs / outcomes. Equity considerations can be listed here (ex. women were unintentionally excluded from the intervention). Also consider environmental and climate effects.
Barriers and facilitators to implementation	Based on the authors' own views, what were the things that facilitated or inhibited implementation. What went well? What did not? Avoid copying text directly from the article as it is likely to be verbose. Summarize in your own words.
Covariates	If any quantitative analysis was considered, list the covariates included. This can help future evaluators determine the data that they are likely to need when doing their own analysis
Outcomes	Present any relevant outcomes that have not been captured in the Outcome description column of the Quant tool. This could include outcomes analyzed using methods excluded based on our study design criteria. Be selective and concise with the excerpts being transcribed here as to ensure accurate and precise descriptions of the outcome. To the extent possible, be sure to include numbers, units, population, and comparators. Include page numbers with every excerpt extracted.
Sources of bias and limitations	Report any author reported bias or limitations of the study

Sustainability comments	Based on the authors' own views, was the project likely to be sustainable? Was it cost effective? Should be implemented in the future? Sustainability can be considered along two dimensions: are effects likely to continue AND are effects worth the investment. Be sure to include information on expected differences between short and long term effects.
Cost effectiveness comments	Report any cost data provided, include the authors' comments on cost data even if quantifications are not provided.
Other	Provide any other relevant information, notes or comments on the study or data extracted .

Appendix 2 of protocol: Rapid critical appraisal tool

Table A1.2: Rapid risk of bias tool

Risk of bias category	Question	Explanation
Confounding	Did the identification strategy achieve a valid counterfactual? (pg #s)	Yes if covariates are balanced at baseline, parallel trend assumption was not rejected, placebo tests were conducted to verify robustness, or any other test of the identification strategy assumptions is reported.
Attrition	Is there unbalanced attrition that has not been dealt with in the analysis?	No if attrition rate in both control and treatment groups is less than or equal to 5% and is differential (ie unequal between treatment and control conditions); OR less than or equal to assumed in power calculations AND convincing statistical techniques have been used to identify and address the attrition bias, AND it is random (balance between attritors and non-attritors); OR cross-sectional quasi-experimental or fixed effect study designs are used.
Spill-overs/ contamination	Was the study adequately protected against spill-overs/ contamination? (pg #s)	Yes if the intervention is unlikely to spill-over to comparisons, groups are isolated from other interventions which might affect the outcomes, intervention delivery could not affect the performance of the groups in different ways (eg. blinding), AND an appropriate analysis was used to estimate the effect of assignment (ITT).
Outcome measurement	Could the measurement of the outcome be different between the study arms or be affected by knowledge of assignment?	No if outcomes were not measured at different time periods or using different methods, OR outcome assessors were blinded, OR outcomes were not self-reported by participants, OR are unlikely to be influenced by knowledge of administration of the intervention, OR outcomes come from administrative records.

Reporting	Reporting	Yes if there is no evidence that outcomes were selectively reported, AND all reported results for the outcome domain correspond to all intended analyses OR There is only one possible way in which the outcome domain can be analyzed OR researchers have provided the reasons for any inconsistencies (not related to the nature of the results).
Summary	What are the key risks of bias identified, and what are their potential implications for interpreting the effects?	Add a brief summary that focuses on the likely implications for interpreting the effects: are the effects likely to be suppressed due to the noted sources of bias? Exaggerated? Are there substantial quality issues or lack of clarity that should be considered? What should the reader keep in mind, to contextualize the findings?

Appendix 3: Methodological details

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

The approach and methodology presented below are 3ie's standard approach to rapid evidence assessment developed in line with other types of evidence synthesis. Our work is rooted in the rigorous methodologies of systematic reviews (Campbell, 2017) and amended to account for time and resource limitations (Barends et al. 2017). The primary adjustment to the standard synthesis approaches made in this REA is a reduction in the scope of the search. The search was limited to studies already included in 3ie's living Food Systems and Nutrition Evidence Gap Map. Other references on rapid evidence assessment methodology have already been provided in the references to the protocol (Appendix A2.4.1) and are not duplicated here.

A3.1Criteria for including and excluding studies in the review (PICOS)

The inclusion criteria for this REA are largely driven by the inclusion criteria for the Food Systems and Nutrition Evidence Gap Map. Since the map is the only database searched, only studies meeting the inclusion criteria for the map could be included in this rapid evidence assessment. In Table A3.1, we summarize the inclusion criteria for the Food Systems and Nutrition Evidence Gap Map and indicate modifications made for this work.

Table A3.1: PICOS adapted from the Food Systems and Nutrition EGM for the purposes of this rapid evidence assessment.

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

Notes: Additional details provided in online Appendix A to the Food Systems and Nutrition EGM (Moore et al., 2021).

Types of study participants

Only studies considering 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, the year of publication) are included. The exception to this is if a country held high-income status for only one year before reverting to low- or middle-income status. These are included even if the intervention began in the high-income year. This applied to Argentina (2014, 2017), Venezuela (2014), Mauritius (2019), and Romania (2019). If the study is conducted in a high-income country but measures impacts on people, firms, or institutions in an low- or middle-income country, it is included. For example, studies that measure impacts of New Zealand's immigration visa lottery on residents of Tonga would be included.

Types of interventions

Food systems interventions are defined based on the framework developed by High Level Panel of Experts and extended by the International Food Policy Research Institute (HLPE 2017; Brauw et al. 2019). The framework is made up of three domains; food supply chain, food environment and consumer behaviour. See Table A3.2 for full intervention list. Only studies considering food production interventions are selected for additional synthesis. For the purpose of making meaningful groups within this REA, we define eligible interventions as: soil and water conservation; the provision of agricultural inputs; agricultural education; agricultural insurance; land titling and markets; sustainability farming certificates; agricultural credit and savings; and contract farming. This grouping is developed by aggregating and dis-aggregating some of the intervention groups from the Food Systems and Nutrition Evidence Gap Map. For example, the map specifies different interventions for farmer field schools and agricultural extension programs, but we combine these.

Several studies identified during our search involved the adoption of agricultural practices without a program supporting adoption (ex. adoption of improved seed varieties without an intervention advocating for adoption). These studies are excluded as no external intervention is conducted.

Table A3.2: Intervention framework

Domain	Intervention category	Intervention
Food supply chain	Food production	Provision of improved water access and management systems
		Provision of free or reduced-cost access to improved seed varieties
		Provision of free or reduced-cost access to fertilizer
		Provision of free or reduced-cost access to pesticides/herbicides
		Provision of free or reduced-cost access to livestock
		Provision of free or reduced-cost access to other/unspecified agricultural inputs
		Provision of mechanical equipment
		Education/information- Farmer field schools
		Education/information- Agricultural extension programs
		Education/information- Information guidance
		Education/information- other educational programs
		Other efforts to improve the production system- Insurance
		Other efforts to improve the production system- Contract farming
		Other efforts to improve the production system- Market support
		Other efforts to improve the production system- Land markets and management
		Other efforts to improve the production system- Agricultural credit/savings
		Other efforts to improve the production system- Other
	Food transport / storage	Support for creating storage structure at farms
		Trade regulations
		Implementation of distribution centres
		Improved transportation from farms to markets
		Education regarding improved storage and distribution techniques
		Cold chain initiatives
	Processing and packaging	Fortification
		Packaging
		On farm, post-harvest processing
		Provision of good or services to support food processes of business models

		Education regarding improved processing and packaging techniques
	Food loss and waste management	Private food donation
		Use of and education regarding the use of spoiled, near spoiled, or traditionally uneaten food
		Composting
Food Environment	Food provision / price reduction	Designations of space and zoning laws
		Direct provision of foods
		Provision or use of supplements
		Cash-for-food programs
		Governmental price manipulations (excluding tariffs)
		Advertising regulations
		Innovative store design
		Labelling regulations
	Quality and safety	Food safety regulations
Consumer behaviour	Efforts to increase women's decision-making power	Efforts to increase women's decision-making power
	Behaviour change communication	Peer support / counsellors
		Professional services (dieticians / nurses)
		Community meetings
		Classes
		Healthy food social marketing campaigns
		Door-to-door campaigns

Types of outcome measures

This REA included food security and nutrition outcomes, defined according to the Food Systems and Nutrition Evidence Gap Map, measured at least 10 years after the beginning of an intervention. See Table A3.3 for full list of outcomes. Outcome groups for which we identified no long-term evaluations are dropped from all visualizations to facilitate presentation. Only studies measuring impacts on income, crop production, climate, or the environment are selected for additional synthesis.

Table A3.3: Outcome framework

Stage of theory of change	Outcome group	Outcome sub-group
Intermediate	Economic	Income
		Assets
		Output value
		Prices received for goods
		Other socioeconomic indicators
		Tax revenue
		Purchasing behaviour
	Agricultural	Water-related
		Animal husbandry
		Plant/crop production
		Land related
		Quality of agricultural inputs
		Agricultural cooperatives
	Bio-nutritional	Food nutrient content
		Caloric requirements
		Nutrient bioavailability
	Advertising and labelling	Exposure to advertisement
		Advertisement topics
		Accuracy of advertisement
	Food distribution	Import / export
		Movement of food
		Location of foods in stores
		Food distribution centres
	Climate / environment	Climate impact
		Non-food waste produced
	Food loss	Time food remains unspoiled
		Food spoilage
		Food loss
	Intrinsic motivators	Consumer preferences

		Perceptions
		Knowledge
	Women's empowerment	Decision making
		Ownership
		Control of resources
		Self-esteem
		Time use
		Other women's empowerment outcome
	Regulations	Violations
		Fines
		Other regulation outcome
	Economic, social, and political stability	Economic, social, and political stability
	Time use	Time use
	Behaviour change	Behaviour change
	Other steps taken due to non-compliance	Other steps taken due to non-compliance
Final	Anthropometric	Linear growth
		Weight
		Relative weight
		Mid-upper arm circumference
		Birth outcomes
		Anthropometric other
	Developmental	Physical
		Other developmental outcomes
	Micronutrient status	Iron
		Iodine
		Vitamin A
		Zinc
		Other micronutrient status outcome
	Diet quality / adequacy	Breastfeeding
		Dietary diversity

		Insufficient diet
		Micronutrient intake
		Other diet quality and adequacy
	Food safety	Food toxins
		Food borne illness
		Other food safety outcome
	Food affordability / availability	Food access
		Food availability and supply
		Affordability
		Food insecurity measures
		Food stressed households

Types of comparators

Studies using pre-intervention data, an alternate intervention, a control group, or business as usual comparators, including pipeline and waitlist controls, are included. Studies with no comparator are excluded.

Types of study design

This REA considers experimental, quasi-experimental, and cost-effectiveness studies. As systematic reviews are unlikely to report results separately for outcomes 10 or more years after the intervention, they are not considered in this REA. Although it is possible that a few systematic reviews consider long-term outcomes, they may choose a different threshold. Given the low likelihood of relevance, we choose to exclude these. The following study designs are included in this REA:

- Randomized controlled trial
- Regression discontinuity design
- Controlled before-and-after studies, including
    - Instrumental variable
    - Fixed-effects models, including difference-in-differences and any mathematical equivalents
    - Matching techniques, including propensity-weighted multiple regression
- Interrupted time series

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

Date, language, and form of publication

Inclusion is restricted to studies in English from the year 2000. We include only published complete studies and not protocols for the evidence synthesis.

A3.2 Search strategy

No new search is conducted for this REA. Instead, the results of the living Food Systems and Nutrition EGM, are the basis of the search strategy. The EGM was first published in January 2021 and is currently updated with new studies every four months. Studies included through the search conducted in July 2022 are reviewed for this REA.

A3.3 Selection of studies and data extraction

Screening

Study screening adopted a rapid synthesis approach. A single reviewer recorded the time period over which each study in the Food Systems and Nutrition EGM measures outcomes. A subset of about 100 studies were checked by a second reviewer for quality assurance. All included studies were also checked by a second reviewer. The studies considering outcomes 10 years or more after intervention initiation were selected for additional data extraction. If, at

any time in this screening process, studies were determined to be ineligible for the Food Systems and Nutrition EGM as a whole, they were excluded from both this work and the underlying EGM.

During the data extraction phase, the second set of inclusion criteria, restricting the synthesis to agricultural interventions and income, crop production, climate, or environment outcomes was applied.

Data extraction and coding procedures

Bibliographic, geographic, and other descriptive data had already been extracted from all studies included in the Food Systems and Nutrition Evidence Gap Map. Additional data related to effect sizes, moderators, barriers and facilitators of long- and short-term impact, sustainability and equity implications, and other considerations for practitioners was extracted from studies included in the evidence synthesis (Appendix 1 of the Protocol above).

In the first phase of data extraction, we identified the effects to be extracted by listing the types of interventions, outcomes, and methods of each included study. Approximately half of this mapping was conducted in duplicate. We reviewed the extracted information to develop decision-rules regarding which outcomes to extract in full. We were careful to engage with each study on its own terms. The primary concern at this stage was shortlisting outcomes before we began analysis to avoid selection bias. Only effects relating to the impacts of agricultural interventions on income, crop production, climate, or the environment were selected for the second phase of data extraction.

Definitions of primary and secondary outcomes for extraction were specified with a focus on including heterogenous effects. If the same outcome was measured with multiple variables, coders extract composite measures. If measures were reported in absolute amounts and per capita/acre amounts, the latter was preferred. If income was reported separately for agricultural and non-agricultural income, both were extracted. If authors present multiple models, we use the one that is preferred by the author themselves. If studies report sub-group analysis based on sex, age, socio-economic status, or length of the study period, outcomes for all sub-groups were extracted.

After agreeing on the types of outcomes, the outcomes were extracted in a second phase. Coders were informed of the indicators to extract for each study and data extraction was conducted in duplicate. After data extraction, the two coders met to discuss their extraction and reconcile any differences.

Qualitative insights from each included study are extracted by one nominated team member based on pre-agreed standards. Information relating to the hypothesized mechanism of action for the intervention; unintended and adverse effects; implementation; outcomes not eligible for inclusion in this work; sustainability; cost; implementing partners; and funders were extracted (Appendix 1 of the Protocol above).

A3.4 Critical appraisal

Included impact evaluations were appraised using a rapid critical appraisal tool (Appendix 2 of the Protocol above). After the independent data extraction was completed, the coders

evaluated each *outcome* from the shortlisted studies on the following criteria along with the justification for their scoring:

- Confounding – If the identification strategy achieved a valid counterfactual
- Attrition – If there was an unaddressed issue of unbalanced attrition in the study
- Spillovers/ Contamination – If the study adequately addressed potential spillovers
- Outcome Measurement – If the measurement of outcome could be affected by the knowledge treatment group
- Reporting – If all outcomes and intended analysis were in line with the described methods

Every outcome from the evaluation study was put through this five-point test, after which two coders made an independent assessment of the Risk of Bias. If any outcome scored a negative on one of the five, coders assigned it as having a potential for Risk of Bias. Once their decisions were made, the pair meet to reconcile their critical judgments and data extractions. For outcomes and studies where there was a difference of opinion, each makes their case and agree on each criterion and extraction category. Therefore, each outcome was critically appraised by two independent analysts who later convene to cross check their informed, statistical opinions.

A3.5 Analytical approach

Quantitative analysis

A meta-analysis was conducted for effects on income, crop production, climate, and the environment. Meta-regression and subgroup analysis were used to examine variation in effects by intervention type, geographic region, length of the study period, and study methods.

To compare the effect sizes, we converted all of them to a single metric, Cohen's d. We then converted all Cohen's *d* to Hedges *g* to correct for small sample sizes using the following formula (Ellis, 2010):

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

Where *n* denotes the sample size of the treatment (n_T) and control (n_C) groups.

We chose the appropriate formulae for effect size calculations in reference to, and dependent upon, the data provided in included studies. For example, for studies reporting means (X) and pooled standard deviation (SD) for treatment (T) and control or comparison (C) at follow up only (p+1), we used the following formula:

$$d = \frac{\bar{X}_{T_{p+1}} - \bar{X}_{C_{p+1}}}{SD_{p+1}}$$

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

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

Where the intervention is expected to change the standard deviation of the outcome variable, we used the standard deviation of the control group only.

For studies reporting means ($\underline{X}$) and standard deviations (SD) for treatment and control groups at baseline (p) and follow up (p+1), the following formula was used:

$$d = \frac{\Delta\bar{X}_{p+1} - \Delta\bar{X}_{p}}{SD_{p+1}}$$

For studies reporting mean differences ($\Delta\underline{X}$) between treatment and control and standard deviation (SD) at follow up (p+1), the following formula was used:

$$d = \frac{\Delta\bar{X}_{p+1}}{SD_{p+1}} = \frac{\bar{X}_{T_{p+1}} - \bar{X}_{C_{p+1}}}{SD_{p+1}}$$

For studies reporting mean differences between treatment and control, standard error (SE) and sample size (n), the following formula was used:

$$d = \frac{\Delta\bar{X}_{p+1}}{SE\sqrt{n}}$$

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

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

where *n* denotes the sample size of treatment group and control. We used the following where total sample size information (*N*) is available only (as suggested in Polanin, 2016):

$$d = \frac{2t}{\sqrt{N}}$$

$$Var_d = \frac{4}{N} + \frac{d^2}{4N}$$

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

$$SE = \sqrt{N}\ \frac{(upper\ CI - lower\ CI)}{3.92}$$

If the study does not report the standard error, but reports *t,* we extracted and used this as reported by the authors. If an exact p-value is reported but no standard error or *t*, we used the inverse of the two-tailed Student's T distribution to obtain *t* based on the exact p-value and the sample size minus 1 (n-1). If the precise p-value and standard error are not reported, then we assumed the following for the t statistics:

$$p - value\ >\ 0.1: t\ =\ 0.5$$
$$0.1\ \geq\ p - value\ >\ 0.05: t\ =\ 1.645$$
$$0.05\ \geq\ p - value\ >\ 0.01: t\ =\ 1.960$$
$$0.01\ \geq\ p - value:\ >\ 0.001: t\ =\ 2.576$$
$$p - value:\ \geq\ \ 0.001: t\ =\ 3.291$$

Where outcomes are reported in proportions of individuals, we calculated the Cox-transformed log odds ratio effect size (Sanchez-Meca et al., 2003):

$$d = LogOddsRatio\ \frac{\sqrt{3}}{\pi}$$

We fit a random effects meta-analyses model when we identified two or more studies that we assess to be sufficiently similar. We assessed heterogeneity using the DerSimonian-Laird estimator by calculating the *Q* statistic, I^2, and τ^2 to provide an estimate of the amount of variability in the distribution of the true effect sizes (Borenstein et al. 2009). If sufficiently similar studies were not identified, effect sizes were presented individually and interpreted with caution. This was the case for some sub-group analysis used to explore potential drivers of variation.

To improve the independence of effects within outcome categories (income, crop production, climate, and the environment) only one effect size was used in the meta-analysis from each intervention for each outcome category. Other outcomes are discussed in the quantitative analysis to explain variation in results. If multiple interventions are considered within a single study, effect sizes for all were extracted, even if they use the same comparator.

Qualitative analysis

We performed a thematic barriers and facilitators analysis to identify any implementing element which may or may not facilitate program success, as well as any nuances about the context of each included study following the method by Thomas and Harden (2008). Specific context-related information which can help to understand and explain the direction of the meta-analysis effects were included to give an overall view of how interventions work.

First, two independent coders read the full text of the studies included in the evidence synthesis and extracted insights into the following areas of inquiry: intervention description, sustainability, unintended consequences, adverse effects, hypothesized mechanism of action, and cost evidence. Insights were generally drawn from the authors' own conclusions about what worked well, or did not, in the included interventions. Because none of the included evaluations were mixed-methods, qualitative data is not available to validate authors' conclusions. Insights were paraphrased and organised in Excel with each insight recorded in a unique row under the corresponding area of inquiry.

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

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

While drawing findings from the studies, we are aware that they may be specific to the study region and may not translate as general conclusions. This does not however, take away from the thematic understanding of what worked and didn't work for the interventions.

A3.6 Data presentation

We present a series of tables and figures describing the evidence base of impact evaluations measuring outcomes 10 years or more after the beginning of implementation (research questions 1-3). These include a breakdown by year of publication, time since implementation began, intervention and outcomes considered, and country of the intervention. We then present the results of the meta-analysis, focusing on the underlying factors which may explain variation in results (research question 4). Finally, we narratively describe the facilitators and barriers to impact, focusing on how these may function differently in the short- and long-term (research question 5).

A3.7 Limitations of the methods

Due to the rapid nature of this work, results should be interpreted more cautiously than those of a systematic review. Relying on the Food Systems and Nutrition Evidence Gap Map may result in some relevant studies being omitted from this evidence assessment. Although the

EGM is rigorous and broad, this broad nature means that some studies related to the long-term outcomes may have been missed. In particular, the Food Systems and Nutrition Evidence Gap Map only contains one study from each set of linked studies. If one of the linked studies that was excluded from the map contained long-term information, this would not have been identified. Linked studies are those that are published on the same intervention and present similar analysis. Often, this occurs when a working paper and a journal article are published on the same intervention, generally by the same authors. Only one publication among a set of linked studies is retained to avoid over-representing the evidence base. We expect the risk of missing relevant linked studies to be low as the original Food Systems and Nutrition EGM included the most recent study published in any set of linked studies. The most recent study is expected to be the one that considers the longest time period. However, during the update period, 20 additional linked studies have been identified. In these cases, the originally included study was retained.

This REA considers the sustained effects of interventions after they start, regardless of the period during which they occurred. Interventions that took place for a short period of time and those that are still ongoing are included, potentially adding heterogeneity to results.

The issue of significant heterogeneity limits interpretation and findings. The interventions considered in the evidence synthesis are highly variable, as are the populations considered and methods employed. As such, average treatment effects may not be meaningful when considering the expected effects of specific types of interventions or outcomes for specific groups. Interpretation of average reported effect estimates should be done cautiously. Instead, the focus should be on understanding drivers of variation in effects.

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

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

Table A4.1 Outcome set 1

Interventions	Outcomes							
	Economic	**Agricultural**	**Climate / environment**	**Anthropometric**	**Behaviour change**	**Bio nutritional**	**Developmental**	**Diet quality / adequacy**
Agricultural credit / savings	3	1	0	1	0	0	0	1
Agricultural extension programs	9	5	0	0	3	1	0	1
Agricultural information / guidance	0	1	0	0	0	0	0	0
Agricultural insurance	2	2	1	0	0	0	0	0
Cash-for-food programs	1	1	0	1	0	0	1	0
Classes	1	0	0	1	0	0	0	0
Cold chain initiatives	1	1	0	0	0	0	0	0
Composting	0	0	0	0	0	0	0	0
Contract farming	0	1	0	0	0	0	0	0
Direct provision of foods	5	0	0	1	2	0	0	1
Distribution centres	0	1	0	0	0	1	0	0
Farm to market transport	2	1	0	0	0	0	0	1
Farmer field schools	3	3	0	0	2	0	0	1
Fertiliser access	2	3	0	0	1	0	0	0
Food safety regulations	0	0	0	0	0	0	0	1
Fortification	0	0	0	2	0	0	3	0

Government price manipulations (excl. tariffs)	1	0	0	0	0	0	0	1
Improved seeds	2	1	0	0	1	0	0	0
Land markets and management	1	2	1	0	0	0	0	0
Market support	1	0	0	0	0	0	0	0
On farm, post-harvest processing	0	1	0	0	0	1	0	0
Other Agricultural education programs	4	3	0	1	0	0	0	1
Other agricultural inputs	2	2	0	1	0	0	0	1
Other production system improvements	1	1	0	0	0	0	0	1
Pesticide/herbicide access	0	1	0	0	1	0	0	0
Professional services (dieticians/nurses)	0	0	0	2	0	0	0	0
Provision of free or reduced-cost access to livestock	1	1	0	0	0	0	0	0
Provision or use of supplements	0	0	0	6	0	0	1	0
Trade regulations	2	2	1	0	1	0	0	0
Water access/management	6	6	1	0	1	0	0	2
Women's empowerment efforts	2	1	0	1	0	0	0	0
Total unique count[1]	**40**	**32**	**4**	**15**	**10**	**1**	**5**	**7**

1. Total number of unique studies may be less than the sum of a column as a single study can be in several cells.

Table A4.2 Outcome set 2, interventions repeated

Interventions	Outcomes: Food affordability / availability	Food distribution	Food safety	Intrinsic motivators	Micronutrient status	Women's empowerment	Total unique count[1]
Agricultural credit / savings	1	0	0	0	1	0	3
Agricultural extension programs	3	0	0	0	0	2	11
Agricultural information / guidance	0	0	0	0	0	0	1
Agricultural insurance	0	0	0	0	0	0	3
Cash-for-food programs	1	0	0	0	0	0	2
Classes	1	0	0	1	0	0	3
Cold chain initiatives	0	0	0	0	0	0	1
Composting	0	0	0	0	0	1	1
Contract farming	0	0	0	0	0	0	1
Direct provision of foods	2	0	0	0	1	1	8
Distribution centres	1	0	0	0	0	0	1
Farm to market transport	0	0	0	0	0	0	2
Farmer field schools	2	0	0	1	0	1	3
Fertiliser access	1	0	0	0	0	0	3
Food safety regulations	0	0	1	0	0	0	1
Fortification	0	0	0	0	2	0	4
Government price manipulations (excl. tariffs)	0	0	0	0	0	0	1
Improved seeds	0	0	0	0	0	0	2

Land markets and management	0	0	0	0	0	0	3
Market support	0	0	0	0	0	1	1
On farm, post-harvest processing	1	0	0	0	0	0	1
Other Agricultural education programs	1	0	0	0	1	1	5
Other agricultural inputs	1	0	0	0	1	0	2
Other production system improvements	0	0	0	0	0	0	3
Pesticide/herbicide access	0	0	0	0	0	0	1
Professional services (dieticians/nurses)	0	0	0	1	0	0	3
Provision of free or reduced-cost access to livestock	1	0	0	0	0	0	1
Provision or use of supplements	0	0	0	0	1	0	8
Trade regulations	0	1	0	0	0	0	3
Water access/management	0	0	0	0	0	1	10
Women's empowerment efforts	0	0	0	0	0	2	3
Total unique count[2]	**9**	**1**	**1**	**3**	**5**	**7**	**78**

1. Total unique count is across Table A3.1 and A3.2. The total may be less than the sum of a single row because a single study can be in several cells
2. Total number of unique studies may be less than the sum of a column as a single study can be in several cells.

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

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

Table A5.1 Outcome set 1

		Outcomes							
		Economic	**Agricultural**	**Climate / environment**	**Anthropometric**	**Behaviour change**	**Bio nutritional**	**Developmental**	**Diet quality / adequacy**
Interventions	Behaviour change communication	0	0	0	0	0	0	0	0
	Food loss and waste management	0	0	0	0	0	0	0	0
	Food production	14	16	0	0	6	1	0	2
	Food provision / price reduction	1	0	0	1	2	0	0	0
	Food transport / storage	2	2	0	0	0	1	0	0
	Processing and packaging	0	1	0	1	0	1	1	0
	Quality and safety	0	0	0	0	0	0	0	0
	Women's empowerment in the food system	0	0	0	0	0	0	0	0
	Total unique count[1]	**16**	**17**	**0**	**2**	**8**	**1**	**1**	**2**

1. Total number of unique studies may be less than the sum of a column as a single study can be in several cells.

Table A5.2 Outcome set 2, interventions repeated

Interventions	Food affordability / availability	Food distribution	Food safety	Intrinsic motivators	Micronutrient status	Women's empowerment	Total unique count[1]
Behaviour change communication	0	0	0	1	0	0	1
Food loss and waste management	0	0	0	0	0	0	0
Food production	4	0	0	1	0	3	23
Food provision / price reduction	0	0	0	0	0	0	4
Food transport / storage	1	1	0	0	0	0	4
Processing and packaging	1	0	0	0	0	0	2
Quality and safety	0	0	0	0	0	0	0
Women's empowerment in the food system	0	0	0	0	0	0	0
Total unique count[2]	**4**	**1**	**0**	**2**	**0**	**3**	**31**

1. Total unique count is across Table A4.1 and A4.2. The total may be less than the sum of a single row because a single study can be in several cells
2. Total number of unique studies may be less than the sum of a column as a single study can be in several cells.

Appendix 6: Meta-regression and subgroup analysis

Using meta-regression and subgroup analysis, we consider if reported long-term effects from included studies vary by intervention type, geographic region, period of analysis, methods, or risk of bias. While subgroup analysis provides estimates of reported effects for each group, meta-regression allows us to determine if these effects are statistically different from one another. Generally, we find that, although effects are statistically significant for some subgroups and not others, these effects are not statistically different from one another. This apparently contradictory result can be observed when confidence intervals on average treatment effects are large and overlapping, with some including zero and others do not. Given that results are not statistically different from one another, the statistical significance of findings for some subgroups and not others should be interpreted cautiously.

A6.1 Income

Subgroup meta-analysis shows that the reported effect on income is positive and statistically significant for soil and water conservation interventions (k = 4) and agricultural insurance interventions (only based on one estimate from Li & Wang,[3] 2022; Table A6.1). Effects are not statistically significant for other subgroups: geographic region, period of analysis, use of controls, evaluation methods (except for Li & Wang, 2022 which used fixed effects), treatment effect estimated, or risk of bias.

Despite these statistically significant effects for specific subgroups, effect estimates do not vary by intervention type, region, study period, or risk of bias. We grouped the interventions into those considering agricultural education (k = 6), soil and water conservation (k = 4), and other interventions (provision of agricultural inputs, agricultural insurance, land titling, and sustainable certification programs, k = 5). Studies using difference-in-difference statistical approaches (k = 5) report smaller average effects than those using statistical matching (k = 8) or other methods (fixed effects and instrumental variables, k = 2). Effects measured as average treatment effects (k = 13) were statistically significantly greater than those measured as average treatment effect on the treated (k = 2).

[3] Work by Niu and colleagues (2022), Lv (2020), and Li and Wang (2022) is not presented in the main report because they are based on analysis conducted at the province level and, therefore, very small sample sizes (n = 31 based on 31 provinces in China). However, they are included in these analysis in order to be comprehensive.

Table A6.1: Moderator analysis results for the long-term effects of agricultural interventions on income

	Meta-regression			Subgroup meta-analysis		Number of reported effects
	Effect estimate	**Standard error**	**Reference group**	**Effect estimate**	**95% Confidence interval**	
Intervention type[1]						
Soil and water conservation interventions	*-0.37*	*0.48*	Effects from other interventions excluding agricultural education (n=5)	0.12**	0.02 to 0.21	4
Agricultural education interventions[2]	*-0.52*	*0.43*	Effects from other interventions excluding soil and water conservation (n=5)	-0.02	-0.13 to 0.09	6
Reference group: Effects from other interventions excluding soil and water conservation and agricultural education (n=5):						
Provision of agricultural input			N/A as reference group	0.20*	-0.03 to 0.42	1
Land titling				0.17	-0.06 to 0.41	1
Sustainable certification programs				-0.13*	-0.29 to 0.02	2
Agricultural insurance program				3.15***	2.29 to 4.00	1

			Geographic region[3]			
Asia	*0.35*	*0.41*	Effects from studies in Sub-Saharan Africa (n=7)	0.35	-0.09 to 0.78	6
Latin America and the Caribbean	*0*	*0.58*	Effects from studies in Sub-Saharan Africa (n=7)	0.07	-0.18 to 0.32	2
Sub-Saharan Africa			N/A as reference group	0.02	-0.06 to 0.11	7
			Period of analysis			
Period less than 16 years[4]	*0.18*	*0.46*	Effects with a period of 16 years or more	0.09	-0.04 to 0.23	12
Period of 16 years or more			N/A as reference group	0.04	-0.13 to 0.21	3
			Use of control variables			
Use control variable	*0.20*	*0.46*	Other effects (n=3)	0.1	-0.14 to 0.24	12
No use of control variable			N/A as reference group	0.01	-0.17 to 0.18	3
			Evaluation methods[5]			
Difference-in-differences	*-1.17***	*0.54*	Effects based on fixed effects and instrumental variable methods (n=2)	-0.04	-0.24 to 0.16	5
Statistical matching	*-1.0**	*0.52*	Effects based on fixed effects and instrumental variable methods (n=2)	0.07*	0 to 0.14	8

Reference group: Effects based on fixed effects and instrumental variable methods (n=2):						
Fixed effects			N/A as reference group	3.15***	2.29 to 4.00	1
Instrumental variable				-0.15*	-0.33 to 0.03	1
			Treatment effects			
Average treatment effect	*1.16***	*0.43*	Effects estimated as average treatment effects on the treated (n=13)	1.63	-0.28 to 4.55	2
Average treatment effect on the treated			N/A as reference group	0.03	-0.05 to 0.11	13
			Risk of bias			
High risk of bias	*0.01*	*0.74*	Effects with low risk of bias (n=1)	0.07	-0.05 to 0.20	14
Low risk of bias			N/A as reference group	0.17	-0.06 to 0.41	1

Note: p value * <0.1 ** <0.05 ***<0.01. The number of observations in the meta-regressions is always 15.

1. Meta-regression results here are estimated regressing two dummy variables on the calculated effect sizes. The first dummy variable takes the value 1 if the effect size is based on soil and water conservation interventions and 0 otherwise. The second dummy variable takes the value 1 if the effect size is based on agricultural education interventions and 0 otherwise. The reference group is the effect sizes from the group of other interventions (n=5).

2. Three effects included in the agricultural education subgroup originate from interventions which also include other types of interventions (activities linked to provision of credit).

3. Meta-regression results here are estimated regressing two dummy variables on the calculated effect sizes. The first dummy variable takes the value 1 if the effect size is from a study in Asia and 0 otherwise. The second dummy variable takes the value 1 if the effect size is from a study in Latin America and the Caribbean and 0 otherwise. The reference group is the effect sizes from studies in Sub-Saharan Africa (n=7).

4. 16 years is the mean time period considered by the included studies.

5. Meta-regression results here are estimated regressing two dummy variables on the calculated effect sizes. The first dummy variable takes the value 1 if the effect size is based on a statistical matching method and 0 otherwise. The second dummy variable takes the value 1 if the effect size is based on a difference-in-differences method and 0 otherwise. The reference group is the effect sizes from studies based on fixed effects and instrumental variable methods (n=2).

A6.2 Crop production

The reported effects are statistically significant for specific intervention types, for interventions in Latin America and the Caribbean, and for low-risk of bias estimates (Table A6.2). In addition, when the sample is restricted to studies considering periods of analysis less than the mean of the sample, 12 years, the effects are also positive and statistically significant ($\text{SMD} = 0.16; 95\%\ \text{CI: } 0.08 \text{ to } 0.25; \text{p} < 0.001;$ k = 10). However, the statistically insignificant effects of studies considering outcomes 12 or more years after intervention initiation may be related to the inclusion of both farmer certification interventions in this period (Akoyi & Maertens, 2018; Ibanez & Blackman, 2015). Both had statistically significant, negative effects. Reported effects based on estimations without any control variables are on average positive and statistically significant ($\text{SMD} = 0.29; 95\%\ \text{CI: } 0.15 \text{ to } 0.45; \text{p} < 0.001;$ k = 4). We find that the subgroup of estimates derived from difference-in-differences ($\text{SMD} = 0.07; 95\%\ \text{CI: } 0.02 \text{ to } 0.12; \text{p} = 0.006;$ k = 3) and statistical matching ($\text{SMD} = 0.22; 95\%\ \text{CI: } 0.05 \text{ to } 0.39; \text{p} = 0.01;$ k = 8) approaches are positive and statistically significant. Similarly, average treatment effect and intent-to-treat effect estimates are statistically significantly greater than zero ($\text{SMD} = 0.28; 95\%\ \text{CI: } 0.10 \text{ to } 0.46; \text{p} = 0.002$ k = 6; $\text{SMD} = 0.08; 95\%\ \text{CI: } 0.06 \text{ to } 0.11; \text{p} < 0.001;$ k = 3, respectively).

However, in the meta-regressions, reported effects are not statistically significantly different across intervention types, period of the analysis (less than 12 years relative to 12 years or more), region of interventions, methods, and risk of bias (high relative to low risk of bias). We grouped the interventions into those considering agricultural education (k = 6), soil and water conservation (k = 2), and other interventions (provision of agricultural inputs, agricultural insurance, land titling, sustainable certification programs, and contract farming interventions, k = 8). Agricultural education interventions and soil and water conservation interventions do not have statistically different reported effects relative to the group of other interventions. Reported effects from Latin America and the Caribbean (k = 5) and Asia (k = 2) are not statistically significantly different from those found in Sub-Saharan Africa (k = 9). We find similar results across the evaluation method used (difference-in-differences, k = 3; statistical matching, k = 8; instrumental variable, reference group, k = 5) and across the type of effect measured (average treatment effect, k = 6; Intent-to-treat effect, k = 3; average treatment effect on the treated, reference group, k = 7).

Table A5.2: Moderator analysis results for the long-term effects of agricultural interventions on income

	Meta-regression			Subgroup meta-analysis		Number of reported effects
	Effect estimate	**Standard error**	**Reference group**	**Effect estimate**	**95% Confidence interval**	
Intervention type[1]						
Soil and water conservation interventions	*-0.15*	*0.31*	Effects from other interventions excluding agricultural education (n=8)	-0.11	-0.29 to 0.07	2
Agricultural education interventions[2]	*0.12*	*0.21*	Effects from other interventions excluding soil and water conservation (n=8)	*0.16***	*0.09 to 0.22*	*6*
Reference group: Effects from other interventions excluding soil and water conservation and agricultural education (n=8):						
Provision of agricultural inputs			N/A as reference group	*0.37***	*0.15 to 0.66*	*1*
Land titling				0.18	-0.33 to 0.69	2
Sustainable certification programs				-0.45***	-0.76 to -0.14	2
Agricultural insurance				1.76***	0.87 to 2.64	1
Contract farming				0.07	-0.06 to 0.21	2

Geographic region[3]						
Asia	*0.46*	*0.39*	Effects from studies in Sub-Saharan Africa (n=9)	0.77	-1.08 to 2.62	2
Latin America and the Caribbean	*0.07*	*0.21*	Effects from studies in Sub-Saharan Africa (n=9)	0.12***	0.05 to 0.20	5
Sub-Saharan Africa	N/A as reference group			0.06	-0.13 to 0.26	9
Period of analysis						
Period less than 12 years[4]	*0.32*	*0.20*	Effects with a period of 12 years or more	0.16***	0.08 to 0.23	12
Period of 12 years or more	N/A as reference group			-0.12	-0.54 to 0.29	4
Use of control variables						
Use control variable	*-0.26*	*0.20*	Other effects (n=4)	0.03	-0.07 to 0.12	12
No use of control variable	N/A as reference group			0.29***	0.14 to 0.45	4
Evaluation method[5]						
Difference-in-differences	*-0.02*	*0.28*	Effects based on instrumental variable method (n=5)	0.07***	0.02 to 0.12	3
Statistical matching	0.22	*0.23*	Effects based on instrumental variable method (n=5)	*0.22***	0.05 to 0.39	8

Instrumental variable			N/A as reference group	0.08	-0.34 to 0.49	5
			Treatment effects[6]			
Average treatment effect	*0.34*	*0.20*	Effects estimated as average treatment effects on the treated (n=7)	0.28***	0.10 to 0.46	6
Intent-to-treat effect	*0.05*	*0.25*	Effects estimated as average treatment effects on the treated (n=7)	0.08***	0.06 to 0.11	3
Average treatment effect on the treated			N/A as reference group	-0.01	-0.28 to 0.25	7
			Risk of bias			
High risk of bias	*0.11*	*0.25*	Effects with low risk of bias (n=3)	0.14	-0.04 to 0.31	13
Low risk of bias			N/A as reference group	0.08***	0.06 to 0.11	3

Note: p value * <0.1 ** <0.05 ***<0.01. The number of observations in the meta-regressions is always 16.

1. Meta-regression results here are estimated regressing two dummy variables on the calculated effect sizes. The first dummy variable takes the value 1 if the effect size is based on soil and water conservation interventions and 0 otherwise. The second dummy variable takes the value 1 if the effect size is based on agricultural education interventions and 0 otherwise. The reference group is the effect sizes from the group of other interventions (n=8).

2. Three effects included in the agricultural education subgroup originate from interventions which also include other types of interventions (two with the provision of agricultural inputs and one with soil and water conversation).

3. Meta-regression results here are estimated regressing two dummy variables on the calculated effect sizes. The first dummy variable takes the value 1 if the effect size is from a study in Asia and 0 otherwise. The second dummy variable takes the value 1 if the effect size is from a study in Latin America and the Caribbean and 0 otherwise. The reference group is the effect sizes from studies in Sub-Saharan Africa (n=9).

4. 12 years is the mean time period considered by the included studies.

5. Meta-regression results here are estimated regressing two dummy variables on the calculated effect sizes. The first dummy variable takes the value 1 if the effect size is based on a statistical matching method and 0 otherwise. The second dummy variable takes the value 1 if the effect size is based on a difference-in-differences method and 0 otherwise. The reference group is the effect sizes from studies based on an instrumental variable method (n=5).

6. Meta-regression results here are estimated regressing two dummy variables on the calculated effect sizes. The first dummy variable takes the value 1 if the effect size was estimated as an average treatment effect and 0 otherwise. The second dummy variable takes the value 1 if the effect size was estimated as an intent-to-treat effect and 0 otherwise. The reference group is the effect sizes estimated as average treatment effect on the treated (n=9).

Appendix Table 1: Summary of included studies

First author	Year	Title	Country	Evaluation method	Intervention type	Outcome type	Intervention description
Abdoulaye	2013	A matching approach to analyze the impact of new agricultural technologies: Productivity and technical efficiency in Niger	Niger	Statistical matching	Provision of agricultural inputs and agricultural education	Production	Improved sorghum technology packages providing a moderate level of inorganic fertilizers, improved sorghum cultivar, fungicide and agronomic recommendations.
Abebe	2014	The impact of soil and water conservation program on the income and productivity of farm households in Adama District, Ethiopia	Ethiopia	Statistical matching	Soil and water conservation	Production Income	Soil and water conservation intervention providing educational and in-kind support for conserving, developing and rehabilitating degraded agricultural lands to increase productivity and food security.
Akoyi	2018	Walk the talk: Private sustainability standards in the Ugandan coffee sector	Uganda	Instrumental variable	Certification	Production Income	The Triple Utz-Rainforest Alliance-Common Code of Conduct for Coffee scheme provided private sustainability certificates to coffee growers, high-quality inputs, and education to increase biodiversity and natural resource conservation. The Double Fairtrade-Organic scheme provided certificates focusing on empowering farmers. The Rainforest Alliance and Organic scheme

							meant to increase biodiversity and natural resource conservation. The interventions provided farmer-to-farmer input sharing and extension services.
Baiyegunhi	2019	Impact of outsourced agricultural extension program on smallholder farmers' net farm income in Msinga, KwaZulu-Natal, South Africa	South Africa	Statistical matching	Agricultural education	Income	Outsourced extension program disseminating information, raising awareness and training farmers to improve technological choices and support diversification.
Besley	2015	Long-run impacts of land regulation: Evidence from tenancy reform in India	India	Statistical matching	Land titling	Income	Land tenancy reform laws provided land rights, such as minimum terms of lease; the right of purchase of non-resumable lands; the right to mortgage land for credit; mandatory recording of tenant names; limitations on the landlord's right of resumption; caps on rent; temporary protection against eviction or prohibition of eviction.
Chankrajang	2015	Partial land rights and agricultural outcomes: Evidence from Thailand	Thailand	Instrumental variable	Land titling	Production Environment	Partial property titling (SPK 4-01) intervention provided upgraded rights and granted formal titles to farmers.

Datta	2014	Evaluating impacts of watershed development program on agricultural productivity, income, and livelihood in Bhalki Watershed of Bardhaman District, West Bengal	India	Statistical matching	Soil and water conservation	Income Environment	Watershed development intervention lead by community members who organized into self-help groups, planting trees and attending meetings to create awareness about the importance of soil and water conservation. A microfinance institution for the watershed committee was developed and more soil and water conservation structures (mostly reservoirs) were created along with income generating activities in a second phase.
De Los Santos-Montero	2017	Productivity effects and natural resource management: Econometric evidence from POSAF-II in Nicaragua	Nicaragua	Statistical matching	Agricultural education	Production	PROSAF-II encouraged the adoption of sustainable farming technologies and practices (ex. terracing and integrated pest management) primarily in agroforestry and forest management systems for individual farms. It focused on forest planting and regeneration.
Deschamps-Laporte	2013	The impact of extension services on farming households in Western Kenya: A propensity score approach	Kenya	Statistical matching	Agricultural education	Production Income	The National Agriculture and Livestock Extension Programme provided institutional set-up for farmers, extension services, promotion of technical packages, collaboration with other NGOs, and gender equity work.
Funsani	2016	Farmer input support programme and household income: Lessons	Zambia	Statistical matching	Provision of agricultural inputs	Production Income	National program providing reduced cost fertilizer and maize seeds to small scale farmers.

		from Zambia's Southern Province					
Gibbons	2016	Money for Wine? Complementarities in the provision of private and public goods to wine producers	Argentina	Difference-in-difference	Provision of agricultural inputs and agricultural education	Production	PROVIAR provided specialists advise on the implementation of good agricultural and manufacturing practices, commercial development of markets and the availability of agricultural insurance. It also provided hail-resistance nets, wood, wire, and improvement of irrigation. PROSAP provided irrigation infrastructure and financed initiatives that support the competitiveness of small- and medium-producers. Extension program providing a combination of PROVIAR and PROSAP.
IFAD	2018	Impact assessment report: The Agricultural Sector Development Programme-Livestock and the Agriculture Service Support Programme, Tanzania	Tanzania	Statistical matching	Agricultural education	Income	The Agricultural Sector Development Programme – Livestock and Agricultural Service Support Programme provided capacity building and training activities through farmer field schools.
Ibanez	2015	Environmental and economic impacts of growing certified organic coffee in Colombia	Colombia	Difference-in-difference	Certification	Production Income	Organic coffee certification program providing certificates to those who discontinued the use of chemical inputs and adopted conservation and pollution prevention practices.

Kumar	2011	Access, adoption, and diffusion: understanding the long-term impacts of improved vegetable and fish technologies in Bangladesh	Bangladesh	Difference-in-difference	Credit and agricultural education	Income	Fish intervention by Banchte Shekha providing long-term leases of fish ponds managed by groups of women. The Mymensingh Aquaculture Extension Programme provided training to better off households and training with credit to poor households who owned individual fish ponds. Vegetable intervention by the Gono Kallayan Trust providing credit and training in small-scale vegetable growers promoting improved vegetable varieties.
Li	2021	Analysis on the effect of farmer income of policy based agricultural insurance	China	Fixed	Agricultural insurance	Income	Intervention increasing a policy-oriented agricultural insurance subsidy.
Lv	2020	Empirical analysis on the effect of agricultural insurance on production: Based on panel data of 31 provinces and cities in China from 2008 to 2018	China	Instrumental variable	Agricultural insurance	Production	Intervention providing agricultural insurance.
Maia	2016	Impact of microcredit on small-farm agricultural production:	Brazil	Statistical matching	Credit	Production	The National Program for Strengthening Family Farming provided production financing, infrastructure and municipal service financing, training and professionalization, research and extension funding. This work only evaluated the effect of accessing the microcredit.

		Evidence from Brazil					
Melesse	2015	Does land registration and certification boost farm productivity? Evidence from Ethiopia	Ethiopia	Statistical matching	Land titling	Production	The Ethiopian land registration and certification program provided every rightful holder a certificate of usufructs.
Niu	2022	Agricultural insurance and agricultural fertilizer non-point source pollution: evidence from China's policy-based agricultural insurance pilot	China	Difference-in-difference	Agricultural insurance	Environment	Pilot of policy-based agricultural insurance providing low-cost agricultural insurance.
Romero	2021	Can a territorial use right for fisheries management make a difference for fishing communities?	Chile	Difference-in-difference	Land conservation	Income	Artisanal fishers' organizations were granted exclusive use rights through the Management and Exploitation Areas for Benthic Recources. They had the responsibility of planning extraction, surveillance, and biological control. The program aimed to encourage sustainable fishing practices.
Ruml	2020	Effects of marketing contracts and resource-providing contracts in the African small farm sector: Insights	Ghana	Instrumental variable	Contract farming	Production	Resource contract: agreements between a company and farmer where the company provides cultivation assistance which included planting material, tools, machinery, and agrochemical inputs provided on credit. Farmers were obligated to sell all the fruit bunches harvested on the contracted plot to the company.

		from oil palm production in Ghana					Marketing contract: agreements between a company and farmers, in which an annual fixed price and regular pick-ups of the harvested produce are specified
World Bank	2009	Republic of Niger impacts of sustainable land management programs on land management and poverty in Niger	Niger	Statistical matching	Soil and water conservation	Production	Natural resources management project providing activities supporting sustainable land management, largely related to tree planting.

Appendix Table 2: Effects of included studies

First author	Year	Country	Intervention type	Outcome	Standardized effect estimate (Confidence Interval)	Interpretation
Abdoulaye	2013	Niger	Provision of agricultural inputs and agricultural education	Sorghum yield (kilogramme/ hectare)	0.27 (0.12;0.42)	Increase in production
			Provision of agricultural inputs and agricultural education (subgroup effect: Effect in 2012)		0.36 (0.10;0.62)	Increase in production
			Provision of agricultural inputs and agricultural education (subgroup effect: Effect in 2011)		0.36 (0.10;0.62)	Increase in production
			Provision of agricultural inputs and agricultural education (subgroup effect: Effect in 2010)		0.56 (0.30;0.82)	Increase in production
Abebe	2014	Ethiopia	Soil and water conservation	Teff yield (kilogrammes/ hectare)	0.05 (-0.27;0.37)	No change in production
				Total income in the local currency	0.03 (-0.31;0.37)	No change in income
Akoyi	2018	Uganda	Certification (Utz-Rainforest Alliance-4C,	Total income per capita in the local currency	0.20 (0.01;0.39)	Increase in income

			8-year effect)			
			Certification (Fairtrade-Organic)	Total income per capita in the local currency	-0.15 (-0.33;0.03)	No change in income
			Certification (Utz-Rainforest Alliance-4C, 8 year effect)	Coffee yield (kilogramme/hectare)	1.10 (0.90;1.30)	Increase in production
			Certification (Fairtrade-Organic)	Coffee yield (kilogramme/hectare)	-0.59 (-0.78;-0.41)	Decrease in production
Baiyegunhi	2019	South Africa	Agricultural education	Net farm income in the local currency	0.23 (0;0.45)	Increase in income
Besley	2015	India	Land titling	Wages in the local currency	0.17 (-0.06;0.41)	No change in income
Chankrajang	2015	Thailand	Land titling	Major rice yield (kilogrammes/ hectare)	-0.13 (-0.62;0.35)	No change in primary production
				Second rice yield (kilogrammes/ hectare)[1]	0.69 (0.18;1.21)	Increase in secondary production
				Share of land with acid soil	0.28 (-0.23;0.79)	No change in the environment
Datta	2014	India	Soil and water conservation	Crop income per hectare in the local currency	0.23 (-0.03;0.48)	No change in income

				Cropping intensity (% change)	0.04 (-0.21;0.29)	No change in the environment
De Los Santos-Montero	2017	Nicaragua	Agricultural education (agroforestry)	Value of production in USD per hectare	0.47 (0.26;0.69)	Increase in production
			Agricultural education (forestry)		0.26 (0.10;0.42)	Increase in production
Deschamps-Laporte	2013	Kenya	Agricultural education	Maize yield (kilogrammes/ acre)	0.11 (-0.04;0.27)	No change in production
				Total crop revenue in the local currency per acre	-0.07 (-0.22;0.08)	No change in income
Funsani	2016	Zambia	Provision of agricultural inputs	Average income 2013-2015 in the local currency	0.20 (-0.03;0.42)	No change in income
			Provision of agricultural inputs	Average maize yield 2013-2015 (tons/hectare)	0.37 (0.15;0.60)	Increase in production
			Provision of agricultural inputs	Average maize yield 2014-2015 (tons/hectare)[1]	0.36 (0.14;0.59)	Increase in production
			Provision of agricultural inputs	Average maize yield 2013-2014 (tons/hectare)[1]	0.28 (0.05;0.50)	Increase in production

Gibbons	2016	Argentina	Provision of agricultural inputs and agricultural education (subgroup effect: Both intervention)	Grape yield (log, kilogrammes/hectare)	0.06 (0.03;0.09)	Increase in production
			Water conservation and agricultural education (subgroup effect: Both intervention)		0.07 (0.04;0.10)	Increase in production
			Provision of agricultural inputs and agricultural education		0.09 (0.07;0.12)	Increase in production
			Water conservation and agricultural education		0.07 (0.05;0.10)	Increase in production
IFAD	2018	Tanzania	Agricultural education	Total income in the local currency (log)	0.03 (-0.06;0.12)	No change in income
Ibanez	2015	Colombia	Certification	Income from coffee production in thousands in the local currency	-0.10 (-0.36;0.17)	No change in income
				Maize yield (kilogrammes/acre)	-0.27 (-0.54;0)	Decrease in production
Kumar	2011	Bangladesh	Credit and agricultural education (long-term leases of fish ponds)	Total income in the local currency per capita (log)	-0.14 (-0.44;0.16)	No change in income
			Credit and agricultural education	Total income in the local currency per capita (log)	-0.01 (-0.31;0.29)	No change in income

			(HH owned individual fish ponds)			
			Credit and agricultural education (improved vegetable varieties)	Total income in the local currency per capita (log)	-0.28 (-0.59;0.03)	No change in income
			Credit and agricultural education (improved vegetable varieties, 2 year effect)	Total income in the local currency per capita (log)	0.22 (-0.08;0.53)	No change in income
			Credit and agricultural education (long-term leases of fish ponds, 3 year effect)	Total income in the local currency per capita (log)	0.36 (0.06;0.65)	Increase in income
			Credit and agricultural education (HH owned individual fish ponds, 6 year effect)	Total income in the local currency per capita (log)	-0.34 (-0.65;-0.04)	Decrease in income
Li	2021	China	Agricultural insurance		3.15 (2.29;4.00)	Increase in income
			Agricultural insurance (subgroup effect: Severe drought area)	Net business income per capita	5.28 (4.16;6.40)	Increase in income
			Agricultural insurance (subgroup effect: Second lowest wealth quantile)		4.23 (3.25;5.21)	Increase in income

			Agricultural insurance (subgroup effect: Middle wealth quantile)		10.02 (8.19;11.85)	Increase in income
			Agricultural insurance (subgroup effect: Second highest wealth quantile)		6.39 (5.12;7.67)	Increase in income
			Agricultural insurance (subgroup effect: Highest wealth quantile)		4.49 (3.48;5.51)	Increase in income
Lv	2020	China	Agricultural insurance	Value of agricultural production in the local currency per capita	1.76 (0.87;2.64)	Increase in production
Maia	2016	Brazil	Credit (subgroup effect: Northern region)	Value of crop production in the local currency (log)	0.03 (0.02;0.03)	Increase in production
			Credit (subgroup effect: Northern-Eastern region)		0.08 (0.08;0.09)	Increase in production
			Credit (subgroup effect: South-Eastern region)		0.07 (0.07;0.08)	Increase in production
			Credit (subgroup effect: Southern region)		0.08 (0.08;0.09)	Increase in production
			Credit (subgroup effect: Centre-West region)		0.06 (0.05;0.07)	Increase in production
Melesse	2015	Ethiopia	Land titling		0.40 (0.22;0.58)	Increase in production

				Value added from agriculture in the local currency per timad (log)	0.38 (0.21;0.56)[2]	Increase in production
Niu	2022	China	Agricultural insurance	Pollution intensity of the agricultural fertilizer non-point source[3]	-0.85 (-1.65;-0.06)	Increase in pollution
			Agricultural insurance (subgroup effect: High-risk area of disasters)		-0.70 (-1.40;0)	Increase in pollution
Romero	2021	Chile	Soil conservation	Total income in the local currency (log)	0.17 (0.09;0.25)	Increase in income
Ruml	2020	Ghana	Contract farm (resources contract)	Palm oil yield (tons/acre)	0.14 (-0.04;0.32)	No change in production
			Contract farming (Resources contract) (subgroup effect: Small plot size)	Palm oil yield (tons/acre)	0.75 (0.44;1.06)	Increase in production
			Contract farming (Resources contract) (subgroup effect: Medium plot size)	Palm oil yield (tons/acre)	1.36 (1.05;1.67)	Increase in production
			Contract farming (Resources contract) (subgroup effect: Large plot size)	Palm oil yield (tons/acre)	0.06 (-0.29;0.40)	No change in production
			Contract farm (Marketing contract)	Palm oil yield (tons/acre)	0.00 (-0.17;0.18)	No change in production

			Contract farming (Marketing contract) (subgroup effect: Small plot size)	Palm oil yield (tons/acre)	-0.06 (-0.36;0.24)	No change in production
			Contract farming (Marketing contract) (subgroup effect: Medium plot size)	Palm oil yield (tons/acre)	0.06 (-0.23;0.35)	No change in production
			Contract farming (Marketing contract) (subgroup effect: Large plot size)	Palm oil yield (tons/acre)	-0.23 (-0.58;0.11)	No change in production
World Bank	2009	Niger	Soil and water conservation	Value of crop production in USD	-0.16 (-0.28;-0.04)	Decrease in production
				Total income in USD per capita	0.03 (-0.09;0.15)	No change in income

1. Secondary outcome, not used in main analysis
2. Secondary, intention-to-treat analysis
3. All values are standardized so that negative is reflects a negative outcome to allow for comparability across studies. This indicates an increase in pollution

Department for Evaluation

norad.no/evaluation