BUILDING RESILIENCE

A meta-analysis of Oxfam's resilience Effectiveness Reviews

Photo: Women from Dola village, India, construct a pond to irrigate their vegetable gardens. Credit: Tom Pietrasik

ROB FULLER AND JONATHAN LAIN

INDEPENDENT CONSULTANTS FOR OXFAM GB

Oxfam's Effectiveness Reviews evaluate the impact of the organization's projects on the lives of those they are intended to help. This paper uses statistical meta-analysis to summarise the results of all 16 Effectiveness Reviews carried out under the theme of resilience between 2011 and 2015. We find that the projects evaluated had a significant positive impact overall, as measured by an index of contextually-appropriate indicators of resilience. However, there are important differences in impact between regions of the world. The meta-analysis also reveals a difference in the level of resilience by household head gender, and provides some insights into the resilience measurement approach.

www.oxfam.org.uk/effectiveness

CONTENTS

1 INTRODUCTION

International development actors are increasingly investing in efforts to evaluate the impact of their work, with the aims both of understanding their effectiveness and generating useful learning on how to make programmes more effective. Most impact evaluations take the form of in-depth studies of particular projects or programmes, implemented in particular contexts. While the resulting reports are often useful for understanding effectiveness on a local level, producing generalized insights or learning remains difficult. In this paper we seek to bridge this gap through the use of statistical meta-analysis, a technique that pools data from multiple studies, in order to seek patterns across those studies and to allow for the analysis of questions which are not answerable with any one study in isolation.

A conventional meta-analysis (such as those carried out by the Cochrane Collaboration or the International Initiative for Impact Evaluation, 3ie) involves consolidating multiple studies of a single type of intervention, with the aim of drawing general conclusions about the effectiveness of that intervention.[1] In contrast, this meta-analysis consolidates data on the effectiveness of a single organisation, Oxfam GB, across 16 projects that aimed to build resilience in rural communities. These data come from Oxfam's 'Effectiveness Reviews', a series of impact evaluations carried out each year since 2011 on randomly selected projects. The projects evaluated were implemented in various locations around the world and were diverse in their scale and activities, but they were evaluated against a common objective – that of building resilience to shocks, stresses and uncertainty – using a consistent quasi-experimental methodology.

In this paper, we present the results of the meta-analysis and discuss what can be learned from those results. Overall, the projects evaluated are found to have had a significant positive effect on resilience. The aggregated project effect across the 16 evaluations is estimated to be approximately 0.4 standard deviations in terms of Oxfam's 'resilience index', a composite index of context-specific indicators of resilience. There are significant positive effects across each of the three components of this resilience index – absorptive capacity, adaptive capacity and transformative capacity. Surprisingly, there does not appear to be any connection between the size of impact and the scale of the project, nor the time-frame over which the interventions are implemented. However, there are large differences between regions of the world: projects evaluated in Asia are estimated to have much greater impact, on average, than those in Africa and Latin America. We investigate whether these regional differences could be caused by differences in the nature of the projects in different regions, or whether they result from the measurement approach or the evaluation methodology itself. We conclude that it seems unlikely that these factors could fully explain the difference in results between Asia and the other two regions.

We also find evidence of an important gender difference in resilience: female-headed households were generally found to have lower scores on the resilience index than were male-headed households. This gender difference is small in magnitude, but it is consistent across the components of resilience and across the contexts where the evaluations were carried out. If female-headed households are generally less resilient than male-headed households, it is important also to consider whether the activities of resilience-building projects tend to have greater impact among female-headed or male-headed households. In fact we find no evidence that the projects evaluated overall had greater or lesser impact among female-headed households than among male-headed households.

The meta-analysis also provides useful insights about the approach used to measure resilience in these evaluations. Firstly, we investigate whether there is potential for using a standard set of resilience indicators across all evaluations, rather than adapting the indicators to the local context in each case. We find that such a standardized resilience index would produce lower

estimates of projects' impact on average – but it could arguably provide a more accurate measure of resilience, less at risk of biasing the selection of indicators towards the priorities of any particular programme. However, this also means that a standardized index would be less sensitive to the more immediate changes that are likely to arise from programme activities. Secondly, we test whether households that we assess as being more resilient in fact generally suffer less when large-scale crises occur. We find evidence that households with higher scores for the resilience index reported losing fewer livestock during recent crises (droughts or floods), but that there is no relationship between the resilience index and losses of crops. The result for livestock losses provides at least some reassurance that the resilience index is a meaningful measure of households' ability to deal with shocks.

This paper is structured as follows. Section 2 introduces the evaluations included in the meta-analysis, the evaluation methodology, and the approach used to measure resilience. Section 3 presents the results of the meta-analysis and examines differences in the results by region and by gender of the household head. Section 4 discusses ways in which the meta-analysis can help to inform the resilience measurement approach. Section 5 concludes by discussing what can be learned from the results of this meta-analysis, both for programme design in general and for the measurement of resilience specifically.

2 DATA

2.1 SELECTION OF EVALUATIONS

The results analysed in this review come from a series of 'Effectiveness Reviews' – impact evaluations carried out each year by Oxfam GB on a sample of mature or recently closed projects. Between the start of this initiative in 2011 and the time of this meta-analysis in 2016, Oxfam GB carried out 16 such Effectiveness Reviews of projects that were seeking to build resilience at a household or community level.[2] The 16 projects evaluated were selected largely at random from among all of Oxfam GB's community-level projects that were seeking to build resilience and that met a particular budget threshold.[3]

The 16 projects are listed, with a brief description of the activities of each, in Appendix 1. Results from all 16 of these impact evaluations are included in this meta-analysis. In contrast to a conventional meta-analysis, no additional criteria were used to determine whether a study is eligible to be included in the meta-analysis. All 16 Effectiveness Reviews applied a similar methodology, so criteria relating to the quality of the methodology would discriminate little between them. All of the Effectiveness Reviews carried out under Oxfam's 'resilience' theme since 2011 are included in this meta-analysis, so there is little potential for the results to be affected by publication bias.

2.2 EVALUATION METHODOLOGY

Each of the impact evaluations involved carrying out a household survey, comparing outcomes for programme participants or beneficiaries to non-participants with similar pre-project characteristics. Surveys were carried out either towards the end of each project's implementation period, or up to two years after implementation ended. In some cases, the evaluations attempted to assess impact only among households that had directly participated in project activities; in other cases, the impact of activities on a community as a whole was considered. In all 16 evaluations the respondents interviewed for comparison purposes were sampled from different communities to those in which the project was implemented, in order to minimize the possibility that the comparison respondents may have benefited indirectly from the project activities.[4]

The way in which comparison respondents were selected in each evaluation was intended to replicate, as far as possible, the methods that had been used to select the project participants or beneficiaries at the start of the project. The degree to which evaluations were successful in this respect varied. In some cases, project sites or individual participants were selected through an idiosyncratic targeting process that could not be replicated closely in the comparison areas. The consequence of this is that in these evaluations we cannot exclude the possibility that there are unobservable differences between the intervention and comparison groups that may bias the estimates of project effects in a positive direction.

In addition to a careful selection process for the comparison group, each of the evaluations used propensity score matching (PSM) to control for observable differences between the intervention and comparison groups. The variables used for matching include demographic characteristics, pre-project wealth indicators and, in some cases, indicators of pre-project engagement in particular livelihoods activities. Since pre-project data were not available, data on pre-project wealth indicators and (where applicable) on pre-project livelihoods activities were based on information recalled by respondents during the single, post-project, survey. The use of

recalled data is likely to be subject to error, which has the potential to bias our comparison of outcomes if the distribution of the errors differs systematically between the intervention and comparison groups. However, we do not have evidence on the direction or size of any such recall error, so we are unable to determine whether this would be likely to lead to project effects being under- or over-estimated.

The approach used to implement PSM is described in more detail in Appendix 2. The output of the PSM models is an estimate of the average effect of the particular project on those in the intervention group (known as the 'average treatment effect on the treated', ATET).

It should be stressed that the studies included in this meta-analysis evaluated the impact only of project activities that were carried out at a household or community level. Many of the projects evaluated also sought to contribute to resilience at a higher level – for example, through building the capacity of local or regional governments to adapt to climate change or prepare for disasters, or through influencing policy or practice at a systemic level. The success of these higher-level activities is not assessed in the 16 evaluations, nor in this meta-analysis.

2.3 MEASUREMENT OF RESILIENCE

Oxfam defines resilience as 'the ability of women and men to realize their rights and improve their wellbeing despite shocks, stresses and uncertainty' (Jeans *et al.*, 2016). In contrast to the literature that emphasizes that resilience is a property of a system, and one that should not be assumed to promote the welfare of individuals within that system (for example, Levin *et al.*, 1998), this definition focuses specifically on the welfare of individual women and men.[5] In this understanding, building resilience is seen as an intermediate outcome on the path to achieving realisation of rights or improvements in wellbeing.

Given Oxfam's definition of resilience, the ideal approach to evaluating the impact of a resilience-building intervention would be to allow time to elapse after the project or programme had ended, and observe whether the participants or beneficiaries had, in fact, been able to realize their rights and improve their wellbeing over time, in spite of whatever shocks and stresses they had encountered. However, Oxfam's Effectiveness Reviews – in common with most other impact evaluations – are attempting to evaluate the impact of resilience-building interventions only shortly after those interventions were carried out (at most 18 months after the end of a project). In most cases, we would not expect to see increased resilience leading to improvements in wellbeing over such a short time.

The outcome measures in these evaluations were therefore based on indicators of resilience – that is, characteristics that are believed to be important for people's ability to realize their rights and improve their wellbeing in the future, despite shocks, stresses and uncertainty. Part of the preparatory work for each of the evaluations involved drawing up a set of resilience indicators that were considered to be appropriate in the local context. This process typically involved discussions with programme implementation staff and local partners, and, in several cases, qualitative fieldwork. Some of the indicators were specific to the particular context, while others were identified as important in all or most of the contexts where evaluations were carried out. Examples of indicators that were used in most of the evaluations include diversification of income sources; access to productive assets, savings and credit; and the strength of social support networks. The table in Appendix 3 lists the indicators that were used most commonly across the 16 evaluations.[6]

Given that the evaluations were based on household surveys, indicators of resilience could only be included if they were observable at a household level. Clearly there are many factors that contribute to the resilience of a household that are not directly within that household's control. For example, the strength of social relations in the community, the health of local ecosystems, the robustness of essential infrastructure, the ability of local and national government to deliver

services, and the wider policy environment all have important effects on the ability of women and men to respond to shocks, stresses and uncertainty. Those characteristics may be outside the direct control of any particular household, but indicators related to them were included in the resilience index if they could be measured at a household level. For example, several of the evaluations asked survey respondents to give a subjective assessment of the ability of community leads and/or local or national government to respond to crises appropriately. However, it is recognized that using household-level indicators to measure changes that are expected to take place outside the household is not ideal.[7]

For each household surveyed, a resilience index score was created by aggregating data across the various indicators of resilience. The approach used to compile the resilience index is adapted from the Alkire-Foster method used by the Oxford Poverty and Human Development Institute for the measurement of multidimensional concepts, such as poverty and women's empowerment (Alkire and Foster, 2011). This methodology involves converting each indicator into binary form, using thresholds appropriate in the local context for a household to be considered to be faring reasonably well with respect to the characteristic in question. These thresholds were provisionally identified through conversation with local consultants, Oxfam staff and partner organisation staff. These were then verified or adjusted once the data for each evaluation were available, to ensure that important variation was not obscured by dichotomising the variables in this way.

Once thresholds have been defined for each of the resilience indicators, the overall index of resilience is calculated as the proportion of resilience indicators in which the household meets the threshold. In the absence of any clear basis on which to allocate weights between the various indicators, each indicator is assigned equal weight in the resilience index.[8]

Oxfam has recently adopted a framework to conceptualize resilience in terms of three capacities – the capacities to absorb, adapt and transform (Jeans *et al.*, 2016, following Martin-Breen and Anderies, 2011, Béné *et al.*, 2012, and others).[9] The ways in which Oxfam understands these three capacities are summarized in Box 1. In order to assess the extent to which Oxfam projects have been contributing to building each of those capacities, indices for each capacity were constructed for this meta-analysis. This involved allocating each indicator of resilience to the capacity to which it seems most closely related. There are, of course, no hard boundaries between the various categories, and many of the indicators could be classified under two of them or under all three. For example, having savings was considered to be primarily an indicator of absorptive capacity, since savings may be used to weather a crisis, even though savings could certainly also be used to fund an adaptive investment. It was particularly challenging to identify indicators of transformative capacity in the evaluations, but some of the characteristics that were measured may at least indicate that household members have the potential to contribute to transformative change. For example, indicators of participation in community groups and in community decision-making were included on the basis that a community in which more people are participating in such activities may have greater potential to make significant change.

Indices for each of the three capacities were created in a similar way to the overall index of resilience. That is, the indices were defined as the proportion of indicators of the relevant capacity in which each household reached the threshold. Again the indices were standardized before using the PSM models to estimate the effects of each project on each capacity.

2.4 META-ANALYSIS METHODOLOGY

The procedures used for the meta-analyses in this paper follow the guidance provided by Borenstein et al. (2009), Higgins and Green (2011) and Waddington et al. (2012). The resilience indices and other outcome measures discussed in this paper are standardized before analysis to ensure that they can be meaningfully aggregated across contexts. Standardisation is carried out by dividing by the standard deviation of the outcome variable as estimated from the pooled sample of intervention and comparison observations. The outcomes in this paper are therefore reported in terms of Cohen's d, a measure of standardized mean difference between groups.[10] The interpretation of standardized effect sizes may not be intuitive, but it may be helpful to be aware that Cohen (1992) characterized standardized effects of 0.2, 0.5 and 0.8 as being 'small', 'medium' and 'large' respectively, with a 'medium' effect being one that is 'likely to be visible to the naked eye of a careful observer' (p. 156).

We use meta-analysis models with random effects, which take into account that the effect of the projects on the resilience index and other outcome measures may vary between contexts. This seems appropriate, given the wide variation in the types of interventions evaluated and the various environments in which these interventions were carried out. In each meta-analysis model, the I^2 statistic, which represents the proportion of the variation between studies that is attributable to heterogeneity, provides a test of this assumption. For example, in Figure 1 below, the high value for the I^2 statistic of 96 percent (reported in the last line of the chart) provides strong evidence that projects' effects on the resilience index vary between evaluations.[11]

The analysis in this paper follows common practice in weighting each evaluation according to the inverse of the variance of its outcome estimates, modified to account for the heterogeneity between studies. The sample sizes used in the 16 evaluations are similar, so the standard errors of the outcome estimates are reasonably homogeneous: this leads to each evaluation being given approximately equal weight in most of the meta-analyses. (For example, the weights allocated to each evaluation in the meta-analysis of project effects on the resilience index can be seen on the right-hand side of Figure 1.) However, the projects evaluated varied widely in scale, as can be seen in Appendix 1. The consequence of this is that the aggregate estimates derived from the meta-analysis do not represent the effect on the *average* project participant or beneficiary household.[12] However, as we will see in Section 3.1, there is no evidence of a relationship between the scale of a project and the size of its effect on participant or beneficiary households. The meta-analysis is therefore considered to provide a reasonable guide to the size of the effect across the 16 projects as a whole.

Duvendack et al. (2012) provide a warning about including quasi-experimental studies – of which Oxfam's Effectiveness Reviews are an example – in a meta-analysis. They have three concerns: firstly, that the imperfect identification strategies used in quasi-experimental methods may bias the conclusions; secondly, that there may be heterogeneity between studies in the methodology applied in different studies and the treatment effects being estimated; and thirdly, that 'researcher allegiance' can lead to publication bias or other forms of positive bias in results. The second of these concerns does not apply in our case: all the evaluations applied a common PSM approach, and they all estimate the average treatment effect on the treated. However, the other two points warrant some discussion.

The concern about bias in the underlying evaluations is clearly relevant to this meta-analysis: the results discussed here are valid only insofar as the identification assumptions made in each of the underlying evaluations are valid. As discussed in Section 2.2, the evaluation teams attempted to select the intervention and comparison observations in a way that would minimize both observable and unobservable differences between them. Remaining observable differences were controlled for using PSM models at the analysis stage, but the extent to which they were successful in controlling for unobservable characteristics cannot be known.

With respect to the third of Duvendack et al.'s concerns, we can at least eliminate the possibility of publication bias: all 16 of Oxfam's Effectiveness Reviews carried out under the 'resilience' theme up to early 2016 are included in this meta-analysis. Whether there is potential for the results to be affected by subtler forms of 'researcher bias' is more difficult to assess. However, the evaluators, despite being employed by or contracted by Oxfam, have a high degree of autonomy within the organisation and have frequently published results that are less favourable than programme implementers may have expected.

What can be said with confidence is that the Effectiveness Review data are the most robust data available on the impacts of Oxfam's resilience-building work. This meta-analysis, therefore, represents the organisation's best attempt to systematically analyse and learn from those impacts.

3 META-ANALYSIS RESULTS

3.1 OVERALL IMPACT ON RESILIENCE

Figure 1 shows the results of a meta-analysis for the estimated impact of the 16 different projects on the resilience index among participant households. For each evaluation, the difference in the resilience index between intervention and comparison households was estimated through PSM. In Figure 1, and other plots in this paper, the estimated effect size of each evaluation is shown as a point, with the horizontal bars representing the corresponding 95 percent confidence interval. The diamond shapes represent the 95 percent confidence intervals for the average effects across evaluations, aggregated through meta-analysis.

Overall, the 16 projects evaluated are estimated to have resulted in an increase in the resilience index of 0.40 standard deviations, on average, with a 95 percent confidence interval ranging from 0.18 to 0.63 standard deviations.

As we noted in the previous section, the projects evaluated varied widely in scale, but they are each given approximately equal weight in the meta-analysis. Weighting the projects equally in this way may seem counter-intuitive: in particular, it may be objected that projects that have focused their resources on working with a small number of participants are likely (all else being equal) to achieve greater impact on the average participant than those that have sought to work with a much larger number of participants. In fact, the data do not provide evidence for any such relationship between the scale of the project and the effect size.[13] For example, four of the projects evaluated had worked with particularly small numbers of participants (fewer than 200 households in each case): excluding these four evaluations from the meta-analysis results in an estimate of the aggregate effect that is almost unchanged from that shown in Figure 1.

It can be seen from the list of projects in Appendix 1 that there was also considerable variation in the *duration* of the projects evaluated. It may be natural to assume that projects of a longer duration would have greater impact, but again the data provide no evidence of any such relationship.[14] The likely explanation for this is that the duration is of a project is closely related to the types of interventions carried out. Many of the shorter-term projects were focused on reducing vulnerability to natural hazards – particularly flooding – through activities such as establishing early-warning systems or providing training on how to respond to an imminent threat. In contrast, longer-term projects generally had more emphasis on strengthening the resilience of livelihoods activities, such as through supporting households in diversifying their livelihoods activities, a longer-term endeavour for which the outcomes are less certain. Projects that focused on disaster-risk reduction tend to show more positive results than projects those that mainly worked on building livelihoods resilience.[15]

Figure 1: Random effects meta-analysis for the overall index of resilience

Country	Year	Number of indicators		Effect size (95% CI)	% Weight
Africa					
Mali	2011/12	13		-0.09 (-0.19, 0.01)	6.61
Niger	2011/12	12		-0.05 (-0.25, 0.15)	6.36
Ethiopia	2012/13	31		0.21 (-0.06, 0.47)	6.12
Kenya	2012/13	20		0.16 (-0.03, 0.34)	6.41
Zambia	2012/13	27		0.33 (-0.02, 0.68)	5.74
Mali	2013/14	16		0.07 (-0.06, 0.21)	6.54
Niger	2013/14	22		-0.04 (-0.22, 0.15)	6.39
Chad	2014/15	19		0.36 (0.22, 0.50)	6.52
Subtotal (I-squared = 78.0%, p = 0.000)				0.10 (-0.03, 0.23)	50.68
.					
Asia					
Indonesia	2011/12	15		0.38 (0.20, 0.56)	6.42
Pakistan	2011/12	14		1.14 (1.02, 1.27)	6.56
Nepal	2012/13	23		0.81 (0.36, 1.26)	5.25
Nepal	2013/14	21		1.61 (1.30, 1.91)	5.95
Pakistan	2013/14	28		0.60 (0.43, 0.78)	6.44
Thailand	2014/15	16		0.51 (0.36, 0.66)	6.50
Subtotal (I-squared = 94.8%, p = 0.000)				0.83 (0.49, 1.17)	37.12
.					
Latin America					
Nicaragua	2014/15	23		0.52 (0.21, 0.84)	5.91
Bolivia	2015/16	18		0.09 (-0.13, 0.31)	6.29
Subtotal (I-squared = 79.9%, p = 0.026)				0.29 (-0.14, 0.72)	12.20
.					
Overall (I-squared = 95.9%, p = 0.000)				0.40 (0.18, 0.63)	100.00

Standardised mean difference between intervention and comparison group

3.2 IMPACT ON ABSORPTIVE, ADAPTIVE AND TRANSFORMATIVE CAPACITIES

As described in Section 2.3, Oxfam has recently adopted a framework to conceptualize resilience in terms of absorptive, adaptive and transformative capacities. Indices were created for each of the three capacities by allocating each of the resilience indicators to the capacity to which it was most closely linked, and then calculating the proportion of the corresponding indicators in which each household met the binary threshold. Again, project effects were estimated by PSM. The meta-analysis for projects' effects on the three capacities is shown in Figures 2 to 4.

The results show a positive effect, on average, for each of the three capacities, with estimated effect sizes of 0.37, 0.24 and 0.23 standard deviations for absorptive, adaptive and transformative capacities respectively. This suggests that Oxfam projects are building resilience across all three capacities. It should be reiterated that the index of transformative capacity relies on a generous interpretation of which characteristics can be considered 'transformative' – this includes indicators such as whether household members participate in community groups and whether they have attended any community meetings. A full assessment of projects' impacts on transformative capacity – as well as on systemic aspects of absorptive and adaptive capacities – would require a different approach to evaluation and data collection.

Figure 2: Random effects meta-analysis for index of absorptive capacity

Country	Year	Number of indicators		Effect size (95% CI)	% Weight
Africa					
Mali	2011/12	4		-0.13 (-0.24, -0.03)	6.53
Niger	2011/12	6		-0.10 (-0.31, 0.11)	6.31
Ethiopia	2012/13	14		0.16 (-0.11, 0.44)	6.11
Kenya	2012/13	10		0.21 (0.00, 0.42)	6.32
Zambia	2012/13	12		0.30 (-0.05, 0.64)	5.85
Mali	2013/14	7		-0.07 (-0.20, 0.07)	6.49
Niger	2013/14	11		0.11 (-0.09, 0.31)	6.33
Chad	2014/15	8		0.29 (0.14, 0.45)	6.44
Subtotal (I-squared = 76.5%, p = 0.000)				0.08 (-0.05, 0.21)	50.38
.					
Asia					
Indonesia	2011/12	5		0.29 (0.12, 0.45)	6.43
Pakistan	2011/12	6		1.09 (0.97, 1.21)	6.51
Nepal	2012/13	9		0.66 (0.15, 1.16)	5.18
Nepal	2013/14	10		1.70 (1.49, 1.92)	6.30
Pakistan	2013/14	13		0.37 (0.23, 0.51)	6.47
Thailand	2014/15	5		0.59 (0.43, 0.75)	6.43
Subtotal (I-squared = 97.0%, p = 0.000)				0.78 (0.37, 1.20)	37.31
.					
Latin America					
Nicaragua	2014/15	10		0.41 (0.12, 0.71)	6.04
Bolivia	2015/16	7		0.07 (-0.15, 0.30)	6.27
Subtotal (I-squared = 68.6%, p = 0.074)				0.23 (-0.10, 0.56)	12.32
.					
Overall (I-squared = 96.7%, p = 0.000)				0.37 (0.12, 0.62)	100.00

Standardised mean difference between intervention and comparison group

Figure 3: Random effects meta-analysis for index of adaptive capacity

Country	Year	Number of indicators		Effect size (95% CI)	% Weight
Africa					
Mali	2011/12	9		-0.07 (-0.17, 0.04)	6.82
Niger	2011/12	5		0.06 (-0.14, 0.27)	6.38
Ethiopia	2012/13	11		0.15 (-0.14, 0.43)	5.94
Kenya	2012/13	7		0.06 (-0.11, 0.24)	6.55
Zambia	2012/13	9		0.17 (-0.10, 0.44)	6.02
Mali	2013/14	6		0.19 (0.08, 0.29)	6.82
Niger	2013/14	8		-0.19 (-0.39, 0.00)	6.44
Chad	2014/15	8		0.21 (0.08, 0.34)	6.74
Subtotal (I-squared = 69.9%, p = 0.002)				0.07 (-0.03, 0.18)	51.71
.					
Asia					
Indonesia	2011/12	8		0.28 (0.10, 0.45)	6.55
Pakistan	2011/12	7		1.08 (0.95, 1.21)	6.73
Nepal	2012/13	9		0.41 (0.10, 0.72)	5.74
Nepal	2013/14	6		0.13 (-0.41, 0.68)	4.21
Pakistan	2013/14	11		0.61 (0.43, 0.80)	6.50
Thailand	2014/15	10		0.36 (0.21, 0.51)	6.67
Subtotal (I-squared = 93.7%, p = 0.000)				0.50 (0.18, 0.82)	36.39
.					
Latin America					
Nicaragua	2014/15	11		0.42 (0.09, 0.75)	5.62
Bolivia	2015/16	7		0.00 (-0.22, 0.23)	6.28
Subtotal (I-squared = 75.7%, p = 0.043)				0.19 (-0.21, 0.60)	11.90
.					
Overall (I-squared = 93.8%, p = 0.000)				0.24 (0.07, 0.42)	100.00

-2   -1.5   -1   -.5   0   .5   1   1.5   2

Standardised mean difference between intervention and comparison group

Figure 4: Random effects meta-analysis for index of transformative capacity

Country	Year	Number of indicators	Effect size (95% CI)	% Weight
Africa				
Niger	2011/12	1	-0.14 (-0.33, 0.06)	7.07
Ethiopia	2012/13	6	0.17 (-0.08, 0.41)	6.49
Kenya	2012/13	3	-0.01 (-0.19, 0.16)	7.31
Zambia	2012/13	6	0.33 (-0.06, 0.71)	4.89
Mali	2013/14	3	-0.01 (-0.17, 0.15)	7.45
Niger	2013/14	3	-0.11 (-0.34, 0.12)	6.67
Chad	2014/15	3	0.17 (0.03, 0.32)	7.66
Subtotal (I-squared = 49.4%, p = 0.065)			0.04 (-0.07, 0.14)	47.53
Asia				
Indonesia	2011/12	2	0.26 (0.08, 0.44)	7.26
Pakistan	2011/12	1	0.21 (0.06, 0.37)	7.55
Nepal	2012/13	5	0.70 (0.30, 1.10)	4.75
Nepal	2013/14	5	1.28 (0.96, 1.60)	5.60
Pakistan	2013/14	4	0.28 (0.11, 0.45)	7.38
Thailand	2014/15	1	0.13 (-0.04, 0.30)	7.41
Subtotal (I-squared = 88.8%, p = 0.000)			0.44 (0.19, 0.69)	39.94
Latin America				
Nicaragua	2014/15	2	0.55 (0.25, 0.86)	5.77
Bolivia	2015/16	4	0.11 (-0.12, 0.33)	6.75
Subtotal (I-squared = 80.9%, p = 0.022)			0.32 (-0.12, 0.75)	12.53
Overall (I-squared = 83.7%, p = 0.000)			0.23 (0.10, 0.36)	100.00

Standardised mean difference between intervention and comparison group

3.3 REGIONAL DIFFERENCES IN EFFECT SIZE

Figures 1 to 4 show that the evaluations carried out in Asia have tended to have more positive results than those carried out in Africa and Latin America. This is clearly the case for the results of the overall resilience index (Figure 1), but the same applies to the results for absorptive and adaptive capacities, and perhaps (with less clarity), to transformative capacity as well.[16] This pattern persists even if the two evaluations with particularly positive results, those conducted in Pakistan in 2011/12 and in Nepal in 2013/14, are excluded from the analysis. The effects on the various indices of resilience for projects in Africa and Latin America are not statistically significantly different from zero. It is difficult to draw a general conclusion about results in Latin America from the two projects evaluated in that region. However, the aggregate results for projects in Africa are more concerning, given that several evaluations have been carried out in the region, and that the confidence intervals are correspondingly smaller.

One possible reason for the apparent differences in outcomes between the regions is in the nature of the shocks (at least the large-scale climate shocks) that they typically experience. Several of the projects in Asia were carried out in contexts where the primary concern is flooding, a hazard that affects a particular geographic area over a particular period of time, and for which the steps needed to reduce risk are relatively well understood. In contrast, it seems likely that having a positive impact on resilience is more challenging in arid zones that are

subject to prolonged climate stress, where most of the evaluations in Africa were carried out. A related point is that there were differences between the regions in the nature of the projects being evaluated. It is clear from Appendix 1 that many of the projects in Asia have had more focus on disaster-risk reduction interventions, whereas those in Africa and Latin America were mostly seeking to strengthen the resilience of livelihoods activities. We noted in Section 3.1 that projects focused on disaster-risk reduction tended to show more positive results than those focused on livelihoods resilience.

It is also important to consider whether there may be factors inherent in the evaluation process that may bias the comparison out outcomes between Asia and the other regions of the world. One possibility is that changes in outcomes brought about through disaster-risk reduction activities are easier to observe than changes in livelihoods resilience. Indeed, evaluations of projects that had a more risk-reduction focus tended to include indicators – such as access to an early-warning system and participation in risk-reduction committees – that are closely linked to the delivery of project activities and that are reasonably straightforward to measure. In contrast, evaluations that were more focused on livelihoods resilience placed more emphasis on indicators such as diversification in activities, building up assets or savings, and adoption of innovative practices, all of which are less directly connected to the project itself and more likely to be prone to measurement error. Since measurement error tends to attenuate estimates of project effects, this could lead to projects with a disaster-risk reduction focus appearing to have greater impact than those with a livelihoods focus, even if there were in reality no difference between the results of the two types of project.

Some evidence that this may be happening is provided in Section 4.1, where we construct an alternative 'generic' index of resilience using a set of indicators that is more consistent across the 16 evaluations. Using this modified index, there is no significant difference in project effect estimates between disaster-risk reduction and livelihoods resilience projects. However, even using the generic index of resilience, there is still a clear difference in effect estimates between Asia and the other two regions ($p < 0.01$). The different nature of the projects, then, does not fully explain the difference in results between the regions.

Another potential source of bias in comparing projects in Asia to those in Africa and Latin America may arise from the identification strategy used. In particular, the way the comparison respondents were selected was specific to each evaluation. As discussed in Section 2.2, there is potential in these evaluations for unobserved differences between the intervention and comparison groups to bias the estimates of project effects. Such differences may arise, for example, from the way that individual survey respondents are selected. In the case of projects for which the participants were selected (or self-selected) through an idiosyncratic process that could not be fully replicated in the comparison communities, it is quite possible that they tend to differ from the comparison respondents in terms of (for example) their initiative, attitudes to risk or innovation, or social connections. If these unobserved differences are positively associated with resilience, this would bias estimates of project impacts upwards.

However, in the case of all but one of the evaluations in Asia (that conducted in Thailand in 2014/15), respondents in the intervention and comparison communities were sampled at random from among the population of the community as a whole, so there should be little potential for results to be influenced by individual-level unobservable factors in selection. In contrast, three of the eight evaluations in Africa were carried out among a selected group of project participants, as were both of those in Latin America. This suggests that it is unlikely that the possibility of individual-level or household-level unobservable factors helps to explain the more positive results in Asia.

A second potential source of unobservable bias in the estimates of project impact arises from the way that communities themselves were selected for the survey. All the evaluations aimed to ensure that the comparison communities selected for the surveys were as similar as possible to the communities where the project was implemented, in terms of, for example, their agro-ecological conditions and their access to markets, roads, electricity networks and other

infrastructure. However, in each case the project implementers had decided to carry out the project in particular communities rather than in the communities that were later selected for comparison. In most cases, the specific reasons for these local-level decisions – made during the initial stages of planning for a project, years before the evaluation – were unknown to the evaluators. It is likely that these decisions on implementation sites were influenced, in at least some cases, by the willingness of community members to engage in a new project or by the proactiveness of local leaders. Again, these are factors that are likely to be associated with higher resilience in the community, so they could result in estimates of project impacts being biased upwards. If such unobservable factors in selecting implementation sites were generally more important in the evaluations in Asia than in Africa and Latin America, then this would lead to the projects in Asia appearing to be generally more successful. Unfortunately, there is no evidence available from which to judge whether or not this is the case. Such effects would anyway have to be large to fully account for the observed difference between the regions.

In summary, then, the large difference in projects' impact on resilience between Asia and the other two regions may partly be due to the Asian projects' greater emphasis on disaster-risk reduction activities – for which it may be possible to achieve changes in resilience indicators over a shorter time frame, or at least that changes in those indicators may be easier to measure. Differences between the regions in the extent to which unobserved bias may affect the estimates of project impact could also contribute to this difference. However, neither of these factors seem large enough to account for the extent of the observed difference between the results in Asia and those in Africa and Latin America.

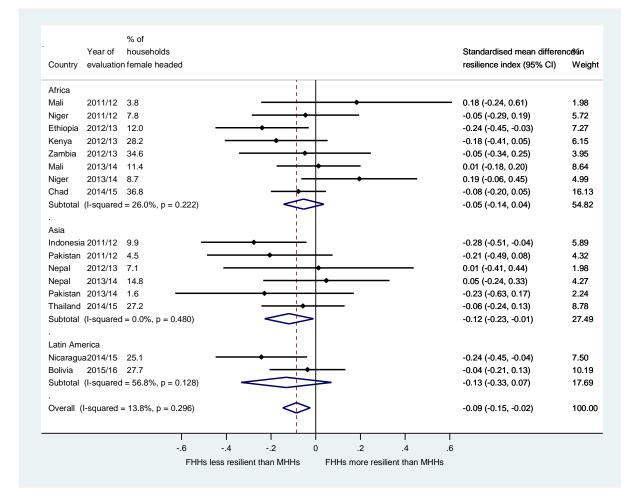# 3.4 DIFFERENCES IN EFFECT SIZE BY GENDER OF THE HOUSEHOLD HEAD

Many of the projects evaluated were mainly or exclusively targeted at building resilience among women. The objective of the evaluations was explicitly to evaluate projects' impacts on resilience at the level of the household, and interviews were generally carried out with only a single member of each household. For this reason, the data available are not revealing about differences in resilience *within* households, and in particular differences in resilience between women and men.

Nevertheless, it is possible to examine differences in the resilience index between households headed by women and men. This comparison is shown in Figure 5. The decision as to which household member to identify as the head was made by the respondent at the start of each interview.[17] Since female-headed households tend to be smaller and poorer than male-headed households, indicators of household composition and the households' economic level are controlled for in making this comparison.[18] The results for some of the individual evaluations in Figure 5 have large confidence intervals, since in some of the datasets there are few female-headed households. However, the overall result is clear: female-headed households have scores on the resilience index that are on average approximately 0.1 standard deviations lower than male-headed households that have similar observable characteristics. The I^2 statistic of 14 percent shows that this difference between female-headed and male-headed households is quite consistent across all the contexts in which the evaluations were carried out. The size of the difference in resilience index between female-headed and male-headed households is consistent across the three capacities of resilience.

In interpreting this finding, it is important to consider whether this difference could be a result of the measurement approach itself – that is, whether the indicators being applied have tended to be biased towards higher scores for male-headed households. This could be the case, for example, if male-dominated livelihoods activities were being promoted in the indicators. Given that many of the projects evaluated had a strong gender focus, or had specifically worked with women, and that the selection of indicators was intended to be appropriate to understand the

impact of the project activities in each case, it seems unlikely that the indicators are generally biased towards giving higher scores to male-headed households.

Figure 5: Meta-analysis for regressions of resilience index on the household being female-headed, controlling for observable characteristics

Country	Year of evaluation	% of households female headed	Standardised mean difference in resilience index (95% CI)	% in Weight
Africa				
Mali	2011/12	3.8	0.18 (-0.24, 0.61)	1.98
Niger	2011/12	7.8	-0.05 (-0.29, 0.19)	5.72
Ethiopia	2012/13	12.0	-0.24 (-0.45, -0.03)	7.27
Kenya	2012/13	28.2	-0.18 (-0.41, 0.05)	6.15
Zambia	2012/13	34.6	-0.05 (-0.34, 0.25)	3.95
Mali	2013/14	11.4	0.01 (-0.18, 0.20)	8.64
Niger	2013/14	8.7	0.19 (-0.06, 0.45)	4.99
Chad	2014/15	36.8	-0.08 (-0.20, 0.05)	16.13
Subtotal (I-squared = 26.0%, p = 0.222)			-0.05 (-0.14, 0.04)	54.82
.				
Asia				
Indonesia	2011/12	9.9	-0.28 (-0.51, -0.04)	5.89
Pakistan	2011/12	4.5	-0.21 (-0.49, 0.08)	4.32
Nepal	2012/13	7.1	0.01 (-0.41, 0.44)	1.98
Nepal	2013/14	14.8	0.05 (-0.24, 0.33)	4.27
Pakistan	2013/14	1.6	-0.23 (-0.63, 0.17)	2.24
Thailand	2014/15	27.2	-0.06 (-0.24, 0.13)	8.78
Subtotal (I-squared = 0.0%, p = 0.480)			-0.12 (-0.23, -0.01)	27.49
.				
Latin America				
Nicaragua	2014/15	25.1	-0.24 (-0.45, -0.04)	7.50
Bolivia	2015/16	27.7	-0.04 (-0.21, 0.13)	10.19
Subtotal (I-squared = 56.8%, p = 0.128)			-0.13 (-0.33, 0.07)	17.69
.				
Overall (I-squared = 13.8%, p = 0.296)			-0.09 (-0.15, -0.02)	100.00

-.6   -.4   -.2   0   .2   .4   .6

FHHs less resilient than MHHs   FHHs more resilient than MHHs

This result is investigated further in Fuller and Lain (2017). The difference between female-headed and male-headed households is reduced once the effects of the projects being evaluated are controlled for, but it appears to be robust to alternative specifications and reasonably robust to the possibility of unobserved heterogeneity between the groups. Approximately half of the measured difference in resilience between female-headed and male-headed households appears to be accounted for by observable characteristics (including household composition, economic level and participation in the project), with the remaining half accounted for by a combination of unobserved differences and, it appears, an inherent difference in resilience by the gender of the household head.

If female-headed households are generally less resilient than male-headed households, the natural next question to ask is whether the activities of resilience-building projects tend to have greater impact among female-headed or male-headed households. This question is analysed in Figure 6, where the results represent the estimated difference between female-headed and male-headed households in the size of projects' effects on the resilience index.[19] Overall, there is no indication of any pattern regarding whether female-headed or male-headed households have benefited more. Two of the projects (in Kenya and Pakistan) appear to have had less impact among female-headed households than male-headed households, while two others (in Zambia and Niger) had greater impact among female-headed households. It is notable that the project in Niger had specifically worked with vulnerable women – a proportion of whom were household heads – whereas the typical male-headed household benefited only from the wider community-level interventions. On the other hand, the project in Kenya was also targeted mainly

at women as direct participants, and the project in Zambia does not seem to have had any particular prioritisation of gender issues. These cases seem to provide little guidance, then, in understanding whether a project is likely to have higher impact among female-headed or male-headed households.

Figure 6: Meta-analysis for the differential effect on female-headed households against male-headed households, from propensity-score weighted regression models of project impact

Country	Year	Proportion of households female headed		Difference in effect on resilience index (95% CI)	% Weight
Africa					
Mali	2011/12	1.8%		-0.17 (-1.15, 0.80)	2.76
Niger	2011/12	6.9%		0.03 (-0.56, 0.63)	5.08
Ethiopia	2012/13	11.9%		-0.20 (-0.65, 0.26)	6.43
Kenya	2012/13	33.4%		-0.46 (-0.91, -0.01)	6.48
Zambia	2012/13	38.8%		0.28 (-0.04, 0.60)	7.88
Mali	2013/14	12.7%		0.19 (-0.20, 0.58)	7.14
Niger	2013/14	6.0%		0.72 (0.09, 1.35)	4.80
Chad	2014/15	39.5%		0.21 (-0.02, 0.44)	8.87
Subtotal (I-squared = 49.5%, p = 0.054)				0.09 (-0.12, 0.31)	49.43
.					
Asia					
Indonesia	2011/12	9.4%		0.20 (-0.26, 0.65)	6.44
Pakistan	2011/12	3.9%		-0.20 (-0.87, 0.48)	4.48
Nepal	2012/13	4.1%		-0.35 (-0.98, 0.29)	4.74
Nepal	2013/14	14.2%		0.14 (-0.04, 0.31)	9.38
Pakistan	2013/14	1.4%		-1.09 (-1.52, -0.67)	6.71
Thailand	2014/15	26.9%		0.27 (-0.20, 0.73)	6.35
Subtotal (I-squared = 83.8%, p = 0.000)				-0.16 (-0.59, 0.26)	38.09
.					
Latin America					
Nicaragua	2014/15	30.5%		-0.38 (-0.86, 0.10)	6.17
Bolivia	2015/16	30.1%		0.13 (-0.33, 0.60)	6.30
Subtotal (I-squared = 56.0%, p = 0.132)				-0.12 (-0.62, 0.38)	12.47
.					
Overall (I-squared = 69.9%, p = 0.000)				-0.03 (-0.22, 0.16)	100.00

-1.5   -1   -.5   0   .5   1   1.5

Greater impact among MHHs   Greater impact among FHHs

4 EXPLORING THE RESILIENCE MEASUREMENT APPROACH

4.1 MOST COMMON INDICATORS OF RESILIENCE

We stressed in Section 3 that indicators of resilience are selected so as to be appropriate for the local context in each case. Nevertheless, there are particular indicators that have been identified as relevant in many contexts and so used in many of the evaluations. Such indicators include diversification of income sources, diversification in crops cultivated, savings, access to credit, access to productive assets, strength of social support networks, attitudes to change and awareness or understanding of climate change. Appendix 3 provides a list of the indicators that were included most often in the 16 evaluations. The precise definitions of these indicators and the binary thresholds used to assess them varied between evaluations, but the concepts that they are attempting to measure are consistent across evaluations.

To inform the selection of resilience indicators for future evaluations, it is useful to examine which of these common indicators are most closely associated with the results of the resilience index. For example, a few of the most common indicators – notably the strength of social support networks, attitudes to change, awareness or understanding of climate change, access to extension or other state services, and participation in community groups – appear to be good predictors of impact on the resilience index as a whole. For each of these indicators, in all the evaluations in which a positive project effect was found for the specific indicator (with $p < 0.1$), a positive project effect was also found on the resilience index as a whole. This suggests that these indicators may be *sufficient* predictors of a project having a positive effect on the resilience index. However, that does not imply they are *necessary* conditions: there are multiple evaluations in which the project had a significant effect on the resilience index, but no effects on those indicators.

The indicators that do seem to be necessary conditions for impact on the resilience index are those that tend to be more directly linked to the project activities – for example, the adoption of specific improved production activities or disaster-preparedness activities promoted under the project, access to disaster early-warning information, or participation in community or district-level decision-making. The main exception to this is the indicator that arguably measures adaptive capacity most directly: that is, whether households have adopted any innovative practices in the recent past (other than those promoted by the project). The analysis in Appendix 3 suggests that this indicator is a sufficient condition and (almost) a necessary condition for impact on the overall resilience index.

The fact that several of the indicators of resilience have been used in many of the evaluations raises the question of whether resilience could be measured purely by using a standard set of indicators that are consistent across contexts. If so, that would eliminate the need to tailor the measurement approach to the local context in each case. To investigate this, a new, 'generic' index of resilience was created, based only on the indicators that were included in at least eight of the 16 evaluations. There are 14 such indicators (the first 14 indicators listed in the table in Appendix 3). Only two of the evaluation datasets include measures relating to all 14 of those indicators, but on average the evaluations measured 10 of the 14. In each evaluation dataset,

the generic index of resilience was calculated as the proportion of indicators in which the household scored positively among all the indicators from the list of 14 for which data were available.

Despite being composed of only a subset of the original set of indicators, the new 'generic' index of resilience was found to have a high correlation with the full resilience index: the correlation coefficient was 0.82 on average.[20] It is therefore natural to consider how accurate the generic index would have been in estimating the effects of projects on resilience. Figure 7 shows a meta-analysis for the estimates of project effects on the generic index, using the same PSM models that have been used to generate the results earlier in this paper. Comparing Figure 1 and Figure 7, it is clear that the generic index tends to produce lower estimates of each project's impact. In particular, the 95 percent confidence interval for the effect across all the evaluations is reduced from 0.40 to 0.19 standard deviations.

The lower estimates of project impact found when using the generic resilience index are probably a consequence of the generic indicators focusing mostly on higher-level outcomes and being less closely related to specific project activities. On that basis, it could be argued that the generic index represents a *more* accurate picture of how resilience has changed as an outcome of those activities, despite the loss of the contextually-specific indicators. However, we should be cautious in making this claim, or in concluding that the set of 14 generic indicators considered here can provide a good overall measure of resilience in all situations. We have used the same set of evaluations to identify the most common indicators and then to test their correlation with the overall resilience index; a better test would be to examine the correlation of those indicators with the overall resilience index in new datasets.

Figure 7: Random effects meta-analysis for 'generic' resilience index

Country	Year	Number of indicators		Effect size (95% CI)	% Weight
Africa					
Mali	2011/12	6		-0.12 (-0.23, -0.02)	6.86
Niger	2011/12	4		-0.22 (-0.41, -0.02)	6.50
Ethiopia	2012/13	14		0.05 (-0.19, 0.29)	6.26
Kenya	2012/13	9		0.09 (-0.08, 0.27)	6.59
Zambia	2012/13	13		0.02 (-0.35, 0.38)	5.48
Mali	2013/14	9		-0.00 (-0.16, 0.16)	6.66
Niger	2013/14	12		-0.28 (-0.47, -0.09)	6.50
Chad	2014/15	11		0.14 (-0.01, 0.30)	6.68
Subtotal (I-squared = 65.0%, p = 0.006)				-0.04 (-0.15, 0.06)	51.52
.					
Asia					
Indonesia	2011/12	7		-0.09 (-0.26, 0.07)	6.63
Pakistan	2011/12	7		0.83 (0.70, 0.95)	6.80
Nepal	2012/13	14		0.66 (0.26, 1.06)	5.24
Nepal	2013/14	13		0.93 (0.44, 1.42)	4.61
Pakistan	2013/14	13		0.41 (0.25, 0.58)	6.62
Thailand	2014/15	9		0.45 (0.29, 0.60)	6.66
Subtotal (I-squared = 93.8%, p = 0.000)				0.51 (0.19, 0.82)	36.56
.					
Latin America					
Nicaragua	2014/15	10		0.32 (-0.03, 0.66)	5.58
Bolivia	2015/16	12		0.13 (-0.10, 0.35)	6.34
Subtotal (I-squared = 0.0%, p = 0.369)				0.18 (-0.01, 0.37)	11.92
.					
Overall (I-squared = 93.4%, p = 0.000)				0.19 (0.01, 0.37)	100.00

Standardised mean difference between intervention and comparison group

4.2 IS THE RESILIENCE INDEX ASSOCIATED WITH ABILITY TO COPE BETTER WITH SHOCKS, STRESSES AND UNCERTAINTY?

As discussed earlier, the ideal approach to evaluating projects' impact on resilience would be to observe how project participants are affected by and able to respond to shocks, stresses and uncertainty for some period after the end of the project. The evaluations included in this meta-analysis adopted the approach of identifying drivers of resilience and creating an index of resilience because it is normally necessary to evaluate impact on a short time-scale. Nevertheless, several of the evaluations were carried out in areas where there had been a large-scale crisis that had affected most households in the area – either a drought or a flood – during the project's lifetime or shortly after the end of the project. In the eight cases in which this had occurred, data were collected on losses that households had experienced due to those crises. A natural question to ask, then, is whether the scores in the resilience index were associated with households experiencing lower losses from these crises.

There are four important reasons why examining losses suffered in crises is not an ideal test of the accuracy of the resilience index. Firstly, the crises all occurred at some time prior to the survey, whereas the resilience index is attempting to measure how well households can cope with crises in the future. The effects of past crises may have rendered households either better able or less able to cope with future crises – for example, through prompting people to make a proactive adaptation, or through the loss of productive assets. Secondly, most of the crises on which data were collected occurred during the lifetime of the projects being evaluated, and in many cases projects provided short-term humanitarian support to participants to enable them to cope with those crises: this could affect the losses that households suffered without necessarily affecting their resilience to future crises. Thirdly, increased resilience should not only mean that a household is able to avoid losses during a crisis, but also to recover more quickly. Finally, and most fundamentally, the resilience index seeks to measure households' ability to deal with shocks, stresses and uncertainty of all kinds; any particular crisis provides only a partial test of this.

With these caveats in mind, the relationship between the resilience index and losses suffered due to crises was investigated using data from the eight evaluations for which that was possible. Figures 8 and 9 show the results of meta-analyses for regressions of losses of crops and livestock respectively on the resilience index, after controlling for demographic characteristics and baseline wealth. Crop losses were measured using a subjective question in which respondents were asked to state what proportion of the harvest they had expected was lost due to the particular crisis, on a five-point scale. Livestock losses were measured simply as the number of heads of livestock lost during the crisis, controlling for the number of livestock owned at baseline.[21]

The results shown in Figure 8 do not provide any evidence of a negative correlation between the resilience index and crop losses experienced (except in the case of the evaluation in Nepal in 2012/13). However, Figure 9 shows that there is a significant relationship overall between the resilience index and the number of livestock lost across the five evaluations in which data on livestock losses were collected. Specifically, a one standard deviation increase in the resilience index is associated with a 0.05 standard deviation reduction in the number of livestock lost (with the 95 percent confidence interval of the reduction ranging from 0.02 to 0.09). The I^2 statistic of zero implies that this correlation is of a consistent size across those five evaluations. The 'generic' resilience index discussed in Section 4.1 is found to have a correlation with livestock losses that is similar in magnitude.

Figure 8: Meta-analysis for regressions of crop losses experienced during crises on the resilience index, controlling for gender of household head, number of adults in the household and pre-project wealth indicators

Country	Year	Effect on crop losses (95% CI)	% Weight
Pakistan	2011/12	0.07 (0.01, 0.13)	17.57
Ethiopia	2012/13	-0.13 (-0.33, 0.07)	15.60
Nepal	2012/13	-0.43 (-0.50, -0.37)	17.56
Zambia	2012/13	0.03 (-0.08, 0.14)	17.10
Nepal	2013/14	0.22 (-0.01, 0.46)	14.92
Pakistan	2013/14	0.07 (-0.02, 0.17)	17.25
Overall (I-squared = 97.1%, p = 0.000)		-0.03 (-0.26, 0.19)	100.00

-.4   -.2   0   .2   .4
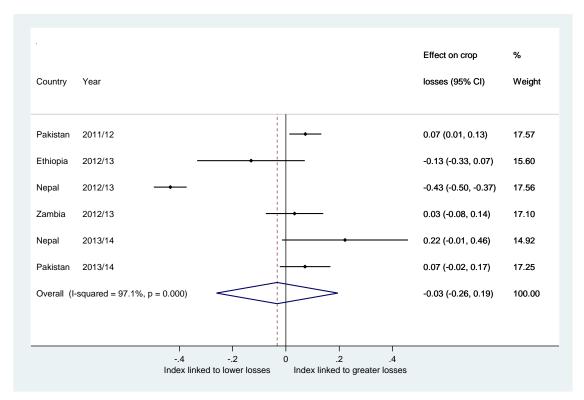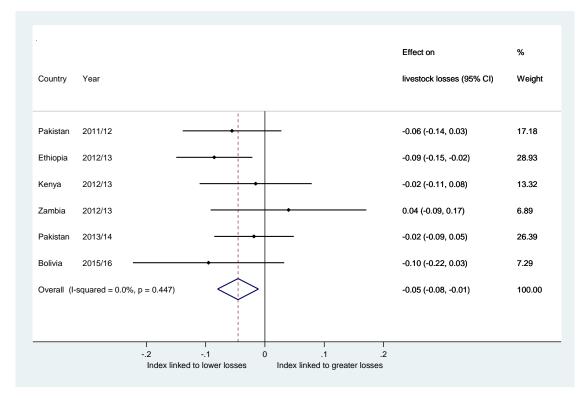Index linked to lower losses   Index linked to greater losses

Figure 9: Meta-analysis for regressions of livestock losses experienced during crises on the resilience index, controlling for gender of household head, number of adults in the household and pre-project wealth indicators

Country	Year	Effect on livestock losses (95% CI)	% Weight
Pakistan	2011/12	-0.06 (-0.14, 0.03)	17.18
Ethiopia	2012/13	-0.09 (-0.15, -0.02)	28.93
Kenya	2012/13	-0.02 (-0.11, 0.08)	13.32
Zambia	2012/13	0.04 (-0.09, 0.17)	6.89
Pakistan	2013/14	-0.02 (-0.09, 0.05)	26.39
Bolivia	2015/16	-0.10 (-0.22, 0.03)	7.29
Overall (I-squared = 0.0%, p = 0.447)		-0.05 (-0.08, -0.01)	100.00

-.2   -.1   0   .1   .2
Index linked to lower losses   Index linked to greater losses

In spite of the provisos mentioned above, this finding may provide some confidence that the resilience index is accurately reflecting households' ability to deal with large-scale crises. Indeed, given that the measurement of crop losses relied on respondents' subjective

assessments, we have reason to be more confident in the positive result related to livestock losses than in the ambiguous result relating to crop losses.

However, the correlation between the resilience index and losses of livestock experienced in five specific contexts cannot provide a full test of the use of the resilience index for predicting households' ability to improve their wellbeing in the future. It remains important to seek opportunities to carry out more comprehensive tests of the resilience index against people's actual experience of shocks, stresses and uncertainty, and their effect on wellbeing over time.

5 CONCLUSIONS

This meta-analysis provides Oxfam with several insights into the impact that its community-level projects have on the resilience of beneficiary households that were not observable from the individual evaluation reports. The projects evaluated are estimated to have had a significant positive effect on resilience overall, as well as on each of the three underlying components of resilience: absorptive, adaptive and transformative capacity. (A caveat to this finding is that the assessment of transformative capacity in these evaluations relies on quite a generous interpretation of the transformative potential of some of the indicators analysed.) However, it is important to note that these results are driven by the generally positive results from the evaluations of projects in Asia. The difference in results between Asia and Africa is clear even if the two Asian projects with particularly positive results are excluded from the analysis. This regional difference does not appear to be fully explained by differences in the nature of the projects implemented, nor by the measurement approach or the evaluation methodology.

Female-headed households appear to be less resilient than male-headed households, even after controlling for observable differences between them, such as household size and wealth indicators. The difference in the value of the resilience index between female-headed and male-headed households is small in size, but it is consistent across contexts and across the three capacities of resilience. However, there is no clear pattern as to whether projects tend to have more positive effects for female-headed households or male-headed households.

The meta-analysis also provides some insights about the approach adopted to measuring resilience in these evaluations. Firstly, higher scores for the resilience index are found to be negatively correlated with reported losses of livestock during floods or drought, although no such correlation was found with a subjective measure of crop losses. The association with lower losses of livestock suggests that the resilience index is reflecting households' ability to prepare for and/or respond to large-scale crises. However, this is only a very partial test of the appropriateness of the index for measuring resilience: importantly, the crises suffered were in the past, while the resilience index is forward-looking and relates to a broader range of shocks, stresses and uncertainty. It will be important to seek opportunities to carry out further tests of the resilience index against people's experience of crises and how well they are able to secure their rights and improve their wellbeing over time.

Section 5.1 of this paper suggests that there may be potential for focusing on a standard set of indicators of resilience that are common across contexts. The 'generic' resilience index defined in this meta-analysis was found to be closely correlated with the full, context-specific resilience index. However, it tended to produce lower estimates of projects' impact, probably because it focused more on higher-level outcomes than on indicators related more directly to project activities. One advantage of using a generic resilience index is that it would avoid the temptation to base the selection of resilience indicators on the same assumptions as the design of the project, and then to conclude that a project had built resilience without questioning whether those assumptions about the drivers of resilience are correct.

In reviewing the results of this meta-analysis, it should be reiterated that the results relate only to the impact that Oxfam's work at a community level has had on resilience of beneficiary households, rather than to the wider, systemic impacts of Oxfam's work. Most of the reviews have included some indicators of higher-level changes that could potentially be observed at a household level – such as household members' participation in community meetings as an indicator of change in the inclusiveness of community-level decision-making structures. The next challenge for Oxfam will be to identify additional indicators of transformative capacity that can be included in future household surveys, as well as to broaden the evaluation tools being applied, in order to provide an understanding of the organisation's impact on resilience at a systemic level.

APPENDIX 1: EVALUATIONS INCLUDED IN THE META-ANALYSIS

Region	Country	Year of evaluation	Main project activities	Number of beneficiary households	Project duration
Africa	*Mali*	*2011/12*	Support to cotton farmers in adopting improved farming practices and diversifying income sources	2,936	4 years
Africa	*Niger*	*2011/12*	Support to pastoralists in marketing livestock, renovation of water sources	845	3 years
Africa	*Ethiopia*	*2012/13*	Drought recovery and preparedness, livestock vaccination, provision of water-storage facilities, rehabilitation of grazing land, setting up early-warning system	4,000*	2 years
Africa	*Kenya*	*2012/13*	Establishing pastoralist field schools and village community banks, training community animal-health workers, setting up village land-use planning committees	1,000*	3 years
Africa	*Zambia*	*2012/13*	Setting up early-warning system for flood and drought, supporting livelihood diversification	197	3 years
Africa	*Mali*	*2013/14*	Cash transfers to vulnerable women, training and support in agricultural production	488	2 years
Africa	*Niger*	*2013/14*	Provision of water sources, rehabilitation of grazing land, training and support in vegetable production	1,658*	5 years
Africa	*Chad*	*2014/15*	Drought recovery and preparedness, rehabilitation of water sources, training and support in vegetable production and seed replication	681	3 years
Asia	*Indonesia*	*2011/12*	Community-based disaster risk management	5,470*	3 years
Asia	*Pakistan*	*2011/12*	Community-based flood risk management, constructing emergency shelters and water	3,198*	3 years

Region	Country	Year of evaluation	Main project activities	Number of beneficiary households	Project duration
			harvesting facilities, training in agricultural production and livestock management		
Asia	Nepal	2012/13	Encouraging crop diversification, providing improved seeds, implementing water conservation schemes	173	3 years
Asia	Nepal	2013/14	Community-based flood-risk management, setting up early-warning system	974	2 years
Asia	Pakistan	2013/14	Establishing community-level farmer organisations, supporting crop and dairy production, facilitating access to state service providers and social safety nets	1,184*	2 years
Asia	Thailand	2014/15	Training and support in production of organic rice and in livelihood diversification	1,554	6 years
Latin America	Nicaragua	2014/15	Training and support in agricultural production, livestock management and natural resource management	177	3 years
Latin America	Bolivia	2015/16	Construction of 'camellones', a land-management system intended to protect agriculture and fish farming against drought or flood	135	5 years

* Estimated number of households in intervention communities.

APPENDIX 2: METHODOLOGY USED FOR PROPENSITY SCORE MATCHING

The size of project effects in the 16 evaluations included in this meta-analysis are estimated using propensity score matching (PSM).[22] The principle of PSM is to match households in the intervention group to those in the comparison group, based on their similarity in terms of pre-project observed characteristics. Following the guidance provided by Caliendo and Kopeinig (2008), variables were used for matching only if they were thought to influence selection into the project but not be affected by participation in the project. The specific characteristics used in each evaluation varied, but typically included:[23]

1. Indicators of the size and composition of the household.

2. Indicators of the gender, age and education level of the head of household (as defined by the respondent).

3. Indicators of the household's pre-project wealth level, based on respondents' recollections of wealth indicators (ownership or assets and housing characteristics) from a period before implementation of the project.[24]

4. Indicators of the livelihoods activities or sources of income that the household was engaging in before the project, if these were thought to be relevant to the participation decision and if recalled data were thought to be reliable.

5. Indicators of a household's access to infrastructure or markets, based on estimated travel time to the nearest market or nearest major road.

It would clearly be difficult to find exact matches for each treated household based on all of those different characteristics. Instead, these characteristics were used to estimate the *propensity score* for each household: that is, the probability that a household is in the intervention group, conditional on all the matching variables. Rosenbaum and Rubin (1983) demonstrated that, if the intervention and comparison groups are balanced in terms of their propensity scores, then they are also balanced in terms of each of the matching variables.

Within each evaluation dataset, propensity scores were estimated by regressing actual intervention status on the selected matching variables, and then using the resulting model to predict the probability of intervention.[25] Analysis was then restricted to the area of common support, the region in which the propensity score distributions of the treated and comparison groups overlapped. Observations outside the area of common support were dropped from the analysis. In some cases, this results in a small number of treated observations being dropped, meaning that the intervention group was not a fully representative sample of the project participants or beneficiaries. In only one of the evaluations were more than 10 percent of the treated observations dropped for this reason.

Within the area of common support, a kernel matching procedure was used to match each treated observation with a weighted average of the comparison observations. Greater weight was given to comparison observations with propensity scores closer to the propensity score of the treated observation. The rate at which weights given to the comparison observations declined with distance from the treated observation was adjusted to minimize the observed bias of the overall model in each case. The mean standardized bias in the matched model (suggested as a criterion for matching quality by Rosenbaum and Rubin (1985) and Caliendo and Kopeinig (2008)) was less than five percent in all but two cases, and was only slightly higher than five percent (specifically, 5.1 percent and 5.2 percent) in those cases. The median standardized bias in the matched model was less than five percent in all cases. The pseudo-R^2 of the matched model was less than 0.015 in all cases, indicating that there is little difference in the distribution of the matching variables between the intervention and comparison groups.

The estimates of the effect of each project on the resilience index are provided by the average treatment effect on the treated in the corresponding matching model, calculated as the difference in average outcomes between the matched treated and comparison groups. Standard errors of the estimates were bootstrapped using 1,000 repetitions to account for the variation caused by the estimation of the propensity scores and the determination of the common support. In datasets where sampling had been carried out at the cluster level, standard errors were clustered during bootstrapping for cases in which this was found to increase the size of the estimated standard errors.

APPENDIX 3: NECESSARY AND SUFFICIENT INDICATORS OF IMPACT ON RESILIENCE INDEX

The table on the following page lists all the indicators of resilience that were included in at least six of the 16 evaluations. The indicators are not defined or measured in exactly the same way in each of the evaluations, but they are attempting to measure the same characteristic in a way that is appropriate to the local context.

We attempt to identify whether any of the indicators are driving results for the dimensions or capacities by examining whether there is a connection between finding a statistically significant effect in terms of each of the individual indicators and finding a statistically significant effect on the overall index of resilience. This comparison is facilitated by columns A to D of the table:

- Column A shows the number of evaluations in which the interventions were found to have a positive effect (statistically significant with $p < 0.1$) on both the overall index of resilience and the specific indicator.

- Column B shows the number of evaluations in which a positive effect was found on the overall index of resilience, but not on the specific indicator.

- Column C shows the number of evaluations in which a positive effect was found on the specific indicator, but not on the overall index of resilience.

- Column D shows the number of evaluations in which no positive effect was found, either on the specific indicator or on the overall index of resilience.

We can investigate whether any of the indicators are predictors of change in the overall index by comparing the figures in columns A, B and C of the table.

If having an effect on a specific indicator were a *necessary* condition for having an effect on the index, there would be no evaluations in which a positive effect was found on the resilience index but not the specific indicator.[26] That is, the figure in column B would be zero. Even if the figure in column B was not exactly zero, the fact of that figure being small relative to the number of evaluations that found a positive effect on both the index of resilience and the indicator (column A) would imply that an intervention is unlikely to have an effect on resilience overall without having an effect on that indicator. Comparing the figures in columns A and B of the table identifies several such indicators that seem to be necessary conditions for an effect on the overall resilience index: these are identified in the right-hand column of the table.

For a particular indicator to be a *sufficient* condition for change in the overall resilience index, all the evaluations in which a positive effect is found on the overall resilience index should also find a positive effect on the particular indicator.[27] That is, the figure in column C of the table should be zero, and the figure in column A should be greater than zero. This applies to several indicators, identified in the second-to-last column of the table.

Necessary and sufficient indicators of impact on the resilience index

Indicator	Number of evaluations in which indicator is included	Comparison of result between indicator and index of resilience				Evidence of being a sufficient indicator of impact on index	Evidence of being a necessary indicator of impact on index
		Significant positive effect on indicator and on index (A)	Significant positive effect on index, not on indicator (B)	Significant positive effect on indicator, not on index (C)	No significant positive effect on indicator or index (D)		
Diversification in sources of income	16	0	9	0	7		
Savings	15	2	7	1	5		
Access to credit	14	1	8	0	6		
Social support networks	13	5	4	0	4		
Ownership of or access to productive assets	13	0	7	0	6		
Attitude to change	13	2	6	0	5	Yes	
Diversification in crops produced	12	3	3	1	5		
Awareness or understanding of climate change	12	3	5	0	4	Yes	
Access to irrigation or water for agriculture	10	1	6	0	3		
Income from regular employment, social transfers or remittances	10	1	5	0	4		
Access to extension or other state services	10	4	2	0	4	Yes	
Participation in community groups	10	2	3	0	5	Yes	
Crisis support from local or national government	8	1	3	0	4		
Adoption of innovative practices (other than those promoted by the project)	8	4	1	0	3	Yes	Yes

Access to safe drinking water	7	1	4	1	1		
Participation in community-level or district-level decision-making	7	4	0	1	2	Yes	Yes
Adoption of improved production practices that were promoted by the project being evaluated	6	2	1	2	1		Yes
Adoption of disaster preparedness practices	6	2	1	0	3	Yes	Yes
Access to early-warning information	6	5	0	0	1	Yes	Yes
Awareness of local community disaster preparedness or disaster management plan	6	5	0	0	1	Yes	Yes

All indicators that appear in six or more of the 16 evaluations are shown in this table. 'Significant positive effects' are those with a positive difference between the matched mean and comparison groups that are statistically significant at at least the 10 percent level.

REFERENCES

Alkire, S. and J. Foster. (2011). *Counting and Multidimensional Poverty Measurement. Journal of Public Economics*, 95(7–8), 476–87. https://doi.org/10.1016/j.jpubeco.2010.11.006

Béné, C., R. Godfrey Wood, A. Newsham and M. Davies. (2012). *Resilience: New Utopia or New Tyranny? Reflection about the Potentials and Limits of the Concept of Resilience in Relation to Vulnerability Reduction Programmes*. Institute of Development Studies Working Paper 405. Retrieved 6 February 2017, from http://www.ids.ac.uk/publication/resilience-new-utopia-or-new-tyranny. https://doi.org/10.1111/j.2040-0209.2012.00405.x

Borenstein, M., L.V. Hedges, J.P.T. Higgins and H. Rothstein. (2009). *Introduction to Meta-Analysis*. Chichester, UK: Wiley. https://doi.org/10.1002/9780470743386

Caliendo, M. and S. Kopeinig. (2008). *Some Practical Guidance for the Implementation of Propensity Score Matching. Journal of Economic Surveys*, 22(1), 31–72. https://doi.org/10.1111/j.1467-6419.2007.00527.x

Cohen, J. (1992). *A Power Primer. Psychological Bulletin*, 112(1), 155–9. https://doi.org/10.1037/0033-2909.112.1.155

Duvendack, M., J.G. Hombrados, R. Palmer-Jones and H. Waddington. (2012). *Assessing 'what Works' in International Development: Meta-Analysis for Sophisticated Dummies. Journal of Development Effectiveness*, 4(3), 456–71. https://doi.org/10.1080/19439342.2012.710642

Filmer, D. and L.H. Pritchett. (2001). *Estimating Wealth Effects without Expenditure Data – or Tears: An Application to Educational Enrollments in States of India. Demography*, 38(1), 115–32.

Fuller, R. and J. Lain. (2015). *Measuring Resilience: Lessons Learned from Measuring Resilience in Oxfam's Large-N Effectiveness Reviews*. Oxfam Discussion Paper. Retrieved 6 February 2017, from http://policy-practice.oxfam.org.uk/publications/measuring-resilience-lessons-learned-from-measuring-resilience-in-oxfams-large-583601

Fuller, R. and J. Lain. (Forthcoming 2017). *Are Female-Headed Households Less Resilient? Evidence from a Meta-Analysis of Oxfam's Impact Evaluations*.

Higgins, J.P.T. and S. Green (eds.). (2011). *Cochrane Handbook for Systematic Reviews of Interventions*. Version 5.1.0. Retrieved 6 February 2017, from http://www.handbook.cochrane.org

Higgins, J.P.T., S.G. Thompson, J.J. Deeks and D.G. Altman. (2003). *Measuring Inconsistency in Meta-Analyses. BMJ*, 327(7414), 557–60. https://doi.org/10.1136/bmj.327.7414.557

Hughes, K. and H. Bushell. (2013). *A Multidimensional Approach to Measuring Resilience*. Oxfam GB Working Paper. Retrieved 6 February 2017, from http://policy-practice.oxfam.org.uk/publications/a-multidimensional-approach-to-measuring-resilience-302641

Hughes, K. and C. Hutchings. (2011). *Can We Obtain the Required Rigour without Randomisation? Oxfam GB's Non-Experimental Global Performance Framework*. International Initiative for Impact Evaluation Working Paper 13. Retrieved 6 February 2017, from http://www.3ieimpact.org/en/publications/working-papers/working-paper-13/

Jeans, H., S. Thomas and G. Castillo. (2016). *The Future Is a Choice: The Oxfam Framework and Guidance for Resilient Development*. Retrieved 6 February 2017, from http://policy-practice.oxfam.org.uk/publications/the-future-is-a-choice-the-oxfam-framework-and-guidance-for-resilient-developme-604990

Levin, S.A., S. Barrett, S. Aniyar, W. Baumol, C. Bliss, B. Bolin, P. Dasgupta, P. Ehrlich, C. Folke, I.-M. Gren, C.S. Holling, A. Jansson, B.-O. Jansson, K.-G. Mäler, D. Martin, C. Perrings and E. Sheshinski. (1998). *Resilience in Natural and Socioeconomic Systems. Environment and Development Economics*, 3(02), 221–62. https://doi.org/10.1017/S1355770X98240125

Martin-Breen, P. and J.M. Anderies. (2011). *Resilience: A Literature Review*. Bellagio Initiative Background Paper. Retrieved 6 February 2017, from http://opendocs.ids.ac.uk/opendocs/handle/123456789/3692

Pain, A. and S. Levine. (2012). *A Conceptual Analysis of Livelihoods and Resilience: Addressing the 'Insecurity of Agency'*. Humanitarian Policy Group Working Paper. Retrieved 6 February 2017, from https://www.odi.org/publications/6976-resilience-livelihoods-agency-socio-economic-security

Rihoux, B. and C.C. Ragin (eds.). (2009). *Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques*. Thousand Oaks, California: SAGE Publications.

Rosenbaum, P.R. and D.B. Rubin. (1983). *The Central Role of the Propensity Score in Observational Studies for Causal Effects. Biometrika*, 70(1), 41–55. https://doi.org/10.1093/biomet/70.1.41

Rosenbaum, P.R. and D.B. Rubin. (1985). *Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score. The American Statistician*, 39(1), 33–8. https://doi.org/10.1080/00031305.1985.10479383

Waddington, H., H. White, B. Snilstveit, J.G. Hombrados, M. Vojtkova, P. Davies, A. Bhavsar, J. Eyers, T.P. Koehlmoos, M. Petticrew, J.C. Valentine and P. Tugwell. (2012). *How to Do a Good Systematic Review of Effects in International Development: A Tool Kit. Journal of Development Effectiveness*, 4(3), 359–87. https://doi.org/10.1080/19439342.2012.711765

NOTES

1 The Cochrane Collaboration and 3ie catalogue their reviews at http://www.cochrane.org/evidence and http://3ieimpact.org/en/evidence/systematic-reviews respectively.

2 This initiative is described further in Hughes and Hutchings (2011). In addition to the 16 resilience Effectiveness Reviews, other Effectiveness Reviews have sought to evaluate projects' impacts on household livelihoods, women's empowerment, citizen's voice, policy influencing, or humanitarian response. Reports describing the full results for each of the Effectiveness Reviews can be found at http://www.oxfam.org.uk/effectiveness.

3 It should be noted that the results of the 16 Effectiveness Reviews are not fully representative of Oxfam's resilience-building work. In a small number of cases, projects that were randomly selected for an Effectiveness Review were not evaluated because they were considered not to be mature enough to show significant impact, or because a good-quality impact evaluation had been conducted in the recent past. Among the 16 projects that were evaluated, in many cases it was necessary to select specific project components or specific geographic areas to be included in the Effectiveness Review. These decisions were generally made on the grounds of evaluability (for example, based on locations where suitable comparison sites could be identified) and on the potential for maximising learning (for example, through focusing the evaluation on project components that were being considered for scale-up), rather than with the aim of maximising representativeness.

4 In this paper, we refer to 'intervention' and 'comparison' groups rather than 'treatment' and 'control' groups. We believe that the term 'comparison group' is more appropriate than 'control' in the context of quasi-experimental evaluations.

5 Pain and Levine (2012) have a useful discussion of this distinction.

6 Further details of the process of identifying indicators and the challenges involved are discussed in Fuller and Lain (2015).

7 Indeed, in most of the evaluations some characteristics were identified that were recognized as important contributors to resilience, but which could not be measured with sufficient robustness in a household survey; those aspects of resilience were not assessed in the evaluation. This applied frequently with respect to ecosystem health, for example.

8 Some of the original evaluations used alternative aggregate measures of resilience, and in some cases different weights were assigned to different indicators, based on evaluators' judgements. For this meta-analysis, the original analysis has been modified to use a consistent aggregation process, as described here, and to give equal weight to each indicator. In some cases, the selection of indicators and the way the indicators are constructed were also modified from the original analysis, to provide a more consistent approach to measuring resilience. In addition, some of the PSM models used to estimate the project effects were improved for this meta-analysis. For these reasons, the effect sizes used in this meta-analysis do not always match those reported in the original evaluation reports. In all but two cases these amendments produced only minor changes in the estimated effect size or in its statistical significance.

9 In fact the measures of resilience used in the 16 evaluations were developed using an older framework, based on an understanding of resilience in terms of five dimensions (described in Hughes and Bushell, 2013). The categorisation of indicators between the three capacities in the new framework was made specifically for this meta-analysis.

10 For analyses involving the estimation of project effects, standardisation is carried out by dividing by the standard deviation from the matched sampled, derived from the corresponding PSM model. For cases in which the relationship between two variables is tested (that is, for the analysis in Sections 3.4 and 4.2),

standardisation is carried out by subtracting the sample mean and dividing by the sample standard deviation across the entire sample.

11 The I^2 statistic is defined as $(Q − df)/Q$, where Q represents Cochran's heterogeneity statistic, the sum of the squared deviations of each study's estimate from the meta-analysis estimate, and df is the degrees of freedom. The interpretation of this statistic is discussed in Higgins et al. (2003) and Higgins and Green (2011).

12 An alternative approach would be to weight each evaluation by the size of the sample frame in the intervention group for that evaluation, so that the overall effect represents the average effect across all those targeted as beneficiaries by the 16 projects. Unfortunately, there does not seem to be any guidance in the literature on how to combine a random-effects model with externally defined weights.

13 A meta-regression for the size of the effect on the scale of the project produces a coefficient close to zero that is not statistically significant.

14 A meta-regression for the size of the project effect on the duration of the project produces an estimated coefficient that is negative but not statistically significant.

15 A meta-regression of project effect on a dichotomous variable representing whether the project had a primary focus (as coded at the time of the meta-analysis) on disaster-risk reduction or livelihoods resilience produced an estimated coefficient on that variable that is statistically significant with $p < 0.05$.

16 The differences between Asia and the other regions are confirmed through use of meta-regression models for the effect size, with a binary variable indicating whether the evaluation was carried out in Asia as the sole covariate. The coefficient on the binary variable in the meta-regression for the overall resilience index is estimated at 0.68 standard deviations, and is strongly statistically significant ($p < 0.01$). The corresponding coefficients for the indices of absorptive and adaptive capacity are both positive and statistically significant with $p < 0.01$, while that for transformative capacity is positive and statistically significant with $p < 0.1$.

17 No guidance was provided to respondents in identifying the head of household, other than that the head should be a current household member. (In particular, the head could not be someone, such as a migrant worker, who spends most of their time living outside the household.) It is possible that, if the project interventions had had a significant impact on intra-household relations (for example, through empowering women), then this may have affected which individual was identified as the head of household at the time of the survey. If so, this could result in bias between the intervention and comparison groups in the gender of the head of household. It is thought that any such bias would be small, though it is not possible to check this with the data available.

18 This meta-analysis is based on regression models of the form

$$Y_i = \beta_0 + \beta_1 F_i + X_i\beta + \varepsilon_i$$

for the ith household, where i is a household identifier, Y_i is the resilience index, F_i is a dichotomous variable that takes the value of 1 for female-headed households and 0 for male-headed households, and X_i is a vector of observable household characteristics (wealth indicators, number of adults living in the household, education level of the household head, and the age and squared age of the household head). The coefficient β_1 is estimated separately for each evaluation, and then is used as input into a meta-analysis model, using random effects with inverse-variance weighting.

19 This meta-analysis is based on regression models of the form

$$Y_i = \beta_0 + \beta_1 F_i + \beta_2 T_i + \beta_3 F \times_i T_i + X_i\beta + \varepsilon_i$$

for the ith household, where the notation is the same as that described in the previous footnote, and, in addition, T_i is a dichotomous variable that takes the value of 1 for households included in the intervention group for each evaluation, and 0 for households in the comparison group, and X_i is a vector of observable household characteristics. In this case the coefficient of interest is β_3: this coefficient is estimated separately for each evaluation, and then is used as input into a meta-analysis model, using random effects with inverse variance weighting.

20 This figure was derived by calculating the correlation within each dataset, then converting the correlation coefficients to Fisher's z scale, as suggested by Borenstein et al. (2009), and then conducting a random-effects meta-analysis among them.

21 The variables measuring losses were standardized before analysis (as were the resilience indices), since the losses suffered were of different scales in different situations. Households that were not farming or that did not own livestock in the pre-project period are excluded from these calculations. In most cases, only a minority of respondents (21 percent or less) were excluded for this reason. The exceptions are for the evaluations in Zambia and Bolivia, where only 35 percent and 22 percent respectively owned any livestock before the crisis. The small number of observations over which losses are calculated is the reason for the large confidence intervals in the estimates for Zambia and Bolivia in Figure 9.

22 To ensure they used a consistent approach, some of the matching models were modified from those used to derive the estimates reported in the original evaluation reports.

23 An exception is the evaluation carried out in Thailand, where matching variables were based largely on pre-project engagement in livelihoods activities, with less emphasis on demographic characteristics and pre-project wealth indicators.

24 Wealth indices were generated through principal component analysis, following the approach of Filmer and Pritchett (2001). Households surveyed were then divided into quintiles according to the value of the pre-project wealth index, thus providing a set of four binary variables on which to match (after omitting one of the quintiles as the base category).

25 In several of the evaluations, PSM was carried out separately in two geographical regions or among two subgroups. In these cases, the results of the two regions or subgroups were aggregated using a fixed effects meta-analysis model, weighted by the sampling weights.

26 Formally, an indicator is a necessary condition for an effect on resilience overall if the set of evaluations that find a positive effect on the overall resilience index is a subset of the set of evaluations that find a positive effect on the particular indicator. These definitions for necessary and sufficient conditions correspond to those used in Qualitative Comparative Analysis (Rihoux and Ragin, 2009). Carrying out formal QCA is not possible with our data, primarily because of the low number of 'cases' (evaluation datasets) and the high number of 'conditions' (indicators).

27 That is, an indicator is a sufficient condition for an effect on resilience overall if the set of evaluations that find a positive effect on the particular indicator is a subset of the set of evaluations that find a positive effect on the overall resilience index.

Oxfam Effectiveness Reviews

For more information, or to comment on this report, email opalenquiries@oxfam.org.uk.

This research report was written to share research results, to contribute to public debate and to invite feedback on development and humanitarian policy and practice. It does not necessarily reflect the policy positions of the publishing organizations. The views expressed are those of the author and not necessarily those of the publishers.

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

www.oxfam.org.uk/effectiveness