ePact

Strengthening evaluation effectiveness and impact

Longitudinal Monitoring and Independent Impact Assessment of CLP-2

Final Evaluation Report – Volume II: Technical Companion and Methodological Annexes

Paul Jasper, Denis Nikitin, Stephanie Brockerhoff, Ferdous Jahan, Michele Binci, Alastair Haynes, Martina Garcia, Alexandra Doyle, Elisabeth Resch, and Tahera Ahsan

Oxford Policy Management

June 2016

Oxford Policy Management

itad

e-Pact, is a consortium led by Oxford Policy Management and co-managed with Itad.

Acknowledgements

We gratefully acknowledge the collective efforts of a number of enthusiastic people who contributed their time, effort and wisdom to this endeavour.

This work would not have been possible without the substantial support we received from the staff of the Chars Livelihood Programme (CLP) and in particular the CLP Innovation, Monitoring, Learning and Communication (IMLC) division.

In addition, the authors would like to acknowledge the contributions of our field team who implemented the qualitative fieldwork and provided substantial research support, particularly Mamun-ur-Rashid and Omar Faruque Siddiki and their team at the Development Research Initiative (dRI).

We would also like to thank the UK Department for International Development (DFID) and the Australian Department for Foreign Affairs and Trade (DFAT) for the cooperation throughout these projects.

Finally, we would like to express our gratitude to all participants of two workshops held in Dhaka during the implementation phase of this evaluation, who provided very useful and valuable comments and insights for our work.

All errors are, naturally, our own.

Table of contents

List of tables and figures

Abbreviations

ATP	Asset Transfer Programme
ATT	Average Treatment effect on the Treated
BBS	Bangladesh Bureau of Statistics
BDT	Bangladeshi Taka
BMI	Body Mass Index
BRAC	Bangladesh Rural Advancement Committee
CBC	Chars Business Centre
CLP	Chars Livelihood Programme
CPHH	Core participant households
DAC	Development Assistance Committee
DFAT	Department for Foreign Affairs and Trade
DFID	Department for International Development
dRI	Development Research Initiative
FGD	Focus Group Discussion
HFA	Height-for-Age
HIES	Household Income and Expenditure Survey
KIIs	Key informant interviews
IMLC	Innovation, Monitoring, Learning and Communication
IMOs	Implementing organisations
LSP	Livestock Service Provider
NGO	Non-Governmental Organisation
NN	Nearest Neighbour
OECD	Organisation for Economic Co-operation and Development
OLS	Ordinary Least Squares
OPM	Oxford Policy Management
OTUP	Other Targeted Ultra Poor

PRIME	Programmed Initiatives for Monga Eradication
PSM	Propensity Score Matching
RDCD	Rural Development and Cooperatives Division
SEQAS	Specialist Evaluation and Quality Assurance Service
STUP	Specifically Targeted Ultra Poor
TOC	Theory of Change
TOR	Terms of Reference
TUP	Targeting the Ultra Poor Programme
WASH	Water, Sanitation and Hygiene
WFA	Weight-for-Age
WFL	Weight-for-Length
VDC	Village Development Committee
VfM	Value for Money
VSLs	Village Savings and Loans groups

1 Introduction

1.1 Objectives of this report

This document has four main objectives. First, it aims to be a technical and methodological companion to Volume I of this evaluation report (Longitudinal Monitoring and Impact Assessment of CLP-2: Final Evaluation Report – Volume I; OPM, 2016a), which presents our main findings. Second, it gives additional background information about the process of defining the scope and implementing our evaluation. Third, it presents additional results and information that were not included in Volume I. Finally, this report responds to any additional requirements posed by DFID's internal quality assessment system (SEQAS). The target audience for this report is those key stakeholders who are interested in the technical and methodological details of the evaluation. For a non-technical discussion of the key findings, please refer to Volume I (OPM, 2016a).

1.2 Structure of this report

The remainder of this report is structured as follows: Section 2 describes the background to this evaluation, with a particular focus the evaluation objectives, evaluation questions, the context of the evaluation, the design process, the departures from the original Terms of Reference (TOR) and the evaluation team. Section 3 provides the background to the evaluation, discussing the activities CLP undertakes (3.1) as well as the activities of other programmes being implemented in CLP-2 areas (3.2). Section 4 discusses the mixed methods approach to this evaluation and how generalisable the findings are.

Section 5 presents the technical and methodological details of our qualitative research component. Section 5.1 discusses the objectives and research questions of the qualitative analysis while 5.2 provides the data collection methods and 5.3 the strategies used for sampling.

Section 6 provides a technical discussion of our quantitative research component, with Section 6.1 providing a general overview of the three data analysis approaches we employed. Section 6.2 summarises the data used, discusses the methodology behind both the sampling weights and the poverty indicators created for this analysis, and present our analysis of sample attrition. Section 6.3 briefly discusses our main findings from a descriptive analysis of the quantitative data.

Section 6.4 presents the propensity score matching (PSM) models used for this analysis. This provides technical detail related to the methodology used for PSM (6.4.1). Section 6.4.2 summarises the PSM model in the context of CLP, whilst Section 6.4.3 discusses the various estimation strategies that can be conducted, given the structure of the available data. Section 6.4.4 then discusses the possible limitations to PSM analysis and what steps have been taken to address these. Finally, Section 6.4.5 presents our results for multiple estimation strategies for each area of research that the quantitative component tackled. This includes the results related to balancing for each variable in each strategy, as well as the average treatment effect on the treated (ATT) for both kernel and nearest neighbour approaches to matching.

Section 6.5 then discusses the panel analysis. Section 6.5.1 provides detailed technical discussion surrounding panel analysis, whilst Section 6.5.2 discusses how we built the panel models. Section 6.5.3 then presents the results for the panel analysis for the different areas of interest as well as the robustness checks that were implemented. Section 6.5.4 sets out the possible limitations associated with the panel analysis and how these have been addressed.

Section 6.6 then looks at the cost–benefit analysis, stating the objectives of this analysis (6.6.1), what data the analysis are used (6.6.2), a technical discussion surrounding the methodology used (6.6.3) and both the absolute and relative cost-effectiveness (6.6.4).

2 Background to the evaluation

2.1 Evaluation objectives and questions

Volume I (OPM 2016a) summarises the objective of this evaluation, in addition to presenting the key evaluation questions that are answered by the evidence presented in that report. It is important to mention here that, according to what was agreed in the inception phase, this evaluation pursued two main objectives:

- First, to assess the effectiveness of CLP-2 along several key dimensions as specified by the Organisation for Economic Co-operation and Development (OECD) Development Assistance Committee (DAC) criteria; and
- Second, to draw lessons from this to inform future programme design.

Achieving these objectives and answering the evaluation questions also implied adding to the evidence base with respect to programmes that aim at reducing extreme poverty. For the context of the current evaluation, this particularly relates to producing findings with respect to the challenge of reducing poverty in an environment of extreme vulnerability, such as is the case with the chars in Bangladesh.

In terms of the first objective, the four relevant OECD DAC criteria to assess the effectiveness of CLP-2 were defined as follows:

Effectiveness: To what extent were the objectives/targets of CLP-2 met?

Impact: What changes in key indicators can be attributed to CLP-2?

Sustainability: Do these changes persist in time?

Efficiency: To what extent does CLP-2 represent good value for money (VfM)?

As was explained in detail in our Inception Report (OPM 2015), the fifth OECD DAC criterion, **relevance**, has not been addressed by this evaluation. Sub-questions related to this DAC criterion referred to local economic development and the effects of CLP-2 on non-participants and other char dwellers. However, after consultation with CLP and DFID Bangladesh in the inception phase, the evaluation team came to the conclusion that this evaluation will not be able to focus on questions related to this criterion separately. Annex B provides more details on the thinking behind this decision.

This evaluation has addressed questions that relate to each one of the above criteria. Again, Annex B delineates how the evaluation questions relate to the OECD DAC criteria. For ease of presentation, these questions have been restructured around areas of investigation in Table 1 below.

Table 1 Key evaluation questions

	Area of investigation	Evaluation questions	Quantitative evidence?	Qualitative evidence?
1	Graduation	How many members of core participant households (CPHHs) met the graduation criterion developed by CLP-2? What are the major factors that cause graduating households to become non-graduated?	Yes	Yes
2	Poverty	What is the number of people from CPHHs who were lifted out of extreme poverty – as defined by CLP's specific poverty line? What was the impact of CLP-2 on this number and on the poverty gap?	Yes	
3	Livelihoods	How have livelihoods of CPHHs changed in the following areas: income, expenditures, savings, and assets? What was CLP's impact on this?	Yes	Yes
4	Sustainability	To what extent is the graduation of people from CPHHs according to CLP's graduation criterion sustainable? How sustainable are other observed impacts over time and what are the perceptions about the future sustainability of these impacts beyond the CLP implementation phase? What are the major factors that drive sustainable graduation?	Yes	Yes
5	Efficiency	To what extent does CLP-2 represent good VfM? To what extent was the targeting appropriate? What was the level of inclusion error? To what extent did the programme target various social groups such as the disabled and elderly?	Yes	
6	Perceptions around changes in the local economic context	How have livelihoods and the local economic context changed as a result of CLP-2?		Yes
7	Others: **empowerment**, nutrition practices, food security, malnutrition	How have indicators related to these areas changed? What was CLP's impact on these indicators?	For some areas	For some areas

Comments with regards to our evaluation questions

The set of questions presented in Table 1 is an amended version of the evaluation questions initially suggested by DFID and DFAT in the TOR and presented in the Inception Report (and also Annex A). Here, we want to make some qualifying notes with respect to these questions:

- First, from a qualitative perspective, the term 'graduation' is not separable from the general 'wellbeing' of CLP participant households. This means that in this evaluation the question related to 'factors which respondents felt to be inhibiting or supporting participants' to benefit from CLP-2 is interpreted widely to refer to factors that impact welfare improvement of participants.

- Second, it important to specify that we interpret all impact questions (across the different areas of investigation) as strictly referring to the attribution question – i.e. what changes in participant households' lives can be attributed to CLP-2? This question could only be addressed quantitatively and is therefore the major focus of our quantitative research component.

- Finally, it is important to mention that the category of 'others' (number 7) has been included in the table above in order to capture the fact that – based on communications with stakeholders, the client, and initial quantitative or qualitative results – some research areas initially suggested in the TOR have dropped in relevance while other interesting findings and research areas have newly emerged since the start of this evaluation. Most importantly, the following needs to be noted:

- First, our quantitative descriptive statistics report (OPM 2016b) captures indicators that touch a variety of areas that are not included in the table above but were included in the initial TOR and Inception Report, such as, for example, changes in the social capital of participant households. We presented these results in a workshop in Dhaka in January 2016 and in OPM (2016b) (see Annex F for the executive summary to the related report). After this workshop, several decisions were made with respect to which of these areas should be investigated further and whether that was to be done quantitatively or qualitatively.

- Most importantly, it was decided that the evaluation should generally focus its efforts on questions of graduation, poverty, livelihoods, sustainability, efficiency, and the local economic context – i.e. the first six areas of interest presented in the table above – which meant that the 'other' areas were not considered as the overall focus of this evaluation.

- In addition, following explicit requests from stakeholders, it was decided that our qualitative research component should present results with respect to empowerment questions relating to char dwellers generally, and women in particular.

- Moreover, following conversations both at the workshop in January 2016 and a follow-up workshop in April 2016, it was decided that results around food security and, in particular, nutrition practices and malnutrition were not considered to be of significant relevance for this evaluation, given that DFID is currently implementing another evaluation of the Direct Nutrition Intervention. We still present some quantitative results with respect to these areas of interest here but these have not been covered in Volume I (OPM 2016a).

- It was also decided to not further investigate the question of 'social capital'. This was because, first, the quantitative data did not allow for a nuanced analysis of this area of investigation and, second, analysing this topic comprehensively from a qualitative perspective would have required a separate study and different research design.

- Finally, it was decided that questions with respect to vulnerability were not be treated separately from the other areas of investigation but rather as a cross-cutting topic. Note that our discussion around sustainability, savings and factors that affect how well participant households can benefit from CLP-2 addresses the issue of household vulnerability explicitly.

Component-specific objectives

Table 1 also indicates the type of evidence that was used by this evaluation to answer the different evaluation questions. Different types of evidence were produced by the quantitative and qualitative components of this evaluation, based on the different specific objectives that each of those components had within the context of this evaluation. These method specific objectives that exist in addition to the overall objectives of our evaluation are driven by the strengths of the different methodological approaches and by the availability of quantitative data.

For the **qualitative component**, the specific objectives were as follows:

1. **Explore the perceived reasons for the differential levels of wellbeing of participant households** and whether certain household characteristics affect this. This meant producing evidence with respect to areas of investigation 1 and 4 and the question of which factors caused graduating households to become non-graduated or to stay graduated.

2. Investigate **perceptions of the sustainability of the level of wellbeing** achieved. This meant producing evidence with respect to area of investigation 4.

3. Explore **perceptions regarding changes in local markets and the effectiveness of CLP's market development approaches**. This meant producing evidence with respect to area of investigation 6.

4. **Triangulate quantitative findings and investigate questions arising from preliminary quantitative findings**, i.e. explain how and why some salient trends or unexpected findings observed in the quantitative data emerged. The focus for this was on savings, empowerment and livelihood changes. This meant producing evidence with respect to areas of investigation 3 and 7.

For the **quantitative component**, the specific objectives were as follows:

1. **To give descriptive estimates for changes** in the lives of participant households with respect to areas of investigation 1, 2, 3 and 7 and to **assess how much of those changes can actually be attributed to CLP-2.** Note that this included creating quantitative measures of poverty. In addition, addressing the attribution component meant answering the 'impact' questions. Moreover, evidence produced here feeds into our approach to answering the efficiency questions (area of investigation 5).

2. **Second, to analyse how changes have materialised over time and to what extent they persist** several years after the core implementation period of CLP-2. This meant addressing questions under area of investigation 4, using evidence from our descriptive and impact analyses.

2.2 Evaluation process

The process of defining the scope and implementing this evaluation has consisted of several phases in which the OPM team has closely consulted with key stakeholders – mainly DFID, DFAT, and CLP – in order to incorporate their needs, perspectives and views into the approach.

The technical proposal

The main parameters of our proposed methodology – i.e. the mixed methods framework, the combination of quantitative and qualitative evaluation approaches, and the key options for quantitative and qualitative design – were outlined in a technical proposal and were based on OPM's assessment of the best practices of for such evaluations and its own first-hand experience and expertise in implementing comparable evaluations.

The inception phase

However, in order to further adapt the evaluation design to the specific CLP context, the priorities of the DFAT and DFID Bangladesh team, to incorporate the specific insights of the very experienced CLP team, and to initiate the quantitative-qualitative integration at the earliest stages of the project, **OPM has engaged in extensive consultations with DFID Bangladesh, DFAT and CLP in the inception phase**. Important elements of this consultative approach in the inception phase were:

- An in-depth review of CLP-2 and field visits to Bogra and the chars during a first inception mission in September 2015. This mission included:
  - Extensive consultations with key individuals of the CLP team: the team leader, the IMLC, and heads of other key CLP divisions (market & livelihoods, infrastructure, human development, etc.).
  - A one-day field visit to the chars, where participant households were visited and the different components of the intervention observed on the ground.
  - Extensive consultation with IMLC team members on monitoring data, data collection, management and analysis.
  - Two meetings with DFID and DFAT in order to discuss preliminary ideas about the evaluation and changes to the original proposed methodology.
- **A separate qualitative mission in October 2015 by OPM's qualitative analysis experts in order to develop a light-touch Theory of Change (TOC) that would then inform quantitative and qualitative research methodology.**[1] This mission included:
  - Intensive discussions with the different CLP teams and iDE about their areas of responsibility and expertise. The idea was to: (1) gain a detailed understanding of the different interventions, how they evolved over time and the lessons learned during this process; (2) to understand the TOC guiding the individual interventions; and (3) collect programme documentation that would help with the development of a TOC.
  - A one-day visit to implementing partners to discuss lessons learned, challenges experienced, perceptions around the sustainability of interventions, and the TOC guiding the selection of interventions within their areas of responsibility.
  - Development of a light-touch TOC based on discussions with CLP and iDE and the programme documentation reviewed. The TOC was then discussed with the CLP partnership coordinator for an initial validity check at the end of the mission. This light-touch TOC can be found in Annex C.

The result of this inception phase was an inception report submitted to and agreed on by DFID and DFAT in January 2016. This included some initial quantitative data analysis results, a detailed

[1] The light-touch TOC attempted to better understand CLP and its complex interventions, as well as helping us to structure our thinking and the design of the evaluation. It does not attempt to explain the causal links within different interventions or make claims about the strength of these causal links (see Annex C).

quantitative and qualitative research methodology, and our work plan for the implementation phase.

The implementation phase

The implementation phase consisted of several key activities. First was the descriptive quantitative analysis exercise, which produced OPM (2016b) (see Annex F for the executive summary of this report). Second came the inferential component of our quantitative analysis, i.e. implementing the PSM and panel regression approaches. Third was the qualitative fieldwork and, fourth, the qualitative data analysis. The final stage was the reporting phase, in which our evidence was compiled in order to produce the final evaluation report. It should be noted that during all of these phases the team worked across workstream in an integrated manner in order to ensure that findings emerging in one workstream could immediately inform the work of the other.

Two key workshops were implemented in this phase in which we presented both preliminary evaluation results and discussed their implications with key stakeholders:

- First was a quantitative-qualitative workshop in January 2016 for stakeholders to discuss the proposed approach to the qualitative work and to present **findings from the descriptive quantitative analysis.** This workshop included:
    - A presentation and discussion of the quantitative descriptive analysis results, with a focus on their implications for further quantitative and qualitative work.
    - A presentation and discussion of the suggested qualitative field work and research approach. This included a review of the research questions that the qualitative work would focus on, a review of qualitative sampling of sites and selection of respondents, and a revision of the timeline for this analysis.
    - Discussion on refocus of qualitative research following the findings from the quantitative descriptive analysis.
- **A workshop in April 2016 presenting the preliminary findings of the evaluation.** This workshop included:
    - A presentation and discussion on the objectives, methods and results from our mixed methods approach.
    - Presentation of preliminary findings relating to our evaluation questions**.**
    - An in-depth discussion with stakeholders on these findings and resulting preliminary recommendations – including feedback for further investigation.

It is important to mention here that, throughout the evaluation, there has been a continuous interaction of the OPM evaluation team with the CLP team. This was especially the case during the early quantitative data cleaning and preliminary analysis stages, with the aim of understanding the availability and quality of data that could be used for the quantitative components of this evaluation.

2.3 The context of CLP-2 and this evaluation

This evaluation was implemented as an ex-post evaluation of CLP-2, i.e. towards the end of the second phase of this programme, in order to learn from CLP's experiences and to inform future livelihood programming, which is imminent in Bangladesh. The timing of the evaluation was explicitly chosen so that findings could be produced before CLP operations ended and staff members could be informed about those. In addition, the timing was chosen so that results could feed into future livelihood programming in the country.

Such programming will continue to be crucial in the future: Bangladesh retains characteristics of a traditional low income country – with high income poverty and pockets of extreme poverty – but has made tremendous progress in improving social indicators in the last decade. There is an increasing policy focus on both economic development but also emerging areas such as skills and urban development. Climate- and disaster-risk remain a major challenge. (World Bank 2015)

Development spending has mirrored this. Bangladesh has seen the implementation of some very large and well-known programmes that aimed at decreasing extreme poverty and improving the livelihoods of very poor individuals sustainably. For example, DFID has provided substantial funding for the Economic Empowerment of the Poorest programme, also known as Shiree, which has been running for about eight years and will end in September 2016. Similarly, DFID funded the Urban Partnership for Poverty Reduction programme, which ended in January 2016. Finally, BRAC has been implementing the Targeting the Ultra Poor programme for over ten years in Bangladesh. The implementation of CLP-2 must be seen within the context of this concerted effort to sustainably tackle extreme poverty in Bangladesh.

At the same time, this evaluation must be seen within the context of existing and continuous research on the effectiveness of such programmes. Most recently, research on the impact and sustainability of positive effects of the 'graduation approach' has been published by Banerjee et al. (2015) and Bandiera et al. (2016). In particular Bandiera et al. (2016) make reference to the positive effects of livelihoods programmes in Bangladesh. Our evaluation feeds into this existing evidence base by adding knowledge about the effects of CLP-2, an intervention that transferred the multifaceted poverty graduation approach into the explicitly challenging context of the chars and into a population of households that faced high environmental vulnerability, exclusion, and extreme poverty.

Our evaluation provides insights into how this approach has worked in this challenging context. As described above, major livelihoods programmes in Bangladesh are about to end or have ended recently. New livelihood programming is imminent – and insights gained via our evaluation will feed into the design and implementation of new programmes. Our communication and dissemination plan (section 2.5) ensures that key stakeholders for this, including the GoB, will be informed of our findings.

2.4 Departures from the original TOR

The evaluation design and implementation process mentioned above has led to four main departures from what was specified in the original TOR and the Inception Report of this evaluation (see Annex A for the original TOR)[2]:

- First, the timeline has been amended, in order to respond to additional requirements and inputs received from stakeholders. Two particular points need to be mentioned here. First, the final set of quantitative data from CLP-2 was received by the evaluation team only in mid-February 2016. All data cleaning, quality assessment, and analysis using that data could only be implemented after that. Second, the workshop implemented at the end of April 2016 produced some additional questions that then also had to be addressed by the evaluation team.

- Second, the evaluation questions addressed have been updated, based on an assessment of stakeholder needs, interests, preliminary findings and data availability (see Section 2.1 above for a description of how and why questions changed and Annex B for a more comprehensive and disaggregated presentation of the updated evaluation matrix).

- Third, based on these changes, the scope of the qualitative research was expanded to include areas of investigation that were previously not planned (see Section 2.1 for more detail).

- Finally, the range of methods employed to quantitatively assess programme impact evolved to include panel analysis over and above a PSM approach. The strategies informing the PSM analysis have been expanded during the implementation of this programme. Section 6.4.3 provides more detail on the different PSM estimation strategies we employed.

2.5 Communication and dissemination

2.5.1 Communication strategy

Our communication strategy pursued three primary objectives: (1) to inform DFID and DFAT – and to the extent necessary other stakeholders – of the progress of the evaluation; (2) to coordinate field activities with CLP and other stakeholders, e.g. community leaders when conducting focus group discussions (FGDs) and key informant interviews (KIIs); and (3) establish a process for transparent, fluid and timely sharing of evaluation findings between OPM and CLP throughout the evaluation.

Objectives (1) and (2) were achieved by making available the latest work plans, mission announcements, post-mission debriefings and/or mission completion reports (upon agreement with DFID and DFAT), as well as draft reports, early, thus allowing sufficient time for DFID, DFAT and other stakeholders to provide comments. **In addition, our two workshops (i.e. the quantitative-qualitative workshop in January 2016 and the presentation of the preliminary integrated qualitative/quantitative findings at the advanced stage of our analysis in April 2016) served as communication milestones.**

[2] Please also refer to the Inception Report (OPM 2015) for more detail.

The former event presented the early descriptive findings and our fieldwork design in the context of the integrated mixed methods approach; the timing of the workshop was such that it allowed incorporating the feedback from the stakeholders into the qualitative fieldwork and further mixed methods design. The latter event aimed to inform and solicit feedback on the key findings of our evaluation, again leaving time to incorporate stakeholder feedback into our final report.

To meet objective (3), OPM has made every effort to share with the CLP team report drafts and critical and/or counterintuitive findings and take into account the CLP team's feedback. Such information sharing has attempted to fill possible knowledge gaps about the details and history of the implementation of CLP's interventions. It has also enabled early detection of potentially unfounded or inaccurate assumptions about the way CLP interventions were designed or implemented, thus improving the quality and efficiency of our analytical work.

2.5.2 Dissemination

The purpose of dissemination is raising awareness of the findings of the evaluation among the development partners, the policy-making community and the public, as well as promoting policy dialogue on CLP's development effectiveness and the lessons it offers.

Our dissemination approach centres on the public dissemination of evaluation results and making available the final report to the development community through DFID's and DFAT's standard dissemination channels.

Four key products will be used for dissemination purposes:

- A non-technical full final report (Volume I; OPM 2016a);
- A technical final report (Volume II, the current report);
- A four-page summary of the main evaluation findings; and
- A PowerPoint presentation that summarises the findings and methodologies employed in a non-technical and visual way.

All final products will have gone through a peer-reviewing and stakeholder consultation process, so that the content and format of these products are aligned with the needs of DFID, DFAT, CLP and the Government of Bangladesh.

We will then work together with stakeholders to disseminate these products to a wider community of government staff, donor organisations, key national non-governmental organisations (NGOs) and other international partners. **However, it is important to note that no specific in-country event is currently planned for this dissemination**.

Communications with the media on the methodology, content and findings of the evaluation will be handled in close cooperation with CLP, DFID and DFAT.

2.6 The evaluation team and evaluation management

Our management and team structure throughout the evaluation has closely followed what has been suggested in the TOR, our proposal and the Inception Report. The project was managed by

Paul Jasper of OPM, which ensured the smooth running of administrative issues and communication flows within the team and with respect to stakeholders. It is important to emphasise here again that this evaluation was managed in close consultation with stakeholders. This included sharing of findings and recommendations with CLP, DFID, DFAT, and the GoB, and taking into account different views and comments from those stakeholders.

The core technical team members consisted of the team leader (Denis Nikitin), who provided both overall guidance and significant technical inputs both quantitatively and qualitatively. It also consisted of a national qualitative expert (Ferdous Jahan) and an international qualitative expert (Stephanie Brockerhoff), who were responsible for the qualitative components of the evaluation. Moreover, two quantitative experts (Paul Jasper and Dr Michele Binci) led the quantitative component. Finally, an independent national coordinator (Tahera Ahsan) supported the team with respect to coordinating national activities and, in particular, the qualitative fieldwork.

In addition, we have had a pool of OPM experts available for quantitative and qualitative research insights and quality assurance: Sarah Keen and Elisabeth Resch implemented the cost–benefit analysis. Alexandra Doyle, Martina Garcia Aisa, Purava Joshi and Alastair Haynes provided support with respect to the quantitative evaluation component, while Saltanat Rasulova provided qualitative support. Alex Hurrell provided overall strategic guidance and reviewing support.

The evaluation team was able to work freely and without interference throughout the implementation of this evaluation.

3 Background information to CLP-2

Volume I provided a very brief summary description of the programme we are evaluating. We use this section to provide more background information with respect to the different programme components of CLP-2. Please note that a related **light-touch TOC** is presented in Annex C. In addition, we use this section to provide a brief summary to list other programmes that were operating in the area of CLP-2 operations.

3.1 CLP-2 activities

CLP-2 was a complex programme that involved a variety of interventions. What follows aims at giving a concise description of the activities undertaken under CLP-2, which could broadly be categorised into four areas: infrastructural activities, livelihood activities, market development activities, and human development activities.

Infrastructure

The infrastructure activities under CLP-2 were comprised of two major components: the infrastructural development activities and the infrastructural employment programme (IEP).

The key activity under the infrastructural development component was **homestead plinth-raising.** In order to prevent flood damage, the dwellings of CLP-2 participants were raised on an earthen plinth at least 60cm above the highest known flood level. In addition to homestead plinth-raising, infrastructure development activities also included building **tube wells** (above the flood line, protected by a concrete platform, at a depth of at least 40 feet, 30 feet distance from a latrine and with a pumping head) for access to improved drinking water for participant households. Note that access to these tube wells was not limited to participant households but given to neighbouring households as well. Finally, improved **latrines** were also installed under the infrastructural activities programme of CLP-2. For the installation of these latrines, subsidies were given to all village households, irrespective of whether they were core or non-core households, to install a concrete slab fitted with a plastic pan and water seal.

The second component of the infrastructural activities, **the IEP,** was closely linked to plinth-raising. The IEP was designed for the poorest dwellers in the char islands who were most vulnerable during the hunger seasons. In order to create employment opportunities that could create income during these periods, plinth-raising was implemented in two seasons: January to June (the dry season) and September to December (*monga* season). Implementing partners were required to hire labour from the same or adjacent villages so as to offer job opportunities to char dwellers.

Livelihoods

The livelihood activities under CLP-2 had three broad sub-categories: the asset transfer programme (ATP), the homestead gardening programme, and livestock maintenance and training.

As part of the **ATP**, households in the operational cohort were given a choice to invest a grant for purchase of an income-generating asset worth BDT 17,500, whereby they could choose between acquiring cattle, a land lease or other assets such as sewing machines, rickshaws, etc. The assets were procured and then transferred to the households according to their preference. Note that these assets were assigned to the women in participant households.

A large majority of households chose to procure livestock. Additionally, these households were then provided with **training on proper management of livestock** as well **vaccination** and **deworming** of the livestock. There was also an **artificial insemination** project, which aimed to increase the productivity of the next generation of livestock and to maximise income from cattle.

The CPHHs under CLP-2 **also received a stipend for an 18-month period to prevent distress sales of their assets**. The households received BDT 600 a month for the first six months after asset purchase to aid the proper maintenance of assets and household income support, as well as BDT 350 a month for the remaining 12-month period for family income support.

Under **homestead gardening**, CLP-2 provided households with training for growing vegetable gardens and pits as well as composting training. Agricultural inputs such as quality seeds, compost and tree saplings were also provided, as was any additional technical support required. Poultry keeping was also promoted under this category of activities, with full field training provided for the development of poultry farms, and income-generating opportunities were provided for participants who trained to become poultry vaccinators.

Market development

In early 2013, CLP-2 revisited its initial approach to market development and developed the CLP-2 Market Development Strategy. This strategy branded itself as using the 'making markets work for the poor' approach (M4P), which was targeted at two markets – the **meat market (beef)** and **dairy cow milk market**, with associated fodder markets as cross-cutting markets.

The **Integrated Meat and Fodder Market Development Project (IMFP)** was aimed at the meat and fodder market in the char areas. This project included five major interventions: Char Business Centre (CBC) strengthening, Livestock Business Group strengthening, strengthening of local trading systems, strengthening supply chains and strengthening technology practices and commercialisation. Activities under these interventions included, for example, formation of business groups, training and linkage meetings held for the steady supply for feed and fodder, with initiatives also being taken to improve the supply network.

The second component of the market development strategy was the **Milk Market Development Project**, which was targeted at household milk producers in milk business groups with the objective of increasing cow milk productivity and profit. As with the IMFP, activities under this project were aimed at increasing productivity directly at the household level, while also improving market access and sustainable market development.

Human development

The human development interventions under CLP-2 could be sub-categorised into **social development, social protection, health, education, and village savings and loans groups** (VSLs).

Social development enabled the formation of social development groups and adolescent groups to promote health and sanitation messages, creating a community promoting the welfare of the villages, as well as governing bodies for monitoring and maintenance of CLP activities once CLP finishes.

Under **social protection**, community safety net programmes were developed whereby support was provided to the poorest households in the community via the combined contributions of households who were comparatively better-off.

The **education programme** has established learning centres providing school education as well as that contained an adult literacy component for the many individuals in the region who lack basic literacy.

The **health programme** was mainly based on providing primary health care and family planning services through satellite clinics conducted fortnightly. The clinics continued for 18 months for every cohort, and non-CPHHs in the region were also allowed to access these services. CLP recruited village health workers who were women and who lived in the community, allowing them to attend to patients at all times.

The **VSL** groups were formed to create a safe place for saving and borrowing for the participants through operating special community-based microfinance activities. The group size ranged between 15 and 25 participants and was managed through a group management committee within each group and by the underlying community.

3.2 Activities of other programmes being implemented in CLP-2 operation areas

As described above and in Volume I (OPM 2016a), CLP-2 has been operating in an area of Bangladesh that is very difficult to access. Many participant households live on very remote chars that are rarely visited from the mainland. Service provision by public or private institutions is infrequent. This also seems to have an impact on the incidence of other programmes being implemented in the area; our evaluation has found some evidence of there being other programmes in these areas but no programme could match the scale and scope of CLP-2.

The most relevant to our evaluation is the market for the chars (M4C) programme implemented by SwissContact, aimed at improving market conditions on the chars. We mention this programme and its interaction with CLP-2 explicitly in our local economic context analysis in Volume I.

We have also found evidence for other – smaller-scale – projects implemented in CLP-2 areas. For example:

- Shouhardo project – implemented by JSKS in 2010 in Gopaljhar. This project built shallow tube wells for irrigation purposes.
- In 2010 SHIREE distributed pumpkin seeds for cultivation in Gopaljhar.
- BRAC has also been implementing microfinance activities in Gopaljhar, although only for a limited period of time.
- The Sabolombi Society has been building sanitary latrines and tube wells in Afjalpur, through a programme running from 2012 to 2015.

Via our qualitative research component and conversations with stakeholders and implementing organisations (IMOs), we aimed at getting an understanding of the scope of these programmes. We found that none were large enough in scope to impact our answers to our evaluation questions significantly, with the one key exception being the M4C programme by SwissContact that was thus explicitly taken into account in our qualitative analysis. From a quantitative perspective, similarly, we do not expect any of these projects to significantly skew our results, given the limited scope and timeframe during which they were implemented.

4 The overall mixed methods approach

We present our mixed methods approach in Volume I (OPM 2016a). For the purposes of the comprehensiveness of this technical report, we reintroduce this approach here as well.

This evaluation has been implemented with the explicit goal of integrating quantitative and qualitative research. We have aimed to use the strengths of both quantitative and qualitative approaches in order to provide comprehensive answers to our evaluation questions. This has meant that the design and implementation of the quantitative and qualitative workstreams had to be carefully sequenced and adapted to each other, an approach that applied to all phases of this evaluation.

During the inception stage, the goal of integration was to bring both preliminary qualitative and preliminary quantitative data and results to bear on the overall design of the evaluation and planning of evaluation activities. As described above, this meant implementing two inception missions: one with a general focus on starting the evaluation process and setting priorities for the evaluation and one with a qualitative focus, which included working on a light-touch TOC. In addition, quantitative data analysis was frontloaded into the inception phase in order to inform the design of the qualitative research. **As a result of this, we adopted a research design that focused deliberately on triangulating some of our quantitative evidence, but also exploring some 'quantitative blind spots' qualitatively.**

During the implementation and analysis phase, the goal was to enable tight feedback loops between concurrent qualitative and quantitative work to support the fine-tuning of quantitative analysis and qualitative instruments and research design. As described above, one particular activity served as anchor point for this: a quantitative-qualitative workshop held in Dhaka in January 2016, in which descriptive quantitative findings were presented and resulting areas of interest for further qualitative research identified. The qualitative fieldwork started after that workshop, which meant that it could be adapted in light of the insights generated through the workshop.

During the reporting stage, the focus was on developing a narrative of CLP performance that draws on both the qualitative and quantitative findings. To this end, an intensive week-long workshop was held by the team in Dhaka in April 2016, where results related to the evaluation questions were discussed both from qualitative and quantitative perspectives and integrated findings were presented to stakeholders in-country. In particular, insights from the qualitative work were checked by reference to the quantitative data in order to provide further robustness to our conclusions. Report drafting was also implemented concurrently by the quantitative and qualitative workstreams and drafts shared among team members to ensure cross-method feedback. The result was a continuous iterative process – described in Figure 1 below – that brings each workstream to bear on the other through sequential revisions of qualitative results in light of the most recent quantitative findings, and vice versa.

4.1 How generalisable are our findings?

The consequence of this mixed methods approach was that, depending on the area of investigation, different types of evidence were employed to answer our evaluation

questions (see Section 2.1.). This also resulted in different levels of confidence with respect to the generalisability of findings, depending on the areas looked at.

In our context, generalisability refers to the ability to make statements for CLP-2 and participants as a whole based on our study sample, both quantitatively and qualitatively. With some methodological caveats necessarily kept in mind, the following generally holds: first, findings based on quantitative analysis are generalisable to the population of CLP-2 participant households, given that they are based on data coming from a representative sample of those households. Second, findings based on qualitative analyses are not generalisable in the same sense, given that sampling is not representative. Third, however, some qualitative evidence can be very strong and hence indicate that findings hold more widely, based on a combination of our sampling approach and repeated identification of similar results. For example, it could be that across all research sites and data collection instruments respondents indicated that one certain trend holds, which would give us a strong indication that this trend might hold more generally in the study area. Finally, when combining findings from both approaches, our qualitative research can build on the representative nature of quantitative results and we can hence make strong statements of a more general nature.

In practice, this means that, for our evaluation questions, the following holds:

- We have strong, robust and generalisable findings for questions related to the impact of CLP-2, which are based on a complex set of econometric analyses of CLP-2 household data. This includes findings around how CLP-2 affected poverty, income, consumption, cash savings and asset values that households held. Note that our efficiency analysis is also partly based on these findings.

- Similarly, we have strong and generalisable descriptive findings on changes in the lives and livelihoods of participant households, based on quantitative data analyses. This includes our analysis around CLP-2 graduation rates, saving patterns and how they vary over time.

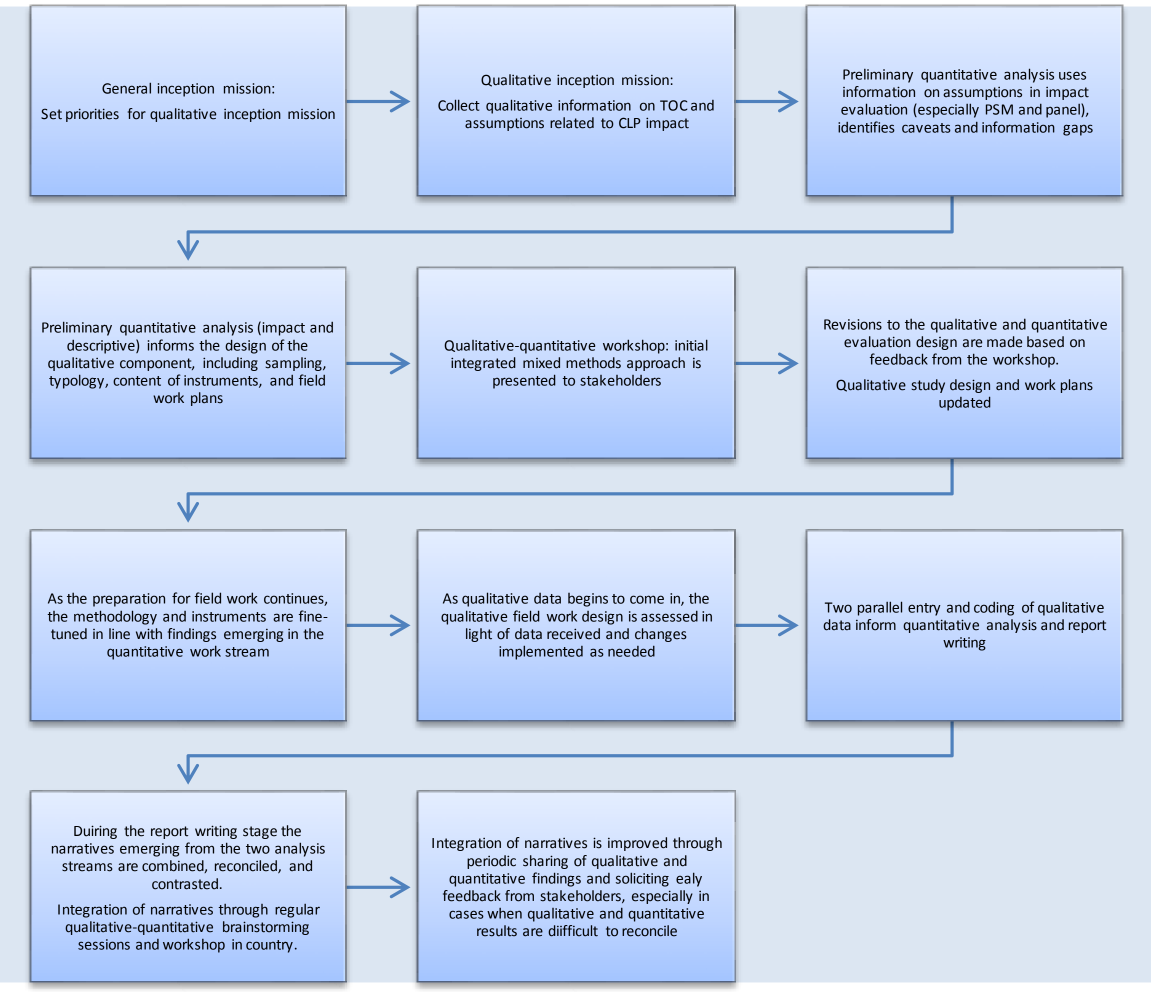
- In addition, some of our findings around the factors that support or inhibit CLP participant households from benefitting from the programme build on a strong combination of both quantitative and qualitative analyses and therefore hold more generally.

- Moreover, our qualitative analysis of the savings behaviour of participant households is corroborated by quantitative results and hence also supports general statements.

- Two areas of research have been addressed almost exclusively from a qualitative perspective in this evaluation: first, perceptions around changes in the local economic context and, second, questions around the empowerment of char dwellers. The evidence presented in Volume I with respect to those is therefore not generalisable to the wider CLP-2 population. In addition, note that qualitative sampling was not geared towards investigating these questions but rather to questions around variations in the level of wellbeing of participant households. However, the findings that we do present are the result of an intensive and rigorous analytical process and therefore do give a detailed indication of dynamics that possibly hold more widely.

- Finally, our findings on sustainability build on a combination of quantitative and qualitative analyses. We mention caveats with respect to sustainability explicitly, but the findings that we do present have a strong grounding in both our research approaches.

Figure 1 Overview of the mixed methods integration process



5 Qualitative component

The following section provides additional information on the qualitative component of the evaluation of a kind that is not contained in Volume 1 (OPM 2016a), although some of the information contained here can also be found in Volume I in an abridged and more accessible form. The aim of this section is to enable the reader to better understand the design, approach, analysis and limitations of the qualitative research. The aim is to be as transparent as possible and enable the reader to make an independent assessment of the quality of the design and execution of the qualitative analysis. A table with limitations resulting from the qualitative design can be found in Section 2.4 in Volume I.

5.1 Objectives and key research questions

Following discussions with DFID and with stakeholders during the inception phase and the production of the descriptive analysis of the quantitative data, it was decided that the qualitative analysis would pursue the following four main objectives:

1. **Explore the perceived reasons for differential levels of wellbeing** of participant households and whether certain household characteristics affect this.[3]

   The aim was to understand perceptions regarding: (1) what factors affect the ability of CLP participants to do well; (2) what factors lead households to fall back into poverty after initially doing well; (3) how the livelihoods of participants have changed following their involvement in CLP; and (4) whether certain household characteristics such as gender affect the likelihood of doing well and the sustainability of the change in wellbeing.[4]

2. **Investigate perceptions of the sustainability of the level of wellbeing achieved.**

   The aim here was to better understand perceptions of: (1) what factors leads to the sustainability of graduation; (2) which of the set of interventions or institutions are still in place and why; and (3) what perceptions CLP participants and non-participants have as to why this is the case. Here emphasis was given to understanding the broader scope of the sustainability of CLP that goes beyond just the household level and includes community-level factors.

3. **Explore perceptions of changes in the local economic context and the effectiveness of CLP's market development approaches**.

   The aim was to explore the following potential themes: (1) participant and non-participant views on the market development initiatives; (2) whether their attitudes towards livelihood

[3] This question is linked to the overall evaluation question of why some households are better able to graduate than others. However, as graduation is a constructed indicator, the qualitative research focuses on wellbeing as a proxy for graduation, as the term 'graduation' does not carry an understandable meaning for CLP participants and most IMOs.

[4] Whilst the scope of the evaluation did not allow us to deliberately explore issues around violence against women, dowries and child marriage, the researchers were trained to pay particular attention to these issues and explore them if they arose. As this was quite frequently the case, the findings on these topics feature prominently in the analysis despite originally not being part of the research objectives.

strategies and options have changed; (3) whether services and markets are available; and (4) whether they perceive the economic environment as having changed for the better. In addition, we also investigated: (5) the perceptions of traders and livestock service providers (LSPs) of CLP market development approaches and interactions with participants.

4. **Triangulate quantitative findings against qualitative data**, and explain how and why the salient trends or unexpected findings observed in quantitative data emerge.

   The aim was to: (1) triangulate findings from the quantitative analysis; and (2) provide additional explanations on why certain phenomena were observed (e.g. differences in the graduation level of participants, differences of savings levels over time, level of women's empowerment, level of economic development, etc.).

These research objectives were translated into the broad thematic research areas and research questions outlined in Table 2 below. This table was also used as the basis for developing the research guides for the different research tools and respondents. In each research location, four FGDs and eight KIIs were held.[5] In addition, KIIs were held with implementers of market linkages programmes operating in the chars. Whilst most FGDs and KIIs touched on a wide range of key research questions, the amount of time spent exploring the different questions differed significantly across groups and respondents. The aim of the table is to allow the reader to see to what extent findings were triangulated across the different respondent groups.

[5] In each research location, the team conducted FGDs with men from CLP participant households, female CLP participants, men from non-participant households, and women from non-participant households separately. In addition, KIIs were held with village leaders, Upazila Nirbahi Offices, chairman, VSL members, VDC members, traders, LSPs and IMOs. At the national level, KIIs were conducted with representatives from iDE, SwissContact and CLP's Markets and Livelihood unit.

Table 2 Research areas and key research questions

Research area	Key research questions	Source of information								
		FGDs		KIIs						
		Men and women from participant households	Men and women from non-participant households	Local elites	VDC and VSL members	IMOs	Traders	LSPs	District officials and politicians	National stakeholders
Perceptions on why participants differ in their ability to benefit from CLP participation (Objective 1)	What factors affect the ability of CLP participants to graduate?	x	x	x	x	x	x	x	x	x
	What factors lead households to fall back into poverty after initially graduating?	x	x	x	x	x	x	x	x	x
	How have the livelihoods of participants changed following their involvement in CLP?	x	x	x	x	x	x	x	x	x
	What are the main shocks faced by households, individuals and different social groups? How do shocks affect different households, individuals and social groups? Have the main shocks faced by households changed over time?	x	x	x	x	x	x	x	x	x
	What strategies to prevent, mitigate and cope have people adopted in order to deal with shocks? If eroded, what do people do?	x	x	x	x	x	x	x	x	x
	What effect do shocks have when they occur? Do households, individuals and different social groups experience shocks differently?	x	x	x	x	x			x	

	Do certain household head characteristics such as gender affect the success of graduation and its sustainability? Do some groups benefit more than others, e.g. male headed households?	x	x	x	x	x	x	x	x	x
	How do people conceptualise saving and decide on the value, method and saving goal? Why do some households save more than others, especially over time?	x	x	x	x	x				
Perceptions of the sustainability of the changes in wellbeing and the institutions created by CLP-2 (Objective 2)	What factors lead to the sustainability of graduation?	x	x	x	x	x	x	x	x	x
	How do people save and invest in assets? How has this changed over time and why?	x	x	x	x	x			x	
	Which of the set of CLP interventions or institutions are still in place and why do people think this is the case?	x	x	x	x	x	x	x	x	x
	To what extent have community-level barriers been sustainably removed or diminished for participants and non-participants?	x	x	x	x	x	x	x	x	x
	What household-level or community-level characteristics affect whether graduation takes place and is sustainable?	x	x	x	x	x	x	x	x	x
	What factors determine whether institutions and interventions continue to exist or show impact beyond the intervention cycle?	x	x	x	x	x	x	x	x	x
	Has CLP changed the perception and interaction of local government with char dwellers, especially women? If yes, how and why?	x	x	x	x	x			x	x
	What was the role of local government in the project and can they play a role in ensuring the sustainability of interventions?	x	x	x	x	x			x	x
	What are the main livelihood activities undertaken in the community? Do they differ for different households, individuals and social groups?	x	x	x	x	x	x	x	x	x
	What are the constraints and challenges associated with these livelihood activities?	x	x	x	x	x	x	x	x	x

Perceptions of changes in the local economy and of the market linkages approaches (Objective 3)	What is the perception of the market development initiatives by participants and non-participants, CBCs, LSPs and traders?					x			x	x
	What livelihood support programmes are available in the community? How do they affect people's livelihood strategies?	x	x	x	x	x			x	
	Have the attitudes towards livelihood strategies and options changed since CLP? Is there a difference between participants and non-participants?	x	x	x	x	x	x	x	x	x
	Do community members perceive the economic environment as having changed for the better?	x	x	x	x	x	x	x	x	
	What opinions and attitudes do traders and LSPs hold towards community members?						x	x		
	Are participants willing to pay for services?	x	x	x	x	x	x	x		x
	Has the number of traders and services increased in the chars?	x	x	x	x	x	x	x	x	x
	What is the relationship between customers and traders in the chars?	x	x	x	x	x	x	x	x	x

5.2 Data collection methods

A combination of **FGDs, KIIs and household-level case studies** was used for data collection.

In each research site, FGDs took place with participants and non-participants separately in order to better understand whether perceptions about the programme differed between groups. This was particularly important when investigating perceptions related to the local economy, the sustainability of CLP institutions and interventions, and whether demonstration effects had an impact. In addition, FGDs were organised with men and women separately. This allowed the team to explore issues around voice, empowerment and intra-household dynamics and whether certain household members experienced CLP differently in more detail and in an environment in which respondents were more likely to speak freely. In total, four FGDs took place per research site with these multiple contrasting groups, producing information that illuminates the distinctive perspectives, experiences and views of different stakeholders in the evaluation.

KIIs were conducted with selected participants who were identified because they could provide rich information on the workings of the programme, community dynamics, the local economy or the market linkages components of CLP-2. KIIs were conducted with village leaders, Upazila Nirbahi Offices, chairmen, VSL members, VDC members, traders, LSPs and IMOs. At the national level, KIIs were conducted with representatives from iDE, SwissContact and CLP's Markets and Livelihood unit.

In addition, researchers conducted two household-level case studies with successful and unsuccessful CLP participants and developed char profiles for each location. The profiles were based on observations of the environment and conversations outside the data collection process. The aim was to provide a rich description of the environment and record any additional findings and observations made during the fieldwork that might be relevant for data analysis.

FGDs

Focus groups are well adapted to those cases where the evaluation topics and issues to be addressed might provoke divergent opinions but where discussion may lead to a deeper and more considered viewpoint. The rationale is that the interaction of respondents will stimulate a richer response or new and valuable insights.

FGDs were organised with specific goals in mind and followed broadly similar structures, timeframes and procedures, which were set out in the FGD guides and refined during the training and piloting with the national research team. Discussion guides were developed for the FGDs that were tailored towards the research questions relevant to the different groups. The FGDs also used some unstructured elements, in order to ensure that unanticipated findings were elicited. Groups comprised six to 10 participants, which ensured that groups were large enough to ensure that no two individuals could dominate the discussion and small enough to ensure that all participants could still contribute in a meaningful way. The findings of the FGDs were triangulated through comparison of findings from one focus group with other focus groups held with different participants from the same group, as well as through comparison with findings from the KIIs to maximise the reliability of the findings. All FGDs followed a similar structure but explored slightly different themes in accordance with the evaluation question table above and the characteristics of the participants. FGDs started with an introduction by the facilitator and note taker where the programme and reason for the research were explained, participants' informed consent was requested, and the process explained. In addition, some basic characteristics of participants and their households were collected in order to ensure that the selection criteria for participants were applied in the field (see Section 5.3.1). The key research questions provided an organising structure for prompts and follow ups that allowed the team to investigate certain points in more detail.

KIIs

KIIs allow the researcher more time to explore issues in greater detail with an individual who holds specific information of relevance to the analysis, while also being a useful tool if the pool of individuals is not large enough to enable the organisation of a FGD.

Semi-structured interview guides were used for the KIIs that were organised around the key research questions. This ensured a degree of standardisation whilst at the same time allowing the team's qualitative researchers enough flexibility to pick up on interesting themes, topics and concerns as they emerged. The guides were developed based on the key areas of interest and key research questions presented above.

As a rule, key informants with different positions and perspectives bring their own sets of interpretative biases to the analysis. In this type of qualitative research – where there is no single absolute truth and where difference (rather than standardisation) is actively sought – trustworthiness in interpretation can nonetheless be strengthened by cross-checking or triangulating the views and analysis of different key informants and focus groups. The qualitative design ensured this through the wide range of key informants interviewed and the cross-checking of answers with the quantitative findings and the probing of similar topics in the FGDs.

5.3 Strategies for ensuring rigour

A major methodological challenge in qualitative research is the definition and achievement of 'rigour'. Qualitative research is often accused of being open to research bias or anecdotal impressions, impossible to reproduce and difficult to generalise (Mays and Pope 1995). In order to ensure a robust evaluation methodology, OPM applies a range of **strategies for ensuring rigour when conducting qualitative methods**. Rigour is conceptualised as the trustworthiness of qualitative research (Lincoln and Guba, 1985) and OPM follows a protocol of ensuring rigour throughout the research by implementing specific strategies at each stage of the evaluation process – design, sampling, data collection, analysis and writing up. The main aim of these strategies is to minimise a single researcher bias and to be transparent in demonstrating the research process as well as data analysis.

Throughout the following sections, the adopted strategies for ensuring rigour will be discussed as they relate to sampling, fieldwork and analysis. In addition, the team also adopted an approach based on reflexivity throughout the project. In other words, in addition to conducting research in a way that meets the principles of how to conduct ethical research to avoid bias, which included an awareness of previously held views, researchers' backgrounds, etc., the team also continuously reflected on whether the research design and approach were the best way of tackling the research questions. Frequent discussions within the qualitative team, with other qualitative researchers not part of the project team and with the wider project team were sought, with the aim of ensuring the credibility of the research approach and execution and of providing independent peer review of the process and findings.

5.3.1 Sampling

Methodological rigour in qualitative research is not best established through a statistically representative sample because results cannot be quantified and aggregated in the same way as quantitative data can be. Rather, as in quantitative research, rigour in qualitative research can be achieved through a 'systematic and self-conscious research design, data collection, interpretation and communication' (Mays and Pope 1995, p. 110). As a rule, the sampling strategy should be driven by the research objectives.

Qualitative research sites were purposefully selected using extreme case sampling. This approach was adopted because the team needed to meet research objective 1 – gaining an in-depth understanding of the perceived reasons for why levels of wellbeing differ between CLP participant households. As mentioned above, objective 1 is linked to the overall evaluation question of why some households are better able to graduate than others. However, as graduation is a constructed indicator, the qualitative research focuses on wellbeing as a proxy for graduation as the term 'graduation' does not carry an understandable meaning for CLP participants and most IMOs. In order to get as close as possible in meaning to the CLP-2 graduation criteria, researchers were trained to describe wellbeing along the dimensions that comprise the indicator.[6] Purposely selecting research sites with a comparatively high and low likelihood of finding households that did well from CLP participation ensured a high degree of variation in respondents' experiences and allowed for the collection of rich information on why graduation outcomes differ. Moreover, this enabled the team to explore what contextual factors beyond the households differed across locations. The advantage of extreme case sampling for research sites is that it yields rich data on a specific phenomenon of interest (in our case the likely failures and successes of graduation) and is particularly well suited in cases where the time to conduct research is limited. In addition to meeting research objective 1, the sampling strategy had to also ensure that qualitative data were collected on the same questions and for the same cohorts as was the case for the quantitative impact analysis, in order to allow for the triangulation of findings.

Given the relatively broad scope of the qualitative research objectives, it was decided to select six sites from three cohorts, striking a balance between having an adequate amount of research locations and stretching the team and resources too thinly for gaining nuanced and triangulated findings. We believe that this provided sufficient data to be able to make meaningful observations within the limited available time span whilst at the same time allowing for sufficient in-depth research per site so as to meet the research objectives.

It should be noted that extreme case sampling does not lend itself to making generalisations across the universe of CLP participants. This sampling approach therefore further limits our ability to generalise, a point very clearly stressed in Volume I (OPM 2016a) where the results were presented.

The following four steps were followed for sampling the qualitative research locations and respondents:

[6] A full description of the graduation criteria can be found in Table 6 in Volume I (OPM 2016a).

- Step 1: Selection of three cohorts

  Cohorts were selected in order to meet two objectives: first, in order to investigate the sustainability of interventions and institutions over time, at least one of the first two programme cohorts had to be included. Second, in order to allow for triangulation of quantitative findings, cohorts that the quantitative impact analysis (i.e. the PSM component) focused on had to be included.

  As a result, cohorts 2.2, 2.4 and 2.5 were selected. As described in Section 6.4.2, the PSM component of the impact analysis could only estimate cohort-specific impacts for cohorts 2.4, 2.5 and 2.6. The team therefore decided to conduct research in sites where cohorts 2.4 and 2.5 were rolled out in order to ensure we would be able to triangulate related quantitative findings. Cohort 2.6 was excluded, as implementation is very recent and was rolled out in a shortened timespan. The area in which cohort 2.2 was rolled out was chosen, as the interventions came to an end a relatively longer time ago and the team could therefore explore the post-intervention sustainability of graduation across a longer time horizon. We chose not to go to cohort 2.1, as we understand that CLP-2 was still finalising some of the details of the intervention during the roll-out of cohort 2.1.

- Step 2: Selection of six upazilas

  Survey data was used to select six upazilas using 'extreme case sampling' (sampling high and low performing upazilas from the cohorts with respect to the likelihood of finding households that meet the CLP-2 graduation criterion. Note that this graduation criterion is a composite indicator that consists of 10 underlying graduation criteria). This step was implemented as follows: first, using the CLP-2 programme graduation criterion, the quantitative survey data were used to produce an estimate of the proportion of sampled households that meet this criterion, by upazila and cohort. For each of the cohorts listed above (2.2, 2.4 and 2.5) the upazilas with the highest and lowest proportion of households meeting the CLP-2 graduation criterion were identified. In a second step, if there was a tie across upazilas in the proportion of households meeting the graduation criterion for any cohort, we looked at the number of underlying graduation criteria that were met, on average, by sampled households in any upazila. We then selected the cohort where sampled households met, on average, the highest or lowest number of these underlying graduation criteria.

- Step 3: Selection of six chars

  Six chars were selected within the sampled upazilas, again using 'extreme case sampling'. As no survey data were available, IMOs were asked to select a char with either very high average levels of wellbeing or very low average levels of wellbeing. As graduation is an indicator constructed on the basis of the quantitative monitoring data it does not carry any meaning with IMOs or government administrators and officials. The qualitative research team therefore had to refer to the 'high' and 'low' levels of wellbeing of CLP participants, as described above for this sampling step. The information provided by IMOs was then verified by upazila administrators and by the team upon arrival on the char in order to avoid selection biases. In addition to finding chars that were high or low performing with regard to the likelihood of finding successful CLP participant households, the team also required the chars to be of a sufficiently large size in order to ensure that enough respondents were available. This was important for ensuring that it was actually possible to conduct the in-depth case studies of the locations that the qualitative design relied upon.

- Step 4: Selection of respondents

Participants were purposely sampled in order to generate responses from a small number of individuals and groups that are representative (not statistically) of the groups the programme tried to affect. The sampling of participants was purposive and stratified. This meant the team targeted the same specific segments of communities across all research locations. Unlike for the selection of research locations, respondents were not sampled on the basis of whether they were successful or unsuccessful CLP participants but in order to meet other criteria set out in Section 5.2. Whilst respondents were identified ahead of time, the research team retained a degree of flexibility to allow them to potentially react to emerging issues of interest by conducting additional FGDs or KIIs that provided information of relevance in a certain location. This was particularly important given that research sites were sampled from different programme cohorts.

Respondents for FGDs with members from CLP participant households were selected following two strategies: first, potential participants were identified with the help of IMOs, VDC members and community leaders (depending on the context). Second, the team talked to all suggested participants individually to check for selection biases and diversity within the sample. This was possible due to the fact that the first day in each char was purely used for rapport building and mobilisation. Participants for the male and female non-participant FGDs were selected with the help of community leaders who were asked to select participants from different socio-economic backgrounds and of different genders. Whether these criteria were followed was verified by the researchers whilst taking the register prior to beginning the discussion. Snowball sampling was used as a last resort in order to mobilise participants if the other approaches failed. Most KIIs were with specific individuals and thus required no active sampling, with the exception of the village leader. The criteria for an individual being a village leader was whether char dwellers reported seeking out this individual for advice and for conflict resolution.

5.3.2 Brief profile of research locations

In order to preserve the confidentiality of respondents, the selected chars are referred to by a pseudonym throughout the report and in the table below, i.e. the cohort they are from and whether they perform well at reaching CLP-2 graduation criteria (i.e. are high performing) or not so well (i.e. are low performing). The table below provides an overview of selected key characteristics of our research sites and a brief profile for each location for the reader's reference.

Table 3 Overview of qualitative research sites and selected characteristics

	Cohort 2.2		Cohort 2.4		Cohort 2.5	
	High performing site	Low performing site	High performing site	Low performing site	High performing site	Low performing site
Upazila in which char is located	1	2	3	4	5	6
Proportion of households meeting CLP-2 graduation definition (reaching six or more out of 10 indicators)[7]	100%	15%	100	57.7	100	66.7
Number of CLP-2 graduation criteria met (out of 10)[8]	7.7 out of 10	3.9 out of 10	9 out of 10	6 out of 10	7.5 out of 10	6 out of 10
Connectivity with mainland	Relatively good in dry and rainy season	Very poor in dry season and moderate in rainy season	Relatively good	Relatively bad, especially in dry season	Relatively good	Not very remotely located, but nonetheless difficult to reach in dry season due to sandy paths
Exposure to floods and erosion	Flash floods	Severe risk	Relatively low risk	Severe risk	Low risk	Severe risk
Increase in traders visiting char	High	High	High	Low	Low	Low
Uptake of LSP services post-CLP	High	Moderate	High	Moderate	High	Low

Cohort 2.2 – high performing site

The char is situated 37km from one upazila, 15km from another upazila and 3km from the nearest union. The char rose in 1971 and is relatively protected from river erosions due to an embankment built in 1972. As a result of its low susceptibility to river erosion, the land has become increasingly fertile and higher due to silt deposits. Flooding is rare and mostly comes in the form of flash floods. There are no roads in this village and during the dry season walking is the main mode of reaching the mainland. Lack of transportation and connectivity with the mainland is perceived as one of the main reasons the village is lagging behind in terms of achievements in education, health, etc. There are three primary schools located on the char and most char dwellers send their children to school. However, for higher education the children have to travel to the mainland. There are no health centres on the char. When they are ill, people have to travel to the nearest community clinic, which is approximately 3km away. Access to mainland markets is poor. However, the number of traders visiting this char has increased since CLP started.

Agriculture is the main mode of livelihood in this village. Around 40% of people reportedly own agricultural land, whilst others lease land. The land here is fertile and a range of crops can be grown here, including wheat, corn, paddy, tobacco and jute. The char and the village were also part of other development programmes prior to CLP and have received training in various

[7] These figures present averages at the upazila level and refer to the second stage of the qualitative sampling process.

[8] These figures present averages at the upazila level and refer to the second stage of the qualitative sampling process.

agricultural and rearing practises, which further popularised farming. Some people also work as day labourers outside the village. Typically, almost every family has one or two male family members who work outside the village or migrate in search of work. At present, women also work outside the home either on their own field or as agricultural day labourers. Women now cultivate the land if their husbands or other male family members are currently not in the village.

Cohort 2.2 – low performing site

The char is situated in a union that is surrounded by the Brahmaputra and Jamuna rivers. Out of the 10 wards that make up the area, nine have been affected by river erosion of different degrees of magnitude. This happens every year and also includes the ward within which the char is located. The area is also susceptible to flooding during the monsoon period when crop fields tend to be flooded, thereby affecting the livelihood of farmers in the area. In 2015, there was a large flood that led to significant river erosion and about 1,400 families in the union lost their houses. Most of the people from the char were affected by this river erosion. Many CLP participants lost their houses in the river erosion and had to live in temporary houses next to the main road. The public transport system – in the form of both roads and waterways – to get to the char is very poor. In order to reach the char, people have to travel to a quay that is about 2.5 hours away from the nearest town and then take a boat to the char, which takes about 1.5 hours. During the dry season, people can use horse-drawn carriages to reach the char but during the rainy season the char can only be reached by boat.

There are no schools on this char except one religious school and most students have to travel to another village close by in order to attend primary school. There are no high schools or colleges in the union and students have to go to boarding schools and colleges in faraway areas to access higher education.

Roughly 40% of the land in the union remains uncultivated due to the high sand content of the soil. In the remaining 60%, mostly maize, chili, nut, jute and paddy is grown. Paddy cultivation requires higher irrigation costs due to the sandy soil and hence most people prefer cultivating maize and chili and nuts. Agriculture is the primary occupation on this char and many dwellers either farm their own land or lease land for farming. Those who do not own land tend to migrate to Dhaka, Rangpur and Dinajpur during the flood season in order to work as rickshaw pullers or day labourers. On the char, people who do not own land work as agricultural day labourers on other char dwellers' land. Prior to CLP, women used to only do household chores. However, this has now changed and women currently work in the fields along with their husbands to save on labour costs. Some women also work as agricultural day labourers, whilst others work as tailors. There are only two shops on this char and both are run by former CLP participants.

Cohort 2.4 – high performing site

The char is located in the most remote area of the upazila and is surrounded by the Brahmaputra River. The area is susceptible to both river erosion and floods. Whilst floods are more frequent, erosion causes more damage to people's livelihoods. There are two housing projects in the area, one of which was started after CLP started. They provide housing for victims of river erosion from various chars as well as landless families from the area. Currently, 155 families out of the 160 families living in this accommodation project are former CLP participants who faced erosion before CLP started operations in the area. The houses built under the project are semi-structured and have tube wells and toilets, meaning CLP therefore did not have to provide such facilities in these cases. Boats are the main mode of transport to and from the char.

The char has one primary school, but teacher attendance is infrequent due to the distance they have to travel and the difficulty in reaching the school. During floods the school remains closed. As a result, many children of char dwellers are actually enrolled at school on the mainland. However, these pupils face difficulties in going to school during the rainy season as they have to cross the river in order to get to school, which poses a risk many parents are unwilling to accept. The union has one secondary school and no colleges. Families who can afford to do so have to send their children to live in faraway areas in order to continue their education. There is a pharmacy on the char and a local village doctor. In addition, trained midwives live on the char. However, for serious illnesses or emergencies residents have to go to the upazila health complex at the mainland hospital.

The main crop farmed on this char is maize, which is relatively well suited to the sandy soil and easier to grow under these conditions than paddy and jute. Paddy requires high irrigation costs. Farmers therefore do not typically chose to grow it and only harvest it in limited quantities twice a year. In the past, most of the land was uncultivated. However, this changed roughly eight years ago when char dwellers received training on irrigation and the planting of maize. As a result, maize is now widely cultivated and is viewed as a profitable crop by residents on this char. Nuts, sweet pumpkin and other pulses are also cultivated in the sandy soil area. Most men on this char grow maize and many CLP participants used the money saved or gained from the sale of CLP assets for purchasing land for maize cultivation. Men who do not have work during the rainy season tend to migrate to Dhaka, Chittagong or Feni to work as rickshaw pullers and day labourers. Younger men tend to migrate while elder men remain on the char and farm and look after the livestock. Whilst women used to only engage in household chores, CLP changed this and women are now looking after the livestock provided by the programme, producing milk or working on the land.

Cohort 2.4 – low performing site

The char is 5km long and 2km wide and lies in the Jamuna River. There are three paras (neighbourhoods) in the village. The char rose out of the river about 20 to 30 years ago and houses around 270 households and a total population of approximately 1,200 to 1,700 people. The char is affected by river erosion every year, including six months after CLP started operations. Every year several families are affected by erosion and loss of property, resulting in relocation costs. Floods also happen at regular intervals. Moderate floods typically affect 85% of households, whilst severe floods affect all houses on the char. Most of the plinths built by CLP were affected by the river erosion that occurred six months after they were installed. During floods mobility becomes comparatively easier as people can travel by boat but households face acute shortages of food and work as a result of the flooding. During the dry season it takes two hours to reach the mainland while in the rainy season boats are the only mode of transport and also require two hours.

There is only one primary school on the char and it is in a frail condition with poor attendance of teachers due to the difficulty of reaching the char. Two guest teachers, who live on the char, were recruited to cover for the absent teachers. There are very few students who go to secondary school and those who do are sent to boarding schools on the mainland.

The char has four grocery shops where people can buy basic necessities. Some families who sell fish or other commodities go to the mainland market where they then also purchase their daily necessities. Around 95% of the inhabitants of this village are landless. They lease lands from the local landowners and build houses. Fishing is the main source of income for people living here. In addition, some lease land for farming, work as day labourers, work in handloom industries or as day labourers or drive vans in the cities. When there is shortage of fish in the river, the men migrate in search of jobs. Women on this char continue to mostly focus on household chores and do not really work outside their houses, with the exception of growing vegetables and rearing cattle and poultry. The main crops grown are paddy, jute, sesame, nut and red gram. However, the constant exposure to erosion has drastically reduced the land available for farming.

Cohort 2.5 – high performing site

This char is around 1.5km long and 1km wide and is located at the Brahmaputra River crossing in the north and northeast side. During the rainy season, flooding and erosion are common on the north side of the char and there are floods every year that damage crops and affect access and interaction with the mainland. In some parts of the char, the soil is too sandy for farming.

Some families on the char live in government demesne lands (khas land) while some families lease their land from landowners. If houses are affected by floods or erosion, the owner has to reimburse a portion of the amount paid for leasing the land. There are no health facilities on the char and people have to go to the mainland for treatment and to buy medicines. The char dwellers can reach the mainland by walking 5 to 6km during the dry season or by using boats during the monsoon period. However, transportation is difficult, especially during medical emergencies when pregnant women who have to be moved in auto rickshaws or carried. There are no primary schools on the char and the nearest primary school is located near the market 5 to 6km away from the char. During the dry season students walk to school but during the monsoon period they have to use boats. For higher education children have to travel and in many cases live in the nearby towns to continue their education.

Many char dwellers in this area migrate abroad, particularly to the Middle East, and most of the male population works in agriculture or as day labourers. Some farm on their own lands while other lease land for farming. The main crops grown on this char are potato, eggplant, corn, red gram, nuts, masalai pulse, sweet potato, banana, pumpkin, guavas, chili, mustered, koun, nuts and black cumin. Jute is only grown during the rainy season. Due to the high sand content of the soil, paddy cannot be grown here. During the monsoon period when opportunities for work are limited, men migrate to the cities to earn a living. Most of the women on the char are housewives and are involved in rearing domestic animals and homestead gardening. However, some women work in the fields and grow chili and potatoes and collect crops.

Cohort 2.5 – low performing site

This char is relatively new and people from neighbouring chars that were eroded have recently moved here. The char disappeared in 2000 and reappeared in 2010. Re-settlement started from 2012 but parts of the char were again eroded in 2014. In addition, the severe floods that year also greatly affected the livelihood of people living here. Floods are a frequent phenomenon and low-lying homesteads are flooded. Only CLP houses are higher as they are raised on plinths and so are not eroded and remain safe from flood waters. Whilst reaching the mainland is easier during the rainy season, people then also tend to suffer from food shortages and a lack of work. In addition, people have to bear the additional cost associated with repairing their houses and buying new furniture to replace flood-damaged items. People spend the entire year carefully tending to the land and increasing its productive capacity, all of which is then destroyed by floods that carry off the top soil. The char only has one grocery shop and this is closed most of the time; even when it is open, it only sells a very limited variety of goods. As a result, people typically have to go to the market in order to buy goods. During the dry season this takes one hour over sandy paths. During the rainy season, the mainland can be reached in 15 minutes by boat.

The nearest market is the biggest bazar amongst all nearby unions. In addition, the bazar is close to the Indian border and hence attracts a lot of different traders. Whilst links to the Bangladeshi mainland are poor, char dwellers report feeling well integrated into trade with India.

There is only one primary school on the char and this is in a very poor condition. Students face difficulty continuing their education due to bad roads and transport systems. In 2014, 293 households lived on this char. Following the erosion that year, many households had to move away. Currently there are only 65 households on the char. Roughly 80% of people work as agricultural day labourers and work in other districts. Only during the three months of the rainy season do these men remain on the char. Some char dwellers also migrate to work in garment and cement factories. About 10% of char dwellers are engaged with agriculture, not as day labourers but by cultivating their own land or through sharecropping or leased land. The main crops produced on the char are kalai pulses, nuts, wheat, etc. Whilst all women work on household chores, homestead gardening and rear poultry and livestock within their households, only about 10% of women work outside their household as day labourers. In households that cultivate their own land or through share cropping/land leasing, women also help in the farming, for instance through sowing seeds, etc.

5.3.3 Training and fieldwork

Rigour and the avoidance of bias in the qualitative fieldwork was achieved through extensive training and the involvement of different individuals in the field teams, so that the teams can provide checks on each other. In addition, the teams kept records of their activities, so that they could be linked to the transcripts and analysis.

<u>The team</u>

The qualitative research was conducted by a team comprised of 17 people (eight men and nine women), 16 of whom are Bangladeshi nationals. Professor Ferdous Jahan from Brac University led the qualitative research, supported by Stephanie Brockerhoff and Tahera Ahsan from OPM. The data collection and analysis in Bangla was conducted by a team of 14 researchers from the dRI, led by Mamun-ur-Rashid and Omar Faruque Siddiki, who were overall responsible for the fieldwork. The data collection was conducted by 12 researchers who were divided into three teams of two men and two women each.

The team was deliberately assembled so as to be composed of members with a rich understanding of Bangladesh, CLP and other related programmes and international experience. CLP participants are women who live on chars. As a result, a lot of the respondents in the research were women. At the same time, char society is relatively patriarchal and conducting fieldwork in these remote and hard-to-reach areas can be challenging and – at times – unsafe. In order to be able to address both these challenges and ensure that respondents are able to work with researchers they are most likely to feel they can trust and provide truthful answers to, the fieldwork teams were composed of an equal number of men and women. In addition, the qualitative experts leading the team were all women; this provided a good balance in relation to the larger project team, which was more male dominated.

Training

A five-day training session was held at dRI's office in Dhaka and led by Ferdous Jahan and Stephanie Brockerhoff. During the first three days the objectives and key research questions were presented to the team and further developed, drawing on the experience of the 12 dRI researchers and their two team leaders, all of whom have extensive experience of conducting this type of research in Bangladesh. Following these discussions, the draft research guides were then further refined. In addition, the training also covered basic ethical principles of conducting research in a respectful manner and an extensive discussion of the biases held by researchers in the room, what impact this might have on data collection and analysis, and how to work toward preventing them affecting fieldwork and analysis. Finally, research protocols for sampling and fieldwork were developed. All research guides and protocols were tested during a one-day pilot and this was followed by another day of training during which final changes were made to the research guides and protocols based on the findings from the pilot.

Fieldwork

Three teams were formed, with one individual in each selected as a team leader. Each team had two male and two female researchers. The two dRI team leaders visited the different teams and provided additional support, as well as feeding information on findings and challenges experienced in the field back to the qualitative team. The fieldwork model was based on the assumption that five days would be spent in each research location. Every team was equipped with voice recording devices, note taking equipment, at least two laptops and an internet dongle. Informed consent was sought from all respondents and if permission was granted the KIIs and FGDs were recorded. Researchers explained that findings would be treated confidentially and that all answers would be anonymised.

Researchers were asked to write comprehensive notes (near transcripts) for each KII, FGD and case study and these notes had to be completed before a team moved to the next research site in order to avoid confusion of data. Once completed, the comprehensive notes were sent to the qualitative experts who would review them and ask for further clarification and detail if needed. The researchers recorded the discussions as they took place, using the language that is used and recording areas of particularly strong agreement and disagreement. In addition, they took note of shifts of attitudes within the room or other observable phenomena that cannot be captured in recordings but that might affect the data (e.g. if an individual respondent is very dominant during a discussion). Notes and recordings were transcribed and sent back to Dhaka for translation. This was important, as it enabled the qualitative experts to keep abreast of the fieldwork, discuss issues and new findings with the team, and start the process of reviewing and translating the notes.

Sometimes, however, the remoteness of chars and the lack of electrification led to delays in submitting fieldwork reports. In addition, some chars were very difficult to reach, which made finding accommodation time consuming and challenging and affected the time available for conducting research, as researchers typically had to either stay on a char overnight or leave during daylight for security reasons. In one community, the researchers had to stay with a char household due to the extreme remoteness of the char. The host households were paid an amount determined through a negotiation between the researchers and the host households.

Ethical considerations in carrying out research with vulnerable groups

Conducting qualitative fieldwork requires high ethical standards in order to ensure that expectations are not raised, confidentiality is maintained and respondents are never forced to participate or encouraged to speak about subjects that may be traumatising. We drew on our experience of qualitative fieldwork to ensure that these standards were met, and sought further review where appropriate, to adhere to ethical protocols in line with the OECD DAC principles of accuracy and credibility. It should be noted that, as these communities were part of CLP-2, they have already grown used to researchers regularly coming into the community as part of the programme's own data collection for monitoring purposes. As such, the presence of researchers is not a new and unknown feature.

There are a number of ethical issues to consider when planning and facilitating the participation of people, especially vulnerable groups, in research. They include:

- **Ensuring that participants are selected in such a way that there is no deliberate exclusion on the basis of stigma, access or gender and that cultural and community norms have been understood and considered in the selection process.** This was ensured both through the sampling strategies for respondents and by having an experienced gender-balanced team of researchers who have extensive experience of conducting research on the chars.

- **Ensuring voluntary participation and seeking permission for the research to go ahead. People have the right not to participate and to end participation at any stage. People should be asked to volunteer to participate and where possible pressure should be avoided.** The importance of these principles was discussed at length during the training and researchers ensured that no pressure was exerted on participants during the fieldwork. The first day of fieldwork was allocated to mobilisation in order to ensure the team had sufficient time to gain buy-in from char dwellers and answer questions about the study's purpose.

- **Clearly explaining the purpose and limits and the time required to conduct the research. The fieldwork should be designed and organised in such a way as to minimise the demand on participants' time and disruption to their daily routine.** The research was stretched over five days in order to maximise the team's ability to arrange discussions and interviews at a time convenient to char dwellers.

- **Ensuring respect of participants. Participants might come from vulnerable groups and the research has to be carried out with full respect. Power differentials will exist between community members and researchers and between the different participants.** These issues were explicitly dealt with during training in order to ensure respectful behaviour on the part of the research team and heightened sensitivity towards these issues. FGD groups were organised keeping power differentials and gender conflicts in mind. In addition, the gender composition of the teams ensured that FGDs with women were held by female researchers in order to increase the likelihood of receiving truthful answers in a protected and respectful environment.

- **Use of appropriate language (terminology, dialect and language) in order to ensure that views are correctly captured and understood.** The choice of wording was discussed both during training and piloting.

- **Guaranteeing the physical security of participants at all times. This entails ensuring that the environment is safe, that two facilitators are present at all times, and – where possible – that a supervisor is present during fieldwork operations.** The FGDs and KIIs were conducted in teams and the communities were informed of the identity of the dRI team leader and provided with their contact details so as to be able to reach them in case of need.

- **Protecting and respecting the right to privacy of participants. This includes ensuring anonymity and confidentiality of record keeping and report writing, and ensuring that participants fully understand that what they say in KIIs and FGDs will remain anonymous.** Prior to starting the FGDs and KIIs, the researchers explained both concepts. All information in the report is anonymised and information is treated confidentially.

5.3.4 Analysis of data

Rigour in the analysis of the qualitative data comes from three principal sources. First, findings were triangulated against different data sources, both qualitative and quantitative. Second, different members of the team conducted and discussed the analysis, reducing the possibility of individual researcher bias. Finally, the analysis was subject to peer review.

Frequent debriefs in the field to ensure reflexivity, discuss findings and emerging new trends. The researchers were trained to conduct regular debriefs in the evenings, during which findings were discussed and grouped under the key research questions. In addition, emerging trends were analysed by the group and the research guides adapted in response if needed. In the daily debriefs, each team would present their findings of the day, which were then discussed and recorded by the group. Research gaps that needed addressing during the next day were identified and any possible biases on the part of researchers discussed on a daily basis. This information was then relayed to the qualitative experts who decided whether certain issues needed further discussion and whether additional emerging themes should be explored in more detail. Following this decision-making process, the team leaders were informed and relayed the information to their team. This was done in order to ensure that there was a mutual exchange of ideas, findings and new themes across the different teams and research locations.

Analysis of initial findings at a three-day workshop with dRI researchers. Immediately following the return from the field, researchers met with Ferdous Jahan to discuss the emerging findings from the fieldwork. The team worked by research area and discussed findings for each of the key research questions. In particular, researchers used the daily debriefs produced and their comprehensive notes from the fieldwork to flesh out findings for each research question and differences in findings between cases and groups. Possible explanations and further areas for systematic investigation were also highlighted during this process. These findings were then documented in a report and translated into English in order to be shared with the rest of the team.

There are two crucial benefits to holding a workshop with the researchers who conducted the fieldwork: first, whilst the analysis is perhaps less rigorous than systematic coding it allows the team to benefit from the wealth of information the researchers who actually conducted the fieldwork hold. In addition, it ensures that no nuances are lost during the translation process and the document produced can be used as a validation tool for the later, systematic analysis. Second, with all teams having jointly conducted an initial analysis of their findings it is then easier later on to validate the findings from the systematic analysis with the team.

Systematic coding of comprehensive notes in English. Following the three-day workshop, all comprehensive notes from all research locations were translated into English and a rigorous coding system by evaluation question and contextual information was developed. It is important to note that uploading data into qualitative analysis software (in our case Dedoose) does not in itself analyse the data. Whilst some rudimentary data visualisation tools are available, these do not support a thorough analysis of the data. Using this software is merely an effective way of storing and organising the data so that they are easily accessible for the analysis process. The best coding practice ensures that coding labels themselves are constantly reviewed so that, if information in certain codes begins to overlap, those codes can be merged as ideally codes should be mutually exclusive. Coding aims to classify all of the data so they can be compared systematically with other parts of the data. The team initially started developing codes by research area. Each key research question was assigned the status of a parent code under which codes were later developed based on findings within the data. The idea was to not be too prescriptive but rather to allow findings to emerge from the data. Two team members worked on the coding of the transcripts. As mentioned above, Dedoose was used and was chosen because it allows team members to work on the transcripts simultaneously and in different locations. In addition, codes created by one researcher are immediately visible to the other researcher, thereby further minimising the potential for different coding behaviour. In order to ensure consistency, both team members initially coded the same two transcripts and compared the codes created and applied. After the completion of coding for one research location per researcher, the team met to discuss the codes created. Codes used were grouped together into categories and clearly defined until agreement on code application was achieved. As expected, additional codes emerged throughout the coding process and the categories had to be frequently refined and adapted.

Thematic analysis of data. The team analysed themes at both the explicit (semantic) and to a lesser extent at the interpretative (latent) level (see Braun and Clarke, 2006). The aim was to both confirm and negate the existing hypotheses that gave rise to the key research questions and at the same time to also explore new and unexpected findings as they emerged in the data.

Applied thematic analysis requires that researchers interpret the data collected, i.e. the textual record of the transcribed and translated KIIs and FGDs. It does not rely on counting words or phrases but identifies and describes the implicit and explicit ideas that were organised into themes. The set of hypotheses that form the evaluation or research questions provided the initial set of themes that were used for the confirmatory analysis. These initial themes were used to organise the data into groups (using Dedoose) that are relevant to these themes and which either confirm or deny the hypotheses. The validity of each piece of data is considered in the light of the context it came from (for instance, the knowledge that the person cited is likely to have about the subject, the incentives they may have to respond in particularly ways, and triangulation from other qualitative or quantitative sources). The researchers then assessed the balance of these groups and whether the conclusions supported the initial hypotheses. The first stage of this analysis was conducted in the de-brief sessions in the field described above and by researchers who actually conduct the qualitative fieldwork, to help ensure that errors of interpretation were minimised. The analysis was led by our qualitative expert, who is a Bangladeshi national as thematic analysis requires a good and detailed understanding of the data and context, the richness of which can at times be compromised by translation. The experienced qualitative fieldworkers and the qualitative expert all speak Bangla and know the chars well, which minimised this concern for the project and allowed the team to refer back to the original recording or comprehensive notes in Bangla to verify findings when necessary.

In addition, new themes were generated from the qualitative data through exploratory analysis on the basis of unexpected ideas. These new themes were used to develop new hypotheses and codes that were tested in similar ways to those set out above under applied thematic analysis. This exploratory analysis is a useful tool for ensuring that all relevant themes and hypotheses are identified and explored.

Validation of findings with researchers and triangulation of findings with quantitative team members during the one-week workshop in Dhaka. Following the analysis of the qualitative data, draft conclusions were shared with the qualitative researchers who conducted the fieldwork. Findings were validated and areas of disagreement discussed. If needed, the data were reanalysed in order to ensure that no errors of interpretation had occurred.

As a final step, the conclusions from the qualitative research that focused on the same questions as the quantitative impact analysis were triangulated against the findings from the quantitative research and from the existing body of evidence. Conclusions that were inconsistent with these other data sources, but which were clearly substantiated in the qualitative data, have been flagged throughout the report.

6 Quantitative component

6.1 Introduction: three data analysis approaches

As described in Section 2.1, the specific objectives of the quantitative component to this evaluation are as follows:

- **To give descriptive quantitative estimates for changes** in the lives of participant households and to **assess how much of those changes can actually be attributed to CLP-2**; and
- **To analyse how changes have materialised over time**.

We employed several different quantitative data analysis methods to address these questions, namely **descriptive statistical analysis, PSM and panel regression analysis**. The descriptive analysis provides a set of summary statistics, measured over time, for all indicators relevant to our evaluation questions, whilst the PSM and panel analysis give answers to the attribution question problem. Both descriptive analyses and the attribution identification approaches provide evidence that we then use to answer questions around how changes materialise over time and whether these are sustained.

We have started the quantitative analysis work with the comprehensive descriptive analysis of relevant indicators. The executive summary of the descriptive statistics report (i.e. OPM 2016b) is included in Annex F to this report. In a nutshell, these analyses have shown positive trends along a variety of dimensions of participant households' lives. However, despite giving a strong indication of CLP-2's effects, such descriptive analyses are no rigorous proof for attributable programme impact.

To identify programme impact, we therefore implemented PSM and panel regression analyses. Both approaches have different strengths and limitations with respect to this goal, which will be discussed in detail in the relevant methodological sections below.

In summary, the PSM approach allows us to gain insight into programme impact by explicitly addressing selection bias and providing a counterfactual to participant households. It is important to mention here that, given the structure of the data available to us, we are not limited to just one PSM analysis per outcome measure. Rather, we are able to implement several different PSM estimation strategies as well as robustness tests by using data collected at different points in time and from different cohorts (see Section 6.4.3 for a more detailed discussion).

On the other hand, using panel regression analyses we can take into consideration the dynamic changes in the data with regard to changes over time and across cohorts, allowing us to understand the intervention's impact across time. Also, given that we have observations for each household over several time periods, the panel analysis can also attempt to control for unobservable characteristics of households that may influence the outcome variables – something the PSM approach does not do.

How do we combine results from different data analysis approaches?

It is important to note that these two quantitative methods should be viewed as complementary rather than as substitutes. More specifically, we use the panel analysis to cover analytical gaps left unanswered by the PSM estimation, which can be investigated by the inter-temporal fixed effects panel analyses. As described above, PSM explicitly addresses the problem of selection bias, without being able to control for unobservable household characteristics. Panel regression analyses, on the other hand, allow us to take into account temporal effects and control for unobservable household-level fixed effects. The panel regression framework also allows for easy comparison of treatment effects between different sub-populations of CLP participant households.

Because the two approaches use different analytical models and build on different sets of data, we cannot expect them to yield the same estimation results. However, it is possible to carefully triangulate findings by comparing the sign on the treatment estimates (i.e. both PSM and panel estimate positive/negative impacts), the level of significance and the approximate magnitudes.

In general, to be confident in the conclusions that we draw from our quantitative analyses, we will expect that two things should hold when comparing results across methodologies: first, across panel analysis and PSM, estimates of changes should go into the same direction. This means that, for example, if we find that a cohort increases income significantly two years after baseline using PSM, we would expect to also find an increase using panel analysis. Second, we would also expect to make conclusions about these findings with similar levels of confidence. For example, if the positive increase in income is very significantly different from zero when looking at the PSM results, then we would expect something similar from the panel analysis results.

Conversely, we would draw conclusions less confidently if results contradict each other (e.g. if we find a negative change using PSM and a positive one from our panel analysis). Similarly, finding statistically significant changes using one methodology and not with the other would lead to more cautious conclusions of what changes CLP-2 was responsible for.

The remainder of this section

The remainder of this section will go into greater detail on data and the methodologies used for these analytical approaches, as well as presenting the associated results. We will first present the data used in Section 6.2 and give a brief summary of our descriptive analysis. We then continue to present the PSM (Section 6.4) and panel analysis (Section 6.5) methods, including associated results.

6.2 Data

The key datasets that are used for the quantitative impact evaluation component are the CLP-2 annual survey datasets (2010–2015). Detailed information on the different survey rounds, cohorts, thematic areas, and number of households sampled by the annual CLP survey are provided in Table 4 below.

Table 4 Annual CLP survey timeline

Cohort	Thematic Areas	May-10	Oct-10	Oct-11	Oct-12	Oct-13	Oct-14	Oct-15	T	Cohort size	Intended N	Final N at baseline
2.1	Demographic Information	Baseline							5	5,004	410	405
	Income, Expenditure, Assets, Savings, Food Security	Baseline							5			
	WASH								3			
	Nutrition	Baseline							4			
	Women's empowerment								3			
	Graduation								3			
2.2	Demographic Information		Baseline						5	11,109	410	410
	Income, Expenditure, Assets, Savings, Food Security		Baseline						5			
	WASH								3			
	Nutrition		Baseline						4			
	Women's empowerment								3			
	Graduation								3			
2.3	Demographic Information			Baseline					4	17,435	410	424
	Income, Expenditure, Assets, Savings, Food Security			Baseline					4			
	WASH								3			
	Nutrition			Baseline					3			
	Women's empowerment								3			
	Graduation								3			
2.4	Demographic Information				Baseline				3	6,309	410	451
	Income, Expenditure, Assets, Savings, Food Security				Baseline				3			
	WASH				Baseline				3			
	Nutrition				Baseline				2			
	Women's empowerment				Baseline				3			
	Graduation				Baseline				3			
2.5	Demographic Information					Baseline			2	13,579	410	441
	Income, Expenditure, Assets, Savings, Food Security					Baseline			2			
	WASH					Baseline			2			
	Nutrition								1			
	Women's empowerment					Baseline			2			
	Graduation					Baseline			2			
2.6	Demographic Information						Baseline		1	13,590	410	464
	Income, Expenditure, Assets, Savings, Food Security						Baseline		1			
	WASH						Baseline		1			
	Nutrition						Baseline		1			
	Women's empowerment						Baseline		1			
	Graduation						Baseline		1			

Two survey rounds took place in 2010, in May and October. After that, survey data have been collected yearly from 2011 to 2015. In 2014 data were collected through two separate questionnaires (livelihoods and a food security components) rather than a single questionnaire as in previous years. Differences between intended and final sample sizes are due to the fact that survey implementers were given a list of 500 households per cohort and a final goal of interviewing at least 410 of them. This sample size was picked by CLP-2 staff without the involvement of the OPM evaluation team. In some cases, enumerators did interview more than 410 households at baseline, making the final sample size larger than initially intended.

6.2.1 The annual household survey datasets

The annual household survey datasets contain data from surveys that covered CLP-2 participants since the programme started, i.e. since early 2010. The CLP survey followed a household panel structure, which means that a sample of the same households were tracked over time and information from the same households collected every year. In total, over 2,000 CPHHs have been covered across all cohorts and years.

Intended sample sizes for each cohort were determined by IMLC without the involvement of the OPM evaluation team. Households were sampled following a simple random sampling approach. For each cohort, at baseline a random list of households was handed to the survey implementing partner, from which the total intended number of households should have been interviewed. Note that this approach of simple random sampling, if implemented correctly, ensures that the sample of households available for each cohort at baseline is representative of that cohort.

However, given that the sample was drawn by cohort and given that cohorts vary in size considerably, the data are not automatically representative of CLP participants as a whole. For example, as can be seen in Table 4 above, cohort 2.1 comprised about 5,000 participant households, whereas cohort 2.5 comprises over 13,500 households. For both, an intended sample of 410 households was drawn randomly – which means that the 410 households of cohort 2.1 in the sample are taken as representative of 5,000 households overall and the 410 households of cohort 2.5 are representative of over twice as many households. When combining the two samples from cohort 2.1 and 2.5 in order to given an estimate that holds for all households across cohorts 2.1 and 2.5, observations from the different samples therefore need to be given different weights. In this example, the observations from cohort 2.5 must be given higher weights than observations from cohort 2.1 in order to produce correct statistics with respect to all participant households in cohorts 2.1 and 2.5. The same applies for sampled households from all other cohorts and the CLP-2 participant household population as a whole. We discuss this in detail in Section 6.2.2 below.

Table 5 below presents the total number of households that were present in the household dataset for each year, disaggregated by cohort. This allows us to see the number of households for which data were collected at baseline for each cohort and how this changes for each cohort across the years. Over time, some households have dropped out of the sample. This attrition is discussed in more detail in Section 6.2.3.

Table 5 Actual number of sampled households in the household datasets

	2010	2011	2012	2013	2014	2015
2.1	405	374	360	329	355	283
2.2	410	399	356	341	374	320
2.3	0	424	402	374	388	357
2.4	0	0	451	401	404	353
2.5	0	0	0	441	420	364
2.6	0	0	0	0	453	405
Total	**815**	**1,197**	**1,569**	**1,886**	**2,394**	**2,082**

Our various analytical approaches throughout the evaluation predominately use household-level information, but it should be noted that when looking at malnutrition indicators the analysis is conducted at an individual level. The BMI indicator contains observations for both mothers and children, whilst the weight-for-height, weight-for-age and height-for-age variables only exist for children. These individual-level datasets are made up of all the mothers and children present in the households that were selected for the household sample.

6.2.2 Sampling weights

As described above, due to the differing sizes of each of the cohorts in the CLP-2 data and the fact that it is desirable to produce results that are representative at both the cohort and overall CLP participant population level, it is necessary to create and use sampling weights. This requires information on the number of households in each cohort, the number of households in the sample drawn and the final number of households included in the final sample used in the analysis. A simple random sampling approach was used to sample the households that therefore makes the creation of sampling weights relatively straightforward.

Table 6 shows the number of participant households for each respective cohort as well as the proposed number of households that were to be sampled. This table demonstrates the difference in the total number of participant households between the different cohorts, highlighting the importance of weights when overall CLP-level estimates are calculated. Each household has the same probability of being selected within the same cohort, so for this analytical level weights are not required. However, if we are to calculate overall CLP estimates, we need to take into account that households in each sample are representative of a different number of participants, as can be seen in the second column of Table 6.

In order to draw the sample of each cohort, CLP first identified the full population of participant households. Then, of these CLP participants, 500 households were sampled and reported on a list. This list was then provided to the survey firm with the aim of reaching the numbers in the final column from Table 6. These numbers were the intended sample sizes.

Table 6 Sampling information for each CLP-2 cohort as well as CLP-1 participants

Cohort number	Total number of participant households	Total number of CPHHs that were sampled for the annual survey
2.1	5,004	410
2.2	11,109	410
2.3	17,435	410
2.4	16,309	410
2.5	13,579	410
2.6	13,590	410
Total	77,026	

The numbers in the final column of Table 6 were not followed exactly though, with different numbers of households being interviewed in each cohort. This is reflected by the number of households that can be seen in the datasets and was shown in Table 5, which gave the numbers of actually sampled households. The number of households in our data is lower than the intended sample size in the 2010 data, while intended sample sizes were exceeded at baseline for cohorts 2.3 to 2.6. Attrition, however, reduces these samples over time.

On the basis of the information above, it is possible to create weights for the CLP-2 data. This is done by the following calculation:

$$Weight = \frac{Number\ of\ HHs\ in\ sampling\ list}{Total\ no.of\ participant\ HHs\ in\ each\ cohort} \times \frac{Final\ number\ of\ HHs\ in\ sample}{Number\ of\ HHs\ approached}$$

The above formula represents the basic approach to creating weights. We have information on the number of households in the sampling list, in each cohort and in the final sample. However, we do not have specific information on the exact number of households approached. We therefore need to assume at this stage that the number of households approached is equal to the number of households on the sampling list. This means that the two values included in the formula above cancel each other out; as a result, the weights for each year are simply the number of households in the final sample over the number of participant households. Using this information applied to the above formula, weights are created for each cohort in each year. Note that we also normalise these weights so that their sum is equal to the number of households in our sample, which is done for the purposes of statistical inference.

It is evident that, in order to create these weights, we are making certain assumptions. First, we are assuming that all 500 households were approached in each year, which is unlikely to be true. We also assume that the households that are lost due to attrition are no different from the households that remain, which again is not necessarily true. Note that we deal with attrition specifically in the next subsection. Finally, we also assume that the random sampling protocol was upheld; however, this may not be true as convenience sampling may have occurred. Nonetheless, given the sizeable differences between the cohorts, not using any weights at all will be bound to create an even larger bias. Hence, creating sampling weights, despite the associated assumptions, still represents our best option.

As mentioned above, it is also apparent that attrition is an issue, with households dropping out of the sample over the years. The way the weights are calculated does not take this into consideration. Section 6.5.1 discusses how we amended the weights for the panel analysis to take into consideration the non-random attrition present in the 2015 data. In what follows, we present a summary of the problem of attrition in the present dataset.

6.2.3 Attrition

Attrition is not a problem *per se*, as long as dropping out happens randomly. This would mean that the remaining sample of households is as representative of each cohort as the baseline sample. If, however, particularly poor or vulnerable households drop out of the sample over the years this could bias estimates of indicators in later years of data collection. It is important to note that we are referring to attrition from a sampling point of view, rather than from a programme perspective.[9] Table 7 below compares the baseline values of a range of key indicators for households that are still present in the sample in 2015 versus households that have been lost between baseline data collection and October 2015. Note that the table presents a selection of indicators – we assessed differences across a wide range of indicators, for which mostly no significant difference could be found.

Table 7 presents the most notable differences and aims at giving a general overview regarding some key indicators and key differences between households. This analysis was conducted for the descriptive report and has been updated with the 2015 data.

Table 7 Difference in baseline indicators (October 2014) for cohort 2.6 between those households present in the sample in 2015 and those lost since baseline

Baseline comparison (October 2014) for Cohort 2.6				
		All	Kept	Lost
Total value of all household assets at 2014 prices	Estimate	4,340.9	4,452.6	3,398.5**
	N	453	405	48
Average household size	Estimate	3.9	4	3.3***
	N	453	405	48

[9] i.e. those households that drop out of the sample. This is relevant from an analytical point of view.

Whether the household head is female	Estimate	10.4	9.1	20.8**
	N	453	405	48
The dependency ratio of the household. (0-14)+(65+)/(15-64)	Estimate	92	94.3	72.7**
	N	443	396	47
Notes: This table uses baseline data (October 2014) for cohort 2.6. It compares indicators across households that have follow-up data in 2015 ('Kept') to households for which follow-up data is not available ('Lost'). Asterisks indicate statistically significant differences between the two groups at significance level of p<.01 (***), p<.05 (**), and p<.1 (*).				

Table 7 shows that, for cohort 2.6, we find some significant differences in the demographic make-up of households that remain in the sample and those that are lost. In total, 48 households dropped out of the sample between 2014 and 2015. This attrition rate is high given the short time difference, representing over 10% of the original sample for cohort 2.6 at baseline. Our sensitivity analysis indicates that this attrition tends to affect a specific type of household.

In particular, for cohort 2.6 we find that at baseline the households that do drop out between 2014 and 2015 are significantly more likely to be female-headed (21% vs 9% in the remaining sample), more likely to be smaller (3.3 household members vs four household members in the remaining sample), and own assets of significantly less value (BDT 3,400 vs BDT 4,400 in the remaining sample). We can observe similar trends for the other cohorts. This could have potential implications for our analysis and the confidence that we have in our results with regards to the sustainability of the intervention.

The fact that less well-off households are dropping out means that the remaining sample in 2015 is, on average, composed of better-off households; importantly, this means that our estimates, especially those related to poverty and graduation, could be biased upwards. This could be concerning for our panel regression analyses, which take into account data from several years at once.

We address this issue in our analysis by implementing two additional robustness analyses: first by using inverse probability weights in a separate specification and second by running additional analyses for a subset of years only – up until 2014, which allows to assess the impact of excluding the year most severely affected by attrition. This approach and its results are discussed in more detail in Section 6.5.1. Overall, we conclude that attrition is not biasing our results significantly.

6.2.4 Poverty indicators

For the purposes of this evaluation, we have considered two measures of poverty: consumption poverty and asset poverty. Consumption poverty here is assessed using information on the consumption levels of households, which are measured using a consumption aggregate at household level that is expressed in monetary units per capita. We use this measure for two key reasons: first, consumption poverty is a widely used measure to assess economic welfare. In fact, it is what the official poverty statistics in Bangladesh measure. Second, this a key measure that we were required to develop by DFID and DFAT.

Asset poverty is assessed using an asset value index calculated using a methodology described below. We use this measure in addition to consumption poverty because, first, we consider asset poverty as a robustness check to estimates based on consumption, i.e. we expect consumption poverty to be correlated with asset poverty. Second, however, asset and consumption poverty reflect different dimensions of being poor. For instance, not all households that can be considered consumption poor are asset poor and vice versa. Using both measures, we therefore aim at capturing both dimensions of poverty and therefore giving a more comprehensive picture of the economic welfare of CLP-2 participant households. Finally, we also look at asset poverty in order to address the issue of seasonality in consumption data. This is because household consumption can be considered much more volatile and responsive to seasonality effects, while the asset holdings of households are less volatile.

To measure consumption and asset poverty we rely on the poverty headcount ratio. For consumption poverty, we also construct the poverty gap index. There are two main reasons for using these indicators: first, these measures have been proposed by DFID and DFAT in the TOR. Second, both the poverty headcount ratio and the poverty gap index are measures that are used widely in the literature on poverty internationally. For example, the Bangladesh Bureau of Statistics (BBS) uses the consumption poverty headcount ratio to given an estimate of poverty in Bangladesh.

The poverty headcount ratio is defined as the share of CPHHs that fall below a set poverty line, for example below a set level of *per capita* consumption of household members. For consumption poverty, the poverty gap index measures the gap between the consumption levels of those households identified as poor in this way and the poverty line, as a proportion of the poverty line. The poverty gap is by definition zero if the household is above the poverty line and is expressed as a percentage of the poverty line for all other households. Higher values indicate that households are ‘further away’ from the consumption levels needed to reach the poverty line.

6.2.4.1 Our approach to constructing indicators of consumption poverty and asset poverty

Consumption poverty

In terms of **consumption poverty**, we use the lower Household Income and Expenditure Survey (HIES) poverty line for rural Rangpur (BDT 1235.66 *per capita* in 2010 prices) to calculate both the poverty headcount ratio and the poverty gap index.[10] This means that we construct a measure of extreme poverty, which is based on an explicit requirement of DFID and DFAT.

It is important to emphasise that the use of this poverty line for the sample of CLP participants requires that the measure of CLP participant consumption is comparable to that collected by the HIES. However, our review of the CLP annual survey questionnaires indicated that this is not the case and that the two measures are not directly comparable. This is due to differences in survey design and seasonality effects, which are discussed below.

Differences in survey design. Some items that are part of the HIES consumption dataset are not recorded in the CLP data. Most importantly, CLP collected information on the value of food expenditure of households but did not capture the value of self-produced goods that were consumed by households. Since the share of self-produced food in poor households’ consumption tended to be larger than in non-poor households, relying on purchased goods consumption alone would lead to underestimating the consumption of the poorer households, thereby overestimating the overall level of poverty. In addition, for items that can be matched between the two surveys,

[10] The 2010 HIES calculated two poverty lines: the upper poverty line is BDT 1486.66 *per capita* and the lower poverty line is BDT 1235.66 *per capita*. The cut-off value for our regression is 1486.66*1.2=1783 BDT *per capita* in 2010 prices for the upper poverty line. This corresponds to the poorest 43% of the total population and poorest 50% of the rural population of Bangladesh.

different recall periods were often used. This results in differential levels of accuracy of the data collected between the two surveys.

Moreover, the HIES builds in features that aim at increasing the accuracy of consumption reporting. For instance, HIES prompts respondents to ask about specific expenditure/consumption items (such as specific types of fruit), which helps recall, but CLP uses aggregate categories (e.g. 'fruit'). HIES also uses a two-week recall for food consumption with consumption for each day reported separately (everyday non-food items are asked about for each week over a two-week period), which is not the case in the CLP survey.

Seasonality effects. While the HIES survey is collected over the entire year on a rolling basis, the CLP data are collected over one month (usually in October/November), which makes them sensitive to seasonal variation in consumption. Furthermore, the CLP survey is usually collected around the time of Eid ul-Adha, which further exacerbates the seasonality issue. Households' consumption patterns are likely to vary depending on whether this festivity lies in the future or has already passed.

To address these issues of the compatibility of the CLP and HIES consumption measures, we rely on an imputed measure of consumption for CLP CPHHs in order to compute both poverty headcount ratios and the poverty gap index. This imputed measure of consumption allows us to compare CLP CPHHs with other households that live in the areas CLP is operating in. The imputation is implemented using CLP consumption data. We recognise that this measure can still be susceptible to seasonality, given that there is inherent seasonality bias in the CLP consumption data. Therefore, we also construct a counterpart measure of welfare based on an asset index.

The calculation of the imputed consumption in the CLP data is straightforward: using HIES data, we first regress HIES consumption aggregates on consumption variables that can be found in HIES and in the CLP data. The resulting coefficients are then applied to the values of the matching variables in the CLP data so as to impute a level of aggregate consumption in CLP data that is comparable to HIES. We implement this approach using the data from rural Rangpur, which is done to ensure comparability between the HIES sample and the CLP-2 data and to ensure that we can appropriately apply the lower rural Rangpur poverty line.[11]

Asset poverty

To construct the **asset poverty** headcount ratio, we first build an index of asset values that households hold. The construction of this asset index requires, in a first step, a transformation of the raw reported asset values in the CLP-2 data because asset variables are standardised in the HIES data. Therefore, before constructing the asset index, we apply the same standardisation procedure to the CLP asset data, using means and standard deviations from the HIES dataset. In a second step, we then construct an asset index using principal component analysis on the HIES data and applying resulting weights to the standardised CLP data. The asset poverty line used to construct the headcount ratio is drawn at the 30th percentile of the rural Rangpur asset index distribution in the HIES data.

[11] We implemented the same analysis using a different sample of HIES households in our descriptive analysis stage. The results are reported in the descriptive statistics report (see Annex F for the executive summary). Our strategy of estimating consumption poverty is robust to this change in household samples.

A more detailed methodology for constructing imputed consumption and the asset index is outlined in Annex D.

Combining measures of consumption and asset poverty

Used together, these two welfare measures allow us to identify two categories of poor households that are of interest for our analysis: those who are *asset and consumption poor* and those who are *consumption poor but asset 'rich*'. The former could be considered chronically poor given that such households are identified as poor both in terms of consumption and in terms of the assets they own. The latter could be considered as transient poor, given that they are not considered to be poor in terms of asset ownership but only in terms of household consumption. Given CLP's focus on asset transfers it is important to consider these transient poor as they represent a group of households that remain vulnerable in the short term despite the transfer of assets and that may require additional emergency or transition support to sustainably increase their consumption.

We have analysed these categories of households in more detail in the descriptive statistics report that was produced in the first phase of this evaluation (see OPM 2016b and also Annex F for the executive summary). For the purposes of the impact assessment, we consider asset and consumption poverty separately. The main purpose is, as described above, to give a robustness check to the results achieved with consumption poverty: if both decrease due to CLP-2 programme effects, we have very robust evidence of significant improvements in participant households' economic welfare. In addition, however, we also describe differences in impact estimates across those measures. The reason for this is that, as mentioned above, CLP-2 is an ATP, and we hence would possibly expect larger decreases in asset poverty, compared to consumption poverty.

6.3 Descriptive analysis

As described above, the first phase of quantitative data analysis in this evaluation was a descriptive exercise. The main objectives were to provide a comprehensive set of summary statistics for all indicators that are relevant to CLP's interventions and to provide a preliminary descriptive assessment of CLP-2 performance. In order to do so, we looked at demographic, child nutrition, poverty and household expenditure indicators, as well as the aggregate graduation variable and the indicators that contribute to the graduation variable. For a more detailed discussion of the descriptive analysis conducted, see Annex F, which contains the executive summary from the descriptive report, or the descriptive report itself (OPM 2016b). It is important to note that the descriptive analysis cannot identify any causal relationship behind the results it presents.

Some interesting areas of discussion arose from the descriptive analysis. First, our estimates showed that, in October 2015 for cohorts 2.1 to 2.6, CLP reached a graduation rate of about 90%. However, this graduation rate varies across cohorts, with cohorts 2.1 and 2.2 achieving lower rates than the others. This overall aggregate graduation indicator also hides the disparity in performance of the composite indicators, with households struggling to meet the criterion relating to cash savings, for instance. Additionally, some of the indicators that make up the aggregate graduation indicator are binary variables, which consist of two individual indicators. These binary indicators can hide the different performance in the underlying indicators. For instance, for some cohorts we observed a decrease in the proportion of households meeting the food consumption-related criteria. However, if this indicator is broken down into the individual variables, we can see the decrease is driven mainly by households not eating three meals a day, while the indicator measuring diet diversity does not change.

With regards to poverty, our descriptive analysis found that CLP is linked with a decrease in poverty both in terms of consumption and asset poverty. Unsurprisingly, we find that household expenditure increases over time for CLP participants, which is mainly driven by increases in non-food expenditure. However, these improvements do not seem to translate into changes in terms of child nutrition, which may be explained by the fact that households tend to increase non-food consumption whilst spending on food remains flat.

Finally, our analysis finds that, for some indicators, positive trends continue after the end of the CLP core intervention period. For example, the number of income sources continues to increase. This could potentially be an indication of a sustainable effect of CLP interventions. However, along with our other findings, this highlights the need for a more comprehensive and rigorous analysis of CLP to identify causal relationships, something which the quantitative analysis components will provide.

6.4 PSM

6.4.1 Methodology

As described in Section 6.1, we have used PSM in order to answer attribution-related questions in this evaluation, such as 'what impact did CLP-2 have on certain outcome variables?' The key problem that needs to be resolved in order to answer such questions is the selection bias problem. Households that did receive CLP-2 support in a certain year could be systematically different from households that did not receive such support – and hence simple comparisons of indicators across such groups could be invalid.

PSM methods solve this problem by constructing appropriate counterfactuals to treatment outcomes by matching and comparing outcomes for units in the treatment group with control units that are as similar as possible to each other along a set of relevant observable characteristics. PSM is a two-stage analysis that employs the propensity score as a 'comparator metric'. Hence, the first stage of any PSM analysis is to compute a valid propensity score for each unit of observation. The second stage is to then compare outcomes across units with similar scores – with the option of using several different comparison algorithms. **Annex G describes in detail the design and implementation of the first and second stages in the PSM analytical setting** for the current evaluation.

It is important to note that, for PSM to work appropriately, the comparator metric constructed in the first stage needs to be valid. For that to be the case, it needs to be calculated using variables that are not influenced by the intervention. The intuition behind this is as follows: as described above, PSM aims at reducing bias by controlling for background characteristics that are correlated with both treatment and outcome measures. In effect, the aim is to account for any differences in outcome between treatment and control groups that are not due to the treatment but due to some other systematic dissimilarities between the groups. Using variables that are themselves influenced by the treatment to do this would mean that one would compare units that are similar to each other after receiving the treatment, hence mistaking dissimilarities between groups that are due to the intervention for structural dissimilarities. We address this problem here by constructing the propensity score using baseline data from cohorts only, which means that no data are influenced by the programme.

Key assumptions: common support and conditional independence

There are two key assumptions that need to hold for PSM to be a valid approach to estimating treatment effects: the common support assumption and the conditional independence assumption.

The common support assumption states that the estimated or predicted probabilities for all individuals of belonging to the treatment or control group must lie between zero and one, i.e. individuals in both groups must have a positive non-zero probability of belonging to either the treatment or control group and the distribution of those probabilities across the two groups must be such that comparable individuals across the groups can be found.

The conditional independence assumption posits that, once observable characteristics have been accounted for, the outcome measure is not related to the treatment status anymore, other than via the effect of the programme. In essence, this assumption states that once we control for observable characteristics appropriately, treatment status can be treated as if it was assigned randomly. If treatment assignment is non-random, the concern is that control and treatment groups are not comparable. For example, it could be that individuals with a certain characteristic that also affects the outcome measure are more likely to select into treatment. Differences in the outcome measure between treatment and control groups would then not only be due to the treatment effect, but also due to the systematic difference in this characteristic. As described above, PSM deals with this problem by comparing outcome measures across treatment and control groups only for individuals that are similar, i.e. by controlling for the important characteristics that are related to both treatment status and the outcome measure. The conditional independence assumption simply states that all important characteristics have been taken care of. This means that any bias that arises due to participation in the programme has been dealt with. Note that this includes biases that arise due to unobservable factors – PSM cannot control for these and the assumption is that once observable characteristics have been dealt with no unobservable bias remains.

The validity of any PSM approach crucially depends on how well the approach reduces any imbalance between treatment and control groups. Under conditional independence – i.e. independence of the treatment assignment from outcome measures when controlling for covariates – the propensity score is a valid balancing score. Conditioning on this score appropriately means that bias will be removed between control and treatment groups. Hence, treatment and control groups will be balanced, i.e. they will have similar covariate distributions. Resultantly, this means that, across a variety of different characteristics, the treatment and control groups will be similar to each other.

However, if after conditioning on the propensity score the treatment groups remain unbalanced, then, again assuming that conditional independence holds, the estimation of the propensity score might not have been specified correctly. Similarly, if after matching the imbalance between samples persists, then the way in which we conditioned on the propensity score might not have been appropriate. **This means that a crucial component of any PSM approach is to assess how balanced the treatment and control groups are along key covariate dimensions, after matching.**

What treatment effect does PSM estimate?

It is important to emphasise that PSM works by looking for control units that can be compared to treatment units, and not the other way round. This means that it is assumed that treatment units are a given and control units need to be identified. Through finding matches for the treatment units (households that have received the CLP intervention) in the pool of control units (households that are yet to receive the CLP intervention), the resulting estimate of the treatment effect is therefore the ATT. Extrapolating this estimate beyond the population for which the treatment sample is representative is therefore not possible.

The following sections outline our approach to employing PSM in the context of the present evaluation. We first present the necessary considerations for applying PSM in the context of CLP and the various estimation and robustness approaches we used. We then comment on the weaknesses of PSM and the strategies we used to address these, including assessing the credibility of the conditional independence assumption. Finally, the results of each of the estimation strategies and robustness checks performed on the available data are presented.

6.4.2 Considerations for applying PSM in the context of CLP-2

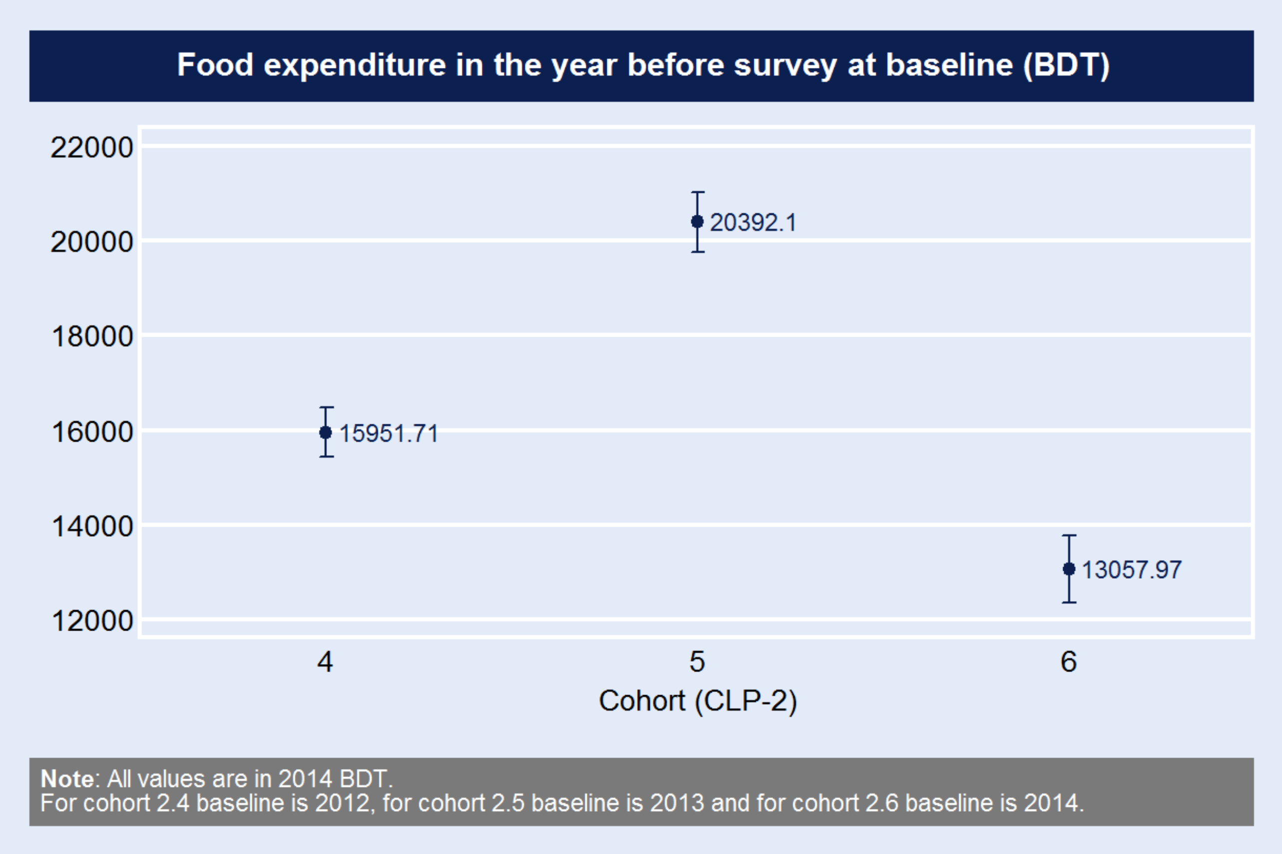
We employed PSM as our starting point to answer questions of programme impact, i.e. what changes that are observed in indicators can be directly attributed to CLP-2 and what is the magnitude of these changes?[12] However, given the available data from CLP-2, it is important that certain considerations are taken into account when applying a PSM approach to this data structure.

In the present context, data only exist for participants or future participants of CLP-2. This means that, strictly speaking, the probability of treatment for each household and for all household members in the sample is equal to one. **Therefore, for the purposes of this evaluation, the propensity score should rather be seen as a metric that captures systematic differences between the cohorts that relate to how the treatment effects of the CLP-2 intervention materialise in each cohort.** This could be, for example, due to systematic differences that are related to the timing of the treatment. Even though participant selection and sampling methods did not change across time, our conversations with CLP staff indicated that for later cohorts some villages were revisited and 'new' participants selected. This could mean that, for example, participants that did not meet the criteria for selection into the programme in the first instance were selected into a later cohort, which could lead to some systematic differences across cohorts.

In fact, our descriptive analysis indicates that such systematic differences do exist across cohorts. For instance, in Figure 2 below, we compare the average total yearly food expenditure of cohorts 2.4, 2.5 and 2.6 at baseline (data from 2012, 2013 and 2014, respectively). We also plot 95% confidence intervals. It is clear that the averages are significantly different from each other in all three cases, even though we are looking at pre-treatment values for all three cohorts.

[12] The changes we refer to are between baseline and post-treatment for the cohorts included in each strategy. For strategy 1a and strategy 2 the changes are between baseline and T2, whereas for strategy 1b the changes are between baseline and T1.

Figure 2 Comparison of total yearly food expenditure at baseline across cohorts



Irrespective of the reasons for the existence of these differences across cohorts, if the observed variation is also correlated with the effect of CLP-2 interventions, then comparing results across cohorts without controlling for covariates does not yield a valid estimate of this effect. As described above, the PSM approach allows to control for this bias.

In this context, it is necessary to emphasise again the importance of the conditional independence assumption and its reliance on observable variables. **If there are unobservable factors that bias the impact estimate, PSM will not be able to capture this.** Although it is not possible to directly test the conditional independence assumption, there are a variety of post-estimation balancing tests that we have implemented in order to assess the comparability of treatment and control units after matching. If the observables are balanced in certain thematic areas, then it is likely that unobservable characteristics are also balanced in the same thematic areas. Furthermore, if these tests confirm that samples are balanced across a variety of dimensions, then it is more likely that conditional independence will hold.

For our analysis most of the strategies presented below estimated treatment effects that are cohort-specific. We know that the sampled data we have access to for each of the cohorts are representative for that cohort only. We have addressed this by estimating treatment effects for several specific cohorts, comparing these results, and also triangulating these with the results from our panel regression analyses.

Implications of the data structure

For the purposes of our PSM exercise, the unit of analysis is mainly the household, with the exception of the malnutrition outcome variables. This means that, when referring to 'matching cohorts' in what follows, we actually mean 'matching households' across different cohorts.

For the reasons set out in our Inception Report (OPM 2015), and given the results from our assessment of the available data, we limited the analysis to data collected from 2012 onwards. This left four years of data collection available to us.

Table 8 exemplifies this situation. Annual survey data were collected four times from October 2012 onwards. Limiting our analysis to these four years of data collection means that we were able to use three rounds of baseline data: for cohort 2.4 (collected in October 2012), cohort 2.5 (collected in October 2013) and cohort 2.6 (collected in October 2014). The table below also shows that we have a diverging number of post-baseline survey rounds available for the different cohorts: three for cohort 2.4 (T1-T3), two for cohort 2.5 (T1-T2) and one for cohort 2.6 (T1).

Table 8 Summary of annual data collection

	Main month of data collection						
	May-10	Oct-10	Jun-11	Oct-12	Oct-13	Oct-14	Oct-15
Cohort 2.1	BL		T1	T2	T3	T4	T5
Cohort 2.2		BL	T1	T2	T3	T4	T5
Cohort 2.3			BL	T1	T2	T3	T4
Cohort 2.4				BL	T1	T2	T3
Cohort 2.5					BL	T1	T2
Cohort 2.6						BL	T1

Notes: BL refers to the baseline year of data collection for each cohort. T1 to T5 refer to the number of years after baseline.

It is important to emphasise here again that the CLP-2 intervention phase lasts for 18 months. This means that for each of the cohorts, strict post-intervention data (i.e. data collected after CLP interventions finished as compared to post-baseline data) are only collected from T2 onwards. In the current context (i.e. for data collected from 2012 onwards), this means that post-intervention data are available for cohorts 2.1, 2.2, 2.3, 2.4 and 2.5 in our analysis, but not for cohort 2.6.

The particular structure of this data means that when implementing analyses and interpreting results, three different potential structural effects that influence indicator measurements have to be taken into account:

- **Year-specific effects**: these include effects such as macroeconomic shocks that are due to some particular event in a certain year. These effects mean that, when comparing variables across years, differences may partly be due to differences in yearly effects. For example, when comparing Cohort 2.4 in T1 (2013) with cohort 2.5 in T1 (2014), diverging yearly effects could affect detected differences in indicators between the two cohorts.

- **Cohort-specific effects**: these are due to the unique characteristics of cohorts. These effects mean that, when comparing variables across cohorts, differences may partly be due to these underlying structural differences in cohorts. For example, when comparing cohort

2.5 at baseline with cohort 2.4 in T1 (both in 2013), diverging cohort-specific characteristics can play a role. As described above, PSM aims at dealing with these structural differences.

- In addition, however, it is possible that within the context of the present evaluation there may be **cohort-specific treatment effects.** Even without any bias in estimation procedures, it could be that CLP-2 intervention treatment effects materialise differentially across cohorts and that this effect diverges from the overall average treatment effect across all CLP-2 participants. This could, for example, be due to the fact that CLP interventions were implemented differently across cohorts and years – CLP-2 interventions did not look exactly the same for all cohorts at all times.

- **Intervention time effects**: these are due to the fact that the CLP-2 intervention is a continuous intervention with an implementation schedule that varies across cohorts. This means that, even though T1 is 12 months into the intervention for all cohorts, the extent of this implementation is not uniform for all cohorts. In addition, the length of post-intervention time periods will influence the persistence of intervention effects. This means that, when comparing variables within one year, e.g. 2014, and across cohorts, differences in these variables could partly be due to the diverging intensity of treatment.

No single estimation methodology is able to deal with all of these structural effects at the same time, given that we can only implement PSM on specific subsets of the data. Therefore, we decided to implement a set of different PSM analyses, which we term 'estimation strategies' in what follows, exploiting the data structure presented above. Estimated treatment effects have been compared across these strategies, yielding a range within which treatment effects for different cohorts and years lie. In what follows, we present the results of all of these estimation strategies. The following paragraphs explain the rationale behind each of the different estimation strategies separately. Note that – in addition to the different PSM estimation strategies – we also implement panel regression analyses and compare results across these approaches.

6.4.3 Estimation strategies

The paragraphs that follow show how we selected different subsets of the data when implementing PSM in order to address the structural effects mentioned above. This will exemplify the relative merits and drawbacks of each of the strategies implemented.

Strategy 1a and strategy 1b: First stage using baseline across years and second stage within one year – two cohorts

In order to deal with the potential issue of the contamination of covariate measurements with treatment effects, we exploit the fact that for each cohort there are also baseline data available, albeit from different years of data collection. Assuming that at baseline the data on participant households are indeed not contaminated by programme effects, calculating the propensity score in the first stage using this baseline data and the full set of available variables, and then matching households using this score, should not be undermined by this problem of contamination.

Table 9 Data set-up for first stage using baseline across years and second stage within one year – two cohorts

	Main month of data collection			
	Oct-12	Oct-13	Oct-14	Oct-15
Cohort 2.1	T2	T3	T4	T5

Cohort 2.2	T2	T3	T4	T5
Cohort 2.3	T1	T2	T3	T4
Cohort 2.4	BL (first stage A)	T1	T2 (treatment A)	T3
Cohort 2.5		BL (first stage B)	T1 (treatment B)	T2
Cohort 2.6			BL (control)	T1
Notes: BL refers to the baseline year of data collection for each cohort. T1 to T5 refer to the number of years after baseline.				

Table 9 demonstrates how such two such strategies – formerly summarised under strategy B in the Inception Report (OPM 2015) – were implemented in the current context.

The first strategy uses baseline data for cohort 2.6 (from October 2014) as a control and cohort 2.4 (from October 2012) as the treatment cohort to calculate the propensity score and to match comparable households. At the second stage, the treatment effect was estimated using the outcome measures at baseline for the control cohort (2.6) and at T2 for the treatment cohort (2.4). **This is referred to as strategy 1a. Note that the result of this is an estimate for the ATT at T2 for cohort 2.4.**

As was set out above, the benefit of employing this estimation strategy is that the first stage matching procedure is not contaminated by any CLP-2 intervention effects. In addition, because we estimate treatment effects using outcome measures for both cohorts within one year, year-specific effects should not play a role in the second stage. However, **cohort-specific differences in how treatment effects materialise are not controlled for using this approach: we estimated the specific treatment effect for cohort 2.4 only – this cannot strictly be extrapolated to other cohorts.**

We also implemented this analysis with cohort 2.5. At the first stage, the baseline data for cohort 2.6 (from October 2014) as control and for cohort 2.5 (from October 2013) as treatment cohort were used to calculate the propensity score and to match comparable households. Note that at the second stage, however, the treatment effect was estimated using the outcome measures at baseline for the control cohort (2.6) and at T1 for the treatment cohort (2.5). Therefore, **the specific treatment effect for cohort 2.5 will not take into account the full CLP-2 intervention period of 18 months but rather gives us an estimate for the ATT at T2 for cohort 2.5.** This approach will be referred to as strategy 1b.

Strategy 2: First stage using baseline across years and second stage within one post-intervention time period but across years – one control cohort and two pooled treatment cohorts

In an alternative strategy, formerly labelled strategy C, we implemented analyses using pooled treatment data from two cohorts. Instead of matching households from the control cohort with households from one other cohort, we used the pooled baseline data for two cohorts to implement the first stage matching procedure. We then used the follow-up data at the same post-intervention time period to estimate impact in the second stage.

Table 10 below shows how this strategy was implemented. At the first stage, baseline data for cohorts 2.4 and 2.5 were pooled and matched to the baseline data for cohort 2.6. At the second stage, outcome data were compared between baseline data from cohort 2.6 and data collected at T2 for the matched households of cohort 2.4 (October 2014) and cohort 2.5 (October 2015).

Table 10 Data set-up for first stage using baseline across years and second stage within one post-intervention time period – one control cohort and two pooled treatment cohorts

	Main month of data collection			
	Oct-12	**Oct-13**	**Oct-14**	**Oct-15**
Cohort 2.1	T2	T3	T4	T5
Cohort 2.2	T2	T3	T4	T5
Cohort 2.3	T1	T2	T3	T4
Cohort 2.4	BL (first stage)	T1	T2 (treatment)	T3
Cohort 2.5		BL (first stage)	T1	T2 (treatment)
Cohort 2.6			BL (control)	T1

Notes: BL refers to the baseline year of data collection for each cohort. T1 to T5 refer to the number of years after baseline.

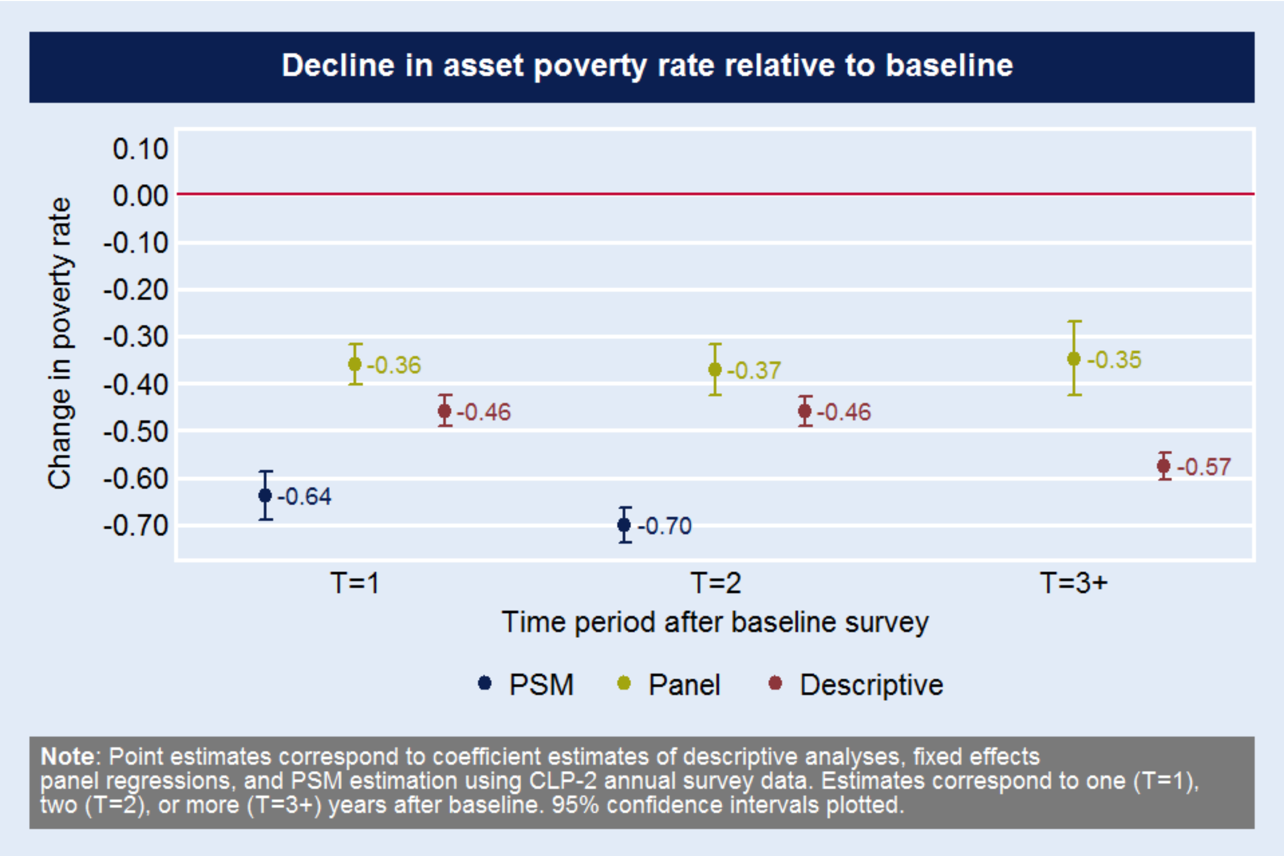
The benefit of implementing this analysis is that cohort-specific treatment effects should be less prevalent here, given that we are comparing outcomes across more than two cohorts. In addition, the matching procedure should produce a more balanced sample, as there is a larger sample available to match to. The drawback of this strategy is that it relies on time periods after the CLP-2 intervention to estimate the treatment effect, which means that outcome data are used across two different years. This means that year-specific effects could play a role in the size of the estimated treatment effect. **Note that this strategy therefore gives us pooled ATT for cohorts 2.4 and 2.5 at T2.**

Combining results across strategies and presentations for Volume I

It is important to reiterate here that the results from these different estimation strategies should be seen as complementary rather than as substitutes. First, as explained above, these strategies give estimates that hold for different subsets of the population of CLP participant households and for different time periods within the implementation period of CLP-2; thus, point estimates are likely to differ across strategies. However, in order to make conclusive statements about the impact of CLP-2, we expect that estimates across the different strategies will point in the same direction and have similar levels of significance. Our results below show that this is the case in all our analyses.

It is also important to note here that, for presentational purposes, we show the results from two estimation strategies in our presentational graphs in Volume I (OPM 2016a). For instance, Figure 3 below is taken from Volume I and shows results with respect to the CLP-2 impact on poverty rates. The PSM estimates in there correspond to the following two strategies: first, for the estimates at T1 the results from strategy 1b. This means that estimates presented there are, strictly speaking, ATT estimates for cohort 2.5 at T1. Second, for the estimates at T2, these are the results from strategy 2. This means that this is an estimate of the average ATT for cohorts 2.5 and 2.4 at T2. As was stressed above, we cannot extrapolate this directly to the overall population of CLP participant households. Note that we rely on a combination of our robustness checks below together with the panel results to make such general statements.

Figure 3 Example presentation of results – impact estimates on poverty rate



6.4.3.1 Robustness checks

We implement several robustness checks in order to increase the confidence that we have in our results.

Robustness check 1: Estimating pseudo-treatment effects for the first stage using baseline across years and second stage for both control and treatment within same post-intervention time period

In order to assess the quality of our matching procedure, we also implemented two 'robustness check' estimations. The first one, which was termed 'Strategy D' in the Inception Report (OPM 2015), helped to evaluate how problematic differential year or cohort effects could be in the current context. In this strategy we again matched cohorts using baseline data in the first stage. In the second stage, however, we assessed differential impact by comparing outcome data for both cohorts at the same time period after baseline.

Table 11 below exemplifies how this was implemented in the present case. At the first stage, the baseline data from cohort 2.4 and cohort 2.5 were used to perform propensity score estimations. Then, at the second stage, impact was estimated using data from T2 for both cohorts. If treatment effects are constant across cohorts and year-specific effects negligible, the estimated impact should not be different from zero. However, if our estimation finds a significant impact, then we would count this as evidence for differential impacts between cohorts or year effects. This also means that we would be less confident in the external validity of results from estimation strategies that only employ two cohorts (i.e. strategies 1a and 1b above).

Table 11 Data set-up for estimating pseudo-treatment effects for the first stage using baseline across years and second stage for both control and treatment within same post-intervention time period

	Main month of data collection			
	Oct-12	**Oct-13**	**Oct-14**	**Oct-15**
Cohort 2.1	T2	T3	T4	T5
Cohort 2.2	T2	T3	T4	T5
Cohort 2.3	T1	T2	T3	T4
Cohort 2.4	BL (first stage treatment)	T1	T2 (treatment)	T3
Cohort 2.5		BL (first stage control)	T1	T2 (pseudo-control)
Cohort 2.6			BL	T1
Notes: BL refers to the baseline year of data collection for each cohort. T1 to T5 refer to the number of years after baseline.				

Robustness check 2: Estimating pseudo-treatment effects with first and second stage at baseline for both control and treatment

The second robustness check strategy is simple and follows the set-up described by Imbens and Rubin (2015, p. 482 ff.), where impact is estimated using baseline data only. If our matching procedure produces well-balanced samples at baseline, then we should not be able to see any significant difference in baseline outcome measures between treatment and control groups. In fact, if we do find significant differences, this could point to the fact that our matching is not performing well and that conditional independence might not hold (*Ibid*., p. 484).

This robustness check has been implemented by checking the balance of the outcome indicators at baseline. Table 12 below exemplifies the data structure for such a strategy. In each of the strategies presented above, we estimated propensity scores using baseline data for cohort 2.4 (or cohort 2.5, depending on the strategy), and then we compared outcome measures at baseline using the same data. As was outlined above, if we did find a significant impact this would point to the fact that the first stage matching procedure is not performing well – some systematic difference between treatment and control cohorts that influences outcome measures has not been controlled for.

Table 12 Data set-up for estimating pseudo-treatment effects for the first and second stage at baseline for both control and treatment

	Main month of data collection			
	Oct-12	**Oct-13**	**Oct-14**	**Oct-15**
Cohort 2.1	T2	T3	T4	T5
Cohort 2.2	T2	T3	T4	T5
Cohort 2.3	T1	T2	T3	T4
Cohort 2.4	BL (first and second stage)	T1	T2	T3
Cohort 2.5		BL (first and second stage)	T1	T2
Cohort 2.6			BL	T1
Notes: BL refers to the baseline year of data collection for each cohort. T1 to T5 refer to the number of years after baseline.				

6.4.4 Caveats – Addressing weaknesses in the analysis

As with any such technique, there are a variety of weaknesses that are inherent to the PSM approach. This section briefly presents how we dealt with these weaknesses.

Reliance on the conditional independence assumption

As already discussed, PSM analyses rely crucially on the conditional independence assumption. This assumption cannot be tested directly (Imbens and Rubin 2015, p. 479 ff.), but a set of analyses can be implemented to assess how plausible this assumption is in the present context. The first one is to assess the balance of covariates after matching, which we implement below.

In addition, another analysis that has been used to assess this has been presented above in section 6.4.3.1, i.e. estimating pseudo-effects at baseline for treatment and control groups. Assuming that conditional independence holds after matching, PSM analyses should not pick up any treatment effect here, i.e. the outcome variable should also be balanced at baseline.

Finally, it should be noted that our panel analysis will also explicitly take into consideration unobservable characteristics that PSM cannot control for and that might invalidate conditional independence.

Standard errors of estimated treatment effects

Calculating standard errors of estimated treatment effects using PSM methods is not straightforward. As Caliendo and Kopeinig (2005, p. 18) put it, 'The problem is that the estimated variance of the treatment effect should also include the variance due to the estimation of the propensity score, the imputation of the common support, and possibly also the order in which treated individuals are matched'. These estimations increase the variation of the treatment effect estimates over and above normal sampling variation. In the literature, there is no consensus on how to take this into account.

A popular approach to solve this problem is to bootstrap standard errors for the estimated treatment effect (see Lechner 2002). Each bootstrap draw re-estimates both the first and second stages of the estimation. This produces *N* bootstrap samples for which the ATT is estimated. The distribution of these means approximates the true sampling distribution, and therefore the standard errors of the population mean (Caliendo and Kopeinig 2005, p.18). We followed this approach and implemented bootstrapping, using 300 repetitions, to estimate the standard errors of our estimated treatment effects. Note that, for the sake of completeness, we show both the bootstrapped and the non-bootstrapped standard errors below.

It is also important to note that there is no clear direction in which estimated standard errors should change due to bootstrapping. On the one hand, the additional variation taken into account should increase standard errors. On the other, bootstrapping generally makes estimates more precise, which tends to decrease standard errors. Overall, the direction of the change is not uniform. In fact, our results show that, with bootstrapping, standard errors in some instances are smaller and in some larger than without bootstrapping.

Contamination of baseline data

A large part of the PSM analysis assumes that baseline data, i.e. the data on which the first stage propensity score estimation procedure is implemented, is not contaminated by the

CLP intervention. It is not clear whether this assumption always holds. For instance, if a village consists of households that pertain to two different cohorts, it could be that some households have already benefitted from village-level interventions, even at baseline. In addition, it could be that spill over effects materialise from treated to non-treated households. It is difficult to test this assumption using the available data. In general, we have therefore assumed that this is the case for the large majority of households in our datasets. We have cross-checked this assumption with CLP and the qualitative information also supports it.

Population-level treatment effect estimates

As described above, estimated treatment effects are cohort-specific. For instance, where we used the data from cohort 2.6 (baseline) and cohort 2.4 (baseline and follow-up at T2) to implement PSM, we estimated the ATT, i.e. the treatment effect for cohort 2.4. Given that we assume that the sample for each cohort is representative for the overall set of participants in that cohort, we can also assume that this estimated treatment effect is representative for this cohort.

The issue is, though, whether the estimated treatment effect could be extrapolated to other cohorts or even the population of CLP-2 participants overall. In general, it is possible to argue that this can be the case at least for the cohorts employed in the estimation procedure (e.g. cohorts 2.4 and 2.6 above), as these cohorts are used as comparators and to estimate the treatment effect. However, it is important to emphasise that this is not necessarily the case. The results from matching in one direction (i.e. using 2.4 as treatment and 2.6 as control) will not necessarily be the same as matching in the other direction. The reason for this is that the matching procedures above always match one unit of observation from the treatment group with one or more from the control group. Hence, matching is not symmetrical and depends on treatment vs control assignment.

In order to address this, we implemented a variety of different estimation strategies (see section 6.4.3) that yielded treatment estimates for different cohorts under different assumptions and data structures (such as, for example, pooling baseline data). As mentioned above, strategy 1b is our only PSM estimate for treatment effects at T1, whilst strategy 2 is our preferred estimation model for treatment effects at T2 as it pools together two cohorts. As mentioned above, we also present these results in Volume I (OPM 2016a) for the PSM estimation.

6.4.5 Results

This section presents the results obtained from applying PSM to CLP-2 data collected from 2012 onwards (for the reasons explained in section 6.4.2). In what follows, we present balancing results and ATT estimates for all estimation strategies and outcome variables that we looked at in the context of this evaluation. This includes the following variables:

- Poverty indicators: We present PSM results for the poverty headcount (both for consumption and asset poverty), the poverty gap index (in percentage and BDT) and imputed consumption levels in BDT.
- Expenditure indicators: Results for the *per capita* expenditure in BDT, *per capita* annual food expenditure and *per capita* annual non-food expenditure.
- Income and savings indicators: Results for total income, *per capita* income and *per capita* cash savings, all in 2014 prices, are presented.

- Asset indicator: Results for the total value of household assets.
- Food security indicator: Results for the number of meals eaten in the past week are presented.
- Malnutrition indicators: Finally, we present results for BMI, Height-for-Age (HFA), Weight-for-Age (WFA) and Weight-for-Length (WFL).

For each of the outcome variables of interest (except from malnutrition indicators), we implemented estimation strategies 1a, 1b and 2, as well as robustness test 1, as discussed in section 6.4.3. In the case of the malnutrition variables (BMI, HFA, WFA and WFL), given that these types of data were only collected for the survey years 2012 and 2014, we implemented only strategy 1a.

At the first stage analysis, a different set of covariates was selected for each one of our outcome variables and each one of the strategies implemented. Regarding the second stage analysis, the same matching algorithms (Kernel function and Nearest Neighbour (NN) with forced common support) were implemented in all cases. Annex G discusses the first stage and matching algorithm selection in more detail, including giving a rationale for covariate selection.

Poverty outcome variables

1. ***Consumption poverty –defined as the proportion of households falling under the lower rural Rangpur poverty line***

For each outcome variable we present two sets of results in the tables that follow below. These tables have the following format. The first graph on the left-hand side shows how individual variables balance before and after matching. The x-axis displays the standardised bias, which is the percentage difference of the sample means in the treated and non-treated (unmatched or matched) subsamples as a percentage of the square root of the average of the sample variances in the treated and non-treated groups (Rosenbaum and Rubin 1985). It is worth noting that for some covariates there are no crosses and dots present on the balancing graphs. This is due to the fact that the covariates are variables that predominately take the value zero (frequently dummy variables), and the mean before and after matching for both treatment and control groups is equal to zero. As a result, the standardised bias cannot be calculated and therefore these results are not present. Looking at Table 13 below, as an example, the unmatched samples display large imbalances with standardised bias being present across many of the covariates of interest. However, once matching takes place, the standardised imbalances are removed.

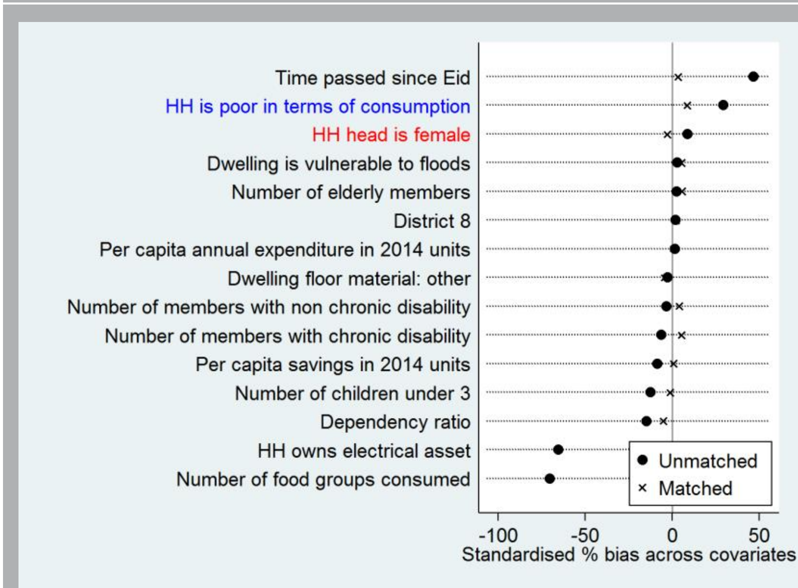
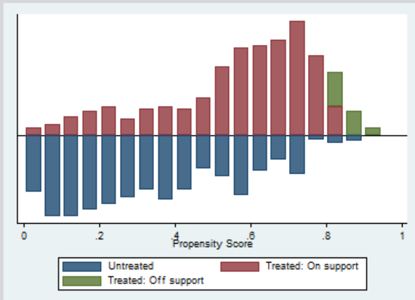
Note also that, in all cases, balance has also been achieved for the outcome variable of interest. This points to the fact that the propensity score estimation procedure is performing well, and any systematic difference between treatment and control cohorts that influences outcome measures has been controlled for.

The second graph, on the right-hand side, shows the distribution of propensity scores across treatment and control groups. This graph visually confirms that, after dropping observations that are off common support, both treatment and control groups contain observations with propensity scores across the full range of the distribution, which is an indication for overall balance. Although the distributions of propensity scores across treatment and control groups would ideally be symmetric, the presence of some level of skewness does not put at risk our estimation procedure, as indicated by the balance achieved for each covariate and the overall values of Rubin's R and B after matching. In addition, it is possible to observe that, once we match using more observations (i.e. strategy 2) the level of skewness is reduced. This is explained by the fact that a larger common support indicates a larger overlap in the characteristics of treated and untreated units, making it easier to find adequate matches across groups. The fact that results don't differ significantly across those estimation strategies provides further evidence that the skewed distribution of estimated propensity scores is no cause for concern.

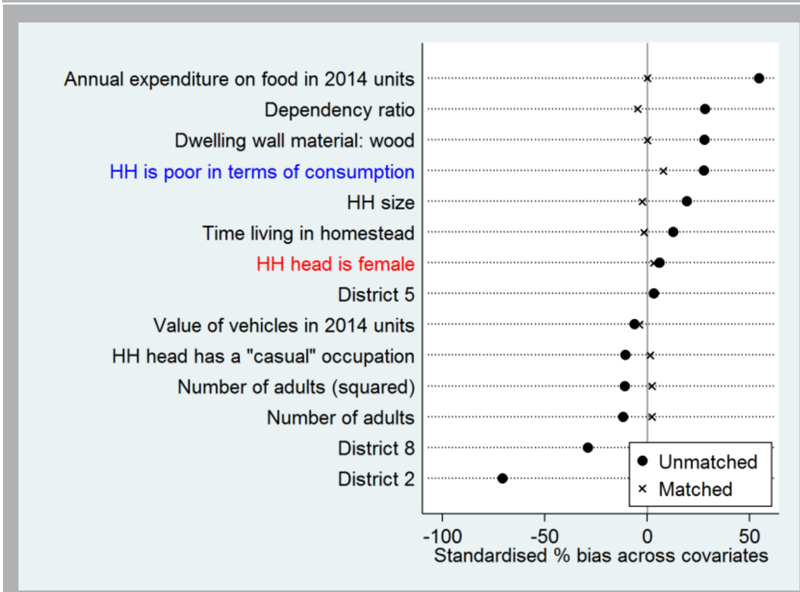
The remaining rows on the right hand side display information related to the PSM model. The bandwidth and level of trimming for the optimal PSM model can be found in the first two rows. For example, strategy 1a's optimal model has a bandwidth of 1 and a trimming value of 8. This is then followed by the number of observations on common support in the next row, and then the Rubin's R and Rubin's B values both before and after matching. Generally, a Rubin's B score under 25 after matching is desirable, whilst a Rubin's R score between 1 and 1.25 is the preferred range after matching (Rubin 2001). The unmatched samples are particularly unbalanced, which is demonstrated by the significantly high Rubin's B scores (115.14, 117.08, 101.47 and 94.83 for strategies 1a, 1b, 2 and robustness test 1, respectively). However, the Rubin's B scores after matching, which are all below 25, show how matching removes the previous imbalances.

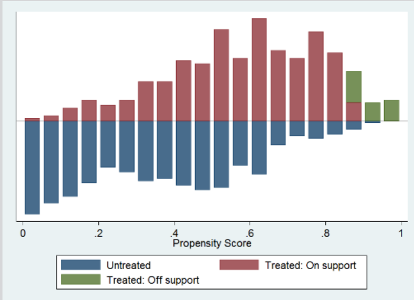
After that, estimated ATTs are presented, both for the kernel matching approach with bootstrapped and non-bootstrapped standard errors and for the NN matching specification, which acts as a robustness test. Given that it is not definitively clear how to produce standard errors for PSM we present both bootstrapped and non-bootstrapped standard errors for robustness purposes. For example, Table 14 displays these results for consumption poverty.

Table 13 Consumption poverty: Selected variables and balancing across strategies

<table>
<tr><th>Variables</th><th colspan="3">Balancing</th></tr>
<tr><td colspan="4">Strategy 1a*</td></tr>
<tr><td rowspan="6">
</td><td colspan="3">
</td></tr>
<tr><td colspan="2">Bandwidth</td><td>1</td></tr>
<tr><td colspan="2">Trimming</td><td>8</td></tr>
<tr><td colspan="2">N on common support</td><td>707</td></tr>
<tr><td>Rubin's B
Rubin's R</td><td>[before matching]</td><td>115.14
0.67</td></tr>
<tr><td>Rubin's B
Rubin's R</td><td>[after matching]</td><td>18.14
0.92</td></tr>
</table>

Strategy 1b

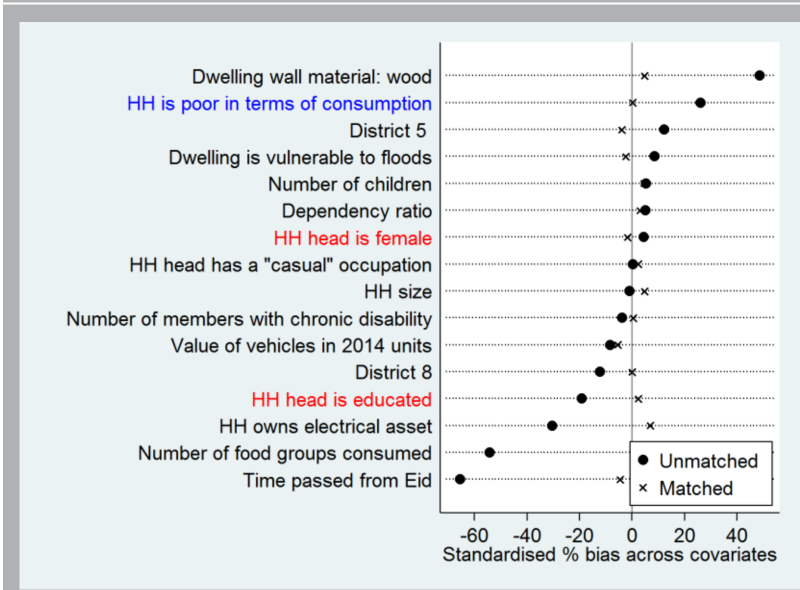


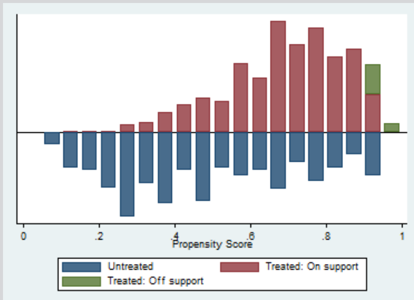


Bandwidth		2
Trimming		8
N on common support		730
Rubin's B Rubin's R	[before matching]	117.08 0.6
Rubin's B Rubin's R	[after matching]	15.14 1.19

Variables	Balancing

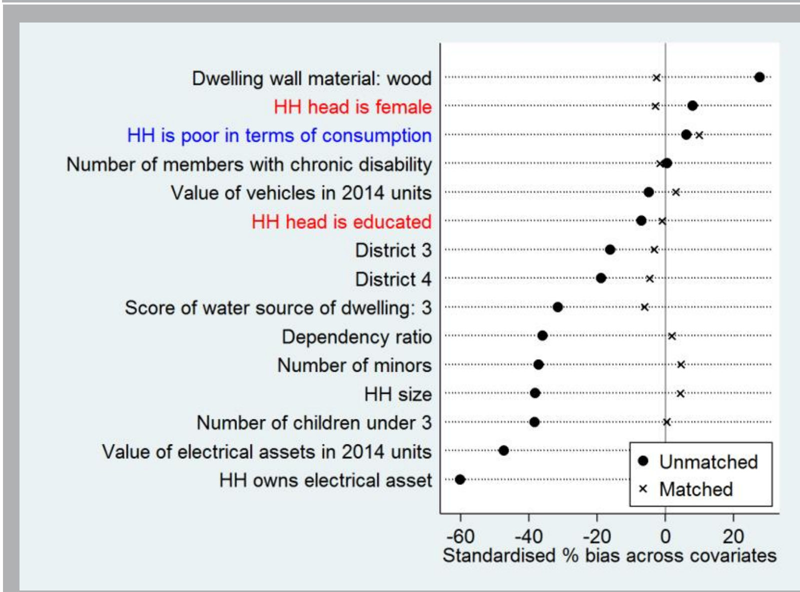
Strategy 2*

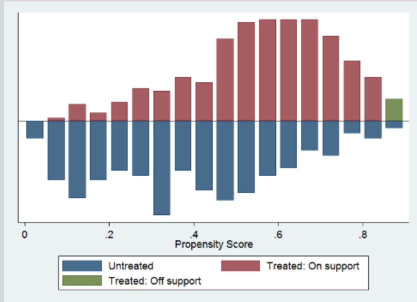




Bandwidth		3
Trimming		5
N on common support		1070
Rubin's B Rubin's R	[before matching]	101.47 0.52
Rubin's B Rubin's R	[after matching]	15.22 1.19

Robustness Test 1





Bandwidth		1
Trimming		3
N on common support		673
Rubin's B Rubin's R	[before matching]	94.83 0.58
Rubin's B Rubin's R	[after matching]	15.44 0.75

*Notes: outcome variable; theory-locked variables; *Baseline value of outcome variable has been used as matching variable*

Table 14 below presents the estimated ATT for all the strategies implemented. In the first place, a Kernel function has been used as the main algorithm to estimate treatment effects (see Annex G for details). Both bootstrapped and non-bootstrapped standard errors were calculated. Note that Volume I (OPM 2016a) plots these results for strategy 1b and 2 – using bootstrapped standard errors.

Furthermore, as an additional robustness check, average treatment effects were also estimated using a NN algorithm to ensure the results are not dependent on the matching algorithm chosen for the estimation.

Table 14 shows an ATT on consumption poverty prevalence of -0.53 percentage points for cohort 2.4 at T2 (strategy 1a). In the case of cohort 2.5 and T1, this treatment effect decreases up to -0.41 percentage points (strategy 1b). This reduction in the size of the treatment effect for cohort 2.5 could be explained by the fact that strategy 1b does not take into account the full CLP-2 intervention period of 18 months, since the treatment effect was estimated using the outcome measure at T1 (rather than T2) for the treatment cohort. Note that all estimates are significantly different from zero and negative – thus indicating strong reductions in poverty rates among CLP-2 participant households.

The results for strategy 2 show an ATT of -0.52, a figure slightly lower than the one obtained with strategy 1a. The difference between both figures can be explained by the fact that either cohort- or year-specific effects might have had an effect on our estimations. On the one hand, strategy 1a is implemented by comparing outcome measures for two cohorts within one year, which means that the estimated ATT is cohort-specific but should not be affected by year-specific effects. On the other hand, strategy 2 is implemented by comparing outcomes across more than two cohorts in different years, meaning that year-specific effects could play a role in the size of the estimated treatment effect, but cohort-specific effects less so (see section 6.4.3 for more detail on the different estimation strategies).

However, the fact that both amounts are similar seems to point to the fact that year- and cohort-specific effects may not be especially relevant in this particular case. Note that the results for strategy 2 and for strategy 1a both have overlapping confidence intervals. In fact, the results from robustness test 1 help to evaluate how problematic differential year or cohort effects could be in the current context. The estimate that we present is indistinguishable from zero, which counts as further evidence for similar impacts between cohorts or years for the cohorts involved in this estimation strategy.

Estimating ATT by using a NN algorithm provides additional robustness to our results. The magnitude of all our Kernel estimates seem to be confirmed by applying this alternative algorithm in the second stage analysis.

Table 14 Consumption poverty: ATT

	Strategy 1a (T1)	Strategy 1b (T2)	Strategy 2	Robustness 1
ATT (Kernel) SE (bootstrapping) *SE (no bootstrapping)*	-0.53 (0.037) *(0.043)*	-0.41 (0.041) *(0.040)*	-0.52 (0.024) *(0.031)*	-0.08 (0.049) *(0.046)*
ATT (NN)	-0.53 (0.044)	-0.39 (0.045)	-0.49 (0.038)	-0.07 (0.066)

2. Asset poverty – defined as the proportion of households lying under the asset poverty line[13]

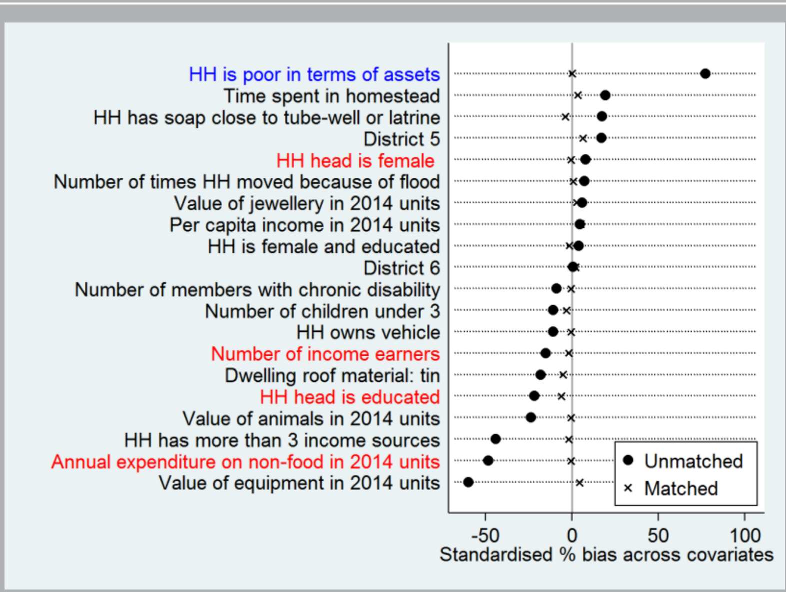
As in the previous case, the results and graphs in Table 15 below show that the matching process removes the imbalance for almost all covariates in the four strategies implemented. After matching, and as indicated by the many crosses close to zero, standardised bias across covariates is considerably reduced. In addition, after matching, Rubin's B lies under 25, and the distribution of propensity scores between treatment and control groups is comparable. Furthermore, balance is also achieved for the outcome variable of interest, pointing to the fact that the first stage matching procedure is performing well.

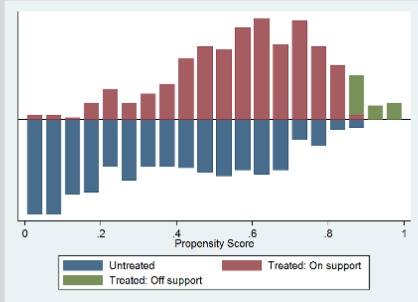
[13] Please refer to section 6.2.4 for a discussion on the creation of the consumption and asset poverty lines.

Table 15 Asset poverty: Selected variables and balancing across strategies

Variables	Balancing

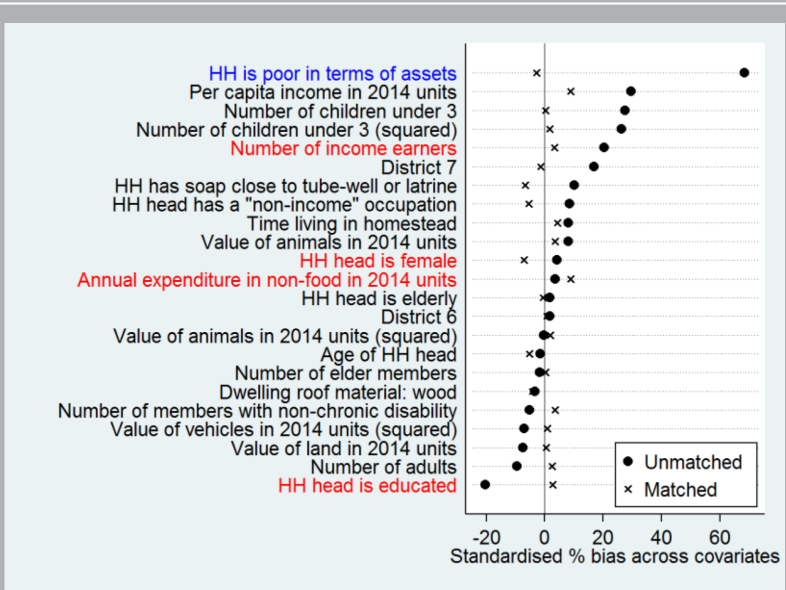
Strategy 1a*

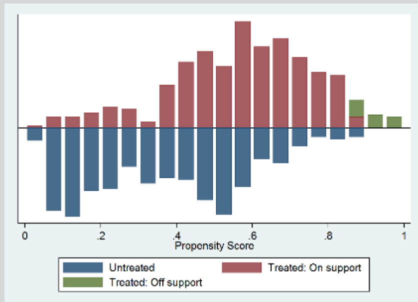




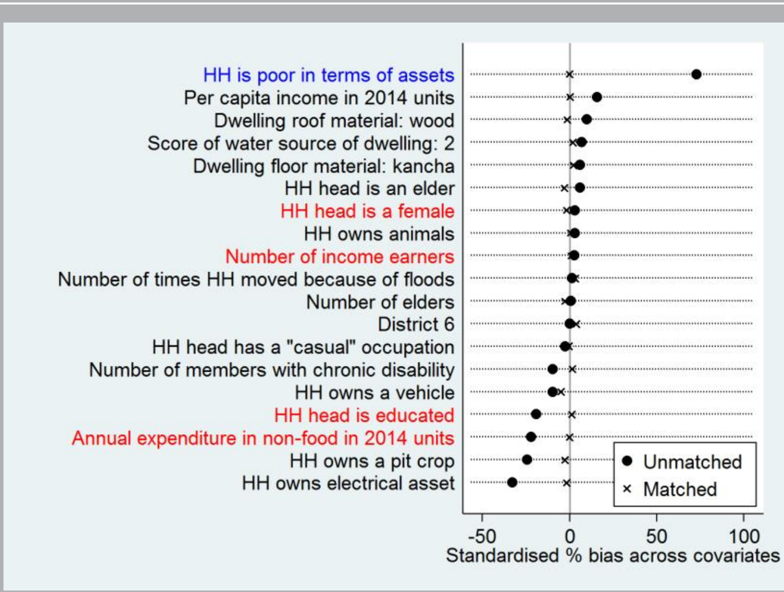
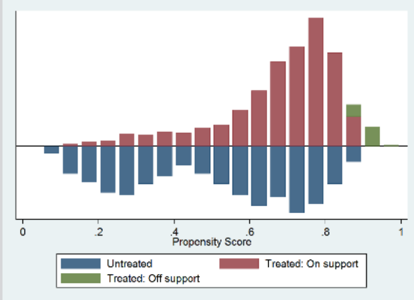
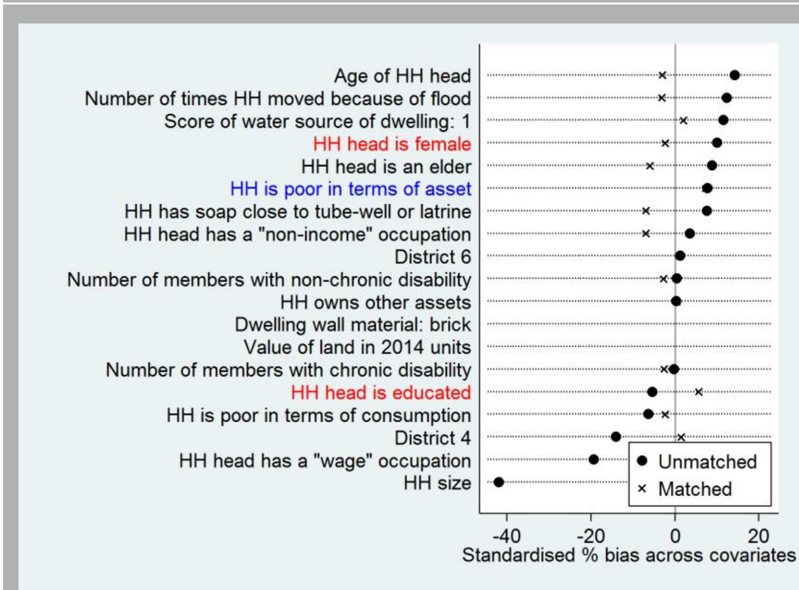
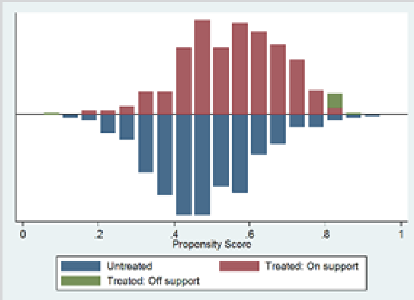
Bandwidth		6
Trimming		8
N on common support		810
Rubin's B Rubin's R	[before matching]	114.42 0.49
Rubin's B Rubin's R	[after matching]	18.52 1.02

Strategy 1b





Bandwidth		3
Trimming		5
N on common support		828
Rubin's B Rubin's R	[before matching]	102.66 0.76
Rubin's B Rubin's R	[after matching]	18.8 1

<table>
<tr><th>Variables</th><th colspan="3">Balancing</th></tr>
<tr><td colspan="4">Strategy 2*</td></tr>
<tr><td rowspan="7">
</td><td colspan="3">
</td></tr>
<tr><td colspan="2">Bandwidth</td><td>3</td></tr>
<tr><td colspan="2">Trimming</td><td>5</td></tr>
<tr><td colspan="2">N on common support</td><td>1167</td></tr>
<tr><td>Rubin's B
Rubin's R</td><td>[before matching]</td><td>90.97
0.52</td></tr>
<tr><td>Rubin's B
Rubin's R</td><td>[after matching]</td><td>10.65
1.01</td></tr>
<tr><td colspan="3"></td></tr>
<tr><td colspan="4">Robustness Test 1*</td></tr>
<tr><td rowspan="6">
</td><td colspan="3">
</td></tr>
<tr><td colspan="2">Bandwidth</td><td>3</td></tr>
<tr><td colspan="2">Trimming</td><td>3</td></tr>
<tr><td colspan="2">N on common support</td><td>734</td></tr>
<tr><td>Rubin's B
Rubin's R</td><td>[before matching]</td><td>60.81
1.01</td></tr>
<tr><td>Rubin's B
Rubin's R</td><td>[after matching]</td><td>15.42
1</td></tr>
<tr><td colspan="4">Notes: outcome variable; theory-locked variables; * Baseline value of outcome variable has been used as matching variable</td></tr>
</table>

With regards to the estimation of treatment effects, results in Table 16 indicate an ATT on asset poverty prevalence of -0.66 percentage points for cohort 2.4 (strategy 1a) in T2, a slightly larger decrease than the ATT observed on consumption poverty prevalence. In the case of cohort 2.5, this treatment effect decreases up to -0.64 percentage points (strategy 1b) at T1. As already explained, this reduction in the magnitude of the treatment effect for cohort 2.5 should be explained by the fact that strategy 1b does not take into account the full CLP-2 intervention period of 18 months.

The results for strategy 2 show an ATT of -0.70, a figure slightly larger than the one obtained with strategy 1a. The difference between the figures can be explained by the potential relevance that either cohort- or year-specific effects might have in our estimations. Nevertheless, both estimates are similar and not statistically different from each other, pointing again to the fact that year- and cohort-specific effects may not be especially relevant in this particular case. Results for robustness test 1 confirm this hypothesis, showing a negligible treatment impact of 0.04 percentage points of asset poverty prevalence. The magnitude of all our Kernel estimates seem to be confirmed by applying a NN algorithm in the second stage analysis.

Table 16 Asset poverty: ATT

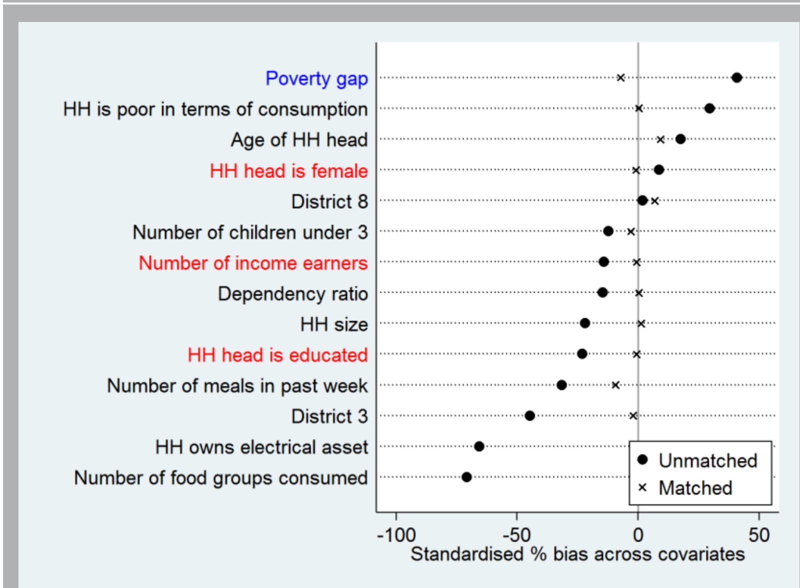
	Strategy 1a	Strategy 1b	Strategy 2	Robustness 1
ATT (Kernel) SE (bootstrapping) *SE (no bootstrapping)*	-0.66 (.026) *(.039)*	-0.64 (.026) *(.040)*	-0.70 (.018) *(.033)*	0.04 (.032) *(.032)*
ATT (NN)	-0.65 (.040)	-0.66 (.043)	-0.70 (.036)	0.07 (.040)

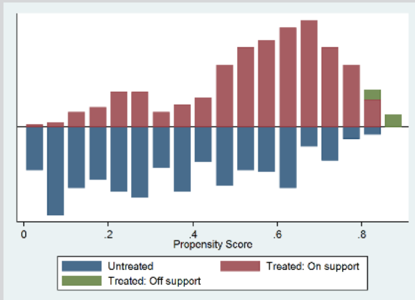
The following sections present both set of results (balancing of selected covariates and estimated ATT) for the remaining outcome variables of interest. All graphs and figures should be interpreted in the same manner as has been done in the above paragraphs.

3. Poverty gap: Poverty gap using consumption poverty line (percentage)

Table 17 Poverty gap (%): Selected variables and balancing across strategies

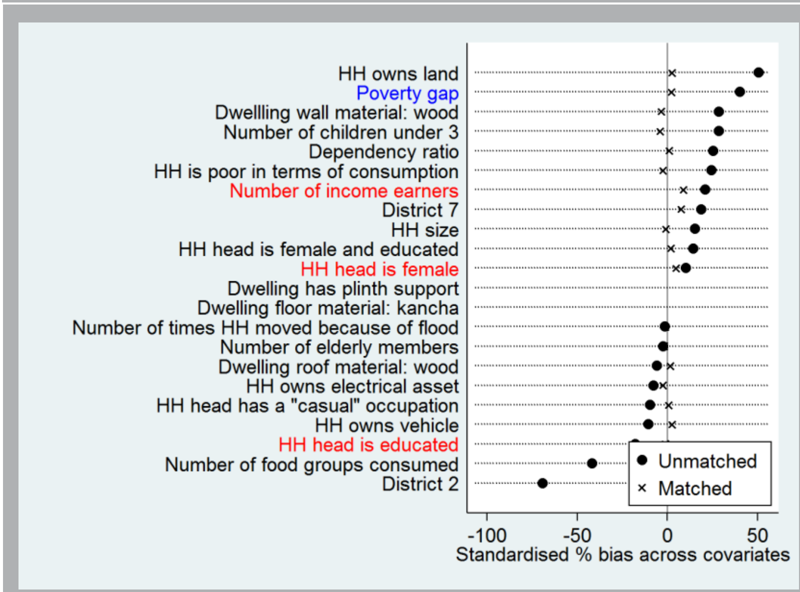
Variables	Balancing
Strategy 1a	

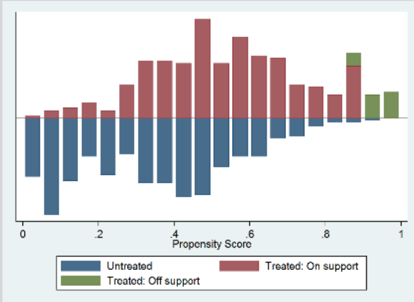




Bandwidth		4
Trimming		3
N on common support		724
Rubin's B Rubin's R	[before matching]	104.91 0.54
Rubin's B Rubin's R	[after matching]	17.93 0.99

Strategy 1b

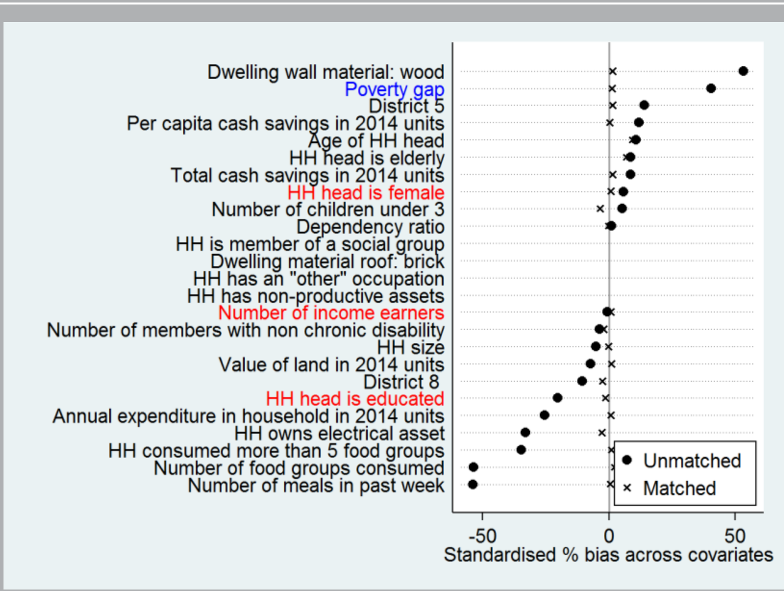


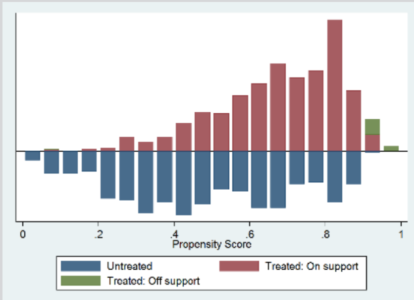


Bandwidth		4
Trimming		8
N on common support		690
Rubin's B Rubin's R	[before matching]	108.2 0.89
Rubin's B Rubin's R	[after matching]	15.82 1

Variables	Balancing

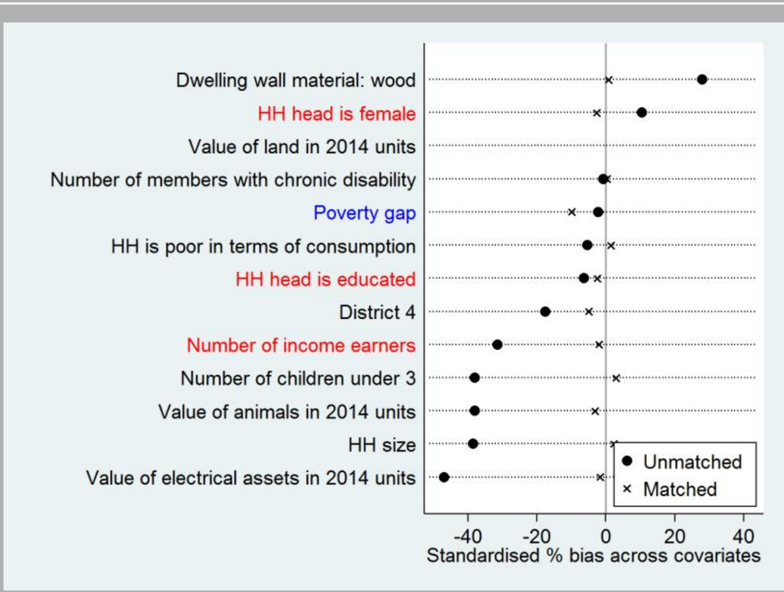
Strategy 2

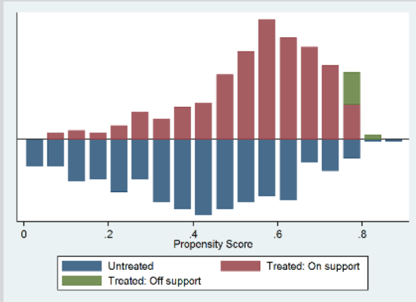




Bandwidth		2
Trimming		3
N on common support		1017
Rubin's B Rubin's R	[before matching]	96.54 0.58
Rubin's B Rubin's R	[after matching]	13.16 1.05

Robustness Test 1





Bandwidth		4
Trimming		5
N on common support		680
Rubin's B Rubin's R	[before matching]	83.54 0.44
Rubin's B Rubin's R	[after matching]	17.05 1.03

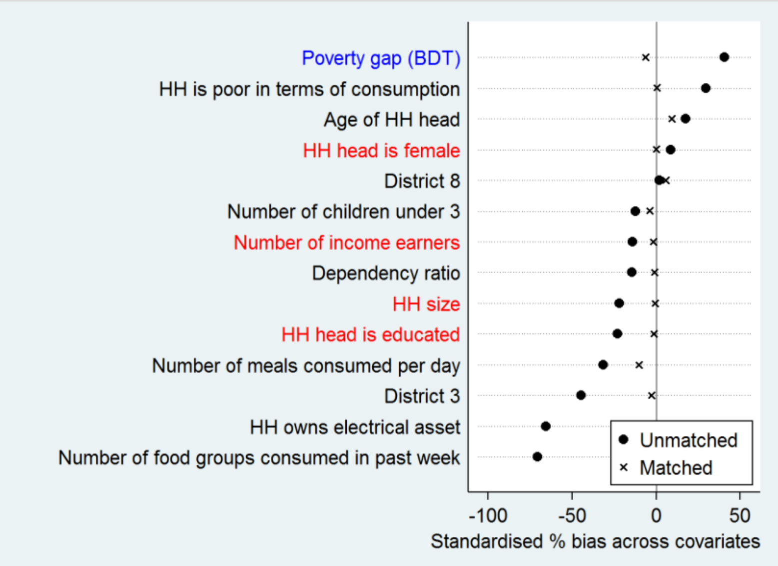
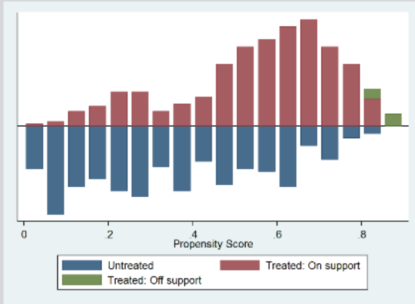
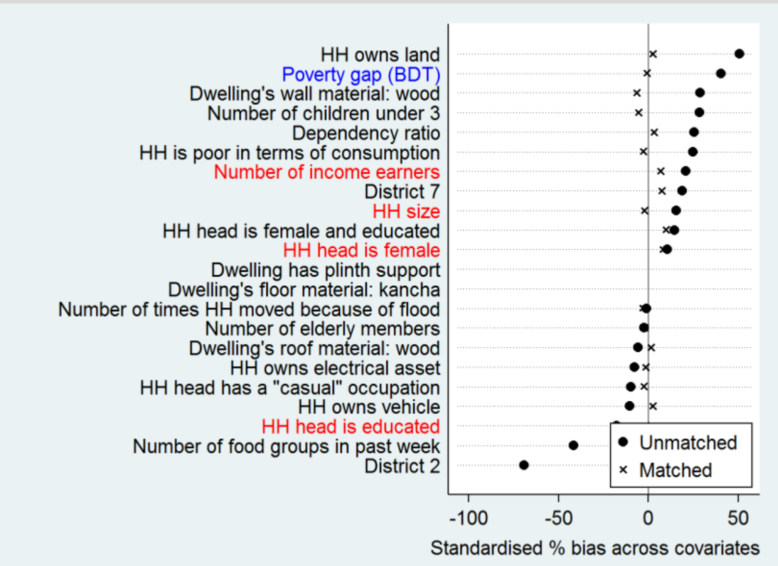
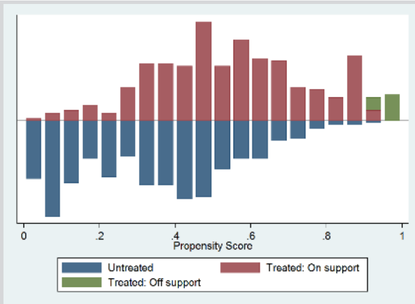
Notes: *outcome variable; theory-locked variables*

Table 18 Poverty gap (%): ATT

	Strategy 1a	Strategy 1b	Strategy 2	Robustness 1
ATT (Kernel) SE (bootstrapping) *SE (no bootstrapping)*	-0.18 (.008) *(.009)*	-0.15 (.009) *(.01)*	-0.17 (.007) *(.008)*	0 (.006) *(.006)*
ATT (NN)	-0.18 (.012)	-0.16 (.014)	-0.17 (.01)	-0.01 (.009)

4. Poverty gap: Poverty gap using consumption poverty line (in BDT)

Table 19 Poverty gap (BDT): Selected variables and balancing across strategies

<table>
<tr><th>Variables</th><th colspan="3">Balancing</th></tr>
<tr><td colspan="4">Strategy 1a</td></tr>
<tr><td rowspan="6">
</td><td colspan="3">
</td></tr>
<tr><td colspan="2">Bandwidth</td><td>6</td></tr>
<tr><td colspan="2">Trimming</td><td>3</td></tr>
<tr><td colspan="2">N on common support</td><td>724</td></tr>
<tr><td>Rubin's B
Rubin's R</td><td>[before matching]</td><td>104.94
0.54</td></tr>
<tr><td>Rubin's B
Rubin's R</td><td>[after matching]</td><td>18.69
1.07</td></tr>
<tr><td colspan="4">Strategy 1b</td></tr>
<tr><td rowspan="6">
</td><td colspan="3">
</td></tr>
<tr><td colspan="2">Bandwidth</td><td>4</td></tr>
<tr><td colspan="2">Trimming</td><td>5</td></tr>
<tr><td colspan="2">N on common support</td><td>699</td></tr>
<tr><td>Rubin's B
Rubin's R</td><td>[before matching]</td><td>108.2
0.89</td></tr>
<tr><td>Rubin's B
Rubin's R</td><td>[after matching]</td><td>19.02
1.08</td></tr>
</table>

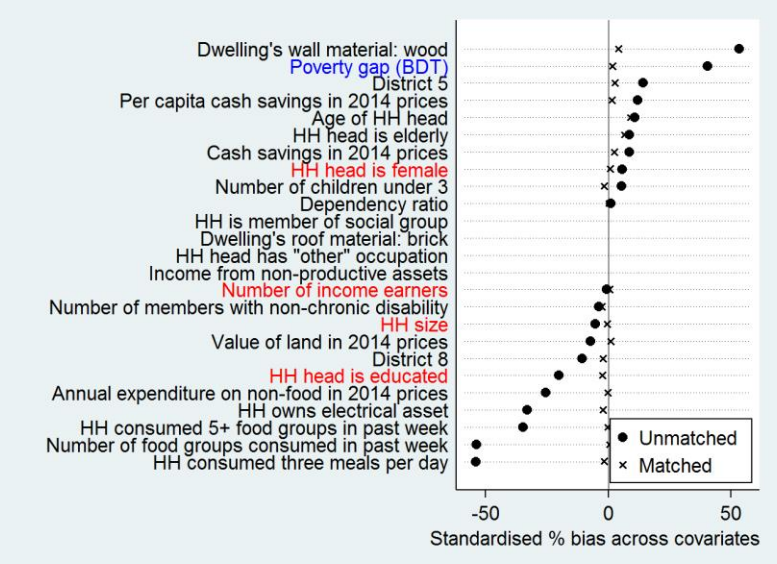
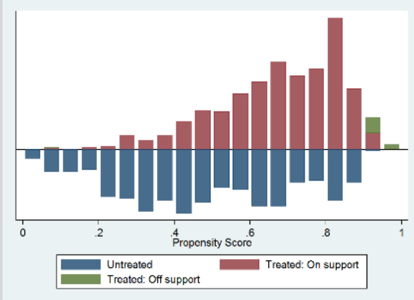
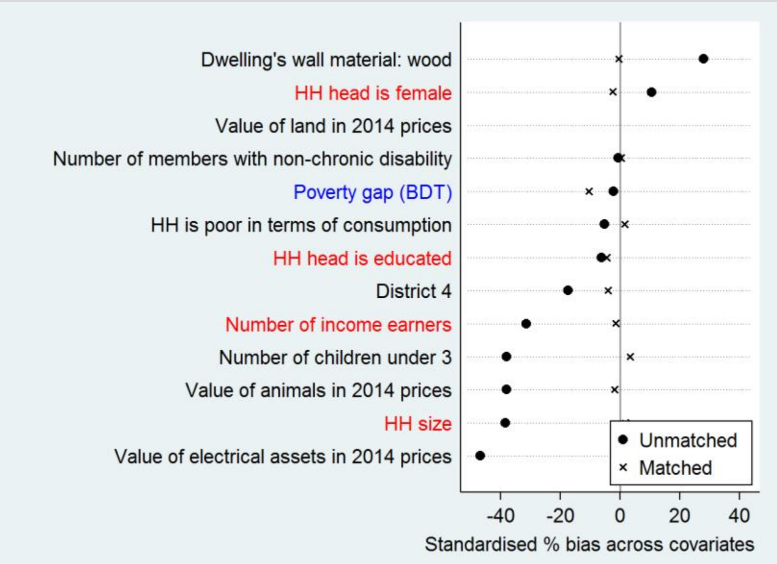
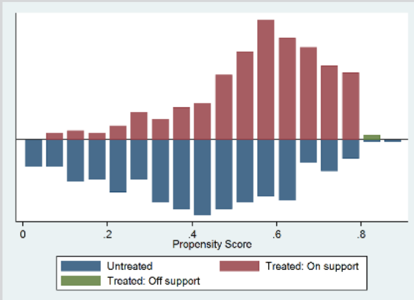
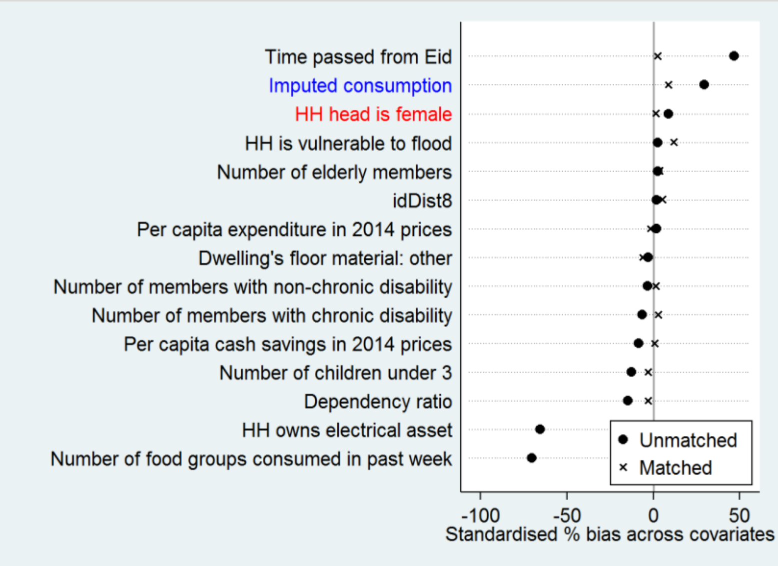
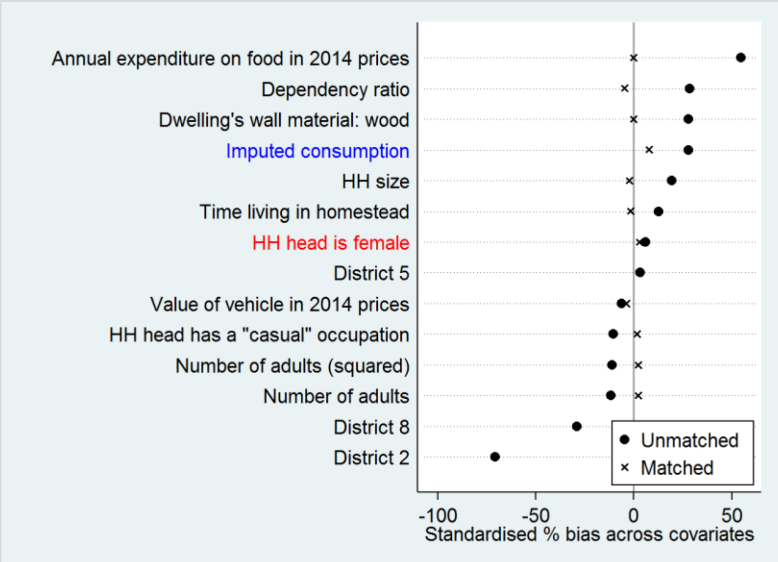
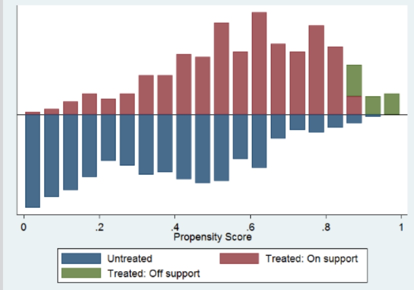
Variables	Balancing		
Strategy 2			
			
	Bandwidth		6
	Trimming		3
	N on common support		1017
	Rubin's B Rubin's R	[before matching]	96.54 0.58
	Rubin's B Rubin's R	[after matching]	13.28 1.12
Robustness Test 1			
			
	Bandwidth		2
	Trimming		3
	N on common support		694
	Rubin's B Rubin's R	[before matching]	83.54 0.44
	Rubin's B Rubin's R	[after matching]	16.98 1.09
Notes: *outcome variable; theory-locked variables*			

Table 20 Poverty gap (BDT): ATT

	Strategy 1a	Strategy 1b	Strategy 2	Robustness 1
ATT (Kernel) SE (bootstrapping) *SE (no bootstrapping)*	-297.54 (13.794) *(14.5)*	-250.51 (15.586) *(16.806)*	-283.85 (11.068) *(12.945)*	-8.8 (9.961) *(9.825)*
ATT (NN)	-306.92 (19.398)	-270.08 (21.646)	-290.45 (16.724)	-9.77 (14.81)

5. *Imputed consumption in 2014 prices*

Table 21 Imputed consumption: Selected variables and balancing across strategies

<table>
<tr><th>Variables</th><th colspan="3">Balancing</th></tr>
<tr><td colspan="4">Strategy 1a</td></tr>
<tr><td rowspan="6">
</td><td colspan="3">
</td></tr>
<tr><td colspan="2">Bandwidth</td><td>3</td></tr>
<tr><td colspan="2">Trimming</td><td>3</td></tr>
<tr><td colspan="2">N on common support</td><td>723</td></tr>
<tr><td>Rubin's B
Rubin's R</td><td>[before matching]</td><td>115.14
0.67</td></tr>
<tr><td>Rubin's B
Rubin's R</td><td>[after matching]</td><td>20.12
0.92</td></tr>
<tr><td colspan="4">Strategy 1b</td></tr>
<tr><td rowspan="6">
</td><td colspan="3">
</td></tr>
<tr><td colspan="2">Bandwidth</td><td>2</td></tr>
<tr><td colspan="2">Trimming</td><td>8</td></tr>
<tr><td colspan="2">N on common support</td><td>730</td></tr>
<tr><td>Rubin's B
Rubin's R</td><td>[before matching]</td><td>117.08
0.6</td></tr>
<tr><td>Rubin's B
Rubin's R</td><td>[after matching]</td><td>15.14
1.19</td></tr>
</table>

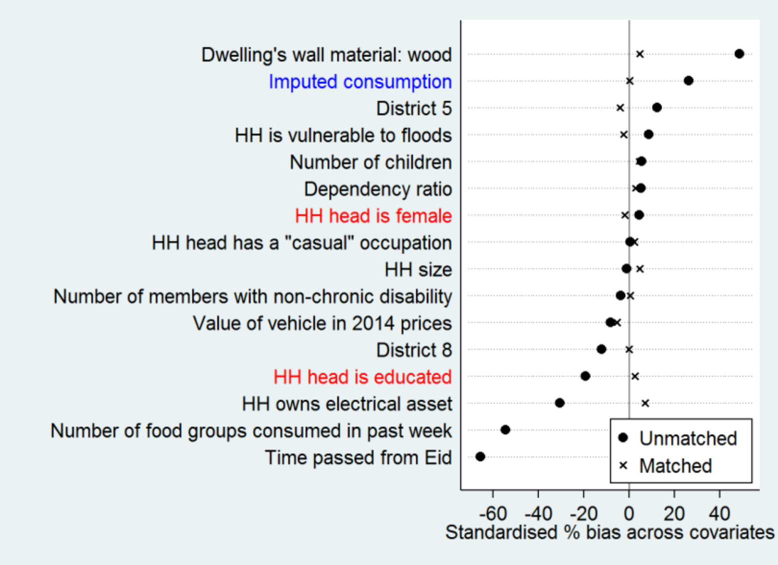
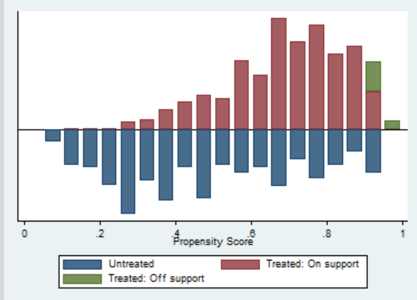
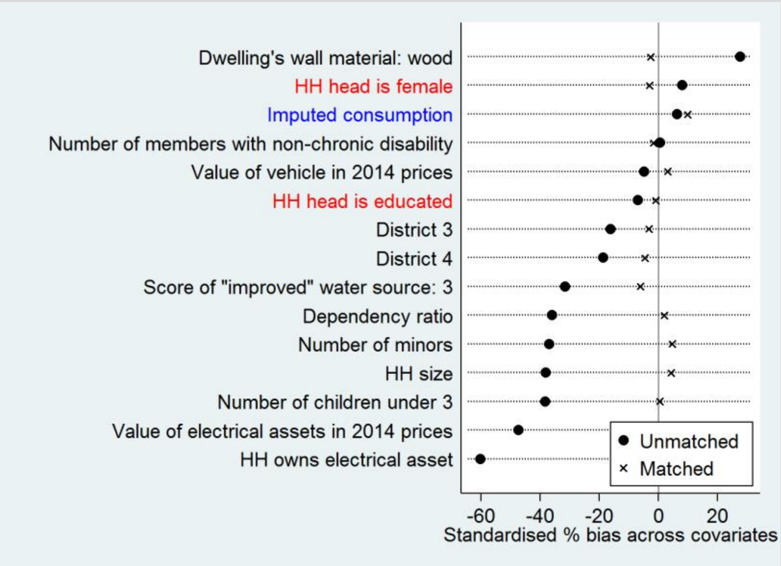
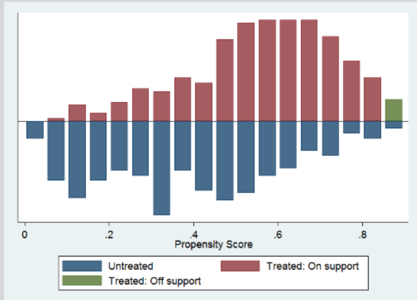
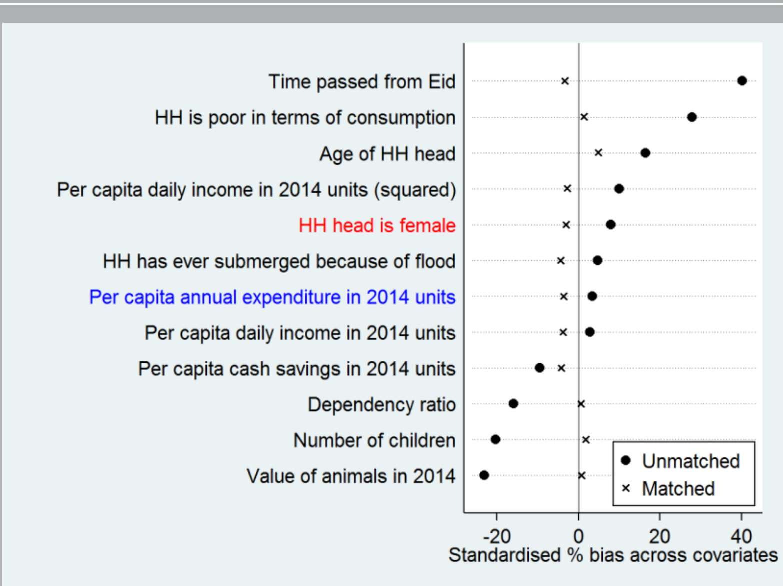
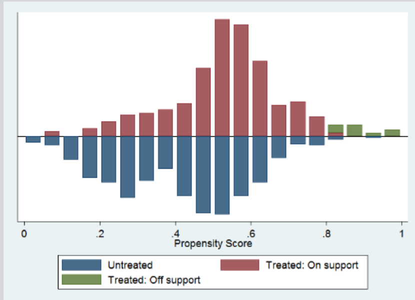
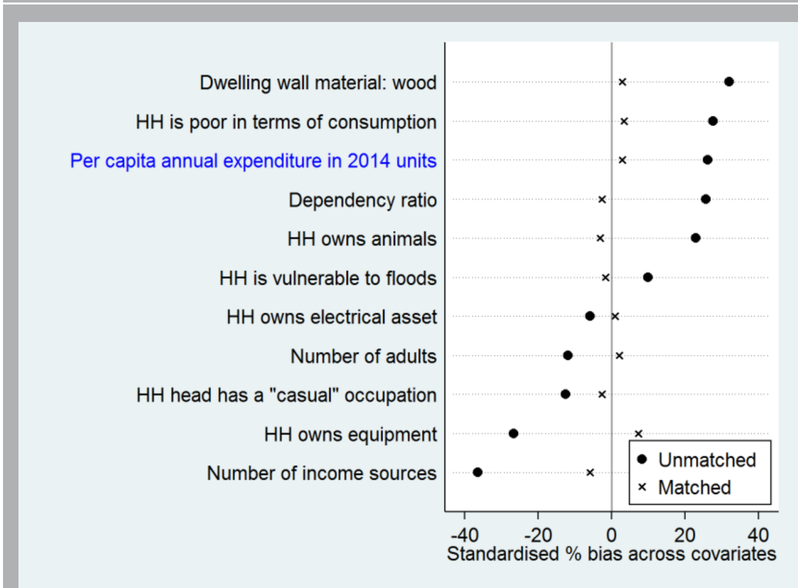
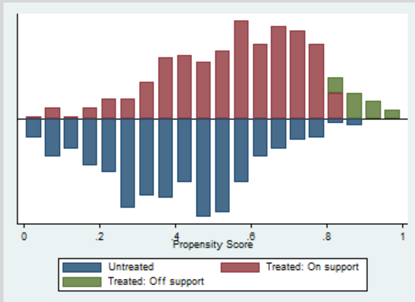
<table>
<tr><th>Variables</th><th colspan="3">Balancing</th></tr>
<tr><td colspan="4">Strategy 2</td></tr>
<tr><td rowspan="6">
</td><td colspan="3">
</td></tr>
<tr><td colspan="2">Bandwidth</td><td>3</td></tr>
<tr><td colspan="2">Trimming</td><td>5</td></tr>
<tr><td colspan="2">N on common support</td><td>1055</td></tr>
<tr><td>Rubin's B
Rubin's R</td><td>[before matching]</td><td>101.47
0.52</td></tr>
<tr><td>Rubin's B
Rubin's R</td><td>[after matching]</td><td>15.22
1.19</td></tr>
<tr><td colspan="4">Robustness Test 1</td></tr>
<tr><td rowspan="6">
</td><td colspan="3">
</td></tr>
<tr><td colspan="2">Bandwidth</td><td>1</td></tr>
<tr><td colspan="2">Trimming</td><td>3</td></tr>
<tr><td colspan="2">N on common support</td><td>673</td></tr>
<tr><td>Rubin's B
Rubin's R</td><td>[before matching]</td><td>94.83
0.58</td></tr>
<tr><td>Rubin's B
Rubin's R</td><td>[after matching]</td><td>15.44
0.75</td></tr>
<tr><td colspan="4">Notes: outcome variable; theory-locked variables</td></tr>
</table>

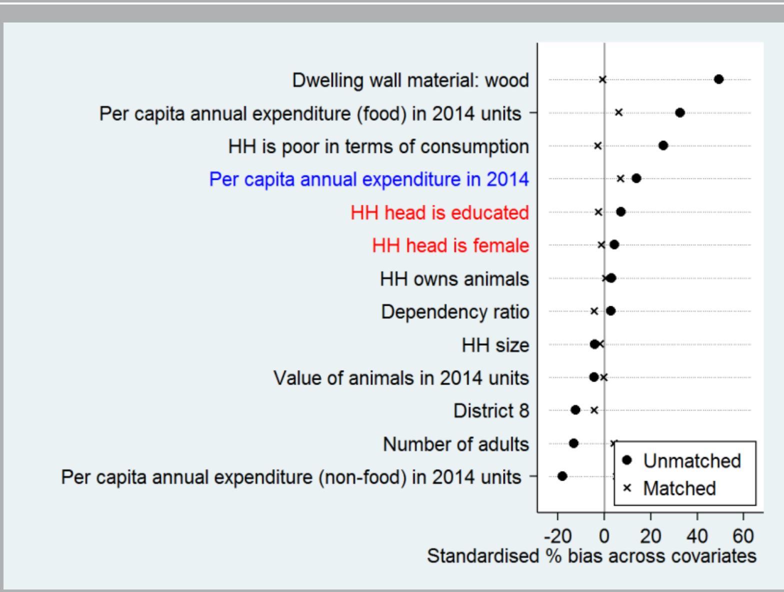
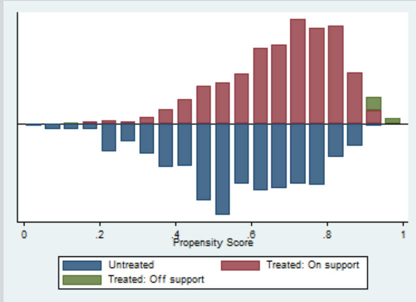
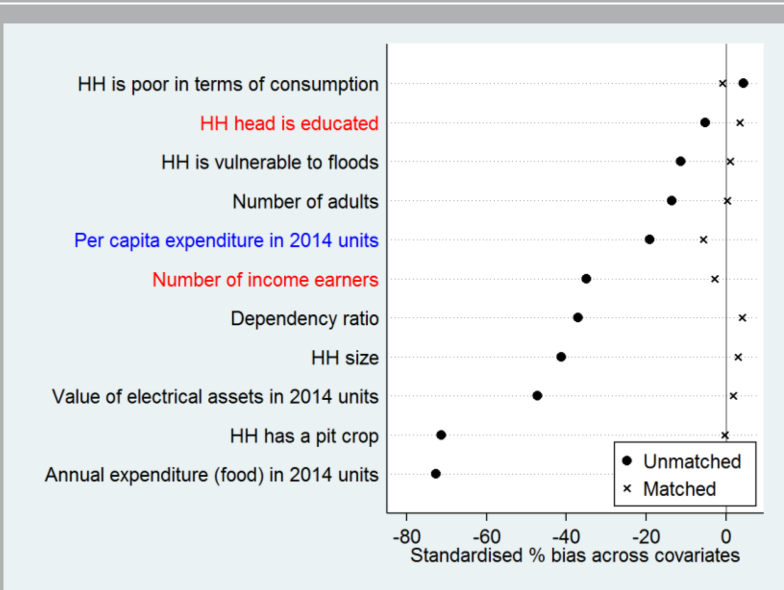
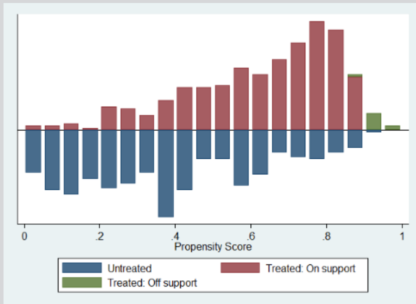
Table 22 Imputed consumption: Average Treatment Effect on the Treated (ATT)

	Strategy 1a	Strategy 1b	Strategy 2	Robustness 1
ATT (Kernel) SE (bootstrapping) *SE (no bootstrapping)*	524.91 (31.105) *(37.98)*	396.00 (29.411) *(37.208)*	484.33 (22.247) *(26.884)*	93.85 (40.427) *(37.463)*
ATT (NN)	494.4 (45.17)	394.24 (40.655)	453.01 (37.604)	108.27 (53.281)

6. *Per capita expenditure in 2014 monetary units*

Table 23 Expenditure: Selected variables and balancing across strategies

<table>
<tr><th>Variables</th><th colspan="3">Balancing</th></tr>
<tr><td colspan="4">Strategy 1a</td></tr>
<tr><td rowspan="6">
</td><td colspan="3">
</td></tr>
<tr><td colspan="2">Bandwidth</td><td>2</td></tr>
<tr><td colspan="2">Trimming</td><td>5</td></tr>
<tr><td colspan="2">N on common support</td><td>766</td></tr>
<tr><td>Rubin's B
Rubin's R</td><td>[before matching]</td><td>73.25
1.04</td></tr>
<tr><td>Rubin's B
Rubin's R</td><td>[after matching]</td><td>12.51
1.11</td></tr>
<tr><td colspan="4">Strategy 1b*</td></tr>
<tr><td rowspan="6">
</td><td colspan="3">
</td></tr>
<tr><td colspan="2">Bandwidth</td><td>4</td></tr>
<tr><td colspan="2">Trimming</td><td>8</td></tr>
<tr><td colspan="2">N on common support</td><td>765</td></tr>
<tr><td>Rubin's B
Rubin's R</td><td>[before matching]</td><td>90.89
1.02</td></tr>
<tr><td>Rubin's B
Rubin's R</td><td>[after matching]</td><td>12.83
1.39</td></tr>
</table>

Variables	Balancing		
Strategy 2			
 	 		
	Bandwidth		1
	Trimming		3
	N on common support		1120
	Rubin's B Rubin's R	[before matching]	77.63 0.81
	Rubin's B Rubin's R	[after matching]	10.35 1.56
Robustness Test 1			
 	 		
	Bandwidth		1
	Trimming		3
	N on common support		720
	Rubin's B Rubin's R	[before matching]	112.6 0.51
	Rubin's B Rubin's R	[after matching]	9.4 0.94

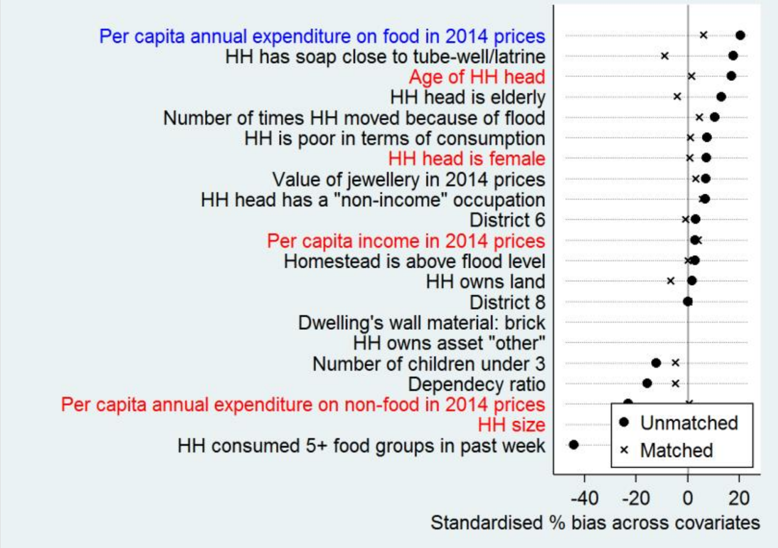
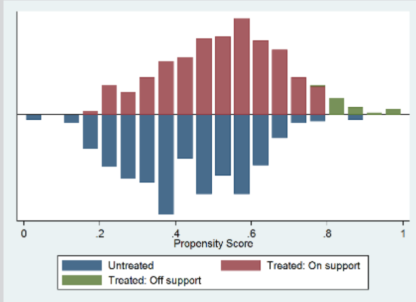
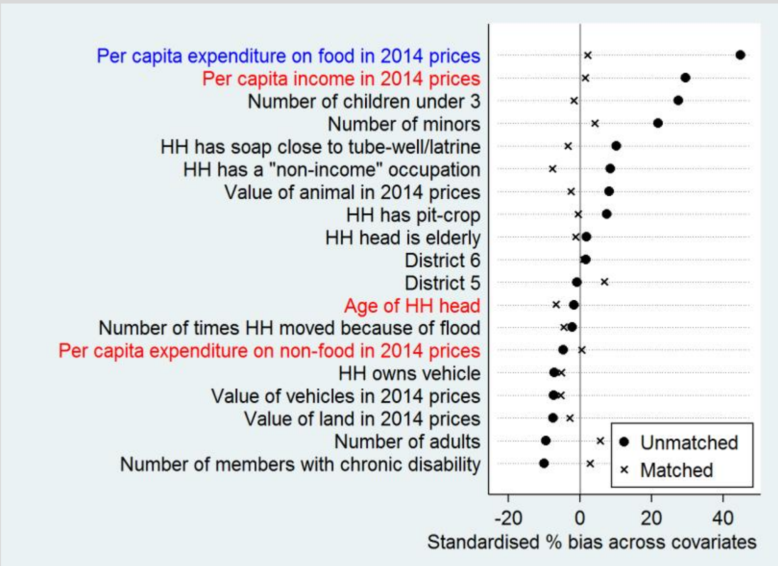
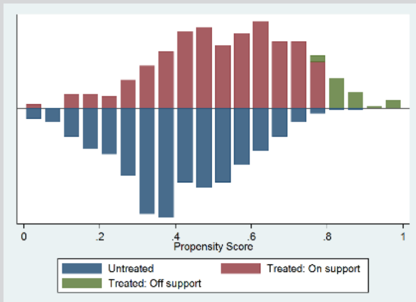
*Notes: outcome variable; theory-locked variables; * Baseline value of outcome variable has been used as matching variable*

Table 24 Expenditure: ATT

	Strategy 1a	Strategy 1b	Strategy 2	Robustness 1
ATT (Kernel) SE (bootstrapping) *SE (no bootstrapping)*	10,149.71 (512.673) *(524.163)*	6,874.74 (348.833) *(355.15)*	9,664.06 (316.324) *(346.622)*	611.54 (888.421) *(839.411)*
ATT (NN)	10,207.23 (521.36)	6,844.24 (382.665)	9,687.73 (352.861)	746.57 (1041.572)

7. Per capita annual expenditure on food

Table 25 Food expenditure: Selected variables and balancing across strategies

<table>
<tr><th>Variables</th><th colspan="3">Balancing</th></tr>
<tr><td colspan="4">Strategy 1a</td></tr>
<tr><td rowspan="7">
</td><td colspan="3">
</td></tr>
<tr><td colspan="2">Bandwidth</td><td>3</td></tr>
<tr><td colspan="2">Trimming</td><td>5</td></tr>
<tr><td colspan="2">N on common support</td><td>763</td></tr>
<tr><td>Rubin's B
Rubin's R</td><td>[before matching]</td><td>70.07
1.04</td></tr>
<tr><td>Rubin's B
Rubin's R</td><td>[after matching]</td><td>14.43
1.01</td></tr>
<tr><td colspan="4">Strategy 1b</td></tr>
<tr><td rowspan="6">
</td><td colspan="3">
</td></tr>
<tr><td colspan="2">Bandwidth</td><td>2</td></tr>
<tr><td colspan="2">Trimming</td><td>8</td></tr>
<tr><td colspan="2">N on common support</td><td>816</td></tr>
<tr><td>Rubin's B
Rubin's R</td><td>[before matching]</td><td>80.12
1.43</td></tr>
<tr><td>Rubin's B
Rubin's R</td><td>[after matching]</td><td>20.06
1.01</td></tr>
</table>

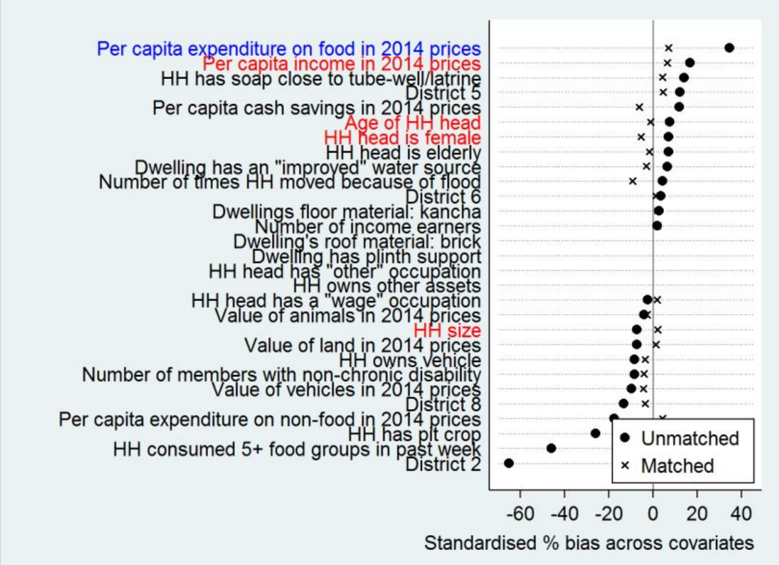
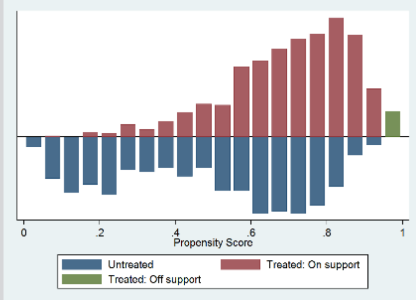
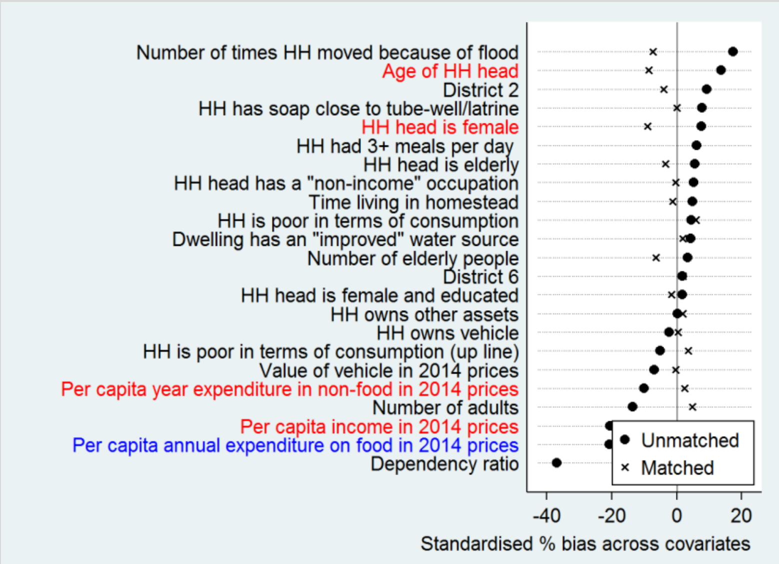
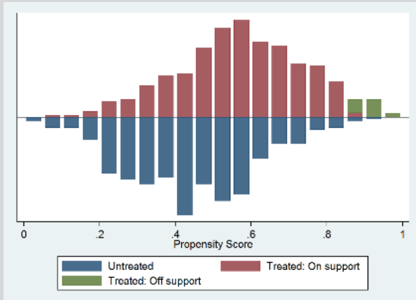
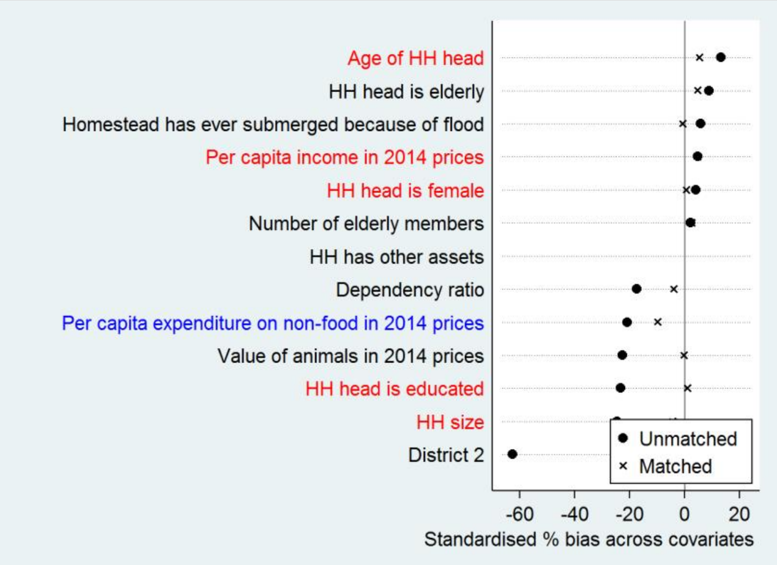
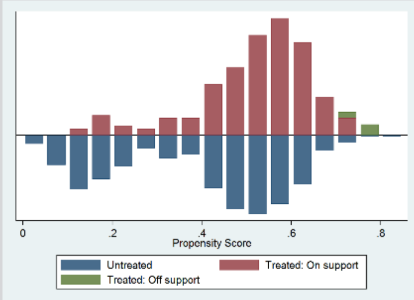
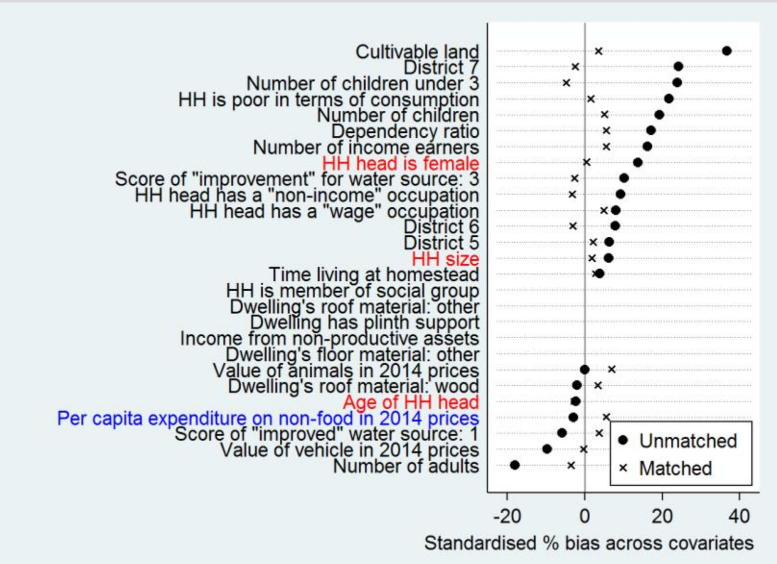
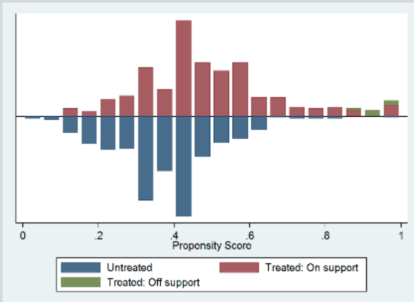
<table>
<tr><th>Variables</th><th colspan="3">Balancing</th></tr>
<tr><td colspan="4">Strategy 2</td></tr>
<tr><td rowspan="6">
</td><td colspan="3">
</td></tr>
<tr><td colspan="2">Bandwidth</td><td>2</td></tr>
<tr><td colspan="2">Trimming</td><td>3</td></tr>
<tr><td colspan="2">N on common support</td><td>1101</td></tr>
<tr><td>Rubin's B
Rubin's R</td><td>[before matching]</td><td>101.4
0.53</td></tr>
<tr><td>Rubin's B
Rubin's R</td><td>[after matching]</td><td>21.36
1.06</td></tr>
<tr><td colspan="4">Robustness Test 1</td></tr>
<tr><td rowspan="6">
</td><td colspan="3">
</td></tr>
<tr><td colspan="2">Bandwidth</td><td>4</td></tr>
<tr><td colspan="2">Trimming</td><td>5</td></tr>
<tr><td colspan="2">N on common support</td><td>714</td></tr>
<tr><td>Rubin's B
Rubin's R</td><td>[before matching]</td><td>78.34
1.05</td></tr>
<tr><td>Rubin's B
Rubin's R</td><td>[after matching]</td><td>17.81
1.0</td></tr>
<tr><td colspan="4">Notes: outcome variable; theory-locked variables</td></tr>
</table>

Table 26 Food expenditure: ATT

	Strategy 1a	Strategy 1b	Strategy 2	Robustness 1
ATT (Kernel) SE (bootstrapping) *SE (no bootstrapping)*	1330.76 (117.366) *(125.229)*	812.49 (119.516) *(122.546)*	977.77 (102.681) *(110.679)*	269.62 (185.834) *(162.759)*
ATT (NN)	1187.38 (153.175)	539.26 (151.923)	872.8 (140.308)	48.26 (229.373)

8. Per capita annual expenditure on non-food

Table 27 Non-food expenditure: Selected variables and balancing across strategies

<table>
<tr><th>Variables</th><th colspan="3">Balancing</th></tr>
<tr><td colspan="4">Strategy 1a</td></tr>
<tr><td rowspan="6">
</td><td colspan="3">
</td></tr>
<tr><td colspan="2">Bandwidth</td><td>4</td></tr>
<tr><td colspan="2">Trimming</td><td>3</td></tr>
<tr><td colspan="2">N on common support</td><td>806</td></tr>
<tr><td>Rubin’s B
Rubin’s R</td><td>[before matching]</td><td>79.36
0.41</td></tr>
<tr><td>Rubin’s B
Rubin’s R</td><td>[after matching]</td><td>23.24
1.0</td></tr>
<tr><td colspan="4">Strategy 1b</td></tr>
<tr><td rowspan="6">
</td><td colspan="3">
</td></tr>
<tr><td colspan="2">Bandwidth</td><td>4</td></tr>
<tr><td colspan="2">Trimming</td><td>3</td></tr>
<tr><td colspan="2">N on common support</td><td>692</td></tr>
<tr><td>Rubin’s B
Rubin’s R</td><td>[before matching]</td><td>70.05
2.23</td></tr>
<tr><td>Rubin’s B
Rubin’s R</td><td>[after matching]</td><td>19.87
1.16</td></tr>
</table>

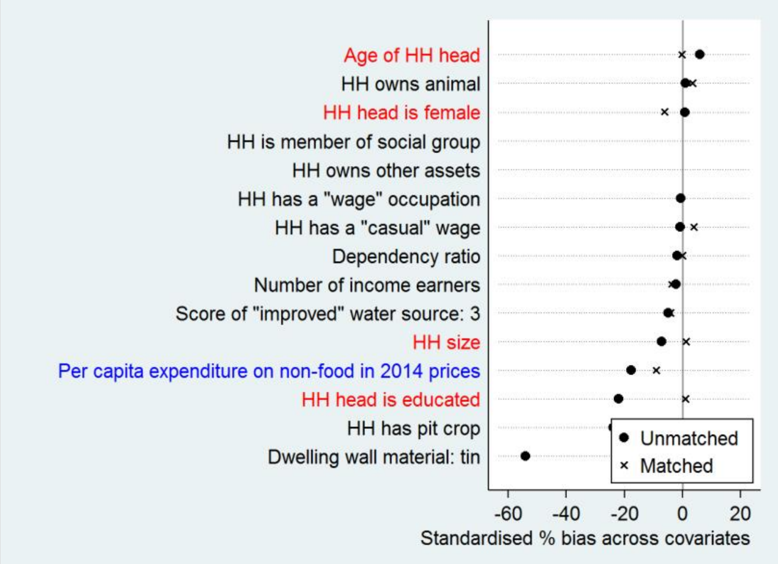
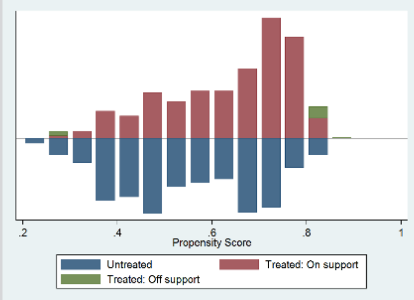
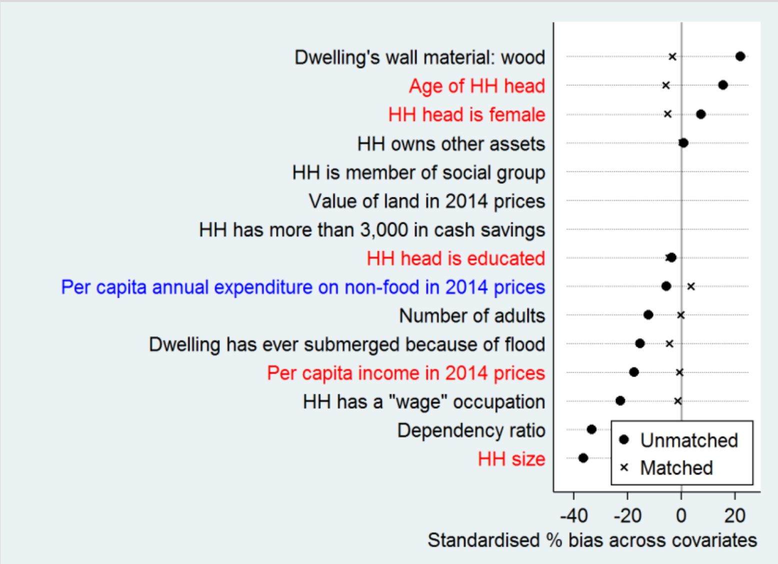
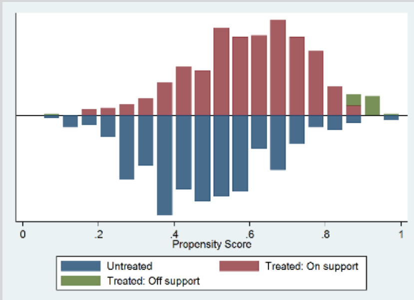
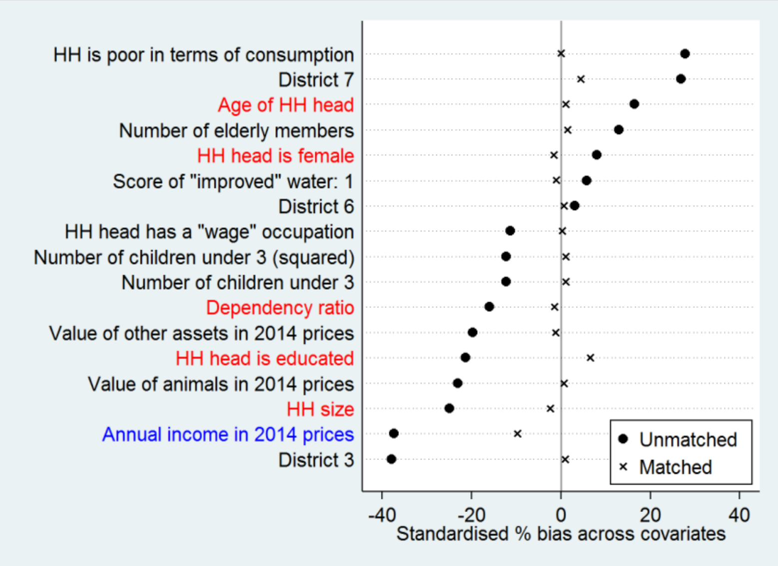
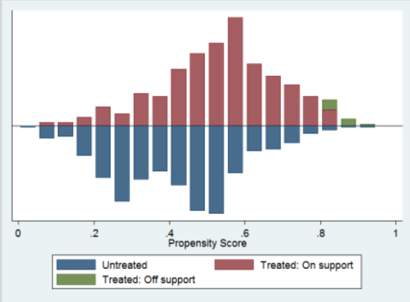
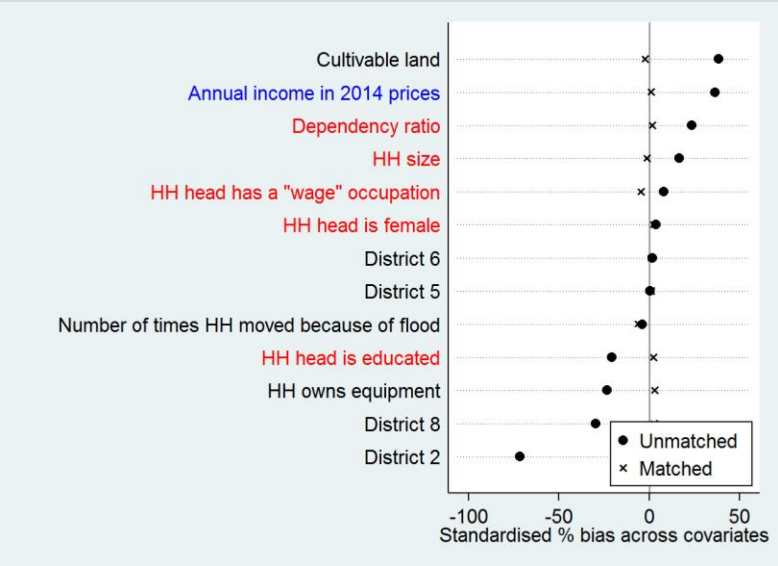
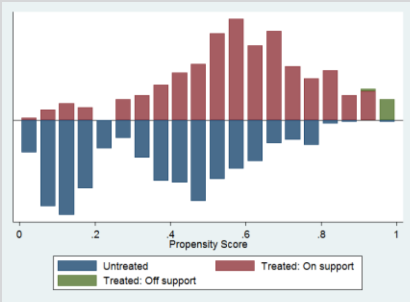
<table>
<tr><th colspan="2">Variables</th><th colspan="3">Balancing</th></tr>
<tr><td colspan="5">Strategy 2</td></tr>
<tr><td colspan="2" rowspan="6">
</td><td colspan="3">
</td></tr>
<tr><td colspan="2">Bandwidth</td><td>4</td></tr>
<tr><td colspan="2">Trimming</td><td>3</td></tr>
<tr><td colspan="2">N on common support</td><td>1103</td></tr>
<tr><td>Rubin's B
Rubin's R</td><td>[before matching]</td><td>65.56
0.83</td></tr>
<tr><td>Rubin's B
Rubin's R</td><td>[after matching]</td><td>15.24
1.04</td></tr>
<tr><td colspan="5">Robustness Test 1</td></tr>
<tr><td colspan="2" rowspan="6">
</td><td colspan="3">
</td></tr>
<tr><td colspan="2">Bandwidth</td><td>3</td></tr>
<tr><td colspan="2">Trimming</td><td>5</td></tr>
<tr><td colspan="2">N on common support</td><td>662</td></tr>
<tr><td>Rubin's B
Rubin's R</td><td>[before matching]</td><td>77.39
1.0</td></tr>
<tr><td>Rubin's B
Rubin's R</td><td>[after matching]</td><td>12.57
1.03</td></tr>
<tr><td colspan="5">Notes: outcome variable; theory-locked variables</td></tr>
</table>

Table 28 Non-food expenditure: ATT

	Strategy 1a	Strategy 2	Strategy 1b	Robustness 1
ATT (Kernel) SE (bootstrapping) *SE (no bootstrapping)*	8868.43 (447.056) *(447.469)*	6122.88 (345.03) *(352.062)*	8379.12 (344.602) *(319.495)*	-104.23 (845.734) *(731.29)*
ATT (NN)	8863.17 (453.195)	6165.76 (357.543)	8322 (327.893)	-401.69 (937.382)

9. Total income in 2014 monetary units

Table 29 Total income: Selected variables and balancing across strategies

<table>
<tr><th>Variables</th><th colspan="3">Balancing</th></tr>
<tr><td colspan="4">Strategy 1a</td></tr>
<tr><td rowspan="6">
</td><td colspan="3">
</td></tr>
<tr><td colspan="2">Bandwidth</td><td>2</td></tr>
<tr><td colspan="2">Trimming</td><td>3</td></tr>
<tr><td colspan="2">N on common support</td><td>784</td></tr>
<tr><td>Rubin's B
Rubin's R</td><td>[before matching]</td><td>75.77
0.74</td></tr>
<tr><td>Rubin's B
Rubin's R</td><td>[after matching]</td><td>17.95
1.14</td></tr>
<tr><td colspan="4">Strategy 1b*</td></tr>
<tr><td rowspan="6">
</td><td colspan="3">
</td></tr>
<tr><td colspan="2">Bandwidth</td><td>2</td></tr>
<tr><td colspan="2">Trimming</td><td>3</td></tr>
<tr><td colspan="2">N on common support</td><td>829</td></tr>
<tr><td>Rubin's B
Rubin's R</td><td>[before matching]</td><td>109.01
0.85</td></tr>
<tr><td>Rubin's B
Rubin's R</td><td>[after matching]</td><td>12.32
1.15</td></tr>
</table>

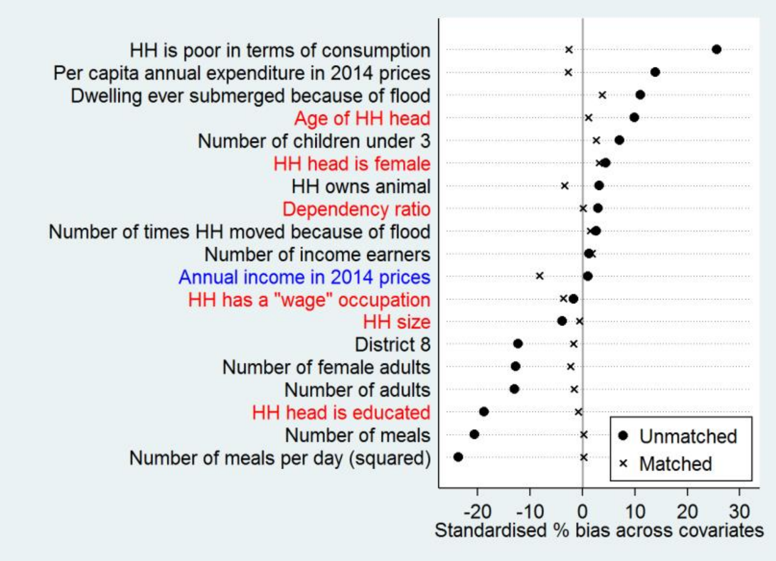
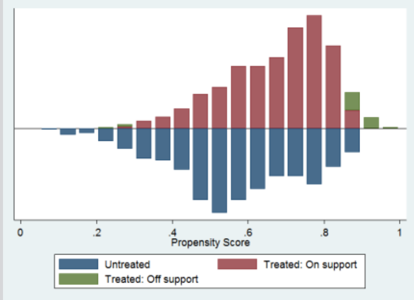
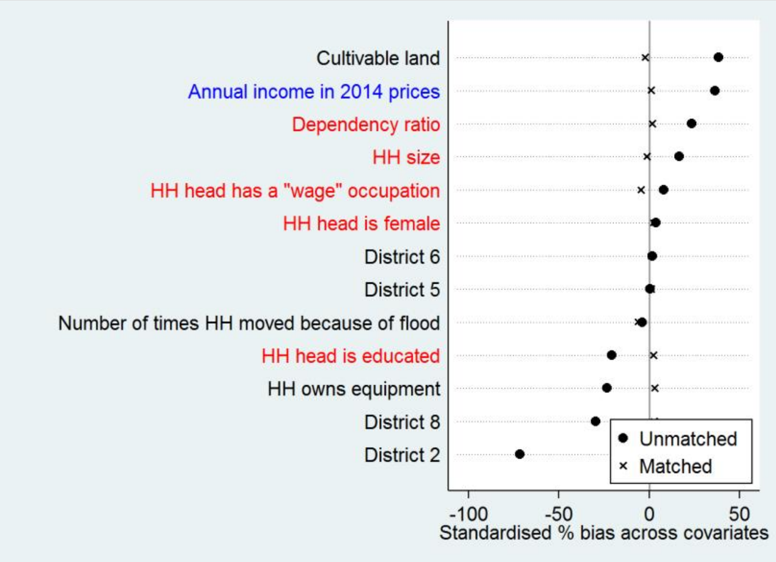
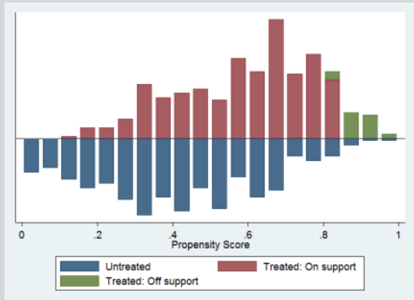
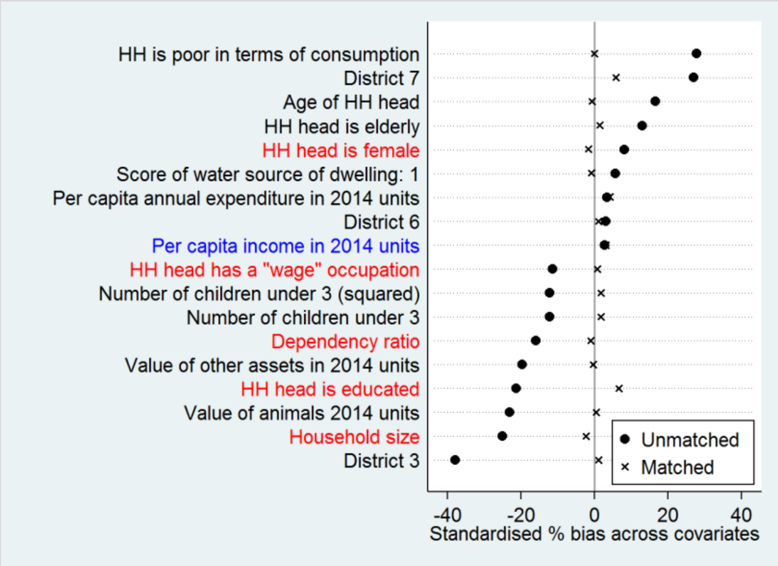
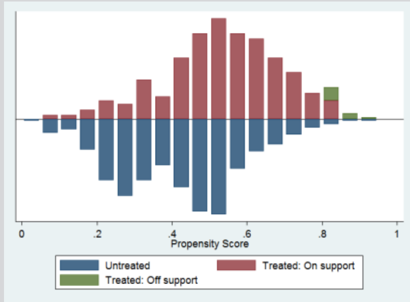
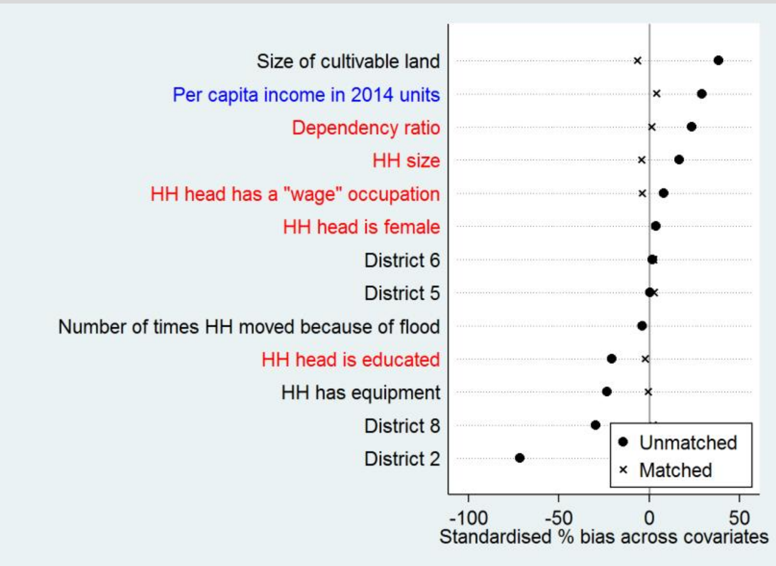
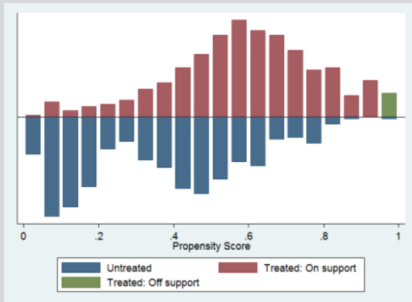
<table>
<tr><th>Variables</th><th colspan="3">Balancing</th></tr>
<tr><td colspan="4">Strategy 2</td></tr>
<tr><td rowspan="7">
</td><td colspan="3">
</td></tr>
<tr><td colspan="2">Bandwidth</td><td>2</td></tr>
<tr><td colspan="2">Trimming</td><td>5</td></tr>
<tr><td colspan="2">N on common support</td><td>1141</td></tr>
<tr><td>Rubin's B
Rubin's R</td><td>[before matching]</td><td>74.26
0.76</td></tr>
<tr><td>Rubin's B
Rubin's R</td><td>[after matching]</td><td>19.4
1.6</td></tr>
<tr><td colspan="4">Robustness Test 1*</td></tr>
<tr><td rowspan="6">
</td><td colspan="3">
</td></tr>
<tr><td colspan="2">Bandwidth</td><td>2</td></tr>
<tr><td colspan="2">Trimming</td><td>8</td></tr>
<tr><td colspan="2">N on common support</td><td>731</td></tr>
<tr><td>Rubin's B
Rubin's R</td><td>[before matching]</td><td>95.54
0.54</td></tr>
<tr><td>Rubin's B
Rubin's R</td><td>[after matching]</td><td>18.86
1.21</td></tr>
<tr><td colspan="4">Notes: outcome variable; theory-locked variables; * Baseline value of outcome variable has been used as matching variable</td></tr>
</table>

Table 30 Total income: ATT

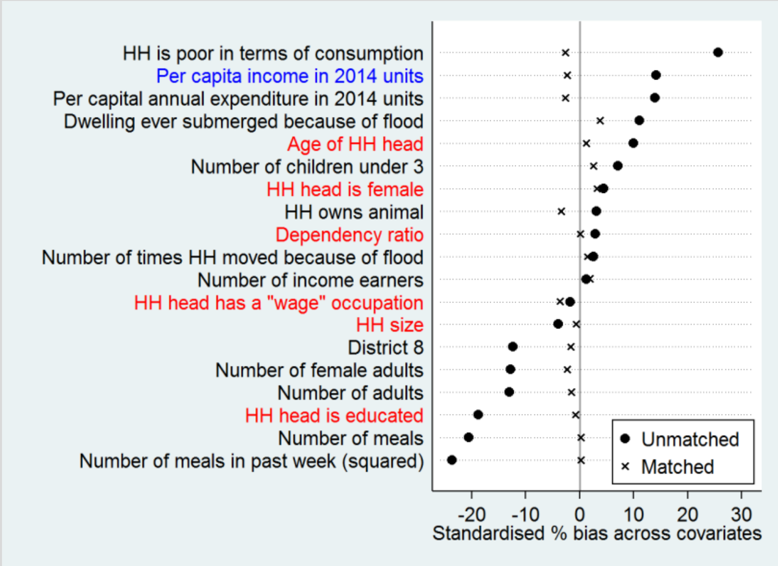
	Strategy 1a	Strategy 1b	Strategy 2	Robustness 1
ATT (Kernel) SE (bootstrapping) *SE (no bootstrapping)*	41623.86 (1585.408) (1670.878)	32674.28 1863.304 *(1967.016)*	41802.24 (1188.934) *(1280.263)*	*-27.64* (2936.407) *(2866.374)*
ATT (NN)	42990.85	32408.53	40843.87	1615.95

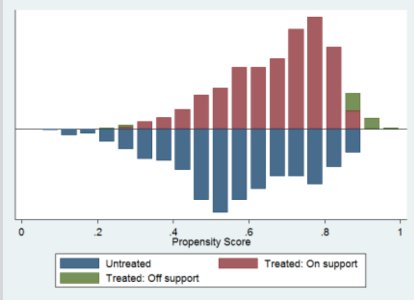
10. Per capita income in 2014 monetary units

Table 31 *Per capita* income: Selected variables and balancing across strategies

<table>
<tr><th>Variables</th><th colspan="3">Balancing</th></tr>
<tr><td colspan="4">Strategy 1a</td></tr>
<tr><td rowspan="6">
</td><td colspan="3">
</td></tr>
<tr><td colspan="2">Bandwidth</td><td>2</td></tr>
<tr><td colspan="2">Trimming</td><td>3</td></tr>
<tr><td colspan="2">N on common support</td><td>784</td></tr>
<tr><td>Rubin's B
Rubin's R</td><td>[before matching]</td><td>73.43
0.81</td></tr>
<tr><td>Rubin's B
Rubin's R</td><td>[after matching]</td><td>13.59
1.64</td></tr>
<tr><td colspan="4">Strategy 1b*</td></tr>
<tr><td rowspan="6">
</td><td colspan="3">
</td></tr>
<tr><td colspan="2">Bandwidth</td><td>2</td></tr>
<tr><td colspan="2">Trimming</td><td>3</td></tr>
<tr><td colspan="2">N on common support</td><td>829</td></tr>
<tr><td>Rubin's B
Rubin's R</td><td>[before matching]</td><td>110.95
0.85</td></tr>
<tr><td>Rubin's B
Rubin's R</td><td>[after matching]</td><td>12
0.92</td></tr>
</table>

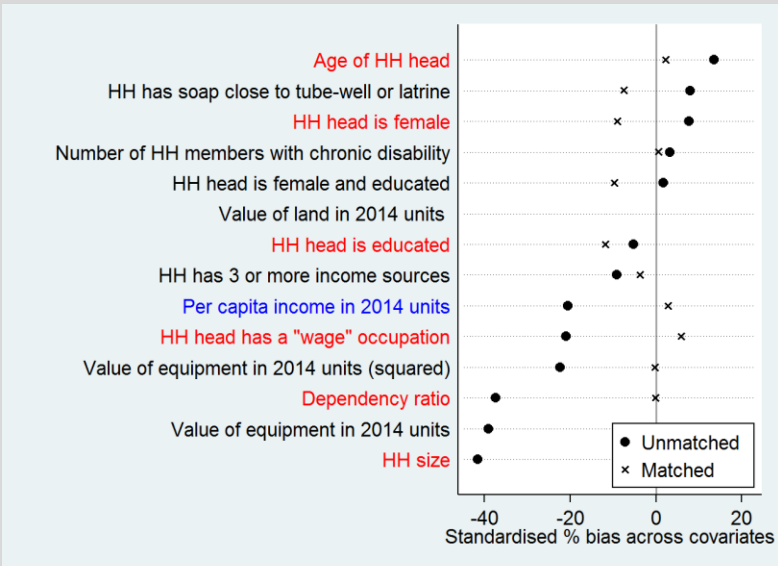
Variables	Balancing		
Strategy 2			

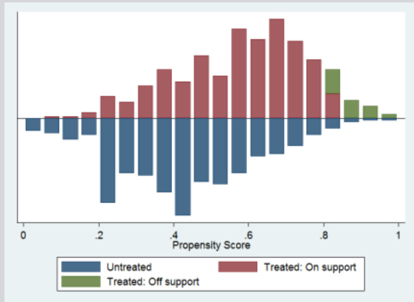




Bandwidth		2
Trimming		5
N on common support		1141
Rubin's B Rubin's R	[before matching]	71.96 0.71
Rubin's B Rubin's R	[after matching]	10.73 1.01

Robustness Test 1*





Bandwidth		2
Trimming		8
N on common support		731
Rubin's B Rubin's R	[before matching]	85.54 0.71
Rubin's B Rubin's R	[after matching]	18.66 0.85

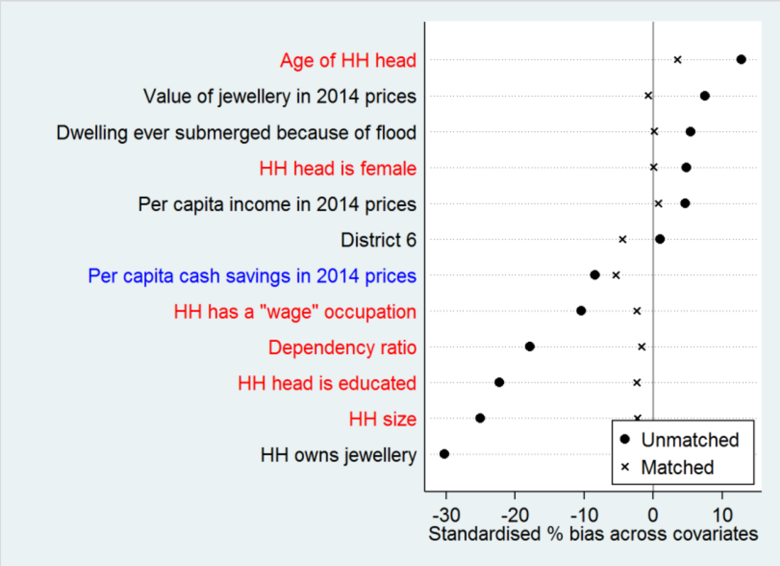
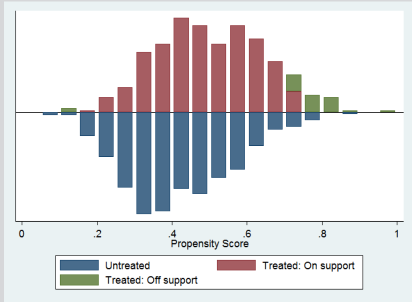
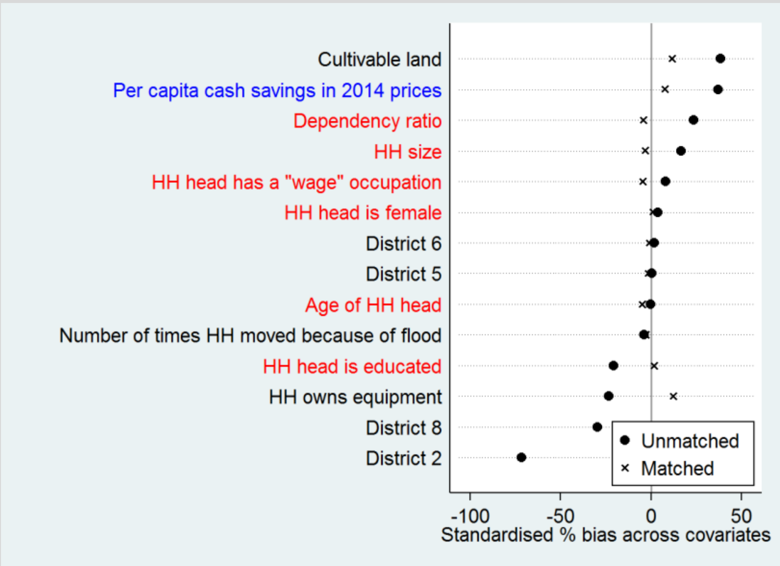
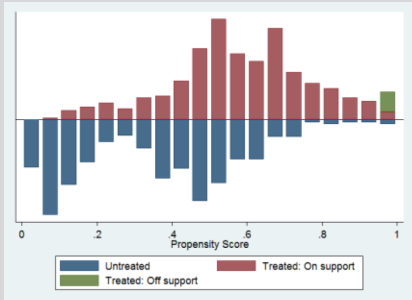
*Notes: outcome variable; theory-locked variables; * Baseline value of outcome variable has been used as matching variable*

Table 32 *Per capita* income: ATT

	Strategy 1a	Strategy 1b	Strategy 2	Robustness 1
ATT (Kernel) SE (bootstrapping) *SE (no bootstrapping)*	11,638.73 (522.155) *(528.369)*	8,547.82 (499.768) *(507.416)*	10,988.14 (352.165) *(371.338)*	253.58 (210.734) *(189.173)*
ATT (NN)	11,774.65 (537.823)	8524.61 (520.945)	10962.38 (378.466)	204.17 (219.191)

11. *Per capita* cash savings in 2014 monetary units

Table 33 *Per capita* cash savings: Selected variables and balancing across strategies

<table>
<tr><th>Variables</th><th colspan="3">Balancing</th></tr>
<tr><td colspan="4">Strategy 1a</td></tr>
<tr><td rowspan="6">
</td><td colspan="3">
</td></tr>
<tr><td colspan="2">Bandwidth</td><td>4</td></tr>
<tr><td colspan="2">Trimming</td><td>8</td></tr>
<tr><td colspan="2">N on common support</td><td>818</td></tr>
<tr><td>Rubin's B
Rubin's R</td><td>[before matching]</td><td>63.31
1.09</td></tr>
<tr><td>Rubin's B
Rubin's R</td><td>[after matching]</td><td>9.24
1.17</td></tr>
<tr><td colspan="4">Strategy 1b*</td></tr>
<tr><td rowspan="6">
</td><td colspan="3">
</td></tr>
<tr><td colspan="2">Bandwidth</td><td>4</td></tr>
<tr><td colspan="2">Trimming</td><td>3</td></tr>
<tr><td colspan="2">N on common support</td><td>829</td></tr>
<tr><td>Rubin's B
Rubin's R</td><td>[before matching]</td><td>109.53
0.81</td></tr>
<tr><td>Rubin's B
Rubin's R</td><td>[after matching]</td><td>21.53
0.93</td></tr>
</table>

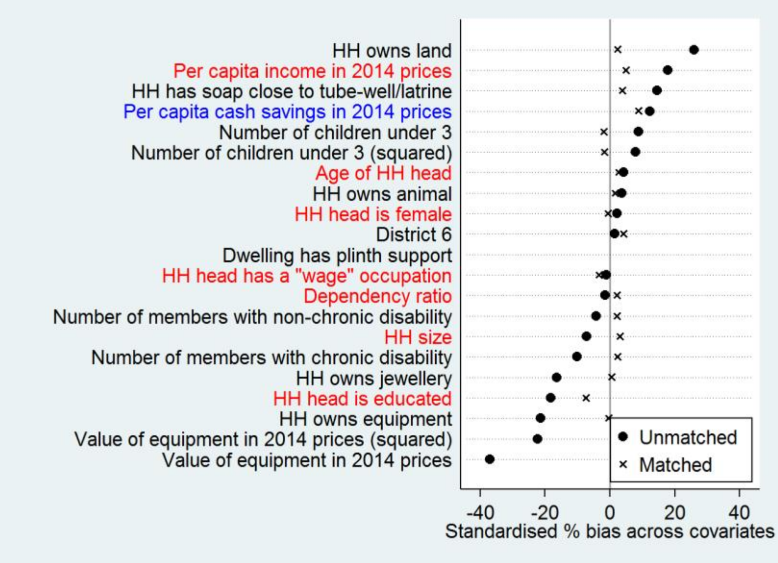
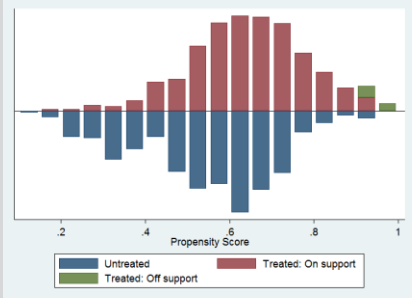
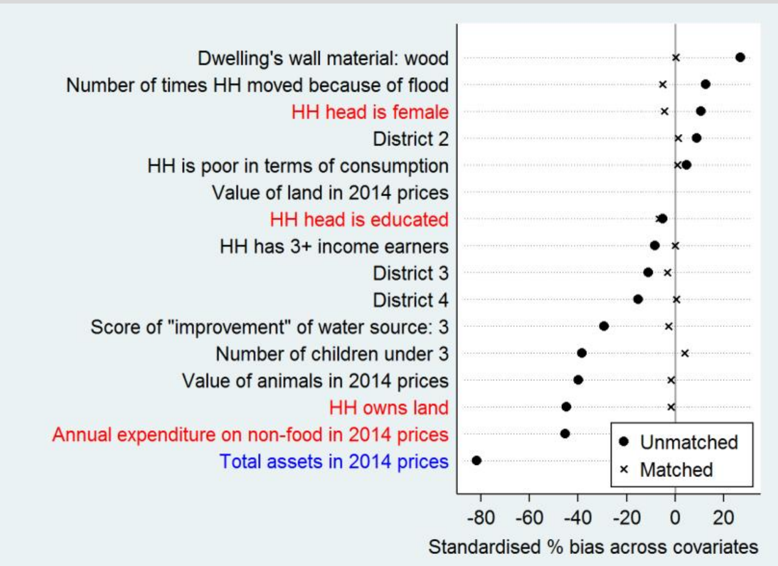
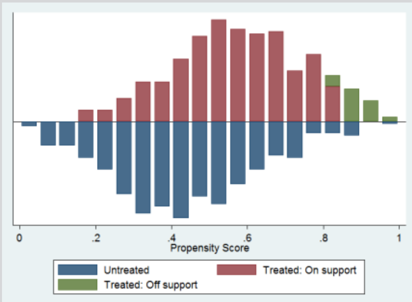
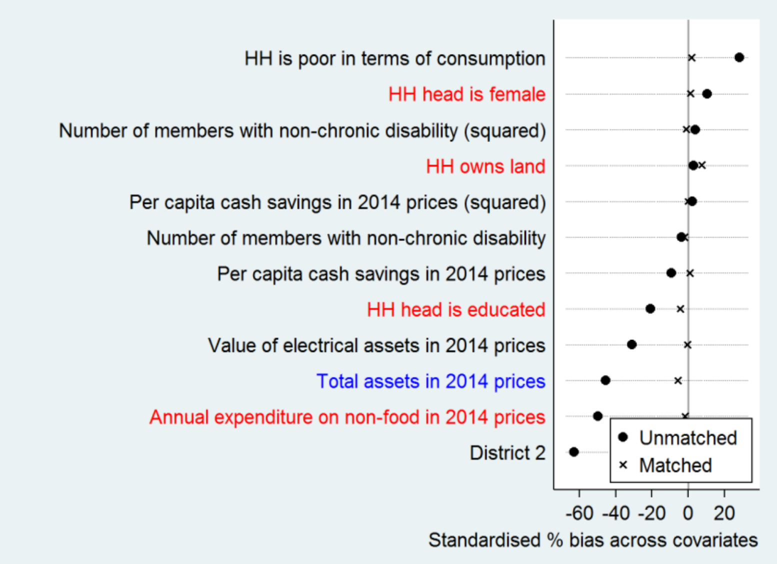
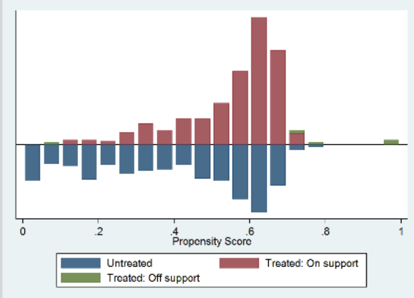
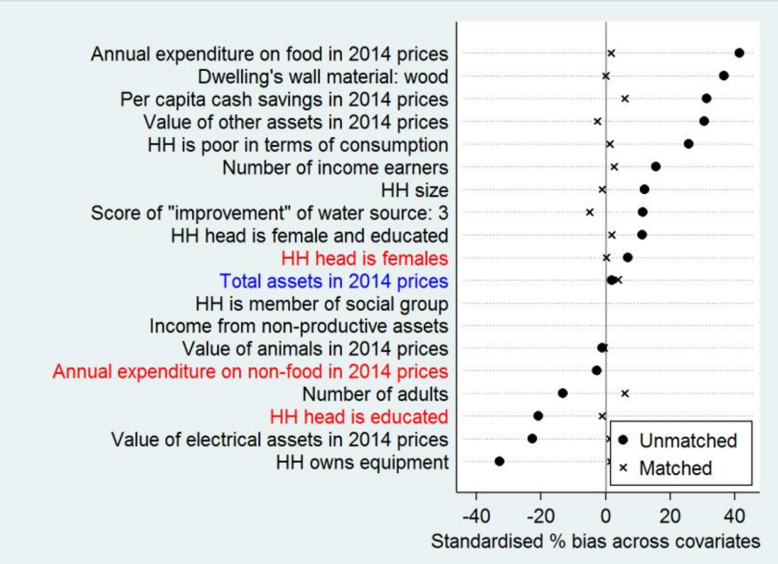
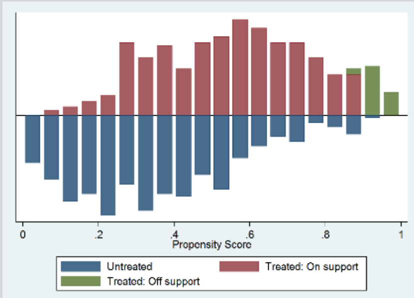
Variables	Balancing		
Strategy 2			
			
	Bandwidth		2
	Trimming		3
	N on common support		1133
	Rubin's B Rubin's R	[before matching]	69.53 1.04
	Rubin's B Rubin's R	[after matching]	19.4 1.13
Robustness Test 1**			
			
	Bandwidth		4
	Trimming		8
	N on common support		732
	Rubin's B Rubin's R	[before matching]	83.87 0.84
	Rubin's B Rubin's R	[after matching]	13.33 1.53
*Notes: outcome variable; theory-locked variables; * Baseline value of outcome variable has been used as matching variable; ** No balance has been achieved for the outcome variable*			

Table 34 *Per capita* cash savings: ATT

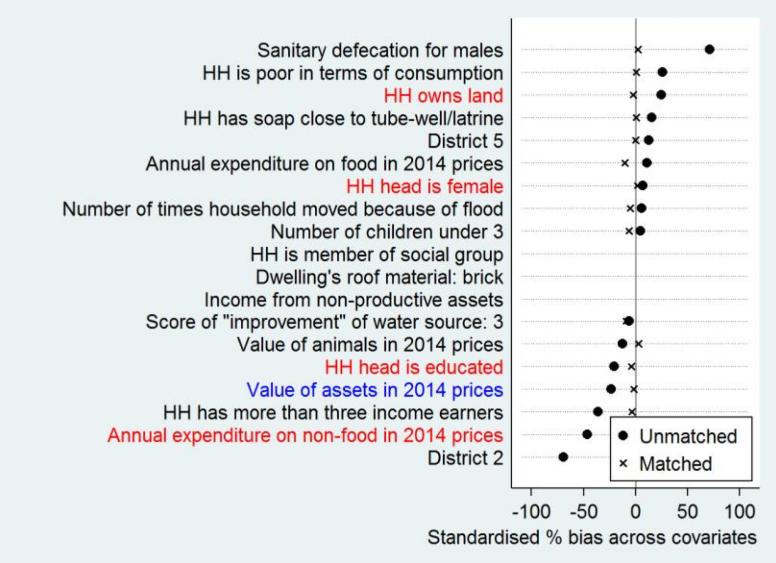
	Strategy 1a	Strategy 1b	Strategy 2	Robustness 1
ATT (Kernel) SE (bootstrapping) *SE (no bootstrapping)*	1624.48 (113.855) *(118.291)*	1485.28 (160.203) *(158.117)*	1457.7 (79.159) *(80.179)*	241.98 (179.211) *(176.837)*
ATT (NN)	1622.33	1489.08	1460.78	372.14

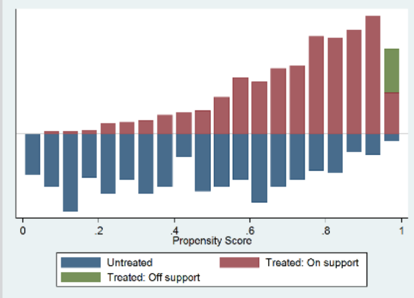
12. Total assets HH owns

Table 35 Total assets: Selected variables and balancing across strategies

<table>
<tr><th>Variables</th><th colspan="3">Balancing</th></tr>
<tr><td colspan="4">Strategy 1a</td></tr>
<tr><td rowspan="6">
</td><td colspan="3">
</td></tr>
<tr><td colspan="2">Bandwidth</td><td>4</td></tr>
<tr><td colspan="2">Trimming</td><td>3</td></tr>
<tr><td colspan="2">N on common support</td><td>795</td></tr>
<tr><td>Rubin’s B
Rubin’s R</td><td>[before matching]</td><td>56.62
0.23</td></tr>
<tr><td>Rubin’s B
Rubin’s R</td><td>[after matching]</td><td>17.33
1.05</td></tr>
<tr><td colspan="4">Strategy 1b</td></tr>
<tr><td rowspan="6">
</td><td colspan="3">
</td></tr>
<tr><td colspan="2">Bandwidth</td><td>2</td></tr>
<tr><td colspan="2">Trimming</td><td>8</td></tr>
<tr><td colspan="2">N on common support</td><td>733</td></tr>
<tr><td>Rubin’s B
Rubin’s R</td><td>[before matching]</td><td>106
0.79</td></tr>
<tr><td>Rubin’s B
Rubin’s R</td><td>[after matching]</td><td>22.55
1.15</td></tr>
</table>

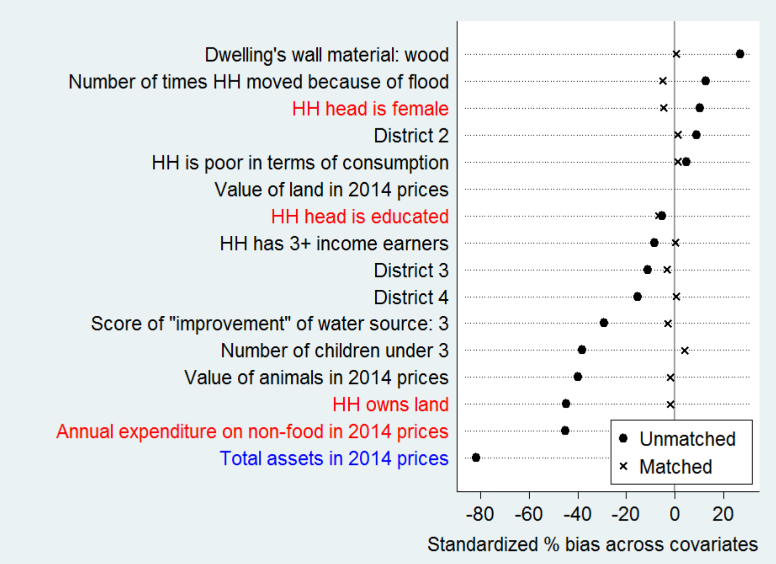
Variables	Balancing		
Strategy 2			

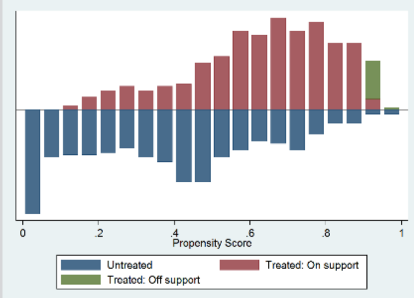




Bandwidth		3
Trimming		5
N on common support		1011
Rubin's B Rubin's R	[before matching]	127.22 0.68
Rubin's B Rubin's R	[after matching]	21.95 1.04

Robustness Test 1





Bandwidth		4
Trimming		5
N on common support		731
Rubin's B Rubin's R	[before matching]	111.57 0.31
Rubin's B Rubin's R	[after matching]	11.41 1.07

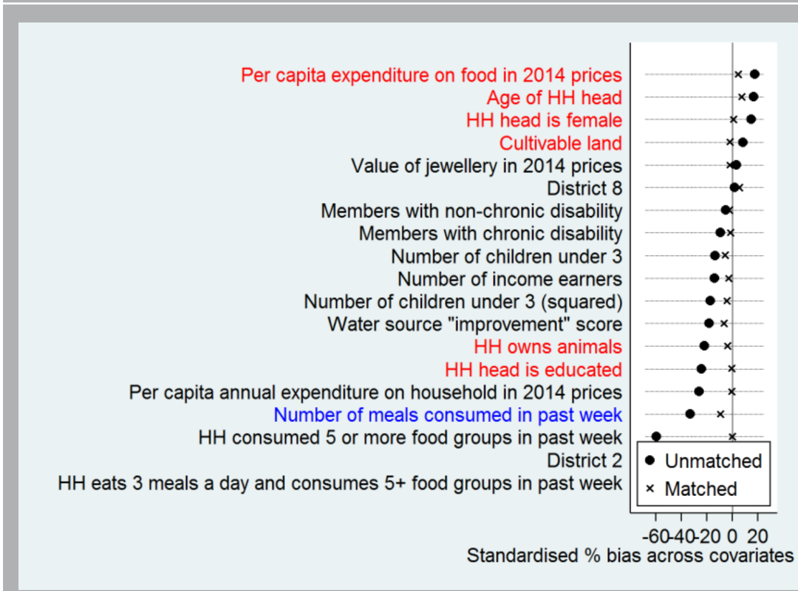
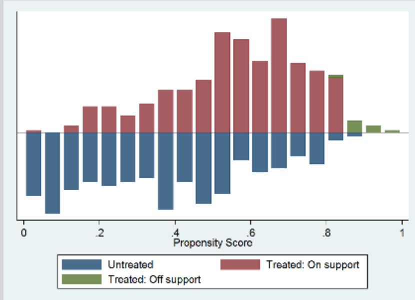
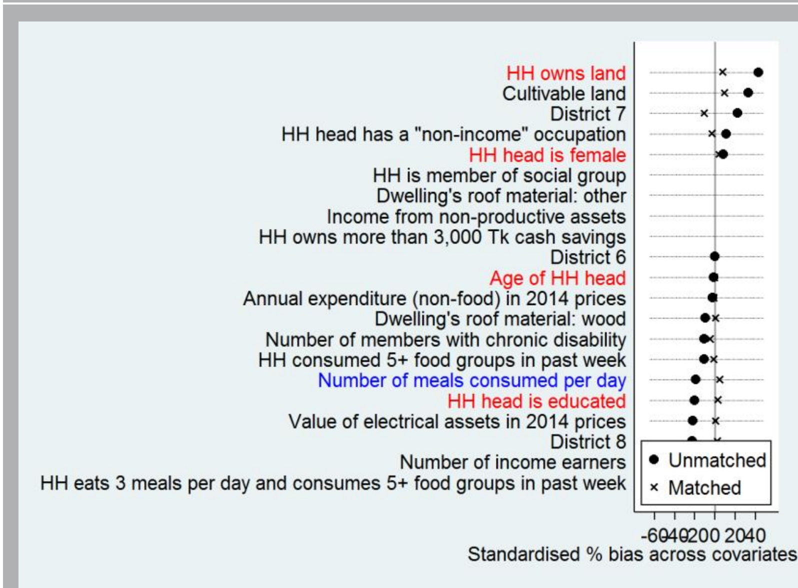
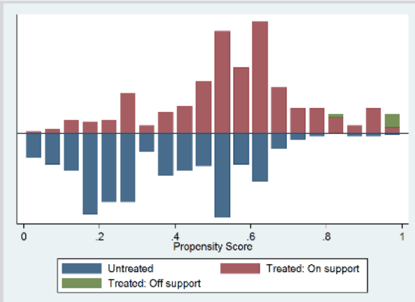
Notes: *outcome variable; theory-locked variables*

Table 36 Total assets: ATT

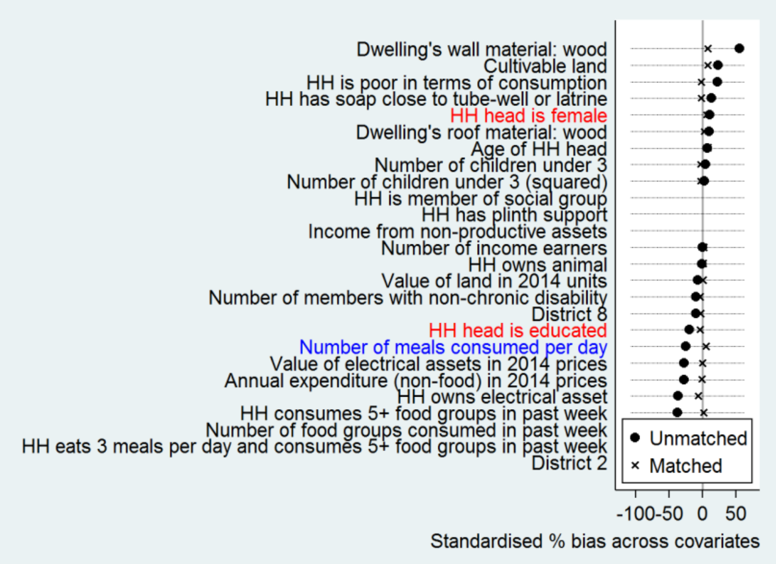
	Strategy 1a	Strategy 1b	Strategy 2	Robustness 1
ATT (Kernel) SE (bootstrapping) *SE (no bootstrapping)*	58,605.12 (3,107.592) *(3,011.586)*	39,860.3 (1,220.592) (1,323.116)	62,812.75 (2,093.979) *(2,074.327)*	-1,803.82 (4,057.374) (4,224.789)
ATT (NN)	58,834.27 (3,005.884)	39,633.29 (1,276.81)	63,021.07 (2,051.387)	-1,500.1 (5,253.02)

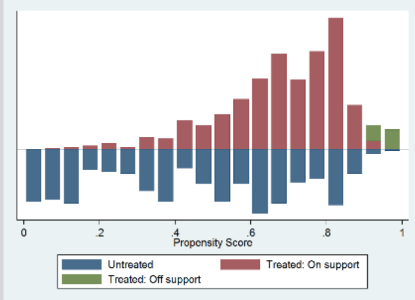
13. Number of meals consumed in past week

Table 37 Meals: Selected variables and balancing across strategies

<table>
<tr><th>Variables</th><th colspan="3">Balancing</th></tr>
<tr><td colspan="4">Strategy 1a</td></tr>
<tr><td rowspan="6">
</td><td colspan="3">
</td></tr>
<tr><td colspan="2">Bandwidth</td><td>4</td></tr>
<tr><td colspan="2">Trimming</td><td>3</td></tr>
<tr><td colspan="2">N on common support</td><td>759</td></tr>
<tr><td>Rubin's B
Rubin's R</td><td>[before matching]</td><td>104.68
0.49</td></tr>
<tr><td>Rubin's B
Rubin's R</td><td>[after matching]</td><td>19.16
1.03</td></tr>
<tr><td colspan="4">Strategy 1b</td></tr>
<tr><td rowspan="6">
</td><td colspan="3">
</td></tr>
<tr><td colspan="2">Bandwidth</td><td>3</td></tr>
<tr><td colspan="2">Trimming</td><td>3</td></tr>
<tr><td colspan="2">N on common support</td><td>733</td></tr>
<tr><td>Rubin's B
Rubin's R</td><td>[before matching]</td><td>100.15
0.86</td></tr>
<tr><td>Rubin's B
Rubin's R</td><td>[after matching]</td><td>16.24
1.01</td></tr>
</table>

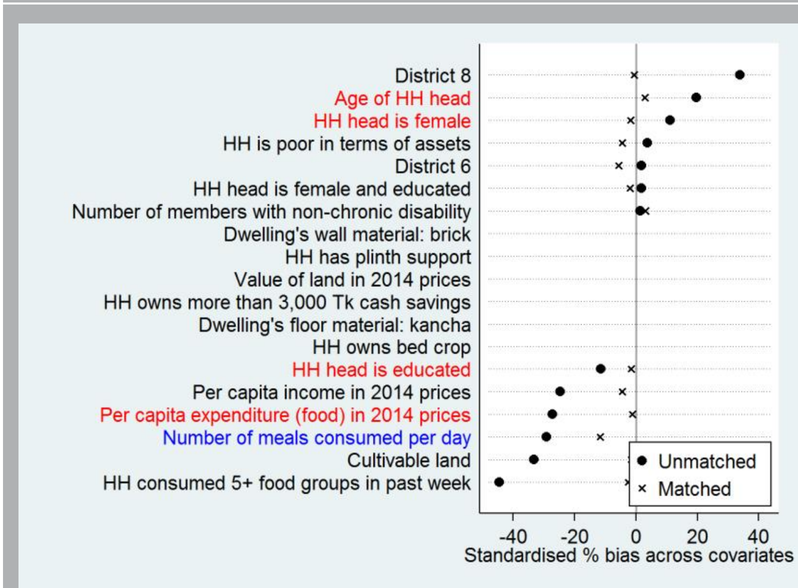
Variables	Balancing
Strategy 2	

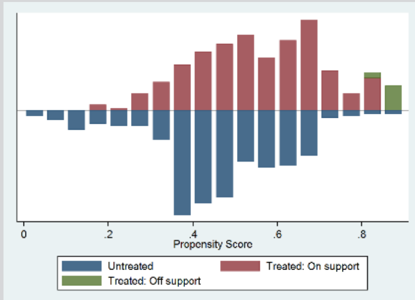




Bandwidth		9
Trimming		5
N on common support		1014
Rubin's B Rubin's R	[before matching]	106.26 0.34
Rubin's B Rubin's R	[after matching]	22.58 1.06

Robustness Test 1





Bandwidth		2
Trimming		5
N on common support		635
Rubin's B Rubin's R	[before matching]	71.1 0.88
Rubin's B Rubin's R	[after matching]	16.69 1.19

Notes: outcome variable; theory-locked variables

Table 38 Meals: ATT

	Strategy 1a	Strategy 1b	Strategy 2	Robustness 1
ATT (Kernel) SE (bootstrapping) *SE (no bootstrapping)*	2.21 (.193) *(.188)*	2.05 (.206) *(.213)*	2.56 (.168) *(.172)*	-0.41 (.066) *(.072)*
ATT (NN)	2.18 (.278)	2.27 (.284)	2.40 (.242)	-0.44 (.075)

Malnutrition outcome variables

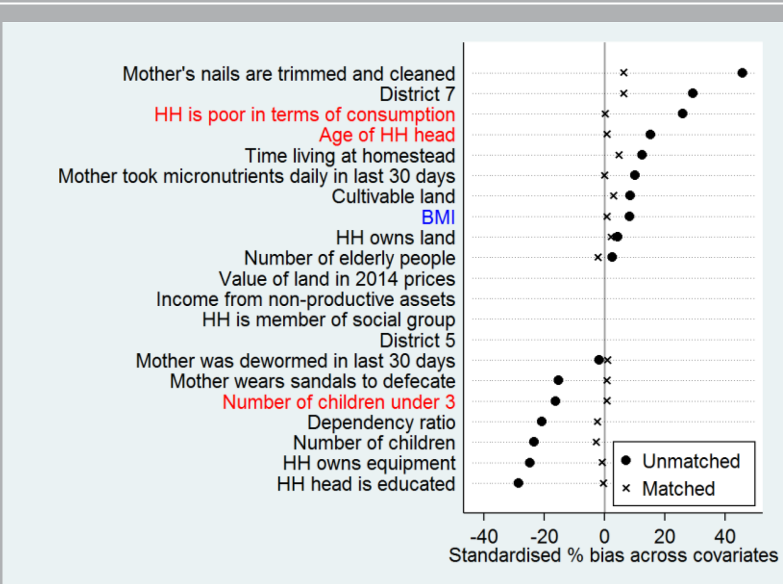
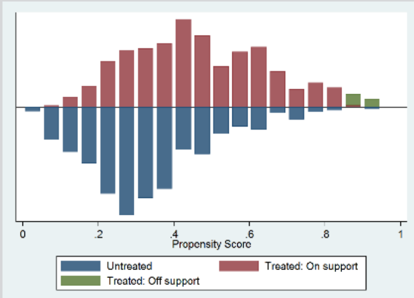
In the case of the malnutrition outcome variables (i.e. BMI, HFA, WFA and WFL), we only implemented Strategy 1a. This was due to the fact that nutrition-related data were only collected for survey years 2012 and 2014.

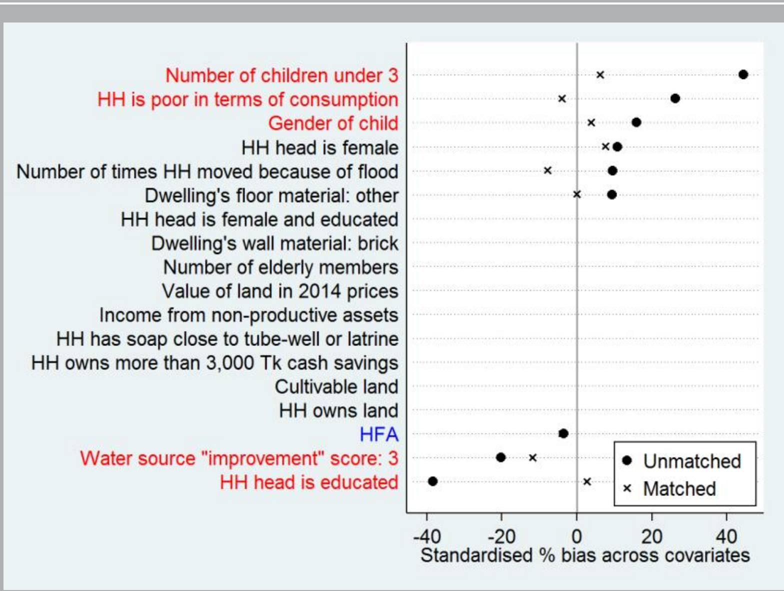
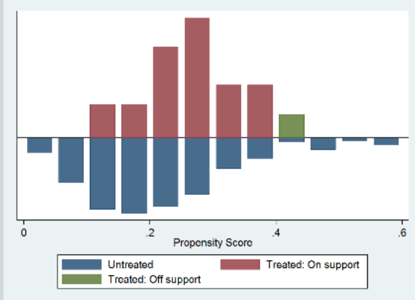
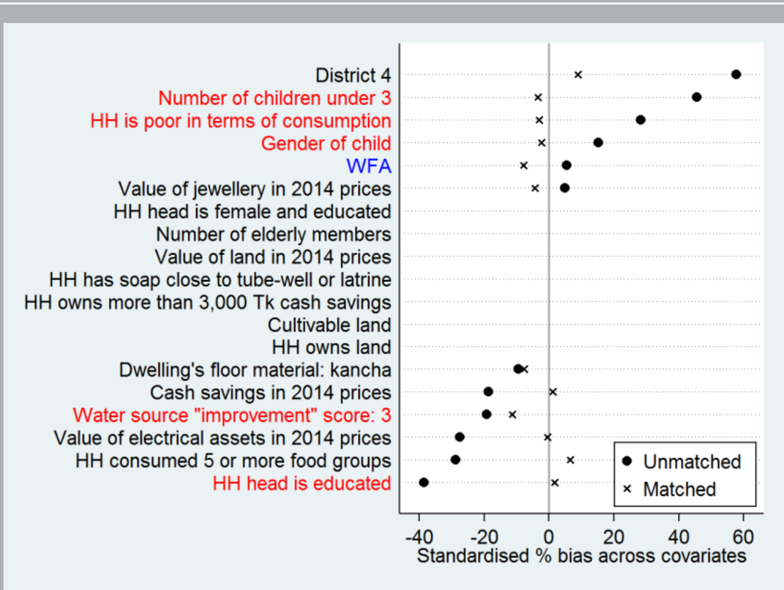
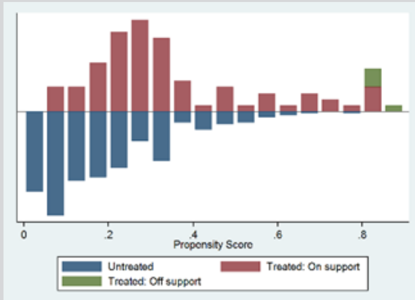
The results and graphs in Table 39 below show that the matching process removes the imbalance for almost all covariates in the case of all outcome variables. After matching, standardised bias across covariates is considerably reduced, Rubin's B lies under 25, and the distribution of propensity scores between treatment and control groups is highly comparable. Furthermore, balance is also achieved for the outcome variable of interest, pointing to the fact that the first stage matching procedure is performing well.

With regards to the estimation of treatment effects, results in Table 40 show a positive ATT on BMI of 0.49 for cohort 2.4. In the case of HFA, WFA and WFL, the estimated treatment effects are -0.23, -0.29, and -0.05, respectively. Out of these four outcome variables, only the ATT on BMI is borderline significant, with the others being insignificant.

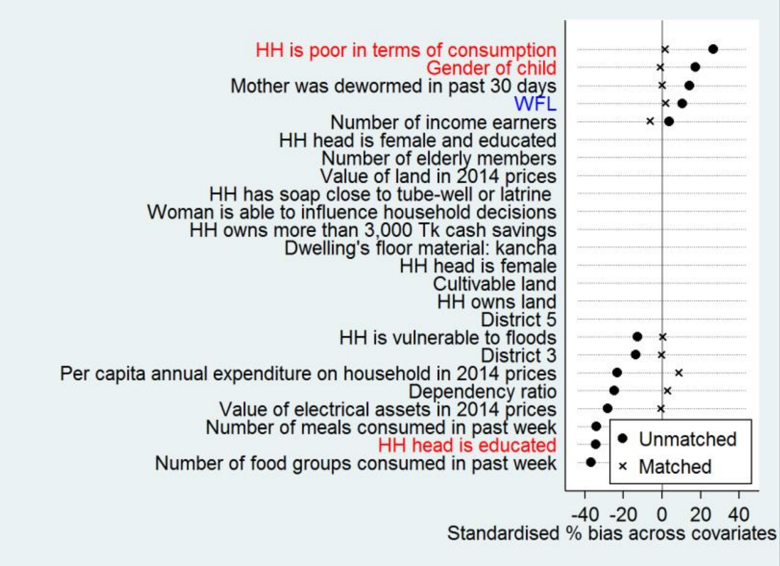
Given the impossibility of implementing strategy 1b or strategy 2, or indeed robustness test 1, no statement can be made with regards to the effect that cohort- or year-specific effects may have in our estimations. The magnitude of all our Kernel estimates seem to be confirmed by applying a NN algorithm in the second stage analysis.

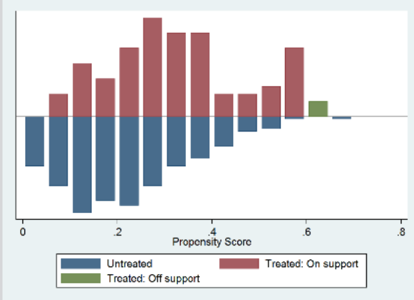
Table 39 Malnutrition outcomes: Selected variables and balancing under Strategy 1a

<table>
<tr><th>Variables</th><th colspan="3">Balancing</th></tr>
<tr><td colspan="4">BMI</td></tr>
<tr><td rowspan="6">
</td><td colspan="3">
</td></tr>
<tr><td colspan="2">Bandwidth</td><td>6</td></tr>
<tr><td colspan="2">Trimming</td><td>3</td></tr>
<tr><td colspan="2">N on common support</td><td>1091</td></tr>
<tr><td>Rubin's B
Rubin's R</td><td>[before matching]</td><td>82.42
1.17</td></tr>
<tr><td>Rubin's B
Rubin's R</td><td>[after matching]</td><td>11.75
1.14</td></tr>
<tr><td colspan="4">Notes: outcome variable; theory-locked variables</td></tr>
</table>

<table>
<tr><th>Variables</th><th colspan="3">Balancing</th></tr>
<tr><td colspan="4">HFA</td></tr>
<tr><td rowspan="6">
</td><td colspan="3">
</td></tr>
<tr><td colspan="2">Bandwidth</td><td>4</td></tr>
<tr><td colspan="2">Trimming</td><td>8</td></tr>
<tr><td colspan="2">N on common support</td><td>361</td></tr>
<tr><td>Rubin's B
Rubin's R</td><td>[before matching]</td><td>68.93
0.36</td></tr>
<tr><td>Rubin's B
Rubin's R</td><td>[after matching]</td><td>23.14
0.96</td></tr>
<tr><td colspan="4">WFA</td></tr>
<tr><td rowspan="6">
</td><td colspan="3">
</td></tr>
<tr><td colspan="2">Bandwidth</td><td>4</td></tr>
<tr><td colspan="2">Trimming</td><td>5</td></tr>
<tr><td colspan="2">N on common support</td><td>366</td></tr>
<tr><td>Rubin's B
Rubin's R</td><td>[before matching]</td><td>61.4
0.09</td></tr>
<tr><td>Rubin's B
Rubin's R</td><td>[after matching]</td><td>23.8
1.08</td></tr>
<tr><td colspan="4">Notes: outcome variable; theory-locked variables</td></tr>
</table>

Variables	Balancing
WFL	





Bandwidth		2
Trimming		3
N on common support		325
Rubin's B Rubin's R	[before matching]	48.94 0.05
Rubin's B Rubin's R	[after matching]	19.1 1.08

Notes: outcome variable; theory-locked variables

Table 40 Malnutrition outcomes: ATT

	Strategy 1a - BMI	Strategy 1a - HFA	Strategy 1a - WFA	Strategy 1a - WFL
ATT (Kernel) SE (bootstrapping) *SE (no bootstrapping)*	0.49 (.226) *(.253)*	-0.23 (.174) *(.161)*	-0.29 (.136) *(.158)*	-0.05 (.153) *(.146)*
ATT (NN)	-0.21 (.361)	-0.31 (.308)	-0.38 (.239)	0 (.179)

6.5 Panel regressions

As mentioned in the preceding section, one of the key limitations when using PSM is that unobserved confounding factors that could introduce bias in estimating treatment effects cannot be controlled for as the matching procedure is based on observed covariates only. In our case, this would mean that there are unobservable characteristics that differ between households in a cohort at baseline and cohorts at follow-up (other than the treatment status) that are correlated with the outcome measure and cannot be controlled for by employing PSM.

A suitable approach to dealing with such unobservable differences would be to use a difference-in-differences estimator, which would also be applicable in a PSM context. However, a standard difference-in-differences panel analysis cannot be adopted here. This is because, in order to implement such an estimation, we would need two non-treatment rounds of data collection for any cohort to be able to compare that to households in a cohort that has both a non-treatment and a treatment round of data. Unfortunately, given that the analysis population does not contain pure controls, we are not able to do this.

As an alternative, pooling data across cohorts and implementing a regression analysis that exploits the panel structure of this data can provide us with the opportunity to **deal with the unobservable characteristics of households and, in addition, to understand the pattern and sustainability of the intervention's impact across time.**

In addition to the PSM analysis, we therefore implemented a fixed effects panel estimation to cover these analytical gaps and also as a further check on the plausibility of the effects identified using PSM. We consider panel regressions to be important additions to this analysis, which can further be used to triangulate findings. Note, however, that due to differences in the methodology, data and covariates used compared to the PSM strategies, we do not expect our estimates to be exactly the same across the two methodologies. As emphasised above, the results produced across the different methods should be seen as complementary rather than as substitutes.

It is important to emphasise that we also implemented the panel analysis as a means to disentangle the effects of the programme over time and assess the sustainability of the programme. To this end, we were able to estimate treatment effects one period after treatment (corresponding to strategy 1b in the PSM estimation), two periods after treatment (corresponding to strategy 2 in the PSM estimation) and three or more periods after treatment (which does not have a corresponding PSM estimate).

Finally, the panel estimation also enabled us to determine whether there were differential treatment effects for certain outcomes. For example, we were able to determine whether a household's propensity to graduate the programme differed based on whether the household head was female or not. This allowed us to understand, in a more nuanced way, the determinants of programme success and it may be helpful in the future to better target interventions.

Note that, as a final step, we compared the panel results to the ATT estimated using PSM, the results of which are shown and discussed in Volume I (OPM 2016a).

6.5.1 Methodology

In general, panel methods can be applied to data in which observations exist for the same units of analysis over several time periods. Given that we use annual CLP data from 2012 to 2015, we have multiple observations for households in cohorts 2.1 to 2.6, although the number of time periods for which data exist varies by cohort. As mentioned above, the benefit of having several observations for each unit is that, by appropriate modification of the data, it is possible to eliminate some unobserved biases that might affect treatment estimates. For instance, if we assume that households in some cohorts have idiosyncratic but time-invariant features that are correlated with an outcome measure, hence introducing bias, employing First-Differencing or Fixed Effects estimations could address this bias.

To exemplify how this worked in our analytical setting, consider the following equation as the base specification for our panel approach:

$$y_{it} = \alpha_0 + \alpha_1 T_{1it} + \alpha_2 T_{2it} + \alpha_3 T_{3it} + \boldsymbol{\beta}\mathbf{X_{it}} + \gamma_t \mathbf{D_t} + c_i + u_{it}. \quad (2)$$

Here, y_{it} represents the outcome variable for household i in year t. We also include a set of dummy variables to capture the treatment effect of the intervention over time. Omitting the dummy variable at baseline (T_{0it}), we include T_{1it} which indicates if an observation of household *i* is in the first period after treatment, T_{2it} which indicates the second period after the intervention has been implemented and finally T_{3it} for the third time period and all subsequent time periods, after treatment. [14] $\mathbf{X_{it}}$ is a vector of the covariates for each household and year that differs for each outcome variable. $\mathbf{D_t}$ is a vector of year dummies including dummies for 2012, 2013 and 2014, and c_i represents the household fixed effect. Finally, u_{it} is the error term. Running this regression in a fixed effects framework controls for year- and household-specific unobservable fixed effects, i.e. time-invariant household-specific effects and year-specific effects.[15]

We are interested in the estimates on the time period dummies: the estimated coefficient on the treatment dummy, $\widehat{\alpha_1}$, can be considered as an estimate of the treatment effect in time period 1, i.e. T1. $\widehat{\alpha_2}$ is an estimate of the treatment effect after two time periods, i.e. T2, and $\widehat{\alpha_3}$ after three or more time periods (i.e. in all time periods after the core period of intervention), i.e. T3+. These are also the estimates presented in Volume I.

The strict exogeneity assumption

The key assumption for this approach to be valid and the estimate on the treatment coefficients to actually be unbiased is the **strict exogeneity assumption**. This assumption posits that, in the specification above, the error term u_{it} has conditional mean zero and is uncorrelated in each time period with the regressors, i.e. with all the explanatory variables in every time period included in the right-hand side of the above equation. This is called the strict exogeneity assumption because it implies that all these explanatory variables are exogenous and therefore determined outside of the estimation model.

In order to test that this assumption holds, we follow Wooldridge (2002) and run several estimation checks. First, we run the same specification as in equation (2) but include lead values for each covariate in $\mathbf{X_{it}}$. The inclusion of the lead regressors allows us to determine whether the assumption of strict exogeneity is satisfied and hence whether fixed effects is unbiased. If strict exogeneity holds, the estimated coefficient on each lead regressor will not be statistically significant, thus implying that the regressors in leading time periods are not correlated with the error term. We ran this specification for each outcome variable to test the exogeneity assumption in all of our panel settings and do not find statistically significant estimated coefficients on the lead variables.[16] This provides robust evidence that the strict exogeneity assumption holds.

In addition, we can compare estimates from a fixed effects estimation to an estimation using first differences, in that large differences in the estimated coefficients would indicate that exogeneity might be violated (*Ibid*., p. 284 ff.). Usually, we could compare these estimates when we have more than one time period available and we would expect the difference in estimates to be solely attributable to sampling errors. When estimates show large differences across estimation methodologies, however, we should be concerned about the strict exogeneity assumption. Note that both fixed effects and first-differencing would be biased in such a case. We ran a first

[14] The first time period after the intervention has begun is a measure of impact while the household is still participating in the programme. The second time period after the intervention is a measure of impact after programme participation has been completed by the household.

[15] See Wooldridge (2002, p. 247 ff.) for a technical discussion of panel methods.

[16] These results are available upon request.

difference estimation for each outcome variable (results available upon request), found no large differences, and therefore again did not find evidence that would lead us to be concerned about the failing of the strict exogeneity assumption.

Fixed effects versus random effects

To further justify our choice of fixed effects over a random effects estimation, we employ the Hausman (1978) test (Wooldridge 2002, p. 288 ff.). Employing the Hausman test allows us to assess whether there are indeed household-level time-invariant effects that are correlated with other variables in the model – and hence that a fixed effects estimation procedure is valid. The alternative would be to implement a random effects estimation procedure where household-level idiosyncratic features are not assumed to be correlated with other regressors. In such a situation, using fixed effects is still unbiased but is a less efficient estimator than a random effects estimation.

As mentioned above, the key difference between the two approaches is that fixed effects assumes that time-invariant unobservable household characteristics are correlated with the regressors and therefore bias our estimates while the random effects approach assumes that there is no such correlation. The estimated degree of correlation between the unobserved characteristics of our observations and the regressors in our model informs the choice between the two methodologies. We employed the Hausman test in the panel regression for each outcome and in all cases are able to reject the null hypothesis that unobservable time-invariant household characteristics are uncorrelated with the regressors. This therefore supports our choice of fixed effects, which is thus deemed the more efficient estimator in this context.

Weights

In order to ensure representativeness of our coefficient estimates, we run all fixed effects estimations using sampling weights. However, it is important to note here that such estimations cannot be implemented using weights that vary year by year. We solve this by applying weights constructed for the year 2015 to all other years in the sample. Note that we run the same analyses without weights as well and present the results below. Section 6.2.2 outlined how we constructed weights in this analysis.

Attrition

Finally, attrition of households across the years, which was discussed in Section 6.2.3, was found to be a problem in the current panel analysis. Although attrition was low and largely random between most survey rounds, we found significantly higher incidences of attrition between 2014 and 2015. Such attrition is not necessarily a problem *per se* – if dropping out happens randomly and does not happen with very large incidence – as this would mean that the remaining sample of households is as representative of each cohort as the baseline sample.[17] However, as discussed above, we found that this attrition does indeed tend to affect a specific type of household in our data.

Following one approach outlined in the literature on econometric techniques to control for attrition, we therefore use inverse probability weights for each household in one specification to test how

[17] If a very large proportion of the sample drops out, the decrease in the power to detect programme impact could be problematic.

robust our results are to this correction. Inverse probability weights weight households according to their respective probabilities of dropping out of the sample in t+1 (Wooldridge 2002, p.587 ff.). In our case, these weights are calculated in several steps. We first create a dummy variable indicating whether a household drops out of the sample in the next time period (t+1). We then run a probit regression with this dummy variable as the outcome variable and use a set of regressors, or observable household characteristics, thought to be correlated with the probability of dropping out. These regressors were chosen based on theoretical priors as well as using forward and backward selection algorithms to determine which regressors were most significantly correlated with attrition. We use the estimated regression coefficients to predict each households probability to drop out of the sample.

This predicted probability of household attrition obtained from the probit regression becomes the weight for each household. Finally, we calculate the inverse of this weight to obtain our inverse probability weights. Using inverse probability weights ensures that households that have a higher probability of dropping out of the sample are given more weight in the regression in order to compensate for underrepresentation due to attrition.

Note that we use pooled Ordinary Least Squares (OLS) with year, district and cohort dummy variables to implement this because fixed effects panel regressions cannot be implemented with varying weights across years. Our inverse probability weights are multiplied by the 2015 sampling weights in order to ensure representativeness. As said before, we do this in order to assess whether correction for attrition in this way changes our results.

These results are shown in column (5) for each panel regression in the tables below. Although the point estimates differ from our fixed effects estimates, the confidence intervals overlap in the large majority of cases. We therefore conclude that we should not be concerned about significant biases introduced by attrition in our panel estimations.

As a final robustness check with respect to attrition, we also run all our analyses using fixed effects estimations but excluding the latest data collection year – 2015. This is because we observe the largest attrition between 2014 and 2015. If our concern that some form of systematic attrition biases our results was correct, we would see changes in the point estimates when excluding 2015 data. However, when running this analysis, we do not observe large differences in our estimates. Again, our conclusion is that we should not be worried about large biases introduced by attrition into our panel estimations. We do not include the results of this analysis here, but as before these results are available upon request.

6.5.2 Building regression models

In order to build our panel regression models for each specification, we begin by assigning each of the variables into analytical categories from which appropriate covariates are selected to ensure that the models are properly specified. We include generic analytical categories that are common to the specification for each outcome variable, such as categories of variables relating to household head characteristics, household demographics, household education and poverty indicators, as well as analytical categories that are outcome-specific. Note that these analytical categories were defined based on theoretical priors about which groups of variables are likely to be related to the outcome.

This systematic approach to building each panel specification ensured that variables from each of the relevant analytical categories were included in each specification in order to achieve exogeneity. This condition for unbiased estimation is violated if important analytical categories that determine the outcome and are correlated with the regressors remain excluded from the analysis and are hence included in the error term. Therefore, for each outcome variable of interest, our full panel specification includes a set of covariates derived from the analytical categories. We also include a variable that controls for re-sweep activities by CLP-2, given that those were likely to be correlated with the outcome but unrelated to the initial treatment. We also included a factor variable indicating the number of periods post-baseline, which would give us our main treatment estimates. Finally, we included year dummy variables for the years 2012, 2013 and 2014. We also use robust standard errors in the estimation to ensure that we do not have serial correlation in the error terms.

As an example, the regression for the consumption poverty line outcome variable was constructed by including variables from the following generic analytical categories: household head characteristics, household demographics, household education and household economic welfare (in this case, we include the number of income earners per household). In order to ensure that the regression is properly specified, we also include variables from the following analytical categories: household income, household social status, household savings, household assets, housing quality and flood vulnerability. Although variables such as the gender of the household head and education of the household head may usually be considered time-invariant, and hence fall away in a fixed effects framework, we are able to control for these variables as, for some observations, they do indeed change over time. However, should they drop out of the regression, this will not bias our results or affect our interpretation as they are not the key variables of interest for impact evaluation.

In order to conduct the panel analysis, we use data from CLP annual surveys in 2012, 2013, 2014 and 2015 (as discussed in more detail in Section 2.3.1.1) and omit data collected in 2010 and 2011. We exclude these earlier data due to the change in survey methodology in subsequent years; this change in survey methodology affects our ability to compare data across different years of data collection and, in some instances, some indicators cannot be constructed using the 2010 and 2011 data (e.g. certain components of the graduation indicator).

6.5.3 Results and robustness checks

The panel regression results in Table 41 to Table 49 show the impact of the CLP intervention on different outcomes of interest. For each outcome, there are five columns of output. In each case, column (1) shows the output from a simple fixed effects regression that does not include any covariates. This simple regression shows, in all cases, that there is a statistically significant impact of treatment on the outcome of interest in all time periods after the intervention was implemented.

Column (2) shows the output from each fixed effects regression including the set of regressors specified for each outcome but excluding year fixed effects. Again, for each outcome variable, the treatment effect is robust to the inclusion of a full set of controls and in all cases remains significant at either the 5% or 1% level. Column (3) includes year fixed effects in the fixed effects framework in order to control for effects related to the year of the intervention rather than the intervention itself. For example, as CLP staff become more experienced in terms of implementing the programme, we may see more beneficial treatment effects for which we wish to control. For each outcome, the treatment effects remain robust to the inclusion of year fixed effects and the significance level of treatment effects across time remains at either the 1% level or 5% level for all outcomes. Note, however, that estimates do differ when including year effects, which means that, generally speaking, outcomes across cohorts are affected by such year effects. This is important – it means that when looking at simple descriptive statistics, trends that we observed are potentially due to such year effects.

For all outputs in columns (1) to (3) the regressions are neither weighted using the sample weights nor the inverse probability weights and hence do not account for attrition. In column (4), we present the results for the full fixed effects regression (i.e. including a full set of controls and year dummy variables) and weight the analysis using the 2015 sample weights. These results are therefore representative of the population of CLP recipients rather than only the sample covered in the annual CLP surveys. **These are the results that are presented in Volume I of this evaluation report (i.e. OPM 2016a).**

Our results do not change significantly when we include the sample weights and remain within the confidence intervals of the unweighted analysis. Finally, in column (5) we control for attrition by running a pooled OLS regression and including the full set of controls, year dummy variables, cohort dummy variables and district dummy variables and use the inverse probability weights multiplied by the 2015 sample weights.

As mentioned earlier, poorer households are more likely to drop out of the sample and we therefore expect those results where we do not control for attrition to be biased upwards. In general, we do not find large differences between the results in column (4) and column (5) but the treatment effects in column (5) are attenuated as expected. **These results show that, despite non-random attrition, our results remain robust and our conclusions regarding the impact of CLP on various outcomes hold even in the face of attrition.**

As an example, looking at the consumption poverty outcome indicator, the results of which are shown in Table 41, we see a significant reduction in the probability of a household falling below the poverty line across all five specifications. The treatment effects for all specifications and for all periods of time are significant at the 0.1% level, indicating that the CLP intervention does reduce the incidence of consumption poverty amongst those taking part in the programme. The inclusion of controls and year fixed effects in specifications (2) and (3) attenuates the estimated treatment effect in all time periods, but the positive effect of the programme on poverty remains.

In terms of the impact of the programme over time, the treatment effects are larger in time period 2 and largest in time period 3, indicating that the treatment effect is larger the longer a household is part of the programme and also that it is sustainable over time. Overall, the treatment effects from the weighted regression in column (4) indicate that a household's probability of falling below the consumption poverty line is 26.9% lower after participating in the programme for one year, 33.3% lower six months after programme participation has ended and 36.6% lower, on average, in all subsequent time periods.

Table 41 Panel regression: Consumption poverty

	Consumption poverty using lower bound					Imputed consumption *per capita* in 2014 prices				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
T=1	-0.372***	-0.299***	-0.262***	-0.269***	-0.230***	367.2***	291.1***	275.1***	282.5***	250.1***
	(0.0152)	(0.0217)	(0.0228)	(0.0235)	(0.0237)	(12.05)	(16.55)	(17.20)	(17.34)	(16.97)
T=2	-0.503***	-0.428***	-0.316***	-0.333***	-0.290***	465.2***	389.2***	306.1***	319.2***	280.7***
	(0.0172)	(0.0252)	(0.0304)	(0.0315)	(0.0334)	(14.12)	(20.07)	(22.97)	(23.45)	(24.15)
T=3+	-0.646***	-0.545***	-0.335***	-0.366***	-0.329***	576.9***	478.5***	317.9***	338.1***	302.0***
	(0.0184)	(0.0293)	(0.0457)	(0.0475)	(0.0524)	(14.04)	(22.54)	(33.27)	(34.73)	(36.24)
Constant	0.989***	0.410***	0.231**	0.278***	0.499***	1290.1***	1852.6***	1974.8***	1926.9***	1699.5***
	(0.0121)	(0.0654)	(0.0748)	(0.0767)	(0.0411)	(9.751)	(61.57)	(64.34)	(62.61)	(32.04)
Controls	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Year fixed effects	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Cohort dummies	No	No	No	No	Yes	No	No	No	No	Yes
District dummies	No	No	No	No	Yes	No	No	No	No	Yes
Observations	7549	7393	7393	7393	6588	7549	7393	7393	7393	6588
Adjusted R-squared	0.181	0.225	0.235	0.242	0.287	0.230	0.288	0.310	0.321	0.369

Standard errors in parentheses
p<0.05 ** p<0.01 *** p<0.001.
Note that the sample size used to conduct the panel analysis is the sum of the total number of households sampled for each cohort multiplied by the respective years of data collection from 2012 to 2015, which sums up to over 7,000 observations (a detailed breakdown of the overall household sample can be found in Table 5). The number of observations presented in columns (1), however, decreases in columns (2), (3), and (4), due to the existence of missing values in covariates included in these specifications. In the case of columns (5), this reduction is larger because 'permanent attrition' must be assumed for this control specification: once households drop out of the sample in one year they do not 'return' into the sample. Such a restriction has not been applied to the other specifications.

Table 42 Panel regression: Poverty gap

	Poverty gap - lower bound					Poverty gap in PC – lower bound in BDT (2014)				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
T=1	-0.138***	-0.125***	-0.118***	-0.120***	-0.114***	-230.2***	-209.4***	-197.2***	-201.3***	-190.1***
	(0.00363)	(0.00438)	(0.00448)	(0.00457)	(0.00442)	(6.061)	(7.330)	(7.492)	(7.642)	(7.386)
T=2	-0.163***	-0.149***	-0.128***	-0.131***	-0.124***	-272.0***	-248.7***	-213.3***	-218.6***	-207.0***

	(0.00375)	(0.00485)	(0.00553)	(0.00562)	(0.00568)	(6.263)	(8.111)	(9.243)	(9.395)	(9.489)
T=3+	-0.188***	-0.169***	-0.131***	-0.136***	-0.132***	-313.8***	-282.9***	-219.5***	-227.4***	-221.4***
	(0.00388)	(0.00532)	(0.00760)	(0.00769)	(0.00792)	(6.493)	(8.898)	(12.70)	(12.85)	(13.24)
Constant	0.219***	0.140***	0.109***	0.118***	0.142***	366.0***	234.6***	183.0***	197.7***	237.5***
	(0.00286)	(0.0119)	(0.0125)	(0.0127)	(0.00817)	(4.777)	(19.87)	(20.85)	(21.22)	(13.65)
Controls	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Year fixed effects	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Cohort dummies	No	No	No	No	Yes	No	No	No	No	Yes
District dummies	No	No	No	No	Yes	No	No	No	No	Yes
Observations	7549	7393	7393	7393	6588	7549	7393	7393	7393	6588
Adjusted R-squared	0.402	0.426	0.436	0.451	0.441	0.402	0.426	0.436	0.451	0.441

Standard errors in parentheses
p<0.05 ** p<0.01 *** p<0.001.
Note that the sample size used to conduct the panel analysis is the sum of the total number of households sampled for each cohort multiplied by the respective years of data collection from 2012 to 2015, which sums up to over 7,000 observations (a detailed breakdown of the overall household sample can be found in Table 5). The number of observations presented in columns (1), however, decreases in columns (2), (3), and (4), due to the existence of missing values in covariates included in these specifications. In the case of columns (5), this reduction is larger because 'permanent attrition' must be assumed for this control specification: once households drop out of the sample in one year they do not 'return' into the sample. Such a restriction has not been applied to the other specifications.

Table 43 Panel regression: Graduation

	Graduation from CLP				
	(1)	(2)	(3)	(4)	(5)
T=1	0.753***	0.727***	0.626***	0.600***	0.590***
	(0.0108)	(0.0119)	(0.0150)	(0.0152)	(0.0158)
T=2	0.998***	0.969***	0.773***	0.761***	0.757***
	(0.0115)	(0.0136)	(0.0232)	(0.0227)	(0.0247)
T=3+	1.148***	1.087***	0.685***	0.686***	0.685***
	(0.0142)	(0.0183)	(0.0373)	(0.0365)	(0.0403)
Constant	-0.208***	-0.466*	0.351	0.403	0.0204
	(0.00803)	(0.231)	(0.229)	(0.224)	(0.0558)
Controls	No	Yes	Yes	Yes	Yes

Year fixed effects	No	No	Yes	Yes	Yes
Cohort dummies	No	No	No	No	Yes
District dummies	No	No	No	No	Yes
Observations	7676	7519	7519	7519	6715
Adjusted R-squared	0.532	0.540	0.559	0.591	0.535

Standard errors in parentheses
p<0.05 ** p<0.01 *** p<0.001.
Note that the sample size used to conduct the panel analysis is the sum of the total number of households sampled for each cohort multiplied by the respective years of data collection from 2012 to 2015, which sums up to over 7,000 observations (a detailed breakdown of the overall household sample can be found in Table 5). The number of observations presented in columns (1), however, decreases in columns (2), (3), and (4), due to the existence of missing values in covariates included in these specifications. In the case of columns (5), this reduction is larger because 'permanent attrition' must be assumed for this control specification: once households drop out of the sample in one year they do not 'return' into the sample. Such a restriction has not been applied to the other specifications.

Table 44 Panel regression: Income

	Per capita annual income at 2014 prices					Total household annual income over the past year at 2014 prices				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
T=1	9095.1***	6465.3***	5597.9***	5449.0***	5736.6***	33343.6***	23124.8***	19271.3***	18500.4***	19546.0***
	(238.6)	(345.2)	(367.4)	(370.8)	(356.2)	(779.0)	(1108.3)	(1197.9)	(1216.1)	(1290.5)
T=2	11592.9***	8045.4***	6415.1***	6265.2***	6464.3***	43938.8***	28927.4***	22478.3***	21915.8***	23521.3***
	(277.7)	(417.0)	(496.0)	(510.1)	(510.7)	(928.0)	(1353.5)	(1724.9)	(1826.8)	(1957.2)
T=3+	12625.6***	7274.1***	4434.5***	4368.2***	4703.8***	49514.4***	27541.1***	17277.6***	16847.4***	19129.4***
	(300.8)	(547.3)	(746.3)	(773.0)	(794.3)	(1041.0)	(1722.5)	(2644.1)	(2786.4)	(3090.0)
Constant	4403.1***	90.44	7257.0	8145.3*	12330.9***	15840.8***	-42764.9***	-15922.1	-17290.8	-840.6
	(193.8)	(3525.7)	(3767.4)	(3956.6)	(1212.2)	(661.2)	(11077.8)	(11594.7)	(11896.4)	(4000.9)
Controls	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Year fixed effects	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Cohort dummies	No	No	No	No	Yes	No	No	No	No	Yes
District dummies	No	No	No	No	Yes	No	No	No	No	Yes

Observations	7824	7657	7657	7657	6800	7824	7657	7657	7657	6800
Adjusted R-squared	0.249	0.335	0.339	0.355	0.385	0.295	0.390	0.397	0.416	0.451

Standard errors in parentheses
p<0.05 ** p<0.01 *** p<0.001.
Note that the sample size used to conduct the panel analysis is the sum of the total number of households sampled for each cohort multiplied by the respective years of data collection from 2012 to 2015, which sums up to over 7,000 observations (a detailed breakdown of the overall household sample can be found in Table 5). The number of observations presented in columns (1), however, decreases in columns (2), (3), and (4), due to the existence of missing values in covariates included in these specifications. In the case of columns (5), this reduction is larger because 'permanent attrition' must be assumed for this control specification: once households drop out of the sample in one year they do not 'return' into the sample. Such a restriction has not been applied to the other specifications.

Table 45 Panel regression: Savings

	Total value of household savings at 2014 prices				
	(1)	(2)	(3)	(4)	(5)
T=1	4709.3***	2378.8***	2130.5***	1962.1***	1797.8***
	(239.0)	(297.1)	(308.4)	(306.2)	(313.2)
T=2	5477.6***	2591.0***	2488.0***	2413.0***	2392.0***
	(232.2)	(323.5)	(403.5)	(425.9)	(473.9)
T=3+	5531.4***	2439.0***	2988.8***	2703.0***	2723.0***
	(251.1)	(415.2)	(653.0)	(685.9)	(777.3)
Constant	-1264.8***	-9670.9***	-10330.7***	-10142.3***	-2488.1**
	(173.2)	(2834.8)	(2955.1)	(3009.0)	(809.2)
Controls	No	Yes	Yes	Yes	Yes
Year fixed effects	No	No	Yes	Yes	Yes
Cohort dummies	No	No	No	No	Yes
District dummies	No	No	No	No	Yes
Observations	7824	7818	7818	7818	6871
Adjusted R-squared	0.097	0.156	0.161	0.164	0.178

Standard errors in parentheses
p<0.05 ** p<0.01 *** p<0.001
Note that the sample size used to conduct the panel analysis is the sum of the total number of households sampled for each cohort multiplied by the respective years of data collection from 2012 to 2015, which sums up to over 7,000 observations (a detailed breakdown of the overall household sample can be found in Table 5). The number of observations presented in columns

(1), however, decreases in columns (2), (3), and (4), due to the existence of missing values in covariates included in these specifications. In the case of columns (5), this reduction is larger because 'permanent attrition' must be assumed for this control specification: once households drop out of the sample in one year they do not 'return' into the sample. Such a restriction has not been applied to the other specifications.

Table 46 Panel regression: Assets and asset poverty

	Total value of all household assets at 2014 prices					Asset poverty using lower bound				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
T=1	41827.9***	38152.3***	29607.6***	29422.4***	29449.7***	-0.503***	-0.438***	-0.358***	-0.360***	-0.330***
	(897.2)	(1921.6)	(2261.2)	(2402.1)	(2670.0)	(0.0147)	(0.0197)	(0.0203)	(0.0211)	(0.0216)
T=2	63755.4***	58295.5***	42639.1***	42751.1***	41639.1***	-0.636***	-0.566***	-0.369***	-0.371***	-0.341***
	(1620.5)	(2456.2)	(3630.7)	(3850.1)	(4338.5)	(0.0156)	(0.0213)	(0.0255)	(0.0269)	(0.0290)
T=3+	84475.4***	75784.7***	46598.6***	48093.9***	47736.4***	-0.806***	-0.700***	-0.343***	-0.348***	-0.326***
	(2118.0)	(3115.2)	(5770.8)	(6177.1)	(7005.2)	(0.0174)	(0.0252)	(0.0381)	(0.0399)	(0.0440)
Constant	-2527.9*	-29369.5	44476.5	52913.9*	-38854.2***	0.941***	1.460***	0.647**	0.584**	1.009***
	(1177.8)	(22506.6)	(23354.0)	(24854.0)	(8877.7)	(0.0116)	(0.207)	(0.207)	(0.218)	(0.0660)
Controls	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Year fixed effects	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Cohort dummies	No	No	No	No	Yes	No	No	No	No	Yes
District dummies	No	No	No	No	Yes	No	No	No	No	Yes
Observations	7824	7658	7658	7658	6801	7824	7818	7818	7818	6871
Adjusted R-squared	0.243	0.261	0.273	0.285	0.294	0.306	0.318	0.345	0.352	0.322

Standard errors in parentheses
p<0.05 ** p<0.01 *** p<0.001.
Note that the sample size used to conduct the panel analysis is the sum of the total number of households sampled for each cohort multiplied by the respective years of data collection from 2012 to 2015, which sums up to over 7,000 observations (a detailed breakdown of the overall household sample can be found in Table 5). The number of observations presented in columns (1), however, decreases in columns (2), (3), and (4), due to the existence of missing values in covariates included in these specifications. In the case of columns (5), this reduction is larger because 'permanent attrition' must be assumed for this control specification: once households drop out of the sample in one year they do not 'return' into the sample. Such a restriction has not been applied to the other specifications.

Table 47 Panel regression: Total expenditure

	Per capita total expenditure over the last year at 2014 prices				
	(1)	(2)	(3)	(4)	(5)
T=1	7707.6***	4833.8***	4255.5***	4150.1***	3872.5***
	(201.8)	(295.1)	(322.7)	(334.0)	(314.8)
T=2	9881.9***	6052.5***	4843.8***	4691.8***	4199.2***
	(258.0)	(385.5)	(472.1)	(491.6)	(487.6)
T=3+	11241.2***	5563.8***	3183.9***	3090.6***	2700.4***
	(280.8)	(509.0)	(722.7)	(759.4)	(777.1)
Constant	4764.2***	4551.0	10773.7***	11444.8***	12853.8***
	(176.8)	(3119.0)	(3269.2)	(3416.6)	(1166.2)
Controls	No	Yes	Yes	Yes	Yes
Year fixed effects	No	No	Yes	Yes	Yes
Cohort dummies	No	No	No	No	Yes
District dummies	No	No	No	No	Yes
Observations	7824	7657	7657	7657	6800
Adjusted R-squared	0.230	0.324	0.327	0.336	0.369

Standard errors in parentheses
p<0.05 ** p<0.01 *** p<0.001.
Note that the sample size used to conduct the panel analysis is the sum of the total number of households sampled for each cohort multiplied by the respective years of data collection from 2012 to 2015, which sums up to over 7,000 observations (a detailed breakdown of the overall household sample can be found in Table 5). The number of observations presented in columns (1), however, decreases in columns (2), (3), and (4), due to the existence of missing values in covariates included in these specifications. In the case of columns (5), this reduction is larger because 'permanent attrition' must be assumed for this control specification: once households drop out of the sample in one year they do not 'return' into the sample. Such a restriction has not been applied to the other specifications.

Table 48 Panel regression: Expenditure

	Per capita food expenditure over the last year at 2014 prices					*Per capita* non-food expenditure over the last year at 2014 prices				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)

T=1	1173.5***	798.6***	1030.1***	1071.6***	1189.1***	6534.0***	5160.9***	4265.8***	4048.7***	3702.6***
	(61.68)	(81.77)	(86.93)	(89.16)	(90.99)	(183.5)	(254.0)	(283.8)	(293.0)	(295.3)
T=2	986.1***	635.9***	936.6***	1037.1***	1201.2***	8895.7***	7091.9***	5353.3***	4983.4***	4484.3***
	(73.88)	(101.6)	(122.6)	(128.5)	(135.2)	(239.2)	(329.2)	(431.1)	(453.5)	(471.6)
T=3+	1350.5***	610.0***	789.7***	979.3***	1148.7***	9890.7***	6885.2***	3835.4***	3386.0***	3096.4***
	(79.10)	(131.7)	(190.0)	(199.6)	(213.0)	(260.5)	(444.5)	(686.7)	(724.1)	(758.6)
Constant	4783.8***	7925.5***	7537.1***	7731.0***	4961.8***	-19.66	-4682.8	2359.4	3139.7	6574.4***
	(52.42)	(1165.7)	(1214.7)	(1323.6)	(330.8)	(163.5)	(2727.3)	(2839.6)	(3037.8)	(1047.8)
Controls	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Year fixed effects	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Cohort dummies	No	No	No	No	Yes	No	No	No	No	Yes
District dummies	No	No	No	No	Yes	No	No	No	No	Yes
Observations	7824	7657	7657	7657	6800	7824	7824	7824	7824	6876
Adjusted R-squared	0.064	0.186	0.207	0.218	0.308	0.210	0.248	0.255	0.264	0.276

Standard errors in parentheses
p<0.05 ** p<0.01 *** p<0.001
Note that the sample size used to conduct the panel analysis is the sum of the total number of households sampled for each cohort multiplied by the respective years of data collection from 2012 to 2015, which sums up to over 7,000 observations (a detailed breakdown of the overall household sample can be found in Table 5). The number of observations presented in columns (1), however, decreases in columns (2), (3), and (4), due to the existence of missing values in covariates included in these specifications. In the case of columns (5), this reduction is larger because 'permanent attrition' must be assumed for this control specification: once households drop out of the sample in one year they do not 'return' into the sample. Such a restriction has not been applied to the other specifications.

Table 49 Panel regression: Nutrition

	Number of food groups consumed in the last week					Number of meals consumed in last 7 days				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
T=1	0.841***	0.751***	0.702***	0.737***	0.664***	2.129***	1.888***	1.886***	1.935***	1.830***
	(0.0382)	(0.0541)	(0.0573)	(0.0587)	(0.0590)	(0.0739)	(0.0907)	(0.0924)	(0.0962)	(0.0905)
T=2	1.123***	1.047***	0.848***	0.888***	0.769***	2.430***	2.090***	2.089***	2.114***	2.044***
	(0.0438)	(0.0613)	(0.0748)	(0.0781)	(0.0849)	(0.0751)	(0.0944)	(0.0938)	(0.0971)	(0.0869)
T=3+	1.406***	1.314***	0.931***	1.005***	0.867***	2.602***	2.117***	2.040***	2.097***	2.065***
	(0.0459)	(0.0742)	(0.112)	(0.117)	(0.133)	(0.0775)	(0.101)	(0.101)	(0.107)	(0.0969)
Constant	4.508***	4.439***	5.033***	5.066***	4.896***	18.40***	19.84***	20.22***	20.36***	18.99***
	(0.0305)	(0.583)	(0.616)	(0.643)	(0.173)	(0.0584)	(0.784)	(0.827)	(0.958)	(0.236)
Controls	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Year fixed effects	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes
Cohort dummies	No	No	No	No	Yes	No	No	No	No	Yes
District dummies	No	No	No	No	Yes	No	No	No	No	Yes
Observations	7808	7542	7542	7542	6657	7808	7387	7387	7387	6583
Adjusted R-squared	0.151	0.157	0.170	0.184	0.178	0.335	0.356	0.359	0.372	0.342

Standard errors in parentheses
p<0.05 ** p<0.01 *** p<0.001
Note that the sample size used to conduct the panel analysis is the sum of the total number of households sampled for each cohort multiplied by the respective years of data collection from 2012 to 2015, which sums up to over 7,000 observations (a detailed breakdown of the overall household sample can be found in Table 5). The number of observations presented in columns (1), however, decreases in columns (2), (3), and (4), due to the existence of missing values in covariates included in these specifications. In the case of columns (5), this reduction is larger because 'permanent attrition' must be assumed for this control specification: once households drop out of the sample in one year they do not 'return' into the sample. Such a restriction has not been applied to the other specifications.

Analysing treatment effects across sub-populations

In addition to the above analyses, we conducted a second set in which we included interaction terms in a number of the panel regressions in order to determine differential treatment effects by certain household characteristics. The choice of these household characteristics was based on qualitative research conducted prior to this analysis. The key point here was to assess whether we could find supporting quantitative evidence to our qualitative findings about 'game changers' with respect to how participant households could benefit from CLP-2. We considered only binary household characteristics here.

In order to implement this, we interact the dummy variable of interest (e.g. female-headed household) with the post-baseline dummy variables (treatment in each time period) in a fixed effects framework including the same regressors as the main specification as well as year dummies and household fixed effects. Equation (3) exemplifies how this was implemented in our analytical setting:

$$y_{it} = \alpha_0 + \alpha_j \boldsymbol{T_{jit}} + \theta_1 \boldsymbol{T_{jit}} * S_{it} + \mu_1 S_{it} + \boldsymbol{\beta}\mathbf{X_{it}} + \gamma_t \mathbf{D_t} + c_i + u_{it}. \quad (3)$$

As before, y_{it} represents the outcome variable for household i in year t. We also include the vector of dummy variables to capture the treatment effect of the intervention over time ($\boldsymbol{T_{jit}}$). We interact the treatment vector with the differential variable, S_{it}, and include this variable in its level form. Again, $\mathbf{X_{it}}$ is a vector of covariates for each household and year which differs for each outcome variable. $\mathbf{D_t}$ is a vector of year dummies including dummies for 2012, 2013 and 2014, and c_i represents the household fixed effect. Finally, u_{it} is the error term. As before, we weight the analysis using the 2015 sample weights and use robust standard errors to control for serial correlation in the error terms.

We are interested in the marginal difference between households with different characteristics, i.e. the estimate of $\widehat{\theta_1}$, and therefore we compare the marginal impacts of the interaction between the variable of interest and the post-baseline variables rather than simply looking at the magnitude of the estimated coefficients. As an example, we analyse differential treatment effects on the consumption poverty line with regards to flooding vulnerability, female-headed households, households that paid a dowry and households with a disable household member. The results of this analysis are plotted and presented in Section 4.2.1 in Volume I (OPM 2016a).

6.5.4 Limitations

When using longitudinal data, attrition is inevitable and is usually a key concern. Although we found very low rates of attrition in the data prior to 2014, we found unusually high levels of attrition between 2014 and 2015. As has already been mentioned, attrition was also non-random; this causes a problem of sample selection, which could bias our estimates. Although we were able to use inverse probability weights to address attrition and implement a sensitivity check by excluding 2015 data from our analysis, this issue should be kept in mind when interpreting the results. However, given that our results were not drastically changed when we control for attrition, we do not think the bias introduced by attrition is severe and hence this does not compromise the validity of our results.

Second, there is always a concern that the key assumption of strict exogeneity underlying unbiased estimates in a fixed effects model is violated. Ideally, we would like to have data from a randomised control trial, in which case we can be sure that strict exogeneity holds by design. Given that this is not the case for CLP-2, there is a concern that there is selection bias and that this assumption does not hold. However, we do implement tests for the plausibility of this exogeneity assumption (see section 6.5.3) and find robust evidence that it is indeed likely to be satisfied and that this should therefore not be a concern.

Finally, in order to fully assess the sustainability of the programme, it would be ideal to disaggregate the treatment effect for each time period in the CLP annual survey data, i.e. for time periods 1, 2, 3, 4 and 5 separately. For cohorts 2.1 and 2.2, this would mean that we have six rounds of data, which extend to five periods after implementation. However, due to the staggered nature of implementation and data collection, we do not have enough observations for later time periods (i.e. time periods 4 and 5) to be able to disaggregate to this level (see Table 5 for the numbers of observations of each cohort).

6.6 VfM/Cost-effectiveness analysis

As is also discussed in section 8 in Volume I, this section updates recent VfM studies of CLP by incorporating our estimated impact results into cost-effectiveness ratios. Here we provide more details on how we arrived at our results, and on the benchmarks we compared our results to. We also give a more detailed analysis of the conclusions reached.

6.6.1 Objectives

This analysis explores cost-effectiveness, which is one aspect of VfM. The project provides VfM in several regards, which are discussed in the following pages. However, the question cannot be answered exclusively from our findings as this study builds on recent analyses of CLP-2, which were wider in scope than the analysis presented in this section. Previous VfM analyses of CLP-2 (White 2014; Wylde *et al*. 2015) estimated the projected beneficial impact on participant households and compared them to the cost incurred and projected this until project end. These previous analyses were wider in their scope, assessing the '3Es' of economy, efficiency and effectiveness and offering an overall assessment of cost-effectiveness.[18]

The scope of our analysis is narrower in that (a) we consider only estimated impacts on participant households from the quantitative study rather than a wider set of projected benefits and (b) we look at a limited set of cost-effectiveness indicators. Specifically, we assess the cost-effectiveness of CLP-2 by comparing costs to increases in income, increases in consumption and reductions in the poverty gap. As outlined in the Inception Report (OPM 2015), the key cost-effectiveness measures reported are: (i) cost per 1 Taka of increased income; (ii) cost per 1 Taka of increased consumption; and (iii) cost to reduce the extreme poverty gap by 1 Taka.

6.6.2 Data

Cost data were obtained from CLP for the financial years 2010/11 to 2015/16 of the programme. For all the indicators, we used the direct programme cost, following the approach of White (2014)

[18] A fourth E referring to 'equity' is sometimes added to assess the extent to which benefits are distributed fairly. However, neither Wylde *et al*. (2015) nor White (2013) address equity.

and Wylde *et al*. (2015). It is common for cost–benefit and cost-effectiveness analyses to use direct programme cost incurred by participants and place them along the benefits – in this case income and consumptions increases and poverty gap reductions; the benefit is placed against the expenditure to achieve it. The table below shows the direct costs that were used to compute the cost-effectiveness ratios for this analysis.

Table 50 Direct programme cost to core participant households

Direct programme costs:	FY 2010/11	FY 2011/12	FY 2012/13	FY 2013/14	FY 2014/15	FY 2015/16
CBHH (78,026)						
Livelihoods and infrastructure	£5.31	£6.88	£6.16	£6.08	£6.66	£1.84
IMOs' delivery cost of livelihoods and infrastructure	£0.86	£1.15	£1.25	£1.25	£1.11	£0.74
Social group formation and discussion	£0.70	£1.27	£1.49	£1.50	£1.58	£0.65
Monga social protection payments	£0.06	£0.06	£0.03	£0.03	£0.03	£0.02
Market development programme	£0.19	£0.37	£0.34	£0.47	£0.33	£0.27
Education, governance, youth training, flood relief	£0.21	£0.17	£0.09	£0.04	£0.18	£0.04
Healthcare and nutrition	£0.39	£0.20	£0.42	£0.48	£0.47	£0.32
GBP direct programme cost (Nominal)	**£8.66**	**£11.48**	**£11.32**	**£11.70**	**£12.43**	**£4.74**
GBP direct programme cost (Real)	**£8.66**	**£11.48**	**£11.32**	**£11.70**	**£11.71**	**£4.38**
BTD direct programme cost (real)	BDT 787	BDT 1,174	BDT 1,183	BDT 1,142	BDT 1,149	BDT 383
Net present value (NPV)	BDT 5,686 million **(Discount rate 7.5 %; Direct programme cost discounted – base year 2014)**					

The impact estimates used for this analysis are the panel result of our quantitative impact assessments and are at the household level. These can be found in Table 51 below. As with all our impact estimates, these results represent the change in income, consumption and poverty gap relative to baseline and are *not* year-to-year increases. For each time period after baseline, they show the estimated levels of each indicator for participant households compared to a counterfactual in which these households did not participate in CLP-2. For the present analysis, these levels are taken to be representative of the CLP-2 participant population. In one scenario, these levels are taken to be *sustained* over the relevant time periods, without assuming further increases.

We use the panel estimates because they render the best estimates for all three time periods and could be used for T1 through T5. Moreover, they are generalisable across cohorts. The PSM estimates are cohort-specific and thus we only have impacts for T1 and T2. Therefore, we cannot model the same scenarios (and certainly not the same timelines) for them.

Table 51 Panel results for cost-effectiveness indicators

	Panel		
	T=1	T=2	T=3 (avg. T3-T4-T5)
Change in total household income relative to baseline (BDT)	18,500	21,916	16,847
Change in *per capita* household consumption per month relative to baseline (BDT)	286	319	338

Change in poverty gap relative to baseline (BDT)	-201	-219	-227

For inflation, we use the GDP deflator from the World Economic Outlook Database (IMF 2016). As the social discount rate, we have used 7.5%, following the example of White (2014) and Wylde *et al*. (2015). While the source of this figure is not explicitly discussed, a review of White (2014) suggests that it is specific to Bangladesh and either specific to or suggested by DFID. We cannot say whether the social discount rate is real or nominal. However, compared to other studies and the typical social discount rate for World Bank cost–benefit analyses, it should be noted that it is rather high. Therefore, CLP-2 under this analysis may appear less cost-effective than programmes in other studies since future impacts were discounted at a higher rate.

Note that, for the discount and inflation rate, we neglected the different timing of the CLP-2 financial year and the time period between surveys, when results were recorded. The financial year (for cost) runs between April and March. We adjusted the currency conversion rate to these time periods. The period between which impacts were recorded runs between October and September. Inflation and social discount rates were not broken down by months; instead, the difference was not taken into account.

2014 was used as base year for costs and for impacts. All impact estimates are stated in 2014 monetary units. Impacts were projected for periods between 2010/11 and 2025 (and 2010/11 until 2018 for the alternative scenario), and aggregated. The costs were stated in nominal terms and in GBP. These were deflated to 2014 terms and converted into BDT. Costs and impact estimates were discounted at a social discount rate of 7.5%.

6.6.3 Methods

In order to assess the cost-effectiveness of CLP-2 we use the following indicators. Indicator (i) is a ratio that compares direct programme cost with the increases in the income of participant households. Indicator (ii) compares direct programme cost to increases in household consumption. Differences in (i) and (ii) capture, amongst other things, time lags or the way increases in income translate into increases in consumption. Both indicators were computed in a manner consistent with the recent VfM analyses (i.e. White 2014; Wylde *et al.* 2015).

Ratio (iii) compares the direct programme cost of the programme for CPHHs with reductions in the poverty gap per household. While less straightforward than comparing costs to income increases, the ratio between costs and reductions in the poverty gap is indicative of targeting efficiency (since the poorer the participant household or the more numerous the poor households among programme recipients, the larger the gap reduced) and operational efficiency (since it is determined by the direct cost of transfers).

6.6.4 Results

6.6.4.1 Absolute cost-effectiveness

Table 52 below reflects the absolute cost-effectiveness under two scenarios – a base scenario and an alternative scenario. The base scenario assumes that estimated impacts for the first five years after programme start are sustained over a period of 15 years.[19] Increases in income and consumption were estimated for the first five years after baseline for participant households (T1, T2 and T3+, with the latter representing an average of years 3, 4, and 5 after the beginning of the programme). In this base scenario, we assume that the higher levels in income and consumption that are due to CLP-2 in the third period of programme participation (T3+) would be sustained for the following 12 years.[20] For example, the difference in income relative to baseline achieved at T3 (BDT 16,847 in Table 51 above) will persist for the next 12 years. In contrast, our alternative scenario assumes that there is no future sustained impact on income or consumption at all. The only increased income and consumption levels counted are those that were estimated in the impact evaluation between 2010/11 and 2015/16 and in the subsequent years the difference between participants and non-participants dissipates (i.e. becomes equal to zero).

Under both scenarios, the CLP is cost-effective in the sense that the amount of additional income incurred by participants exceeds programme costs. Under the base scenario, the costs to increase income and consumption by 1 Taka were BDT 0.452 and BDT 0.499, respectively. Under the alternative scenario, the cost to increase income by 1 Taka was BDT 0.902 and the cost to increase consumption by 1 Taka was BDT 1.06 (see Table 52 below).

Table 52 Cost to increase income/consumption by 1 Taka in the base and alternative scenarios

	Impact estimate	Base scenario	Alternative scenario
Timeline		(2010/11 - 2024/25)	(2010/11 – 2015/16)
		T1, T2, T3 (avg. T3-T4-T5) until T15	T1, T2, T3 (avg. T3-T4-T5)
Cost per 1 Taka of increased income	*Panel*	0.452	0.902
Cost per 1 Taka of increased consumption	*Panel*	0.499	1.06
NPV direct programme cost (Taka MM)		5,950	5,950
NPV income increases aggregated (Taka MM)	*Panel*	13,173	6,600
NPV consumption increases aggregated (Taka MM)	*Panel*	11,927	5,595

As far as increases in consumption are concerned, the programme appears less cost-effective in absolute terms under the alternative scenario. Wylde *et al.* (2015) suggested that increases in consumption following an asset transfer are neither automatic nor immediate. Household members must cultivate the asset and transform it into increased income and then increased consumption. This link between asset transfer and increases in consumption is expected to be weakened by the use of income increases for something other than consumption, and the (expected) time lag between asset transfer and impact in the form of consumption increase (Wylde *et al.* 2015).

[19] The timeline of 15 years ranging from 2010/2011 until 2024/25 was chosen to ensure the comparability of our ratios to those of previous VfM analyses, e.g. Wylde *et al* (2015) simulated a cost–benefit analysis for 2007/08-2022/23 (16 years) and White (2014) used the timeline 2010/2011 – 2026/27 (17 years).

[20] This is in addition to the three years that have already elapsed since the end of CLP participants' active enrolment in the programme at the time of the evaluation, thus bringing the total cost–benefit assessment period to 15 years.

6.6.4.2 Relative cost-effectiveness

Benchmarking these results with similar programmes puts them into perspective and gives an indication of relative cost-effectiveness. Comparing costs against benefits (as above) is one way to assess cost-effectiveness. However, when given the choice between several livelihoods and cash transfer programme designs, an analysis of relative cost-effectiveness is valuable.

Our ratios of cost to increase income by 1 Taka fall within the range of previous cost–benefit studies of CLP. In Table 53, the ratio between cost per 1 Taka of increased income is compared to the *ex ante* expectations of CLP and to five cash and ATPs in Bangladesh. Previous VfM studies have conducted cost–benefit analyses, comparing expected programme cost to expected benefits.[21] At BDT 0.452 (in the base scenario), the cost per 1 Taka of increased income was higher than the cost–benefit ratios estimated by Wylde *et al*. (2015) for CLP, which is partly due to the wider scope of their cost–benefit analysis, which defined benefits not only as income increases but also income support, investment earnings, employment in public works, health cost averted, and averted losses due to plinths. Compared to the cost–benefit ratio of White's (2014) analysis (0.55), the cost per income increase ratio was lower under the base scenario.

Compared to the cost–benefit estimates for other livelihood programmes in Bangladesh, CLP is less cost-effective in increasing incomes. As explained above, this could be due to the wider scope of the cost–benefit ratios used elsewhere. Our cost-effectiveness ratio is higher than the cost–benefit ratio by Wylde *et al.* (2015). The CLP cost–benefit ratio (0.213) by Wylde *et al.* (2015) is close to that of the Specifically Targeted Ultra Poor programme (STUP) (0.194) – a livelihoods programme with an approach that is in many ways comparable to the CLP.

Table 53 Benchmarking of cost to increase income by 1 Taka

	Ex post analysis	*Ex ante* analysis						
	CLP-2	CLP[22]	Shiree	STUP	OTUP	PRIME	Pension	CLP-2
Cost per 1 Taka of increased income	0.452 (Base Scenario) 0.902 (Alternative Scenario)							
Cost–benefit ratio		0.213	0.135	0.194	0.067	0.020	0.185	0.55
Studies, time periods, time lines	2010/11 – 2024/25 – 15 years	Wylde *et al.* (2015) – Cost–benefit analysis for 2007/08 to 2022/23 – 16 years						White (2014): 2010/11 to 2026/27 – 17 years

International comparisons of CLP's cost to increase consumption point to the relatively low cost-to-impact ratio of the CLP. Table 54 places CLP-2's ratio of cost per increase in consumption against that of six other pilot projects globally and against another ATP in Bangladesh. A study on ATPs in six different countries (Banerjee *et al.* 2015) provides a

[21] The recent Wylde *et al.* (2015) study used a microsimulation model to estimate the impact of the livelihood programmes, including CLP-2, on poverty reduction. A cost–benefit analysis was conducted to understand for every Taka spent on the programme, how much monetised benefit is generated in terms of the income support, investment and public works earnings, improved health and losses averted through plinth-raising. The study compares CLP-2 to five cash transfer programmes in Bangladesh. White's (2014) study is built similarly but with a different time line and an exclusive focus on CLP-2. In order to compare those studies to the present one, the same discount factor but different time periods (2010/11 – 2024/25; 2007/08 – 2022/23) were used.

[22] The timeline of Wylde *et al.* (2015) for CLP starts at 2007/08. It is assumed that they used cost and benefit data from both CLP 1 and CLP-2.

comparison to the cost per increase in consumption. Under the base scenario, the costs to increase consumption by 1 Taka were lower for CLP-2 than for livelihood programmes in Ghana, Pakistan, Honduras and Peru. The cost-effectiveness of Ghana's programme was comparable to CLP's under the alternative scenario.

The cost-effectiveness of Bangladesh's Targeting the Ultra Poor programme (TUP) in increasing consumption is significantly better that that of the CLP, but its target populations and the mix of interventions is not comparable. TUP achieved a significantly lower cost per consumption increase than CLP-2 (0.153 Taka versus 0.499 to 1.063 Taka). It must be recognised, however, that TUP includes two different programme – the aforementioned STUP, with participants and interventions that are similar to CLP, and TUP, with participants who tend to be better-off economically and who receive different types of support (with a greater share of assistance focused on micro-loans and self-help). Moreover, Bandiera *et al.* (2016) applied a lower social discount rate (5%) and assumed that changes in consumption would be sustained perpetually rather than for a limited number of years (13 in our case).

Table 54 Benchmarking of cost to increase consumption by 1 Taka

	CLP-2	TUP	Ethiopia	Ghana	Honduras	India	Pakistan	Peru
Cost per increase in consumption	0.499 (Base Scenario) 1.063 (Alternative Scenario)	0.153[23]	0.549	1.075	2.32	0.327	0.787	0.980
Discount rate	7.5%	5%	7%,					
Study		Bandiera *et al.* (2016)	Banerjee *et al.* (2015)					

The cost to reduce the poverty gap by 1 Taka is compared to results for CLP-2 and five other cash transfer programmes from the previous VfM studies by White (2014) and Wylde *et al.* (2015). Without distinguishing between short- and longer-term effects, White (2014) concluded that the cost to reduce the poverty gap by 1 Taka was BDT 2.08, using DFID's recommended approach for cash transfers. Following his steps, this present analysis rendered a value of BDT 1.38. However, White's methodology is based on transfers rather than impacts, which renders it less comparable.

Comparing the ratio to Wylde *et al.*'s analysis (2015) is difficult as their methodology of computation is unknown. International benchmarking suggests values in the range between 1 and 8 are common for this ratio (Wylde *et al*. 2015: xiii). At a cost of BDT 0.968 to reduce the poverty gap by 1 Taka under the base scenario and a cost of BDT 2.082 to reduce the poverty gap by 1 Taka under the alternative scenario, this present analysis found the cost of CLP-2 to reduce the poverty gap by 1 Taka to be well within this range.

6.6.4.3 Is a comparison with these programmes valid?

There are limits to the extent to which our analyses can be compared to other studies and programmes and these need to be taken into account in interpreting the results.

[23] We exclude increase in household assets from the benefits, and assumed costs were discounted to Year 0. Bandiera *et al.* (2016) use estimated consumption increases for four years and assume perpetuity of this consumption increase. Also, we turned around the benefit–cost ratio to a cost–benefit ratio. The social discount rate = 5.

As benchmarks for the cost to increase income, we have selected the asset programmes Economic Empowerment of the Poorest (Shiree) and BRAC's TUP and cash transfer programmes Programmed Initiatives for Monga Eradication (PRIME) that were suggested by Wylde *et al.* (2015).[24] Moreover, we compared our results to White (2014). Cost–benefit analyses are wider in scope, including defined benefits not only as income increases but also income support, investment earnings, employment in public works, health cost averted, and averted losses due to plinths. As our analyses only account for income increases, ratios are expected to be lower. Moreover, at 15 years, our analyses' timeline is shorter than that of previous studies, which used timelines of 16 or 17 years.

For the cost to increase consumption, we looked at a livelihoods project in Bangladesh (Bandiera *et al*. 2016) and other livelihoods. Bandiera *et al.* (2016) analyse the aforementioned TUP, a Bangladesh livelihoods programme that differs in some regards from CLP-2. For example, it does not include an infrastructure component – a costly factor in CLP-2. Moreover, Bandiera *et al* (2016) assume that increases in consumption are sustained perpetually instead of merely for a limited number of years (15 years, in our case). Finally, they use a social discount rate of 5% while we use a social discount rate of 7.5%. Likewise, Banerjee *et al.* (2015) assume that consumption effects of year 3 are sustained in perpetuity, while their figures include an increase in asset values under benefits (which ours does not).

The results for CLP improve when a lower discount rate is used as future benefits get discounted at smaller rates, as can be seen in Table 55.

Table 55 CLP VfM with different social discount rates

Panel data	*Social discount rate (%):*	Direct cost per income increase (BDT)	Direct cost per consumption increase (BDT)
Base scenario	5	0.411	0.45
Alternative scenario	5	0.881	1.04

White (2014) computes the poverty impacts of transfers following his own approach based on a footnote in the DFID guidance on VfM in social transfers (White *et al.* 2013) as follows:

> *Where the full HIES dataset is unavailable but poverty indices have been published, the impact of transfers on poverty headcount and gap can be approximated by assuming a linear consumption distribution below the poverty line and a uniform participant distribution of transfers. The change in poverty headcount due to adding average transfer to the consumption distribution line can then be calculated as* $tq \big/ 2\frac{P_1}{P_0}z$ *and the change in poverty*

24 These appear as STUP (Specially Targeted Ultra Poor) and OTUP (Other Targeted Ultra Poor) in the comparison table.

gap as $q\left(t - {t^2}\Big/{4\frac{P_1}{P_0}z}\right)$*, where t = average transfer, q = no. of poor participants, P0 = poverty headcount index, P1 = poverty gap index, and z = poverty line."*

The assumption of a linear consumption distribution below the poverty line allows White to apply basic geometry to see the effect of raising consumption by the amount of the transfer on the poverty headcount and poverty gap, taking account of the need to exclude the portion of transfers that ends up going to people above the poverty line. Testing his results against microsimulation findings in comparable studies, we find the results to be a reasonable approximation. However, note that international comparisons of the cost to reduce the poverty gap rank anywhere between 1 and 8 (Wylde *et al.* 2015) – a wide range.

Moreover, it should be noted that actual distribution below the poverty line is unlikely to be linear. In this case, much depends on what one includes as a 'transfer' and whether and how one models the poverty impacts of returns to investment in assets. Following White's (2014) steps, this present analysis rendered a value of 1.38. However, due to the shortcomings of this methodology, our estimate of the cost to reduce the poverty gap, which was discussed above, is more reliable.

Conclusion

Overall, from our analysis it appears that CLP-2 does provide VfM. Under both scenarios, the amount of additional income its participants incur exceeds the programme costs. However, when compared domestically, to similar programmes in Bangladesh, the cost-effectiveness of CLP-2 in terms of the programme's ability to increase income appears to be lower. Differences in the social discount rate, the timeline for the analysis, the scope of the programmes being compared and the assumptions made about the sustainability of the increase in income and consumption all limit the comparability of these studies. For instance, Table 55 highlights how the cost-effectiveness changes if the social discount rate is changed. Overall – however – we conclude from our analysis that CLP-2 did provide reasonable benefits for the costs incurred.

References

Bandiera, O., Burgess, R., Das, N., Gulesci, S., Rasul, I., and Sulaiman, M. (2016) 'Labor markets and poverty in village economies'. *Development Discussion Paper* No. EOPP 058). London: STICERD.

Banerjee, A., Duflo, E., Goldberg, N., Karlan, D., Osei, R., Parienté, W., Shapiro, J., Thuysbaert, B., and Udry, C. (2015) 'A multifaceted program causes lasting progress for the very poor: Evidence from six countries'. *Science* 348 (6236). May 2015.

Braun, V. and Clarke, V. (2006) 'Using thematic analysis in psychology'. *Qualitative Research in Psychology*, 3: 77–101.

Caliendo, M. and Kopeinig, S. (2005). 'Some practical guidance for the implantation of propensity score matching'. *IZA Discussion Papers*, No. 1588.

Cataldo, F., Kielmann, K. and Seeley, J. (2011) *Introduction to Qualitative Research Methodology*. Research for Development.

Guest, G., MacQueen, K. and Namey, E. (2010) *Applied Thematic Analysis*. Sage Publications.

Hausman, J. A. (1978) 'Specification Tests in Econometrics'. *Econometrica* 46, 1251–1271.

Imbens, G. and Rubin, D. (2015) *Causal Inference for Statistics, Social, and Biomedical Sciences*, Cambridge University Press.

IMF. (2016). World Economic Outlook Database, October 2015. Retrieved 3 June, 2016, from https://www.imf.org/external/pubs/ft/weo/2015/02/weodata/index.aspx

Krippendorff, K. (2004) *Content Analysis: An Introduction to Its Methodology*. Sage Publications.

Lecher, M. (2002) 'Some practical issues in the evaluation of heterogeneous labour market programmes by matching method'. *Journal of the Royal Statistical Society, A,* 165, 59–82.

Lincoln, Y. and E. Guba (1985) *Naturalistic Inquiry*. Beverly Hills: Sage

Mays, N. and Pope, C. (1995) 'Rigour and Qualitative Research'. *British Medical Journal* 311(6997): 109–112.

OPM (2015*) Longitudinal Monitoring and Independent Impact Assessment of CLP-2: Inception and Design Report.* December 2015.

OPM (2016a) *Longitudinal Monitoring and Independent Impact Assessment of CLP-2: Final Evaluation Report – Volume I.* May 2016.

OPM (2016b) Jasper, P., Nikitin, D., Haynes, A. and Garcia, M. *CLP-2 Evaluation: Descriptive Analysis Report*. February 2016.

Rosenbaum, P.R. and Rubin, D.B. (1985) Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician.* 1985: 39:33–38.

Rubin, D. (2001) 'Using Propensity Scores to Help Design Observational Studies: Application to the Tobacco Litigation'. *Health Services & Outcomes Research Methodology* 2, 169–188.

White, P. (2014) *Chars Livelihood Programme, Bangladesh: developing measures of cost-effectiveness*, Maxwell Stamp PLC for DFID Bangladesh.

White, P., Hodges, A., and Greenslade, M. (2013) *Guidance on measuring value for money in social transfer programmes – second edition toolkit and explanatory text.* Department for International Development (DFID).

Wooldridge, J.M. (2002) *Econometric Analysis of Cross Section and Panel Data.* Cambridge, MA: MIT Press, 2002.

World Bank (2015) *Bangladesh Country Snapshot.* October 2015.

Wylde, E., Khondker, B., and Freeland, N. (2015) *Cost Effectiveness of Selected Livelihood Interventions in Bangladesh.* Department for International Development (DFID).

Annex A Original TOR

Terms of Reference

Longitudinal monitoring and independent impact assessment of CLP-2

1. Introduction

The Chars Livelihoods Programme Phase 2 (CLP-2) is an integrated poverty reduction programme co-funded by UKaid through the Department for International Development (DFID) and the Australia's Department of Foreign Affairs and Trade (DFAT). It operates in one of the world's most vulnerable and challenging locations – the island Chars in north western Bangladesh – and aims to improve the livelihoods, income and food security of extremely poor people living in this area. The Rural Development and Co-operatives Division of the Government of Bangladesh's Ministry of Local Government, Rural Development and Co-operatives sponsors the project, which is implemented through Maxwell Stamp Plc.

The first phase of the programme (CLP 1) was implemented between 2006 and March 2010; the second phase of the programme (CLP 2) began in April 2010 and runs until April 2016[25]. CLP 2 aims to:

- lift more than 65,000 households (more than a quarter million people) out of extreme poverty, based on a set of multidimensional graduation indicators;
- protect 300,000 people from risk by raising 77,000 homesteads on to plinths at least 60 cm above the historical high flood level. It will provide more than 580,000 people with access to a sanitary toilet and over 400,000 people with access to safe water;
- prevent food insecurity of more than a quarter million people by providing integrated asset transfer and 2 million person-days of paid work during the annual "*monga*" periods;
- promote livelihoods for more than 300,000 people (over 78,000 families) by transferring productive assets (e.g. livestock, seeds, saplings) directly to women from the poorest[26] families. These "core participants" will also receive an 18 month support package including: cash and vouchers; latrines; food supplements; and community group formation and training[27].

[25] With a possibility of further (no-cost) extension until June 2016.

[26] Key eligibility criteria for core package: landless; asset less; jobless; women-headed household

[27] For more detailed information see Section 11 and Development Tracker: http://devtracker.dfid.gov.uk/projects/GB-1-114175/documents/

CLP-2 uses rigorous selection criteria and a series of selection processes to ensure appropriate targeting of the programme interventions[28]. CLP 2 has taken a phased approach to implementation, with participants entering the programme through one of 6 cohorts. Cohort 2.1 entered the programme in May 2010, while the final cohort, 2.6, entered the programme in September 2014. Comprehensive baseline data has been collected for each cohort, with follow-up data collection taking place annually. Data has not been systematically collected for non-participants.

2. Objectives

Within the **broader aim** of building an evidence base to reduce extreme poverty and support pro-poor and inclusive economic growth and social development, both in Bangladesh and globally, the evaluation aims to:

2.1 Assess the effectiveness of CLP 2 in:

(i) achieving its main objectives (impacts and outcomes); and

(ii) sustaining developmental impact by strengthening household resilience.

2.2 Draw lessons from CLP's experience to inform delivery of similar programmes, both within Bangladesh and globally;

3. Indicative Scope

An evaluability assessment of CLP in November 2013, conducted by the Economic Policy Research Institute, recommended the undertaking of an operational review of CLP 1 and CLP 2, combined with a longitudinal monitoring exercise for CLP 2. An optional third component was also proposed, comprising a non-experimental (quasi-experimental) impact assessment of CLP 2, building a counterfactual from cohorts that have not yet joined the programme.

Due to time constraints, and an urgent requirement for findings to feed into design of future livelihoods programmes, the operational review was contracted separately, began in November 2014 and will be completed early in 2015[29].

This document therefore provides Terms of Reference for an **independent expert team** to manage an **impact assessment of CLP 2** combining the <u>longitudinal monitoring exercise</u> and a <u>non-experimental (quasi-experimental) impact assessment</u>[30].

[28] See following link for more detail: http://clp-bangladesh.org/wp-content/uploads/2014/08/selection-brief_final.pdf

[29] Where timings overlap, the Independent Expert Team may interact with the operational review team, but will otherwise have access to the review's findings

[30] In addition to the operational review, it should be noted that an independent impact assessment of CLP 1 was carried out in 2011. This assessment developed an indicative theory of change for CLP 1, however this was not incorporated into CLP 2, which was designed and began implementation before the assessment was published. CLP 2 is also part

The longitudinal monitoring and attribution analysis for the impact assessment will draw largely on **existing baseline and follow-up data** collected by CLP. Since the completion of the evaluability assessment, programme implementation has continued and all 6 cohorts are now receiving the intervention. While some further **quantitative** data collection may be possible, this would have to take place within the constraints of the programme implementation cycle and would not be able to draw on a pure 'control' group. Collection of additional **qualitative** data is expected, both to inform the quantitative analysis and to provide insights into any trends that are identified. A summary of the baseline and follow-up data collected for each of the CLP 2 cohorts is included in Annex 2.

The indicative scope of the evaluation is outlined below:

- **During the inception phase:** Validate CLP's existing **quantitative data** and review the usability of existing **qualitative** data to finalise the scope of work. Develop an approach and implementation plan for the combined longitudinal monitoring and impact evaluation. Draw up a communications plan for the evaluation.
- Drawing largely on existing data collected by CLP, carry out a **longitudinal monitoring exercise** to assess the resilience of developmental impact over time, relative to a baseline multi-dimensional index of developmental impact.
- Implement a **non-experimental (quasi-experimental) assessment** of the programme operations and impacts using CLP's existing administrative and other data sets (both quantitative and qualitative), additional qualitative data and possibly new quantitative data collection using appropriate sampling methods as necessary and approved in the implementation plan[31].
- Ensure that quantitative data analysis is informed by and triangulated with **a qualitative assessment** of the programme operations and impacts using, for example, focus groups discussions, in-depth interviews, key stakeholder consultations, analysis of the existing qualitative data base and other approaches as appropriate and approved in the implementation plan;

of an ongoing impact evaluation looking at the added value of nutrition interventions to livelihoods programmes (along with two other DFID-funded livelihoods programmes, UPPR and EEP). The Innovation, Monitoring and Learning Division of CLP also undertakes research studies, which can be found on the programme website: http://clp-bangladesh.org/publications/research-reports/page/2/

[31] NB: the final cohort of participants (2.6) entered the programme and began to receive CLP support from September 2014. Any plans for additional data collection will need to work within these constraints, given that the suggested approach outlined in section 5 envisaged cohort 2.6 as the control group.

- Produce, edit and publish a draft and final fully integrated quantitative-qualitative evaluation report**.**

In addition, the Independent Expert Team will be expected to:

- Consult broadly with key stakeholders to define the key objectives of the evaluation;
- Meet regularly to coordinate and triangulate approaches, both remotely and in person as appropriate;
- Provide updates to the evaluation steering group, including a presentation of quantitative findings prior to carrying out the remaining qualitative work and authoring the final integrated evaluation report.

4. Key questions and themes:

The longitudinal monitoring and non-experimental (quasi-experimental) impact assessment should assess the **effectiveness** of CLP 2 in achieving its objectives; the **impact** of the programmes activities; and the extent to which strengthened household resilience leads to **sustainability** of the programme impacts.

Key evaluation questions focus on the impact and outcome level indicators within the programme log-frame (see Annex **5**), however it will be important to also capture important changes and impacts (both intended and unintended) not explicitly mentioned in the log-frame.

An indicative set of questions is outlined below, but it is expected that these will be refined and rationalised during the inception phase (and may be influenced by the review of the usability of data). The assessments will address a number of questions and themes, which may be modified during the inception phase. In addition to the questions outlined below, cross-cutting themes such as gender should also be considered.

Indicative Evaluation Questions: grouped under OECD/thematic areas

A. Sustainable Impact

I. How many people from the core participant households (CPHHs) have been lifted out of extreme poverty– based on Rajshahi (or Rangpur) rural lower poverty line - through the programme? To what extent is this graduation sustainable? To what extent the impact is **attributable to the CLP**? To what extent has CLP 2 contributed to reducing the poverty gap?

II. To what extent has the CLP: (a) reduced malnutrition, particularly for females and under five children; and (b) improved food security of the CLP participants

B. Effectiveness

III. Using the graduation criteria developed by CLP 2, how many people of CPHHS have been helped to graduate out of extreme poverty? To what extent is this graduation sustainable? What are the major factors that drive sustainable graduation? What are the major factors that cause graduating households to become non-graduated – i.e. to fall back towards extreme poverty? To what extent is the graduation attributable to the CLP?

IV. To what extent has the CLP improved livelihoods of CPHHs in the following areas: increased income, expenditure and savings, improved asset base?

V. To what extent has CLP-2 improved nutrition practices (breastfeeding, micronutrient consumption) among targeted mothers and adolescent girls?

VI. To what extent have market linkages contributed to increases in the profits of all business group members?

VII. Based on the CLP's empowerment scorecard, to what extent has the CLP contributed to enhancing the status of participating women and girls and empowering them socio-economically?

VIII. In what ways and to what extent has the CLP reduced the vulnerability of participating households?

IX. In what ways and to what extent has the CLP increased the overall well-being of its CPHHs and their families?

X. To what extent has CLP-2 improved social capital among char dwellers?

C. Efficiency

XI. To what extent does CLP-2 represent good value for money?

XII. To what extent was the targeting appropriate? What was the level of inclusion and exclusion error? To what extent did the programme target various social groups such as the disabled and elderly?

D. Relevance

XIII. To what extent did the programme contribute to local economic development (local economy)? To what extent have there been spill-over effects and benefits to non-participating Chars-dwellers?

XIV. To what extend different service providers/ organizations approached to provide or facilitate public rights, services and resources to Chars?

5. Approach and Methodology

During the Inception Phase, the **Independent Expert Team,** in consultation with DFID Bangladesh and the Evaluation Steering Group (comprised of representatives of DFID, DFAT and Government of Bangladesh)**,** will be expected finalize the scope of work, developing a detailed methodology, evaluation framework and implementation plan. Based on the evaluability assessment carried out in 2013 by the Economic Policy

Research Institute, the preferred approach for the impact evaluation is outlined below, comprising two main components: the longitudinal monitoring exercise and the non-experimental (quasi-experimental) impact assessment. While these are primarily quantitative in nature, the integration of qualitative data into the evaluation is expected and required.

Close collaboration with the CLP 2 programme will be essential, with time spent in Bogra expected, and a further option available to 'embed' part of the evaluation team within the programme to implement components of the quantitative-qualitative evaluation.

A. Longitudinal monitoring exercise

One of the most important questions facing developmental social protection programmes is the resilience and sustainability of impact. In some programmes, the development outcomes erode rapidly over time, while in others the positive effects remain stable and in yet others the developmental impact continues to grow even after the programme intervention ends. CLP has collected and will continue to collect indicators measuring important dimensions of developmental impact, including: i) Poverty, vulnerability, hunger and food security; ii) Health and nutrition; iii) Assets and livelihoods; iv) Social capital; v) Gender (including intra-households); vi) Systemic changes.

The **longitudinal monitoring exercise** will assess the resilience of developmental impact over time, relative to a baseline multi-dimensional index of developmental impact. By tracking index values (and the sub-components) over time, the study can assess how resilient are the developmental impacts created by the programme. While this component is not specifically designed to rigorously attribute these impacts to the programme intervention, similar studies of other programmes have generated useful evidence on resilience and sustainability with this type of monitoring approach (see annex **4** for an example). The existing and planned data collection activities will support a longitudinal analysis of programme impact over a horizon of up to **six years**. The cohort 2.1 would potentially offer (based on the preliminary analysis of CLP data) a baseline from 2010 and an end-of-treatment end line in 2011/12 followed by longitudinal follow-up indicators in 2013/14/15/16 for a total of 6 years of data. Subsequent cohorts would provide incrementally fewer years depending on whether or not there is any post 2016 follow-up (for a list of cohorts with total number of households and timeline, please see Annex **3**). Combined with an impact assessment component, this process can also undertake to rigorously attribute the outcomes to the programme's interventions.

The evaluation approach will require a detailed longitudinal monitoring plan which the independent expert team will develop, in consultation with CLP, based on global models for quantifying resilience in developmental social protection programmes.

The analysis of data and meta-data provided by CLP documents a series of consistent development indicators over time for a number of programme cohorts, including information collected after participants have exited from the programme. The independent expert team will define an analytical framework assessing multiple dimensions of developmental impact, summarising the diverse indicators into a single index which is tracked over time by households within programme cohorts. Progress in improving overall programme performance over time can be measured across cohorts. The resilience of developmental impact can be tracked within cohorts over time, particularly after the programme interventions have ended. In addition to developing the required database structures for monitoring and analysis in consultation with CLP, the independent expert team will provide the analysis required to assess management responsiveness and learning as well as programme resilience and the dynamic deepening of developmental impact.

B. Quasi-experimental impact assessment

The impact assessment, which would rely on quasi-experimental approaches such as Propensity Score Matching and appropriate Regression[32], may require the application of statistical techniques to estimate a counter-factual to the observed outcomes. An appropriate enhancement to the design could also measure **local economy effects**. The methodological approach would aim to rigorously attribute (or estimate the attribution of) programme impacts

CLP has collected baseline data which can support a non-experimental impact assessment with two recommended comparison groups. The study employs the 2011 programme cohort (2.3) as the comparison group (i.e. target group) as well as the September 2014 programme cohort (2.6) as control group. For a list of cohorts with total number of households and timeline, please see Annex **3**. As cohort 2.6 began receiving the programme interventions shortly after the baseline was collected, the quantitative assessment will largely be reliant on data collected by CLP.

Any approach to assess CLP's causal impacts must address the problem of the counterfactual: what outcomes would have been observed had the CLP participants not received the programme benefits. All rigorous impact assessment strategies are designed to identify a method for constructing a proxy for these counterfactual outcomes using information on non-participants. This requires controlling for the effects of any confounding economic and contextual factors that make programme participants systematically different from an average non-participant, such as the relative poverty of

[32] Although any regression approach would have to deal with the serial correlation inherent in any analysis of a cohort over time.

participants in targeted programs, exposure to economic shocks, or differences in household characteristics (e.g. demographics, skill levels, or social networks), and affect the impacts of the programme. Impact estimates that imperfectly control for these confounders suffer from "selection bias".

The proposed plan will combine the CLP's control with matching methods that construct a comparison group by "matching" treatment group households to comparison group households based on observable characteristics that influence programme participation. The impact of the programme is then estimated as the average difference in the outcomes for each treatment group member from a weighted average of outcomes in each similar comparison group member from the matched sample.

The component will include a qualitative assessment (review of existing CLP qualitative data and fresh FGD, interviews, other appropriate methods as identified by the expert team during the inception period) to triangulate key quantitative findings and also to answer questions related to transformational changes, women empowerment, local economy effect, changing vulnerability context etc.

6. Outputs/deliverables

The independent expert team will produce the following outputs:

- An inception report for the project, which identifies the evaluation's major objectives based on broad consultations, the evaluation's methodological approach and a detailed evaluation framework and work plan. This will include a clear and rigorous attribution strategy employing quasi-experimental approaches. The inception report should make clear how qualitative and quantitative aspects of the evaluation will be fully integrated. A communications plan should also be delivered at the end of the inception phase;
- A design-to-implementation plan for the evaluation, including a fieldwork plan;
- A set of databases containing the underlying data used to construct the indicators of developmental impact, and the Stata do-files or SPSS programmes that create the summary indicators. In addition, a brief report addressing any methodological and data issues identified;
- A de-briefing presentation on early findings at the end of field assessment. In addition, a number of presentations (face-to-face and/or electronically) to the evaluation steering committee on progress update periodically (e.g., monthly/quarterly), including a presentation of quantitative findings;
- Draft and final reports of the integrated quantitative-qualitative evaluation;

- In addition to the main report and its' executive summary , a stand-alone 4 page summary with a short statement describing the purpose of the evaluation, the brief methodology, key conclusions, priority findings and recommendations. The Executive Summary and 4 page summary should both be written using non-technical language that is appropriate for wider audiences.[33]
- A PowerPoint or other-format presentation of the key results, and participation in dissemination activities as determined by the evaluation steering committee.

7. Recipients

The primary recipient of this service will be Government of Bangladesh through the Rural Development and Cooperatives Division (RDCD). DFID Bangladesh and Australia DFAT will be direct users of the study as the co-funding agencies of the programme and findings will be shared with key stakeholders in Bangladesh and globally for improving design and delivery of any similar projects/programmes including any future phases of DFID/DFAT sustainable livelihoods and social protection programmes.

8. Communication and Dissemination

An effective approach to communication and dissemination of findings will be important in ensuring they reach a wide audience and that the uptake of key recommendations is maximised. It is expected that a number of different approaches and channels will be used to reach different audiences and relevant stakeholders. One such planned channel is a dissemination event to be hosted by the Government of Bangladesh in collaboration with UK-DFID, Australia's DFAT (and CLP/Maxwell Stamp). The study team will be expected to attend this event to present and share the key findings with the key development partners and relevant stakeholders in Bangladesh (Government, Development Partners, NGOs, Media, Think Tank/research bodies/ Development Practitioners etc.).

The dissemination activities could include policy briefs and other products to maximise the contribution of the evidence-building process to global learning and policy influencing. The independent expert team will make recommendations to the steering committee in terms of proposed activities and provide a proposed work plan. The budget for the dissemination activities will be developed in line with the recommended options.

The 4-page summary will be used as a communication tool and may be shared both during and after the dissemination event. The final report will be available on the websites of the Maxwell Stamp/CLP, GoB, UK-DFID and DFAT for public access. Moreover, it is expected that the impact assessment methods and findings will inform and contribute to the global evidence base on best practices of social protection and rural livelihoods programmes for the extreme poor.

[33] For example : shorter sentences and paragraphs, limiting the use of Latin phrases, using less technical language

9. Timeline: Key deliverables/dates[34]:

The Independent Expert Team will be engaged for a period of up to 10 months, between June 2015 and April 2016[35] with work on the evaluation likely to begin in late June/early July 2015 following completion of the necessary contractual arrangements and team mobilisation. The indicative timeframe for the activities and outputs for the inception and implementation phase is summarised in Table 2 below.

Table 2. Project activities, timeframes and deliverable outputs

<u>A. Inception Phase</u>

Activity	Timeframe	Output
Commissioning and mobilisation of the Independent Expert Team	by June 2015	Research team structure
Expert Team Arrive in Field/Bangladesh for in-country mission/inception commencement , starting with an initial meeting with the Steering group	Early July 2015	
Developing a draft approach and implementation plan with methodology and research tool; review of existing admin data (both quantitative and qualitative).	Early August 2015	Draft Inception report (with draft Implementation Plan, methods, instrument formats)
Finalisation of the implementation plan /design based on comments from the Steering Group.	Mid-August 2015	Final Inception Report[36]

<u>B. Implementation Phase</u>

Activity	Timeframe	Output
Implementation of qualitative and quantitative research and necessary field work; produce early findings and shared with the Steering Group periodically.	August – November 2015	Monthly/quarterly progress reports/presentations and debriefing the Steering Group about early findings in November 2015
Data processing, analysis and reporting	December 2015	Draft Final Report
Revisions and completion of analysis based on comments from the Steering Group and fact /figures	Feb-March 2016	Final Report

[34] Dates are subject to change if any natural disaster and/or political unrest occur during the time.

[35] With possibility of slight no-cost extension until June 2016.

checking inputs from CLP; proof reading, and final submission ;		
Present Final report in the dissemination meeting	March/April 2016	Presentation on the Final Report

10. Team Composition, Roles and Responsibilities

The Independent Expert Team engaged to undertake this evaluation will need to have the skills and expertise to:

- Design an appropriate quantitative-qualitative impact and operational assessment to meet the objectives agreed in consultation with key stakeholders, including a theory of change; key questions; the high-level methodological approach (which addresses the appropriate scope for qualitative-quantitative integration) and policy linkages, in line with the accompanying ***Evaluation Plan***;
- Establish a strong and collaborative working relationship with the CLP 2 programme team and develop a solid understanding of the characteristics of the programme;
- Provide independent analysis of CLP quantitative data, applying appropriate statistical and matching techniques to assess both sustainability and attribution of results;
- Effectively collect and use qualitative data that can inform the quantitative analysis and explain identified trends.
- Consult broadly with key stakeholders to validate the key results and produce an integrated evaluation report.

The independent expert team will include, but may not be solely limited to, a Team Leader/integrated evaluation expert, an independent quantitative expert, quantitative and qualitative specialists and a three independent experts and a coordinator. The description of each of the roles, as well as the required skills and experience of each team member, is outlined in Table 3. Gender balance within the team is expected.

Table 3.

Role	Description	Skills and Experience
Integrated qualitative-quantitative evaluation expert	The integrated qualitative-quantitative evaluation expert will co-ordinate the longitudinal monitoring exercise and impact	• Experience leading large-scale integrated qualitative-quantitative evaluations, as well as experience in qualitative

and team leader (international):	assessment and will be responsible for designing and overseeing the effective integration of qualitative and quantitative methods. The expert will provide oversight, management and technical support to the evaluation team, including coordinating the strategy and design. For quality control purposes the expert will analytically corroborate all quantitative results, as well as participate in selected qualitative exercises and review the qualitative analysis. The expert will take primary responsibility for editing the reports in conjunction with the other external experts. As team leader, the expert will make required visits to Bangladesh with remote conferencing as required.	evaluations and quasi-experimental or non-experimental impact assessments with rigorous attribution strategies, and use of administrative data in evaluations; • Extensive experience with social protection (at least 10 years), livelihoods programmes and the monitoring and evaluation of social protection programmes. • Demonstrate substantial experience (through field missions) in Bangladesh/South Asia with monitoring and evaluating programmes similar to CLP, BRAC's CFPR or other similar programmes in Bangladesh/South Asia. • Demonstrate extensive experience as a team leader in international projects.
Independent quantitative expert (international):	The independent quantitative evaluation expert will be responsible for the quantitative research (both the non-experimental impact assessment and the longitudinal monitoring exercise) working with the team leader. The expert will be supported by the study coordinator for data management and analysis. The expert will establish and analyse the quantitative longitudinal databases and will assess the resilience of developmental impact over time, relative to a baseline multi-dimensional index of developmental impact. The expert	• Extensive experience with quantitative methodologies including studies that involve qualitative/quantitative integration; • Extensive experience in longitudinal monitoring and quasi-experimental quantitative impact evaluations and use of administrative data in evaluations, particularly in the areas of social protection and livelihoods. • Experience in project management and/or oversight.

	will, be responsible for drafting the quantitative components of the integrated report working with the Team Leader. The quantitative evaluation expert will make required visits to Bangladesh with remote conferencing as required. He/she will work closely with CLP's Innovation, Monitoring and Learning (IML) Division in Bogra.	
Qualitative Specialist (national):	The qualitative evaluation specialist will support the independent expert team leader to design and conduct qualitative research and ensure that the project reflects Bangladesh's social and policy context. The specialist will work closely with the Quantitative experts to ensure effective use and integration of qualitative and quantitative approaches. The Specialist will work with the team leader to draft the qualitative components of the integrated report.	• Extensive experience with operational reviews and other qualitative evaluations, particularly in the context of Bangladesh's livelihoods and social protection programmes. • Experience and understanding of the social dimensions of poverty in Bangladesh and the associated policy context. • Experience in project management and/or oversight. • Demonstrated report-writing experience, and very good English.
Independent Study Coordinator (National – could be hired by the Independent Expert Team through a local consultancy firm. The firm could also provide any additional research/data management, logistical support	The independent study coordinator will coordinate meetings among the independent expert team and between the independent expert team and other stakeholders, maintaining key reporting structures and facilitate communications and information flow among the key stakeholders. The independent team coordinator will also assist the independent expert team with specific research tasks, in particular providing support to the quantitative expert in data	• Demonstrated experience coordinating large-scale evaluations with multiple stakeholders; • Experience in quantitative and/or qualitative evaluation, particularly in the context of Bangladesh's livelihoods, social protection or other programmes. • Proven expertise in creating and managing databases for development outcome and impact analysis. Expertise should include either MS Access or other forms of SQL-based databases.

to the Study Team and dissemination).	management and analysis. The study coordinator will participate in all meetings of the independent expert team.	• Experience of using SPSS/STATA to evaluate large datasets. • Excellent organisational and communications skills.

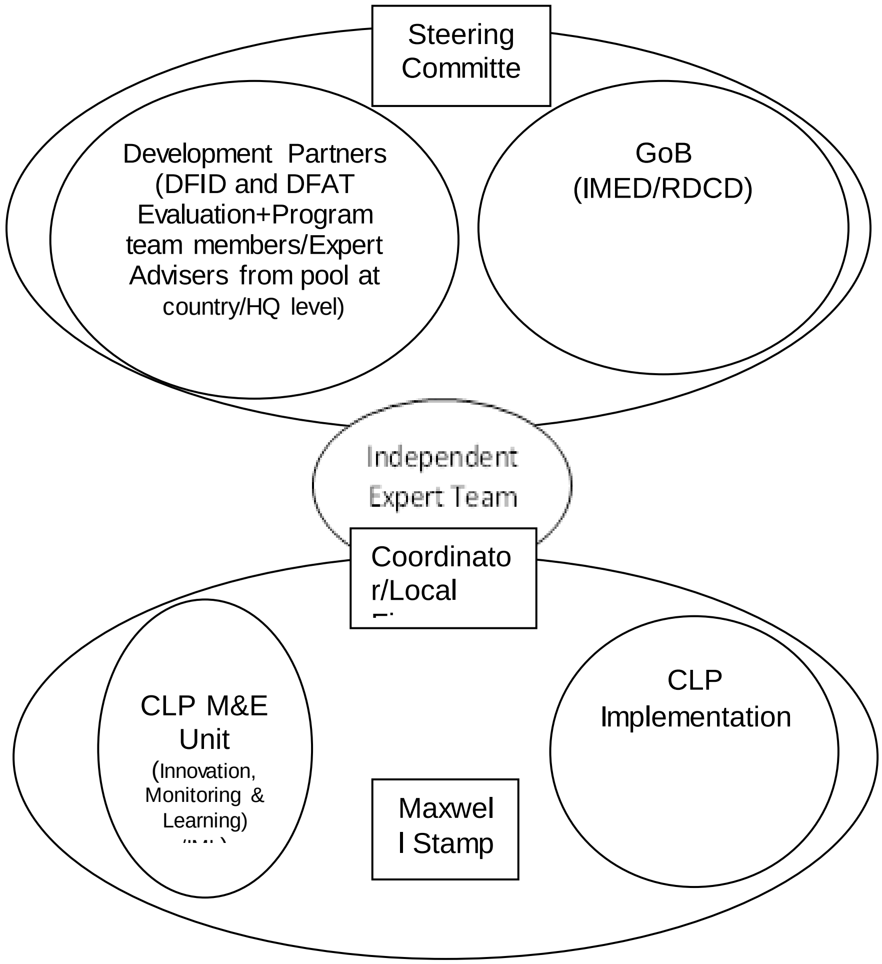
11. Governance structure

The evaluation project will be directed by a steering committee of key stakeholders, including development partner representatives, and the Government of Bangladesh. **The independent expert team will liaise with the steering group** in the design and conduct of the **impact assessment and** will work closely with both CLP's M&E and implementation units. Development Partner (DFID, DFAT) representatives may include Evaluation/Social Protection Experts (staff member or consultants) to provide technical and quality assurance inputs throughout the process. Figure 2 below illustrates the proposed organisational and governance structure.

The steering committee will

- Broaden the policy constituency of the independent expert team;
- Provide additional technical expertise supporting specialised components of the study;
- Peer review the research inputs and report of the independent expert team.

Figure 2. Organisational and governance structure for the evaluation



12. Contract payment structure

DFID's preferred method is to link payments to milestones (payment by results). Bidders should propose a payment plan using payments by results linked to programme outputs that incentivises the achievement of results and value for money.

DFID reserves the right to scale up/ back the contract to respond to changing requirements. The contract will be awarded for 10 months DFID may choose to extend the contract by up to a further 3 months in the case of unforeseen circumstances if there is a value for money rationale and acceptable programme and supplier performance to date.

13. Break Points

There will be a break clause in the contract at the end of the inception phase where the supplier will submit an Inception Report. DFID will review this report and if it is satisfactory will confirm the full contract and move to Implementation Phase.

13. Duty of care/ Logistics/Security/Health
A Summary Risk Assessment Matrix and Circumstance Matrix and Duty of Care Policy-Implementation Guidance Note, and an **'Information Note and Requirements for all Visiting DFID Staff and Consultants'** are attached for your information. Responsibility for the well-being of the supplier's Personnel rests solely with the Service Provider. The Service Provider will be responsible for the provision of suitable security arrangements for them and any business property/equipment that will be used during the course of this assignment. DFID shall forwards any updates to the guidance mentioned above or notify the consultant of any changes to the security situation, as and when these are received during the course of the work.

There is an option to embed one or two members of the team within CLP's offices. If requested, the CLP may arrange necessary field logistics support in completing the field visit (e.g. boats to/from Chars during field work, identifying Chars etc.). However, the Independent Evaluation Team will be expected to provide their own transport:

- whilst in Dhaka
- to/ from Bogra
- to/ from the Chars during field work

The proposal may include a local consultancy firm for data management, study coordination and logistical support.

14. UK-DFIDB Coordination/
The overall coordinator for this study will be DFIDB Poverty and Livelihoods Adviser and he/she will be the focal person for tracking implementation of any recommendations from the evaluation DFIDB Extreme Poverty Team Leader and Evaluation lead will provide technical support/quality assurance inputs throughout the process, as required. Karishma Zaman, DFIDB Programme Manager, will be the project officer of this study and key contact on any logistics.

15. Level of Efforts and Budget
The budget for the project should be in the range of £300,000 to £500,000. DFID will be expecting bidders to demonstrate excellent Value for Money when budgeting for this programme and should only include costs that are necessary to deliver the programme outputs.

16. Risk assessment

Evaluations are intrinsically risky, facing a complex set of challenges including the contracting of expert personnel, the complexities and time requirements of procurement processes, the challenges of Bangladesh's country context compounded by the realities of the Chars areas, <u>mobility restrictions due to political instability (e.g. strike) </u>and an intrinsic risk to any data-dependent exercise for which results are uncertain. This project is rated as medium risk in the absence of specific mitigation activities. Table 3 describes the main areas of risk and identifies mitigation opportunities, particularly in terms of ensuring qualified personnel are contracted, that CLP country risk mechanisms are leveraged and evaluation approaches are diversified. Procurement arrangements such as accountable grant mechanisms may be possible to reduce procurement risk. The risk of the project is rated as low if the identified risks are effectively mitigated. However, the bidders/suppliers will need to submit their own risk assessment during the submission of proposal.

Table 3. Risk assessment and mitigation opportunities

Risk description	Probability if **not** mitigated	Mitigation opportunity	Probability if mitigated
Personnel risk: competency, bias	Medium	Ensure highly **specific** competencies, secure technical inputs from independent expert team	Low
Procurement risk: **delays**, constraints, administrative burden	Medium	Adopt appropriate mechanism after weighing trade-offs: GEFA, OJEU tender, accountable grant, individual contracts	Low
Country risk (including mobility restrictions due to political unrest)	Medium/High	DFID and CLP mechanisms	Low
Evaluation risk	Medium low	Diversify evaluation approaches	Low

Annex 1: Background information and references

A) The Chars Livelihoods Programme-II (April 2010-April 2016)

CLP is a £81.7 million programme supported by DFID and Australia's DFAT and hosted by the Rural Development and Cooperatives Division of Government of Bangladesh. In its second phase , CLP-2's purpose is to improve the livelihoods, incomes and food security of up to one million extremely poor people (including the non-core participants) living on island *Chars* in the north west of Bangladesh.

The CLP is an integrated approach to sustainable livelihoods, delivering a package over 18 months, typically including: (i) raising homesteads onto plinths 2 feet above the high flood level and ensuring access to clean water and a hygienic toilet; (ii) financing a productive asset (people usually decide to buy livestock, particularly cattle) and a small cash stipend; (iii) training in health, household financial management and nutrition (complemented by a direct nutrition supplement); and (iv) ensuring access to basic health care and to markets for selling their produce. After 18 months, most participants are able to sustain and improve their livelihoods with limited further support. The programme then moves on to target other extremely poor households. For further details, please visit: < http://www.clp-bangladesh.org/>

B) References/ Reading Material

i. CLP website < http://www.clp-bangladesh.org/> ; monthly/periodic progress reports;
ii. CLP 1and 2 Programme Memorandum and design documents
iii. CLP-1 Project Completion Report 2010
iv. CLP-1 Final Report 2010
v. White, P (April 2014) 'Chars Livelihood Programme, Bangladesh: developing measures of cost-effectiveness'
vi. White, P. (May 2013) Chars Livelihood Programme, Bangladesh: support for development of a VfM strategy and work plan (Unit cost assessment).
vii. CLP (2012a) *Achieving VfM within the Chars Livelihoods Programme.* Brief, Chars Livelihood Programme, February
viii. CLP (2012b) *The CLP's Approach to Reducing Leakage*, Brief, Chars Livelihood Programme, April
ix. CLP (2012c) *The CLP's Graduation Criteria*. Brief, Chars Livelihood Programme, September
x. DFID (2011) *DFID's approach to value for money (VfM)*, DFID, London.
xi. Hodges, A., P. White and M. Greenslade (2011) *Guidance for DFID country offices on measuring and maximising value for money in cash transfer programmes – toolkit and explanatory text.* DFID, London. October
xii. NAO (2011) *Transferring Cash & Assets to the Poor* National Audit Office, London
xiii. White, P. and M. Greenslade (2013) *Guidance on measuring and maximising value for money in social transfer programmes – toolkit and explanatory text.* Second edition. DFID, London (forthcoming)
xiv. Information Note for all Visiting DFID Staff and Consultants
xv. CLP Annual Reviews/Output to Purpose Review 2007-15
xvi. CLP-1 Impact Assessment + Management Response
xvii. CLP-2 Design Cost Benefit Analysis (Financial Appraisal)
xviii. CLP Research papers /briefs/ studies on Disaster Resilience, Graduation, Food and Nutrition Security, Women Empowerment, Sustainability of Community Based Organisations; cash transfer using mobile phone; village savings and loan group, market development (available on CLP website).
xix. CLP Operational Review Report (March 2015)
xx. CLP Annual Survey Questionnaires (Template)
xxi. Cross-programme Cost Effectiveness Study ToR and Draft Report
xxii. CLP Graduation note on cohorts 2.1-2.4

Annex 2: Summary of CLP Annual Surveys

	Survey 2010	Survey 2010	Survey 2011	Survey 2012	Survey 2013	Survey 2014
Month	**May**	**October**	**June**	**October**	**October**	**October**
Cohorts included	Baseline 2.1 Follow up CLP 1	Baseline 2.2 Tier 2.2	Baseline 2.3 Follow up CLP 1, 2.1, 2.2 & Tier 2.2	Baseline 2.4 Follow up CLP 1, 2.1, 2.2, Tier 2.2 & 2.3	Baseline 2.5 Follow up CLP 1, 2.1, 2.2, Tier 2.2, 2.3 & 2.4	Baseline 2.6 Follow up CLP 1, 2.1, 2.2, Tier 2.2, 2.3 , 2.4 & 2.5
Demographic Information	✓	✓	✓	✓	✓	✓
Income	✓	✓	✓	✓	✓	✓
Expenditure	✓	✓	✓	✓	✓	✓
Assets	✓	✓	✓	✓	✓	✓
Savings	✓	✓	✓	✓	✓	✓
Food Security	✓	✓	✓	✓	✓	✓
WASH	✓ (limited)	✓ (limited)	✓ (limited)	✓	✓	✓
Nutrition	✓	✓	✓	✓	✓ (very limited)	✓
Women's empowerment	✓ (very limited)	✓ (very limited)	✓ (very limited)	✓	✓	✓
Graduation	✓ (very limited)	✓very limited)	✓ (very limited)	✓	✓	✓

Annex 3: Cohort wise households distribution with time line

Cohort Number	Cohort Administrative Start Date	Cohort Assistance Start Date*	Cohort End Date	Administrative Cohort Length	Assistance Cohort Length	Number of CPHHs
2.1	01/04/2010	15/05/2010	31/12/2011	21.01	19.56	5,004
2.2	01/07/2010	30/09/2010	30/06/2012	24.00	21.01	12,109
2.3	01/07/2011	30/09/2011	30/06/2013	24.00	21.01	17,435
2.4	01/07/2012	30/09/2012	30/06/2014	23.97	20.98	16,309
2.5	01/07/2013	15/09/2013	30/06/2015	23.97	21.47	13,579
2.6	01/07/2014	01/09/2014	29/02/2016	19.99	17.95	13,768
						78,204

* All assistance projects do not start at the same time. Usually the group formation and weekly group meetings start first followed by livelihood orientation. Some activities (homestead gardening, asset purchase, IEP work, etc.) start immediately after that while some activities (VSL, market development, etc.) start after couple of months or even later. Moreover, start date largely depends on the completion of baseline survey and therefore varies from IMO to IMO.

Annex 4: Examples of Independently Refereed Evaluation/Impact Assessment

The evaluation design provides maximum value-for-money by leveraging CLP's expertise with the credibility provided by an independent process. CLP understands best the complex interactions that represent the intervention activities. The expert team will serve as independent referees of the evaluation analysis, assuring an objective and credible evaluation. Similar models have been adopted both within Bangladesh (for example, BRAC's Research and Evaluation Department merges elements of self-evaluation with rigorous and credible independence) and internationally (for example, South Africa's Department of Social Development and South African Social Security Agency interacted extensively with an independently contracted evaluation team, influencing the design and research direction without compromising the study's credibility.[37]).

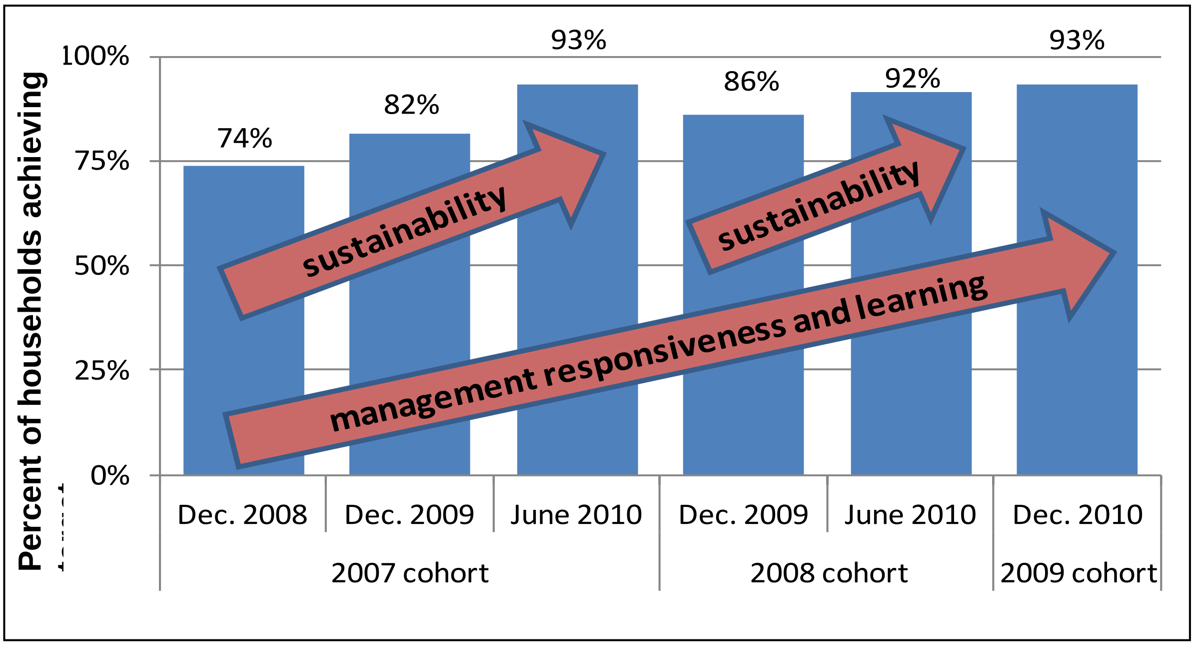
For example, the OECD's 2013 ***Development Co-operation Report*** highlights the analysis of BRAC's ***Challenging the Frontiers of Poverty Reduction (CFPR)*** programme in sustaining and expanding developmental impact, even after participants have ended their direct participation in the programme support activities. The following diagram illustrates the kind of result this component may demonstrate.

Figure 3 illustrates continuing increases in a multiple indicator index of developmental outcomes[38] for three groups of participants in the BRAC's CFPR programme from 2007 to 2009. Participant groups consistently improved outcomes year after year across a range of developmental outcomes, including food security, livelihoods diversity, productive assets, human capital, and other developmental areas. Even after BRAC's provision of developmental benefits ended, programme participants increased their productive assets, improved their livelihoods and strengthened their households' social development (measured through education, health and gender empowerment indicators) and economic opportunities (Das and Misha, 2010; Akhter *et al*., 2009; Samson, 2012a). The increases in the developmental index year-over-year for each of the 2007 and 2008 cohorts represent the sustainability of the programme's impact. The increases over time across cohorts represent on-going improvements in the programme's design and implementation.

[37] The evaluation was ranked in the top 3 out of a set of over a hundred comparable studies by an independent referee panel commissioned by UNICEF.

[38] Including socio-economic indicators related to food security, robustness and diversification of livelihoods, access to quality housing, water and sanitation, savings, school attendance, etc.

Figure 3. Dynamic deepening of developmental impact in BRAC's CFPR Programme



Source: Samson, M. (2012a), "Exit or developmental impact? The role of 'graduation' in social protection programs", 23 August 2012 (cited in OECD 2013)

Annex B OPM evaluation matrices

This annex presents, in a disaggregated manner, the different questions that this evaluation has sought to answer. As described in Table 2 in the introduction to Volume I (OPM, 2016a), these questions were restructured and reordered after the inception phase. We present them here in a disaggregated manner in order to relate them explicitly to the OECD DAC criteria.

It is important to understand that the selection of these questions was based on extensive consultations with DFID, DFAT, and CLP during the inception phase and, importantly, after the quantitative/qualitative workshop held in Dhaka in January 2016.

As mentioned before, this evaluation pursued a two-fold general objective: (i) to evaluate the performance of CLP-2 along several key dimensions as specified by the OECD DAC criteria; and (ii) to draw lessons from this.

Beyond these general objectives, we have structured the activities and tasks of the evaluation as follows:

A. activities and tasks around overarching OECD DAC criteria that need to be addressed by the evaluation; and

B. activities and tasks around specific evaluation questions that relate to each of the criteria.

The OECD DAC criteria can be specified as follows:

- **Effectiveness**: To what extent were the objectives/targets of CLP-2 met?
- **Impact**: What changes in key indicators can be attributed to CLP-2?
- **Sustainability**: Do these changes persist over time?
- **Efficiency**: To what extent does CLP-2 represent good VfM?

As described in the Inception Report (OPM 2015), the fifth OECD DAC criterion, **relevance**, has not been addressed separately by this evaluation. Sub-questions related to this DAC criterion referred to local economic development and the effects of CLP-2 on non-participants and other char dwellers. However, after consultation with CLP and DFID Bangladesh, we have come to the conclusion that this evaluation will not be able to focus on questions related to this criterion. This is due to several reasons: first, it is clear that the quantitative data available to us were data on participants, i.e. not on other char dwellers. Hence, given that no primary data collection for quantitative data was to be implemented for this evaluation, we were not able to assess effects on non-participants or the local economy using such data. Similarly, the availability of secondary data was limited, so again local economy effects or effects on non-participants could not be assessed using such data. Finally, the focus of our qualitative research was not on topics related to relevance. However, it is clear that our qualitative research did reveal some findings about issues related to spillovers, non-participants, and changes in the local economy. These were highlighted in Volume I of the Final Evaluation Report as well (OPM 2016a).

The tables below delineate, explicitly, the specific sub-questions that this evaluation has addressed related to the different OECD DAC criteria, and the different evaluation components that addressed these questions. Note that these questions differ sometimes from the original TOR, given that extensive consultations were conducted during the inception phase on the relevance and evaluability of certain questions.

Please note that, in particular, quantitative evidence related to effectiveness was also presented in the descriptive statistics report, the executive summary to which can be found in Annex D.

Table 56 Questions related to effectiveness

No.	Sub-question	Qualitative assessment?	Quantitative assessment?	Comments
1.	How many members of CPHHs met the graduation criteria developed by CLP-2?		Yes	
2.	What is the number of people from CPHHs who were lifted out of extreme poverty – as defined by the CLP-specific poverty line?		Yes	
3.	What are the major factors that drive sustainable graduation?	Yes	Yes	(Qualitative research looked into perceptions of sustainability of graduation and the institutions created.)
4.	What are the major factors that cause graduating households to become non-graduated – i.e. to fall back towards extreme poverty?	Yes	Yes	(Qualitative research looked into reasons for why households struggle to graduate or fall back into poverty.)
5.	How have livelihoods of CPHHs changed in the following areas: income, expenditures, savings, and assets?	Yes	Yes	
6.	How have nutrition practices changed among targeted mothers and adolescent girls?		(Yes)	Note that this was partly addressed in the quantitative descriptive report produced by OPM. It was decided, together with stakeholders, to not investigate this further.
7.	(How has the vulnerability of participating households changed?)	(Yes)	(Yes)	Note that we address vulnerability as a cross-cutting topic that affects other areas of interest. The descriptive statistics report does present results related to indicators such as plinth-raising, though.
8.	How have livelihoods and the local economic context changed as a result of CLP-2?	Yes		Note that this question was added at the inception phase.
9.	How have livelihoods and the local economic context changed with CLP-2?	Yes		
10.	How has empowerment, and in particular female empowerment changed with CLP-2?	Yes		

We define questions related to impact as ‘attribution questions’. Table 57 below explicitly lists these questions. Please also note that questions related to social capital, female empowerment, food security, and nutrition were included here in the inception phase – but were excluded after consultation with the client and stakeholders. This was partly based on the fact that our efforts were re-focused on the key areas of interest mentioned in Volume I. In addition, as mentioned above, nutrition and food security questions were excluded here given the limited evidence available and given the fact that DFID has implemented a separate evaluation of the direct nutrition intervention in Bangladesh. Female empowerment questions were excluded after the descriptive analysis phase, given the limited usefulness of the empowerment indicators that are available. Social capital questions were excluded after the descriptive statistics phase because annual survey data were of limited use to measure change, other than by looking at the participation in social development groups, which was addressed descriptively.

Table 57 Questions related to impact – all addressed quantitatively

No.	Sub-question	Comments
	What was the causal effect of CLP-2 on:	
1.	• Number of people from CPHHs lifted out of extreme poverty – as defined by an agreed-on poverty line?	Note that we implemented this using two poverty lines: consumption and asset poverty.
2.	• The poverty gap?	
3.	• Malnutrition, particularly for females and children under the age of five?	It was decided not to include this in the main summary report (Volume I), given that the evidence base on which the response would be based was very slim. We have presented results in the current report (Volume II), however.
4.	• How livelihoods of CPHHs have changed in the following areas: income, expenditures, savings, and assets?	

Table 58 Questions related to sustainability

No.	Sub-question	Quantitative assessment?	Qualitative assessment?	Comments
1.	To what extent is the graduation of people from CPHHs according to CLP's graduation criterion sustainable?	Yes	(Yes)	Note that the qualitative component focused on perceptions of sustainability.
2.	How sustainable are other observed impacts over time and what are the perceptions about future the sustainability of these impacts beyond CLP implementation phase? What are the major factors that drive sustainable graduation?	Yes	Yes	Qualitative research focused on perceptions of 'future sustainability'.

Table 59 Questions related to efficiency

No.	Sub-question	Quantitative assessment?	Qualitative analysis	Comments
1.	To what extent does CLP-2 represent good VfM?	Yes		Cost–benefit analysis.
2.	To what extent was the targeting appropriate?	Yes		
3.	What was the level of inclusion error?	Yes		
4.	To what extent did the programme target various social groups, such as the disabled and elderly?	(Yes)		Not explicitly addressed in Volume I – some descriptive statistics presented in the descriptive statistics report.

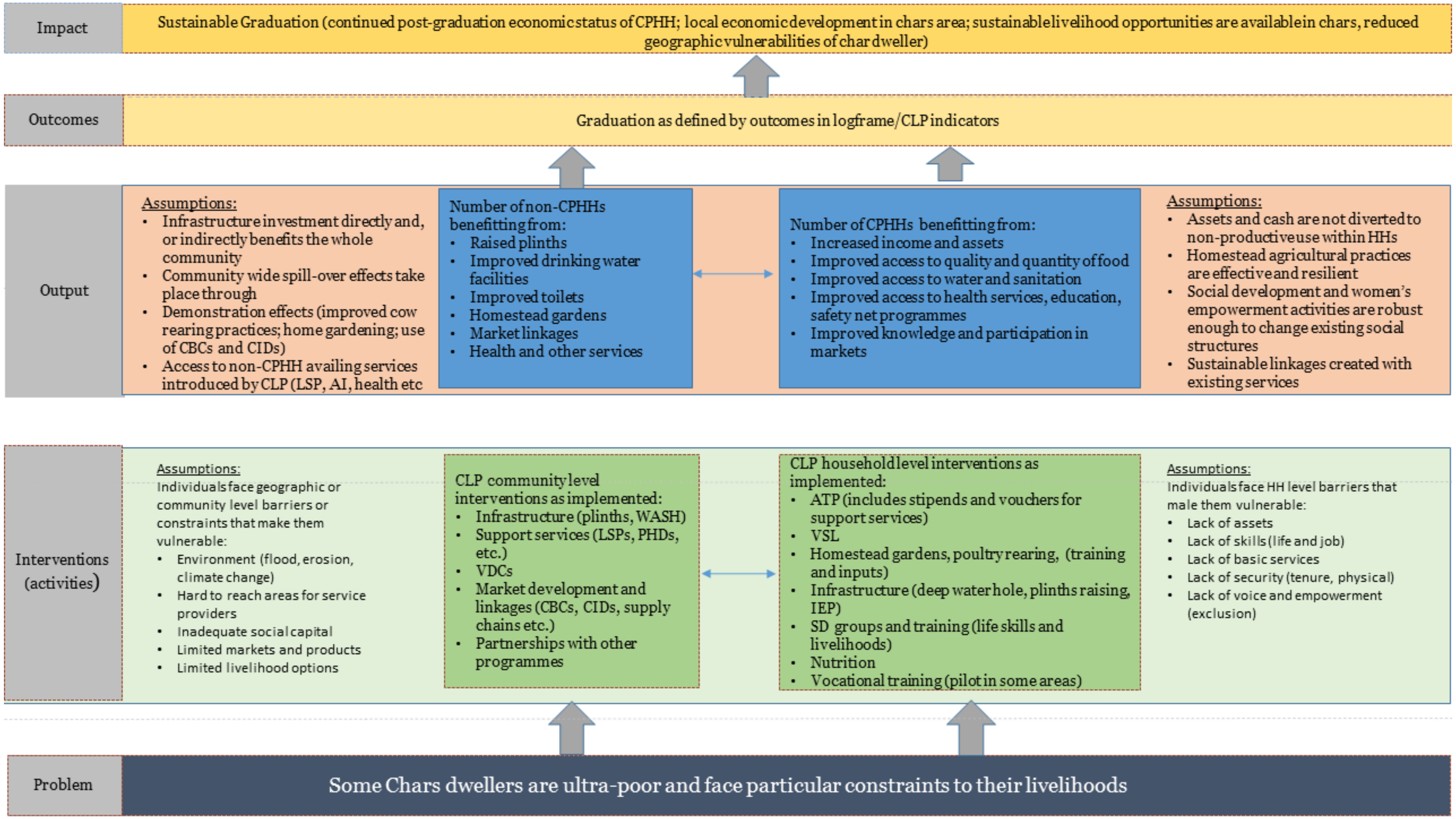
Annex C A light-touch TOC

As described above, CLP is focused on poverty alleviation in hard-to-reach, remote char areas of northern Bangladesh. Retaining the core element of an ‘asset transfer’ to extremely poor households, the programme has followed a ‘learning by doing’ approach in its programming during the 11 years of its existence. This means that the programme has evolved and new interventions and inputs have been added to the original design, as the debates in the livelihood discourse progressed and lessons from monitoring and evaluations of CLP itself emerged. For example, whilst CLP started as a mainstream livelihood programme, market linkages gradually became important components of its design in order to protect livelihood assets and increase the economic returns of the assets in the long run.

In order to better structure the evaluation design, a light-touch TOC was developed for this report. The TOC was developed on the basis of interviews with programme and implementing staff, the programme logframe, and the review of key programme documentation. **It is important to note that this TOC should not be considered as the programmatic TOC of CLP**. It does not explain the causal links within the different interventions that comprise CLP, nor does it make claims about whether causal links are strong or weak. **Rather, the TOC is an attempt by the evaluation team to better understand CLP and its many complex interventions, and it allows us to structure our thinking and design of the evaluation.**

Figure 4 below depicts the high-level TOC that was developed to guide the evaluation design. The diagram is based on the team’s understanding of CLP. Due to the complexity of the programme, certain aspects and interventions had to be simplified in order to be able to capture them. **The TOC divides interventions on the basis of whether they are aimed at the household or the community level.** The rationale for this decision is that char dwellers face both individual- or household-level barriers and constraints, as well as geographic or community-level barriers that expose their livelihoods to vulnerability and shocks. The underlying assumptions made by the programme at the intervention and output level are also specified below. The qualitative component of the evaluation will probe some of the underlying assumptions made by the programme, especially as they relate to the sustainability of graduation. In addition, the research will seek to better understand both community-level and household-level interventions.

Figure 4 A light-touch TOC



Our understanding of the TOC can be summarised in the following way:

A. Problem statement

In the dynamics of erosion and accretion in the rivers of Bangladesh, the sandbars that emerge as islands within the river channels, or as land attached to the riverbanks, often create new opportunities to establish new settlements and to pursue agricultural activities. Once vegetated, such lands are commonly called 'chars' – small islands in Bangla. Although the riverine chars in Bangladesh offer significant amounts of land for settlement and cultivation, living and working conditions there are very harsh due to intensive drought, flood and occasional erosion caused by river currents. Most residents of chars are either poor or extremely poor.

Char dwellers face certain constraints that are inherent in the environment in which they live, that hinder economic development and prevent them from accessing secure livelihoods. These constraints or barriers are present both at the community and at the household level and make them vulnerable to shocks. Household-level constraints are typically a lack of assets, job skills, basic services, and voice and empowerment.

In the char areas there are few fixed assets because every year flood and occasional erosion caused by river currents devastate the assets of char dwellers. As residents are dependent on agriculture, they also lack alternative livelihood skills and opportunities. In addition, as the chars are hard to reach, people living there only have access to very limited basic services, such as education, health and other safety net programmes. Because of low levels of education, skills and assets, poor char residents in general – and women in particular – possess inadequate social capital, voice and empowerment.

In addition to the household-level constraints, there are additional geographic or community-level barriers that make accessing sustainable livelihoods a challenge. These include frequent floods and erosion, a lack of presence of service providers, limited markets and products and severely restricted livelihood options. The absence of roads and a good transport system makes the area hard to reach, which affects economic development, service provision and mobility. During the dry season roads are sandy and passage is difficult, whilst the rainy season sees roads becoming flooded, with transport switching to boats. Despite the reliance on agricultural production, adequate agricultural inputs are hard to come by.

B. Programme input

CLP activities or interventions are designed to tackle the barriers or constraints char dwellers face both at the communal and household levels, thereby reducing their vulnerabilities to shocks and creating options for them to access sustainable livelihoods. Household-level programme inputs are the ATP (which includes stipends and vouchers for support services), VSLs, homestead gardens, poultry rearing, (training and inputs), infrastructure (deep water holes, plinth-raising, the IEP) social development groups and training (life skills and livelihoods), nutrition, and vocational training (piloted in some areas). Community-level inputs are infrastructure (plinths, WASH), support services (LSPs, Partners in Health Development, etc.), VDCs, market development and linkages (CBCs, CIDs, supply chains etc.), and partnerships with other programmes.

C. Programme assumptions

At the individual or household level, CLP assumes that if the assets and cash provided by the programme are not diverted to non-productive use within the households and the homestead agricultural practices are effective and resilient, social development and women's empowerment activities will be robust enough to change existing social structures, and if sustainable linkages are being created with existing service at household level, households will become less vulnerable and poor.

At the same time, it is assumed that at the community level, infrastructure interventions will benefit the whole community. Consequently, community-wide spillover effects will take place. Demonstration effects will lead to overall local economic development as the participants and non-participants will learn better cow-rearing practices, carry out homestead gardening and obtain and benefit from services and market mechanisms introduced by CLP (LSPs, artificial insemination, healthcare, CIDs, CBCs etc.).

D. Output

Outputs at the household level of the CLP programme are the number of CPHHs benefitting from increased income and assets, improved access to quality and quantity of food, improved access to water and sanitation, improved access to health services, education, safety net programmes, and improved knowledge of, and participation in, market linkages. The outputs at the community level are the number of non-CPHHs benefitting from raised plinths, improved drinking water facilities, market linkages, health and other services, improved toilets, and potential use of homestead gardens through spillovers. Finally, the overall output of the programme is reducing the household- and community-level vulnerability of inhabitants of the chars.

E. Outcome

CLP aims to graduate households out of poverty. At the beginning of 2014 the programme finalised a set of 10 underlying graduation criteria and a methodology to assess graduation. To graduate, a household must meet (any) six or more underlying criteria within three months of completing the 18-month cycle. These underlying criteria relate to: 1) income/expenditure/ consumption; 2) nutrition; 3) asset base; 4) status of females; 5) vulnerability; and 6) access to services.

The criteria are: 1) household has had more than one source of income during the last 30 days; 2) household eats three meals a day and consumed five or more food groups in the past week; 3) household has access to improved water; 4) household has access to a sanitary latrine with an unbroken water seal; 5) presence of ash/ soap near to water point or latrine; 6) productive assets worth more than BDT 30,000; 7) participant is able to influence household decisions regarding sale/ purchase of large investments, e.g. cattle; 8) homestead is above known flood level; 9) household has cash savings of more than BDT 3,000; and 10) household is a member of a social group.

F. Impact

CLP expects that after the completion of its interventions, sustainable livelihood opportunities will be available in chars and that overall local-level economic development in char areas will be visible, which will in turn reduce the geographic vulnerabilities of char dwellers. As a result, the programme expects the reduction in poverty levels to be sustained well beyond the time span of the interventions.

Annex D More detailed analysis of changes to the local economic context

As discussed in Volume I (OPM 2016a), the greatest variation across our research sites was observed in relation to changes that occurred in the local economy. It is therefore useful to present a more detailed and nuanced analysis of the findings as they pertain to this area of research, for readers who are interested in understanding a little bit more about the variation and what drives it.

D.1.1 Changes in the local economic context

The following section presents findings from the qualitative research on changes to the local economic context and char dwellers' engagement with markets and service providers. Given the nature of the quantitative data, this section relies exclusively on findings from the qualitative research. As explained in Section 2 of Volume I and in the section on the qualitative methodology in this volume (Section 5.3) our ability to generalise beyond the research sites is limited and this section instead seeks to provide a rich understanding of how and why changes have occurred and what dynamics are currently at play. In addition, the qualitative data on supply side interventions should be viewed as less reliable, as our sampling approach does not allow for extensive triangulation of views by key informants, which decreases the robustness of the findings (see Section 2.4 in Volume I).

D.1.2 Motivation and expectations

CLP attempted to change the local economic context in the chars through a combination of demand and supply side interventions – with the latter a less widespread and more recent addition to the CLP-2 programme. The existence of well-functioning markets matters because, if they are present, they can increase both the impact of the asset transfer and of the additional income available to households and communities, who are then able to sell their produce and products in a responsive market.[39] The idea was to improve demand for functioning markets through demand side interventions aimed at increasing the impact of the asset transfer through improved rearing and agricultural practices, whilst simultaneously ensuring that markets operate in the chars that allow people to sell at a fair price. The expectation was that CLP would sustainably change the way people farm and rear cattle, leading to higher production volumes and superior produce. It was expected that the increased availability of quality produce would in turn attract market actors and lead to changes in the local economy. In addition, market development initiatives were expected to create value chains that would enable char dwellers to produce for the wider economy. Addressing both demand and supply side constraints faced by communities would also increase the likelihood of achieving sustainable improvements to people's wellbeing (see Annex C for the TOC).

[39] CLP-2 Quarterly Report, January–March 2013 and KII interviews with IMOs at research sites.

D.1.3 Key interventions and their sustainability

D.1.3.1 Changes in farming and livestock rearing

The asset transfer and the training provided on how to employ better rearing and agricultural practices created demand for services amongst char dwellers. Expanded agricultural production led to increased demand for agricultural inputs, such as seeds and fertilisers, while improved rearing practices and provision of cattle led to demand for cattle and poultry feed and medical services for livestock, especially for CLP-trained LSPs. At the same time, the introduction of better rearing and farming practices increased the quality and quantity of produce and products available for sale on the chars

Participants in all locations mentioned diversification and improvements in the production of agricultural products as a key result of CLP training on improved farming techniques and the provision of quality seeds. The agricultural improvements were mentioned more widely by female participants, mainly in the context of the development of homestead gardens, which function as a source of income as well as providing households with the means to improve their nutritional intake. Female CLP participants were particularly appreciative of the new skills shared through CLP. For example, female CLP participants in all research locations mentioned an increase in the production volume of agriculture produce such as crops and vegetables due to their applying the techniques taught by CLP.[40] Additionally, CLP provided saplings and seeds as part of the homestead gardening intervention, which the women used to grow more vegetables for their own consumption, as well as to sell in the markets. While CLP did not explicitly push any crops, they did provide advice to the people on how to improve the farming of the crops already being cultivated. Female CLP participants in four research locations explicitly attributed the increased vegetable production to having been taught how to prepare soil beds and being provided with appropriate amounts of fertilisers, as well as using improved irrigation techniques.[41]

However, the research sites differ in the opportunities for agricultural production provided by the land and their availability to women. In cohort 2.2 high and low performing, the vegetable cultivation has increased but there is a greater focus on crop cultivation due to the availability of fertile soil, which allows households to make higher profits from crop cultivation.[42] This has led to shops being set up on the chars that sell seeds and fertilisers, which in turn has further improved access to inputs for farmers.

> *'Now people can grow a huge amount of crops. There are plenty of corn, nuts, and jute. There are some shops selling fertilisers and seeds in this char. For this reason people can get the seeds and fertilisers easily and they can also purchase this on credit.'* [FGD with female participants in 2.2 low performing site]

---

[40] FGD with female CLP participants in cohorts 2.2, 2.4, and 2.5 high and low performing sites.

[41] FGD with female CLP participants in cohorts 2.2, 2.4, and 2.5 high and low performing sites.

[42] FGD with female CLP participants in cohort 2.2 low performing and FGD with men from CLP participant households in cohort 2.2 high performing sites.

In contrast to this, in cohorts 2.4 and 2.5 low performing the soil boasts a high sand content, which limits the variety of crops that can be cultivated. This was one of the factors mentioned by CLP participants as making it harder for them to improve their livelihoods through farming. However, CLP participants also reported that CLP provided them with training and seeds for crops which were tailored to their environment, thereby enabling them to expand and build their livelihoods through farming, despite the less favourable soil.[43] Female CLP participants in cohort 2.5 high performing tended to focus on rearing livestock – especially poultry – and less on farming.[44] A possible explanation for this could lie in the comparatively strict cultural and social norms in the area, which make it harder for women to engage in agricultural activities such as crop cultivation.

Furthermore, **improved techniques and knowledge about agricultural practices has also enabled CLP participants to cultivate a larger variety of vegetables and crops**. Multi-cropping and crop rotation was not previously widely used in the chars.[45] These changes in farming techniques were repeatedly mentioned as a major factor which contributed to the increase in production volume and enabled households to achieve subsistence in consumption whilst at the same time producing a surplus that could be sold to traders and at markets.

> *'I did not use to sell gourd but this year I sold gourd of 300 tk. I could cultivate more gourds because of the training received. In the past, I was able to produce enough for consumption, but not enough to sell. Because of CLP I can now sell part of my harvest and gain an income.'* [FGD with men from a participant household in 2.5 low performing site]

The adoption of multiple harvest cycles per year increased the income of farmers, which has led to improvements in livelihoods. Higher earnings from farming was mentioned more frequently by male respondents. While women are often responsible for the cultivation and harvest, men are typically still in charge of selling the produce, especially if they are sold directly to markets or sold directly from crop fields.[46]

Most CLP participants in our research sites chose cattle as the asset provided by CLP. Within the budget provided by CLP, CLP participants had the option of purchasing cows, bulls, goats or poultry – depending on the livelihoods they wished to pursue and household composition and dynamics. In cohort 2.2 low performing and in cohort 2.4 high and low performing CLP participants preferred cows and wanted to sell milk. In cohort 2.5 high performing, cohort 2.2 high performing and cohort 2.5 low performing, female CLP participants reported preferring to purchase bulls for fattening and selling in the market. In cohort 2.2 high performing the main

[43] FGD with female CLP participants in cohort 2.4 low performing site and KII with IMO in cohort 2.5 low performing site.

[44] FGD with female CLP participants in cohort 2.5 high performing site and KII with IMO in cohort 2.5 high performing site.

[45] FGD with female CLP participants in cohorts 2.2 and 2.5 low performing and in cohort 2.4 high and low performing sites. FGD with men from CLP participant households in cohort 2.2 high performing and cohort 2.5 low performing sites. FGD with female non-CLP participants in cohort 2.5 high performing site.

[46] FGD with men from CLP participant households in cohort 2.2 high performing, cohort 2.4 high performing and cohort 2.5 low performing sites.

livelihood activities pursued are farming and women are involved in tilling the land. As a result, most CLP participants prefer raising bulls, which they are able to fatten and sell in a shorter period of time – after which they invest the money in land, thereby further improving their main income-generating strategy.[47] Male CLP participants in cohort 2.5 low performing expressed a preference for bull fattening due to the shorter turnaround and the higher profit that can be made from raising and selling bulls.[48] In all areas, both male and female CLP participants mentioned higher returns from their cattle and poultry after adopting the rearing techniques taught by CLP. In areas where CLP participants opted for cows, milk production increased as a result of the rearing practices promoted by CLP. CLP participants highlighted the importance of timely and adequate feeding of cows in the form of fresh grass and green fodder for increasing milk production volume and quality.[49] In addition, the existence of the CLP plinths reduced the exposure of cattle to waterborne diseases during flooding periods, further improving the productive value of the asset.

D.1.3.2 Access to LSPs and willingness and ability to pay for their services post-CLP

LSPs were working in all our research sites, and in all locations but one more than one LSP operated. As CLP participants were initially incentivised to make use of the services of LSPs through the provision of vouchers provided by CLP, a key research question was whether CLP participants would be willing and able to pay for their services following the phasing out of CLP-2. **In all of the qualitative research sites CLP participants and non-CLP participants continued to use the services of LSPs, even after CLP ended.[50]** In the high performing chars of cohorts 2.2, 2.4 and 2.5 the qualitative research found strong evidence of behavioural change in respect to utilising the services of LSPs amongst both men and women from CLP participant households, who continued to pay for both curative and preventive services even after CLP-2 ended.[51] Char dwellers also reported having a good relationship with the LSPs, which they defined through the quick response times from LSPs when called upon, as well as reasonable fees and prices for medicines and services, along with the provision of services on credit. LSPs and char dwellers reported that people were charged according to their ability to pay. In addition, some respondents felt that LSPs provided better services if they resided or hailed from the area or village in which they were working. The willingness to pay for preventive and curative services for animals in the high performing chars, as well as the efficiency and commitment of LSPs, played a large role in improving the wellbeing of households.

> *'The relationship with veterinarians is very good. They come if we phoned them and tell them about a problem. In the past, cows used to die due to a lack of treatment. But after*

[47] FGDs with women and men from CLP participant households in cohort 2.2 high performing.

[48] FGD with men from CLP participant households in cohort 2.5 low performing.

[49] FGD with female CLP participants in cohort 2.4 low performing. FGD with men from CLP participant households in cohort 2.4 high performing.

[50] FGD with female CLP participants in cohorts 2.2, 2.4, and 2.5 high and low performing. FGD with men from CLP participant households in cohorts 2.2, 2.4 and 2.5 high performing. KII with LSPs in cohorts 2.2, 2.4, and 2.5 high and low performing.

[51] FGDs with female CLP participants in cohorts 2.2, 2.4, and 2.5 high performing. FGDs with men in CLP participant households in cohorts 2.2, 2.4, and 2.5 high performing.

> *CLP, cows die less often because they can be treated in time. At present the veterinarians are paid for medicine. If people don't have money in hand instantly that they will pay later. If they don't have money they can pay after two or three days.'* [FGD with men from CLP participant households in 2.4 high performing site]

In the low performing chars of cohorts 2.2, 2.4 and 2.5 female CLP participants reported that they continued to use the services of LSPs post-CLP. However, these respondents also reported that since the end of the CLP voucher scheme they could no longer afford preventive services, such as vaccinations, at all times and hence only contacted LSPs when their cattle were sick.[52] There was some evidence that some respondents had not understood that the voucher system meant that LSPs were not CLP employees and would have to be paid for their services. One LSP reported having to convince CLP participants that he did not receive a CLP salary and was dependent on them paying for his services in order to be able to continue visiting the chars. Nonetheless, CLP participants have internalised the importance and benefits of timely preventive and curative services and hence continue to pay for the services of LSPs despite the termination of the voucher system.[53]

In cohort 2.5 low performing female CLP participants said that the LSP did not respond or visit as frequently as during CLP and had also started charging higher fees, which they were unable to afford.[54]

> *'During CLP the CLP LSP was available regularly. However, now we do not see him here. We used to have vouchers and the LSP announced at the mosque that he can treat cows and provide vaccinations in exchange for vouchers. Now [that the programme is over] he does not want to respond to the call of char people. He asks us to instead bring the cows to his practice on the mainland. We are able to purchase deworming or similar types of medicines at the pharmacy but we need the LSP for vaccinations. If the CLP LSP does come, we have to pay high visiting fees of around 150/200 taka. If we call another LSP working here he comes and we have to give him 400/500 taka. In addition to this fee we also have to pay for medicine. Due to CLP training, we now know about several diseases of cows. So after, seeing any symptom of disease we contact the LSP as fast as we can.'* [FGD with men from participant households in cohort 2.5 low performing site]

Char dwellers in all locations showed an increased awareness of the diseases their cattle could contract and the associated symptoms, allowing them to call LSPs before their cattle got seriously ill. The LSPs were trained not just to provide the services, but also to educate the CLP participants regarding the symptoms of diseases and also regarding better care techniques for the animals.

[52] FGD with female CLP participants in cohorts 2.2, 2.4, and 2.5 high performing.

[53] FGDs with women and men from CLP participant households in cohorts 2.2, 2.4, and 2.5 high performing. FGDs with women and men from non-participant households in cohorts 2.2, 2.4 and 2.5 high performing.

[54] FGD with female CLP participants in cohort 2.5 low performing.

> *'The improved health and reduced mortality of cattle has been extremely impressive and has also shown lasting impact and strong spillover effects.'* [KII with CLP Manager of Markets and Livelihood Unit]

Finally, **LSPs in the research locations also benefitted greatly from the CLP training and the access to new groups of clients.** LSPs claimed to be able to provide better service as a result of training on diseases and medicines. The expansion of their customer base has enabled them to improve their income and livelihoods, which has further acted as a catalyst for providing even better services to the char dwellers. LSPs in some of the high performing locations mentioned that they would like to receive more training in order to be able to provide additional services.[55] However, **the accessibility of chars remains a challenge for LSPs**. In the low performing locations, LSPs do not always provide timely service to the chars, due to the difficulty of reaching them and because it is at times more profitable to stay and work in the mainland market.[56] LSPs across all locations reported poor road and communication linkages as a major hindrance of providing timely services to char dwellers.[57]

D.1.3.3 Access to and engagement with traders and input dealers

In theory, demand side interventions will eventually also lead to changes in markets, as traders become aware of the existence of high quality products. **In our research sites, a mixed picture emerged as to whether this was the case.**

In three of our research sites, access to and engagement with traders underwent a significant change after the advent of CLP. Women and men from CLP participant households in cohort 2.2 high and low performing and in cohort 2.4 high performing mentioned a substantial increase in the number of traders and wholesalers visiting their chars in order to buy their crops. This change was attributed to the increased quality and volume of goods produced.[58] As a result, men from CLP participant households in two chars reported being able to bargain for better prices due to the increased competition between traders and wholesalers. In addition, farmers found it easier to access agricultural inputs due to sellers of seeds visiting the chars and small shops opening up on the chars themselves. Furthermore, milk collectors were now visiting the chars in larger numbers. However, female CLP participants in two chars complained that while they were able to sell their milk at a higher price than before CLP, they were still forced to sell below the price they could have earned at the market. Whilst CLP participants lamented this fact, they were nonetheless aware of the time and money saved by not having to carry their products to the market.

The remaining three research sites were located closer to markets, which counter-intuitively seems to have led to a scenario where fewer traders and wholesalers service the chars, leading to lower returns on the increased and improved produce produced. As the chars in cohort 2.4 low performing and in cohort 2.5 high and low performing are located

[55] KII with LSPs in cohorts 2.2, 2.4, and 2.5 high and low performing.

[56] FGD with men from CLP participant households in cohort 2.5 low performing.

[57] KII with LSPs in cohort 2.2, 2.4, and 2.5 high and low performing.

[58] FGDs with women and men from CLP participant households in cohort 2.2 high and low performing and in cohort 2.4 high performing.

nearer to markets, char dwellers tend to transport their products to the markets on the mainland, where they sell them. As a result, there are very few traders who visit the chars – and they only do so if substantial surplus has been produced.[59] In these chars char dwellers seemed to focus on smoothing their consumption. The qualitative data suggest that in these cases the lack of traders who service the chars meant that competition for produce was low, leading to lower prices and diminished bargaining power. Whilst bringing produce to markets can lead to higher prices, it can also have the reverse affect when lower prices are accepted in order to avoid having to transport produce back to the chars.

Whilst the number of actors in the market affects the bargaining power of char dwellers, CLP also taught CLP participants to actively negotiate prices. CLP participants were provided with information on market prices and reported bargaining with traders. Some traders also reported that char dwellers were now demanding higher prices, which they were mostly willing to pay due to the improved quality and quantity sold. One milk collector explained that he could afford to pay higher prices for the milk as he no longer had to service several chars in order to collect the necessary quantity of milk.

D.1.3.4 CLP's market development initiatives

Whilst CLP has always included an 'enterprise development' component that sought to create additional income-generating activities for core participants, the decision to deliberately focus on supply side market linkages was only taken in 2012, when the M4P approach was adopted as the guiding framework for CLP's market development component. The CLP-2 Quarterly Report, January–March 2013 states: '*The emphasis…was on securing and sustaining the gains which have been achieved through CLP's asset transfer project to strengthen livelihoods. As the majority of households select cattle as their preferred asset, the CLP strategy was to facilitate change using M4P principles in mainly three livestock related market sectors in which char households currently operate. These are the dairy cow milk market, meat market (mainly beef, but also goat, sheep and poultry), and the associated fodder market.'*[60] As a result, the coverage of the market linkages initiatives is limited and only covers a fraction of all of the areas in which CLP was active.

IMOs in all of our research sites said that more widespread and deliberate supply side interventions would have increased the impact and sustainability of CLP. The view commonly expressed was that functioning markets would have enabled the development of truly sustainable livelihoods for char dwellers who would then have been in a position to benefit from these markets due to their improved production capacity and bargaining skills. The lack of widespread attempts to change markets was listed as a shortcoming of CLP.

> *'As char area is situated in the remote area, we hardly found bazaars surrounding the char. If a small bazaar is found, necessary goods are insufficient there. Cow, poultry farm, and vegetable cultivation methods have been taught from CLP. So, bazaar was needed for selling produced product and for buying animals' food. At present, char*

[59] FGD with female CLP participants in cohort 2.5 low performing.

[60] CLP-2 Quarterly Report, January–March 2013.

people have been connected with bazaar even though bazaar hasn't been established directly.' [KII with IMO in 2.4 high performing site]

'Absence of market development is a big challenge. Because of this they sold their product or material into the char's bazaar with a very low price. If there exists market linkage then they can sell it to the town and can make a large amount of profit'. [KII with IMO in 2.4 low performing site]

FGDs with male and female CLP participants and non-CLP participants echoed these findings. In our research sites, char dwellers expressed dissatisfaction with market characteristics and their ability to access markets, even in cases where the local economic context had changed as a result of increased demand.

Where active, the M4P market development framework was used to develop the milk, meat and fodder markets in selected areas and cohorts.[61] **The idea was to facilitate the development of a value chain extending from producers all the way to the private sector that links char dwellers to markets.** This value chain was meant to be underpinned by the understanding that a mutually beneficial business opportunity existed that could improve the livelihoods of all actors along the value chain. The creation of CBCs was promoted, as well as of CIDs. The aim was to improve the capability of producers, dealers and traders and to promote organised production groups that were trained to engage with and identify business opportunities. KIIs with implementers of the M4P approaches reported that these initiatives proved to be successful and sustainable, even though their reach was limited due to the geographical restriction of the market linkages component of CLP2.[62] Whilst the approach described above had initially been designed for the meat market, it was eventually also adopted for the milk markets.

'The opportunities in the market are identified through linkages that we create between those production groups and local-level actors that are providing the inputs or the output marketing for whatever is being produced. We then combine that with relatively light-touch engagement around broader government structures, collective action structures or CBC structures. That was relatively light-touch in the sense that we weren't aiming for extreme formalisation. More so that there were spaces for governance around the producer groups that were operating there.' [KII with iDE]

Key informants reported that they were successful in persuading traders to travel to the chars and collect and source produce, whilst CBCs facilitated the supply to the large market actors which had been attracted to the chars as part of the M4P market linkages components. Several large private sector companies are now active in the sites where the market linkages components were rolled out and to date the change seems sustainable, as companies continue to conduct their business and have started to expand to surrounding areas, even after CLP phase-out.[63]

[61] As discussed in Section 2.4 our sample did not include chars on which the M4P components of CLP were active.

[62] KII with CLP Manager: Markets and Livelihood Unit. KII with iDE.

[63] KII with CLP Manager: Markets and Livelihood Unit.

In addition, some initiatives led to unexpected outcomes. Whilst the fodder market was a commercial failure in the sense that CLP participants initially used the produce to feed their cattle in an attempt to produce the highest possible volumes of milk and meat, it can also be viewed as a success as it provided char dwellers with excellent fodder for milk and meat production and the potential for future sales to the mainland.

> *'… [W]e were hoping to a see a little bit of direct commercialisation, the chars producing it [fodder] and selling it to the mainland. We found that instead char producers very quickly in terms of meat and milk production, that they wanted to take as much as these green fodder that they were producing and fold it directly into their cattle to make sure they were producing at the highest volumes possible of meat and milk. For us, from an economic standpoint, this was an even better outcome if you're looking at equity around the chars themselves. We've already seen that kind of idea of selling out to the mainland was starting in a couple of areas because the volumes were growing so large, you didn't have enough demand for actually the own consumption of cattle. For us, we considered that a success in the fodder side as it was working fairly sustainably.'* [KII with iDE]

Overall, CLP-2's market development initiative led to the establishment of 120 milk producer groups – organising roughly 3,000 producers – and the formation of 70 CBCs, 19 of whom cater for both the milk and meat and fodder markets. Almost all of the CBCs remain fully functional and 25 CBCs have now been officially registered, while 19 are in the process of registering.[64] In addition, linkages were made with other M4P programmes in the chars. CLP worked with SwissContact's M4C programme, which uses United Finance to create seasonal loan products specifically tailored to the needs of the char people. As part of this project, United Finance has pledged to provide loans totalling BDT 5 billion over four years, to a total of 15,000 char residents.[65]

D.1.4 Women and markets

The qualitative research found that **women typically do not visit markets on the mainland**. Men and women felt that if women were to visit markets they would be breaking with existing social norms.[66] Some men cited the poor condition of roads as a justification for why women were unable to access markets. Respondents viewed men as responsible for purchasing all necessities of the households, including personal items of clothing for women. However, some women reported now feeling empowered to ask their husbands for specific items, whilst they had previously merely accepted their husband's choices.[67] **On the chars themselves, substantial changes have taken place and women now actively engage with traders visiting the chars and small shops or markets on the chars.** For example, women engage with traders when selling their produce to traders and milk collectors, including being

---

[64] KII with CLP Manager: Markets and Livelihood Unit.

[65] KII with SwissContact.

[66] FGDs with women and men from participant and non-participant households in cohorts 2.2, 2.4, and 2.5 high and low performing.

[67] FGDs with female CLP participants in cohort 2.5 low performing and cohort 2.2 high performing.

responsible for bargaining over the price of the produce.[68] It therefore seems that women have started to engage with markets, but that this engagement is limited to market actors that visit the chars.

> *'The women from chars do not go to bazaars, as they feel shame. If there are no men in the house they send children or neighbours to make purchases or go shopping. This practice has not changed and is the same as before. But they do go to the bazaar or hospital in emergencies like sickness or to meet with a doctor or buy medicine from a pharmacy…Moreover during Eid or other festivals some women go to market. Except for these occasions, they do not need to go market or bazaar.'* [FGD with female participants in 2.5 low performing site]

D.1.5 Spillover to non-CLP participants

As discussed, CLP participants widely reported adhering to the livestock rearing practices and the improved farming techniques promoted by CLP. Similarly, the services of LSPs continued to be used – although with an emphases on curative rather than preventive services.

The qualitative research found strong evidence that non-CLP participants had adopted the farming and livestock rearing practices promoted by CLP. Given the perceived success of these interventions it is perhaps unsurprising that non-CLP participants chose to adopt the new practices. For example, female non-CLP participants in cohort 2.2 high and low performing and in cohort 2.5 high performing mentioned emulating CLP homestead gardening and rearing practices for cattle and poultry practised by the CLP participants. They also reported having benefitted from increased production of, and income from, their produce, which led to an improvement in their livelihoods.[69] Non-CLP participants explained that they either copied the behaviour of CLP participants or listened in during the training sessions. In addition, non-CLP participants were also able to benefit directly if the number of traders and wholesalers servicing a char improved, as they too were now better able to sell their produce.

Similarly, non-CLP participants reported making use of the services of LSPs. However, non-CLP participants from chars that were doing relatively worse in terms of changes in the wellbeing of residents reported often not being able to pay for the services of LSPs or having to opt only for curative services when the cattle had a serious illness, due to issues of affordability.

> *'Those who are not CLP members go to … another doctor if their cows are ill. If they call CLP doctors they intend to visit the member's cows first then others so that they call other doctors rather than wait. If not a major problem they just say it to the doctors and take medicines but if major they call the doctor home. If the cows get problems or if people come to know that any cow died in other villages they call the doctor and give vaccines. Otherwise they do not go to the doctor.'* [FGD with men from non-participant households in 2.5 low performing site]

---

[68] FGDs with female CLP participants in cohort 2.2 high and low performing sites.

[69] FGDs with female non-CLP participants in cohort 2.2 high and low performing sites and in cohort 2.5 high performing site.

Annex E Imputing aggregate consumption and constructing the asset index – Methodological annex

In this annex we present the procedures employed to impute a measure of aggregate consumption and to construct an asset index, both of which are comparable to HIES data.

E.1 Imputing consumption

In the first step, we use a standard linear regression framework to estimate a model of consumption using the HIES 2010 data:

$$Y = \alpha + \beta_1 A + \beta_2 F + \beta_3 C + \beta_4 D + \varepsilon$$

The dependent variable (Y) is total household consumption in the HIES dataset, expressed in 2014 prices in order to match the prices of the CLP dataset. Independent variables include four broad categories: ownership of select assets (A); food consumption (F), measured in the number of days key categories of food had been eaten in the seven days prior to the survey; non-food consumption (C); and demographic characteristics (D).

Several considerations guided the choices of these categories:

- variables had to be present in the HIES data and in CLP datasets;
- variables had to be collected in all rounds of the CLP surveys;
- they had to be defined consistently across all CLP rounds and in the HIES survey; and
- variables had to behave reasonably consistently in the HIES and CLP data. For this reason, a model where the explanatory variables would be monetary values of purchased food consumption was decided against since the share of purchased goods in total consumption on the chars is likely significantly less than in the HIES data overall. This is for three main reasons: the extreme poverty rates on the chars, the prevalence of subsistence farming, and the inaccessibility of the chars. Furthermore, due to the proximity of Eid ul-Adha to the CLP survey data collection period, extreme variability of food expenditure values might significantly increase the error of imputed consumption, even if the model doesn't exhibit any problems in the HIES data. Therefore, we preferred to employ variables that express the actual intake of food over the week prior to the survey as explanatory variables related to food consumption.

The resulting models are balanced in terms of various types of correlates of consumption, i.e. asset ownership, value of consumption of select goods, demographic characteristics of households, and food intake, and can be expected to be less influenced by seasonality. See Table 60 below for the entire list of variables included.

We estimate the model above using two subsamples of the HIES data: (a) the sub-sample of rural households in Rangpur only; and (b) the sub-sample of rural households whose per capita consumption does not exceed 20% more than the consumption at the upper rural poverty line. In the first case coefficients reflect consumption patterns specific to rural Rangpur. In the second case, the coefficients estimated will predict best the consumption in the bottom half of the rural population of Bangladesh. Regression (b) also includes division dummies to control for the effects of geographical location at the level of the division. Table 60 below presents the results for both models.

Table 60 Consumption regression results

Independent variables	Coefficient estimates	
	Specification (a)	Specification (b)
Non-food consumption		
Transport	2.624544***	1.118390***
Cosmetics	8.013490***	4.415195***
Tobacco	1.974194**	1.246780***
Food consumption: N days in last week		
Pulses	190.429837**	116.317162***
Milk	89.726769*	77.459394***
Meat	388.528794**	489.695538***
Poultry	781.979172***	468.301463***
Egg	-42.766726	185.086663***
Fish	289.307736***	221.037949***
Potato	12.894166	-7.960417
Vegetables (other than potato)	-15.281364	78.827253***
Sugar	242.064103***	68.003267***
Fruit	312.393787***	67.833375***
Assets		
Cattle	381.652769***	150.899391***
Fishing net	480.51332	-130.282252
Bicycle	20.947073	231.100542***
TV	741.367822	480.367168***
Mobile phone	1512.904611***	503.349376***
Dining furniture	66.975161	43.243847**
Bedroom furniture	310.646469***	41.629047***
Sewing	345.522155	54.760037
Kitchen items	-7.367821	24.883081***
Demographic		
Household size	708.329416***	1200.992402***
N children under five	-106.379672	-298.977399***
Female head	84.853695	10.161465
Division dummies		
Division code 20	n.a.	814.385867***
Division code 30	n.a.	174.687318***

Division code 50	n.a.	487.432096***
Division code 60	n.a.	83.924987
Constant	-984.443065*	-690.830713***
R-squared	0.752579	0.842088
N of observations	896	3670

Most of the estimated coefficients presented in the table above have intuitive signs and magnitude. However, some coefficients on select assets and food items have negative signs. In the multivariate context in which we are operating, this is a normal occurrence since each coefficient represents a variable's contribution to explaining consumption, holding all other variables constant.

Both specifications predict consumption well (in-sample) as indicated by the high R-squared, although model (b) performs significantly better that the rural Rangpur model. Table 61 below examines the performance of the two models in predicting aggregate consumption of the population of interest, i.e. the rural poor in Rangpur. It shows that specification (b) outperforms specification (a) among the poor (both the using the lower and upper poverty lie), i.e. its higher R-squared identified above is not driven by wealthier households. In addition, Figure 5 sows that the distribution of predicted consumption is more compact for model (b) than for model (a).

Table 61 Correlation between predicted and observed total household consumption among rural households below the lower and upper poverty lines in Rangpur

	Poverty line	
	Lower poverty line	Upper poverty line
Specification (a)	0.8809	0.8778
Specification (b)	0.9388	0.9248

Figure 5 Observed and predicted consumption using specification (a) and (b)

Table 62 below shows the proportion of correctly and incorrectly predicted poor households when using the coefficients estimated above for models (a) and (b). This test is run using the full HIES data and the lower poverty line.

As can be seen, model (b) identifies 73% of lower poor (by full consumption aggregate) correctly in the HIES data, while model (a) identifies 69% of lower poor correctly.

Table 62 Share of observed consumption poor identified correctly or incorrectly by the two consumption models

	Lower poverty line	
	% of poor identified correctly	% of poor identified incorrectly
Specification (a)	**68.95**	31.05
Specification (b)	**73.49**	26.51

The results show that specification (b) performs better in terms of within sample fit (R-squared), in terms of predicting aggregate consumption for the rural poor in Rangpur, and in terms of correctly identifying extreme poverty. In addition to displaying an overall better fit and a better fit among the poor, it also performs estimation over a much larger sample (3,670 versus 896 observations), which improves the stability of coefficients.

Hence, specification (b), which estimates coefficients using the sample of rural poor with consumption levels not exceeding 20% over the upper poverty line, will be the preferred specification.

The coefficients estimated using specification (b) are then taken to the CLP data and employed to predict aggregate consumption there. Using this predicted consumption, we can estimate poverty rates that can be compared to HIES data.

E.2 Asset index

The asset index represents another way of capturing household welfare by aggregating data on the ownership of various assets. While there are various ways of constructing an asset index – the simplest being a count of all assets owned by a household – we compute a principal components-based index, which aggregates different assets using weights that reflect their importance in the overall ownership of assets.

An asset index is a commonly used, non-monetary welfare measure.[70] Unlike consumption, which is sensitive to variations in prices, seasonality, disposable income, etc., the asset index captures accumulated wealth levels and access to productive assets, which reflects households' long-term wellbeing prospects, due to its ability to handle shocks and generate income. Low values of the asset index signal chronic poverty.

The output below shows the weights assigned to individual variables making up the assets, using principle component analysis. The weights are based on the first principal component factor. All the factor loadings have positive values, suggesting that the factor captures their relative contributions to welfare. Note that these weights are calculated using HIES data and will be transferred to CLP data to construct the asset index.

Table 63 Weights for variables included in the asset index

Variables	Factor loadings	Variables	Factor loadings
Cattle	0.16095	Bicycle	0.1629
Goats or sheep	0.11271	TV	0.14034
Poultry	0.01772	Mobile	0.16287
Plough and yoke	0.12918	Dining furniture	0.01695
Tube well	0.18421	Bedroom furniture	0.16694
Sprayer	0.15228	Sewing	0.02032

[70] For instance by the Demographic and Health Surveys.

Country boat	0.11567	Kitchen items	0.10549
Fishing net	0.13294	Power equipment	0.13202
Radio	0.06113		

We consider only one specification of the asset index calculated for all rural Rangpur households. Therefore, when the asset index is constructed in the CLP sample using the weights above, the population of reference will be rural households in Rangpur. The strategy we pursued earlier in estimating consumption for the poorer (in terms of consumption) subset of the population to increase accuracy does not work for an asset index because it is not linked to consumption. While it may produce asset weights that more accurately capture the composition of assets among the poor, in the absence of a common metric (e.g. consumption) it will only allow us to rank the poor, but it will not allow us to rank the poor vis-à-vis the wealthy.

For the asset index we set the poverty line at the 30th percentile (-.5647719) of the rural Rangpur distribution. This poverty line is arbitrary but roughly corresponds to the rural Rangpur lower poverty rate of 29.5% – thus we consider the correspondence between the poorest 30% of the sample in the HIES data by consumption and asset poverty. Table 64 demonstrates that approximately 54% of consumption poor are also asset poor, i.e. face chronic poverty, while 46% display characteristics of transient poor – that is, their consumption is below the poverty line but the fact that they own assets can suggest that they have means of improving their welfare, either by putting their assets to productive use or by liquidating them.

Table 64 Share of asset poor by consumption poverty status

		Asset poverty		
		Non-poor	Poor	Total
Observed consumption poverty	Non-poor	79.78	20.22	100
	Poor	46.35	**53.65**	100
	Total	69.97	30.03	100

Annex F Descriptive report: Executive summary[71]

Introduction

CLP-2 aims to reduce extreme poverty across communities that live on temporary sand islands ('chars') in northern Bangladesh along the rivers Jamna, Padma, and Teesta, with the aim of reaching up to one million extremely poor people in the region. Building on existing data and research reports, DFID has contracted OPM to implement an evaluation of CLP-2.

This report is the first output that has been produced in the context of this evaluation. It serves three main purposes: first, to present estimates of key indicators related to CLP activities and hence make a preliminary descriptive assessment of CLP-2 performance; second, to further inform the qualitative and quantitative impact analysis research that will be implemented within the context of this evaluation; and, finally, to present the approach that the evaluation team will adopt with respect to analysing the poverty status of participant households.

This report has to be seen in conjunction with other activities being implemented within the context of this evaluation: it is the first deliverable produced within the implementation phase of this evaluation and is the result of an inception phase that involved close collaboration with local partners. It also serves as one of the main quantitative inputs into the qualitative research. Finally, results were discussed with local partners in a workshop held at the end of January 2016 in Dhaka.

The specific target audience of this report includes the implementing partner – the CLP-2 team and Maxwell Stamp – on the one hand, and the OPM evaluation team – DFID, and DFAT – on the other hand.

The data used in this report

The key datasets used for this descriptive analysis are the CLP-2 annual survey datasets (2010–2014). These datasets contain data from surveys that covered CLP-2 participants from when the programme started, i.e. early 2010. There were two survey rounds in 2010 (May and October) and then a yearly survey round from October 2011 through to 2015 (2015 data were not available at the time of writing). The CLP survey followed a household panel structure. For each cohort that received CLP assistance (there were six cohorts in total), a random sample of households was drawn and data were collected from that sample on a yearly basis, with a baseline data collection exercise for each cohort before the intervention started. In total, about 2,500 CPHHs have been covered across all cohorts. All of the analysis conducted for this report was conducted using sampling weights.

Results

Sample attrition

Before presenting substantial results we analyse whether sample attrition, i.e. the dropping out of households from the survey, could be a concern for our analysis. We show that, depending on the cohort looked at, 5% to 16% of households within cohorts dropped out of the sample

[71] This is the executive summary from OPM (2016b).

between baseline data collection and October 2014. In addition, comparing the households that dropped out to the households for which follow-up data are available showed that, in the majority of dimensions analysed, these two groups of households do not differ significantly from each other. Overall, we conclude that the potential for bias due to attrition is low.

Demographics

We also compare cohorts across a group of demographic variables, in order to assess whether these cohorts differ systematically in regard to some important background characteristics. The variables looked at were household size, number of children in households, dependency ratio of households, the proportion of female-headed households, and the average age of household heads.

We conclude that no single cohort appears to significantly differ from all the other cohorts on all dimensions that were analysed. However, there are differences between individual cohorts in some dimensions, and some of these are quite significant. These systematic differences could, in theory, introduce bias when naively comparing outcome indicators across cohorts in order to infer programme impact. The results therefore emphasise the need for a more comprehensive impact assessment, where we control for systematic differences in covariates that are potentially related to the outcome measure.

Aggregate graduation

The concept of 'graduation' out of poverty relates to the idea that individuals or households can be pushed, sustainably, out of extreme poverty onto a path of improved livelihoods and wellbeing. CLP-2 considers a household to be 'graduated' if it meets six out of 10 underlying criteria related to a household's wellbeing within three months of completing CLP-2's 18-month core intervention period. The aim is to reach a graduation rate of 85% across households.

Our estimates show that in October 2014, and across cohorts 2.1 to 2.5, CLP is reaching a graduation rate of about 85%. This is mainly driven by high graduation rates among cohorts 2.3 and 2.4. Cohorts 2.1 and 2.2 have significantly lower graduation rates of around 70%. It is difficult to interpret these differences in the combined graduation indicator, but there is evidence that these differences could be due to a learning effect, with programme implementation being adapted and improved over time, from which later cohorts benefitted.

Despite the fact that, on average, CLP has achieved a high graduation rate, this does not mean that all cohorts perform well on all underlying dimensions of graduation. In fact, our analysis shows that in some areas performance is consistently higher than in others. On average, households have difficulties meeting the criteria related to cash savings, income sources, asset values, and access to improved water.

In general, we also find that focusing on binary graduation indicators also hides interesting trends in the original variables that these indicators are based on or are closely related to. We therefore present detailed analyses of underlying indicators as well.

It is important to mention here that in all of these indicators we observe significant improvements in estimates between baseline and follow-up years, which means that there is a strong indication of positive effects of CLP interventions across many indicators.

Income

On average, about 54% of all households across cohorts 2.1 to 2.5 meet the criterion of having more than one income source in 2014. There is a positive trend, within cohorts, of an increasing proportion of households meeting this criterion across the years of data collection, even after the end of CLP's core intervention period. This trend is also reflected in an increase in the average yearly income of households across time. For example, for cohort 2.1 the proportion of households meeting the criterion of more than one income sources increases from about 30% to 52% from 2012 to 2014. Similarly, average yearly income increases from about BDT 47,000 to BDT 60,000 within the same time period.

Food consumption

About 78% of all households across cohorts 2.1 to 2.5 in 2014 meet the criterion of having eaten three meals a day and consumed food from five or more food groups in the week previous to the survey. For some cohorts, this proportion has decreased from 2013 to 2014. Our analysis shows that this is mainly due to households not eating three meals a day – the proportion of households eating food from five or more food groups in the previous week does not decrease. This indicates that dietary diversity among participant households has not decreased but the number of meals eaten might have done so.

Water and sanitation

About 58% of all households from cohorts 2.1 to 2.5 had access to an improved water source in 2014. This indicator has been increasing over time for cohorts, even after the core intervention period ended. We were informed by CLP that this is mainly due to a re-sweep policy, whereby improved water sources were installed retrospectively for earlier cohorts.

In addition, around 70% of households in cohorts 2.1 to 2.5 had access to a sanitary latrine in 2014. The proportion of households meeting this criterion decreases for some cohorts from 2013 to 2014. Our analysis shows that this can partly be explained by the breaking of water seals that form part of the definition of sanitary latrines in the CLP context. This raises questions about the sustainability of installing water seals in sanitary latrines.

About 99% of all households (cohorts 2.1 to 2.5) meet the criterion of having ash or soap close to their water point or latrine in 2014. Within cohorts, the proportion of households meeting this criterion does not vary much and stays relatively constant at about 99% to 100% in non-intervention years.

Asset base

In 2014, about 62% of all households in cohorts 2.1 to 2.5 meet the criterion of owning productive assets valued at over BDT 30,000. We also find increases in the average and variance of the total value of household productive assets across years for the different cohorts – again, even in non-intervention years. Over time, it seems that households shift from livestock assets to land in post-intervention time periods.

Status of females

The graduation indicator related to the status of women is whether women are able to influence investment decisions in the household. In 2014, about 99% of all households from cohorts 2.1 to 2.5 meet this criterion, with dramatic increases from baseline to follow-up data collection periods. Looking at a wider group of decision-making areas, we find similar increases in decision-making power of women associated with the CLP intervention, but with very few changes after this ended.

Vulnerability

We assess vulnerability both from a financial (savings) and environmental risk perspective. In 2014, about 79% of all household from cohorts 2.1 to 2.5 have homesteads that are above the highest known flood levels. This indicator is closely related to a core component of CLP interventions, where participant homesteads are erected on a raised plinth that aims to protect households from flooding. Our analysis finds a slight decrease in households meeting this criterion from 2013 to 2014 for cohorts 2.1, 2.2, and 2.4, which raises questions about the sustainability of this intervention.

In terms of savings, only about 35% of all households in cohorts 2.1 to 2.5 meet the criterion of having BDT 3,000 or more as cash savings, making this the indicator with the lowest compliance rate overall. Interestingly, we find quite significant drops in households meeting this criterion within cohorts 2.1 and 2.2 from 2012 to 2013. When looking at average cash savings, it becomes clear that this is not due to a decrease in cash savings, but rather due to an increase in the variation of cash savings across households. The distribution of cash savings widens over time and hence pushes more households under the threshold of BDT 3,000. This means that, over time, we observe an increasing unequal distribution of cash savings and changing saving behaviours across households.

Group membership

In 2014, about 80% of all households across cohorts 2.1 to 2.5 report being members of a social group. Within cohorts, we see an increase in the proportion of households meeting this criterion, even after the CLP-2 intervention period ended. In addition, though, we observe that the proportion of households meeting this criterion is higher for more recent cohorts than for earlier cohorts.

Poverty

One of the key objectives of this report is to present our approach to measuring poverty and poverty-related indicators in this evaluation. The guiding principle for our poverty analysis is to achieve comparability between a measure of consumption for the CLP households and the consumption aggregate used to calculate official poverty statistics by the BBS, based on the national HIES data.

For the purposes of this evaluation, we consider two measures of consumption poverty and one of asset poverty. Consumption poverty is assessed using information on consumption levels of households. Asset poverty is assessed using an asset index. To measure consumption poverty we rely on the poverty headcount ratio and the poverty gap index. The poverty headcount ratio is defined as the share of CPHHs that fall below a set poverty line. The poverty gap index measures the gap between the consumption levels of those households identified as poor in this way and the poverty line, as a proportion of the poverty line.

We use the lower HIES poverty line for rural Rangpur (BDT 1,235.66 per capita in 2010 prices) to calculate both the poverty headcount ratio and the poverty gap index. To do this, though, we require that the measure of CLP participant consumption is comparable to that collected by the HIES. This, however, is not the case due to differences in survey design and seasonality effects in the CLP data.

We therefore rely on an imputed measure of consumption for the CLP CPHHs in order to compute both poverty headcount ratios and the poverty gap index. The imputation is implemented using simple regression analysis. We also construct the asset index based on principal component analysis applied to HIES data. The asset poverty line is drawn at the 30th percentile of the rural Rangpur asset index distribution in the HIES data

Overall, we find that in October 2014 about 45% of all households from cohorts 2.1 to 2.6 fall under the lower rural Rangpur poverty line. For non-baseline and non-intervention cohorts, this value lies at about 31% to 38%. Similarly, we find that about 28% of all households fall under the asset poverty line in 2014. The CLP intervention is associated with significant decreases in these poverty rates across all cohorts. At baseline, between 84% and 98% of all households fall under the consumption poverty line and around 90% fall under the asset poverty line. Similarly, the poverty gap index decreases substantially across time in parallel with CLP-2 interventions.

Household expenditure

We also look at total average yearly expenditure on food and household goods separately across cohorts. Interestingly, we find no systematic pattern of increased spending on food. Across cohorts there was an increase in food spending from 2012 to 2013, but then a decrease from 2013 to 2014. When looking at household expenditure on household goods, we do find systematic and significant increases in expenditures between baseline and follow-up years and an increase in expenditures across years generally. This indicates that most of the increase in consumption that households show during and after the CLP intervention period is found in non-food items.

Child nutrition

Finally, we assess whether child malnutrition prevalence changes during the CLP intervention period. We find no indication that this is the case. For all indicators looked at (stunting, wasting, underweight) there are no systematic differences between baseline and follow-up values, or across time periods. This could be related to the fact that households do not change their spending behaviour in relation to food.

Discussion

In this report, we have presented the main results of our initial descriptive analyses of CLP-2 participants. The results shown here help us to assess, in a descriptive way, how far lives of CLP participants have changed. A more comprehensive assessment, which will take into account evidence from our qualitative and quasi-experimental exercises, will follow in separate reports. The evidence presented here will feed into those exercises.

CLP-2 is on track to reach 85% of participant households ‘graduating’ – as defined by its own composite graduation indicator (not taking into account time of measurement).

This is mainly due to high graduation rates for more recent cohorts. It is difficult to interpret these differences across cohorts in the combined graduation indicator, but there is evidence that they could be due to a learning effect, whereby programme implementation has been adapted and has improved over time.

The composite graduation index hides significant variation in performance related to underlying indicators.

Despite the fact that, on average, CLP has reached 85% graduation in October 2014, this does not mean that all cohorts perform well on all underlying dimensions. In fact, our analysis shows that in some areas performance is consistently higher than in others. Households have difficulty meeting the criteria related to cash savings, income sources, productive asset values, and access to improved water.

Focusing on binary graduation indicators also hides interesting trends in the original variables that these indicators are based on or are closely related to.

Throughout this analysis, we have not only analysed the binary graduation indicators, but also variations in the underlying or related variables. We find interesting results in this regard:

- First, we find that both the graduation indicator related to income sources and the average value of yearly income increases across cohorts and years, even after the CLP intervention period ends.
- Second, we find that a decrease in the proportion of households meeting the food consumption indicator across cohorts between 2013 and 2014 is mainly due to the fact that fewer households eat three meals a day, and is not due to a less diverse diet.
- Third, we find that a broken water seal could, at least in some instances, explain why the proportion of households that have access to sanitary latrines decreases for some cohorts from 2013 to 2014.
- Fourth, we find that it is not only the average value of assets that households own that increases over time, but also its variation.
- Finally, we show that while there is a decrease, for some cohorts, in the proportion of households that have cash savings of over BDT 3,000, the average value of cash savings does not decrease significantly over time. Rather, the distribution of average cash savings widens significantly, which pushes some households under the BDT 3,000 threshold.

Thus, exposed to similar interventions some households are more successful in increasing the value of their assets or accumulating cash savings than others. We hope that our qualitative research will shed a light on these differential asset and cash savings accumulation strategies.

We also find that some graduation indicators related to specific outcomes should be defined differently if they are to serve as meaningful measures of improvement in these outcomes.

For instance, indicators related to women's empowerment show very little variation between cohorts and it appears that participation in CLP results in an immediate radical change in women's status. At the same time, the savings graduation cut-off threshold of BDT 3,000 may not be consistent with the use of monetary savings as a risk mitigation strategy. Our qualitative research will investigate this further.

The CLP intervention is also associated with a significant decrease in poverty, both measured in terms of consumption and asset ownership.

We find that the CLP intervention is associated with large decreases in poverty rates, both in terms of consumption levels and in terms of asset ownership. In addition, the poverty gap index decreases significantly with CLP interventions. Not surprisingly, we also find that household expenditure (just using CLP data) increases over time for CLP participants – and that this is mainly driven by increases in non-food expenditure.

However, the positive changes in living conditions of CLP households do not translate into changes in the nutritional status of children in these households.

It seems that positive developments in other areas, such as the decrease in poverty rates, does not translate into changes in terms of child nutrition. Part of the explanation for this may be linked to the fact that households tend to increase expenditure on non-food items while the spending on food remains flat.

Across cohorts, we also find that for some indicators positive trends continue after the end of CLP's core intervention period of 18 months. This could be an indication of a sustainable effect of CLP interventions – although more rigorous evidence is needed to substantiate this.

For example, households continue to increase their income sources after the end of CLP's intervention period. Similarly, poverty continues to decrease even after CLP should officially have ended its interventions. This finding is encouraging since it may point to the sustainability of CLP's impact. However, it is not possible to say, simply by looking at these indicators from a descriptive and quantitative perspective, whether these trends indicate sustainable effects of CLP or are the result of certain other factors, e.g. macroeconomic effects, that are driving at least part of these results.

Together with some differences in the background characteristics of cohorts, these findings underline the need for a more comprehensive and rigorous analysis of CLP's effects.

In addition to the issue of external factors that could be influencing the observed results, we also observe some small systematic differences in demographic indicators across cohorts. These systematic differences could, in theory, introduce bias when naively comparing outcome indicators across cohorts in order to infer programme impact. Using panel and PSM techniques, we will aim to ensure that we control for these systematic differences and for yearly fixed effects, so that estimates of CLP impact are more robust than simple comparisons of descriptive statistics. We will supplement this with insights from our qualitative research.

Annex G PSM: First stage selection, matching algorithm and balance diagnostic

In this technical annex we illustrate the analytical steps that have been implemented in this evaluation in order to achieve robust impact estimates using the PSM approach. Firstly, it was important to specify a correct estimation model of the propensity score in the first stage; secondly, a matching method was selected and implemented for the second stage. Finally, balance was assessed across treatment and control groups in order to gauge how well PSM was performing. These following sections separately discuss these different analytical stages.

First stage model selection

To estimate the propensity score in the first stage we followed the procedure suggested by Imbens and Rubin (2015, p. 281 ff.). The underlying model specification for this procedure is a logistic regression or probit specification for the first stage. This means that the propensity scores are estimated by first specifying treatment and control assignment as a binary variable that has the values 0 (for control) and 1 (for treatment). The estimated scores are then modelled as the fitted values that are derived from a logit or probit estimation, with the binary treatment variables as dependent variable and the covariates across which balance is supposed to be achieved as the regressors.

To be more concrete, in the case of a logistic regression specification, the binary response variable is modelled as follows:

$$\Pr(T = 1 \mid X_i) = \frac{e^{f(X_i)}}{1+e^{f(X_i)}}, (1)$$

where $\Pr(T = 1 \mid X_i)$ is the probability of the treatment indicator (T) being equal to one, conditional on the covariates (X_i) for unit i. The function $f(X)$ is normally modelled linearly, i.e. is of the form $f(X) = X\beta$. The coefficients of this function (β) are estimated using maximum likelihood techniques. The fitted values, i.e. the predicted probabilities that follow from this procedure, are the propensity scores for each unit of observation.

The key question for the first stage is which covariates to include in $f(X)$ so that this procedure produces a valid estimate of the propensity score. Following the procedure described in Imbens and Rubin (2015) for selecting covariates, we implemented a three-step approach to make this decision in this evaluation:

1. **Select a set of basic covariates based on substantive grounds**:

   The starting point for the PSM analysis was to select a set of variables that were likely to be relevant for this analysis. 'Relevant' in this case meant that we had to select all variables that were theoretically expected to be correlated with treatment status and treatment effects, thereby introducing bias in a simple comparison of treatment outcomes between control and treatment groups. This requires a theoretically substantiated understanding of the relationships that are being analysed.

   In the present case, selecting these variables was difficult. Note that all households in our sample have the same theoretical treatment status – all households are CLP-2 participants. The only difference is that we are looking at baseline versus follow-up data for different cohorts. Hence, the variables to be selected at this first step were variables that are important indicators of differences between cohorts in characteristics that are plausibly related to the outcome indicator and the way in which treatment effects materialise. We made such a selection for each of the outcomes we looked at.

2. **Increase the set of covariates based on algorithmic approaches:**

   In addition, however, we employed variable selection algorithms to identify variables that vary significantly between control and treatment groups. There are a variety of methods available to do this. Our approach was to implement stepwise first stage regressions.

   There are two stepwise regression approaches that can be employed for this: backward and forward stepwise regression. The underlying idea behind both approaches is to check each covariate, step-by-step, for significant correlation with the outcome and treatment assignment variable. We are looking for such a correlation because variables that possibly bias our impact estimates will have some relation to both the treatment status and the outcome we are looking at.

   Backward selection starts with the full set of covariates, i.e. a regression including all variables, and then discards the term that is least significantly correlated with the dependent variable. It continues to do so until all variables that are uncorrelated with the dependent variable are discarded. Forward selection, instead, starts with an empty set of covariates, i.e. a regression on a constant, and then checks the significance of each covariate if it is included in the regression. It then adds the most significantly correlated variable to the model. This step is repeated until all significant covariates are included in the model.

   Both for backward and forward estimation a threshold p-value for what is considered to be significant needs to be specified. For backward selection, this means setting the level for identifying whether all variables that are uncorrelated with the outcome variable have been discarded: if the p-value of the least significant variable remaining is under the threshold, i.e. all the variables still included in the model are even more significant, the procedure stops. For forward selection, this means setting the level for identifying whether all significant covariates have been included in the model: if the p-value of the most significant variable to be added is equal to the threshold, i.e. the significance level of all variables that have not yet been included in the model is equal or below the threshold, the procedure stops. Setting this threshold therefore influences the variables that are selected in stepwise regressions.

   We implemented both backward and forward selection using different thresholds and selected variables based on whether they were selected in all of the different specifications or not.

3. **Increasing the set of covariates with polynomial and interaction terms using algorithmic selection**

   In a third step, we employed the same method of stepwise regressions to augment our set of covariates by quadratic terms or interactions of variables that had already been selected. The rationale behind this is the fact that balance might only be achieved if the propensity score is estimated using non-linear transformations of the variables selected above (Imbens and Rubin 2015, p. 287). Again, the stepwise regression approach helped to decide which of these non-linear terms were significant predictors of differences across control and treatment groups, and should therefore be controlled for.

The result of this process was the identification of an optimal selection model comprising a set of covariates that were included in the first stage estimation of the propensity score. This three-step approach was conducted for every estimation strategy for each of the outcome variable. Balance, however, also depends on the matching algorithm used in the second stage of the PSM analysis.

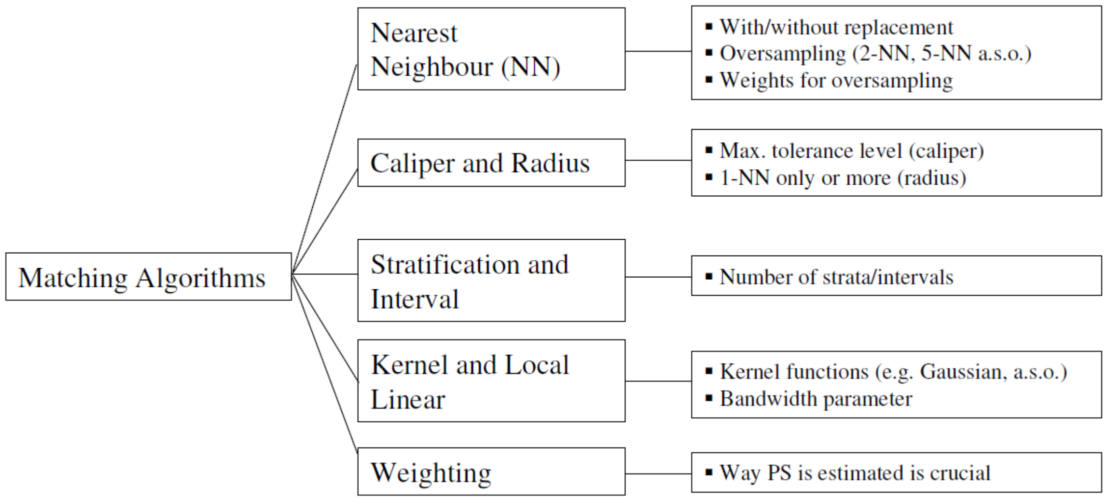
Second stage algorithm selection

There are a variety of algorithms available to implement the second stage of PSM, i.e. to match control and treatment units to each other based on the propensity score estimated in the first stage. Figure 6 below shows algorithm options and sub-options for each of these possibilities. It is beyond the scope of this report to explain in detail the technicalities of each of these approaches. [72] For all approaches the goal is to find appropriate, i.e. sufficiently similar, control group members for treatment group members. Differences between these approaches can be defined along three main dimensions: first, which estimated propensity scores are considered to be valid for inclusion in the analysis? Second, what is the appropriate range of propensity scores that define control comparators for treatment units? Finally, how are these comparators used when estimating the treatment effects?

The first dimension relates to the fact that within both control and treatment groups there could be estimated propensity scores that lie either at the upper or lower bound of the distribution, i.e. close to 0 or 1. For such values, there might not be an appropriately similar propensity score in the respective comparison group. However, for matching to work appropriately, there must be comparable propensity scores in both control and treatment groups – the so-called common support condition. Hence, matching algorithms employ cut-offs or trimming procedures by which some proportion of observations with propensity scores that are not comparable are dropped from the analysis.

[72] See Caliendo and Kopeinig (2005) for a summary overview.

Figure 6 Matching algorithms selection



NN: Nearest Neighbour, PS: Propensity Score

Note: Figure taken from Caliendo and Kopeinig (2005, p. 9).

The second dimension relates to how units in the control group with propensity scores close/similar to a treatment group observation are treated. For instance, kernel matching, as used in our main impact estimation for the PSM model, is a non-parametric matching estimator that uses the weighted averages of all individuals in the control groups to create the counterfactual outcome. The weights are determined by the distance between each individual from the control group and the participant observation for which the counterfactual is estimated. Therefore, higher weights are given to persons closer in terms of the propensity score of a treated individual (Caliendo and Kopeinig (2005), p.10–11). Alternatively, NN matching with just one unit looks for the one control observation that has the closest propensity score to a treatment unit and compares the outcome measure for those observations. NN matching with more than one neighbour looks for several control units with similar propensity scores and compares the treatment outcome to an average of these neighbours. Caliper matching is similar to NN matching but does not include a fixed number of neighbours. Instead, the comparators are selected based on a maximum difference in propensity scores allowed.

Finally, the third dimension refers to how, once comparator units are found, the outcome measures are compared across treatment and control. For example, with NN matching and more than one neighbour simple averages are calculated. Similarly, with kernel functions a form of weighted averages are calculated to estimate treatment effects.

Selecting the appropriate matching algorithm for a PSM exercise is not straightforward and requires careful analysis of how well-balanced samples are after employing algorithms with certain sub-specifications. In general, however, our selection of models was based on the fact that discriminating between models poses a bias/variance trade-off in the estimated treatment effect. For instance, in the extreme case of NN matching with just one neighbour, it could be that the NN is actually quite far away in terms of propensity scores and hence a bad match. If this happens often, this could introduce bias into the estimation procedure. A solution to this could be to implement matching using several comparators in a caliper matching setting.

However, this could decrease the number of available matches, which could increase the variance of the treatment estimate.

Kernel matching with appropriate trimming and enforcement of common support is a good compromise between these different approaches and was therefore selected as our main matching algorithm.[73] In order to find the optimal estimation model we used different kernel matching algorithms with different bandwidths and trimming levels. These different results were then compared with respect to the best balancing properties, with the best performing approach being selected as the optimal. This was again conducted for each estimation strategy for each of the outcome variables. Additionally, as a robustness check, we present results using a different the NN matching algorithm in Section 6.4.5.

Assessing balance

In regard to selecting the appropriate models and matching algorithms it was key to assess how balanced samples were after matching. To do this, we compared matching models along a variety of dimensions. First, we assessed individual covariate balance across samples by looking at the standardised difference in means across treatment and control groups both before and after matching. This standardised difference is the difference in group averages over the square root of the average of the sample variances. If samples are balanced, this difference should be small and matching should reduce this standardised difference as compared to the unmatched samples.

In addition, we performed t-tests to assess whether differences across treatment and control groups were statistically significant. If balance is achieved with PSM, differences between treatment and control groups should be negligible and therefore should not be significantly different from zero.

We also looked at the variance ratios of covariates of treated over control measures. If there is perfect balance across samples, then covariates should be distributed equally and hence this ratio should be equal to one.

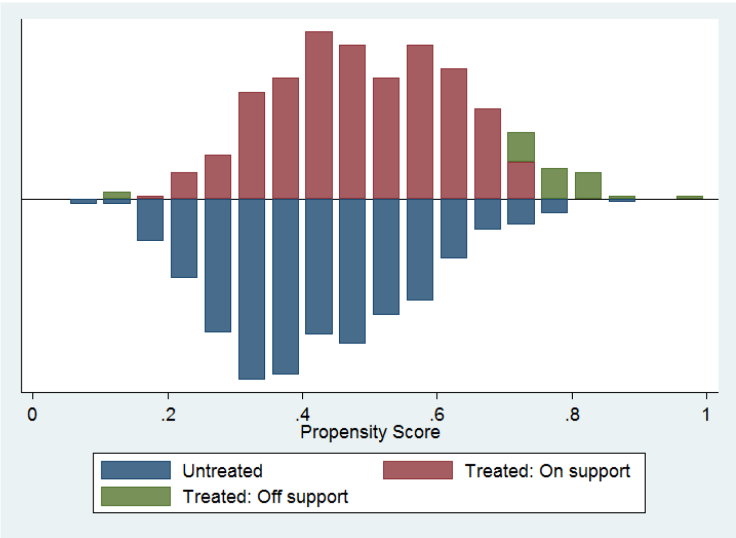
All of these measures give an indication of whether specific covariates are balanced across treatment and control groups. To assess overall variance we looked at two statistics that summarise covariate balance in the sample at hand: Rubin's B and Rubin's R. Rubin's B reflects the absolute standardised difference of the means of the propensity score in the treated and control groups (unmatched and matched). Rubin's R is the ratio of the treated to control variances of the propensity scores. Rubin (2001) suggests that the value of B should lie below 25 and that R should lie between .5 and 2 for overall balance to be sufficient. Together, Rubin's B and Rubin's R provide a reliable indication of the trade-off between bias and variance across the treatment and control groups, as it changes before and after the matching procedure. However, individual-level balance should always be assessed as the overall balance is only an approximation of goodness of fit.

[73] See Caliendo and Kopeinig (2005, p. 10 f.) for a short summary of the pros and cons of different matching techniques.

Matching procedures were implemented using the psmatch2 package in Stata (14.1) and balancing tests were carried out using the pstest package, which provides the resultsfor all of the statistics mentioned above[74].

Finally, we also looked at the distribution of propensity scores graphically. Ideally, propensity scores should be distributed equally across treatment and control groups. Very skewed/diverging distributions could be an indication that balance has not been achieved successfully. Figure 7 below provides a good example of evenly distributed propensity scores for both the treatment and control groups. Please see Section 6.4.5 for all the results from the PSM analysis.

Figure 7 Distribution of propensity scores for per capita cash savings using Strategy 1a



[74] See http://fmwww.bc.edu/repec/bocode/p/pstest.html for details.