

External evaluation of mobile phone technology-based nutrition and agriculture advisory services in Africa and South Asia

Inception Report

First version: 12 August 2015

Revised version: 24 March 2017

Prepared by the Institute of Development Studies, International Food Policy
Research Institute and Gamos

Authors: Barnett, I., Batchelor, S., Gilligan, D., Haddad, L., Hidrobo, M., Ledlie, N.,
Palloni, G., Scott, N. and Shyam, T.

Acknowledgements

This report was produced jointly by the Institute of Development Studies (IDS), the International Food Policy Research Institute (IFPRI) and Gamos. Contributing authors include (in alphabetical order): Inka Barnett, Simon Batchelor, Dan Gilligan, Lawrence Haddad, Melissa Hidrobo, Natasha Ledlie, Giordano Palloni, Nigel Scott and Tara Shyam. The first version of the report was completed in August 2015; this second version incorporates changes made only to the quantitative design and methods between August 2015 and March 2017, with all other sections remaining the same as in the original version. Some sensitive sections of the report have been removed prior to external publication.

We would like to extend our sincere gratitude to the GSM Association (GSMA) team for their ongoing cooperation and support for the evaluation, as well as for their feedback on an earlier draft version of this report. The team would also like to thank the Department for International Development (DFID) and Oxford Policy Management (OPM) teams for their continued support and contributions towards this project and comments on drafts of this report.

We are also grateful to Professor Laura Camfield from the University of East Anglia, Professor Jenny C Aker from Tufts University, Chris Garforth from the University of Reading and Vivienne Benson from IDS who reviewed and commented on an earlier draft of this report. We are also grateful to Jessica Hoel of Colorado College (formerly at IFPRI) and Megan Rounseville from IFPRI who contributed to the inception phase of the project.

Disclaimer

This report has been prepared by the e-Pact consortium through the Global Evaluation Framework Agreement for the named client, for services specified in the Terms of Reference and contract of engagement. e-Pact is a consortium led by Oxford Policy Management and co-managed with Itad. e-Pact cannot be held responsible for errors or any consequences arising from the use of information contained in this report. The information contained in this report shall not be disclosed to any other party, or used or disclosed in whole or in part without agreement from the e-Pact consortium. For reports that are formally put into the public domain, any use of the information in this report should include a citation that acknowledges the e-Pact consortium as the author of the report. This confidentiality clause applies to all pages and information included in this report. This material has been funded by UK aid from the United Kingdom (UK) government; however, the views expressed do not necessarily reflect the UK government's official policies, e-Pact or any other contributing organisation.

Data management

All intellectual property rights in any materials produced from the evaluation (including publication of research findings and all associated reports and data) remain the property of IDS and associated sub-contracted collaborators. IDS and all sub-contracted partners undertaking data collection have specific arrangements in place for handling data generated from the project in accordance with the Data Protection Act (1998), which includes the processing and storage of any sensitive personal data and maintenance of privacy. DFID has unlimited access to any material produced from the evaluation. In order to promote use and uptake of the evaluation findings and in line with DFID's Enhance and Open Access Policy, the evaluation team is committed to ensuring all major report outputs and associated data generated from this project are made publicly available in an accessible format.

The report should be referenced as follows: Barnett, I., Batchelor, S., Gilligan, D., Haddad, L., Hidrobo, M., Ledlie, N., Palloni, G., Scott, N., and Shyam, T. (2017) 'External evaluation of mobile phone technology-based nutrition and agriculture advisory services in Africa and South Asia: Inception Report'. Brighton: IDS.

This project is being led by the Institute of Development Studies (IDS) together with GAMOS and the International Food Policy Research Institute (IFPRI) as part of the e-Pact consortium led by Oxford Policy Management (OPM) co-managed with Itad. The IDS project manager is Jessica Gordon [j.gordon@ids.ac.uk]. The report authors are: Inka Barnett, Simon Batchelor; Dan Gilligan, Lawrence Haddad, Melissa Hidrobo, Natasha Ledlie, Giordano Palloni, Nigel Scott and Tara Shyam. For further information please contact j.gordon@ids.ac.uk.

The contact point for the client is Louise Horner [l-horner@dfid.gov.uk]. The client reference number for the project is PO6420.

e-Pact	Level 3, Clarendon House 52 Cornmarket Street Oxford OX1 3HJ United Kingdom	Tel +44 (0) 1865 207300 Fax +44 (0) 1865 207301 Email admin@opml.co.uk Website www.opml.co.uk
--------	--	--

Executive summary

The GSM Association (GSMA), working with a wide range of mobile network operators and civil society organisations, is launching a series of nutrition-focused m-health and m-agriculture initiatives in South Asia and sub-Saharan Africa. GSMA refers to nutrition-enhanced initiatives collectively as ‘m-nutrition’. This report summarises the plans for an impact evaluation of two of these nutrition-enhanced initiatives: mHealth in Tanzania and mAgri in Ghana. The evaluation consists of three integrated components: a quantitative impact evaluation, a qualitative evaluation focusing on implementation fidelity, pathways of impact and external validity, and an evaluation of the sustainability of the business model behind the mNutrition initiative. The business model evaluation compares the two initiatives described above with a third, mHealth in Ghana, which is closer to the GSMA core commercial model, and additionally, possibly to retain a view on Bangladesh, mAgri to generate more heterogeneity in conclusions.

The quantitative impact evaluation relies on a randomised encouragement design in Ghana and a randomised controlled trial in Tanzania. A randomised controlled trial using a pure control group to generate a counterfactual is not possible in Ghana given the nationwide launch and reach of the m-nutrition services. An encouragement design relies on the random assignment of efforts to substantially increase take-up of the mNutrition project at community level and measures impact using the difference in take-up between encouragement and non-encouragement communities. The encouragements will be conducted using a multifaceted approach including SMS (short message service), voice SMS, and household-level interactions to promote use of the mNutrition mobile platform. In the Ghana mAgriculture evaluation, encouragement messages will be targeted at farming households with a mobile phone. In the Tanzania mHealth evaluation, the roll-out of the programme has been much more limited geographically than originally planned, making it possible to identify a region of limited previous exposure to the programme that could be used to conduct a randomised controlled trial as the programme is rolled out in that region on a pilot basis. In Tanzania, households selected to be part of the study sample in communities randomly assigned to the treatment arm will be offered access to the mNutrition mobile platform at no cost; in control communities, the sampled households will not be offered access to the mNutrition mobile platform. The messages included in the mHealth platform will be targeted at pregnant women and households with children under 12 months of age.

The qualitative design is interwoven with the quantitative impact evaluation and consists of three qualitative data collection events: an initial exploratory study, a qualitative midline and a qualitative explanatory follow-up study. The initial exploratory study comprises an in-depth context analysis and aims to inform the quantitative survey design and the choice of an effective encouragement strategy. The qualitative midline consists of several community-level case studies and sets out to explore ‘how’ and ‘why’ the mNutrition product may (or may not) lead to the desired changes in nutrition and/or agriculture behaviours and practices in the specific country contexts. The qualitative midline also includes an analysis of the implementation processes of the mNutrition product. The qualitative follow-up comprises several mini-case studies that further elaborate, validate and follow up findings that emerge from the triangulation of the quantitative and qualitative impact evaluation data.

The business model evaluation considers the wider imputed benefits from the value-added service for the range of stakeholders involved. It will relate the model to the GSMA theory of change and consider the effectiveness of the customer journey.

Our communications strategy maps the audiences, communications channels and activities that will support and guide an ongoing engagement with key stakeholders to actively involve them throughout the process and to facilitate effective uptake of the evaluation findings and lessons learned.

Important stakeholders include government officials in Tanzania and Ghana who will be encouraged to utilise the evidence produced to decide whether the mNutrition product is likely to make a difference to nutrition behaviours, practices and outcomes in their countries and, therefore, whether or not to spend public funds to enhance or sustain them. The findings and lessons learned in this impact evaluation will also be communicated to: (a) external donors and agencies to inform decision-making on whether or not to invest further resources in overcoming start-up costs to set up mobile phone-based advisory services and under which circumstances; (b) GSMA and mobile operators to learn whether mNutrition drives new clients to their services and whether an enhanced business supports a sustainable business model for m-nutrition; and (c) community health and agricultural workers to inform their decision on whether or not to use and promote mobile phone-based advisory services to change behaviours and promote better behaviours.

Table of contents

Executive summary.....	3
List of tables, figures and boxes.....	7
List of abbreviations and acronyms.....	8
1. Introduction.....	10
2. Proposal for selecting impact evaluations.....	13
2.1 Updated information on costs of conducting the impact evaluations.....	13
2.2 Confirmation of use of encouragement design for identifying impact in Ghana.....	15
2.3 Options for modifying scope and cost of impact evaluations.....	15
2.4 Proposal for conducting two impact evaluations for the existing budget – Option 1a: Tanzania mHealth and Ghana mAgri.....	15
3. Quantitative impact evaluation design.....	19
3.1 Ghana mAgri impact evaluation design.....	19
3.1.1 mAgri programme and the local context.....	19
3.1.2 Research questions.....	21
3.1.3 Impact evaluation design.....	21
3.1.4 Outcomes of interest.....	25
3.1.5 Sampling strategy and data collection.....	27
3.1.6 Partners.....	31
3.1.7 Risks and their mitigation.....	31
3.2 Tanzania mHealth impact evaluation design.....	32
3.2.1 mHealth programme.....	32
3.2.2 Research questions.....	34
3.2.3 Impact evaluation design.....	34
3.2.4 Outcomes of interest.....	38
3.2.5 Sampling strategy and data collection.....	39
3.2.6 Partners.....	43
3.2.7 Risks and their mitigation.....	44
4. Qualitative impact evaluation design.....	45
4.1 Objectives.....	45
4.2 Approach.....	45
4.2.1 Sequential mixed-method design.....	45
4.2.2 Qualitative evaluation studies.....	48
4.3 Methodology.....	51

5. Business model evaluation design	60
5.1 Introduction	60
5.2 Research scope	61
5.3 Research variables	62
5.4 Stakeholders	62
5.5 Revenues and brand	63
5.6 Maturity and scale	65
5.7 Customer experience	67
5.8 Emerging issues.....	68
5.9 Methodology.....	69
5.9.1 Data collection methods	69
5.9.2 Model analysis	72
5.9.3 Notes on methodology	73
5.9.4 Risks and their mitigation	73
6. Communications strategy	75
6.1 Statement of purpose	75
6.2 Communications goal and objectives	75
6.3 Target audiences for the evaluation	75
6.4 Preliminary mapping of target audiences.....	76
6.5 Target audience channels for communications.....	79
6.6 Monitoring and evaluation of the communication activities	83
7. Governance, management and ethics	84
7. 1 Governance	84
7.2 Evaluation partners and personnel.....	84
7.3 Management, communications, intellectual property and data.....	86
7.3.1 Management.....	86
7.3.2 Internal reporting and communications.....	86
7.3.3 Intellectual property and data	87
7.4 Ethical considerations	87
References	90
Annex 1 Terms of Reference.....	94

List of tables and figures

Table 3.1 Required sample sizes for different outcome variables and scenarios.....	29
Table 3.3 Required sample sizes for children’s height-for-age z-scores.....	41
Table 3.4 Required sample sizes for women’s dietary diversity.....	41
Table 3.5 Required sample sizes for exclusive breastfeeding.....	42
Table 3.6 Required sample sizes for children’s dietary diversity.....	42
Table 3.7 Data sources for power calculation assumptions.....	42
Table 4.2 Data collection events for the initial qualitative baseline, per mNutrition intervention and country.....	55
Table 4.3 Data collection events for the qualitative midline, per mNutrition intervention and country.....	57
Table 4.4 Data collection events for the qualitative follow-up, per mNutrition intervention and country.....	58
Table 5.1 Key stakeholder groupings for the four propositions under study during the inception phase.....	62
Table 5.2 mHealth Ghana stakeholders and types of service.....	63
Table 5.3 Distribution of revenue in Ghana mHealth?.....	65
Table 5.4 Key variables for the four propositions under study during the inception phase.....	66
Table 5.5 mHealth Ghana, stakeholders’ experience.....	66
Table 6.1 Country stakeholders.....	79
Table 6.2 Target audiences and channels for communication.....	81
Figure 4.1 Sequential mixed-method design to study the impact of the mNutrition products.....	47
Figure 5.1 Illustrating the iterative nature of business model development for value-added services.....	60
Figure 5.2 Response to question on business drivers for value-added services.....	64
Figure 5.3 'Revenue' vs 'brand' – What’s in it for the MNO?.....	64
Figure 5.4 Funding curve for Agri VAS.....	65
Figure 5.5 Illustrating the potential pitfalls of customer experience.....	67

List of abbreviations and acronyms

ANCOVA	analysis of covariance
ATE	average treatment effect
ATT	average treatment effect on the treated
BDM	Becker-DeGroot-Marschak
BMI	body mass index
CLE	community listing exercise
CMO	context-mechanisms-outcome
CSO	Civil society organisation
CSR	corporate social responsibility
DAC	Development Assistance Committee
DFID	Department for International Development
DHS	Demographic and Health Survey
FAC	Future Agricultural Consortium
FGDs	focus group discussions
GAIN	Global Alliance for Improved Nutrition
GSMA	GSM Association
HANCI	Hunger and Nutrition Commitment Index
HAZ	height-for-age Z scores
HDDS	Household Dietary Diversity Score
ICYF	Infant and Young Child Feeding Practices
IDI	in-depth interviews
IDS	Institute of Development Studies
IFPRI	International Food Policy Research Institute
ISSER	Institute of Statistical, Social and Economic Research
ITT	intent to treat
JHU	Johns Hopkins University
LATE	local average treatment effect
MDD-W	Minimum Dietary Diversity-Women
MDE	minimum detectable effect
M&E	monitoring and evaluation
MNO	mobile network operator
MOU	memorandum of understanding
NGO	non-governmental organisation
OECD	Organisation for Economic Co-Operation and Development
OPM	Oxford Policy Management
pp	percentage point
RCT	randomised controlled trial
SD	standard deviations
SMS	short message service
SUN	Scaling Up Nutrition Network
ToR	terms of reference
USSD	Unstructured Supplementary Service Data
VAS	value-added service
WAZ	weight-for-age
WDDS	Women's Dietary Diversity Score
WHO	World Health Organization
WHZ	weight-for-height

WN Wazazi Nipendeni
WTP willingness to pay

1. Introduction

mNutrition is a global initiative supported by the Department for International Development (DFID), organised by GSMA and implemented by in-country mobile network operators (MNOs) to use mobile technology to improve the health and nutritional status of children and adults in the developing world. The potential to utilise mobile technology to change attitudes, knowledge, behaviours and practices around health and agriculture for improved nutritional status has been recognised for some time, but to date there have been no rigorous evaluations of m-services at scale. A consortium of researchers from Gamos, the Institute of Development Studies (IDS) and the International Food Policy Research Institute (IFPRI)¹ will conduct a rigorous mixed-methods evaluation to estimate the impact of mNutrition on children and adults and to understand how the context and the components of the mNutrition intervention shape its impact.

mNutrition will be implemented through mAgri and mHealth programmes in 14 countries throughout sub-Saharan Africa and South Asia. The nutrition content aims to promote behaviour change around key farming decisions and practices and around maternal and other household practices that are likely to result in improved nutritional health within a household. The evaluation design is expected to measure the impact, cost-effectiveness and commercial viability of mNutrition, using a mixed-methods evaluation design. In line with the requirements in the terms of reference (TOR),² the evaluation will address the following research questions:

- What are the impacts and cost-effectiveness of mobile phone-based nutrition and agriculture services on nutrition, health and livelihood outcomes, especially among women, children and the extreme poor?
- How effective are mobile phone-based services in reaching, increasing the knowledge, and changing the behaviour of the specific target groups?
- Has the process of adapting globally agreed messages to local contexts led to content which is relevant to the needs of children, women and poor farmers in their specific context?
- What factors make mobile phone-based services effective in promoting and achieving behaviour change (if observed), leading to improved nutrition and livelihood outcomes?
- How commercially viable are the different business models being employed at country level?
- What lessons can be learned about best practices in the design and implementation of mobile phone-based nutrition services to ensure (a) behaviour change and (b) continued private sector engagement in different countries?

The evaluations will be conducted on two programmes, Ghana mAgri and Tanzania mHealth, which were selected from four possible programmes to evaluate: Ghana mAgri, Ghana mHealth, Tanzania mHealth and Bangladesh mAgri. In order to satisfy the objectives of the TOR, the evaluation will include a quantitative component, a qualitative component, a business model analysis, and a communications strategy for study findings.

¹ See Chapter 7 for management and governance arrangements, as well as for descriptions of each of the organisations.

² See Annex 1.

The quantitative impact evaluation relies on a randomised encouragement design in Ghana and a randomised controlled trial in Tanzania. A randomised controlled trial using a pure control group to generate a counterfactual is not possible in Ghana given the nationwide launch and reach of the m-Nutrition services. An encouragement design relies on the random assignment of efforts to substantially increase take-up of the mNutrition project at community level and measures impact using the difference in take-up between encouragement and non-encouragement communities. The encouragements will be conducted using a multifaceted approach including SMS (short message service), voice SMS, and household-level interactions to promote use of the mNutrition mobile platform. In the Ghana mAgriculture evaluation, encouragement messages will be targeted at farming households with a mobile phone. In the Tanzania mHealth evaluation, households selected to be part of the study sample in communities randomly assigned to the treatment arm will be offered access to the mNutrition mobile platform at no cost; in control communities, the sampled households will not be offered access to the mNutrition mobile platform. The messages included in the mHealth platform will be targeted at pregnant women and households with children under 12 months of age.

Two further research questions in the TORs cannot be undertaken with existing resources without compromising the impact evaluation, the process evaluation and the business model evaluation. The two questions are:

1. Are mobile phone-based services a cost-effective way to register and identify at-risk populations to target with health support?
2. Are mobile phone-based services a cost-effective way for community health workers to improve the quality and timeliness of data surveillance (a core set of nutrition-related indicators)?

To address these questions we would need to conduct a separate impact evaluation study on the effectiveness of a mNutrition product specifically developed and targeted at community health workers (who are usually responsible for community-based surveillance (Question 2) and identification of at-risk populations (Question 1)). According to GSMA, its primary focus is currently on the development of a mNutrition project targeted at consumers (i.e. pregnant women). The development of an additional product specifically for healthcare workers will depend on country-specific interest in such a tool and will be determined in the future. Consequently, it will not be possible (even if the budget would be available) to address the two questions at this point.

This inception report first describes the process undertaken and information gathered during the inception phase, resulting alterations as compared to the original proposal, the quantitative impact evaluation options proposed, and the reasons for selecting the two programmes to be evaluated.

The quantitative evaluation design chapter lays out, for the Ghana mAgri and Tanzania mHealth evaluations in turn, the research questions to be addressed, the product to be evaluated, and the experimental evaluation designs for measuring causal impacts of the programmes.

The qualitative evaluation design chapter describes how the qualitative component of the evaluation complements the quantitative component. This chapter details the objectives and research questions for the qualitative evaluation, which includes providing a better understanding the contexts of both programmes and analysing the process of nutrition behaviour change and its underlying mechanisms within these contexts.

The business model evaluation chapter lays out the research scope for assessing the commercial viability of the business models of the two programmes, as well as the best practices in these services that ensure behaviour change and continued private sector engagement. Ghana's mHealth and Bangladesh's mAgri have also been tentatively included in this plan.

Finally, the communications chapter outlines the planned strategy for engaging with key stakeholders throughout the project timeline in order to facilitate the uptake of evaluation findings and the lessons learned.

2. Proposal for selecting impact evaluations³

2.1 Updated information on costs of conducting the impact evaluations

During the inception phase, we met with many of the main stakeholders in all three countries for all four projects, gathered information about the interventions and context, and worked intensively on strengthening and refining the evaluation designs. Through this process, we learned that the cost of designing and conducting the mNutrition impact evaluations to provide the quality of evidence required and to meet the standards for rigour in causal design and quality of evidence that we believe DFID and mNutrition stakeholders expect is substantially higher than was budgeted for in the proposal. These additional costs arise in terms of labour effort needed to conduct the evaluations, the cost of data collection for the quantitative evaluations, and the cost of conducting the encouragements in some cases where we have learned that the evaluation budget will need to fund the encouragement SMS messages and community promotion campaigns, because detailed integration with the implementing partners on encouragement promotions may not be feasible.

The main reasons that the costs of conducting the quantitative evaluations are higher than budgeted in the proposal are the following:

- With the exception of mAgri in Ghana, implementation plans for rolling out the mNutrition platforms were not as well developed as we expected during our inception visits. Key features of the implementation plans were still being developed, including which existing mHealth and mAgri platforms would host the new nutrition content, which NGOs would help to enrol beneficiaries in the service, the price structure for key aspects of the service, and the timing of the availability of the nutrition content in the platforms. This has drawn out the process of finalising the impact evaluation designs, which will take substantially more time and effort after the inception phase than originally planned.
- The complexity of the mNutrition platforms and the large number of partners involved in implementation is greater than we anticipated on the basis of the TOR. This raises the labour cost to the evaluation teams of coordinating the evaluation with multiple implementing partners on the ground.
- The implementation risk – that mNutrition platforms may start later than scheduled or may have designs that differ from what we were originally told – is higher than we anticipated, which increases the evaluation team’s cost of monitoring the implementation for compliance with the evaluation design. This may require additional effort to redesign subcomponents of the evaluation to be consistent with implementation.
- The sample sizes needed to provide the promised level of statistical power and confidence level are larger than estimated at the proposal stage. This is true in part because the sample size estimate provided in the response to questions from DFID on the proposal were based, as they often are, on a randomised controlled trial (RCT) design with perfect compliance. However, these estimates should be adjusted for the partial take-up rates that characterise encouragement designs and RCTs that are not efficacy trials. Rather than

³ Note that some sections have been removed prior to publication due to their sensitive nature

comparing 100 per cent take-up in the treatment arm and 0 per cent take-up in the control arm, as in an RCT with perfect compliance, the encouragement design compares high take-up (say, 50-75 per cent) to low take-up (say, 5 per cent).⁴ This smaller difference in take-up rates in an encouragement design reduces statistical power in the study. Similarly, though the take-up gap in an RCT with imperfect compliance is likely to be larger than it would be under a randomised encouragement design, it almost certainly will be less than 100 per cent. Formally, the necessary sample size for either an encouragement design or an RCT with imperfect compliance compared to the necessary sample size for an RCT with perfect compliance increases by the inverse square of the difference between high and low take-up (Duflo *et al.* 2007: pp. 3895-3962). For example, in a randomised encouragement design with 50 per cent take-up in the encouragement arm and 5 per cent take-up in the non-encouragement arm the sample size inflation factor is:

$$\left(\frac{1}{0.50 - 0.05}\right)^2 = 4.9$$

That is, a difference in take-up rates of 45 per cent between the encouragement and non-encouragement would require a sample size almost five times larger than in an RCT with perfect compliance. We have adjusted the evaluation and sample designs to improve on statistical power where possible (e.g. limiting age of children under study to 0-35 months, the age of highest potential to respond to treatment; shorter questionnaires), but increases in sample size will be needed for all outcomes. Anthropometry, in particular, requires large samples of children to be able to detect a sufficiently small minimum detectable effect (MDE); our current mHealth design retains anthropometry as a primary outcome, but impacts (.25 standard deviation of HAZ – height-for-age Z scores) and mNutrition take-up rates (45-70 per cent) will need to be large to be detectable (as described in Sections 3.1.5 and 3.2.5). That is, given the sample sizes that can be afforded for this study, our power calculations indicate that the project will need to have relatively large effects on anthropometry and relatively high take-up rates in order for the study to detect a significant impact on anthropometry. When revising the sample size estimates, we also updated the data and samples used to be more current and representative. This changed some of the parameters used in the direction of increasing sample sizes in some cases.

- The proposal budget for the impact evaluation assumed that the cost of running any encouragement messaging as part of the impact evaluation would be absorbed by the mNutrition implementing partners under their own budgets for rolling out mNutrition. During the inception phase, we learned that our encouragement promotion campaigns are more complementary to the plans to roll out the implementation of mNutrition in some cases than in others, and that some implementing partners are more able than others to support our encouragement designs for learning purposes. For example, implementing partners working with Vodafone for mAgri in Ghana have detailed plans for promotion campaigns and are somewhat concerned about how adding the encouragement could increase the cost and effectiveness of their plans. As a result, we need to include an

⁴ There may actually be different levels of take-up of the mNutrition project, ranging from being exposed to the project by receiving an SMS message containing mNutrition content, to signing up for a stream of information through the mNutrition platform. SMS messages intended exclusively to promote using or signing up for the mNutrition platform come under the label of promotion. The encouragement design creates experimental variation in the amount of promotion a person receives. The objective in an RCT is to have 100 per cent exposure to the treatment in the treatment arm and 0 per cent in the control arm.

encouragement budget to provide resources to run or support the encouragements for some projects.

2.2 Confirmation of use of encouragement design for identifying impact in Ghana

We have also considered whether alternative impact evaluation designs could be effectively implemented in Ghana at substantial cost savings, but we have concluded that a refined version of the encouragement design provided in the original proposal continues to provide the best possible design to obtain rigorous impact estimates in this context, for several reasons. First, an encouragement design is one of the best methods for measuring impact of interventions that are universally available (Duflo and Saez 2003; Duflo *et al.* 2007). This is because an encouragement design provides randomised (experimental) variation in the probability of taking up the intervention that can be exploited to provide identification of causal effects. RCTs using a pure control group are not feasible in Ghana because it is not possible to systematically exclude large areas from the implementation as would be required to randomly assign some communities to a control group. Methods based on ‘matching on observables’ (e.g. propensity score matching, covariate matching and propensity score weighted regression) are also not preferred in this setting.

For the mHealth and mAgri interventions on which mNutrition platforms will be built, the current rate of take-up of the interventions is very low, between 2 and 5 per cent. Locating these beneficiaries for a matching analysis (probably through community census exercises repeated throughout the study) would be costly. More importantly, finding a set of households to serve as the counterfactual for these households may be very challenging because, at these low levels of take-up, mNutrition participants are likely to be fairly unique. Also, matching generally provides less convincing evidence because matching methods rely on the untestable (and often strong) assumption that observable variables account for all differences between the treatment and comparison group due to selection effects. Although matching is sometimes used when experimental methods (e.g. RCTs or encouragement designs) cannot be used, experimental approaches allow researchers to better control the design of the study to reduce bias in impact estimates.

2.3 Options for modifying scope and cost of impact evaluations

Given these concerns, it is not feasible to complete all four impact evaluations for the current budget. Therefore, we developed a menu of options for DFID to consider, which we presented to DFID in a meeting of partners on 4 June 2015 at DFID headquarters.

2.4 Proposal for conducting two impact evaluations for the existing budget – Option 1a: Tanzania mHealth and Ghana mAgri

During a meeting at DFID headquarters on 4 June 2015, and in subsequent communication, representatives from DFID, IFPRI, Gamos and IDS discussed the options presented. Everyone recognised that any option for conducting three or four impact evaluations would require substantial additional resources. On 9 June 2015, DFID indicated that it would not be possible to increase the budget for the impact evaluations, so that some version of Option 1 would need to be implemented. This section presents a proposal for conducting the impact evaluations

under Option 1a: Tanzania mHealth and Ghana mAgri. We are willing to consider alternative projects to include in a two-project evaluation, but we believe that evaluating Tanzania mHealth and Ghana mAgri provides the best chance of conducting rigorous evaluations of high-quality mNutrition projects. In the next two sections of this inception report, we describe the Ghana mAgri and Tanzania mHealth projects in detail and present detailed impact evaluation plans. In this section, we provide additional information about the Ghana mHealth and Bangladesh mAgri projects in order to inform the selection of projects to be evaluated.

In deciding which two mNutrition projects to propose, we started with the principle that the study should include one mHealth and one mAgri project, in order to provide lessons for both types of interventions. We then used two criteria to decide which mHealth and mAgri projects to select: (i) strength of the overall project in terms of design and potential for implementation, and (ii) potential for completing a rigorous impact evaluation.

For mHealth, the Tanzania and Ghana projects have some similar design features. In both countries, GSMA is working with local partners to add mNutrition content on top of existing mHealth platforms targeted at pregnant women and mothers of children in the first year of life. In Tanzania, this platform is the 'Healthy Pregnancy, Healthy Baby' SMS programme run by Wazazi Nipendeni (WN). In Ghana, GSMA is in discussion with several local content providers, including Grameen Foundation (maternal and child health), Call-A-Doctor (call centre), Microinsure (financial transactions) and Audry Pack (maternal and child health). GSMA in Ghana has identified a content aggregator that will help to link mNutrition content to the mHealth platforms of several of these local content providers. Grameen Foundation's 'Mobile Midwife' platform is a leading candidate to be included in mNutrition and offers similar services to those provided by WN in Tanzania.

However, there are a number of important differences between the Tanzania and Ghana mHealth projects that make Tanzania rank above Ghana on the first criterion, strength of the mHealth project:

- **Scale:** In Tanzania, WN has signed up 800,000 unique subscribers since 2012 and it currently has 100,000 active users. This demonstrates the potential of WN to roll out an attractive platform that is likely to be taken up by users. This is a major advantage over the existing Ghana mHealth platforms, all of which appear to be substantially smaller. Grameen Foundation's Mobile Midwife, for example, has roughly 300 community health workers supporting its platform, but this likely means it has substantially fewer active subscribers.
- **Negotiations with partners:** The roll-out of mNutrition is at a more advanced stage in Tanzania than in Ghana. As noted in the Tanzania mHealth section below, WN has already negotiated with all of the MNOs in Tanzania to 'zero-rate' (provide to the consumer for free) their original package of health messages, and they were negotiating with the MNOs in June 2015 on the business model to add the mNutrition content provided by GAIN (the Global Alliance for Improved Nutrition). The price of the new mNutrition content has not yet been set, but is expected to be zero-rated. In Ghana, GSMA had identified an aggregator, Mobile Content, to build the mNutrition platform, but during our inception visit they were at an early stage of negotiating with the local content partners to serve as hosts for the new mNutrition content. GSMA already has agreements with MTN and Airtel to carry this content, but the participation of Grameen Foundation and other local providers is essential to ensuring that the product is known and will be taken up. This suggests that

the introduction of mHealth in Ghana could be delayed or that the number of households adopting the mNutrition service may be much smaller than in Tanzania.

- **Preparedness of project design:** The mHealth projects in both countries still have outstanding decisions to make about important design features of their platforms, but it appears that the project design in Ghana lags behind that in Tanzania. For example, the profiling section of a skeleton platform we were shown in Ghana did not have all of the appropriate age groups identified (e.g. children age 0-6 months, 6-24 months, etc) and there was uproar from partners at a meeting we attended over a proposal to introduce mNutrition using 'freemium' pricing (free for one month only) with an aggressive opt-out requirement that could lock unsuspecting consumers into paying for the service. We were told later that this pricing strategy would not be utilised, but it suggested to us that decisions on these important features were still at a very early stage.

Regarding the second criterion for selecting a project to evaluate – feasibility of conducting a rigorous evaluation – both mHealth projects could still be evaluated, but we expect to have better coordination of the study with the implementation of the programme in Tanzania. In our meetings to discuss plans for the evaluation in Tanzania, WN was very supportive of the evaluation plans and anxious to help make the evaluation a success. WN indicated that it may be able to help coordinate roll-out of the project in a way that would assist the research design. In Ghana, GSMA was also supportive of the plans for the evaluation, but we were not able to get similar buy-in from the local content providers, some of which were still negotiating plans for mNutrition with GSMA and so did not want to discuss the evaluation. Also, in Ghana there is no uniform plan for marketing, because marketing will be the responsibility of the individual content providers and the MNOs, so the plans for supporting roll-out of the project could vary substantially by carrier or location. This would complicate the development of an effective encouragement design unless the study was implemented entirely in the catchment area of one local provider. When the details of programme implementation differ substantially across study areas, the evaluation assumption of a 'stable treatment' is violated and measured effects are usually weaker. For all of these reasons, we believe that the Tanzania mHealth project is a better choice to include in the evaluation than Ghana mHealth.

Next, we considered the mAgri projects. Here, we consider the Ghana mAgri project is a stronger project, with feasibility for evaluation similar to the Bangladesh mAgri project. Regarding the strength of the product, as described in the section below on Ghana mAgri, mNutrition content will be included as part of the Vodafone Farmer Club intervention. The substantial strengths of this project include its use of agricultural content provided by Esoko, a well known and established mAgri platform which has well-developed plans for a substantial roll-out, and it includes a service that provides free calls between members of the Farmer Club. The Bangladesh mAgri project, on the other hand, is being developed by MNO Grameenphone with content and roll-out support provided by WIN-Miaki. WIN-Miaki will aggregate content from Helen Keller, WorldFish, and Asanhiya Mission, provide it to Grameenphone, and then use these partners as distributors of their product. WIN-Miaki operates call centres that would help to support the project, but its current mAgri service, the 27676 helpline, has a very small subscriber base, so the new WIN-Miaki product will be rolled out in relatively untested markets where there is less experience in providing services to farmers through mAgri platforms. Moreover, its content is not built on the strength of a product such as Esoko in Ghana.

We consider the feasibility of evaluating the Ghana mAgri intervention to be strong. Vodafone is a strong MNO partner with a large subscriber base that will be adding mNutrition messages to an established mAgri product. Although the 'aggregator' approach to mAgri in Bangladesh –

which pools content from multiple providers and redistributes it through their platforms – is of interest, it presents greater risks to implementation of the service due to differing interests or implementation capacity of local content providers. The quality of the evaluation would also be weaker if the service differed substantially across providers. Also, through recent communication, we have begun to develop stronger ties with the GSMA staff for mAgri in Ghana, and we received positive initial feedback on our plans from Vodafone. In Bangladesh, WIN-Miaki was also open to the evaluation, though its plans were not as developed as those of Vodafone in Ghana. **For these reasons, we believe that the Ghana mAgri project is a better choice to include in the evaluation than Bangladesh mAgri.**

We have suggested that Tanzania mHealth and Ghana mAgri are the strongest projects to study, with strong feasibility for an evaluation, but we must also consider the cost of conducting these evaluations. Although Tanzania mHealth has the most expensive fieldwork budget of the four potential studies, and Ghana mAgri will be more expensive than Bangladesh mAgri, we can nevertheless complete the Tanzania mHealth and Ghana mAgri evaluations within the total budget already provided in the contract, for labour, travel and fieldwork. For mHealth, the Tanzania study presents the best case for completing a rigorous evaluation of a high-quality integrated mHealth and mNutrition platform. An important justification for recommending Tanzania mHealth over Ghana mHealth is that the large current subscriber base for WN in Tanzania demonstrates capacity for successful large-scale implementation of an mHealth platform. The fieldwork for the Tanzania mHealth evaluation costs more than for the mHealth evaluation in Ghana, but we believe the risks to a high-quality evaluation are lower for the Tanzania mHealth study. For mAgri, replacing Ghana with Bangladesh in mAgri would save in fieldwork costs, but we do not believe that these savings justify switching our proposal to evaluate Ghana mAgri. As described above, the Ghana mAgri platform proposes substantially lower risks to an effective evaluation than the one in Bangladesh.

3. Quantitative impact evaluation design

3.1 Ghana mAgri impact evaluation design

3.1.1 mAgri programme and the local context

Goals and objectives

The mAgri programme aims to enhance the livelihoods and quality of life of developing-world smallholder farmers by improving access to information, financial services and supply chain solutions, delivered via mobile. mNutrition aims to promote behaviour change around key farming decisions and practices via mobile nutrition content. The objective of mNutrition in mAgri is to create and scale commercially sustainable mobile services, enabling smallholder farmers to improve their nutrition, yields and incomes. The stated GSMA targets are the following (GSMA M4D 2013):

- At least 20 per cent of registered households that act on information and advice report consuming at least **four food groups** on a daily basis for at least nine months of the year as a result of more diverse agricultural output, increased income and/or behaviour change in terms of nutrition.
- At least 50 per cent of registered households that act on information and advice report a 25 per cent increase in **agricultural productivity**.
- At least 50 per cent of registered households that act on information and advice report increases in **agricultural income** of 20 per cent.

Product

The product to be delivered and evaluated in Ghana is the Vodafone Farmer Club. The service is a bundled solution offering agricultural information in addition to voice and SMS services. The value-added services components include:

- discounted SMS and Voice SMS;
- free calls to other Farmer Club users;
- weekly agricultural content provided by Esoko that includes:
 - **price information:** wholesale and retail prices per market, prices available weekly on 35 markets and daily on ten most important regional capital markets; 56 commodities available
 - **weather information:** daily SMS weather predictions including rainfall
 - **farm tips/extension information:** disease awareness and prevention, best agricultural practices (livestock and crop)
 - **nutrition tips:** information on nutritional value of crops, storing, processing and preservation
- caller-ring-back tones;
- access to Farmer Helpline and Vodafone Customer care.

The Vodafone Farmer Club product is available only through a dedicated Farmer Club SIM and is activated upon subscribing monthly to the service. This offer is strictly for farmers who will be identified by retail staff and special agents. The cost for the service was originally 2 cedis/month, but in order to reach the intended number of subscribers, promotional offers of the service for free started in October 2016.

The mode of content will be SMS for weather and price information and voice messages for agriculture and nutrition information. While SMS will be in English, voice messages will be

available in five local languages. In total, three agriculture messages per month and one nutrition message⁵ per month will be sent to the subscriber, along with weather and market information.

Targeting

The primary targets for Vodafone's Farmer Club are smallholder farmers with access to mobile telecommunications and farmers involved in small-scale farming and livestock. The target market Vodafone expects to attract is about 450,000 Farmer Club users by 2016/17.⁶ The solution will target eight regions in Ghana: Eastern Region, Western Region, Ashanti Region, Central Region, Northern Region, Volta, Brong-Ahafo and Greater Accra.

Marketing and roll-out

Some 71 districts have been selected to roll out Farmer Club product. These districts were selected based on cell tower sites (3G connectivity) and crops (e.g. not selecting districts where households are mostly fisherman because there is no market information on fish).

The global launch for the Farmer Club product was on 26 May 2015 and the full commercial launch on 10 June. Initial aggressive marketing occurred in one strategic trading centre in the North and South Zones of Ghana. Thereafter, a lower level of marketing followed in each region. A range of the following 'permanent' agents enrolled customers:

- ambassadors – roughly one per district;
- retail agents;
- community shops;
- input dealers.

Additional marketing included a roadshow that was present at strategic events, such as National Farmers Day, organised by the Ministry of Food and Agriculture and by non-governmental organisations (NGOs).

Context

Nutrition: Child stunting is 28 per cent nationally in Ghana and higher in the Northern (32%), Eastern (38%), and Central (34%) and Upper East (36%) regions (Ghana Statistical Service *et al.* 2009). Anaemia prevalence is also very high, with rates of 76 per cent in children 6-59 months of age and 62.4 per cent in women (Ghana Nutrition Landscape Analysis n.d.). Varied and high-quality diets are key to addressing child and maternal undernutrition. The percentage of children 6-23 months who consume the minimum diet diversity of four food groups is 46.8 per cent and on average women consume four food groups (Kothari and Nouredine 2010).

Mobile penetration in Ghana has risen dramatically in the past ten years, increasing from less than 20 subscriptions per 100 people in 2005 to 108 subscriptions per 100 people in 2013 (World Bank 2010). According to the Ghana Living and Standards Survey (GLSS Round 6 2014), mobile phone penetration in 2013 was 80 per cent in Ghana, with 70 per cent of rural households reporting owning a phone and 88 per cent of urban households reporting owning a phone (GLSS Round 6 2014). However, access to mobile phones in Ghana varies dramatically by region, socioeconomic status and gender. In USAID's Feed the Future zone of influence (districts in Northern, Upper West and Upper East Regions)

⁵ Plans to increase the number of nutrition messages to three per month in 2017 are currently being negotiated.

⁶ This figure was revised down to 200,000 in 2016.

only 38 per cent of males and 41 per cent of females report having a mobile phone in the household (USAID 2012). Access to mobile phones in these regions is also lower among females, with only 14 per cent saying they own most of the phones, while 57 per cent of males saying they own most of the phones.

Literacy in Ghana: According to the GLSS Round 6, adult literacy rates in rural areas are quite low, with only 41.7 per cent of adults knowing how to read or write in English or any Ghanaian language.⁷ Among rural women, rates are even lower, at 31.4 per cent. These low rates have implications on the design of the Farmer Club product and its ability to reach an illiterate population.

Agriculture in Ghana: A little over half (51.5%) of households in Ghana own or operate a farm. Farming is predominantly rural, with 82.5 per cent of rural households involved compared to 26.6 per cent of urban households.⁸ The proportion of females involved in agriculture is 41.2 per cent, and there is virtually no difference in urban and rural areas. The main crop harvested is maize, followed by cocoa and groundnut/peanut.

3.1.2 Research questions

The research questions that will be addressed through the experimental design described in more detail below relate to the first two questions stated in the TOR and the specific objectives of GSMA. In particular, the design will address the following questions:

- How effective is the Farmer Club at increasing the knowledge and changing the behaviour of farmers?
- What are the impacts of the Farmer Club product on household's dietary diversity, agricultural income, and production?

In addition to these two research questions, the impact evaluation will address two additional questions:

- What is the demand for the Farmer Club product and can framing about the agriculture or nutrition objectives of the product affect household's willingness to pay?
- Does targeting women have differential impacts on knowledge, behaviour and final outcomes than targeting men?

3.1.3 Impact evaluation design

Design

Given that the Farmer Club will be available to all farmers in the 71 districts of Ghana, an RCT within these districts where we randomly assign some individuals or communities to a true 'control' group that does not have access to the service is not an option. Moreover, comparison of farmers within the 71 districts to those outside the 71 districts would not lead to causal estimates because farmers within and outside the 71 districts are likely to be very different. Vodafone purposefully chose the 71 districts based on their access to a 3G cell tower and on their crop cultivation to ensure that farmers would be able to receive the messages and that the Esoko price information was relevant to them. Consequently, farmers not in the 71 districts are likely to have less access to a 3G cell tower and engage in different farming activities. Thus, in order to estimate the causal impact of the Farmer Club product, we propose implementing a randomised encouragement design. The encouragement design

⁷ GLSS Round 6, August 2014.

⁸ Ibid.

will not restrict access to the Farmer Club product, but instead will work by randomly assigning some communities to receive additional marketing and promotion of the programme.

The additional marketing and promotion to encourage take-up and continued use is likely to be a combination of price discounts, SMS blasts, and door-to-door marketing to farmers in selected communities throughout the evaluation period. The initial qualitative scoping study as part of the qualitative work stream of this impact evaluation will explore different potential encouragement strategies and identify the most effective approach to encourage take-up (see Chapter 4 for further details).

During the door-to-door marketing, we will promote the product using a short advertisement script on the value added of the service. Households will be randomly assigned to receive one of two scripts: 1) a script that focuses on the agriculture value added of the product (Vodafone's current script), or 2) a script that augments the agriculture focus with additional information about the nutrition value added of the product. Comparing outcomes from the two scripts will indicate whether emphasising the nutrition component of the programme leads to larger changes in impact of the programme on outcomes such as household diets. Lastly, we will randomly target either a male or female from each household to receive the advertisement scripts and free subscription to Farmer Club. Comparing outcomes between male- and female-targeted households will indicate whether the gender of the person receiving the messages affects the household's utilisation of the information provided.

Thus the encouragement design will be composed of the following five groups:

- **Comparison group (Group 1)** – enumeration areas that are not receiving the extra marketing or promotion;
- Encouraged group – enumeration areas that receive the extra marketing and promotion in the form of door-to-door marketing, blast SMS to farmers, and price discounts:
 - **Encouraged male, agri group (Group 2a)** – Households that receive marketing scripts that focus on the agriculture value added of the product and target a male household member;
 - **Encouraged male, agri+nutrition (Group 2b)** – Households that receive marketing scripts that focus on the agriculture and nutrition value added of the product and target a male household member;
 - **Encouraged female, agri group (Group 2c)** – Households that receive marketing scripts that focus on the agriculture value added of the product and target a female household member;
 - **Encouraged female, agri+nutrition group (Group 2d)** – Households that receive marketing scripts that focus on the agriculture and nutrition value added of the product and target a female household member.

Assignment to the different intervention groups will occur in two stages. The first stage that assigns enumeration areas to either the comparison group (Group 1) or encouraged group (Group 2a, Group 2b, Group 2c, Group 2d) will occur at the enumeration area level. We chose to randomise at the enumeration area level as opposed to household level because it is likely that individuals will discuss what they learn from the Farmer Club with other community members; thus, even individuals who do not directly use the service may be exposed to the information through their community members and therefore cannot be considered 'untreated'. The second stage of randomisation that assigns households to either Group 2a, 2b, 2c or 2d will occur at household level for households in the encouraged enumeration areas.

The proposed design will allow us to answer our specific research questions by making the following comparisons:

- Comparison of **combined encouraged group** (Groups 2a, 2b, 2c, 2d) with **comparison group** (Group 1): What is the absolute impact of the Farmer Club on household outcomes and behaviour change relative to the comparison group?
- Comparison of **encouraged male group** (Group 2a and 2b) with **encouraged female group** (Group 2c and 2d): What is the relative impact of targeting women on household outcomes and behaviour change?
- Comparison of **encouraged agri group** (Group 2a and 2c) with **encouraged agri+nutrition group** (Group 2b and 2d): Does framing the Farmer Club as an agriculture and nutrition programme lead to differences in a household's willingness to pay compared to framing as only an agriculture programme.

As described in more detail below, the estimation methodology will compare differences in our outcomes of interest across the encouraged and comparison groups, using data collected in baseline and endline surveys. For both baseline and endline, we will collect detailed information on: (1) final outcomes on which we expect to see impacts; (2) intermediate outcomes that may explain pathways of impact, such as changes in behaviour, knowledge and practices; and (3) outputs such take-up rates and factors that may affect take-up rates and use of the product. The quantitative data collection and analysis strategy, as well as the interpretation of the analysis findings, will be informed by findings from the qualitative data collection rounds.

Estimation strategy

Because the encouragement is randomly assigned, we will use the systematic variation in take-up of the product to measure the causal impact of the programme as the difference in outcomes between encouraged and comparison communities. The random assignment of enumeration areas will be stratified by region and occur within each of the regions chosen for the impact evaluation. Stratification will guarantee that, within each region, each intervention arm is roughly equally represented. Random assignment ensures that baseline characteristics of children, households and communities will be similar across encouraged and comparison communities, minimising bias in impact estimates due to unobserved heterogeneity or selection.⁹ Similarly, coverage of other similar agriculture and nutrition interventions, past and current, and access to services should be balanced across the encouraged and comparison communities as a result of randomisation, which should limit the effect of confounding variables on the impact estimates. As a result, average differences in outcomes across the groups after intervention can be interpreted as being truly caused by, rather than simply correlated with, the interventions.

The estimation methodology will compare differences in outcomes for interest across the comparison and encouraged groups and within the different encouraged groups, using data collected in baseline and endline surveys. The baseline survey will be conducted before the extra encouragement is implemented, and the endline survey will occur two years after the baseline. To estimate the impact of the Farmer Club product, we will use a combination of analysis of covariance (ANCOVA), single difference and double difference techniques, depending on the outcome of interest. We will analyse our outcomes primarily with ANCOVA models, using difference-in-difference models and single difference models as robustness checks. ANCOVA models are more flexible than typical difference-in-difference models when autocorrelations are low because it allows us to estimate rather than impose the

⁹ A post-randomisation test will be conducted to ensure that the intervention arms are balanced across key characteristics.

autocorrelation in each outcome.¹⁰ Moreover, there are substantial power gains of using ANCOVA models over difference in difference when autocorrelation is low, which is likely to be the case with many of our outcome variables.¹¹

Using the ANCOVA model, we will estimate the intent to treat (ITT) estimate because treatment is encouragement to participate in the Farmer Club rather than actual participation in the programme. The average treatment effect (ATE) is the actual effect of the product. However, because compliance is not perfect, and not all who are encouraged will take up the product, we are measuring the ITT effect. In addition to the ITT estimate, we will estimate the local average treatment effect (LATE) of the Farmer Club product using instrumental variable techniques, which use the random variation in encouragement as an instrument for take-up of the product. LATE estimates the effect of the Farmer Club product on those whose treatment status was affected by the encouragement.

For comparison of the **combined encouraged group** (Groups 2a, 2b, 2c, 2d) with the **comparison group** (Group 1), the exact empirical specification on the ANCOVA parametrisation in its simplest form is the following:

$$Y_{1hv} = \beta_0 + \beta_1 Encouraged_v + \beta_2 Y_{0hv} + \varepsilon_{hv}$$

where Y_{1hv} is the outcome of interest at endline for household h from enumeration area v , Y_{0hv} is the outcome of interest at baseline, and $Encouraged_v$ is an indicator for whether or not enumeration area v received the extra encouragement. β_1 measures the differences in outcomes of the encouraged versus comparison enumeration areas, and thus the impact of the Farmer Club product.

For comparison of the **encouraged female group** (Group 2c, 2d) with the **encouraged male group** (Group 2a, 2b), the exact empirical specification on the ANCOVA parametrisation is the following:

$$Y_{1hv} = \beta_0 + \beta_1 Male_{vh} + \beta_2 Female_{hv} + \beta_3 Y_{0hv} + \varepsilon_{hv}$$

where $Male_{vh}$ is an indicator for whether household h in enumeration area v targeted the Farmer Club service to a male and $Female_{hv}$ is an indicator for whether the Farmer Club service was targeted at a female. β_1 measures the impact of the Farmer Club product when it is targeted at males and β_2 the impact when it is targeted at females. To test whether the ITT estimators are statistically different across male and female groups, we conduct Wald tests of equality of the two estimates.

Heterogeneity

The absolute and relative impacts measured for the Farmer Club may depend on baseline characteristics of the study sample. In particular, the two regions where we plan to conduct the study, Central Region and Upper West Region, are very different in terms of seasons, agriculture and nutrition. Consequently, we plan to measure heterogeneity of impact by region, following Bruhn and McKenzie (2009). For the first-stage randomisation, we will stratify the sample of enumeration areas by region and will randomise assignment to the encouraged (Groups 2a, 2b, 2c, 2d) or comparison group (Group 1) across these two region

¹⁰ McKenzie, D. (2012) 'Beyond baseline and follow-up: the case for more T in experiments', *Journal of Development Economics* 99(2): 210-221.

¹¹ The ratio of the difference in differences variance to the ANCOVA variance is $2/(1+\rho)$. So when $\rho=.25$, with a single baseline and follow-up, the sample size needed is 60 per cent higher with difference in differences than with ANCOVA to get the same power.

strata. This will help to assure even coverage of the intervention arms across regions, and will facilitate subgroup analysis.¹²

The programme may also have different effects for households where women own their own mobile phone, as opposed to those who share with their husband or other household members. Consequently, we plan to also measure heterogeneity of impact by whether or not females in the household own their phone.

Gender

Gender is incorporated across all stages of the research, from promoting gender-specific encouragements to conducting gender-related analysis on impact. For the evaluation this entails ensuring that the survey instruments are designed and implemented in such a way that males and females are included in quantitative data collection and that, as appropriate, data collected can be disaggregated by gender, analysis can be disaggregated by gender of the Farmer Club user, and outcomes can be disaggregated by gender of household head.

Limitations (spillovers, etc.)

One limitation of the study design is that we will not be able to distinguish the impact of the encouragement from the impact of the Farmer Club product. Individuals may change their nutrition knowledge, attitudes and behaviours because of the encouragements, independent of their use of the services being advertised. With this design, we measure the impact of the full bundle, both the encouragement and use of the services. Similarly, our encouragement could lead to a different type of user signing up for a product than would otherwise sign up. While we recognise this limitation, we also believe that any successful mobile information platform will have a strong marketing/encouragement component, and thus measuring the impact of the full bundled product (product + encouragement) is still of interest.

Spillovers are another potential limitation. Given that the first stage of randomisation (encouraged or comparison groups) will be conducted at the enumeration area level, we expect minimal bias in our impact estimates due to spillovers or sharing of information across intervention groups. Prior to randomisation, eligible enumeration areas will be grouped together into clusters of enumeration areas in cases where two or more enumeration areas are very close together in terms of distance. Enumeration area clusters will then be randomised within region for the purposes of the evaluation. All enumeration areas within a cluster will receive the same encouragement intervention. This approach should minimise the potential for spillovers across intervention groups.

For the second stage, randomisation (Groups 2a, 2b, 2c, 2d) will occur across households within the same enumeration areas. In order to minimise spillovers, we will cluster households in the same compounds or concessions prior to randomisation, and all households within the cluster will receive the same household-level treatment. Women still might share information with other women outside of their compound cluster, and we will try to quantify how much this is occurring by collecting data on information-sharing across households in our endline survey. The issue of information-sharing will also be explored further in the qualitative midline data collection.

3.1.4 Outcomes of interest

Primary outcomes

The primary outcomes of interest are those related to GSMA's stated goals:

¹² We recognise that we may not be powered to detect impacts in the subgroups, but even coverage will facilitate analysis in the case that we are powered.

1. household and women's dietary diversity;
2. agricultural productivity;
3. agricultural income.

Household and women's dietary diversity: Household dietary diversity indices will be constructed using information collected on all food the household consumed in the seven days prior to the survey. This information will then be used to construct a Household Dietary Diversity Score (HDDS), which combines food items consumed into 12 food groups. The HDDS indicates a household's economic access to food, and thus included in the score are items that require household resources to obtain, such as condiments, sugar and sugary foods, and beverages (Kennedy, Ballard *et al.* 2011).

Individual food consumption of women will also be collected and used to construct the Women's Dietary Diversity Score (WDDS) and the Minimum Dietary Diversity-Women (MDD-W). The WDDS is composed of nine food groups and reflects the nutritional quality of the diet (*Ibid.*). The MDD-W reflects consumption of at least five of ten food groups and the greater likelihood of women meeting their micronutrient needs than women consuming foods from fewer food groups.

Agricultural productivity and income: A detailed agriculture module will be implemented in order to construct accurate measures of productivity and income. The agriculture module will include questions on crops, inputs, yields, sales and prices.

The findings from the quantitative outcome indicators will be combined with findings from the qualitative data to redefine, contextualise and add different dimensions (including participants' understanding and perspectives on dietary diversity, agricultural productivity and income, as well as changes in these). This triangulation of quantitative and qualitative indicators will help to increase the validity of the primary outcomes and add different dimensions.

Secondary outcomes

Secondary outcomes include a large set of variables to measure: (1) intermediate outcomes, such as nutrition knowledge and behaviour, and knowledge and practice of farming techniques; and (2) project outputs and factors that affect take-up and use of the product. Below are some examples under each set of variables.

1. *Nutrition knowledge and behaviour:* knowledge of vitamin A and iron rich sources of food; knowledge of nutrient-dense crops, use of homestead food production and production of nutrient-dense, diverse crops; raising of small livestock for consumption; household food preservation techniques;
2. *Knowledge and practice of farming techniques:* improved crop handling, management and storage, fertiliser use;
3. *Take-up and use of Farmer Club:* number of households in sample subscribing to Farmer Club, number of months subscribing, most used components, female access to messages, sharing of messages with others within and across households.

In addition to these intermediate outcomes, we are interested in collecting and analysing data on intra-household dynamics, especially related to female status and decision-making. Therefore, we will collect data on nutrition knowledge and behaviour and farming knowledge and behaviour for males and females within the same household. This will allow us to see if there are any differences between males and females in their knowledge. In addition, we will collect indicators on female control over productive assets and female roles in decision-

making. Qualitative observations and in-depth interviews will further explore whether, how and why the intervention might have influenced intra-household dynamics and female status within the household.

The findings from the two qualitative data collection rounds will be used to inform the development of the baseline and endline surveys and to triangulate and contextualise the findings.

Willingness to pay

Willingness-to-pay studies are designed to ascertain the maximum amount that an individual will pay for a particular product or service. Willingness to pay (WTP) can be measured through either: 1) revealed preferences, by observing prices being paid in markets or expenditures being made on the product or service; or 2) stated preferences, by directly asking individuals what prices they would be willing to pay for the product or service. In our encouraged group, we will measure WTP from revealed preferences using a 2-step procedure where we use the Becker-DeGroot-Marschak (BDM) method in the first step. In BDM, the farmer will be asked how much they are willing to pay for the Farmers' Club product, then a random price is drawn. If their bid is greater than or equal to the price, they are offered the product at the randomly drawn price. If their bid is below the price, they are not offered the product. In the second stage, regardless of the outcome of the first stage, there is another opportunity to receive the Farmer Club at an additional discount. Farmers again select a random price, but this time all farmers are offered the service for free. The two stages are necessary in order to first elicit a farmer's WTP and then to offer the product for free to all farmers in the encouraged group.

3.1.5 Sampling strategy and data collection

Sampling strategy

The study will be undertaken in five districts in the Upper West Region and five in the Central Region, for a total of ten districts across two regions. The ten districts selected are based on Esoko market price information for crops, and on low Farmer Club subscription rates. From each selected district we will randomly select 20-21 enumeration areas from a list of enumeration areas within a 10-mile radius of a cell phone tower.¹³ Urban enumeration areas without a cell phone tower will be dropped from the sampling list. A total of 207 enumeration areas (104 in the encouragement arm and 103 in the comparison arm) will be part of the study.

In each enumeration area we will randomly sample 19 farmer households, for a total sample of 3,933 households at baseline and 3,736 at endline (assuming a 5% attrition rate across rounds). The inclusion criteria into the sample are that households must: 1) be a farming household; 2) own a mobile phone; 3) not be a current member of Farmer Club; and 4) have at least one female member age 15-60 years old. The last criterion ensures that we can measure women's dietary diversity (a primary outcome) in all our sample households. In order to know which households meet our sampling criteria, a census in our selected enumeration areas will be conducted. The census will also allow us to calculate current take-up rates of the Farmer Club in each enumeration area before the start of the encouragement.

Sample size

Our sample size calculations are conducted for the first stage estimation, which compares the pooled encouraged groups to the comparison group. Given that randomisation for this

¹³ Urban areas that are made up of more than one enumeration area were clustered together for randomisation.

comparison is done at enumeration area level, the sample sizes needed to detect impact are more demanding because we need to consider that the error term may not be independent across individuals in the same enumeration area. In other words, outcomes of individuals in the same enumeration area may be correlated. Although sample sizes are likely to be more demanding for the first comparison, we also report the power for the given sample size for the second comparison across encouragement arms, in order to ensure that we are also powered for the second comparison.

For consistency across designs, and because it is one of the primary outcome indicators, we design the study to detect impacts on dietary diversity scores for women aged 15-49 years (number of food groups 0-9). This indicator has been validated by the World Health Organization (WHO) as a good predictor of diet quality and micronutrient density (Ruel *et al.* 2013). We obtain means, standard deviations (SDs) and intracluster correlations¹⁴ of women's dietary diversity index from the Demographic and Health Survey (DHS) 2008. We also design the study to detect impacts on agricultural production using yields of cocoa, which is the second largest crop in Ghana. We obtain mean yield, SDs and intracluster correlations from the GLSS Round 6 (2014).

We show calculations for two different minimum detectable effects.¹⁵ In particular, we show options for a 25 per cent increase over the mean value, in line with Akrofi *et al.* (2010) and Thorne-Lyman *et al.* (2010). We also show options for a 15 per cent increase for women's dietary diversity and a 35 per cent for cocoa yields. We use conventional levels and set the power¹⁶ at 80 per cent and the significance level at 0.05.

We put these designs and assumptions together using standard sample size calculations for cluster randomised controlled trials:¹⁷

β

$$n_c = 2\sigma^2 \left[\frac{(z_{\alpha/2} + z_{\beta})^2 [1 + (m - 1)\rho]}{d^2} \right]$$

Where n_c =required sample size per arm, $z_{\alpha/2}$ =1.96, z_{β} =0.85, m =number of observations per cluster, ρ =intracluster correlation coefficient, σ =population SD of outcome variable and d =minimum detectable effect.

We conduct the power calculations for ANCOVA models, which take into account the autocorrelation from baseline to endline. For women's dietary diversity index we assume an autocorrelation of 0.1, which is taken from data in Uganda, and for cocoa yields we assume an autocorrelation of 0.3, which is similar to autocorrelations in Ghana for food expenditures. We assume 5 per cent attrition from baseline to endline.

Because we use an encouragement design, this formula needs to be amended from its use in RCTs in one crucial way: in a standard RCT we set d to the size of the effect we intend to detect between the 100 per cent treated group compared to the 100 per cent untreated (control) group. In an encouragement design, some individuals assigned to the encouraged group may not take up the treatment, while some individuals assigned to the comparison group may take up the treatment. This in essence dilutes the difference in effect sizes

¹⁴ The intracluster correlation is the fraction of the total variance of an outcome that can be explained by the within-cluster variance.

¹⁵ The minimum detectable effect is the smallest true treatment effect that a research design is likely to identify as statistically significant.

¹⁶ Power is the ability of a test to detect an effect, if the effect actually exists.

¹⁷ For example, Hemming *et al.* 2011.

between groups, which will increase the required sample size. With an encouragement design, d should be set to (take-up rate in encouraged group – take-up rate in comparison group) \times (expected effect on treated individual compared to untreated individual).

We show two different assumptions about take-up rates in comparison and encouraged areas. In the first, we assume a 5 per cent take-up of mNutrition in comparison areas and a 75 per cent take-up in encouraged areas, leading to a 70 percentage point (pp) take-up gap, which is very optimistic. In the second scenario, we again assume a 5 per cent take-up in comparison areas and a 50 per cent take-up in encouraged areas, resulting in a 45 pp take-up gap. Although these take-up rates are high, we hope to implement a strong encouragement informed by the qualitative fieldwork on which strategies are the most effective in increasing take-up rates in the specific context of the intervention.

Table 3.1 shows the required number of households (baseline and endline) for different outcomes of interest, possible effect sizes, and possible gaps in take-up rates between encouraged and comparison areas.

Table 3.1 Required sample sizes for different outcome variables and scenarios

Women: number of food groups consumed (mean=3.88, SD=1.45, ICC=.17, autocorrelation=.1)							
		Effect size					
		15%			25%		
		Number of clusters	Baseline	Endline	Number of clusters	Baseline	Endline
Take-up gap	70 pp	86	1,611	1,534	31	582	554
	45 pp	207	3,898	3,712	75	1,407	1,340
Cocoa yields (kg) (mean=956.61, SD=1190.39, ICC=.23, autocorrelation=.3)							
		Effect size					
		25%			35%		
		Number of clusters	Baseline	Endline	Number of clusters	Baseline	Endline
Take-up gap	70 pp	395	7,455	7,100	202	3,805	3,624
	45 pp	954	18,024	17,166	487	9,198	8,760

Although sampling 18,024 or 9,198 households at baseline is not likely to be possible given the budget, we strongly recommend sampling 3,898 households at baseline and 3,712 at endline. This will ensure that we are able to detect an effect on women's dietary diversity of 15 per cent under the more realistic take-up gap of 45 pp. It also ensures that we are able to detect an impact on cocoa yields if the take-up gap is 70 pp and the effect size is 35 per cent. In order to achieve 3,898 households at baseline, 207 clusters are necessary (104 in the encouraged group and 103 in the comparison group). With 207 clusters, we will need to sample 19 households per cluster at baseline for a total sample of 3,933 households at baseline.

With this sample size, we are also powered at 97.5 per cent to detect impacts of 15 per cent (assuming 45 pp take-up gap) on women's dietary diversity and we are powered at 81.5 per cent to detect impacts of 35 per cent (assuming 45 pp take-up gap) on cocoa yields when

comparing encouraged male targeted households with encouraged female targeted households.

Surveys

For the baseline and endline surveys, we plan to collect detailed information on: (1) outcomes on which we expect to see impacts; (2) intermediate factors that may explain pathways of impact, such as changes in behaviour, knowledge and practices; and (3) factors that may affect take-up rates and use. Quantitative data will be collected from intervention arms using a combination of surveys at household level and with both males and females. The design and exact questions and modules used will be informed by findings from the qualitative fieldwork.

The data collected at all levels will be panel data and thus the same households, males and females will be surveyed in each round. A baseline survey will be conducted before the start of the designed encouragement and an endline survey two years later. This schedule allows for a longer two-year period of intervention, which is necessary given the lag between planting and harvest seasons, and potentially slow take-up or learning. While some attrition of households is likely between baseline and follow-up, we will collect detailed information on households' location and their contact information in order to minimise attrition rates.

The surveys will include modules on demographics and education for all household members, as well as details on household food consumption, gender-disaggregated assets, household decision-making and intra-household bargaining, nutrition knowledge and practice, agricultural production, agricultural income, agricultural knowledge and practice, local agricultural markets, use of and access to mobile phones and other mAgri platforms, and savings and credit. These data will allow construction of dietary diversity measures, measures of nutrition knowledge and practices, and analysis of agricultural inputs and outputs, yields and income.

The endline survey will collect information on the same indicators included in the baseline survey, in addition to a detailed module on mNutrition usage, exposure, sharing of information, and impressions; usage, exposure and impressions of other mAgri programmes; and potential other topic areas identified as relevant by the qualitative midline data collection, such as factors influencing intervention take-up and outcomes and potential unanticipated and unintended (both negative and positive) outcomes.

3.1.6 Partners

The partners for this evaluation are GSMA, Vodafone, Esoko, and the Institute of Statistical, Social and Economic Research (ISSER) in Ghana.

GSMA is the organising partner working directly with Vodafone to incorporate mNutrition into its Farmer Club bundle.

Vodafone is the MNO offering the Farmer Club product and in charge of the marketing and roll-out of the product.

Esoko is the content provider, providing information on weather, prices and nutrition. The nutrition information will be developed by GAIN. In addition, Esoko provides the call centre for farmers.

ISSER is the local data collection partner that will be responsible for collecting baseline and endline data. ISSER is a research unit of the University of Ghana and has been recognised as a centre for comprehensive and sustained research and training in the social sciences. ISSER has worked previously with IFPRI and IDS on other research projects.

3.1.7 Risks and their mitigation

The biggest risk to this evaluation design is that the gap between take-up rates of the encouraged (Groups 2a, 2b, 2c, 2d) and comparison groups (Group 1) may not be as large as we expect it to be. Because the required sample size is related to the difference in take-up rates by an inverse square, small changes in the take-up gap have large effects on statistical power and thus our ability to detect treatment effects. This risk can be mitigated by choosing an encouragement that is likely to generate a large gap in take-up rates; for example, we will conduct door-to-door promotions and provide the product for free in addition to sending SMS messages to farmers to encourage them to use the product. Initial qualitative work to explore users' perceptions, preferences and attitudes towards different encouragement strategies will also help to ensure that an effective encouragement strategy is chosen.

The second risk is if initial take-up of encouraged groups is high, but use and renewed subscriptions are low. Given that behaviour change is complex and difficult (see the landscaping review conducted as part of the inception phase of this impact evaluation, Barnett *et al.* 2016), repeated exposure to behaviour change messages via mobile phones may help to reinforce and increase effectiveness of the messages. If users initially sign up for the product but then do not renew their subscription, it is not likely that we will see an impact. Thus a discount for a prolonged period, and SMS messages and nudging for a prolonged period, are necessary to encourage continuous use. The qualitative midline will explore people's perceptions, attitudes and opinions on the prolonged encouragement.

A third risk is that our partner organisations will not be interested in the evaluation or coordinating marketing strategies/encouragement. While we have tried to design the study in such a way that it interferes little with Vodafone's current marketing strategy, the company's general cooperation is key to the success of the evaluation. In order to mitigate this risk, we have started to engage with Vodafone early in the process and will continue to engage with the company throughout the project.

Additional risks are present in this impact evaluation that would exist regardless of what specific design is chosen, but they are important to keep in mind. First, the mAgri landscape

is swiftly changing in Ghana, with many new services and products being offered. It is possible that another organisation will introduce a competing product in our study areas that interferes with the take-up of this product. We will mitigate this risk by remaining in contact with Vodafone, Esoko and GSMA, and at least try to remain aware of what other groups are doing. We will also try to collect data on competing products available during our baseline survey. Second, the qualitative data collection (both the initial scoping study and the midline) will also collect data on the institutional, sociocultural, environmental and economic contexts of the interventions, and on the presence, awareness and use of competing mobile phone-based products, and will explore how these factors may enhance and/or hinder take-up.

3.2 Tanzania mHealth impact evaluation design

3.2.1 mHealth programme

Goals and objectives

The GSMA mHealth programme brings together the mobile industry and health stakeholders to improve health outcomes. The objective of mNutrition in mHealth is to provide vulnerable pregnant women and caregivers of children under the age of five years old with access to a suite of basic health and nutrition information and services, delivered through a single consolidated access point on their mobile phones. The stated targets for mNutrition in mHealth are the following:

- 5% reduction in **stunting** in children under five years old;
- 5% improvement of body mass index (**BMI**) in pregnant mothers;
- 15% of those registered in mNutrition services change **nutrition behaviour**.

GSMA hopes to work with partners that already have mobile health platforms in place in Tanzania and elsewhere to embed new nutrition content into previously existing services. GSMA hopes that this strategy will reach the greatest number of people with the greatest cost efficiency.

Components/product

Content providers

GAIN will contract a local content provider to adapt the global factsheets developed by CABI and the wider content development consortium for the Tanzanian context. The products of the local content provider are required to be 'open access' and available to any platform that wishes to use the content.

Local mobile health platforms

Wazazi Nipendeni (WN)

GSMA has identified WN, under the mHealth Tanzania Partnerships initiative, as a strong potential partner. Since November 2012, WN has run the 'Healthy Pregnancy, Healthy Baby' SMS programme, which sends 150 SMS messages in Swahili to women up to 16 weeks post-partum on a range of pregnancy and early childhood issues. Nutrition is a small component of the original service. WN has negotiated with all of the MNOs in Tanzania to 'zero-rate' (provide to the consumer for free) the original package of messages as a part of their corporate social responsibility programmes. WN was already planning to expand its programme to 300 messages delivered to caregivers of children up to five years old. Since 2012, WN has signed up 800,000 unique subscribers, of whom 100,000 are active in the network currently.

In addition to its original service, WN is close to signing a memorandum of understanding (MOU) with GAIN to start an add-on mNutrition service with the GAIN content. WN will design and manage the platform. It was due to begin negotiations with MNOs on the business model in June 2015 and expected that customers would be able to sign up for 'packages' of messages (e.g. late pregnancy, birth to age one). The price of the packages is expected to be 'zero-rated'.

Additional platforms

Because the content developed by the local content provider will be open access, new and additional entrants are possible and probable. For example, a small firm called Green Telecom is intending to develop an independent mNutrition platform available to AirTel subscribers. We chose to focus our impact evaluation work with WN because it is the most established, is the only organisation to have signed MOUs with respect to building an mNutrition platform, and has a solid reputation in the field with government and citizens.

Targeting

WN is targeted at pregnant women and their caregivers (husbands, etc) and the primary caregivers of children under five. It is available to women nationally, on all phone networks.

Marketing and roll-out

WN is advertised nationally through a multi-media campaign, and also works with health clinics and community health workers to sign up pregnant women.

The new mNutrition add-on will be advertised through push SMS to WN customers. The new mNutrition service will be launched in late 2016.

Context

Nutrition outcomes: 53 per cent of children aged 24-35 months are stunted; 39 per cent of women in rural areas are anaemic, of whom 10.2 per cent are moderately or severely anaemic.¹⁸

Nutrition behaviours: 81 per cent of children 0-1 month old are exclusively breastfed, while only 23 per cent of children 4-5 months old are exclusively breastfed; 72 per cent of rural women with children under three years old reported eating vitamin A-rich foods in the past 24 hours, while only 32 per cent reported consuming iron-rich foods.¹⁹

Mobile penetration: 63 per cent of all households have access to a mobile phone, including 50 per cent of rural, unbanked, poor households.²⁰

Literacy: 61 per cent of women aged 15 and over are literate, while 76 per cent of similarly aged men are literate (World Bank 2010).

The initial qualitative scoping study and the qualitative midline aim to further study the socio-economic, political, institutional and environmental contexts within which the mNutrition intervention will be implemented in Tanzania.

¹⁸ DHS 2010

¹⁹ Ibid.

²⁰ InterMedia FITS study of households in Tanzania. N=2,980 households, interviewed April-May 2012.

3.2.2 Research questions

The research questions that will be addressed through the experimental design described in more detail below relate to the first two questions stated in the TOR and the specific objectives of GSMA. In particular, the design will address the following questions:

- How effective is mNutrition at increasing the nutrition knowledge and changing the behaviour of pregnant women and caregivers of children under three years old?
- How effective is mNutrition at increasing knowledge of Infant and Young Child Feeding Practices (ICYF)?
- What are the impacts and cost effectiveness of the mNutrition product on women's and children's dietary diversity?
- Does mNutrition improve ICYF?
- Does mNutrition lead to improvements in the nutritional status of children under three years old?

3.2.3 Impact evaluation design

Design

Because the new mNutrition service is layered on top of existing programmes, it will be difficult or impossible to evaluate the impact of mNutrition on nutrition outcomes independent of the impact of the existing programmes. For the main evaluation, we propose partnering with WN to evaluate the impact of the original WN service plus the new mNutrition service. We believe this is of interest to DFID because the evidence on the impact of mobile-based behaviour change communication programmes is limited, particularly of programmes implemented at scale. While we cannot evaluate the impact of mNutrition alone, we will be able to evaluate the impact of a large mobile platform health service.

For the main evaluation, we will use a cluster-randomised controlled trial on a panel of households, mothers and children.

Cluster-randomised controlled trial

WN and mNutrition will be available to all mobile phone users in Tanzania, and thus it will be difficult to guarantee that individuals randomly assigned to a control group do not receive access to the bundled programme during the study period. However, by selecting a study region – Iringa – where WN has no existing relationships with health clinics or other NGOs, and therefore where use of the basic WN product is extremely low, we can limit the potential take-up of the combined WN and mNutrition programme in control group areas. Further, we will leverage fieldwork activities to ensure that take-up of the mNutrition programme is high in treatment areas. We will do so by having our field team ask for consent to receive the WN and mNutrition content on mobile phones when they visit sampled households in treatment areas. WN will then begin sending the WN and mNutrition content to the mobile phones of consenting individuals as quickly as possible. In effect, this amounts to a door-to-door offer of the programme for sampled households in treatment areas. We expect this tactic to be highly effective at inducing households in treatment areas to take up the WN and mNutrition product.

In practice, the differences between the cluster-randomised controlled trial discussed above and the randomised encouragement design that was previously proposed are minimal. We expect the door-to-door offer of the WN and mNutrition programme to generate a take-up gap that is high relative to those discussed in the context of the randomised encouragement design. As a result, we expect to have substantially higher statistical power than we would

have under a randomised encouragement design. Under conservative assumptions, we expect the difference in take-up between treatment communities and control communities to be 70 percentage points, which corresponds to the large take-up gap under the randomised encouragement design. In determining the necessary sample size to achieve reasonable minimum detectable effects (MDEs), we therefore rely on the power calculations conducted using the 70 percentage point take-up gap.

To reduce the possibility of spillovers between households assigned to the treatment and control groups, we plan to randomly assign treatment at community level – where we define communities as villages in the three rural districts of Iringa region. We think it is likely that individuals may discuss what they learn from the products with other community members; thus, were we to randomise at individual level, even individuals assigned to be ‘untreated’ would be exposed to the service through treated individuals in their community.

Panel elements

In many studies, the ideal design consists of measuring the same individuals at baseline and endline, because controlling for baseline measures improves the statistical power with which one can measure impacts at endline. This study has two panel elements. The first panel sample will be composed of households with children 0-11 months old at baseline. This means that two years later, at endline, we will survey the same mother and child, who will be 24-35 months old. Consequently, for older children (24-35 months at endline), we will have baseline outcome measures and can estimate the impact of mNutrition using ANCOVA models. In this panel element, our primary outcome of interest is child anthropometry. The second panel sample will be composed of households with a pregnant woman. At endline two years later, the children in these households will be 18-24 months old. Combining the mothers in both of these panels, we can estimate the impact of mNutrition on women’s dietary diversity using ANCOVA models.

Cross-section elements

While panel designs are attractive for their improvements to statistical power, some of the main outcomes of interest in this study are outcomes that change with the lifecycle, thus it is not possible to control for the measure baseline. For example, proper feeding practices differ for six-month old children and 30-month old children, thus it would not make sense to control for feeding practices at six months old when estimating the impact of mNutrition on children at 30 months old. Instead of a panel design for these types of outcomes, we will use the same samples of people but estimate effects by OLS (ordinary least squares) using only the endline data as a cross-section. While random assignment ensures that the error term in this OLS regression will be uncorrelated with our encouragement treatment, we will include controls for the lifecycle phase of children to increase the precision of our estimates. We will estimate the impact of WN and mNutrition by comparing endline outcomes between those in encouraged communities and comparison communities. For the pooled cross-section of children 18-35 months old at endline, our primary outcome of interest is children’s dietary diversity.

Estimation strategy

Because the offer of treatment is randomly assigned, we will use the systematic variation in take-up of the product to measure the causal impact of the programme as the difference in outcomes between treatment and comparison communities. Random assignment ensures that baseline characteristics of children, households and communities will be balanced across treatment and comparison villages, minimising bias in impact estimates due to unobserved heterogeneity or selection. Similarly, coverage of other interventions, past and current, should be roughly equal across the treatment and comparison villages as a result of randomisation, which should mitigate the effect of confounding variables on the impact estimates. As a result, average differences in outcomes across the groups after intervention

can be interpreted as being truly caused by, rather than simply correlated with, the interventions.

To estimate the impact of WN and mNutrition on young children and their mothers, we will use a combination of ANCOVA, single-difference and double-difference techniques depending on the outcome and population of interest. ANCOVA models (McKenzie 2012) control for individuals' outcome variables at baseline and are more flexible than typical difference-in-difference models when autocorrelations are low because they allow us to estimate rather than impose the autocorrelation in each outcome. We will analyse our panel elements primarily with ANCOVA models, using difference-in-difference models and single-difference models as robustness checks. For the cross-section elements of our design, we will use single differences controlling for baseline household covariates. If randomisation is successful and our treatment and control groups are balanced across a broad range of characteristics, then single difference, double difference and ANCOVA should lead to similar results. Thus, using any parameterisation, our impact estimates should have very low bias.

Cross-section element estimation strategy

To estimate the impact of WN and the mNutrition add-on on outcomes such as exclusive breastfeeding and child dietary diversity, we will use a single-difference estimator, controlling for household characteristics at baseline and estimate ITT.²¹ The regression specification is then:

$$Y_{1ihc} = \beta_0 + \beta_1 Treatment_c + \delta_x X_{0ihc} + \varepsilon_{ihc}$$

where Y_{1ihc} is the outcome of interest for individual i in household h from community c at time 1, $Treatment_c$ is an indicator for whether or not community c was assigned to the treatment group, and X_{0ihc} is a vector of household characteristics at baseline. β_1 measures the differences in outcomes of the treatment versus comparison communities, and thus the impact of the WN and mNutrition services.

Panel element estimation strategy

To estimate the impact of WN and the mNutrition add-on on outcomes such as child height-for-age and women's dietary diversity, we will use an ANCOVA specification, controlling for the respondent's outcome at baseline. As in the cross-section design, we will estimate the ITT. The regression specification is then:

$$Y_{1ihc} = \beta_0 + \beta_1 Treatment_c + \delta_x X_{0ihc} + \delta_y Y_{0ihc} + \varepsilon_{ihc}$$

The regression specification is the same as in the cross-section, but because it includes Y_{0ihc} , the baseline level of the outcome for the person of interest, it is an ANCOVA specification. β_1 measures the differences in outcomes of the treatment versus comparison communities, and thus the impact of the WN and mNutrition products.

Heterogeneity

The absolute and relative impacts measured for mNutrition may depend on baseline characteristics of the study sample. In particular, the programme may have different effects

²¹ This research design estimates the ITT, rather than the average treatment effect (ATE), the local average treatment effect (LATE), or the average treatment effect on the treated (ATT), because here treatment is the offer to participate in WN and mNutrition rather than actual participation in the programme. By assuming that the offer of treatment itself has no impact on outcomes, except through its impact on the likelihood that a subject participated in the WN and mNutrition programme, we can also estimate the ATT for those individuals who were induced to participate in the programme by the offer of treatment.

for women who are pregnant at baseline and for women with low baseline levels of knowledge about IYCF practices and other nutrition-related behaviour. Consequently, we plan to measure heterogeneity of impact by whether the respondent was pregnant at baseline and whether the respondent had below the median level of IYCF knowledge among women in the sample at baseline.

Gender

Gender is incorporated across all stages of the research, from promoting gender-targeted interventions to conducting gender-related analysis on impact. For the evaluation, this entails ensuring that the survey instruments are designed and implemented in such a way that both males and females are included in quantitative data collection and that, as appropriate, data collected can be disaggregated by gender; analysis can be disaggregated by gender of the phone owner, and outcomes can be disaggregated by gender of household head.

Spousal Information

Households participating in the study will overwhelmingly consist of a husband and a wife, with a small fraction of polygamous households. Given the rapid spread of mobile technology in sub-Saharan Africa, and Tanzania in particular, we expect that in roughly 50 per cent of surveyed households, both husband and wife will have their own mobile phone. This creates an opportunity for us to learn about how information flows within households and how household members form beliefs, both of which are critical open questions for policy-makers hoping to generate behaviour change through the provision of information.

In households from treatment villages where both the husband and wife have their own mobile phones, we will randomly assign half the households to receive the WN and mNutrition messages on both mobile phones, while the other half will receive the content on just the wife's mobile. Because this 'spousal information' treatment is randomly assigned, we can identify the effect of sending duplicate messages to the mobile phone of the husband, as well as to the mobile phone of the wife, by simply comparing outcomes across these two sub-treatment arms. We will test for differences between the standard treatment arm and the spousal treatment arm for each of the primary and secondary outcomes. In addition, we will use information from the baseline and endline surveys to identify the mechanism through which the spousal treatment may be differentially affecting beliefs, behaviours and nutrition outcomes.

Limitations

The biggest limitation inherent in evaluating this type of programme is that an individual's 'treatment status' may change over the course of the study. A pregnant woman may initially sign up for WN and mNutrition and receive messages for some time, then lose or break her phone, run out of battery or credit, or choose to unenrol from the services. As a result, our estimates of the effect of WN and mNutrition may be a lower bound relative to what their impact would be with perfect compliance among individuals who take up the programme. However, it is implausible to expect perfect compliance from users, suggesting our research design will provide the more policy-relevant parameter.

Given that the randomisation will be conducted at community level, we expect minimal bias in our impact estimates due to spillovers or sharing of information across intervention groups. However, to some extent, we can test whether there are likely to be information spillovers in our sample. By taking advantage of the random location of treatment and control villages, we can explore whether there are differential changes in primary and secondary outcomes between control villages close to and far from treatment villages. Though this test may be somewhat underpowered, if there are information spillovers we would expect that control villages spatially closer to a treatment village will be more likely to be affected by them. However, given the rural and somewhat isolated nature of the villages

that will be included in our sample, we do not expect to find evidence of information spillovers.

3.2.4 Outcomes of interest

Primary outcomes

The primary outcomes of interest are those related to GSMA's stated goals:

1. Women's dietary diversity;
2. Infant and Young Child Feeding Practices (IYCF);
3. Anthropometrics of children (height-for-age z-scores, HAZ).

Women's dietary diversity: Information about the food consumption of women will be collected and used to construct the WDDS. The WDDS is composed of nine food groups and reflects the nutritional quality of the diet (Kennedy, Ballard *et al.* 2011).

IYCF: a set of optimal feeding practice indicators that relate to breastfeeding and complementary feeding. IYCF practices have been shown to be a major factor in child survival, growth and development (UNICEF 2011).

Anthropometrics of children: We will weigh and measure all children aged 0-11 months at baseline and all children aged 18-35 months at endline. All measures will be converted to z-scores using the WHO reference populations. HAZ is a measure of chronic malnutrition, with stunting ($HAZ < -2$) reflecting cumulative retarded growth. Weight-for-height (WHZ) is a measure of acute malnutrition, with wasting ($WHZ < -2$) reflecting a deficit in tissue and fat mass. Weight-for-age (WAZ) is a composite indicator of HAZ and WHZ, and thus captures both transitory and chronic aspects of malnutrition.

Findings from the quantitative outcome indicators will be combined with findings from the qualitative data to redefine, contextualise and add different dimensions (including participants' understanding and perspectives on dietary diversity, agricultural productivity and income, as well as changes in these). This triangulation of quantitative and qualitative indicators will help to increase the validity of the primary outcomes and add different dimensions.

Secondary outcomes

Secondary outcomes include a large set of variables to measure: (1) take-up and use of WN and mNutrition, and (2) improvements in nutrition and health knowledge and behaviour. Below are some examples of such variables.

1. **Take-up and use of WN and mNutrition product:** Number of women subscribing to the services, number of months actively using mNutrition, most used components, sharing of information within the household, health insurance;
2. **Improvement in nutrition and health knowledge and behaviour:** Knowledge of vitamin A- and iron-rich sources of food, responsive feeding techniques, malaria prevention, growth monitoring, improved water and sanitation practices;
3. **Spousal information and nutrition-related knowledge and behaviours:** How does sending the WN and mNutrition content to the mobile phone of the primary female and her husband differentially affect the other primary and secondary outcomes?

The findings from the two qualitative data collection rounds will be used to inform the development of the baseline and endline surveys and triangulate and contextualise the findings.

3.2.5 Sampling strategy and data collection

Sampling strategy

The sample will be composed of pregnant women and mothers with children under the age of 12 months. To be included in the sampling frame, these women must reside in a household where at least one household member owns a mobile phone and there must be at least one household member who is literate in Swahili. This latter requirement is to ensure that the household is able to access the WN and mNutrition information, which will be sent through SMS messages in Swahili. In order to know which households are composed of pregnant women and women with a child less than one year old with ownership of a mobile phone and at least one household member literate in Swahili, a community listing exercise (CLE) will be conducted. The CLE will also allow us to calculate take-up rates of target women in each village.

We propose randomly sampling households in the following manner. We will begin by removing urban villages within Iringa region. From the remaining rural villages, we will randomly select 180 villages to be included in the study. Because of the importance of the CLE for identifying eligible households and the time-sensitive nature of the treatment – pregnant women will only benefit from the messages targeted at pregnant women while they are pregnant – we will conduct the CLE and baseline survey simultaneously, with the CLE preceding the baseline survey by roughly two weeks. The field team responsible for conducting the CLE will send data to the research team in batches every two weeks. Using that data, and taking as given the treatment assignment from any previously received batches, we will randomly assign villages to the treatment group and control group in a manner as to maximise the relative design efficiency of the allocation with respect to balancing the number of pregnant women across the two treatment arms. At the end of the fieldwork, we will have allocated 90 villages to each treatment arm, and the two treatment arms will be well balanced with respect to the number of pregnant women. In each village we will attempt to randomly sample 17 households, 11 households with children 0-11 months old, and six households with pregnant women, for a total sample of 3,060 households (1,980 households with children 0-11 months + 1,080 households with pregnant women) at baseline and 2,907 households at endline (assuming a 5% attrition rate of the panel across rounds).

Because fertility rates are likely to vary greatly across villages within our study region, we expect that we will not be able to identify 11 households with a child aged 0-11 months and six households with a pregnant woman in every village. To account for this, in villages with below the required number of eligible households, we will sample all eligible households. In addition, we will over-sample from villages where more than the required number of eligible households are found, ensuring that, across all villages in the sample, the average number of households with children aged 0-11 months is 11, and the average number of households with a pregnant woman is six.

Sample size

We have designed the study to detect impacts on anthropometrics. We additionally show power calculations for dietary diversity scores for pre-schoolers 18-35 months of age (minimum dietary diversity – the percentage of children who consumed food from four or more food groups) and women 15-49 years old (number of food groups, 0-9). Both of these indicators have been validated by WHO as good predictors of diet quality and micronutrient

density for these two interventions' windows of opportunity (Ruel *et al.* 2013). We also show calculations for exclusive breastfeeding of children 0-5 months old.

We conduct the power calculations for ANCOVA and single-difference models, depending on our outcome variable. The assumptions about relevant statistics are shown in the power calculation tables below (Tables 3.3-3.6), while their data sources are listed in a subsequent table (Table 3.7). We use conventional levels and set the power at 80 per cent and the significance level at 0.05. We assume 5 per cent attrition from baseline to endline.

We show calculations for two different minimum detectable effects.²² For HAZ we show options for .25 and .3 SD increase of the mean value. For women's dietary diversity, we show options for a 25 per cent increase over the mean value, in line with Akrofi *et al.* (2010) and Thorne-Lyman *et al.* (2010). We also show options for a 15 per cent increase. For exclusive breastfeeding and child dietary diversity indicators, we show 15 and 25 percentage point increases.

Because take-up rates in treatment and control areas may vary depending on household characteristics, we show two different assumptions about take-up rates in comparison and treatment areas. In the first, we assume 5 per cent take-up of mNutrition in comparison areas and a 75 per cent take-up in treatment areas, leading to a 70 pp take-up gap. In the second scenario, we again assume a 5 per cent take-up in comparison areas and a 50 per cent take-up in treatment areas, resulting in a 45 pp take-up gap. We believe the 70 pp take-up gap is conservative given our research design, while the 45 pp take-up gap corresponds to the difference in take-up that might expected from a randomised encouragement design.

We put these designs and assumptions together using standard sample size calculations for cluster RCTs:²³

$$n_c = 2\sigma^2 \left[\frac{(z_{\alpha/2} + z_{\beta})^2 [1 + (m - 1)\rho]}{d^2} \right]$$

Where n_c =required sample size per arm, $z_{\alpha/2}$ =1.96, z_{β} =0.85, m =number of observations per cluster, ρ =intracluster correlation coefficient,²⁴ σ =population SD of outcome variable and d =minimum detectable effect. For binary outcomes we approximate the variance (σ^2) using the following formula of the proportions π_1 and π_2 :

$$\sigma^2 \approx 1/2[\pi_1(1 - \pi_1) + \pi_2(1 - \pi_2)]$$

Because we use a randomised controlled trial with potentially imperfect compliance, this formula needs to be amended from its use in RCTs in one crucial way: in a standard RCT with perfect compliance, we set d to the size of the effect we intend to detect between the 100 per cent treated group compared to the 0 per cent treated (control) group. In our design, some individuals assigned to the treatment group may not take up the treatment, while some individuals assigned to the comparison group may take up the treatment. This in essence dilutes the difference in effect sizes between groups, which will increase the required sample

²² The minimum detectable effect is the smallest true treatment effect that a research design is likely to identify as statistically significant.

²³ For example, Hemming *et al.* 2011.

²⁴ The intracluster correlation is the fraction of the total variance of an outcome that can be explained by the within-cluster variance.

size. d should therefore be set to (expected take-up rate in treatment group – expected take-up rate in comparison group) x (expected effect on treated individual compared to untreated individual) (Duflo *et al.* 2007: 3895-3962).

Tables 3.3-3.6 show the required number of households (baseline and endline) for different outcomes of interest, possible effect sizes, and possible gaps in take-up rates between treatment and comparison areas. We assume ten households in our panel of households with children 24-36 months old at endline (based conservatively on a baseline sample of 11 households with children in this age range and a 5% attrition rate), and six households with children 18-24 months old at endline. We estimate that one out of every ten households with children aged 0-11 months at baseline will also have a pregnant woman in the household, which will increase the number of pregnant women in the endline sample to 1,278.

Table 3.3 Required sample sizes for children’s height-for-age z-scores

Panel: 10 obs/cluster							
Anthropometry (HAZ): children 24-35 months							
(ANCOVA; mean=-1.47, SD=1.38, ICC=.09, autocorrelation=.7)							
		Effect size					
		.25 SD			.3 SD		
		Clusters	Baseline	Endline	Clusters	Baseline	Endline
Take-up gap	70 pp	181	1895	1804	126	1317	1254
	45 pp	437	4583	4364	303	3182	3030

Table 3.4 Required sample sizes for women’s dietary diversity

Panel: 16 obs/cluster							
Dietary diversity (number of food groups consumed): women 18-35 months post-partum							
(ANCOVA; mean=2.42, SD=1.35, ICC=.19, autocorrelation=.1)							
		Effect size					
		15%			25%		
		Clusters	Baseline	Endline	Clusters	Baseline	Endline
Take-up gap	70 pp	212	3552	3382	77	1279	1218
	45 pp	512	8587	8178	185	3098	2950

Table 3.5 Required sample sizes for exclusive breastfeeding

Cross-section: 6 obs/cluster							
Exclusive breastfeeding (indicator): women 0-5 months post-partum							
(OLS; mean=.498, ICC=.05)							
		Effect size					
		<i>15 pp</i>			<i>25 pp</i>		
		Clusters	Baseline	Endline	Clusters	Baseline	Endline
Take-up gap	<i>70 pp</i>	155	973	926	57	355	338
	<i>45 pp</i>	370	2329	2218	135	849	808

Table 3.6 Required sample sizes for children's dietary diversity

CROSS-SECTION: 16 obs/cluster							
Dietary Diversity (Minimum Dietary Diversity): Children 18-35 months old							
(OLS; mean=.109, ICC=.07)							
		Effect size					
		<i>15 pp</i>			<i>25 pp</i>		
		Clusters	Baseline	Endline	Clusters	Baseline	Endline
Take-up gap	<i>70 pp</i>	55	910	866	24	393	374
	<i>45 pp</i>	116	1939	1846	49	811	772

Table 3.7 Data sources for power calculation assumptions

Outcome	Reference population	Data source for mean	Data source for SD	Data source for ICC	Data source for auto-correlation
Height-for-age Z-score	Children 24-35 months	Team's calculations, DHS 2010 data	Team's calculations, DHS 2010 data	Team's calculations, DHS 2010 data	Team's best estimate with data from Guatemala and Bangladesh
Women dietary diversity (number of food groups)	Women 15-49	Team's calculations, DHS 2010 data	Team's calculations, DHS 2010 data	Team's calculations, DHS 2010 data	Team's best estimate with data from Uganda
Exclusive breastfeeding	Children 0-5 months	DHS 2010 report	N/A	Team's best estimate	N/A
Child dietary diversity (percentage reaching minimum dietary diversity)	Children 18-35 months	Team's calculations, DHS 2010 data	N/A	Team's calculations, DHS 2010 data	N/A

Although sampling more than 8,000 households is not possible given the budget, we strongly recommend sampling 11 households per cluster with children aged 0-11 months at

baseline (1,980 such households), which is enough to provide more than 1,881 children who will be 24-36 months at endline. This will ensure that we are able to detect an effect on children's HAZ if the effect size is .25 SD when the take-up gap is 70 pp. In addition to these households, we will sample six households with pregnant mothers per cluster, for a total of 17 households per cluster at endline. Assuming 5 per cent attrition, this leads to a sample of 3,060 at baseline. This sample will allow us to detect 25 per cent improvement in women's dietary diversity, 25 pp improvement in exclusive breastfeeding, and a 15 pp improvement in the percentage of children reaching minimum dietary diversity, all if the take-up rate gap is 70 pp.

Surveys

For the baseline and endline surveys, we plan to collect detailed information on: (1) outcomes on which we expect to see impacts; (2) intermediate factors that may explain pathways of impact, such as changes in behaviour, knowledge and practices; and (3) factors that may affect take-up rates. Quantitative data will be collected from intervention arms using multiple survey instruments at household, mother and child levels. The design and exact questions and modules used will be informed by the findings of the qualitative fieldwork.

A baseline survey will be conducted before the start of the intervention and an endline survey two years later. This schedule allows for a longer intervention period, which is necessary to observe impacts on HAZ.

The baseline survey will include modules on demographics and education for all household members, as well as details on household decision-making and female status, nutrition knowledge and practice, IYCF practices, women's dietary diversity, child's dietary diversity, health practices, women and child health, and use of and access to mobile phones.

The endline survey will take place approximately two years after the baseline survey, and will collect information on the same indicators included in the baseline survey, in addition to: detailed modules on cell phone usage; mNutrition usage, exposure and impressions; and usage of, exposure to and impressions of other mHealth programmes. We will also collect information on other topic areas that were identified as relevant for the intervention take-up and outcomes by the qualitative midline data collection. The quantitative endline survey will follow up on potential unanticipated and unintended (both negative and positive) outcomes that might be identified and captured in the qualitative midline data collection.

3.2.6 Partners

GSMA is the lead organising partner.

TFNC (Tanzania Food and Nutrition Centre) is the government oversight group.

GAIN is the content creator.

Wazazi Nipendi (WN) is the implementing partner.

OPM is the data collection firm.

3.2.7 Risks and their mitigation

The biggest risk to this evaluation design is that initial take-up in treatment areas is high, but use is low. Given that behaviour change does not occur quickly, prolonged messaging and information is needed. If a user initially signs up for the product but then does not use it, it is not likely that we will see an impact. Thus SMS messages and nudging for a prolonged period may be necessary to encourage continuous use.

Another risk is that our partner organisations (WN and the organisations that deliver the encouragement – either WN, the MNOs, health clinics or NGOs) will not implement the randomly assigned treatment properly, either by mistake or by choice. We will mitigate the possibility of mistakes by working closely with our partners and helping to supervise the process. It is more difficult to mitigate the risk that a partner may choose to invalidate the randomisation (e.g. by advertising heavily in comparison areas), because they find it more profitable to move into new areas than to wait for the completion of the study. We can mitigate this risk by working with our partners to understand and negotiate with their interests. We have also mitigated this risk by choosing a study region where our partner organisations have little existing presence.

It is also possible that our assumptions about baseline means, SDs and intracluster correlation coefficients of our outcomes of interest are incorrect. We have mitigated this risk by using several data sources to inform our estimates, and believe we have made conservative assumptions. It is also possible that we will not be able to recruit enough pregnant women per cluster. We will mitigate this risk by confirming our population estimates with local groups before beginning surveying and by over-sampling from villages with more eligible households than required to ensure, on average, that we achieve the target number of households with pregnant women and the target number of households with a child aged 0-11 months.

Additional risks are present in this impact evaluation that would exist regardless of what specific design is chosen, but are important to keep in mind. First, the mobile health landscape is swiftly changing in Tanzania, both with this specific programme and with complementary and competing organisations. It is likely that the WN mNutrition product will continue to develop over the year-long evaluation period, and possible that it changes so much that an evaluation of 'the intervention' is not meaningful. We will mitigate this risk by communicating with our partner organisations about the goals and necessary conditions of our evaluation. It is also possible that another organisation will introduce a competing product in our study areas that interferes with the measurement of impact of this product. We will mitigate this risk by remaining in contact with the government officials who approve such programmes, and at least try to remain aware of what other groups are doing. The qualitative data collection (both the initial scoping study and the midline) will also collect data on the institutional, sociocultural, environmental and economic contexts of the interventions, and explore how these factors may enhance and/or hinder take-up. During the baseline, the quantitative surveys will also collect data on competing products that were identified in the qualitative scoping study.

Lastly, it is possible that the MNOs may at some point remove their support for the WN and mNutrition programmes and the mNutrition product would no longer be cheaply available in Tanzania. We rely on our partners at WN and GSMA to negotiate with the MNOs for the continued success of the programme.

4. Qualitative impact evaluation design

4.1 Objectives

The qualitative component of the evaluation aims: (1) to provide an in-depth understanding of the different contexts within which the mNutrition interventions – i.e. mHealth (Tanzania) and mAgri (Ghana), referred to as mNutrition in this chapter – products are embedded, and which are likely to affect the take-up and outcomes of the products; and (2) to explore processes of change and their underlying mechanisms to explain how and why (or why not) mobile phone-based services lead to change of agricultural and/or nutritional behaviours within the different contexts. The qualitative component therefore allows for an exploration of links along the causal chain between activities, outputs, intermediate outcomes and impacts described in the theory of change for the mNutrition project. As a part of this, the qualitative investigation will also unearth and document as yet unarticulated programme assumptions that may have the potential to affect programme outcomes for each of the products and countries.

The qualitative component will address the following research objectives stated in the TOR and the specific objectives of GSMA:

- What factors make mobile phone-based services effective in promoting and achieving behaviour change (if observed) leading to improved nutrition and livelihood outcomes?
- Has the process of adapting globally agreed messages to local contexts led to content that is relevant to the needs of children, women and poor farmers in their specific contexts?
- What lessons can be learned about best practices in the design and implementation of mobile phone-based nutrition services to ensure (1) behaviour change and (2) continued private-sector engagement in different countries?

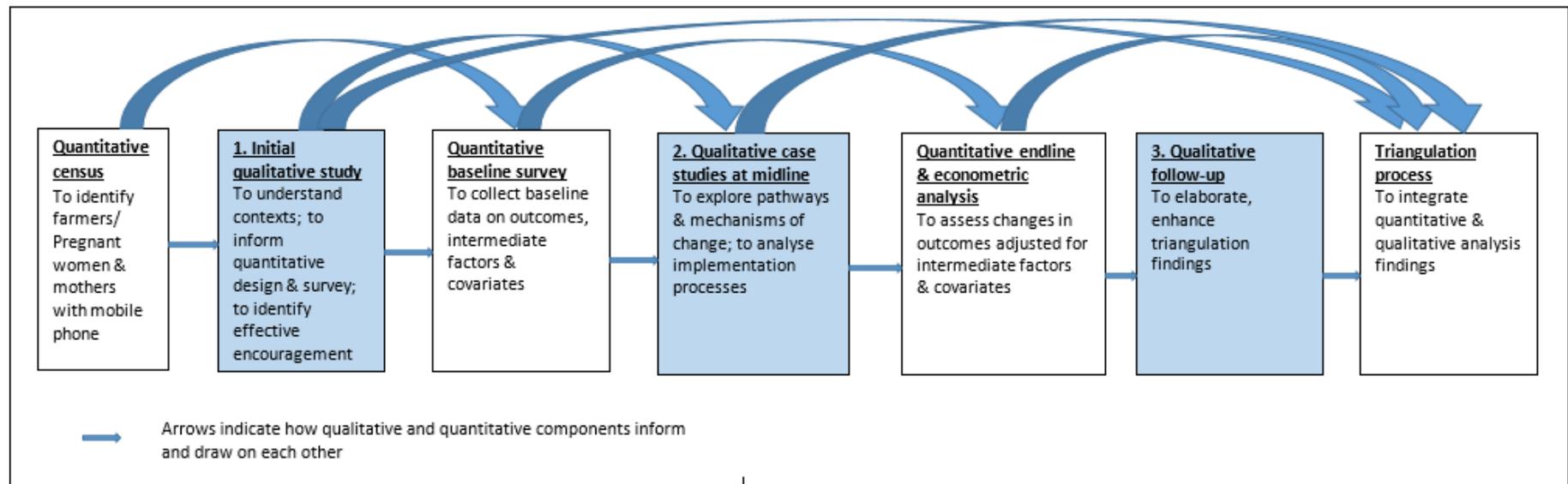
4.2 Approach

4.2.1 Sequential mixed-method design

The qualitative evaluation component will be closely linked and integrated with the quantitative encouragement design at all stages of the evaluation to inform, enhance and explain the design, data collection and analysis of the quantitative component. Qualitative work will be carried out in a sub-sample of the quantitative study communities and will focus (as the quantitative impact evaluation) on mHealth in Tanzania and mAgri in Ghana. Qualitative and quantitative approaches will be used in a sequential manner (see, for example, Creswell 2013). This will begin with a quantitative census to identify the target groups for the quantitative data collection and to guide the sample selection for the exploratory initial qualitative study. Through in-depth interviews, focus group discussions (FGDs) and literature reviews this study will inform and contextualise the quantitative baseline. The findings from the quantitative baseline surveys will be explored further using several community-level qualitative case studies at the midline of this impact evaluation. These case studies will also inform the quantitative endline. In the final step of the impact evaluation, qualitative and quantitative findings will be compared and integrated into a systematic triangulation process to capture how mobile phone based-services could improve child and maternal nutrition and agricultural practices in the respective country and programme contexts. A triangulation matrix will be employed to link qualitative and quantitative findings and better understand the different levels and complexities of the intervention impact (Mason 2006; Institute of Medicine and National Academies 2014). A

complementary approach will be used in order to integrate seemingly 'contradictory' findings from qualitative and quantitative data (Slonim-Nevo and Slonim 2009). This will help to further understand the realities of the intervention within the different settings (May 2010). As part of the triangulation, qualitative researchers will return to the field to conduct mini-case studies to follow up and further elaborate the findings that will emerge from the combination of the qualitative and quantitative impact evaluation streams. In line with the quantitative study, the qualitative study will also retain a special focus to study the effects of mNutrition interventions on women and analyse factors such as ownership of productive assets and decision-making, which influence the outcomes and behaviour related to nutrition. Figure 4.1 demonstrates how qualitative and quantitative approaches will be used sequentially in this evaluation.

Figure 4.1 Sequential mixed-method design to study the impact of the mNutrition products



4.2.2 Qualitative evaluation studies

The qualitative component will consist of three qualitative data collection events: an initial exploratory qualitative study, in-depth case studies at midline, and qualitative follow-up mini-case studies following the quantitative endline. We assume our budget for the qualitative evaluation is the same as outlined in the original budget. The available budget will be split between the two interventions to be evaluated.

Initial exploratory qualitative study

Contextual analysis

A comprehensive analysis of social, institutional, political and environmental factors that may influence the implementation, take-up, use of the mobile phone product and outcomes of the different mNutrition products will be conducted. Within this wider objective, the contextual analysis will specifically explore:

1. The acceptability of, familiarity with and use of mobile phone technology by the target groups – farmers (male and female), mothers and pregnant women – of the mNutrition products;
2. Factors that may affect the operation of and/or access to a mobile phone and behaviour change SMS by the target group (e.g. electricity and network coverage, intra-household dynamics with specific focus on gender relations that may determine access to and use of mobile phone);
3. Current information-seeking behaviours related to nutrition and agriculture as part of the existing information and knowledge economy (e.g. where do people get nutrition and agricultural information from? Whom do they trust? Are there any services and products, including other mobile phone-based services, that compete with the mNutrition products?);
4. Social, economic and environmental factors that may influence whether (or not) the behaviour change messages distributed as part of the mNutrition products are taken up and put into practice.

A better understanding of the ‘real world’ context within which different mNutrition products are embedded is important because context has been shown to be intrinsically involved in the causal processes that may bring about (or not) the desired impact of an intervention (Maxwell 2004). Contextual analysis is also important to understand differences in the impact of the products across the different settings and can thus help to enhance external validity of the evaluation findings (Mohr 1999). The findings of the contextual analysis will inform the development of the quantitative baseline survey and the definition of indicators to measure outcomes of the mNutrition products.

Inform the choice of context-specific encouragement strategy

To encourage high levels of take-up of the mNutrition intervention by the treatment groups of the quantitative study sample (see Sections 3.1.3 and 3.2.3) an effective encouragement strategy is important. FGDs and interviews with key informants (e.g. project implementation staff) and potential users of the mNutrition product will help to explore different incentives strategies (e.g. financial incentives, different additional marketing strategies) and assess their perceived effectiveness in promoting take-up of the mNutrition products. Which strategies are perceived as most effective may be country specific and depend on contextual factors such as the existing nutrition/agriculture information and knowledge economy (e.g. are there other trusted sources of information on nutrition or is there a lack of information?) or whether or not there are other mobile phone-based services (e.g. people may already

receive regular SMS on other topics, and if such uses affect the delivery and absorption of these messages). For example, in Bangladesh the government uses mobile phone messaging to communicate with citizens. As a consequence, many people are tired of 'constant' SMS and often perceive the messages as spam. In order to resolve such tensions (if any), a good encouragement strategy must be able to identify and deliver what people want in order to lead healthier lives with better nutritional value.

Qualitative case studies at midline

Case studies to explore causal pathway and mechanism

Using qualitative case studies (Woolcock 2013) we will set out to explore 'how' and 'why' the mNutrition products may (or may not) lead to the desired changes in nutrition and/or agriculture behaviours and practices in the specific country contexts. The qualitative studies thus attempt to 'open up the black-box' (see, for example, White 2009) and shed light on the mechanisms and processes through which the products may/may not trigger behaviour change. Drawing on (and if necessary refining and extending) the existing theory of change and the contextual analysis, pathways through which changes in the targeted behaviours may occur will be explored and context-specific barriers and facilitators will be identified. Participatory qualitative approaches – such as matrix ranking (Mukherjee 1999: 13-15) in which subscribers will score, rank and then discuss the different elements of the mNutrition intervention – will provide multiple interpretative perspectives on participants' perceptions, potential frustrations and experiences with the products. We will also draw on theoretical behaviour change models such as the theory of reasoned action/planned behaviour (Ajzen and Fishbein 1980) and the social cognitive theory (Bandura 1989) to identify additional potential avenues and factors that explain change/absence of change in behaviour and practices (Bandura 1989).²⁵ To examine the underlying mechanisms that influence whether or not behaviour change messages are translated into actual behaviour change, and for whom and under what circumstances this may happen, a realist evaluation approach (Pawson and Tilly 1997) will be used. The realist approach will allow us to systematically explore and test why/why not the messages are translated into practice. The assumption of realist evaluations is that underlying mechanisms (M) triggered by the programme interact with the social, political, personal context (C) to produce varying outcomes (O): 'mechanisms + context = outcomes' or CMO (context-mechanisms-outcome). Consequently, the regular reception of behaviour change messages via mobile phone (as well as other channels that different mobile phone products might use) may result in different reactions and behaviours by users (outcomes) (Pawson 2013).

The qualitative case studies also aim to reveal potential unanticipated (positive and negative) outcomes and consequences of the mNutrition products at individual, household and community levels. The capture of unintended consequences is important, as these may limit the ability to detect and explain the outcomes of the products, as well as help to explain the lack of behavioural change (Morell 2005).

The box below shows an indicative list of questions we will address in the qualitative case studies. The final list will be developed together with the qualitative teams in the different countries and will also be informed by findings from the initial contextual analysis and the quantitative baseline survey.

²⁵ See the landscaping review conducted for this impact evaluation for a critical review of behaviour change theories (Barnett *et al.* 2016).

Questions to be addressed in the qualitative case studies

- What value do participants place on the behaviour change messages? How relevant, useful and novel is the content of the messages?
- Which elements of the mNutrition intervention are perceived as most effective in promoting behaviour and practice change? Which elements are least effective? Why?
- When, how often and why do participants use the mNutrition services (e.g. read the SMS)? Do they plan to continue using the service? Why? Why not?
- Who accesses the mNutrition services and, in particular, do women and very poor households access the service? How do they access the service (e.g. if they don't have access to a mobile phone)?
- What are their attitudes towards and opinions of the use of mobile phone technology to deliver nutrition/agricultural behaviour change messages (e.g. do they trust/value information received by mobile phones)?
- What are the underlying individual-, household- and community-level barriers and facilitators that influence take-up of the messages?
- Is the knowledge from the messages translated into practice? Why? Why not? What are the facilitators, enablers, barriers? What role does context play?
- Are the behaviour change messages shared with other households and/or community members? Why? How? How often? How do they use the messages?
- How could the mNutrition service be improved further?

Process analysis

The qualitative midline will also include an in-depth analysis of the intervention implementation process. The aim of this process analysis is to study how well the mNutrition products are actually implemented and whether there are any shortcomings in the design and/or implementation that may affect the outcomes (e.g. technical difficulties, problems with the targeting of messages with relevant content). The process analysis will also be linked with the ongoing monitoring and evaluation (M&E) activities of the mNutrition products²⁶ to avoid duplications and gain deeper insights into the programme operations, implementation and service delivery. A thorough analysis of the implementation processes is essential to be able to determine whether a potential absence of impact could be because mobile phone-based services are not effective in promoting behaviour change (design failure) or due to problems with the implementation (implementation failure).

The process analysis will also be able to detect potential changes, developments and local adaptations of the products and/or the implementation of the products. Capturing potential changes is important as it may influence the outcomes and impact and need to be considered in the final analysis stage of this evaluation.

Informing the analysis of the quantitative baseline and development of the quantitative endline

The findings from the qualitative midline will enrich the analysis of the quantitative baseline surveys and also inform development of the quantitative endline (e.g. by suggesting additional indicators and measures of primary outcomes, such as additional and country-

²⁶ Aline conducts the M&E for the mAgri interventions and Johns Hopkins University for the mHealth interventions.

specific dimensions of dietary diversity based on users' perceptions of a diverse diet, and unintended consequences). Qualitative data will also deepen the understanding, validity and reliability of the quantitative outcome indicators (i.e. dietary diversity, agricultural practices, disaggregated nutritional outcomes for women) by providing different dimensions (including participants' perspectives and understanding) and contextualisation for these indicators.

Informing and enhancing the sustainability of the mNutrition model

To gain a better understanding of the potential effect of the encouragement strategy used in the quantitative design on outcomes and impacts, the qualitative midline will explore: (1) why participants signed up for the mobile phone-based service; and (2) what role the incentive played in their decision to sign up and continue or discontinue using the service.

Qualitative follow-up study

Mini-case studies

To further elaborate, validate and follow up findings that emerge from the triangulation of the quantitative and qualitative impact evaluation data following the quantitative endline, mini-case studies will be employed. These studies will draw upon the rapport established with the communities during the exploratory and qualitative midline study. These rapid explanatory studies will be conducted in the same communities and with a sub-sample of the mNutrition users interviewed in the previous phases. mNutrition users who still actively use the mNutrition product, as well as users who discontinued use, will be interviewed to explore barriers and motivators of use (in particular, long-term use). During the follow-up studies some of the initial findings from the triangulation will also be presented to the mNutrition users to check and validate the accuracy of the evaluation teams' interpretations of the data (interpretive validity) (Maxell 1992).

4.3 Methodology

Data collection methods

To gain multiple perspectives into the contexts, implementation processes, perceptions and experiences with the mNutrition products, qualitative data will collect information from different sources (e.g. pregnant women and mothers who signed up for mNutrition and those who did not, farmers who signed up and those who did not, health professionals, agriculture extension workers, implementation staff of the mNutrition products) using multiple data collection methods (e.g. in-depth interviews, focus groups, participatory ranking matrix, programme documents²⁷ and primary and secondary literature reviews).

In-depth interviews (IDIs) will be used for the contextual analysis and to gain deeper insights into mechanisms and processes through which the products may/may not trigger behaviour change. The interviews will provide detailed information about m-service users' thoughts, views, perceptions and experiences with the intervention and its impact. They will also allow for deep insights into factors at individual and household levels, including gender roles and dynamics and intra-household decision-making, and uncover how these factors might interact with and shape the intervention outcomes. Semi-structured interviews with key informants²⁸ (e.g. health professionals, agriculture extension workers, and national and local implementation staff) will provide a detailed understanding of contextual issues, implementation processes and challenges.

²⁷ Documents will include mNutrition product descriptions by GSMA/local MNOs, implementation guidelines and marketing material.

²⁸ We categorise key informant interviews as a sub-category of in-depth interviews.

Focus group discussions (FGDs) will be used for the context analysis to identify effective encouragement strategies, to explore enablers of and barriers to uptake of the mNutrition message and change of behaviours, and for the process analysis. The focus groups offer dynamic group settings where people encourage and stimulate each other to consider different views, and power relations come to the fore in group settings. In order to ensure that different views and perceptions are captured and represented, the focus groups will comprise different categories of users including men/women, older/younger and poorer/richer users, and care will be taken to mix them appropriately for a representative sample, bearing in mind the number of m-service users (long term and short term) in a given community, and power relations within the community. These focus groups will be held in a sequential manner, and at different times of the evaluation study, to analyse the different factors that shape nutritional outcomes and behaviour change in the intervention areas.

Participatory matrix ranking will be used during the FGDs to visually score and discuss m-service users' appreciation of different aspects of the mNutrition intervention and to explore their perceived contribution to change.

Flexible semi-structured topic guides will be developed to guide the IDIs (with key informants and users) and FGDs. Pilot-testing of the guides will take place and the guides will be modified as appropriate. Development of the topic guides will be informed by literature reviews on determinants of nutritional and agricultural behaviours and practices in Ghana and Tanzania.

As part of the initial context analysis, we will conduct a review of the literature on nutrition- and agriculture-related behaviours in both Ghana and Tanzania, with the aim of identifying factors that influence, hinder or motivate beneficial behaviours. The literature analysis will also aid in explaining the institutional and political context for the programme interventions and the emerging causal links that will be tested in the exploratory, midline and endline qualitative study.

As part of the process analysis, we will review programme documents from the mNutrition interventions in order to assess how far actual implementation corresponds with the planned implementation process.

Table 4.1 presents an outline of the different qualitative methods that will be used.

Table 4.1 Overview of qualitative methods, sample and purpose

Method		1. Initial qualitative study	2. Qualitative midline	3. Qualitative follow-up
IDI	Purpose	<p>To understand the social, economic, political context</p> <p>To explore acceptability, attitude, beliefs and actual use of mobile phone technology</p> <p>To identify an effective country-specific encouragement strategy</p>	<p>To explore views, perceptions and judgements of the mNutrition products and how and why they may/may not change behaviours</p> <p>To explore actual use and barriers to/enablers of use of the mNutrition product</p> <p>To explore challenges and realities of the implementation process</p>	<p>To explore why/why not the mNutrition product leads to behaviour change (e.g. identify barriers including barriers to long-term use, enablers of use)</p> <p>To identify reasons for discontinued use</p> <p>To follow up on any 'contradictory findings' that may emerge from the quantitative endline study</p>
	Sample	Pregnant women/mothers/farmers with mobile phone access (potential users), local key informants	Active mNutrition users, implementation staff, key informants	Active mNutrition users, mNutrition users who discontinued use
FGD	Purpose	<p>To explore contextual factors within the communities that might affect take-up and behaviour change</p> <p>To explore current behaviours related to nutrition/agriculture and underlying reasons</p> <p>To explore acceptability, attitude, beliefs and actual use of mobile phone technology</p>	<p>To explore the effect of encouragement strategy on take-up and behaviour change</p> <p>To understand barriers, motivators, facilitators of take-up of the behaviour change messages and actual</p>	<p>To explore why/why not the mNutrition product leads to behaviour change (e.g. identify barriers including barriers to long-term use, enablers of use)</p> <p>To identify reasons for discontinued use</p> <p>To follow up on any 'contradictory findings' that may emerge from the</p>

		<p>To explore current information-seeking behaviours related to nutrition/agriculture</p> <p>To identify an effective country-specific encouragement strategy</p>	<p>behaviour change</p> <p>To assess potential spillover of the mNutrition behavioural change messages to other community members</p> <p>To explore unintended consequences (positive and negative)</p>	quantitative endline study
	Sample	Pregnant women/mothers/farmers with mobile phone access (potential users), community members	Active mNutrition users, community members	Active mNutrition users/discontinued users, non-users
Participatory ranking matrix	Purpose	Not used	To score and discuss mNutrition users' appreciation of different aspects of the mNutrition products and their contribution to behaviour change	To score and discuss mNutrition users' appreciation of different aspects of the mNutrition products and their contribution to behaviour change
	Sample		Active mNutrition users	Active mNutrition users

All qualitative research methods will adhere to strict ethical standards (as described in Chapter 7 of the inception report, and including OECD DAC²⁹ quality standards and IFPRI and IDS ethical guidelines). Qualitative data collection will only take place once full and informed consent is obtained from respondents. Confidentiality and the anonymity of participants will be protected at all times and the real names of people and locations will be replaced with pseudonyms.

²⁹ See OECD (2010)

Sample

1. Initial (exploratory) qualitative study

Site selection

The sample selection for the initial qualitative study will be informed by the quantitative census. Six villages/communities/health facility catchment areas per mNutrition intervention per country will be purposefully selected to reflect different geographical areas: urban and rural settings for mHealth and agricultural areas (e.g. cocoa, corn, rice) for mAgri.

Participant selection

At community level, participants will be purposefully sampled (drawing on the quantitative census as a starting point) to illustrate characteristics of different relevant sub-groups and to allow comprehensive understanding of the contextual issues that may affect the products in different settings. Table 4.2 presents the number of qualitative data collection events planned at each site for the initial qualitative study.

Table 4.2 Data collection events for the initial qualitative baseline, per mNutrition intervention and country

		Per country and per mNutrition intervention							
		Site 1	Site 2	Site 3	Site 4	Site 5	Site 6	National	Total
IDIs									
	Local key informants	1-2	1-2	1-2	1-2	1-2	1-2	-	6-12 (plus 3-6 IDIs with national key informants) = 9-18
	Pregnant women/mothers with mobile phone	2	2	2	2	2	2	-	12
	Farmers with mobile phone (mAgri only)	2	2	2	2	2	2	-	12
Total									33-42
FGDs									
	Pregnant women/mothers with mobile phone	1	1	1	1	1	1		6
	Farmers with mobile phone (mAgri only)	1	1	1	1	1	1		6
	Other community members (e.g. men, elderly)	1	1	1	1	1	1		6
Total									18
Total data collection events									51-60

2. In-depth qualitative study at midline

Site selection

For the qualitative midline studies three communities per mNutrition intervention will be purposefully selected from the quantitative treatment sites (Group 2) and to reflect the different encouragement arms (2a and 2b with gender encouragement). Two communities will be selected to represent 'typical communities' (Bamberger *et al.* 2012a) (one typical community will be located in a 2a treatment site and one in a 2b treatment site – see quantitative design) based on data collected in the quantitative baseline survey. Criteria to identify a typical community will include, for example, distance to the next town and access to public services. Conducting in-depth qualitative research in typical communities will provide deeper insights into the use and impact pathways of the mNutrition product on average communities. The third community (located in a 2a or 2b treatment site) will be selected using extreme case sampling (Yin 2013) and will reflect an atypical community (e.g. a community that is geographically difficult to reach and thus excluded from many services and NGO activities). Qualitative research in the extreme case community will help to understand the use and effectiveness of the mNutrition product among excluded populations.

Participant selection

mNutrition users within the selected communities will be identified from the quantitative baseline survey. Using a snowball sampling approach (Bamberger *et al.* 2012a), identified users will be asked to identify other active users in the community.³⁰ We assume that active mNutrition users will be aware of other active users living in the same community, for example, because subscriptions to the product allow users to call each other free of charge. All active users will be invited to participate in an in-depth interview about their experiences with and perceptions of the product, take-up of the behaviour change messages, and barriers to and facilitators of take-up and behaviour change. Active users will also be invited to attend FGDs that will use participatory matrix ranking to explore different components of the mNutrition product and their perceived effectiveness in more detail. To study how and why the mNutrition product may bring about change among some groups of active users but not others, we will conduct FGDs with men/women, older/younger and poorer/richer users separately. The concept of saturation (Morse 1995) will be employed as a guiding principle for the sample selection whereby additional mNutrition users will be identified and interviewed until new data no longer provide any new insights. To explore community-wide spillover of the mNutrition messages we will also conduct FGDs with purposefully selected community members who do not use the mNutrition product (male/female, old/young, poor/rich). The sample for these FGDs will be purposefully selected to ensure diversity and representation in the groups. This will allow us to explore different potential pathways for spillover within the community. To investigate implementation processes, in-depth interviews with project implementation staff and other key informants in the community will be conducted. Table 4.3 summarises the number of qualitative data collection events planned at each site during the qualitative midline study.

³⁰ In case the number of active subscribers that can be identified per case community is fewer than eight users, we will aim to identify additional active users in neighbouring communities that are also treatment sites in the quantitative impact evaluation.

Table 4.3 Data collection events for the qualitative midline, per mNutrition intervention and country

		Per country and per mNutrition intervention			
		Case 1	Case 2	Case 3	Total
IDIs					
Active mNutrition users		10~20 ^a	10~20 ^a	10~20 ^a	30-60
Project implementation staff (e.g. ambassadors)		2-3	2-3	2-3	6-9 (plus interview with 1-2 national implementation staff) = 7-10
Key informants (e.g. midwives, agriculture extension workers)		2	2	2	6
FGDs with participatory ranking matrix					
Active mNutrition users		3~4	3~4	3~4	9-12
FGDs					
Community members		2	2	2	6
Total data collection events					58-94

^a The aim is to interview as many active users of the mNutrition product as can be identified and until data saturation is reached.

3. Qualitative follow-up

Site selection

The qualitative mini-case studies will be conducted in the three case study locations selected for the midline.

Participant selection

mNutrition users within the selected communities will be identified from the quantitative endline survey. The aim is to select users who are still actively using the mNutrition product at the time of the quantitative endline, as well as users who have discontinued use since the midline. Table 4.4 presents the number of qualitative data collection events planned at each site for qualitative mini-case studies during the follow-up.

Table 4.4 Data collection events for the qualitative follow-up, per mNutrition intervention and country

		Per country and per mNutrition intervention			
		Case 1	Case 2	Case 3	Total
IDIs					
Active mNutrition users		3-6	3-6	3-6	9-18
mNutrition users who discontinued use		2-3	2-3	2-3	6-9
FGDs eventual with participatory ranking matrix					
Active mNutrition users		3~4	3~4	3~4	9-12
Non-users of mNutrition		1-2	1-2	1-2	3-6
Total data collection events					27-45

Data analysis

The qualitative data will be analysed using a directed content analysis approach focused on the main qualitative evaluation questions (Patton 2002). Data analysis will start with open coding of several interviews and the development of an initial coding scheme that will guide the coding of the remaining data. The coding scheme will be adjusted and modified as necessary during this process. To increase the rigour of the data analysis, analysis will be carried out independently by different qualitative researchers and results will be combined. Qualitative data analysis software (Nvivo) will be used to manage and aggregate coded data.

For the realist evaluation component, initial theoretical models ('middle-range-theories') of how the mobile phone-based messages may lead to behaviour change will be developed in the first year of the evaluation based on behaviour change theory, the existing theory of change, review of the literature, and knowledge and experiences of the evaluation team. These preliminary theories will be further developed based on the qualitative (and quantitative) data from the different data collection rounds. Eventually, a set of CMO configurations will be constructed. In the final stage of the analysis, the CMOs that provide the most robust explanations for why/why not behaviour change occurred in the different contexts will be determined.³¹

Different strategies will be employed to enhance the trustworthiness and rigour of the qualitative data and will be employed throughout the qualitative evaluation (Mays and Pope 1996; Bamberger *et al.* 2012b). For example, the adoption of well-recognised qualitative tools and analysis approaches by a team of experienced qualitative researchers with high familiarity with the contextual environments will enhance the credibility of the qualitative data. Frequent peer-debriefing sessions between the qualitative researchers during data collection and qualitative data analysis will facilitate reflexivity and ensure that researchers' biases could be removed to the extent possible to aid data interpretation. Triangulation via the use of different qualitative methods will help to reduce potential systematic bias. This will include the use of different qualitative data sources, different qualitative informant groups, and

³¹ Further details on the methodology that will be used can be found in Marchal *et al.* (2010).

qualitative study sites. In-depth documentation of the qualitative research methods will further increase the integrity and trustworthiness of the qualitative data.

Qualitative and quantitative findings will be integrated using systematic triangulation to complement, explain and enrich interpretations of the findings. During the triangulation we will also check for differences in findings and interpretations from the different methods. These differences may also contain useful insights into different perspectives, different values, tensions within the programmes and so on, and may be a fruitful source of insights.

5. Business model evaluation design

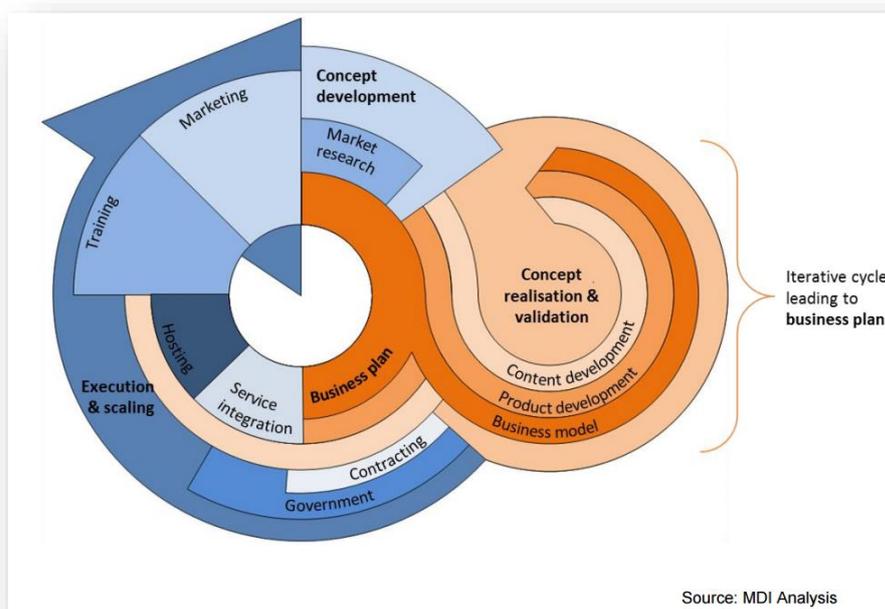
This chapter responds to the following two questions in the project ToR:

- **How commercially viable are the different business models being employed at country level?**
- What lessons can be learned about best practices in the design and implementation of mobile phone-based nutrition services to ensure a) behaviour change and b) **continued private-sector engagement in different countries?**

5.1 Introduction

As discussed in the landscape document,³² there are a considerable number of permutations and variables for the associated business models. Simple direct income versus expenditure rarely applies to mobile telephony. The modern landscape includes a number of different stakeholders, all of whom may undertake revenue sharing, and the issue of brand management and indirect revenue streams plays an important and often overriding role. The complexity of developing a new value-added service is suitably illustrated by the MDI Analysis diagram reproduced by GSMA (2013) in Figure 5.1 below.

Figure 5.1 Illustrating the iterative nature of business model development for value-added services



Source: © GSMA Intelligence (2013)

In addition to this complex landscape, the GSMA mHealth stream is ambitious in scope, particularly in its engagement with stakeholders. In its core aspiration it is attempting to consolidate existing and emerging mHealth propositions into a single gateway, aggregating

³² See landscaping analysis conducted as part of the inception phase of the external impact evaluation (Barnett *et al.* 2016).

the offerings so that the consumer experiences a single entry point, with the MNOs contributing zero-rated traffic for basic health services. This is ambitious and for that reason it requires a much more involved process of negotiation and is challenging to implement.

Of the four impact surveys considered during the inception phase, the arrangements for the mHealth offering(s) in Ghana are the closest to the GSMA aspiration. GSMA and its collaborators are currently discussing the details, and during field visits there were ongoing negotiations between the convenors and the network of stakeholders. During the visit there was a tight iterative loop as details were proposed, pushed back against and revised. Negotiations of the details are ongoing, so we can identify elements of the business model that we will need data on, but cannot yet state exactly the indicators and data required.

That being said, our choice of Tanzania mHealth provides a simpler delivery model (and less ambitious). WN has not yet negotiated the zero-rating with the MNO, therefore even here the exact business model remains in development. WN has funded its work by combinations of donor and corporate social responsibility (CSR) funding, and in discussions some key stakeholders believe this is a sustainable and viable way forward for the medium term. This model would not fall under the title of 'commercial sustainability' model (where revenue equals – and exceeds – expense), but is rather a 'sponsorship' model. There was an aspiration by WN to move towards a consumer payment model. We also note that a new managing company is being established for WN, which may influence the future development of business models.

The two mAgri offerings of the inception phase were more defined, due in part to the grant-making process employed. For Ghana mAgri, Vodafone Farmers Club has given verbal confirmation of 2 cedis per month to the consumer and articulated a clear roll-out strategy – i.e. a clear direct revenue model. However, for Bangladesh mAgri the product pricing is still under discussion, and the expenditure on marketing remains fluid. Therefore, again, we can identify the elements or domains we need to gather data on, but cannot yet solidify exact data specifications.

The TOR for our study also specify that the impact of the services and products be compared with more conventional or traditional communication channels such as radio and extension agents. For health, the GSMA has commissioned the Futures Group to undertake an exercise in calculating the cost benefit of the systems in terms of impact. This group is using the LiST model to calculate benefits in terms of reduction in lives lost. It has only been recently commissioned and we have not yet been able to determine the exact data it will be using to feed into the model. However, we are liaising with the group to ensure we support its processes and do not duplicate them. Nevertheless, the intention of this work package commissioned by GSMA is to strengthen the case to various governments that mHealth has a role to play in the landscape of the health system, and to this end the work is likely to fulfil its purpose.

5.2 Research scope

While the quantitative study proposes studying two services, Tanzania mHealth and Ghana mAgri, we believe we can offer insights into the business model for the health proposition in Ghana within the budget. This is not ideal, as the modelling would not be linked to an evidence-based impact study. However, if GSMA agreed, given that Ghana mHealth is an innovative model of commercial sustainability and the team would be making regular visits to Ghana for the mAgri side, it would be possible within budget to interview key stakeholders, get cost and consumer data from GSMA, and seek to answer the highlighted research questions above for Ghana mHealth.

In the following outline of proposed research, we also retain a view on Bangladesh mAgri, knowing that while we have proposed Ghana mAgri, the evaluations team’s discussion with DFID³³ indicated that Bangladesh was a country of interest and that transfer of models from Africa to South Asia was of particular interest. We are aware that DFID may yet choose to redirect us to a different pairing, and/or may choose to ask for inclusion of Bangladesh mAgri in the business models.

5.3 Research variables

The quantitative study will provide impact data for the overall service(s) and propositions for two of the studies. The replicability of this impact will likely depend on various business factors, each of which we will attempt to document. Of the key variables, we identify the following as important, and note that there is sufficient variation across the studies to suggest relevant comparative insight.

5.4 Stakeholders

Table 5.1 documents the key stakeholders and their role in each proposition.

Table 5.1 Key stakeholder groupings for the four propositions under study during the inception phase

	Ghana mHealth	Tanzania mHealth	Ghana mAgri	Bangladesh mAgri
Government endorsement	Needed to endorse service	Endorsed by existing arrangement	Covered by existing regulations	Covered by existing regulations?
MNOs	(Ideally) across all MNOs (to be determined)	Across all MNOs (to be confirmed)	Single MNO	Single MNO
Aggregator	Key to proposition	Single service	Single service	Single service
Service providers	Varied and diverse	Single service	Single service	Single service
(Potential) professional users	Anyone in health service	Anyone in health service	Anyone in farmers clubs	Anyone in farmer extension
(Potential) consumers	Pregnant women and caregivers	Pregnant women and caregivers	Anyone	Anyone

Looking more into Ghana mHealth stakeholders

As we have already noted, the Ghana mHealth aspiration is the bringing together of many stakeholders, aggregating their services into a single consumer offering. Therefore, even within the single service of ‘Ghana mHealth’, there are a range of stakeholders and products. Table 5.2 shows a breakdown of mHealth Ghana stakeholders and service types.

³³ 4 June 2015, DFID Offices Whitehall.

Table 5.2 mHealth Ghana stakeholders and types of service

Provider	Service	Type			Funding
		Information	Transaction	Market	
MC.com	Information (free or paid)	✓			Commercial
Grameen*	Information plus support	✓	✓		Non-profit
Votomobile	Information	✓			Non-profit
Audrey Pack*	Packs given to women	✓			Non-profit
MicroEnsure	Collect insurance premiums (broker)		✓	✓	Non-profit
CERsGIS*		✓			Academic?
Zero Mothers Die (UNICEF)					Non-profit
VAS2Nets*	Technical support?				Commercial
Vantage Medical		✓			Commercial

Note: * Providers yet to commit to the partnership at the time of the country visit.

It was made clear that each service provider will operate their own business model. Further discussions will be needed with each partner to gather detail on what those models are.

5.5 Revenues and brand

As the landscape document states, many value-added services are not necessarily driven by direct revenues. In 'Fighting Smart' (Amdocs 2012),³⁴ a survey of MNO leaders in Asia noted that direct revenue was not a key reason cited as to why they focus on value-added services (see Figure 5.2).

³⁴ <http://www.amdocs.com/products/digital-services/documents/amdocs-fighting-smart-white-paper.pdf>

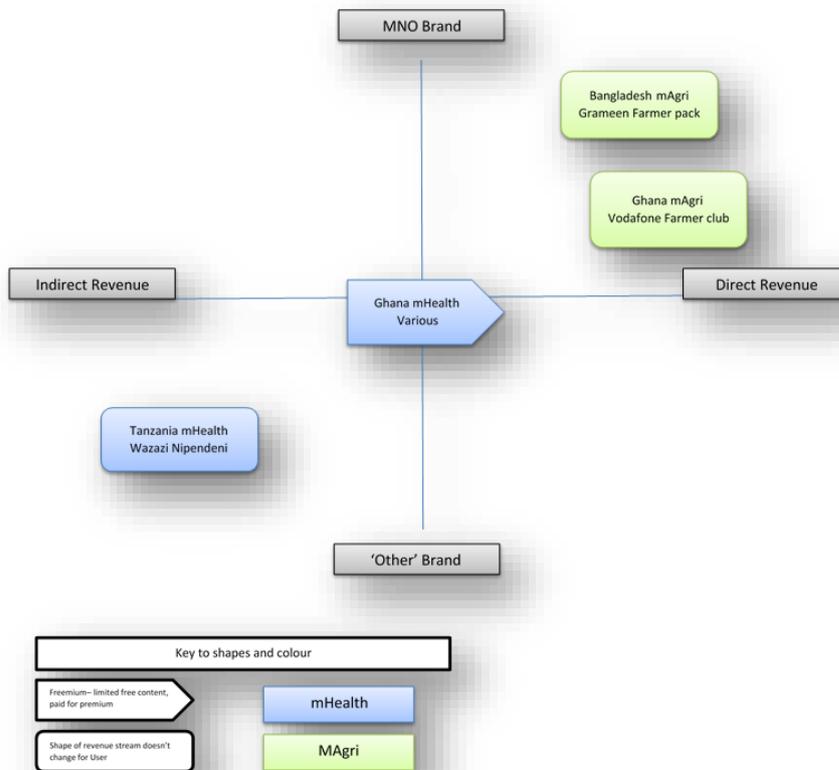
Figure 5.2 Response to question on business drivers for value-added services



Source: authors' own, adapted from Amdocs (2012)

In the various propositions, there is a variety of types of revenue and branding. The four studies will give insight into the relative roles of revenue and brand support.

Figure 5.3 'Revenue' vs 'brand' – What's in it for the MNO?



Source: authors' own

Figure 5.3 illustrates that Ghana mHealth might be the most fluid, depending on which stakeholders sign up to the proposition and how. The following table is based on information gleaned from the visit. However, at the time of the visit, the structure of the offering was still under discussion, so Table 5.3 is given for illustrative purposes only.

Table 5.3 Distribution of revenue in Ghana mHealth?

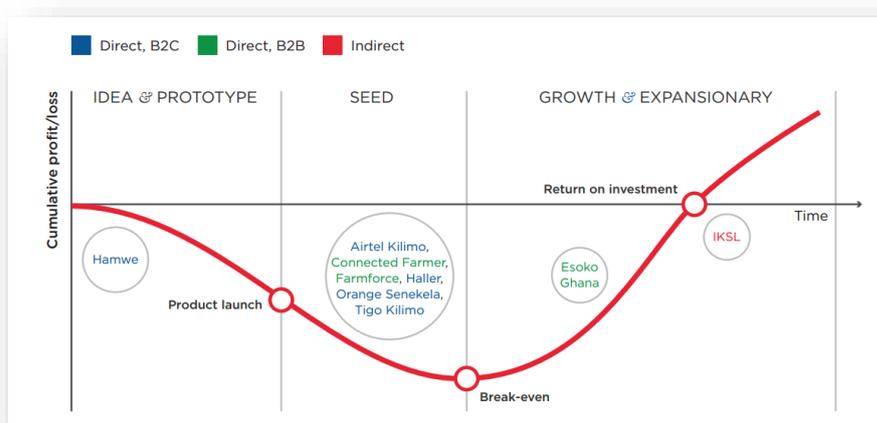
Provider	Service	Revenue
MC.com	Information (free or paid)	Part of GSMA grant; proportion of revenue
Grameen	Information plus support	Proportion of revenue
Votomobile	Information	Proportion of revenue
Audrey Pack	Packs given to women	Products donated by manufacturers
MicroEnsure	Collect insurance premiums (broker)	From premiums collected
CERsGIS	?	?
Zero Mothers Die (UNICEF)	?	Donor
VAS2Nets	Technical support?	?
Vantage Medical	?	Proportion of call revenue

Returning briefly to the mAgri propositions, both of these are clearly branded for a single MNO. As such, they could potentially reduce churn and increase ‘stickiness’, and perhaps even increase the average revenue per user for the MNO. Our modelling will consider these.

5.6 Maturity and scale

It takes time to grow a viable value-added service. In GSMA (2015) the authors illustrate how few Agri VAS (value-added service) propositions have yet to show a return.

Figure 5.4 Funding curve for Agri VAS



Source: © GSMA Intelligence (2015)

The scale of the x-axis is not defined. The propositions within the study show different degrees of maturity. As discussed above, WN in Tanzania has a product that has been

available since 2009. The addition of nutritional content will be used to expand its existing user base. Esoko, partner to Vodafone Ghana, also has an existing service, and the Vodafone Farmer Club proposition seeks to expand this. A particular challenge for this service will be the rapid expansion of the helpline from the few hundred existing customers to the tens of thousands planned for.

Table 5.4 Key variables for the four propositions under study during the inception phase

	Ghana mHealth	Tanzania mHealth	Ghana mAgri	Bangladesh mAgri
Existing service?	Yes but not aggregated	Yes	Yes	No
Degree of expansion	Challenge of multiple stakeholders	Additional customer type	Considerable growth in numbers	Start-up
Key capacity development	Cooperation of stakeholders	Integration of content	Call centre growth	Service design
Roll-out plan defined?	Not yet	Not yet	Defined	Draft
Timetable available?	Yes	Aspirational	Yes	Yes

Although most service providers have a track record, once again the detail of mHealth Ghana illustrates how their contribution to the proposed offering will be new in at least some respects. For example:

- Grameen has developed the Mobile Midwife service over many years, but it has only been operational in a handful of districts; going to national scale is new (and relations with health clinics need to be understood).
- Audrey Pack is operational in a number of countries, but not yet in Ghana.
- MicroEnsure has provided a service for Airtel for a couple of years, but the costing of the service to be offered under mHealth will change (so that customer pays).
- Vantage Medical operates call centres, but will scale up under mHealth.

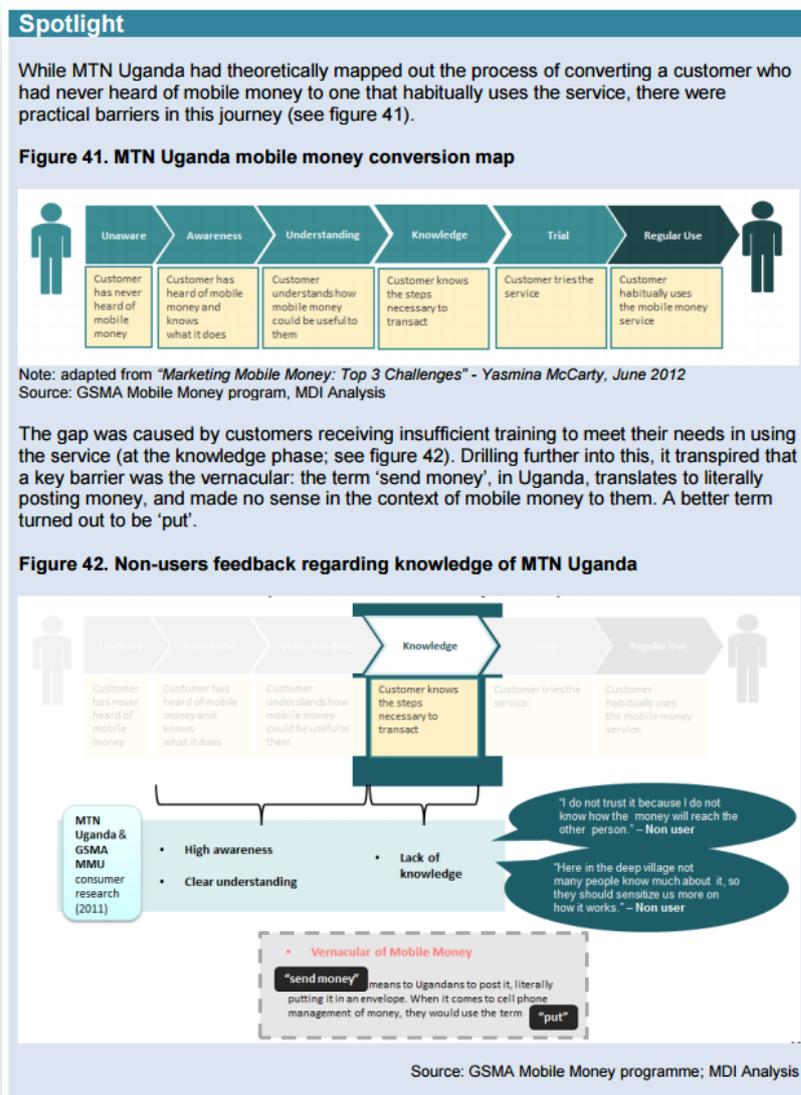
Table 5.5 mHealth Ghana, stakeholders' experience

Provider	Service	Stage of development
MC.com	Information (free or paid)	Nascent
Grameen	Information plus support	Mature, not yet at scale
Votomobile	Information	Nascent
Audrey Pack	Packs given to women	Nascent (in Ghana)
MicroEnsure	Collect insurance premiums (broker)	Young (still evolving)
CERsGIS	?	Nascent
Zero Mothers Die (UNICEF)	?	?
VAS2Nets	Technical support?	?
Vantage Medical	?	Young (?)

5.7 Customer experience

For each of these propositions the customer will have his or her own journey. Their experience may define whether the proposition finds traction or not. Recent lessons learned have shown that long multi-layered menus that take longer to navigate than a single USSD (Unstructured Supplementary Service Data) session, may well cause customers to stop using the service. Most of the propositions under study take into consideration the lessons learned as captured and documented by GSMA. We will seek to monitor the role of customer feedback and its influence on the propositions.

Figure 5.5 illustrating the potential pitfalls of customer experience



Source: © GSMA Intelligence (2013)

Relating the customer journey to the theory of change

The programme has a theory of change developed for the mAgri side by Aline/Firetail, commissioned by GSMA to undertake 'regular' monitoring and evaluation of the programme.³⁵

Data to be collected by Aline will be relevant to understanding customer behaviour and its effect on business model(s). For instance, data relating to the access chain in the diagram will include customer profiles (indicating whether the service is reaching small resource-poor farmers and indicating the gender of users). Similarly, data from repeat users will help in understanding aspects of the service design that drive repeat use of the service. Data will yield insights into a range of benefits to stakeholders, such as reduction of churn. Indeed, the data will include insights into the opportunity to sell more mobile products to the customers, which is a key value proposition.

A view on the efficiency and sustainability of the business model can be formed by working with Aline/Firetail and GSMA, shared data from their M&E activities, supplemented by specific data requests from the OPM team and situated in qualitative data from the stakeholders.

5.8 Emerging issues

Against this backdrop of multiple stakeholders, there are a number of incentive issues which will have to be considered when detailing the business modelling.

Types of partners: There is an ideological difference between service providers with for-profit goals and those with social benefit goals. The clearest example of this is in the approach to be taken at the end of a free trial period. One option is for customers to opt in (i.e. the default no-action pathway is to unsubscribe from the offering), with no payment. The other option is to use autorenewal where customers need to opt out (i.e. the default is to subscribe to the service and payments are automatically deducted).

Role of the aggregator: The aggregator plays a pivotal role in the functioning of the service (designing and operating front-end gateway for all partner services, dealing with billing and passing customers on to partners' services). If the aggregator makes money by taking a proportion of premium service subscriptions, then it has no incentive to ensure quality of the free service. If anything, it could be argued there is a perverse incentive to deliver a poor-quality free service in order to maximise the marginal benefit, and hence the attractiveness, of premium services. GSMA, in its role as programme coordinator, has the ability to ensure that quality content is made available for free to 'bottom of the pyramid' customers but GSMA's involvement is of limited duration, and the exit strategy is that country partnerships will become self-sustaining. In this instance, the longer-term viability of quality, free content may well turn out to be dependent on the balance of power between members.

Willingness to pay vs cross-subsidisation: Among partners (in Ghana) views differed over whether customers were able or willing to pay for information services. Without detailed information on customer demographics, it may be that some partners are servicing the bottom of the pyramid (e.g. Grameen?) whereas others are servicing half-way down the pyramid (e.g. VAS2Nets in Nigeria); both types are providing services to 'poor' consumers. If premium customers are expected to subsidise 'poor' customers, then the ratios need to be explored (e.g. if only one in ten customers pay for a service, then they will need to pay ten

³⁵ The theory of change is not authorised for publication

times the 'cost' of the service. Given that information services are of marginal value, this is unlikely to add up).

Gives and gets: For MNOs, the primary value of mobile for development services is often to be found in reducing churn. Offering a service that consumers like increases their 'stickiness' – they are less likely to move to competitors' networks. However, this is only the case if a service is unique to a given network, or if it operates only on a given network. A typical path for value-added development is for one network to introduce an innovative service, then for others to introduce similar services if it proves to be successful. The innovator may subsequently enjoy some advantage as a first mover (e.g. Safaricom with mPesa). The mHealth offering in Ghana has secured a national shortcode (#247), meaning that customers on any network can access the service. Whereas this increases the potential reach of the service, it is not clear what benefits MNOs expect to receive from the service.

Government roles: GSMA has approached the Ghana Health Services, which it hopes to persuade to endorse the mHealth service. The credibility this would lend to the service is expected to be an important part of marketing the service. At the time of the visit, the design of the service and the partnership had yet to be finalised, so GSMA was not in a position to submit final paperwork to the Department. The Department is subject to many requests from mHealth pilot projects seeking its endorsement/collaboration in order to achieve scale. It sees value in the gateway model that will provide nationwide access to multiple health services.

Regulation: Issues relating to telecommunications regulation were discussed (e.g. whether 'autorenewal' was permitted in Ghana, taking money from an account after a subscription period has expired). Offerings should be scrutinised to see whether any services provided by partners are also subject to regulation (e.g. call centres). There may also be tax incentives for investment by MNOs in certain activities (e.g. CSR).

B2B (business to business): At the time of the visit, the offering was being designed around services to be offered to, and paid for by, individual customers. MicroEnsure³⁶ is interesting because it acts as a broker for the insurance industry, as it does not have an insurance licence. In this respect, it is linking consumers to business. In M4D (mobile for development) more generally, there is growing interest in exploring B2B business models, and as the design of the offering may well continue to evolve during the lifetime of the evaluation, this is something to monitor.

5.9 Methodology

5.9.1 Data collection methods

The business modelling component will consist of rolling ongoing mixed data collection drawing on commercial data aggregated by GSMA and qualitative interviews with both stakeholders and clients of the services.

³⁶ Launched January 2014 and reached 1.2 million registrations before cutting back. Cost of basic cover was met by Airtel – comprehensive health cover includes life, disability and health. As it was so successful, the cost increased to the point where Airtel wanted to cut back. They sent out an SMS asking people to send a reply SMS or they would be dropped. Nevertheless, the free service still exists. They launched a premium service, which gives customers double the cover, and got 20,000 registrations in three weeks. MicroEnsure is an intermediary, without an insurance licence, so policies are underwritten by mainstream insurance companies.

Consumer-facing data

For consumers, it will draw on qualitative events outlined in Section 4.2.2: an initial exploratory qualitative study, in-depth case studies at midline, and qualitative follow-up mini-case studies following the quantitative endline. It will also draw on monitoring data, to be gathered by Aline (mAgri), and on operational data from MNOs, to be gathered by GSMA.

The parameters of the consumer-facing qualitative interviews are given in Section 4.2.2, and in brief include:

- 1) the acceptability, familiarity and use of mobile phone technology by the target groups;
- 2) factors that may affect the operation and/or access to a mobile phone and behaviour change;
- 3) current information-seeking behaviours;
- 4) social, economic and environmental factors.

Working with members of the qualitative team, the qualitative discussions shall include not only (actual and intended) expenditure on the services but attitudes to other services provided by mobile network operators (indicating propensity to churn) and to alternative services offered by other providers (such as face-to-face extension).

These insights and data will be supplemented with the monitoring data from Aline, operational data from MNOs, and the VAS providers own collection of data around the customer journey. The customer journey will be crucial to uptake of the service and is likely to change over the period of the study, being modified by the service provider in the light of feedback. The business modelling will contextualise the customer journey with the insights of the initial exploratory qualitative study, and then modify the model in the light of changes over time and insights derived from GSMA M&E data and the in-depth user-oriented case studies.

In particular, the quantitative study sample depends on encouraging high levels of take-up of the mNutrition intervention by the treatment groups (see Sections 3.1.3 and 3.2.3). This encouragement will be informed by the qualitative study and will be used to explore different incentive strategies (e.g. financial incentives, different additional marketing strategies) and assess their perceived effectiveness in promoting take-up of the mNutrition products. While this does not necessarily affect the business modelling, as different encouragement strategies are introduced, it is likely that the consumer-facing details of the business model and both the direct and imputed benefits might 'evolve'. Different marketing strategies will require different expenditure, and financial incentives may lower direct income. The balance sheet of the service including the imputed benefits, may change over time, and the business modelling will take this into account in its reporting.

Implementing agents

The qualitative data of Section 3 will include interviews with key informants (e.g. project implementation staff, health professionals, agriculture extension workers, and national and local implementation staff) who may provide a significant overview of consumer behaviour. These staff may be direct employees of an agency rolling out the service or salaried staff of 'other' stakeholders. Their view of how the service has enhanced their work will likely to be key to understanding imputed benefit.

Institutional stakeholders

We are using this term as a catch-all for all stakeholders other than customers – MNOs, VAS providers, implementing NGOs, extension agencies, government, etc. As described above, there will potentially be a wide variety of stakeholders engaged in the provision of the

service, each with their own focus and desire for imputed benefits. For these stakeholders, we will conduct in-depth interviews, carried out as opportunity arises. As demonstrated during the inception phase, it is often difficult to undertake long interviews with key individuals from private sector and senior government, and we will be opportunistic in seeking suitable timeslots. In the inception phase, the Cape Town Mobile 360 conference saw many of the stakeholders in one place, and the opportunity was used to gather strategic information. In the main study, it is likely we will use GSMA events and related workshops, and events organised by third parties, to undertake timely interviews.

The inception phase was also marked by the need to interview stakeholders at a timely point in their thinking. During inception, a number of interviews were conducted where the respondent had not yet formulated their thinking and could not give detail to the service. Alliances will come and go during the timespan of the study, and it will be important to ensure that business modelling interviews are made at a timely point in the stakeholder's journey.

There is no definitive calculation for sample size when seeking business model information. We will use 'snowball sampling' (Pole and Lampard 2002). Respondents who have already been sampled will be asked to direct us to other potential respondents. We note the challenges of snowball sampling with respondents not tied into social networks, but believe that in the case of institutional stakeholders the danger of missing key informants is minimal.

We anticipate that a minimum of 24 interviews per year will be conducted for the two main interventions. The distribution of these among stakeholder types will become evident as the programmes are implemented. For the Ghana health business modelling (unsupported by quantitative data, and with limited consumer qualitative data) we anticipate an overlap of stakeholders with some efficiencies in travel and time taken for other relevant stakeholders.

The box below sets out an indicative list of questions we will address in the stakeholder interviews. A final checklist approach will be determined by the evolution of the products and the information previously received from either the respondent or other stakeholders.

Questions to be addressed in the stakeholder interviews

- What is the history of the institutions involvement in the service roll-out?
- How does the rolling-out of the service benefit the respondent's institution?
- Is this view shared by all in the institution, particularly senior management?
- In terms of direct costs, what has been the institution's investment in this innovation?
- In terms of indirect cost, how much manpower, thought time, distraction has the development of this service taken?
- The service has a slightly different shape in some locations – is this a benefit or a distraction?
- What direct benefits does your institution gain from this service offering?
- What indirect or imputed benefits does your institution gain?
- What are some of the other indirect or imputed benefits regarding the service being a public good?
- To what extent does the service complement existing policies and procedures?
- What would be needed (in the future) for the institution to commit to scaling up their investment in, and use of, the service?

Quantitative data

In addition to the qualitative data indicated above, we will seek to work with GSMA to obtain the following data in an aggregated, non-commercially sensitive manner.

Network operators:

- numbers of subscribers (monthly, to track rate of uptake);
- usage statistics – e.g. frequency, duration;
- revenue generated;
- subscribers disaggregated by geographic area (assess rural/urban split, and assess effectiveness of encouragement);
- breakdown of different types of subscribers (where applicable) – e.g. free versus premium.

Aggregators:

- frequency of use of each of the services in the product;
- data disaggregated by subscriber profiling information;
- revenue generated;
- costs.

Service developers:

- frequency of use of services (to assess how people are using services);
- disaggregated by free/premium (where applicable);
- revenue generated;
- costs.

This data will be collected by the stakeholders, and GSMA will be asking them for such data. We expect to obtain this data only through GSMA and not directly.

5.9.2 Model analysis

As described in the proposal the model must by necessity be more than a simplified income/expenditure profit-loss model. Other related services may give indirect or imputed benefits. The inception phase has confirmed this is vital to understanding the business premise.

We will use Osterwalder and Pigneur's (2010)³⁷ inductive approach to business model generation as a general starting point. The analysis will include, where possible, a mapping of the key partners, documenting the resources required by each stakeholder to sustain the model, documenting the key activities needed to sustain the model, a view of customer segmentation, and a commentary on the value proposition from each stakeholder's point of view. Where possible, direct and indirect costs will be documented; where that is not possible, commentary will be made. Where possible, direct and imputed benefits (to the institution) from each stakeholder's point of view will be given; where that is not possible, commentary will be made.

It is highly unlikely that the service will demonstrate a simple profitable revenue model. In the GSMA logical framework, while commercial sustainability is important, the definition of 'commercial sustainability' includes imputed benefits. The business model analysis will take this into account.

³⁷ Osterwalder, A. and Pigneur, Y. (2010) Business model generation: a handbook for visionaries, game changers, and challengers. John Wiley & Sons.

Cost-effectiveness analysis

Drawing on published secondary data on other similar or equivalent service provisions and where possible unpublished data identified within each country (Tanzania and Ghana) on similar or equivalent provisions, we will seek to compare the service with alternative possibilities. We will use stakeholder interviews to determine whether unpublished data exists.

As discussed in Section 5.2, the GSMA has commissioned the Futures Group to undertake an exercise in calculating the cost benefit of the systems in terms of impact. This group is using the LiST model to calculate benefits in terms of reduction in lives lost. We will liaise with the group to ensure we support its processes, and the updated business model can feed into the then publicly available LiST for comparison with other like services.

Process analysis

The business modelling team will work closely with the quantitative and qualitative teams to inform the in-depth analysis of the intervention implementation process. As discussed in Section 4.2.2, the aim of this process analysis is to study how well the mNutrition products are actually implemented and whether there are any shortcomings in the design and/or implementation that may affect the outcomes (e.g. technical difficulties, problems with the targeting of messages with relevant content). The analysis of data will include the then known data on direct and indirect costs and on direct and imputed benefits.

5.9.3 Notes on methodology

In addition to the primary data-gathering described above, secondary documentation will likely include mNutrition product descriptions by GSMA/local MNOs, implementation guidelines and marketing material.

All qualitative research methods (including stakeholder interviews) will adhere to strict ethical standards (as described in Chapter 7 of the inception report, and including OECD DAC quality standards and IFPRI and IDS ethical guidelines). Data collection will only take place once full and informed consent is obtained from respondents. Confidentiality and anonymity of participants will be protected at all times and the real names of people and locations will be replaced with pseudonyms.

5.9.4 Risks and their mitigation

The business modelling will depend on the cooperation of GSMA and the associated stakeholders. We will seek to continue an ongoing relationship with all key players. The reduction from four studies to two presents a very real risk for the cooperation.

Commercial sensitivities were flagged as a risk in the proposal and have been discussed extensively during the inception phase. By offering the insights of the impact study, many stakeholders, including the MNOs and the aggregators, understood that they would get beneficial insights into their users, the way they use the product and the consumer experience from the baseline and ongoing qualitative work. Their cooperation was encouraged by the expected return of valuable data. The team emphasised to all stakeholders the benefits of baseline and ongoing qualitative work to refine and adjust their product. It will now be necessary to approach Ghana mHealth stakeholders for business model information without the 'carrot' of the quantitative study.

Anticipating the reviewers' assumption that the Ghana mAgri study may offer insights into Ghana mHealth (enough to be a carrot for the stakeholders), it should be noted that while

Ghana mAgri consumers may well be targeted for Ghana mHealth services, the quantitative study is unlikely to yield insights valuable to the Ghana mHealth stakeholders. The idea of the mAgri proposition is to encourage users to purchase and use a dedicated SIM card, to allow them to receive information and free calls. mHealth will be channelled through other SIM cards. Only if Vodafone decided to join in the mHealth proposition would consumer insights from one be relevant to the other. Indeed, the mAgri proposition in its current form is likely to slow take-up of the mHealth proposition in those areas.

The risk of non-cooperation within GSMA is also compounded by expectations developed during the inception phase that the baseline, endline and qualitative studies would contribute strongly to GSMA's M&E of the programme. Johns Hopkins University (JHU), which is undertaking M&E for eight other countries, was asked not to be present in Ghana in order to avoid excessive sampling of consumers. GSMA will now need to renegotiate the terms of the M&E with JHU to include Ghana, and while this is not unreasonable it is likely to lessen the friendly terms the inception team established with the GSMA health team.

If business modelling of the Bangladesh mAgri were requested, without the associated quantitative study, the established GSMA M&E through Aline would provide basic user and usage data. In undertaking the business modelling of Bangladesh, one risk is the possible disappointment of the Bangladesh stakeholders, who were keen to support an evidence-based impact study; the challenge would be to manage such disappointment.

6. Communications strategy

6.1 Statement of purpose

This communications strategy will map the audiences, communications channels and activities that will support and guide ongoing engagement with key stakeholders to actively involve them throughout the process and to facilitate take-up of the evaluation findings and lessons learned.

6.2 Communications goal and objectives

Goal

To communicate the findings of the mNutrition impact evaluation and the lessons learned in order to enhance understanding, maximise ownership, and facilitate the use of the evaluation to inform decision-making on the efficacy of mobile phone-based services as a tool to change behaviour.

Communications objectives

1. Effectively report empirical evidence on the impact of mobile phone-based nutrition and agricultural advisory services to DFID and other stakeholders.
2. Actively communicate with key stakeholders during various stages of the evaluation to share emerging findings and lessons learned (e.g. on the design and implementation of mobile phone-based advisory services for nutrition and agriculture).
3. Promote stakeholder involvement throughout the evaluation to increase the likelihood that the findings will be used to inform and influence decision-making on the use of mobile phone services for behaviour change communication in Ghana, Tanzania and globally.

Strategic direction

To become a trusted source of empirical evidence for DFID and other key stakeholders on mobile phone technology for development and in particular for the delivery of behaviour change communication in the area of nutrition and agriculture.

Communications activities will also link to other ongoing work on mobile phone use for real-time nutrition monitoring at IDS as part of DFID's accountable grant (Barnett *et al.* 2014; Barnett and Gallegos 2013; Barnett and Edwards 2014). This work is led by one of the Co-Principal Investigators of the mNutrition impact evaluation (Dr Inka Barnett).

6.3 Target audiences for the evaluation

A mapping of the evaluation stakeholders will take place during the first year, in order to fully understand the interests, priorities and information needs of the different stakeholders and to enable us to be more targeted in our communications and engagement.

During the stakeholder mapping the communications team will work with stakeholders to understand how best to engage and enhance utilisation of the evaluation findings (Datta 2012). This will include an exploration of: (a) the information needs and requirements of each stakeholder; (b) the level of detail required (e.g. brief summary or comprehensive report); (c) the preferred format and delivery channel; (d) the desired frequency of communication; and (e) the best timing and critical deadlines to communicate findings to influence decision-making processes.

The primary recipient of this work will be DFID. We will communicate frequently with DFID throughout the evaluation to: (a) understand and be able to deliver on specific information needs; and (b) keep DFID informed of the progress and preliminary findings as they emerge.

Other key stakeholders include GSMA and its national members (including local MNOs), national governments (and here in particular the Ministry of Health and Agriculture), international agencies and donors, as well as community-level health and agriculture extension workers.

During the stakeholder mapping we will identify key agents of change within each target audience, and engage and share findings with these groups/individuals on an ongoing basis. We will seek to have frequent communication with these 'change agents' to develop a constituency for the evaluation from the beginning, build up trust and confidence in the evaluation, be able to detect potential contextual changes early on and adapt the communication of the findings accordingly. All of this will increase the chances for uptake and use of the findings. We will work under the premise that engagement is 'two-way' (Benequista and Wheeler 2012) and engage with their platforms and through their channels, while inviting them to engage with the channels we use.

Engagement and communications will be targeted at global and country levels. This will help us to promote the utilisation of the findings on multiple levels and ensure long-term impact. Target audiences have been separated by global and country levels, and further detail around specific local-level audiences will be identified during the preliminary stakeholder mapping in each of the evaluation countries.

Findings of the impact evaluation will be communicated to the beneficiaries (i.e. the subscribers and active users of the m-nutrition product) and other study participants (who did not subscribe) at several stages of the evaluation as recommended by Groves 2015.³⁸ Interactive two-way conversation will be employed as part of the qualitative focus group discussions in the mini-case studies in the third round of the qualitative data collection. Beneficiaries and the qualitative evaluation team will discuss, refine and validate the findings of the initial qualitative and quantitative data analysis together. At the final dissemination stage of the impact evaluation all beneficiaries will be informed of the findings via a text message. Findings will also be communicated to local mNutrition staff, health workers and agriculture extension workers to share with local communities and subscribers.

6.4 Preliminary mapping of target audiences

Global level

Primary global stakeholders

DFID

DFID wants to understand the impact and cost-effectiveness of mobile phone technology advisory services for nutrition and agriculture and how to construct sustainable business models for these services. This information is important to inform funding decisions for new mobile phone-based advisory services to improve nutrition in developing countries. Apart from these primary aims, the impact evaluation will also provide valuable learning for DFID interests beyond nutrition and agriculture – for example, with regard to the use of mobile phones for behaviour change communication in general.

³⁸ Groves, L. (2015) Beneficiary feedback in evaluations, DFID's Evaluation Department: London, <https://www.gov.uk/government/publications/beneficiary-feedback-in-evaluation>

GSMA

For the impact evaluation, the evaluation team will work alongside GSMA and local MNOs during the implementation and first two years of operation of the mNutrition services. Ongoing communication with GSMA during this entire phase is important for the evaluation team to get an in-depth understanding of the mNutrition product that will be delivered via GSMA's national partners. The evaluation team will also communicate lessons learned on the design and implementation and from the business model analysis of the mNutrition product to GSMA to ensure the long-term improvement of the mNutrition service. GSMA and local MNOs are likely to be especially interested in whether or not the mNutrition product will attract new clients and in the findings on how to construct a sustainable business model for m-Nutrition.

International agencies and donors

An increasing number of international agencies and donors have started to integrate mobile phone technology in their nutrition, health and agriculture programmes. However, there is currently little rigorous empirical evidence on the impact, sustainability and cost effectiveness of these services and programmes (see, for example, findings from the landscaping review conducted during the inception phase of the impact evaluation).³⁹ The findings from this impact evaluation will inform donors' decision-making on their investments in such initiatives, initiate further mobile phone-based services to change behaviours and encourage the development of sustainable business models to fund them.

Other relevant global stakeholders

Academic institutions

We will engage with academic institutions and the broader academic community interested in nutrition and agriculture in resource-poor settings and/or ICTs for development to initiate critical discussion of the evaluation findings. Critical reflections on the evaluation by the wider academic community will help to enhance policy-makers' and other key stakeholders' confidence in the rigour of the evaluation and subsequently promote uptake of the evaluation. Through sharing evidence from this evaluation, we seek to stimulate future engagement around the impact of mobile phone technology as a potentially effective tool for behaviour change.

Civil society organisations (CSOs)

We will seek to build relationships with key civil society representative organisations to promote active engagement and uptake of the evaluation findings and lessons learned throughout the evaluation. Strong and clear evidence will inform CSO engagement and advocacy with government and partners, and in turn influence the decision- and policy-making process. The Scaling Up Nutrition (SUN) Network will be a key platform with which to engage and share findings. The communications team will also tap into existing CSO networks – from the Future Agricultural Consortium (FAC) to the Hunger and Nutrition Commitment Index (HANCI) – in order to communicate findings and promote uptake.

Media

The media are important not only in sharing key findings with a broad range of stakeholders, but also in influencing the general perception of the value and sustainability of the use mobile phone technology for behaviour change communication. We will make use of IDS's media relationships on a global scale, and also work with partners and stakeholders to build relationships in-country.

³⁹ See landscaping analysis conducted as part of the inception phase of the external impact evaluation (Barnett *et al.* 2016).

Country level

The communications strategy will be targeted at two mNutrition interventions: the mHealth product in Tanzania and the mAgri product in Ghana (in accordance with the focus of the impact evaluations). Table 6.1 presents a preliminary selection of country-specific key stakeholders. These stakeholders were identified during the initial country visits that informed the development of the evaluation design during the inception phase. A detailed stakeholder map for each country will be developed during the first year of the evaluation as described above.

Primary country-level stakeholder

National governments in Tanzania and Ghana

One important group of stakeholders in Tanzania and Ghana will be government officials from the Departments of Health, Agriculture and Nutrition. Evidence and lessons learned from this impact evaluation will help them to decide whether the mNutrition intervention is likely to make a difference to nutrition behaviours, practices and outcomes in their countries and whether or not to spend public funds to enhance or sustain them.

Local health and agriculture service providers

Findings on the level of effectiveness of mobile phones in delivering behaviour change communication and initiating behaviour change are particularly relevant for health and agricultural extension workers concerned with improving health/nutrition and agricultural practices at community level. Evidence from the impact evaluation will help them to decide whether or not to promote and use mobile phone-based services.

Beneficiaries and evaluation participants

Communicating the findings of the impact evaluation to beneficiaries and other evaluation participants is ethical and also important in order to validate and redefine the findings. It will take place at different stages of the evaluation process, namely during the analysis and the final dissemination stage.

Table 6.1 Country stakeholders

Audience type	Ghana mAgri impact evaluation	Tanzania mHealth impact evaluation
Government departments	Ministry of Food and Agriculture Family Health Division, Nutrition Department Ministries of Local Government and Rural Development Ministry of Lands and Natural Resources Ministry of Health Ministry of Environment, Science, Technology and Innovation Ministry of Communications	Tanzania Food and Nutrition Centre (TNFC) (government oversight group) Ministry of Health and Social Welfare Ministry of Communication, Science and Technology Ministry of Livestock and Fisheries Development Ministry of Agriculture, Food Security and Co-operatives Ministry of Community Development, Gender and Children
Community-level service providers	Agricultural extension workers	Community health workers
National MNOs	Vodafone GSMA Ghana Esoko	Airtel Vodacom
Multilateral/bilateral organisations	USAID	GAIN
Media	Ghanaian Chronicle Ghanaian Times Ghanaweb AllAfrica	The Citizen All Africa
CSOs	The Ghana Nutrition Improvement Project Ghana Coalition of NGOs in Health ACDI/VOCA Ghana Center for Democratic Development	Partnership for Scaling up Nutrition in Tanzania (PANITA) Save the Children Farm Radio International
Academic institutions	University of Ghana, Department of Nutrition and Food Science University for Development Studies, Tamale	Institute of Development Studies University of Dar es Salaam Ifakara Health Institute

6.5 Target audience channels for communications

The stakeholder mapping exercise in the first year of the evaluation will identify in further detail the specific information requirements of the different stakeholders. Table 6.2 breaks down the audiences, their information needs, channels for communicating with them and the expected outcomes.

We will collaborate with partners and stakeholders to produce outputs for the dissemination of the evaluation findings. IDS, IFPRI and Gamos already host and engage with a broad range of networks, knowledge platforms and blogs. Some of these include: DFID HDRC/PEAKS; MQSUN; Eldis; Eldis Communities; the Global Nutrition Report, Debate Graph; Secure Nutrition; Food and Nutrition Technical Assistance III Project (FANTA); CMAM Forum; Global database on the Implementation of Nutrition Action (GINA); Africa Child Health; Africa Portal; AllAfrica.

Examples of relevant blogs

Global Food for Thought; Humanitas Global Development blog; Save the Children Blogs; 1,000 Days; Action Against Hunger Blog; Development Horizons; Transform Nutrition; Future Agricultural Consortium; ZimbabweLand.

Examples of relevant online forums and networks

We will engage with a number of knowledge platforms and networks across relevant themes, including nutrition and research uptake. These platforms include:

- **Nutrition:** Food Security and Nutrition (FSN) Network; Global Forum on Food Security and Nutrition (FAO); International Malnutrition Taskforce – Malnutrition Forum; Child Health and Nutrition Knowledge Network (CHNKN); Emergency Nutrition Network; UN Standing Committee on Nutrition (UNSCN);
- **Research uptake:** EBPDN (Evidence-Based Policy in Development Network); ResUpMeetUp; Knowledge Brokers' Forum; Research to Action; LSE Impact Blog; ON Think Tanks; WonkComms; Research Impact @O3; Mobilize this!; Policy Action Network; K-star network; 3ie.

There are some key events that members of the team will participate in, or have participated in, to share findings and lessons widely. These include: the Nutrition for Growth Event following the Rio Olympics in 2016; annual United Nations General Assembly meetings every September; Making Cents Annual Conference; and the Mobiles for Development Annual Summit.

Table 6.2 Target audiences and channels for communication

Audience	Intended outcomes	Channel/method of communication	Information needs
DFID	<ul style="list-style-type: none"> To inform future design, implementation and decisions on whether or not to fund new projects/continue funding existing projects on mobile phone-based advisory services DFID to modify programmes to reflect lessons learned 	<ul style="list-style-type: none"> Regular update meetings Lessons Learnt Workshops Global meetings (e.g. Mobile for Development Annual Summit; Nutrition for Growth – Rio 2016) Evaluation reports Impact stories Newsletter Knowledge platforms 	<ul style="list-style-type: none"> Robust evidence on the effectiveness of mobile phone-based behaviour change communication for nutrition and agriculture Lessons learned on the design and implementation of mobile phone-based services Business case analysis and how to create a sustainable business model for mNutrition
GSMA/MNOs	<ul style="list-style-type: none"> To inform future design and implementation of ICTs for development programmes/projects GSMA to modify programmes to reflect lessons learned 	<ul style="list-style-type: none"> Regular update and information-sharing meetings with global GSMA representatives and local MNOs Lessons Learnt Workshops Social media Mobile for Development Annual Summit 	<ul style="list-style-type: none"> Evidence on the effectiveness of mobile phone-based behaviour change communication for nutrition and agriculture Lessons learned on the design and implementation of mobile phone-based services Whether or not mNutrition attracts new clients Business case analysis and how to create a sustainable business model for mNutrition
Governments	<ul style="list-style-type: none"> Governments use evidence to revisit and/or inform their own mHealth/mNutrition programmes and decisions on whether or not to fund new programmes/continue funding existing programmes 	<ul style="list-style-type: none"> Lessons Learnt Workshops One-to-one meetings Evaluation briefings National key events Blogs Newspapers Roundtable Impact stories Nutrition for Growth – Rio 2016 United Nations General Assembly 	<ul style="list-style-type: none"> Robust empirical evidence on the impact, cost effectiveness and sustainability of mobile phone-based advisory services to inform decision-making on whether or not to fund/continue to fund services Lessons learned on what works and what does not work with regard to behaviour change messages
International agencies and donors	<ul style="list-style-type: none"> Increased understanding of the usefulness, challenges and opportunities of mobile phone technology for behaviour change Informed decision-making with regard to 	<ul style="list-style-type: none"> Social media National and global events Blogs Newspapers Roundtable 	<ul style="list-style-type: none"> Robust empirical evidence on the impact, cost effectiveness and sustainability of mobile phone-based advisory services to inform decision-making on whether or not to

	<p>support of and investments in mobile phone-based advisory services for nutrition/agriculture behaviour change</p>	<ul style="list-style-type: none"> • Evaluation briefings • Knowledge platforms • Nutrition for Growth – Rio 2016 • United Nations General Assembly • Making Cents Annual Summit 	<p>fund/continue to fund services</p> <ul style="list-style-type: none"> • Lessons learned on what works and what does not work with regard to behaviour change messages
Academic institutions	<ul style="list-style-type: none"> • Critical discussions, increased understanding and engagement of the academic community with the challenges and opportunities of mobile phone technology for behaviour change communication in resource-poor settings 	<ul style="list-style-type: none"> • Peer-reviewed journal publications • Social media • Blogs • Knowledge platforms • Seminars/webinars • Personal contacts • Impact stories • Newsletter 	<ul style="list-style-type: none"> • Robust empirical evidence on the impact, cost effectiveness and sustainability of mobile phone-based advisory services for nutrition and agriculture
CSOs	<ul style="list-style-type: none"> • Findings are used to inform design and planning of projects and projects • Advocates use findings and evidence to influence policy-makers with regard to the use of mobile phone technology for behaviour change communication 	<ul style="list-style-type: none"> • Peer-reviewed journal publications • Social media • Blogs • Newspapers • Interactive ‘how to’ on evaluation communications (PDF) • Seminars/webinars • Personal contacts • Impact stories • Infographics • Nutrition for Growth – Rio 2016 • United Nations General Assembly • Making Cents Annual Summit 	<ul style="list-style-type: none"> • Evidence and learning to inform planning and delivery of ICT and development-related sensitive programming • Strong evidence to inform CSO engagement and advocacy with government and partners • Complex findings presented in an accessible and shareable format
Beneficiaries/evaluation participants	<ul style="list-style-type: none"> • Beneficiaries/participants use findings to decide whether or not to continue/subscribe to mNutrition product 	<ul style="list-style-type: none"> • SMS • Focus group discussions • Via health workers and agriculture extension workers 	<ul style="list-style-type: none"> • Easily accessible information on the effectiveness of the mNutrition product
Media	<ul style="list-style-type: none"> • Findings are shared broadly with key stakeholders • Broad coverage of case studies in key media outlets 	<ul style="list-style-type: none"> • Social media • Email • Face to face • Global events • IDS website • Impact stories • Nutrition for Growth – Rio 2016 	<ul style="list-style-type: none"> • Better understanding of the need for evaluations in implementing ICT programmes • Strong appropriate evidence for advocacy to government and general public

The Lessons Learnt Workshops and Roundtables will be valuable spaces both to develop stakeholder capacity and to deepen the evaluation team's understanding of the intervention impact. Capacity-building will be around understanding and sharing knowledge of approaches to evaluation methodologies and effective behaviour change communication with mobile phone technology. Additionally, as there is currently a lack of evidence around how to communicate findings from evaluations, we will produce a guide on impact evaluation communication (Jones *et al.* 2009). The purpose of this is to develop the capacity and knowledge of others to enable them to deliver successful and sustainable communications as part of an impact evaluation to ensure effective utilisation.

All the activities will be funded through the budget for communications outputs and support, Lessons Learnt workshops, and travel costs.

6.6 Monitoring and evaluation of the communication activities

This communication work will embed clear M&E indicators to assess and measure engagement and impact of the different activities. Our strategy for M&E will be further developed and outlined during year one.

The M&E aims to assess whether the mNutrition impact evaluation findings and the lessons learned throughout the evaluation process have influenced key stakeholders either in their decision-making or in their behaviour/attitude towards mobile phone-based advisory services. M&E will seek to monitor impact at global and national levels.

We will use three methods in undertaking the M&E:

1. **Initial impact:** where mNutrition evaluation outputs will be published, measurement of publication downloads, media hits, social media and blog traffic;
2. **Medium-term impact:** involvement of the mNutrition evaluation team in policy meetings at national and international levels, invitations issued to team members for participation in external academic and policy seminars and conferences;
3. **Long-term impact:** the use of mNutrition evaluation findings to inform future projects on the use of mobile phone technology for behaviour change communication in developing country contexts, the level of demand for additional research, evidence of influence in policy statements and documents, media coverage and blog-postings.

Through stakeholder mapping and engagement, we will introduce a survey to ascertain stakeholder understanding of and perspectives on the use of mobile phones for behaviour change activities. The communications team will design a brief survey to look at knowledge and attitudes among a set of key stakeholders at the start of the programme, in order to track change across the lifetime of the mNutrition impact evaluation. The survey will be used to find out how key stakeholders have engaged with the mNutrition evaluation findings and lessons learned, how useful they have found the outputs, how the project has influenced their thinking and practice, and if there are examples of the utilisation of evaluation findings.

The survey will also ask respondents to identify key issues related to agriculture, nutrition, ICTs, and development in general (in close collaboration with the context analysis as part of

the qualitative impact evaluation stream), and to rank issues and debates to enable the project team to monitor the way in which debates are changing. It is anticipated that this survey will engage with, broadly, the same set of individuals over the programme's lifespan. The survey tool will be designed by the communication team with input from the evaluation team and Southern evaluation partners as appropriate, and will be conducted over the phone/web communications platforms.

A report will be produced employing this data and it will seek to assess the impact of the communication activities for this evaluation.

7. Governance, management and ethics

7.1 Governance

The evaluation is funded by DFID under the structure of its Global Evaluation Framework Arrangement (GEFA), with Oxford Policy Management (OPM) leading the GEFA administrative management.

IDS is overseeing the evaluation process and provides a project management function, including facilitation of an independent peer review process for major outputs before they are submitted to OPM. In addition, all outputs from the evaluation will be reviewed by DFID advisers from the Agricultural Research team and Policy Division, in addition to others as determined by DFID. Finally, an independent external specialist evaluation and quality assurance service (SEQAS) contracted by DFID will provide specialist technical advice and recommendations on the evaluation design and quality. SEQAS will provide feedback and recommendations on this inception report and it is expected that a similar review will be conducted for the baseline and endline reports.

Any contributions and suggestions made from DFID regarding the evaluation design or final products will be considered carefully by the evaluation team and action taken as needed. However, care will be taken not to jeopardise the independence of the evaluation.

7.2 Evaluation partners and personnel

Institute of Development Studies (IDS)

IDS is a leading global charity for research, teaching and information on international development. Our vision is a world in which poverty does not exist, social justice prevails and economic growth is focused on improving human well-being. IDS believes that research knowledge can drive the change that must happen in order for this vision to be realised.

Founded in 1966, IDS enjoys an international reputation based on the quality of our work and our commitment to applying academic skills to real-world challenges. IDS' theory of change is at the heart of what we do. IDS thinks that knowledge should be generated by sound methodology and in partnership with other development and non-development actors.

The Institute is home to approximately 100 researchers, 70 knowledge services staff, 65 professional staff and about 200 students at any one time. But the IDS community extends far beyond, encompassing an extensive network of over 360 partners, 2,100 alumni and hundreds of former staff across the development community worldwide.

Among other areas, IDS has particular experience and expertise in the area of nutrition and our work with partners aims to inform and shape current debate. IDS research and research uptake work is contributing to the effort to accelerate malnutrition reduction by showing how

to define, support and evaluate enabling environments for sustained nutrition progress. IDS has been working with partners worldwide, ensuring that the momentum and political will to put an end to undernutrition is sustained.

International Food Policy Research Institute (IFPRI)

IFPRI, an international non-profit organisation, conducts research to provide policy solutions that reduce poverty and end hunger and malnutrition throughout the developing world in an environmentally sustainable manner. For 40 years, IFPRI has worked with policy-makers, academics, NGOs, the private sector, development practitioners, and others to undertake research, capacity strengthening, and policy communications activities. These activities are designed to address the broad range of economic development and poverty reduction issues that bear on IFPRI's mission.

The Institute is one of 15 organisations worldwide that make up the CGIAR. IFPRI's vision is a world free of hunger and malnutrition. IFPRI's mission is to provide research-based policy solutions that sustainably reduce poverty and end hunger and malnutrition.

More than half of IFPRI's research activities focus on sub-Saharan Africa and approximately one-third on Asia, particularly South Asia. IFPRI's work is carried out by four research divisions and one communications division. Research is undertaken at community, national, regional and global levels across six strategic research areas (Ensuring Sustainable Food Production; Promoting Healthy Food Systems; Improving Markets and Trade; Transforming Agriculture; Building Resilience; Strengthening Institutions and Governance) with a cross-cutting gender theme supported by the cross-cutting themes of policy communications, partnerships, and capacity strengthening.

IFPRI's researchers come from various professional and cultural backgrounds. Most of them are agricultural economists, while others are nutritionists, geographers, engineers, anthropologists, sociologists and political scientists. IFPRI has more than 500 staff – one-third of whom are senior researchers – from 50 countries. More than 35 senior staff are outposted to IFPRI's regional or project offices in Asia (Beijing, Dhaka, Islamabad and New Delhi) and sub-Saharan Africa (Abuja, Accra, Addis Ababa, Dakar, Kampala and Lilongwe).

The Institute has been successful in building relationships with a broad range of constituencies – policy-makers, national agricultural research institutes, developing-country universities, NGOs, international and regional organisations, and local communities. IFPRI headquarters are in Washington, DC.

Gamos

Gamos Ltd works with the social factors surrounding development interventions, particularly the adoption of technology. Gamos has a strong track record in conducting field-level assessments as part of a range of sectoral interventions, particularly infrastructure (ICTs) and broader aspects of livelihoods (e.g. household poverty, security, vulnerability and HIV/AIDS). It is currently a member of three enabling frameworks with UKAid.

Having worked with the telecommunications private sector for more than 15 years, Gamos brings deep experience of working in the mTech sector in a development context, together with expertise on how to evaluate mTech interventions. Last year Gamos conducted an mAgric landscaping study for the International Sustainability Unit of HRH Prince Charles with the GSMA as advisers to the process. Gamos also has experience of baseline knowledge, attitude and practice surveys as part of wider M&E programmes – for instance, for Tearfund's IMPACT (Improving Parent and Child Outcomes) programme in Malawi and Nigeria, which focuses on vulnerable mothers, promoting maternal health, maternal healthcare and the care of infants through strategic use of mobile phones. Gamos is

currently engaged in an ESRC-funded project researching ICTs and the changing health knowledge economy in Bangladesh. The programme is exploring the complexity of seeking health information and the growth of ICTs as a potentially disruptive technology that could lead to substantial changes in the health knowledge economy.

7.3 Management, communications, intellectual property and data

7.3.1 Management

This evaluation has three principal co-investigators, one from each of the three collaborating organisations – IDS, IFPRI and Gamos. The overall management and coordination of the evaluation will be the responsibility of IDS, with responsibilities for specific sub-components of the evaluation divided between IDS, IFPRI and Gamos. The team will follow a timeline for delivering outputs as agreed with DFID.

IDS has appointed a part-time project manager responsible for leading on the general management and coordination of the evaluation programme activities, ensuring effective internal communication between partners and externally, and reporting to OPM on behalf of all partners on the evaluation's progress on a six-monthly basis; OPM, in turn reports to DFID.

All the evaluation partners have significant experience in managing and/or contributing to these types of complex mixed-method multi-partner evaluations. Strong leadership from the three principal co-investigators, as well as effective cross-team working and transparency, are considered central to the overall management of the evaluation. The core multidisciplinary evaluation team incorporates project directors, research fellows, research assistants, a programme manager and a communications officer.

The evaluation team will engage the support of trusted and established in-country partners identified during the development of the proposal to carry out fieldwork activities in Ghana and Tanzania. While the project budget does not include funding for capacity-building activities, country-level mutual learning will be incorporated into designated Lessons Learnt Workshops, as well as through ongoing communication and engagement with the fieldwork partners.

7.3.2 Internal reporting and communications

IDS will convene frequent virtual meetings for the evaluation partners in order to ensure open and ongoing communication and encourage close engagement between the quantitative, qualitative and business aspects of the evaluation. IDS will also collate progress from the evaluation partners every six months to submit a report to OPM. OPM will include this in their reporting to DFID.

The IDS evaluation project manager will hold primary responsibility for internal and external communications for the evaluation. A web-based file-sharing system has been created to facilitate cross-partner information-sharing, and provide a central repository for all key documentation, communications and deliverables relating to the project. Partner organisations contributing jointly to a specific evaluation component are responsible for maintaining close working relationships and direct channels of communications between partners, in addition to internal information exchange.

As well as regular reporting via OPM, IDS, along with the three principal investigators, will maintain contact directly with DFID on operational issues likely to affect programme partners, and on any major modification to the design of the evaluation or its timeline.

7.3.3 Intellectual property and data

Intellectual property

Utilising the direct flow-down provisions from the head contract between OPM and DFID, each party will retain all intellectual property in any material it develops during the course of the agreement and will grant to each other party, OPM and DFID, a worldwide non-exclusive, irrevocable, royalty-free licence to use all the material, including the reproduction, publication and sub-licence of any of these materials.

Data

The evaluation team will ensure that all its data storage devices have password-protected security in place so that any data generated, processed, stored or transmitted for the purposes of developing the evaluation cannot be accessed by unauthorised persons. The team will also ensure that its data storage devices are protected by the latest versions of anti-virus definitions from an industry-accepted anti-virus software vendor. Finally, the evaluation team will not store, copy, disclose or use data except as necessary for the performance of this evaluation.

7.4 Ethical considerations

The evaluation team is highly experienced in dealing with ethical concerns in difficult contexts, and has in place well-developed sets of ethical principles, norms and codes of practice, which will be closely followed and monitored throughout the project.

As a guiding principle, the evaluation will be conducted in a professional and ethical manner, with strict respect for principles of integrity, honesty, confidentiality, voluntary participation, impartiality and the avoidance of personal risk. Adherence to these guiding principles will be overseen by IDS in collaboration with its Research Ethics Committee.

The evaluation team will adhere to the appropriate ethical guidelines for development evaluation, including:

1. OECD (2010) DAC Quality Standards for Development Evaluation on ethical standards, which states:

'Evaluation abides by relevant professional and ethical guidelines and codes of conduct for individual evaluators. Evaluation is undertaken with integrity and honesty. Commissioners, evaluation managers and evaluators respect human rights and differences in culture, customs, religious beliefs and practices of all stakeholders. Evaluators are mindful of gender roles, ethnicity, ability, age, sexual orientation, language and other differences when designing and carrying out the evaluation.'

2. DFID's 'Ethics Principles for Research and Evaluation' (2011:2) which states:

'Research and evaluation should usually be independent of those implementing an intervention or programme under study. Independence is very important for research and evaluation; in fact evaluations in DFID can only be classified as such when they are led independently. Involvement of the stakeholders may be desirable so long as the objectivity of a study is not compromised and DFID is transparent about the roles played. Any potential conflicts of interest that might jeopardise the integrity of the methodology or the outputs of research/evaluation should be disclosed. If researchers/evaluators or other stakeholders feel that undue pressure is being put on them by DFID officials, such that their independence has been breached, this should

be reported to the Head of Profession for Evaluation who will take appropriate action.'

The review and continuing oversight of any extensive research into human subjects performed as part of the quantitative component of the evaluation, has been delegated to the Institutional Review Board (IRB), based at IFPRI (which is leading on this component). IFPRI is guided by the ethical principles regarding all research involving humans as subjects, as set forth in the report of the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research [*Ethical Principles and Guidelines for the Protection of Human Subjects of Research ('The Belmont Report')*], regardless of whether the research is subject to federal regulation or with whom conducted or source of support (i.e. sponsorship).

The mission of the IRB is to protect human subjects by complying with the code of US federal regulations established by the Office of Human Research Protection (OHRP) and the international guidelines for ethical research. All research involving human subjects and conducted by IFPRI researchers must be reviewed and approved by the IFPRI IRB prior to the beginning of fieldwork.

All IFPRI researchers are required to complete the Collaborative Institutional Training Initiative (CITI) training programme on ethics for social science researchers.

Local IRB approval has not been necessary, as there is no ethical review mechanism in place for minimal-risk surveys that do not involve collection of blood, urine, stool, saliva or other biological specimens.

The evaluation involves the collection of potentially sensitive data that may affect the privacy of subjects, including children. Data will be collected by trained enumerators on areas including anthropometric indicators for children aged 0–24 months, as well as knowledge, behaviours and practices of children's mothers, childcare providers and access to mobile phone services. Since the livelihood interventions are targeted at rural households, respondents are likely to be economically and educationally disadvantaged. The potential risk of stress or discomfort through participation in the evaluation is deemed minimal, although care will be taken to minimise risk exposure.

Mechanisms that will be set in place to ensure compliance with ethical principles throughout the evaluation include:

- Members of the survey supervision staff will meet with village leaders to describe the overall scope, purpose and duration of the study and seek approval to approach specific households.
- Interviews will take place only when full and informed consent is obtained (through signed consent forms when appropriate or through verbal agreement). The evaluators will describe to participants in the clearest possible terms the content and purpose of the study, possible harm or discomfort that it may entail, and the degree of anonymity and confidentiality that will be provided. Respondents will be able to end participation at any time, and interviewers will be instructed to finish the interview if a respondent becomes unduly distressed. Efforts will be made to conduct interviews privately and at flexible times.
- If, at any stage, any of the evaluators consider that security or emotional comfort of respondents or interviewers may be in question, more sensitive questions will be re-evaluated and interviews may be terminated.
- Anonymity will be ensured in the datasets by changing names and removing personal data from reports prior to data entry. Processing of data will be made using

anonymous files. Names and personal details that may lead to the identification of participants will be removed and/or changed to the largest possible extent that will not compromise the integrity of the project.

- Public use data will include no identified individuals.

References

- Akrofi, A., Brouwer, I.D., Price, L.L. and Struik, P.C. (2010) 'Home Gardens Contribute Significantly to Dietary Diversity in HIV/AIDS Afflicted Households in Rural Ghana', *J Hum Ecology*, 31(2): 125-134.
- Ajzen, I. and Fishbein, M. (1980) *Understanding attitudes and predicting social behavior*, Englewood Cliffs, NJ: Prentice-Hall.
- Bamberger *et al.* (2012a) 'Qualitative impact evaluation approaches', in: *Real World Evaluations: Working under Budget, Time, Data and Political Constraints*, Thousand Oaks, CA: Sage.
- Bamberger *et al.* (2012b) 'Addressing threats to validity in qualitative evaluations', in *Real World Evaluations*, Thousand Oaks, CA: Sage.
- Bandura, A. (1989) 'Social cognitive theory', in E.T.H. Paul, A.M. Van Lange and Arie W. Kruglanski (eds), *Handbook of Theories of Social Psychology*, Vol. 6, pp. 1–60, London: Sage.
- Barnett I., Scott N., Batchelor S. and Haddad L. (2016) 'Dial "N" for Nutrition? A Landscape Analysis of what we know about mNutrition, mAgriculture and mDevelopment', *IDS Working Paper*, Volume 2016, Number 481, Brighton: IDS.
https://opendocs.ids.ac.uk/opendocs/bitstream/123456789/12642/1/Wp481_Online.pdf
- Barnett, I. and Gallegos, J. (2013) 'Using Mobile Phones for Nutrition Surveillance: A Review of Evidence', *IDS Evidence Report 1*, <https://www.ids.ac.uk/publication/using-mobile-phones-for-nutrition-surveillance-a-review-of-evidence>
- Barnett, I. *et al.* (2014) 'A Mixed-Method Impact Evaluation Design of a Mobile Phone Application for Nutrition Service Delivery in Indonesia', *IDS Evidence Report 79*, www.ids.ac.uk/publication/a-mixed-method-impact-evaluation-design-of-a-mobile-phone-application-for-nutrition-service-delivery-in-indonesia
- Barnett, I. and Edwards, D. (2014) 'Mobile Phones for Real-time Nutrition Surveillance: Approaches, Challenges and Opportunities for Data Presentation and Dissemination', *IDS Evidence Report 75*, www.ids.ac.uk/publication/mobile-phones-for-real-time-nutrition-surveillance-approaches-challenges-and-opportunities-for-data-presentation-and-dissemination
- Batchelor, S. (2012) 'Changing the Financial Landscape of Africa: An Unusual Story of Evidence-informed Innovation, Intentional Policy Influence and Private Sector Engagement', *IDS Bulletin*, 43: 84–90,
<http://opendocs.ids.ac.uk/opendocs/bitstream/handle/123456789/6012/Draft%20-%20Changing%20the%20Financial%20Landscape.pdf?sequence=1>
- Benequista, N. and Wheeler, J. (2012) 'Cartographers, Conciliators and Catalysts: Understanding the Communicative Roles of Researchers', *IDS Bulletin* 43.5: 45-52.
- Bruhn, M. and McKenzie, D. (2009) 'In Pursuit of Balance: Randomization in Practice in Development Field Experiments', *American Economic Journal. Applied Economics* 1(4): 200-232.

Creswell, J.W. (2013) *Research design: Qualitative, quantitative and mixed method approaches*, Thousand Oaks, CA: Sage.

Datta, A. (2012) 'Deliberation, Dialogue and Debate: Why Researchers need to Engage with Others to Address Complex Issues', *IDS Bulletin* 43.5: 9-16.

DFID (2011) 'Ethics principles for Research and Evaluation'.
www.dfid.gov.uk/Documents/publications1/evaluation/dfid-ethics-prcpls-rsrch-eval.pdf

Duflo, E., Glennerster, R. and Kremer, M. (2007) 'Using Randomization in Development Economics Research: A Toolkit', *Handbook of Development Economics, Volume 4* (Elsevier B.V.), pp. 3895-3962.

Duflo, E., and Saez, E. (2003) 'The Role of Information and Social Interactions in Retirement Plan Decisions: Evidence from a Randomized Experiment', *The Quarterly Journal of Economics* 118(3): 815–842.

EIU (2010) *Fighting Smart: Strategy options for telecoms operators*,
http://www.economistinsights.com/sites/default/files/EIU_Fighting%20smart_Final_WEB.pdf

Ghana Nutrition Landscape Analysis, GAIN presentation, n.d.

Ghana Statistical Service (GSS), Ghana Health Service (GHS) and ICF Macro (2009) *Ghana Demographic and Health Survey 2008*, Accra, Ghana: GSS, GHS, and ICF Macro.

GLSS Round 6, August 2014

Groves, L (2015) *Beneficiary feedback in evaluations*. DFID's Evaluation Department: London, <https://www.gov.uk/government/publications/beneficiary-feedback-in-evaluation>

GSMA (2013) 'Scaling Mobile for Development', *Analysis Paper*,

<https://gsmaintelligence.com/research/?file=130828-scaling-mobile.pdf&download>

GSMA (2015) 'GSMA Intelligence Agri VAS: market opportunity and emerging business models', *Analysis Paper*,

<https://gsmaintelligence.com/research/?file=3d4627c21554cbaae585d8a9dc022a2e&download>

GSMA M4D (2013) mNutrition Initiative, October 2013 presentation

Hemming, K., Girling, A., Sitch, A., Marsh, J. and Lilford, R. (2011) *BMC Medical Research Methodology*, 11:102.

Institute of Medicine of the National Academies (2014) *Evaluation design for complex global initiatives: workshop summary*. Washington, DC: National Academies Press. Available from: www.nap.edu/catalog.php?record_id=18739

Jones, N., Jones, H., Steer L. and Datta, A. (2009) 'Improving Impact Evaluation Production and Use', *ODI Working Paper 300*, London: Overseas Development Institute.

Kennedy, G., Ballard, T. *et al.* (2011) 'Guidelines for Measuring Household and Individual Dietary Diversity', Rome, Food and Agriculture Organization.

Kothari, M. and Noureddine, A. (2010) 'Nutrition Update 2010, Calverton', Maryland, USA: ICF Macro.

- Marchal, B., Dedzo, M. and Kegels, G. (2010) 'A realist evaluation of the management of a well-performing regional hospital in Ghana', *BMC Health Services Research*, 10: 24.
- Mason, J. (2006) 'Six strategies for mixing methods and linking data in social science research', *Real Life Methods working paper*, University of Manchester.
- Maxwell, J.A. (1992) 'Understanding and validity in Qualitative Research', *Harvard Educational Review* 62(3):279-300.
- Maxwell, J.A. (2004) 'Using qualitative methods for causal explanation', *Field Methods* 16(3), 243–264.
- May, V. (2010) *Realities Toolkit #12: What to do with contradictory data?*, National Centre for Research Methods, University of Manchester.
- Mays, N. and Pope, C. (1996) 'Rigour in qualitative research', in N. Mays and C. Pope (eds). *Qualitative research in health care*, London: BMJ Books.
- McKenzie, D. (2012) 'Beyond baseline and follow-up: the case for more T in experiments', *Journal of Development Economics* 99(2): 210-221.
- Mohr, L.B. (1999) 'The qualitative option of impact analysis', *American Journal of Evaluation* 20(1): 69-84.
- Morell, J.A. (2005) 'Why are there unintended consequences of program action, and what are the implications for doing evaluation?', *American Journal of Evaluation* 26: 444-6.
- Morse, J.M. (1995) 'The significance of saturation', *Qualitative Health Research*, 5: 147-9.
- Mukherjee, N., *et al.* (1999) 'Project benefit-impact matrix from the Maldives – a method for participatory evaluation', *PLA Notes*, 35: 13–15.
- OECD (2010) 'Quality Standards for Development Evaluation'. *DAC Guidelines and Reference Series*. <https://www.oecd.org/development/evaluation/qualitystandards.pdf>
- Osterwalder, A. and Pigneur, Y. (2010) *Business model generation: a handbook for visionaries, game changers, and challengers*, John Wiley & Sons.
- Patton, M.Q. (2002) *Qualitative research and evaluation methods*, Thousand Oaks, CA: Sage.
- Pawson, R. (2013) *The Science of Evaluation: a Realist Manifesto*, London: Sage.
- Pawson, R. and Tilly, N. (1997) *Realistic Evaluation*, London: Sage.
- Pole, C.J. and Lampard, R. (2002) *Practical social investigation: Qualitative and quantitative methods in social research*, Pearson Education.
- Ruel, M.T., Harris, J. and Cunningham, K. (2013) *Diet quality in developing countries*, Volume 2, Chapter 18, in *Diet Quality: An Evidence-Based Approach*, Victor R. Preedy, Lan-Anh Hunter, Vinood B. Patel (eds), New York, Heidelberg, Dordrecht, London: Springer, pp 239-261.
- Slonim-Nevo, V. and Slonim, I. (2009) 'Conflicting findings in mixed methods research: An

illustration from an Israeli study on immigration', *Journal of Mixed Methods Research*, 3(2): 109-128.

Thorne-Lyman, A.L., Valpiani, N., Sun, K., Semba, R.D., Klotz, C.L., Kraemer, K., Akhter, N., *et al.* (2010) 'Household dietary diversity and food expenditures are closely linked in rural Bangladesh, increasing the risk of malnutrition due to the financial crisis', *The Journal of Nutrition* 140, 1: 182S-188S.

UNICEF (2011) *Infant and Young Child Feeding: Programming Guide*, May 2011.

USAID (2012) *Ghana Feed the Future Baseline Survey 2012*.

Woolcock, M. (2013) 'Using case studies to explore the external validity of 'complex' development interventions', *Evaluation* 19(3): 229-48.

White, H. (2009) 'Theory-based impact evaluation: Principles and practice', *Working Paper 3*, New Delhi: International Initiative for Impact Evaluation.

World Bank (2010) *World Development Indicators*.

Yin, R. (2013) *Case study research: Design and methods* (5th edn), Thousand Oaks, CA: Sage.

Annex 1 Terms of Reference

Section 4, Annex

A

Call-down contract

Terms of reference

PO 6420: External evaluation of mobile phone technology-based nutrition and agriculture advisory services in Africa and South Asia

Introduction

DFID (Research and Evidence Division) wishes to commission an external impact evaluation of mNutrition, a mobile phone technology-based nutrition and agricultural advisory service for Africa and South Asia. mNutrition is a programme supported by DFID that, through business and science partnerships, aims to build sustainable business models for the delivery of mobile phone technology-based advisory services that are effective in improving nutrition and agricultural outcomes.

mNutrition is primarily designed to use mobile phone-based technologies to increase the access of rural communities to nutrition and agriculture-related information. The initiative aims to improve knowledge among rural farming communities, especially women, and support beneficial behaviour change as well as increasing demand for nutrition and agriculture extension services. The mNutrition initiative launched in September 2013 will work in ten countries in Africa (Cote d'Ivoire, Ghana, Malawi, Mozambique, Nigeria, Tanzania, Kenya, Rwanda, Uganda, Zambia) and four countries in South Asia (Bangladesh, India, Pakistan and Sri Lanka). The desired impact of mNutrition will be improved nutrition, food security and livelihoods of the poor.

Mobile phone-based services have been endorsed by WHO as an effective strategy for behaviour change and for driving adherence to anti-retroviral treatment protocols (Horvath, Azman, Kennedy and Rutherford 2012). There is currently scant evidence on the impact and cost-effectiveness of mobile phone technology-based services for nutrition and agriculture and on the sustainability of different business models for their provision. A rigorous evaluation of mobile phone technology-based nutrition services would add significantly to the current evidence base. An external evaluation team managed by the evaluator, independent of the programme delivery mechanism, will conduct an assessment of the impact, cost effectiveness and sustainability of mobile phone technology-based information and behaviour change messages for nutrition and agriculture.

Background to mNutrition

Introduction

Undernutrition is a major challenge to human and economic development globally. It is estimated that almost one billion people face hunger and are unable to get enough food to meet their dietary needs. Agriculture is a major source of livelihood in many poor countries and the sector has a potentially critical role in enhancing health, specifically maternal and child health and nutritional status. A well-developed agriculture sector will deliver increased and diversified farm outputs (crops, livestock, non-food products) and this may enhance food and

nutrition security directly through increased access to and consumption of diverse food, or indirectly through greater profits to farmers and national wealth. Better nutrition and health of farmers fosters their agricultural and economic productivity. Current agricultural and health systems and policies are not meeting current and projected future global food, nutrition and health needs.

Despite major investment in agricultural and nutrition research and its uptake and application, there is significant social and geographic inequality in who benefits from these investments. Furthermore, in many developing countries, public extension systems for agriculture, health and nutrition are inefficient, have limited capacity and have a poor track record of delivery, especially in terms of supporting women and girls and the most marginalised populations (Alston, Wyatt, Pardey, Marra, and Chan-Kang 2000; Anderson 2007); IFPRI 2010; Van den Berg and Jiggins 2007).

Several research and mobile network operators (MNOs) are testing a range of information and communication technology (ICT) solutions for improving access to a wide range of information and advisory services. Mobile phone-based technologies are among the most promising ICT strategies, although current initiatives in nutrition are relatively small and fragmented.

What is mNutrition?

Enhancing access to the results of nutrition and agricultural research and development is potentially critical for improving the nutrition, health and livelihoods of smallholders and rural communities. mNutrition will harness the power of mobile phone-based technologies and the private sector to improve access to information on nutrition, health and agricultural practices, especially for women and farmers (both male and female). Specifically, mNutrition will initiate new partnerships with business and science to deliver a range of services including:

- an open-access database of nutrition and agriculture messages for use in mobile phone-based communication (for example, information and behaviour change messages on practices and interventions that are known to have a direct impact on nutrition or an indirect impact via for example agriculture);
- a suite of mobile phone-based nutrition and agriculture information, extension and registration services designed to: improve knowledge and generate beneficial behaviour change in nutrition and agriculture; increase demand for nutrition, health and agriculture goods and services; register and identify target populations for support; and, using real-time monitoring, support the conduct of nutrition risk assessments by community health workers.

The impacts of mNutrition are expected to include improved nutrition, food security and livelihoods of the poor, especially women in ten countries in Africa (Cote d'Ivoire, Ghana, Kenya, Malawi, Mozambique, Nigeria, Rwanda, Tanzania, Uganda and Zambia) and four countries in South Asia (Bangladesh, India, Pakistan and Sri Lanka). This impact will result from the increased scale and sustainability of mobile phone-based nutrition and agricultural-based information services, delivered through robust public private partnerships in each country.

mNutrition has two major outcomes. One outcome will be cost-effective, sustainable business models for mobile phone-enabled nutrition and agriculture services to three million households in ten countries in Africa and four countries in South Asia that can be replicated in other countries. Linked to this outcome, the second outcome will expect these services to result in new knowledge, behaviour change and adoption of new practices in the area of agriculture and nutrition practices among the users of these mobile phone-based services.

These outcomes will be achieved through four outputs:

- Improved access to relevant mobile-based health, nutrition and agricultural advisory services for three million poor people and community health workers across ten sub-Saharan African and four South Asian countries;
- Launch and scaling of mobile phone-based health, nutrition and agricultural advisory services targeted at poor people and community health workers;
- Generation and dissemination of high-quality research and evidence on the impact, cost effectiveness and sustainability of mobile phone-based advisory services in nutrition and agriculture in South Asia and sub-Saharan Africa; and
- Development of locally relevant content for mobile phone technology-based agriculture and nutrition services meeting demands from users and community health workers.

In terms of promoting behaviour change and/or adoption of new practices, mNutrition will seek to achieve changes in one or more of the following areas:

- Adoption of new agricultural practices that are nutrition sensitive, improve agricultural productivity and utilise post-harvest technologies;
- Changes in nutrition practices in either one or several knowledge domains including improved maternal nutrition practices during pregnancy; infant and young child feeding practices; and micro-nutrient supplementation to children at risk (i.e. vitamin A, zinc and oral rehydration solution (ORS)).

mNutrition started implementation in September 2013. For the two countries selected for the impact evaluation (Tanzania and Ghana), mobile network operators and content providers have been identified through a competitive process undertaken during the first half of 2014. The MNOs and content providers started developing and launching their services during the 4th quarter of 2014 and early 2015. The mobile phone-based advisory services are expected to run at least until the 3rd quarter of 2018.

mNutrition project coordination

DFID support to mNutrition will be channelled to GSMA, as well as directly to this associated independent external impact evaluation. GSMA is a global body that represents the interests of over 800 mobile operators. GSMA already works with the major mobile operators across Africa (including Airtel, MTN, SafariCom/VodaCom), with a collective mobile footprint of more than 67 per cent of total African connections. GSMA has a number of existing development initiatives, including mHealth and mFarmer, that are part of GSMA's Mobile for Development, which brings together mobile operator members, the wider mobile industry and the development community to drive commercial mobile services for underserved people in emerging markets. GSMA will provide technical assistance to mobile phone operators and support new partnerships with content providers to develop and scale up new nutrition and agriculture message services. GSMA will ensure sharing of best practices and promote wider replication and uptake of effective business models.

Objective and main questions

The objective of this work is to conduct an external evaluation of the impacts and cost effectiveness of the nutrition and agriculture advisory services provided by mNutrition compared to alternative advisory services available in the two selected countries (Ghana and Tanzania), with particular attention paid to gender and poverty issues. The impact assessment is required to answer the following questions that relate to impact, cost effectiveness and commercial viability:

- What are the impacts and cost effectiveness of mobile phone-based nutrition and agriculture services on nutrition, health and livelihood outcomes, especially among women, children and the extreme poor?

- How effective are mobile phone-based services in reaching, increasing the knowledge and changing the behaviour of the target groups?
- Has the process of adapting globally agreed messages to local contexts led to content which is relevant to the needs of children, women and poor farmers in their specific context?
- What factors make mobile phone-based services effective in promoting and achieving behaviour change (if observed) leading to improved nutrition and livelihood outcomes?
- How commercially viable are the different business models being employed at country level?
- What lessons can be learned about best practices in the design and implementation of mobile phone-based nutrition services to ensure a) behaviour change and b) continued private sector engagement in different countries?

Further evaluation questions related to other aims of mNutrition will be addressed in at least one country (either Ghana and/or Tanzania):

- Are mobile phone-based services a cost-effective way to register and identify at-risk populations to target with nutrition support?
- Are mobile phone-based services a cost-effective way for community health workers to improve the quality and timeliness of data surveillance (a core set of nutrition-related indicators)?

The content for the mobile phone-based advisory services will be based on international best practices and widely endorsed protocols (i.e. by the World Health Organization) and evidence-based nutrition-sensitive agricultural practices identified by international experts. Through an iterative multi-stakeholder process, international and country experts will localise and adapt the content to make it relevant to the target audience in the 14 countries. The adapted content and nature of messages is expected to vary across different target audiences within and across countries. The main purpose of assessing the relevance of the content is not to evaluate the overall health and nutrition content but on how this content has been localised and adapted and to what extent the needs of the target groups within their particular context have been met.

In assessing the commercial viability, it is recognised that evaluating the sustainability/long-term financial viability of the mobile phone-based advisory services will be difficult, as mobile network operators may not be willing to provide this potentially commercially sensitive information. Therefore, GSMA will provide support through its access to aggregated confidential financial results of the mobile network operators providing the service. GSMA will provide a financial summary report on the commercial viability of the business models without compromising the commercial sensitivity of the data for the mobile network operators. The evaluator will assess and validate commercial sustainability through an analysis of the aggregated information provided by GSMA and additional qualitative business analysis approaches.

The evaluator has the option of proposing refinements of the existing evaluation questions during the inception phase as part of developing the research protocol. These suggestions will be considered by the steering committee and an independent peer review during the review of the research protocol as part of the inception phase.

Output

The output of this work will be new and robust evidence on the impact, cost effectiveness and commercial viability of mobile phone-based advisory services focusing on nutrition and agriculture delivered by public and private partners, and including the development of robust methodological approaches to impact assessment of phone-based advisory services.

Recipient

The primary recipient of this work will be DFID, with the beneficiaries being GSMA, governments, international agencies, foundations, MNOs and other private companies and civil society involved in policies and programmes in nutrition and agriculture that are aimed at improving nutritional, health and agricultural outcomes. The findings of this impact evaluation are intended as global public goods.

Scope and timeline

The scope of this work is to:

- Develop a research protocol for the external evaluation of mNutrition;
- Design and undertake an external evaluation of mNutrition in two countries: Ghana and Tanzania;
- Contribute to the communication of the learning agenda, evaluation strategy and evaluation results.

The evaluation will be in two of the 14 mNutrition target countries: Ghana and Tanzania. These countries have been selected based on the phased start-up of mNutrition programme activities. The focus and approach in the two respective countries will be different, allowing for a comparison of the effectiveness of approaches applied. In Tanzania, mNutrition will focus on mobile phone technology-based nutrition and health services and on registration and identification of target population. In Ghana, the mobile phone technology will focus on nutrition- and agriculture-sensitive services.

In terms of coverage in number of people being targeted for these services, in total three million people will be reached through mNutrition, including two million for nutrition-sensitive agriculture advisory messages in four Asian and at least two African countries, and about one million beneficiaries for mobile phone-based nutrition services in ten countries in sub-Saharan Africa.

The evaluation contract period will be September 2014 to 31 December 2019. The development of the research protocol must be completed by month 4 for review and approval by DFID. Full details on tasks and deliverables are provided in sections below.

Statement on the design of the mNutrition evaluation

The evaluation design is expected to measure the impact, cost effectiveness and commercial viability of mNutrition, using a mixed-methods evaluation design and drawing on evidence from two case study countries and the M&E system of the programme. Overall, the proposed design should ensure that the evidence from the two case study countries has high internal validity and addresses the priority evidence gaps identified in the Business Case. Being able to judge the generalisability/replicability of lessons learned from the programme is of equal importance and so a credible approach to generalisation and external validity will be an important component of the overall evaluation design. The final evaluation design and methodology to generate robust evidence will be discussed in detail with DFID and GSMA before implementation.

For assessing cost effectiveness, the evaluator will further fine-tune their proposed evaluation approach and outline their expectations in terms of data they will require from implementers. A theory-based evaluation design, using mixed methods for evaluating the impact has been proposed. During the inception phase, the evaluator will put forward a robust evaluation design for the quantitative work, either an experimental or a quasi-experimental method, with a clear outline of the strengths and limitations of the proposed method relative to alternatives. During the inception phase, the evaluator is also expected to identify clearly what will be the implications of the design for implementers in terms of how the overall programme would be designed and implemented and for evidence to be collected in the programme's monitoring

system. The evaluator will also assess the degree to which it is realistic to assess impacts by early 2019 for a programme where implementation started mid-2015 and, if there are challenges, how these would be managed.

The evaluator, in its six-monthly reports, will be required to provide information to feed into the DFID Annual Review and Project Completion Report of mNutrition.

Gender and inclusiveness

The impact evaluation will pay particular attention to gender and other forms of social differentiation and poverty issues. From current experiences, it is clear that access to and use of mobile services is differentiated along a range of factors, including gender, poverty, geographic marginalisation, education and illiteracy levels. Therefore, the impact evaluation will look at and analyse differentiated access to and potential utilisation of mobile phone-based services for improved nutrition and agricultural production. Based on the findings, it will identify opportunities and challenges in having an impact on women in general and more specifically the poor and the marginalised.

Tasks

The evaluator will perform the following tasks:

A. Finalise a coherent and robust evaluation approach and methodology based on their proposal (inception phase)

- Conduct a landscape analysis of existing experiences in mobile phone-based services for nutrition and agriculture based on available publications and grey project documents to identify additional critical lessons and priorities for evidence gathering and programme design and implementation;
- Ensure that gender issues and poverty issues are well integrated into the impact evaluation design;
- Develop robust sampling frameworks, core set of indicators and research protocols that allow the consistent measurement and comparison of impacts across study countries, taking into account differences in business models and programmes as needed;
- Work closely with the mNutrition programme team in GSMA to familiarise them with impact assessment methodology, discuss evaluation approaches, identify and agree on data provided by programme monitoring system and possible modifications to design;
- Identify risks to the evaluation meeting its objectives and how these risks will be effectively managed;
- Review existing evaluation questions and if deemed relevant propose refinement of existing questions and/or add other questions;
- Prepare a research protocol, including an updated workplan, project milestones and budget. The research protocol will be subject to an independent peer review organised by DFID; and
- Develop a communication plan.

B. Implement and analyse evaluations of impact, cost effectiveness and commercial viability in accordance with established best practices

- Based upon the agreed evaluation framework, develop and test appropriate evaluation instruments, which are likely to include data collection forms for households, community health workers, service providers including health and agricultural services, content providers and private sector stakeholders including mobile network operators. Instruments will involve both quantitative and qualitative methods;

- Register studies on appropriate open-access study registries and publish protocols of studies where appropriate;
- Conduct baselines and endlines, qualitative assessments and business model assessments in both of the two impact evaluation countries;
- Conduct and analyse the evaluations and present findings in two well-structured reports addressing the evaluation questions. The reports should follow standard reporting guidelines as defined by, for example, the Equator Network. Primary findings should be clearly presented along with a detailed analysis of the underlying reasons why the desired outcomes were/were not achieved;
- The evaluating organisation or consortium may sub-contract the administration of surveys and data entry, but not the supervision of those tasks, study design or data analysis; and
- The country-specific mixed-methods evaluation reports, cost effectiveness and business models studies and final evaluation report will be subject to an independent peer review organised by DFID.

C. Contribute to the communication of the learning agenda, impact evaluation strategy and evaluation results

- Develop a communication plan outlining the main outputs and key audiences;
- Conduct Lessons Learnt Workshops in each of the two impact evaluation countries and key dissemination events; and
- Assist in communicating the results of the evaluation and contribute to the development and communication of lessons learned about mobile phone-based extension approaches in nutrition and agriculture.

Deliverables

The evaluator will deliver the following outputs:⁴⁰

During the design and study inception phase of maximum four months:

- A publishable landscape analysis report highlighting lessons learned from existing initiatives on mobile phone-based advisory services related to nutrition and agriculture by month 4;
- An updated workplan with project milestones and budget by end of month 1 (possibly adjusted based on the approved research protocol by month 4);
- A communication plan outlining the key outputs, audience and timeline for review and approval by month 4; and
- A full research protocol by month 4 for review and approval. The research protocol should be registered with appropriate open-access study registries;

Interim reports:

- Four biannual progress reports for the external evaluation as a whole, and for each country evaluation, against milestones set out in the workplan;
- Two desk reviews submitted by June 2016;
- Two baseline quantitative reports submitted by April 2017;
- Two baseline qualitative reports submitted by February 2017;
- Two cost-effectiveness reports – one submitted by March 2017;
- Two business model reports – one submitted by March 2017;
- Two mixed-methods baseline reports completed by September 2017;
- Two midline qualitative reports submitted by March 2018.

⁴⁰ The exact timeframe of deliverables will be agreed upon during the design phase as appropriate.

All survey data collected during the evaluation provided in a suitable format to DFID for public release.

At project's end:

- Two endline quantitative reports submitted by June 2019;
- Two endline qualitative reports submitted by August 2019;
- Two cost-effectiveness reports submitted by July 2019;
- Two business model reports submitted by July 2019;
- Two evaluation reports submitted by October 2019;
- At least one article, based on findings from the country evaluation reports, published in a research journal;
- A shared lesson learned paper published and at least one presentation highlighting key lessons for similar initiatives of promoting mobile-based technologies for providing extension services and the promotion of uptake of technologies by December 2019.

Research protocol and all final reports will be independently peer reviewed. This will be organised by DFID. Outputs are expected to be of suitable quality so that a synthesis of findings can be published in a leading peer-reviewed journal.

Coordination and reporting requirements

A mNutrition Advisory Group (AG) will be established for the programme which will a) provide technical oversight and b) maximise the effectiveness of the programme. The AG will meet on a bi-annual basis and will comprise of representatives of DFID, NORAD and GSMA and independent technical experts. The evaluator will be managed by DFID on behalf of the mNutrition AG. The evaluator will work closely with the mNutrition programme team in GSMA and its specific country implementing partners. The evaluator will:

- Ensure coherence and lesson learning across all pilot impact assessments on the key evaluation questions and indicators identified;
- Incorporate a clear code of ethics;
- incorporate plans for open access publications and public access to datasets.

The evaluator will work closely with the mNutrition project management team, in particular in the design of the overall evaluation framework and the evaluation plan for the specific project components and the countries selected for the evaluation. Collaboration and regular communication between evaluator and mNutrition project management team and implementing partners in selected case study countries is crucial, as the evaluation design may have implications for project implementation and vice versa. The mNutrition project management team will lend support in communication as requested by the evaluator or AG. The evaluator will report directly to DFID, who will manage the evaluation on behalf of the mNutrition AG. The main point of contact for technical matters is Louise Horner, Livelihoods Adviser, and the contact for all other project-related issues is Hugh McGhie, Deputy Programme Manager. The mNutrition AG will be the arbiter of any disputes between the evaluation function and the overall programme implementation.

At the end of each six months, the evaluator will submit a brief report outlining key achievements against the agreed deliverables. Pre-agreed funding will then be released provided that deliverables have been achieved.

In addition to the six-monthly reports outlined above, the evaluator will provide information to feed into the DFID Annual Review of mNutrition. The six-monthly reports will be a key source

of information used to undertake the Annual Review and Project Completion Report for the programme. These reviews will be led by the Livelihoods Adviser and Deputy Programme Manager, in consultation with the mNutrition AG. All reviews will be made available publicly in line with HMGT transparency and Accountability Requirements.

Mandatory financial reports include an annual forecast of expenditure (the budget) disaggregated monthly in accordance with DFID's financial year April to March. This should be updated at least every quarter and any significant deviations from the forecast notified to DFID immediately. In addition, the evaluator will be required to provide annual audited statements for the duration of the contract.

Contractual arrangements

The contract starts in September 2014 and will run till end of December 2019 subject to satisfactory performance as determined through DFID's Annual Review process. Progression is subject to the outcome of this review, strong performance and agreement to any revised work plans or budgets (if revisions are deemed appropriate).

A formal break clause in the contract is included at the end of the inception period. Progression to the implementation phase will be dependent on strong performance by the evaluator during the inception period and delivery of all inception outputs, including a revised proposal for the implementation period. Costs for implementation are expected to remain in line with what has been agreed upon for this contract, with costs such as fee rates fixed for contract duration. DFID reserves the right to terminate the contract after the inception phase if it cannot reach agreement on the activities, staffing, budget and timelines for the implementation phase.

DFID reserves the right to scale back or discontinue this assignment at any point (in line with our terms and conditions) if it is not achieving the results anticipated. The evaluator will be remunerated on a milestone payment basis. DFID has agreed an output-based payment plan for this contract, where payment will be explicitly linked to the evaluator's performance and effective delivery of programme outputs as set out in the ToR and approved workplan. The payment plan for the implementation phase will be finalised during the inception period.

Open access

The evaluator will comply with DFID's Enhanced and [Open Access Policy](#). Where appropriate, the costs of complying with our open-access policy should be clearly identified within your commercial proposal.

Branding

The public has an expectation and right to know what is funded with public money. It is expected that all research outputs will acknowledge DFID support in a way that is clear and explicit, and which fully complies with DFID Branding Guidance. This will include ensuring that all publications acknowledge DFID's support. If press releases on work which arises wholly or mainly from the project are planned, this should be in collaboration with DFID's Communications Department.

Duty of care

The evaluator is responsible for the safety and well-being of their personnel (as defined in Section 2 of the contract) and third parties affected by their activities under this contract, including appropriate security arrangements. The evaluator is responsible for the provision of suitable security arrangements for their domestic and business property. DFID will share available information with the evaluator on security status and developments in-country where appropriate.

The evaluator is responsible for ensuring appropriate safety and security briefings for all of their personnel working under this contract and ensuring that their personnel register and receive briefings as outlined above. Travel advice is also available on the Foreign and Commonwealth Office website and the evaluator must ensure they (and their personnel) are up to date with the latest position.

The evaluator has confirmed that:

- The evaluator fully accepts responsibility for security and duty of care;
- The evaluator understands the potential risks and has the knowledge and experience to develop an effective risk plan;
- The evaluator has the capability to manage their duty of care responsibilities throughout the life of the contract.