Measuring changes in resilience as a result of the SUR1M project in Niger

Christophe Béné, Alex Riba and Dave Wilson

BRACED

BRACED Knowledge Manager: Impact Evaluations

Table of contents

List of figures

List of tables

About the authors

Christophe Béné is Senior Policy Advisor at the International Center for Tropical Agriculture (CIAT) based in Cali, Colombia. He has 15+ years of experience in inter-disciplinary research, advisory work, and project evaluation focusing on poverty reduction, food security and more recently resilience interventions.

Alex Riba is a statistician with Statistics for Sustainable Development (Stas4SD). He has more than 20 years of experience teaching statistics and conducting research at Universitat Politecnica de Catalunya, Barcelona Tech. Alex's areas of expertise are experimental design, statistical process control, six-sigma statistical techniques and analysis of surveys.

Dave Wilson is a principal consultant and leads the climate change theme at Itad and coordinates the BRACED Knowledge Manager's evaluation activities. Dave has over 10 years' experience of managing and evaluating climate change and natural resource programmes in the UK and overseas. @davewilson_itad

Acknowledgements

This report was written by Chris Béné (International Center for Tropical Agriculture) and Alexander Riba (Stats4SD), with additional comments from David Wilson (Itad) and Derek Poate (Itad). The authors are grateful to Manzo Rio-Rio Aminou and the whole M&E team of the CRS Niamey office as well as the enumerators who conducted the field survey and data collection in Niger. The continuous support of Carlos Barahona (Stats4SD) is also gratefully acknowledged. The work is part of Evaluation Activity (EA3) implemented by Itad and funded by DFID under the BRACED programme.

Photo credit: Front cover. Chris Béné

Acronyms

ATT	Average Treatment Effect for the Treated
ANOVA	Analysis of Variance
BRACED	Building Resilience and Adaptation to Climate Extremes and Disasters programme
CCA	Climate Change Adaptation
CCAFS	Climate Change, Agriculture and Food Security programme
CI	Confidence Interval
CMDRR	Community-Managed Disaster Risk Reduction
CRS	Catholic Relief Services
CSAP	Climate-Smart Agricultural Practice
CSO	Civil Society Organisation
DAC	Development Assistance Committee
DFID	UK Department for International Development
DRR	Disaster Risk Reduction
EA	Evaluation Activity
EWG	Early Warning Group
EWS	Early Warning System
GoN	Government of Niger
HFIAS	Household Food Insecurity Access Scale
HDDS	Household Dietary Diversity Score
ICF	International Climate Fund
IP	Implementing Partner
KM	Knowledge Manager
KPI	Key Performance Indicator
LZ	Livelihood Zone
M&E	Monitoring and Evaluation
NGO	Non-Governmental Organisation
NRM	Natural Resource Management
OECD	Organisation for Economic Co-operation and Development
PRC	Participatory Radio Campaign
PSM	Propensity Score Matching
RCT	Randomised Control Trial
SILC	Savings and Internal Lending Community
SUR1M	Scaling-Up Resilience to Climate Extremes for over 1 Million People in the Niger River Basin
T/C	Treatment/Control
UK	United Kingdom
UNDP	United Nations Development Programme

Executive summary

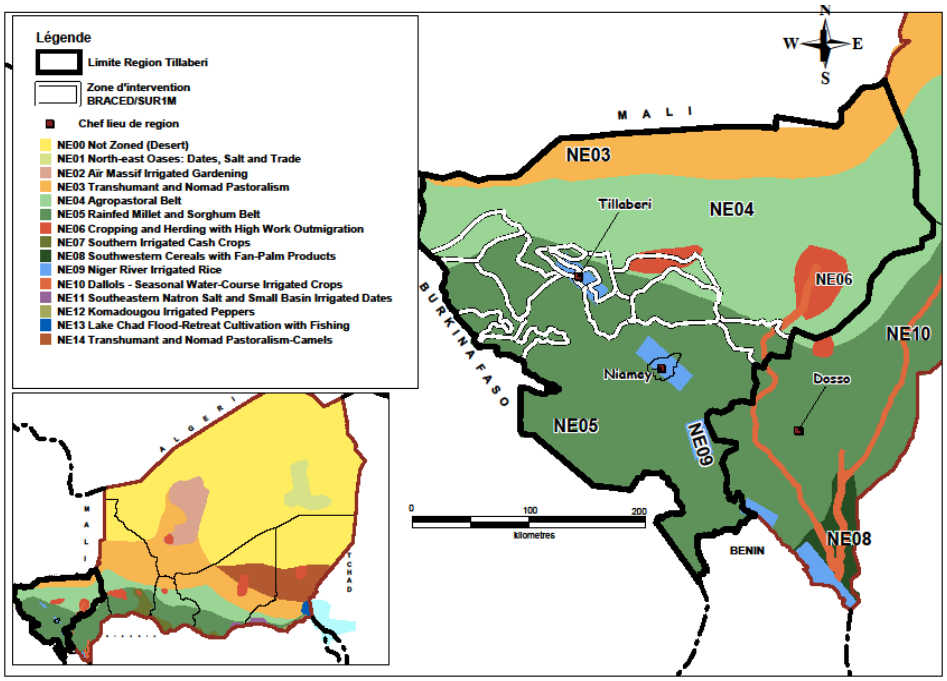
Background

The three-year, £110 million UK Department for International Development (DFID)-funded Building Resilience and Adaptation to Climate Extremes and Disasters (BRACED) programme aims to build the resilience of up to 5 million vulnerable people against climate extremes and disasters. It was launched in January 2015 and supports over 120 organisations in 15 consortia across 13 countries in East Africa, the Sahel and Asia. Understanding the extent to which these projects are able to strengthen the resilience of the households, communities and organisations they work with is critical in ensuring that successful approaches are scaled and replicated. The overall purpose of this evaluation is to determine **to what extent household-level resilience has increased as a result of BRACED interventions**.

This report presents the results of the BRACED Impact Evaluation of the SUR1M project in Niger. The SUR1M (Scaling-Up Resilience to Climate Extremes for over 1 Million People in the Niger River Basin) project was implemented by Catholic Relief Services (CRS) from 2015 until the end of 2017 with a series of local and international partners. It aims to strengthen the resilience of the local population to climate extremes. It is one of the 15 resilience projects funded by the UK Department for International Development (DFID) under the three-year Building Resilience and Adaptation to Climate Extremes and Disasters (BRACED) programme in Africa and Asia.

To achieve its objective, the SUR1M project had been implementing a series of activities around community-level self-help and savings; re-greening, natural resource management and climate-smart agricultural practices; development of local seed multiplication and input distribution systems; community-managed disaster risk reduction (DRR) and early warning groups (EWGs); participatory radio campaigns; public–private linkages; and strengthening institutional capacity of local governance bodies.

Figure A. Map of the areas in Niger where the SUR1M project is operating



Evaluation design

The evaluation presented in this report was commissioned by DFID and implemented by the BRACED Knowledge Manager. It draws on recent conceptual advances made in understandings of resilience in the context of food security, where resilience is understood as 'the ability of individuals, households, communities, institutions or higher-level systems to adequately deal with shocks and stressors' (the terms 'adequately' referring to the ability to avoid short- and longer-term negative impacts).

In the absence of reliable resilience data collected at the beginning of the project (baseline) by the SUR1M project, an ex-post (endline) survey was conducted where the responses to shocks/stressors (intermediary outcome), ability to recover from shock/stressors (final outcome) and well-being (long-term impact) of households that benefited from the SUR1M project (treatment group) were compared with those of non-beneficiary households (control group) with similar demographics and socioeconomic background but living out of the areas where the project had been operating.

The impact evaluation was constructed around three hypotheses. **Hypothesis 1:** At the shocks/stressors level, SUR1M beneficiaries were expected to show lower propensity to adopt detrimental (coping) responses and higher propensity to adopt positive (adaptive/transformative) responses than non-beneficiaries; **Hypothesis 2**: At the recovery level, SUR1M beneficiaries were expected to show faster recovery rates than households in the control group (everything else being equal). **Hypothesis 3**: At the long-term impact level, SUR1M beneficiaries were expected to show higher levels of well-being than households in the control group (everything else being equal).

The evaluation exercise was implemented in 11 communes in the western part of Niger where CRS had been implementing the SUR1M project. Villages were identified as 'high intensity' or 'low intensity' based on the number of activities implemented in these villages. High-intensity villages were defined as villages where at least 4 activities (and up to 13) had been implemented by the SUR1M project at the time of the evaluation. Low-intensity villages on the other hand were villages where only the SUR1M radio broadcast had been implemented. In total, 1,498 households (812 control households in 27 low-intensity villages and 686 treatment households in 22 high-intensity villages) were surveyed in accordance with the original evaluation framework.

Analysis

A series of preliminary descriptive analyses were performed to compare the beneficiary and non-beneficiary groups. Some general household characteristics were found to be similar between the two groups, including age of head of household, size of household and even income and assets, but others differed. In particular (despite our efforts to ensure the two groups to be comparable), households in the treatment group (SUR1M beneficiary) reported being exposed to a higher number of shocks/stressors than those in the control group. The nature of these shocks/stressors also differed slightly. Overall, however, control and treatment households both reported being affected by the same combination of major shocks/stressors, including seasonal drought; serious illness; armyworm attacks; food and general commodities price peaks; and seasonal food shortages.

Further analysis shows that, although treatment households reported being more exposed to shocks/stressors, they appeared to be less likely to engage in some detrimental responses (including reducing food consumption and changing the type of food consumed) than the control group and more likely to engage in adaptive/transformative strategies (both on- and off-farm activities).

Key findings

Hypothesis 1: This first key result, which directly addresses our first hypothesis, was obtained without accounting for household characteristics. When controlling for such characteristics through the use of specific models, analysis still shows that the SUR1M project increases the probability of households

engaging in adaptive/transformative strategies for 7 out of the 10 shock–response combinations that were tested. This increase is statistically significant in 6 out of these 7 combinations. On the other hand, results are less clear-cut for the detrimental coping strategies. For these, only 6 out of the 24 shock–response combinations that were tested show a decrease in the likelihood to engage in detrimental strategies for the treatment households. Among these 6 cases, 5 were significant. For the rest of the 18 combinations of shock responses tested, the SUR1M project does not seem to have a negative or a positive impact on households' probability of adopting coping strategies..

Hypothesis 2: The second hypothesis (at the recovery level) was tested for the five most important shock/stressors, using a self-reported indicator of household capacity to recover from past events. A specific statistical technique (Propensity Score Matching) was used to compare these recovery indicators between treatment households and a pool of comparable control households. The results indicate that **treatment households reported a statistically significantly higher capacity to cope with a range of the most common shocks when compared with the control group.** Table A (Table 13 in the main report) shows the details of the difference in self-reported capacity to recover between the treatment and the control households calculated for the matched households (the ATT

Table A. Results of the PSM test for the household resilience index computed for the five most reported shocks

	Mean treated	Mean control	ATT	Std.Err.	*t* [(1)]	
Seasonal drought	9.50	8.17	1.326	1.156	1.15	
Serious illness	11.65	9.79	1.857	1.276	1.46	*
Armyworm attack	15.60	9.76	5.840	2.905	2.01	**
Price peaks	Not enough matching pairs					
Soudure	8.27	6.94	1.331	0.819	1.62	*

Note: [(1)] One tailed *t*-critical value at ***1% = 2.326; **5% *t*-critical value = 1.645; *10% *t*-critical value = 1.282. Value of t-statistics is based on bootstrapped standard errors (100 iterations).

Hypothesis 3: The third hypothesis (at the long-term impact level) focused on household well-being. At this level, we hypothesised that, overall, if the SUR1M project were successful at building the resilience of the beneficiaries, then these beneficiaries should display a higher level of food security and nutritional security (used as indicators of well-being). The comparison of the two indicators shows no specific difference between the beneficiaries of the SUR1M beneficiary and non-beneficiary households, suggesting that at this stage the SUR1M project does not have a strong long-term impact on the well-being of its beneficiaries.

Overall conclusion

Overall the SUR1M project contributes positively to strengthening households' capacity to better handle shocks and stressors (household resilience) by strengthening their ability to engage in adaptive/transformative strategies when faced with shocks and stressors. However, the SUR1M project seems to have only limited effect on these same households' adoption of detrimental coping strategies. Likewise, no statistical difference was observed –at least at the present time- at the long-term impact level on food (in)security and dietary diversity. One possible explanation for this last result is that the two-year period during which the SUR1M operated in Niger was too short for its outputs to transform into positive outcomes and lead to significant impact on the long-term well-being of the beneficiaries despite the positive results observed at the resilience index level.

Additional lessons and recommendations derived from this evaluation will be combined with lessons and recommendations generated by the two other EA3 reports (Myanmar and Ethiopia) and presented together in a summary report for policy makers (forthcoming)

1. Introduction

1.1. Building Resilience and Adaptation to Climate Extremes and Disasters: An overview

The three-year, £110 million UK Department for International Development (DFID)-funded Building Resilience and Adaptation to Climate Extremes and Disasters (BRACED) programme aims to build the resilience of up to 5 million vulnerable people against climate extremes and disasters. It was launched in January 2015 and supports over 120 organisations in 15 consortia across 13 countries in East Africa, the Sahel and Asia.

Understanding the extent to which these projects are able to strengthen the resilience of the households, communities and organisations they work with is critical in ensuring that successful approaches are scaled and replicated. With risks from climate-related disasters increasing and impacts set to plunge an additional 100 million people into poverty by 2030 (Hallegatte et al., 2016), it is an urgent imperative to understand what makes people, households, communities, markets, organisations and countries better able to anticipate, absorb and adapt to climate extremes.

1.2. Quantifying and attributing changes in household resilience under BRACED

All BRACED projects seek to increase the resilience of people vulnerable to climate extremes. This outcome is reported on for International Climate Fund (ICF) Key Performance Indicator 4 (KPI4) and is typically captured by a number of variables relating to types of assets and capacities and the ability of people to utilise these in the face of climate extremes (see Box 1).

Box 1. International Climate Fund Key Performance Indicator 4

Each project is mandated by the funder (DFID) to report against ICF KPIs, the most relevant of which for this evaluation is *ICF KPI4: Number of people with improved resilience as a result of ICF support*. Each implementing partner (IP) reports a number against this indicator. The methods by which they arrive at this number varies, with some IPs using baseline, midline and endline household surveys, some using smaller panel surveys (which trace a subset of individuals surveyed at the baseline stage) and others using more qualitative measures. This variation and aggregation of all findings to a single reportable number means that KPI4 alone does not offer sufficient insight into what works to strengthen resilience or even whether changes have benefited different groups in different ways. We used KPI4 outcome measures as an entry point with the IPs, each of which had a list of indicators or scorecard in place to measure resilience changes for this purpose. Indeed, we used 'KPI4' or 'KPI4 index' interchangeably with the specific resilience measurement indices used by each IP. Under this evaluation, these indices have been refined based on additional analysis working in partnership with each of the IPs. In this way, we aim to go beyond 'headcounts' to offer more nuanced and rich analysis of data generated through the course of this evaluation.

However, the nature of resilience-strengthening activities and the fact that they are not operating in a vacuum with outcomes potentially influenced by other (confounding) factors makes it difficult to attribute quantitative changes in the resilience outcome solely to a particular BRACED project or intervention. Simply comparing baseline and end-of-project data does not solve this problem as the observed change may owe to project and external effects. What is missing is a counterfactual – the resilience indicator for project beneficiaries in the absence of the project – that would then enable the attribution of changes only to BRACED interventions.

In an attempt to address the question of attribution, the BRACED Knowledge Manager (KM) has been working with two of the 15 BRACED projects: **SUR1M, Niger (the subject of this technical report), and**

Myanmar Alliance, Myanmar. Together, we have designed and implemented impact evaluations to determine the extent to which household resilience has changed as a result of the project interventions. Known as Evaluation Activity 3 (EA3), this is led by the KM working in close collaboration with the IPs. Each IP has an evaluation designed to be relevant for the context in which it is operating, its implementation plans and its existing monitoring and evaluation (M&E) framework while offering some coherence across all three.[1] All involve a large sample household survey and quantitative calculations of changes in resilience as a result of project activities via different methods (BRACED KM Evaluation Plan, 2015).

1.3. Purpose and structure of this report

This report presents the results of the BRACED Impact Evaluation in Niger also known as Evaluation Activity 3 (EA3). It is designed to be a standalone document but will also inform a summary document that will bring together results from other country studies and translate these into lessons and recommendations (forthcoming).

The report is organised into the following sections. After this introduction, Section 2, Background and Context, describes the objective, activities and intervention zones of the SUR1M project in Niger and Mali. Section 3, on the Evaluation Framework, details the general evaluation approach adopted for this work, including the choice of a quasi-experimental approach, the conceptual framework and the evaluation hypotheses. Section 4, Methodology, summarises the design of the survey, including the approach to sampling and the selection of the treatment vs. control villages. Section 5, Results, presents the data from the household survey and headline findings and insights by the team on that basis; these key findings are revisited in the discussion, in light of the theory of change of the project. Section 6, Conclusion, wraps up the work. The separate Technical Annex presents the more detailed material.

1.4. Note to the reader

This technical report is aimed at those interested in resilience measurement from government, non-governmental organisations (NGOs) and academia, and M&E practitioners as well as funders and commissioners of evaluations. It is also intended for other IPs within the BRACED programme that may be considering similar approaches for future resilience-strengthening projects under or outside of BRACED. All attempts have been made to explain technical concepts and results as clearly as possible but the inclusion of technical, statistical or evaluation terms is unavoidable and therefore some understanding of research and evaluation methods will be advantageous.

1.5 Dissemination Plan

This report has a pre-defined 'learning and uptake plan' which sets out the audience and opportunities to reach and engage with them. In addition to direct submission to DFID advisers within the BRACED team and the Climate and Environment department, a workshop to socialise findings and co-create/validate recommendations has been planned. The report itself will be shared via the usual communication channels – co-hosted across the BRACED and Itad websites, strategic social media accounts and other channels recommended by the BRACED communications lead, Thomson Reuters Foundation. The authors will also seek opportunities to present at regional and global fora and

[1] Full details of the evaluation can be found in the BRACED KM Evaluation Plan (http://www.braced.org/resources) and detailed design document (available on request).

conferences. A BRACED webinar to share key results of the evaluation with as wide an audience as possible will be hosted in summer 2018.

2. Background and context of the SUR1M project

2.1. Background

In the Sahel, increasingly frequent droughts, floods and other shocks exacerbated by climate change impede household livelihood, asset-building and systems-strengthening. This undermines the ability of households and communities living in these regions to overcome the next shock, resulting in increased poverty and vulnerability. In Niger and Mali, where the SUR1M project operates, more than 60% of the population is poor or extremely poor (GoN, 2012; CRS, 2013). A large proportion of these households are agro-pastoralists, particularly vulnerable to climate extremes as a result of their reliance on rain-fed agriculture and low annual rainfall.

Recurring prolonged dry spells and severe floods in recent years have negatively affected these populations' cereal production in terms of yields, plant growth and grain quality, contributing to widespread and chronic food shortages with devastating impacts on food security and the viability of rural livelihoods (Gubbels, 2012). Extreme climate events exacerbate these structural vulnerabilities by triggering negative coping strategies, such as unsustainable tree-cutting for charcoal, selling animals and other assets to buy cereals at unfavourable prices and participation in artisanal mining, all of which exacerbate vulnerabilities and further undermine long-term resilience (Jalloh et al., 2013).

2.2. The SUR1M project

In the context described above and within the overarching framework of the BRACED programme, the SUR1M (Scaling-Up Resilience to Climate Extremes for over 1 Million People in the Niger River Basin of Niger and Mali) project implemented by Catholic Relief Services (CRS) with a series of local and international partners[2] aims at strengthening the resilience of the population of the Niger River Basin to climate extremes.

2.2.1. Objective

The main objective of the SUR1M project is to 'increase resilience to climate extremes by increasing disaster risk preparedness and climate change adaptation in the face of droughts and floods, deepening mitigation practices, and building critical assets' (CRS, 2013). To achieve this, SUR1M is intended to scale up impact by reinforcing the disaster risk management capacity of 30 communes in Niger and Mali through an integrated strategy, combining technical assistance; performance-based grants for gender-responsive disaster risk reduction (DRR) planning, natural resource management (NRM) and climate change adaptation (CCA); and mass media campaigns.

2.2.2. Activities

SUR1M is organised around a series of activities:

Savings and Internal Lending Communities (SILCs): In Niger and Mali, SUR1M improves access to financial services through the implementation of SILCs. SILCs comprise 15–30 self-selecting members (averaging 75% women) and build capacity in financial knowledge and practices through a safe savings mechanism with an internal loan fund, increasing women's social capital, voice and financial stability.

[2] These are CAFOD, Caritas, Radios Rurales Internationales, the Climate Change, Agriculture and Food Security (CCAFS) programme and the United Nations Development Programme (UNDP).

Re-greening, NRM and climate-smart agricultural practices (CSAPs): SUR1M promotes a number of locally adapted and proven re-greening CSAPs, soil and water conservation and farmer-managed natural regeneration techniques appropriate for both men and women. Many of these techniques require low financial inputs, strengthen the natural resource base, can be taken to scale and contribute to household food security and resilience. The combination of diverse techniques, including Conservation Agriculture with Trees and Bio-reclamation of Degraded Land, increases micronutrient-dense food production, income and women's land access at scale.

Development of local seed multiplication and input distribution systems: SUR1M provides training to local seed multipliers and distributors, in partnership with private agro-dealers. These agro-dealers acquire and provide improved base seed to multipliers and buy back the bulk of their certified seed production, giving farmers consistent access to improved seeds and other inputs through local suppliers.

Community-managed disaster risk reduction (CMDRR) and early warning groups (EWGs): CRS with local partners establishes commune and community-level CMDRR and EWGs in SUR1M target areas. Investing in DRR has both direct and indirect economic benefits. These can include improved governance, women's participation, basic service delivery, access to services and other benefits that ultimately lead to sustainability of DRR investments.

Participatory radio campaigns (PRCs): SUR1M has developed a coordinated information education communications strategy that stimulates demand for improved services, access to technologies and good governance while delivering key behaviour change communication messages. Radio is the primary channel of dissemination, as it has the largest reach and the ability to share messages with 70–80% of rural households, taking into account both men's and women's schedules.

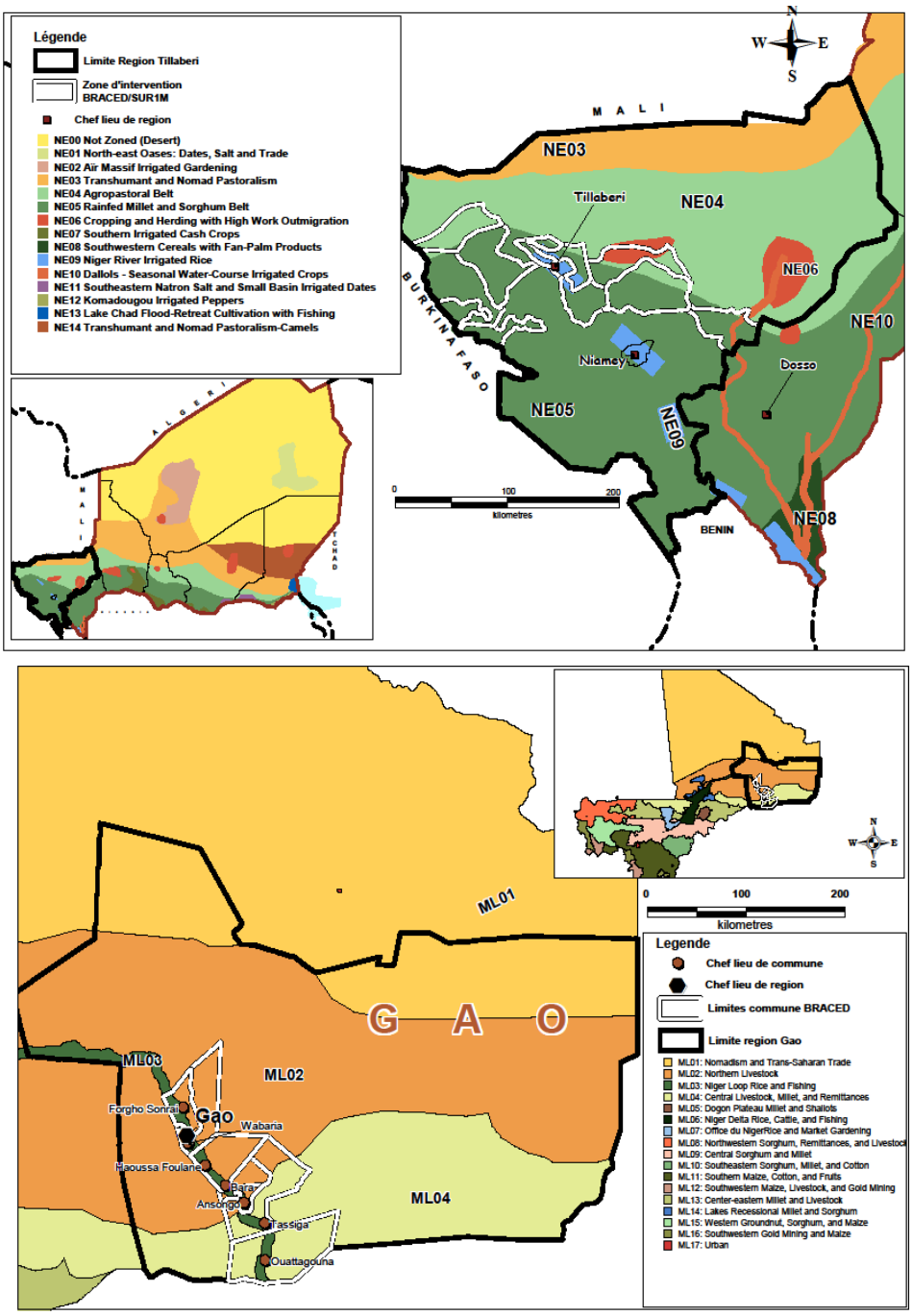
Public–private linkages: SUR1M leverage and strengthens Sahel-wide investment opportunities and small-scale public–private linkages through the participation of national chambers of commerce in steering committees and roundtables between communes, private firms, multilateral agencies and others to enhance market and value chain opportunities. Annual 'commune resilience days' provide opportunities to raise project visibility and directly link producers and buyers to private service providers, local seed producers, agro-dealers, veterinary services, mobile phone companies and entrepreneurs and to promote new products such as time- and energy-saving technologies and inputs.

Local governance: In an effort to respond to the challenge and build off the opportunities of decentralisation, SUR1M provides performance-based grants to communes to strengthen their institutional capacity and promote gender-responsive CCA and DRR in the Commune Development Plans. Additionally, SUR1M provides tools and train civil society organisations (CSOs) in advocacy and intermediation so they can influence local governance.

2.2.3. Intervention zone

The SUR1M project aims to reduce the risk of exposure to disasters (droughts and floods) of about 1 million people in 19 communes, including 12 in Niger and 7 in Mali, all in the Niger River Basin. In Niger the communes are distributed in the departments of Ouallam, Tillaberi, Tera, Bankilare and Ballayara in Tillaberi region. In Mali, the communes are located in the circles of Ansango and Gao in Gao region (Figure 1).

Figure 1. Maps of the two regions in Niger (top) and Mali (bottom) where the SUR1M is operating



The initial target groups represent 80% of the total population living in this part of the Niger River Basin, who share cultural, ecological, political and socioeconomic factors and vulnerabilities, including cyclical food insecurity, underlying gender barriers and weak local governance. Aiming to reduce climate change vulnerability and build resilience in this highly vulnerable zone, SUR1M leverages traditional, positive, coping mechanisms and promotes innovative, locally adapted and proven strategies across the two countries.

3. Evaluation framework

3.1. Purpose and scope of this impact evaluation

In line with the mandate of the BRACED KM, the focus of this evaluation is on learning, not on accountability. For this reason, the results are not explicitly organised according to standard evaluation norms (e.g. the Organisation for Economic Co-operation and Development (OECD) Development Assistance Committee (DAC) criteria, OECD, 2010), although criteria of effectiveness and impact are thoroughly addressed. The overall purpose of the evaluation is to determine the extent to which household-level resilience has increased as a result of BRACED interventions.

Using household-level composite indices, we aim to measure changes in resilience for beneficiaries and to determine whether project interventions have made a statistically significant difference to resilience over the period in question. We recognise that this could be positive, negative or neutral (i.e. no change detected). This will provide answers to the core evaluation question: *To what extent has beneficiary resilience increased as a result of BRACED interventions?*

Adopting such a focused approach allowed us to develop and apply a (statistically) rigorous quasi-experimental evaluation framework (see below) and go one step further compared to most other resilience analyses currently found in the literature. It however also implies that we did not explore every possible dimension of the SUR1M interventions. In particular since our unit of analysis is the household, the evaluation does not allow to investigate intra-household or gender dynamics, nor community or higher-level policy environment.

3.2. Choice of the quasi-experimental method

A central issue in the design of evaluations for programme such as BRACED is that establishing cause and effect in a classic, linear sense (i.e. Intervention X results in observable effect Y) is challenging for two major reasons. First is the complexity of the programme and in particular the fact that BRACED interventions are not single, stand-alone, interventions delivered in a perfectly stable and controlled environment, but rather a combination of different interventions delivered together as a 'package' (with exact composition varying depending on location and households needs), and implemented in a 'real-life' context with many other external factors also affecting the life and livelihoods of these beneficiaries.

Additionally, since the BRACED interventions were not assigned at random over a large eligible population but rather explicitly targeted at specific households/communities identified as vulnerable to climatic shocks, the random assignment of subjects/households required for true experimental designs was not applicable. In such conditions, a quasi-experimental design was preferred over a more robust but more constraining Randomised Control Trial (RCT).

Several types of quasi-experimental designs exist, each with different objectives, strengths, weaknesses and applications: (i) Instrumental Variables (IV), (ii) Regression Discontinuity, (iii) Difference-in-Difference and (v) Matching (see Table 1). Among these, no single design is considered best; instead, the choice usually depends on a number of factors, including the purpose of the evaluation, how long the evaluation has been planned, if data from before the programme started (baseline) is available, whether there are data or resources for data collection at several measuring points along the way in the programme or whether there may be a possibility to generate a control or a relevant comparison group for the intervention group (de Janvry et al., 2011; Gertler et al., 2016).

Table 1. Choice of the impact assessment framework in the context of the SUR1M project

Type of impact assessment design and generic conditions for utilisation[1]	**Applicability in the case of the SUR1M project**
Instrumental Variables (IV) - evaluating programmes when not everyone complies with their assignment	Could have been considered in the case of the SUR1M project (as some household enrolled did not participate in all activities on a regular and continuous basis). Was discarded, however, as it was hard to identify an IV that satisfied at the same time the exogeneity condition (IV should not be correlated with the characteristics of the beneficiary and control groups) and the relevance condition (IV should affect participation rates in the treatment and control groups differently).
Regression Discontinuity – evaluating programmes that use an eligibility index	Not applicable in the case of the SUR1M project as the selection criterion was not a well-defined quantifiable index (such as e.g. poverty or hectare of land) but the combination of a geographic and local participatory targeting exercise.
Difference-in-Difference – evaluating programmes when the rule of assignment is less clear	Was not applicable in the case of the SUR1M project as insufficient quality baseline data was collected in the control and treatment groups.
Matching – constructing an artificial comparison group	Given the data and design constraints (absence of baseline including a control group), this appeared to be the most appropriate approach in the case of the SUR1M project. However, using matching techniques raises issues of internal validity[2] (see discussion).

Notes: (1) Derived from Gertler et al. (2016); (2) internal validity means that the estimated impact of the programme is net of all other potential confounding factors – or, in other words, that the comparison group provides an accurate estimate of the counterfactual, so we are estimating the true impact of the programme (Gertler et al., 2016, p.71).

In preparing for the endline survey and in discussion with the SUR1M project, the baseline data in relation to resilience was found to be unreliable in terms of how the project had subsequently rolled out activities. This meant that the planned Difference-in-Difference design was not possible and an alternative approach was sought. Regression Discontinuity was not applicable as the selection criterion for the SUR1M beneficiaries was not a well-defined quantifiable index (such as e.g. poverty or hectare of land) but the combination of a geographic and local participatory targeting exercise (see 'selection of control/treatment areas' section). The IV technique was also discarded because of the difficulty in identifying the exogenous and relevant instrumental variable (see Table 1 for detail). In these conditions, an ex-post evaluation framework based on matching techniques was developed – see Section 4 on Methodology.

3.3. Conceptual framework

3.3.1. Generic theory of change of a resilience project

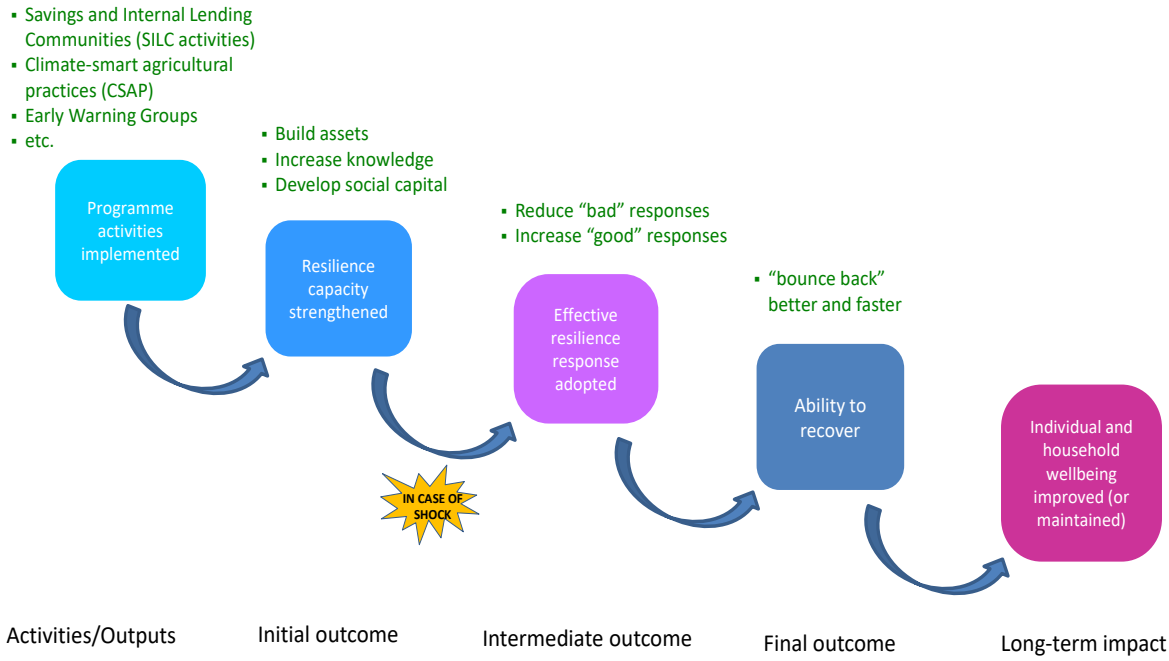
Many different conceptualisations of resilience are available in the literature (see e.g. Frankenberger and Nelson, 2013 for a review) but very few of them are formulated with an operational impact evaluation framework in mind. In this study, we follow Béné et al. (2015), who proposed one of the few resilience M&E frameworks specifically designed in relation to development objectives. Their work partially draws on recent conceptual advances made in the understanding of resilience in the context of food security (see e.g. von Grebmer et al., 2013; Constas et al., 2014). In their framework, resilience is defined as 'the ability of individuals, households, communities, institutions or higher-level systems to adequately deal with shocks and stressors', where the terms 'adequately' refers to the ability to avoid short- and longer-term negative impacts (Béné et al., 2015, p.6).

One of the key principles that underlie recent conceptualisations of resilience is recognition that resilience should not be seen as the final goal of a development programme but instead as an intermediate outcome required for achievement of a more fundamental goal related to a longer-term developmental ambition, typically a measure of well-being (e.g. food security, health/nutrition status, poverty). This means that programmes should not have resilience as their ultimate objective. Rather,

the ultimate goal of development programmes/projects should remain the improvement of people's well-being. A second important principle that emerges from recent progress made in conceptualising resilience for development is that resilience is fundamentally about ability and capacity (we defined resilience above as the *ability* of individuals, households, communities, institutions or higher-level systems to adequately deal with shocks and stressors).

Framed into a theory of change, this understanding of resilience implies that the activities of a resilience programme should be designed to achieve a first initial outcome in the form of strengthening the **resilience *capacities*** (absorptive, adaptive, transformative) of the target population. This initial outcome should then lead to the intermediate outcomes, which is the adoption of **appropriate responses** in the face of shocks and stressors. These appropriate responses are leading to the *actual* improved resilience of the target populations (understood as the ability of these populations to handle shocks). This can be measured by assessing the effective **recovery** (the 'bouncing back better' element as presented in DFID's initial resilience framework (DFID, 2011)), which eventually is expected to lead to the programme's ultimate goal – that is, improving the **well-being of the target population**. Figure 2 presents this generic theory of change.

Figure 2. Generic theory of change of a resilience intervention



Source: Derived from Béné et al. (2015).

The process of formulating such a theory of change is also useful as it brings measurement requirements into focus. In particular, it highlights some of the key components that need to be included in the M&E system.

3.3.2. Assessment framework and hypotheses

Based on the generic theory of change presented in Source: *Derived from Béné et al. (2015).*, we formulated three assessment hypotheses that will need to be tested within a treatment/control framework – one at the intermediate outcome level, one at the final outcome level and one at the ultimate impact level.

Hypothesis 1 (intermediate outcome level): If the SUR1M project has been effective in building beneficiary households' resilience capacities, the types and/or frequencies of response(s) put in place by treatment households are expected to differ from those adopted by control households under the same circumstances. More specifically, we expect to observe:

- SUR1M beneficiaries to show lower propensity to adopt detrimental (coping) responses
- SUR1M beneficiaries to show higher propensity to adopt positive (adaptive/transformative) responses

} Hypothesis (1)

Hypothesis 2 (final outcome level): Those more appropriate responses are expected to lead to stronger resilience (higher ability to 'bounce back' adequately) in the face of adverse events. More specifically we expect to observe:

- SUR1M beneficiaries to show higher level of recovery rate than households in the control group (everything else being equal)

} Hypothesis (2)

Hypothesis 3 (long-term impact level): The higher ability to 'bounce back' adequately in the face of adverse events should translate into higher well-being indicators. More specifically, we expect to observe:

- SUR1M beneficiaries to show higher level of food and nutrition security than households in the control group (everything else being equal)

} Hypothesis (3)

3.3.3. Impact pathways and formulation of the statistical models

To help formulate the statistical models that will be used to test these three hypotheses, the impact pathway of the SUR1M project at both outcome and impact levels was elaborated, drawing on a series of discussions with CRS staff.

Intermediate outcome-level pathway (Hypothesis 1)

At the intermediate outcome level, the main outcome of the project is assumed to be related to households' response to specific adverse events. The choice of these responses is assumed to be influenced by a series of different household characteristics and external factors. At the basic level, some of the household demographic and socioeconomic characteristics (e.g. age, gender, education, size of household) are assumed to be potentially important determinants of these responses (in theory we could expect, for instance, that younger and older heads of household may differ in their choices when faced with the same shock/stressor. Similarly, we could expect men and women to respond differently to the same stressors or shocks) (e.g. Kumar and Quisumbing, 2014). Other household characteristics, such as income and asset levels, social capital (e.g. 'social connection') and access to information/knowledge, are also very likely to influence households' choice (e.g. del Ninno et al., 2001; Hoddinott, 2006; Woodson et al., 2016).

The other major component that is assumed to influence households' choice is the perception that people have about the severity of shocks/stressors. Shocks/stressors have by definition some impact on households' assets and income, but also on household members' emotions and cognitive processes, and therefore also influence people's choice of response.

Translating this impact pathway into a model equation, we have:

$$Resp_{shock} = f(HH_{char}; shock_{char}; SUR1M) \quad (1)$$

where $Resp_{shock}$ stands for the probability of adopting one particular type of response in relation to a particular shock; HH_{char} represents the sets of households' various demographic and socioeconomic characteristics; $shock_{char}$ refers to the type, severity and impact of adverse events affecting the

households; and *SUR1M* is the dummy variable representing the SUR1M project effect. The objective is to test the significance of the SUR1M variable for particular types of response through the estimation of logit models[3] structured around equation (1). The advantage of using the logit model is that a logit model allows us to test for the influence of the SUR1M project while controlling for potential confounding factors such as the ones mentioned above. Finally, note that the test is to be carried out for each combination shock/stressor response.

Final outcome-level pathway (Hypothesis 2)

At the final outcome level, we assume that the outcome (households' ability to recover) reflects the combination of three main dynamics: (i) the direct effect of the initial adverse event, (ii) the 'mitigating' effect of the responses put in place by the households and (iii) the effect of the external support received by the household – including the humanitarian help (if any) delivered in the aftermath of a severe event. Note that, for the effect of the responses, the term 'mitigating' has been put in inverted commas. The reason for this is that, although households effectively adopt the responses household with the objective of mitigating the immediate effect of the adverse event, these responses may themselves induce other (short- or long-term) negative effects. The detrimental effect of some particular coping strategies (e.g. distress selling of assets, reduction of expenses or food consumption) has long been recognised and documented (e.g. Sinha et al., 2002; Dercon et al., 2005; Hoddinott, 2006; Kazianga and Udry, 2006). This detrimental long-term outcome, however, can also be observed in the case of adaptive (or even transformative) responses, leading to what is referred to in the climate change literature as 'maladaptation' (Barnett and O'Neill, 2013; Macintosh, 2013).

The dependent variable tested under the second assessment hypothesis is the ability of households to recover from specific shocks or stressors. This ability was measured through a self-assessed recovery index estimated at the household level through a series of recall questions and psychometric techniques (self-reporting evaluation using a Likert scale) – see details below. Translating this impact pathway into a model equation, we have:

$$Recov = f(shock_{char}; Resp; support) \quad (2)$$

where *Recov* is the recovery rate representing the (self-assessed) ability of households to recover from past events; $shock_{char}$ refers to the type and severity of adverse events affecting the households; *Resp* is the type of response put in place by the households; and *support* is a variable capturing the level of support received by these households. The objective is to compare the recovery rate *Recov* between the treatment and the control groups (SUR1M beneficiaries versus non-beneficiaries). This corresponds to testing Hypothesis 2 at the final outcome level. In order to conduct this comparison, the approach adopted in this analysis has been matching techniques – see detail below. Besides, since resilience is shock-/stressor-specific, the comparison is carried out for each shock reported by the households.

Long-term impact level pathway (Hypothesis 3)

The intended long-term effect of resilience interventions is the improvement (or at least the maintenance) of the well-being of households and communities that have been affected by various shocks and stressors. Hypothesis 3 captures this idea by conjecturing that the level of well-being of the SUR1M beneficiaries should be at least as good – or possibly better – than the well-being level of households/communities that did not benefit from the project activities. In our case, the well-being was measured through food security and diet diversity indicators recorded at household level.

[3] Logit models are regression models in which the dependent variable is a probability, for instance the probability of occurrence of a particular event – in our case of households engaging in one particular type of response.

Arguably, the levels of these two well-being indicators have not been influenced only by the impact of the SUR1M activities. They also reflect some other factors and determinants such as income or assets levels, familial situation, access to health services and other key development factors that are recognised in the general literature as critical to the improvement of people's well-being. In order to control for the impact of these other confounding factors, we used (again) a matching technique approach applied to the following model:

$$Well = f(HH_{char}; shock_{char}; Resp; support) \quad (3)$$

where *Well* is an indicator measuring the household well-being; $shock_{char}$ refers to the type and severity of adverse events affecting the households; *Resp* is the type of response put in place by the households; and *support* is a variable capturing the level of support received by these households.

3.3.4. Matching techniques

Both Hypothesis 2 and Hypothesis 3 are tested using matching techniques. The use of matching techniques is justified here by the methodological difficulty of trying to compare two groups that are not comparable in the first place. Indeed, since SUR1M beneficiaries are households that were initially selected by the project because they were recognised to be poor and vulnerable to climate-related shocks (thus were not randomly selected), comparing them with non-beneficiaries means that differences between these two groups can reasonably be expected, in particular in our case in their ability to recover from shocks/stressors. Whether this difference is the result of the SUR1M project (treatment) or reflects some other (initial) characteristics (e.g. aversion to risk, level of income or assets, education, social network, etc.) is difficult to determine. Matching techniques have been developed to address this issue of selection bias (Rosenbaum and Rubin, 1983; Heckman et al., 1997; Dehejia and Wahba, 2002).

3.4. Resilience index

In theory, a rigorous impact-level assessment of resilience would require measuring the instantaneous change in a household's well-being, welfare (income, assets), food security or nutrition indicators following shocks/stressors (Béné et al., 2015). Since none of this information had been collected at the time of implementation of the project, we had to base the assessment on a 'simpler' resilience indicator that was easier to construct in an ex-post context. For this, we used an approach that had recently been field-tested in four countries (Ghana, Fiji, Sri Lanka and Vietnam) (Béné et al., 2016). In this approach, resilience outcomes are assessed using psychometric techniques (self-reporting evaluation using a Likert scale) whereby households are asked to assess the degree of recovery they managed to achieve for each adverse event experienced in the previous two years. The self-assessment process is based on two distinct recall questions: (i) self-recovery from past events; and (ii) self-recovery compared with the rest of the community. For each question, respondents select appropriate answers from a 6-point Likert scale system for question (i) and a 5-point Likert scale system for question (ii). The index is then constructed by combining the answers at the household level, producing a self-assessed level of resilience (or ability to recover) – noted *Recov* in model 2 above – which is specific to shocks/stressors.

3.5. Well-being index

Many different indicators of household (or individual) well-being exist in the literature (see e.g. Gough et al., 2007), several of which are built around happiness or quality of life (see e.g. Diener et al., 1985). In our case, however, given the overall context of the Sahel, it was deemed more appropriate to consider dimensions of well-being that are more closely related to the level of household food and

nutrition security. We therefore used the standard Household Food Insecurity Access Scale (HFIAS) (Coates et al., 2007) and Household Dietary Diversity Scores (HDDS) (Swindale and Bilinsky, 2006) as the basis for the two indicators of well-being. Furthermore, these make particular sense in the context of a resilience project where food and nutritional security are usually two of the first well-being indicators that may be affected by shocks and stressors.

For the HFIAS, we used a modified version of the original score proposed by Coates and her colleagues. Instead of the standard 9-question score system, we used a simplified 3-question-based score system[4] (Annex 1 provides the detail of the questionnaire). In our case, the HFIAS score would then be a categorical vector (gradient) varying from 0 to 9, where 0 indicates a very severe household insecurity level and 9 an 'adequate' food security level.[5] The HDDS was computed according to the standard approach suggested in Swindale and Bilinsky, and we then used the 1 to 12 gradient as a basic indicator of nutrition security where low scores indicate poor nutritional security at the household level and higher scores indicate better nutritional situations. Items in the 12 food groups have been adjusted to reflect local product and food habits (see Annex 2). Local key informants reviewed the final version of both the HFIAS and the HDDS questionnaires.

3.6. Deviation from original design

The original design for this evaluation was approved by DFID in September 2015 and is captured in the "Detailed Design Document"[6]. That document sets out the intended design based on discussions with DFID and the Implementing Partner who were at the point of beginning baseline and implementation work when the design was taking place. This section describes and justifies deviation from the design as originally conceived. While overall, the quantitative work has largely followed the original design (see Limitations in Section 6.1 for details), there was a change in the approach to qualitative data collection.

Assessing intervention effectiveness:

The detailed design document indicates that in addition to the core question of whether household resilience had increased as a result of the BRACED project, the evaluation team would seek to address the 'How and why have particular intervention packages led to observed results and changes?' Ultimately, this question was picked up by the BRACED Evaluation Activity 2 work which focused at the intervention level so the focus here shifted to a project level assessment in line with evaluation question 1 and explored via the three hypotheses presented in 3.3.2.

While testing these three specific evaluation hypotheses we also made the deliberate choice not to explore another question potentially relevant for the impact pathway of the programme, which is: which of the SUR1M activities -or combination of activities- are more likely to be associated with specific responses at the household level? One could indeed hypothesize that certain activities could be more 'successful' at influencing the adoption of particular responses than others and this findings would be important; for instance if a correlation could be found between one specific activity and the propensity amongst households to adopt some particular detrimental coping strategies, or, in

[4] The justification for this simplification was the need to keep the overall number of questions manageable. As the quasi-experimental evaluation survey was administered at the same time as the CRS endline survey (see next section), we needed to ensure the total number of questions included in both questionnaires was 'reasonable' and did not require an unacceptable time from the household heads.

[5] For coherence reasons and in order to ensure the two indicators vary in the same direction when they improve (worsen), the 'sense' of the HFIAS gradient has been inverted: in the original version a 0 would indicate an adequate level of food security and higher levels would indicate a degradation in that food security.

6 The Detailed Design Document contains sensitive financial and organisational information but a redacted version can be made available upon request.

contrary, with the propensity to engage in adaptive or even transformative strategies, this information would obviously be extremely useful from a programmatic perspective. Unfortunately, because the SUR1M project was designed in such a way that several activities (in some cases more than 10) were proposed in each village and not every household in one given village was participating in all these activities (in fact even within the same household, members may not have been participating to the same combination of activities), the design of the SUR1M project does not allow to test such hypothesis. The potential number of combination of activities to test does indeed become rapidly exponential and the size of the groups of households benefiting from one specific combination diminishes at the same exponential rate, making a rigorous statistical test impossible or requiring an unmanageable sample size. Had SUR1M programme proposed only one (or a small number) of activities, it would have been possible to explore this additional hypothesis.

Qualitative triangulation

Reduction in KM led qualitative work: The original plans for qualitive work which were made in September 2015 before the baseline work began assumed that the KM would lead this work. As the project progressed and at the point of planning for the final data collection round in July 2017, it became clear that the lead IP was mandated to collect qualitative data for its final evaluation report (a contractual milestone with the Fund Manager). Any work that we had planned would therefore have been duplicative and instead we worked closely with CRS to help define questions and lines of enquiry which they then contracted a third party to manage. The end product was the independently produced final evaluation report. Whenever possible, relevant information from the SUR1M programme endline survey (CRS 2018) which was completed during the same period by the same team of enumerators (see Section 4), was used to triangulate our findings.

4. Methodology

4.1. Selection of control/treatment households and sampling approach

4.1.1. Sampling approach

In the absence of reliable baseline data for the majority of the variables of interest for our evaluation (e.g. resilience index, shocks/stressors/responses, etc.), we adopted an ex-post treatment versus control approach where the responses (preliminary outcome), ability to recover from shock/stressors (final outcome) and well-being (final impact) from the treatment group (households that benefited from the project by being direct recipients of the SUR1M activities) are compared with the responses, ability to recover and well-being of control (non-beneficiary) households. The appropriate choice of control households is critical in order to reduce the potential effect of selection bias.[7] The control group needs to be as comparable as possible with the treatment group in terms of socioeconomic and demographic characteristics as well as exposure/vulnerability to shocks and stressors.

Initially, the SUR1M beneficiaries had been selected by the project through a targeting process combining both geographic and local participatory targeting, whereby the households (identified through participatory targeting and voluntary membership at the community level) were selected within communes considered the most exposed to extreme events (geographic targeting).

For security reasons, it was agreed that the EA3 evaluation would be conducted in Niger only – that is, no villages/communes were surveyed in Mali as part of this quasi-experimental evaluation. For logistical and financial reasons, it was also decided that the EA3 survey would 'piggyback' on the SUR1M endline survey that was to be carried out at the same time and in the same regions as the EA3 evaluation. This means that the same enumerators who were conducting the SUR1M endline survey collected the information and data for the EA3 evaluation. As a consequence, the overarching sampling strategy developed for this work reflects a compromise between the needs of the endline and the needs of the EA3 surveys. Within this dual objective, the overall sample size was computed to be statistically representative of project beneficiaries and to detect a change of 0.10 in proportion of beneficiaries residing in households with moderate or severe Hunger Scale scores between survey rounds (baseline and endline surveys) (CRS, 2017). This means that a minimum of 1,322 households were to be sampled.[8] On the other hand, the matching techniques that were to be used to test the EA3 Hypotheses 2 and 3 required the group of control households to be slightly larger than the group of treatment. A 60:40 ratio was therefore adopted between these treatment and control households, with 60% control and 40% treatment.

The overall sampling was structured as a two-stage cluster sampling. The sampling was first stratified by 'high-intensity' and 'low-intensity' villages based on electronic records of activities implemented in each village. This information served as the sampling frame and the basis for the first stage, consisting of a random selection of villages (clusters) with probability proportional to village size. The second stage involves the use of systematic random sampling to select a set number of households from the project beneficiary lists at the village level.

4.1.2. Selection of control/treatment households

The distinction between beneficiary and non-beneficiary was made at the village level. Villages were selected purposively as either 'high-intensity' or 'low-intensity' villages. High-intensity villages were

[7] Selection bias occurs when intervention participants are non-randomly drawn from the beneficiary population, and the criteria determining selection are correlated with outcomes.

[8] Details of the calculation are provided in the CRS endline inception report (CRS, 2017).

defined as villages where at least 4 activities (and up to 13) had been (or were being) implemented by the SUR1M project at the time of the evaluation. Low-intensity villages, on the other hand, were villages where only one activity – the SUR1M radio broadcast – had been implemented. In theory, these villages are villages where no SUR1M activities were carried at all. However, since it is not possible to ensure that households in villages located in the radius of the radio broadcast did not receive the project's messages, it was hypothesised that these villages did receive and benefited from at least the SUR1M radio broadcast activity. Avoiding this spillover effect[9] and being able to select control villages that had not benefited from any activity in the SUR1M project at all would have required selecting villages/communes further away from the intervention zone. However, since we also needed the control households to be as comparable with the treatment households as possible (in terms of socioeconomic, livelihood but also shock/stressor variables), we needed to select control villages sufficiently close geographically to the treatment villages. In these conditions, selecting control villages likely to be within the radius of the radio broadcast was the best compromise we found between the two constraints of (i) maximising comparability and (ii) reducing to a minimum the potential spillover effect.

4.1.3. Surveys

The survey was conducted by 24 trained enumerators divided into four teams, under the supervision of four supervisors and the CRS-Niger monitoring, evaluation and learning leader. The data collection took place during a five-week field mission (October to November 2017) after the enumerators and field supervisors had been provided with a three-day training to familiarise themselves with the questionnaire and the objective of the survey. Data collection was carried out during face-to-face interviews with the head of household and information was recorded using electronic devices (tablet pc).

4.2. Statistical models

4.2.1. Intermediate outcome model (Hypothesis 1)

In order to test the first evaluation hypothesis, a series of logit models structured around the generic equation (1) were estimated. These models were tested individually for the five most frequent shocks/stressors identified by the households, combined with the five most frequent strategies/responses used by those households to mitigate the impact of these shocks/stressors. Twenty-five models (5×5) were therefore constructed of the form:

$$Resp\_A_S = \alpha_1 + \sum_i \alpha_{2,i} HH_{i,j} + \alpha_3 sub\_res_j + \sum_k \alpha_{4,k} S\_char_{k,j} + \alpha_5 SUR1M_j + \varepsilon_j \quad (3)$$

where $Resp\_A_S$ is the probability of households engaging in Response *A* as an attempt to mitigate the impact of a shock/stressor *S*, $HH_{i,j}$ is the set of household characteristics *i* recorded for the pool of households *j* that have reported being affected by *S*, $S\_char_{k,j}$ is the set of characteristics *k* (severity, frequency, impact) of the shock/stressor *S* reported by each household *j* and $SUR1M_j$ is the dummy variable representing the SUR1M effect (*SUR1M*=0 for control; *SUR1M*=1 for treatment). The objective is to test the statistical significance of α_5. A significant α_5 would indicate that the SUR1M project has an impact on the probability of households engaging in $Resp\_A_S$.

4.2.2. Final outcome model (Hypothesis 2)

The objective in this part of the analysis is to estimate the ability to recover of the SUR1M beneficiaries and compare this with that of the non-beneficiaries. Since this ability is expected to be related to the characteristics (severity, duration, nature) of the shocks that affect people, those comparisons were

[9] Spillover effects refer to situations where members of the control group are affected by the intervention.

conducted separately for each subgroup of households that reported having been affected by the same types of shocks/stressors. More specifically, we focused on the five most frequent shocks/stressors reported by the households and we compare the household ability to recover to each of these shocks using the self-assessed resilience index as computed in Section 3.4.

To run this comparison, we relied on matching techniques. The fundamental idea of matching techniques is to use statistical techniques to construct an artificial comparison group that can be used as a counterfactual. In our case we used the Propensity Score Matching (PSM) tests.

The first step in the completion of the PSM test involves estimating a logit model that predicts the probability of each household being enrolled in the SUR1M project as a function of observed household and community covariates (known as the propensity score), using a household sample that contains both beneficiaries and non-beneficiaries of the SUR1M project.

By then matching beneficiaries and non-beneficiaries on the basis of the probability that the former are enrolled in the project, the approach basically allows for a combination of the observables (in our case household and community characteristics) to control for factors affecting both treatment and non-treatment (Rosenbaum and Rubin, 1983). The underlying assumption is that, after controlling for these observables, SUR1M beneficiaries have the same average outcome as non-beneficiaries would have, had these beneficiaries not been included in the project, thus creating a constructed control and treatment group.

The propensity scores are then used as a device to 'balance' the observed distribution of covariates across the treated and the untreated groups, with the objective of trying to reach equality of the mean and standard deviation of the observed characteristics across the SUR1M beneficiary sample and the non-beneficiary comparison group sample.

If this condition is satisfied (this test is called the 'balancing propensity' test; Rosenbaum and Rubin, 1983; Heckman et al., 1997), then the next step involves testing the 'match'. This means using the propensity scores estimated in the first instance to identify non-beneficiary matching households that compare with a beneficiary household (i.e. with the closest propensity score values). Several techniques are available for this. We used the 'nearest neighbour matching' algorithm.

Once the matching is done for each beneficiary household, the impact estimate (average treatment effect for the treated, or ATT) is constructed by computing the difference in outcome for each matching pair (the treated unit – a SUR1M household – and its nearest neighbour – a non-SUR1M beneficiary) and then the mean difference across pairs. Standard errors of the impact estimates are estimated by bootstrap using 100 replications.

4.2.3. Long-term impact model (Hypothesis 3)

The general approach adopted for testing Hypothesis 3 (the difference between the well-being indicators of SUR1M beneficiaries and of non-beneficiaries) is similar to the approach adopted to test Hypothesis 2 in the sense that we also use a PSM test. This time, however, the test is not run on subgroups of households affected by specific shocks but on the entire beneficiary group (treatment) compared with the entire non-beneficiary group (control). The justification for this is that the objective here is to determine if, in the long run, the SUR1M beneficiaries manage to maintain (or possibly improve) their general well-being despite the series of combined and overlapping shocks and stressors that continuously affect their life and livelihoods. In other words, we are interested through this third hypothesis in the overall 'picture' and not the shock-specific situation adopted in the first two hypotheses.

We recall that the two indicators of well-being that have been chosen for this analysis are the HFIAS (Coates et al., 2007) and the HDDS (Swindale and Bilinsky, 2006). The analysis consisted therefore of two PSM tests run in parallel, the first one comparing the HFIAS of SUR1M beneficiaries with that of non-beneficiaries while controlling for confounding factors and the second one doing the same comparison but with the HDDS.

4.3. Limitations and potential bias

A series of limitations – or 'compromises' – in the methodology could in theory have affected some aspects of the evaluation and should be recognised. First, the evaluation was implemented in Niger only, with no fieldwork conducted in Mali (the other country in this project), owing to security and access issues. We are therefore not assessing the whole project but only the Niger component or treatment arm. This was made clear at design stage and was the intention. A question may be raised as to whether the results as we observed them would have been the same ones if some data had been collected in Mali. It is difficult to answer this question with certainty. It is likely that the detailed list and the ranking of shocks specific to the region where the SUR1M project has been operating in Mali may have differed slightly from the list identified in Niger, but this in itself would be confirming the local-specific nature of shocks and stressors.

Another compromise introduced in the methodology is related to the fact that, for logistical and financial reasons, the EA3 field survey was coupled to and administered at the same time as the SUR1M endline. One of the consequences of this was that the number of questions included in the survey instrument for this assessment was slightly reduced to ensure the overall number of questions remained 'acceptable' for respondents. We ensured, however, that all the questions that were necessary to generate the information required for the evaluation were included in the questionnaire.

The third potential limitation relates to the choice of the control group and the control of the wider environment. In theory, a control group should be chosen to reduce as much as possible any risk of spillover effect.[10] The villages included as control were, however, not totally outside the zone of influence of the SUR1M project, and in particular were located within the PRC activity zone. At the same time, the control group needed to be as comparable with the treatment group as possible in terms of socioeconomic, livelihood but also shock/stressor variables. Selection of a control group within the radius of the radio broadcast was the best compromise we could find. The actual impact of the PRC activity is difficult to quantify, but we might expect that this resulted in an underestimation of the actual effect of the project. In other terms, our results are probably slightly conservative. The second potential issue regarding the choice of the control group relates to the presence of other NGOs in the zone of intervention of the SUR1M project. The activities implemented by these NGOs have potential effects which could add to, complement, or affect the impact of the SUR1M project –and as such could potentially 'contaminate' the evaluation analysis and in particular the comparison between the treatment and the control groups. The evaluation was however designed to account for that possible effect. Questions were included in the questionnaires which were specifically formulated in order identify the presence and potential impact of those 'external' actors', so that this information could then be included in the logit models used for the testing of the three hypotheses and their effect controlled for.

[10] We recall that the spillover effect refers to situations where members of the control group are affected by the project intervention even though they are not deliberated targeted.

Finally, the last important potential limitation relates to the way households' level of resilience was measured, and in particular the fact that it was generated through self-assessed indices. While self-assessed indices could be subject to some form of bias, the questions included in the module on resilience were carefully crafted and had already been tested in several low-income countries prior to this evaluation (Béné et al., 2016). Furthermore, since the test (Hypothesis 2) was based on the *relative* values between the two groups and not on the absolute values of these self-reported resilience indexes, if a bias had been introduced its effect would be removed by the test procedure – assuming of course that control and treatment households would be subject to the same level of bias. There is no reason to assume otherwise, however.

5. Findings

This section presents the results of the quasi-experimental assessment in a series of related sub-sections that address the three key hypotheses presented in Section 3. Results and analysed data are presented in tables and figures with key findings highlighted in bold and then discussed further in the subsequent paragraphs.

5.1. Comparability of the households to shocks

5.1.1. General exposure to shocks

A total of 49 villages in 11 communes were surveyed, 22 high intensity and 27 low intensity (see Tables 3.1 and 3.2 in Annex 3 for the complete list of villages included in this evaluation). These villages/communes are located in 4 livelihood zones (LZs): LZ04, LZ05, LZ06 and LZ09. In total, 1,498 households (812 control and 686 treatment) were surveyed, a proportion that is broadly in accordance with the initial sampling design agreed.

A series of preliminary basic analyses were performed to compare the treatment (T) and control (C) groups in the different LZs. In Niger, interventions were carried out in LZ04 = agro-pastoral belt; LZ05 = planted millet and sorghum belt; LZ06 = cropping/herding with high work outmigration; and LZ09 = Niger river irrigated rice. A closer look at those LZs (see Fig.1 above) reveals, however, that LZ09 is included in LZ05 and LZ06 is part of LZ04. In the rest of the analysis, the four initial livelihood zones were therefore aggregated into two groups: LZ04+LZ06 and LZ05+LZ09.

Some general household characteristics were found to be similar between the two groups across the four zones. These include number of active members per household (4.25 and 4.39 members, T/C respectively (p = 0.382); age of household head (49 and 47 years, T/C respectively); and even income (p = 0.768). We also compared the mean value of the household assets, accounting for treatment and control and for the sex of the household head using a two-way ANOVA. The comparison shows no significant difference in asset levels between treatment and control households (p = 0.7894), but a difference between male and female (***p* < 0.001**). However, the interaction between the two factors was not significant (p = 0.2570).

When the main livelihood activities between LZs and treatment/control were compared using a two-way ANOVA, no significant differences were found, meaning that treatment and control groups do not differ significantly in terms of type of livelihood across the four LZs. The vast majority of the households are involved in farming activities (T 88.2% and C 88.2%), followed by trading/commerce (T 4.7% and C 3.1%) and inactive (T 1.3% and C 3.6%).

However, and despite our initial effort to ensure that the control and treatment groups were comparable by selecting them in the same geographic areas, some other characteristics were found to differ (at least at the inter-group average level). **In particular, the level of exposures to shocks and stressors (estimated by the number of shocks/stressors reported by households for the 24 months prior to the survey) was found to be different, with households in the treatment group exposed to a higher number of shocks/stressors than those in the control group** (***p* < 0.001**).

A two-way ANOVA test was then carried out, looking at the potential effect of the sex of the head of household (male/female) and the T/C status on the level of exposure to shocks/stressors. The comparison shows a difference between treatment and control (***p* < 0.001**) – confirming the result above – but no difference between sex ($p = 0.189$) or in the interaction ($p = 0.980$), meaning that the sex of the head of household does not seem influence the level of exposure to shock.

A comparison between levels of exposure to shocks/stressors for treatment and control accounting for the LZs (two-way ANOVA) showed a significant difference between treatment and control (***p* <0.001**), LZs (LZ04+LZ06 and LZ05+LZ09, ***p* < 0.001**), and their interaction (***p* < 0.001**), meaning that the difference (T-C) in the exposure to shock/stressors is different between the two aggregated LZs.

Table 2 Covariant and idiosyncratic shocks and stressors – comparison between T and C groups (t-test results)

Type of shock	n_Control(1)	n_Treat(1)	p-value	PC : PT
Covariate				
Seasonal drought (not enough rain during normal rainy season)	456	413	.115	
Sudden flood owing to heavy rain	66	91	.001 ***	PT > PC
River flood	4	8	.159	
Very high temperature	5	13	.031 *	PT > PC
Locusts attack/outbreak	61	75	.024 *	PT > PC
Armyworm attack	258	161	<.001 ***	PC > PT
Large livestock disease / death	56	36	.196	
Epizootic among medium-size livestock (goat, sheep, etc.)	75	122	<.001 ***	PT > PC
Small livestock epizootic (chicken, duck, etc.)	52	92	.001 ***	PT > PC
Commodity (food) price peaks	120	138	.007 **	PT > PC
Idiosyncratic				
Sudden death/disability of head of family	30	42	.030 *	PT > PC
Sudden death/disability of family member (other than head)	84	65	.664	
Accident (physical injury)	6	3	.442	
Serious illness, including diarrhoea and/or malaria of family member	350	324	.118	
Family separation (divorce/abandonment)	6	17	.009 **	PT > PC
Attack by thieves	5	13	.031 *	PT > PC
House fire	1	10	.005 **	PT > PC
Soudure (seasonal food shortage)	150	84	<.001 ***	PC > PT
Other	45	64	.005 ***	PT > PC

Note: (1) Number of households reporting these shocks/stressors. One household can report several different shocks/stressors. Total number of shocks/stressors reported 3,601; control = 1,830; treatment = 1,771

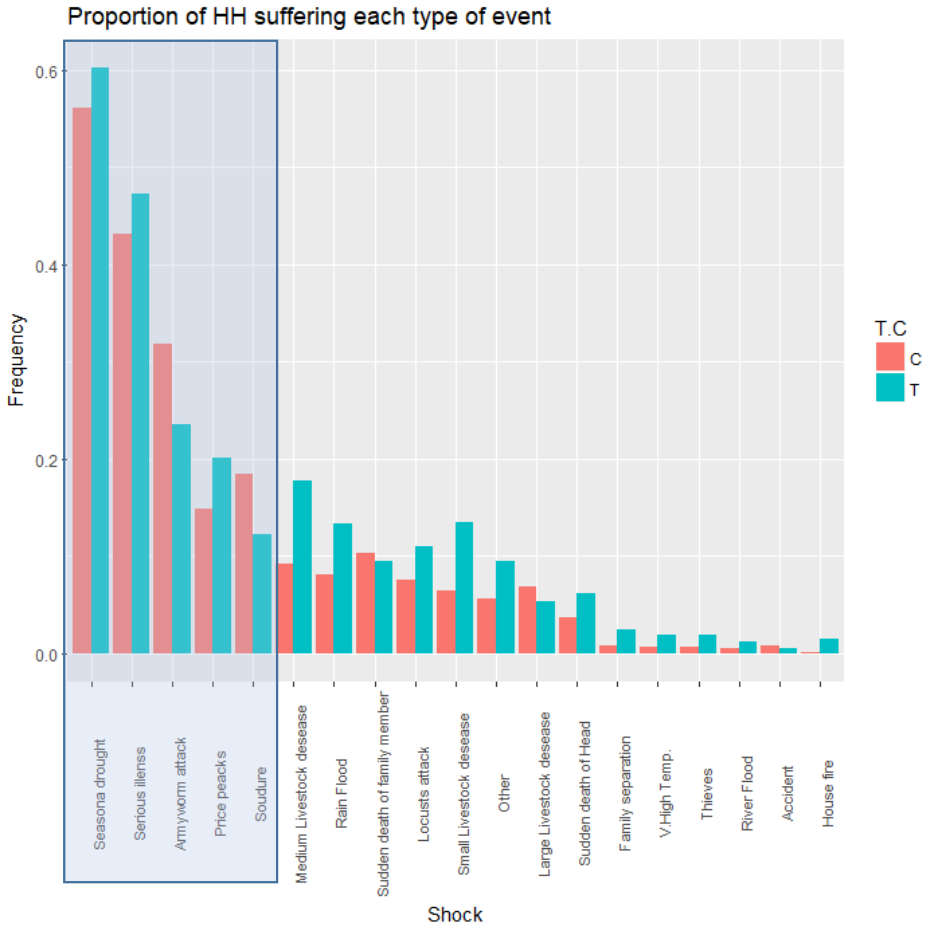
* = significant at 5% ; ** = significant at 1% ; *** = significant at 1‰

The nature of these shocks/stressors is also slightly different, with control households more frequently affected by particular shocks/stressors such as armyworm attacks or seasonal food shortages while treatment households seem to be more exposed to other types of covariant shocks such as flash floods after strong rain, food price peaks, epizootics among small livestock or even some idiosyncratic events[11] such death of the main breadwinner (Table 2). On the other hand, both groups reported

[11] Idiosyncratic events are events (shocks or stressors) that affect individual or households independently (e.g. illness) – that is, without affecting their neighbours, while covariant shocks/stressors are events that affect a whole groups/community at the same time (e.g. a flood).

similar levels of exposure to seasonal drought (covariant) and to loss of large livestock or serious illness (idiosyncratic).

Figure 3 Adverse events (shocks/stressors) that have affected households during past two years



Note: Highlighted in the blue rectangle are the five shocks/stressors on which the rest of the analysis will focus.

Figure 3 represents these results graphically. It shows that, with the exception of armyworm attacks and price peaks, the proportion of treatment households reporting adverse events is always slightly higher than the proportion of control households reporting the same types of events – illustrating the earlier result that households in the treatment group seems to be more exposed to shocks/stressors than households in the control group.

Figure 3 also shows that **the five adverse events most frequently reported by households during the surveys are (i) seasonal drought (C 57% and T 60%), (ii) serious illness (C 42% and T 47%), (iii) armyworm attacks (C 32% and T 24%), (iv) food and general commodity price peaks (C 15% and T 20%) and (v) seasonal food shortage (C 18% and T 16%).**

5.1.2. Shocks and gender

There appears to be no significant difference between male- and female-headed households and the number of shocks experienced. A series of five independent two-way ANOVA tests were carried out for the five most reported shocks (seasonal drought; serious illness; armyworm attacks; price peaks; seasonal food shortages), looking at the possible independent effects, and interaction between, the sex of the head of household (male/female) and the enrolment in the SUR1M project (T/C), on the number of shocks reported. Only seasonal drought shows some significant differences

for both factors and their interaction (p = 0.012). The other four shocks have no significant differences for the two factors.

Two additional two-way ANOVA tests were performed with the aggregate counts of the five most reported shocks, looking at the possible independent effects of, and interaction between, the sex of the head of household (male/female) and enrolment in the SUR1M project (T/C). Both factors show no significant difference on either factor (p = 0.4346 for T/C; p = 0.2601 for sex of household head).

5.1.3. Direct effects on households

The data suggests that households in the treatment group are slightly less affected[12] by adverse events than households in the control group, as illustrated across a range of variables and even in the context of greater exposure to climate shocks.

When comparing the direct (immediate) effect of adverse events on households' life and livelihoods, the figures on income losses show a significant difference between the treatment and the control groups (t-test; $p < 0.001$), with households in the control group more often reporting shock-induced income losses than households in the treatment group (Table 3a).

Table 3 Implication of adverse events on households (estimated through income or asset losses or intra-household conflicts/tension

	(a) Income loss				(b) Asset loss				(c) Intra-household tension			
Group	Obs.	%	[95% CI.]		Obs.	%	[95% CI.]		Obs.	%	[95% CI.]	
T	615	62.76	58.81	66.60	618	43.85	39.90	47.86	615	8.12	6.08	10.56
C	736	68.61	65.12	71.95	739	53.85	50.18	57.50	736	12.09	9.82	14.67
***p*-value**		**< 0.00001 *** **				**0.00024 *** **				**0.01653 ** **		

* = significant at 5%; ** = significant at 1%; *** = significant at 1‰. CI = confidence interval.

Figure 4 (Self-reported) assessment of how households handled shock/stressors the last time they occurred (in percentage of households)

[12] We make here the distinction between the exposure to shocks/stressors (proxied by the number of adverse events reported by individual households) and the effect of these events (proxied by income loss, assets loss and intra-household tension).

Note: A Pearson Chi-2 test confirmed that treatment households are better able to handle shocks/stressors than control households.

A similar pattern is observed for losses in assets induced by adverse events (Table 3b) and with intra-household tension created by adverse events (Table 3c). This finding is in line with the results shown in Figure 4, where we observe (graphically) that control households have found it slightly harder to cope with the latest shock that has affected them than households in the treatment group. This graphical observation was confirmed by a Pearson Chi-2 test ($p < 0.001$). These different results are particularly striking when we consider the earlier finding that showed that treatment households are more *exposed* to adverse events.

5.2. Responses to shock and stressors

5.2.1. Detrimental coping strategies

In this section, we first explore the specific responses – both negative ('coping strategies') and positive ('adaptive/transformative responses') – adopted by households when faced with shocks. These results will provide some preliminary results that relates to our Hypothesis 1.

Starting with negative responses, we look more specifically at the probability of households engaging in five coping strategies that are commonly considered in the literature to be detrimental (Corbett, 1988; Devereux, 1993; Davies, 1996; Coates et al., 2006). These are (i) reduced food consumption; (ii) changing the type of food consumed; (iii) reduced family expenses; (iv) taking a loan; and (v) selling assets.

The data indicates that the most frequently adopted coping strategy in the communities surveyed is one that consists in 'taking a loan': more than 60% of households (across both control and treatment groups) have engaged at least once in such a strategy over the past two years. The next most frequent coping strategy is 'reduced family expenditure' (48%), followed by 'reduced food consumption' (45%), 'selling assets' (40%) and 'changing the type of food consumed' (31%).

It is notable that, in four out of five cases, the probability of engaging in negative coping strategies is systematically higher among the control group than among the treatment group (Table 4), and in two out of the four cases the difference is statistically significant: for 'reduced food consumption' and 'changing the type of food consumed'.

These results suggest that, for the five coping strategies, the treatment group never displays a higher propensity to engage in detrimental responses than the control group (even if they had initially reported to be more exposed to adverse events –cf. section 5.1.1)**.** At worst, treatment households display the same propensity to engage in detrimental coping strategies as the control households; at best they show a lower propensity (in two out of the five cases).

Table 4 Percentage of households engaging in various negative coping strategies when faced by shocks or stressors (for the top five shocks/stressors)

	(a) Reduced food consumption				(b) Changing type of food consumed			
Group	Obs	% (*)	[95% conf. interval]		Obs.	% (*)	[95% conf. interval]	
T	268	24.41	21.893	27.061	177	16.73	14.528	19.118
C	409	29.57	27.177	32.056	304	22.03	19.820	24.362
Total	677	27.29	25.542	29.086	468	19.67	18.092	21.327
Difference		-5.16	-8.663	-1.667		-5.29	-8.470	-2.128
***p*-value**		**0.0041 *** **				**0.0012 *** **		

	(c) Reduced level of family expenditure				(d) Taking loans			
Group	Obs.	% (*)	[95% conf. interval]		Obs.	% (*)	[95% conf. interval]	
T	325	31.49	28.666	34.424	386	36.17	33.288	39.141
C	392	30.29	27.798	32.879	520	38.68	36.076	41.354
Total	717	30.82	28.952	32.747	906	37.58	35.64	39.545
Difference		1.20	-2.583	4.980		-2.51	-6.399	1.370
***p*-value**			**0.534**				**0.2055**	

	(e) Selling assets			
Group	Obs.	% (*)	[95% conf. interval]	
T	255	24.57	21.975	27.303
C	342	25.95	23.600	28.405
Total	597	25.34	23.593	27.147
Difference		-1.38	-4.912	2.148
***p*-value**			0.4439	

Note: (*) percentage of number of times this specific strategy was reported among all recorded responses.
* = significant at 5%; ** = significant at 1%; *** = significant at 1‰.

The analysis has focused so far on the *types* of negative coping strategies adopted by households. The *duration* for which these coping strategies are employed is, however, an equally important dimension of resilience. For four out of the five coping strategies considered, the questionnaire therefore also included questions about this duration.[13]

Results (Table 5) show that, when both treatment and control households do engage in detrimental coping strategies, control group households appear to deploy some of those strategies for longer. In particular in the case of 'taking a loan', the duration of adopting this strategy is significantly longer (p = 0.001) for the control group than for the treatment group.

Table 5 Average time (number of days) during which households engage in coping strategies

Coping strategy	mean T	mean C	diff T-C	min CI	max CI	*p*-value
Reducing amount of food	111.01	122.58	-11.57	-34.71	11.57	0.3263
Changing type of food	141.62	130.15	11.47	-19.97	42.9	0.4733
Reducing level of family expenditure	148.91	135.73	13.18	-9.43	35.79	0.2527
Taking loans	137.12	189.06	-51.94	-83.51	-20.37	**0.0012 *** **

* = significant at 5%; ** = significant at 1%; *** = significant at 1‰.

Control group households appear to resort to even riskier detrimental coping strategies than treatment group households. In addition to the duration of adoption of these coping strategies, more detailed information was collected on the ways the strategies were applied. For instance, in the case of loans, respondents were also asked who provided these loans. The results (Table 6) suggest an important contrast between the two groups. While the major sources of borrowing for the treatment group are friends/neighbours and family, as well as NGOs, the control households rely mainly on the

[13] The only coping strategy for which such duration of adoption was not investigated was 'selling assets', which is usually a one-off strategy.

owners of local shops (more than 38% of the time) and local money-lenders (in slightly less than 10% of cases). In fact, in this case the difference between control and treatment is significant ($p = 0.0001$). Other important – and statistically significant – sources of loans include NGOs ($p < 0.001$) and the community-based saving groups (SILC) ($p = 0.04$).

Table 6 Sources of loans from which households borrow money in response to the impacts of shocks/stressors

Source	T	C	diff	95% CI		*p*-value
1. Friends / neighbours	35.48	32.78	2.7	-15.48	20.891	0.7676
2. Family	25.8	17.22	8.58	-7.777	24.945	0.2554
3. Local money loaners	0.00	8.889	-8.889	-13.046	4.731	**0.0842**
4. Local shop owners	9.68	38.33	-28.65	-41.256	-16.056	**0.0018** ***
5. Community saving group	6.45	1.11	5.34	-3.442	14.123	**0.0440** *
6. Local cooperative	0.00	0.56	-0.56	-1.641	0.53	0.6774
7. local NGOs	22.58	0.00	22.58	7.862	37.299	**<.0001** ***

* = significant at 5%; ** = significant at 1%; *** = significant at 1‰. In bold are proportions that are significant at 10%.

There appears to be no significant difference between treatment and controls groups in the specific assets sold as part of the strategy that consists in 'selling assets'. The data shows that the main type of assets sold as a response to a shock or a stressor is medium-size livestock (e.g. goats), in more than 50% of cases, followed by small livestock (e.g. chickens) (about 22% of the cases) and large livestock (18% of cases) (Figure 5a). Other assets mentioned were jewellery and agricultural tools. Pieces of land were reportedly in less than 1% of cases. For none of these cases does the difference between the treatment and the control groups appear significant (*t*-tests).

Figure 5 Shock responses in terms of asset sales (a) and expenditure (b)

Figure 5a. Nature of assets sold in response to impact of shocks/stressors

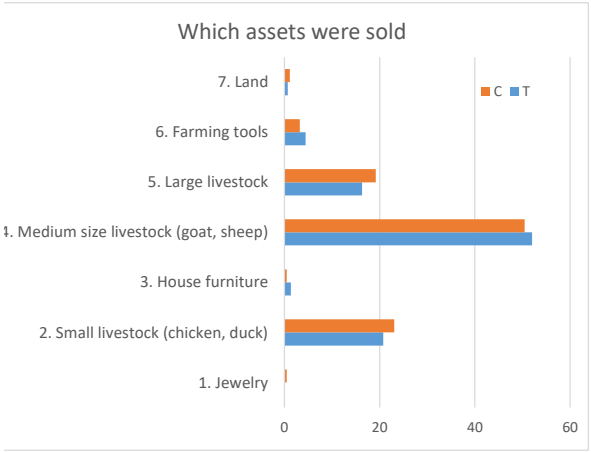
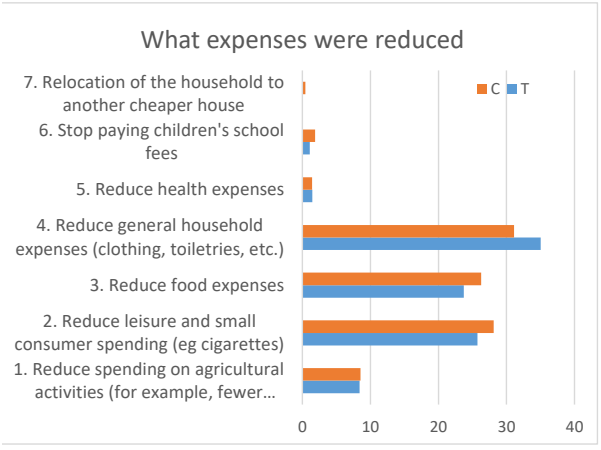


Figure 5b. Types of expenditure reduced as a consequence of shocks/stressors



In the same way, the data reveals no statistical difference between treatment and control groups and the types of reduction in expenses they make to cope with shocks (Figure 5b). The main types of expenditures reduced are 'general' (e.g. clothes, toiletries, etc.) (in about 33% of cases), followed by leisure and small consumption (cigarettes, etc.) (27% of cases), food (about 25%) and agricultural inputs (e.g. use of fertiliser) (8%). Reduction of expenditure related to school (e.g. school fees) was mentioned in only about 2% of cases.

5.2.2. Adaptive or transformative strategy

Analysis has so far focused on the negative/detrimental coping strategies adopted by households in the face of shocks and stressors. However, in order to assess whether the SUR1M project has

effectively increased beneficiaries' resilience, it is important to explore whether it has also strengthened those households' ability to engage in more positive (i.e. adaptive or transformative) responses. Households were asked to list any new on- and off-farm activities or strategies they had started (or at least tried out) in anticipation of or in response to shocks or stressors.

Table 7 shows the most frequently adopted on-farm adaptive strategies (both treatment and control households) are 'saving seeds in advance' and 'purchasing improved seeds', followed by 'saving money'. Treatment and control groups do differ in some of these on-farm strategies. Control households show a (statistically) higher propensity to engage in (1): 'saving seeds in advance' and (10) 'working as agricultural labourers', while treatment households show a (statistically) higher propensity to engage in strategies (5) 'getting more information about early warning systems (EWS)', (6) 'applying conservation agriculture techniques' and (7) 'using drought-tolerant seeds'.

Table 7 Adoption of on-farm adaptive/transformative strategies

Response (in %)	T	C	diff	95% CI		*p*-value
1. Save seeds in advance	16.84	28.25	-11.41	-17.62	-5.21	**0.000** ***
2. Store water in advance to irrigate	3.83	4.13	-0.3	-3.2	2.6	0.839
3. Use improved seeds	24.74	22.86	1.88	-4.42	8.19	0.559
4. Construct dikes to protect my fields	1.02	0.63	0.39	-0.94	1.71	0.579
5. Try to get more information about EWS	7.65	0.95	6.7	3.86	9.54	**0.000** ***
6. Apply conservation agriculture techniques	13.01	5.71	7.3	3.09	11.5	**0.001** ***
7. Cultivate drought-resistant crops	10.97	7.3	3.67	-0.55	7.89	**0.096**
8. Move activity to a new site	1.02	0.33	0.69	-0.47	1.88	0.268
9. Save money	16.32	12.06	4.26	-0.87	9.36	0.109
10. Work as a labourer (on somebody else's field)	3.57	11.75	-8.18	-12.18	4.17	**0.000** ***
Total = 707	392	315				

* = significant at 5% ; ** = significant at 1% ; *** = significant at 1‰.

As far as new off-farm activities are concerned (Table 8), the most frequently adopted strategies (for both control and treatment households) is general 'diversification of income-generating activities (out of agriculture)' followed by 'urban migration', 'making food reserves' and 'starting home gardening'. Within this, the control households show a higher propensity to engage in (6) 'starting home gardening' while treatment households display a higher propensity to engage in (7) 'engaging in fishing' and (9) 'contacting the extension services'.

Table 8 Adoption of off-farm adaptive/transformative strategies

Response (in %)	T	C	diff	95% CI		*p*-value
1. Started other non-agricultural activities to diversify sources of income	30.65	31.98	-1.33	-8.75	6.08	0.723
2. Made food reserves	15.58	17.00	-1.42	-7.34	4.50	0.636
3. Reinforced the structure of my roof	0.26	0	0.26	-0.25	0.77	0.422
5. Some household members have invested in raising small animals	6.75	6.48	0.27	-3.69	4.24	0.892
6. Some other household members started a home garden	14.55	24.69	-10.14	-16.58	3.72	**0.001** ***
7. Some household members are fishing	6.23	0	6.23	3.82	8.65	**0.000** ***
8. Some household members migrate to urban centres	15.84	18.21	-2.37	-8.41	3.67	0.435
9. Getting in touch with extension services	8.31	0.81	7.5	4.53	10.48	**0.000** ***
10. Building dikes	1.81	0.81	1	-0.73	2.75	0.296
Total = 632	385	247				

* = significant at 5% ; ** = significant at 1% ; *** = significant at 1‰.

We then explored whether treatment and control households differed overall in their respective propensities to engage in these adaptive/transformative responses. **Tables 9 and 10 show that, for both on- and off-farm strategies, SUR1M beneficiary households do display a statistically significant (at the 99% level) higher propensity to engage in positive strategies than control households.**

Table 9 Comparison of propensity of households to engage in on-farm adaptive/transformative strategies in anticipation of or response to shocks/stressors

Group	Obs.	Proportion	[95% CI]	
T	1003	25.52	22.85	28.34
C	1270	19.21	17.08	21.49
Total	2273	22.00	20.31	23.77
Difference		6.31	2.85	9.77
p-value		**0.0003 *** **		

* = significant at 5% ; ** = significant at 1% ; *** = significant at 1‰.

Table 10 Comparison of propensity of households to engage in off-farm adaptive/transformative strategies in anticipation of or response to shocks/stressors

Group	Obs.	Proportion	[95% CI]	
T	991	27.14	23.40	30.03
C	1270	15.90	13.94	18.03
Total	2261	20.83	19.17	22.56
Difference		11.24	7.82	14.66
p-value		**0.000 *** **		

* = significant at 5% ; ** = significant at 1% ; *** = significant at 1‰.

These different results contribute to our preliminary attempt to better understand households' responses to shock and stressors. They all tend to suggest that Hypothesis 1 may be (at least partially) valid. However, these preliminary results are based on simple descriptive statistics that do not account for some potential confounding factors. What we propose in the next section is to test more rigorously this Hypothesis 1 by controlling for these potential confounding factors.

5.3. Testing the preliminary outcome-level effect (first evaluation hypothesis)

The first evaluation hypothesis aims to evaluate the effectiveness of the SUR1M project at the preliminary outcome level. If the SUR1M project has been effective, then

- SUR1M beneficiaries are expected to show lower propensity to adopt detrimental (coping) responses and
- SUR1M beneficiaries are expected to show higher propensity to adopt positive (adaptive/transformative) responses

} Hypothesis (1)

We propose testing these two sub-hypotheses through logit models of the form:

$$Resp\_A_S = \alpha_1 + \sum_i \alpha_{2,i} HH_{i,j} + \sum_k \alpha_{4,k} S\_char_{k,j} + \alpha_5 SUR1M_j + \varepsilon_j \quad (3)$$

where $Resp\_A_S$ is the probability of households engaging in Response *A* as an attempt to mitigate the impact of a shock/stressor *S*, $HH_{i,j}$ is the set of household characteristics *i* recorded for the pool of households *j* that have reported being affected by *S*, $S\_char_{k,j}$ is the set of characteristics *k* (severity, frequency, impact) of the shock/stressor *S* reported by each household *j* and $SUR1M_j$ is the dummy variable representing the ER effect (*SUR1M*=0 for control; *SUR1M*=1 for treatment). The objective is to test the statistical significance of α_5.

5.3.1. Detrimental coping strategies

In order to keep the analysis focused on the most relevant shock–response combinations, we apply the logit models only to the five most reported shocks/stressors and the five most frequent responses (i.e. 5x5=25 models). As an illustration, Annex 4 shows the results of the model estimated for the combination 'seasonal drought'–'reduced food consumption'. **The estimation shows that being a beneficiary of the SUR1M project reduces statistically the probability of engaging in the strategy 'reduced food consumption' (p = 0.081).**

The other 24 logit models are estimated using the same approach. Table 11 summarises the results obtained for the 25 logit models estimated across the 5 major shocks/stressors and 5 main types of response, focusing exclusively on the dependent variable of interest for our Hypothesis 1 (the SUR1M dummy variable), while details of the estimated models and their goodness of fit are presented in Annex 5.

Results in Table 11 show that the SUR1M projects reduces the probability that households will engage in detrimental coping strategies for 24% of the shock–response combinations (6 out of the 25 models tested). This reduction is statistically significant in 5 out of these 6 models (at a 10% significance level)**.** For those 5 statistically significant cases, we then calculate the marginal effects dy/dx[14] to determine the *magnitude* of the ER project's effect. **Results show that participation in the SUR1M project reduces the probability of engaging in detrimental coping strategies by 8% to 39% – depending on the shock–response combination (Table 11).**

For the other 18 combinations of shock responses, the SUR1M project does not seem to have either a negative or a positive impact on households' probability of adopting coping strategies (p > 0.25). Finally, for one case, the model was discarded because the number of observations was extremely small (40 observations only).

Table 11 Results of all 25 logit models testing the effects of the SUR1M project on households' detrimental responses to the five most reported shocks/stressors

	Seasonal drought	Serious illness	Armyworm attack	Price peaks	*Soudure* (seaonal food shortage)
Response					
Take loan	**Significant positive effect (likelihood of taking a loan decreased by 8%)**	No effect	No effect	No effect	No effect
Reduce level of family expenses	No effect	No effect	No effect	No effect	No effect
Reduce food consumption	**Significant positive effect (likelihood of reducing consumption decreased by 9%)**	No effect	No effect	No effect	No effect
Sell assets	No effect	No effect	**Significant positive effect (likelihood of selling assets decreased by 16%)**	No effect	No effect

[14] The marginal effect dy/dx indicates the change in a household's probability of engaging in a particular coping strategy, if it becomes enrolled in the ER project (i.e. dummy variable *ER* changed from 0 to 1).

Change type of food consumed	**Significant positive effect (likelihood of changing type food decreased by 9%)**	(1)	No effect	**Significant positive effect (likelihood of changing type food decreased by 39%)**	Positive effect, but not statistically significant

Note: (1) The model testing the combination 'serious illness–change in type of food' was discarded because of the small number of observations (40 observations across the two groups (13 controls and 27 treatments)). Bold text indicates a statistically significant positive effect (i.e. the SUR1M project makes it less likely that a household would engage in that particular negative coping strategy) ($p < 0.10$); blue text indicates a positive but not significant effect ($0.10 < p < 0.25$); green text indicates no effect ($p > 0.25$); and red text indicates a significant negative effect ($p < 0.10$). The marginal effect dy/dx indicates the change in a household's probability of engaging in a particular coping strategy, if it becomes enrolled in the SUR1M project (i.e. dummy variable SUR1M changed from 0 to 1), expressed in %.

5.3.2. Adaptive/transformative strategies

The second sub-hypothesis of Hypothesis 1 relates to the adoption of 'positive' adaptive and/or transformative strategies.[15] To test this sub-hypothesis, we applied an approach similar to that adopted for the detrimental coping strategies above. Two series of logit models were run: one considering the five most reported shocks combined with the different on-farm strategies and one with the same five most reported shocks combined with off-farm activities. Table 12 shows the results in a format similar to that of Table 11, while details of the estimated models and their 'goodness of fit' are presented in Annex 6.

Table 12 Result of all 10 logit models testing the effects of the SUR1M project on households' adaptive/transformative responses to the five most reported shocks/stressors

Shock/stressor	Seasonal drought	Serious illness	Armyworm attack	Price peaks	*Soudure*
Responses					
Adoption of on-farm adaptive/ transformative strategies	No effect	**Significant positive effect (likelihood of adopting on-farm activities increased by 4%)**	**Significant positive effect (likelihood of adopting on-farm activities increased by 10%)**	**Significant positive effect (likelihood of adopting on-farm activities increased by 29%)**	**Significant positive effect (likelihood of adopting on-farm activities increased by 12%)**
Adoption of off-farm adaptive/ transformative strategies	**Significant positive effect (likelihood of adopting off-farm activities increased by 9%)**	**Significant positive effect (likelihood of adopting off-farm activities increased by 7%)**	No effect	Positive effect, but not statistically significant	No effect

Note: Bold text indicates a statistically significant positive effect (i.e. the SUR1M project makes it more likely that a household would engage in that particular positive strategy) ($p < 0.10$); blue text indicates a positive but not significant effect ($0.10 < p < 0.25$); and green text indicates no effect ($p > 0.25$). The marginal effect dy/dx indicates the change in a household's probability of engaging in a particular strategy, if it becomes enrolled in the SUR1M project (i.e. dummy variable SUR1M changed from 0 to 1), expressed in %.

[15] At this stage we did not attempt to make any form of conceptual distinction between adaptive and transformative strategies.

The results in Table 12 strongly suggest the SUR1M project has a positive effect on the propensity households have to engage in adaptive/transformative strategies in response to or in anticipation of a shock/stressor. In 7 out of the 10 cases, beneficiaries show an increase in their propensity to engage in these positive strategies compared with the control, and in 6 of these 7 the increase is statistically significant ($p < 0.10$). For these 6 statistically significant cases, the marginal effects dy/dx show participation in the SUR1M project increases the probability of engaging in adaptive/transformative strategies by 4% to 29% – depending on the shock–response combination considered (Table 12).

5.4. Testing the final outcome-level effect (second evaluation hypothesis)

The second evaluation hypothesis aims at evaluating the effectiveness of the SUR1M project in relation to the final outcome. At that level, if the SUR1M project is effective, then

- SUR1M beneficiaries are expected to show faster recovery rates than households in the control group (everything else being equal) } Hypothesis (2)

To test this hypothesis, we compare control and treatment households' resilience index[16] using a PSM approach.[17] The PSM approach assumes that, after controlling for all observable household characteristics, non-beneficiaries have the same average ability to recover from shocks as beneficiaries, had they not participated in the SUR1M project.

As for the intermediate outcome evaluation, the PSM models were run only for the five most commonly reported shocks/stressors. We tested the difference between self-reported ability to recover from past shocks/stressors (resilience index) between treatment and control groups. The final estimations of the PSM model are shown in Table 13; Annex 7 presents intermediary calculations and models.[18]

Treatment households reported a statistically significantly higher capacity to cope with a range of the most common shocks when compared with the control group. Table 13 shows the details of the difference in self-reported capacity to recover between the treatment and the control households calculated for the matched households (the ATT)[19] – and their *t*-statistical significance. We used a one-tailed test – as our hypothesis was that self-reported ability to recover for the treatment group is expected to be larger than for the control group. For the price peak shock, not enough matching pairs were found to be able to run the PSM test. For the four other shocks, the results show that the treatment group systematically displays a higher resilience index than the control group. In three out of these four cases, this difference is statistically significant, suggesting that our evaluation Hypothesis 2 is confirmed by the empirical data.

[16] The details of the resilience index computation were provided in Section 3.4.
[17] In our case the STATA_ package (version 13) and the command "psmatch2" (as developed by Leuven and Sianesi, 2003) were used. The matching was implemented using the nearest-neighbour matching algorithm.
[18] These intermediary calculations include the five logit models that were estimated to test the balancing property (one for each type of shock/stressor). The analysis shows that the balancing property was satisfied for each model.
[19] The ATT is the difference in mean (average) outcomes between treatment and control. In our case, it represents the difference in the self-reported capacity to recover between the treatment and control households calculated for the matched households.

Table 13 Results of the PSM test for the household resilience index computed for the five most reported shocks

	Mean treated	Mean control	ATT	Std.Err.	*t* [(1)]	
Seasonal drought	9.50	8.17	1.326	1.156	1.15	
Serious illness	11.65	9.79	1.857	1.276	1.46	*
Armyworm attack	15.60	9.76	5.840	2.905	2.01	**
Price peaks	Not enough matching pairs					
Soudure	8.27	6.94	1.331	0.819	1.62	*

Note: [(1)] One tailed *t*-critical value at ***1% = 2.326; **5% *t*-critical value = 1.645; *10% *t*-critical value = 1.282.
Value of t-statistics is based on bootstrapped standard errors (100 iterations).

5.5. Testing the long-term impact-level effect (third evaluation hypothesis)

The third evaluation hypothesis sought to evaluate the effectiveness of the SUR1M project in terms of long-term impact by looking at the effect on two household well-being indicators: food security and diet diversity. More specifically, Hypothesis 3 was:

- SUR1M beneficiaries are expected to show higher level of food security and diet diversity than households in the control group (everything else being equal) } Hypothesis (3)

To test this hypothesis, we compare the food security and diet diversity levels of the control and treatment households using a PSM approach. This time however –and in contrast with Hypothesis 2 above- the test is not run on subgroups of households affected by specific shocks but on the entire beneficiaries group (treatment). Results are shown in Table 14, while intermediary calculations and models are presented in Annex 8.[20]

Table 14 Results of the PSM test for the household food security (HFIAS) and nutritional security (HDDS) scores

	Mean treated	Mean control	ATT	Std.Err.	*t* [(1)]	
HFIAS	8.493	8.769	-0.275	0.253	-1.09	
HDDS	5.512	5.260	0.252	0.280	0.90	

Note: [(1)] one tailed *t*-critical value at ***1% = 2.326; **5% *t*-critical value = 1.645; *10% *t*-critical value = 1.282.
Value of *t*-statistics is based on bootstrapped standard errors (100 iterations).

SUR1M households are not characterised by a level of food security or dietary diversity higher than that of the non-beneficiaries. Our results indicate that for neither of the two well-being indicators (the HFIAS and the HDDS) does the treatment group display a higher ATT value than the control group, suggesting evaluation Hypothesis 3 is not confirmed.

[20] These intermediary calculations include the logit models that were estimated to test the balancing property. The analysis shows that the balancing property was satisfied for each model.

6. Lessons and conclusions

The three-year SUR1M project was implemented in Niger and Mali from 2015 to 2017 by CRS with the support of several other national and international partners. The project was one of 15 carried out in Asia and Africa by IPs as part of the £110 million DFID-funded BRACED programme.

The aim of the EA3 was to explore and evaluate the impact of the SUR1M project on the resilience of its direct beneficiaries. For this, a quasi-experimental approach was adopted, built on an ex-post treatment versus control framework. The evaluation framework was structured around the project's generic theory of change and designed to evaluate three elements: household responses to shocks/stressors (intermediate outcome); ability of these households to recover from adverse events (final outcome); and the resulting level of well-being (long-term impact) as observed for the treatment group (households that benefited from the SUR1M project) and compared with the control group (households that did not benefit from the SUR1M project).

For this, three sets of evaluation hypotheses were formulated. At the intermediate outcome level, we hypothesised that the types and duration of the strategies/responses put in place by the beneficiaries of the SUR1M project to respond to shocks/stressors would differ from those put in place by the non-beneficiaries. At the final outcome level, we hypothesised that ability to recover from shocks/stressors would be higher for the beneficiaries of the SUR1M project than for the non-beneficiaries. Finally, at the long-term impact level, we hypothesised that the well-being level of the SUR1M beneficiaries would also be higher than that of non-beneficiaries.

6.1. Preliminary findings

Preliminary analysis revealed that there was no clear distinction between the treatment and the control groups in terms of socioeconomic and demographic variables (age, size of household, main activities, income or asset level, etc.). Nor was there a distinction in terms of the nature of the adverse events that had affected their lives and livelihoods in the two years prior to the survey. However, the level of exposure to these various shocks/stressors differed between the two groups. In particular, households in the treatment group appear to have reported higher exposure to shocks/stressors than control households. The difference was found to be statistically significant. Two possible scenarios can be advanced to explain this.

1. The households and production system most vulnerable to shocks and stressors are effectively in the treatment villages. This would confirm the appropriateness of the initial geographical targeting strategy adopted by CRS to identify the intervention villages for the project, which was done in consultation with the municipal authorities and consisted in prioritising interventions for the most vulnerable villages (no bias in the results).
2. Households in treatment villages through EWS and CMDRR project activities have learned how to collect information on, monitor, analyse and respond to shocks, so they are the most likely to report on these shocks (bias introduced by the project).

Which of these two scenarios is correct (or even a combination of both) is difficult to confirm.

The same data also reveals that households in the control group found it slightly harder to handle shocks than households in the treatment group. Also, the proportion of control households reporting losses of income and assets and/or increases in intra-household tension was higher than the proportion in treatment households. These different preliminary results could be interpreted as initial evidence that SUR1M activities have had positive effects on the beneficiaries of the project since these

beneficiaries seem to do better than non-beneficiaries even though they may have been exposed to more shocks/stressors. These conclusions are based on an incomplete 'picture', however, as they involved looking only at the effect of the SUR1M project and not accounting for the fact that other confounding factors may also have influenced these outcomes. Further analyses were necessary to account and control for these additional confounding factors. Such analyses were completed in relation to the three-level hypotheses mentioned above.

6.1.1. Intermediate outcome: More impact on positive than on negative shock responses?

At the intermediate outcome level, results show that beneficiary households seem to be less likely to engage in negative coping strategies than non-beneficiaries (Table 4). More specifically, among the five most reported negative strategies, SUR1M households display a lower propensity to engage in four of them and the difference is significant for two out of these four cases: 'reducing food consumption' and 'changing the type of food consumed'. In addition to the nature of coping strategies adopted, we also investigated the duration of these coping strategies. The data shows no difference between the beneficiaries of the SUR1M project and the control, except for on 'taking loans/borrowing money', for which control households show a statistically longer period of adoption than treatment households (Table 5). These control households are also more likely to borrow from local shop-owners and local money-lenders – a strategy recognised in the literature to be risky as it is often associated with relatively high interest rates (e.g. Guérin et al., 2014).

The investigation did not focus only on detrimental responses, however. It also included the analysis of more 'positive' adaptive/transformative responses. In this regard, the data shows that SUR1M beneficiary households do display higher propensity to engage in positive strategies (for both on- and off-farm activities) than treatment households (Tables 9 and 10).

Overall, these different findings indicate that SUR1M households seem to respond more 'appropriately' than non-beneficiaries to shocks and stressors. These results are in this respect in line with our first hypothesis. In order to determine whether these various differences could be more assertively attributed to the SUR1M project – and therefore to test our first evaluation hypothesis – we then used a series of logit models. Results from these models (Tables 11 and 12) show that accounting for other co-variables somewhat tempers the initial findings: in particular for negative/detrimental coping strategies, the SUR1M project did reduce the probability of households engaging in these detrimental coping strategies in (only) 24% of the cases that were tested (6 out of 25 cases in total), with the difference being statistically significant in 5 out of these 6 positive cases (i.e. 20% of the total cases tested). For these 5 cases, the probability of engaging in detrimental coping strategies is reduced by 8% to 39% (depending on the shock–response combination considered). For all the other combinations tested, the SUR1M project appears to have neither a positive nor a negative effect on the propensity of the beneficiaries to engage in negative coping strategies.

The analysis regarding the adaptive/transformative responses reveals more assertive results. In 7 out of the 10 combinations tested, the beneficiaries of the SUR1M project show a positive difference with the control, and in 6 out of these 7 cases the difference is statistically significant. For these 6, the probability of engaging in adaptive/transformative responses is increased by 4% to 29%.

Overall, it seems therefore that Hypothesis 1 is partially confirmed by the empirical data. While the findings for the detrimental coping strategies show a mixed result, the positive influence of the project on the adoption of adaptive/transformative responses is clearer. Overall, this difference in the models (between detrimental and adaptive/transformative responses) could mean the SUR1M project is more successful at helping households engage in adaptive/transformative responses than at preventing them from engaging in detrimental coping strategies.

6.1.2. Higher-order outcome: Positive effects for treatment households

At the final outcome level, preliminary results initially indicated that, although the SUR1M beneficiaries were on average exposed to a larger number of shocks than non-beneficiaries (Section 5.1), they also reported being able to better handle these shocks (Figure 4). In order to determine whether this difference could be attributed to the SUR1M project rather than to other factors – and therefore to test our second evaluation hypothesis more rigorously – we used a statistical matching technique (PSM). The PSM confirmed the validity of Hypothesis 2. For all four shocks for which the hypothesis was formally tested, the models indicate that the beneficiaries display a higher ability to recover than the control households, and the difference was statistically significant in three out of these four cases (Table 13).

6.1.3. Impact-level: Unclear effects on household long-term status

The third component of the evaluation focused on the final, long-term impact level. At this level, we hypothesised that, overall, if households are more resilient to shocks and stressors, they should be more effective at 'protecting/buffering' their well-being against the impacts of these shocks/stressors than less resilient people would be. In the context of this evaluation, this means that, if the SUR1M were successful at building the resilience of the beneficiaries, then these beneficiaries should display a higher level of food and nutritional security (used as indicators of well-being). This was our evaluation Hypothesis 3. The comparison of the two indicators – carried out using PSM techniques – shows no specific difference between the beneficiaries of the project and the control households, suggesting that, at this stage and after two years of implementing activities, the SUR1M project does not have a clear long-term impact on the well-being of its beneficiaries.

This observation resonates quite well with the SUR1M Endline report, which concludes from the tables summarizing the results of the SUR1M project against the KIP4 « De l'analyse de ce tableau, on peut dire que globalement le projet a atteint ses objectifs, mais que ces résultats sont encore fragiles et méritent d'être consolidés. (CRS 2018, p.46)[21].

6.2. Overall conclusions

Putting together the results of these three evaluation hypotheses in light of the initial theory of change (Figure 2) reveals an overall coherent picture. The SUR1M project showed indicators of success for both absorptive/coping and adaptive/transformative responses but seems to have been more successful at fostering the positive adaptive/transformative responses than at preventing the adoption of negative detrimental responses. This in itself may not be too surprising. Looking at the list of the most adopted adaptive/transformative strategies, be these for on- or off-farm activities (e.g. using improved seeds, diversifying income-generating activities, starting home gardens, etc.), we can notice that, even if they are successful, these new activities may have a positive effect only in the medium to long term. In the meantime, households would still need to be able to buffer the immediate impacts of shocks. In other words, the wide resort to coping strategies is almost inevitable especially for households (including of SUR1M beneficiaries) which appear still to be very vulnerable to many diverse shocks and stressors. In that context the nature of the SUR1M activities (mainly oriented toward long-term developmental objectives as opposed to more short-term humanitarian interventions) might have unintentionally contributed to this result.

The effects of the SUR1M project seem to 'kick in', however, and the results of the models used to test Hypothesis 2 provide strong evidence in this regard. For the four shocks for which the self-assessed households' resilience index was computed (which are among the five most reported

[21] "From the analysis of this table, we can say that overall the project has achieved its objectives, but that these results are still fragile and deserve to be consolidated".

shocks/stressors), the results shows that the households that benefited from the SUR1M project systematically display a higher resilience index than the control group – and this resilience score is statistically higher for three of them: serious illness, armyworm attacks and seasonal food shortages. The only shock for which the resilience index of the SUR1M beneficiaries is not statistically significant (even if it is higher than that of the control group) is seasonal drought, which is the most frequently reported shock.

Treatment households may recover more quickly

Additional evidence derived from other parts of the evaluation corroborates the hypothesis that the resilience of the SUR1M beneficiaries may start surpassing that of the control groups (even if the SUR1M beneficiaries reported being initially exposed to a larger number of shocks/stressors). One of these pieces of evidence relates to the time during which households are 'forced' to engage in detrimental strategies after a shock. We recall here that resilience is often conceptualised in the literature (especially the engineering literature) as recovery speed ('the faster a system returns to its initial pre-shock state, the more resilient it is'). Although this specific definition raises many problems (not least because it is too mechanical and rules out the possibility of a 'bounce back better' or 'learning' scenario), it reminds us that effectively one way to quantify resilience is to measure the time it takes for a system to recover. In this context, it is therefore worth pointing out that, in the case of 'taking a loan/borrowing money', the data indicates that the SUR1M beneficiaries were systematically reporting a significant shorter period (137 days ≈ 4.5 months) compared with the 180 days (6 months) reported by households in the control group. When we also remember that borrowing money was reported as being the most frequently adopted coping strategy – with more than 60% of households across both the control and the treatment groups having engaged in such a strategy over the past two years – this reduction of 1.5 months (25%) may represent a significant improvement in the status of the households.

Impact-level results may take longer to mature

On the other hand, the data did not permit the drawing of any rigorous conclusions about improvement in the long-term well-being of the SUR1M beneficiaries. At this level, we had hypothesised that, if households are more resilient to shocks and stressors (which seemed to be the case here for the SUR1M beneficiaries), they should be more effective at protecting their well-being against the impacts of these shocks/stressors. The data, however, did not show any difference between the control and the treatment groups. A possible explanation for this somewhat counterintuitive result is that not enough time has elapsed since the beginning of the project and that, even if beneficiaries of the SUR1M project are effectively characterised by a stronger ability to deal with shock and stressors, this higher resilience has not yet materialised in the form of higher well-being. In the best case scenario, the SUR1M project had been operational for 28 months in the villages where it started first (May 2015). In other villages, it had been operational for less than 18 months. It is likely that change in well-being need more time than this to take place.

This last point may also explain why the beneficiaries of the SUR1M project also do not show a higher level of income or assets than the control group. Income and assets are also often considered a good proxy of well-being. In the present case, the data reveals (see Section 5.1) that, for both income and asset values, the treatment group was not different from the control group, despite displaying higher level of resilience. This absence of difference may be because a significant increase in income and/or assets may take more than just two seasons to materialise.

In sum, when presented together, the results of the three evaluation hypotheses reveal an overall picture that remains coherent. It also highlights the need to plan for an ex-post evaluation several

months after the end of the project, with the double assumption that (a) the effect of the project will continue after the closure of the SUR1M project and (b) this additional time will be enough for the improved resilience to lead to improved food and nutritional security – thus confirming the last evaluation hypothesis of this analysis.

References

Barnett, J. and O'Neill, S. (2010). Maladaptation. *Global Environmental Change* 20: 211–13.

Béné, C., Frankenberger, T. and Nelson, S. (2015). Design, Monitoring and Evaluation of Resilience Interventions: Conceptual and Empirical Considerations. Working Paper 459. Brighton: IDS.

Béné, C., Al-Hassan, R. M., Amarasinghe, O., Fong, P., Ocran, J., Onumah, E., Ratuniata, R., Van Tuyen, T., McGregor, J. A. and Mills, D.J., (2016). Is Resilience Socially Constructed? Empirical Evidence from Fiji, Ghana, Sri Lanka, and Vietnam. *Global Environmental Change* 38: 153–70.

Coates, J., Frongillo, E. A., Lorge Rogers, B., Webb, P., Wilde P. E. and Houser R. (2006). Commonalities in the Experience of Household Food Insecurity across Cultures: What Are Measures Missing? *Journal of Nutrition* 136(5): 1438S–48S.

Coates, J., Swindale, A. and Bilinsky, P. (2007) Household Food Insecurity Access Scale (HFIAS) for Measurement of Household Food Access: Indicator Guide. Washington, DC: Food and Nutrition Technical Assistance Project, Academy for Educational Development.

Corbett, J. (1988). Famine and Household Coping Strategies. *World Development* 16(9): 1099–12.

CRS (Catholic Relief Services) (2013). BRACED Full Proposal Application. Niamey: CRS.

CRS (2017). BRACED/SUR1M Final Evaluation – Inception Report. Niamey: CRS.

CRS (2018). Evaluation finale du projet BRACED/ Scaling-Up Resilience to Climate Extremes for 1 Million People in the Niger River Basin of Niger and Mali rapport provisoire, Bureau de Recherches, de Formation, d'Ingénierie et de Réalisations Agro-Sylvo-Pastorales, 111p.

Davies, S. (1996). *Adaptable Livelihoods. Coping with Food Insecurity in the Malian Sahel*. London: MacMillan Press.

De Janvry, A., Dunstan, A., and Sadoulet, E. (2011). Recent Advances in Impact Analysis Methods for Ex-post Impact Assessments of Agricultural Technology: Options for the CGIAR. Report prepared for CGIAR Standing Panel on Impact Assessment. Rome: Independent Science and Partnership Council Secretariat.

Dehejia, R. H. and Wahba, S. (2002). Propensity Score Matching Methods for Non-Experimental Causal Studies. *Review of Economics and Statistics* 84(1): 151–61.

Del Ninno, C., Dorosh, P. A., Smith, L. C. and Roy, D. K. (2001). The 1998 Floods in Bangladesh: Disaster Impacts, Household Coping Strategies, and Response. Research Report 122. Washington, DC: IFPRI.

Dercon, S., Hoddinott, J. and Woldehanna, T. (2005). Shocks and Consumption in 15 Ethiopian Villages, 1999-2004. *Journal of African Economies* 14(4): 559–85.

Devereux, S. (1993). Goats before Ploughs: Dilemmas of Household Response Sequencing during Food Shortages. *IDS Bulletin* 24(2): 52–59.

Diener, E., Emmons, R. A., Larsen, R. J. and Griffin, S. (1985). The Satisfaction with Life Scale. *Journal of Personality Assessment* 49 (1): 71–75.

DFID (Department for International Development) (2011). Defining Disaster Resilience: A Dfid Approach Paper. London: DFID.

Frankenberger, T. and Nelson, S. (2013) Background Paper for the Expert Consultation on Resilience Measurement for Food Security, TANGO International, sponsored by FAO and WFP, Rome, 19–21 February.

Gertler, P. J., Martinez, S., Premand, P., Rawlings, L. B. and Vermeersch, C. M. J. (2016). *Impact Evaluation in Practice*, second edition. Washington, DC: IADB and World Bank.

Gough, I., McGregor, J. A. and Camfield, L. (2007). Introduction: Conceiving Wellbeing in Development Contexts. In I. Gough and J. A. McGregor (eds.) *Wellbeing in Developing Countries: New Approaches and Research Strategies*. Cambridge: Cambridge University Press.

GoN (Government of Niger) (2012). Profil de moyens d'existence – zone agropastorale – Département de Tillabéri. Niamey: GoN.

Gubbels, P. (2012). A New Drumbeat for the Sahel, *Humanitarian Exchange* 55, September, 3–6. http://ec.europa.eu/echo/files/aid/countries/Sahel_humanitarianexchange055_1.pdf

Guérin, I., Morvant-Roux, S. and Villarreal, M. (eds) (2014). Microfinance, Debt and Over-Indebtedness: Juggling with Money. New York: Routledge.

Hallegatte, S. et al. (2016). *Shock Waves: Managing the Impacts of Climate Change on Poverty*. Climate Change and Development. Washington, DC: World Bank.

Heckman, J. J., Ichimura, H. and Todd, P. E. (1997). Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme. *Review of Economic Studies* 64: 605–54.

Hoddinott, J. (2006). Shocks and Their Consequences across and within Households in Rural Zimbabwe. *Journal of Development Studies* 42(3): 301–21.

Jalloh, A., Nelson, G. C., Thomas, T. S., Zougmoré, R. and Roy-Macauley, H. (2013). *West African Agriculture and Climate Change: A Comprehensive Analysis*. Washington, DC: IFPRI.

Kazianga, H. and Udry, C. (2004). Consumption Smoothing? Livestock, Insurance and Drought in Rural Burkina Faso. Discussion Paper 898. New Haven, CT: Yale University, Economic Growth Center.

Kumar, N. & Quisumbing, A. (2014). Gender Shocks and Resilience. Brief prepared for the Conference onf Building Resilience for Food and Nutrition Security, Addis Ababa, 15–17 May.

Leuven, E. and Sianesi, B. (2003). 'PSMATCH2 Stata Module to Perform Full Mahalanobis and Propensity Score Matching, Common Support Graphing, and Covariate Imbalance Testing'. http://ideas.repec.org/c/boc/bocode/s432001.html

Macintosh, A. (2013). 'Coastal Climate Hazards and Urban Planning: How Planning Responses Can Lead to Maladaptation. *Mitigation and Adaptation Strategies for Global Change* 18(7): 1035–55.

OECD (Organisation for Economic Co-operation and Development) (2010). *Quality Standards for Development Evaluation DAC Guidelines and Reference Series*. Paris: OECD.

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

Sinha, S., Lipton, M. and Yaqub, S. (2002). Poverty and 'Damaging Fluctuations': How Do They Relate? *Journal of Asian and African Studies* 37(2): 186–243.

Swindale, A. and Bilinsky P. (2006). Household Dietary Diversity Score (HDDS) for Measurement of Household Food Access: Indicator Guide (v.2). Washington, DC: FHI 360/FANTA.

Woodson, L., Frankenberger, T., Smith, L., Langworth, M. and Presnall, C. (2016). The Effects of Social Capital on Resilience Capacity: Evidence from Ethiopia, Kenya, Uganda, Niger and Burkina Faso. Report 2: Strengthening the Evidence Base for Resilience in the Horn of Africa. Nairobi: ILRI and TANGO International.

BRACED Building Resilience and Adaptation to Climate Extremes and Disasters

This material has been funded by UK aid from the UK government; however the views expressed do not necessarily reflect the UK government's official policies.

UKaid from the British people

itad