



Department  
for International  
Development



# Evaluating social transfer programmes

**Guidance for DFID country offices**

**June 2012**

**By Ranil Dissanayake, Zoe Stephenson and Matthew Greenslade**

# DFID guidance for evaluating social transfer programmes

## Contents

|   |    |
|---|----|
| 1. Why this guidance? .....   | 3  |
| 2. Evaluation and the project cycle .....   | 4  |
| 3. Evaluation in the business case – 11 steps.....  | 4  |
| 4. Purpose of the evaluation .....  | 5  |
| 5. Key users.....   | 6  |
| 6. Timing of evaluation activity.....   | 7  |
| 7. Key evaluation questions – what do you want to find out? .....                         | 8  |
| 8. Evaluation design and methods.....   | 16 |
| 9. How evaluation approach fits with the evidence base .....                              | 28 |
| 10. Use of monitoring data.....   | 29 |
| 11. Role of stakeholders.....   | 29 |
| 12. Budget .....  | 30 |
| 13. Contracting the evaluation team .....   | 31 |
| 14. Strategy to communicate findings .....  | 32 |
| Key contacts.....   | 33 |
| Key links and references.....   | 33 |
| Extended impact evaluation case studies of DFID-supported social transfer programmes..... | 34 |
| Glossary .....  | 42 |

## How to use this guidance

This guidance is for DFID advisers and staff, but will also be useful for contracted evaluation teams and international partners. The note focuses on DFID systems but we will at a later date explore internationalising the guidance with others. Not all of the technical detail needs be absorbed. Parts of Section 8 on methods are quite technical - the table under paragraph 8.9 summarises the main approaches for impact evaluation with examples). The guidance is a reference document to be used when issues and decision points for evaluations arise and parts of it may be best used by evaluation teams themselves. But it is important to be aware of the scope of evaluation methods that are available and currently being used. It is worth thinking about evaluations early if possible, so the right group of stakeholders can be involved from the start, and a baseline survey can be commissioned if need be for an impact evaluation. It is also important to be clear from the start on whether an evaluation is needed at all: the benefits should be weighed against the often significant costs. Evaluation methods have been reviewed in the guidance but the picture is changing all the time. We are planning to peer-review the guidance and issue a second version, most likely in 2013. To access evaluation reports, links and references are given at the end of the guidance, as are longer case studies of some current transfer evaluations.

## Acknowledgements

This guidance has benefited greatly from significant comments and contributions from a wide range of DFID and external experts, to whom we are very grateful, in particular: Alicia Martin (European Commission), Andy Hinsley (DFID), Ashu Handa (University of North Carolina), Benjamin Davis (FAO), Caitlin Tulloch (J-PAL), Christine Bockstal (ILO), David Seidenfeld (American Institutes of Research), Elisa

Cavacece (Irish Aid), Elizabeth Robin (DFID), Fred Merttens (OPM), Heather Kindness (DFID), Jessica Petitprez (DFID), Luca Pellarano (OPM), Matthias Rompel (GIZ), Naomi Hossain (IDS), Nick York (DFID), and Philip White (University of East Anglia).

## 1. Why this guidance?

**1.1 DFID is expected to do more and better evaluation of its programmes**, and the responsibility for commissioning evaluations now lies with staff in operational divisions rather than within DFID's Evaluation Department. This guidance aims to help staff feel confident to commission and manage high quality impact and process evaluations of social transfer programmes (programmes which transfer cash, food and assets).<sup>1</sup> It should be used in conjunction with general DFID guidance on evaluation<sup>2</sup> and aims to ensure evaluations commissioned by DFID reach international best-practice standards and in doing so supports the needs of partner governments and the international community.

**1.2 This guidance will inform DFID engagement with evaluation processes led by others (co-funders or implementing partners)**, as well as those led by DFID. In working on interventions in which DFID is a funding or implementing partner, working within existing aid effectiveness principles remains paramount. DFID needs to account for its spending but this will most often mean supporting partner government evaluation systems and processes. It is national policy objectives that programmes will usually be measured against, and it is national monitoring, evaluation and statistical systems that evaluations should support.

*"We will ensure much greater transparency and scrutiny of aid spending to deliver value for money for British taxpayers and to maximise the impact of our aid budget"*

*The Coalition: Our Programme for Government*

**1.3 Evaluations contribute to a rapidly evolving global evidence base on social transfers that is informing policy internationally.** DFID can contribute by using the full range of appropriate evaluation designs, and integrating rigorous quantitative and qualitative methods, including participatory approaches to ensure beneficiaries' voices are heard.

**1.4 This guidance complements the existing DFID VfM guidance on cash transfer programmes**, which is available on the Cash Transfers and Social Protection theme site, along with the most recent DFID evaluation

<sup>1</sup> Particular issues arising for public works programmes, for example valuing assets created, are not covered in this document but will be considered for the next version of the guidance.

<sup>2</sup> DFID guidance on evaluation is available internally at <http://dfidinsight/Other/Departments/EvaluationDepartment/Evaluationguidancetraining/index.htm> and can be made available externally if required.

reports on social transfer programmes, and externally.<sup>3</sup> While this guidance is aimed at DFID staff and systems, we will explore ‘internationalising’ it with our international partners.

**1.5 We need to maintain and build on the existing good quality of DFID social transfer programme evaluations.** DFID social transfer programme evaluations were recently praised by the UK National Audit Office (NAO) and UK Parliamentary Accounts Committee (PAC)<sup>4</sup>. But the PAC recommended that evaluations of DFID-funded social transfer programmes look more at long term effects and assess the impact of different transfer levels and NAO urged the collection of better information on costs and efficiency to assess value for money (VfM) with greater rigour.

**1.6 The following sections cover the main stages of evaluation design and implementation.**

## **2. Evaluation and the project cycle**

**2.1 Evaluations should be planned from the first stages of project design.** Wherever possible evaluators will be contracted before the programme begins and ideally the evaluators will be able to interact with the implementers as they make key programme design decisions. This will enable:

- *Sharing discussions on the theory of change, objectives and purpose* from early stages of the project design.
- *Opportunities for interaction* between evaluation and programme staff at the design stage, enabling use of robust evaluation designs which have implications for programme design (e.g. random selection of participants and phasing of programme rollout across sites).
- *Planning for timely evaluation* that will deliver qualitative and quantitative information when needed to feed into key decisions.
- *Collection of baseline data* before the programme begins - a key requirement for many impact evaluation designs.

## **3. Evaluation in the business case – 11 steps**

**3.1 All business cases now need to consider whether an evaluation will be conducted.** It is not necessary, or desirable, to evaluate every programme, but the decision needs to be clearly justified. Many DFID country offices have developed evaluation strategies identifying criteria for deciding if an evaluation is needed. Common criteria include: a weak evidence base, a contentious intervention, stakeholder interest, an innovative or pilot programme, and a high financial value.

---

<sup>3</sup> <http://teamsite/sites/policydivision/SocialProtectionNetwork/default.aspx>; externally <http://www.dfid.gov.uk/Documents/publications1/guid-dfid-cnty-offs-meas-max-vfm-csh-trsfr-progs.pdf>  
<sup>4</sup> NAO 2011 :[http://www.nao.org.uk/publications/press\\_notice\\_home/1012/10121587.aspx](http://www.nao.org.uk/publications/press_notice_home/1012/10121587.aspx); PAC 2012: <http://www.publications.parliament.uk/pa/cm201012/cmselect/cmpublic/1695/169502.htm>

**3.2 If an evaluation is planned then basic information needs to be provided in the business case on these points:**

1. Purpose of the evaluation
2. Key users
3. Timing of evaluation activities
4. Key evaluation questions – what do you want to find out?
5. Evaluation design and methods
6. How evaluation approach fits with the evidence base
7. Use of monitoring data
8. Role of stakeholders
9. Budget
10. Contracting evaluation team
11. Strategy to communicate findings

**3.3 This guidance provides brief generic information on each of these points but is more detailed on their application to evaluations of social transfers.** Only basic information is needed in the business case but DFID staff should think more deeply about each of these issues as they move forward with an evaluation.

## **4. Purpose of the evaluation**

**4.1 Not every programme needs to be evaluated.** Decide carefully if an evaluation is warranted – why do you want to do it? Evaluations may be relevant at the individual level, but may not be a priority strategically.

**4.2 Evaluations have three broad purposes** (and the scope and cost of the evaluation is likely to vary depending on which of the objectives prevail):

- **Accountability.** Many social transfer programmes are high-spend so there is a clear need to account to UK HMG and UK tax payers. In many cases we co-fund with partner Governments, so there will also be scrutiny from Parliament, media, tax-payers and the wider public in the country itself. If a key driver is to demonstrate whether the programme has ‘worked’ then an evaluation of impact will be important.
- **Learning and information for the local context.** Evaluation results should inform learning for programme improvement, extension or expansion, and decisions about whether and how best to roll out the programme (though there will also be learning opportunities for fully rolled out programmes). They may also feed into government or other partners’ decisions about whether to co-fund. A good evaluation leads to an understanding of why a programme does or does not have an impact, it is not necessarily limited to showing a ‘positive’ effect.
- **Learning for the global evidence base.** Evaluations influence wider approaches to social transfer programme design and implementation by contributing robust evidence. Aspects of social transfer programmes that we need to learn more about are highlighted in section 7.

***Purpose of the evaluation: Kenya's Hunger Safety Net Programme, monitoring and evaluation***

The Hunger Safety Nets Programme (HSNP) is a pilot social protection project. The HSNP is intended to reduce dependency on emergency food aid by sustainably strengthening livelihoods through cash transfers. Monitoring and evaluation aims to inform programme scale-up as well as the government's social protection strategy more generally. The evaluation includes a large-scale community-randomised impact evaluation conducted over 3 years, an assessment of targeting performance of three alternative targeting mechanisms, and qualitative research to assess targeting and impact issues that are less easily captured in the quantitative survey.

*For more details see the detailed case studies at the end of the guidance.*

## **5. Key users**

**5.1 The value of an evaluation depends on how well its findings are communicated and used** (see section 14 on communications). Some of the potential key users of evaluations (for any instrument, not just social transfers) are:

- Partner government's ministers and officials
- Programme managers and staff
- Beneficiary groups
- Local NGOs/CSOs
- Co-funding donors
- DFID country office staff, senior managers and Evaluation Department
- International researchers and organisations
- International NGOs
- Other partner governments
- Other DFID and development partner country offices

**5.2 Potential users can be engaged in a number of ways:**

- helping choose evaluation questions to ensure their needs are covered
- being part of a steering committee
- being consulted on evaluation methods
- being given timely information, for example within budget cycles or at donor decision points
- being given appropriately formatted reports (technical and non-technical reports may be required)
- getting early receipt of valuable information, such as baseline survey reports

**5.3 What constitutes credible evidence may depend on the audience.**

Beneficiary perspectives are an important, credible and perhaps under-used evaluation source to examine causal mechanisms and inform improvements in implementation. DFID is increasingly using participatory research methods (see section on participatory approaches at paragraph 8.34). DFID annual reviews now include a specific question on beneficiary views.

## 6. Timing of evaluation activity

**6.1 Evaluations should be planned to report, if possible, when information is needed to inform key decisions.** Evaluations need to fit programme and national policy and political cycles to maximise their usage. Formative or mid-term evaluations can also provide information to guide adjustments in implementation.

**6.2 Evaluations should be planned early.** Evaluation design should, if possible, be simultaneous with programme design. This will help avoid delays to the start of the programme while a baseline study is put in place or a compromised evaluation design. Experience suggests it can take up to six months between signing an evaluation contract to starting the baseline study.<sup>5</sup> As contracting via the Official Journal of the European Union (OJEU) process can take six months or more itself, around a year may be needed between tendering an impact evaluation to starting the baseline.

### ***Timing your surveys carefully: cash transfers in Zambia***

Timing of data collection is important because livelihoods in rural Zambia are tied to the farming cycle with certain periods of the year being more bountiful than others. Data need to be collected during specific times to control for these cyclical factors. Similarly, it is necessary to collect data before the rain season because roads become impassable in rural areas. It was not possible to collect data during the recent census or 2011 presidential elections, to avoid confusion about the purpose of the study and because of the potential for respondent fatigue. The evaluators, the American Institutes of Research, had to remain in constant communication with the government to plan the timing of data collection.

*Source: David Seidenfeld, American Institutes of Research. For more details see the case studies at the end of the guidance.*

**6.3 Within DFID, use of the new Global Evaluation Framework Agreement (GEFA) will reduce this time** as only mini-competitions will be required. It will comprise a number of pre-qualified suppliers who can be commissioned for evaluation work. Commissioners will be able to go to the framework and select suppliers from there based on their skills matrices and conduct a mini competition between them. GEFA should be up and running August 2012.

**6.4 There are a number of evaluation-related tasks that can take up significant staff time.** Some of the activities that could affect evaluation planning and timing are:

- Drafting evaluator ToRs
- Engaging stakeholders' on evaluation questions, via evaluators through their ToRs or directly

---

<sup>5</sup> Experience shared at a February 2012 cash transfers evaluation workshop in Kenya run by the Transfer Project (<http://www.cpc.unc.edu/projects/transfer/events/2nd-annual-research-workshop>) by the American Institutes of Research who are conducting an impact evaluation of social transfer programmes in Zambia.

- Managing OJEU processes or, within DFID, the new Global Evaluation Framework Agreement process (see paragraph 6.3)
- Reviewing bids
- Reviewing evaluator's inception report and finalising evaluation design
- Reviewing baseline, mid-term and end line reports
- Ensuring quality is assured through Evaluation Department's quality assurance (see below); other methods may also be appropriate, on which Evaluation Department can advise (this is mandatory)
- Drafting a management response to the evaluation (this is mandatory)
- Publishing (also mandatory), disseminating and discussing evaluation findings with users

#### ***Evaluation Department Quality Assurance (QA) of ToRs and final reports***

Evaluation Department will fund an external expert to provide comments on draft ToR, baseline surveys and final evaluation reports. They can access people with significant social transfer evaluation experience, including experience of randomised control trials, and we strongly recommend that you make use of this service. The QA reviewers use structured QA templates which you may also find helpful when drafting ToR. Please contact Evaluation Department or an evaluation adviser.

## **7. Key evaluation questions – what do you want to find out?**

**7.1 Evaluation questions will determine the choice of methods.** Questions need to drive methods, not the other way round. Some questions will focus on impact, some on process. Many social transfer evaluations will address both of these. A comment from the UK National Audit Office in a recent report was that DFID-supported programmes need better timely information on operational efficiency.<sup>6</sup>

**7.2 There will be many competing demands for evaluation questions.** But it is important to tightly prioritise questions to be addressed, because of limited space on survey forms. Identification of questions will be informed by:

- The programme purpose and theory of change
- Stakeholders needs, including beneficiaries
- Information needed at key decision points
- Global evidence gaps on social transfers
- OECD DAC evaluation criteria: relevance, effectiveness, efficiency, sustainability and impact (explained in DFID general evaluation guidance<sup>7</sup>)

### **Evaluating impact**

**7.3 Possible social transfer programme impacts that may be measured are set out below.**

---

<sup>6</sup> See footnote 4.

<sup>7</sup> See footnote 2.



### Impacts at various levels

| Individuals  | Households   | Communities   | Wider economy and polity   |
|--|--|---|--|
| <ul style="list-style-type: none"> <li>• Child health</li> <li>• Nutrition</li> <li>• Vaccination</li> <li>• Morbidity</li> <li>• School Enrolment</li> <li>• Attendance</li> <li>• Performance</li> <li>• Labour</li> </ul> | <ul style="list-style-type: none"> <li>• Consumption</li> <li>• Expenditure</li> <li>• Income generation</li> <li>• Asset accumulation</li> <li>• Food security</li> <li>• Migration</li> <li>• Production</li> <li>• Household decision making</li> </ul> | <ul style="list-style-type: none"> <li>• Local economies</li> <li>• Recipient networks</li> <li>• Social cohesion</li> <li>• Reciprocity</li> </ul> | <ul style="list-style-type: none"> <li>• Fiscal probity</li> <li>• Exchange rates</li> <li>• Social attitudes towards poverty</li> <li>• Inequality</li> <li>• Voting patterns</li> <li>• Political priorities</li> <li>• Social contracts</li> <li>• Institutions</li> <li>• State-society interaction</li> </ul> |

### Causal questions

7.4 **Causal questions in impact evaluation can take two basic forms.** One asks if the intervention resulted in a desired impact – the answer to which is either ‘yes’ or ‘no’. A second more challenging question is to explain the causal link, to demonstrate how the intervention caused the impact. Below are four key causal questions addressed in impact evaluations, the first two addressing the yes/no question and the second two the explanatory question. Different evaluation designs and methods will be better suited to each of these questions and this is further explored in section 8.

#### **Four key causal questions addressing impact**

1. To what extent can a specific (net) impact be attributed to the intervention?
2. Did the intervention make a difference?
3. How has the intervention made a difference?
4. Will the intervention work elsewhere?

### What are the specific areas of interest for social transfer impact evaluations?

7.5 **Key areas of interest for impact evaluations in countries where DFID works in sub Saharan Africa and South Asia, and globally, is where there is currently least evidence of impact.** A review of evidence for cash transfer programmes (including programmes with food and asset transfer components), in Africa and developing countries generally, is in DFID’s cash transfers literature review available of the Cash Transfers Social Protection theme site and externally and shown in summary below.<sup>8</sup>

<sup>8</sup> [www.dfid.gov.uk/r4d/PDF/Articles/cash-transfers-literature-review.pdf](http://www.dfid.gov.uk/r4d/PDF/Articles/cash-transfers-literature-review.pdf)

**Gradation of evidence for outcomes of cash transfers** (Source: DFID cash transfers literature review, 2011)

***Raising living standards of the poor***

- *directly* reduces poverty, hunger and inequality
- helps households sustain and improve livelihoods in the face of vulnerability and shocks

very consistent evidence

***Human development / human capital***

- improves quantity and quality of food consumption (child nutrition and development)
- helps households make use of health and education services (meeting access costs, reducing need for child labour and school dropout)

growing body of positive evidence

***Economic development and inclusive growth***

- facilitates structural reforms supporting long-term growth
- helps households to escape low risk, low productivity poverty traps
- frees up household savings for investment
- raises household spending with local multiplier effects and (in MICs) potential for fiscal stimulus role

strong logic, to date limited evidence

***Empowerment and gender equality***

- empowers women within households and communities
- empowers poor individuals and households to make their own decisions for improving their lives.

***Climate change and natural disasters***

- help reduce and mitigate risk of environmental shocks (e.g. through public works and diversification) and cope with shocks that do occur

***Facilitating social cohesion and state-building***

- reduce inequalities that contribute to social fragmentation, crime and political instability

**7.6 Areas where evidence of a positive impact is relatively rare are economic development and inclusive growth; empowerment and gender equality; climate change and natural disasters; and facilitating social cohesion and state-building.** There is significantly more evidence for impacts on living standards, food security and accessing health education services.

**7.7 Some questions that could be asked in transfer programme impact evaluations** in areas where there is a clear global evidence gap are:

- How do social transfers affect
  - investment in productive assets and activities;
  - local economic activity and markets;
  - empowerment of women;
  - empowerment of other excluded individuals;
  - dignity of beneficiaries and stigma
  - reduction and mitigation of risk of environmental shocks;
  - coping with environmental shocks;
  - social cohesion and networks;
  - informal social protection provision; and

- o political processes?

### **Empowerment of women**

Evaluations need to look at the impact on gender relations: are programmes reinforcing women's responsibilities for unpaid care or reproductive work or encouraging more gender equitable sharing of household responsibilities including child care. Part of this will be measuring household distribution effects. Evidence from Latin America, where the evidence base for social transfers is strongest, is inconclusive on this issue.

**7.8 Whether it is logical to address global information gaps depends on context and resources, but the contribution to global knowledge should at least be considered.** Some impact evaluations in Africa are also pioneering new areas. **Sexual activity** is being looked at in Kenya, Malawi and Zimbabwe); and **psycho-social status** and **mental health** in Kenya and Zambia (discussed in the April 2012 edition of the Journal of Development Effectiveness).

### *Economic impact*

**7.9 The economic impact of programmes at the individual, household, community and wider economy levels is important, especially given the often argued links between transfers and growth.** This is an area where the volume of evidence is still relatively light. Possible variables to look at include participation in labour markets, use of time, productive activities, non-farm business activities, access to credit (these may already be covered by national Living Standards Measurement Studies) and variables relating to climate change and resilience, and social networks of exchange.

**7.10 DFID's Research and Evidence Division is supporting a programme, managed by the Food and Agriculture Organisation, looking at the economic impacts of cash transfers,** called From Protection to Production (P2P). Countries involved include Lesotho, Malawi, Zambia, Zimbabwe and Ghana. An experimental impact evaluation of the Malawi Social Cash Transfer project found that unconditional cash transfers to ultra-poor households generated investments in agricultural assets, reduced adult participation in low skilled labour, and limited child labour outside the home while increasing child involvement in household farm activities.<sup>9</sup>

**7.11 The FAO From Protection to Production project is also evaluating local economy-wide multiplier impacts through computable general equilibrium (CGE) modelling.** This is being used as a complement rather than a substitute for impact evaluation.<sup>10</sup>

<sup>9</sup> See April 2012 Journal of Development Effectiveness at <http://www.tandfonline.com/toc/rjde20/current>

<sup>10</sup> See the reference in footnote 9 for a paper on CGE modelling of the local economic impact of social transfer programmes in Malawi and Ghana.

### ***The expansion of social transfer evaluations in Africa***

Most impact evaluations of social transfer programmes in developing countries in the recent past have been in Latin America, but this picture is changing fast. Apart from a large number of on-going evaluations in South Asia and elsewhere, a large number of impact evaluations are happening or are planned in sub-Saharan Africa. Recent, on-going and future cash transfer impact evaluations in Africa, listed in the Journal of Development Effectiveness, April 2012 are:

#### **Malawi** SCT

- Mchinji pilot, 2007–2009
- Expansion, 2012–2013

#### **Mozambique** PSA

- Expansion, 2008–2009

#### **Zambia**

- Monze pilot, 2007–2010
- Expansion and child grant, 2010–2013

#### **Ghana** LEAP

- Pilot, 2010–2012

#### **Tanzania** TASAFA

- Pilot, 2010–2011

#### **Uganda** SAGE

- Begins in 2012

#### **Burkina Faso**

- Pilot, 2008–2010

#### **Kenya** CT-OVC

- Pilot 2007–2011
- Expansion, 2012–2013

#### **South Africa** CSG

- Retrospective, 2010

#### **Ethiopia**

- PNSP, 2006–2010
- Tigray SPP, 2012–2014

#### **Lesotho** CGP

- Pilot, 2011–2012

#### **Zimbabwe** SCT

- Begins in 2012

#### **Niger**

- Begins in 2012

### ***The changing context for cash transfer evaluations in Africa***

This means that the context within which impact evaluations are being carried out is changing fast. As the introduction to the April 2012 Journal of Development Effectiveness, devoted to evaluating cash transfer programmes in Africa, explains. *“There are a number of unique features of the context in sub-Saharan Africa that have shaped the nature of CT programmes in the region. The region is characterised by a much greater level of risk and vulnerability, in large part due to the devastating effect of HIV/AIDS, which finds its global epicentre in Southern Africa. Beyond the pandemic the region suffers from more generalised levels of poverty, less developed markets and greater political instability. In contrast with Latin America and the Caribbean, at least in the short term, livelihoods and exit from poverty are inextricably linked to small holder agriculture and the informal economy, and less to engagement with the formal wage economy. Moreover, public institutions tend to be weaker in sub-Saharan Africa. Governments have less fiscal space available for social protection, and consequently, despite the Paris Declaration, multilateral and bilateral donors play a much stronger role in terms of promoting and determining policy – including requiring impact evaluations of pilot programmes.”*

Source: The Journal of Development Effectiveness, 3 April 2012, at

<http://www.tandfonline.com/toc/rjde20/current>

### ***Long term impacts***

#### **7.12 DFID was criticised by the UK Public Accounts Committee for not looking at long term impacts in social transfer programmes sufficiently.<sup>11</sup> Potential questions to look at are:**

<sup>11</sup> “Research following-up recipients over a number of years after their participation has so far only been done on one transfer programme. While the results of this longer term evaluation are encouraging,

- What are the long term impacts of social transfer programmes in terms of for example health, education and labour market outcomes?
- How do these impacts vary by gender?
- How persistent are impacts on transfer recipients after the recipient individual or household has graduated from eligibility?

7.13 **Examining these issues may require information gathering a number of years after programme participation**, or they may be gathered during an extended period of programme participation. Post-programme follow up surveys may be required. Additional datasets may need to be looked at, covering job status and incomes for an entire cohort of the population as well as transfer recipients. It may be possible to integrate specific modules on social transfers in national representative surveys that are collected at household level by national statistical offices.

7.14 **Resources and logistical arrangements will need to be set aside for this in the programme plan** and a judgement will need to be made on whether the effort and resources required is proportionate. Whether to measure long term impacts is a decision for development partners (usually partner governments) working with DFID country offices. If long term impacts are not measured the reasons for this decision should be explained.

7.15 **Some DFID supported programmes already measure or are planning to measure long term impacts.** Measuring long term impacts is built into programme planning in Zambia, is happening and about to be reported on in Kenya, and has already been done in Bangladesh and Ethiopia (see below).

7.16 **In addressing whether social transfer programmes targeted at women have an impact on gender relations in the longer term**, it may be worth measuring:

- the value assigned to women and to reproductive or domestic work by the state;
- men's involvement in childcare;
- women's capacities to engage in paid work;
- whether any programme conditions require harder work by women exceeding the benefits they receive in return.

***Measuring long term impacts in Ethiopia's Productive Safety Net Programme***  
 A report on a recent study of the Ethiopia Productive Safety Net Programme (PSNP) concluded "The authors find that it is only after 5 years of participation in the PSNP that the impacts on food security really kick in... If the average food gap is around 3-3.5 months per year, 5 years participation in the PSNP public works project reduces

---

overall there is less evidence for longer-term impacts."  
<http://www.publications.parliament.uk/pa/cm201012/cmselect/cmpublic/1695/169502.htm>

the gap by just over one month, or about one third of the food gap... Too often in development donors and governments give up because the short-term impacts are underwhelming. So the government and donors deserve credit for taking the medium term view with the programme (and with the evaluation)."<sup>12</sup>

### *Transfer level*

**7.17 The UK Public Accounts Committee also criticised DFID for not assessing the impact of different transfer levels.** “Larger amounts transferred to recipients are likely to have a more substantial and long-lasting effect and may therefore be better value for money, but the Department has not tested this hypothesis.”

**7.18 There may be scope for testing transfer levels through experimentation.** This should be considered, or a reason given for not taking this approach. Transfer levels may be more a design issue than for ex-post appraisal. In design, transfers values for new and existing programmes are not plucked out of the air. They are usually informed by the cost of average household monthly food/consumption needs. They are also generally a compromise between getting high impact and reaching the largest number of people.

### ***Assessing transfer levels in Ghana’s Livelihood Empowerment Against Poverty programme***

A baseline survey and modelling of impact for the Livelihood Empowerment Against Poverty programme (LEAP) published February 2012 found the value of the transfer at 7% of mean per capita expenditure to be one of the lowest in the world - the value of transfers in Kenya, Zambia, South Africa, Mexico and Colombia are 20-30% of mean consumption. So the expected impact of the programme was predicted to be low. This directly led to a decision by government of Ghana to triple the transfer size.

**7.19 Where it is possible and politically acceptable DFID is supporting the testing transfer levels** by development partners (for example in Ethiopia) and testing different levels of support for different groups (for example in Bangladesh). Where it is not possible politically (for example in Kenya and Zambia) it is not taking place. Potential questions to ask are:

- How do transfer levels affect programme impact?
- Is there evidence of poverty trap thresholds in the local context?
- Impact will tend to increase with transfer level, but is there an optimal transfer level that maximises programme returns in terms of the benefit to cost ratio?

**7.20 Assessing transfer levels can be done in programme appraisals** as well as in evaluations. Transfer levels can be assessed through ex-ante modelling as well as through experimentation using baseline survey data, as happened in Ghana.

<sup>12</sup> Lawrence Haddad on 22 May 2012, at [www.developmenthorizons.com](http://www.developmenthorizons.com). He is referring to the paper: Guush Berhane, John Hoddinott, Neha Kumar, Alemayehu Seyoum Taffesse (2011), *The impact of Ethiopia’s Productive Safety Nets and Household Asset Building Programme: 2006-2010*, International Food Policy Research Institute.

## Conditionality

**7.21 Another important research question is on conditionality**, for which potential evaluation questions are:

- How do effects of conditional and unconditional transfers differ (on both the outcome on which the transfer is conditional, such as school attendance, and other outcomes)
- Do these issues vary by gender?
- Does the impact of conditional transfers depend on the condition or could a similar impact be achieved by, for example, an effective information campaign?
- Is there a potential role for softer conditionalities?
- How do transfer levels interact with conditionality?

**7.22 As mentioned, the evidence on conditionality and other areas, for cash transfer programmes, is reviewed in DFID's cash transfers literature review** (see box) available of the Cash Transfers Social Protection theme site and externally.<sup>13</sup>

**7.23 This discussion of areas to look at and questions to ask in social transfer impact evaluations is not an exhaustive list.** It is intended to stimulate thought and discussion, but there may well be other areas of interest including those linking with the operational side, for example how the targeting process affects impact. Nor is it intended that impact evaluations should cover all areas mentioned. Not all questions can be answered and as mentioned, impact evaluations will have to vigorously prioritise.

### **Recent impact evaluations in DFID-supported social transfer programmes in Bangladesh and Kenya**

#### **Bangladesh**

*Impact evaluation for the Chars Livelihoods Programme 1 (CLP 1) finished November 2011*

**Findings:** 12,490 households (46,712 individuals) lifted above the extreme poverty threshold; seven-fold increase in value of productive assets among sampled households; raising of 90,684 homesteads above flood levels. DFID Bangladesh is responding to the methods used in the evaluation.

*Impact evaluation using differences-in-differences (see discussion of methods below) of Challenging the Frontiers of Poverty Reduction II (CFPR II) July 2011*

**Findings:** 42% increase in real per capita incomes for treatment group compared to 16% in control group; doubling of asset values for treatment group (after taking out programme assets transfers) compared to 60% increase for control; but no impact on education, at least in the short run.

#### **Kenya**

*Impact evaluation of the Hunger Safety Net Programme (HSNP) 2009-12 using a*

<sup>13</sup> [www.dfid.gov.uk/r4d/PDF/Articles/cash-transfers-literature-review.pdf](http://www.dfid.gov.uk/r4d/PDF/Articles/cash-transfers-literature-review.pdf)

*3,000 household randomised control trial. Final report expected July 2012*

**Findings:** after 1 year of programme, no significant impact yet on average consumption levels and poverty rates amongst beneficiary households compared to baseline, but decline in expenditure for control group suggests programme acting as safety net in context of severe drought in programme districts. No effect yet on accessing health and education

*Impact evaluation (new) (four year follow up) of the Cash Transfers Orphans and Vulnerable Children programme (CT-OVC) being published July 2012*

**Findings:** positive impacts on school enrolment in 2010 evaluation being maintained and positive impact on grade attainment is emerging for older students; new impacts also being found on young child health

Sexual behaviour: those that had not had sex in baseline, less likely to have had sexual debut, had fewer partners in last year and fewer unprotected acts

Limited economic impacts: small impacts on household productive activities and assets and no effect on wage income

New innovative areas being measured: subjective well-being indicators moving in positive direction, and psycho-social status improving (depressive symptoms reduced) especially for 15-19 age group

**Locating reports:** all published evaluation and baseline reports are available on the Cash Transfers and Social Protection theme site at and where they are not available on the web can be provided to anyone external on request.

## Evaluating process

7.24 **Many social transfer evaluations also ask important questions about process** (more ideas for measuring operation efficiency are on page 13 of DFID's VfM guidance for transfer programmes<sup>14</sup>):

- **targeting:** appropriateness of targeting criteria; inclusion and exclusion errors; transparency of process; barriers to access for the most vulnerable; local and national politics of targeting;
- **implementation:** fidelity to plan; variation across programme sites; factors affecting implementation; capacity of key implementers;
- **costs:** costs of inputs; indirect costs to recipients, communities and government; cost per unit of transfer or as a proportion of total programme cost i.e. cost-efficiency;
- **governance:** grievance procedures; local accountability processes; process for minimising rent seeking and leakage.

## 8. Evaluation design and methods

### Identifying the evaluation question

8.1 **Evaluation questions are the key driver in determining the choice of appropriate design and methods.** For example experimental and quasi experimental designs will be well suited to questions about whether the transfer programme had an impact and the scale of that impact. Theory based and participatory designs will be suitable for questions about how the social transfer had an effect and the causal mechanisms through

<sup>14</sup> See footnote 4.



which it worked. The table below provides a brief summary of some of the designs that may be most suitable for specific types of impact evaluation questions.<sup>15</sup>

| <b>Key Evaluation Question</b>   | <b>Possible designs</b>   |
|--|---|
| <i>To what extent can a specific impact be attributed to the intervention?</i> | Experiments<br>Statistical studies<br>Hybrids with 'Case' based and participatory designs   |
| <i>Has the intervention made a difference?</i>                                 | Experiments<br>Theory based evaluation (e.g. Contribution Analysis) see paragraph 8.38<br>Case-based designs (e.g. Qualitative Comparative Analysis) see paragraph 8.40 |
| <i>How has the intervention made a difference?</i>                             | Theory based evaluation<br>Participatory approaches   |
| <i>Can this be expected to work elsewhere?</i>                                 | Participatory approaches<br>Natural experiments<br>Synthesis studies  |

### Other factors affecting evaluation design

**8.2 Programme attributes are another key factor affecting programme design.** Some examples are given below of how programme attributes and other factors may have implications for evaluation design.

|   |   |
|---|---|
| <i>Programme Attributes</i> <sup>16</sup> | Complex programmes such as those with multiple components, inter-dependent causal strands and impacts which are hard to predict, are less amenable to experiments |
| <i>Sub-groups</i>                         | Evaluations interested in a specific subgroup (e.g. disabled people) need either large samples or surveys targeted to the specific group                          |
| <i>Politics</i>                           | Sometimes randomisation or phased rollout is politically unfeasible   |
| <i>Strategic priority</i>                 | Relatively high priority evaluations may justify expensive designs  |
| <i>Timing of programme</i>                | Sometimes partners are (legitimately) unwilling to delay implementation for a baseline survey   |
| <i>Data availability</i>                  | Sometimes data is too poor to know even who to sample for the evaluation  |
| <i>Users of results</i>                   | Which methods do potential users find credible?   |

**8.3 Ethical concerns may be a factor.** See DFID's Ethical Principles and further discussion of ethical issues in DFID Evaluation Handbook (chapter 4). There may also be potential harmful social tensions associated with creating a control or comparison group which partner governments wish to

<sup>15</sup> Broadening the Range of Designs and Methods for Impact Evaluations, Stern et al (2012), DFID Working Paper 38

<sup>16</sup> For a fuller discussion of programme attributes (including complexity) and their implications for evaluation design see chapter 5 of *Broadening the Range of Designs and Methods for Impact Evaluations*, Stern et al (2012), DFID Working Paper 38

avoid. Also, a trade off may then be faced between generating better evidence on impact and causality (which is in turn useful for improving project roll-out) and earlier coverage of individuals and communities in need.

**8.4 A participatory approach or ethos can be applied to evaluations.** This recognises the world looks different to different people, and explicitly seeks and values the perspectives of the people with least money and power. It involves seeking opportunities to involve beneficiaries throughout the evaluation cycle – e.g. in setting evaluation questions, discussing the theory of change, collecting and analysing data and interpreting findings and dissemination. A participatory *approach* to the overall evaluation design should be distinguished from the use of specific participatory *tools*, which are mostly qualitative in nature (see later discussion in this section).

**8.5 Secondary data sources for evaluations should be considered.** It may be possible to draw on national surveys or to include modules for transfer programmes on national survey questionnaires.

#### ***Focusing on sub-groups: Malawi Social Cash Transfer Programme***

The Malawi Social Cash Transfer Programme (SCT) programme objectives include reducing poverty and hunger in vulnerable households and increasing child school enrolment. At present, approximately 103,000 individuals benefit from the programme of which nearly two-thirds are children and nearly half are orphaned children – a sample large enough to generate robust results.

*For more details see the case studies at the end of the guidance.*

#### **Evaluation of impact**

**8.6 Evaluation of impact is not necessarily the same as impact evaluation (IE),** although the terms are often used interchangeably, and so the latter can have different definitions depending on the context. It can include any evaluation that systematically and empirically investigates the impacts produced by an intervention. More commonly the term is used more narrowly to refer to experimental designs which quantify impact and use a counterfactual to support claims that the intervention alone caused certain impacts (known as attribution claims).<sup>17</sup>

**8.7 Different approaches to impact evaluation use different bases for causal inference (connecting causes to effects).** These include:

- The difference between two cases which are identical in all ways except for the cause or intervention – the inference basis for experimental or quasi-experimental approaches;
- Association and analysis of multiple combinations of causes that lead to an effect – the inference basis for ‘configurational’ approaches;

<sup>17</sup> See *Impact Evaluation in Practice*, World Bank, Paul J. Gertler, Sebastian Martinez, Patrick Premand, Laura B. Rawlings, Christel M.J. Vermerrsch at <http://www.worldbank.org/pdt>; *Impact Evaluations and Development: NONIE Guidance on Impact Evaluation*, 2009 at [www.worldbank.org/ieg/nonie](http://www.worldbank.org/ieg/nonie); *Introduction to Impact Evaluation*, Patricia Rogers 2012.

- Identification and tracing of mechanisms that explain effects – the inference basis for ‘theory based’ and ‘realist’ approaches.

**8.8 Each approach has its strengths and weaknesses.** For example experimental approaches are good at answering the question ‘Has this particular intervention made a difference here?’ but weak on generalisation (external validity) and an explanation of *how* an intervention has had an effect, particularly where causal factors are inter-dependent.

### Impact evaluation approaches and examples from DFID country offices

| Method                    | Description  | Example of DFID country offices using in transfer evaluations   |
|---------------------------|--|---|
| <i>Experimental</i>       | Random selection of a treatment and control group, for example through Randomised Control Trials (para 8.18)   | Randomised Control Trial: Zambia, Kenya (HSNP) (see case studies at end of guidance)  |
| <i>Quasi-experimental</i> | Creation of a non-random control group, for example through <ul style="list-style-type: none"> <li>• Regression Discontinuity Design (para 8.23),</li> <li>• Propensity Score Matching (para 8.28)</li> <li>• Difference-in-Difference Estimation (para 8.30)</li> </ul> | Difference in Differences: Bangladesh (see box under para 7.23)<br>Differences in Differences with Propensity Score Matching: Kenya (CT-OVCs)*<br>Propensity Score Matching: Ghana*<br>Regression Discontinuity Design: Uganda* |
| <i>Non-experimental</i>   | No control group, impact investigated by other methods, for example through <ul style="list-style-type: none"> <li>• participatory approaches (para 8.34),</li> <li>• theory-based approaches (para 8.38)</li> <li>• case based approaches (para 8.40)</li> </ul>        | Participatory approaches: Kenya, Mozambique, Uganda, Yemen, OPTs (see box under para 8.36)  |

\* see case studies at end of guidance

**8.9 Impact evaluation designs can be categorised as experimental, quasi-experimental or non-experimental.**<sup>18</sup> Experimental methods involve random selection of a treatment and control group; quasi-experimental methods create or simulate a non-random ‘comparison’ group; and non-experimental designs do not use a control or a comparison group but investigate impact through other methods. The creation of the comparison or control group is a fundamental part of experimental and quasi-experimental designs, and a statistically rigorous way of creating a counterfactual. However while experimental and quasi-experimental methods are popular, and have been widely used for social transfer evaluations, they are not the only valid approach to impact evaluation.

<sup>18</sup> Some researchers do not use the category of quasi experimental and define evaluations as either randomised (experimental) or not (non-experimental). Other researchers view the term non-experimental as an inappropriate label for a wide range of approaches to causal inference drawing on decades of research in the social sciences.

Other significant non-experimental approaches include theory based, case based and participatory approaches.

- 8.10 To conduct most social transfer evaluations ‘hybrid’ designs are needed.** It is rare that a single design or method will be enough, particularly if the evaluation poses questions about both whether and how the intervention has had an impact.
- 8.11 All designs are likely to draw on both quantitative and qualitative methods of data collection.** While the heart of many experimental and quasi-experimental designs is quantitative data collection and analysis to assess the existence and size of programme impact, they also draw on qualitative data to look at causal pathways and help explain why impacts are or are not being seen and to throw light on whether results are replicable by analysing the programme in social, environmental and political terms. Key contributions of qualitative methods are: addressing questions of how and why the intervention had an effect; exploring outcomes which are less amenable to quantitative research methods, such as empowerment, discrimination, social conflict and social ties; and posing open-ended questions which create the possibility of discovering information we had not thought to look for.
- 8.12 For all designs there are key issues the evaluation team needs to address** including the approach to causal inference; mechanisms for quality assurance; threats to validity<sup>19</sup> and ethical issues. If a number of designs are possible, the evaluators should spell out the implications of each in terms of robustness of findings, cost and any operational issues (for example a delay to the start of the programme).
- 8.13 External validity refers to the extent to which the findings of the evaluation can be generalised to other contexts.** Experimental and similar impact evaluations are normally very context specific, and their findings have weak external validity. This can be increased by replication in new contexts, increasing confidence that results can be generalised further. Another way of increasing external validity is to use various arms of a transfer to examine alternative possible causal mechanisms in detail and use qualitative methods to inform judgments about how context specific these are. This may be a fruitful approach when the sample and extent of the programme is large, and can support many ‘sub-experiments’.
- 8.14 For experimental and quasi-experimental methods there are additional issues which the evaluation team needs to address:** selection bias (when those receiving the intervention differ in some systematic way from those in a control/comparison group); attrition;

---

<sup>19</sup> Validity threats include: internal, external, construct and statistical conclusion validity. See Ton G (2012). The mixing of methods: A three step process for improving rigour in impact evaluations. *Evaluation* 18 (1)

spillovers; and proposed sample sizes and how these vary for different effect sizes and associated power calculations.

**8.15 The trade-off between the sample size for sufficient power and the cost of the evaluation needs to be weighed up.** To do this, the evaluation terms of reference should ask the evaluation team, as part of their initial work, to explain sample size options and the power implications. This should include power calculations. Within DFID, findings can be discussed with the quality assurance panel in Evaluation Department and with statistical advisers.

**8.16 Attrition occurs when individuals or households move or are otherwise fundamentally altered during the course of the evaluation,** causing them to drop out of the study. While mitigation should be built into any evaluation plan, the fear is that attrition from the ‘control’ or ‘comparison’ group will be higher since there is no transfer acting to anchor the individual to the intervention area. If this effect is large enough, it may diminish the control group to the extent that the evaluation power is compromised – this should be addressed during the design phases.

### **Experimental designs for impact evaluation**

**8.17 Social Transfers are often well suited to experimental methods.** They tend to have a large number of direct beneficiaries (for random sampling) and so can support experimental evaluations based on surveys of large numbers of people – both those affected by the programme (the treatment group) and those not directly affected (the control group). Programmes that are just starting and where phased expansion will occur are particularly amenable to an experimental design (allowing random programme assignment across communities). For existing programmes, a new expansion wave is also an opportunity to design an experiment.

### *Randomised Control Trials (RCTs)*

**8.18 RCTs support high levels of confidence in attributing impact to the intervention and quantifying that impact,** and are less likely to produce biased or wrong results than quasi-experimental methods. RCTs use a random allocation process to choose treatment and control groups. Given a large enough overall sample, this random allocation should ensure that there are no significant baseline differences between the two groups – so in theory no baseline survey is needed. However in practice a baseline survey is generally done, enabling verification of the equivalence between treatment and control group and providing information on the starting point of both groups. Any difference in measured outcomes at the end of the surveyed period can be attributed to the treatment with a good degree of confidence. When coupled with additional qualitative investigation on key variables in the theory of change, insights into causal mechanisms can also be provided. RCTs require large sample sizes and

are therefore relatively expensive due to surveying costs (see paragraph 12.1 on budgets).<sup>20</sup>

**8.19 Typical social transfer RCTs randomly assign the transfer to individuals within a community (or to communities within a region),** as has been done in Zambia and Kenya in the Hunger Safety Net Programme (see case studies at the end of the guidance). The people (or communities) receiving the transfer are the treatment group; those not receiving any transfer are the control group. Evaluators conduct baseline surveys on key variables and points of interest regarding the causal mechanism. Follow-up surveys are undertaken and data on the key variables of interest analysed. If there is a systematic difference in the value of these variables between the treatment and control groups the difference in values can be assigned to the intervention. However randomization may not be perfect and significant baseline differences may remain between treatment and control groups. In such cases, interpretation of results may involve difference-in-difference analysis and econometric regression (discussed below).

**8.20 RCTs are a powerful way to answer intervention-specific questions about effectiveness, however as with all designs they work best under certain conditions.** These include:

- Where a ‘control’ group can be identified and contamination between treatment and comparison groups can be controlled;
- Where there are sufficient numbers to support statistical analysis and to provide mean differences or other statistical comparisons between the ‘treated’ and the control group;
- Where there is an interest in the success of a specific intervention in a particular setting rather than in wider generalisation.

**8.21 Experimental designs can raise ethical issues since they depend on some eligible people not receiving the transfer (at least for a period of time).** One response to this concern is to roll out the programme to other eligible individuals if the impact evaluations indicates that it has a positive effect. But ethical issues do not always arise. In Zambia a random selection of households and communities received popular support, being seen as more fair and transparent than leaving decisions to individuals at risk of rent-seeking behaviours.

### **Quasi-experimental designs for impact evaluation**

**8.22 Quasi-experimental evaluations create a comparison group through non-random methods.** The advantage of quasi-experiments is they avoid the logistical difficulties of randomisation. However even at their

---

<sup>20</sup> If a pre-existing source of data removes the need for additional surveying, RCTs need not be very expensive. This should be investigated if randomisation is possible in programme design but cost is an issue. Randomised designs also do not always have higher sample needs than non-experimental designs. Evaluators should be asked about sample size requirements for different designs. Cost differences should be investigated if this is an issue.

best, all have lower internal validity than RCTs. If the project is very novel or expensive, or of high political importance, additional confidence in evaluation results may be worth the extra cost and difficulty of randomisation.

### *Regression Discontinuity Design (RDD)*

- 8.23 One quasi-experimental approach is regression discontinuity analysis**, as has been used in Uganda (see case study at the end of the guidance). If eligibility for the transfer is set at a defined point, for example a certain income threshold or a proxy means score, and baseline data shows that the individuals or communities just above and just below the threshold are very similar, it is possible to analyse the ‘jump’ in the variables of interest across the threshold for eligibility for the transfer. For example, if those people just eligible for the transfer have, post-intervention, a 5% higher rate of educational achievement than those just ineligible for the transfer, and all other factors are roughly similar, we can attribute this difference to the intervention, though with less confidence than with an RCT.
- 8.24 Regression discontinuity analysis will not be a robust methodology under all circumstances.** If non-recipients change their behaviour by copying recipients, then though the transfer is effective, no ‘jump’ in outcomes will be recorded. For example, if treatment households send their children to school more regularly due to a transfer, there may be a social pressure for non-recipient households to also send their children to school more regularly so as not to be left behind. Here, RDD would find no difference between the households and wrongly conclude the transfer is ineffective, when it is very effective, even among non-recipients.
- 8.25 Where positive spillovers exist, as above, regression discontinuity will underestimate the impact of the transfer.** If there are negative spillovers (i.e. non-recipients do worse in an outcome when recipients exist) then regression discontinuity analysis will overestimate the impact. Because of the spillover issue some evaluators are reluctant to use RDD for social transfer programmes.
- 8.26 Another limitation of regression discontinuity design is internal validity.** Because this method focuses on data around a cut-off line, we cannot be certain that the observed results are true of recipients some way away from this cut off line. For example, if we focus on people just around a certain income threshold for eligibility and find only a small impact of the transfer, we cannot be sure the impact of the transfer isn’t much larger for poorer recipients further below the income threshold.
- 8.27 Regression discontinuity approaches also depend on close adherence to the ‘discontinuity’.** Since the approach depends on analysing individuals or households on either side of a cut-off point, this cut-off must be strictly adhered to. If in practice implementers allow some people just above the cut-off line to access the transfer (or disallow some

people just below the cut-off the transfer), evaluation results will be biased in a similar way to that observed when spillovers exist.

### *Propensity Score Matching (PSM)*

**8.28 Propensity score matching seeks to create a comparison group for the given treatment group.** It has been used in Kenya, in the OVC programme, and Ghana (see case studies at the end of the guidance). It involves analysing individuals from the treated and non-treated populations based on their observable characteristics and ‘matching’ individuals (or groups) from the non-treated and treated groups. This process of matching allows us to create a comparison group that appears similar to the treatment group in all respects apart from the transfer. Analysis is carried out to assess the effect of the transfer by comparing the matched-comparison and treatment groups.

**8.29 PSM can be a powerful evaluation design** and may be fruitful if an RCT is impossible. However, it has drawbacks: it requires a great deal of data to perform the matching; it requires a baseline survey; and it also requires that the data available do not omit any systematic factors that might cause a difference between the treatment and comparison individuals. If there are unobserved factors (such as motivation) that are key in determining the outcomes PSM could lead to biased estimates of the impact of the intervention.

### *Difference-in-Difference (D-in-D) Estimation*

**8.30 Difference-in-difference is a commonly used method of estimating programme impact and compares a treatment and non-randomly selected comparison group, before and after the intervention.** It was used in a recent evaluation in Bangladesh, see box below paragraph 7.26. D-in-D essentially compares the size of change in the outcome variable of interest between two sites (the treatment and comparison groups) and tests whether this difference is statistically significant. This design requires the existence of baseline data. It will not create a perfect comparison group – there will be baseline differences so careful assessment of the baseline survey will be needed. D-in-D, sometimes in the context of multivariate regression (which controls for other systematic differences), is then used to judge whether or not the programme had a significant impact.

**8.31 As for other quasi-experimental methods, this design has weaker explanatory power than an RCT.** If the evaluation compares a comparison area to the treatment area and uses simple difference-in-difference analysis, the comparison area needs to be chosen extremely carefully. Poor choice of comparison sites can fundamentally compromise the quality of the evaluation.<sup>21</sup> To compare pre- and post-treatment outcomes using regression analysis, a well-specified model is needed that

---

<sup>21</sup> This criticism has been made of the recent evaluation of the Millennium Village Project published in the Lancet in May 2012. See, for example, <http://blogs.worldbank.org/impactevaluations/the-millennium-villages-project-impacts-on-child-mortality>



captures all systematic effects on the variable of interest. If there are deficiencies in the model (and hence the counterfactual), the evaluation results may be open to question.<sup>22</sup>

## Non-experimental designs for impact evaluation

**8.32 Non-experimental designs have the power to explore how and why social transfer programmes have worked or not worked.** These are key questions which have been under-explored by evaluations using experimental or quasi-experimental designs. Non-experimental designs focus on explaining and exploring impact rather than quantifying it. They complement rather than replace experimental or quasi experimental evaluations.

**8.33 DFID recognises that non-experimental impact evaluation designs have been under-utilised and sometimes not conducted to a high standard.** DFID is therefore creating a fund to support greater use and refinement of these designs for impact evaluation. The designs DFID hopes to support include: *Participatory approaches*; *Theory Based Approaches*, employing methods such as Contribution Analysis, Process Tracing, Realist Evaluation and Congruence Analysis; *Case Based Approaches* employing methods such as Ethnography, Grounded Theory and Qualitative Comparative Analysis.<sup>23</sup>

### *Participatory approaches*

**8.34 Participatory approaches are extremely valuable for increasing our understanding of how beneficiaries and non-beneficiaries perceive the impact of social transfer programmes,** both positive and negative, and how they see its causal mechanisms working. They are important for exploring how power relations, social networks, reciprocity, social cohesion, gender dynamics and other intra-household dynamics have been affected by social transfers.

**8.35 They can also be a valuable means for empowering poor people and promoting the inclusion of their voices in political processes.** All social transfer impact evaluations should consider including participatory methods.

**8.36 However participatory approaches have been somewhat neglected in social transfer evaluations and this is a key area where DFID needs to improve.** The perceptions of beneficiaries about programme design, implementation and impact have not been systematically included in evaluation of social transfers, particularly in low income country contexts. The risks of not including beneficiaries'

---

<sup>22</sup> This also applies to randomised evaluations that use economic regressions due to imperfections in the randomisation process, which are common.

<sup>23</sup> More information can be found on these in Elliot Stern et al (2012) *Broadening the Range of Designs and methods for Impact Evaluations*, and Howard White and Daniel Philips (2012) *Addressing attribution of cause and effect in small n impact evaluations: towards an integrated framework*.

perspectives systematically in evaluation processes include: reduced programme efficiency and effectiveness, undervalued impacts, lost opportunities to strengthen empowerment and accountability and limited communication of results to non-technical audiences, particularly the public.

#### ***Participatory assessment of five DFID social transfer programmes***

DFID is conducting a study across five DFID programmes (Kenya, Mozambique, Uganda, Yemen and OPTs) using participatory techniques to gain more in depth understanding of cash transfer programme design, implementation and impact from the perspectives of beneficiaries, and their communities. The study will contribute to the evidence base and also lead to guidance on approaches for effective beneficiary and community participation in future programme monitoring and evaluation.

**8.37 Many of the methods used to gather data in a participatory way are qualitative.** For example tools for depiction of timelines, understanding social networks, ranking importance of different factors, power mapping, and tracing life histories. Newer methods include qualitative comparative analysis, most significant change, outcome mapping, and Sensemaker. There is also a growing body of work to demonstrate that participatory tools can be used to generate statistical data. For more information on participatory statistics contact the University of Reading Statistical Services Centre. Institutional mapping exercises are useful for understanding how social transfers fit within the universe of other sources of support.

#### ***Theory based approaches***

**8.38 Theory based approaches build on the assumption in the philosophy of science that plausible causal claims depend on the identification of mechanisms that make things happen.** Programme mechanisms take the step from asking whether a programme works to understanding what it is about the programme that makes it work. As well as the intervention's explicit activities there are other conditions variously understood as 'contextual' or 'helping conditions' that need to be in place. Mechanisms are not therefore 'context free' – they are contingent and depend on various starting conditions, supporting factors and predispositions to be sure that they will operate.

**8.39 These designs look for the connection between cause and effect through in-depth theoretical analysis,** rather than by demonstrating regularity or inevitability. Instead of comparing cases generative causality explains how causal factors interact. The causal explanation is not a matter of one element (X), or a combination of elements (X1.X2) asserting influence on another (Y); it is the association as a whole that is explained.

#### ***Case based approaches***

**8.40 Case based approaches emphasise context,** locating variables in the context of the 'case' and conducting within-case analysis alongside comparisons across cases. Cases may be policy interventions, institutions,

events or even countries during a particular historical period. This represents a shift from focusing causal analysis on variables taken out of their specific context and has opened up major new opportunities for causal analysis that are still little utilised in evaluation practice.

**8.41 These approaches are interested in generalising that goes beyond the single case - though not in 'universalising'.** They tend to generalise 'under certain conditions' and identify clusters or subsets of cases about which it is possible to make similar causal inferences. Case based causal inference may 'test' a theory and also contribute to 'theory building'.

**8.42 Qualitative Comparative Analysis (QCA) is a case-based approach which combines in-depth case studies with the identification and interpretation of causal patterns in the cases.** Each of the cases examined is conceived of as consisting of a configuration of relevant characteristics or conditions which produce a particular outcome. QCA sets out to compare the different combinations of conditions and outcomes pertaining to each case, with the goal of discovering what configurations of conditions lead to what outcomes, and which of those conditions are key in producing certain outcomes. It bridges qualitative and quantitative research and uses software to strip away elements not causally related to the outcome and identify the conditions which lead to either the presence or absence of an outcome.

### Process evaluation

**8.43 Process evaluations will often be conducted to complement an impact evaluation.** They can be used to understand how a social transfer programme has been implemented and experienced by different stakeholders, identify what worked well and what proved difficult, explore variations in implementation across different sites and assess what conditions are critical for success. They can also help investigate the extent to which implementation may have compromised impact. They can therefore help us to understand if lack of impact is due to a failure of implementation or a flaw in the theory underlying the intervention. They draw on qualitative and quantitative data including review of programme documentation, interviews with implementers and other stakeholders, and participatory evaluation approaches with beneficiaries. Suggestions on what may be measured in process evaluations is at paragraph 7.27 above.

### Mixing methods

**8.44 Mixing methods (i.e. qualitative and quantitative methods for data collection and analysis) is valuable for all types of evaluation designs.** We mix methods to gain better understanding, and more specifically for the following purposes:

- *Development:* using one method to inform the development/design of another (e.g. qualitative work to support good questionnaire design).

- *Complementarity*: using different methods to add information, explore different questions or different dimensions of the same questions, and aid interpretation of results produced by other methods. In particular, process problems (such as poor transfer delivery) can affect impact, but an impact evaluation alone might wrongly assume that the transfer has little effect, when the problem was that it wasn't delivered properly.
- *Triangulation*: when one method corroborates or contradicts, refines or 'adjusts' the findings of another. If multiple methods or sources all point towards the same result, more confidence can be placed in it.
- *New ideas*: when different methods provide contradictory information, this divergence can be fruitful in helping us re-think our assumptions, and reach new insights.

### **Mixing Methods in Uganda Social Assistance Grants for Empowerment impact evaluation**

Impact Evaluation of the pilot SAGE cash transfer programme adopts a mixed method approach.

- A quantitative household panel survey uses RDD to estimate impact on material deprivation, economic security, social exclusion and access to services.
- The qualitative research focuses on impact on empowerment and social cohesion.
- The operational effectiveness assessment will draw on both indicators generated through the quantitative and qualitative research and indicators provided by the SAGE programme's internal operational monitoring systems.

*For more details see the case studies in the annex*

**8.45 Qualitative and quantitative components of an evaluation should be developed together to interact and influence each other.** To quote a recent investigation of the connection between quantitative and qualitative analysis, "while quantitative methods produce data that can be aggregated and analysed to describe and predict relationships, qualitative research can help to probe and explain those relationships and to explain contextual differences in the quality of those relationships. Qualitative research is able to use social analytical frameworks to interpret observed patterns and trends—including analysis of socially differentiated outcomes—and to analyse poverty as a dynamic process rather than a static outcome".<sup>24</sup>

## **9. How evaluation approach fits with the evidence base**

**9.1 The evidence base for a planned intervention will inform the selection of evaluation questions.** As discussed in section 7, the evidence base for social transfer interventions requires strengthening in some areas. If there are knowledge gaps or weak and contested evidence there is call for a more ambitious, externally valid evaluation. This benefits policy makers and implementers within DFID and beyond, enabling better future programme design.

<sup>24</sup> See GSDRC Issues Paper on Quantitative and Qualitative Methods in Impact Evaluation, <http://www.gsdrc.org/docs/open/EIRS4.pdf>

## 10. Use of monitoring data

10.1 **Below are some factors to consider in collecting monitoring data (this is not an exhaustive list).** Social transfer evaluations will typically require significant data collection. Whether the evaluation considered is randomised or not, generally a baseline survey, a mid-term survey, a completion survey and a follow-up survey may need to be conducted; further, national data collection or compilation may also be necessary. Data also need to be quality assured and checked for consistency. Evaluators should seek to minimise the burden of data collection.

- Be clear about **what is being measured** and what information is needed;
- Ensure that the **minimum data requirement** to answer the questions posed is collected;
- If reliable data is collected in a useable format through **pre-existing mechanisms**, consider using these, or investing to strengthen existing processes, especially when run by national institutions;
- **Use the data collection process to support national processes** for data collection and monitoring and evaluation as far as possible;
- Where possible integrate data into an automated **Management Information System (MIS)**. An ICT-based single registry with recipient ID and appropriate delivery systems can facilitate not only evaluation but real-time monitoring of the programme.<sup>25</sup>

10.2 **Programme outputs need to be monitored and evaluated as well as impact**, so that shortcomings in programme design can be distinguished from shortcomings of implementation. In some cases the main reason for lack of impact is poor implementation, and this needs to be identified clearly.

10.3 **It may also be worth systematically monitoring aspects of the changing context in which the programme is being implemented.** Possible links between the changing context and the programme can be explored and findings then fed into evaluation and learning. This is particularly useful for evaluating the policy influence component of programmes and without it the attribution of change to the intervention risks becoming myopic and unreliable.

## 11. Role of stakeholders

11.1 **Internal and external stakeholders should be engaged in generating and answering evaluation questions.** These include partner government's ministers and officials, programme managers and staff,

---

<sup>25</sup> For more on connecting monitoring and evaluation data see <http://www.interaction.org/sites/default/files/Linking%20Monitoring%20and%20Evaluation%20to%20Impact%20Evaluation.pdf>.

beneficiary groups, local NGOs/CSOs, co-funding donors, DFID country office staff, senior managers and Evaluation Department, and international researchers and organisations (if global information gaps are being addressed, especially if they are providing support or funding).

**11.2 Strong local ownership of the evaluation is important.** This will allow lessons to be assimilated by the Government more easily. There may also be important opportunities to help build government capacity and therefore sustainability. Most often it will be partner government evaluation systems and processes that we are seeking to support and the development of which is one of the ultimate objectives.

## **12. Budget**

**12.1 Evaluation budgets of around 1-5% of programme spend are usually appropriate according to the general DFID evaluation guidance.** Within DFID Evaluation Department can advise on budgets as part of the quality assurance service they provide for evaluation terms of reference. They do not operate a hard and fast rule on costs because they should and do vary significantly with the type of programme.

**12.2 Evaluation costs may be higher for pilot schemes and reduce as the programme scales up.** They may also vary:

- If you need to collect survey data at household level, either because you want to run an impact evaluation or it makes sense for monitoring outcomes;
- If it is an area where there are no existing evaluations so the methodology has to be thought through from scratch;
- If there are no existing evaluation frameworks and indicator sets to draw on, so you have to design them from scratch;
- If the evaluation is being done in a country where security and or logistics are particularly difficult for getting around.

**12.3 To take one example of a social transfer impact evaluation from Africa,** the opinion from the American Institutes of Research who are managing an evaluation in Zambia, at a February 2012 workshop in Kenya workshop on impact evaluation,<sup>26</sup> was that data collection in Africa can cost up to \$350k per survey round. Assuming three surveys (one baseline and two follow ups) with \$500k for other costs, the overall cost can reach \$1.5m or more with follow-up surveys to measure long run impacts. Such costs can be justified as if the intervention lacks a strong supporting evidence base (which will usually be the case). It increases the importance of producing high quality outputs in a usable format for all stakeholders, as well as high quality terms of reference, and contracting and management arrangements.

---

<sup>26</sup> <http://www.cpc.unc.edu/projects/transfer/events/2nd-annual-research-workshop>

## 13. Contracting the evaluation team

13.1 **Terms of reference for an evaluation need to provide substantial detail on the approach and design.** For simple evaluations terms of reference can be drafted within country offices then sent for quality assurance to DFID's Evaluation Department. Other approaches to quality evaluation may be appropriate on which Evaluation Department can advise. For a complex evaluation with a range of possible questions and designs, offices may first need to commission an 'approach paper' or 'design paper' to explore design options and methodological issues. The DFID Evaluation Department Handbook provides example terms of reference for evaluations and for approach and design papers, and the DFID Cash Transfer and Social Protection theme site provides links to terms of reference for DFID-funded social transfer evaluations.<sup>27</sup>

13.2 **Independence is important.** Evaluations cannot be done by staff involved with the programme. In terms of commissioning, it is recommended that partner governments and semi-government agencies commission evaluations if possible rather than donors, for sustainability reasons, even if there is a cost on how quickly it's done. Contracted companies responsible for implementing programmes should not commission evaluations because of conflicts of interest.

13.3 **Suitable governance arrangements are needed to ensure the quality, integrity and credibility of an evaluation.** These should be detailed in the evaluation terms of reference. Possible arrangements include:

**Management Group.** This is a **decision making** body often created for large multi-donor evaluations. The group will consider budget, evaluation design, proposed fieldwork and draft reports. It is likely to include people with evaluation experience from the donor(s), partner government and CSOs but not anyone directly involved in implementation. It may operate alongside a reference group.

**Reference Group.** This is an **advisory** group whose role is to provide a route for stakeholder engagement, ensure an evaluation's future utility and utilisation, and helps build support for an evaluation. Such groups may be useful where there are a range of stakeholders who do not have evaluation expertise but have a strong interest in the evaluation.

**Steering Group.** This is a **hybrid** of a Management Group and a Reference Group which combines the functions of advice and decision-making. Some, but not all, representatives should have evaluation experience and external stakeholders should be included.

---

<sup>27</sup> See 'Key links and references' at the end of the guidance for internal DFID sites – documents can be made available externally if this is of use to partners.

13.4 **None of these groups should include those directly involved in the implementation of the intervention being evaluated.** All should ideally involve a DFID evaluation specialist from outside the commissioning office or department. DFID’s Evaluation Department can advise on potential candidates.

## 14. Strategy to communicate findings

14.1 **A strong strategy for communicating findings is likely to be required.** The terms of reference should address usability by stakeholders, and should ask for succinct and tailored evaluation outputs and clear visual presentation of data. Social transfers are a rapidly evolving field and new evaluations should aim at wider lesson learning as well as feeding back on individual programme performance. A communications strategy for various audiences should be considered (see table). Examples of evaluation reports can be found on the DFID Cash Transfers and Social Protection theme site and outside DFID, for example the J-PAL and Transfer Project sites – references are in ‘Key links and references’ below.

### Key stakeholders and communication needs<sup>28</sup>

| Country Government  | DFID Management                                   | Project Implementers   | Wider Research Community   |
|---|---|--|--|
| Process and impact reports for technocrats and politicians  | Simple, short summaries of impact and VfM metrics | Early information on process to inform programme refinements         | Academic research, use international knowledge sharing platforms |
| <b>Other country offices</b>  |   | <b>Beneficiaries</b>   |  |
| Dissemination of lessons through adviser retreats, communities of practice and social protection theme site |   | Results and proposed improvements disseminated, perhaps through CSOs |  |

<sup>28</sup> International Knowledge Sharing Platforms include, for example, the ILO Social Security Inquiry, the GESS platform (<http://www.social-protection.org/gimi/gess/ShowWiki.do?wid=9>) or the Transfer Project (<http://www.cpc.unc.edu/projects/transfer>).



## Key contacts

Zoe Stephenson, Evaluation Adviser, Evaluation Department

Matthew Greenslade, Economic Adviser, Poverty and Vulnerability Team

Evaluation queries should be discussed with **evaluation and results advisers** that are now present in many DFID offices and in regional directorates. They can advise on all areas including the most appropriate ways to procure high quality evaluation services.

## Key links and references

### *Within DFID*

**Social transfer evaluation reports**, ToRs and baseline studies for DFID supported programmes are available on DFID's Cash Transfers Social Protection theme site (internal only) at <http://teamsite/sites/policydivision/SocialProtectionNetwork/default.aspx> and can be provided externally on request.

The theme site also contains **DFID's VfM guidance** for cash and asset transfer programmes and **DFID's literature review** of evidence on cash transfer programmes

**DFID's general evaluation guidance** is available internally at

<http://dfidinsight/Other/Departments/EvaluationDepartment/Evaluationguidancetraining/index.htm>

### *Outside DFID – evaluation reports*

Details of social transfer impact evaluations in Africa are available from the **Transfer Project**, at <http://www.cpc.unc.edu/projects/transfer> and in the April 2012 edition of the Journal of Development Effectiveness at <http://www.tandfonline.com/toc/rjde20/current>

J-PAL has a database of cash transfer evaluations using randomised control trial methods at [www.povertyactionlab.org/search/apachesolr\\_search/transfers?filters=type:evaluation](http://www.povertyactionlab.org/search/apachesolr_search/transfers?filters=type:evaluation)

A **World Bank** Independent Evaluation Group (IEG) review of the impact of social safety nets is at <http://ieg.worldbankgroup.org/content/ieg/en/home/reports/ssn.html>

### *Outside DFID – advice and support*

For **World Bank** guidance on evaluating social transfer programmes see *Impact Evaluation in Practice*, World Bank 2011, Paul J. Gertler, Sebastian Martinez, Patrick Premand, Laura B.

Rawlings, Christel M.J. Vermersch at <http://www.worldbank.org/pdt>

Support on impact evaluations available from **3ie** is at [www.3ieimpact.org](http://www.3ieimpact.org)

Additional resources are available from the World Bank **Network of Networks on Impact Evaluation** including a freely available impact evaluation guide *Impact Evaluations and*

*Development: NONIE Guidance on Impact Evaluation*, 2009 at [www.worldbank.org/ieg/nonie](http://www.worldbank.org/ieg/nonie)

On qualitative approaches to evaluation see Patricia Rogers at [www.betterevaluation.org](http://www.betterevaluation.org)

## Extended impact evaluation case studies of social transfer programmes

### ***Ethiopia: Tigray Social Cash Transfer Pilot Programme***

This cash transfer programme initiated by Tigray regional state and UNICEF aims to reduce poverty, hunger and starvation in extremely poor, labour constrained households, with particular attention on elderly and child headed households, female headed households, and households with people with disabilities. The programme also seeks to enhance household access to essential social welfare services such as health care and education via access to schools. A total of 169,540 (4% of the regional population) are expected to benefit directly or indirectly during the pilot phase. The initial disbursement took place in September 2011, and transfers will continue for two years as a pilot. On average 14 USD per month will be paid every four months by beneficiary cards. The transfers will vary according to household size: 6, 9, 12 and 15 USD will be given to households with 1, 2, 3 and 4 members respectively, with an additional 1.5 and 3 USD for each child enrolled in primary and secondary school respectively.

One urban and one rural woreda (district) were purposefully selected for the pilot: Abi Adi town and Hintalo-Wajirat woreda respectively. All residents of Abi Adi that met selection criteria were eligible for the programme, while in Hintalo-Wajirat only 7 of 22 purposefully selected tabias (ward) were included. Subsequently an 8th tabia was included, where payments had not yet begun as of June 2012. The programme will be targeted via local community care coalitions (CCCs) at the Tabia level, which act as an entry point and support mechanism for especially vulnerable households in the community. The CCCs first identify households that meet the extreme poverty ('poorest of the poor' and not able to meet essential needs) and labour constrained (no able-bodied labour aged 19-64, or dependency ratio of greater than 3) criteria, and then rank these households by neediness. Beneficiaries are then selected from this ranked list.

The objective of the impact evaluation is to provide evidence as to the viability of the expansion of the programme to other regions of the country. The evaluation strategy aims to evaluate the impacts of the cash transfers on the socioeconomic conditions of the beneficiary families, such as food and non-food consumption, as well as on a number of human development indicators such as school enrolment and attendance, health and nutrition outcomes, and child labour, and employment effects and its contribution to economic growth. The evaluation will also focus on targeting and operational effectiveness.

As is evident from the operational implementation of the pilot, the impact evaluation design is quasi-experimental. Treatment households will be randomly selected from beneficiary lists. Comparison households will be taken from treatment communities. These will be randomly drawn from those eligible households not selected into the program. For both the treatment and comparison groups, elderly-, child- and female-headed households, as well as

households with a disabled member, will be oversampled. A final group of non-eligible household will be randomly drawn tabia lists.

Besides the non-random assignment of the programme, a second challenge of this impact evaluation is that that transfers began in September, 2011, nine months prior to baseline. Where possible, this will be addressed by the inclusion of retrospective questions in the baseline survey. Overall, given these challenges, the consortium has proposed using Difference in Difference combined with PSM as well as possibly RDD to identify impact. After the baseline in June, 2012, shorter monitoring surveys will be fielded later in 2012, 2013 (three times) and 2014, with a final household survey in June, 2014.

(Source: Ben Davis, Food and Agriculture Organisation.)

### ***Ghana: Livelihood Empowerment Against Poverty programme***

The Ghanaian Livelihood Empowerment Against Poverty (LEAP) programme provides cash and health insurance to extremely poor households to improve short-term poverty and encourage long-term human capital development. LEAP started a trial phase in March 2008 and then began expanding gradually in 2009 and 2010, and currently reaches over 35,000 households across Ghana with an annual expenditure of approximately USD11 million. The programme is fully funded from general revenues of the Government of Ghana, and is the flagship programme of its National Social Protection Strategy. On average 22.5 USD per month is paid quarterly. The transfers vary according to household size: 15, 18, 21 and 27 USD will be given to households with 1, 2, 3 and 4 members respectively. In addition, beneficiary households are linked free of charge to Ghana's national health insurance programme.

The LEAP programme operates in all 10 regions of rural Ghana. Within regions, districts are selected for inclusion based on the national poverty map; within districts, local government offices choose communities based on their knowledge of relative rates of deprivation.

The impact evaluation takes advantage of a nationally representative household survey implemented during the first quarter of 2010. The initial treatment sample of 700 households were randomly drawn from the group of 13,500 households that were selected into the programme in the second half of 2009, and are located in 7 districts across 3 regions. These households were interviewed prior to receiving any indication that they had been selected for LEAP. The baseline survey instrument was a reduced version of the national household survey instrument, and the national survey sample and the treatment household sample were surveyed at the same time. The strategy is to draw the control households from the national survey using PSM techniques. A comparison group of 'matched' households was selected and re-interviewed after 2 years, in April-May, 2012, along with LEAP beneficiaries to measure changes in outcomes across treatment and comparison groups.

Handa and Park (2012) carried out PSM analysis to identify the comparison group from the sample and assessed the appropriateness of this comparison

group for the evaluation strategy. Their results show that for the outcome variables of the original evaluation design (demographic and child welfare), the PSM strategy works well. For the purposes of the DFID-supported From Protection to Production project (see paragraph 7.12), however, this sample of control households is not optimal. There is little that can be done to change this sample, since the PSM corresponds to the original objectives of the evaluation. One practical solution would be to rerun the PSM ex-post once we have obtained the collected second round data, hoping that there is an improvement in the matching. A second strategy will be to re-estimate the matching using alternative methods, such as propensity score weights. (Source: Ben Davis, Food and Agriculture Organisation.)

### ***Kenya: Hunger Safety Net Programme***

The Hunger Safety Nets Programme (HSNP) is a pilot social protection project conducted in the Arid and Semi-Arid Lands (ASALs). The ASALs are extremely food insecure areas which have experienced recurrent food crises and food aid responses for decades. The HSNP is intended to reduce dependency on emergency food aid by sustainably strengthening livelihoods through cash transfers. OPM is responsible for the monitoring and evaluation (M&E) component of the programme, with the intention of informing programme scale-up as well as the government's social protection strategy more generally. The M&E involves: a large-scale rigorous community-randomised controlled impact evaluation household survey (RCT) conducted over 3 years; an assessment of targeting performance of three alternative targeting mechanisms (Social Pension; Dependency Ratio; Community-based Targeting); qualitative research (including key informant interviews and focus group discussions) to assess targeting and impact issues that are less easily captured in the quantitative survey; and on-going operational and payments monitoring to aid implementation of the programme. Findings are communicated to the HSNP Secretariat, Government of Kenya and DFID on a regular basis to inform and advise policy revisions and development. (Source: Fred Merttens, OPM)

### ***Kenya: Cash Transfers for Orphan and Vulnerable Children***

The Kenya Cash Transfer Programme for Orphans and Vulnerable Children (CT-OVC) is the Government's flagship social protection programme, reaching over 125,000 households as of the end of 2011, with the ultimate goal of providing coverage to 300,000 households or 900,000 OVC. The monthly value of the transfer is 20 USD and is disbursed on a bi-monthly basis.

The impact evaluation used a randomized cluster longitudinal design, with the baseline quantitative survey fieldwork carried out in mid-2007. The underlying justification for the design was that the programme could not be spread out to all eligible locations at the same time, and as a result sites whose entry was expected to happen later were considered as control sites. Approximately 2,750 households were surveyed in seven districts across Kenya. Two-thirds of these households later began receiving cash transfers while the remaining households served as the control group. Within each district, two locations were chosen randomly to benefit from the programme intervention and two

were selected as control. These households were re-interviewed (first round) two years later, between May and July 2009, in order to assess the impact of the programme on key welfare indicators. The re-interview success rate was approximately 83 percent. The second round follow up study was again conducted between May and August 2011.

Unfortunately the 2007 and 2009 rounds of data collection were very limited on economic activities, including labour market participation. The 2011 second round follow up added a more detailed economic activity module (including wage labour, self-employment, crop and livestock activities, risk preferences, etc) to capture the potential investment and productive activity benefits of the programme. For some of the household level analysis, such as ownership of livestock and agricultural implements, baseline data were collected. Thus for many of the outcome variables of interest to the project, we have only one data point (no baseline).

The randomization in Kenya was not as robust as in the case of Lesotho (below) due to the fewer units over which the randomization took place. Difference in Difference alone combined with propensity score techniques when necessary has been used to analyse these data. In cases where the outcome variables are not available at the baseline, we will be forced to rely exclusively on the PSM technique using cross sectional data (with some baseline predictors) to identify the effect of treatment. Since we do have some variables over which we can use Difference in Difference estimators, such as livestock, we can test the robustness of propensity score techniques on these data by comparing propensity score results for livestock to the DD.  
(Source: Ben Davis, Food and Agriculture Organisation.)

### ***Lesotho: Child Grant Programme***

The Lesotho Child Grants Programme provides an unconditional cash transfer to poor and vulnerable households. The primary objective of the CGP is to improve the living standards of orphan and vulnerable children (OVC) including nutrition and health status and increased school enrolment. The CGP is targeted at poor households with children (<18), including child-headed households—there is no other special definition of “vulnerable” children. As of February 2012 the program reached 9,915 households (covering 28,000 children), with a planned scale up of 5,000 additional families annually, reaching 25,000 in 2014. The monthly value of the transfer is 14 USD and is disbursed on a quarterly basis.

The quantitative analysis for the Lesotho study is an experimental design impact evaluation. Participation in the program was randomized at the level of the electoral district (ED). First, all 96 EDs in four community councils were paired based on a range of characteristics. Once these 48 pairs were constructed, 40 pairs were randomly selected to be included in the evaluation survey. Within each selected ED, 2 villages (or clusters of villages) were randomly selected, and in every cluster a random sample of 20 households (10 potentially called to enrolment and 10 potentially non-called to enrolment) were randomly selected from the lists prepared during the targeting exercise. After the baseline survey data were collected in all evaluation EDs, public

meetings were organized where a lottery was held to assign each ED in each of the pairs (both sampled and non-sampled) to either treatment or control groups. Selecting the treatment electoral districts after carrying out the baseline survey helped to avoid anticipation effects.

The baseline household survey was carried out prior to the distribution of the first transfers to treatment households; a follow up panel survey will take place two years later in 2013. A total of 3,102 households were surveyed; 1,531 programme eligible households (766 treatment and 765 control) to be used for the impact evaluation analysis, with the remaining 1571 programme non-eligible households to be used for targeting analysis and spillover effects. Besides the household survey, community and business enterprise questionnaires were implemented. The impact evaluation is being implemented by Oxford Policy Management (OPM) and Sechaba Consultants.

The method of randomization described above, including the relatively large number of units of randomization, reduced the likelihood of systematic differences between treatment and control households. The baseline analysis report shows that the randomization was quite successful, with few significant differences between the households in treatment and control groups on key selection indicators. As to be expected in any randomization process, a small number of significant differences did emerge between treatment and control households, including some dimensions of food security, social networks, land cultivation, and livestock and land ownership. The household and individual level impact analysis will thus use Difference in Difference estimators and, when necessary, propensity score methods.

(Source: Ben Davis, Food and Agriculture Organisation.)

### ***Malawi: Social Cash Transfer Programme***

The Malawi Social Cash Transfer Programme (SCT) was initiated in 2006 in the pilot district of Mchinji, providing small cash grants to ultra-poor, labour-constrained households. The SCT programme objectives include reducing poverty and hunger in vulnerable households and increasing child school enrolment. The SCT is currently operational in 7 districts and reaches over 26,000 ultra-poor and labour-constrained households and is expected to serve 300,000 households by 2015. At present, approximately 103,000 individuals benefit from the programme of which nearly two-thirds are children and nearly half are orphaned children. The programme is executed by the Government of Malawi through the District Councils by Social Welfare Officers. On average 12 USD per month are paid quarterly. The transfers vary according to household size: 4, 7, 10 and 13 USD will be given to households with 1, 2, 3 and 4 members, respectively. In addition, 1.5 USD will be given on a monthly basis to households for each child enrolled in primary school and 3 USD for each child enrolled in secondary school.

An impact evaluation of the programme pilot in Mchinji in 2007-08 was implemented by Boston University and the Center for Social Research at the University of Malawi. A new impact evaluation is planned to accompany the next round of programme expansion. The final impact evaluation design is

awaiting the tendering of the contract for the evaluation, which will be finalized by July, 2012. As such, it is unclear what level of randomization will be possible. The baseline will take place in the fourth quarter of 2012, with a follow up survey one year later.

(Source: Ben Davis, Food and Agriculture Organisation.)

### ***Uganda: Social Assistance Grants for Empowerment***

OPM is engaged in designing and implementing an impact evaluation of the pilot Social Assistance Grants for Employment (SAGE) cash transfer programme, which aims to support vulnerable households under the Uganda government's Expanding Social Protection programme. The impact evaluation adopts a mixed method approach, employing a quantitative household panel survey and qualitative research to assess impact and evaluate operational effectiveness of the programme across 8 out of 14 pilot programme districts. The quantitative household panel survey uses a Regression Discontinuity Design to estimate impact and will comprise three rounds of survey over four years including baseline. The Impact Evaluation focuses on the core impact areas of reduced material deprivation, increased economic security, reduced social exclusion and increased access to services. It includes impact on consumption expenditure and nutritional status of under-fives. The qualitative research focuses on impact on empowerment and social cohesion. The operational effectiveness assessment will draw on both indicators generated through the quantitative and qualitative research and indicators provided by the SAGE programme's internal operational monitoring systems in order to feed into the programme's Learning Framework.

(Source: Fred Merttens, OPM)

### ***Zambia Social Cash Transfer Programmes***

In 2010 Zambia's Ministry of Community Development and Social Services (MCDSS) began implementing the Child Grant cash transfer programme (CGP) in three districts – Kalabo, Kaputa and Shongombo – with the highest rates of mortality, morbidity, stunting, and wasting among children under 5. The CGP includes all households with a child under five years of age. Eligible households receive 55,000 kwacha a month (equivalent to US\$11) irrespective of household size, an amount considered sufficient to purchase one meal a day for everyone in the household for one month. The goal of the programme is to reduce extreme poverty and the intergenerational transfer of poverty. The objectives of the programme are to (1) supplement and not replace household income; (2) increase the number of children enrolled in and attending primary school; (3) reduce the rate of mortality and morbidity among children under 5 years old; (4) reduce stunting and wasting among children under 5 years old; (5) increase the number of households owning assets such as livestock; and (6) increase the number of households that have a second meal a day.

UNICEF-Zambia contracted the American Institutes for Research (AIR) to design and implement a randomized controlled trial for a three-year impact evaluation of the programme. The baseline was carried out in September-October 2010, with follow ups planned for 2012 and 2013. Baseline data collection occurred in Zambia's lean season (September through February),

when people have the least amount of food left from the previous harvest and hunger is at its greatest. The study includes 2,515 households, 1228 treatment and 1287 control. Communities were randomly assigned to treatment (incorporated into the programme in December 2010) or control (to be brought into the programme at the end of 2013). The baseline data collection began before communities were randomly assigned to treatment or control groups. Analysis of the baseline data shows that randomization appears to have worked well; greater detail on the randomization process can be found in Seidenfeld and Handa (2011).

(Source: Ben Davis, Food and Agriculture Organisation.)

### ***Zambia: Cash Transfers Programmes***

American Institutes of Research (AIR) has been contracted to evaluate Zambia's cash transfer programmes. AIR is implementing two separate cluster randomized controlled trials to measure the impacts of Zambia's two cash transfer programs - one that targets households with children under five in the country's poorest districts, the other targets extremely vulnerable and labour constrained households. It is one of the largest cash transfer impact evaluations in Africa and is one of the few with sufficient power to detect effects for children under two years, with sample sizes of 2,500 households and 3,000 households respectively. This is a 3 year impact evaluation with baseline and three follow ups at 2 years, 2.5 years, and 3 years out. The outcome measures cut across many policy areas including health, labour, education, nutrition, early childhood development, consumption, productivity, and women's empowerment.

The size and scope of this study poses several challenges that AIR has had to address. For example the national census and presidential elections were scheduled to occur during the same time as data collection. Timing of data collection is important because livelihoods in rural Zambia are tied to the farming cycle with certain periods of the year being more bountiful than others. Data need to be collected during specific times to control for these cyclical factors. Similarly, it is necessary to collect data before the rain season because roads become impassable in rural areas. However, we could not collect data during the census or presidential elections to avoid confusion about the purpose of the study and respondent fatigue. To overcome this challenge, AIR had to work closely with the Zambian government and remain in constant communication with the government about the scheduling delays for these two activities. Ultimately, AIR had to remain flexible and adapt to changes in the government's schedule, responding quickly when windows of opportunity became available.

(Source: David Seidenfeld, American Institutes of Research.)

### ***Zimbabwe Social Cash Transfer (SCT) Programme***

To address household poverty as a key driver of child vulnerability in Zimbabwe, the revised National Action Plan for Orphans and Vulnerable Children (NAP II) 2011-2015 and its accompanying pooled funding mechanism (the Child Protection Fund) have included social cash transfers as a major programme component, accompanying other interventions in child protection and access to social services. The Fund is a multi-donor pooled



funding mechanism managed by UNICEF in partnership with other partners.

District selection for the prioritization of the phased cash transfer programme has been based on a cross-analysis of the Poverty Assessment Survey (2003), the Nutrition Survey (2010) and the Zimbabwe Vulnerability Assessment Committee (2010) to determine a proxy for prioritizing specific districts. The overall objective of the new impact evaluation is to generate policy-relevant evidence on the impact of the cash transfer scheme of the NAP II on key child health, education, HIV, equity, nutrition, protection and livelihood (or household economy) outcomes. The first transfers were made in December 2011; the impact evaluation will be applied on successive rounds of implementation of the pilot phase of the programme. The average volume of transfers is set as USD 20 per household per month. For individual households the transfers vary in accordance with the size of the household. Households with one member get 10 USD whereas households with two, three and four (or more) members get 15, 20 and 25 USD respectively.

The final impact evaluation design is awaiting the tendering of the contract for the evaluation, which should be finalized by the end of March, 2012. The baseline will take place during the third quarter of 2012, with two follow ups at one year intervals.

(Source: Ben Davis, Food and Agriculture Organisation.)

## Glossary

**Attrition:** occurs when individuals being monitored for the evaluation die, move away or otherwise become unavailable for further monitoring. Too much attrition or attrition from just one sub-group within the evaluation can affect the power and validity of results.

**Evaluation:** DFID adopts the OECD DAC definition of evaluation – ‘The systematic and objective assessment of an on-going or completed project, programme or policy, its design, implementation, and results in relation to specified evaluation criteria’. Evaluation involves measuring objectively what we did, what happened as a result, and why. The key features of DFID evaluations are independence, transparency and methodology.

**External validity:** the extent to which evaluation results can be generalised to other situations outside of the evaluation area. Without understanding the causal pathways through which change has occurred, and making a judgement as to whether these might be replicated elsewhere, external validity of pure impact evaluations may be low.

**Internal validity:** the extent to which the results of an evaluation can be applied to different contexts within the locus of the evaluation. A result with high internal validity is likely to be generally true across all the individuals and communities being evaluated. Evaluation results with more limited internal validity may apply only to a subset of them.

**Monitoring:** a continuous internal process throughout the project cycle, conducted either by managers or by beneficiaries, to measure the progress of development interventions against pre-defined objectives and plans.

**Regression Analysis / multivariate regression analysis:** a statistical tool which estimates the impact on a variable of interest of a number of potentially explanatory variables. It produces estimates of the effect of each explanatory variable on the variable of interest if all other explanatory variables are held constant. For social transfer evaluations it can be useful to help isolate the effect of the intervention (which will be one explanatory variable) when randomisation or quasi-experimental design has been imperfect and a number of baseline differences exist between treatment and control / comparison groups exist.

**Sample size:** the number of individuals or communities being monitored (as treatment and control / comparison) as part of the evaluation. The smaller the effect of an intervention, very often, the larger is the necessary sample size to be able to detect that and attribute that effect with statistical significance to the intervention.

**Selection bias:** occurs when the selection of treatment and comparison / control groups for some reason skews the evaluation results for reasons not connected with the intervention. For example, if selection of the treatment group focuses on students likely to have high academic ability, and the comparison group on students of average ability, an evaluation may show that the treatment group perform better in examinations in response to the programme intervention, when the actual cause is unrelated to programme attributes.

**Spillovers:** These occur when non-treated individuals or communities benefit or suffer from the treatment being given to other communities or individuals.