Oxford Policy Management

Independent Evaluation of African Risk Capacity (ARC)

Final Inception Report

Zoë Scott, Claire Simon, Jesse McConnell, Paula Silva Villanueva

January 2017

This assessment is being carried out by Oxford Policy Management. For further information contact felicity.lequesne@opml.co.uk.

Oxford Policy Management Limited
Registered in England: 3122495

Level 3, Clarendon House
52 Cornmarket Street
Oxford, OX1 3HJ
United Kingdom

+44 (0) 1865 207 300
Fax +44 (0) 1865 207 301
Email admin@opml.co.uk
Website www.opml.co.uk

Executive summary

Overview of African Risk Capacity

The African Risk Capacity (ARC) was established by the African Union (AU) in 2012 as an African-owned, index-based weather risk insurance pool and early response mechanism that combines the concepts of early warning, disaster risk management, and risk finance. ARC's mission is to develop a pan-African natural disaster response system that enables African governments to meet the needs of people at risk to natural disasters (ARC 2016). ARC is comprised of two entities: ARC Agency and ARC Limited. The ARC Agency is the capacity building, educational, and advocacy arm of ARC, responsible for making AU Member States and the broader public aware of ARC's mission and goals. Engagement of countries with ARC includes a 9-12 month capacity building programme on the elements of early warning, risk modelling (particularly *Africa RiskView* (ARV), ARC's proprietary software application which combines historical rainfall data with vulnerability data to estimate drought-related response costs and define triggers for the parametric insurance), contingency planning, disaster risk management and risk financing. ARC Limited is a sovereign-level mutual insurance company that provides weather-related insurance coverage to Member States.

The expected impact of ARC is, firstly, through a pooled insurance model, it should offer African countries competitive pricing for insurance products. At the national level, it should improve the ability of governments to better anticipate, plan, and respond to disaster risk by strengthening capacities, awareness, and action around DRM. Finally, at the local level, vulnerable households should be more resilient to disasters through the receipt of timely support.

Overview of the ARC Independent Evaluation

In 2015, the UK Department for International Development (DFID) awarded Oxford Policy Management (OPM) the contract for an Independent Evaluation of ARC from 2015 to 2024. There are two components to the evaluation - a two-stage formative evaluation; and a two-stage impact evaluation (see Table 2). A pilot impact evaluation will also be carried out during the formative stage. The purpose of the overall evaluation is:

- To identify and feed lessons learnt into the management of the ARC programme. This will be the focus of the formative evaluation, which will consider ARC's effectiveness and performance.
- To test if risk pooling and transfer is a cost effective way to incentivise contingency planning and ensure rapid responses to drought and other extreme weather events. The impact evaluation will consider the value of contingency planning and early responses in minimising the impact of (and accelerating recovery from) extreme weather. It will consider where, when, why and how ARC is or is not effective with the aim of contributing to the global evidence base.
- Provide accountability to the UK taxpayer for DFID's investment in ARC, demonstrating evidence that informs continued DFID investment in the programme and similar initiatives.

Table 1: ARC evaluation phases

Evaluation Stage	Description	Timing
First Formative	To test early stages of the ToC and provide an assessment of whether ARC is on the right trajectory towards achieving its outcomes.	Feb-Oct 2017

Second Formative	To test the learning adaptation cycle of ARC and continue to build evidence as to whether ARC remains on the right trajectory towards achieving its outcomes.	Feb -Oct 2019
Pilot	To design and test the set of tools to be used in the impact evaluation phase of the evaluation	Nov 2019 – Jan 2020
First Impact	Assess the ARC programme's contribution to the outcomes identified in the TOC.	Feb - Oct 2020
Second Impact	Assess the ARC programme's contribution to the outcomes identified in the TOC.	Feb - Oct 2024

As specified in the TOR, an additional amount of up to £500K may be available for additional work to complement the formative and impact evaluations. We have proposed an additional quantitative household survey using data from OPM's evaluation of the Hunger Safety Nets Programme (HSNP) in Kenya as a baseline, plus alternative options for DFID's consideration.

Progress to date

The Inception Phase for the ARC evaluation was conducted between November 2015 and January 2017. The original timeline in the TOR specified an Inception Period of 6 months. However, this was considerably delayed due to a protracted contracting period and a break of several months following the ARC Agency's development of a Strategic Framework that differed substantially from the draft Theory of Change. The primary objectives of the Inception Phase were to develop the ARC Theory of Change and the Evaluation Questions (EQ), in collaboration with DFID, ARC Agency and ARC Ltd. Methods to answer the EQs have also been developed, along with a detailed workplan and budget for the first formative evaluation and tentative plans for the other phases. A Stakeholder Engagement and Communications Plan has also been created and reviewed by DFID.

Stakeholder engagement and communications

Our strategy identifies four key stakeholder groups: 1) core evaluation stakeholders, namely DFID, ARC Agency and ARC Ltd., with whom the Evaluation Team intends to work in close partnership; 2) African stakeholders; 3) multi-lateral and bi-lateral donors; and 4) the wider international risk management community. The team will use a variety of communication methods to aid wide, effective dissemination to these different stakeholder groups:

- **Digital presence** – a page on the OPM website, reports disseminated to other key websites (including DevTracker) and guest blogs and opinion pieces
- **Written outputs** – reports, policy briefs, a literature review, a journal article and regular email newsletters to contacts. Some outputs will be translated into French.
- **Graphics** – infographics, illustrations and an animation for use in reports and presentations, to engage varied audiences
- **Events** – presentations at conferences, webinars, national and international workshops

The strategy sets out the timing for these activities, as well as detail on access, data storage, gender and M&E. All the evaluation outputs will be made public to benefit the international community. For more detail see Section 2.

ARC's Theory of Change

To develop the TOC, the Evaluation Team first conducted an extensive desk review of ARC documentation to better understand the problem, ARC's hypothesis and proposed solution. A draft TOC was then presented in a two-day workshop with ARC staff. We also conducted a rapid literature review to identify which links in the logical chain are well supported by past evidence and which were less well substantiated. A revised version was then presented to key stakeholders within ARC and with donor partners for review and comment. A more detailed summary of the TOC is provided in Section 3 and the full document is included as Annex A.8.

The TOC model identifies how ARC investments, activities and outputs will follow three different pathways leading to specific longer-term changes:

- **Pathway 1**: When a disaster hits, the timely insurance pay-out, coupled with the effective implementation of contingency plans, will enable governments to respond more effectively to support vulnerable households. By ARC '**supporting timely and effective response' by Governments**, vulnerable beneficiary households will reduce their loss of assets and livelihoods following a disaster.

- **Pathway 2**: Through dialogue and capacity building, ARC can positively influence the policy and practices of Member States around disaster risk management (DRM). By ARC '**influencing policy and practice of member states'**, African Union (AU) member countries will be better able to anticipate, plan, finance and respond to climate-related disasters in a timely, effective manner.

- **Pathway 3**: Through dialogue and coordination with the broader DRM community in Africa, ARC will be '**creating increasing value/demand for ARC products and services**' among non-member states. In time, as membership grows ARC Agency will become self-sustaining and ARC Ltd. will pay capital back to donors.

Achievement of each of these pathways rests on a set of assumptions, outlined in Annex A.2. If these pathways occur, the long-term impacts are anticipated to be that Africa will have a functional and vibrant pan-African response system that enables African governments to better meet the needs of people at risk to natural disasters. ARC will also have demonstrated that the returnable capital model of official development assistance (ODA), is a workable innovative finance mechanism that leverages ODA to access financial/insurance market capital to increase development impacts. Ultimately with this new system, African countries can remain on their development path despite weather shocks and stresses.

Evaluation questions

The evaluation questions were developed using an iterative and collaborative approach. First, we generated a set of questions based on the ARC theory of change and the broader context in which the ARC programme is positioned. This process yielded four high-level questions that link back to the ARC context and to each of the three pathways for change outlined in the TOC:

- **Q1 ARC Contex**t: To what extent does ARC's institutional setup and outputs lead to the adoption and effective use of ARC insurance products? Can this be improved?

- **Q2 Pathway 1:** To what extent has ARC contributed to in-country timely and effective responses that protect affected households' livelihoods and prevent asset loss and food insecurity?

- **Q3 Pathway 2:** To what extent has ARC influenced AU member states' capacity to anticipate, plan, finance and respond to climate related disasters generally, and more specifically in making best use of ARC?

- **Q4 Pathway 3:** Do participating governments and other stakeholders value ARC's risk pool and technical assistance? Why?

Under each of these primary questions are 3-6 summary questions that help inform the higher order question. We mapped each of these summary questions to the OECD Development Assistance Committee (DAC) Criteria for Evaluating Development Assistance to ensure they cover and inform all of the OECD-DAC evaluation criteria dimensions related to relevance, efficiency, effectiveness, sustainability, and impact. We also provided an indication of when in the evaluation process we might be able to provide some initial evidence to answer the question (e.g. first or second formative or impact assessment). For more details, see Section 4 and Annex A.3.

Evaluation approach and methodology

We have chosen a **theory-based design** given the complexity of this evaluation. For example, complexity is manifested in (i) high levels of uncertainty around how a programme will evolve and where and when it will achieve results; (ii) high degrees of interdependence across multiple stakeholder levels; (iii) emergent conditions in implementation of the programme and in the manifestation of droughts (and other natural disasters); and (iv) the co-evolutionary nature of applying ARC's contingency planning frameworks into implementation as interacting and adaptive agents self-organise. Given the explicit learning and adaptation objectives discussed in the TOR; and given the inherent challenges in implementing an experimental evaluation design for a complex programme such as ARC, a theory-based approach provides the most rigorous, thorough, and appropriate model.

Contribution Analysis (CA) is an approach that fits well with theory-based evaluations, as it involves an in-depth analysis of a theory of change from inputs to outcomes to establish a plausible contribution. The structured but flexible nature of this type of analysis lends itself to the complexities and uncertainties inherent in the ARC programme. Typically, the 'impact statement' of a contribution analysis approach emerges through the creation of a 'contribution story' rather than the result of a measured 'impact'.

Evaluation design

In response to the above design considerations, our proposed design framework identifies three workstreams falling under the theory-based paradigm, and one workstream based on an experimental design:

- **Workstream 1: Country case studies**. The bulk of the evaluation data will come from twelve country case studies, spread across the formative and impact phases of the evaluation. Case studies provide an opportunity to conduct in-depth interviews with a wide range of stakeholders about their specific experiences with ARC and review relevant documentation. Focus groups and participatory impact assessment are tools that will also be used in the impact phases. In some instances, we will look at the same country more than once to assess change; in others we will review different countries selected purposefully based on specific experiences and contextual changes.

- **Workstream 2: ARC organisational review**. As many of the evaluation questions relate to ARC Agency processes, policies, procedures and capacities, it will be necessary to collect data directly from ARC staff and related governance institutions, primarily through KIIs and documentary review. Analysis will focus on dimensions such as ARC models, contingency

planning processes, M&E processes, governance structure, historical and current capacity and training.

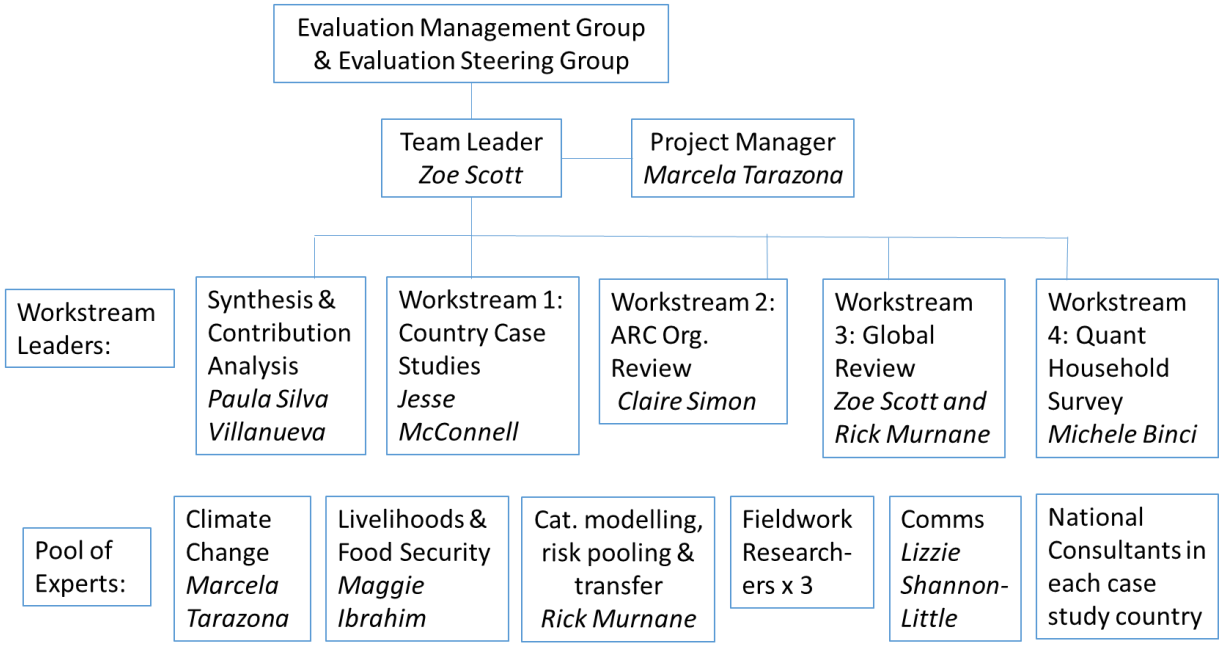
- **Workstream 3: Global review**. In order to collect data from international industry experts on DRMF and non-member states (to validate Pathway 3 which relates to continent-wide perceptions of ARC), we will conduct key informant interviews (KIIs) with selected international disaster insurance specialists, and conduct a Perceptions Survey with a broad range of representatives from ARC and non-ARC member countries. Under this workstream we will also conduct a Baseline Context Assessment, using international indicators.

- **Workstream 4: Quantitative household survey**. This is the optional extra based on an experimental design involving a household survey. See section 7 for details.

Data from the different workstreams, collected over different phases of the evaluation, will feed into the overall contribution analysis and be synthesised into the evaluation reports.

Evaluation governance and management

The diagram below sets out a revised team structure, comprising 6 OPM staff members and 6 external consultants:



As shown in the diagram above, the governance structure now incorporates DFID's Evaluation Management Team and a new Evaluation Steering Group which is being established at the time of writing. The Team Leader and Project Manager will be responsible for ensuring the risk matrix (presented in Section 10) is regularly updated and that conflicts of interest are managed appropriately. The evaluation team will draw on its experience of conducting qualitative and quantitative fieldwork to ensure that ethical standards are met, and seek further review where appropriate, to adhere to ethical protocols in line with the OECD-DAC principles of accuracy and credibility and DFID's Ethics Guidance for Research and Evaluation. Our approach will focus on ensuring informed consent is gained from all participants, anonymity is maintained and the safety of all participants is ensured.

The project manager will ensure that the evaluation progresses against the proposed workplan and within the budget set out in Annex A.9. We can confirm that the evaluation methodology proposed above can be successfully implemented within the original budget ceiling for this contract.

Table of contents

List of tables and figures

List of abbreviations

ARV	Africa RiskView
ARC	African Risk Capacity
ARC Agency	African Risk Capacity Specialised Agency of the African Union
ARC Ltd	African Risk Capacity Insurance Company Limited
AU	African Union
CoP	Conference of Parties
CP	Contingency Plan
DFID	Department for International Development of the UK Government
DRM	Disaster Risk Management
DRMF	Disaster Risk Management Finance
DQAF	Data Quality Assessment Framework
ESG	Evaluation Steering Group
EW	Early warning
EQ	Evaluation Questions
FGD	Focus Group Discussion
FIP	Final Implementation Plan
IMF	International Monetary Fund
KKI	Key Informant Interviews
ODA	Official Development Assistance
OECD DAC	Organisation for Economic Cooperation and Development's Development Assistance Committee
OPM	Oxford Policy Management
PIA	Participatory Impact Assessment
RCT	Randomized Control Trial
TOC	Theory of Change
TOR	Terms of Reference
QA	Quality Assurance

1 Introduction to the ARC Evaluation

1.1 Description of ARC

The African Risk Capacity (ARC) was established by the African Union (AU) in 2012. It is an African-owned, index-based weather risk insurance pool and early response mechanism that combines the concepts of early warning, disaster risk management, and risk finance. ARC's mission is to develop a pan-African natural disaster response system that enables African governments to meet the needs of people at risk to natural disasters (ARC, 2016). ARC has developed an integrated solution for African governments that includes the configuration of weather risk models, pooled risk insurance, and a capacity building programme in disaster risk and its management through risk financing, early warning and contingency planning. The hope is that the combined package of goods and services will provide better early warning information and allow African governments to better plan, prepare and respond to weather risk emergencies, ultimately protecting vulnerable households through the rapid provision of support to disaster-affected people.

ARC Strategic Framework, April 2016

There is evidence to suggest that responding rapidly to a disaster by getting affected households' relief before they need to reduce food consumption or sell off assets can help protect lives and livelihoods (DFID 2013). The ARC business case posits that insurance risk pooling and transfer linked to effective early warning and contingency planning is a cost effective way to incentivise planning and ensure a rapid response to drought and other weather events such as floods and tropical cyclones (DFID, 2013). Insurance that is triggered by parametric models can disburse funds quickly, often before other funding becomes available. With contingency plans in place, governments can respond quickly, addressing the needs of those impacted by disaster.

ARC is comprised of two entities: ARC Agency and ARC Limited. The ARC Agency is the capacity building, educational, and advocacy arm of ARC, responsible for making AU Member States and the broader public aware of ARC's mission and goals. The ARC Secretariat serves African Union member states, based on its mandate from the Conference of the Parties[1] and through strategic guidance from the ARC Governing Board. A key role of the ARC Agency is to engage countries to join the insurance platform. This engagement includes an extensive capacity building programme on the elements of early warning, risk modelling, disaster risk management and risk financing. ARC Limited is a sovereign-level mutual insurance company

[1] The Conference of the Parties includes the 32 African union member states signatory to ARC Treaty.

that provides weather-related insurance coverage to Member States. Beyond the provision of insurance contracts, ARC Limited engages reinsurers and capital markets to ensure competitive prices for the transfer of risk. Combined, these two entities support the ARC mission.

ARC Agency currently relies on grant funding from the UK (DFID), Sweden, Germany, USA and the Rockefeller Foundation (with Canada, France and the EU also considering grants from 2017 onwards). In 2014, the UK and Germany provided the equivalent of £30m each to capitalise the ARC Insurance Company (ARC Ltd), enabling it to sell insurance. African member countries contribute capital over time alongside their premiums, which will enable the UK and German funds to be returned after 20 years.

To be eligible to take out an insurance contract with ARC Ltd, Member States must sign a Memorandum of Understanding with the ARC Agency and successfully complete a 9-12 month long capacity building programme. Through this programme, government experts complete training in three areas: risk modelling, contingency planning, and risk transfer (ARC 2016). The risk modelling component involves the customization of *Africa RiskView* (ARV), ARC's proprietary software application. With ARC support, a country, through ARV, defines a country's weather risk profile by combining a crop water requirement satisfaction based model for drought with data on vulnerable and affected populations so as to estimate drought-related response costs. Countries are then able to determine their risk transfer needs and select the parameters which will determine the triggers for its insurance policy. While ARV currently focuses on drought, ARC is in the process of developing similar applications to model flood and cyclone risks.

The contingency planning component supports the development by the new member country of a contingency plan of how an ARC pay-out would be used. Where possible, ARC encourages governments to explore if and how on-going social safety net programmes might be scaled up in the event of a disaster so as to leverage existing operational delivery mechanisms. A viable contingency plan is a requirement to take out insurance. Completed plans are reviewed by an independent committee of technical experts who make recommendations to the ARC Agency Governing Board for approval. In the event of an insurance pay-out, a country is required to tailor the contingency plan into a Final Implementation Plan (FIP) for the specific disaster and re-submit for Board approval. In the risk transfer component of the capacity building programme, government finance and disaster management experts learn about risk transfer and risk financing decisions, of which ARC insurance is one of several options. The ARC programme emphasises building an understanding of how risk transfer via insurance fits into a broader risk management portfolio. The ARC approach is to work in partnership with each country to better understand and develop their 'disaster risk management priorities, institutional and policy environment, and existing programmes and priorities' (ARC, 2016).

The expected impact of this programme comes at many levels. First, through a pooled[2] insurance model it should, on a continental scale, offer African countries competitive pricing for insurance products. At the national level, it should improve the ability of governments to better anticipate, plan, and respond to disaster risk by strengthening capacities, awareness, and action around disaster risk management. The ARC programme also seeks to create links to existing social protection programmes, improving ongoing nation-wide resilience efforts. Finally, at the local level, through coverage from ARC insurance products, vulnerable households will be better protected from disaster risk through the receipt of timely support.

[2] A risk pool is a mechanism where individual risks are transferred and combined. That pool then takes on the risk profile of the group rather than the risk profile of each individual risk, combining the uncertainty of individual risks into a calculable risk for the group.

1.2 Overview of the evaluation

In November 2015, the UK Government's Department for International Development awarded Oxford Policy Management (OPM) the contract for an Independent Evaluation of the African Risk Capacity (ARC) from 2015 to 2024. OPM's proposal was submitted as part of e-Pact, a consortium led by OPM and co-managed with ITAD. The work conducted during the Inception Phase has been completed by the Evaluation Team, which comprises several internal staff members at OPM and a few selected external consultants (see Section 9 for more details on the team).

As set out in the TOR, there are two components to the evaluation - a two-stage formative evaluation; and a two stage impact evaluation (see Table 2). A pilot impact evaluation will also be carried out during the formative stage. The purpose of the overall evaluation is:

- To identify and feed lessons learnt into the management of the ARC programme. This will be the focus of the formative evaluation, which will consider ARC's effectiveness and performance.
- To test if risk pooling and transfer is a cost effective way to incentivise contingency planning and ensure rapid responses to drought and other extreme weather events. The impact evaluation will consider the value of contingency planning and early responses in minimising the impact of (and accelerating recovery from) extreme weather. It will consider where, when, why and how ARC is or is not effective with the aim of contributing to the global evidence base.
- Provide accountability to the UK taxpayer for DFID's investment in ARC, demonstrating evidence that informs continued DFID investment in the programme.

Table 2: ARC evaluation phases

Evaluation Stage	Description	Timing
First Formative	To test early stages of the ToC and provide an assessment of whether ARC is on the right trajectory towards achieving its outcomes. The focus is on: (i) ARC's institutional setup; (ii) ARC activities and outputs and related assumptions; (iii) ARC and early country adopters' experiences in implementation. The objective of this phase is to establish Baselines in relation to some of the country case studies, the ARC Organisational Review and elements of the Global Review. The purpose will be to identify lessons learned and make recommendations to improve the operational efficiency and effectiveness of the ARC suite of programmes, products and services.	Feb-Oct 2017
Second Formative	To test the learning adaptation cycle of ARC and continue to build evidence as to whether ARC remains on the right trajectory towards achieving its outcomes. The focus is on: (i) Evidence on uptake in learning by ARC to improve its institutional framework, products and activities; (ii) Evidence of uptake in learning by early adopters; (iii) Evidence of increasing interest & demand in ARC by member and non-member stakeholder	Feb - Oct 2019
Pilot (part of	To design and test the set of tools to be used in the impact evaluation phase of the evaluation.	Nov 2019 – Jan 2020

the formative)		
First Impact	Assess the ARC programme's contribution to the outcomes identified in the ToC. Depending on the selection of case study countries, some should be able to serve as a baseline (see section 5.4 for more detail).	Feb – Oct 2020
Second Impact	Assess the ARC programme's contribution to the outcomes identified in the ToC. Depending on the selection of case study countries, some should be able to serve as an endline (see section 5.4 for more detail). This will also include the Endline ARC Organisational Review and Global Review.	Feb – Oct 2024

The evaluation has a deliberately long timeline. The reasons for this are twofold. First, to enable early engagement with ARC Agency and to allow time for learning from the formative evaluation. Second, to give adequate opportunity for ARC's impact to unfold and become evident through iterative pay-outs across several different countries. Per the TOR, DFID's decision points on future investment are likely to be in 2018 and 2025 – as much as possible we have scheduled outputs to fit in with this timeline.

A pilot is a required milestone in the ToR, and we believe that it will be a critical step in the overall evaluation. Following best practice, OPM usually conducts pilots as part of impact evaluations and typically finds them to be a very important step in the process. In the case of ARC, the evaluation is divided into two quite distinct parts (formative and impact) with different objectives, different research questions and different methods for each. There is therefore a need to take stock after the formative evaluation of the evidence gaps that need to be filled during the impact component and to revise all the research tools across the workstreams, including creating new tools.

A component of the Impact phases is data collection at the household level using Focus Group Discussions and Participatory Impact Assessment (these are not part of the formative evaluation – see table 8). Given the importance to DFID of the evaluation being able to provide robust evidence as to the extent of ARC's contribution at the household level, this is a critically important element of the evaluation that will be introduced at this stage. A pilot is necessary to assess the appropriateness of this tool and the sample size, giving us an opportunity to make any adjustments to how we approach data collection at the household level before the main data collection phase begins. Some findings may emerge as reliable from the pilot but this cannot be confirmed in advance. The timing of the pilot will also have to remain flexible as it will need to occur soon after a payout in a country.
As specified in the TOR, an additional amount of up to £500K may be available for additional work to complement the formative and impact evaluations. During the Inception Phase, the Evaluation Team presented their idea to DFID to use the data from OPM's evaluation of the Hunger Safety Nets Programme in Kenya as a baseline for a follow-up survey that would seek to quantitatively assess the impact of ARC at the household level. Plans for this element of the evaluation are set out in section 7. We strongly recommend that DFID invest in this additional extra element, as we believe the addition of a robust quantitative impact evaluation would enhance the overall findings from the qualitative work.

The scope and ambition of ARC – to create a functional and vibrant pan-African response system to weather risk and natural disasters – is broad. As such, there will be some limitations on what the evaluation can and cannot cover. First, it will not be possible to review ARC activities in every country in which ARC operates or plans to operate. We will use purposeful sampling for our country case studies (see Section 5.10) to provide a representative sample,

covering as many scenarios and as broad a geographic scope as possible within budget. Second, the ambition of ARC is to include additional coverage models for disease outbreaks. This type of model is not within scope. Our evaluation will look primarily at drought as this model and insurance product is already available. We will also look, to a lesser extent and where possible, at models for rapid onset weather disasters such as floods and cyclones, assuming these products become operational in the timeframe of the evaluation. Third, the *ARC Strategic Framework* (April 2016) discusses plans to develop an additional two products in the form of a replica coverage plan for other humanitarian actors working in Member States and an Extreme Climate Facility fund to Member States to boost climate adaptation measures. In conjunction with DFID we agreed that assessing the value of these non-core products was outside the scope of this evaluation. Finally, while our evaluation will inform the question '*how well do ARC risk models function*?' we have agreed with DFID that it is out of scope to do a deep technical dive into the mechanics of these models. This work has already been done for *Africa RiskView* and presumably will be done for the rapid onset models.

1.2.1 Inception phase activities

The Inception Phase for the ARC evaluation was conducted between November 2015 and January 2017. The primary objectives of the Inception Phase were to develop the ARC Theory of Change and the Evaluation Questions that are to guide the various stages of the evaluation. These documents are provided in full as an addendum to this report and summarised in Sections 3 and 4, respectively. These products were developed in a collaborative manner with inputs from the key ARC stakeholders: DFID and other donors, and the ARC Agency and ARC Ltd staff. During Inception, the Evaluation Team also mapped out the methods to be used to answer the Evaluation Questions (see Section 5) and developed a detailed work plan and budget for the first formative evaluation and tentative work plans and budgets for the second formative and impact evaluations (see Section 11 and Annex A.9). The OPM team also developed a Stakeholder Engagement and Communications Strategy, provided in full in the addendum to this report and summarised in Section 2.

Specific activities undertaken during the Inception period include:

- Clarification meeting with DFID in London, including a presentation of the quasi-experimental design for a quantitative impact evaluation at household level (further developed in this report).
- 2-day workshop with all Evaluation Team members in Oxford in Nov 2015.
- Establishment of regular communication with key stakeholders, including
    - between DFID and OPM (fortnightly calls) and various meetings with DFID in London and Oxford to discuss main issues, priorities, overview of stakeholders or key outputs
    - regular email communication with ARC Agency and contact with ARC Ltd. (including later sharing of TOC and EQs and incorporating feedback).
- Clarification balancing the role of close engagement with ARC while maintaining independence (Nov 2016).

- Development of a draft Theory of Change for ARC based on wide range of documents (Nov 2015 – January 2016).
- Development of impact evaluation options and discussion with DFID (November 2015)
- Visit by a team member to ARC Agency in South Africa to build TOC in a participatory manner (January 2016).
- Preparation of key informant interviews (Jan 2016)
- Rearrangement of project in light of emergence of Strategic Framework (Feb-May 2016)
- Significant additional development work on the TOC to incorporate the ARC Strategic Framework (May-July 2016)
- Conduct of key informant interviews feeding into the TOC work (July 2016)
- Further revision to the TOC following feedback from external reviewers, ARC Agency, ARC Ltd and DFID (August 2016)
- Development of evaluation questions based on the TOC, split into 'Headline', 'Summary' and 'Detailed' questions. Review by external reviewers, ARC Agency, ARC Ltd and DFID (September 2016).
- Development of evaluation methods and some related tools, and creation of an evaluation framework (October 2016)
- Further development of the methodology for a quantitative impact evaluation element, building on OPM's HSNP data (see section 7).
- Compilation of a first draft of the literature review (September / October 2016)
- Generation of a stakeholder mapping and development and review of a Stakeholder Engagement and Communications Strategy, incorporating feedback from DFID (October 2016).
- Drafting of one-page flyer for the evaluation (November 2016)
- Ongoing revision of the workplan and budget.

The original timeline in the TOC specified an Inception Period of 6 months. However, this was considerably delayed when it emerged in January 2016 that ARC Agency had developed a Strategic Framework that differed substantially from the draft theory of change (largely based on the Business Case). As such, DFID agreed to a break in the project so that DFID and other donor partners could discuss with the ARC team the different frameworks and agree on a single unified framework. This break lasted from February to May 2016 at which point the Evaluation Team had to revise the Theory of Change document.

1.2.2 Departures from the TOR

The original TOR date from 2014 and OPM's proposal was submitted in April 2015. Ideas for the evaluation have therefore evolved considerably from the time the TOR were written, to the time of writing this Inception Report.[3] That said, the evaluation's broad purpose, scope and objectives were not altered during the Inception phase. Key departures from the TOR are outlined below.

- **Evaluation questions.** The TOR give a long list of suggested evaluation questions. Given that the methodology for the evaluation was set as a theory-based evaluation using contribution analysis as the primary analytical framework, it was agreed between the Evaluation Team and DFID to use the Theory of Change as the basis of evaluation questions. DFID acknowledged that this process would lead to a different list of questions to those provided in the TOR, but that it was not necessary for the Evaluation Team to go through and map the revised questions with those in the TOR, giving an explanation for any differences. It was deemed that this would be an unnecessary and time-consuming step, and DFID were happy with the robustness of the process that had been undertaken to develop the EQs (described more fully below and in Section 4).

- **Timelines.** All the timelines given in the TOR have been substantially delayed, with the exception of the final impact evaluation, which we still propose to take place in 2024. The delay is due in part to delays in contracting (OPM did not receive a signed contract until October 2015) and a period of several months in mid-2016 when the evaluation was put on hold due to the introduction of the ARC Strategic Framework. This issue is discussed in more detail above in Section (1.2.1). From a managerial perspective the implication was that team members had to find alternative work, leading to lack of availability and uncertainty. Considerable organisation was required to re-mobilise the team in line with project requirements. This caused a delay of several weeks as, even after work was resumed, the TOC had to be substantially further developed to integrate the ARC framework.

In addition to the above changes, several conversations have taken place between the team and DFID to clarify one further point:

- **Value for Money**. Several evaluation questions in the TOR relate to VfM and cost-effectiveness. DFID advised the Evaluation Team that ARC intends to commission an update to its existing Cost-Benefit Analysis and also that DFID hopes to commission analysis that focuses specifically on the cost-effectiveness of a range of different risk financing options and suppliers, including ARC, so this area has been de-prioritised in this evaluation. However, the team felt it would be impossible to make a full assessment of ARC without considering the affordability of ARC insurance to member governments in comparison with other options available to them (as part of the country case studies), or considering perceptions of cost-effectiveness (as part of the global case study), so these areas remain within scope. However, the team will not be in a position to conduct a full cost-effectiveness comparison with other insurance initiatives and options available to African governments. This approach was accepted by DFID.

[3] The exact date when the TOR was written by DFID is unknown to the Evaluation Team, but given the tender dates, this was a minimum of two years' ago.

1.2.3 A note on terminology

The TOR refers to an Impact Evaluation. However, the term 'impact evaluation' is not always consistently used or well understood. Some use the term broadly to encompass multiple qualitative and quantitative methodologies, whilst others use the term narrowly with the assumption that it refers to a specific experimental or quasi-experimental methods.

The Organisation for Economic Cooperation and Development's Development Assistance Committee (OECD-DAC) defines **impact** as 'positive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended' (OECD-DAC, 2002: 24). And **evaluation** is defined as 'the systemic and objective assessment of an on-going or completed project, programme or policy, its design, implementation and results' (OECD-DAC, 2002: 21).

These definitions imply that an **impact evaluation** goes beyond describing the *results* of an intervention, to understanding how the intervention *caused* these results (BetterEvaluation), placing the causal attribution of the observed results with the intervention. Similarly, other prominent international agencies provide definitions of impact evaluation that are rooted in the *causal attribution of results to an intervention* (the World Bank and 3ie in Stern *et al*, 2012: 6).

The technical underpinnings of these definitions tend to be methods-led (i.e. requiring strict adherence to a specific method), usually with an associated 'assumption that experimental and quasi-experimental methods are the best or default method' (Stern *et al*, 2012: 6). However, as Stern *et al* notes, while these experimental methods may be useful in some circumstances, they are not necessarily useful or appropriate to all circumstances (2012: 7). As a result, they redefine a use of impact evaluation, still starting with the original OECD-DAC definition of impact, but diverting the claims of attribution to the following: 'assessing the direct and indirect causal *contribution* claims of these interventions' (Stern *et al*, 2012: 11-12, emphasis added; also see Mayne, 2012).

For the purpose of this report, we have continued to use the terms set out in the evaluation report, particularly 'formative' and 'impact' to describe the different phases of the overall evaluation process. This is to aid clarity and link better with the TOR. However, we have also used the term 'impact assessment' to differentiate our proposed qualitative methods – albeit still rigorous and credible - from the optional 'impact evaluation' experimental design for a quantifiable causal attribution of impact presented in Section 7.

2 Stakeholder Engagement and Communication Plan

A full Stakeholder Engagement and Communications Strategy was developed by the Evaluation Team and reviewed by DFID as part of Inception Phase activities. This section summarises the main points while the strategy document is submitted as an addendum to this report.

2.1 Strategic objectives and challenges

One of the objectives of the evaluation as specified in the TOR is to 'identify lessons learnt that are relevant for ARC, and lessons to inform other risk pooling and transfer initiatives, i.e. insights on enabling and constraining factors, critical actions and gaps which could affect future programmes or in other contexts' (DFID_b [no date]: 2). The evaluation therefore intends to have an impact on ARC programming but also on policy debates across Africa, and in other low income countries. This section sets out the strategy for achieving that impact through stakeholder engagement at various levels and through various forms, given that a high quality evaluation is the primary output.

Supply-driven evaluations can face challenges in gaining traction for the uptake of findings, and because information for policy decision-making is often wanted long before genuine outcomes and impact can be determined. In the case of the ARC evaluation, we aim to overcome these problems by pursuing a partnership approach with ARC Agency for the formative evaluation. We will focus on relationship building with ARC in order to understand their particular constraints and strengths, but also to ascertain the best way of 'packaging' findings so that they are relevant and lessons learnt can be quickly adopted by management.

The unusually long timeframe of the overall evaluation presents us with an opportunity to build relationships over time, but it also introduces some challenges. For example, it can be difficult to keep people engaged over such a long period, and hard to manage contacts in different stakeholder organisations when turnover means engaging with an ever-changing set of individuals. There is also a need to recognise that because of the length of the evaluation, our outputs (released periodically throughout the evaluation period) ultimately will play a role in ARC's overall theory of change. Pathway 3 of the Theory of Change (see Section 3) relates to how well ARC has been able to build demand for their products and services – our evaluation findings, whether positive or negative, will no doubt play a role in either building or challenging support for ARC.

2.2 Target audiences and engagement levels

There are four primary target audiences for ARC evaluation findings: core evaluation stakeholders, African policy stakeholders, donor agencies, and the wider risk management community. Each of these audience groups will have different uses for the evaluation findings and so we propose different types of engagement for each group:

- **Partnership**: Broad, two-way engagement, continual sharing of knowledge and decision-making on process and actions;

- **Participation:** two-way engagement within limits of what is appropriate for an independent evaluation;
- **Consultation:** Limited two-way engagement, structured opportunities for interaction;
- **Push communication:** One-way engagement. Information is broadcast to stakeholders or targeted at particular stakeholder groups using various channels e.g. email updates, newsletters, presentations etc.;
- **Pull communication:** One-way engagement. Information is made available and stakeholders choose whether to engage with it, for example accessing resources from a website.

The various audience groups and planned interactions are summarised in Table 3 and are more fully described in the strategy document in the addendum to this report.

Table 3: Target audiences and engagement levels

Audience Group	Members	Evaluation Usage	Engagement approach
Core Evaluation Stakeholders	DFID	To assess the effectiveness of their investment in ARC against their own strategic priorities.	Partnership
	ARC Group (ARC Agency & ARC Ltd)	To improve their institutional set up, products and activities.	Participation
African stakeholders	African policy makers and government technicians who are directly involved with DRM	For non-member states to provide useful information in assessing whether to join ARC; For ARC Member States, to provide useful information to help assess the value of their membership.	Consultation Push communication
	Non-governmental environment who interact in African countries with government and ARC around DRM issues	To build understanding and knowledge exchange, improving NGOs ability to hold their governments to account.	Consultation Push communication
	Beneficiaries (or potential beneficiaries) of the policies i.e. citizens;	To be channelled to citizens through NGOs and the media as appropriate for informational purposes, with the aim of improving accountability.	Pull communication
Donor Agencies and IFIs	Multi-lateral and bi-lateral donors who support ARC Agency or have capital invested in ARC Ltd or are considering investments	To assess the appropriateness of their investments in ARC and as background information for policy decisions about future investments.	Consultation Push communication
Wider international risk management community	Practitioners, policy-makers and academics across related fields such as insurance, disaster risk management, humanitarian response, social protection and monitoring & evaluation stakeholders outside Africa.	To draw lessons learned that may be transferable to other contexts.	Pull communication

Source: OPM

2.3 Communication methods

The ARC evaluation should have impact on global audiences: African policymakers, international donor agencies and global thought-leaders in disaster risk insurance policy and practice. Therefore, the type of engagement will depend on the stakeholder's interest in the evaluation, their level of influence on uptake of the findings, and their relationships. Outreach to all audiences depends on identifying the right stakeholders, engaging with them, communicating the evaluation findings in the right way for those audiences; and crucially having a quality evaluation method which stands up to scrutiny.

A cost-efficient and effective way of reaching a wide audience is to attend and contribute to large-scale events and smaller-scale meetings being hosted by other organisations, including annual conferences and one-off workshops. We have already planned to 'piggy-back' ARC Agency's Governing Board meeting in January 2017, to present the final methodology and approach to the Board. Throughout the evaluation we will seek to use opportunities such as ARC Regional Workshops to both consult with stakeholders and widen our dissemination options. We will develop a list of appropriate meetings and expect this list will be updated throughout the evaluation.

2.3.1 Communication and Dissemination plan

Below we set out the main communication activities that the team will employ to reach various target audiences throughout the course of the evaluation. DFID will review and approve all outputs prior to dissemination. We will also abide by the UKaid Branding Guidelines. We have divided the communications activities into different methods. For a more detailed description on the timing, objectives of each activity, and links to different audience groups see Annex A.1.

Table 4: Communication and dissemination plan

Activities / Outputs	Channels / frequency
Digital content and channels	
Project page/ OPM website	OPM continuously update webpage as reports are finalised
Other websites	OPM to provide other disaster risk and development focused websites and communities of practice with key evaluation reports. The report will first be published on DFID's website (https://devtracker.dfid.gov.uk/) from where it will cascade to other sites such as Preventionweb. Reliefweb, ELDIS and ALNAP. A link with ARC's own website will also be established.
Blog posts and opinion pieces	Blog articles on interesting findings and experiences within or across the case study countries. To be published on OPM website or submitted to other relevant platforms that might include: Guardian Poverty Matters; Development Horizons, Ideas4Development, Evidence-Based Policy in Development Network (EBPDN), Poverty to Action (Duncan Green), The Economist.
Graphic outputs	
Info-graphics and	Various infographic software packages are readily available on the internet and OPM has a dedicated Research Uptake function who can advise on these. An

illustrations	external design company will be used for the production of any illustrations.
Animation	It will be necessary to work with an external company on the production of a short (approx. 3-minute video). This can be uploaded to YouTube and used widely in dissemination events as well as being a link embedded in documents.
Written reports	
Internal update reports	Bi-annual
Literature review	A first version of the rapid literature review will be made available at the end of the Inception Phase. Given the length of the evaluation timeline, and the rapidly increasing literature in this area, it may make sense to treat the literature review as a living document that is updated every few years, budget allowing.
Phase 1 Formative evaluation report	This report will summarise key findings from the fieldwork across the case study countries and include recommendations aimed at the national and international levels.
Phase 2 Formative evaluation report	This report will summarise key findings from the fieldwork across the case study countries and include recommendations aimed at the national and international levels. It will build on and extend the findings presented in the Phase 1 report.
Pilot impact evaluation report	A short report with some national level findings, but with an emphasis on clarifying the approach, scope and any limitations of the future impact evaluation.
Impact evaluation reports	These in-depth reports will present main findings and supporting evidence.
Briefing notes	These products would accompany the main outputs from the evaluation, and provide summary key findings. Drafts to be reviewed by DFID prior to publication.
Journal article	A one-off journal article could be planned for a critical juncture in the evaluation. This could focus on findings in a particular country or group of countries, or be scheduled for the end of the evaluation period to provide an overview of the findings.
Live discussions and reports	
Conferences / meetings	These will include attendance at ARC events in Africa, for example regional workshops and national / international conferences hosted by other organisations working on disaster risk insurance where budget allows.
Webinar	An online webinar could be hosted by OPM or another relevant organisation working on disaster risk insurance. Piggy-backing a relevant event being hosted by another organisation can be a rapid, cost-effective way to increase reach.
Workshops with the evaluation management group[4]	Discussions may be hosted by OPM or DFID, streamed online but also physical meetings.
In-country workshops and presentations	In each case study country, where possible, we will conduct a key stakeholder meeting / workshop where we can present national level findings for discussion and validation, but also (depending on timing) present preliminary findings from other ARC countries).
Promotional Materials	

[4] This group includes the DFID Project Manager, DFID Evaluation Manager, DFID internal Evaluation Advisor

Flyer/ one-pager	Distributed prior to key informant interviews, at public events and private meetings, throughout the life of the project, when relevant.
Slides presenting overview	Summary findings from each stage of the evaluation (Inception, formative 1, formative 2, impact 1 and impact 2) that can be adapted as required for different presentations.
Email newsletter	Short, email updates sent regularly (likely to be approximately three times a year depending on publication of reports to be disseminated) signposting key activities and outputs during the period.

Source: OPM

2.4 Ensuring broad access and representation

In designing the communication and dissemination plan above, the team has taken into consideration the needs of different groups to ensure that access to the evaluation findings is as broad as possible. Budget limits put some constraints on the team in this regard, but to facilitate access we will translate some outputs into French and other languages as deemed appropriate (for example, policy briefs). Although we do not consider illiterate groups to be amongst our key audiences, we have also proposed to develop a number of infographics and illustrations that we believe will enhance the visual appeal of our outputs, and aid understanding for those who may not have English as a first language.

When engaging stakeholders in an evaluation it is necessary to consider issues related to gender and equality. This is true when seeking to engage stakeholders as well as in attempts to communicate findings and disseminate outputs. Most of the communication methods mentioned in the plan above are gender-neutral (e.g. journal articles, email newsletters) but where an event is organised by the Evaluation Team (e.g. national workshops) we will endeavour to ensure an appropriate gender balance amongst both invitees and presenters. In developing the graphics, we will ensure that these are gender-sensitive and that men are women are both depicted, and in culturally-sensitive ways.

2.5 Communications workplan

As noted above, a good deal of work around stakeholder engagement has taken place during the Inception phase. In the first three months of the implementation period, the evaluation team will commence work on the more significant aspects of this strategy, alongside the commencement of the formative evaluation. Initial steps include:

- Create database, schedule and template for email newsletters. Consider how to create a regular flow of information that will not overwhelm recipients but will keep up momentum and engagement. This newsletter will carry DFID / UKaid branding and will explain that the evaluation is commissioned by DFID.

- Engage with DFID and ARC to identify suitable recipients for the newsletter. Consider how to link with contacts from other OPM projects e.g. the Shock Responsive Social Protection Research has a database of over 300 contacts for their newsletter, many of whom may be interested in joining an ARC evaluation email list.

- Circulate flyer / one-pager to potential stakeholders to publicise the evaluation and alert them to a sign-up mechanism for the newsletter.

- Create a schedule for ARC events and key meetings that would be appropriate to attend. Consider how the schedule could interact with fieldwork in order to achieve efficiency savings.

- Create schedule and likely invite list for the evaluation management group workshops and webinars. Alert invitees of scheduled dates.

- Review up-coming conferences and international events that could be relevant for engaging stakeholders or presenting findings.

- Research websites, communities of practice and knowledge management repositories that could host evaluation outputs.

- Create project page on OPM website and keep current with updated links and content.

- Create a schedule for blogposts and engage OPM communications team for support in this area.

OPM's internal communications team has provided guidance and input into the process of developing this strategy. They will continue to be available to assist the evaluation team with the implementation of the strategy, and additional media and communications consultants well-known to the company can also be contracted if needed. Budget has been made available for this purpose throughout the evaluation.

2.6 Monitoring and evaluation of the strategy

Progress with the implementation of the strategy will be reported on in the regular biannual reports to DFID during the Implementation phase. We will be able to record, track and report on a number of performance indicators related to the strategy including:

- Average number of people opening the email updates during the period
- Number of additional people signing up to the mailing list
- Number of email updates sent during the period
- Detail of interactions with ARC Agency and ARC Ltd
- Number of organisations / websites sent outputs during the period
- Team attendance at conferences / workshops
- Number of blogs / papers produced

As well as the quantitative information mentioned above, we will also seek to record and report on other information we receive during the course of the evaluation, particularly around additional communications mechanisms or the design of outputs as requested by different stakeholders. In this way, we expect this stakeholder engagement and communications strategy to be a 'living' document that is regularly updated, as the strategy evolves and is refined over time.

2.7 Confidentiality

We will ensure confidentiality of information, privacy and anonymity of all study participants. We fully understand our responsibility to ensure that their confidentiality is maintained and personal

information is protected. This will be operationalised by ensuring that all datasets are anonymised, in the sense that all names or other identifying information of respondents are removed before the data is shared publicly. Any audio recordings of the FGDs and individual interviews will be made with participants' consent, and then transcribed and translated if necessary. The confidentiality and anonymity of FGD participants and key informants will be respected and maintained at all times by ensuring that nothing which is recorded can be ascribed to a particular individual, and the transcripts and recordings will be accessible only to members of the evaluation team.

2.8 Ownership and data storage

In line with standard DFID practice, all reports resulting from the evaluation will be open access. DFID will own the intellectual property. Subject to prior approval by the DFID Evaluation Manager of all the evaluation reports and associated evaluation dissemination products and where they will be published, OPM will be permitted to publicise and use the reports. This is expected to include hosting the reports on OPM's external website and disseminating to other websites and organisations.

All the datasets with documentation from the quantitative impact evaluation will be made public to enable national researchers, masters and PhD students and other stakeholders to use the data to conduct further analysis and research. We believe this is a key way of maximising the impact of the evaluation, as well as allowing for future useful research to be conducted on related topics in Africa and internationally.

Nevertheless, releasing the quantitative impact evaluation data into the public domain will have to be a considered process, approached following agreement with DFID. A clear data structure and comprehensive metadata are the keys to making data such as ours accessible to others. Metadata refers to all the information about the data that might be useful to other researchers. It includes information on data collection procedure, data access, and a description of data limitations. Properly prepared metadata should allow researchers to gain a full understanding of the study design and be able to autonomously use the data. In addition, metadata in an appropriate format can easily be used to create an archive containing the results of various studies which can then be browsed using key words.

In order to maximize the use of the quantitative IE datasets, these must be easily accessible and stored where they will be preserved for a long time. The way to ensure this is to deposit the data into an online server facility that allows researchers to easily download the required information. There are numerous institutions that provide such online loci for collecting, preserving, and disseminating the output of their institution. In addition, there are also a number of open repositories that collect research outputs and data from other organisations and institutions in order to enhance knowledge sharing. Once the data is ready for release we will liaise with DFID to agree the process, including which data repository is preferred and how the data should be released (i.e. round by round or all at once at the end of the evaluation). OPM will then prepare the data and submit to the host repository.

3 ARC Theory of Change

The ARC Theory of Change (TOC) draws on existing evidence of past programming as well as research and practice around disaster risk insurance mechanisms and response activities. To develop the TOC, the Evaluation Team first conducted an extensive desk review of ARC documentation to better understand the problem, ARC's hypothesis and proposed solution. Based on this research, we developed an initial high-level draft of the TOC. Using this draft as a basis of discussion we conducted a two-day theory of change workshop with key ARC staff where we clarified and expanded upon the logic. As part of this process, we also conducted a literature review to identify which links in the logical chain are well supported by past evidence and which were less well substantiated. A revised version based on these inputs was then presented to key stakeholders within ARC and with donor partners for review and comment. The final output of this process is the document entitled *The African Risk Capacity Theory of Change* (October 2016). This document is provided as an addendum to this report. In the sections below we summarise the key components of the ARC TOC.

3.1 The problem

Weather extremes, and related slow and rapid-onset disasters such as drought, floods, and cyclones, damage individual and household wellbeing and broader development progress in vulnerable African countries. While African governments want to respond, with limited contingency budgets in place, accessing cash quickly comes at a cost. To pay for the disaster, most governments must either reallocate budgets away from other important programmes which, if unplanned, can be costly and sub-optimal; appeal for humanitarian aid; or borrow with limited access to affordable and timely credit (Dercon, 2015). Both options may be politically untenable and threaten to set back the country's overall development trajectory. While many governments are aware of the risks they face from extreme weather events, they have few tools for quantifying and managing these risks both from an operational and a financing perspective. There is also little capacity in many of these countries to effectively design and implement contingency plans to deal with weather-related emergencies within current government constraints (van Aalst et al, 2013). With limited financial and planned-response capacity, many African governments are unable to act quickly when disaster strikes, exacerbating the impact of the event on already vulnerable citizen populations and potentially further limiting a county's growth trajectory.

Relying on post-disaster humanitarian relief also poses challenges. As currently structured, the international humanitarian system for responding to natural disasters is not as timely or effective as it could be (van Aalst et al 2013). Funding is secured on a largely ad hoc basis after disaster strikes, often coming too late to prevent household losses. Furthermore, international relief organisations tend to operate on small contingency budgets and are therefore likely to be underfunded for any given disaster. Indeed, in December 2015 the total sum of global UN appeals (a proxy for humanitarian need) stood at USD 20 billion of which only about half was funded (Dercon 2015: 17). In addition, there is often a lack of coordination between multiple national and international implementing agencies. In many cases the government can be side-lined in its coordination role due to an imbalance in funding available.

There are also various potential political trade-offs and perverse incentives that may emerge related to international aid. For example, if governments know they can rely on international aid, there may be little incentive to invest in resilience programming or risk reduction/transfer financing (Clarke and Wren-Lewis, 2016). Also, short-term political gain from governments who take decisive action in a post disaster situation can be greater than that gained from longer-term pre-disaster planning. As a result, in the highly emotive post-disaster political climate, any

existing emergency plans may not be followed or they may become negotiation tools that only serve to delay action[5].

3.2 ARC hypothesis

ARC's Key Hypothesis

ARC's mission is to create 'pan-African natural disaster response systems that enable African governments to help meet the needs of people at risk to natural disasters' (ARC 2016). Building on the evidence from other programmes in other regions, the overarching hypothesis is that:

- it is possible to link insurance markets to disaster risk management in Africa in a cost-effective and sustainable way so as to transfer some of the financial impacts of weather risk and natural disasters away from governments and to leverage private finance via insurance markets; and by doing so
- allow governments to better plan, prepare for and, respond more rapidly to extreme weather events and natural disasters; and that
- such a link will protect vulnerable households by providing support to disaster-affected people quickly and in doing so;
- AU countries will continue to grow in spite of shocks and stresses through effective risk management and financing systems

ARC's detailed hypothesis suggests that *if investments are made* in the following areas:

i) Improving government understanding of the of (i) the nature, intensity and likely impact of the climate and weather risks affecting their countries, (ii) the concept of insurance and other financial options for managing disaster risk.
- Designing parametric models for slow and rapid onset events that will trigger the pooled risk insurance mechanism;
- Demonstrating the value of, and preparing AU countries through technical work to join, the weather risk insurance scheme;
- Expanding government capacity to develop national risk management strategies that detail how to prepare and implement contingency plans linked to pay outs on how to respond to weather risk;
- Engaging financial/insurance markets in understanding and accepting the financial risk transferred from African states as a group using the ARC model (e.g. in a way that would not be likely or possible if countries approached the markets individually).

then member states will have measurably greater capacity to plan, prepare and respond to weather-related emergencies – initially for drought, and in the future for floods and tropical cyclones - in a more timely and cost-effective manner. Better planning includes the development of national disaster risk management strategies that include a number of financing options of which ARC insurance is one of many. As a result, member states will be better able to reduce the impact of extreme weather shocks on livelihoods, lives lost, and household asset depletion.

[5] Clarke et al 2014 note that a complicating factor is the distortions to sound decision-making that often are associated with high impact disaster events.

Over time, as a result of the value that ARC demonstrates to its member states, ARC will expand, bringing in additional African countries and extending its coverage, based on demand, to other hazards such as epidemics[6]. Premiums paid by the covered countries will contribute to the capital needed to sustain both the risk transfer institution, ARC Ltd and the capacity building arm, the ARC Agency. It is expected that continued pooling of risk across the continent and diversifying the types of perils insured will result in lower premium rates for coverage and a sustainable and growing interest from reinsurers to provide reinsurance, than if these countries approached the insurance markets individually or on a regional basis. In addition, insurance coverage will allow a country to optimize the budgeted amount of emergency contingency funds and decrease reliance on external aid.

If all these results happen, the long-term impacts are that Africa will have a functional and vibrant pan-African response system that enables African governments to meet the needs of people at risk to natural disasters. ARC will have demonstrated that the returnable capital model of official development assistance (ODA), is a workable innovative finance mechanism that leverages ODA to access financial/insurance market capital to increase development impacts. Ultimately with this new system, AU countries can remain on their development path despite weather shocks and stresses.

3.3 ARC Theory of Change model

The ARC Theory of Change (TOC) maps out the different components of the programme and identifies a conceptual and logical progression of the changes the ARC programme aims to influence if it is to be successful. The TOC is a conceptual model, not a literal representation of a linear process. Its main purpose is to provide the basis for the elaboration of evaluation questions based on the key hypotheses and assumptions.

Figure 1 visually presents ARC's Theory of Change. The TOC breaks down ARC's impact into linked stages. It starts with the problem statement and then outlines the investments and activities that will be made into ARC, namely the establishment and operation of the ARC Agency and ARC Ltd and on-going engagement activities with ARC Member States and others. These investments and activities are expected to produce a set of outputs that include (i) the ARC product offering: innovative financial tools such as ARV as well as tools around contingency planning and (ii) ongoing capacity-building programmes around early warning (EW), contingency planning (CP), disaster risk management (DRM) and disaster risk management finance (DRMF).

These outputs in turn are expected to lead to a series of short-term, intermediate, and longer-term changes. The short-term changes are directly within ARC's sphere of influence and include improving Member State government understanding and technical capacity around EW, CP, DRM, and DRMF, ensuring that contingency plans and ARC insurance contracts are in place, and increasing global awareness of ARC programmes and activates.

If these short-term changes are achieved, the TOC identifies three pathways of change through which the intermediate and longer-term level changes are likely to follow. ARC's TOC posits that, when a disaster hits, the timely insurance pay-out, coupled with the effective implementation of contingency plans (pathway 1), will enable governments to respond in a timely and effective manner, protecting the assets and livelihoods of vulnerable households.

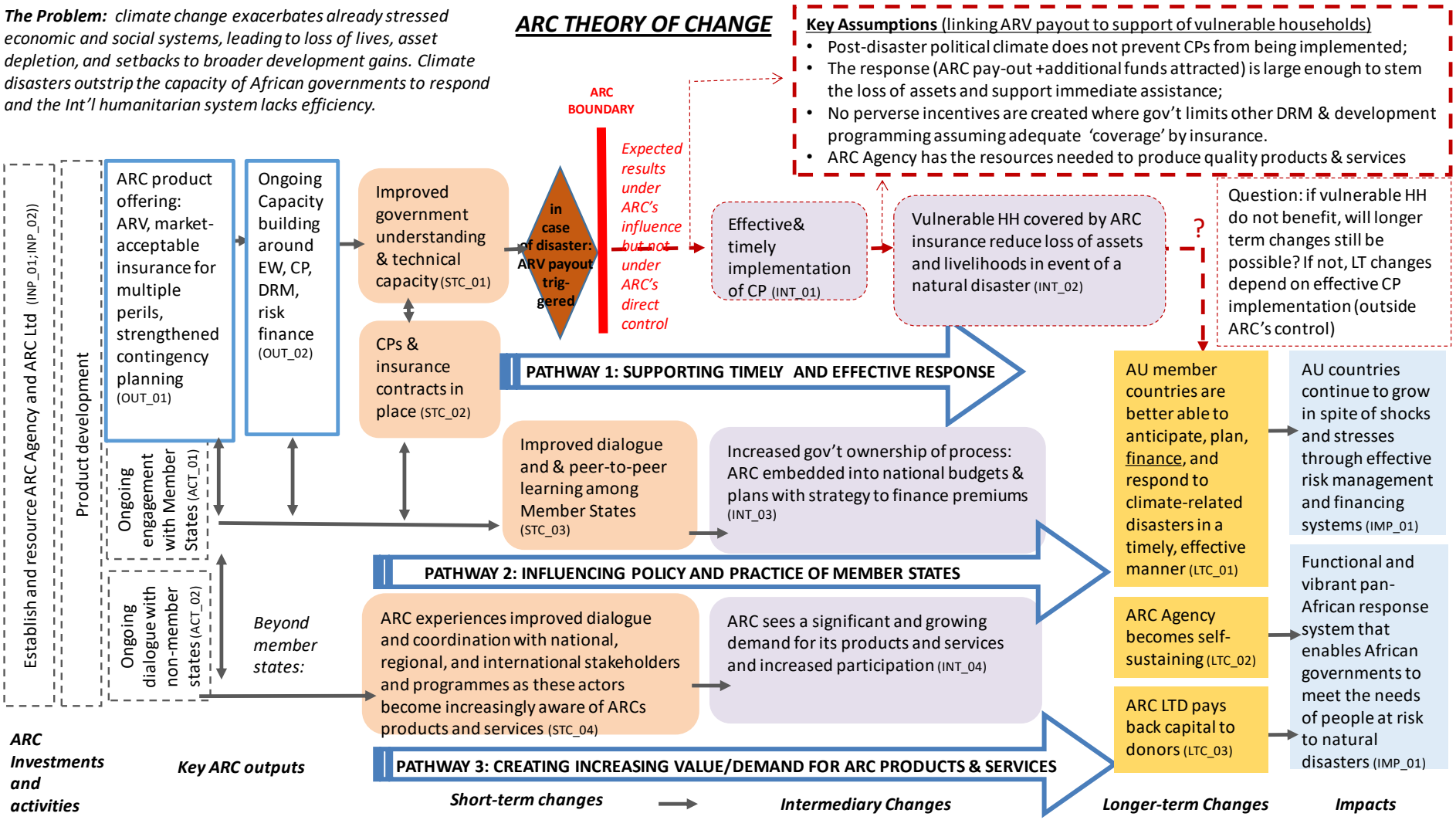
[6] Insurance products around epidemic modelling are outside the scope of this evaluation.

Beyond supporting effective and timely response, a more holistic approach is required to ensure that ARC's other long-term objectives are achieved. Two additional pathways describe these change processes. Here the ARC TOC posits that through dialogue and programing ARC can positively influence the policy and practice of Member States around broader disaster risk management planning (pathway 2). In addition, through dialogue and coordination with the broader community involved in disaster risk management, ARC can create increasing value/demand for ARC products and services among non-member states (pathway 3).

The theory of change analysis then suggests *plausible* links to the long-term changes to which ARC seeks to contribute. These changes relate to Member States being better able to anticipate, plan, finance and respond to weather-related disasters and ARC becoming a financially sustainable organization. If all these results come about, the TOC posits that ARC can achieve its goal of creating a functional and vibrant pan-African response system, enabling governments to better meet the needs of citizens at risk to weather-related disasters and, in doing so, allow AU countries to continue to grow in spite of weather shocks and stresses.

Achievement of each of these pathways rests on a set of assumptions. The key assumptions are found in the diagram, although many more are identified for each link in the causal chain (see Annex A.2 on TOC linkages and assumptions).

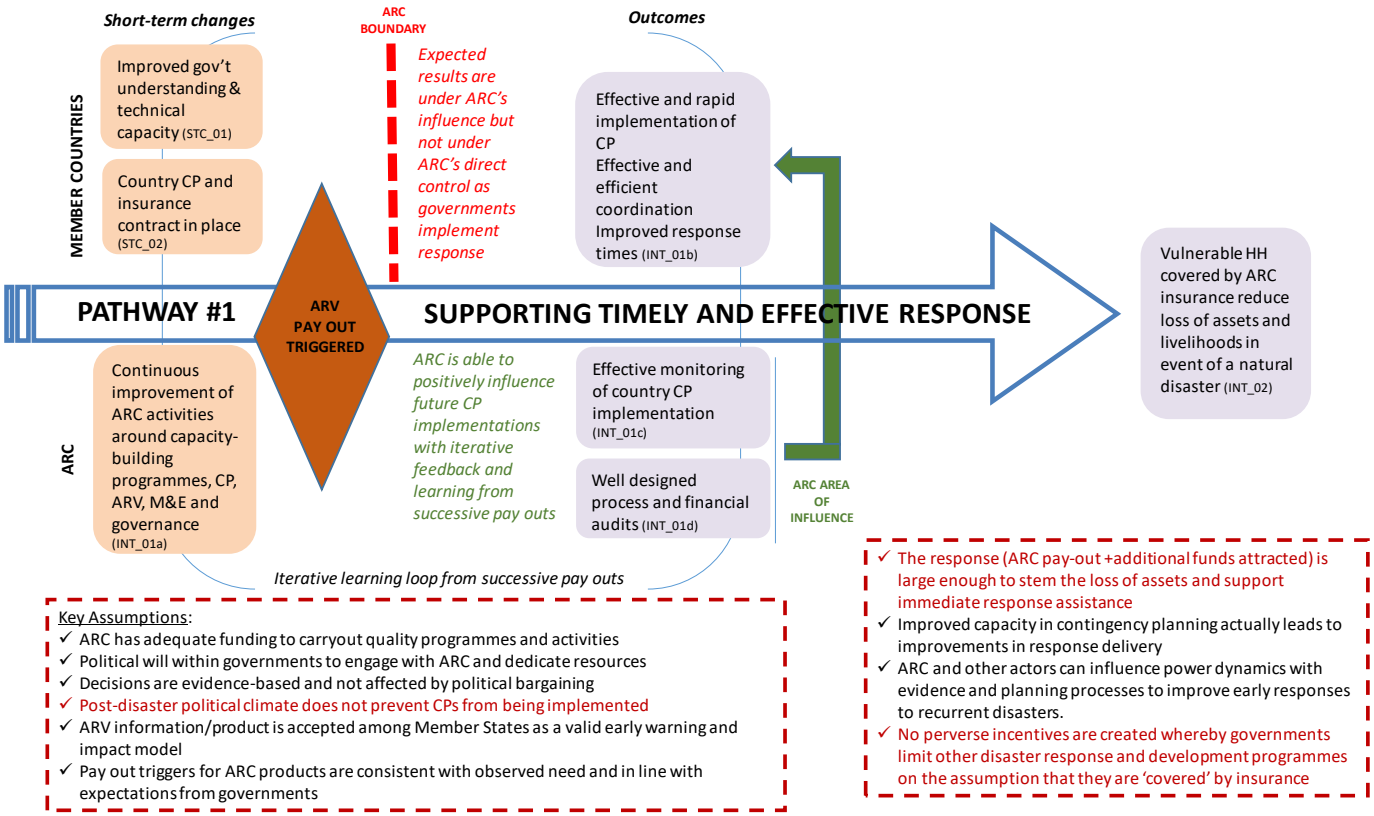
Figure 1: ARC Theory of Change



Below we discuss in more detail the mechanisms of pathway one as we feel they are essential to understanding the Theory of Change. Pathway one describes a key objective of ARC, namely to reduce the loss of assets and livelihoods in vulnerable households affected by weather risk.[7] There is however a gap between what ARC can control (design of contingency plans and pay out to governments) and what is desired (governments, in a timely manner, use pay out to support vulnerable households). We call this divide the 'accountability gap' as ARC cannot be held accountable for reaching an objective that is central to its vision. This gap is visualised in the TOC in several ways: First, the TOC shows the 'ARC boundary line'. Second, the TOC highlights the key assumptions for the successful implementation of contingency plans. Finally, a more detailed description of pathway one (see Figure 2) explains that while any given Member State response to a weather disaster is outside ARC's direct control, ARC is expected to influence successive pay outs through an iterative learning cycle. It is through this learning cycle that ARC expects to narrow the accountability gap.

.

Figure 2: ARC Theory of change pathway one details



[7] ARC vision statement: 'ARC strives to protect the livelihoods of the poor against natural disasters through innovative, cost-effective and sustainable solutions.' ARC Strategic Framework 2016-2020, April 2016, p.5

4 Evaluation questions

The evaluation questions were developed using an iterative and collaborative approach. First, we generated a set of questions based on the Theory of Change and the broader context in which the ARC programme is positioned. This process yielded four high-level questions that link back to the ARC context and to each of the three pathways for change outlined in the TOC.

- **Q1 ARC Contex**t: To what extent does ARC's institutional setup and outputs lead to the adoption and effective use of ARC insurance products? Can this be improved?

- **Q2 Pathway 1:** To what extent has ARC contributed to in-country timely and effective responses that protect affected households' livelihoods and prevent asset loss and food insecurity?

- **Q3 Pathway 2:** To what extent has ARC influenced AU member states' capacity to anticipate, plan, finance and respond to climate related disasters generally, and more specifically in making best use of ARC?

- **Q4 Pathway 3:** Do participating governments and other stakeholders value ARC's risk pool and technical assistance? Why?

Under each of these primary questions are 3-6 summary questions that help inform the higher order question. We mapped each of these summary questions to the OECD Development Assistance Criteria (DAC) to ensure they cover and inform all of the DAC dimensions related to relevance, efficiency, effectiveness, sustainability, and impact[8]. We also provided an indication of when in the evaluation process we might be able to provide some initial evidence to answer the question (e.g. first or second formative or impact assessment). See Table 5, below.

Table 5: ARC Evaluation Questions

#	Question	F1	F2	IM	DAC	ToC Link
ARC Context: outputs and strategies						
1.	**To what extent does ARC's institutional setup and outputs lead to the adoption and effective use of ARC insurance products? Can this be improved?**	**x**	**x**	**x**	**Relevance/ Effectiveness**	***Roll up of summary links[1]***
1.1	How and to what extent do ARCs products/services/activities support on-going engagement and an on-going learning cycle of ARC and ARC Member States within and across countries?	x	x		Relevance/ Effectiveness	ACT_01, OUT_02, OUT_01b, INT_01a, INT_01c, INT_01d
1.2	To what extent does ARC's institutional model (role, governance, financing structure) support the delivery of ARC's outputs?	x	x		Efficiency/ Sustainability	ACT_01, ACT_02, INP_01, INP_02, OUT_01a, OUT_02, STC_03,

[8] The DAC Criteria for Evaluating Development Assistance are the following (available at: http://www.oecd.org/dac/evaluation/49756382.pdf):

1. **Relevance**: the extent to which the aid activity is suited to the priorities and policies of the target group, recipient and donor;
2. **Effectiveness**: a measure of the extent to which an aid activity attains its objectives;
3. **Efficiency**: efficiency measures the outputs – qualitative & quantitative – in relation to the inputs; it is an economic term which signifies that the aid uses the least costly resources possible in order to achieve the desired results;
4. **Impact**: the positive and negative, primary and secondary long term effects produced by a development intervention, directly or indirectly, intended or unintended;
5. **Sustainability**: sustainability is concerned with measuring whether the benefits of an activity are likely to continue after donor funding has been withdrawn.

#	Question	F1	F2	IM	DAC	ToC Link
						STC_04, INT_04, LTC_02
1.3	Are ARC product offerings acceptable for the market? Could such products be offered regardless of donor involvement?	x	x		Relevance / Sustainability	OUT_01a, LTC_02
1.4	How well do ARC's risk models function? Are they improving overtime?		x	X	Effectiveness	OUT_01a, OUT_02
Pathway 1: supporting timely and effective response						
2	**Pathway 1: To what extent has ARC contributed to in-country timely and effective responses that protect affected households' livelihoods and prevent asset loss and food insecurity?**	**x**	**x**	**X**	**Effectiveness/ Impact**	***Roll up of summary links***
2.1	Does the ARC model lead to enough disaster financing for different size slow and rapid onset disasters to make a crucial difference in the livelihoods of households? In what way is ARC's impact limited when other planned sources/mechanisms of financing are not available?		x	X	Impact	INT_02
2.2	Where pay outs have occurred, to what extent have countries implemented contingency plans effectively? What have been the drivers of a successful CP implementation? What have been the barriers to an effective CP implementation?	x	x	X	Effectiveness/ Impact	INT_01b, STC_02
2.3	Is the ARC iterative learning model capturing lessons-learned from various country implementations, leading to future improvements in country response delivery	x	x	X	Effectiveness/ Sustainability	INT_01a
2.4	What evidence is there that pay-outs to governments and the implementation of ARC Contingency Plans has contributed to the protection of livelihoods and food security, and prevented asset loss?	x	x	X	Effectiveness/ impact	INT_02
2.5	Does ARC deliver equally well in both slow and rapid onset situations?		x	X	Effectiveness/ Impact	INT_02
2.6	Is there evidence that links a country's improved DRMF planning to continuous growth?			X	Impact	LTC_01, IMP_01
Pathway two: influencing DRM policy and practice of ARC member states through on-going engagement and capacity building						
3	**To what extent has ARC influenced AU member states' capacity to anticipate, plan, finance and respond to climate related disasters generally, and more specifically in making best use of ARC?**		**x**	**X**	**Effectiveness/ Impact/ Sustainability**	***Roll up of summary links***
3.1	Is there evidence of countries taking action (e.g. creation of broader risk-management platforms, planned budgetary expenditures related to DRM, uptake of insurance or other risk-financing products, etc.) as a result of increased knowledge of DRM and quantified risk? What evidence is there that the change is sustainable?		x	X	Relevance/ Impact/ Sustainability	STC_01, STC_02, INT_01b, INT_03, INT_04, LTC_01, LTC_02
3.2	What combination/network of stakeholders has ARC engaged in the country to support policy and practice change and is this the relevant network for changes to occur?	x	x	X	Relevance/ Effectiveness	INT_03, STC_03, INT_02, STC_04
3.3	Does ARC engagement with Member States lead to tangible commitments by governments in terms of dedicated resources and time?	x	x		Effectiveness/ Sustainability	ACT_01, OUT_02, STC_01, STC_03, INT_01b,
Pathway Three: Increasing Value/Demand around ARC Products and Services						
4.	**Do participating governments and other stakeholders value ARC's risk pool and technical assistance? Why?**	**x**	**x**	**X**	**Relevance**	***Roll up of summary links***
4.1	How relevant is ARC's strategy and role in a country / the region relative to the wider country context and broader DRM architecture?		x	X	Relevance	STC_04
4.2	Are ARC's products and services competitive in the broadening area of DRMF on the continent?		x	X	Effectiveness/ Sustainability	INT_04, STC_02, LTC_03
4.3	Over time, is there evidence of a diverse market of risk financing products available in African Countries? If so, what evidence exists that ARC contributed to this market?			X	Sustainability/ Relevance/ Impact	IMP_02, LTC_03,
4.4	To what extent do member country stakeholders consider ARC as a key actor/partner in supporting effective risk management and risk financing in the country?	x	x	X	Relevance/ Effectiveness	STC_01, STC_02, INT_03,
4.5	To what extent do non-member country stakeholders consider ARC as a key actor/partner supporting effective risk management and risk financing in the country?	x	x	X	Relevance/ Effectiveness	ACT_02,
4.6	What is the nature of the link between ARC pay-outs, successful CP implementation and governments' motivation to engage with ARC?		x	X	Relevance/ Sustainability	INT_02

1. Includes all the links from the summary questions that fall underneath the key question.

Next we transformed all the assumptions (found in Annex A2) into additional questions which we in turn linked back to one of the summary questions to yield a third level of detailed questions. In this way we could ensure that all linkages in the theory of change are covered by a question. The mapping exercise was an extensive process that involved moving, adding, and merging various questions at various levels. The full output of this exercise with all the detailed questions can be found in Annex A.3. An example is shown below in Figure 3 to demonstrate the process of how each detailed question links to one or more levels of the ToC. Per Figure 3, summary question 1.1 '*How and to what extent do ARCs' products/services/activities support on-going engagement and an on-going learning cycle within and across countries?*" has seven detailed questions that are asked at different levels of the theory of change. In this way we can see how the detailed questions show the progression along the pathway to change, leading from inputs and activities to outputs and short-term changes to intermediary and long-term changes, and ultimately to impacts.

Figure 3: Example of detail question mapping to TOC

#	Question	Link to ToC
1. To what extent does ARC's institutional setup and outputs lead to the adoption and effective use of ARC insurance products? Can this be improved?		
1.1	How and to what extent do ARC's products/services/activities support on-going engagement and an on-going learning cycle within and across countries?	ACT_01, OUT_01b, OUT_02, INT_01a, INT_01c, INT_01d
1.1.1	How has ARC engaged with member states to define realistic insurance parameters such as attachment points, limits, and retentions?	ACT_01
1.1.2	How has ARC engaged with member partners on CP and DRMF? To what extent is there coordination? To what extent are these partnerships working?	ACT_01, OUT_02
1.1.3	How is the CP process including the peer review and approval process of both the CP and FIP evolving overtime? Is there evidence that lessons from country implementation is fed back into successive iterations?	OUT_01b
1.1.4	How do governments perceive the CP process? Do they find it useful? What might be improved? Do countries use the CPs during implementation?	OUT_01b
1.1.5	How is ARC's learning from successive iterations of pay-outs fed back into its products & programmes (e.g. CP planning, risk models, training materials, M&E processes, etc.?)	INT_01a
1.1.6	Are ARCs activities around monitoring country response delivery of ARC pay-outs effective? What works? What doesn't? Are lessons being fed back into ARC programmes and activities for successive iterations?	INT_01c, INT_01d
1.1.7	How does ARC address gendered differences? Is this approach successful in addressing gendered differences in capacity and learning?	

The first two levels of questions were shared with key stakeholders within ARC and with collaborating partners and the feedback incorporated into this final design. Note: Some of the questions that were originally identified in the shared version as summary questions have become detail questions and can be found in Annex A.3. We feel that the third level of questions is more for internal use to ensure we cover the broad scope of the evaluation and to help us design the appropriate tools for data collection. However, findings in relation to the detailed questions will also be included in the final reports, as part of presenting findings on the headline questions.

It should also be noted that gender is addressed in two ways in the approach to this evaluation. First, when looking at the ARC institutional setup detailed question 1.1.7 (see Annex A.3) asks "*How does ARC address gendered differences? Is this approach successful in addressing gendered differences in capacity and learning*?" Second, in our methods section, the participatory approaches we suggest, in particular at the local level, are carefully designed to be sensitive to issues of gender and inclusion.

5 Evaluation methodology

5.1 Design Considerations

Two key areas of consideration informed our initial design of the ARC programme evaluation: the characteristics of the *context* in which the programme is implemented; and the *programme design* features that respond to these contextual characteristics. Addressing these two elements determine the most realistic, appropriate, and effective way to evaluate a programme.

Context

The contextual elements around ARC are discussed in detail in the DFID Business Case and are elaborated in the ARC Strategic Framework and ARC Theory of Change documents. For our purposes, we summarise the main elements:

- Evidence suggests that responding quickly to a disaster protects vulnerable household livelihood and assets, allowing them to be more resilient to weather shocks (DFID, 2013: 2, 7);
- African governments struggle to respond quickly to disaster (ARC Theory of Change, Oct 2016: 2);
- The commercial insurance landscape is unconducive to affordable and effective natural risk management as it is costly and data/models scarce (DFID 2013: 7); and the
- Global humanitarian response to crises is typically slow, expensive and outside the critical timeframe for ensuring longer term resilience (DFID 2013: 7).

This context presents a number of potentially non-linear and interrelated facets all of which combine into a complex set of unique characteristics. First, at the local level, where the impact of a drought is most evident, the effects of the drought are uneven (i.e. the effects might be more severe for some constituencies than for others), and the response that a particular country provides might also vary (e.g. a variety of intervention-types implemented by a variety of partners, including local and international agencies, all to varying degrees of effectiveness). At a national government level, the management of a disaster may vary significantly depending on the strength of government disaster planning mechanisms, the flexibility of established distribution channels for implementing interventions, potential political influences on the allocation and distribution of funds for pay-outs (e.g. the prioritisation of particularly constituencies over others), and a wide variety of other institutional factors that might influence the management of a disaster. Finally, at a global level the context is also highly variable. The institutional framework through which the UN and related humanitarian response mechanisms operate is contingent, controversial, political and often slow to react. Furthermore, the insurance market consists of another set of stakeholders with different objectives and goals that must be addressed.

Programme design

African Risk Capacity (ARC), which focuses on capacity building and the provision of early warning, disaster risk management, and risk finance options (namely pooled risk insurance), was established to fill a clear gap and address some of the challenges posed by weather-related disaster risk in Africa. While a more detailed description of ARC can be found in Section 1.1 of this report, here we concern ourselves with the characteristics of ARC that most concern the evaluation design. In this case, these are:

The multiple levels of interrelated relationships within which ARC operates. To be successful, ARC must interact with multiple ministries[9] in multiple countries, all of which have unique contextual, social, political, and economic challenges and competing priorities. ARC must also engage regularly with the broader humanitarian response landscape – most of which are large organisations or multilateral agencies influenced by multiple factors. Furthermore, ARC must engage with the broader insurance and capital markets. And finally, while ARC does not have direct engagement at the local level, their mission is to help protect the livelihoods of beneficiary households.

How to disentangle the contribution of ARC from other interventions. ARC insurance in not intended to cover the full cost of responding to a weather disaster, but rather to arrive early so as to protect vulnerable assets and livelihoods. It is therefore anticipated that other forms of relief will also come into these same communities later or in parallel, perhaps using the same distribution channels (e.g. via an existing social safety net mechanism). As such, it will be a challenge to clearly separate the ARC impact from that of other assistance.

How to measure ARC success at the local level. A key interest to DFID and other donors is the impact at the household level, yet this impact is outside the direct control of ARC since governments control the implementation of relief (see Theory of Change, Section 3). This creates a challenge as to how to best measure ARC's impact.

All of these characteristics of both the context and the programme of ARC itself represent core characteristics of a broader discourse on *complexity* and more specifically, how complexity provides a heuristic framework for understanding development programmes in challenging and emergent environments. This has proven particularly important for development programme evaluations, where an awareness of *contextua*l complexity has introduced a level of *programmatic* complexity that traditional evaluation paradigms have not been able to adequately address (see Weiss 1997, Stame 2004, Treasury Board of Canada Secretariat 2012).

Evaluation design options

Much of the traditional and mainstream evaluation discourse over the years has been dominated by what Scriven (2016: 29) calls the 'long period of quantitative idolatry' – the use of 'positivist'[10] or experimental evaluation design paradigms as the 'gold standard' for evaluating a programme's impact (see Stame 2004, the Treasury Board of Canada 2012, White 2009, and Munter *et al* 2016). Experimental evaluation designs are powerful tools for measuring the expected results of an intervention, and for providing a 'causal link between an intervention and observed results', generally called attribution (Treasury Board of Canada, 2012). However, while attribution is desirable, establishing it is rare.

The true determination of attribution requires an experimental or quasi-experimental evaluation design, which are highly specific in context and programme design with little to no flexibility to allow for change[11]. Moreover, the use of experimental designs comes up against two main challenges: the first is context and programme implementation, much of which relates to the discussion above

[9] Coordination is handled through a government coordinator who acts as a point of contact.

[10] A positivist research paradigm is rooted in science and employs the scientific method, which believes in underlying universal laws and involves a process of collecting data, forming a hypothesis to test a theory, and then testing that hypothesis through more data collection.

[11] Experimental and quasi-experimental evaluation designs are most appropriate under fairly specific contexts, such as when an intervention involves a new approach or where the goal of the evaluation is to test whether the intervention 'works' (Treasury Board of Canada, 2012: 1). Most importantly, experimental evaluations require specific circumstances in which a counterfactual can be identified and used to measure the results of an intervention's 'treatment' group (beneficiaries) against. The difference between the two groups effectively provides the programme's measure of impact. In order for an experimental evaluation design to be effective and credible, both the control and treatment groups need to be very similar (comparable) in their characteristics, and the application of the 'treatment' needs to be random, to control for selection bias. Often, given various contexts and complexities of interventions – particularly those noted for ARC and the descriptions of complexity described above – a counterfactual is simply not possible.

on complexity; and the second is establishing an understanding beyond simply whether impact has been achieved, but *how* and *why* it was successful – the proverbial 'black box' of development evaluation (Munter *et al*, 2016). As Munter *et a*l (2016: 9) state, 'the effort to extend causal claims from merely demonstrating *that* A caused B to explaining *why* A caused B affords opportunities to refine theory and improve interventions'.

With the learning and adaptation objectives in mind for the formative evaluations for the ARC programme (DFID ToR; Scriven 1967), the limited available evidence to date about the effectiveness of risk insurance mechanisms, and the challenges for implementing an experimental evaluation design for a complex programme such as ARC, a theory-based approach provides the most rigorous, thorough, and appropriate model for designing an evaluation in this context and with these objectives (McKinnon & Hole, 2015, Stern et al., 2012).

5.2 Evaluation approach: contribution analysis

A theory of change evaluation approach – within a theory-based framework – involves developing a TOC for the intervention and showing how it is meant to work, including explicit articulations of the theory's assumptions. It is normally based on the views of a range of relevant stakeholders. The evaluation then aims to assess to what degree the theory matches with what is observed (i.e. are the desired outcomes and impact achieved, and are the pathways assumed to lead to these outcomes in fact doing so). This provides for opportunities for adaptive management of the programme, should alterations be necessary, or key assumptions be shown to be faulty and require programmatic adjustments, or shown to be accurate and effective, and perhaps warrant amplification (see McKinnon & Hole, 2015).

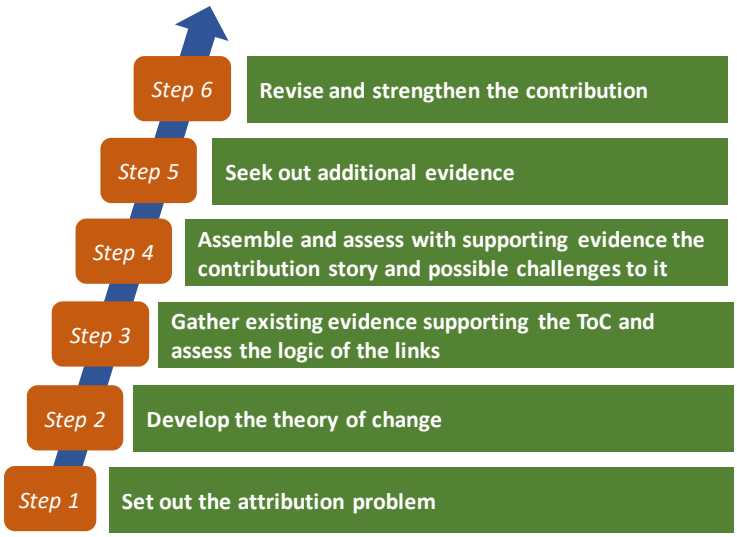
Contribution Analysis (CA) is an approach that fits well with theory-based evaluations, and as such is our selected approach for the ARC evaluation. CA is a theory-based approach that posits 'if an evaluator can validate a theory of change with empirical evidence and account for major external influencing factors, then it is reasonable to conclude that the intervention has made a difference' (Treasury Board of Canada, 2012: 9). To conduct a CA analysis a theory of change is examined in depth from inputs to outcomes to establish a plausible contribution. A plausible contribution can be made if the following criteria are met (White and Phillips, 2012, Mayne 2008):

- The programme is based on a reasoned TOC – the results chain and the underlying assumptions of why the intervention is expected to work are sound, plausible and agreed to by key stakeholders;
- The activities of the intervention are implemented;
- The expected chain of results identified in the TOC occurred, the assumptions held, and the outcomes were observed;
- External factors (context) influencing the intervention were assessed and shown not to have made a significant contribution, or if they did, their relative contribution was recognised.

The structured but flexible nature of this type of analysis lends itself to the complexities and uncertainties inherent in the ARC programme.

Typically, the 'impact statement' of a contribution analysis approach emerges through the creation of a 'contribution story' rather than the result of a measured 'impact'. Conducting a Contribution Analysis consists of six steps as outlined by Mayne (2008) and visually depicted in Figure 4.

Figure 4: Steps of a contribution analysis



Step 1, *setting out the attribution problem*, is clearly articulated in the evaluation TOR. Step 2, *develop a theory of change*, was completed during the Inception Phase and is summarised in section 3. The implementation phase of the evaluation will therefore focus on steps 3 to 6. By the end of Phase 1 of the formative evaluation we expect to have completed Step 3, *gathering existing evidence*, and Step 4, *assemble the initial contribution story*. As we assemble the contribution story we will follow the same processes and use the same data tools across the countries to test and triangulate claims with multiples sources of evidence.

Phase 2 of the formative evaluation will give us the opportunity to conduct Steps 5 and 6. It should be noted however that not all stages of the theory of change will be covered in the formative evaluations as we expect longer-term results and impacts to emerge only after an extended period of programme operation. These latter stages will be captured in the impact evaluations where we will be able to repeat steps 3-6 with more data covering a longer period of time.

We expect to collect robust evidence for the contribution story through our country case studies using document reviews, key informant interviews, and focus groups. However, we also note a few risks. Although CA is highly suitable for this type of evaluation, we anticipate some challenges to the CA analytical framework:

- **CA is not able to validate mechanisms in the TOC that have not yet happened**. However, timeframe poses a challenge to any analytical method. Furthermore, given the long timescale of this evaluation, we do not expect timeframe to be an issue. By the time of the impact evaluation we expect to see results at all stages of the TOC. However, Mayne (2008) suggests that in such a scenario, one can presume the theory of change will play out as expected by checking to see whether the programme is presently on the correct path, and by drawing on evidence from similar interventions that have demonstrated achievement of desired outcomes and impacts under similar circumstances.

- The country case study sample is limited (as all case study based evaluations are inherently limited). Due to budgetary and feasibility constraints, our country case study sample is limited (12 in total) and therefore may not represent the contribution story at large. To mitigate this risk, we will use purposeful sampling of countries for each phase of

the evaluation (see section 5.10). In this way we can seek out countries that might add alternative explanations or challenges to the contribution story.

- Finding evidence for alternative explanations and triangulating this with the evidence collected for the contribution story can be challenging. A key strength of CA is early identification and tracking of alternative explanations. One criticism of CA can be that tools are tailored to seek ARC contribution, but we will carefully design tools and select data points to ensure that the process also lends itself to identifying, tracking and measuring the contribution of alternative explanations. There are several ways in which this occurs. Firstly, the design of the ToC, and therefore the research questions, explicitly state assumptions. If these assumptions do not hold then an alternative explanation will emerge. For example, one of the key ToC assumptions is that the carefully designed ARC CPs will be implemented, despite post-disaster political machinations. If this assumption does not hold across countries, an alternative explanation – that what is needed for success (maybe in addition to ARC or in replacement of ARC) is some type of politically-focused solution – will emerge. Turning the ToC assumptions into questions (see detailed evaluation questions in Annex 3) allows us to explicitly ask about these alternative explanations. Secondly, as part of our implementation process, we will identify up front the key known alternative explanations and track them over time (see section 5.8). The literature review has informed our initial thinking in this area. Thirdly, we will specifically ask KIs about alternative explanations, and will purposively select KIs to include individuals with non-ARC related views and experiences (for example, individuals from programmes implementing other DRM capacity building initiatives). These independent 'expert' key informants will be able to test the explanations and claims of contribution by ARC. Fourthly, at the household level, informants may well not be specifically familiar with ARC, or recognise a payout as having come from them, and so will not be biased to discuss the programme or incentivised to suggest a contribution. Fifthly, we will consider conducting a whole case study in a non-ARC country, which means that the entire case study would be focused on identifying alternative explanations for progress. Sixthly, in the case study countries we will develop activity timelines (see section 5.5) which will focus on both ARC and non-ARC interventions in order to help us understand alternative contributing factors.

- **It is difficult to define the 'net impact' of the programme in light of other influencing actors**. In the case of ARC, social protection systems and humanitarian responses will be on-going alongside ARC-related interventions. It may therefore be difficult to isolate the results of ARC funding from those of other funding sources through contribution analysis alone. Even using a quasi-experimental impact evaluation, untangling the direct contribution of ARC would be a challenge. However, the dynamic nature of CA, which involves careful and methodical triangulation of data, lends itself to answering this question. For example, households will likely not know of 'ARC' as the source of their benefit. Rather they will know that assistance in various forms arrived in time to be helpful. Since CA will triangulate data from KIs in the community, programme implementers, funders and others, there is ample opportunity to measure the contribution claims of ARC versus other explanations (again, see section 5.8). This will be explored further in the design of the impact evaluation pilot, identifying more complimentary evaluation methods as necessary.

Despite these challenges, overall, CA remains the most appropriate analytical framework given the complexity and variability of the ARC initiative. While CA cannot give a definitive statement about 'attribution', CA does specifically seek to highlight other factors that might contribute to the observed outcomes, making the context of contribution from ARC clear and explicit. Indeed, the contribution stories provide a rich explanation of the different mechanisms through which change occurs.

Our broader evaluation framework makes use of mixed methods alongside a set of varying data collection tools, including scope to collect both quantitative and qualitative data. This combination of methods and tools provides a rigorous and robust evaluation framework that will be able to provide substantive insights across the complex levels that constitute the ARC programme.

5.3 Overall evaluation design

In response to the above design considerations, our proposed design framework identifies three workstreams falling under the theory-based paradigm, and one workstream based on an experimental design:

- Workstream 1: Country case studies
- Workstream 2: ARC organisational review
- Workstream 3: Global review
- Workstream 4: Quantitative household survey (optional extra based on experimental design)

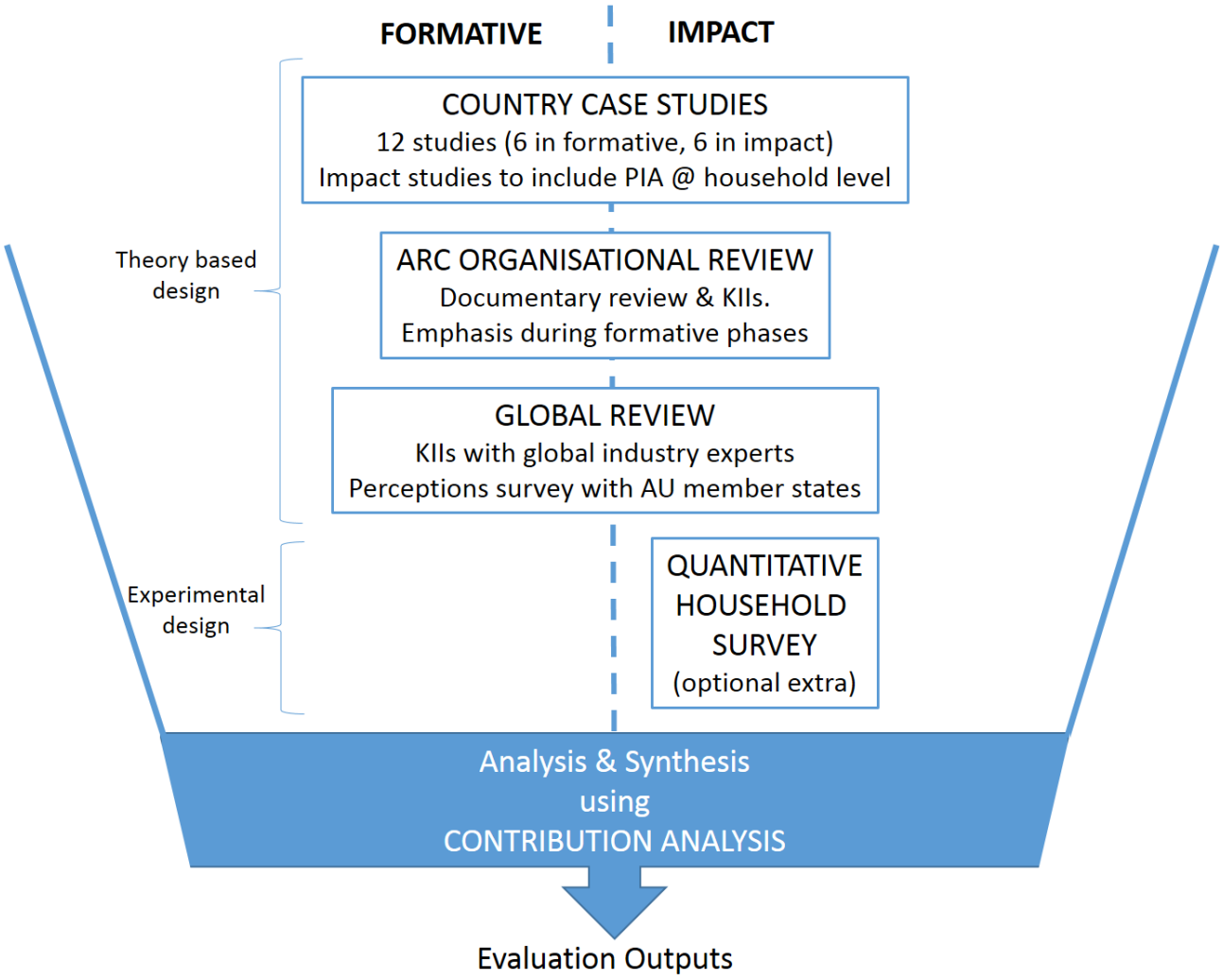
Workstreams 1 and 2 will consider all the ToC pathways. Workstream 3 will also address all pathways, but will particularly focus on pathway 3 (creating demand around ARC products). Workstream 4 will focus entirely on pathway 1 (timely and effective response, including impact for beneficiary households).
Sections 5.3.1 to 5.7 below present greater detail on each of these workstreams. The Evaluation team has created a Methods matrix which maps all the detailed questions against workstreams and phases of the evaluation.[12] This information is summarised in Table 13 as part of the Evaluation Framework.

Workstream 4 relates to the reference in the TOR to additional funds potentially available to conduct a stand-alone impact evaluation that would use a quasi-experimental design to answer a very specific component of the TOC: whether vulnerable households covered by ARC insurance are better protected (as measured by loss of assets and livelihoods) than those not covered by ARC insurance (pathway #1 in the TOC). See Section 7 for detail. If this option is undertaken it will operate under an experimental paradigm, and will make use of a different analytical framework based on quantitative methods and analysis. We should be clear that our theory-based approach will also answer this question but through a different approach, that is, a participatory impact assessment approach (see Annex A4). We highly recommend that DFID should commission this additional impact evaluation work as we believe it would add greater robustness to the evaluation findings.

The figure below sets out the overall evaluation design in diagrammatic form, showing how the different workstreams relate to the different evaluation phases, and how all contribute data to the cross-country, cross-workstream analysis and synthesis process.

Figure 5: ARC Evaluation Design Framework

[12] This is a very lengthy document created for internal team use and so has not been included here. However, the team would be happy to provide it if required.



5.3.1 Sampling

While discussed in more detail within each workstream section, an overview of the sampling strategy for the different phases of the evaluation and for the different workstream tools is provided in table 6. It should be noted that qualitative research sampling differs fundamentally from that of quantitative research sampling insofar as they pursue different objectives: while the latter often seeks a representation of a large population and wants to know how often something occurs within the population, the former is more interested in understanding why or how something occurs, which does not require the same quantifiable representation of the population being studied. Qualitative research sample sizes are therefore not set in fixed rules for determining rigour. Rather, the sample size is determined more by what is wanting to be understood, or what data are being sought in the research (see Varkevisser et al 2003:207). Thus, it is commonly accepted that a sample size can be 'estimated', rather than determined, at the outset of a study, based on possible strata or clusters, and then potentially adapted during the course of data collection. For a more detailed discussion of our sampling strategy see Annex 10: Sampling methodology.

For our purposes here, practical constraints include the time and scope limitations that necessarily contain a case study, as well as budgetary limitations to the amount of data that can be collected and analysed for each evaluation phase. With these considerations in mind, in the table below, we suggest an indicative sample size with an underlying flexibility to adjust as the need arises. The flexibility of the sample size will depend on contextual variations that will be important to consider and integrate into the sampling design, such as different sizes for a large pay-out, broad

geographical coverage, and multiple interventions within a country, versus a small pay-out, small geographical coverage and a single intervention within a country.

Table 6: Estimated sampling framework by evaluation phase

Evaluation phase	Perception surveys	Expert interviews	Number of case studies	KIIs per case study	Local level FGD per case study
Formative 1	20	10	3	30 per case, 90 in total	
Formative 2	20	10	3	30 per case, 90 in total	
Impact 1	20	10	3	30 per case, 90 in total	Ave 10 per case study, 30 in total
Impact 2	20	10	3	30 per case, 90 in total	Ave 10 per case study, 30 in total
Total	80	40	12	480 total	approx. 600-720 individuals

In terms of geographic coverage, we will purposively select a FG sample from the intervention areas. Obviously, broad geographic coverage is one of the aims of our sampling approach but without knowing the size of the intervention areas it is not possible to give exact details at this stage, except to say that geographic coverage with be one of several considerations taken into account when selecting the samples. Please see Annex 10 on Sampling Methodology for futher discussion on this topic.

5.4 Baselines

For a CA approach to work well it helps to start with a clear baseline. The activities of the programme can then be assessed with reference to differences between the baseline and endline, based on the TOC. However, in the case of the ARC evaluation, establishing a comprehensive baseline is not straightforward because:

- There is uncertainty around which countries will join, remain or leave ARC over the evaluation period, making it difficult to select case study countries with confidence;
- ARC has been operational for several years and so historical data may need to be used in some cases;
- The evaluation has been divided into formative and impact phases at set times, which does not fit easily around a baseline / endline approach;

To overcome these issues, we have devised a number of ways to incorporate baselines across the workstreams:

Workstream 1: Country case studies - In both the formative and the impact phases of the evaluation we will consider repeating one or two country studies to assess country progress over time. We may consider more, if feasible and if there is a strong case for doing so. Phase 1 would therefore represent the baseline, and Phase 2 the endline in those countries. Not all the case studies will be revisited in order to allow a broader sample of countries which, through cross-country analysis, will help us to gain an appreciation of ARC's impact across the continent. However, within the 'one-off' case studies we will make use of historical analysis so that we can

construct a robust understanding of changes over time in the country and use this to assist the contribution analysis. For more detail on this see Section 5.5.

Workstream 2: ARC Organisational Review – For this element of the evaluation it is easy to construct a baseline and endline, as the object of enquiry (ARC agency) remains constant. The Organisational Review in Phase 1 of the Formative Evaluation will operate as a baseline. This will then be repeated 18 months later as part of Phase 2 of the Formative evaluation, serving as a first midline, and then again in 2020 as a second midline. The endline will be conducted in 2024 as part of Phase 2 of the Impact Assessment.

Workstream 3: Global Review – The main elements of the Global Review are KIIs with industry experts and a Perceptions Survey with a broad range of stakeholders from ARC-member and non-member states. These will be undertaken as baselines in Formative Phase 1, midlines in Formative Phase 2 and Impact Phase 1, and endline in Impact Phase 2. In addition, we propose to conduct a rapid continent-wide Baseline Context Assessment. This will serve to provide a broad-brush indication of risk and response capacity within governments across the continent, using a standard set of indicators and covering a range of countries. As part of the analysis at each later stage of the evaluation (Formative Phase 2, Impact Phase 1 and Impact Phase 2) we will update the context assessment, with a view to identifying whether there has been a broad shift either regionally or continent-wide. Our contribution analysis will then be able to identify the role that ARC may have played in any changes.

Workstream 4: Quantitative Household Survey – By definition a quasi-experimental design must have a baseline. We discuss the options for establishing a baseline in Section 7.

5.5 Workstream 1: Country case studies

Given the complexity of ARC and the multiple levels of analysis, a case study approach provides an efficient way to gather evidence and structure the analysis. A large part of the ARC evaluation data will come from the country case studies, spread across the formative and impact phases of the evaluation. By making use of a variety of data types and data sources, a case study approach enables evaluators to assess programmes in a way that a large quantitative survey cannot by allowing one to 'study intensely one set (or unit) of something – programmes, cities, countries, worksites – as a distinct whole' (Balbach, 1999: 3). For ARC, case studies will provide an opportunity to conduct in-depth interviews a wide range of stakeholders about their specific experiences.

The country case studies will focus on national and local stakeholders within ARC member countries. Within a country case study, we will use a suite of tools to assess the activities, processes and views at both national and local levels, but mainly focused on key informant interviews. We will conduct three (3) detailed country case studies in each of the formative phases of evaluations and three (3) in each of the impact phases of the evaluation. This gives a total of twelve case studies. In some instances, we might look at the same country more than once so as to assess change; in other instances, we will review different countries selected purposefully based on specific experiences and contextual changes that have occurred that would make a particular country of interest of a case study. For a review of our proposed country selection, see section 5.10.

We intend to select a wide range of countries for case study analysis, including ones who have recently joined the risk pool, been longstanding members, have experienced pay-outs, have left or never joined the risk pool. For this reason, the exact content of each case study will vary, and we will tailor our tools and approach to the specific history and context of the country. For example, in some countries we might focus much attention on the implementation of Contingency Plans following a pay out; but for a country that never joined

the risk pool, we might choose instead to focus on the capacity building element of ARC's activities in country. Other example focal areas could be:

- Performance of ARV in the country
- Different response mechanisms across different countries
- Process of attracting a new country to ARC
- Policy decisions around dropping out of ARC

We expect to conduct the Impact phase case studies shortly after implementation of an ARC payout has been completed, meaning the intervention will still be easily recalled by interviewees. See Table 11 for more information on countries under consideration as case studies and potential focal areas. Overall, we will ensure that sufficient evidence is collected to validate each stage of the TOC.

For the case studies we will conduct a range of activities including:

- Conduct a Stakeholder Mapping analysis – this will be an initial step to identify the key actors with whom we need to collect information. It may include ARC representatives working with / in the country, government officials, civil society representatives, donor agencies, humanitarian agencies, representatives from social protection programmes, ARC beneficiaries at all levels including citizens, academics and independent consultants;
- Create a timeline of ARC's activities and engagement with the country and verify it with ARC and other stakeholders;
- Conduct an institutional review of the DRM institutional infrastructure, including tracking the development of key policies and relevant institutional processes over time, resulting in a timeline of DRM institutional development;
- Map the ARC timeline with the DRM institutional development timeline and investigate any apparent links;
- Review the Contingency Planning process, actual Contingency Plans (historical and current) and their historical implementation if relevant;
- Review budgets and national development plans;
- Interview key stakeholders to gather views on the quality of ARC activities, barriers and drivers;
- Conduct FGDs at the household level and use Participatory Impact Assessment (PIA) to investigate the impact of pay outs (Impact phases only). See Annex A.4.

As mentioned above, different evaluation questions will be relevant for different country case studies, however, the table below provides some examples of the proposed activities, how they will map on to selected evaluation questions, and what evidence may be found to validate the Theory of Change:

Table 7: Formative case study activities and possible evidence

Evaluation question	Relevant evaluation activities & who will be engaged to answer qns..	Examples of evidence to validate the TOC
Is there evidence of countries taking action (e.g. creation of broader risk-management	Review of National Government policies, plans and budgets.	Specific improvements to institutional frameworks and DRM policies leading

platforms, planned budgetary expenditures related to DRM, uptake of insurance or other risk-financing products, etc.) as a result of increased knowledge of DRM and quantified risk? What evidence is there that the change is sustainable? (3.1)	Creation of DRM Institutional Development timeline. Review of Contingency Planning process, comparison of current and historical CPs	to an overall positive change matching the timeline of ARC's involvement. KIIs validating link rather than citing other causes. Clear technical improvements in CPs and greater government involvement in drafting. Premiums included in budgets for future years, regular and increasing mentions of ARC and DRMF in National Plans.
Where pay outs have occurred, to what extent have countries implemented contingency plans effectively? What have been the drivers of a successful CP implementation? What have been the barriers to an effective CP implementation? (2.2)	Historical review of CP implementation (comparison of plan with actual events). KIIs to identify barriers and drivers of implementation.	Significant overlap between CP and actual events. If little overlap then evidence of learning (subsequent CPs take better account of barriers and drivers).
What combination/network of stakeholders has ARC engaged in the country to support policy and practice change and is this the relevant network for changes to occur? (3.2)	Stakeholder Mapping, creation of ARC Activity Timeline, review of participation records for ARC activities	Significant overlap between our stakeholder mapping and the stakeholders mentioned in the ARC Activity Timeline / those participating in ARC activities
Does ARC engagement with Member States lead to tangible commitments by governments in terms of dedicated resources and time? (3.3)	Review of National Government operational plans, budgets, staffing records. Review of participation records for ARC activities.	General increase in budgets and staff resources made available for DRM and DRMF, increase in the number and seniority of government staff engaging in ARC activities.
How relevant is ARCs strategy and role in a country relative to the wider country context and broader DRM architecture? (4.1)	KIIs with donor agencies, civil society, humanitarian organisations and government officials. Review of ARC country strategy. DRM institutional review, ARC activity timeline. Stakeholder mapping.	ARC strategy and activities timeline match closely the needs identified by KIs and gaps identified in the institutional review. Evidence of joint working with other stakeholders.
To what extent do member and non-member country stakeholders consider ARC as a key actor/partner supporting effective risk management and	KIIs with government officials, ARC Activities Timeline, review of participation records for ARC activities	Increasing participation in ARC activities, KIs show consistent / increasing belief that ARC is a key partner.

risk financing in the country? (4.4 and 4.5)		

Where possible, to make sense of the multitude of data from multiple countries that will inform this analysis, we will use rubrics so that different evaluators use the same parameters to assess the quality or strength of the findings. These rubrics will be developed in conjunction with the tools.

5.6 Workstream 2: ARC Organisational Review

The ARC Organisational Review focuses on ARC stakeholders, in particular ARC staff and the governance structures related to the ARC Agency and ARC Ltd. Many of the evaluation questions relate specifically to the processes, procedures, capacities and resources of ARC Agency. As such it will be necessary to collect information, documents, and views directly from the ARC Agency staff and related governance institutions. Since ARC Agency performance appears at all stages of the TOC, in each phase of the formative and again in each phase of the impact assessment we will spend time onsite collecting, verifying, and triangulating data. The first visit will serve as a baseline from which we can build evidence to support the learning cycle of the organisation which is a critical component of the ARC model and Theory of Change. The analysis will focus on several dimensions, such as (i) ARC risk models (ii) contingency planning processes; (iii) ARC M&E processes; (iv) ARC governance structures; (vi) ARC capacity now and in future; and (vii) ARC training materials, to name a few.

The organisational review is simply a systematic method of assessing the value, consistency and strength of the processes, activities and capacities that underpin the ARC organisation. A core component of the ARC theory of change rests on the ability of the ARC Agency to educate Member State governments on the value and importance of preparing for disaster, train them on the tools needed to understand, assess, and finance risk, and ultimately to encourage these governments uptake ARC products. As such, the organisation's ability to provide these outputs is a key component of the evaluation. Indeed, the TOR emphasises the need to review ARC organisational structure, operations, and capacity and provide feedback for learning and improvement, thus the reason for a two phase formative evaluation.

The approach for the organisational review will mainly focus on documentary review and key informant interviews. The output from this exercise will be a detailed account of ARC operations and activities that points out what works and where improvements can be made, ultimately forming a set of recommendations. While the data will be organised by dimension for ease of reporting, the data will be mapped back to key evaluation questions related to ARC capacity and institutional setup, thus supporting the contribution analysis in the overall assessment of the theory of change.

A draft set of dimensions for reporting purposes are listed below.

- Contingency planning approval process
- Outreach programme to countries
- Outreach to other stakeholders
- Capacity-training programme/workshops
- ARC pooled risk models (initially ARV)
- ARC Agency staff capacity
- ARC Ltd staff capacity
- Governance structures
- ARC budgeting and financial reporting processes
- ARC monitoring and evaluation programme

We will collect data on each dimension from multiple sources and record it in a database to facilitate triangulation and analysis. For each dimension we will collect information on strengths and weaknesses of different sources, and record any gaps in data. In the analysis phase we can triangulate these data to get a broad perspective.
During the first phase of the formative evaluation, we will assess the quality and robustness of ARC's own monitoring and evaluation data, and use this where possible. ARC is considering quality control enhancements with respect to ARV customisation in light of the Malawi experience

and will launch a process and governance review. We expect to be able to incorporate these findings into our wider organisational review.

5.7 Workstream 3: Global Review

The TOR requires the Evaluation Team to consider the impact of ARC across all member countries, not just the case study countries. Pathway #3 of the TOC also relates to ARC's ability to attract non-member states and build momentum across the continent for ARC products and services. We therefore cannot just rely on data collected from ARC itself or from country case studies involving ARC member countries to give us a broad enough perspective, or validate pathway #3 of the TOC. We therefore propose to conduct a global review which will pull in other stakeholders related to the ARC programme, who would not otherwise be included in either of the workstreams above.

Essentially, activities under this workstream will fall into three main categories:

1) **Context Assessment** providing a continent-wide, light-touch assessment of risk and response capacity across African countries with high disaster vulnerability (see section 5.4). This will be desk-based, and we will draw on ARC's routine monitoring data, and data from other global DRM initiatives, as much as possible to ensure cost-effectiveness. As such, specific indicators have not yet been identified, but work on this (in collaboration with ARC) will commence in early 2017). As part of the analysis at each later stage of the evaluation (Formative Phase 2, Impact Phase 1 and Impact Phase 2) we will update the context assessment, with a view to identifying whether there has been a broad shift either regionally or continent-wide. Our contribution analysis will then be able to identify the role that ARC may have played in any changes.

2) **Collection of perceptions relating to ARC from AU member states (who may or may not currently be members of ARC)**. Several of the detailed evaluation questions across the TOC relate to stakeholder perceptions of ARC, including from countries who are not participants of ARC. For example:

   - Is there evidence of improved dialogue and coordination with non-member stakeholders and programmes as these actors become increasingly aware of ARCs products and services? (4.5.1)
   - How do different stakeholders view ARC? Is ARC viewed as an influencing agency on the continent?
   - How successful is ARC in marketing their products and services to member states? What encourages governments to join the ARC programme?
   - How well is ARC understood within the context of DRM and DRMF within countries and within the region? (4.1.1)
   - Do member states value and treat ARC differently as an AU/African owned initiative (e.g. rather than via some external agent such as the World Bank)? (4.1.2)

Please see the Evaluation Framework table in section 6 for more detail on which evaluation questions will be included in the perceptions survey.

In our proposal we had previously suggested an online survey, but concerns over likely response levels have led us to revise our strategy. This revised approach will allow us to collect multiple views in a reliable and cost-effective way. The number of surveys we are able to collect at each evaluation phase will vary depending on the timing of relevant workshops and conferences. In our

proposal the online survey was expected to collect data for up to 20 ARC countries and we are confident that a similar number will be possible through this revised approach.

In order to collect as wide a range a range of views as possible, we will 'piggy-back' ARC events and DRMF related conferences in the region and globally, conducting short 'perceptions surveys' with selected individuals which will incorporate closed yes/no questions and ratings exercises. The types of individuals will vary, but we expect to collect data from around 80 respondents, including national country officers, individuals working in DRM / DRMF in Africa, humanitarian organisations and insurers/reinsurers. We will incorporate the ability to disaggregate the data by country, role type and by sector. As mentioned above, quantity is no substitute for quality, and the team's emphasis will be on collecting data from individuals with current, relevant knowledge and experience.

The exact timing of the survey data collection will depend on the timing of relevant events and we expect to collect data on an on-going basis as opportunities arise. Initial possible examples include the ARC Conference of Parties (CoP) in March 2017 and the UNISDR Global Platform in May 2017. It will be important to include non-ARC related conferences so that we can be sure to be collecting information that will contribute to alternative hypotheses.

3) **Collection of views relating to ARC from insurance industry professionals**. There are a number of very specific, detailed evaluation questions that require very detailed technical knowledge, and we are concerned would not be adequately answered by KIIs at the national level as part of the case studies. Examples include:

   - Is ARC cost effective (for donors and member governments) compared to alternative mechanisms for financing the same level of risk? 4.2.4)
   - Is the ARC insurance pool growing enough to have a critical mass of countries and peril coverage options to develop a large enough risk pool to be sustainable? How much diversification benefit is expected from pooling risks that ARC covers? How much diversification benefit has been realised? (4.3.2)
   - Is there sustained interest in ARC products by the capital markets and reinsurers? (4.3.3)

In order to collect data in relation to these more globally focused questions, we propose to conduct a set of interviews with internationally recognised reinsurance industry experts. Our 'Catastrophe Modelling, Risk Pooling and Transfer Expert', Rick Murnane, will lead the identification of suitable key informants, using a global stakeholder mapping approach, and ensuring a broad range of perspectives and backgrounds. We anticipate conducting approximately 20 online interviews during the formative phase of the evaluation, and we will also conduct follow-up interviews during the impact phase.

3) As mentioned above, workstream 4, given that it is an additional extra, is presented separately in Section 7.

5.8 Implementing the Contribution Analysis

Data from the different country case studies within each phase will be added to data from the other workstreams and together will be analysed using a CA approach. At a fundamental level, implementing contribution analysis involves triangulating data from a variety of sources and assessing the contribution of the programme versus some possible alternative explanation. We will implement contribution analysis in a series of steps, explained indicatively below, with further refinement as we create the tools.

First we will need to create an '*evidence database*' that tracks the information we collect from the different workstream that relates to each evaluation question. The design of this database must be conducive to reviewing data from multiple sources so as to build a contribution story. An example

of what this might look like is shown in Figure 6. The database will track against each evaluation question the source type (e.g. documentary, key informant interview, etc.), the level of analysis (e.g. local, national, institutional, global), and the country of reference (or name of organisation if related to the national or global level). We will also keep a secure document of all sources which will be coded to protect the identification of stakeholders but will allow us to understand where the information is coming from so as to provide an assessment of the quality of the source. The final two columns are where we will place the exact information that either supports the ARC contribution story or provides an alternative explanation. It will be information from these fields which we will use to build our contribution stories.

Figure 6: Example of CA evidence database

Eval Q #	Evaluation Question	Source type	Level	Country	Source	Quality of source	Contribution of ARC	Alternate Explanation
2.6.3	Is there evidence that ARC member countries are de-investing in climate change adoption/mitigation strategies in favour of insurance financing?	*Document* *KII* *FGD*	*Local, national; Institution; Global*	*country name or in case of insitution or global, the org. involved*	*code that links back to a document name, or an ID that links back to a person or group*	*Rating score (1-3) on quality of source based on defined rubric - tbd*	information that supports a contribution of ARC	Information that supports an alternative explanation

Much of the information for the contribution analysis will come from key informant interviews that use a consistent set of questions across countries and stakeholder levels. To ensure data triangulation, we will need to create interview tools that are precisely coded so that answers can be easily compared across stakeholders.

Following this methodology, we will be able to identify analytical themes, trace comments back to their source, and be able to give an indication of the strength of the evidence for each theme. Ultimately, we will therefore be able to use the data to develop the Contribution Story, making an assessment of the assumptions and causal pathways underpinning the TOC, and assessing the trajectory towards change.

An initial Contribution Story will be created for the formative phase 1 report, and we expect this will be partial and will explicitly identify areas where further investigation is needed to verify elements of the TOC and of ARC's contribution to change. Each of the subsequent key evaluation reports (formative phase 2, impact phase 1 and impact phase 2) will update this Contribution Story, providing increasing levels of robustness.

One of the key features of CA is the ability to identify and assess the importance of alternative explanations. This step is critical to the evaluation for in the absence of 'attribution', tracking alternative explanations and measuring their importance reduces uncertainty related to the contribution question (Mayne 2001). To unpack and explicitly address these alternative explanations we will use a version of what Lemire et al (2012) describe as a Relevant Explanation Finder (REF). In essence, the REF makes explicit the key alternative explanations to the primary change mechnanisms, coding them and tracking them over time. An indicative example of a possible REF is presented in Figure 7. Each alternative explanation is coded either as a direct rival to the contribution story, thus undermining the story, or as a comingled rival that works with the primary to explain observed outcomes, or as an implementation rival that influences the primary contribution story in either positive or negative ways (Lemire et al 2012). For each of these explanations you identify what must be observed to determine that the rival explanation is in play. Then observations (if they occur) are evaluation to determine the degree of influence the explanation has across a set of pre-defined attributes. Some examples of these attributes described by Lemire et al (2012) are described below:

- Certainty: the degree to which the observed outcome pattern matches the one predicted by the factor or mechanism.

- Robustness: the degree to which the factor or mechanism is identified as a significant contributor across a broad range of data sources.

- Prevalence: degree to which the factor or mechanism occurs across a wide range of implementation environments or target groups.

For example, earlier we presented an example alternative explanation related to the ability of CPs to be implemented in a politically charged post-disaster environment. How we rate the contribution of this attribute might be on prevalence – do we see this occurring in many countries or just a few?

Figure 7: Example of tracking alternative explanations

Description	Type	Level	Identifiers	Degree of Influence			Implication
				Certainty	Robustness	Prevalence	
Effective& timely implementation of CP (INT_01)	Primary explanation	Country/ National	Household recieves benefit within 120 days of payout				
CPs are not implemented on time due to post-disaster political machinations	comingled rival	Country/ National	Household receives money much later than 120 days	High. Households consistently report delayed or no assistance	Low. The reasons for the delays are rarely attributed to policial delays. Often due to implentation issues such as supply sourcing	Low. Only reported for one intervention in one country	The influence of this alternative explanation on the contribution story is low…

During the first formative evaluation we will create a set of the key alternative explanations and track them over time, thus assuring they help fully describe the contribution story.

5.9 Data needs and sources

Much of the data needed for this evaluation will be primary in nature and collected by the Evaluation Team through key informant interviews, focus group discussions, and select surveys. However, secondary data will also be used, and in particular, data collected by ARC, including routine reporting data, will be critical to the evaluation as well. A summary of these key data sources is provided below:

ARC documents:

- ARC financial reports
- ARC M&E reports (on results framework)
- ARC Secretariat Reports and ARC Agency Governing Board records.
- Certificates of Good Standing and ARC Ltd Insurance Policies.
- Technical Review Committee (TRC) and Peer Review Mechanism (PRM) reports.
- Process audit reporting conducted in the event of significant pay-outs by an independent firm.
- Financial audit (for large pay outs that meet certain criteria)

Country data:

- Country hazard, vulnerability, loss, damage and risk information particularly that embedded within ARV (at the national and sub-national levels).
- Country contingency plans (Operation Plans and Final Implementation Plans).
- Country monthly progress reporting during a pay-out and subsequent rollout of the interventions.
- Country reports relating to the institutional and regulatory context for insurance and DRM.

In terms of data to be collected via key informant interviews (KIIs), the ARC programme involves many different levels of stakeholders, from global actors involved in similar risk finance mechanisms, to beneficiary households in member states who are covered by ARC insurance policies (see Table 8). To better understand and demonstrate this stakeholder architecture, we have outlined the various types of stakeholders that will be involved in data collection according to a grouping of levels that constitute the ARC programme and its implementation. These groupings of stakeholders fit into different levels of analyses which help inform our methods.

Table 8: ARC Stakeholders

Level	Stakeholder Group	Examples	Workstream and method of data collection
Global	Regional & global actors	– International experts in EW, DRM, and risk finance – Insurance competitors in the space – Non-member AU states	Global review (KIIs and Perceptions Survey)
Institutional	ARC institutional stakeholders	– ARC Agency staff – ARC Ltd staff – ARC Agency Governing Board – ARC Ltd Governing Board – Peer Review Mechanism (PRM) – Technical Review Committee (TRC)	ARC Organisational Review (KIIs)
National	National government stakeholders	– High-level political actors (e.g. to whom ARC speaks with initially when entering into a MOU) – Ministerial technocrats that belong to various committees in charge of Early Warning, DRMF, CP, ARV configuration – Treasury or whomever is responsible for receiving the insurance pay-out	Global Review (Perceptions survey), Case studies (KIIs and FGDs) ; Stakeholder mapping
	Intervention implementers	– Programme managers and staff who are responsible for implementing interventions identified in CP in the event of a pay-out	Case studies (KIIs and FGDs)
Local	Community stakeholders	– Key informants in the community, traders, chiefs, others who provide insight and context on the impacts of a disaster and the impacts of interventions to provide insight into community resilience to weather risks	Case studies (FGDs)
	Beneficiary	– Those covered by an ARC insurance policy or who received a	Case studies (FGDs) and Quantitative Impact

	households	benefit from an ARC insurance pay-out	Evaluation (household survey)

5.9.1 Data quality

OPM has experience of undertaking comprehensive data quality audits based on data quality assessment frameworks such as the International Monetary Fund's (IMF) Data Quality Assessment Framework (DQAF). However, undertaking such a process within this evaluation project would not be realistic as it is very time intensive. Instead, we will undertake a more selective process for reviewing the data quality of specific sources, based on their importance to the overall evaluation, likely in the form of a standard rubric to allow for more standard comparison across data sources and reviewers. In terms of the secondary data that we expect to access from ARC Agency, the quality will be assessed as part of the ARC Organisational Review. This will include a review of documentation and where appropriate, interviews with selected producers and users of the data. It may also include scrutiny of related data from other sources to check for consistency and coherence. Preliminary review of some documents, for example process reports conducted by an external company, are encouraging and indicate that we will be able to use some of ARC's secondary data. However, we are aware that until very recently there has been an ongoing lack of M&E expertise within the ARC Agency staff team, and so we anticipate there may be gaps in data quality, particularly around routine monitoring data. In future we anticipate that ARC's logframe indicators will be routinely monitored and it may be possible to use some of this information, once the process is being fully implemented. The extent of the gaps will be ascertained during the ARC Organisational Review in Phase 1 of the Formative evaluation.

We anticipate that there may also be gaps in the secondary data generated by country governments, for example around vulnerability mapping, contingency planning or DRM processes. This may impact upon our ability to assess ARC's impact on national DRM policy and processes. However, given that the countries have not yet been selected as case studies, we are not able to provide further detail. Instead, we will include time for review of data quality within the design of each case study, and will seek to overcome any data gaps by collecting relevant primary data, e.g. additional KIIs.

Workstream 4, focusing on a quantitative household survey, has very specific data needs and sources. These are discussed in section 7, along with a robust discussion of data quality across the various options.

5.9.2 Data collection tools

The approach to data collection takes into account the different stakeholder levels described above with the understanding the certain tools and methods will be used across multiple levels but will involve tailored versions. For instance, an important data collection method at all levels is the Key Informant Interview (KII). These semi-structured KII questionnaires will be tailored to the appropriate level, but will have questions in common to allow for the triangulation of data across stakeholder groups. The list of proposed data collection tools and the level at which they will be used is detailed in Table 9.

Table 9: Data collection tools

Data Tool	Description, risks and sampling	Minimum requirement guidelines[13]	Phase	Workstream
Documentary Review	There are many sources of written documentation on various aspects of ARC that are important to the validation of the ToC. Each of these documents will need to be reviewed, coded and linked to the relevant evaluation questions for which it provides evidence. Care will be taken to ensure and record the perceived quality of the document, as judged by the reviewer using a standard rubric. Gaps or inconsistencies will be recorded.	n/a	F&I	Country case studies, ARC Org Review
Key Informant Interviews (KII)	A KII is a semi-structured questionnaire that asks in-depth and probing questions on a range of topics to better understand specific motivations, beliefs, and connections. Key informants will be identified by a stakeholder mapping approach in each country, utilising ARC and DFID knowledge, alongside OPM's and our national consultant's personal networks. Snowballing[14] may also be used to identify additional KIs, particularly if there are areas that require additional probing.	We would anticipate holding approximately 30 KIIs in connection with each case study country.	F&I	Throughout
Focus Group Discussion (FGD)	A FGD is a planned, facilitated discussion among a small group of stakeholders with similar backgrounds or interests. It serves to provide a range of opinions/views on a specific topic of interest. FGDs have to be carefully designed so that political economy factors do not introduce bias into the findings, and to ensure adequate representation amongst the population groups being studied. We plan to design our FGD approach by selecting discussion groups along a number of criteria (beneficiary / non-beneficiary, geographic area, gender, disaster severity, ethnicity, etc.) and plan to conduct approximately 10 FGDs per case study although this will be finalised following the pilot We will therefore prioritise the use of KIIs as a data collection tool except where we can be confident of appropriately homogenous FGs and where the evaluation would benefit from the cost-effectiveness of being able to rapidly collect multiple views in a short space of time. This is most likely to be the	The number of FGDs will depend on the context and will relate to the number of KIIs that are conducted. We expect approximately 10 per case study during the Impact Phase.	I	Country case studies

[13] Quantity is no substitute for quality, and establishing minimum requirements is challenging for a qualitative evaluation. However, we note DFID's request for such information so we have included some guidelines here. It should also be noted that data collection in some countries is likely to be more expensive than in others, which will also affect the activities undertaken.

[14] Snowballing is an approach where additional KIs are identified by asking existing KIs to suggest appropriate, additional people to interview. It can be a very effective way of identifying networks within a country, although care must be taken to ensure broad representation.

	case during the impact assessment at the local level. Sampling methodology will be developed closer to the time, but we will consider both grid sampling and purposeful sampling based on circumstance. Furthermore, if the quantitative impact evaluation is also agreed, the design of the FGDs and broader qualitative data collection will be closely coupled with that of the quantitative design to ensure that the qualitative data provides explanatory power to the quantitative findings.			
Perception Survey	Short (approx. 10-15 minutes) closed-ended questionnaire consisting of a mixture of yes/no and rating exercises that collect a variety of information on stakeholder perceptions of ARC's programmes, products, and activities. Possible venues for data collection are: (i) ARC national and regional workshops; (ii) Other regional and international conferences related to EW, CP, and DRM and DRMF; (iii) In-country from intervention programme managers, government officials, and other relevant in-country stakeholders at the national and community levels.	We anticipate collecting information from approx.. 80 respondents relating to 20 different AU member states.	F&I	Global Review
Participa-tory Impact Assess-ment	Participatory Impact Assessment is a tool that can be used to complement KIIs or FGDs at the local level. We propose using it to enhance data collection at the local level during the impact phase of the evaluation. Please see Annex A4 which gives greater detail on PIA and how it could be used in relation to ARC.	This would complement the FGDs and be used during them.	I	Country Case Studies
Stakehold-er mapping	A process whereby one identifies those who have a stake or interest in the subject being evaluated. Assists the evaluator in understanding the stakeholders, their positions, level of influence, and interest in the subject being evaluated.	Done at the global level, and will be done at the country level for case studies.	F&I	Country Case Studies
Baseline Context Assess-ment	A methodical mapping of AU countries against a standard set of indicators collected primarily from ARC monitoring data and documentary review. It is referred to as a baseline assessment, but will be updated throughout the evaluation to identify shifts and trends. In the context of ARC this tool serves two purposes: to inform the Contribution Analysis on a broad scale and to help identify future country case studies.	We anticipate collecting information relating to 20 different AU member states.	F&I	Global review

Source: OPM

5.10 Selection criteria for case studies

Table 10 shows the countries that are currently actively engaged with ARC[15]. Those with a Board-approved contingency are eligible to take out insurance. Those that took out an insurance policy in a given year become part of that year's risk pool.

Table 10: Countries currently engaged with ARC

Country	Signed MoU	Completed Capacity programme	Board accepted CP	Risk Pool 1 2014-2015	Risk Pool 2 2015-2016	Risk Pool 3 2016-17	Pay-out
Kenya	2012	2013	√	√	√	Policy not taken up	
Mauritania	2012	2013	√	√	√	√	**2015**
Niger	2012	2013	√	√	√	Premium not paid	**2015**
Senegal	2012	2013	√	√	√	√	**2015**
Malawi	2012	2014	√		√		**2017**
Mali	2014	2014	√		√	√	
The Gambia	*?*	2014	√		√	√	
Lesotho	2012	2014					
Mozambique	2012	2014					
Burkina Faso	2012	2014				√	
Zimbabwe	2013	2014					
Madagascar							
Ghana	2016	*Expected end 2016*					
Chad		*Expected end 2016*					
Comoros		*Expected end 2016*					
Cote D'Ivoire		*Expected end 2016*					
Ethiopia							
Nigeria							

Source: OPM and ARC

Using the information above and guided by the evaluation questions, the selection of countries for the case studies will be purposeful and based on several considerations such as:

- Geographical region
- Disaster risks (some rapid onset as well as drought)
- Early adopters and recent signatories
- DRM institutional capacity
- Countries who have continued ARC insurance coverage over multiple years or those who have dropped insurance coverage after one or more years of coverage
- Locations where pay outs have been made and where they have not yet been made

Since many of these factors will evolve overtime, it is hard to predict which countries will be best suited for analysis at this early stage in the evaluation. Countries are joining ARC on an on-going basis, some complete the capacity building process but opt not to take out insurance, and others

[15] As of the time of writing.

might decide not to take out insurance over consecutive years. As such, our selection of case studies needs to remain flexible and iterative.

Below we propose a shortlist of countries for the first formative evaluation and then describe the guiding principles for the selection of countries for the second formative and impact phases of the evaluation. We intend to discuss further with both DFID and ARC Agency before making the final selection. At the time of writing we are in conversation with ARC Agency, and are continually gaining updated information as to countries who are likely to join or leave the risk pool. We are also open to considering splitting a case study into two lighter touch studies if there are specific issues worthy of investigation in particular countries, but their engagement with ARC would not warrant a full case study (for example, a country that has had very limited engagement with ARC, or a country that investigated but chose not to sign up to ARC's capacity building programme).

First Formative

The objective of the formative evaluation is to identify lessons learned and make recommendations to improve the operational efficiency and effectiveness of the ARC suite of programmes, products and services. With this in mind, the focus of the first formative evaluation as related to Member States is on early adopters' experiences in going through the ARC capacity-building programme and, where relevant, experiences in the receipt of a pay-out and subsequent implementation of ARC-funded interventions. We have developed a shortlist of countries, listing the reasoning behind the selections (see Table 11 below). We anticipate a collaborative discussion with DFID and ARC to determine the final list.

Table 11: Shortlist of country selections for first formative

Country	Criteria considered
Mauritania	– Example of a country that received an ARC pay-out where implementation of the CP was considered quite successful – DRM capacity low – West Africa
Niger or Senegal	– Example of a country that received an ARC pay-out where implementation of the CP was considered challenged – DRM capacity low – West Africa
Malawi	– Example of a country that took out an insurance premium, had expectations of a pay out that was not initially agreed, but following a review was agreed at $8.1m (for more see Joint ARC/GoM press release) – Long relationship with ARC but is not expected to take out insurance in the coming year – Southern /Eastern Africa
Kenya	– Example of a country that has been involved throughout, ARC's biggest client, although has recently withdrawn from the risk pool – DRM capacity high – East Africa
Mozambique	– Example of a country that completed the programme but did not take out insurance because of financial and possibly other constraints. ARC is currently re-engaging them so it is a good opportunity to observe this process. – High government buy-in to DRM – DRM capacity high – Southern / Eastern Africa – Slow and rapid onset disasters

Source: OPM

Second Formative

In the second formative evaluation we want to test the learning adaptation cycle of ARC and continue to build evidence as to whether ARC remains on the right trajectory towards achieving its outcomes. The focus therefore of the second formative evaluation as it relates to Member States

is around evidence of uptake in learning by early adopters and how ARC programmes evolve and improve for new countries going through the process. As we do not know exactly which countries will be best suited for study, we propose a set of selection criteria to consider.

Table 12: Country Selections for second formative

Country	Criteria to consider
One repeat country from the first formative	– Did the country take up insurance contract again? – Was there a specific finding from the first formative that should be revisited? – Is there a potential pay-out expected and thus the possibility to observe an implementation? – Geography to ensure representation of all regions
One country that has or is planning to take out a rapid-onset insurance policy	– Is there a country that is taking out both types of insurance (slow & rapid onset)? – Geography to ensure representation of all regions
One country that just recently had a pay-out or is expecting a pay-out	– Is there a potential pay-out expected and thus the possibility to observe an implementation? – Geography to ensure representation of all regions – Level of DRM capacity

Source: OPM

Impact assessments

The impact assessments are too far in the future to definitively select criteria but some of the considerations are explained below:

- If DFID decides to do the optional quasi-experimental household impact evaluation, the country selected for this study should be included as a case study country;
- Selection of at least one country in each geographical region;
- Selection of at least one country that has taken out a rapid-onset insurance;
- Selection of a mixture of high and low capacity DRM countries
- Selection of countries that have a long history with ARC and some that are newer

6 Evaluation framework

The table below acts as a simple summary table, presenting an overview of the entire evaluation. The table links the evaluation questions with the different phases of activity and the four workstreams, as well as showing the data collection methods that will be used to address each question. For each question and for each phase of the evaluation (formative one = F1, formative two = F2 and Impact = IM) we also provide an indication of the level of certainty we have of robustly answering each question based on a rubric that measures the expected strength of the evidence of contribution:

Tentative (T): see evidence of contribution across multiple stakeholders in at least one country
Plausible (P): see evidence of contribution across multiple stakeholders in multiple countries
Likely (L): see evidence of contribution across multiple stakeholders in multiple countries across time.
Highly likely (HL): see *repeated* evidence of contribution across multiple stakeholders in multiple countries across time.

Table 13: Evaluation framework

#	Question	Who engaged	F1	F2	IM	Workstream	Data methods
1.	**To what extent do ARC's institutional setup and outputs lead to the adoption and effective use of ARC insurance products? Can this be improved?**	**ARC staff, National country officers**	**T**	**L**	**HL**	**Primarily ARC Org Review**	**Documentary review, KIIs**
1.1	How and to what extent do ARCs products/services/activities support on-going engagement and an on-going learning cycle within and across countries?	ARC staff, National country officers	T	L	HL	ARC Org review Case studies	Documentary review, KIIs
1.2	To what extent does ARC's institutional model (role, governance, financing structure) support the delivery of ARC's outputs?	ARC staff National country officers	T	L	HL	ARC Org review	Documentary review, KIIs
1.3	Are ARC product offerings acceptable for the market? Could such products be offered regardless of donor involvement?	National country officers DRM experts; reinsurers	T	P	L	Case studies Global review	Documentary review, KIIs, perceptions survey
1.4	How well do ARCs risk models function? Are they improving overtime?	DRM experts, Technical modelling experts Country officers	T	L	HL	ARC Org review Case studies	Documentary review, KIIs
2	**To what extent has ARC contributed to in-country timely and effective responses that protect affected households' livelihoods and prevent asset loss and food insecurity?**	**National/local country officers; Implementation staff; beneficiaries**	**T**	**P**	**L** **HL**	**Primarily case studies and** **Quantitative household survey**	**KIIs, Household Survey, PIA**
2.1	Does the ARC model lead to enough disaster financing for different size slow and rapid onset disasters to make a crucial difference in the livelihoods of households? In what way is ARC's impact limited when other planned sources/mechanisms of financing are not available?	National/local country officers; Humanitarian agencies beneficiaries		T	L HL	Case studies Quantitative household survey	PIA, quant household survey
2.2	Where pay outs have occurred, to what extent have countries implemented contingency plans effectively? What have been the	ARC staff; National/local country officers; Implementation	P	L	HL	Case studies	Documentary review, KIIs, FGDs

#	Question	Who engaged	F1	F2	IM	Workstream	Data methods
	drivers of a successful CP implementation? What have been the barriers to an effective CP implementation?	agencies; beneficiaries					
2.3	Is the ARC iterative learning model capturing lessons-learned from various country implementations, leading to future improvements in country response delivery?	ARC Staff; National country officers; Implementation agencies	T	L	HL	ARC org rev Case studies	Documentary review, KIIs
2.4	What evidence is there that pay-outs to governments and the implementation of ARC Contingency Plans has contributed to the protection of livelihoods and food security, and prevented asset loss?	ARC staff; National/local country officers; Implementation agencies; beneficiaries	T	P	L HL	Case studies Quantitative household survey	PIA,quant household survey
2.5	Does ARC deliver equally well in both slow and rapid onset situations?	ARC staff; National/local country officers; Implementation agencies; beneficiaries		P	L	Case studies	KIIs, FGDs, PIA
2.6	Is there evidence that links a country's improved DRMF planning to continuous growth?	National country officers; DRM/economic experts			P	Case studies Global review	KIIs, documentary review
3	**To what extent has ARC influenced AU member states' capacity to anticipate, plan, finance and respond to climate related disasters generally, and more specifically in making best use of ARC?**	ARC staff; National/local country officers; Implementation agencies		**P**	**HL**	**Primarily case studies**	**Documentary review, KIIs, Context assessment**
3.1	Is there evidence of countries taking action (e.g. creation of broader risk-management platforms, planned budgetary expenditures related to DRM, uptake of insurance or other risk-financing products, etc.) as a result of increased knowledge of DRM and quantified risk? What evidence is there that the change is sustainable?	ARC staff; National/local country officers; Implementation agencies	T	P	HL	Case studies	Documentary review, KIIs, context assessment
3.2	What combination/network of stakeholders has ARC engaged in the country to support policy and practice change and is this the relevant network for changes to occur?	ARC staff; National country officers; DRM experts; humanitarian orgs.	T	P	HL	Case studies Global review	KIIs, Perceptions survey, stakeholder mapping
3.3	Does ARC engagement with Member States lead to tangible commitments by governments in terms of dedicated resources and time?	ARC staff; National country officers	T	P	HL	Case studies	Documentary review, KIIs
4.	**Do participating governments and other stakeholders value ARC's risk pool and technical assistance? Why?**	National country officers; reinsurers; humanitarian orgs.	**T**	**P**	**HL**	**Case studies Global review**	**KIIs, Perceptions survey**
4.1	How relevant is ARCs strategy and role in a country / the region relative to the wider country context and broader DRM architecture?	ARC staff; National country officers; DRM experts; humanitarian orgs		T	L	Case studies, Global review	KIIs, Perceptions survey
4.2	Are ARCs products and services	ARC staff;		T	HL	Case studies,	KIIs

#	Question	Who engaged	F1	F2	IM	Workstream	Data methods
	competitive in the broadening area of DRMF on the continent?	National country officers; DRM experts; humanitarian orgs				Global review	
4.3	Over time, is there evidence of a diverse market of risk financing products available in African Countries? If so, what evidence exists that ARC contributed to this market?	DRM and DRMF experts; insurers/reinsurers; humanitarian orgs			HL	Global review	KIIs, perceptions survey, context assessment
4.4	To what extent do member country stakeholders consider ARC as a key actor/partner in supporting effective risk management and risk financing in the country?	National country officers	T	P	HL	Case studies Global review	KIIs, perception survey
4.5	To what extent do non-member country stakeholders consider ARC as a key actor/partner supporting effective risk management and risk financing in the country?	National country officers	T	P	HL	Global review	Perception survey
4.6	What is the nature of the link between ARC pay-outs, successful CP implementation and governments' motivation to engage with ARC?	National country officers; ARC staff;		T	L	Case studies Global review	KIIs, perceptions survey

7 Options for a quantitative impact evaluation

As noted above and in the TOR, an additional £500K may be available for additional work as part of the evaluation. The Evaluation Team is keen to take advantage of this opportunity to strengthen the robustness of the overall evaluation findings by incorporating a quantitative impact evaluation component to measure the effects of ARC at the beneficiary level. The findings would then be incorporated into our overall analysis and synthesis of all the qualitative and quantitative data using a Contribution Analysis approach. This section outlines our recommended approach as well as alternative options. Discussion on these options continues, with the aim of reaching agreement by mid/late 2017 if possible.

The impact evaluation of ARC is intended to shed light on the effects of the programme at the beneficiary level, namely households and individuals receiving the ARC related pay out. The impact should therefore be measured on household and individual level indicators, which are relevant for the ARC programme. These include for instance food and non-food consumption, livestock and asset retention and other poverty related outcome indicators, with a focus on the ability of households and individuals to withstand the economic shock due to a natural hazard (severe drought). If beneficiaries' experience a reduction in asset depletion thanks to ARC and a severe drought has a less detrimental effect on household welfare than in a non-ARC context, it can be argued that there is a positive impact on programme beneficiaries.

7.1 The problem of the counterfactual

In technical terms, the core and ideal objective of this quantitative impact analysis is to understand whether the ARC transfer has had an effect on the target population and to quantify with statistical confidence the scale of the impact detected. However, there is a key challenge for evaluating the impact of any programme with statistical confidence, which is the identification of an appropriate control group that does not benefit from the programme. This control group can be used as a *counterfactual* to the treatment group that does benefit. In other words, a counterfactual enables us to answer the question: *what would have happened to the treatment group (i.e. households and individuals that benefited from the ARC transfer) had the ARC intervention not taken place*?
A valid control should satisfy the following three conditions: (i) The treatment and control groups should share on average the same characteristics; (ii) Treatment and control groups should react to the programme in the same way if it were offered to both groups; (iii) Treatment and control groups should not be differentially exposed to other interventions during the period of the evaluation. For the HSNP2 impact evaluation, the eligibility criteria used to identify HSNP beneficiaries was used to identify both treatment and control groups. A Regression Discontinuity Design (RDD) was employed to define as treatment those households located just below the assignment threshold, and as control those located just above the threshold, meaning that both groups can be considered equal and comparable.

Randomised Controlled Trials (RCTs) are typically considered the most robust design for quantitative impact evaluations, as RCTs better address the problem of selection bias and provide the most convincing estimate of the counterfactual. By randomising which individuals, households or communities are affected by a policy and which are not, a control group is established that, by construction, is statistically identical (no systematic differences due to selection bias) to the group receiving the policy before the intervention has begun. However, in the case of the ARC impact evaluation, this does not appear to be a feasible option, as the programme's stated aim is to target all eligible ARC beneficiaries with the transfer and no control group can be randomly identified within the population of potential eligible recipients.

This section discusses the alternative impact analysis approaches that we believe can be developed to obtain some measure of the effects of ARC at the beneficiary level, given the programme and budget constraints. In particular, we envisage three different options:

1. A counterfactual analysis based on a quasi-experimental design and a large sample, which takes advantage of data collected for the HSNP2 evaluation in Kenya; this may turn into a treatment only analysis if a control group cannot be robustly identified.
2. A counterfactual analysis based on a quasi-experimental design and a smaller sample, which is centred around two new quantitative surveys in a single ARC country; this may also turn into a treatment only analysis if a control group cannot be robustly identified.
3. A descriptive historical analysis of trends of the treatment group only (i.e. ARC beneficiaries), which would be based on a comparative analysis of new primary survey data and existing secondary data. The latter would serve as baseline data.

The main features of each of these three options are summarised below and explained in detail in Annex 5, with both positive and negative aspects associated with them described. Given the higher degree of robustness from a quantitative impact approach perspective and the evident value for money associated with option 1, it is fair to say that this currently represents our preferred route. However, option 2 and option 3 have strengths that should be taken account. Especially if the potential downsides of option 1, as discussed below, are deemed not acceptable.

7.2 Option 1: Impact evaluation based on HSNP2 baseline

There exists an opportunity to design and develop a quasi-experimental design that could allow us to detect a quantitative measure of impact of ARC on the beneficiary population as well as a disaggregated measures of impact across sub-sample categories of interest. OPM has recently carried out a large household survey in Northern Kenya as part of the impact evaluation of the second phase of DFID's Hunger Safety Net Programme (HSNP2). This survey collected detailed information on almost 6,000 households (5,980) across four counties (i.e. Turkana, Wajir, Mandera and Marsabit) on an extensive range of indicators, including household characteristics, consumption patterns, assets, land, livestock, financial inclusion, business activities and transfers received. We believe that the OPM HSNP2 survey data could be used as the baseline for an inter-temporal impact analysis of ARC, in which the endline would be represented by a new survey implemented by OPM in the same areas of Northern Kenya, once the ARC transfer will have been disbursed.

The main reasons why we consider this approach as both feasible and appealing are:

- We are confident about the high quality of the HSNP2 data, which covers impact areas that are clearly of interest for the analysis of the effects of ARC. This data represents a baseline scenario for the ARC evaluation, as no ARC transfer had occurred yet.

- The counties in Northern Kenya where HSNP2 data was collected are extremely prone to drought risk and its population likely to be the recipient of ARC pay-outs in case ARC is triggered in Kenya. It is therefore likely that the HSNP2 sample is relevant for ARC.

- The large sample size of the HSNP2 survey would provide our estimation model with considerable power to detect relatively small changes in outcome indicators of interest that can be attributable to the ARC intervention (See Annex 5). Heterogeneity analysis of relevant population groups may also be possible, as sub-samples would be relatively large too.

- The likely differentials in implementation of ARC across Northern Kenya in case of drought should allow to identify a 'treatment' group (received ARC transfer) and a 'control' group (did not receive ARC transfer). As this unequal roll-out would not follow a random pattern, a quasi-experimental design (e.g. matching with difference-in-differences) will be employed to

build a valid counterfactual and measure impact attributable to ARC by comparing treatment and control outcomes.

- Finally, it may be possible to directly compare the effects of HSNP2 as measured by the separate OPM impact evaluation with the effects of ARC. The two analysis will be based on the same sample of households and individuals and will cover the same range of impact indicators. The impact attributable to the HSNP2 transfer will therefore be comparable in terms of significance and magnitude to the impact attributable to ARC.

For more details on the proposed quasi-experimental design see Annex A.5.

7.2.1 Possible challenges and limitations

This strategy presents a series of risks and potential issues that will have to be taken into account and will determine whether a quasi-experimental design is indeed feasible:

1. Project coverage is extensive and it is not possible to identify a control group:
    - In this case, a multivariate regression approach could be undertaken by focusing only on the treatment group, with the aim of estimating the influence of a series of explanatory variables on the outcome indicators of interest. The large sample size would allow us to perform a range of sub-sample analyses.

2. The HSNP2 baseline data does not provide us with a viable baseline dataset:
    - Although the HSNP2 baseline survey will be representative of four Northern Kenyan provinces that will most likely to be affected by a drought (data on previous droughts confirms this), it is possible that a scattered ARC project implementation will leave us with a population of ARC beneficiaries that is too small to be derived from the HSNP2 data. In this case, historical data analysis defined as option 3 above and a cross-sectional (based only on one round of data collection) regression analysis could be used to analyse ARC effects;

3. The sub-sample of ARC treatment and controls is too small for the analysis:
    - The PSM approach is based on a first stage propensity score model, which ensures that a balance across the characteristics of treatment and control households is achieved through a matching procedure. The latter can lead to the loss of several observations that cannot be matched because they are not sufficiently similar. Once we know what HSNP2 households can be considered treatments or control for ARC, the resulting matching procedure could leave us with a sample that is too small for performing further heterogeneity analysis.

4. The same HSNP2 households cannot be traced and re-interviewed for ARC:
    - Although we might be able to identify treatment and control areas for the ARC evaluation and derive the corresponding groups within the HSNP2 dataset, we might not be able to re-interview at follow-up the exact same households but only households living in the same treated or non-treated villages. In this case, we would not be able to employ a panel data analysis and we could not control for time-invariant unobservable confounders. However, we could still undertake an inter-temporal analysis based on a Difference-in-differences (Diff-in-diff) approach. This will enable us to control for baseline differences between our treatment and control groups and to remove biases in the estimates of programme impact that could result from permanent observable differences between the two groups. The key assumption underpinning a Diff-in-diff approach is that our treatment and control groups will follow the same time trend over the course of the evaluation period (the

parallel trend assumption), which seems to be likely given the similar context in which they live.

5. Finally, it is possible that Kenya will drop out from ARC:
    - In this case, we would have to abandon this HSNP2 based approach and move the focus on either option 2 or 3. The alternative country would be chosen amongst those that are part of ARC but have not received the ARC transfer yet.

7.3 Option 2: Impact evaluation based on two new surveys

If Kenya is not deemed as a suitable case study for the impact evaluation of ARC, the same budget assigned to this component of the evaluation could be used to undertake two smaller household surveys rather than a single household survey the size of HSNP2. The two surveys would have an approximate sample size of 2,000 households and would represent the baseline and endline stages of a quantitative impact evaluation to be held in a single ARC country.

The main features of this option can be summarised as follows:

- Multipurpose household surveys that will cover a range of relevant indicators for the impact evaluation of ARC at the beneficiary level. This will include as for option 1 modules on household characteristics, food and non-food consumption patterns, assets, livelihoods, livestock, land and financial inclusion, amongst others.
- The survey instruments will be developed by OPM in line with previous quantitative evaluations undertaken on similar areas of interest (e.g. HSNP2 survey). This will translate into a tailored and comprehensive list of impact indicators that will underpin the analysis of beneficiary level outcome levels and ensure an informative assessment of the effects of ARC on targeted households and individuals.
- High quality data will be ensured given the extensive experience and expertise of OPM in developing and implementing household surveys for the purpose of impact evaluations. In particular, OPM has undertaken quantitative surveys in a series of ARC countries[16] before and could rely on its contacts with local survey agencies for mobilising local survey teams relatively quickly and still guarantee high data quality.
- The first round of data will be collected in a pre-treatment situation across areas of the country (a single ARC country) that are likely to be affected by a drought and thus receive the ARC pay-out, when triggered. If programme implementation is not homogeneous, we may be able to identify treatment and control groups at endline once the ARC transfer has been disbursed. If the counterfactual analysis is not possible, we would still be able to undertaken a treatment only analysis of trends.

Similarly to option 1, we are expecting the ARC roll-out to be heterogeneous across the target population due to either programme targeting design or programme implementation issues. If this is the case, we should be able to identify two groups amongst our 2,000 households, which could be considered as a treatment and a control group in a quasi-experimental design. In case a counterfactual based design becomes infeasible due to a homogenous roll-out of ARC across our target population or even a scattered implementation that does not provide us with sufficiently large and balanced sub-groups of treatments and controls[17], we still have the option of undertaking a robust treatment only analysis based on a multivariate regression model. For more details on these different approaches see Annex A.5.

[16] For instance Malawi or Zimbabwe are two ARC countries in which OPM has undertaken household surveys and quantitative impact studies. Our local experience and local contacts would greatly assist in carrying out household surveys in these two countries. We would expect to be able to mobilise quickly and still collect high quality data.
[17] For instance, if out of 2,000 households only 250 are reached (or not reached) by the ARC related pay-out, our treatment group (or control group) would be too small to allow us to undertake a robust counterfactual analysis.

The approach we are proposing would take advantage of the rich quantitative data collected through our survey questionnaires to construct an econometric specification that models household outcomes. Besides, as it will be based on a panel of households (same household interviewed at baseline and endline), rather than repeated cross sections, our treatment only analysis will deal with time-invariant unobservable characteristics of beneficiaries and further isolate ARC's effects.

7.3.1 Possible risks for evaluation viability

It is important to highlight two risks that could undermine the viability of option 2:

1. No (or very few) households captured in our baseline survey are actually reached by the ARC pay-out due to programme targeting design or implementation issues;

2. The country selected for the impact evaluation does not receive the ARC intervention as it is not affected by a sufficiently severe drought or drops out from ARC.

In either case, it would not be possible to carry out any impact evaluation of the impact of ARC at the beneficiary level.

7.4 Option 3: Historical data analysis

Both option 1 and 2 described above entail that the analysis of ARC's impact at the beneficiary level will be undertaken in a single country (i.e. either Kenya to take advantage of HSNP2 or a different ARC country where to administer new baseline and endline surveys). Option 3 offers the opportunity to cover two countries, though it is important to consider that this would represent a quantitative analysis of descriptive trends in ARC-related indicators and not a robust quantitative estimation of impact. In particular, option 3 would rely on the existence of *historical data* in ARC countries concerning welfare and resilience indicators of interest, which can then be used as baseline data to be compared with newly collected primary data after the ARC pay-out is triggered. The idea behind this approach is to take advantage of the wealth of data already gathered on drought affected populations in ARC countries. This data, which normally includes information on consumption, livelihoods and asset retention behaviour of households and individuals facing harsh natural conditions, can provide a pre-ARC picture of the situation (i.e. baseline). Once a sufficiently severe drought leads to ARC being triggered, post-treatment data could then be collected by OPM on a similar range of indicators for a sample taken from the same population (i.e. endline). The comparison of indicator levels from the two data sources (i.e. existing secondary data for baseline and new primary data for endline) would represent a descriptive indication of whether ARC had some positive effects.

The key advantage of this approach is represented by the fact that we can run the two endline surveys, which are feasible within the budget constraints for this component of the evaluation, in two different countries[18]. Our quantitative analysis would not be limited to a single case study as it is the case for option 1 and option 2. As ARC is operating across multiple countries, we appreciate that this is an appealing aspect of this approach. At the same time, the resources for undertaking the endline fieldwork would have to be provided only *if* and *when* the ARC pay-out is actually triggered and there is limited risk of wasting resources. However, the limitations in our ability to measure and attribute impact to ARC under this option should be taken seriously. Whilst descriptive trends can be informative and give us an indication of whether having received the

[18] Due to economies of scales associated with setting up and running a survey, it would not be as affordable to run more than two separate surveys (e.g. three of four surveys of around 1,000 households). Fix costs associated with the training of the enumerators and the piloting and pre-testing of survey instruments imply that the cost of multiple surveys across three of more countries would not be economical. Besides, design and analysis time would also have to be increased to cover the different sample analyses and this would exceed the budget limit.

ARC pay-out is showing some visible effect on outcome indicators (e.g. are food consumption levels considerably higher or asset selling considerably lower after a drought mitigated by the ARC pay-out, when compared to a similar drought that took place in the past without the ARC intervention?), it would not be possible to conclude with confidence that ARC had a positive impact. Therefore, the pros and cons of this option need to be carefully compared to those of option 1 and 2.

7.4.1 Assessment of baseline data availability

An important consideration to be noted in the context of option 3 is the reliance on the existence of available and good quality information to be used as baseline. We have therefore performed a preliminary assessment of existing data sources to determine whether this is the case. Specifically, the purpose of this assessment was to identify the scope, type, accessibility and quality of quantitative data that could be used to populate the historical baseline study for the ARC beneficiary level evaluation as explained above. The criteria used for the assessment are described in more detail in Annex A.5. The countries assessed are those who undertook the capacity-building programme in the second year: Burkina Faso, The Gambia, Malawi, Mali and Zimbabwe.

No panel data that met our criteria was found for these countries. Repeated cross-sectional data appeared to be the 'next best' type. This is not an issue in our case, as the analysis envisaged for option 3 would only use a single round of existing data as baseline. Table 14 below presents a summary of the country data findings. For a more detailed explanation see Annex A.5.

Table 14: Summary of country data findings

Country	Assessment
The Gambia	No appropriate data was identified for The Gambia. A stock-taking report produced by UNDP on The Gambia's National Adaptation Plan Process observes that 'very little research has been done in the Gambia on the linkages between the climate and natural and social processes'[19]- which was confirmed by this assessment
Burkina Faso	Two reports produced by USAID and FAO present good data sets for Burkina Faso. However in general, appropriate (and/or accessible) data is limited compared to the other ARC countries assessed- in particular, there are comparatively few government-led household surveys with relevant indicators, large sample sizes and in recent years.
Malawi	There are strong data sets available for Malawi- in particular the Third Integrated Household Survey, and various other nationally representative, recent household surveys led by the National Statistical Office with support from donors. Data appears to be well presented and easy to access.
Mali	FAO and USAID have produced strong data sets for Mali, as both organisations have a strong focus on resilience in the Sahel region. The available data (from official, nationally-representative surveys like the DHS) is limited in terms of the scope and relevance of its indicators, however, compared to Malawi and Zimbabwe.
Zimbabwe	Zimbabwe has a comparably large number of strong, relevant data sets available. Nationally representative surveys have been produced by the National Statistics Agency and other government institutions, and are relatively comprehensive in scope and sufficiently recent

[19] http://adaptation-undp.org/sites/default/files/downloads/gambia_stocktaking_report_for_nap_and_road-map_for_cambodia_nap_gsp_and_giz_31.pdf

The country-specific assessments presented above indicate a generally positive picture with a large range of relevant and comprehensive studies undertaken on issues of interest for the analysis of the effects of droughts on household and individual level welfare and resilience. Both the broad scope of the surveys undertaken and the reputation of the agencies that have undertaken or sponsored these studies give us confidence on the suitability and reliability of the data collected. It seems reasonable to suggest at this stage that Malawi and Zimbabwe would represent two good case studies to focus on, given the expected high quality of the datasets available to be used as baseline there. At the same time and as already mentioned above, OPM's own experience in both countries would facilitate the endline fieldwork and data collection processes and ensure a sound quantitative analysis. For this to go ahead, of course, both Zimbabwe and Malawi would need to be ARC members (Zimbabwe is currently unable to join without donor support towards the cost of its premiums). Malawi did not renew its insurance for Pool 3, but ARC continues to engage with the country on participation in the future.

8 Ethics

Conducting qualitative and quantitative field work requires high ethical standards to ensure that expectations are not raised, confidentiality is maintained and respondents are never forced to participate or encouraged to speak about subjects that may be traumatising. The Evaluation Team will draw on its wide experience of conducting qualitative and quantitative fieldwork to ensure that these standards are met, and seek further review where appropriate, to adhere to ethical protocols in line with the OECD-DAC principles of accuracy and credibility and *DFID's Ethics Guidance for Research and Evaluation*.

An important consideration when seeking an individual's participation in research, is to ensure that they understand exactly what is being done with the information they have provided. OPM has extensive experience of conducting mixed methods research with vulnerable people and will ensure that the below-described standards are met throughout the impact evaluation.

Informed consent: means that potential respondents are given enough information about the research and researchers ensure that there is no explicit or implicit coercion so that potential respondents can make an informed and free decision on their possible involvement in the fieldwork.

Anonymity: given that some of the research respondents (as part of the household surveys) will share considerable amounts of personal information, it is OPM's responsibility to ensure that their confidentiality is maintained and personal information is protected. This will be operationalized by ensuring that all datasets are anonymised, in the sense that all names of respondents are removed before the data is shared publically.

Ensuring the safety of participants: this means that the environment in which research is conducted is physically safe. The impact evaluation team will seek to achieve this by ensuring that fieldworkers are local to areas in which they are assigned. In addition, fieldwork supervisors will support the fieldwork manager in monitoring local security concerns.

The relationship between our work and the *DFID Ethics Principles for Research and Evaluation* (DFID, 2011) is outlined below.

1. In the countries in which we will conduct case studies, we will obtain formal approval to undertake primary data collection from the relevant government authorities, where appropriate. For the formative evaluation, most Key Informants will be answering questions in their professional capacity, and we will not be collecting biological samples, so it may be that formal ethics approval is not required in many countries. Once the case study countries have been finalised in agreement with DFID, we will be able to start this process.
2. Our analysis will be to sufficiently high standard that the findings can be reliably used for their intended purpose.
3. We will avoid any harm to participants. We will seek to achieve this by ensuring that fieldworkers are local to areas in which they are assigned. In addition, fieldwork supervisors will support the fieldwork manager in monitoring local security concerns. The team will endeavour to ensure that any service disruptions (for example, the operations of cash transfer programmes) are kept to a minimum by ensuring that staff are informed as early as possible of the exact dates of the fieldwork and what KIIs will take place, when and where. The sequencing of interviews and FGDs will also - as far as possible - be organised in cooperation with community members to ensure the smooth running of the research and to minimise disruption to village life.
4. All participation in our evaluation is entirely voluntary. We practice informed consent meaning that potential respondents are given enough information about the research and researchers ensure that there is no explicit or implicit coercion so that potential respondents can make an informed and free decision on their possible involvement in the fieldwork. All

participants are made aware of their right to withdraw from research/ evaluation and withdraw any data concerning them at any point without fear of penalty.

5. We will ensure confidentiality of information, privacy and anonymity of all study participants. We full understand our responsibility to ensure that their confidentiality is maintained and personal information is protected. This will be operationalised by ensuring that all datasets are anonymised, in the sense that all names or other identifying information of respondents are removed before the data is shared publically. If any audio recordings of the FGDs and individual interviews are made, this will be with participants' consent, and then transcribed and translated into English. The confidentiality and anonymity of FGD participants and key informants will be respected and maintained at all times by ensuring that nothing that is recorded can be ascribed to a particular individual, and the transcripts and recordings will be accessible only to the researchers on the team.
6. We will abide by all international human rights conventions and covenants to which the United Kingdom is a signatory, regardless of local country standards. We will also take account of local and national laws in the countries in which we undertake fieldwork.
7. We will respect cultural sensitivities. FGDs and interviews will be carried out in local languages, as relevant, and interpreters will only be used if participants are uncomfortable with using these languages. We will take account of differences in culture, local behaviour and norms, religious beliefs and practices, sexual orientation, gender roles, disability, age and ethnicity and other social differences such as class when planning studies and communicating findings.
8. As discussed in our stakeholder engagement and communication strategy, we will share our results widely. Full methodological details and information on who has undertaken the work will be given. While respecting confidentiality requirements, our primary data will be made public to allow secondary analyses.
9. We will act independently from the programme we are evaluating. We will disclose any potential conflicts of interest that might jeopardise the integrity of the methodology or the outputs of research/ evaluation should any arise.
10. We will ensure that women and socially excluded groups can freely and safely participate in our research.

9 Evaluation management and governance

9.1 Personnel and team structure

The team structure has evolved slightly during the Inception Phase due to the addition of two team members, both of which were agreed with DFID: Zoë Scott and Claire Simon.

Claire Simon joined the team early in the Inception Phase, because of her previous work evaluating ARC, and has worked alongside Paula Villanueva as an additional M&E expert. Claire has over 10 years' experience in quantitative and qualitative evaluation research that includes formulating program strategy, developing baseline, mid-line, and end-line evaluation toolkits, training and managing survey and research teams, analysing data, assessing outcomes, extracting lessons learned, and reporting to stakeholders. Her areas of subject matter expertise include social protection, food security, poverty measurement, disaster risk management, social enterprise and microfinance. She specializes in the design, implementation, and analysis of results-based monitoring and evaluation (M&E) systems for international development programs. Before turning to the field of development assistance, Ms. Simon spent eight years working in software development, managing teams to build client relationship management (CRM) systems for large financial institutions and online retail organizations. Dr. Simon holds a PhD in Geography from the University of Colorado and a MA in International Economics and Finance from Brandies University. Dr. Simon's pen portrait is available in Annex **Error! Reference source not found.**.

Zoë Scott is currently the Research Coordinator for DFID's 2 year research programme on Shock-Responsive Social Protection and was brought into the team to improve overlap between those two project teams in OPM. Since then, Nils Riemenschneider has moved into a different role within OPM and now has reduced time to work on projects. In discussion with DFID, it was agreed that Zoë should replace Nils as Team Leader. Zoë leads OPM's work on Disaster Risk and is a senior consultant with almost 20 years' experience in evaluation, impact assessments, policy research and project management. As the Research Coordinator for DFID's 2 year research programme on Shock-Responsive Social Protection Systems, she provides oversight of qualitative and quantitative data collection in six countries (Mozambique, Lesotho, Mali, Sahel region, Pakistan and the Philippines). She has worked extensively on evaluations and assessments across Africa, including being Team Leader for a qualitative evaluation of a decade of governance reforms in Mozambique for OECD-DAC and managing a mixed methods £1.4 million impact evaluation for DFID in Zimbabwe. She recently completed an evaluation of UNISDR's Global Assessment Report 2007-2015 and she currently chairs the steering committee for the Impact Evaluation of the UK's Disaster Emergency and Preparedness Programme (DEPP).

Zoë is also an experienced Programme Manager, for example she managed the Inception year of a £15 million DFID research programme involving a number of qualitative case studies and coordinated RCTs across multiple countries. She was also the Project Manager and DRM governance technical lead for a large 2 year IFRC research programme on National and Local Capacity Building for Disaster Risk Management (2013-2015) involving qualitative case study research in the Philippines, Ethiopia, Myanmar, Sierra Leone, Pakistan, Mozambique and Haiti. The research involved the design and testing of a DRM M&E framework, later published in the journal *Disaster Prevention and Management*. Zoë was a lead author for the World Disasters Report 2015 and has published widely on disaster risk governance. Her full CV is available in Annex **Error! Reference source not found.**.

We also propose the following additions to the team:

Felicity Le Quesne to replace Constantin Albot in the team. Felicity is a Consultant in OPM's Climate Change and Disaster Risk Portfolio**.** She is an experienced researcher and policy analyst,

and has worked with government on environmental and social development issues in South Africa, India, Brazil and Indonesia. Her recent work at OPM includes an evaluation of the Global Assessment Report on Disaster Risk Reduction, and production of reports on public management of climate finance. She is currently Programme Manager for a DFID-funded £14m research programme on energy and economic growth, a role which includes design and implementation of a communications and research uptake strategy. Prior to joining OPM, she worked for the United Nations Office for REDD+ Coordination in Indonesia where she provided technical inputs to the Government of Indonesia to support policy development on REDD+ and green growth. Felicity has an MPhil in Development Studies from the University of Oxford.

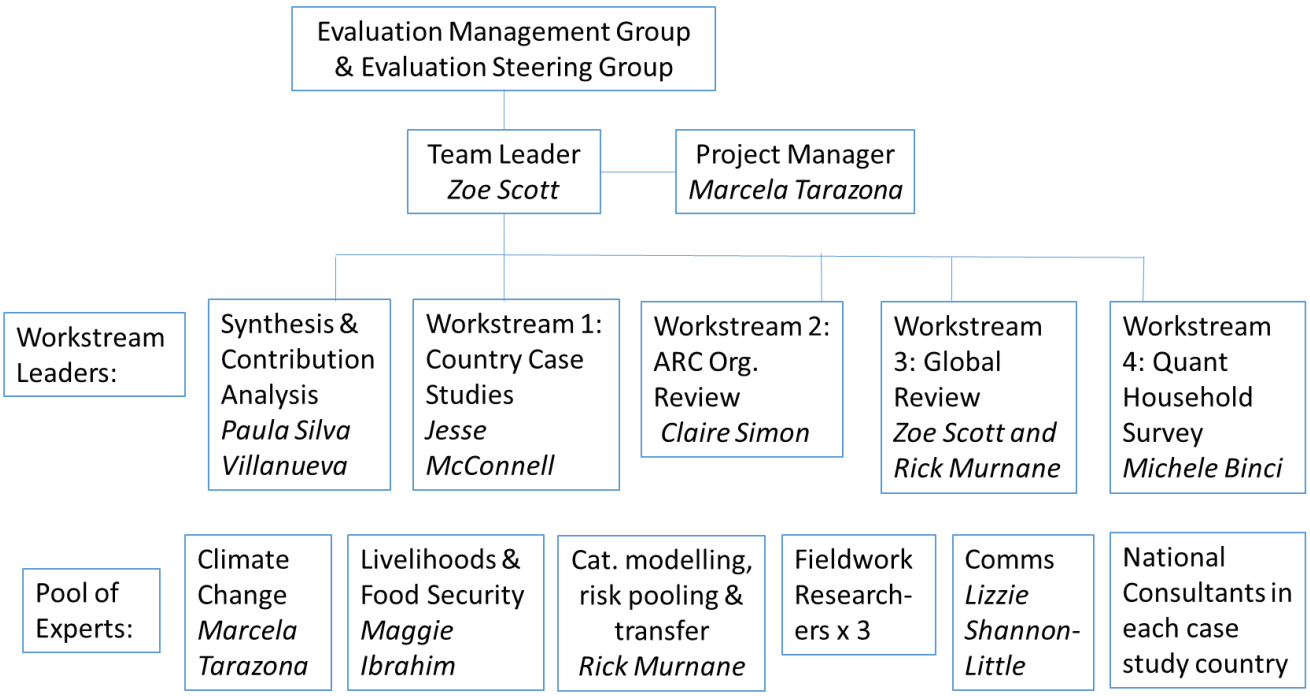
Ashira Perera to share the fieldwork with Felicity. Ashira Perera is a French-speaking Assistant Consultant at OPM and a PhD candidate in Development Economics at the University of Nottingham (U.K). She is currently working on a desk-based review of a nutrient fortification assessment coverage survey tool (FACT) which has been implemented in Nigeria, Senegal and Kazakhstan. She will be involved in the design and implementation of a new FACT survey tool which will be implemented in Nigeria, Burkina Faso and Pakistan in 2017. Ashira is also assisting with the project management of OPM's DFID-funded Kenya Extractives Programme (KEXPRO) during its Inception Phase. Ashira's technical expertise includes the design, collection and quantitative analysis of behavioural experimental and household survey datasets from developing countries, specifically Sri Lanka. Her doctoral research investigates how farmers cope with environmental risk.

Emilie Gettliffe to support fieldwork in Francophone countries. Emilie is a monitoring and evaluation specialist for innovative poverty reduction and economic development initiatives. She has expertise in planning, designing and conducting qualitative research using household assessments, key informant interviews and in-depth interviews. Emilie is bilingual and has experience of qualitative M&E data collection in a range of Francophone African countries (Senegal, Burkina Faso, Guinea etc) as well as Anglophone African countries (Ethiopia, South Africa, Tanzania). She also has a broad background in M&E and research, including household level data collection and also in relation to cash transfers, capacity building initiatives and food insecurity in Africa.

These changes mean that we can guarantee that all team members have sufficient capacity and time available for the next stage of the evaluation. All other team members remain the same as those proposed in the proposal document (pen portraits available in Annex **Error! Reference source not found.**). The team therefore retains broad sectoral expertise, covering disaster risk, climate change, social protection, catastrophe modelling, risk pooling and transfer, livelihoods, food security, and also has wide methodological expertise, including qualitative and quantitative methods, impact evaluations, surveys and knowledge management. The addition of Claire Simon and Zoe Scott allows the team to benefit from knowledge and expertise gained from other related projects: particularly the reviews of ARC undertaken by Kimetrica (conducted by Claire Simon) and case studies on Shock Responsive Social Protection (undertaken by Zoë Scott). The addition of Felicity, Ashira and Emilie allows us to ensure that we have adequate resources to conduct the extensive fieldwork that we have planned, and brings greater French language skills to the team which will be critical for the case studies in particular.

9.2 Governance and management arrangements

The team structure has been revised to make best use of the additional expertise and is now as follows:



The Team Leader will have overall responsibility for project delivery and resourcing and will report to the Evaluation Management Group for all issues concerned with management and delivery of the contract. Within OPM, the Team Leader also reports to Ed Humphreys, Portfolio Leader for Climate Change and Disaster Risk, and Simon Hunt, Managing Director, for the timely delivery of a high quality evaluation within budget.

The Evaluation Management Team (EMT) holds the contract with OPM and oversees day to day delivery of the evaluation and has final approval of all evaluation outputs. The EMT consists solely of DFID employees, including an Evaluation Adviser, a Programme Adviser and a Programme Manager as well as a Senior Responsible Officer and other DFID representatives, as required.

The Evaluation Steering Group (ESG) advises DFID on the relevance, quality and impartiality of the evaluation and encourages uptake of the findings within ARC. Members include representatives of all ARC's stakeholders, and they have been informally engaged during the evaluation design. Draft TOR for this group were developed by DFID, and it is expected that membership will include Representatives from ARC Member States, from the ARC Agency Board, from ARC Ltd, from ARC Agency, from DFID and from other donors including KfW and SDC. Together, this group is expected to represent a broad spread of sectoral and methodological expertise, various perspectives and be free from organisational influence and political pressure. It is expected that the ESG will:

- Comment on the inception report
- Attend Initial Findings Meetings for each of the Evaluation Reports
- Review and comment on the draft version of each Evaluation Report
- Review and agree each of the Final Evaluation Reports
- Agree DFID Management Response to each Evaluation Report
- Join annual update meetings on the progress of the evaluation

Evaluation outputs, including this Inception Report, will also be scrutinised by DFID's Evaluation Quality Assurance Service, which is an independent contracted service that is required to review

all DFID funded evaluation outputs for quality and robustness. The service will report directly to the EMT, providing independent advice.

The OPM Project Manager will assist the Team Leader in project delivery and resource management, ensuring that the evaluation stays on track in terms of workplan and budget. Each Workstream Leader will be responsible for managing their workstream, and will report to the Team Leader.

The core evaluation team is comprised of five experienced internal OPM staff members and five international external consultants. This approach has worked well during Inception, with regular communication between team members via email and weekly team meetings. In each case study country, we will also work with national consultants, who will be recruited once the case study countries have been finalised. All contracts with individuals and subcontracting companies have tightly defined Terms of Reference, which effectively holds experts accountable for their deliverables and ensure clarity around roles and responsibilities.

9.3 Capacity

The main rationale for changing the team structure was to ensure that there was sufficient capacity within the team. The Inception Period has experienced a lengthy delay and this has impacted on individuals' ability to perform future tasks against the revised timetable. With the new team structure and revised workplan in place, the Team Leader has held conversations with each team member to agree their availability and ensure that team members have reserved adequate time in their schedules. Should any unforeseen issues with availability arise, OPM has a wide range of internal staff members and external consultants with skills related to this evaluation, in particular in relation to impact evaluation, disaster risk, climate change and social protection. Drawing on this pool of expertise to supplement the team if necessary is an available option, subject to approval from DFID.

Zoë Scott has considerable experience of leading teams and managing staff for similar sized (and larger) evaluations and programmes. However, there is also considerable leadership capacity elsewhere in the team as we have a number of staff and consultants who have leadership capacity and have acted in Team Leader or Programme Management roles, including Marcela Tarazona, Claire Simon and Paula Villanueva.

In terms of institutional capacity, the evaluation team also benefit from:

- OPM's in-house M&E portfolio and all their related resources including documents on the intranet, reports from previous evaluations, quality assurance (QA) checklists and access to expertise.
- OPM's in-house Climate Change and Disaster Risk portfolio and all their related resources including documents on the intranet, reports from previous projects, technical seminars and access to expertise
- OPM's in-house Communications team who are able to advise on communications strategy development, website text, writing and pitching blogs, editing skills, newsletter development and M&E of communications activities.
- A dedicated Project Finance officer who will work with the Project Manager to ensure that the project stays within budget and that invoicing is handling promptly and efficiently.

- A dedicated Project Administrator who will work with the Team Leader and Project Manager to ensure that all project documentation, contracting, reporting and invoicing is dealt with accurately and efficiently.

In addition, we have a multi-layered QA system – for each individual assignment then the Team Leader provides regular monitoring and oversight of products and processes. We apply the ePact quality checklists, which are based on DAC quality standards. We also have an external QA panel who check every key evaluation product prior to submission to DFID and discuss necessary revisions with the Team Leader.

As shown on the team structure diagram above, in each country we will work with national consultants. Although not a specific requirement of the TOR, wherever possible we intend to make their participation a mutually beneficial experience. We will deliberately aim to build the capacity of local consultants by ensuring that all fieldwork starts with a participatory training orientation, running through the evaluation protocols, tools and methodologies. Each country case study will have a Fieldwork Leader and this person will be assigned as a contact point for the national consultants and will be able to answer any technical questions they may have. In addition, the Fieldwork Leader will identify suitably developmental opportunities for the national consultants, to ensure that skills and expertise are being stretched, rather than just utilised. The Fieldwork Leader will also review all national consultants' outputs and provide written or verbal feedback, with a view towards developing overall M&E skills in the country.

10 Risk management

A risk matrix was supplied in OPM's original tender document. This has been updated and refined during the Inception Phase – see Table 15 below.

Table 15: ARC evaluation risk matrix

Most significant risks	Impact	Likelihood	Mitigation measures
Technical			
Data on previous droughts may lack accuracy and robustness	H	M	We will use multiple sources of information and data to form a sufficient understanding of how drought has so far affected the countries in question.
Difficult to separate the effect of ARC pay-outs from other humanitarian efforts, and identifying a suitable counter-factual	H	M	For the qualitative research we will use contribution analysis to address the issue of attribution. The optional quantitative household survey (option 1) has been carefully designed to allow for the identification of a counterfactual.
Impact relates more to how efficiently govs implement their contingency plans rather than ARC activities and outputs	M	M	In the ToC we acknowledge this risk. We will specifically investigate how well ARC supported the development of realistic, operational contingency plans and how this process improves (or not) over successive pay-outs.
As the timing and location of ARC pay-outs cannot be predicted in advance it will be difficult to develop baselines and counter-factuals	H	H	The design of the optional quantitative household survey has been carefully designed to take into account different options if payouts do not materialise during the course of the evaluation in the case study countries. See Section 7 for more details and mitigation measures. However, the theory-based methods on which the bulk of the evaluation rests demonstrate other ways to develop robust findings around impacts at the household level (see Annex A4 for more details).
ARC will expand to more countries and additional perils but these plans and timelines are not yet confirmed.	M	L	ARC has moved forward with flood risk since the writing of the TOR. As specified in Section 5.10 on country selection criteria, we intend to conduct 1 or 2 case studies to investigate rapid onset disasters. We will remain flexible as to which countries should be considered as case studies and will update the workplan and timescales in liaison with ARC.
Difficult to engage some stakeholders	M	L	In developing our Stakeholder Engagement and Communications Strategy we have conducted a full mapping of likely stakeholders and considered the best options for engaging different groups, including workshops, newsletters and face-to-face meetings. During fieldwork we will use local partners with strong networks and political sensitivity to help us access particularly important key informants. We will also be able to mobilise DFID support in-country if needed.
Difficult to select case study countries	M	H	During Inception OPM has worked to establish criteria for case study selection, see Section 5.10 on country selection criteria. During Implementation we will work with DFID and ARC to identify the most suitable countries that fit the criteria.
Suitable, experienced	H	L	OPM has worked in many of the proposed case

national consultants are not found for the fieldwork.			study countries, and therefore has established relationships with national consultants and diverse networks that can be tapped into. We will ensure that fieldwork is planned well ahead of time to give us ample opportunity to find high calibre national consultants.
Evaluation tools inappropriate for context	M	L	We have built into the evaluation process opportunities for multiple team members to review the evaluation tools, and also for them to be adjusted if necessary following the first case study.
Difficulty in reaching consensus on key findings	H	L	We will ensure a rigorous validation process which takes into account potentially conflicting views and uses triangulation of evidence to produce consensus. We have built considerable time into the budget for validation of findings including national workshops involving senior, evaluation experts and review of reports by stakeholders.
Written outputs not high quality	H	L	The core research team are fluent English speakers with history of writing high level reports. We have an internal QA system involving senior evaluators who will review reports and ensure rigour in research methods.
Evaluation findings do not reach key audiences	H	L	A thorough Stakeholder Engagement and Communications Strategy has been developed during the Inception phase which will ensure that, via a range of comms mechanisms, the evaluation findings will reach different stakeholder groups. OPM has a designated Communications Team who will provide support throughout the life of the evaluation.
Managerial			
The long length of the evaluation means that certain key members of staff may no longer be available to conduct the evaluation.	M	M	It is very likely that there will be some staff turnover during the course of a 10-year project. Our proposal set out our staff retention strategy, which has been followed to date. We have reviewed the availability of the team during the development of the implementation workplan and have gained assurances from key team members of their availability for the tasks they have been assigned. To further mitigate this risk we will review the staff requirements after the first 4 years of the evaluation to ensure the right balance of skills and experience.
Security risk levels in some ARC countries may increase during the evaluation making fieldwork difficult.	H	M	We will, along with our external security contractor Spearfish, regularly monitor the risks in fieldwork countries and will maintain links with on-the-ground organisations providing real-time information. We will ensure the scheduling of trips avoids times of likely unrest (e.g. during elections) and will maintain back-up options for alternative case study countries should the situation on the ground prevent fieldwork in particular locations.
Project costs escalate.	H	L	The contract is a lump sum, so OPM are committed to absorbing any overspends. The Project Manager will work closely with OPM's Project Finance to ensure budgetary oversight is strong.

10.1 Conflicts of interest

We understand conflict of interest in this case to refer to the risk that if the Evaluation Team works as partners with ARC during the formative evaluation the Team will not be sufficiently 'objective' in later stages of the evaluation. The initial stages of the evaluation are formative to enable learning generated to inform how the ARC programme is delivered in order to lead to the desired impact as effectively as possible. To achieve this learning loop, the evaluation team must have a relationship with ARC stakeholders and in particular the ARC team. ARC stakeholders must be engaged in the design of the evaluation (validating the TOC, defining the evaluation questions to be answered), in reviewing evaluation reports, and in workshops that report the evaluation findings and set out/ discuss recommendations for increasing ARC's effectiveness and VFM. Engagement will be through workshops/ meetings/ interviews with ARC staff, participating countries, delivery partners and donors. There will also be interaction with ARC on strengthening M&E systems and data collection.

At the same time, the Evaluation remains an independent and impartial evaluation of the ARC Programme. To ensure that this impartiality and independence is protected, the evaluation team will adhere to the following protocols.

Protocols to protect impartiality and independence of ARC evaluation include:

DFID will:

1. Ensure that there are 1-2 independent members on the ARC Evaluation Steering Committee (they will review all evaluation outputs, and monitor that these protocols are fit for purpose)

The Evaluation Steering Committee will:

2. Record of any disagreements between steering committee members and the final decisions

The OPM Evaluation Team will:

3. Engage as wide a range of ARC stakeholders as possible during each stage of the evaluation (particularly evaluation design, ToC/outcome definition/validation)
4. Record this wide-ranging engagement.
5. Explain the evaluation's methodology and purpose to all ARC stakeholders to:
    a. help them understand the methods chosen and how they will generate robust findings
    b. build buy-in for the learning that will come out of the evaluation
    c. prevent stakeholders from challenging methods later.
6. Record justifications for changes to the evaluation questions
7. Focus on triangulating evidence for findings and recommendations where possible, to increase their robustness and ensure that findings are not just based on one point of view.
8. Ask that ARC stakeholders only comment on factual inaccuracies in draft reports, and do not challenge the findings of the evaluation.
9. Recognise that very regular engagement with the programme delivery partners can unconsciously reduce impartiality. Therefore, the Evaluation team will ensure that at least 1 member of the team has limited engagement with ARC (a non ARC-facing role) and is able to repeatedly challenge the independence/ integrity of the evaluation approaches/findings and also test the approaches/findings on wider OPM staff for feedback that approach remains impartial.
10. Clearly identify in all reporting the role of the evaluators in each step of the evaluation (both formative and impact), pointing out potential conflicts of interest (e.g. impact evaluation measuring the effectiveness of recommendations from the formative evaluation) and how OPM addressed these conflicts.

11 Future workplan and costing

11.1 Workplan

The two tables below set out our proposed workplan for all phases of the evaluation. Please note that timing of fieldwork, particularly the pilot and the Impact Phase case studies, may need to shift considerably to fit with the timing of ARC payouts. It is expected that case studies in the impact phase will be held shortly after the implementation of payouts has been completed.

Table 16: Formative workplan

Formative phase - component 1	Feb-17	Mar-17	Apr-17	May-17	Jun-17	Jul-17	Aug-17	Sep-17	Oct-17	Feb-19	Mar-19	Apr-19	May-19	Jun-19	Jul-19	Aug-19	Sep-19	Oct-19	Nov-19	Dec-19	Jan-20
Formative 1:																					
Planning and design	x	x																			
Case study 1		x	x																		
Case study 2				x	x																
Case study 3						x	x														
ARC org review	x	x	x	x																	
Global study					x	x	x														
Analysis and drafting							x	x													
ESG mtg and draft submission									x												
Revisions and final submission									x												
Formative 2:																					
Case study 4										x	x										
Case study 5												x	x								
Case study 6														x	x						
ARC org review												x	x								
Global study														x	x						
Analysis and drafting																x					
ESG mtg and draft submission																	x	x			
Revisions and final submission																		x			
Pilot IA:																					
Planning and design																			x		
Case study																				x	
ESG meeting and draft submission																					x
Revisions and final submission																					x

Table 17: Impact workplan

Impact phase - component 2	Feb-20	Mar-20	Apr-20	May-20	Jun-20	Jul-20	Aug-20	Sep-20	Oct-20	Feb-24	Mar-24	Apr-24	May-24	Jun-24	Jul-24	Aug-24	Sep-24	Oct-24
Impact Assessment 1:																		
Planning and design	x																	
Case study 1	x	x																
Case study 2			x	x														
Case study 3					x	x												
ARC org review		x	x															
Global study				x	x	x												
Analysis and drafting						x	x											
ESG mtg and draft submission								x	x									
Revisions and final submission									x									
Impact Assessment 2:																		
Case study 4										x	x							
Case study 5												x	x					
Case study 6														x	x			
ARC org review												x	x					
Global study														x	x			
Analysis and drafting																x		
ESG mtg and draft submission																	x	x
Revisions and final submission																		x

11.2 Budget

References / Bibliography

ARC Strategic Framework (2016). Unpublished document, latest version from April 2016

Balbach, E. (1999), *Using Case Studies to do Program Evaluation*. Stanford University & California Department of Health Services

BetterEvaluation website, available at: http://betterevaluation.org/en/themes/impact_evaluation

Clarke D., Dercon S., Dull Disasters, unpublished manuscript (14 December 2015)

Clarke, D. J., and L. Wren-Lewis (2016). *Solving Commitment Problems in Disaster Risk Finance*. The World Bank Group, Washington, DC.

DFID (2013), 'Africa Risk Capacity: Business Case'

DFID_b (no date), 'Africa Risk Capacity: Terms of Reference'

DFID (2011). *DFID Ethics Principles for Research and Evaluation,* https://www.gov.uk/government/uploads/system/.../**dfid**-**ethics**-prcpls-rsrch-eval.pdf,[15 Oct 2016]

Lemire et al (2012) Making Contribution Analysis Work: A Practical Framework for Handling Influencing Factors and Alternative Explanations, Evaluation 18(3) 294 –309.

McKinnon, M. and Hole, D. (2015), Exploring programme theory to enhance monitoring and evaluation in ecosystem-based adaptation projects. In D. Bours, C. McGinn, & P. Pringle (eds), Monitoring and evaluation of climate change adaptation: a review of the landscape. New Directions for Evaluation, 147, 49-60.

Mayne, J. (2008), *Contribution analysis: an approach to exploring cause and effect*, Institutional Learning and Change (ILAC) Initiative, Brief 16 (available at: http://www.cgiar-ilac.org/files/ILAC_Brief16_Contribution_Analysis_0.pdf)

Mayne, J. (2012), 'Contribution analysis: coming of age?' *Evaluation,* 18(3) 207-280. Sage Publications.

Morra, L., and Friedlander, A. (no date), *Case Study Evaluations*, The World Bank Group, available at: https://ieg.worldbankgroup.org/Data/reports/oed_wp1.pdf (15 Oct 2016).

Munter, C., Cobb, P., and Shekell, C. (2016) 'The Role of Program Theory in Evaluation Research: a consideration of the what works clearinghouse standards in the case of mathematics education', American Journal of Evaluation Vol 37(1), 7-26. Thousand Oaks, CA: Sage Publications.

OECD-DAC (2002), *Glossary of Key Terms in Evaluation and Results Based Management*, Paris: OECD (available at: https://www.oecd.org/dac/evaluation/2754804.pdf)

Scriven, M. (1967), The Methodology of evaluation. In R. W. Tyler, R. M. Gagne, and M. Scriven (eds.), *Perspectives of curriculum evaluation)* AERA monograph series on curriculum evaluation, *1*, 405-412). Chicago, IL: Rand McNally.

Scriven, M. (2016), Roadblocks to Recognition and Revolution, *American Journal of Evaluation,* Vol 37(1), 27-44. Sage Publications

Stame, N. (2004), 'Theory-based Evaluation and Types of Complexity', *Evaluation*, Vol 10(1), 58-76. Sage Publications.

Stern, E., Stame, N., Mayne, J., Forss, K., Davies, R., Befani, B. (2012), *Broadening the Range of Designs and Methods for Impact Evaluations*, Department for International Development Working Paper 38: April 2012.

The Economist (2016), Drought insurance in Africa. ARC's covenant: a worthy insurance scheme goes awry. 27 August 2016.

The Treasury Board of Canada (2012), 'Theory-Based Approaches to Evaluation: Concepts and Practices' (available at: www.tbs-sct.gc.ca).

van Aalst, M, Jan Kellett, Florence Pichon and Tom Mitchell (2013), *Incentives in Disaster Risk Management and Humanitarian Response*, Background note for World Development Report 2014

Varkevisser, C., Pathmanathan, I., Brownlee, A. (2003) Designing and Conducting Health Systems Research Projects, Volume 1: proposal development and fieldwork. World Health Organisation and International Development Research Centre.

Weiss, C. (1997), Theory-based evaluation: past, present and future. *New Directions for Program Evaluation*, *76,* 41-56.

White, H. and Phillips, D. (2012) 'Addressing attribution of cause and effect in small n impact evaluations: towards an integrated framework.' Working Paper 15, International Initiative for Impact Evaluation (3ie) (available at: http://www.3ieimpact.org/media/filer_public/2012/06/29/working_paper_15.pdf).

Annexes

All annexes are provided in a separate zip file attachment to this document. The directory of annexes is listed below.

A.1 Stakeholder engagement and communications strategy

A.2 TOC linkages and assumptions

A.3 Detailed evaluation questions

A.4 Participatory impact assessment

A.5 Options for a quantitative household survey (details)

A.6 CVs and pen portraits

A.7 TOR for ARC evaluation

A.8 ARC Theory of Change

A.9 Evaluation Budget

A.10 Sampling methodology