USING INTERNAL EVALUATIONS TO MEASURE ORGANIZATIONAL IMPACT

A meta-analysis of Oxfam's women's empowerment projects – *working paper*

Simone Lombardini
Oxfam GB, Oxford, United Kingdom

Kristen McCollum
International Fund for Agricultural Development, Rome, Italy

This paper presents the results of a meta-analysis examining the overall impact of women's empowerment projects evaluated as part of Oxfam GB's Effectiveness Reviews. Results show a positive and significant impact on the Women's Empowerment Index and mixed results with its individual indicators. We found a statistically significant effect on opinions on women's economic role and their ability to participate and have influence in the community. We did not find evidence of overall changes in power within the household nor with the share of household income. The meta-analysis also found statistically significant overall effects where the individual studies were too under-powered to detect impact. This paper provides an example of how using meta-analysis in the presence of a robust organisational global evaluation framework can enable evidence-based learning, organizational accountability and better programme implementation.

Oxfam Discussion Papers

Oxfam Discussion Papers are written to contribute to public debate and to invite feedback on development and humanitarian policy issues. They are 'work in progress' documents, and do not necessarily constitute final publications or reflect Oxfam policy positions. The views and recommendations expressed are those of the author and not necessarily those of Oxfam.

This is an Accepted Manuscript of an article published by Taylor & Francis Group in Journal of Development Effectiveness on 27/09/2017, available online:
http://dx.doi.org/10.1080/19439342.2017.1377750

For more information, or to comment on this paper, email slombardini@oxfam.org.uk

OXFAM

1 INTRODUCTION

Every year since 2011, Oxfam GB has conducted rigorous evaluations (known as Effectiveness Reviews) on samples of randomly selected mature projects to help the organization understand and evidence whether its work is resulting in positive change. Development projects implemented at an individual, household or community level have been assessed through the use of ex-post quasi-experimental impact evaluation techniques by collecting individual or household survey data from a sample of project participants as well as from a comparison group. The two groups are then analysed using propensity-score matching and multivariate regression to assess the impact of the project and foster project learning. Oxfam GB has committed to publishing all the evaluations, regardless of whether the results are positive or negative. The number of evaluations conducted in a given year is not big enough to draw generalizable conclusions on the overall impact of the organization. However, the accumulation of evaluations of projects selected following a random selection process, avoids the risk of implicitly or explicitly evaluating only the best performing projects, and provides a realistic overview of Oxfam's impact (Hutchings 2014; Hughes and Hutchings 2011).

Meanwhile, in the wider arena of international development research, meta-analysis and systematic reviews are becoming an important tool to promote transparency and strengthen the available evidence on development interventions. Meta-analysis, or 'the statistical combination of results from two or more separate studies', allows researchers to increase the power and precision of research on specific interventions, and to answer additional questions not posed in the original evaluations (Higgins and Green 2011). In cases where individual studies have sample sizes too small to detect statistically significant results, meta-analysis can increase statistical power by combining studies, thus at times revealing findings which were previously undetectable (Higgins and Green 2011; Duvendack et al. 2012). Through the process of a systematic review, researchers can then gather all of the available evidence on intervention and conduct a meta-analysis to have a broader, more thorough, and less biased understanding of the impact of that intervention. Finally, systematic reviews can help to identify future research priorities by identifying areas lacking sufficient evidence, or diagnosing weaknesses in methods or measurement (Mallett et al. 2012).

This paper combines the effect sizes from all the Effectiveness Reviews conducted under the thematic area of Women's Empowerment since 2011, achieving the following goals.

First, the paper assesses and measures the overall impact of Oxfam projects aiming to achieve women's empowerment. It does so by investigating the aggregate projects' impact on a multidimensional index Oxfam has developed for measuring women's empowerment, as well as investigating the most commonly used empowerment indicators. Second, it investigates heterogeneity in the projects' impacts, and explores whether projects with certain characteristics are likely to present a bigger impact. Third, the research assesses the validity of the measurement tools employed in measuring women's empowerment. Fourth, it explores whether meta-analysing the treatment effects in these evaluations allows for detecting aggregated project effects that did not show up in individual evaluations.

To conclude, the paper aims to promote the use of meta-analysis as a tool that can support evidence-based learning, and improvements to programmes by testing projects' assumptions and shedding light on areas that require further improvement and attention, as well as a tool to communicate organizational impact and enable organizational accountability. The paper is structured as follows. Section 2 describes the Effectiveness Review project – how projects are selected and why that is important for the meta-analysis – and presents the measurement approach for women's empowerment. Section 3 describes the protocol used for the meta-analysis, the choice of the effect sizes and an assessment of the risk of bias for the studies included in the analysis. Section 4 presents the results, and Section 5 concludes.

2 OXFAM'S EFFECTIVENESS REVIEWS ON WOMEN'S EMPOWERMENT

2.1 WHAT EFFECTIVENESS REVIEWS ARE, AND HOW THE PROJECTS ARE SELECTED

Since 2011, under Oxfam GB's Global Performance Framework, Oxfam GB has been selecting a sample of mature or closing projects for rigorous evaluation under six thematic areas: Humanitarian Response, Accountability, Livelihood Support, Resilience, Women's Empowerment, and Good Governance. All Oxfam GB projects with a budget greater than £200,000 and that have been implemented for at least 2.5 years are eligible to be selected for the Effectiveness Reviews (ERs). A random sample of projects is selected for each category each year.

While evaluations conducted under the first two thematic areas (Humanitarian Response and Accountability) consider the degree to which interventions meet agreed standards; evaluations carried out under the final four thematic areas (Livelihood Support, Resilience, Women's Empowerment and Good Governance) aim to assess project impact.

The projects considered in this meta-analysis are the projects selected and evaluated under the thematic area of Women's Empowerment.[1] All the projects share the same goal of contributing to greater women's empowerment, however, as Oxfam is a rights-based organization, the design and the implementation of these projects varies significantly.

Table 2.1 describes in more detail each project evaluated and included in this meta-analysis. It is rarely the case for these projects to constitute a single set of specific interventions. In fact, the opposite it is usually true. Projects are often larger and more complex than the Effectiveness Review is able to assess, with the evaluations frequently focusing on specific components of the project. Table 2.2 attempts to provide a visual summary of the activities and components implemented by each project. When reading this table, the reader should be aware of two caveats. First, projects conducted a broader set of activities than shown here. What is represented in Table 2.2 are only the activities that were evaluated. Second, the table does not account for the scale or intensity of implementation of each component. It does not represent whether each component was a one-off intervention or the main focus of the project.

Table 2.1: Projects included in the meta-analysis

Country (year)	Title Project – Short project description	Project Type – Women's Economic Empowerment
Philippines (2011)	The evaluation investigates the impact of the project 'Sustainable Livelihoods Mindanao Project' on women's empowerment indicators, focusing on the component implemented by Oxfam's partner Paglilingkod Batas Pangkapatiran Foundation Incorporated (PBPF). The project seeks to strengthen community governance, improve household food security, and empower women among a group of indigenous peoples. The project started in 2007, and the evaluation was conducted in May 2012.	Yes
Guatemala (2011)	The evaluation investigates the impact of the project 'Guatemala Highlands Value Chain Development Alliance', together with two other projects that provided support to members of producer associations in the Department of Sololá. The primary objectives of these projects have been to improve household food security and strengthen agricultural livelihoods. Project activities included: agricultural extension support, distribution of fertilizer, training on the cultivation of kitchen gardens, provision of productive infrastructure to associations, and support in establishing linkages to private-sector vegetable exporters. Additional activities carried out specifically with female members of the associations, aimed to strengthen the position of women within the associations and within their households. Activities initially started in May 2010, and the evaluation was conducted in February 2012.	Yes
Zambia (2011)	Zambia's 'Copperbelt Livelihoods Project' was implemented by Oxfam's local partner, the Sustainable Agriculture Programme (SAP). It targeted 1,000 small-scale farmers (60% of whom are women) living in ten villages in Kitwe district of Zambia's Copperbelt Province. In addition to empowering women, the project sought to bolster household income and food security and reduce vulnerability through the provision of agricultural inputs and increasing market access. The project started in 2009, and the evaluation was conducted in October 2011.	Yes
Honduras (2012)	The project 'Linking APROALCE to Local, National and International Markets' is the latest in a series of projects carried out since 1998 by Oxfam in partnership with the Organización de Dessarrollo de Corquín (ODECO). The project established and supported community banks, with the intention of providing community members with a source of credit for productive investments. The evaluation was conducted in September 2012.	Yes
Nigeria (2012)	Nigeria's 'Improving Women's Leadership and Effectiveness in Agricultural Governance' project aims to increase women's leadership and participation in agricultural decision-making and governance through building women's skills and capacity in improved production techniques and by influencing local government and community leadership structures to enable greater involvement of women. The evaluation focused on the activities implemented by the local partner organisation Justice Development and Peace Commission (JDPC).	No
Malawi (2012)	The project 'Promoting Sustainable Livelihoods for Women & Vulnerable Groups' was implemented between 2009 and 2013 in collaboration with the Centre for Alternatives for Victimised Women and Children (CAVWOC). It aimed at empowering vulnerable people – especially women – through engagement in production and facilitating access to markets. The project's main activities involved facilitating the creation and capacity-building of village savings and loans (VSL) groups and raising awareness and reducing the acceptability of gender-based violence in the project areas.	Yes
Yemen (2012)	The 'Safe Age of Marriage and Women's Economic Empowerment' project in Yemen aimed to raise awareness about the negative impacts of early marriage and to reduce its prevalence. The activities evaluated aimed at promoting opportunities for self-employment for women in rural communities by providing micro-credit and business skills training.	Yes

	Activities have been carried out since 2006, and the evaluation took place in March 2013.	
Sierra Leone (2012)	The project 'Strengthening and Linking Women-Led Efforts to Promote Women's Property and Literacy Rights' in Sierra Leone was managed by a consortium of partners and implemented in four districts of Sierra Leone. The project conducted trainings on women's rights issues, leadership skills and advocacy skills with women leaders. These women were then supported in carrying out awareness-raising and advocacy on women's property and literacy rights within their communities.	No
Ethiopia (2013)	The project 'Facilitating Access to Financial Services for Women Beekeepers in Ethiopia' was carried out by Oxfam together with the Zembaba Bees Product Development and Marketing Cooperatives Union between May 2010 and April 2013. The project aimed at strengthening women's livelihoods and promoting empowerment, especially by facilitating the formation of women's self-help groups.	Yes
Rwanda (2013)	The project 'Women's Economic Leadership through Horticulture Planting-Material Business' was implemented between August 2011 and March 2014 by Oxfam in conjunction with Duterimbere. The project aimed to strengthen women's capacity for engaging in the production of pineapple and enhance women's socio-economic status at household and community level by providing finance and business services.	Yes
Indonesia (2013)	The project 'Papua Women's Empowerment' aimed to improve women's awareness and participation in the decision-making and implementation of a large-scale national community-driven development programme called PNPM Mandiri/RESPEK. The grant programme provided funding to women's groups to organize women's capacity-building activities. In addition, the project provided support to women's groups in designing project proposals and running group activities. The project was implemented between May 2009 and April 2013, and the evaluation took place in November 2013.	No
Uganda (2014)	The project 'Piloting Gender Sensitive Livelihoods in Karamoja' was implemented by Oxfam and partner organizations between July 2011 and March 2014. The project's goals were to improve the livelihoods of poor women through the support of enterprises and to reduce violence against women (VAW) by promoting awareness and influencing attitude and behavioural change using a Gender Action Learning System (GALS) approach.	Yes
Pakistan (2014)	The aim of the project 'Empowering Small Producers, especially Women, in the Dairy Sector' was to improve livelihoods opportunities, and increase income and employment, as well as raising women's empowerment by improving their economic leadership in the dairy sector.	Yes
Lebanon (2014)	The project 'Women's Access to Justice in the MENA Region' was implemented in three countries: Lebanon, Jordan and Iraq, with a multitude of interventions at individual, community and institutional level. The project started in May 2011 and finished with no-cost extension in July 2014. The evaluation focused on the impact of activities conducted in Lebanon at individual and community level, which included: awareness-raising sessions; free legal consultation to women who decided to bring their case to court; and the provision of awareness-raising sessions to men's forums and community leaders.	No
Armenia (2015)	The objectives of the project 'Women's Economic Empowerment in Rural Communities of Vayots Dzor Region' are increasing household income and promoting women's economic empowerment. The project was implemented from April 2011 to March 2013 and evaluated in 2015 more than two years after its conclusion.	Yes
Mali (2015)	The overall objective of the project 'Girls CAN – Promoting Secondary Education for Girls in West Africa' was to promote the successful transition rate of adolescent girls from primary to secondary school.	No

Table 2.2: Projects' interventions (under analysis)

		Interventions explicitly targeting women				Supporting agricultural interventions		
Country (year)	WEE?	Women's tech or leadership training	Community gender discussions/trainings	Women's productive associations	Savings/credit to women	Producer → commerce strengthening	Ag extension/inputs	Other
Philippines (2011)	Yes	✓	✓				✓	Husband/wife production planning
Guatemala (2011)	Yes	✓		✓		✓	✓	
Zambia (2011)	Yes	✓	✓				✓	
Honduras (2012)	Yes				✓	✓		
Nigeria (2012)	No	✓	✓	✓				Advocacy at local level.
Malawi (2012)	Yes	✓	✓		✓		✓	
Yemen (2012)	Yes	✓	✓		✓			
Sierra Leone (2012)	No	✓	✓					Literacy training
Ethiopia (2013)	Yes	✓		✓	✓	✓	✓	
Rwanda (2013)	Yes	✓		✓	✓		✓	
Indonesia (2013)	No	✓	✓	✓				
Uganda (2014)	Yes	✓	✓	✓	✓			Husband/wife production planning
Pakistan (2014)	Yes	✓		✓		✓		
Lebanon (2014)	No	✓	✓					Legal support
Armenia (2015)	Yes	✓		✓	✓	✓	✓	
Mali (2015)	No	✓	✓		✓			Girls' secondary education

Table 2.3: Indicators of women's empowerment found in each study

Country	Year	Women's Empowerment Index	Opinion on Women's Economic Role	Income Share	Household Decision-Making	Influence in Community	Experience of Violence
Philippines	2011				✔	✔	
Guatemala	2011		✔		✔		
Zambia	2011				✔	✔	
Honduras	2012	✔	✔	✔	✔	✔	
Nigeria	2012	✔			✔	✔	
Malawi	2012	✔	✔	✔	✔	✔	✔
Yemen	2012	✔	✔	✔	✔	✔	✔
Sierra Leone	2012	✔	✔	✔	✔	✔	✔
Ethiopia	2013	✔	✔	✔	✔	✔	✔
Rwanda	2013	✔	✔	✔	✔	✔	
Indonesia	2013	✔	✔	✔	✔	✔	
Uganda	2014	✔		✔	✔	✔	✔
Pakistan	2014	✔		✔	✔		✔
Lebanon	2014	✔		✔	✔		✔
Mali	2015	✔					✔
Armenia	2015	✔	✔	✔	✔		✔

Eleven of the 16 studies are classified as Women's Economic Empowerment (WEE) projects. They aim to achieve greater empowerment through facilitating income-generating activities for women. All the WEE projects in this analysis were implemented in rural areas, and their activities carry a strong agricultural theme. Agriculture-specific components often provided agricultural inputs and facilitated access to markets for both women and men in the communities. While specific interventions change depending on the project, and according to its context, it is possible to identify a few common theories of change between projects. WEE projects encouraged women's productive abilities with the aim of increasing their income potential. Projects conducted activities to support women's groups in accessing markets, provided training to increase knowledge of technical production and marketing, and strengthened access to credit. By increasing income contribution within the household, projects attempted to amplify women's decision-making power in the household. Projects also aimed to change social norms by conducting training and community discussions on gender and creating positions of responsibility within community groups. This, in turn, attempted to change popular attitudes toward women's economic role and was intended to amplify women's decision-making power both within the household and in the community.

In addition, projects that are classified not as economic empowerment, but as women's empowerment more broadly, include the following evaluations.

- Lebanon, which aimed to increase women's access to justice by providing awareness-raising sessions and free legal consultation to women.
- Nigeria, which aimed to increase women's leadership by providing women's skills in influencing local government and community leadership.
- Sierra Leone, which provided trainings on women's rights, leadership and advocacy skills to women leaders.
- Indonesia, which provided capacity-building activities to improve awareness and participation in a large-scale national development programme.
- Mali, which implemented training and activities to promote the successful school transition rate for adolescent girls.

The theories of change employed here are more varied than those of WEE projects, and so they are analysed in a separate subgroup. However, each is expected to contribute positively to the set of common individual indicators (each tested in Section 4), as well as to an overall increase in women's empowerment, whose measurement is explained below.

2.2 HOW IS WOMEN'S EMPOWERMENT MEASURED?

While there is no one overall agreed definition of women's empowerment in the sector, a significant body of the literature points to Kabeer (2001) who defines empowerment as 'the expansion in people's ability to make strategic life choices in a context where this ability was previously denied to them', and to Narayan (2002) who define empowerment as 'the expansion of freedom of choice and action to shape one's life'. Oxfam recognizes women's empowerment as a multidimensional context-specific concept (Oxfam GB 2016). 'Multidimensional' because women who may be empowered in one area of life may not be in others (Malhotra and Mather 2007; Alkire et al. 2013), and 'context specific' because behaviours or attributes that might signify empowerment in one context may have different meanings elsewhere (Malhotra and Schuler 2005). The Effectiveness Reviews consider empowerment as a process whereby women's and girls' lives are transformed from a situation where they have limited power to a situation where their power is enhanced.

In each evaluation, a range of context-specific indicators is identified to capture the characteristics of an 'empowered woman' in the socio-economic context under analysis, within a framework that allows for comparison between studies. The identified indicators are categorized under three levels of change. Changes at personal level refer to changes taking place within the person, which include changes in how the person sees herself, how she considers her role and that of other women in society; their economic role and their confidence in deciding and taking actions concerning themselves. Changes at relational level refer to changes taking place in the relationships and power relations within the woman's surrounding network. This includes changes both within the household and within the community, including markets, local authorities and decision makers. Finally, changes at environmental level refer to changes taking place in the broader context. These can be informal changes, such as social norms and attitudes, and the beliefs of wider society, or can be formal changes in the political and legislative framework.

Identified indicators are aggregated into a composite index by applying a simplified version of the Alkire-Foster (A-F) method (Alkire and Foster 2011) to give an overview result on women's

empowerment. To do this, each indicator is transformed into to a binary variable based on locally informed thresholds, and then all the indicators are averaged across all the characteristics in order to obtain an index ranging from 0 to 1 expressing a measure of empowerment.

This measurement tool builds on the experience of the Women's Empowerment in Agriculture Index (WEAI), developed by the Oxford Poverty and Human Development Initiative (OPHI) and the International Food Policy Research Institute (IFPRI) (Alkire et al. 2013), but differs from the WEAI on a number of points. Firstly, it is context specific, as the empowerment characteristics, indicators and thresholds are defined in each country/evaluation. Secondly, when transforming the indicators into a categorical variable, the index takes a value equal to 1 if the woman is considered 'empowered', contrary to the WEAI where it takes a value equal to 1 if the woman is considered 'disempowered'. Thirdly, unlike the WEAI, which is based on a dual cut-off method (Alkire et al. 2013), the Women's Empowerment Index employed in the Effectiveness Reviews applies only the first cut-off, not the second, giving the index a value ranging from 0 to 1, rather than being either 0 or 1. Finally, contrary to the WEAI, this tool also defines characteristics of empowerment that sit outside the control of the woman, such as social norms or policy and laws that exist at the environmental level.

Please refer to Bishop and Bowman (2014) and Lombardini, Bowman and Garwood (2017) for more detail on the Women's Empowerment Index and how Oxfam measures empowerment.

2.3 HOW ARE THE EVALUATIONS CONDUCTED?

All the projects implemented at individual, household or community level have been mainly assessed using quasi-experimental impact evaluation techniques by Oxfam's Impact Evaluation Advisers. Having all evaluations conducted by the same team of advisers has enabled high levels of methodological harmonization. For each evaluation, Oxfam's Impact Evaluation Adviser (abbreviated as Adviser here) travels to the country and works in collaboration with country staff, staff from the partner organizations involved in the project, and a local consultant. Together, they define the evaluation tools by mapping the project's theory of change, defining project indicators, and identifying relevant characteristics of women's empowerment. As these are not donor driven exercises, they are not bounded by the project's logframe or reporting requirements. Whenever possible, quantitative household surveys are combined with qualitative interviews or focus group discussions in order to ground the measurement of women's empowerment and provide more depth of understanding about the change resulting from the project. The Adviser then works with the country staff, partner organizations, and the consultant to design and test the questionnaire and define the sampling strategy.

A huge amount of thought and effort goes into selecting appropriate comparison groups. This work includes visiting nearby communities with similar characteristics to those where the project was implemented, understanding and attempting to mimic targeting processes, or comparing communities selected for participation in different phases of the project cycle. Particular attention is paid to reducing self-selection bias. This is done by considering observable characteristics (such as geographic and socio-economic status) as well as unobservable characteristics (such as willingness, availability and desire to participate in development projects). Recall questions of objective household measures are used to match individuals in the intervention and comparison group. The matched groups are then compared on the Women's Empowerment Index as well as the individual indicators.

3 META-ANALYSIS PROTOCOL

This meta-analysis includes all 16 Effectiveness Reviews conducted under the thematic area of Women's Empowerment from 2011 to 2016. As mentioned before, the projects evaluated are selected based on a random sampling of all Oxfam projects aiming to improve women's empowerment, with a budget greater than £200,000 and which have been implemented for at least 2.5 years. Because of this selection process and Oxfam's commitment to publish each study conducted, regardless of whether the results are positive or negative, we are able to substantially reduce the risk of selection and publication bias for this meta-analysis.

Data used in this analysis is extracted directly from the Effectiveness Reviews reports, which are publicly available from Oxfam's Policy and Practice website.[2] Oxfam has also made the anonymized household and individual data used for conducting statistical analysis, alongside the questionnaires, publically available. As a result, all calculations and analysis can be readily replicated. Any assumptions made for the data extraction are detailed in the appendix.

3.1 EFFECT SIZE CHOICES

In this meta-analysis, we choose to use Hedges' g for the multidimensional Women's Empowerment Index, and the risk ratio or response ratio (RR) when estimating other outcome variables. The interpretation of these effect sizes and the reasoning for our choices are explained below.

The overall Women's Empowerment Index is measured by the proportion of indicators in which a woman is considered to be empowered. This represents a number ranging from 0 to 1 where 1 describes a woman 'empowered' under all the characteristics, and 0 a woman not empowered in any. For this outcome variable, we employ Hedges' g, which refers to the difference in Women's Empowerment Index between the intervention and matched comparison groups, divided by the pooling weighted standard deviation (Borenstein et al. 2009; Ellis 2010). Effect sizes are expressed in terms of standard deviation units.

The Index itself is composed of individual indicators of which the majority have a binary form. With these indicators, we use the risk ratio for binary indicators (also called the response ratio for continuous or count variables), which allows us to pool binary, count, and continuous indicators between studies, as long as these indicators maintain an intuitive zero-point (Waddington et al. 2012; Borenstein et al. 2009). Additionally, the risk ratio has the benefit of having an intuitive interpretation, as it can be read as a percent change, with overall impacts higher than 1 representing a change in a positive direction, while impacts lower than 1 indicate a change in a negative direction. For a more comprehensive description of the effect size choices and calculations in this paper, please refer to the appendix.

3.2 RANDOM EFFECTS

We choose to use random effects rather than fixed effects for the plots in this report, as the cultural differences between countries as geographically diverse as those used in this meta-analysis are sure to introduce unpredictable, random variation to our impact measurements (Higgins and Green 2011).

However, some of the heterogeneity in impact can be attributed to identifiable variation; for example, differences in the socio-economic status of the target population, varying programme types, or duration/age of the project.[3] While studies included in this meta-analysis are all

methodologically comparable and without publication bias, the *types* of interventions in each programme vary widely, and as a result, the impacts of each programme are often significantly heterogeneous. With this in mind, we attempt to explain heterogeneity using subgroup analysis when reasons are evident. With further Effectiveness Reviews (and thus a larger sample size), subgroup analysis and meta-regression will be able to further distinguish which types of programme contribute most to empowerment.

3.3 RISK OF BIAS FOR THE META-ANALYSIS

Duvendack at al. (2012) caution the use of meta-analysis when studies include quasi-experimental methods. Such studies can create obstacles in a meta-analysis in two potential ways: firstly, the internal validity of the study and its heightened potential for bias compared with a more robust experimental design, such as randomized controlled trials, and secondly for the heterogeneity in the methodology applied. In order to address these concerns, we developed a risk-of-bias table, modelled after a tool developed by Hombrados and Waddington (2012). The results are summarized in Table 3.1.

Table 3.1: Risk-of-bias analysis for included studies

Study	**Risk-of-bias evaluation criteria**							
	I. Mechanism of assignment	II. Group equivalence	III. Free from Hawthorne/ John Henry Effects	IV. Free from spill-overs	V. Free from selective outcome reporting	VI. Free from selective analysis reporting	VII. Free from other bias	VIII. Confidence intervals (sample size, clusters)
Philippines 2011	Unsure	Unclear	Yes	Yes	Yes	Yes	Unclear	No
Guatemala 2011	Unsure	Unclear	Yes	Yes	Yes	Yes	Unclear	No
Zambia 2011	Unsure	Unclear	Yes	Yes	Yes	Yes	Unclear	No
Honduras 2012	Unsure	No	Yes	Yes	Yes	Yes	Unclear	No
Nigeria 2012	Unsure	Unclear	Yes	Yes	Yes	Yes	Unclear	No
Malawi 2012	Unsure	Unclear	Yes	Yes	Yes	Yes	Unclear	No
Yemen 2012	Unsure	Unclear	Yes	Yes	Yes	Unclear	Unclear	No
Sierra Leone 2012	Unsure	Unclear	Yes	Yes	Yes	Yes	Unclear	No
Ethiopia 2013	Unsure	Unclear	Yes	Yes	Yes	Yes	Unclear	Yes
Rwanda 2013	No	Unclear	Yes	Yes	Yes	Yes	Unclear	No
Indonesia 2013	Unsure	Unclear	Yes	Yes	Yes	Yes	Unclear	No
Uganda 2014	Unsure	Unclear	Yes	Yes	Yes	Yes	Unclear	No
Pakistan 2014	Unsure	Unclear	Yes	Yes	Yes	Yes	Unclear	Yes
Lebanon 2014	Unsure	No	Yes	Yes	Yes	Yes	Unclear	Yes
Armenia 2015	Unsure	Unclear	Yes	Yes	Yes	Yes	Unclear	No
Mali 2015	Unsure	Unclear	Yes	Yes	Yes	Yes	Unclear	Yes

Column I investigates whether the allocation or identification mechanism was able to control for selection bias. This means that for evaluations that are not based on randomized controlled trials, participants and non-participants are either matched based on all relevant characteristics explaining participation and outcomes, or else all relevant characteristics are accounted for.

All studies in this analysis are based on Propensity-Score Matching (PSM), which assumes the ability to observe all variables that influence selection into treatment and that these variables are not influenced by receiving treatment (Caliendo and Kopeinig 2008). To comply with this assumption, the studies match intervention and comparison individuals based on a range of recalled observable characteristics from before the beginning of the project. In addition to the statistical matching procedure, Oxfam's Impact Evaluation Advisers actively seek to identify comparison observations that reduce ex-ante observable and unobservable differences between the two groups. This is often done by identifying already pre-existing women's groups not involved in the project, in order to control for self-selection processes that might affect the analysis.

However, as the instructions from the IDCG Risk-of-Bias Tool suggests, it is unlikely that studies not based on randomization or regression discontinuity can score 'yes' on this criterion (as matching on all relevant characteristics is usually only feasible when the programme allocation rule is known and there are no errors of targeting). We marked all but one of the studies as 'unsure', with the only exception being Rwanda 2013, where we believe the study did not sufficiently address the potential for self-selection bias. While the other studies attempted to control for the self-selection processes by identifying already pre-existing groups of women not involved in the project, Rwanda 2013 did not select women from any existing groups, nor did it match on group participation at baseline.

Column II investigates if the method of analysis was executed adequately to ensure comparability of groups throughout the study and prevent confounding. In order to positively score under this dimension, it was necessary for matching to be conducted on relevant baseline characteristics or time-invariant characteristics that the PSM Rosenbaum's test suggests the results are not sensitive to the existence of hidden bias, and that covariates means are equated for treatment and comparison groups after matching. From 2013, all the studies report in the appendices the matching procedure and all the balancing tests for assessing the quality of the match. This significantly increases the confidence in the analysis; however, as the Rosenbaum's test is not reported, we marked them as 'unsure'. The Honduras 2012 and Lebanon 2014 studies seem to be exposed to a greater risk of bias because the comparison group from these studies was selected from a different geographical area from the intervention group. These two studies, along with Rwanda 2013, will be considered 'at risk of bias' when conducting subgroup analyses.

Column III assesses whether the study is affected by Hawthorne and John Henry effects, meaning the process of being observed causing motivation bias. As all data were collected ex-post, we believe data were not at risk of Hawthorne or John Henry Effects.

Column IV explores if the study adequately protected against spill-overs effects (that is, participants and non-participants are geographically and/or socially separated from one another). Control groups were consistently chosen from separate geographical units to avoid risk of spill-over effects, and questionnaires investigated exposure to project activities to assess whether the comparison group has consistently lower exposure to similar interventions.

Column V assesses whether the studies are presenting selective outcome reporting bias. There is no evidence that outcomes were selectively reported because all relevant outcomes in the methods section are reported in the results section and the Effectiveness Reviews process commits to publishing all results whether positive or negative.

Column VI investigates whether the study is free from analysis reporting bias; that is, it examines whether the authors use the most credible method of analysis to address attribution

given the data available, and the study does not suggest the existence of biased exploratory research methods (including 'data mining'). For PSM specifically, when over 10 per cent of participants fail to be matched, sensitivity analysis is used to re-estimate results using different matching methods. We believe in studies being free of selective analysis reporting as well. The only point of caution is represented by Yemen 2012, where only one method is reported and in many cases 10 per cent of the observations are missing.

Column VII investigates whether the results suggest any other sources of bias, which may include concerns about courtesy bias from outcomes collected through self-reporting; data on the baseline collected retrospectively; and information collected using an inappropriate instrument. Firstly, as baseline data were not available, survey respondents were asked to recall some basic information about their household's situation at baseline. These recall data are unlikely to be highly accurate, however recall data were only used for matching purposes, and analysis is conducted under the assumption that these inaccuracies do not differ between treatment and comparison. Secondly, there is also a question of whether there is a risk of social desirability related to the questionnaire aiming to investigate women's opinions. It should be noted that subjective measures are considered in conjunction with indicators of reported behaviour. Moreover, as long as this potential source of bias is affecting both the intervention and comparison groups equally, then this should not lead to significant bias in the estimates.

Column VIII assesses specifically the confidence intervals. To score positively under this characteristic, the analysis should be carried out at the relevant unit of treatment assignment, or account for lack of independence between observations within assignment clusters. The majority of the studies reported in the analysis choose not to cluster. This is because in most of the cases the number of clusters do not exceed 15 or 20 across the intervention and comparison groups. According to Cameron and Miller (2015), with a limited number of clusters, normal corrections for clustering might actually cause the standard errors to be biased downwards, leading to over-rejecting the null hypotheses. The pragmatic approach used in the recent Effectiveness Reviews has been to choose to bootstrap at individual level, or undertake cluster bootstrapping, depending on which approach produced higher standard errors.[4] [5]

Finally, a potential threat for the current analysis is the outcome in question. Women's empowerment is admittedly a challenging concept to measure, and Oxfam GB is engaged in a continual process of learning how to define and measure it. As such, it is important to caution that this meta-analysis assesses the overall impact of Oxfam GB's projects on women's empowerment as it is currently understood and conceptualized through the Women's Empowerment Index. Whether the Index is indeed an effective way to measure such a challenging concept is an ongoing conversation, and continues to be discussed elsewhere (Bishops and Bowman 2014; Lombardini, Bowman, and Garwood 2017). For this meta-analysis, we can complement the Index with findings on commonly used indicators describing characteristics of empowerment.

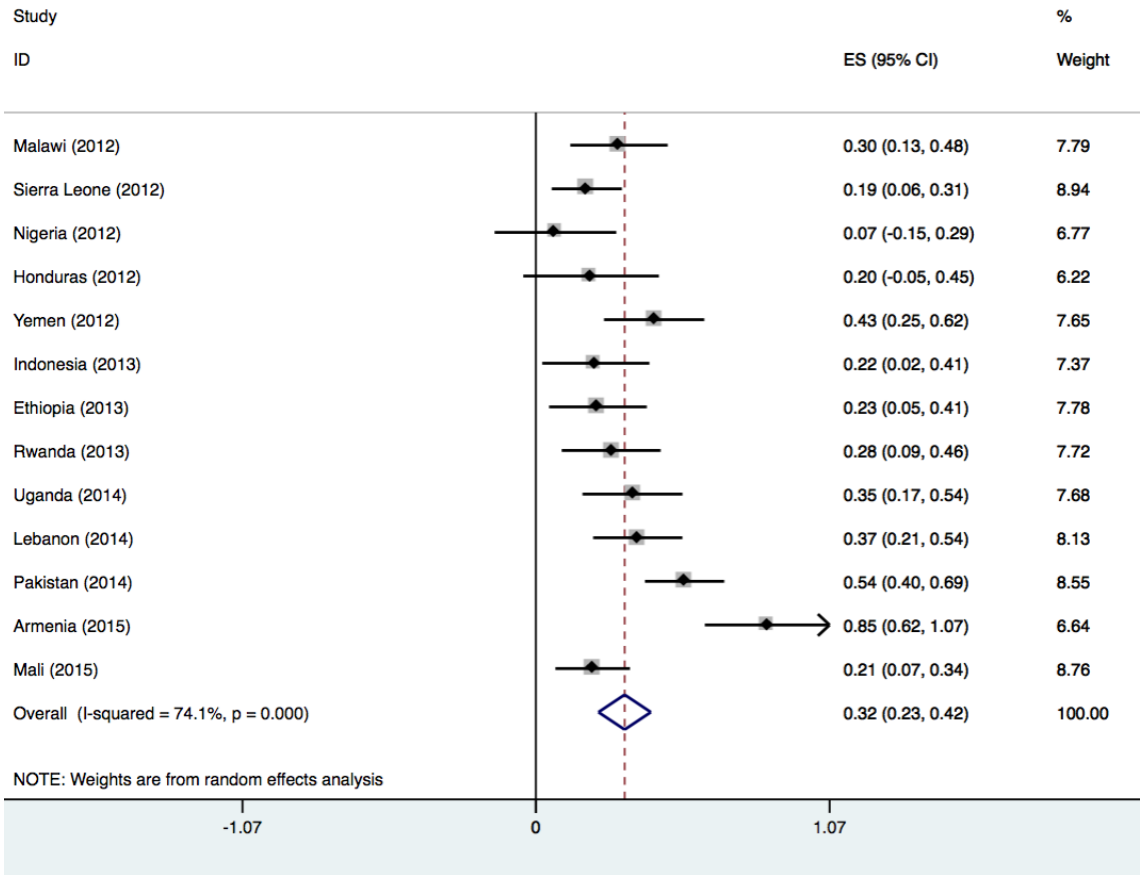
As identified in our risk of bias assessment, we recognize that some studies are subject to different levels of risk of bias. We report in the notes how these biases affect the overall effect sizes in each indicator by conducting meta-analysis with and without the studies in question and comparing results.

4 RESULTS

4.1 THE WOMEN'S EMPOWERMENT INDEX

Results of the meta-analysis on the Women's Empowerment Index suggest that overall, projects contribute to a positive impact on women's empowerment. Figure 4.1 provides a visual representation of the standardized mean difference (Hedges' g) of the 13 projects evaluated under the thematic area of Women's Empowerment between 2012 and 2016.[6] Overall, the impact of development projects on women's empowerment is estimated to have an effect size of 0.32; it appears to be statistically significant at the 5 per cent level, with the real value ranging from 0.23 to 0.42.[7] The effect size is in line with the effect size found in literature looking at micro-credit the impact of self-help groups and women's empowerment (Brody et al. 2015).

Figure 4.1: Forest plot of projects' impact on Women's Empowerment Index (Hedges' g, random effects)



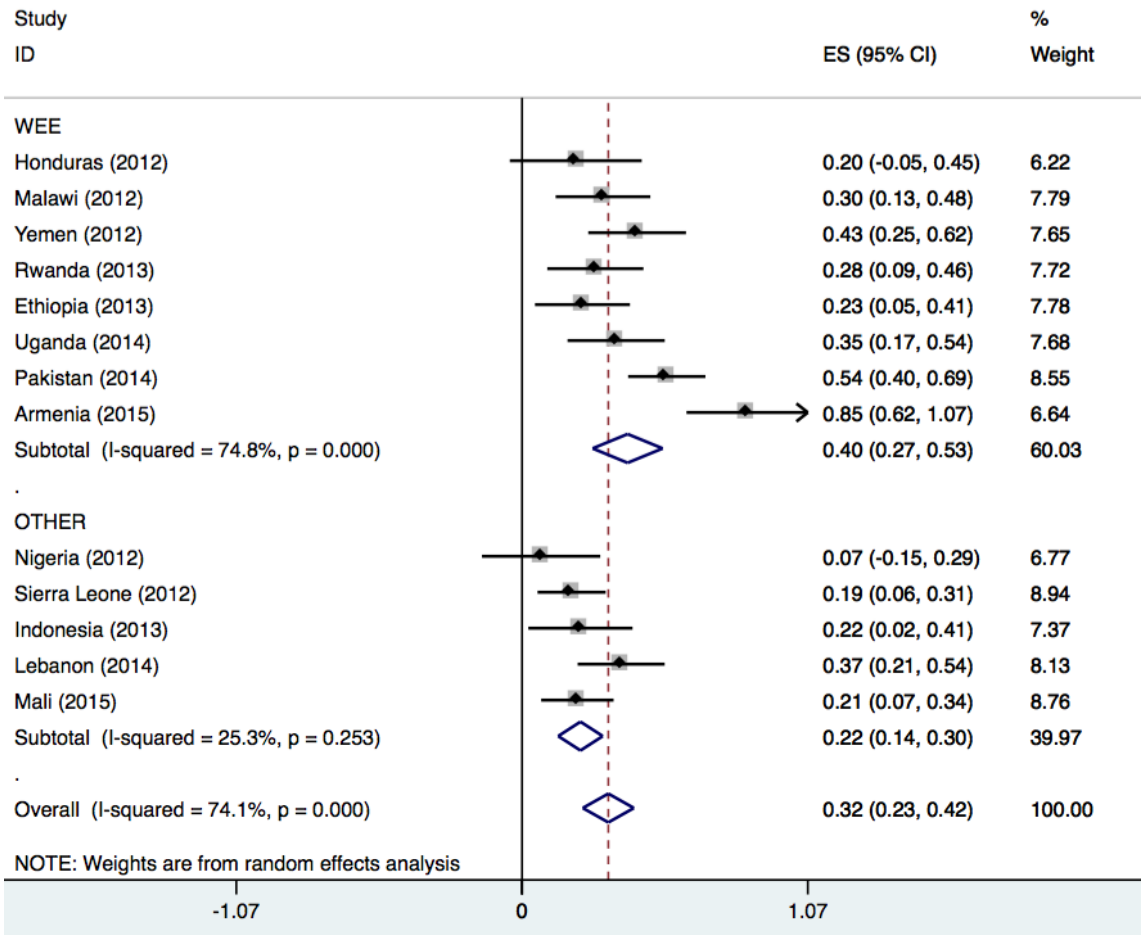
Note: Removing the outlier (Armenia) and the three studies at risk of bias, results become slightly more conservative (0.29, 95% CI: [0.21, 0.36]).

Heterogeneity in the multidimensional index

The studies reported here are all methodologically comparable, and the project selection process used for the Effectiveness Review ensures that publication bias is kept to the minimum. However, the types of interventions in each programme vary widely, and as a result, the impacts of each programme are often significantly heterogeneous (I-squared = 74.1%). We attempt to explore heterogeneity by conducting subgroup analysis on projects' characteristics, such as of expenditure, duration, region of implementation and underlying theory of change.

As presented in Table 2.1 not all the projects evaluated share the same theory of change. Projects can be divided in two main categories: those aiming to achieve women's empowerment via economic empowerment (WEE), and those aiming to achieve empowerment via other theories of change that are not mainly based on economic empowerment. Among the 13 evaluations employing the Women's Empowerment Index, 8 can be categorized as being projects that aim to achieve empowerment via economic empowerment. Figure 4.2 displays the forest plot with the overall effect sizes of projects, disaggregated by project type. Results suggest that on average the overall Women's Empowerment Index is statistically significant in both groups, it appears to be higher for those projects aiming to achieve economic empowerment (0.40; 95% CI: [0.27, 0.53]),[8] compared with projects categorized as 'other' (effect size = 0.22; 95% CI: [0.14, 0.30]). However, the difference between these two groups is not statistically significant.

Figure 4.2: Forest plot of projects' impact on Women's Empowerment Index by project type (Hedges' g, random effects)



Note: Removing the outlier and studies at risk of bias gives more conservative results for both categories, though they remain statistically significant: WEE (0.38, 95% CI: [0.26, 0.49]) and 'other' (0.18, 95% CI: [0.11, 0.26]).

Disaggregating by region gives inconclusive results.[9] Two regions (Latin America and the Caribbean, and Southern Africa) have only one evaluation each. Indeed, region subgroup analyses are hindered by small sample size. We hope with additional impact evaluations in the future to be able to explore further the cause of geographical heterogeneity.

We also examine whether projects that have been implemented for a longer time span are on average correlated with a bigger effect size; or if projects working in more depth with a smaller number of women are positively correlated with bigger effect size. Table 4.1 tests these assumptions, reporting the results of a meta-regression on the effect size regressed against: the overall project expenditure over the number of direct beneficiaries, project duration, and a dummy variable for each Oxfam region (except WAF). Estimates present a positive and significant relationship between expenditure per beneficiary and effect size, and a positive, but not significant, relationship between project duration and effect size.

Table 4.1: Regression output of Women's Empowerment Index on project expenditure per beneficiary, duration of project, and region

Meta-regression	Number of obs	=	13
REML estimate of between-study variance	tau2	=	.008909
% residual variation due to heterogeneity	I-squared_res	=	53.04%
Proportion of between-study variance explained	Adj R-squared	=	64.79%
Joint test for all covariates	Model F(7,5)	=	2.50
With Knapp-Hartung modification	Prob > F	=	0.1654

G	Coef.	Std. Err.	t	P>|t|	95% Conf. Interval
Budget over beneficiaries	0.00022	0.00010	2.16	0.08	-0.00004 - 0.00048
Project duration	0.00946	0.00688	1.38	0.23	-0.00822 - 0.02713
Asia	0.16172	0.12712	1.27	0.26	-0.16505 - 0.48848
HECA	0.21458	0.14008	1.53	0.19	-0.14550 - 0.57467
LAC	-0.18723	0.20392	-0.92	0.40	-0.71143 - 0.33696
MECIS	0.27480	0.11400	2.41	0.06	-0.01825 - 0.56785
SAF	0.13167	0.15491	0.85	0.43	-0.26653 - 0.52988
_cons	-0.34244	0.34793	-0.98	0.37	-1.23683 - 0.55195

Finally, we test for differences in effect sizes resulting from the way the women's empowerment indicator is constructed. It is, in fact, possible that evaluations using a greater number of indicators within the composite index are more likely to include outcome-level indicators, rather than impact-level. If this is the case, it would inflate the impact of the projects, questioning the validity of the measurement tool. We also test whether the sample size in each evaluation is correlated with the estimated effect size. We tested this hypothesis (Table 4.2), but it appears that the correlation is not statistically significant, meaning that there is no evidence to suggest the construction of the index is invalid.

Table 4.2: Regression output of Women's Empowerment Index on number of indicators used in Index and sample size of study

Meta-regression	Number of obs	=	13
REML estimate of between-study variance	tau2	=	.0331
% residual variation due to heterogeneity	I-squared_res	=	78.00%
Proportion of between-study variance explained	Adj R-squared	=	-30.84%
Joint test for all covariates	Model F(2,10)	=	0.01
With Knapp-Hartung modification	Prob > F	=	0.9938

g	Coef.	Std. Err.	t	P>|t|	95% Conf. Interval
Sample size	-0.00004	0.00037	-0.10	0.92	-0.00086 - 0.00078
Number indicators	0.00164	0.01715	0.10	0.93	-0.03656 - 0.03984
_cons	0.31587	0.25993	1.22	0.25	-0.26328 - 0.89502

4.2 IMPACT ON EMPOWERMENT INDICATORS

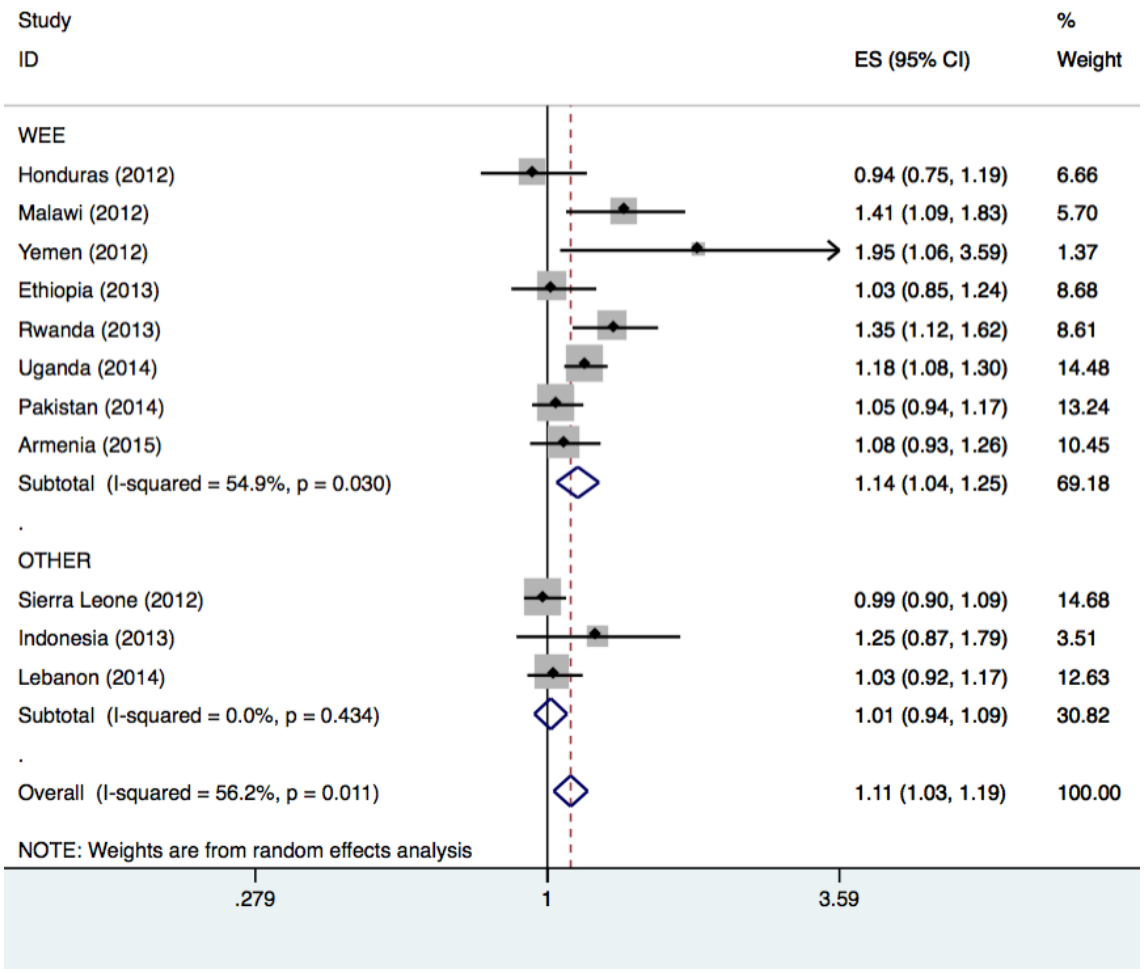
As described in Section 2, the Women's Empowerment Index is composed of a multitude of indicators measuring characteristics of empowerment deemed relevant in the context of the study. In practice, some characteristics of empowerment have been measured in several evaluations (see Table 2.3). In this section we explore the overall effect on some these commonly used indicators. The effect size used for these indicators is based on the risk/response ratio (RR) with random effects.

Estimates are presented for indicators measuring changes taking plac: within the person (women's opinion on women's economic roles), in the power relations within the household (household decision; share of household income) and within the community (influences in the community), and finally indicators related to violence (experience of violence).

Changes within the person

The analysis suggests that these projects have had an overall positive effect (1.11, CI: [1.03, 1.19]) on increasing women's positive opinion on their economic role (Figure 4.3). This indicator measures the extent to which women believe they can play a relevant part in the economic development of the household and the community (see Table 4.3). By conducting activities supporting income potential and changing social norms, projects appear to have had a positive effect on the women's opinion of their economic role. It is probably not surprising that the overall effect size appears to be bigger for WEE projects (effect size = 1.14, 95% CI: [1.04, 1.25]) compared with other type of women's empowerment projects (effect size = 1.01, 95% CI: [0.94, 1.09])[10] as this indicator is specifically concerned with economic development.

Figure 4.3: Forest plot of projects' impact on women's *opinion on women's economic role* by project type (RR, random effects)



Note: Removing outliers and studies at risk of bias, WEE programme overall effect size is 1.12 (95% CI: [1.03, 1.22]), with the 'other' category including only Sierra Leone. Subgroup analysis by region is inconclusive, as there are too few observations in each region. Subgroup analysis by outcome type suggests outcomes may be influenced by the way the indicator is measured: binary indicators give stronger results (1.13, 95% CI: [1.02, 1.24]) than the two studies using a count variable (1.06, 95% CI: [0.97, 1.16]). This difference remains even when excluding the outlier and studies at risk of bias.

Table 4.3: Summary of *opinion on women's economic roles* indicators by study

Country	Outcome type	Cut-off (binary)/Description
Guatemala	Score 0 to 1	Constructed using factor analysis of levels of agreement to 15 statements
Honduras	Binary	1 = if respondent answered 3 (out of 3) statements expressing positive opinions on women's economic roles
Malawi	Binary	1 = if respondent answered 2 (out of 3) statements expressing positive opinions on women's economic roles
Yemen	Binary	1 = if respondent answered 3 (out of 5) statements expressing positive opinions on women's economic roles
Sierra Leone	Binary	1 = if respondent answered 3 (out of 4) statements expressing positive opinions on women's economic roles
Ethiopia	Binary	1 = if respondent answered 3 (out of 3) statements expressing positive opinions on women's economic roles
Rwanda	Binary	1 = if respondent answered 3 (out of 6) statements expressing positive opinions on women's economic roles
Indonesia	Binary	1 = if respondent answered 4 (out of 8) statements expressing positive opinions on women's economic roles
Uganda	Binary	1 = if respondent answered 2 (out of 2) statements expressing positive opinions on women's economic roles
Pakistan	Count	Number of statements in which woman responds positively (of 3)
Lebanon	Binary	1 = if respondent answered 1 (out of 1) statements expressing positive opinions on women's economic roles
Armenia	Count	Number of statements in which woman responds positively (of 3)

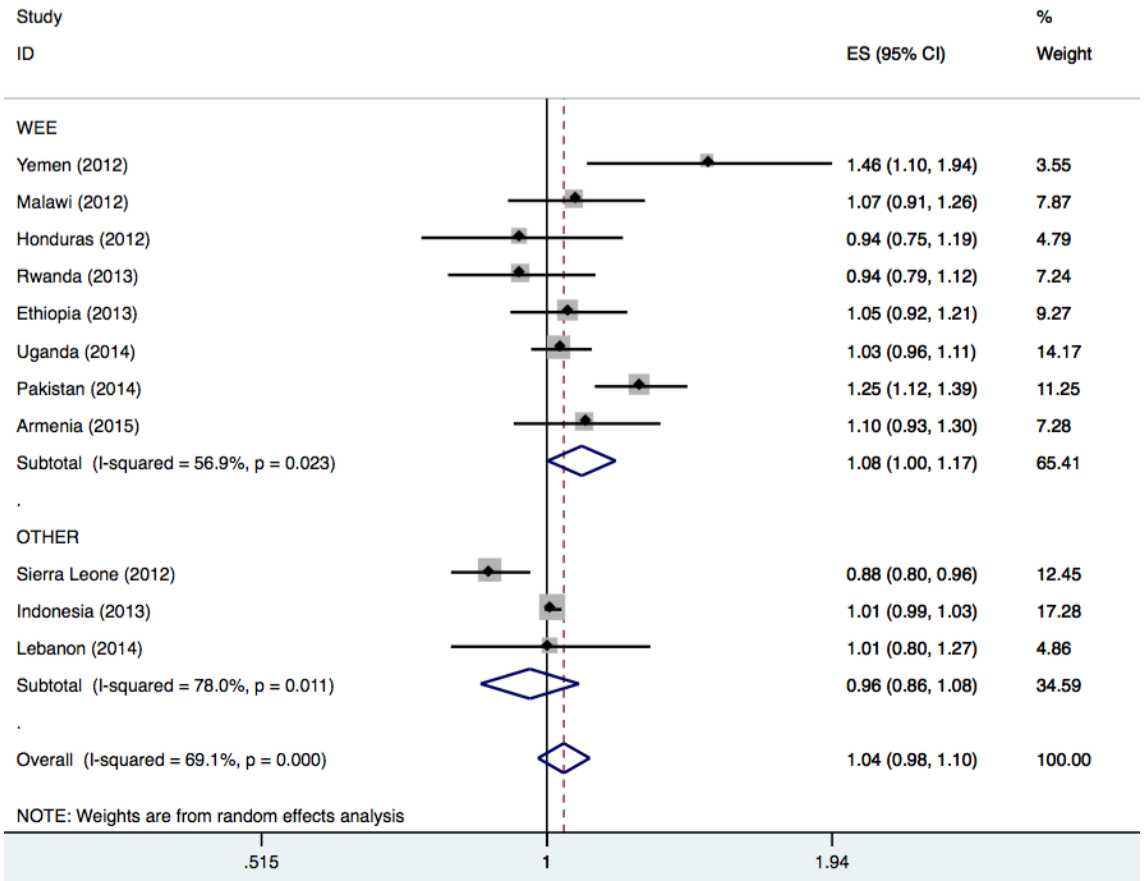
Changes in power relations within the household

Indicators measuring changes taking place in the power relations within the household provide enlightening results. In contrast to *opinion on women's economic roles,* analysis of *share of household income* (Table 4.4) does not give such distinct effects depending on project type. Figure 4.4 shows positive, though not quite statistically significant, results for projects working with women's economic empowerment (effect size = 1.08; 95% CI: [1.00, 1.17]), and no detectable effect for other type of projects (effect size = 0.96; 95% CI: [0.86, 1.08]).[11]

Table 4.4: Summary of *share of household income* indicators by study

Share of household income		
Country	**Outcome type**	**Cut-off (binary)/ Description**
Honduras	Binary	1 = if reported that they personally contributed more than a third
Malawi	Binary	1 = if she estimated that she contributed more than a third of household resources or income
Yemen	Binary	1= if she reported that she engages in some productive activity, and is able to make some if contribution to household finances
Sierra Leone	Binary	1 = if she reported that she personally contributes more than a third of total household income through her work
Ethiopia	Binary	1 = if she reported that she personally contributes at least 40 per cent.
Rwanda	Binary	1 = if she reported that she personally contributes at least half
Indonesia	Binary	1 = if she reported that she personally contributes at least half.
Uganda	Proportion	Proportion of contribution to household income
Pakistan	Proportion	Proportion of contribution to household income
Lebanon	Proportion	Proportion of household income that she personally contributes to the household
Armenia	Proportion	Proportion of household income that she personally contributes to the household

Figure 4.4: Forest plot of projects' impact on women's *share of household income* by project type (RR, random effects)



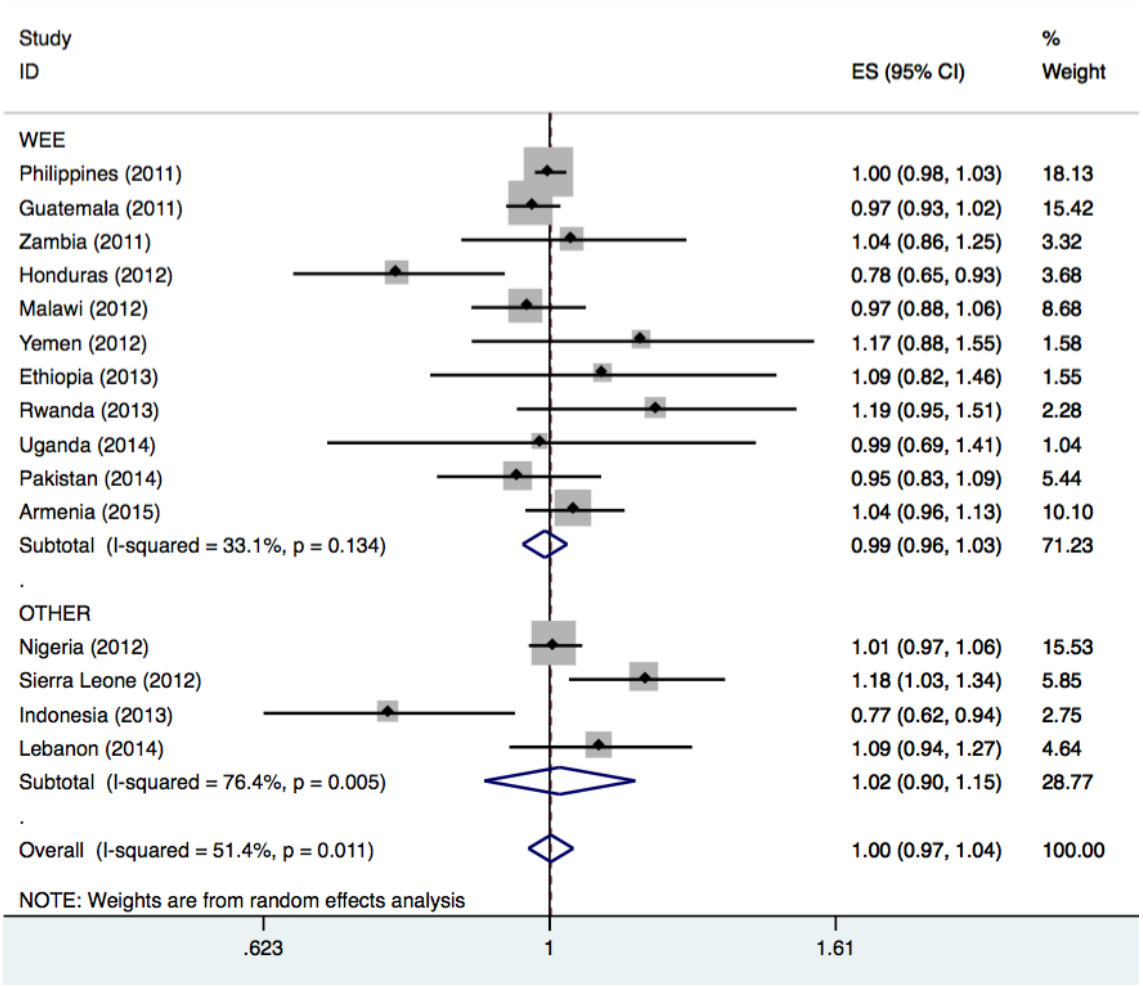
Note: When excluding one outlier (Yemen) and the studies at risk of bias, WEE programmes give significant results with a slightly larger effect size (1.10, [1.02, 1.19]). Subgroup analysis by region is inconclusive, as there are too few observations in each region. The four studies using proportional data seem to give stronger outcomes (1.10, 95% CI: [0.99, 1.23]) than binary studies (1.01, 95% CI: [0.93, 1.09]), though in each case they remain statistically insignificant, even when removing an outlier and studies at risk of bias.

In addition, the indicator measuring *household decision-making* (Table 4.5) does not present overall statistically significant results (effect size = 1.00; 95% CI: [0.97, 1.04]), nor when examining by project type.[12]

Table 4.5: Summary of *household decision-making* indicators by study

Household decision-making (Management)		
Country	**Outcome type**	**Cut-off (binary)/Description**
Philippines	Proportion	Respondent scores from 1 to 5 depending on involvement in each of 24 decisions. Outcome is the total score out of possible score
Guatemala	Proportion	Respondent scores from 1 to 4 depending on involvement in each of 25 decisions. Outcome is the total score out of possible score
Zambia	Proportion	Respondent scores from 1 to 4 depending on involvement in each of 25 decisions. Outcome is the total score out of possible score
Honduras	Binary	Respondent is involved to a medium extent in all household decisions
Nigeria	Proportion	Percentage of household decisions in which respondent has influence
Malawi	Binary	Respondent has influence over at least half of household decisions
Yemen	Binary	Respondent has influence over at least half of household decisions
Sierra Leone	Binary	Respondent has influence over at least half of household decisions
Ethiopia	Binary	Respondent has influence over at least half of household decisions
Rwanda	Binary	Respondent has influence over at least half of household decisions
Indonesia	Binary	Respondent has influence over at least half of household decisions
Uganda	Proportion	Proportion of household decisions in which respondent has influence
Pakistan	Proportion	Proportion of household decisions in which respondent has influence
Lebanon	Proportion	Proportion of household decisions in which respondent has influence
Armenia	Proportion	Proportion of household decisions in which respondent has influence

Figure 4.5: Forest plot of projects' impact on *household decision-making* by project type (RR, random effects)



Note: Results are only slightly different when excluding studies at risk of bias: WEE programmes' overall effect size is 1.00 (95% CI: [0.98, 1.02]) and the remaining three studies in the 'other' category give an overall impact of 0.99 (95% CI: [0.84, 1.17]). Regional subgroup analysis remains inconclusive. There appears to be no difference between indicator measurement in this case, as both proportion indicators and binary indicators give an overall effect size of 1.00. Removing studies at risk of bias in this subgroup analysis results in binary studies giving a slightly higher overall effect size, though it remains statistically insignificant (1.02, 95% CI: [0.88, 1.18]).

This proves to be an important finding both for its contribution to current research in the women's empowerment sector, and for its potential to inform Oxfam's work with women's empowerment moving forward. Research by Duvendack, Palmer-Jones and Vaessen (2014) on micro-credit has previously argued that increased *access* to income is not necessarily sufficient to change power dynamics within the household. This meta-analysis appears to confirm these doubts. While women participants express personal changes in their opinions of their role in the economy of their household and community, these impacts seem not to translate to changes in intra-household power dynamics, as measured through contribution to household income and ability to influence household decisions.

These findings have important programmatic implications that should be considered both within Oxfam's projects and within the wider sector. Projects should be more explicit in defining which dimensions of empowerment they are trying to change. Those projects aiming to change power dynamics within households should, therefore, reconsider their theories of change to address the assumption that access to income automatically leads to stronger bargaining power within the household. For example, the Oxfam GB Gender Justice and Women's Right Programmatic Strategic Framework 2015–2018 suggests the need to promote a more holistic approach to women's empowerment, by working across different dimensions of women's personal, political, social and economic empowerment. This includes supporting women in increasing autonomy

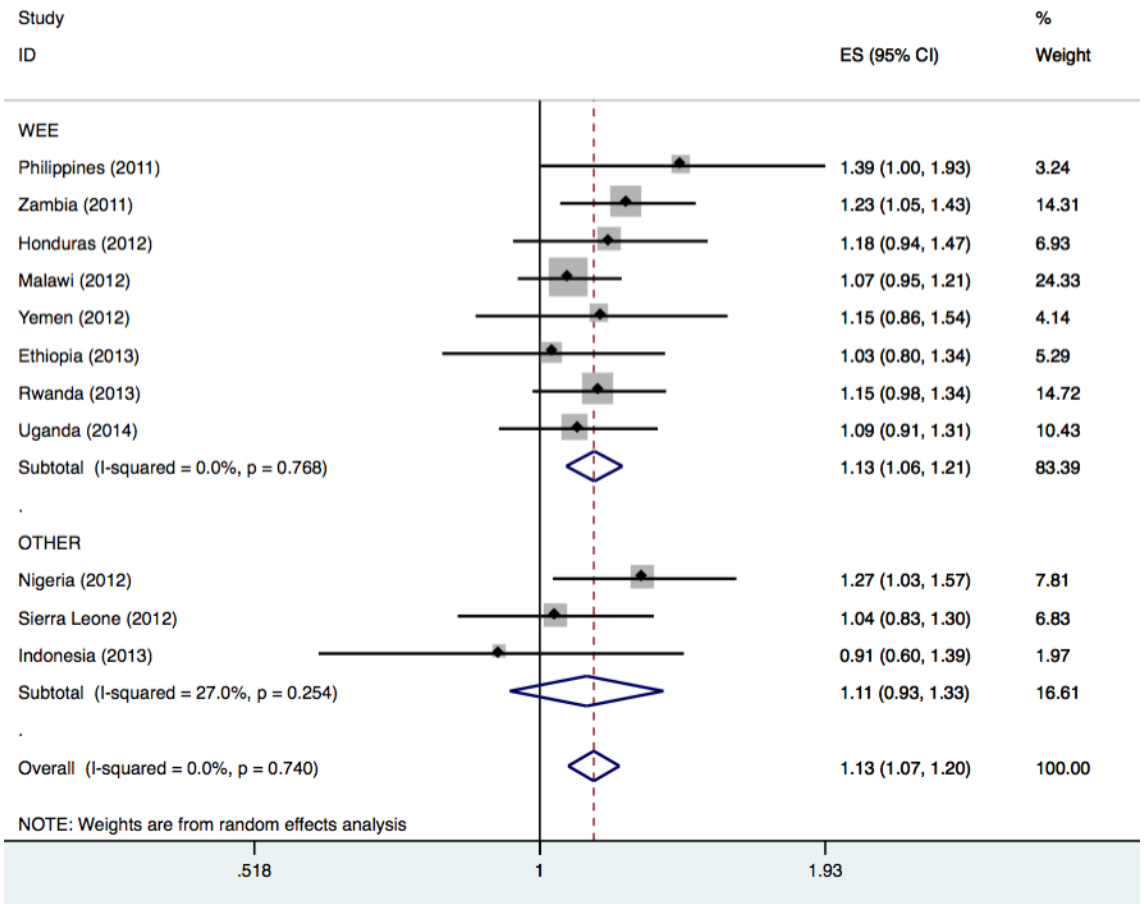
and self-belief to make changes in their own lives, agency and power to organize and influence, freedom from violence, redistribution of heavy and unequal care work, and equal rights with men. The Strategic Framework also recognizes the need to continue exploring effective ways of working with men and boys as key partners and allies in challenging and changing power relations.

Changes in power relations within the community

Indicators measuring changes taking place in the power dynamics of the community present positive and statistically significant results. Group participation and influence within the community do present overall positive and statistically significant results.[13] However, we argue it is more informative to investigate women's influence in the community.[14]

Figure 4.6 provides evidence that there is an overall positive and significant impact on the indicator measuring women's influence within the community (effect size = 1.13; 95% CI: [1.07, 1.20]) [15] (see Table 4.6 for description of indicators in each study). This is possibly due to activities conducted with community groups, where often women were encouraged and supported in taking positions of responsibility.

Figure 4.6: Forest plot of projects' impact on *women's influence in the community* by project type (RR, random effects)



Note: These results exclude the Pakistan study, as it was a very obvious outlier at 7.12 overall effect size. Further removing studies at risk of bias does not change the overall effect size of WEE projects (1.13, 95% CI: [1.04, 1.21]). Subgroup analysis by region shows no obvious patterns. As all studies employed binary outcome variables, subgroup analysis by measurement type is not applicable for this indicator.

Table 4.6: Summary of *influence in the community* indicators by study

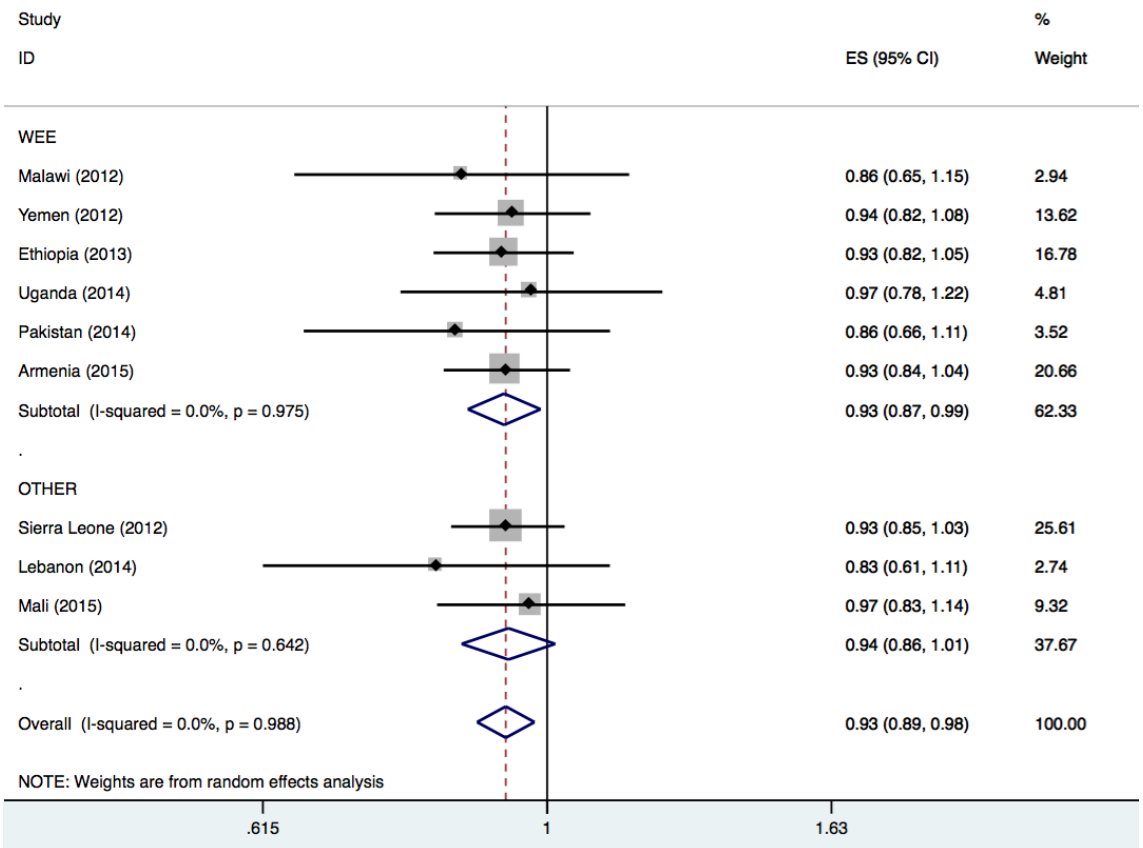
Influence within the community		
Country	**Outcome type**	**Cut-off (binary)/Description**
Philippines	Binary	Measures extent to which woman agrees with 10 statements expressing ability to influence community decisions; 1 = if respondent agreed with an amount above the median score.
Zambia	Binary	1 = if respondent agrees with 60% of statements expressing ability to influence community decisions
Nigeria	Binary	1 = if respondent agrees with 8 (out of 11) statements expressing ability to influence community decisions
Uganda	Binary	1 = if respondent agrees with both of 2 statements expressing ability to influence community decisions
Honduras	Binary	1 = if respondent agrees with 6 (out of 10) statements expressing ability to influence community decisions
Malawi	Binary	1 = if respondent strongly agrees with 6 (out of 7) statements expressing ability to influence community decisions
Sierra Leo-ne	Binary	1 = if respondent agrees with 3 (out of 4) statements expressing ability to influence community decisions
Yemen	Binary	1 = if respondent agrees with 3 (out of 4) statements expressing ability to influence community decisions
Ethiopia	Binary	1 = if respondent agrees with all of 4 statements expressing ability to influence community decisions
Rwanda	Binary	1 = if respondent agrees with 2 (out of 4) statements expressing ability to influence community decisions
Indonesia	Binary	1 = if respondent strongly agrees with 4 (out of 7) statements expressing ability to influence community decisions

When conducting subgroup analysis on project type the effect on *influence in the community* is estimated to be positive and statistically significant with WEE projects (effect size = 1.13; CI: [1.06, 1.21]), but not so with other types (effect size = 1.11, CI: [0.93,1.33]). This may be due to the small sample size of non-WEE projects. Overall, positive results confirm that women are not only participating in their community, but also reporting feeling they could actively contribute to the groups within their community.[16]

Violence against women

Finally, results of projects' overall impact on women's experience of violence require particular consideration. While none of the impact evaluations conducted show a statistically significant negative impact on this indicator, the overall effect size is negative and statistically significant (0.93, 95% CI: [0.89,0.98]) with an I-squared value of 0.0% (indicating that heterogeneity is not influential in the result) (Figure 4.7).[17] While this result may be interpreted that women in the intervention groups reported being exposed to more violence than those in the comparison groups, it is also possible that project participants become more likely to identify and report violence as they become more empowered. We examine these possibilities below.

Figure 4.7: Forest plot of projects' impact on *experience of violence* by project type (RR, random effects)



Note: Removal of the study at risk of bias (Lebanon) does not alter the overall effect size and results remain statistically insignificant for the 'other' category. Regional subgroup analysis remains inconclusive. Analysing differences between measurement methods suggests that the neighbourhood method picks up more negative results, even when excluding the study at risk of bias.

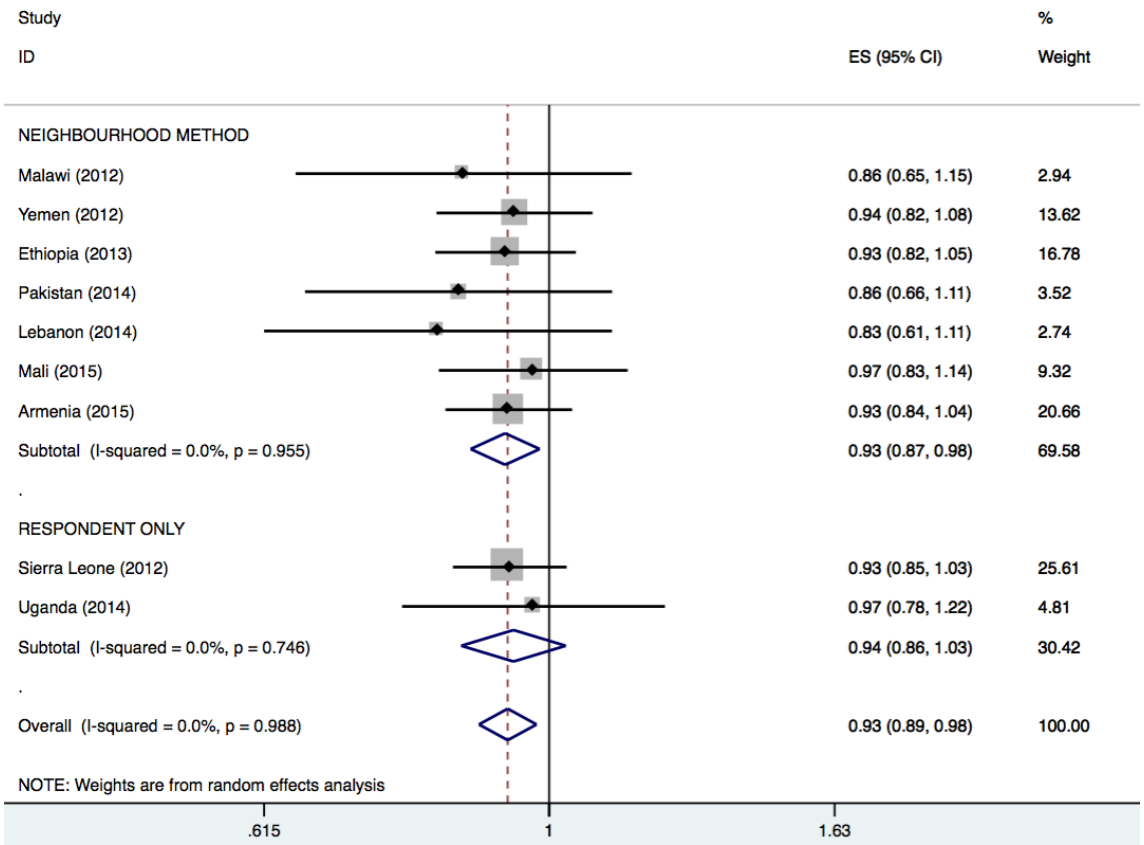
There is no *a priori* agreement on what is the effect of empowerment projects on violence. It may be possible that more empowered women are less vulnerable to violence as a result of their increased bargaining power (Agarwal 1997) and household financial contribution (Vyas and Watts 2009). Equally, it may be possible that violence is used as an instrument to exert authority over women's income (Hidrobo and Fernald 2013) or express frustration if traditional gender roles are threatened (Jewkes 2002). The evidence is mixed. While a number of studies present evidence of a positive relationship between empowerment projects and a reduction of violence (Kim et al. 2007; Hashemi, Schuler and Riley 1996; Panda et al. 2006), there is also a growing body of literature suggesting an increase of violence in the context of empowerment projects (Ahmed 2005; Angelucci 2008; Flake 2005; Krishnan et al. 2010). According to Hughes et al. (2015), Oxfam staff have also noted that some projects that encourage an indigent source of income, or that support women going to public spaces, may cause jealousy and household conflict, particularly when gender roles were highly unequal and men felt their status threatened.

However, it should also be noted that seven out of nine evaluations measured violence employed a measurement tool based on the 'neighbourhood method' (Stark et al. 2010). This approach does not directly ask respondents about their own experience of violence, but instead asks about episodes of violence that happened to a woman close to the respondent (Table 4.7). As such, it may be possible that project participants, who are more interconnected and have higher levels of awareness, may appear to be more exposed to experience of violence than are other women. Figure 4.8 presents a subgroup analysis of measurement type which appears to support this theory. Overall effect size for neighbourhood method studies is 0.93 (95% CI: [0.87, 0.98]) while respondent-only measurement studies result in an overall effect size of 0.94 (95% CI: [0.86, 1.03]). However, the sample size of the respondent-only studies is not yet large enough to draw definitive conclusions on the influence of the neighbourhood method.[18]

Table 4.7: Summary of *experience of violence* indicators by study

Experience of violence in past 12 months				
Country	**Outcome type**	**Cut-off**	**Persons included**	**Type of violence**
Malawi	Binary	1 = did not experience any violence	Women close to the respondent	Psychological, physical, sexual
Yemen	Binary	1 = did not experience any violence	Women close to the respondent	Psychological, physical, sexual
Sierra Leone	Binary	1 = did not experience any violence	Respondent	Psychological, physical
Ethiopia	Binary	1 = did not experience any violence	Women close to the respondent	Psychological, physical, sexual
Uganda	Binary	1 = did not experience any violence	Respondent	Psychological, physical, sexual
Pakistan	Binary	1 = did not experience any violence	Women close to the respondent	Psychological, physical
Lebanon	Count	Number of violent acts experienced (out of 8)	Women close to the respondent	Psychological, physical, sexual
Mali	Binary	1 = did not experience any violence	Girl close to the respondent	Psychological, physical
Armenia	Binary	1 = did not experience any violence	Women close to the respondent	Psychological, physical

Figure 4.8: Forest plot of projects' impact on *experience of violence* by measurement type (RR, random effects)



Even so, it is clear that conducting a meta-analysis was crucial for identifying this trend, as the individual studies were not sufficiently powered to detect statistically significant effects. This

demonstrates the importance and utility of conducting meta-analyses for internal accountability and evidence-based learning within an organization. Although recognizing that empowerment projects can potentially be linked to episodes of violence should not call these projects into question, the results demonstrate the importance of incorporating strategies to minimize, properly measure, and monitor, respond and mitigate unintended negative project effects (see Hughes et al. [2015]).[19]

5 CONCLUSION

This paper provides an example of using meta-analysis to enable evidence-based learning, organizational accountability and better programme implementation in the presence of a robust organizational evaluation framework. Since 2011, Oxfam GB has conducted rigorous impact evaluations, known as Effectiveness Reviews, from a random sample of mature projects in order to learn from its projects and document its impact. This paper presents the results of a meta-analysis of the 16 Effectiveness Reviews that aimed to assess projects' impact on women's empowerment using a quasi-experimental approach. The coordinated nature of the Effectiveness Review process has proved to be an advantage in an otherwise tricky endeavour.

Results from the meta-analysis provide evidence that overall, the impact of development projects on women's empowerment is estimated to have a positive and statistically significant effect size of 0.32, with the real value ranging from 0.23 to 0.42. The analysis attempted, without finding robust evidence, to explain heterogeneity by conducting subgroup analysis by project type, regional implementation, and project characteristics. While not finding conclusive results on project characteristics that are associated with higher levels of empowerment, we hope that with additional impact evaluations in the future, we will be able to explore further the cause of heterogeneity.

This meta-analysis also investigated the overall impact on some of the most commonly used indicators for women's empowerment, providing valuable insights into how empowerment is taking place in these projects and identifying important programmatic challenges, as well as measurement limitations.

Firstly, there is evidence that the projects had an overall positive and significant impact in changing women's opinions on women's economic role. However, there appears to be no evidence of increasing women's contribution to household income, nor of overall positive changes in power dynamics within the household.

Secondly, there is evidence that overall, evaluated projects increased women's participation in community groups and women's influencing in the community.

Finally, we identified an overall negative impact on the indicator measuring women's experience of violence, which was not obvious by looking at the individual impact evaluations. Questions remain on whether a measurement issue drives this result, with more empowered and interconnected women being more willing to report violence, or if it is, in fact, a negative reaction by those actors threatened by changes in power dynamics. In any case, these results call for strengthening measurement on violence and for programmes to more accurately detect and prevent any possible unintended negative impact on violence.

To conclude, despite the fact that women's empowerment is a challenging topic to investigate, a strong and consistent internal evaluation system allowed the conducting of a meta-analysis with enlightening results. This research provides important insights into the programmatic strengths and weaknesses of women's empowerment projects, and of evaluation measurement. Combining evidence from a variety of studies even allowed the capturing findings which, due to lack of statistical power, were not detected in the individual studies.

APPENDIX

CALCULATING HEDGES' G

When investigating the overall Women's Empowerment Index used in this meta-analysis, we choose to employ the standardized mean difference (SMD). As the overall Women's Empowerment Index is a continuous variable from 0 to 1, the SMD is the recommended effect size for comparing outcomes across contexts (Borenstein et al. 2009).

The SMD measures 'the size of the intervention effect in terms of the number of standard deviations in the outcome variable' (Waddington et al. 2012). For example, an effect size of 0.5 would tell us that the impact of an intervention was an increase in the outcome by half of a standard deviation.

Formulas

The effect size for the SMD (*d)* is calculated as follows:

$$d = \frac{\bar{X}_1 - \bar{X}_2}{S_{within}}$$

Where X_1 is the outcome of the intervention group, X_2 is the outcome for the comparison group, and S_{within} is the pooled standard deviation for the groups. In this study, the pooled standard deviation is estimated from the standard error of the mean difference.

It is important to note that *d* has a slight bias, and so we correct these estimates by converting to Hedges' g using an approximation for the factor *J*, as follows:

$g = J \times d \qquad where \qquad J = 1 - \frac{3}{4df-1}$ and *df* is the degrees of freedom used to estimate S_{within}.

The variance can then be calculated as follows:

$$V_g = J^2 \times V_d \qquad where \qquad V_d = \frac{n_1 + n_2}{n_1 n_2} + \frac{d^2}{2(n_1+n_2)}$$

where n_1 and n_2 are the sample sizes for the intervention and comparison groups, respectively.

CALCULATING THE RISK RATIO AND RESPONSE RATIO (RR)

When investigating commonly used indicators describing characteristics of women's empowerment in Section 5, we found that the majority of these outcomes employed binary indicators, each with a context-specific cut-off point. Such variables must be pooled with caution, as the interpretation of the effect size is slightly different depending on the country and cut-off point used. At the same time, to make interpretation of the estimates more easily understandable, some studies chose to use count or continuous variables for these indicators. To this end, we chose an effect size called the risk ratio for binary variables and the response ratio for count/continuous variables (RR), which allows us to pool these types of outcomes (Waddington et al. 2012).

Though we recognize that this effect size is typically used for outcomes with natural scales, we believe that with careful interpretation and a thorough understanding of each indicator, this effect size choice is most appropriate for the data available. With this in mind, the effect sizes in Section 4.2 should be interpreted in terms of percentage increase or decrease in the intervention group's likelihood of achieving or exceeding the cut-off point for empowerment, in cases of binary indicators.

For this reason, the analysis of indicators in the report is accompanied by a table explaining exactly how each component indicator was measured. Effect sizes are primarily examined for their direction (positive or negative) and statistical significance, and their size in comparison with other indicators.

Formulas

The effect size for RR is calculated as follows:

$$RR = \frac{Y_t}{Y_c}$$

where Y_t is the outcome mean of the treatment group and Y_c is the outcome mean of the control group. We then use log transformations for this effect size in order to estimate its standard error, calculated using the following:

$$SE\ \ln(RR) = \frac{\ln(RR)}{t}$$

where *t* is the *t*-value for the difference in outcomes between treatment and control. It is important to note here that by using this formula to calculate the standard errors, we are also able to simultaneously adjust for the lack of clustered SEs found in the studies. Finally, ln(RR) and *SE* ln(RR) for each study are entered into Stata and analysed using the –*metan*– command (Harris at al. 2008). The effect sizes are weighted, and then both measures are exponentiated for correct interpretation (Higgins and Green 2011).

Assumptions

Firstly, noting that each study uses PSM with a kernel matching strategy, we had to assume that each control observation on the common support is used fully and only once. In reality, it is possible that some control observations may be used more than once and with different weights. However, we do not find a reason to believe that this assumption would drastically affect the analysis.

Secondly, when sample sizes were not reported separately for treatment and control after matching, we assumed no observations were lost during the matching procedure, and sample size information was captured from descriptions in the sampling procedure. Thirdly, if outcome means were not reported after matching, we assumed that the treatment group mean was the same for pre- and post-matching. In these cases, we then calculated the control group means by subtracting the post-matching difference from the treatment mean.

ACKNOWLEDGEMENTS

We thank Claire Hutchings, Maren Duvendack, Emily Brown, Hugo Sintes, and Kimberly Bowman for their useful comments. We also thank Hugh Waddington and the editors of this journal, as well as our anonymous reviewers for their support and feedback. All errors are the authors' own.

REFERENCES TO INCLUDED STUDIES

Bishop, David. 2012. “Effectiveness Review: Copperbelt Livelihoods Project, Zambia”. Effectiveness Review Series 2011/12, Oxford: Oxfam GB http://policy-practice.oxfam.org.uk/publications/effectiveness-review-copperbelt-livelihoods-project-zambia-247233

Bishop, David. 2012. “Effectiveness Review: Improving Women’s Leadership and Effectiveness in Agricultural Governance, Nigeria”. Effectiveness Review Series 2012/13, Oxford: Oxfam GB. http://policy-practice.oxfam.org.uk/publications/effectiveness-review-improving-womens-leadership-and-effectiveness-in-agricultu-303455

Caeyers, Bet and Rob Fuller. 2015. “Women's Empowerment in Rwanda: Evaluation of women's economic leadership through horticulture planting material business”. Effectiveness Review Series 2013/14, Oxford: Oxfam GB. http://policy-practice.oxfam.org.uk/publications/womens-empowerment-in-rwanda-evaluation-of-womens-economic-leadership-through-h-550099

Caeyers, Bet and Simone Lombardini. 2015. “Women's Empowerment in Indonesia: Evaluation of Papua women's empowerment”. Effectiveness Review Series 2013/14, Oxford: Oxfam GB. http://policy-practice.oxfam.org.uk/publications/womens-empowerment-in-indonesia-evaluation-of-papua-womens-empowerment-347365

Fuller, Rob. 2013. “Effectiveness Review: Highlands Value Chain Development Alliance, Guatemala”. Effectiveness Review Series 2011/12, Oxford: Oxfam GB. http://policy-practice.oxfam.org.uk/publications/effectiveness-review-highlands-value-chain-development-alliance-guatemala-277752

Fuller, Rob. 2014. “Effectiveness Review: Promoting Sustainable Livelihoods for Women and Vulnerable Groups in Chiradzulu District, Malawi”. Effectiveness Review Series 2012/13, Oxford: Oxfam GB. http://policy-practice.oxfam.org.uk/publications/effectiveness-review-promoting-sustainable-livelihoods-for-women-and-vulnerable-312840

Fuller, Rob. 2014. “Effectiveness Review: Safe Age of Marriage and Women’s Economic Empowerment Yemen”. Effectiveness Review Series 2012/13, Oxford: Oxfam GB. http://policy-practice.oxfam.org.uk/publications/effectiveness-review-safe-age-of-marriage-and-womens-economic-empowerment-yemen-312692

Fuller, Rob. 2014. “Effectiveness Review: Supporting Rural Community Banks, Honduras”. Effectiveness Review Series 2012/13, Oxford: Oxfam GB. http://policy-practice.oxfam.org.uk/publications/effectiveness-review-supporting-rural-community-banks-honduras-312856

Fuller, Rob. 2014. “Effectiveness Review: Strengthening and Linking Women-Led Efforts to Promote Women’s Property and Literacy Rights in Sierra Leone”. Effectiveness Review Series 2012/13, Oxford: Oxfam GB. http://policy-practice.oxfam.org.uk/publications/effectiveness-review-strengthening-and-linking-women-led-efforts-to-promote-wom-312683

Fuller, Rob. 2014. “Women’s Empowerment in Ethiopia: Evaluation of women’s beekeeping and access to financial services”. Effectiveness Review Series 2013/14, Oxford: Oxfam GB. http://policy-practice.oxfam.org.uk/publications/womens-empowerment-in-ethiopia-evaluation-of-womens-beekeeping-and-access-to-fi-336387

Hughes, Karl. 2012 “Effectiveness Review: Enhancing Access and Control to Sustainable Livelihood Assets, Philippines”. Effectiveness Review Series 2011/12, Oxford: Oxfam GB. http://policy-practice.oxfam.org.uk/publications/effectiveness-review-enhancing-access-and-control-to-sustainable-livelihood-ass-247232

Lombardini, Simone and Kanako Yoshikawa. 2015. “Women’s Empowerment in Uganda: Impact evaluation of the project ‘Piloting gender sensitive livelihoods in Karamoja’”. Effectiveness Review Series 2014/15, Oxford: Oxfam GB. http://policy-practice.oxfam.org.uk/publications/womens-empowerment-in-uganda-impact-evaluation-of-the-project-piloting-gender-s-592575

Lombardini, Simone and Kimberly Bowman. 2015. “Women's Empowerment in Pakistan: Impact evaluation of the empowering small scale producers in the dairy sector project”. Effectiveness Review Series 2014/15, Oxford: Oxfam GB. http://policy-practice.oxfam.org.uk/publications/womens-empowerment-in-pakistan-impact-evaluation-of-the-empowering-small-scale-580463

Lombardini, Simone and Marcella Vigneri. 2015. “Women’s Empowerment in Lebanon: Impact evaluation of the project ‘Women’s access to justice in the Middle East and North Africa region'”. Effectiveness Review Series 2014/15, Oxford: Oxfam GB. http://policy-practice.oxfam.org.uk/publications/womens-empowerment-in-lebanon-impact-evaluation-of-the-project-womens-access-to-582081

Lombardini, Simone. 2017. “Women’s Empowerment in Armenia: Impact evaluation of the women's economic empowerment project in rural communities in Vayots Dzor region”. Effectiveness Review Series 2015/16, Oxford: Oxfam GB. http://policy-practice.oxfam.org.uk/publications/womens-empowerment-in-armenia-impact-evaluation-of-the-womens-economic-empowerm-620210

Vigneri, Marcella and Simone Lombardini. 2017. “W omen’s Empowerment in Mali: Impact evaluation of the educational project: ‘Girls CAN – Promoting Secondary Education in West Africa”. Effectiveness Review Series 2015/16, Oxford: Oxfam GB. http://policy-practice.oxfam.org.uk/publications/womens-empowerment-in-mali-impact-evaluation-of-the-educational-project-girls-c-620252

REFERENCES

Agarwal, Bina. 1997. "'Bargaining' and gender relations: within and beyond the household." *Feminist Economics* 3(1): 1–51.

Ahmed, Syed Masud. 2005. "Intimate partner violence against women: experiences from a woman-focused development programme in Matlab, Bangladesh." *Journal of Health Population and Nutrition* 23(1): 95–101.

Alkire, Sabina and James Foster. 2011. "Counting and Multidimensional Poverty Measurement." *Journal of Public Economics* 95: 476–487.

Alkire, Sabina, Ruth Meinzen-Dick, Amber Peterman, Agnes Quisumbing, and Greg Seymour. 2013. "The Women's Empowerment in Agriculture Index." *World Development* 52:71–91.

Angelucci, Manuela. 2008. "Love on the rocks: domestic violence and alcohol abuse in rural Mexico." *B.E. Journal of Economic Analysis and Policy* 8:1935–1682.

Bishop, David and Kimberly Bowman. 2014. "Still learning: a critical reflection on three years of measuring women's empowerment in Oxfam." *Gender & Development* 22:253–269.

Borenstein, Michael, Larry V. Hedges, Julian P.T. Higgins, and Hannah R. Rothstein. 2009. *Introduction to Meta-Analysis*. Hoboken, NJ, USA: John Wiley & Sons, Ltd.

Brody, Carinne, Thomas de Hoop, Martina Vojtkova, Ruby Warnock, Megan Dunbar, Padmini Murthy and Shari Dworkin. 2015. "Economic Self-Help Group Programs for Improving Women's Empowerment: A Systematic Review." *Campbell Systematic Reviews* 2015:11.

Cameron, Colin and Douglas L. Miller. 2015. "A Practitioner's Guide to Cluster Robust Inference." *Journal of Human Resources* 50:317–372.

Caliendo, Marco and Sabine Kopeinig. 2008. "Some practical guidance for the implementation of propensity score matching." *Journal of Economic Surveys* 22:31–72.

Duvendack, Maren, Jorge G. Hombrados, Richard Palmer-Jones and Hugh Waddington. 2012. "Assessing 'what works' in international development: meta-analysis for sophisticated dummies." *Journal of Development Effectiveness* 4:456–471.

Duvendack, Maren, Richard Palmer-Jones and Jos Vaessen. 2014. "Meta-analysis of the impact of microcredit on women's control over household decisions: methodological issues and substantive findings." *Journal of Development Effectiveness* 6:73–96.

Flake, Dallan. 2005. "Individual, family, and community risk markers for domestic violence in Peru." *Violence Against Women* 11: 353–373.

Ellis, Paul D. 2010. *The Essential Guide to Effect Sizes: Statistical Power, Meta-Analysis, and the Interpretation of Research Results*. Cambridge: Cambridge University Press.

Harris, Ross J., Michael J. Bradburn, Jonathan J. Deeks, Roger M. Harbord, Douglas G. Altman and Jonathan A.C. Sterne. 2008. "metan: fixed- and random-effects meta-analysis*." The Stata Journal* 8:3–28.

Hashemi, Syed M., Sidney Ruth Schuler and Ann P. Riley. 1996. "Rural credit programs and women's empowerment in Bangladesh." *World Development* 24:635–653.

Hidrobo, Melissa and Lia Fernald. 2013. "Cash transfers and domestic violence." *Journal of Health Economics* 32: 304–319.

Higgins, J.P.T. and S. Green, eds. 2011. “Cochrane Handbook for Systematic Reviews of Interventions” Version 5.1.0. *The Cochrane Collaboration*. www.cochrane-handbook.org

Hombrados, Jorge G. and Hugh Waddington. 2012. “Internal validity in social experiments and quasi-experiments: an assessment tool for reviewers.” Presented at the Dhaka Colloquium on Systematic Reviews in International Development, Dhaka, December 10–14. http://www.3ieimpact.org/media/filer_public/2012/12/26/jorge_ hombrados_and_hugh_waddington_conference-session12-b_3ie_dhaka_colloquium.pdf

Hughes, Karl and Claire Hutchings. 2011. “Can we obtain the required rigour without randomisation? Oxfam GB’s non-experimental Global Performance Framework.” International Initiative for Impact Evaluation (3ie) Working Paper 13. http://www.ipdet.org/files/Publication-Can_we_obtain_the_required_rigour _without_randomization.pdf

Hughes, Christine, Mara Bolis, Rebecca Fries and Stephanie Finigan. 2015. “Women's economic inequality and domestic violence: exploring the links and empowering women.” *Gender & Development* 23:279-297.

Hutchings, Claire. 2014. “Balancing accountability and learning: a review of Oxfam GB’s global performance framework.” *Journal of Development Effectiveness* 6:425–435.

Jewkes, Rachel. 2002. “Intimate partner violence: causes and prevention.” *The Lancet* 359:1423–1429.

Kabeer, Naila. 2001. “Reflections on the measurement of women’s empowerment.” In *Discussing Women’s Empowerment: Theory and Practice*, Sida Studies No. 3, 17–57. Stockholm: Swedish International Development Cooperation Agency.

Kim, Julia C., Charlotte H. Watts, James R. Hargreaves, Luceth X. Ndhlovu, Godfrey Phetla, Linda A. Morison, Joanna Busza, John D.H. Porter and Paul Pronyk. 2007. “Understanding the impact of a microfinance-based intervention on women's empowerment and the reduction of intimate partner violence in South Africa.” *American Journal of Public Health* 97: 1794–1802.

Krishnan, Suneeta, Corinne H. Rocca, Alan E. Hubbard, Kalyani Subbiah, Jeffrey Edmeades and Nancy S. Padian. 2010. “Do changes in spousal employment status lead to domestic violence? Insights from a prospective study in Bangalore, India.” *Social Science & Medicine* 70:136–143.

Lombardini, Simone, Kimberly Bowman and Rosa Garwood. 2017. “A ‘How To’ Guide to Measuring Women’s Empowerment: Sharing experience from Oxfam’s impact evaluations.” Oxford: Oxfam GB.

Malhotra, Anju and Mark Mather. 2007. “Do Schooling and Work Empower Women in Developing Countries? Gender and Domestic Decisions in Sri Lanka.” *Sociological Forum* 12: 599–630.

Malhotra, Anju and Sidney R. Schuler. 2005. “Women’s Empowerment as Variable in International Development.” In *Measuring Empowerment: Cross-disciplinary Perspectives*. Washington, DC: The World Bank.

Mallett, Richard, Jessica Hagen-Zanker, Rachel Slater and Maren Duvendack. 2012. “The benefits and challenges of using systematic reviews in international development research.” *Journal of Development Effectiveness* 4:445–455.

Narayan, Deepa, ed. 2002. *Empowerment and Poverty Reduction: A Sourcebook*. Washington, DC: The World Bank.

Oxfam GB. 2016. “How are Effectiveness Reviews carried out?” Oxford: Oxfam GB http://policy-practice.oxfam.org.uk/publications/how-are-effectiveness-reviews-carried-out-594353

Panda, Pradeep, Jayoti Gupa, Indika Bulankulame, Nandita Bhatla, Swati Chakraborty and Nata Duvvury. 2006. "*Property Ownership and Inheritance Rights of Women for Social Protection – The South Asia Experience: Synthesis Report of Three Studies*," Washington, DC: International Centre for Research on Women.

Poole, Charles and Sander Greenland. 1999. “Random-Effects Meta-Analyses Are Not Always Conservative.” *American Journal of Epidemiology* 150:469–475.

Stark, Lindsay, Les Roberts, Wendy Wheaton, Anne Acham, Neil Boothby and Alastair Ager. 2010. “Measuring violence against women amidst war and displacement in northern Uganda using the ‘neighbourhood method’.” *Journal of Epidemiology and Community Health* 64:1056–1061.

Vyas, Seema and Charlotte Watts. 2009. “How does economic empowerment affect women's risk of intimate partner violence in low and middle income countries? A systematic review of published evidence.” *Journal of International Development* 21: 577–602.

Waddington, Hugh, Howard White, Birte Snilstveit, Jorge Garcia Hombrados, Martina Vojtkova, Philip Davies, Ami Bhavsar, et al. 2012. “How to do a good systematic review of effects in international development: a toolkit.” *Journal of Development Effectiveness* 4:359–387.

NOTES

1 It should be noted that there is a proportion of Oxfam projects that are also aiming to increase women's empowerment, but for which it is not possible to evaluate using quasi-experimental impact evaluation techniques. These projects typically work with a reduced number of women aiming to improve their political engagement using a rights-based approach. However, the size of these interventions is usually not big enough to allow exploitation of the properties of the central limit theorem. In these cases, evaluations have used non-counterfactual impact evaluation techniques, such as process tracing, and the results of these evaluations are not included in this meta-analysis.

2 www.oxfam.org.uk/effectiveness

3 In our case, random effects analysis yields slightly more conservative results than with fixed effects (results available upon request). However, it is important to note that using random effects analysis is not always a 'conservative' choice – some authors argue that in cases 'when the studies are not methodologically comparable, when their results are discernibly heterogeneous, or when there is evidence of publication bias' it may not be appropriate to combine them at all, as random effects analysis will not ameliorate such differences (Poole and Greenland 1999, 469).

4 http://policy-practice.oxfam.org.uk/blog/2016/12/real-geek-to-cluster-or-not-to-cluster-not-to-cluster

5 The meta-analysis attempts to correct for this with its alternative calculation of the risk ratio standard errors, using the log transformation of the *t*-statistic as explained in the appendix.

6 The reader will notice that during the first year of implementation of the Effectiveness Reviews (2011/12), the measurement tool for women's empowerment was not yet developed. Therefore, while the evaluations were already assessing the impact of the project on empowerment indicators, these were not then combined in one women's empowerment index. Thus, these studies are excluded from the meta-analysis of project impact on the Index.

7 Removing Armenia, which appears to be an outlier, gives a slightly more modest estimate (overall effect size = 0.29, CI: [0.21, 0.36]). Further removing studies at risk of bias does not significantly alter the results.

8 Removing the outlier (Armenia) and studies at risk of bias gives more conservative results for both categories, though impacts remain statistically significant: WEE (0.35, 95% CI: [0.25, 0.46]) and 'other' (0.18, 95% CI: [0.09, 0.26]).

9 Results available upon request.

10 Removing outliers and studies at risk of bias, WEE programme overall effect size is 1.12 (95% CI: [1.03, 1.22]), with the 'other' category including only Sierra Leone. Subgroup analysis by region is inconclusive, as there are too few observations in each region. Subgroup analysis by outcome type suggests outcomes may be influenced by the way the indicator is measured: binary indicators give stronger results (1.13, 95% CI: [1.02, 1.24]) than the two studies using a count variable (1.06, 95% CI: [0.97, 1.16]). This difference remains even when excluding the outlier and studies at risk of bias.

11 When excluding one outlier (Yemen) and the studies at risk of bias, WEE programmes give significant results with a slightly larger effect size (1.10, [1.02, 1.19]), while the 'other' category is confined to Sierra Leone. Subgroup analysis by region is inconclusive, as there are too few observations in each region. The four studies using proportional data seem to give stronger outcomes (1.10, 95% CI: [0.99, 1.23]) than binary studies (1.01, 95% CI: [0.93, 1.09]), though in each case they remain statistically insignificant, even when removing an outlier and studies at risk of bias.

12 Results are only slightly different when excluding studies at risk of bias: WEE programmes' overall effect size is 1.00 (95% CI: [0.98, 1.02]) and the remaining three studies in the 'other' category give an overall impact of 0.99 (95% CI: [0.84, 1.17]). Regional subgroup analysis remains inconclusive. There appears to be no difference between indicator measurement in this case, as both proportion indicators and binary indicators give an overall effect size of 1.00. Removing studies at risk of bias in this subgroup analysis results in binary studies giving a slightly higher overall effect size, though it remains statistically insignificant (1.02, 95% CI: [0.88, 1.18]).

13 Results of meta-analysis for group participation are available upon request.

14 Group participation is an indicator commonly used to recognize social capital (Alkire et al. 2013). However, many projects implement their activities by establishing women's groups in the intervention areas. In such cases, there is an argument that group participation represents more a measure of outcome rather than impact. As a consequence, including it in the composite index may provide an inflated measure of positive impact of the project. Evaluators are therefore encouraged to consider, on a case-by-case basis, whether is it appropriate to include it in the overall index.

15 These results exclude the Pakistan study, as it was an outlier at 7.12 overall effect size.

16 Further removing studies at risk of bias does not change the overall effect size of WEE projects (1.13, 95% CI: [1.04, 1.21]). Subgroup analysis by region shows no obvious patterns. As all studies employed binary outcome variables, subgroup analysis by measurement type is not applicable for this indicator.

17 Removal of the study at risk of bias (Lebanon) does not alter the overall effect size. Regional subgroup analysis remains inconclusive.

18 When excluding the study at risk of bias, subgroup analysis still suggests that the neighbourhood method picks up more negative results.

19 As a direct response to this meta-analysis, in the most recent Effectiveness Reviews of women's empowerment projects, Oxfam has attempted to further understand these findings. Specifically, where surveys can be conducted in private and when cultural context allows, enumerators ask about exposure to violence using both methods: using reports of personal violence and the neighbourhood method.

Oxfam Discussion Papers

Oxfam Discussion Papers are written to contribute to public debate and to invite feedback on development and humanitarian policy issues. They are 'work in progress' documents, and do not necessarily constitute final publications or reflect Oxfam policy positions. The views and recommendations expressed are those of the author and not necessarily those of Oxfam. The final version of this paper will be published in the Journal of Development Effectiveness in September 2017.
http://dx.doi.org/10.1080/19439342.2017.1377750

For more information, or to comment on this paper, email slombardini@oxfam.org.uk

© Oxfam GB September 2017

This publication is copyright but the text may be used free of charge for the purposes of advocacy, campaigning, education, and research, provided that the source is acknowledged in full. The copyright holder requests that all such use be registered with them for impact assessment purposes. For copying in any other circumstances, or for re-use in other publications, or for translation or adaptation, permission must be secured and a fee may be charged. E-mail policyandpractice@oxfam.org.uk.

The information in this publication is correct at the time of going to press.

Published by Oxfam GB. Oxfam GB, Oxfam House, John Smith Drive, Cowley, Oxford, OX4 2JY, UK.

OXFAM

Oxfam is an international confederation of 20 organizations networked together in more than 90 countries as part of a global movement for change, to build a future free from the injustice of poverty. Please write to any of the agencies for further information, or visit www.oxfam.org.

Oxfam America (www.oxfamamerica.org)
Oxfam Australia (www.oxfam.org.au)
Oxfam-in-Belgium (www.oxfamsol.be)
Oxfam Brasil (www.oxfam.org.br)
Oxfam Canada (www.oxfam.ca)
Oxfam France (www.oxfamfrance.org)
Oxfam Germany (www.oxfam.de)
Oxfam GB (www.oxfam.org.uk)
Oxfam Hong Kong (www.oxfam.org.hk)
Oxfam IBIS (Denmark) (www.ibis-global.org)
Oxfam India (www.oxfamindia.org)
Oxfam Intermón (Spain) (www.oxfamintermon.org)
Oxfam Ireland (www.oxfamireland.org)
Oxfam Italy (www.oxfamitalia.org)
Oxfam Japan (www.oxfam.jp)
Oxfam Mexico (www.oxfammexico.org)
Oxfam New Zealand (www.oxfam.org.nz)
Oxfam Novib (Netherlands) (www.oxfamnovib.nl)
Oxfam Québec (www.oxfam.qc.ca)
Oxfam South Africa (www.oxfam.org.za)

OXFAM